Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-18405-7

# ESKVS: efficient and secure approach for keyframes-based video summarization framework

**Parul Saini**[1] · **Krishan Berwal**[2]

Received: 5 July 2023 / Revised: 2 October 2023 / Accepted: 22 January 2024


**Abstract**
Video security has emerged as a critical study issue in multimedia security in recent years as videos are the most effective and widely used multimedia format. They immediately establish a connection with users. It is necessary to prevent this sensitive information from being stolen or destroyed in various domains, including the military, finance, and education. It can be achieved by either hiding its significance, turning it into a secret code through encryption, or doing both simultaneously. Because video data is generated, transmitting it securely is challenging, and there is resource wastage of memory, processing, and bandwidth. Users must spend a lot of time and effort scanning through this enormous amount of video information in search of the required information. Therefore, a secure keyframes-based generic video summarization (VS) model is proposed to generate a secure video summary. First, Secret Keyframes (SKs) are extracted from the video through proposed Probability VS (PBVS) and Extended (E-PBVS). Second, multi-secret image sharing is provided to the SKs by the proposed EBEMSS (Enhanced Blockwise Encryption based Multi Secret Sharing) scheme, which uses a polynomial congruence concept for keyframe security. The proposed model E-PBVS achieved an average F-score of 0.76 and 0.81 on two benchmark (OV and YT) datasets, respectively, showing its effectiveness in producing informative video summaries. Additionally, EBEMSS outperforms the other related security models. The proposed model outperforms with generated summary and security compared to other keyframe-based VS and MSS techniques.

**Keywords** Video security · Multi-secret sharing · Clustering · Encryption · Multimedia security · Secure image

Krishan Berwal contributed equally to this work

✉ Parul Saini
parulsaini.phd2020@nituk.ac.in

Krishan Berwal
k2b@ieee.org

[1] Department of Computer Science & Engg., NIT, Srinagar, Uttarakhand, India

[2] Military College of Telecommunication Engineering, Mhow, MP, India

Published online: 17 February 2024

## 1 Introduction

Images and videos are the most popular and successful multimedia formats because they directly connect with the users. With the development of high-speed Internet and low-cost storage, the amount of data has drastically expanded, with the bulk of multimedia arriving in the form of visual or video data. Protecting this digital content from unwanted access and alteration has become a severe issue in the digital era as internet development technology has been developing exceedingly rapidly and swiftly. Private information such as bank account numbers and medical images is distributed online at significant risk. They were designing quick and trustworthy security systems that can consume high bandwidth data. Images or videos could experience loss or format conversion, which is problematic. The problem of securely sharing sensitive information online has recently taken on critical importance. Protecting the privacy and secrecy of hidden photographs or videos is becoming difficult since they include crucial information [1, 2]. Video hosting, TV program hosting, social networking, and online news websites, including Wistia, SproutVideo, Youtube, Netflix, Amazon Prime, Twitter, LinkedIn, and Facebook, have massive saved video data.

YouTube alone produces more than 10 hours of video every second. Video requires more bandwidth and storage than text and images do. Also, watching such videos involves a lot of human resources. Videos must be obtained and shown more concisely and clearly before being used in various applications. Therefore, efficient methods and technologies are required [3]. For the users, sorting through the massive amount of captured video data to obtain the necessary information is getting more difficult and time-consuming. Utilizing a method to remove the essential frames from the vast quantity of video gathered is crucial to secure the recordings. The main objective of VS is to analyze the video while keeping the keyframes by deleting extraneous or redundant frames [4–7]. Moreover, it speeds up scrolling through a sizable collection of video data and enables organized access to and representation of the video content. Security techniques may be applied to these keyframes assumed to be pictures when the video summary of keyframes is formed [8]. These keyframes are the secret images that need to be secured. The secrecy of image information was previously protected using either steganography or cryptography. The secret image is either encrypted or concealed in a single file, leading to a single point of failure (SPOF). These also need help with key management and dependability. As a result, they cannot offer tolerance for content removal or modification [9–11].

The proposed algorithms by Shamir and Blakley [12, 13] have led to the development of Secret Image Sharing (SIS) schemes to solve the abovementioned issue. In SIS, confidential image information is divided into shadow images using encoding or encryption techniques, ensuring that no individual share can disclose the secret image. These shares are distributed to different parties at various locations through diverse channels, eliminating the Single Point of Failure (SPOF) issue. Secret Sharing Schemes can be utilized for sharing a single secret (SS) or multiple secrets (MS) and are often implemented with various techniques, such as steganography, visual cryptography, watermarking, discrete wavelet transform, threshold schemes, Deoxyribonucleic acid (DNA) Encoding, Chaotic maps, etc [1, 14–16]. A combination of VS and securing VS summary builds up the proposed Secure Technique for the Keyframes-based Video Summarization model (ESKVS). It attempts to give a compact and informative overview by selecting the most valuable video content segments as keyframes. It expedites the processing of videos and the effective and efficient management of videos. In this work, the generated summary is a static summary composed of a keyframe or video storyboard, a group, or a collection of frames [2].

The authors are motivated by the above techniques and decoded to propose PBVS and E-PBVS for keyframe selection based on the SSIM (Structural Similarity Index Metric) clustering and the probability. A frame is selected as a keyframe, and SKs are recovered by an unsupervised deep-learning algorithm. After extracting SKs, proposed EBEMSS, Enhanced BEMSS (Blockwise Encryption based Multi Secret Sharing scheme [16]), which is an (n,n) multi-secret image encryption with a secret sharing scheme using blockwise and polynomial Congruence encryption, is applied to SKs. The main contributions to this work are summarized as follows:

- A novel ESKVS model is proposed for extracting important frames from videos, and their secure secrets are shared with participants using the proposed (n,n) EBEMSS.
- The proposed PBVS extracts candidate frames from the original frames to avoid the redundancy of similar frames from the clusters made by K-means Clustering using the deep high-level features extracted by a deep learning model, whereas E-PBVS extracts winnowed frames through deep feature-based novel SSIM Clustering which uses frame-level processing to similar group frames to avoid the redundancy of frames.
- The proposed PBVS and E-PBVS consider a Frame Probability Score, which uses Euclidean distance between frames within the clusters and the entropy of a frame as a metric to generate the video summary.
- The novel (n,n) EBEMSS algorithm is proposed, which uses the polynomial congruence concept for encrypting the shares of the SKs.
- To our best knowledge, EBEMSS is the first attempt to use the concept of polynomial congruence in the MSS.

The rest of the paper is organized as Section 2 discusses the related work on VS and MSS schemes. The proposed ESKVS model combining proposed E-PBVS and EBEMSS algorithms for VS and MSS scheme is explained in Section 3. Section 4 discusses the details of the performance analysis for the proposed model. Section 5 discusses the experimental results and performance of the proposed framework. A discussion of the performance has been done in Section 6. The overall proposed framework is concluded with future work in Section 7.

## 2 Related work

Several high-quality approaches are urgently required for protecting valuable multimedia information against unauthorized access, illicit monitoring, or modification [17]. Here, VS is done to extract the essential SKs from the video. Many approaches such as feature-based (color, motion, gesture, audio-visual, speech, objects, etc.), clustering-based techniques (K-means clustering, partitioned clustering, spectral clustering), shots selection-based VS, event-based summarization, trajectory-based VS, etc. are used for VS. Since deep learning (DL) emerged, much research has been done on DL-based VS. Based on the DL algorithms; VS can be supervised or unsupervised [18, 19]. We have used an Unsupervised learning keyframe selection approach as there are not enough labeled standard datasets for VS. [20] developed a content-based adaptive clustering technique that uses video color, motion, form, and texture to generate video summaries. Krishan et al. [21] presented an Eratosthenes sieve-based keyframe extraction clustering procedure. Events are remodeled from the extracted keyframes by setting minimum and maximum frame numbers for the event boundaries. Generic Video SUMarization (GVSUM) [22] is proposed in which keyframes are extracted whenever there is a change in the cluster number of the frame.

In the GVSUM approach, K-means clustering is used to cluster the video frames, and whenever there is a change in the cluster number of the frame, a keyframe is selected. A memorability and entropy-based VS framework combined these scores to choose keyframes [24]. Cost-effective VS uses aesthetics features to generate video summaries using deep convolutional neural network (CNN) [25]. Muhammad et al. [26] suggest a surveillance VS architecture for devices with limited resources and lower computational complexity. DeepReS [27], a DL-based VS, is proposed for Resource-Constrained Industrial Surveillance environments and does the coarse and fine refining of video data to produce the summary. Equal frames partition-based VS [8] is proposed using an equal partition-based clustering technique where the entire video is clustered into keyframe groups in the first variation. However, the second variant divides the video into equal-sized frames and then clusters them into keyframes. ESVS (Eratosthenes Sieve-based VS) [3] has been suggested for VS to provide concise and intelligent video abstraction, commonly known as event summary.

The DPCA+HSV (Density Peak Clustering Algorithm + (Hue, Saturation, Value)) [28] approach uses the HSV histogram as the color features for each frame. The recurrent encoder/decoder and the frame selection Long Short-Term Memory (LSTM) are trained using a discrete similarity measure that is learned using the GAN's (Generative Adversarial Networks) discriminator in SUM-GANdpp (Determinantal Point Process) [29] in an unsupervised manner. Muhammad et al. [30] model is based on low-level frame color features mixed with a sub-shots detection strategy, which identifies both sharp and gradual video transitions equally well and is less susceptible to noise or flashes than previous approaches of a similar kind. In [31], a two-stream method that captures motion and visual features from the video is proposed for extracting keyframes using an unsupervised static VS technique. From different CNN layers, feature maps are obtained and fused. Using a window-based peak detection technique, the candidate keyframes are extracted. To get the final keyframes, redundant frames are removed by building a similarity network. This method performs better than other related unsupervised methods.

A different way of dividing a secret among a group of participants, each of whom allots a share of the secret, is called secret sharing. Individual shares are useless and must be merged to rebuild the secret. Secret-sharing involves a combination of information disclosed by each participant to decrypt the key. The secret could be a single or multiple images [1]. Many techniques have been proposed for video security based on cryptographic algorithms and steganography [35–38]. The GUESS(Genetic Uses in video Encryption with Secret Sharing)model [32] was proposed based on the Genetic algorithm and Secret sharing systems. The proposed solution improves the effectiveness of an encryption process by being faster and more precise. E-MOC [33] is an Efficient secret-sharing model for Multimedia On the Cloud, which demonstrates the role of dividing an image into different frames and encrypting every frame to protect user data. The (n, m, l)-MMSIS(Multilevel Multi-Secret Image Sharing) [34] method divides the' m' shares produced by the' n' separate secret pictures among the' m' participants assigned to the' l' distinct levels. The prime numbers are exposed for encryption to protect multimedia data. Prime numbers have the distinctive quality of being complicated to factor. Researchers have also employed prime numbers to safeguard user data [1, 15]. When using prime integers in cryptography, congruence is frequently utilized in place of equality [44]. Zn (Set of Residues) to Z (set of integers) mapping is not one-to-one; hence, any number of Z members can map to a single Zn member.

Although the congruence operator resembles the equality operator, there are some distinctions. The congruence operator maps a member from Z to a member of Zn, while an equality operator maps to itself first. Second, although congruence is mapped many to one, equality is mapped one to one. To encrypt a pixel value of an image, quadratic congruence

**Table 1** A comparative table of the existing VS and security models

| Model | Contributions | Limitations |
|---|---|---|
| Khurana et al. [31] | Keyframe selection based on window-based peak detection | Few extra keyframes |
| Basavarajaiah et al. [22] | Keyframe selection based on change in cluster number | Ignore important keyframes |
| Nair et al. [23] | MultiCNN features based | High complexity |
| Muhammad et al. [26] | Memorability and entropy-based | Low processing rate(18fps) |
| Archana and Malmurugan [39] | Edge detection and LSTM | Highly complex and high computation |
| Muhammad et al. [27] | Coarse and fine refining | Low detection |
| Muhammad et al. [30] | Shot identification | Low level features and low detection |
| Fei et al. [24] | Shot segmentation with importance scores | Redundant frames |
| John et al. [40] | Meaningful shares | Grayscale secrets only |
| Yadav and Singh [41] | Meaningful shares | Lossy quality and high complexity |
| Bhat et al. [14] | Polynomial based | High complexity |
| Bisht and Deshmukh [34] | Multilevel MSS | Specific order of shares need to be maintained at each level |
| Agarwal et al. [42] | Object detection based | Complexity increases with the increase in the number of objects |
| Koppanati et al. [33] | Logistic map based | Limited to single SSS |
| Khan et al. [43] | Chaotic based | Limited to single SSS |

$X^2 \equiv a(mod\ p)$ easily find solutions $X \equiv a^{\frac{p+1}{4}}(mod\ 4)\ and\ X \equiv -a^{\frac{p+1}{4}}(mod\ 4)$ where $a$ is an integer and $p$ is a prime number. However, when the power $n$ and $p$ can take user-defined values, finding the solution for the polynomial congruence $x^n \equiv a(mod\ p)$ becomes very difficult. The proposed model ESKVS follows an unsupervised approach based on SSIM-based clustering, a technique to form groups of a similar pattern of unlabeled data based on their similarities or differences. This work has incorporated the above concept in our proposed EBEMSS security approach.

The related domains of the current study identify some gaps in the VS and security models, as shown in Table 1. The current VS needs assistance with storage problems, frame selection redundancy, and practical information retrieval limitations. Performance could be better since current VS approaches must fully utilize the computing capability of modern GPUs. The lack of extensive datasets with ground truth annotations constrains using VS models. The most significant issue is current video processing methods' ineffectiveness in computation time, storage efficiency, complexity, low communication costs, Etc. Due to the absence of hybrid cryptographic techniques in current video protection, video data is exposed to security risks. These flaws demonstrate the need for novel approaches to improve VS effectiveness, take advantage of GPU capabilities, deal with dataset constraints, and bolster security in the context of video data. The proposed research attempts to build deep learning-based VS approaches and security systems based on these research gaps. Here, we put into practice a concept that offers security to summarized videos or SKs.

## 3 Proposed framework

By choosing the most informative sections of the video information as keyframes, video summarising approaches attempt to provide a brief and complete description. It speeds up the video processing and management of videos effectively and efficiently. The generated summary in this work is static, a group or collection of frames called keyframing or video storyboard [2]. An efficient ESKVS model is proposed to produce a secure static video summary, as shown in Fig. 1.

SKs are extracted by an unsupervised deep learning method using the proposed PBVS and E-PBVS key frame selection algorithm based on the SSIM clustering, probability of a frame being a keyframe, and information present in the frame. After extracting SKs, the proposed EBEMSS algorithm, which is an (n,n) multi-secret image encryption based on BEMSS (blockwise and polynomial congruence encryption [16]), is applied to SKs. The proposed approaches for VS are discussed in Section 3.1, and the proposed security algorithm for SKs is discussed in Section 3.2.

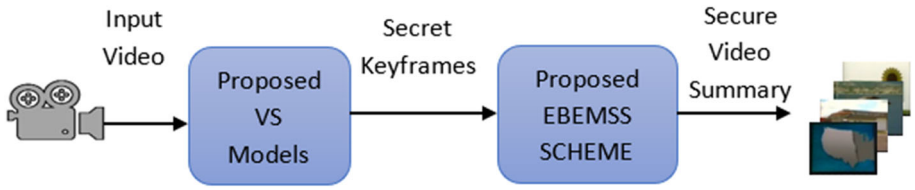


**Fig. 1** Proposed ESKVS model

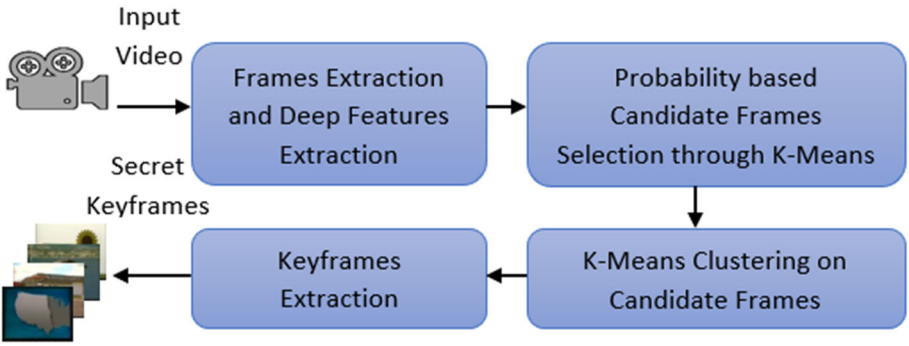


**Fig. 2** Proposed PBVS

### 3.1 Video summarization

VS produces the semantic visual summary of a video. The created static summary contains a portion of the original video frames. The proposed PBVS and E-PBVS method is explained in Figs. 2 and 3. For both the proposed models, an input video is initially divided into colored frames at its original frame rate. In this approach, no pre-sampling is done to maintain the originality of the video, nor have we converted the colored frames into grayscale ones. Learned features have been extracted from the pre-trained Visual Geometry Group (VGG 19) with 16 convolution layers, 3 fully connected layers, 5 MaxPool layers, and 1 SoftMax layer [45].

**Proposed PBVS** In this approach, K-means clustering, an unsupervised iterative learning algorithm that divides frames into clusters that share similarities and are different from the frames belonging to another cluster using the Euclidean distance(ED) as a measure of similarity, is used. The silhouette analysis [22] is used to choose an ideal value for the cluster. Candidate frames are generated, using Algorithm 1, from the clusters following the steps given below-

- K-means clustering is done based on deep features of all frames
- Compute entropy(E) and Euclidean distance(ED) of the frames within each cluster.
- Compute Frame Probability Score (FPS), based on E and ED, of each frame being a keyframe is calculated.

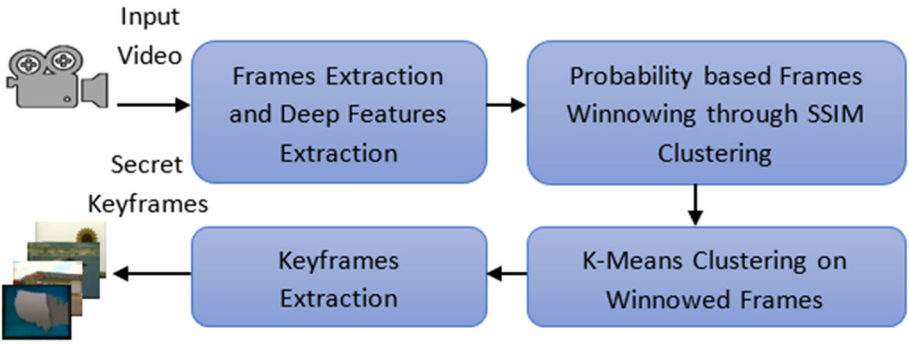


**Fig. 3** Proposed E-PBVS

- FPS is compared with the optimum threshold of each frame being a keyframe.
- Frames with higher probability are selected as Candidate frames.
- K-means Clustering on Candidate Frames.
- Centered frames are selected as Keyframes

SKs are generated from the candidate frames after applying K-means clustering. Here, centroids of the clusters, formed from the candidate frames, are selected as keyframes.

**Algorithm 1** Candidate frames selection.

```
1. Input: k clusters
2. Output: Candidate frames
3. For i=1 to k
4. For each frame in a cluster
5. Compute E and ED
6. Compute FPS based on E and ED
7. Compute threshold
8. if (Prob_frame > Threshold)
   select a frame as a candidate frame
```

**Proposed E-PBVS** : PBVS approach, as explained, works well and is suited for short-duration videos. For long-duration videos, it shows some redundancy in keyframes. For example, a long-duration video sample ( v26 - 3.29 minutes ) has 6270 extracted frames. So, when the number of frames increases, K-means clustering may generate clusters with varying densities and sizes. Also, the same cluster may have redundant frames from different video events, which increases the redundancy in the candidate frames and final extracted keyframes. In this video sample, we get 5413 candidate frames from 6270 extracted frames, which shows a significant amount of redundancy and some redundant frames in the final video summary as shown in Fig. 4.

Also, the video summary length is 13. In contrast, the ground truth summary length is 24 since, for long-duration videos, the optimum number of clusters computed does not match the summary length of the ground truth summary. Hence, after applying the proposed PBVS process, extracted keyframes show redundancy. So, E-PBVS is utilized based on the proposed PBVS and uses deep features based SSIM clustering [1] on all the video frames,

**Fig. 4** Video summary with redundant frames of a long duration video sample

which measures the similarity or dissimilarity between the two frames based on brightness, contrast, and structure present in it. E-PBVS generates the video summary following the steps below-

- SSIM clustering is done based on deep features of all frames using Algorithm 2.
- Compute the E Score of all the frames.
- Compute FPS based on E of each frame within a cluster is a keyframe is calculated.
- FPS is compared with the optimum threshold of each frame being a keyframe.
- Frames with higher probability are selected as Winnowed frames.
- K-means Clustering on Winnowed Frames.
- Again, Compute FPS based on E and ED
- Frames with high FPS are selected as Keyframes

**Algorithm 2** SSIM clustering.

1. Input: n frames
2. Output: k clusters
3. For i=1 to n
4. For each frame in a cluster
5. Compute SSIM
   For each frame F(i) and F(i+1 )
   If SSIM(i) and SSIM (i+1) matched
   Consider F(i) and F(i+1 ) in same cluster
   else
   Consider F(i) and F(i+1 ) in different cluster

As discussed above, the Proposed EPBVS selects the keyframes based on a high probability score. Figure 5 shows the sample video frames with their probability score. It is clear from the figure that the frame with a high probability score is selected as the keyframe.

### 3.2 Encryption based multi secret sharing scheme

The proposed encryption-based (n,n) MSS scheme is an enhanced BEMSS based on the BEMSS scheme [16]. EBEMSS uses a polynomial congruence approach for the encryption of shares, which enhances and increases the level of security of BEMSS [16], which is used for the security of multi-secret images. The proposed EBEMSS takes the n secret images and generates n encrypted shares using encryption, which are then distributed to n participants or parties. All the n shares are combined to retrieve the n secret images. The detailed block diagram of the proposed EBEMSS using image encryption and decryption is illustrated in Figs. 6 and 7. The proposed algorithm follows the steps of BEMSS [16]. The workflow of proposed EBEMSS is as follows:


**Fig. 5** Probability Score of frames of a sample video

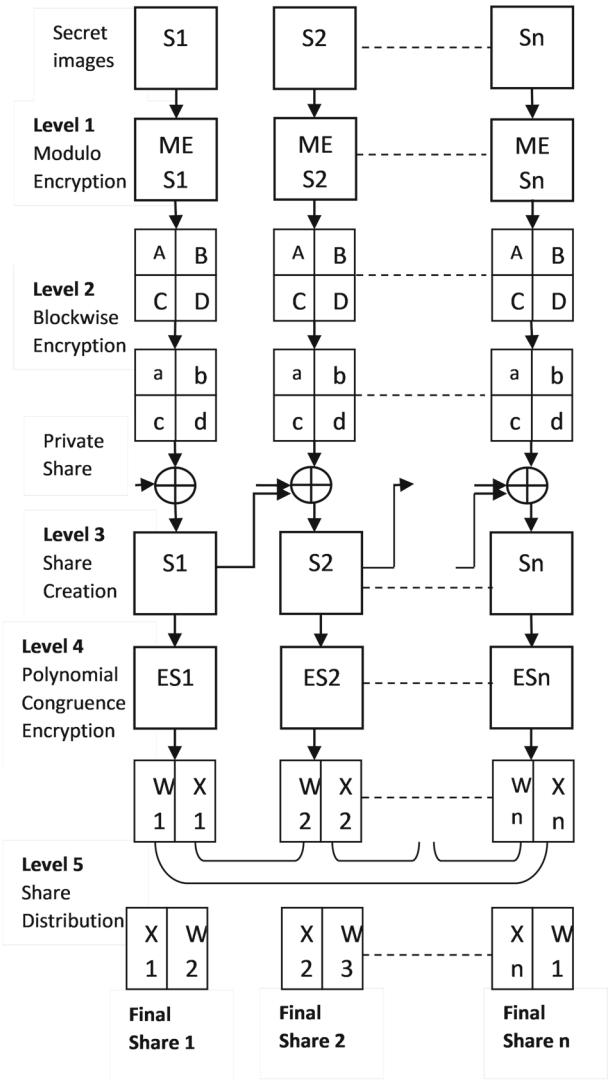


**Fig. 6** Image encryption using EBEMSS scheme

***Image Encryption:***

1. Input: Secret images (keyframes)
2. Modulo Encryption:
   - Generate random value (RMVal in pixel value range(0-255)).
   - Modulo operation on each pixel for encryption.
   - Create Encrypted Secret Image.
3. Blockwise Encryption (BE):
   - Divide the encrypted image into four blocks.
   - Encrypt remaining blocks using Bitwise XOR and block swap.

- Generate Blockwise Encrypted Images.

4. Share Creation:
   - Create temporary shares from BE images.
   - Utilize private share (PS) and XOR for share creation.
5. Polynomial Congruence Encryption
   - Temporary Shares created to follow Algorithm 3 for encryption.
6. Shares Distribution:
   - Combine pairs of temporary encrypted shares for final shares.

***Image Decryption:***

1. Shares Collection:
   - Gather n shares from participants.
   - Split and join shares to create encrypted temporary shares.
2. Polynomial Congruence Decryption:
   - Encrypted temporary Shares follow Algorithm 4 for decryption.
3. Shares Generation:
   - Utilize private share (PS) and XOR for blockwise decrypted share creation.
4. Blockwise Decryption:
   - Separate decrypted shares into four blocks.
   - Utilize XOR and block swapping to recover original blocks.
5. Modulo Decryption:
   - Use inverse modulo encryption for modulo decryption.
   - Decrypt each pixel of images on each plane
6. Output:
   - Lossless Secret images are reconstructed.

Proposed EBEMSS differs from BEMSS in the share distribution shares step, where another layer of encryption is added based on the polynomial congruence. The polynomial congruence approach is a standard power congruence (higher degree) in the form of where '$p$' is prime and exists in the form of

$$x^n \equiv a(mod\ p) \tag{1}$$

Where '$a$' is a positive integer, and exact *gcd(p-1,n)* solutions exist if the below condition is true where *gcd* is the greatest common divisor.

$$a^{((p-1)/gcd(p-1,n))} \equiv 1 mod p \tag{2}$$

To use the above equation for encrypting the pixel value $x$, EBEMSS follows Algorithm 3 and Algorithm 4. Algorithm 3 explains the encryption process, which provides another layer of security in the proposed EBEMSS. Here, encryption is done on the shares, which will

be distributed among the participants for MSS using the polynomial congruence concept. Suppose any pixel value of $x$ of the shared image is 234. Select $p$ as 7643 and $n$ as 849, using $a \equiv x^n (mod\ p)$, we get the encrypted value of $x(a) = 45$. The decryption key is calculated using Algorithm 4, and $x$ is retrieved using $x \equiv a^k (mod\ p)$. In this example, the key is calculated as 7633, which is used in $x \equiv a^k (mod\ p)$ to compute $x$ as 234 to its original value.

## 4 Performance analysis

The performance of the proposed framework is analyzed through VS analysis and MSS analysis, as discussed in Sections 4.1 and 4.2.

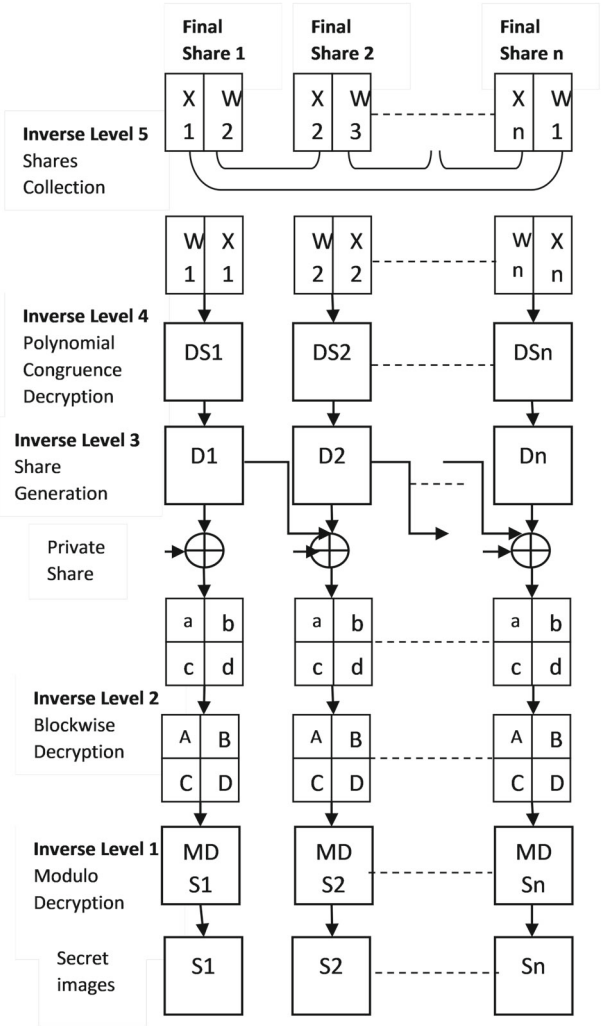


**Fig. 7** Image decryption using EBEMSS scheme

**Algorithm 3** Encryption algorithm.

1. For $i$ =1 to $n$ share images
2. For every pixel of the image, shares
3. Choose $p$ and $n$ such that the solution exists for the above polynomial congruence equation.
4. If not, then choose another value for $p$ and $n$.
5. If solution exists encrypt the x pixel value using
   $a \equiv x^n (mod\ p)$
   to get the encrypted value $a$.
6. If $a$ is out of the pixel value range, it is normalized in the required range.
7. end For

**Algorithm 4** Decryption algorithm.

1. For $i$ =1 to $n$ final share images
2. For every pixel of the final shared image
3. Calculate the original $a$ if it is normalized.
4. Calculate the key to decrypt the $a$ using $p$ and $n$
5. *key = gcd(n, p-1)[1] mod (p-1)*
6. Decrypt the value of a using $x \equiv a^k (mod\ p)$ to get the original value $x$.
7. end For

### 4.1 VS analysis

It is a challenging task to evaluate a video summary as it is different for different users. It depends on the individual interest and opinion to analyze the summary of the video. For qualitative analysis, keyframes are analyzed visually and compared with the ground truth summaries. Three assessment metrics are used for quantitative analysis- Precision ($P$), Recall ($R$), and *F-measure*. *F- measure* is used as a quantitative measure for evaluating the quality of the VS method [2]. *F-measure* is defined as the harmonic mean of precision and recall and calculated as in (3).

$$F - Measure = \frac{2 \times P \times R}{(P + R)} \tag{3}$$

where $P$ is Precision and $R$ is Recall and are calculated as in (4).

$$P = \frac{A \cap B}{B}, R = \frac{A \cap B}{A} \tag{4}$$

Where $A$ is the number of frames in the ground-truth or reference summary and $B$ is the number of frames in the proposed summary. The maximum value of the F measure indicates a more accurate approach. While calculating the precision and recall, we need to compare the keyframes obtained from the proposed summary and the reference summary with which we compare our results.

### 4.2 MSS scheme analysis

The performance of the proposed EBEMSS algorithm is analyzed using histogram, differential, statistical, and computation time analysis.

**Histogram Analysis** is the graphical representation of the distribution of the pixel values of a color image for each component (R, G, and B). On excellent encryption, the distribution of the pixels of an image must be uniform [1, 2].

**Differential Analysis** investigates the relationship between the original and encrypted image when a slight modification is made.

**Statistical Attack** takes advantage of statistical flaws in a proposed algorithm to evaluate its efficiency and security.

**Computation Time (CT) Analysis** For an algorithm to be efficient, it should utilize minimum resources and computation time. Hence, computation time is observed and measured in seconds and compared with other approaches.

The different metrics used in the evaluation techniques are described as follows:

*Differential Analysis* Unified average changing intensity (*UACI*), number of pixel change rate (*NPCR*), and histogram analysis are done to evaluate the differential attacks. [1].

*Unified Averaging Changing Intensity (UACI)* is the average intensity of divergence with a one-pixel between the encrypted and plain image. It is mathematically defined as in (5).

$$UACI = \frac{\sum_{p,q} S(p,q) - S'(p,q)}{255 \times WT_i \times HT_i} \times 100 \tag{5}$$

where *S(p, q)* and *S' (p, q)* represent the encrypted image and modified image, and the width and height of the images are represented by $WT_i$ and $HT_i$, respectively.

*Number of Pixel Change Rate (NPCR)* compares the original and the encrypted image's pixel values. For positive analysis, its value should be more than 99% [1] and is defined as in (6).

$$NPCR = \frac{\sum_{p,q} S(p,q)}{WT_i \times HT_i} \times 100 \tag{6}$$

$$S(p,q) = \begin{cases} 0 \ if \ S(p,q) = S'(p,q) \\ 1 \ if \ S(p,q) \neq S'(p,q) \end{cases}$$

where *B(l,m)* and *B' (l,m)* denote the difference between the pixels of the original and encrypted image.NPCR range is *[0,100]*. For ideal encryption, the rate of NPCR must be close to 100.

*Statistical Analysis* includes Correlation Analysis (*CA*), the Peak Signal To Noise Ratio (PSNR), Structural Similarity Index Metric (*SSIM*), and Information entropy (*IE*) analysis.

*Correlation Coefficient (CC)* detect the similarity among the related pixels of the original and encrypted image [1]. The range of values is -1.0 to 1.0. Its range is -1.0 to 1.0. *CC* among samples p and q, containing n values, is mathematically defined as in (7).

$$CC(p,q) = \frac{\sum_{a=1}^{n}(p_a - \bar{p})(q_a - \bar{q})}{\sqrt{\sum_{a=1}^{n}(p_a - \bar{p})^2}\sqrt{\sum_{a=1}^{n}(q_a - \bar{q})^2}} \tag{7}$$

*Peak Signal To Noise Ratio (PSNR) in decibels* is the ratio of the strength and the noise in the signal. Higher *PSNR* indicates higher image quality [1] and is defined as in (8).

$$PSNR = 10\text{Log}_{10}\frac{Max^2}{MSE} \tag{8}$$

$$MSE = \frac{\sum_{P,Q}[IM_1(p,q) - IM_2(P,Q)]^2}{P \times Q}$$

*MSE* is the mean squared error, *P*, *Q* is the size of image matrix $IM_1$ and $IM_2$, and *Max* is the maximum pixel value for the 8-bit image is 255.

*Structural Similarity Index Metric (SSIM)* measures the similarity or dissimilarity between the two images based on brightness (l), contrast (c), and structure (s) [1] and is

calculated as in (9). Its ranges is 0 to 1 where 0 means dissimilar and 1 means similar images.

$$SSIM(IM1, IM2) = \frac{(2\mu_{IM1}\mu_{IM2} + C1)(2\sigma_{IM1IM2} + C2)}{(\mu_{IM1}^2 + \mu_{IM2}^2 + C1)(\sigma_{IM1}^2 + \sigma_{IM2}^2 + C2)} \tag{9}$$

Where $\mu_{IM1}$, $\mu_{IM2}$, $\sigma_{IM1}$,$\sigma_{IM2}$, and $\sigma_{IM1}\sigma_{IM2}$ are the local means, standard deviations, and cross-covariance for images *IM1* and *IM2*.

*Information Entropy* measures the degree of randomness in an image [1] and is mathematically defined as in (10).

$$E(S) = -\sum_{0}^{255} PB(a_i) \log_2 PB(a_i) \tag{10}$$

where $a_i$ is a discrete random variable, PB($a_i$) is the probability density function of occurrence of $a_i$. The ideal value of information entropy is 8.

## 5 Experimental results

All experiments were carried out in Google Colab with 166.8 GB of disc storage, an Intel Xeon 2.20GHz processor, and a T4 GPU running at 51 GB of RAM. The proposed ESKVS model is examined in this Section using the performance measures given in Section 4.

### 5.1 Video summary

Different techniques for video summarising assess their experimental findings using various datasets. Many did not disclose the implementation specifics to other researchers or make the datasets publicly accessible. Thus, it is difficult to compare a large number of VS approaches. Considering this, videos for performing experiments for proposed PBVS are selected from a benchmark VSUMM(Video Summarization) dataset [46]. There are 50 videos each from Open Video(OV) and Youtube (YT) datasets. Every video is in color, has sound, and is in MPEG-1 format (30 frames per second, 352 by 240 pixels). These videos span various genres (documentary, instructive, ephemeral, historical, lecture), range from one to four minutes, and total about 75 minutes. 250 user summaries are also manually created by 50 individuals, each of whom worked on five videos, for a total of five video summaries for every single video. Since the selected datasets contain five user summaries for each video, it makes quantitative evaluation and Comparison straightforward.

**Qualitative analysis** Ground truth summaries generated by five users for the 50 videos of each dataset are available or accessible in the dataset itself. The proposed PBVS and E-PBVS method is compared with user-generated summaries. The standard summaries of Delaunay Clustering DT [47], STIMO (still and moving video storyboard ) [48], OV(open video) Summary [49], VSUMM1 and VSUMM2 [46] algorithms for OV Dataset [46]. For the YT dataset, they are compared with user-generated summaries and the standard summaries of VSUMM [46], Fie et al. [24], Muhammad et al. [26], and DeepReS [27]. Figures 8 and 9 show that the summary produced by our proposed algorithms PBVS and E-PBVS are closer to user summaries and contain salient frames than others for OV and YT datasets. Also, they select the optimum number of keyframes compared to other approaches, and their summary is very similar to the ground truth summaries.

**Quantitative Analysis** For the quantitative analysis of VS, there are no standard methods. Keyframes extracted are matched with the ground truth summaries generated by users


**Fig. 8** Sample video results of Proposed PBVS and E-PBVS in Comparison with five user ground truth summaries, DT, STIMO, OV, VSUMM1, and VSUMM2 algorithms for OV Dataset

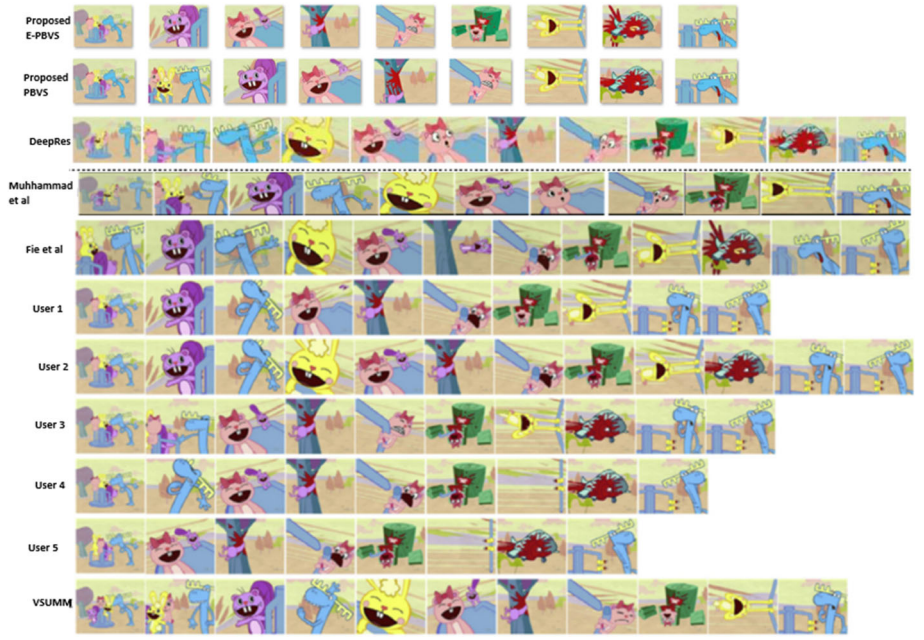


**Fig. 9** Sample video results of Proposed PBVS and E-PBVS in Comparison with five user ground truth summaries, VSUMM, Fie, et al., Muhhammad et al., and DeepReS algorithms for YT Dataset

**Table 2** Comparison of proposed PBVS and E-PBVS with different approaches on YT dataset

| Video | VSUMM [46] | Fie et al [24] | Muhammad et al [25] | Muhammad et al [26] | DeepReS [27] | Proposed PBVS | Proposed E-PBVS |
|---|---|---|---|---|---|---|---|
| Cartoons v11 | 68 | 82 | 80 | 82 | 86 | 86 | 87 |
| Cartoons v12 | 65 | 67 | 80 | 73 | 66 | 70 | 73 |
| News v88 | 67 | 75 | 71 | 78 | 83 | 80 | 82 |
| Home v108 | 52 | 73 | 83 | 73 | 82 | 80 | 81 |
| Average | 63 | 74 | 74 | 76 | 79 | 79 | 81 |

**Table 3** Comparison of different approaches on YT dataset with user summaries

| Video | Best | | | | | | Worst | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | [24] | [25] | [26] | [27] | Proposed PBVS | Proposed E-PBVS | [24] | [25] | [26] | [27] | Proposed PBVS | Proposed E-PBVS |
| Cartoons v11 | 77 | 80 | 77 | 86 | 88 | 94 | 67 | 50 | 67 | 81 | 82 | 84 |
| Cartoons v12 | 76 | 80 | 76 | 69 | 73 | 76 | 65 | 58 | 65 | 53 | 63 | 66 |
| News v88 | 84 | 71 | 84 | 90 | 89 | 87 | 68 | 44 | 68 | 77 | 73 | 67 |
| Home v108 | 79 | 83 | 79 | 92 | 85 | 86 | 57 | 50 | 57 | 72 | 71 | 76 |
| Average | 79 | 79 | 79 | 84 | 84 | 86 | 64 | 51 | 64 | 70 | 72 | 73 |

**Table 4** Comparison of proposed PBVS and E-PBVS with different approaches on OV dataset

| Algorithm | Precision | Recall | F-Measure % |
|---|---|---|---|
| DT [47] | 47.0 | 50.0 | 48.5 |
| STIMO [48] | 39.0 | 65.0 | 48.8 |
| OV [49] | – | – | 44.0 |
| VSUMM1 [46] | 42.0 | 77.0 | 54.4 |
| VSUMM2 [46] | 48.0 | 63.0 | 54.5 |
| EVS [8] | 70.9 | 59.6 | 64.8 |
| ESVS [3] | 69.4 | 61.8 | 65.4 |
| DPCA + HSV [28] | 66.6 | 60.6 | 63.4 |
| SUM-GANdpp [29] | – | – | 72.0 |
| Jin. et al [50] | 51.0 | 87.0 | 62.6 |
| Muhammad. et al [30] | – | – | 67.0 |
| Proposed PBVS | 75.2 | 71.7 | 73.1 |
| Proposed E-PBVS | 78.4 | 73.8 | 76.0 |

available in the VSUMM dataset. The proposed PBVS and E-PBVS performance evaluation uses the F-Measure metric, as discussed in Section 4.1. To investigate the effectiveness of the proposed algorithms, we compare the proposed PBVS and E-PBVS results with other keyframe selection-based VS approaches, including DT [47], STIMO [48], OV Summary [49], VSUMM1, VSUMM2 [46], EVS [8], ESVS [3], DPCA+HSV [28], SUM-GANdpp [29], Jin. et al [50], Khurana et al [31] and Muhammad et al. [30]. For the YT dataset, they are compared with user-generated summaries and the standard summaries of VSUMM [46], Fie et al. [24], Muhhammad et al. [25], Muhhammad et al. [26], and DeepReS [27] for videos in different genres. From Table 2, it is clear that the proposed algorithms obtain 79 and 81 F-Measure and outperform the other existing techniques. Also, Table 3 compares our approaches with user-generated summaries and shows the effectiveness of our system as a high average F-Measure in both the best and worst scenarios. It is evident from Tables 4 and 5 show that our VS approaches obtain higher F-measure scores of 73.1 and 76.0 for the OV dataset and F-measure scores of 76.3 and 81.2 as compared to the other existing methods.

**Table 5** Comparison of proposed E-PBVS with different approaches on YT dataset

| Algorithm | Precision | Recall | F-Measure % |
|---|---|---|---|
| VSUMM1 [46] | 38.0 | 72.0 | 49.7 |
| VSUMM2 [46] | 44.0 | 54.0 | 48.5 |
| AVS [8] | 55.6 | 49.4 | 52.3 |
| EVS [8] | 53.0 | 49.7 | 51.3 |
| ESVS [3] | 58.5 | 50.0 | 53.9 |
| DPCA + HSV [28] | 74.4 | 64.0 | 68.8 |
| SUM-GANdpp [29] | – | – | 60.1 |
| Jin et al. [50] | 42.0 | 74.0 | 50.2 |
| Khurana et al. [31] | – | – | 62.2 |
| Proposed PBVS | 80.2 | 72.7 | 76.3 |
| Proposed E-PBVS | 85.5 | 77.3 | 81.2 |

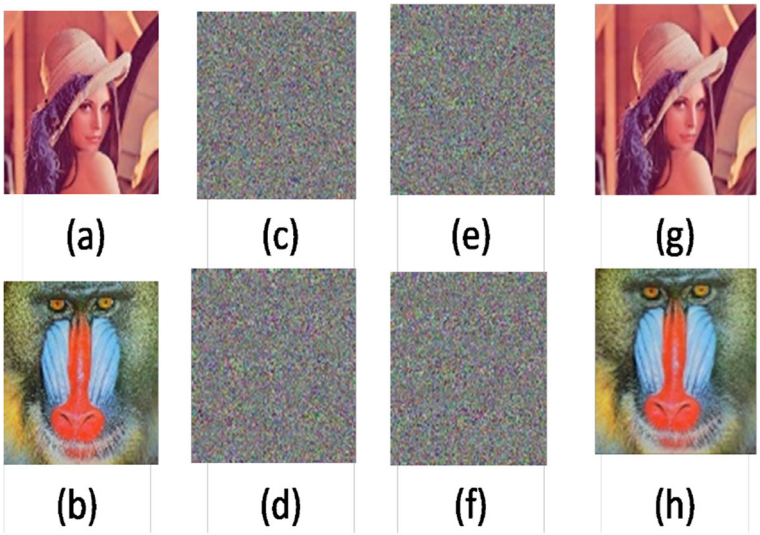


**Fig. 10** Proposed EBEMSS (2,2) : (a,b) Secret Test Images (Lena and Baboon, (c,d) Shares images, (e,f) Encrypted Shares S1 and S2 and (g,h) are lossless recovered Secret images

### 5.2 Encryption based multi secret sharing scheme

For our experiments, we have taken the test images (n=2) as the“ Lena” image (RGB) and “Baboon” image (RGB) of dimensions 512 x 512 before applying the proposed EBEMSS to the SKs generated from the proposed E-PBVS. After experimenting with the proposed EBEMSS on test images, we have implemented the proposed EBEMSS on the SKs extracted from the video and compared it with the BEMSS [16] and other techniques. Sample SKs (n=9) are extracted from the fifth sample video of the VSUMM dataset. Figures 10 and 11 represent the original secret test images and SKs with corresponding encrypted, shared, and lossless recovered images. Figures 12 and 13 show the histograms for the secret test images and SKs and their shares. It shows the distribution of the intensities.

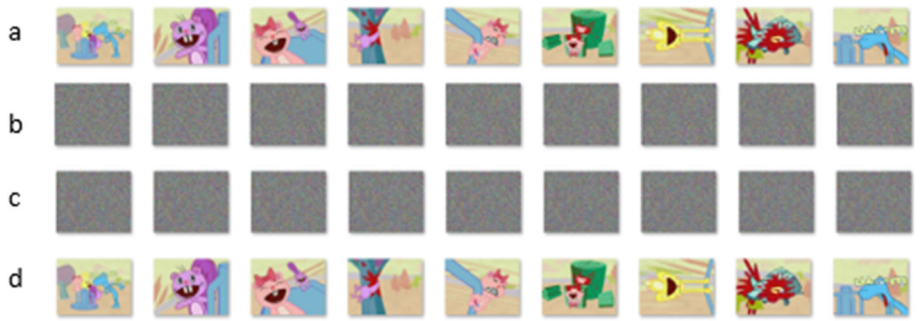


**Fig. 11** Proposed EBEMSS (9,9) : (a) SKs, (b) SKs shares, (c) Encrypted SKs shares, and (d) are lossless recovered SKs

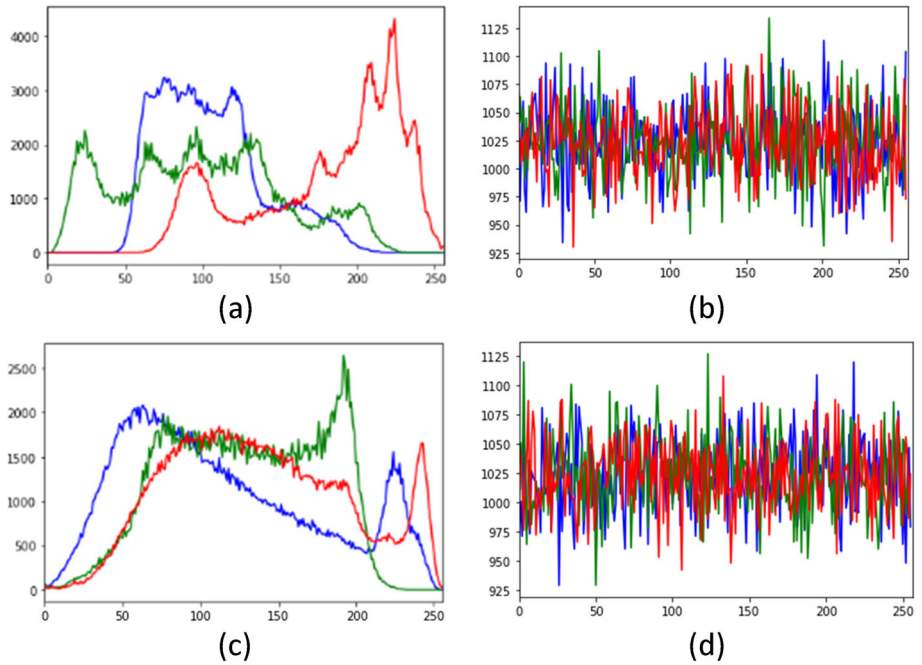


**Fig. 12** Histogram Analysis for Secret Test Images: (a-b) Lena(Original1 and Share1), (c-d) Baboon (Original1 and Share1)

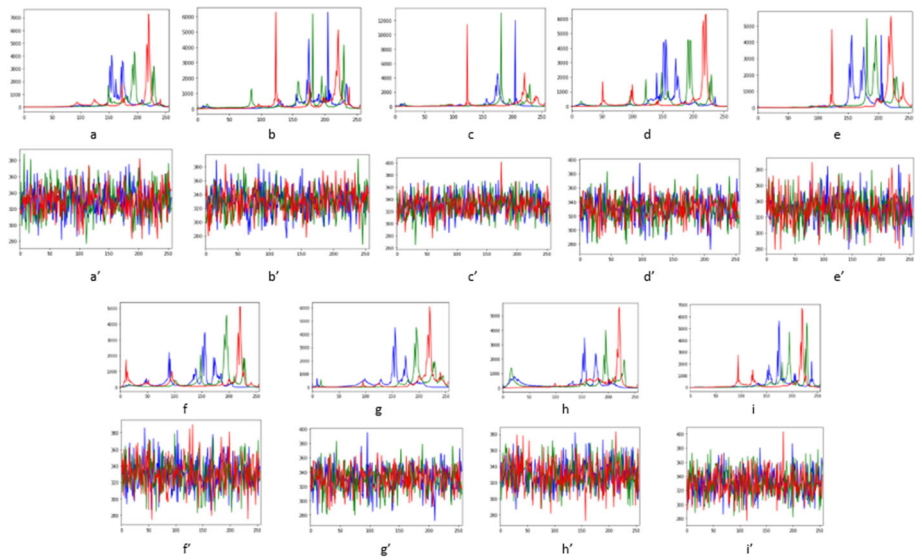


**Fig. 13** Histogram Analysis for SKs: (a-i) Original SKs, (a'-i') Shares of SKs

**Table 6** PSNR, CC, and SSIM values for SKs shares and recovered SKs

| SKs | PSNR | | CC | | SSIM | |
|---|---|---|---|---|---|---|
| | BEMSS [16] | EBEMSS | BEMSS [16] | EBEMSS | BEMSS [16] | EBEMSS |
| k1,Rk1 | 100 | 100 | 0.00056 | 0.000012 | 1 | 1 |
| k2,Rk2 | 100 | 100 | 0.00013 | −0.000003 | 1 | 1 |
| k3,Rk3 | 100 | 100 | 0.00067 | 0.000089 | 1 | 1 |
| k4,Rk4 | 100 | 100 | 0.00021 | 0.000051 | 1 | 1 |
| k5,Rk5 | 100 | 100 | 0.00030 | −0.000021 | 1 | 1 |
| k6,Rk6 | 100 | 100 | 0.00003 | −0.000091 | 1 | 1 |
| k7,Rk7 | 100 | 100 | 0.00023 | 0.0000019 | 1 | 1 |
| k8,Rk8 | 100 | 100 | 0.00014 | 0.000028 | 1 | 1 |
| k9,Rk9 | 100 | 100 | −0.00045 | 0.000075 | 1 | 1 |

**Table 7** PSNR, CC, and SSIM values for the test images

| Images | PSNR | | CC | | SSIM | |
|---|---|---|---|---|---|---|
| | BEMSS [16] | EBEMSS | BEMSS [16] | EBEMSS | BEMSS [16] | EBEMSS |
| I1,SI | 27.91 | 26.83 | −0.00030 | −0.0000011 | 0.0015 | 0.00016 |
| I2,S2 | 27.90 | 27.01 | −0.00011 | −0.0000003 | 0.0045 | 0.00023 |
| I1,E1 | 27.91 | 27.11 | 0.00034 | 0.000021 | 0.0212 | 0.000011 |
| I2,E2 | 27.89 | 26.12 | 0.00032 | −0.00003 | 0.0112 | 0.000009 |
| S1,S2 | 27.89 | 25.03 | 0.00047 | 0.0000021 | 0.0023 | 0.00004 |
| I1,R1 | 100 | 100 | 0.00086 | 0.0000024 | 1 | 1 |
| I2,R2 | 100 | 100 | 0.00049 | 0.0000008 | 1 | 1 |

**Table 8** PSNR, CC, and SSIM Values for SKs of sample video and their shares

| SKs | PSNR | | CC | | SSIM | |
|---|---|---|---|---|---|---|
| | BEMSS [16] | EBEMSS | BEMSS [16] | EBEMSS | BEMSS [16] | EBEMSS |
| k1,Sk1 | 27.78 | 27.11 | −0.00013 | 0.000034 | 0.0088 | 0.00044 |
| k2,Sk2 | 27.99 | 27.67 | −0.00009 | −0.000002 | 0.0083 | 0.000008 |
| k3,Sk3 | 27.76 | 26.55 | 0.00010 | −0.000003 | 0.0097 | 0.000047 |
| k4,Sk4 | 27.88 | 27.14 | 0.00073 | 0.0000056 | 0.0098 | 0.000089 |
| k5,Sk5 | 27.84 | 26.89 | −0.0004 | −0.0000023 | 0.0097 | 0.000004 |
| k6,Sk6 | 27.85 | 26.67 | −0.00087 | 0.0000032 | 0.0094 | 0.00065 |
| k7,Sk7 | 27.90 | 27.12 | −0.0008 | 0.0000029 | 0.0095 | 0.000063 |
| k8,Sk8 | 27.85 | 27.32 | 0.00028 | 0.0000067 | 0.0093 | 0.00009 |
| k9,Sk9 | 27.79 | 26.65 | 0.00023 | 0.0000045 | 0.0007 | 0.000012 |

**Table 9** UACI and NPCR values for SKs of sample video and their shares

| Keyframes | | k1 | k2 | k3 | k4 | k5 | k6 | k7 | k8 | k9 |
|---|---|---|---|---|---|---|---|---|---|---|
| UACI | BEMSS [16] | 33.17 | 33.24 | 33.04 | 33.24 | 31.23 | 32.08 | 34.05 | 33.08 | 32.60 |
| | EBEMSS | 33.89 | 33.04 | 34.12 | 34.26 | 33.87 | 33.32 | 35.41 | 33.74 | 31.12 |
| NPCR | BEMSS [16] | 99.23 | 99.12 | 99.89 | 99.15 | 99.78 | 99.03 | 99.08 | 99.58 | 99.43 |
| | EBEMSS | 99.27 | 99.42 | 99.87 | 99.78 | 99.65 | 99.90 | 99.74 | 99.12 | 99.71 |

**Table 10** UACI and NPCR values for the test images and shares

| Test images | I1 | | I2 | |
|---|---|---|---|---|
| | BEMSS [16] | EBEMSS | BEMSS [16] | EBEMSS |
| UACI | 33.21 | 35.43 | 31.56 | 36.21 |
| NPCR | 99.60 | 99.84 | 99.83 | 99.87 |

The EBEMSS is evaluated using differential and statistical analysis. Tables 6, 7, and 8 show the PSNR, CC, and SSIM values for secret test images and keyframes. Tables 9 and 10 shows UACI and NPCR values among the secret test images, SKs, and their shares. Tables 11 and 12 show the IE values for secret test images and keyframes. For an algorithm to be efficient, it should utilize minimum resources and computation time. Our proposed algorithm based on polynomial congruence in EBEMSS is compared with Anees et al. [54], Ahmed et al. [55], Ahmed et al. [56], and Khan et al. [57]. Also, the proposed EBEMSS is compared with other MSS schemes. Tables 13 and 14 clearly show that the proposed EBEMSS has less computational time of 0.49 and 0.258 seconds than others. Table 15 shows the comparative analysis of the proposed method with similar MSS based on recovered image quality, strategy, use of pixel expansion, type of SSS, complexity, color depth ( B-Binary, G-Grayscale, and C-Color), size of the shares (E-Equal, D- Different) and presence of cover images.

## 6 Discussion

In this Section, we interpret and analyze our results shown in Section 5. The following observations are made from the experiments.

- Tables 2–5 clearly show that the proposed PBVS and E-PBVS outperform the other approaches in terms of F-measure in both the OV and YT datasets. Proposed PBVS and EPBVS achieve the average F-measure of 73.12 and 76.06 on the OV dataset and 79 and 81, respectively, on the YT dataset. Results show the proposed approaches' effectiveness and efficiency compared to recent techniques.
- The histogram of the shared images in Figs. 12 and 13 has a uniform distribution. It does not share any statistical similarities with the histogram of the secret images, which shows significant sharp rises and sharp declines.
- Tables 9 and 10 observed that the value of UACI is around 33%, and NPCR value is above 99%, which shows that the shared secret images or keyframes are highly secure against differential attacks. The proposed technique is fit for an excellent encryption technique for MSS.
- Tables 6, 7, and 8 show that the PSNR value between secret test images or keyframes and their shares and encrypted images is low, which indicates that the shares generated

**Table 11** IE values for SKs and their encrypted shares

| Keyframes | k1 | k2 | k3 | k4 | k5 | k6 | k7 | k8 | k9 |
|---|---|---|---|---|---|---|---|---|---|
| Original | 5.555 | 5.787 | 6.576 | 7.680 | 6.939 | 7.239 | 6.811 | 7.296 | 7.609 |
| Shares | 7.995 | 7.985 | 7.997 | 7.980 | 7.986 | 7.996 | 7.995 | 7.988 | 7.996 |
| Recovered | 5.555 | 5.787 | 6.576 | 7.680 | 6.939 | 7.239 | 6.811 | 7.296 | 7.609 |

**Table 12** IE values for the secret test images and share images

| Test Images | I1 | | I2 | |
|---|---|---|---|---|
| | BEMSS A | EBEMSS | BEMSS A | EBEMSS |
| Original | 7.240 | 7.240 | 7.705 | 7.705 |
| Shares | 7.967 | 7.989 | 7.968 | 7.959 |
| Recovered | 7.240 | 7.240 | 7.705 | 7.705 |

are randomized well and have a higher error rate between them. PSNR value is 100 between the secret test images or SKs and the recovered images, which shows the lossless secret recovery in the proposed EBEMSS. CC values are also near 0, which indicates the dissimilarity between the secrets and shares. It also shows SSIM values are near 0, meaning the two respective images differ in pixel intensity. SSIM value for the secret and recovered images is 1, which means lossless recovery of the secret images and keyframes.

- Table 15 shows the proposed technique provides efficient and effective security using XOR, Modulo, and Polynomial Congruence with no pixel expansion and complexity of $O(n \log^2 n)$.
- Tables 11 and 12 show that the IE is nearly 8 for the secret images and keyframes and their shares, which show the excellent randomness property and highly secure approach.
- Table 13 shows the computation time of the proposed polynomial congruence-based encryption algorithm with other recent techniques, which concludes that the proposed approach takes less computation time than others. Table 14 shows that the proposed EBEMSS takes less computation time than other recent MSS approaches.

Existing security techniques for the video are designed for all the video frames, which increases the complexity and computation. Hence, it validates the need for VS before applying security to the secret data. In this research, we formulated our proposed model ESKVS to solve the given problem. From the experiments and observations, Our model ESKVS, which uses Key frame selection-based VS approaches, has proven effective in creating concise and representative summaries of videos. Also, it uses E-BEMSS for providing security to the significant content rather than the whole chunk of content for easy transmission and less computation overhead. However, like any other approach, they come with certain limitations. Our proposed VS models do not incorporate user preferences or feedback. PBVS model does not work well for long videos compared to E-PBVS, but still, the performance of E-PBVS can be improved by using other or multi-pretrained models. This model does not focus on the audio information, so videos with significant audio-based content, such as interviews or music performances, may not be accurately summarized. Moreover, if security is concerned, scalability (Increase in the number of secrets or participants) and verifiability (Honest Participants) issues can affect the efficiency and performance of the secret-sharing system, especially

**Table 13** CT (seconds) of Proposed polynomial congruence based EBEMSS and its comparisons

| Image | [54] | [55] | [56] | [57] | EBEMSS (Polynomial Congruence) |
|---|---|---|---|---|---|
| Lena | 11.42 | 3.23 | 2.25 | 2.14 | 0.49 |
| Baboon | 11.45 | 3.53 | 2.55 | – | 0.58 |

**Table 14** CT (seconds) per frame of Proposed EBEMSS and other MSS schemes

| MSS Schemes | BEMSS [16] | [34] | EBEMSS |
|---|---|---|---|
| Time(sec) | 0.556 | 0.374 | 0.258 |

in large-scale applications. Also, the security of polynomial congruence-based schemes is highly dependent on the proper selection of coefficients in the polynomial, which might introduce vulnerabilities. Despite these limitations, our proposed model ESKVS remains a valuable technique for providing security to the significant video content (keyframes) only rather than the whole chunk of video content for easy transmission and less computation overhead

## 7 Conclusion

This work proposed an efficient, Secure Technique for the Keyframes-based Video Summarization model (ESKVS) to produce a secure static video summary. Instead of considering the whole video, the most informative keyframes are selected, which speeds up the video processing and management of the videos effectively and efficiently. The proposed key frame selection algorithms (PBVS and E-PBVS) are based on the probability that a frame would be a keyframe and the information present in the frame. SSIM clustering extracts secret key frames using an unsupervised deep learning technique. The EBEMSS (n,n) scheme is proposed for securing the SKs using blockwise and polynomial congruence encryption. We evaluated the proposed PBVS and E-PBVS through F-Measure and EBEMSS through differential, statistical analysis, and computation time. The results demonstrate the feasibility of the proposed PBVS, EPBVS, and EBEMSS scheme, which outperforms all other recent related techniques. The scheme is also appropriate and effective since sharing capacity is maximal, and share sizes are similar to secret images. The proposed framework ESKVS can be used in real-time applications such as surveillance and security, medical imaging, news and journalism, military and defense, e-learning, legal and compliance, and many more where systems produce enormous amounts of video data. Costs associated with storage and transmission can be minimized with secure summarization. It has the potential to dramatically increase the handling and analysis of video data in various fields, resulting in cost savings, quicker decision-making, and improved user experiences. Future research suggests an application-based safe VS model that might incorporate steganography, sequence learning, audio characteristics, and hybrid cryptography to increase the security of private and enlightening video summaries.

**Acknowledgements** The authors thank the DST GoI for sponsoring this work under DST/ICPS/General/2018.

**Author Contributions** All authors have contributed equally to this work.

**Data availability and access** NA

## Declarations

**Ethical and informed consent** NA

**Table 15** Comparative analysis of the proposed EBEMSS with similar MSS

| Schemes | Recovered image Quality/ Strategy | Pixel expansion | Type | Complexity | Color depth | Size of the shares | Cover images |
|---|---|---|---|---|---|---|---|
| Wang et al. [51] | Loseless/ XOR | Yes | (n,n) | O(n) | B,G | E | NO |
| Feng et al. [52] | Recognition/ XOR | No | (n,n) | O(n) | B | E | NO |
| Faraoun et al. [53] | Loseless/ XOR | No | (n,n) | O(n) | B,G,C | E/D | NO |
| Bhatt et al. [14] | Loseless/ Polynimial Extended Visual Cryptography | Yes | (t,n) | $O(n \log^2 n)$ | B,G,C | E/D | YES |
| Chattopadhyay et al. [2] | Loseless/ XOR,Hash | No | (n,n) | $\theta(n)$ | B,G,C | E | NO |
| BEMSS [16] | Loseless/ XOR, Modulo | No | (n,n) | O(n) | B,G,C | E | NO |
| Proposed EBEMSS | Loseless/ XOR,Modulo Polynomial Congruence | No | (n,n) | $O(n \log^2 n)$ | B,G,C | E | NO |

## References


1. Sajitha AS, Rekh AS (2022) Review on various image encryption schemes. Materials Today: Proceedings 58:529–534
2. Saini P, Kumar K, Kashid S, Saini A, Negi A (2023) Video summarization using deep learning techniques: a detailed analysis and investigation. Artif Intell Rev 1–39
3. Kumar K, Shrimankar DD, Singh N (2018) Eratosthenes sieve based key-frame extraction technique for event summarization in videos. Multimed Tools Appl 77:7383–7404
4. Meena P, Kumar H, Yadav SK (2023) A review on video summarization techniques. Eng Appl Artif Intell 118:105667
5. Bozkurt F, Köse C, Sarı A (2018) An inverse approach for automatic segmentation of carotid and vertebral arteries in CTA. Expert Syst Appl 93:358–375
6. Huang C, Wang H (2019) A novel key-frames selection framework for comprehensive video summarization. IEEE Trans Circ Syst Vid Tech 30(2):577–589
7. Kashid S, Awasthi LK, Kumar K, Saini P (2023) NS4: a Novel Security approach for extracted video keyframes using Secret Sharing Scheme. 2023 International conference on computer, electronics and electrical engineering and their applications (IC2E3), Srinagar Garhwal, India, pp 1–6
8. Kumar K, Shrimankar DD, Singh N (2016) Equal partition-based clustering approach for event summarization in videos. In 2016 12th International conference on signal-image technology & internet-based systems (SITIS) pp 119–126. IEEE
9. Saini P, Kumar K, Kashid S, Negi A (2022) MEVSS: Modulo Encryption Based Visual Secret Sharing Scheme for Securing Visual Content. In International conference on deep learning, artificial intelligence and robotics pp 24–35. Cham Springer International Publishing
10. Kashid S, Kumar K, Saini P, Dhiman A, Negi A (2022) Bi-RNN and Bi-LSTM Based Text Classification for Amazon Reviews. In International conference on deep learning, artificial intelligence and robotics pp 62–72. Cham Springer International Publishing
11. Kashid S, Kumar K, Saini P, Negi A, Saini A (2022) Approach of a multilevel secret sharing scheme for extracted text data. In 2022 IEEE students conference on engineering and systems (SCES) pp 1–5. IEEE
12. Shamir A (1979) How to share a secret. Commun ACM 22(11):612–613
13. Blakley GR (1979) Safeguarding cryptographic keys. In Managing requirements knowledge, international workshop on pp 313–313. IEEE Computer Society
14. Bhat K, Reddy KRUK, Kumar HSR, Mahto D (2021) A novel scheme for lossless authenticated multiple secret images sharing using polynomials and extended visual cryptography. IET Inf Secur 15(1):13–22
15. Belazi A, Talha M, Kharbech S, Xiang W (2019) Novel medical image encryption scheme based on chaos and DNA encoding. IEEE access 7:36667–36681
16. Saini P, Kumar K, Kashid S, Dhiman A, Negi A (2022) BEMSS-Blockwise Encryption based Multi Secret Sharing scheme for Securing Visual Content. In 2022 IEEE 9th Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON) pp1–6
17. Negi A, Kumar K, Saini P (2023) Object of Interest and Unsupervised Learning-based Framework for an Effective Video Summarization Using Deep Learning. IETE J Res 1–12
18. Negi A, Kumar K, Chauhan P, Saini P, Kashid S (2022) Resource Utilization Tracking for Fine-Tuning Based Event Detection and Summarization Over Cloud. In International conference on deep learning, artificial intelligence and robotics pp 73–83. Cham Springer International Publishing
19. Negi A, Kumar K, Saini P, Kashid S (2022) Object detection based approach for an efficient video summarization with system statistics over cloud. In 2022 IEEE 9th Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON) pp 1–6. IEEE
20. Sun X, Kankanhalli MS (2000) Video summarization using R-sequences. Real-Time Imaging 6(6):449–459
21. Kumar K, Shrimankar DD, Singh N (2018) Eratosthenes sieve based key-frame extraction technique for event summarization in videos. Multimed Tools App 77:7383–7404
22. Basavarajaiah M, Sharma P (2021) GVSUM: generic video summarization using deep visual features. Multimed Tools App 80:14459–14476
23. Nair MS, Mohan J (2021) Static video summarization using multi-CNN with sparse autoencoder and random forest classifier. SIViP 15:735–742
24. Fei M, Jiang W, Mao W (2017) Memorable and rich video summarization. J Vis Commun Image Represent 42:207–217
25. Muhammad K, Hussain T, Tanveer M, Sannino G, de Albuquerque VHC (2019) Cost-effective video summarization using deep CNN with hierarchical weighted fusion for IoT surveillance networks. IEEE Intern Things J 7(5):4455–4463

26. Muhammad K, Hussain T, Baik SW (2020) Efficient CNN based summarization of surveillance videos for resource-constrained devices. Pattern Recogn Lett 130:370–375
27. Muhammad K, Hussain T, Del Ser J, Palade V, De Albuquerque VHC (2019) DeepReS: A deep learning-based video summarization strategy for resource-constrained industrial surveillance scenarios. IEEE Trans Ind Informat 16(9):5938–5947
28. Zhao H, Wang WJ, Wang T, Chang ZB, Zeng XY (2019) Key-frame extraction based on HSV histogram and adaptive clustering. Math Probl Eng 2019:1–10
29. Mahasseni B, Lam M, Todorovic S (2017) Unsupervised video summarization with adversarial lstm networks. In Proceedings of the IEEE conference on computer vision and pattern recognition pp 202–211
30. Asim M, Almaadeed N, Al-Máadeed S, Bouridane A, Beghdadi A (2018) A key frame based video summarization using color features. In 2018 Colour and visual computing symposium (CVCS) pp 1–6. IEEE
31. Khurana K, Deshpande U (2023) Two stream multi-layer convolutional network for keyframe-based video summarization. Multimed Tools App 1–42
32. Sharma S, Kumar K (2018) Guess: genetic uses in video encryption with secret sharing. In Proceedings of 2nd international conference on computer vision and image processing: CVIP 2017, vol 1. Springer Singapore, pp 51–62
33. Koppanati RK, Kumar K, Qamar S (2021) E-MOC: an efficient secret sharing model for multimedia on cloud. In Conference Proceedings of ICDLAIR2019 pp 246–260. Springer International Publishing
34. Bisht K, Deshmukh M (2021) A novel approach for multilevel multi-secret image sharing scheme. J Supercomput 77(10):12157–12191
35. Himeur Y, Boukabou A (2018) A robust and secure key-frames based video watermarking system using chaotic encryption. Multimed Tools App 77:8603–8627
36. Shyaa GS, Al-Zubaidie M (2023) Utilizing Trusted Lightweight Ciphers to Support Electronic-Commerce Transaction Cryptography. App Sci 13(12):7085
37. Rao V, KV P (2021) DEC-LADE: Dual elliptic curve-based lightweight authentication and data encryption scheme for resource constrained smart devices. IET Wirel Sensor Syst 11(2):91–109
38. Muhajjar RA, Flayh NA, Al-Zubaidie M (2023) A perfect security key management method for hierarchical wireless sensor networks in medical environments. Electronics 12(4):1011
39. Archana N, Malmurugan N (2020) Multi-edge optimized LSTM RNN for video summarization. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-020-02025-8
40. John Blesswin A, Selva Mary G, Manoj Kumar S (2022) Multiple secret image communication using visual cryptography. Wirel Pers Commun 122(4):3085–3103
41. Yadav M, Singh R (2022) Essential secret image sharing approach with same size of meaningful shares. Multimed Tools App 81(16):22677–22694
42. Agarwal A, Deshmukh M, Singh M (2020) Object detection framework to generate secret shares. Multimed Tools App 79(33):24685–24706
43. Khan M, Masood F, Alghafis A, Amin M, Batool Naqvi SI (2019) A novel image encryption technique using hybrid method of discrete dynamical chaotic maps and Brownian motion. PLoS One 14(12):e0225031
44. Forouzan BA, Mukhopadhyay D (2011) Cryptography and Network Security (Sie). McGraw-Hill Education
45. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
46. De Avila SEF, Lopes APB, da Luz Jr A, de Albuquerque Araújo A (2011) VSUMM: a mechanism designed to produce static video summaries and a novel evaluation method. Pattern Recogn Lett 32(1):56–68
47. Mundur P, Rao Y, Yesha Y (2006) Keyframe-based video summarization using delaunay clustering. Int J Digit Libr 6:219–232
48. Furini M, Geraci F, Montangero M, Pellegrini M (2010) STIMO: STIll and MOving video storyboard for the web scenario. Multimed Tools App 46:47–69
49. DeMenthon D, Kobla V, Doermann D (1998) Video summarization by curve simplification. In Proceedings of the sixth ACM international conference on multimedia pp 211–218
50. Jin H, Yu Y, Li Y, Xiao Z (2022) Network video summarization based on key frame extraction via superpixel segmentation. Trans Emerg Telecommun Tech 33(6):e3940
51. Wang D, Zhang L, Ma N, Li X (2007) Two secret sharing schemes based on Boolean operations. Pattern Recogn 40(10):2776–2785
52. Feng JB, Wu HC et al (2008) Visual secret sharing for multiple secrets. Pattern Recogn 41(12):3572–3581
53. Faraoun KM (2017) Design of a new efficient and secure multi-secret images sharing scheme. Multimed Tools App 76(5):6247–6261

54. Anees A, Siddiqui AM, Ahmed F (2014) Chaotic substitution for highly autocorrelated data in encryption algorithm. Commun Nonlinear Sci Numer Simul 19(9):3106–3118
55. Ahmed F, Anees A, Abbas VU, Siyal MY (2014) A noisy channel tolerant image encryption scheme. Wirel Pers Commun 77:2771–2791
56. Ahmad J, Hwang SO (2016) A secure image encryption scheme based on chaotic maps and affine transformation. Multimed Tools App 75:13951–13976
57. Khan M, Masood F, Alghafis A, Amin M, Batool Naqvi SI (2019) A novel image encryption technique using hybrid method of discrete dynamical chaotic maps and Brownian motion. PLoS ONE 14(12):e0225031