Research Article

Urmila Pilania, Rohit Tanwar*, and Prinima Gupta

An ROI-based robust video steganography technique using SVD in wavelet domain

https://doi.org/10.1515/comp-2020-0229
received March 4, 2020; accepted June 1, 2021

Abstract: Steganography is a technique that embeds secret information in a suitable cover file such as text, image, audio, and video in such a manner that secret information remains invisible to the outside world. The study of the literature relevant to video steganography reveals that a tradeoff exists in attaining the acceptable values of various evaluation parameters such as a higher capacity usually results in lesser robustness or imperceptibility. In this article, we propose a technique that achieves high capacity along with required robustness. The embedding capacity is increased using singular value decomposition compression. To achieve the desired robustness, we constrain the embedding of the secret message in the region of interest in the cover video file. In this manner, we also succeed in maintaining the required imperceptibility. We prefer Haar-based lifting scheme in the wavelet domain for embedding the information because of its intrinsic benefits. We have implemented our suggested technique using MATLAB. The analysis of results on the prespecified parameters of the steganography justifies the effectiveness of the proposed technique.

Keywords: Haar-based lifting scheme, singular value decomposition, face detection, region of interest, watermark

*** Corresponding author: Rohit Tanwar,** Department of Systemics, School of Computer Science, University of Petroleum & Energy Studies, Dehradun-248007, Uttrakhand, India,
e-mail: r.tanwar@ddn.upes.ac.in
Urmila Pilania: Department of Computer Science and Engineering, Manav Rachna University, Faridabad-121004, Haryana, India,
e-mail: urmila@mru.edu.in
Prinima Gupta: Department of Computer Science and Engineering, Manav Rachna University, Faridabad-121004, Haryana, India,
e-mail: prinima@mru.edu.in

1 Introduction

Steganography is in existence for a long time. It comes from Greek roots (στεγαν́ος, γραφ-ειν), literally meaning "sheltered writing." Technically, it means embedding information in other information [1]. Depending upon the type of cover file used, steganography is categorized as text, image, audio, and video steganography [2]. Despite the availability of a variety of media for embedding data, the video started gaining popularity because of some benefits inherited from the image steganography. Traditionally, video steganography provides a good capacity to hide secret information. It is conceptualized by decomposing the cover video into frames, and then these frames are converted to image and audio. Ultimately, the secret information can be embedded either in images or in audio [3]. As videos are dynamic, the probability of suspecting the existence of any embedded information is very less as compared to still images and audio [4,5].

A user generally desires to increase the data hiding capacity but unfortunately, robustness and imperceptibility are compromised at the same time. In this article, we have developed a novel robust video steganography approach with enhanced capacity. Singular value decomposition (SVD) technique is used to perform the required compression that eventually contributes to the emergence of data hiding capacity. To get the required robustness, the secret data is embedded in the region of interest (ROI) only instead of the entire frame. The embedding is done in the wavelet domain using lifting wavelet transform (LWT). The preferred use of the wavelet domain for the embedding task provides the additional inherent robustness of the wavelet domain. Moreover, in Haar-based lifting scheme, the number of computations is almost half than the other traditional wavelet schemes. It reduces the storage as well as time requirement using in-place computation of the wavelet transform.

A hybrid technique that collectively utilizes the lifting scheme, ROI, and SVD is proposed, for video files, which makes it useful for real-time applications. This technique can successfully transmit the embedded information and recover it efficiently at the other end.

2 Literature survey

Steganography is an extension of watermarking and encryption techniques. Numerous encryptions, watermarking, and steganography techniques exist with varying characteristics. Some of the parameters that compare the quality of steganography techniques are transparency, imperceptibility, security, computational time for embedding and extraction, robustness, effect on bandwidth, and interoperability.

The author proposed an effective digital watermarking scheme in an earlier study [6] in which the image was transformed from red green blue (RGB) to the YCbCr domain, and then the secret image with watermark is rooted in the *Y* module of YCbCr by discrete wavelet transform (DWT) and SVD. The robustness of the embedded image is calculated in terms of normalized coefficient. A robust image watermarking scheme is proposed by the author in an earlier study [7] by combining the lifting wavelet scheme and SVD, which is used to hide the secret information. Two-level LWT is applied on the cover file along with SVD to hide the secret image. The author in an earlier study [8] proposed an improved technique in terms of robustness and imperceptibility. Audio video interleave video files were used as a cover file. Video files having audio were divided to get frames. Frames are analogous to unmoving images and used in image steganography. When audio is extracted from the video, it is like an audio file, and later it can be used in steganography. Video and audio can be used as a cover file as proposed in an earlier study [9] to obtain high embedding capacity. When video trim is read, several video and audio frames become accessible to embed secret information. Experimental observation showed that the proposed technique gives good quality watermark having good peak signal to noise ratio (PSNR) in realistic execution time.

Embedding of secret information inside transform domain is more time consuming but provides more robustness to secret information. Authors in an earlier study [10] proposed an integer wavelet transform technique for embedding information in HH coefficients along with cropping in 8×8 block on wrap file. After embedding information in HH coefficients, optimal pixel alteration procedure was applied for more robustness of watermarked image and to decrease the dissimilarity error between the cover and watermarked image. This improved the embedding ability with low distortions. The author proposed the LWT technique to convert an image into four subbands. The subbands carrying energy more than the calculated "*Q*" value were chosen for watermark embedding. SVD was also used to hide the digital signature as a watermark. Many image processing and geometric attacks are there which can detect secret information. In an earlier study [11], authors proposed a technique that was reliable in founding ownership and was robust against attacks. A hybrid policy of cryptography, steganography, and digital watermarking was proposed to embed a protected image with a watermark logo into a wrap file.

Working in a similar domain, a lossless dual watermarking scheme by integral wavelet transform (IWT) and SVD was implemented by Harshitha and Vidya [12]. The extracted watermark had a good correlation with the original cover file. Initially, the secret information was embedded on the LL subband of SVD, and the resulting image SVD was embedded in the host watermark image. Embedding of secret information in video increases capacity as well as robustness because of its large size and dynamic nature. Discrete cosine transforms (DCT)–DWT method using several object tracking and error correction codes were recommended by embedding a secret message in items that are in motion in an earlier study [13].

Authors in this study used Haar wavelet transform in the DWT domain. First, the secret information was preprocessed with Bose–Chaudhuri–Hocquenghem and Hamming codes (n, k) to generate programed communication. In the second step, motion-based multiple entities observing procedure was applied on wrap files to discover the ROI of items in motion. In the next step, the secret information embedding procedure was completed by embedding the secret information into DWT and DCT coefficients of all action locations in video depending on forefront masks. The method achieved a PSNR value of *49.03DB* (which is assumed to be of good visual quality). The researchers developed a technique to present a high-level security organization by implementing and designing a multilevel steganography system to embed information in a color video file [14]. This scheme was executed in the frequency domain, using a wavelet transform. Embedding information in the frequency domain was extra supportive than embedding in the time domain, due to the compactness attributes of some transforms and its stoutness. SVD was also used in the planned system for providing more robustness.

3 Proposed work and methodology

The objective is to develop, analyze, and validate a video steganography technique that ensures large data embedding capacity, high security, flexibility, and good imperceptibility without loss of secret information. The hybrid of Haar-based lifting scheme, ROI, and SVD for embedding the secret image in a video file yields the expected values for the above-mentioned evaluation parameters.

Traditional wavelet domain techniques use up and down sampling, which results in loss of information. Moreover, transforming the image in the frequency domain increases calculation difficulty in these techniques. As this technique works on a floating-point number, the reconstruction of the watermark is also difficult. To overcome all these problems, a stable, high-capacity video steganography is proposed that collaborates the face detection and SVD techniques in the wavelet domain. In Haar-based lifting scheme, there is no up and down sampling. It increases cyclic redundancy check sum values resulting in highly robust steganography that provides a good reconstruction of watermark information, increasing smoothness, and decreasing aliasing effects. Additionally, the calculation difficulty of Haar-based lifting scheme is reduced due to the split and merge procedure.

3.1 ROI

Embedding in ROI results in more imperceptibility. In this study, the embedding is done in ROI, which is a human face for this implementation. To locate ROI first, the skin portion in the image is identified, and the features are identified to locate the human face by using Viola–Jones method. Some of the important features of Viola–Jones method, which build it a good face detection method are as follows:

- Robust – higher rate of face detection.
- Real time – process rate of frames per second is high in case of Viola–Jones that is two frames per second at least.
- Face detection only – its main goal is to detect faces not to recognize faces.
- It is able to detect a single face among multiple faces.

Viola–Jones method takes total of four steps to detect a face as shown in Figure 1 [15]:

Haar feature selection: Haar features are represented by rectangular box. These features include eyes, nose, eyebrow, and lips. Eye regions are darker as compared to upper checks and nose bridge. All human faces have these features. If these features are present in video/image, it means face is there.

Creating an integral image: Representation of frames in video and evaluation of rectangular features in fixed time interval is known as an integral image.

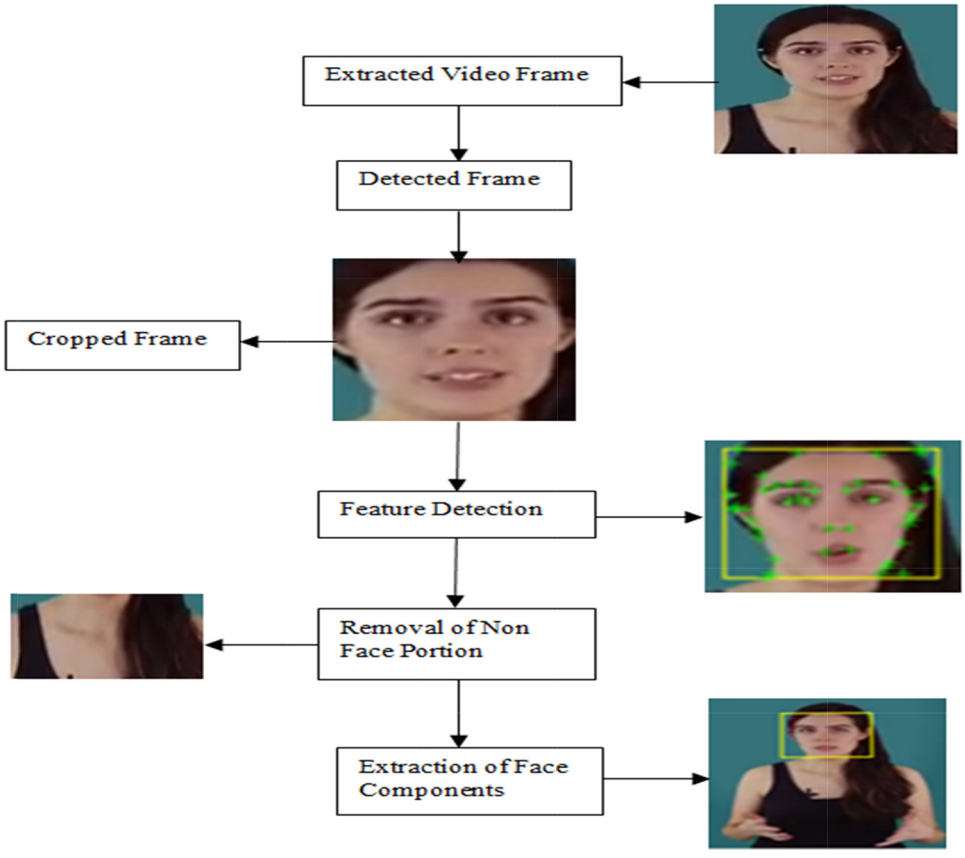


Figure 1: Working of Viola–Jones method.

Adaboost training: It can build a strong classifier for selection of best features and then to train classifiers which use these features. Adaboost build a strong classifier by combining many simple and weak classifiers.

Cascading classifiers: Job of the cascading classifier is to find a face. Cascade classifier consists of a number of levels and each level carry a strong classifier. The features selected by these classifiers are combined into a number of stages where every stage has assured number of features.

Embedding secret information inside the face part of the image is more secure than embedding in other parts of the image. The human face has most of the information in the low-frequency range, thus high-frequency range can be used for embedding information. Skin tone detection finds change among skin and nonskin pixels. Algorithm alters specified pixels into the suitable color region and uses skin classifiers for tagging whether pixels are skin color pixels or not and sets a few restrictions to differentiate between skin or nonskin color pixels as per the work done in an earlier study [16]. According to Kawulok et al. [17], while finding skin pixels, it is important for symbolizing skin color pixels that color is consistent with variations in bright conditions. Bayes rule can be used to find the skin pixel as given in equation (1).

$$P(C_s|V) = \frac{[P(V|C_s)\cdot P(C_s)]}{[P(V|C_s)\cdot P(C_s)] + [P(V|C_{ns})\cdot P(C_{ns})]}, \quad (1)$$

where $P(C_s)$ and $P(C_{ns})$ are priori probabilities and can be calculated by using many pixels in both the classes.

A specific combination of R, G, B values is there to qualify as a skin pixel. Only then, the pixel can be categorized as a skin pixel as given below:

$$R > 95 \,\&\, G > 40 \,\&\, B > 20 \,\&\, \max(R, G, B) - \min(R, G, B) > 15 \,\&\, |R - G| > 15 \,\&\, R > G \,\&\, R > B.$$

Many objects may contain skin-tone colors due to which the skin color detector may consider image surroundings as skin pixels if the atmosphere is not restricted according to Cheddad et al. [18]. The limitations can be overcome by setting boundary conditions for the existence of skin pixels as calculated by Viola et al. [19]. Thus, Viola–Jones algorithm is used in the proposed scheme for locating face as ROI. Figures 2 and 3 show the detected skin portion and face in the chosen video cover file, respectively. Viola–Jones algorithm can find a single face along with multiple faces in a given video, provided these faces should be directly in front of the camera without a tilt.

Fast features of an image are calculated efficiently and in less time. Instead of scaling the entire image, only features are scaled as a part of this algorithm, which enables it to detect other real-world identities such as a table, chair, and so on [18]. The literature reveals that this algorithm is capable of determining faces in the long-wavelength infrared images along with finding images in the visible spectrum. The precision of this algorithm is good enough among different datasets, approaching 97% [17]. Like any other algorithms, it also has some limitations. It is not very effective in detecting tilted or turned faces. It is somewhat sensitive to lighting conditions and may result in different detections of the same face because of overlapping subwindows. In some cases, it puzzled a shirt button as a face while detecting faces and tracking multiple faces [2].

3.2 SVD

SVD is a significant tool in linear algebra that is commonly used in a lot of research fields such as principal components analysis, chosen cover text attack, multivariate analysis, and image compression. SVD creates a characteristic equation whose measures are identical to the number of rows and columns of the given image. It has excellent stability and does not have degradation

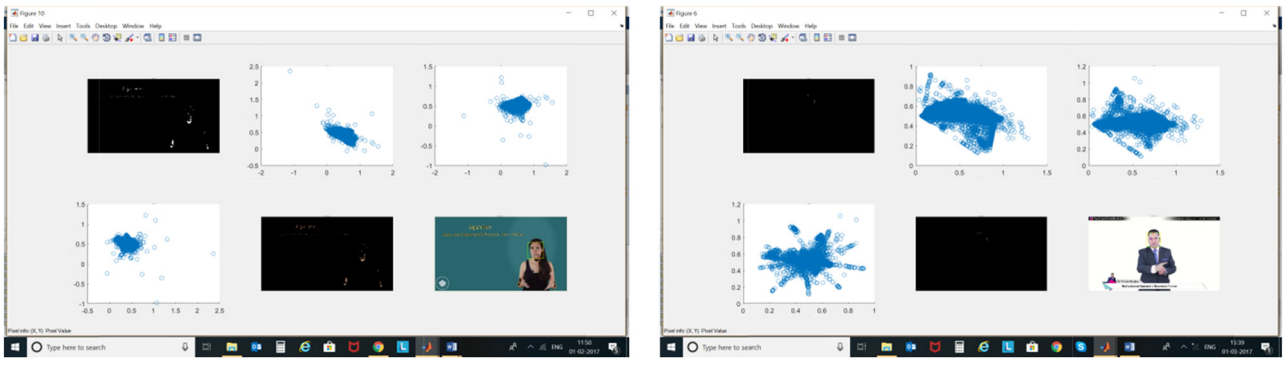


Figure 2: Detection of skin portions.

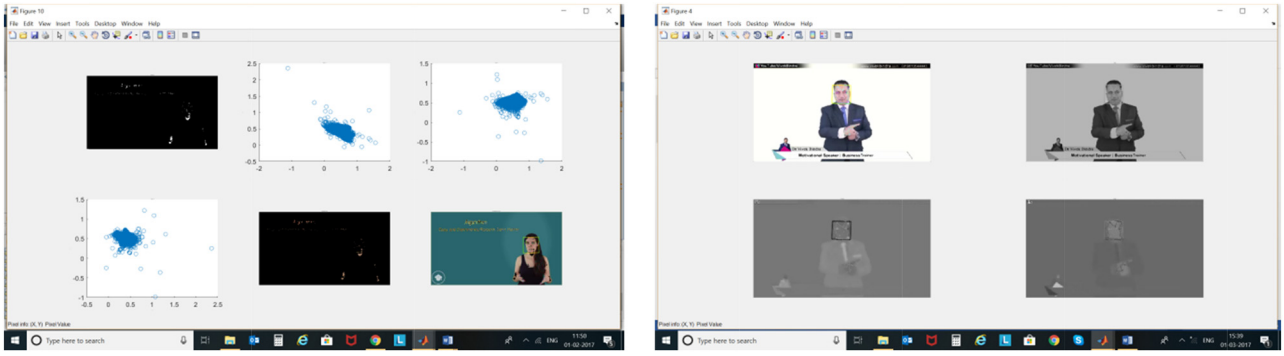


Figure 3: Detection of faces.

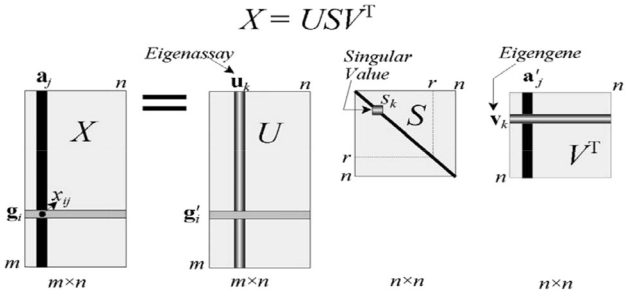


Figure 4: Graphical representation of SVD of matrix "*X*."

because singular values symbolize algebraic image properties, which are inherent and invisible. It also helps to retain fidelity of watermarked image and reconstruct watermark images efficiently. In linear algebra, SVD is a factorization of a real or complex matrix. It also generalizes the eigen-decomposition of a symmetric matrix having positive eigen values to any matrix through a delay of polar decomposition as shown in Figure 4. SVD for image "*X*" as given in Wall et al. [20] produces two orthogonal metrics, and *S* represents a singular matrix. *U* and *V* act as left and right exacting vectors. Matrix *U* and *V* represent geometry understated fundamentals of the image according to Acharya et al. [21]. *U* talks to flat surface points of interest and left singular matrix; and *V* talks to vertical understated elements of the image. Below mentioned commands are used for implementing SVD in MATLAB:

Corner to corner estimations of matrix *S* is prepared in withdrawing request, which means that the significance of passages is withdrawing from first singular value to last value. This module is used in SVD-based compression techniques. Two most important properties of SVD programmed watermarking plans, little variations in singular values do not manipulate the behavior of the picture; singular estimations of the image contain more reliability according to Wall et al. [20].

3.3 Haar-based lifting scheme

Haar developed wavelet transform in the year 1910. Haar wavelet transform is a function that "waves" above and below the *x*-axis based on the properties of (1) varying frequency, (2) limited duration, and (3) zero average value as shown in Figure 5. Haar wavelet transform uses a variable-length window.

In this wavelet, narrower windows are more appropriate at high frequencies and wider windows are more appropriate at lower frequencies wavelet. This transform is based on symmetrical matrices whose fundamentals are "1," "–1," or "0" multiplied by the power of root "2." It is computationally efficient as a transform of N-point vectors needs 2(*N*–1) multiplications. Haar scaling captures information at different frequencies and

```
[U_y, S_y, V_y]=SVD[LL_1]  //SVD of the LL component of the frame in video in which we insert secret image.
Q= size(S_y);              //Variable Q store size of singular matrix Sy to calculate value of PSNR and MSE.
[U_w, S_w, V_w]=SVD[double(Secret_Image)];
//SVD of the secret image which is to be hidden in the cover video frame.
[U_y_wmv, S_y_wmv, V_y_wmv]=SVD(LL_1_wmv);
//SVD of extracted watermarked image.
```

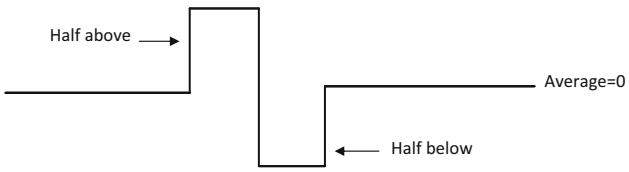


Figure 5: Haar transform.

translation captured information at different locations. The wavelet scaling and wavelet functions are shown in Figure 6.

Algorithm for Haar function:

1. Haar function $h_k(x)$ is characterized on continuous period, where
   $x \in [-1,1]$ and for $k = 0, 1, 2, \ldots N-1$ and $N = 2n$
2. Integer k is exceptionally divided as follows: $2p + q - 1$ where $0 \le p \le n-1$; $q = 0, 1$ for $p = 0$ and $1 \le q \le 2p$ for $p \ne 0$.
3. Let us consider when $N = 4$ (or $n = 2$), we include

k	0	1	2	3
p	0	0	1	1
q	0	1	1	2

Representing k by (p, q), Haar function is given as follows:

$$h_0(x) \equiv h_{0,0}(x) = 1/\sqrt{n},\ x \in [0, 1]$$

$$h_k(x) \equiv h_{p,q}(x) = \begin{cases} 2^{p/2}, & (q-1)/2^p \le x < (q-1/2)/2^p, \\ -2^{p/2}, & (q-1/2)/2^p \le x < q/2^p, \\ 0, & \text{Otherwise } x \in [0, 1]. \end{cases} \quad (2)$$

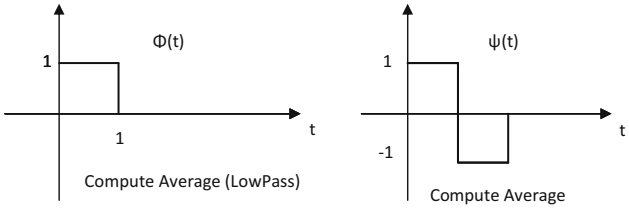


Figure 6: Haar function.

Haar-based lifting scheme works on integer values in both the domains, time, and frequency. Usually, in the wavelet domain, we embed information in areas that the human visual system is not as responsive to, such as high-resolution feature bands HL, LH, and HH [21]. Embedding information in LL areas allows improved toughness although maintaining fine visual excellence. The image can reconstruct without any loss because of the integer nature of coefficients and stored without rounding off errors [7]. Haar-based lifting scheme is faster as compared to real numbers arithmetic in computer registers because longer data length is required by floating-point numbers [10]. Another benefit of using the lifting scheme is reversibility because it maps integer to integer. The main benefit is that it takes less computational time because of filter bank structure, and it is also able to produce a method for filter design [22]. Below mentioned commands are used for implementing Haar wavelet in MATLAB.

```
[LL1, HL1, LH1, HH1]=IWT('Video _Frame, 'Haar')//Wavelet transform of the cover video frame in four
co-efficient that are LL, LH, HL and HH.
Smark =Sy+alpha×Sw//Smark is embedding of secret image inside cover video
frame. Here Sy and Sw are the singular value for the
cover frame and secret image respectively.
Video_frame=IWT(LL1, HL1, LH1, HH1, 'Haar');// Inverse wavelet transform to get the watermarked
image.
Sw_rec=(Sy_wmv-Sy)/alpha;//Sw_rec is the extracted watermark.
```

4 Proposed work

The proposed video steganography technique is described in two modules: (1) embedding process and (2) extraction process. The flowchart in Figure 7 describes the embedding process.

4.1 Embedding process

The embedding process begins by selecting a video file as a cover. Then we divided the video file into frames, which provide multiple options where secret information could be embedded. Viola–Jones method is used to find the ROI. It helped in detecting skin tone and then find the faces inside the video. The human face has most of the information in the low-frequency range thereby enabling the high-frequency range to be used for embedding information. The application of the Haar-based lifting scheme on ROI provides the "HH" component of the cover video

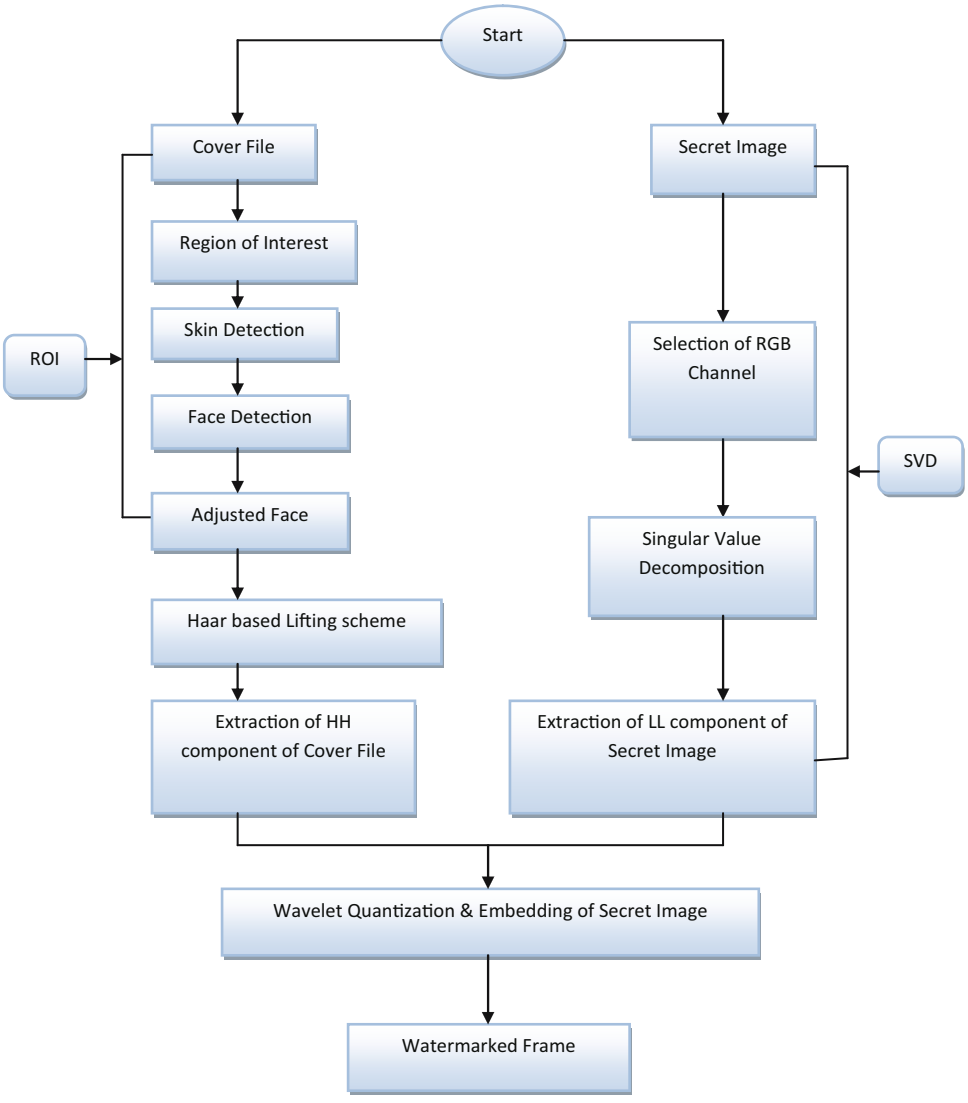


Figure 7: Embedding process.

file. Then the secret information was embedded inside the noisy part of the "HH" component. For playing video inside MATLAB, we selected the RGB channel.

The secret information is in the form of an image. Then we apply SVD on the image and decompose the secret image into LL, HH, LH, and HL components. By using SVD, we found the "LL" component of the secret image, and "LL" component has all useful information in it. Then we superimposed these "LL" components in the "HH" components of the cover file by wavelet quantization. After wavelet quantization, the watermarked frame was produced with almost similar quality as the original cover file. It is now difficult to figure out whether there exists any secret information. Algorithm for the embedding process is as follows:

Algorithm. ROI_Lifting_scheme_Haar_wavelet_SVD

Input: cover_video, extract_frame, secret_image
Output: steg_video

1. temp_video ← copy(cover_video);//copy cover video
2. temp_img ← copy(secret_image); //copy secret image
3. ROI[frm, n] ← extract_frame_using_Viola_Jones (temp_video);//extract frames
4. [LL, LH, HL, HH] ← SVD(temp_img); //apply SVD on image
5. Initialize $i = 1$, $j = 1$, k, count;
6. Do
7. if (j + count) > max(LL) //LL component have max data

```
8.              count ← max(LL)-j;    //avoid overflow
9.          else
10.             continue;
11.         k ← j + count;
12.         temp_frame ←
    Lifting_scheme_Haar_wavelet(frm[i]);
13.         bits_to_hide ← LL[j,k];
14.         embed(temp_frame, bits_to_hide);
15.         updatei ← i + 1;
16.         update j ← k + 1;
17.       while((i ≤ n) AND (j < max(LL));    //all frames
    used or complete message hidden
18. steg_video ← build_video(frm);//construct video
    to transmit
```

4.2 Extraction process

We apply the SVD on the watermarked frame to find its "LL" components for the extraction. After that, the application of inverse wavelet transform on watermarked frame yields its "LL," "HH," "HL," and "LH" coefficients, respectively. In the next step, we obtain a cover image from the "HH" coefficients. After updating the RGB channel, the four components were combined to form frames. Two files are produced in this process; one a secret message, which is an image and another is the cover file. The extraction process is shown in Figure 8 and algorithm for the same is as follows:

5 Evaluation parameters

The effectiveness of a steganography technique is generally measured on three factors: data hiding capacity, robustness, and imperceptibility. We used various parameters as a measure of attainment of these factors, such as capacity, PSNR, and mean square error (MSE), structure similarity index matrix (SSIM), correlation coefficient (CC), and histogram.

5.1 Embedding capacity

The embedding capacity is the maximum number of bits of secret information that can be hidden in the selected cover frame [23,24]. The size of a frame mainly depends on color depth and resolution as shown in equation (3). As video is the combination of multiple frames, the size of video can be calculated as in equation (4).

$$\text{Frame size} = V \times H \times \text{CD} \tag{3}$$

where "V" stands for vertical resolution, "H" stand for horizontal resolution, and "CD" stands for color depth.

$$\text{Video size} = \text{Frame size} \times f \times t \tag{4}$$

where "f" stands for frames per second and "t" is the total duration of video in seconds. Frame size can be measured in terms of pixel resolution.

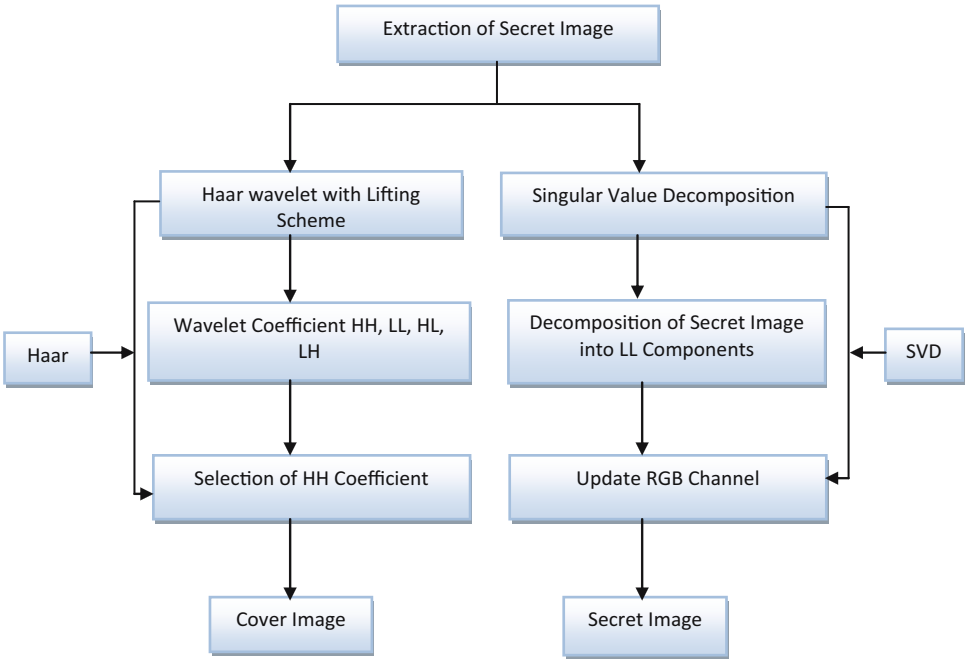


Figure 8: Extraction process.

Embedding capacity of video steganography technique can be calculated with the help of equation (5) when the entire frame is considered for hiding the secret information.

$$\text{Capacity with entire frame} = \frac{n \times f}{F} \tag{5}$$

where "n" is the number of bits embedded in frame, "F" is the size of frame in terms of height and width, and "f" is frame rate. Equation (6) is formed by combining equations (4) and (5) as follows:

$$\text{Capacity with entire frame} = \frac{n \times f \times f \times t}{V} \tag{6}$$

Viola–Jones algorithm is used for finding ROI in the video cover file. Equation (7) calculates the embedding capacity for the ROI as follows:

$$\text{Capacity for ROI} = \left(\frac{P2}{P1}\right) \times \frac{n' \times f}{F} \tag{7}$$

where "n'" is the modified number of bits embedded in ROI of a frame, "$P1$" represents probability of presence of face in a frame and "$P2$" represent probability of finding the face.

Video files are getting popular because of their great embedding ability [25]. The embedding capacity of the proposed technique is calculated for the different cover videos using equation (7). As all the frames in these video contain faces, $P1$ and $P2$ are kept as "1." The method achieves an average embedding capacity of 51.7 *BPS* as shown in Table 1. The embedding capacity is significantly affected by the type of video (number of frames containing faces).

5.2 Imperceptibility

Imperceptibility is a measure of the modification induced in the cover file after embedding the secret message. The steganography technique is highly imperceptible if a third person cannot suspect the presence of any information embedded inside a cover file. However, the probability of suspecting the presence of a secret message inside the cover file is high if imperceptibility is low [26]. The average value of SSIM achieved shows that the technique is highly imperceptible.

Table 1: Embedding capacity

Cover video	Frame size ($W \times H$)	Number of frames	Frames rate (fps)	Capacity (bps)
Video 1	190 × 160	750	21	51.8
Video 2	200 × 190	890	22	51.5
Video 3	210 × 190	899	23	51.8

5.3 Robustness

It is the ability of the embedded message to remain unaltered and undamaged even if the stego file is subjected to various transformations. Image exploitation, such as editing or pivoting, can be performed on the image. These operations may destroy the embedded information. The highest need for steganography algorithms is the robustness against malicious modifications to the cover file [27].

Robustness and imperceptibility are qualitatively measured using PSNR, MSE, and SSIM.

PSNR is a quality measure that tells the extent to which one got succeeded in reconstituting a file after some processing. It approximates the human perception of reconstruction quality. In the case of video steganography, it refers to the extent to which a stego file looks like the original cover file. It is measured in terms of MSE [28] as given in equation (8):

$$\text{PSNR} = 10 \log_{10}\left(\frac{\text{MAX}_\text{I}^2}{\text{MSE}}\right) \tag{8}$$

where, MAX_I = maximum possible power of a signal and and MSE = the power of corrupting noise.

MSE is a measure of the deviation of the reconstituted signal from the original signal. It is measured as an average of the square of the error between the original file and the watermark. It can be calculated using the formula [29] given in equation (9):

$$\text{MSE} = \frac{1}{n}\sum_{i=0}^{i=n}[I(i) - J(i)]^2, \tag{9}$$

where $I(i)$ = cover file and $J(i)$ = watermark.

SSIM is a measure of similarity or dissimilarity between two images. It is used to quantify the originality maintained in an image after some processing has been done. Unlike PSNR and MSE, it measures the differences between visible structures of two images. The acceptable value of SSIM is near to "1," which shows very much similarity between the original message and the extracted message. It can be measured using the formula [30] specified in equation (10):

$$\mathrm{SSIM}(x, y) = [l(x, y)^{\alpha} \cdot c(x, y)^{\beta} \cdot s(x, y)^{\gamma}], \quad (10)$$

where l(x, y) = luminance of samples x and y, $c(x, y)$ = contrast of samples, and $s(x, y)$ = structure of samples, α, β, γ, denote the relative importance of each component and usually set to 1.

CC is used to determine the power of connection between the two images. CC values that lie between "−0.8" and "+0.8" are considered significant [31].

Histogram is a graphical representation of the frequency of pixels intensity values of an image. The x-axis and y-axis, respectively, show the intensity and frequency of these intensities [32].

The ultimate objective of video steganography techniques is to achieve a balance among the data embedding capacity, robustness while maintaining video quality and imperceptibility. The method of data embedding should be robust against attacks to make it useful for practical applications.

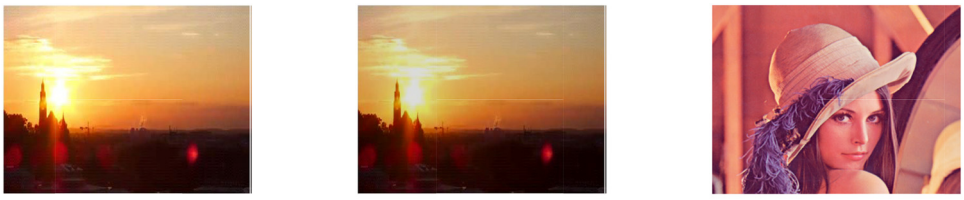
6 Experimental results and discussion

This section provides the details of the implementation environment comprising the tool, libraries, input, and output used to implement and evaluate the proposed research.

Tool used: We have implemented the proposed technique using MATLAB, which is a multiparadigm environment developed for numerical computing primarily. With time, it got equipped with a list of tool boxes developed for specific tasks for file processing. Initially, the command window was the only way to interact for executing files. Later on, an additional feature of the graphical interface benefitted the users. MATLAB 2016 has been used to implement the proposed technique because of its broad applicability, popularity, and effectiveness. We used the corresponding MATLAB libraries to perform specific tasks comprising the technique. We have generated the required graphs in MATLAB to perform further analysis and to validate the results [33,34].

Input and output: The famous "Lena" image is used as a secret message that is to be transmitted securely using the proposed method. We embedded the secret message in three different video files sequentially. The three cover files were having different attributes (no. of frames per second, kbps, size, etc.). The extraction process described in Section 4.2 is used to successfully extract the secret message with acceptable quality.

Figures 9a–c, 10a–c, and 11a–c show (a) the original frame, (b) watermarked frame, and (c) extracted message, respectively, for the three cover video files. It is clear that the method can extract the embedded message without significant distortion irrespective of the video file chosen as cover.



a) b) c)

Figure 9: (a) Cover frame, (b) watermarked frame, and (c) extracted message.

a) b) c)

Figure 10: (a) Cover frame, (b) watermarked frame, and (c) extracted message.

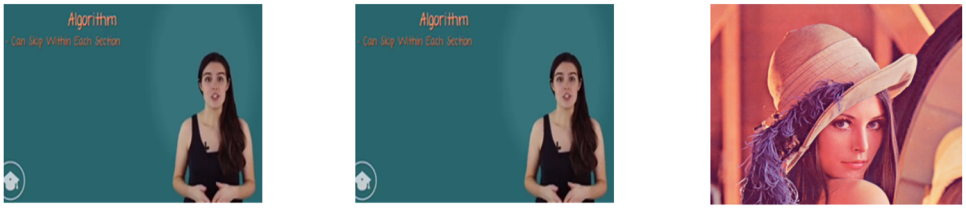


Figure 11: (a) Cover frame, (b) watermarked frame, and (c) extracted message.

Table 2: Experimental results

Cover videos	Test image (64 × 64) Lena			
	PSNR	MSE	SSIM	CC
Video 1	60.12	0.0014	0.97	0.9775
Video 2	63.07	0.0011	0.99	0.9989
Video 3	63.23	0.0011	0.99	0.9989

Moreover, no significant variation of the watermarked frames from the original frames is noticed. The similarity of the original message with the extracted message is justified by the value of SSIM (approaching "1") as shown in Table 2. The attainment values of PSNR, MSE, SSIM, and CC are listed in Table 2, and the values show that the method is effective in maintaining the robustness.

The results are validated and analyzed on specified parameters. From the PSNR and MSE graph in Figures 12 and 13, respectively, it can be concluded that embedding secret information in frames having a face is more robust than embedding in any random frame. Lower the value of the MSE matrix means less is the error between the original cover frame and the watermarked frame. Higher the value of PSNR means the visual quality of the watermarked frame is very good or we can say human eyes

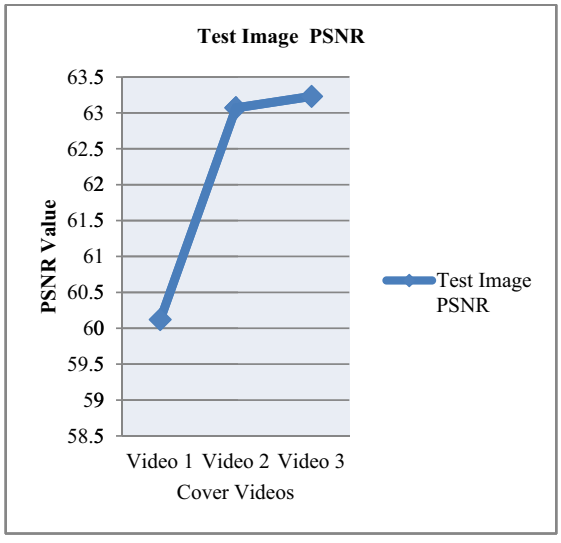


Figure 12: Peak signal to noise ratio (PSNR).

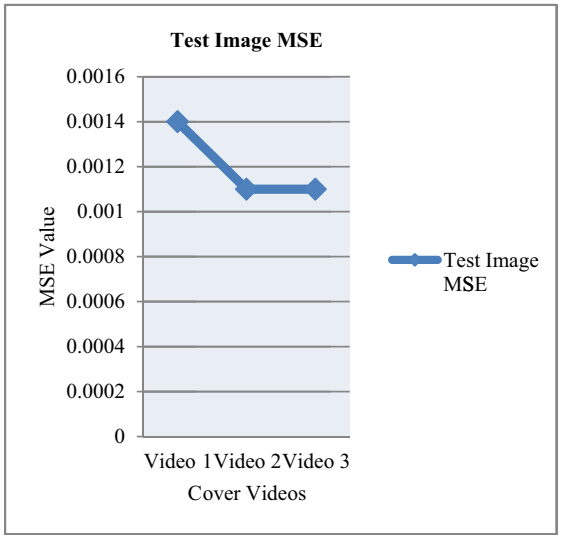


Figure 13: Mean square error (MSE).

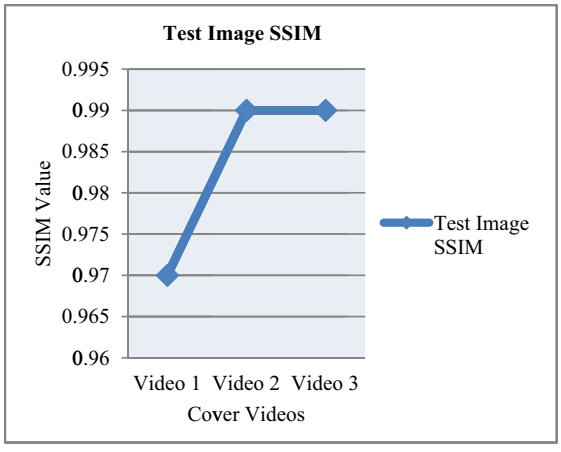


Figure 14: Structural similarity index (SSIM).

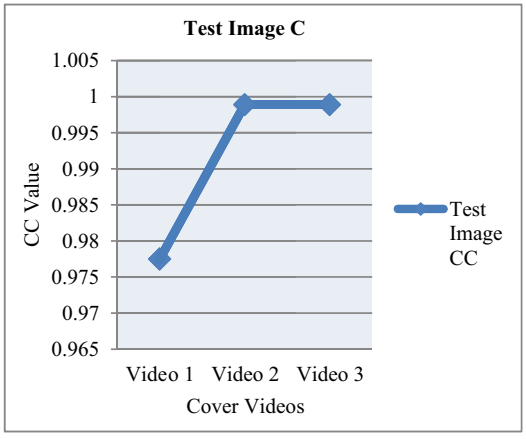


Figure 15: Correlation coefficient (CC).

are not able to distinguish between the original and watermarked frame. SSIM and CC matrixes graph in Figures 14 and 15 also have very good values almost equal to 1, which means watermark is highly imperceptible.

A histogram is a representation of information frequency in the form of graph. In general, histogram has *x*-axis and *y*-axis. In other words, histogram can be defined as the approximate distribution of image information.

With the help of the histogram deviation of the original watermark from the extracted watermark is tested in Figures 16–21. Results show that both the histograms are reasonably similar, which means that the extracted watermarked frame has a good visual quality.

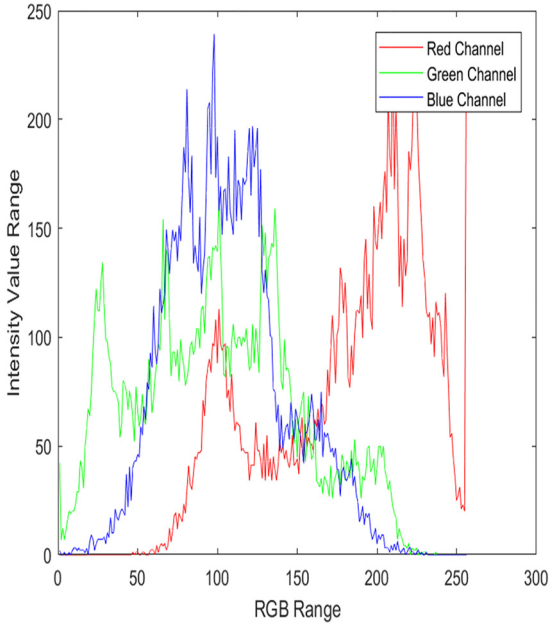


Figure 16: Histogram for original watermark in video 1.

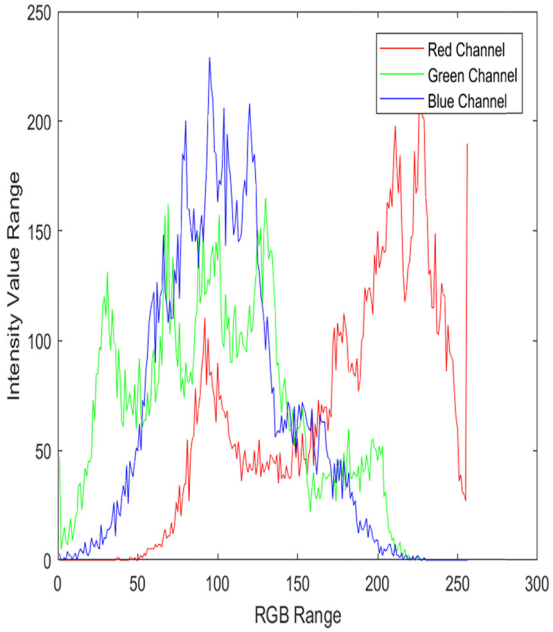


Figure 17: Histogram for extracted watermark in video 1.

6.1 Measures of robustness

Robustness is a measure of the ability of a watermark or stego file to withstand the various attacks, no matter they are intentional or unintentional. It is expected from a standard communication technique that uses secure data transmission that the receiver should be able to extract correct data (almost similar to the embedding) even after being attacked.

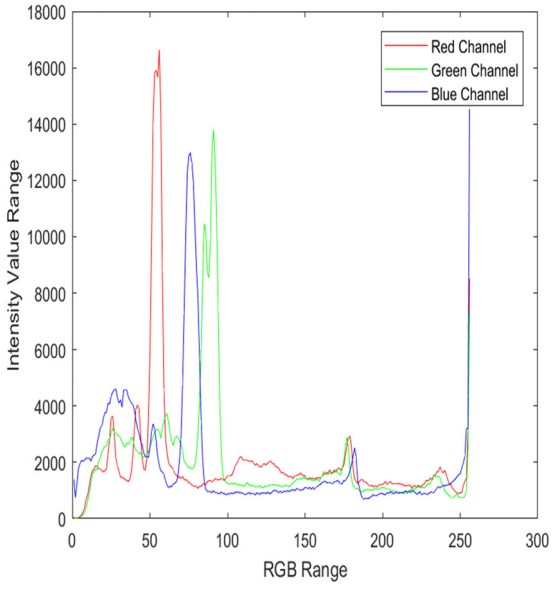


Figure 18: Histogram for original watermark in video 2.

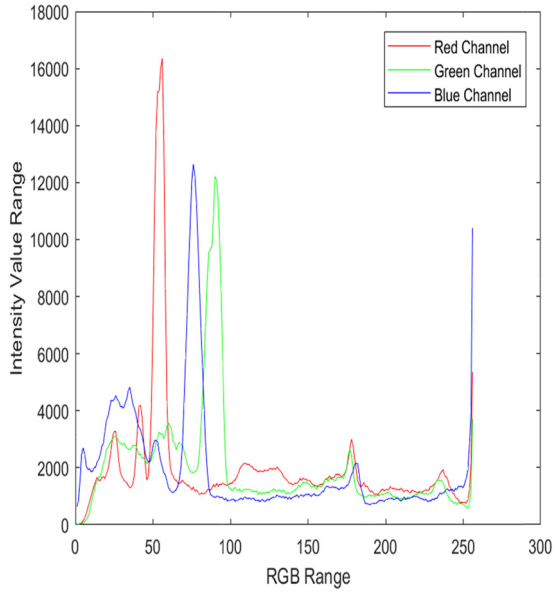


Figure 19: Histogram for extracted watermark in video 2.

There are various methods to estimate the element of robustness in a watermark out of which a few are identified as follows:

- Robustness estimation using PSNR and MSE.
- Robustness testing against well-known stego attacks.

In this section, the above two methods are explored to justify the robustness of the proposed method.

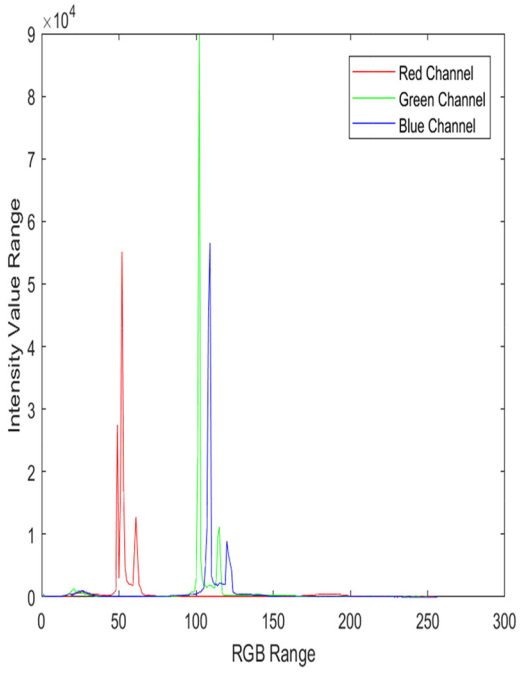


Figure 20: Histogram for original watermark in video 3.

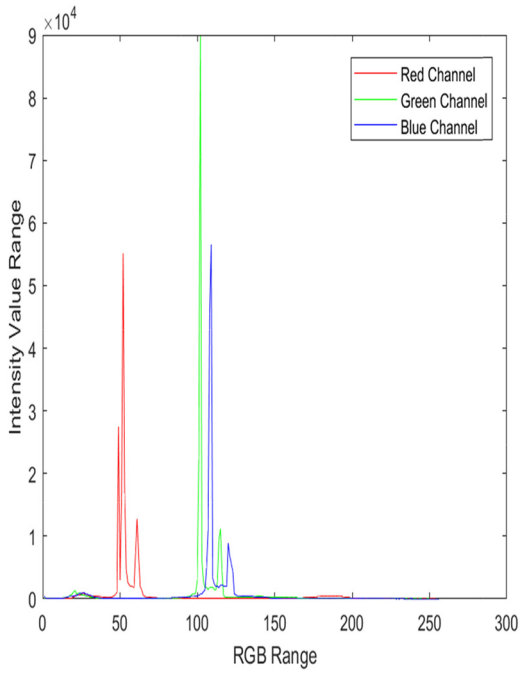


Figure 21: Histogram for extracted watermark in video 3.

6.1.1 Robustness using PSNR and MSE

PSNR is the most common parameter that measures the presence of distortion present in a video frame as compared to the original cover file content. Mathematically, it is measured using MSE. A lower value of MSE results in an increased value of PSNR, and that represents the lack of presence of noise or distortion concerning the original content.

It can be visualized from Table 2 that the proposed method achieves a PSNR value of more than 60 for different choices of cover file. The satisfactory attainment of PSNR supports the robustness of the proposed method in one aspect.

6.1.2 Robustness against compression

The compression is applied to reduce the size in bytes of the multimedia file without modifying the excellence of the file to an objectionable level. By applying compression, we can store more multimedia to the same disk space. Technically, it is achieved by compromising with statistical values of image or video parameters; the extent of that depending upon the type of compression (Lossless or Lossy). There are many ways through which compression can be done [35]. On the other side, it is a big threat to steganography and is capable of destroying the embedded data either knowingly or unknowingly. This led to excessive

Figure 22: (a) Watermarked image, (b) compressed watermarked image, (c) original secret message, and (d) recovered secret message from compressed watermark.

importance on testing the robustness of the stego file against compression. This is usually done by compressing the stego file using different compression methods and at different compression degrees [36]. Figure 22 shows the secret information retrieved after the application of a compression attack.

6.1.3 Robustness against geometric attack

The geometric attack includes operations that disturb the geometry of images, such as scaling, translation, flipping, rotation, cropping, and sharpening, may be detected. Geometric attacks are also able to identify the existence

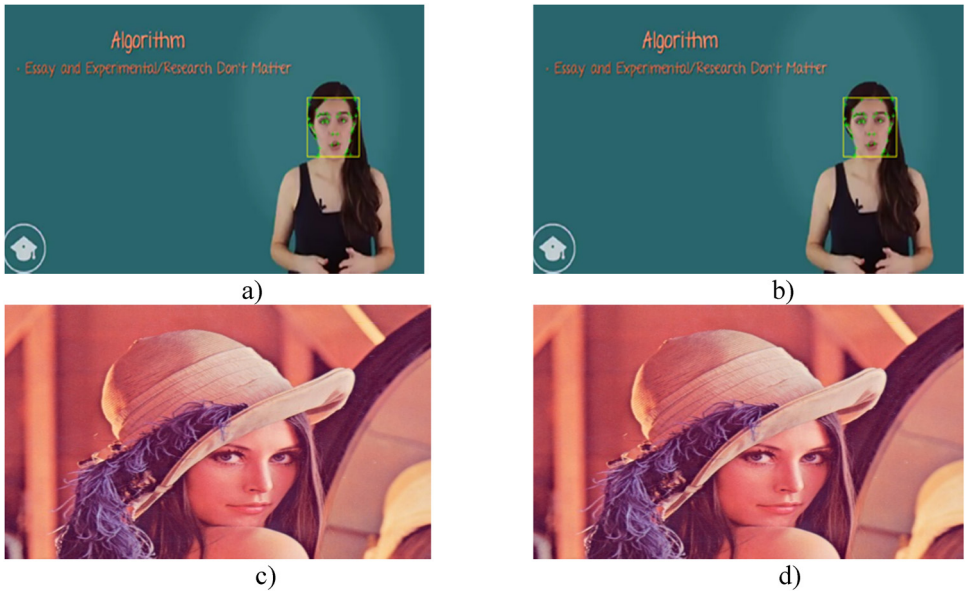


Figure 23: (a) Watermarked image, (b) watermarked image after attack, (c) original secret image, and (d) recovered secret image with sharping attack.

Figure 24: (a) Watermarked image (b) watermarked image with attack, (c) original secret image, and (d) recovered secret image with gamma correction.

of embedded secret messages within the stego image according to Madhavi et al. [2]. Sharpening attack has been considered here for analyzing the robustness. Figure 23 shows the impact of the specified geometric attack on the degree of successful extraction of the embedded data. The figure shows that the secret message was recovered successfully even after the attack.

6.1.4 Robustness against gamma correction attacks

Gamma correction defines a connection between a color value and its intensity on a specific device. For images labeled in an RGB color mode to appear visually accurate, the display device must produce an output intensity directly proportional to the input color value [22]. Any disturbance in that proportion may result in inadequate data extraction. It is shown in Figure 24 that the proposed scheme can withstand the gamma correction attacks and recovers the secret data as per expectation.

7 Conclusion

In this article, we have implemented a novel video steganography technique with enhanced capacity in the wavelet domain using face detection and SVD. It is easy to securely transmit a larger amount of data using this technique. ROI provides a suitable region to embed the secret image. Updated singular values of the watermark in the wavelet domain provided good robustness against regular image processing attacks. Haar-based lifting scheme results in receiving good reconstruction of watermark information, increasing smoothness, and decreasing aliasing effects. Because of split and merge procedure in lifting scheme calculation difficulty reduced by approximately 57%. The proposed technique provides high data hiding capacity along with robustness, imperceptibility of the watermark, low computational cost, and simplicity.

Conflict of interest: Authors state no conflict of interest.

References

[1] A. Kala and K. Thaiyalnayaki, "Robust lossless image watermarking in integer wavelet domain using SVD," *Int. J. Computer Sci. Eng.*, vol. 2, pp. 30–35, 2013.

[2] K. Madhavi, D. Kumar, and S. Mahitha, "A robust and efficient steganography using skin tone as biometric for real time images," *Int. Res. J. Eng. Technol.* (*IRJET*), vol. 5, pp. 3289–3292, 2018.

[3] S. Sharma and D. Somwanshi, "A. DWT, based attack resistant video steganography," *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, New York, NY, United States: Association for Computing Machinery, 2016, pp. 1–5.

[4] S. Hemalatha, and A. D. Shamathmika, *MP4 video steganography in wavelet domain*, IEEE, 2017, pp. 1229–1235.

[5] D. R. I. M. Setiadi, "Improved payload capacity in LSB image steganography uses dilated hybrid edge detection," *J. King Saud. Univ.-Computer Inf. Sci.*, vol. 34, pp. 1–11, 2019.
[6] S. Bhatnagar, S. Kumar, and A. Gupta, "An approach of efficient and resistive digital watermarking using SVD," *International Conference on Advances in Computing, Communications and Informatics* (*ICACCI*), 2014, pp. 2470–2475.
[7] A. Bhardwaj, D. Verma, and V. Verma, "A robust watermarking scheme using lifting wavelet transform and singular value decomposition," *AIP Conference Proceedings*, Vol. 1802, No. 1, AIP Publishing LLC, 2017, p. 020002.
[8] S. Hemalatha, A. Renuka, D. Acharya, and P. Kamath, "A secure image steganography technique using Integer Wavelet Transform," *World Congr. Inf. Commun. Technol.*, vol. 1, pp. 755–758, 2012.
[9] A. Sharp, Y. Yang, D. Peng, and H. Sharif, "A video steganography attack using multi-dimensional discrete spring transform," *IEEE International conference on Signal and Image Processing Applications*, 2013, pp. 182–186.
[10] S. Kejgir and M. Kokare, "Lifting wavelet transform with singular value decomposition for robust digital image watermarking," *Int. J. Computer Appl.*, vol. 39, pp. 10–18, 2012.
[11] P. Patel, and Y. Patel, "Secure and authentic DCT image steganography through DWT-SVD Based Digital Watermarking with RSA Encryption," *Fifth International Conference on Communication Systems and Network Technologies*, 2015, pp. 736–739.
[12] R. Harshitha and S. Vidya, "Robust and high limit watermarking using DWT-IWT," *Int. J. Advance Sci. Resour. Manag.*, vol. 2, pp. 18–21, 2017.
[13] A. Sharp, Q. Qi, Y. Ang, D. Peng, and H. Sharif, "A novel active warden steganographic attack for next-generation steganography," *9th International Wireless Communications and Mobile Computing Conference* (*IWCMC*), 2013, pp. 1138–1143.
[14] A. Hussain, "Multi-level Steganography System Using Wavelet Transform," *J. Eng. Sustain. Dev.*, vol. 2018, pp. 50–61, 2018.
[15] K. Vikram, and S. Padmavathi, "Facial parts detection using Viola Jones algorithm," *2017 4th International Conference on Advanced Computing and Communication Systems* (*ICACCS*), 2017, pp. 1–4.
[16] E. Ghasemi, J. Shanbehzadeh, and B. ZahirAzami, "A steganographic method based on Integer Wavelet Transform and Genetic Algorithm," *2011 International Conference on Communications and Signal Processing*, 2011, pp. 42–45.
[17] M. Kawulok, J. Kawulok, and J. Nalepa, "Spatial-based skin detection using discriminative skin-presence features," *Pattern Recognit. Lett.*, vol. 41, pp. 3–13, 2014.
[18] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "A skin tone detection algorithm for an adaptive approach to steganography," *Signal. Process.*, vol. 89, pp. 2465–2478, 2009.
[19] J. Viola and J. Paul Michael, "Robust real-time face detection," *Int. J. Computer Vis.*, vol. 57, no. 2. pp. 137–154, 2004.
[20] M. Wall, A. Rechtsteiner, and L. Rocha, "Singular value decomposition and principal component analysis," *A practical approach to microarray data analysis*, Boston, MA, Springer, 2003, pp. 91–109.
[21] S. Acharya, H. Kabra, P. Kasambe, and S. Rathod, "Performance evaluation of an integer wavelet transform for FPGA implementation," *International Conference on Nascent Technologies in the Engineering Field* (*ICNTE*), 2015, pp. 1–5.
[22] K. Raja, K. Venugopal, and L. Patnaik, "High capacity lossless secure image steganography using wavelets," *International Conference on Advanced Computing and Communications*, 2006, pp. 230–235.
[23] W. Jue, Z. Min-qing, and S. Juan-li, "Video steganography using motion vector components," *IEEE 3rd International Conference on Communication Software and Networks*, 2011, pp. 500–503.
[24] A. K. Sahu, G. Swain, and E. S. Babu, "Digital image steganography using bit flipping," *Cybern. Inf. Technol.*, vol. 18, pp. 69–80, 2018.
[25] A. A. Hanafy, G. I. Salama, and Y. Z. Mohasseb, "A secure covert communication model based on video steganography," *MILCOM 2008–2008 IEEE Military Communications Conference*, 2008, pp. 1–6.
[26] S. Khosla and P. Kaur, "Secure data hiding technique using video steganography and watermarking," *Int. J. Computer Appl.*, vol. 95, pp. 7–12, 2014.
[27] M. Grangetto, E. Magli, M. Martina, and G. Olmo, "Optimization and implementation of the integer wavelet transform for image coding," *IEEE Trans. image Process*, vol. 11, pp. 596–604, 2002.
[28] M. Dalal and M. Juneja, "A robust and imperceptible steganography technique for SD and HD videos," *Multimed. Tools Appl.*, vol. 78, pp. 5769–5789, 2019.
[29] R. J. Mstafa, K. Elleithy, and E. Abdelfattah, "A robust and secure video steganography method in DWT-DCT domains based on multiple object tracking and ECC," *IEEE Access*, vol. 5, pp. 5354–5365, 2017.
[30] P. K. Paramesh, N. Ranjitha, and S. Swetha, "Video Steganography using MATLAB," *EAI Endorsed Trans. Cloud Syst.*, vol. 310, 2017, p. e3.
[31] H. Shivaram, D. U. Acharya, and R. Adige, "Wavelet transform based steganography technique to hide audio signals in image," *Proc. Computer Sci.*, vol. 47, pp. 272–281, 2015.
[32] M. Sadek, A. Khalifa, and M. Mostafa, "Video steganography: a comprehensive review," *Multimed. tools Appl.*, vol. 74, pp. 7063–7094, 2015.
[33] V. Kumar, and D. Kumar, "Performance evaluation of DWT based image steganography," *IEEE 2nd International Advance Computing Conference* (*IACC*), 2010, pp. 223–228.
[34] K. Ng, S. Liew, and F. Ernawan, "An improved RDWT-based image steganography scheme with qr decomposition and double entropy," *IOP Conference Series: Materials Science and Engineering*, 2020, Doi: 10.1088/1757-899X/769/1/012069.
[35] S. Bhattacharyya and G. Sanyal, "Data hiding in images in discrete wavelet domain using PMM," *Int. J. Electr. Computer Eng.*, pp. 359–367, 2010.
[36] M. Barani, P. Ayubi, M. Valandar, and B. Irani, "A blind video watermarking algorithm robust to lossy video compression attacks based on generalized Newton complex map and contourlet transform," *Multimed. Tools Appl.*, vol. 79, pp. 2127–2159, 2020.