Electronic Letters on Computer Vision and Image Analysis 21(1):78-100, 2022

A multiple secret image embedding in dynamic ROI keypoints based on hybrid Speeded Up Scale Invariant Robust Features (h-SUSIRF) algorithm

Suganthi Kumar* and Rajkumar Soundrapandiyan *

* *School of Computer Science and Engineering, Vellore Institute Technology, Vellore, India*

Received 20th of July, 2021; accepted 19th of May 2022

Abstract

This paper presents a robust and high-capacity video steganography framework using a hybrid Speeded Up Scale Invariant Robust Features (h-SUSIRF) keypoints detection algorithm. There are two main objectives in this method: (1) determining the dynamic Region of Interest (ROI) keypoints in video scenes and (2) embedding the appropriate secret data into the identified regions. In this work, the h-SUSIRF keypoints detection scheme is proposed to find keypoints within the scenes. These identified keypoints are dilated to form the dynamic ROI keypoints. Finally, the secret images are embedded into the dynamic ROI keypoints locations of the scenes using the substitution method. The performance of the proposed method (PM) is evaluated using standard metrics Structural Similarity Index Measure (SSIM), Capacity (C_P), and Bit Error Rate (BER). The standard of the video is ensured by Video Quality Measure (VQM). To examine the efficacy of the PM some recent steganalysis schemes are applied to calculate the detection ratio and the Receiver Operating Characteristics (ROC) curve is analyzed. From the experimental analysis, it is deduced that the PM surpasses the contemporary methods by achieving significant results in terms of imperceptibility, capacity, robustness with lower computational complexity.

Key Words: Video steganography, imperceptibility, capacity, robustness, computational cost, h-SUSIRF, steganalysis.

1 Introduction

The recent year has been very progressive for digital media communication. This exponential growth of digital communication on public networks may pose several technical threats such as illegal and malicious access and modification of the digital contents. Steganography, the science of secret communication is recommended as a sophisticated solution to prevent these risks. In the art of steganography, the steganographers namely sender and receiver exchange secret digital content (secret message) without being suspected by a third party. So, they devise another digital content as a transmission medium through which the secret message is being shared. This transmission medium is known as the cover medium. Once the secret data is embedded into the cover

Correspondence to: <rajkumarsrajkumar@gmail.com>

Recommended for acceptance by <Dr. Anuj Sharna>
https://doi.org/10.5565/rev/elcvia.1470
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

medium it is known as stego-medium and it is sent via a network for communication [1]. The third-party steganalyst observes the exchange of stego-medium between the steganographers and tries to retrieve the secret data illegally. This action by steganalyst is called steganalysis. The primary motive of steganography is to hide the secret message within the cover medium by concealing the very existence of communication from the steganalyst. Based upon the type of cover medium, steganography is classified as (i) text, (ii) audio, (iii) image, (iv) video, (v) protocol, and (vi) DNA steganography. By widespread usage of video-based applications, video steganography has recently got more emphasis [2]. Moreover, the following attributes of videos qualify them as good cover medium (1) videos can accommodate a large amount of secret data, (2) the visual fluctuation of videos provide an imperceptible environment to conceal secret messages (3) The data loss in videos can be managed. A good video steganography algorithm must be designed in such a way that it should fulfil the fundamental steganography requirements which include imperceptibility, capacity, and robustness [3, 4]. In addition to these, there are some more requirements such as reversibility and computational complexity for a better process.

- **Imperceptibility:** The degree of visual quality should be higher. After hiding the secret data, cover medium and stego_medium should look similar. i.e., invisible distortion of the stego_medium.
- **Capacity:** The cover medium should provide enough space to hide the secret data without visual degradation [5].
- **Robustness:** The embedded secret data should not be removed or modified by a third party. And the secret data should be recovered by the receiver without data attenuation.
- **Reversibility:** The lossless or absolute recovery of secret data at the receiver end [6].
- **Computational complexity:** The cost of time and space taken for the successful execution of the steganography algorithm which needs to be lesser for better performance.

In the realm of video steganography, a good amount of research works has been proposed to improve the above-mentioned features and some related works are discussed in the below section.

1.1 Literature review

In recent days, many videos steganography-based researches have been performed [7]. Despite promising results, operations such as finding visually poor perceiving yet stable frame regions for hiding the secret data, achieving a balance between the fundamental requirements are still challenging processes and are seldom investigated in the majority of the state-of-the-art [8].

Lately, a special class of substitution method referred to as the adaptive video steganography method has drawn more attention wherein the domanial pattern of the cover medium is studied before embedding the secret data in spatial or transform domains [9, 10, 11, 12, 13]. This approach helps in determining the suitable regions of cover medium to embed the secret data. These regions are popularly known as ROI. In the case of determining the ROI, the video steganography system gets benefited from the fact that the Human Visual System (HVS) is poor to perceive variation inside the scenes, dynamic, and moving regions. Based on this fact, Luo et al.[14] targeted the HVS to improve the imperceptibility where most relevant neighbors are selected based on sparse representation for embedding. Ramalingam et al.[15] formulated a video steganography scheme based on the scene change detection process wherein the scene change is detected using the difference between DCT coefficients of two consecutive frames. And the secret data is fused in the DWT domain. But this method is unable to provide enough concealing capacity. Then, the motion estimation process has also been exploited for the embedding process. For instance, Cao et al.[16] utilized motion vectors for embedding to improve undetectability.

Similarly, Yao et al. [17] suggested a motion vector-based video steganography. Wherein, the distortion function is formed to find the embedding impact. Concerning the embedding impact, the secret data is hidden using syndrome-trellis codes. This method performs well in terms of imperceptibility, yet it fails to give adequate capacity and robustness. Song et al. [18] developed a reversible video steganography algorithm. Here, the secret data is embedded into the motion vector using the inner product condition between the motion vector and modulation vector. The main disadvantage of this work is the poor robustness against attacks. Further, the intraframe (I-frame) and inter frames (B and P frames) from motion detection algorithms are also employed for embedding. Considering the intraframe distortion drift Liao et al. [19] presented a video steganography framework using the intraframe embedding method. A lower embedding rate is the main delimitation of this work. Therefore, to enhance capacity rate, inter frames are deployed in Yao et al. [20] proposed a reversible video steganography scheme using motion estimation.

Though, embedding secret data in moving regions is the best approach in terms of imperceptibility there is a chance of data loss while compression or transmission. To address this issue, the benefits of keypoints detection algorithms are incorporated with steganography and watermarking domains. Yan et.al [21] initiated a keypoints-based watermarking scheme for a geospatial vector. Manikandan et al. [22] utilized keypoint detection algorithms such as Speeded Up Robust Feature (SURF) and Scale Invariant Feature Transform (SIFT) and histogram shifting method to generate robust watermarking scheme for copyright purpose of images in 5G network. Mstafa et al. [23] embedded the secret data only in the human face region of the video using Kanade-Lucas-Tomasi (KLT) keypoints feature detection approach. Since the human facial regions in videos are often in motion appending secret data is not causing visually perceivable changes. Similarly, Hashemzadeh [24] came up with a motion clue-based video steganography algorithm where the feature points of the segmented frames are studied and tracked using KLT. The spatial and temporal behaviors of identified feature points are employed in determining the ROI to hide the secret data. Incorporating keypoints detection algorithms with steganography methods immensely improves the robustness by resisting all kinds of signal processing and geometrical attacks. But most of the time this process fails to provide adequate payload capacity. Thus, there is still a lot of advancements required to enhance concealing capacity while maintaining the visual quality. Also, when it comes to finding keypoints in video frames for steganography purposes the computational cost is another important requirement that needs to be addressed.

1.2 Contribution of the proposed work

In this work, the proposed h-SUSIRF keypoints detection algorithm-based video steganography method is designed and implemented to improve the robustness and payload capacity in less time. The novel characteristics of this work are summarized below:

- The proposed h-SUSIRF algorithm complements the features of the keypoint detection algorithms SURF and SIFT by attaining, accurate localization, a greater number of keypoints, and lesser computational complexity.
- The process of morphological dilation of the determined keypoints using the h-SUSIRF algorithm increases the capacity.
- Embedding the secret data in dilated dynamic regions using gradient and momentum greatly improves the robustness against steganalysis as it lies in different scales.

1.3 Real-time application scenarios

The proposed video steganography technique can provide adequate capacity to embed multiple secret images with at most robustness while maintaining the visual quality. This method retrieves the secret data precisely at the receiver end. As the PM undergoes a semi-reversible data hiding process only the unaltered part of cover data be retrieved at the receiver end. Thus, in the application perspective, the PM can be deployed in the industries where the utmost secret communication is demanded without considering the cover data through which the secret data is being shared.

- **Medical applications:**The goal of any medical application is to aid patients/doctors with healthcare prerequisites especially when doctor and patient are away from each other. In this respect, the Electronic Patient Record (EPR) must be exchanged securely. This EPR includes clinical health data such as patients details, medical insurance particulars, doctor prescription, laboratory reports, health monitoring data like blood pressure, sugar level, heartbeat rates, and medical images such as CT, MRI, etc. As the proposed video steganography system provides distinct scenes to embed the secret data it is easier to embed a variety of EPR data. For example, if a video consists of some 3 scenes 3 different EPR data can be embedded in a cover video (e.g., CT image, MRI image, and clinical data).

- **Military:** In the defense and law enforcement sectors top-secret communication is the primary aim. Steganography is widely utilized in this field as it is always receptive to the technology advancement for sophisticated and secure communication. The PM can provide a highly robust steganographic framework that can look after the integrity of the secret data. So, the proposed method can be deployed in the military, law enforcement, and defense scenarios.

- **Geographical applications:** The Geographical Information System (GIS) embodies a vector of geospatial data such as coordinates, longitude, latitude, altitude, orientations, location, view volume, different feature types (raster, points, lines, etc.) for graphical map presentation. It is mandatory to transfer the GIS data securely amongst the authorized users as acquiring GIS is expensive. GIS can be used in a variety of applications such as object tracking, navigation, location sharing, and marketing. Incorporating the PM in GIS-based applications can improve data security.

- **Multimedia applications:** In multimedia, the proposed video steganography can be used in an invisible watermarking process for copyright verification purposes. This technique restrains illegal copying and infringements.

- **Smart IDs:** The PM can also be exploited in transmitting or storing the personal information and photographs of individuals securely.

- **Corporate:** In the world of corporate, professional secret leakage in vulnerable communication is the biggest threat and it could lead to a serious data breach. Therefore, the video steganography method is highly recommended for secure and authentic communication.

- **Intelligence agencies:** Intelligence agencies by definition require top-level secrecy and are always in demand of covert communication. The proposed method can be employed in such agencies for the secured covert communications within and without agency.

2 Proposed Method

In this section, the proposed video steganography algorithm based on the dynamic ROI keypoints is discussed. The proposed schemes for the embedding phase and extraction phase are elaborated in the following subsections.

2.1 Embedding phase

The flow diagram of the embedding phase is given in Figure 1. As exhibited in the figure, firstly, the cover video is converted into a sequence of frames $F_1, F_2, F_3, ...F_n \in F_i; i = 1, 2, 3, ...n$ and the frames are partitioned into a group of frames (GoF) using the abrupt Scene Change Detection (SCD) method. In this, every two consecutive frames(F_i, F_{i+1}) are subjected to the SCD process. First, the frames are transformed to the frequency domain using 2D-DCT coefficients [25]. The difference matrix $D(F_i, F_{i+1})$ is formed between the attained transformed

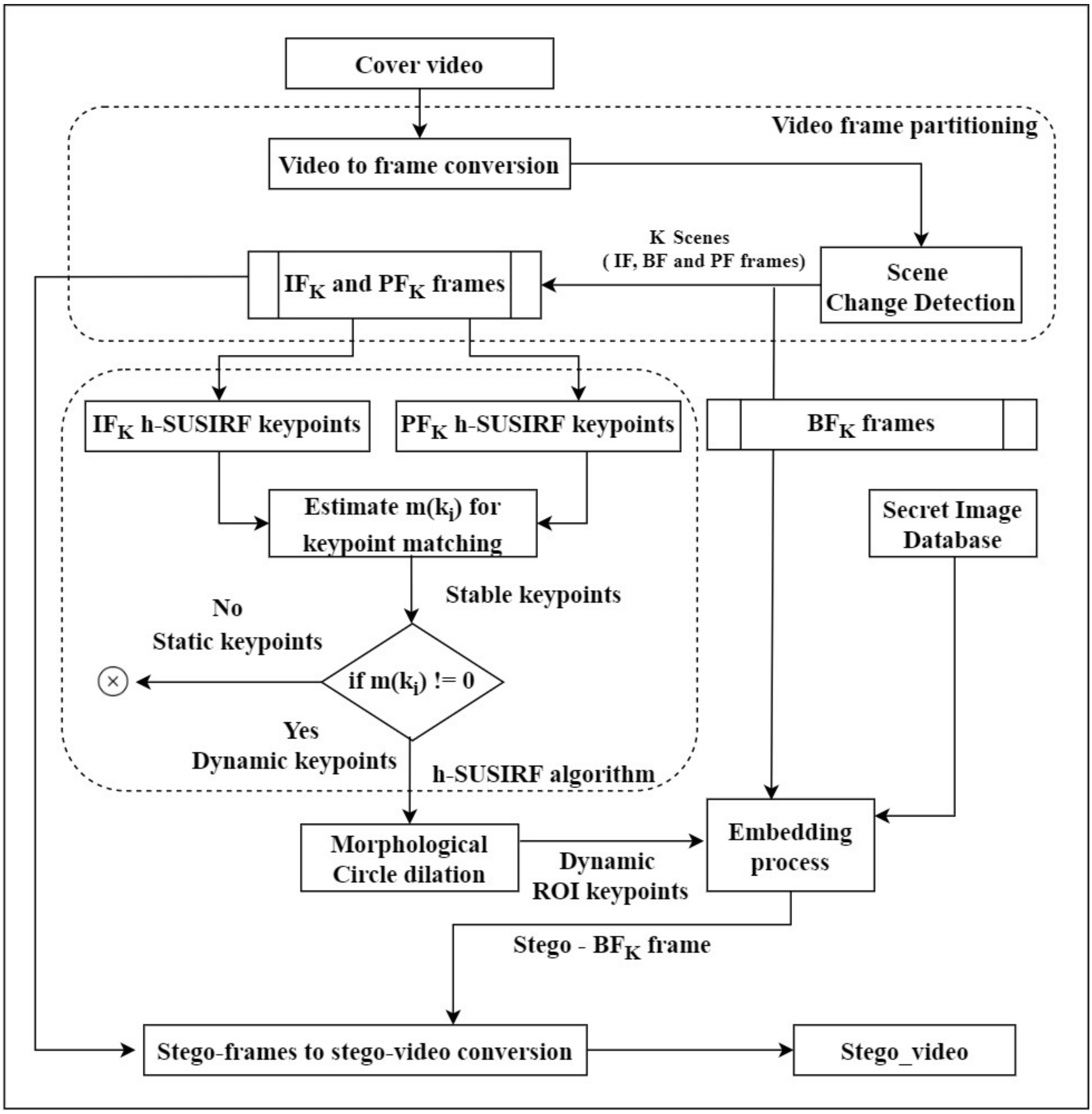


Figure (1) The flow diagram of the embedding phase

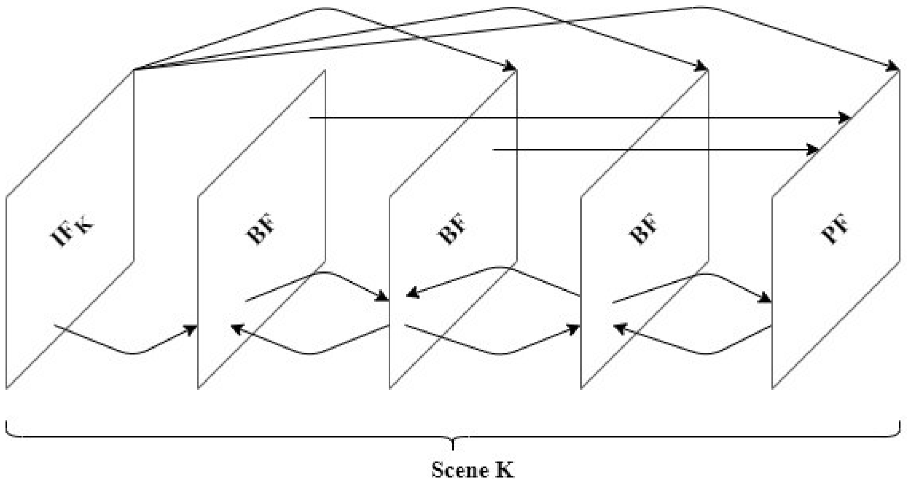


Figure (2) Types of frames and their compatibility with each other

coefficient matrices of the current frame C_{F_i} and its successive frames $C_{F_{i+1}}$ as in Equation 1. The mean square error (MSE_{dct}) rate is estimated using difference matrix as in Equation 2.

$$D(F_i, F_{i+1}) = (C_{F_i}(r,c) - C_{F_{i+1}}(r,c))^2 \tag{1}$$

$$MSE_{dct} = \frac{\sum_R \sum_C D(F_i, F_{i+1})}{R \times C} \tag{2}$$

Here, (r, c) is the row (r) and column (c) details of the pixel and $(R \times C)$ is the size of the frame. Once $MSE_{dct}$is calculated it is being compared with a threshold value (Tdct). The threshold value is set as 0.5 as it segments the scenes accurately [15]. If the MSE_{dct} value is lesser than the threshold value T_{dct}, it is considered that there is no significant difference between the frames. Hence, F_i is added to the same GoF scene (K). If not, it is apparent that there is a major difference between the frames, and the scene change is detected. Thus, the frame is added to the next GoF scene (K=++K). Each GoF scene consists of three types of frames namely IF frames, BF frames, and PF frames. Figure.2 represents the types of frames and their compatibility with each other. Wherein, IF frames denote Intra-Frames which are the initial frames of a new scene. It takes only the backward references. PFs are the Predictive-Frames that are end frames of a scene and takes forward references. Wherein, the intermediatory Bi-predictive-Frames known as BF frames take references from forward as well as backward. Once the GoFs are identified, the embedding phase proceeds with the following steps: (1) dynamic ROI keypoints selection using the proposed hybrid Speeded Up Scale Invariant and Robust Features (h-SUSIRF) (2) the embedding process.

2.1.1 The proposed h-SUSIRF for dynamic ROI keypoints determination

In this phase, the proposed h-SUSIRF is performed only between IF and PF frames to find the dynamic ROI keypoints locations inside every scene. Both the frames undergo the h-SUSIRF process separately. The h-SUSIRF algorithm combines and improvises the processes involved in SIFT and SURF keypoint detection algorithms. SIFT algorithm is highly accurate and most efficient. However, the heavy computation complexity limits the use in real-time applications. SURF is an approximation technique that gets faster and very similarly accurate keypoints compared to SIFT. Yet it finds lesser keypoints than SIFT. Thus, the h-SUSIRF tends to compensate these drawbacks by incorporating the processes of SIFT and SURF keypoint detection algorithms. In the proposed approach, the two detectors, the DDoG approach and fast Hessian matrix-based approach are utilized for keypoint detection. Once these two sources of keypoints are obtained, they are described in terms

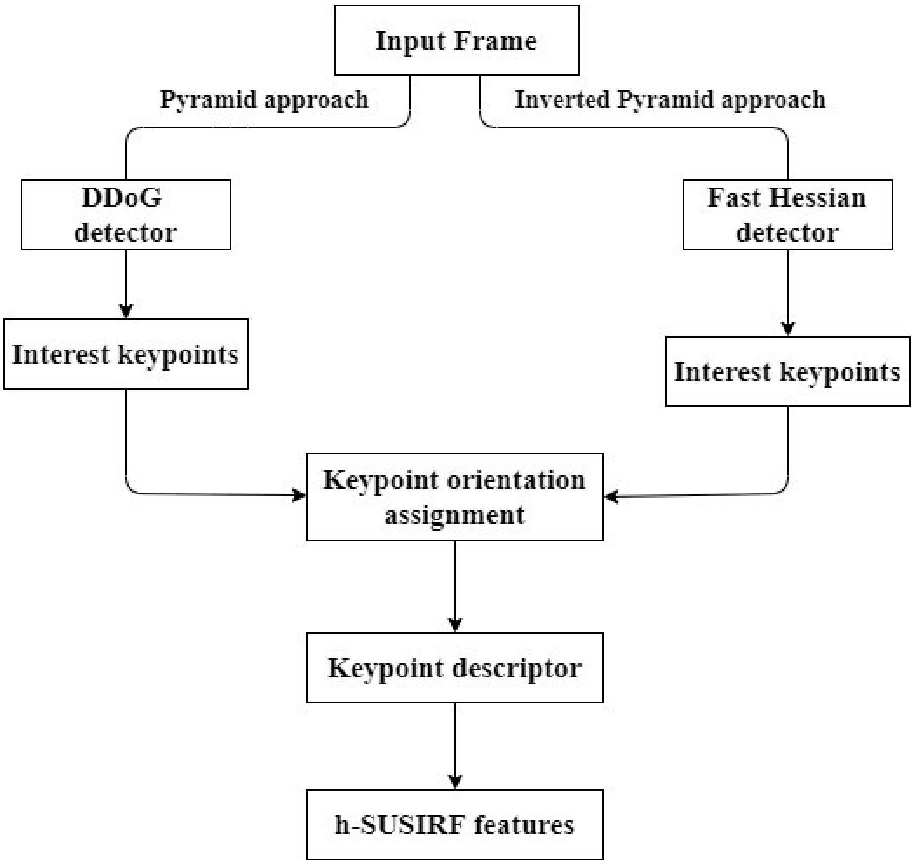


Figure (3) The flow diagram of the h-SUSIRF approach

of SURF features. Figure 3 shows the flow of the proposed h-SUSIRF approach. Initially, the scale space is created by the pyramid approach and the inverted pyramid approach.

In the pyramid approach, first, the 2D frame is converted as a 1D grey-level sequence. Therefore, the scale space is the 1D sequence and it is defined as a function of one variable. The scale-space $L(z, \sigma)$ is attained by convolving the one variable Gaussian function $G(z, \sigma)$ with the 1D frame sequence $IF(z)$ as given in Equation 3.

$$L(z, \sigma) = G(z, \sigma) * IF(z) \tag{3}$$

wherein,z is the 1D frame sequence and σ is the scale rate. And,

$$G(z, \sigma) = \frac{1}{\Pi\sigma^2} e^{\frac{-z^2}{\sigma^2}} \tag{4}$$

The Difference of Gaussian (DoG) is calculated by convolving the 1D frame sequence with the difference of Gaussian scales and it can be computed as follows:

$$D(z, \sigma) = (G(z, \sigma) - G(z, k\sigma)) * IF(z) \tag{5}$$

$$= L(z, \sigma) - L(z, k\sigma)$$

where the factor $'k'$ is constant over all the scales. The second-order difference of DoG (DDoG) is estimated to simplify the construction of the pyramid using the Equation 6.

$$DD(z, \sigma) = D(z, \sigma) - D(z, k\sigma) \tag{6}$$

In the theory of differential and integral calculus, the local extremum point of a function is more or less equal to zero of its first derivative. Thus, Equation 6 is partially differentiated in respect to as given in Equation 7,

$$\frac{\partial D}{\partial \sigma} \approx \frac{D(z,k\sigma)-D(z,\sigma)}{\sigma-k\sigma} = \frac{DD(z,k\sigma)}{\sigma-k\sigma} \tag{7}$$

And thus,

$$DD(z,k\sigma) = D(z,k\sigma)-D(z,\sigma) \approx \frac{\partial D}{\partial \sigma}(\sigma-k\sigma) \tag{8}$$

Here, $k\sigma - \sigma \neq 0$ and so $DD(x,\sigma) = 0$. It shows that the local maximum value obtained by DoG is zero at the first derivative. This step replaces the local extrema and non-suppression region detections from SIFT method by observing the fact that obtaining zero is much easier than finding extreme points. This process considerably reduces the computational complexity. The threshold value is fixed to check the pixels whose absolute value at the second-order derivative corresponds to zero. In this work threshold value is set as 400 [26]. The absolute values from the second-order derivative of each pixel that are lesser than or equal to the threshold value are recorded as extreme points. These extreme points are then given to find their accurate localization using the 2D curve fitting function as follows:

$$DD(Z) = DD + \frac{\partial DD^T}{\partial Z}X + \frac{1}{2}Z^T\frac{\partial^2 DD}{\partial Z}Z \tag{9}$$

where, $X = (z,\sigma)^T$ is the offset from the keypoints. The derivative of the 2D function concerning z and setting it to zero we can get as below,

$$Z^{'} = -\frac{\partial^2 DD^{-1}}{\partial Z^2}\frac{\partial DD}{\partial Z} \tag{10}$$

Substituting 10 into 9,

$$DD(Z^{'}) = DD + \frac{1}{2}\frac{\partial DD^T}{\partial Z}Z^{'} \tag{11}$$

Notably, when $Z^{'}$ is greater than the value 0.5 in z and σ it means that the keypoint is poorly localised and is removed. Also, improve the stability of the keypoints further the contrast of the selecte1d keypoints is also checked. The keypoints with the contrast lesser than $0.04/\tau(\tau - octave)$ are discarded. The remaining keypoints are known as the interest keypoints from the pyramid approach.

Now, the same frame is subjected to the inverted pyramid approach. Initially, the input frame is converted as integral version $I(x,y)$ of the same. This integral frame $I(r,c)$ seemingly reduces the computational complexity. The scale space is approximated by involving various scale levels in the convolution between the integral frame and the second-order derivative of the gaussian matrix from small to large, and hence it is an inverted pyramid approach and it is given in Equation 12.

$$H(r,c,\sigma) = \begin{bmatrix} L(r,c,\sigma)_{xx} & L(r,c,\sigma)_{xy} \\ L(r,c,\sigma)_{yx} & L(r,c,\sigma)_{yy} \end{bmatrix} \tag{12}$$

where, $H(r,c,\sigma)$ is known as Determinant of Hessian (DOH) matrix, $L(r,c,\sigma)_{xx}$,$L(r,c,\sigma)_{xy}$, $L(r,c,\sigma)_{yy}$, $L(r,c,\sigma)_{yx}$ are the convolution of the second order-derivative of Gaussian filter with the integral frame at the location (r,c) in the scale . In the generated scale space, for each interest point corresponding local extremum points are identified using neighbor points interpolation. This interpolation is expanded as Taylor series expansion as in Equation 9. For the local extrema, some non-sensitive points from the scale space are removed and the strong keypoints are detected as interest keypoints.

Now, the interest keypoints from these two approaches are fused and non-maximum suppression is applied to avoid repetition. Now, the refined keypoints are given to the SURF approach at the descriptor level. At this level, both SIFT and SURF approaches perform well. But in the SURF descriptor, the complexity is stripped

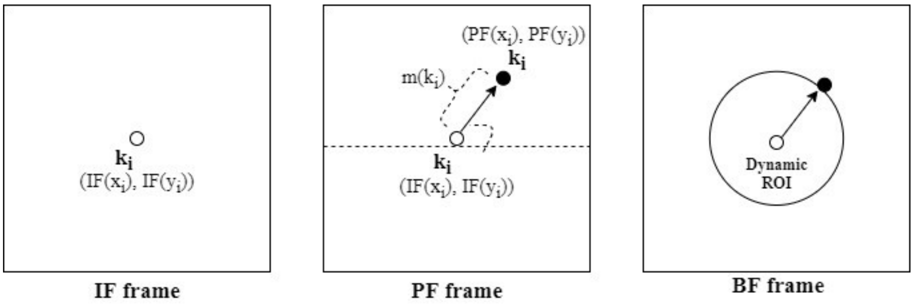
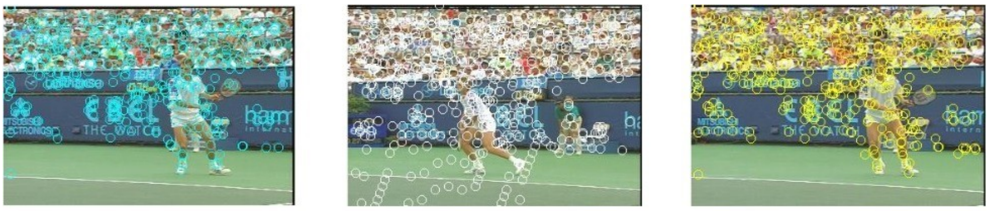


Figure (4) The dynamic ROI keypoints generation



IF_K SUSIRF keypoints PF_K SUSIRF keypoints BF_K Dynamic ROI keypoints

Figure (5) An example of dynamic ROI keypoints generation

down further. This descriptor contains two steps. The first step is the orientation around the interest keypoint circular region is fixed. In the next step, a square region is aligned to select the orientation. This square region is divided into smaller 4×4 subregions. Here, the Haar wavelet responses in the horizontal direction d_r and vertical direction d_c are obtained to define the orientation in the descriptor. Besides, to define the polarity and intensity changes the sum of absolute values of the responses $|d_r|$ and $|d_c|$ are also calculated. So, each 4×4 sub-region contains a four-dimensional descriptor vector $v = [\sum d_r, \sum d_c, \sum |d_r|, \sum |d_c|]$ of length 64. These attained features with descriptors are known as h-SUSIRF features.

Once the h-SUSIRF features of IF and PF frames are determined they are subjected to find their matching rate with each other. Based on the (higher) matching rate the Euclidean distance between the descriptor vectors is estimated. The Euclidean distance is considered as the motion magnitude of keypoints $m(k)$ and it is calculated as given in Equation 13.

$$m(k_i) = \sqrt{(PF(r_i) - IF(r_i))^2 + PF(c_i) - IF(c_i))^2} \tag{13}$$

Here, $(IF(r), IF(c))$ and $(PF(r), PF(c))$ are referred to as the keypoint coordinates of IF frame and PF frame of a scene respectively. Based upon the motion magnitude the dynamic keypoints are separated from the static keypoints $(m(k_i) = 0)$. Once the dynamic keypoints are derived, the motion gradient of these keypoints is determined using the equation below,

$$mg(k_i) = arctan\left(\frac{PF(c_i)^2 - IF(c_i)^2}{PF(r_i)^2 - IF(r_i)^2}\right) \tag{14}$$

Then, since the points around each dynamic keypoints are also having a higher probability of motion magnitude, the neighboring pixels are also taken as dynamic ROI. Thus, the idea of morphological dilation around the dynamic keypoints is utilized to form dynamic ROI keypoints. In the proposed method, a circle dilation is performed with a radius equal to the motion magnitude of the corresponding keypoint. The attained dynamic ROI keypoints locations are being subjected to the embedding process. Figure 4 is the pictorial representation of the dynamic ROI keypoints generation between the IF and PF frame. Figure 5 is an example of the h-SUSIRF keypoints generation between the IF and PF frame of scene K.

Algorithm 1: Embedding phase algorithm

```
Input: Cover video (CV)
Output: Stego-video (SV)

    Step 1. Scene Change Detection
            Find the scenes (GoF) consist of IF, BF, and PF frames using the MSE rate of DCT
            coefficient of every consecutive frame.

    For K=1: number of scenes

      Step 2. DDoG detector ( $\forall IF_K$ and $PF_K$ frames in scene K)
              Convert the frames into the 1D image and convolve with the Gaussian filter in different
              scales to extract the interest keypoints using zero detection at the first derivation.

      Step 3. The inverted pyramid approach ( IFK and PFK frames in scene K)
              Convert the frames into the integral image and convolve with different sized Gaussian
              filters on the same scale to extract the interest keypoints using the fast Hessian matrix
              method.

      Step 4. Non maximum suppression
              Integrate the derived keypoints from Step 2 and Step 3 and remove the repeated keypoints.

      Step 5. Keypoints descriptor
              Lay the quadratic grid with 4  4 square sub-regions over the integrated keypoints and
              find 64 attributes from sample wavelet responses.

      Step 6. The SUSIRF keypoints extraction
              Compare the keypoints descriptors of IFK and PFK frames and find the matching
              keypoints as stable keypoints using Euclidean distance $(m(k_i))$.

      Step 7. Dynamic keypoints selection
              Obtain the dynamic keypoint by abandoning the static ones i.e., $m(k_i) = 0$.

      Step 8. Dynamic ROI keypoints generation
              Form a morphological circle dilation around the dynamic keypoints using motion
              magnitude $m(k_i)$ and motion gradient $mg(k_i)$ .

      Step 9. Embedding process ( $\forall BF_K$ frames in a scene K)
              For i = K
              Secret Image_Database (i) = sImg;
              Convert sImg into binary secret data vector [b];
              Generate the Stego-BFK frame by embedding the [b] into 4LSBs of dynamic ROI keypoints
              intensity values;
              End for

    End for

    Step 10. Stego-video generation
            Fuse all the scenes that consist of IF, stego-BF frames, and PF frame and make stego-video.
```

2.1.2 The embedding process

In this section, the embedding process of secret images into cover frames is explained. In the proposed work each GoF scene consists of a secret image. For example, if a cover video is found with 3 scenes, then there will be 3 different secret images embedded. The secret images are converted as binary vectors before embedding starts. In this phase, the dynamic ROI keypoints locations, and secret binary vectors are the inputs and secret image embedded stego-frames are the outputs. Wherein, the embedding process is being performed only on BF frames and IF, the PF frames remain untouched. This process eases the extraction phase to identify the keypoint locations where the secret data is concealed. Thus, each BF frame in a scene undergoes the embedding process. The R-G-0 substitution method is used in the embedding process. This method is evolved based on the fact that the distortion in the blue component is more perceivable by HVS than the red and the green components [27]. Therefore, only in red and green components of BF frames, the embedding process is performed. Here, R-G-0 indicates the Red-Green-Blue components of a frame. Thus, the secret bits are embedded only in the red and green components, but no secret bit is concealed in the blue component to maintain the visual quality. While embedding the intensity pixel value of the desired keypoint is converted as 8-bit binary data. Up to 4 LSBs are replaced with the secret bits (4-4-0). The secret data hidden cover BF-frames are known as stego-BF frames. These stego-BF frames and non-stego-IF and PF frames are combined and coveted as stego-video for secret communication.

2.2 Extraction phase

The process of retrieving secret data from the stego-video file is carried out on the receiver side. This phase performs the reverse of the data hiding phase. Here, firstly, the stego-video file is sequenced into stego-frames. Next, the scene change detection method is applied to find the scenes. It is noticed that the threshold value $T_{dct} = 0.5$ works well with stego-frames also as it is able to find exactly similar scenes as in the embedding phase. Once the scenes are detected, the dynamic ROI keypoints are determined between IF and PF frames using the proposed h-SUSIRF algorithm. Since the IF and PF frames are not containing any secret bits, it accurately finds the dynamic ROI keypoints in which the secret data is embedded. Then, the extraction process is performed upon BF frames. The secret data is retrieved from 4 LSBs of the dynamic ROI keypoints and stored in secret data vectors. These secret data vectors are then converted as the original secret images. Here, using IF, BF, and PF frames a partially reversed cover video is generated.

3 Results and discussions

This section discusses the dataset description and experimental setup, performance analysis, and comparison with some contemporary methods of the PM, video quality, robustness against steganalysis.

3.1 Dataset and experimental setup

The proposed video steganography algorithm is evaluated using the Xiph.org video database [28] This database provides a sequence of uncompressed YUV4MPEG format videos for research purposes. These videos contain the frames of complex textures such as natural sceneries, human/animal body, objects, clothes; also, highly dynamic things like a cascade waterfall, crowd, and herds and even static elements as the background such as the wall, immobile objects, mountains, floor, non-flipping flower, grass. This database contains around 118 videos with a frame count ranges from 300 to 1200. The proposed method is executed on all the videos from the database. In this paper, visually distinct 6 color cover videos are considered for representation purposes. A detailed description of these considered videos is given in Table 1. The secret images are gathered from the image database Uncompressed Colour Image Database (UCID). This database has around 1338 images that are used for embedding purposes. Figure 6 represents some secret images from the database. In this, after the

Figure (6) Secret Images from UCID database

embedding process, the stego_video is converted into .avi format to assure the integrity of secret data against compression. Further, the PM is executed on an Intel Core i5 processor with a 4 GB RAM configured system using MATLAB 2018a.

Table (1) The descriptions of cover videos

Video(.avi)	Number of Frames ‘n’	Size of the frames ‘$R \times C$’	Number of pixels ‘$3 \times n \times R \times C$’	Camera Type
Playing	400	352 × 288	3(40550400)	Moving camera
Foreman	450	352 × 288	3(45619200)	Moving camera
Holiday	300	352 × 288	3(30412800)	Moving camera
Television	200	352 × 288	3(20275200)	Still camera
Nature	550	352 × 288	3(25344000)	Moving camera
Crew	250	352 × 288	3(25344000)	Still camera

3.2 Performance analysis

The performance of the PM is analyzed and compared with some existing methods in terms of imperceptibility, capacity, robustness against attacks, and computational cost. The metrics used and the experimental results are discussed in the below sections.

3.2.1 Imperceptibility analysis

The perfect steganography system is always expected to conceal a huge amount of secret data with insignificant perceivable differences in the stego-video. The imperceptibility of the PM is evaluated using Structural Similarity Measure Index (SSIM) metric. Based on the features like luminance, contrast, and structure, the SSIM is estimated between the original video and stego-video. To calculate SSIM, each frame of the original video and the corresponding stego-frame of the stego-video go through SSIM calculation as it is defined in the Equation 15

$$SSIM(F_o, F_s) = avg\left(\sum_f \frac{(2\mu_{F_o}\mu_{F_s} + C1)(2\mu_{F_o}\mu_{F_s} + C2)}{(\mu_{F_o}^2 + \mu_{F_s}^2 + C1)(\mu_{F_o}^2 + \mu_{F_s}^2 + C2)}\right) \tag{15}$$

where μ_{F_o}, μ_{F_s} are the mean values of the original frame F_o and the stego-frame F_s, $\mu_{F_o}\mu_{F_s}$ are the standard deviations of the original frame F_o and the stego-frame F_s , $\mu_{F_o}\mu_{F_s}$ is the cross-covariance between the original frame F_o and the stego-frame F_s. The variables $C1 = (K_1G)^2$ and, $C2 = (K_2G)^2$ wherein G is the dynamic range of the pixel values in the frames F_o and F_s , by default K_1 and K_2 are considered as 0.01 and 0.03, respectively. The SSIM is calculated for all frames and the average is taken as the SSIM value for the video. The acceptable value of SSIM varies between 0 (0% match) to 1 (100% Match). The videos such as Playing, Foreman, Holiday and Nature can produce their respective stego-videos with 99% matching similarity. The videos Crew and Nature score 98% and 97% similarity with their respective stego-videos. Because the formerly mentioned videos have a lot of dynamic regions whereas the latter mentioned videos have more static or not so dynamic regions. Thus, the similarity is higher in those videos comparing to the videos Television, Nature. The comparison of SSIM values between the PM and some contemporary works [[23], [24], [29]] is represented in Table 2. From the table, it can be noticed that the SSIM values attained by the method [23] are competing with the PM in the videos Playing and crew. However, the rate of dissimilarity of these videos for the PM is determined as 0.42% and 1.71% which lie under the acceptable range.

Table (2) Comparative analysis of SSIM and C_P values

Videos	[PM]		[23]		[24]		[29]	
	SSIM	C_P	SSIM	C_P	SSIM	C_P	SSIM	C_P
Playing	0.9958	**0.7861**	**0.9971**	0.5481	0.9648	0.6324	0.9651	0.4776
Foreman	**0.9967**	0.7573	0.9921	**0.7996**	0.9864	0.7044	0.9611	0.4293
Holiday	**0.9914**	0.7505	0.9794	0.4646	0.9761	0.7974	0.9595	**0.9193**
Television	**0.9876**	0.8084	0.9817	0.4457	0.9568	0.8657	0.9664	**0.9081**
Nature	**0.9934**	**0.8859**	0.9728	0.7351	0.9816	0.5112	0.956	0.6191
Crew	0.9829	**0.9316**	**0.9882**	0.6705	0.9748	0.5538	0.9722	0.7953

3.2.2 Capacity analysis

The capacity C_P is the rate of secret data that is embedded within the available cover data [30] and is determined as in Equation 16. The capacity is measured in bits-per-pixel (bpp).

$$C_P = \frac{SB_s}{AB_c} \tag{16}$$

where,SB_s denotes the number of Secret Bits in the secret image, AB_c is the Available Bits in the cover medium. A minimum of 0.7bpp capacity is required for a better concealing process. Higher indicates higher

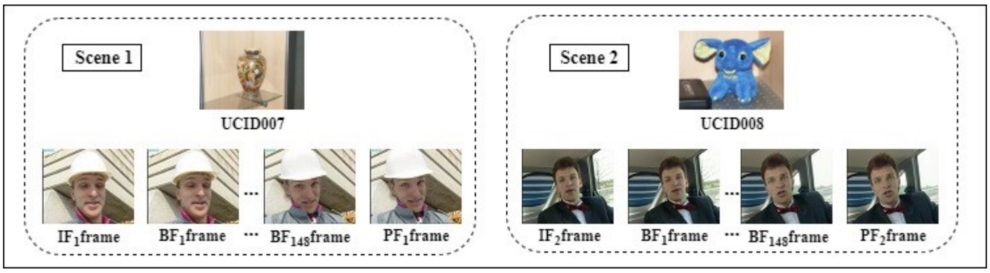


Figure (7) Foreman video scene segments and correspondingly embedded secret images

payload capacity. The videos Playing, Foreman, Holiday provide 70 bpp of the cover region for embedding the secret data. Similarly, Television and Nature give 80 and the video Crew allocate 90 of keypoints. It is observed that the video Crew is found with more scenes comparing to other videos and hence more secret bits are embedded besides its stable backgrounds. Figure 7 shows how the number of scenes affects the payload capacity. Table 2 shows the comparative analysis of values where the PM is competing with the methods [23] and [24] for the videos Foreman and Holiday, Television respectively. This is because, though these three videos are found with more h-SUSIRF keypoints every 3:10 of the identified keypoints get rejected due to their static character. Thus, lesser dynamic ROI keypoints are selected for payload capacity. Yet, the capacity C_P rates are beyond the acceptable range by providing enough space for all the secret images.

3.2.3 Robustness analysis

The proposed method is tested for robustness against attacks such as (1) Null & void, (2) Rotation (22.5), (3) Scaling, (4) Translation (Horizontal: 12, Vertical: 25), (5) Gaussian noise (Variance: 0.04; Mean: 0), (6) Salt & pepper noise (Noise density: 0.07; affects 7 of pixels), (7) Speckle noise (Variance: 0.06; Mean: 0), (8) Median Filter (Filter size: 44), (9) Histogram Equalization, and (10) JPEG compression (QP=80). Thus, each stego-video is being put through these attacks and the secret images are extracted from these attacked videos. The Bit Error Rate (BER) is calculated between the original secret image (SI) and the extracted secret image (eSI) from the stego-video as given in Equation 17.

$$BER = \frac{\sum_{x=1}^{X}\sum_{y=1}^{Y}(SI(x,y) \oplus eSI^{'}(x,y))}{X \times Y} \tag{17}$$

$$SI(x,y) \oplus eSI^{'}(x,y) = 1, \quad if \quad SI(x,y) \neq eSI^{'}(x,y) \tag{18}$$

$$SI(x,y) \oplus eSI^{'}(x,y) = 0, \quad if \quad SI(x,y) = eSI^{'}(x,y) \tag{19}$$

where, $SI(x,y)$ is the original intensity value at the location (x,y) and $eSI^{'}(x,y)$ is the extracted intensity value at the location (x,y) . And (X,Y) is the size of the secret image. The accepted range of BER is 0 to 1. If there is no error in the extracted secret image, the BER value is 0, otherwise, the value is 1 [30]. Lesser the value better the quality. Table 3 represents the BER values comparative analysis between the PM and the methods [23], [24], [29]. These results are generated for the secret images UCID0001 and UCID0002 extracted from the attacked stego-video Playing. From the table, it is observed that the PM has higher resistance as most of its BER values against the considered geometrical and signal processing attacks are lesser (Table 3, highlighted in bold) compared to other methods. Figure 8 depicts the extracted secret images UCID0007 and UCID0008 from the attacked stego-video Foreman and their respective BER values.

Table (3) BER values comparative analysis for the extracted secret images from an attacked video (refer 4.2.3 for attack specifications)

Secret Image	S.No.	Attacks	[PM]	[23]	[24]	[29]
UCID001	1	Null & void	**0.0998**	0.1558	0.1137	0.1968
	2	Rotation (22.5)	**0.2137**	0.3207	0.3246	0.3883
	3	Scaling	**0.2516**	0.3507	0.3287	0.3694
	4	Translation	**0.2872**	0.3208	0.3646	0.3381
	5	Gaussian noise	**0.2501**	0.2986	0.3855	0.4417
	6	Salt& pepper noise	**0.2406**	0.3200	0.3617	0.3701
	7	Speckle noise	**0.2483**	0.3114	0.4127	0.4038
	8	Median Filter	**0.2587**	0.3541	0.3715	0.3978
	9	Histogram Equalization	**0.2124**	0.2781	0.3554	0.3875
	10	JPEG compression	0.2949	**0.2945**	0.3418	0.3595
UCID002	1	Null & void	**0.0069**	0.1245	0.1257	0.1648
	2	Rotation (22.5)	**0.2973**	0.3250	0.3570	0.4500
	3	Scaling	**0.2904**	0.3090	0.3968	0.4314
	4	Translation	**0.2927**	0.3407	0.3868	0.3770
	5	Gaussian noise	0.2829	**0.2696**	0.4085	0.4087
	6	Salt & pepper noise	0.3172	**0.2828**	0.4239	0.4021
	7	Speckle noise	**0.3093**	0.3787	0.4026	0.4271
	8	Median Filter	**0.2886**	0.3211	0.4322	0.3768
	9	Histogram Equalization	**0.2705**	0.3512	0.4203	0.4132
	10	JPEG compression	**0.2954**	0.3164	0.3838	0.4045

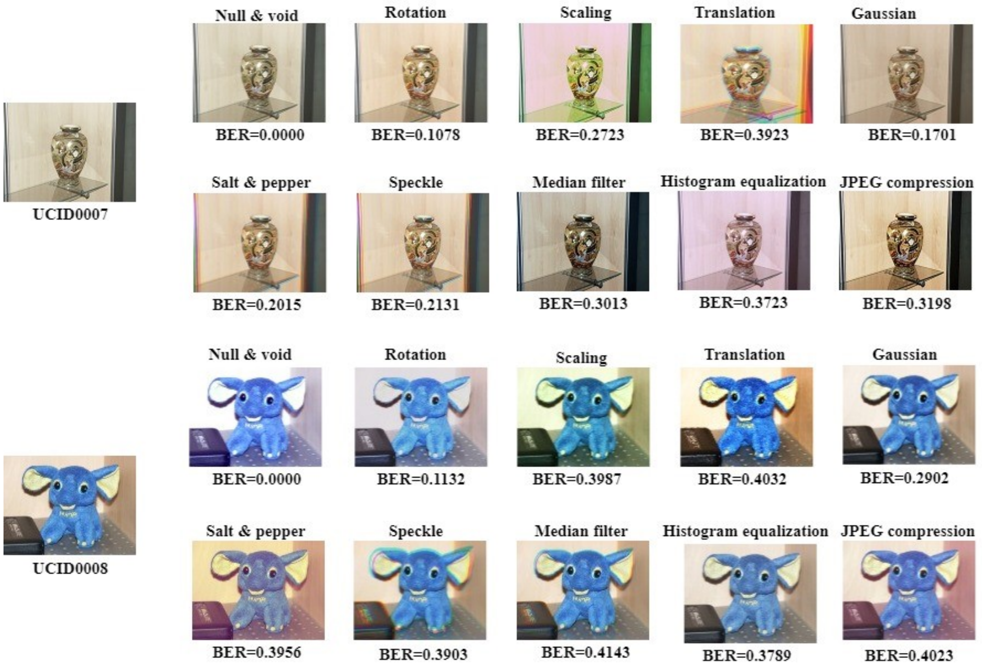


Figure (8) Extracted secret images from the attacked stego-video Foreman and their BER values

3.2.4 Computational cost analysis

The computational cost is another important factor that can impact the performance of video steganography. The lower computation cost is widely desired. The computational cost of the proposed scheme is measured using the execution time which is the accumulation of video coding time, embedding/extraction cost calculation time, and embedding/extraction process time. Video coding time depends upon the type of video encoder. Thus, we analyze only the embedding/extraction cost calculation time and embedding/extraction process time. The embedding cost calculation time of the PM includes the time taken by video frame partitioning and dynamic ROI keypoints selection. Here, the time complexity is measured at the worst-case level big $O(n)$. Wherein $'n'$ is the number of cover frames. The computational complexity for SCD is the linear time big $O(n-1)$. For the dynamic ROI keypoints selection, the time complexity is constant big $O(1)$ because irrespective of $'n'$ only two frames are involved in this process. Then, in the embedding process, the secret data is distributed to the cover keypoints data until it is fully accommodated by the keypoints. Thence, this process is performed based on complexity $O(log(n))$. Now, the overall computational complexity for the PM is linear big $O(log(n))$. i.e.,$O(((n-1)+1)+logn)$. Hence, based on the number of frames the computational complexity of the PM changes linearly. It is assumed that both embedding and extraction modules function in similar complexity pace as they almost follow the same processes. Wherein, the challenging part in both embedding and extraction module is the dynamic ROI keypoint selection which executes in constant time. This process is carried out by the proposed h-SUSIRF algorithm which considerably reduces the time cost. How the h-SUSIRF impacts the embedding rate and time requirement is analyzed as follows,

Impact of h-SUSIRF keypoint detection method: Table 4 explains the effect of the proposed h-SUSIRF algorithm in terms of average capacity rate and time taken for execution. Table 4 represents the average capacity rate, embedding time, and extraction time when the keypoints detection methods SIFT, SURF, FAST, BRISK, and the proposed h-SUSIRF are separately applied in the dynamic ROI keypoints selection module of PM. Here, the capacity rate is taken as the representation of generated keypoints as both are directly proportional to each other. From the table, it can be noticed that though the embedding and extraction time of FAST and BRISK methods are relatively low their capacity rate is also lesser which may lead to poor performance. Wherein, the proposed h-SUSIRF gets more capacity rate (refer to Table 4, highlighted in bold) due to the detection of more keypoints. Figure 9 shows the outcome of SURF and SIFT keypoints features extraction and the proposed h-SUSIRF model. Here, each model detects the scale, orientation, location, metric, and count of the keypoints. From the figure, it can be seen that the count of keypoints features detected by the proposed h-SUSIRF algorithm is higher than SIFT and SURF methods. Also, Table 4 exhibits the time taken for the embedding and extraction process is better than its former methods SIFT and SURF. Hence, it is inferred that the proposed h-SUSIRF method produces a higher capacity rate with minimum computational cost. Therefore, the entire PM can be solved in linear time. In this proposed scheme, the computational time ranges from 2s 10s (in seconds) for the dataset considered. This lower computational cost paves the way for a fast video steganography process.

Table (4) The comparison of average capacity rate (bpp), embedding time, and extraction time of different keypoints detection methods using QP 30 on the proposed method

Keypoint detection method	Average capacity rate (bpp)	Embedding time (in seconds)	Extraction time (in seconds)
SIFT	0.9323	12.9876	12.0019
SURF	0.9096	8.9123	7.9987
FAST	0.6985	6.8871	6.0128
BRISK	0.4567	5.9012	6.0123
h-SUSIRF	**0.9598**	**8.1126**	**7.9832**

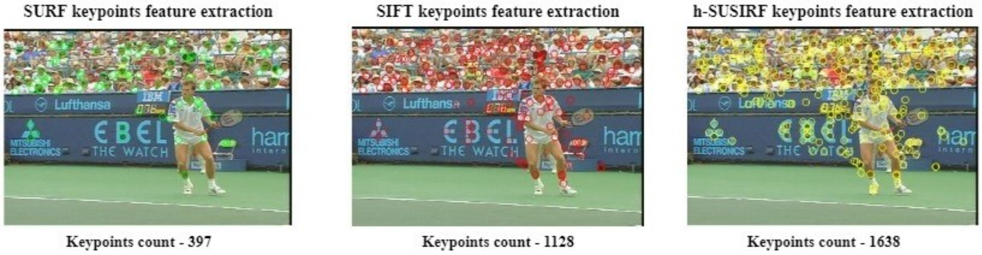

SURF keypoints feature extraction
SIFT keypoints feature extraction
h-SUSIRF keypoints feature extraction
Keypoints count - 397
Keypoints count - 1128
Keypoints count - 1638

Figure (9) The outcome of keypoints features extraction comparison

3.3 Video Quality Measurement (VQM)

The quality of the stego-video or in other words the embedding impact on the cover video is evaluated using the VQM metric. It compares features such as spatial-temporal alignment, valid region estimation, and gain-level offset of the cover video with the stego-video. The value of VQM indicates the fixation behavior of the human eye while viewing the stego-video [31]. It is computed using a linear combination of full reference (FR) quality parameters. The quality parameters include Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). The MSE is the regression cost function that calculates the squared error between the frame of the original video and the corresponding frame of the stego-video. This cost function is expressed as,

$$MSE_f = \frac{1}{R \times C} \sum_{r=1}^{R} \sum_{c=1}^{C} (F_o(r,c) - F_s(r,c))^2 \tag{20}$$

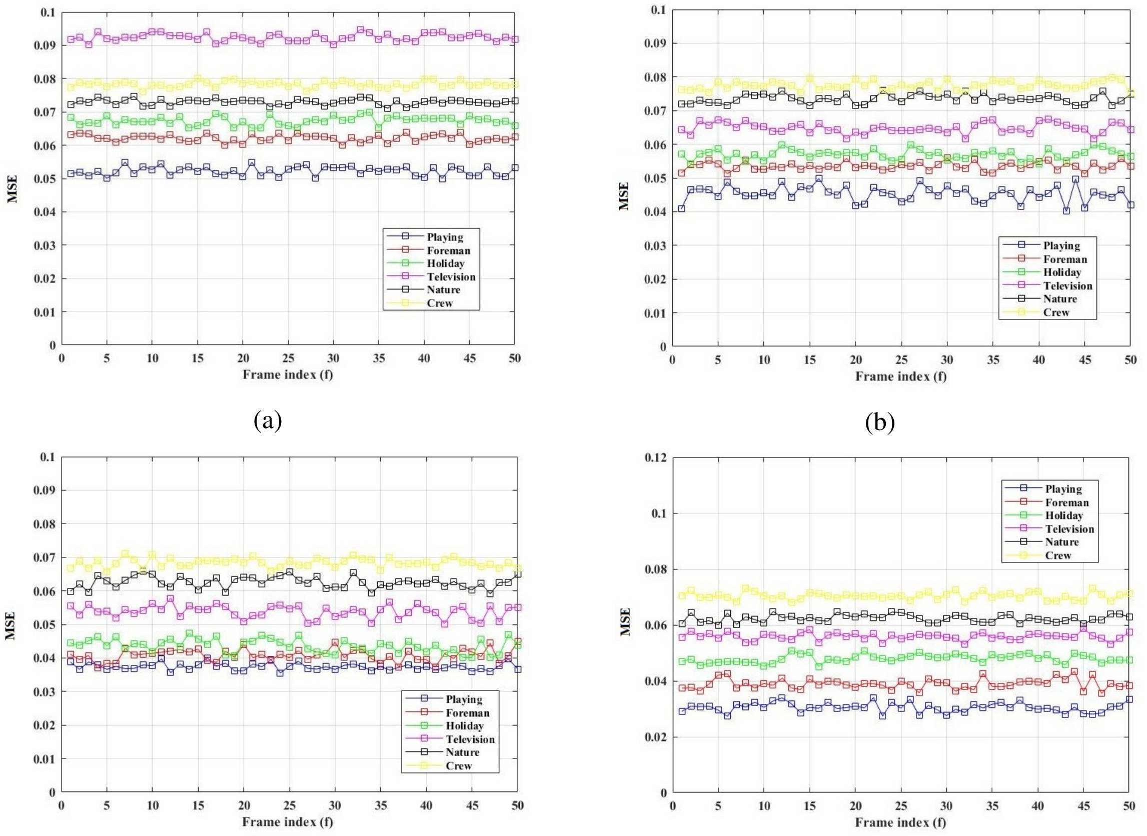
Here, $R \times C$ is the size of the video frame and $F_o(r,c)$ is the intensity pixel value located at the coordinate (r,c) in the original frame, and $F_s(r,c)$ is the intensity pixel value located at the coordinate (r,c) in the stego-frame. MSE is calculated for all the frames in the video $(\forall f \in F)$. The lesser value of MSE_f implies better quality. Figure 10 compares the MSE values of the video frames for the methods Mstafa et al. [23], Hashemzadeh et al. [24], Yao et al. [29], and the PM. In this graphical portrait, the first 50 frames of each video are considered for representation purposes. And the same amount of payload is given for each video. From the graph, it is observed that the MSE of frames exposed to the algorithm by Mstafa et al. [23] ranges from 0.051 to 0.094. Similarly, for Hashemzadeh et al. [24] and Yao et al. [29] the MSE of frames varies from 0.042 to 0.079 and 0.035 to 0.072 respectively. The PM gets the MSE from 0. 025 to 0.066. From this analysis, it can be inferred that the embedding impact on frames caused by the PM is minuscule. Next, the PSNR computes the statistical difference between the cover video and the stego-video in decibels (dB). Hence, each MSE value of the video is given to the next level cost function PSNR and the average is taken as given in Equation 21.

$$PSNR = avg\left(\sum_{f} 10log_{10}\frac{i^2}{MSE_f}\right) \tag{21}$$

where i is the maximum intensity range value (i.e.,255) of the frame. The ideal value of PSNR is greater than 30dB. Using PSNR, the cost function of VQM is expressed as,

$$VQM = \frac{1}{1 + e^{0.17*(PSNR - 25.66)}} \tag{22}$$

The VQM requires a lower value (Average 0) to indicate both are equal frames and so better video visual quality [32]. Table 5 represents the PSNR and VQM values of the PM and the considered existing methods. Here, the PM attains the average of PSNR, and VQM values as 49.3223, and 0.0225 respectively. It shows that the PM produces better quality stego-videos. Though the existing methods generate good quality videos the PM gives better quality videos comparatively (Table 5, in bold font).



(a)

(b)

(c)

(d)

Figure (10) The comparative analysis of the MSE values for the methods (a). Mstafa et al. [23], (b). Hashemzadeh et al. [24], (c). Yao et al.[29], and the [PM].

Table (5) The comparative analysis of PSNR and VQM metrics

Methods	Metrics	Playing	Foreman	Holiday	Television	Nature	Crew
[23]	PSNR	38.4099	41.4673	42.6739	38.8822	34.3084	32.674
	VQM(objective)	0.1027	0.0637	0.0525	0.0955	0.1869	0.2328
[24]	PSNR	35.2264	48.499	36.7624	34.0874	42.9255	36.6009
	VQM (Objective)	0.1643	0.0202	0.1315	0.1927	0.0504	0.1347
[29]	PSNR	42.2045	38.3383	33.7078	31.6786	35.4043	41.7991
	VQM (Objective)	0.0566	0.1038	0.2029	0.2644	0.1602	0.0604
[PM]	PSNR	53.0912	55.1231	42.9876	47.4431	51.9147	45.3763
	VQM (Objective)	**0.0093**	**0.0066**	**0.0499**	**0.0241**	**0.0114**	**0.0338**

3.4 Steganalysis

Steganalysis, on the contrary to steganography, is the process of detecting embedded secret information using steganalysis tools or techniques in media files. To analyze the efficiency of the PM against steganalysis, some current steganalysis techniques such as Fan et al. [33], Tasdemir et al. [34], Su et al.[35], and Wang et al.[36] are deployed. The steganalysis scheme proposed by Fan et al.[33] uses cross-correlation features of video frames and a hash function for steganalysis. Tasdemir et al. [34] employ high-pass filters for steganalysis where temporal dependency with spatial dependency is utilized to improve the accuracy. Again, Su et al. [35] proposed a video steganalysis technique based on the spatial-temporal detector (ST_D) using histogram distribution for local motion intensity and texture complexity. Wang et al. [36] proposed the MV-based steganalysis method by reducing the effect of statistical characteristics originated from videos. These steganalysis techniques are applied on the PM and the other contemporary steganography methods considered and the probability of detection $'P(\mathit{detect})'$ is evaluated as follows,

$$P(\mathit{detect}) = 1 - P(\mathit{error}) \tag{23}$$

$$P(\mathit{error}) = P(x_0)P(y_1|x_0) + P(x_1)P(y_0|x_1) \tag{24}$$

$$= \frac{1}{2}P(y_1|x_0) + \frac{1}{2}P(y_0|x_1); \forall P(x_0) = P(x_1) = \frac{1}{2} \tag{25}$$

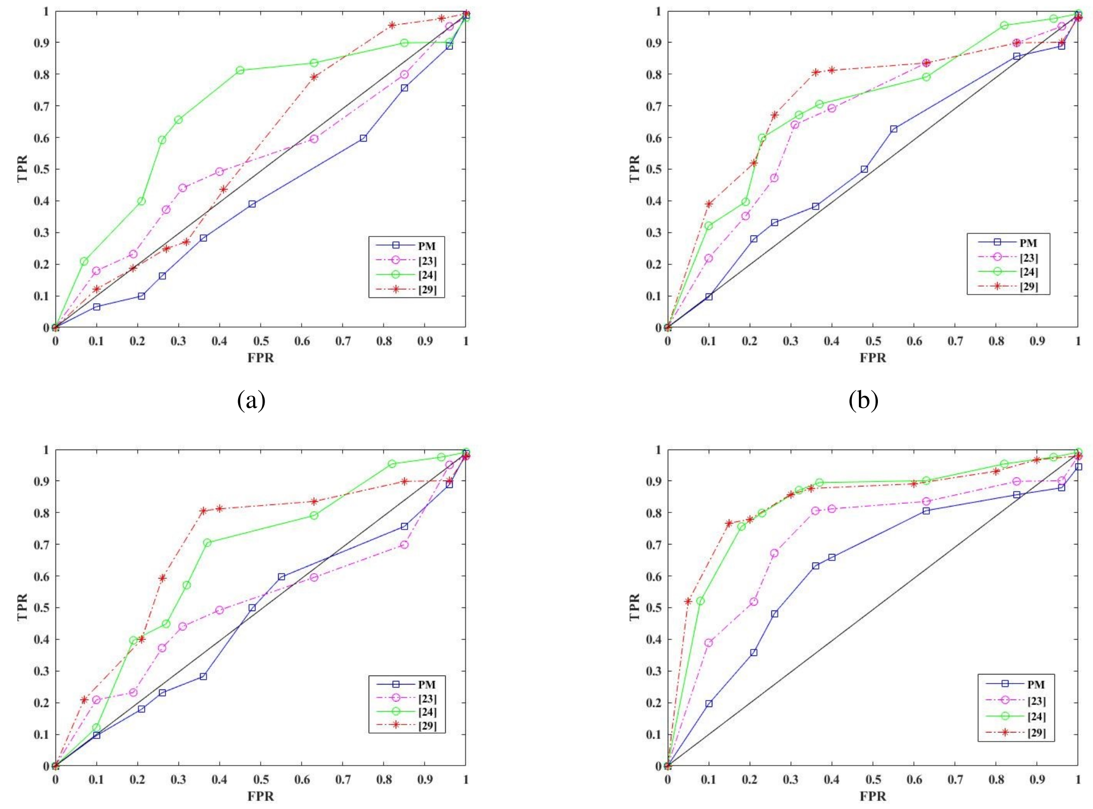
$$= \frac{P(y_1|x_0) + P(y_0|x_1)}{2} \tag{26}$$

$$= \frac{P_{FP} + P_{FN}}{2} \tag{27}$$

where, x_0 and x_1 are the actual frames belong to class 0 and 1 respectively; y_0 and y_1 are the detected frames belong to class 0 and 1 respectively. And P_{FP} and P_{FN} are the false positive and false negative as given in [37]. According to steganalysis $P(\mathit{detect}) \geq 0.5$ is considered as good guessing whereas $P(\mathit{detect}) = 1$ shows 100 accuracy. Based on this probability the Receiver Operating Characteristics (ROC) is interpreted using True Positive Rate (TPR) and the False Positive Rate (FPR). Wherein, the Area Under Curve (AUC) represents the average steganalysis accuracy rate. Figure 11 is the graphical comparison between the PM and the contemporary methods against considered steganalysis schemes. The intercept or downward moment of the curve with the line at $45°$ is the worst case of steganalysis. In this case, it can be observed from the graphs that the PM is highly resistant against the steganalysis schemes considered as the ROC curve of PM is found with a lot of false predictions by the schemes. Hence, the PM is undetectable against recent steganalysis techniques. Though motion vectors and hash-based techniques are used in the steganalysis schemes, the utilization of dynamic keypoints region of the PM improves the efficacy against distinct steganalysis schemes.

4 Conclusion and perspectives

In this paper, first, the significance of steganography, and some recent works in the video steganography domain have been expressed. Then, to aid the PM concepts, preliminary concepts such as 2D-DCT transformation, SIFT, and SURF keypoint detection techniques are presented. Next, the proposed embedding and extraction modules are elaborated. In the embedding phase, a hybrid version of SIFT and SURF methods; h-SUSIRF keypoint detection algorithm is proposed to select dynamic ROI keypoints from the partitioned IF and PF video frames with minimum computational cost. Wherein, multiple secret images collected from UCID databased are allocated to each partition. Then embedding process is performed only in the red and green components of BF frames using the LSB substitution method and sent for communication. At the extraction module, the exact



(a) (b)

(c) (d)

Figure (11) The ROC representation of the PM and the contemporary methods [[23], [24], [29]] against steganalysis schemes (a). Fan et al. [33], (b).Tasdemir et al. [34], (c). Su et al. [35] and (d). Wang et al. [36]

secret images are retrieved from these BF frames and the cover video regions are semi reversed.

The experimental results of PM are analyzed and compared with some contemporary works in terms of imperceptibility, capacity, and robustness. From the results and analysis, it can be inferred that the PM significantly outperforms the existing methods; the average values of SSIM, C_P, and BER are 0.9896, 0.7654, and 0.2145 respectively. Then, the quality of the stego-videos is tested using VQM, and the computational cost to generate those videos is examined. The proposed h-SUSIRF reduces the computational complexity to a good extent. Also, the robustness of the stego-videos against some recent steganalysis techniques is investigated and compared with the existing methods using the ROC curve. The generated ROC curves of the PM (Figure 10) imply that the PM is resilient against any kind of feature, motion, and neural network-based steganalysis. The PM can be applied in any domain where highly secure communication plays a pivotal role. Also, it can easily be adapted in real-time application scenarios as the computational complexity is relatively low.

There are two limitations in the PM: (1) security and (2) semi-reversibility. Since the embedding process is performed in the spatial domain at times security can be a challenging task. Similarly, the PM can reverse only partial cover video regions at the receiver end. So this work might be unsuitable for the scenarios where a fully recovered cover video is demanded. Thus, these two breakdowns can be optimized in the future by employing a reversible embedding process in a wavelet or transform domains.

References

[1] Kumar, S., Soundrapandiyan, R., ”A multi-image hiding technique in dilated video regions based on cooperative game-theoretic approach”, *Journal of King Saud University-Computer and Information Sciences.*

[2] Yeh, H. L., Gue, S. T., Tsai, P., Shih, W. K., ”Reversible video data hiding using neighbouring similarity”, *IET Signal Processing* 8(6), 579-587, (2014).

[3] Sadek, M. M., Khalifa, A. S., Mostafa, M. G., ”Video steganography: comprehensive review”, *Multimedia tools and applications* 74(17), 7063-7094, (2015)

[4] Muhammad, K., Ahmad, J., Rho, S., Baik, S. W., ”Image steganography for authenticity of visual contents in social networks”, *Multimedia Tools and Applications* 76(18), 18985-19004, 2017.

[5] Rabie, T., Baziyad, M., ”The pixogram: Addressing high payload demands for video steganography”,*IEEE Access* 7, 21948-21962, (2019).

[6] Mstafa, R. J., Elleithy, K. M., ”A video steganography algorithm based on Kanade-Lucas-Tomasi tracking algorithm and error correcting codes”, *Multimedia Tools and Applications*, 75(17), 10311-10333, (2016).

[7] Zhao, H., Dai, Q., Ren, J. C., Wei, W., Xiao, Y., Li, C.,”Robust information hiding in low-resolution videos with quantization index modulation in DCT-CS domain”, *Multimedia Tools and Applications*, 77(14), 18827-18847,(2018).

[8] Liu, Y., Liu, S., Wang, Y., Zhao, H., Liu, S.,”Video steganography: A review”, *Neurocomputing*, 335, 238-250, (2019).

[9] Lakshmi, M., Arjun, K. P., Sreenarayanan, N. M., Arya, K. A.,”Reversible Data Hiding in Videos for Better Visibility and Minimal Transfer”, *Procedia Technology*, 25, 256-263, (2016).

[10] Mstafa, R. J., Elleithy, K. M., Abdelfattah, E.,”A robust and secure video steganography method in DWT-DCT domains based on multiple objects tracking and ECC”, *IEEE Access*, 5, 5354-5365, (2017).

[11] Chung, K. L., Chiu, C. Y., Yu, T. Y., Huang, P. L.,"Temporal and spatial correlation-based reversible data hiding for RGB CFA videos", *Information Sciences*, 420, 386-402, (2017).

[12] Rajalakshmi, K., Mahesh, K.,"Robust secure video steganography using reversible patch-wise code-based embedding", *Multimedia Tools and Applications*, 77(20), 27427-27445,(2018).

[13] Banerjee, A., Jana, B.,"A Secure High-Capacity Video Steganography Using Bit Plane Slicing Through (7, 4) Hamming Code", *In Advanced Computational and Communication Paradigms*, (pp. 85-98). Springer, Singapore, (2018).

[14] Luo, T., Jiang, G., Yu, M., Xu, H., Gao, W., "Sparse recovery based reversible data hiding method using the human visual system", *Multimedia Tools and Applications*, 77(15), 19027-19050, (2018).

[15] Ramalingam, M., Isa, N. A. M.," A data-hiding technique using scene-change detection for video steganography", *Computers & Electrical Engineering*, 54, 423-434, (2016).

[16] Cao, Y., Zhang, H., Zhao, X., Yu, H.,"Covert communication by compressed videos exploiting the uncertainty of motion estimation", *IEEE Communications Letters*, 19(2), 203-206,(2014).

[17] Yao, Y., Zhang, W., Yu, N., Zhao, X.,"Defining embedding distortion for motion vector-based video steganography",*Multimedia tools and Applications*, 74(24), 11163-11186, (2015).

[18] Song, G., Li, Z., Zhao, J., Hu, J., Tu, H.,"A reversible video steganography algorithm for MVC based on motion vector",*Multimedia Tools and Applications*, 74(11), 3759-3782, (2015).

[19] Liao, X., Guo, S., Yin, J., Wang, H., Li, X., Sangaiah, A. K.,"New cubic reference table-based image steganography",*Multimedia Tools and Applications*, 77(8), 10033-10050, (2018).

[20] Yao, Y., Zhang, W., Yu, N.,"Inter-frame distortion drift analysis for reversible data hiding in encrypted H. 264/AVC video bitstreams",*Signal Processing*, 128, 531-545, (2016).

[21] Yan, H., Li, J., Wen, H.,"A key points-based blind watermarking approach for vector geo-spatial data. Computers",*Environment and Urban Systems*, 35(6), 485-492 (2011).

[22] Manikandan, V. M., Masilamani, V.,"Histogram shifting-based blind watermarking scheme for copyright protection in 5G",*Computers & Electrical Engineering*, 72, 614-630, (2018).

[23] Mstafa, R. J., Younis, Y. M., Hussein, H. I., Atto, M.,"A new video steganography scheme based on Shi-Tomasi corner detector", *IEEE Access*, 8, 161825-161837, (2020).

[24] Hashemzadeh, M.,"Hiding information in videos using motion clues of feature points",*Computers & Electrical Engineering*, 68, 14-25, (2018).

[25] Biswas, R., Bandyapadhay, S. K.,"Random selection-based GA optimization in 2D-DCT domain color image steganography", *Multimedia Tools and Applications*, 1-20, (2019).

[26] Lowe, D. G.,"Distinctive image features from scale-invariant keypoints", *International journal of computer vision*, 60(2), 91-110, (2004).

[27] Manisha, S., Sharmila, T. S., "A two-level secure data hiding algorithm for video steganography", *Multidimensional Systems and Signal Processing*, 30(2), 529-542, (2019).

[28] Dataset: https://media.xiph.org/video/derf/

[29] Yao, Y., Yu, N.,"Motion vector modification distortion analysis-based payload allocation for video steganography",*Journal of Visual Communication and Image Representation*, 74, 102986, (2021).

[30] Rajkumar, S., Malathi, G.,"A comparative analysis on image quality assessment for real time satellite images", *Indian J. Sci. Technol*, 9(34), (2016).

[31] Pinson, M. H., Wolf, S.,"A new standardized method for objectively measuring video quality",*IEEE Transactions on broadcasting*, 50(3), 312-322, (2004).

[32] Narwaria, M., Da Silva, M. P., Le Callet, P.,"HDR-VQM: An objective quality measure for high dynamic range video", *Signal Processing: Image Communication*, 35, 46-60, (2015).

[33] Fan, M., Liu, P., Wang, H., Sun, X.,"Cross correlation feature mining for steganalysis of hash based least significant bit substitution video steganography", *Telecommunication Systems*, 63(4), 523-529, (2016).

[34] Tasdemir, K., Kurugollu, F., Sezer, S.,"Spatio-temporal rich model-based video steganalysis on cross sections of motion vector planes",*IEEE transactions on Image Processing*, 25(7), 3316-3328, (2016).

[35] Su, Y., Yu, F., Zhang, C.,"Digital video steganalysis based on a spatial temporal detector",*KSII Transactions on Internet and Information Systems (TIIS)*, 11(1), 360-373, (2017).

[36] Wang, P., Cao, Y., Zhao, X., "Segmentation based video Steganalysis to detect motion vector modification", *Security and Communication Networks*, 2017.

[37] Solanki, K., Sarkar, A., Manjunath, B. S.,"YASS: Yet another steganographic scheme that resists blind steganalysis",*In International Workshop on Information Hiding*, (pp. 16-31), Springer, Berlin, Heidelberg, (2007, June).