



Efficient Memory Handling Model with Consistent Video Frame Duplication Removal with Precise Compression



Shaik Sameerunnisa^{1*}, Jones Jabez²

¹ School of Computing, Sathyabama University, Chennai 600119, India

² Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai 600119, Tamilnadu, India

Corresponding Author Email: sksameera236@gmail.com

<https://doi.org/10.18280/ria.370215>

ABSTRACT

Received: 24 December 2022

Accepted: 10 February 2023

Keywords:

video frame deduplication, video compression, video segmentation, memory handling, duplicate frame removal

Massive amounts of videos are being made and shared online as mobile devices and social networks gain popularity in recent years. The enormous expansion in the amount of video data created has made storing and quickly searching it all quite difficult. Because many movies are duplicates or near-duplicates in practice, recognizing these copies has become a critical strategy for decreasing the amount of storage with duplicate removal models. Video compression is an important part of Internet video delivery for efficient memory management. Deep learning's growth has sparked a revival in video compression, with many frameworks offering comparable or even higher performance than traditional video codecs presented in recent years. Despite the advancement in rate-distortion, these models are substantially slower and need more memory, limiting their practical application. The exponentially increasing volume of video data created has presented enormous problems to video deduplication technologies. People are interested in uploading and sharing information in photo and video formats in this digital era. This expansion has resulted in increased storage capacity, which contains a large amount of redundant multimedia material. Many deduplication algorithms are being rapidly developed nowadays, although they are often slow and have rather imprecise identification processes. Deduplication is one of the emerging ways for coping with redundant data stored in several locations. When more than a copy of the same data is detected, a single copy is preserved, and the other data is replaced by pointers pointing to the preserved copy and also duplicate frames will be removed by segmenting the video for memory efficiency. Storage can be utilised to effectively store a large amount of other data. While there are many other types of deduplication algorithms, picture and video deduplication strategies and implementations receive a lot of attention since they are difficult to implement. In this research a Consistent Video Frame Duplication Removal with Precise Compression (CVFDR-PC) model for efficient memory handling is proposed. This research provides a versatile and efficient video frame deduplication framework with compression model that effectively handles the memory. The proposed model when contrasted with the existing methods exhibit better performance levels.

1. INTRODUCTION

In recent years, video deduplication technology has been met with increasing difficulty as the amount of video data being created has increased dramatically. Numerous deduplication methods are now under active development, but they are slow and often make mistakes during the identification process [1]. Till now, there has been little research into the efficient similarity ranking approach and the generic hashing architecture for video deduplication. Videos are primarily textually labelled on commercial websites that are increasing day by day in this digital world [2]. These videos tags don't do anything to monitor the content or stop copyright violations. Such infractions have been found using watermarking and material copy detection methods [3]. The watermarking method looks for a specific mark in a movie to determine whether it is copyrighted [4]. By contrasting the fingerprinting of the query video and the fingerprints of a copyrighted videos, other method identifies the duplication.

A fingerprint is a condensed representation of a video's signature that is resistant to changes in the frames and discriminatory enough to differentiate between videos. The attributes used for fingerprint often ensure that the ideal match in the signatures space remains largely constant also after various noise attacks, in contrast to the noisy robustness of the content protection schemes, which is not generally guaranteed [5]. Consequently, the fingerprinting strategy has had better results.

A duplicate video is one that comprises only a subset of the frames from the original movie and, in certain cases, the frames themselves have been modified and reordered in time. The assumption that a duplicated video only contains frames from a single video has been used in many previous efforts on duplication detection. In, it is demonstrated that 32% of the relevant videos returned for just a set of 24 queries in YouTube, Google Video, and Yahoo Video are duplicates. Every video content in the collection is said to have on average of five identical copies; there were 81000 clips

totaling 2800 hours of material in the database [6]. Additionally, in the 10-leading retrieval of information for some popular video search engines, there were more than five duplicate results. The duplicate frame removal process is shown in Figure 1.

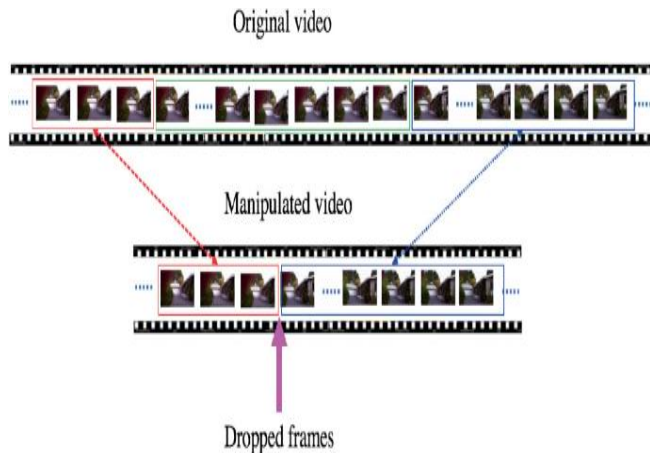


Figure 1. Duplicate video frame removal process

Massive amounts of video content are being produced and shared online as a direct result of the widespread adoption of smartphones and the rise of social media platforms. As the volume of video content continues to increase at an exponential rate, it becomes increasingly difficult to both store and quickly search through. Many films are duplicates or near-duplicates in practise, therefore finding them is a crucial strategy for saving space and processing power [7]. Numerous methods have emerged in recent years for automatically spotting such duplicates in large datasets based on their shared content. By quantizing the attributes of the key frames into visual words and employing an inverted file index to conduct speedy search, a million-video-scale near-duplicate video removal system [8] has been constructed. In contrast to previous studies, a duplicate video frame model methodology based on consistent frame deduplication model is proposed in this research.

Data storage and delivery options are greatly expanded with the incorporation of video into the modern Internet [9]. Video deduplication involves removing duplicated scenes or frames both from within a video and between videos. Compare two videos that are otherwise identical except that the final frame of one video features the same action as the first. Thus, the second video's final frame is unnecessary and might be substituted by a pointer [10]. This not only increases the reliability of locating duplicate photographs, but also efficiently reduces the amount of storage space they take up. In order to achieve video deduplication, the original videos must be split into frames for processing.

The frames in a variable-sized frame can be the same size as those in a fixed-sized frame, or they can be a variety of sizes. Each frame is given a fingerprint and then compared against others of the same or other films, much like the image deduplication process [11]. Data structures like inverted indexes, B+ trees, etc., can be used to index the data and keep the frames in the correct sequence [12]. In order for the films to be retrieved in their proper sequence and seen by the users with minimal fuss.

To efficiently use the memory to store big data of videos, efficient compression model is required that maintains the quality videos without any loss. When data is compressed,

the number of bits used to represent it is lowered [13]. The elimination of superfluous details is the primary goal alongside frame compression. As a result, fewer bits are needed to describe the compressed version of the image compared to the original. Since less data needs to be stored, more photos can be kept in the same amount of space and transferred more quickly, saving both time and bandwidth [14]. Data compression is the process of reducing the amount of data by reducing the number of bits used to represent the data from its original, uncompressed form.

Data compression, image compression, audio/video compression, and telecommunications bandwidth compression are all examples of compression [15]. To take use of the inherent statistical redundancy in real-world photographs, various image compression methods have been developed [16]. Video compression is a procedure that modifies the frames components in order to reduce its overall size while maintaining its visual quality. The term spatial correlation refers to the relationship between the average signal strength at a receiver and the direction from which the incoming signal was received. By dispersing the signal in such a way that its multipath portions are acquired from many multiple geometric directions, rich multipath propagation reduces spatial correlation. The hash intensity is considered as threshold value based on the image pixel similarity levels. The approach described here can reduce both the space needed to store the image and the bandwidth allotted to transmit it. The video frame compression process in general is shown in Figure 2.

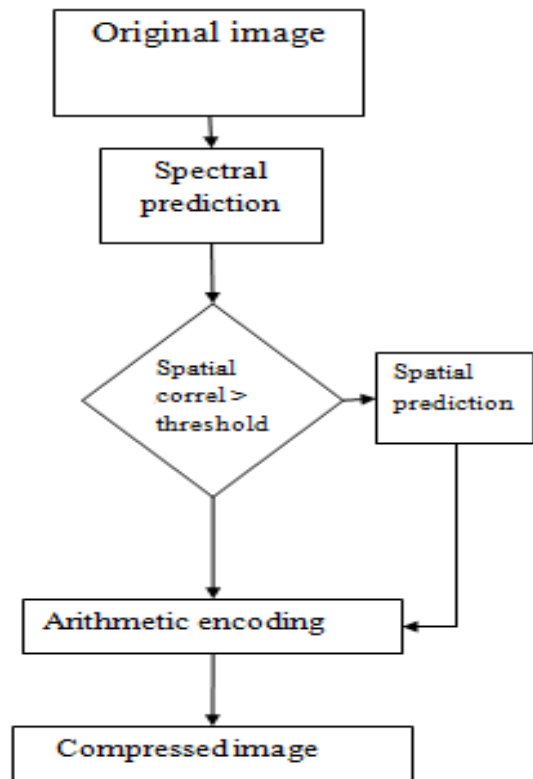


Figure 2. Image compression process

Depending on the desired quality and data throughput, compression techniques can demand a lot of computing power anywhere from 100 mega operations per second (MOPS) to more than a tera operations per second [17]. In contrast to conventional methods, Convolution neural video

deduplication algorithms can identify the similarities across videos by acquiring the visual features from a succession of frames (frame by frame) [18]. On the other hand, videos have high-level temporal consistency between frames and reduced visual content in each frame [19]. As a result, the CNN algorithm is ineffective due to the absence of a model for the temporal coherence between succeeding video frames [20]. The RGB (Red, Green, and Blue) frames' salient material is handled by the two-stream CNN model's first input, while the optical flow's salient content is handled by its second input [21]. The two teaching stream outputs are then merged to create the final video. Additionally, the idea behind the utilization of salient material is to use particular areas of a camera shot rather than employing all of the temporal or spatial aspects of data.

2. LITERATURE SURVEY

2.1 Memory management

One of the key components of cognitive computing systems is neural associative memory (AM). It learns to remember information and retrieves it using the information's actual substance. The ability to increase memory capacity while utilizing the lowest amount of hardware or energy resources is one of the main problems of creating AM for intelligent devices. But past works demonstrate that memory capacity grows incrementally, or in the square root of the sum of synaptic weights. Recursive synapse bit reuse, a synapse architecture is suggested by Guan et al. [2] to address this issue, provides nearly linear scaling in memory storage with total synapse bits. Additionally, compared to the traditional model, this methodology is capable of handling input data which are strongly correlated. The author assessed the model using Hopfield neural networks (HNNs), which have synaptic weight data storage of 5-327 KB. The model can improve the memory of Nervous system as big as 30× so over conventional ones. The implementation of HNNs at very large scales in 65 nm reveals that our suggested approach can reduce area and energy dissipation by up to 19 and 232, respectively, as compared to the traditional paradigm. The size of the network is anticipated to increase these savings.

File-level deduplication and block-level deduplication are the two main types of deduplication techniques proposed by El-Moursy et al. [3]. Using a hash value computed for the file, the file-level technique ensures that a file with the same hash value is only kept if every bit of it matches the original file. All other duplicates are removed as well. Although simple to execute, the resulting deduplication is subpar. The deduplication cannot be achieved if even a single byte in the file is altered. To do deduplication in the block-level approach, each file is first split up into many blocks. This technique is superior than the previous one and works well for partial content changes that require deduplication.

Related efforts on deduplication include access control in secure deduplication analyzed by Tan et al. [4], evidence of ownership for deduplication, facilitating privacy-preserving fuzzy deduplication, securing approved deduplication, and distributing deduplication is analyzed. It appears that the vast majority of existing deduplication systems are geared toward the analysis of text and images, whereas only a select few works have investigated video. The issue of private video deduplication has not had a secure framework to tackle it

until now that is analyzed.

In the age of deep learning, video semantic segmentation, which tries to produce temporally consistent segmentation findings, is still an extremely difficult task. Yuan et al. [5] enhanced previous methods in two ways. The crossing and self-attention networks (CSANet) is shown at the level of network architecture. Contrary to earlier techniques, CSANet is built to gather spatial context only within current frame in addition to propagating temporal characteristics from nearby frames. This has been proven to significantly increase the consistency and durability of the retrieved deep features. On the level of the loss function, the author additionally put forth the inter-frame joint learning technique, which guarantees the cross-attention module will concentrate on context regions that are semantically associated, enabling the selected features at various frames to be cooperatively enhanced. The author demonstrated that the suggested strategy can achieve cutting-edge efficiency just on Cityscapes and CamVid benchmarks by merging the two aforementioned unique ideas.

When it comes to social media image streams, Chen et al [7] proposed a new method for detecting and categorising near-duplicates. This issue is interpreted as one of clustering the streamed photos so that each cluster has a set of images that are nearly identical to one another. The suggested method is a two-tiered hierarchy, with the first level using a global feature descriptor-based similarity check to reduce the search space and the second level using Locality Sensitive Hashing (LSH) to detect similarity using a small number of photos and speed up the processing time. There are two stages to the method. In the first step, known as training, a large collection of random photos from various social media platforms is used to generate a training dataset. The moment-based features used in this clustering technique are also invariant to translation, rotation, and scale, making them ideal for creating a visual lexicon. The next phase is a near-duplicate detection technique that occurs online. Here, an LSH-based similarity check between the test picture and the identified candidate images is carried out. The candidate images are chosen in such a way that each candidate image represents a unique grouping (cluster) of near-duplicate images.

2.2 Image quality management

Choi et al. [9] developed a face spoof detection method to identify instances of the same person appearing in multiple different photo formats. The algorithm is able to recognise people by their faces and normalise them. The analysis of image distortion is the subject of this investigation. For the purpose of detecting image distortion in the spoof face images, four different features were developed (specular reflection, blurriness, colour moments, and variation in colour). Classifiers trained from various sets of training data are compared with the retrieved characteristics of the normalised images. If the input image is identical to all the examples in the classifier, the software will conclude that it is a forgery.

An efficient and successful global context verification technique for image copy detection was presented by Zhuo et al. [10]. The core features of the system are feature matching using the Scale Invariant Feature Transform (SIFT), OR-GCD extraction, and SIFT match verification. The first one uses Bag of Words (BOW) quantization to match SIFT features between pictures to obtain preliminary SIFT matches.

OR-GCD extraction defines the global context of each matched SIFT function. The extracted and concatenated binary vectors are then utilised to verify the matched SIFT features by comparing their respective OR-GCDs, and the verification result can be applied to the calculation of picture similarities for copy detection. The proposed method is extended to accommodate the task of detecting partial-duplicate images. The identification of possible duplicated regions is a vital step in the detection of partial-duplicate images. As a means of weeding out geometrically contradictory matches for the possible location of duplicated regions, the SGV technique is proposed. This method considers not only the spatial links between local features, but also their characteristic associations.

For visual sensor networks, Liu et al. [13] developed a method to efficiently and accurately remove images that are almost identical. This is due to the abundance of visual data, such as digital photographs and movies, that are sent on visual sensor networks, produced by visual sensor nodes, in this case the camera nodes. There are a lot of near-duplicate pictures among them, which is a huge waste of space. In this step, the photos are sliced into a 33 matrix before being turned into hash code. When trying to get rid of near duplicates, the first and most important step is to group them into clusters. Then, to finish off the near-duplicate elimination, they propose a new seed image selection approach based on the PageRank algorithm. This algorithm can reliably identify the most relevant photos to use as seed images, while simultaneously removing any unnecessary ones. The well-known inverted index file generation method known as Locality Sensitive Hashing (LSH) is used to match high-dimensional characteristics. A second comparison is performed between the original characteristics to verify a match once the hash values of a user query function are retrieved from the inverted index file entries.

By simultaneously examining the unique qualities of backdrop and foreground redundancy and introducing two innovative fixed points in a complimentary way, Maninis et al. [14] improved the inter stable quality of surveillance videos. One way to lessen background redundancy is to use a block level backdrop reference frame (BRF). The suggested method uses instance segmentation to make background block selection easier and integrates semantics into the compression to ensure that the resulting BRF is free from foreground interference. In contrast, a foreground frame (FRF) is created using the Security Prediction Generative Adversarial (SP-GAN), which makes use of previously reconstructed frames, optical circulation prediction, and BRF to deduce the foreground objects of the coded frame. The author included the suggested plan into the HM-16.6 software and add BRF and FRF to the list of reference images.

2.3 Image compression models

Motion predictive coding, a very efficient approach for video compression, can scarcely be taught into a neural network, which poses a significant obstacle to learning-based video compression. The idea of Pixel-Motion CNN (PMCNN), which incorporates motion extension and SVM based networks, is presented in this research by Zhong et al. [17]. In order to efficiently perform predictive coding within the learning network, PMCNN can simulate spatiotemporal coherence. The author investigated to learn framework for

compression algorithms using binarization and iterative analysis/synthesis as additional components on top of PMCNN. The experimental results show how effective the suggested plan is. Despite not using entropy coding or complicated settings, the author nonetheless demonstrated better performance than MPEG-2 and reach results that are comparable to those of the H.264 codec.

The technologies for video and image coding have made tremendous strides in recent years. The growing rate of image and video data, however, is far outpacing the improvement in compression ratio as a result of the widespread use of picture and video acquisition devices. It has been well acknowledged, in particular, that pursuing additional coding performance enhancement within the conventional hybrid coding framework presents increasing hurdles. A unique and promising approach to image and video compression is offered by the deep convolution neural network by Xuan et al. [18], which has helped the neural net resurge in recent times and has seen considerable success in the domains of artificial intelligence and signal processing. The author offered an organized, thorough, and current evaluation of machine - learning image and video reduction approaches in this study. For photos and videos, respectively, the genesis and advancement of machine - learning compression approaches are shown. More particularly, cutting-edge video coding methods that make use of the HEVC framework and deep learning are introduced and analyzed. These methods significantly improve the performance of modern video coding.

3. PROPOSED METHOD

A video is a collection of frames that play one after another. It's well knowledge that video files include a lot of redundant information because adjacent frames are highly connected. Keeping this much duplicate information takes up unnecessary space [22]. File-level deduplication might not be applicable here because the frames are not identical but are strongly connected and locally similar [23]. By removing unnecessary frames from videos, a lot of storage room can be saved. Over the past few years, researchers have examined the impact of block granularity on the efficiency of block-level deduplication. Although storage and security performance are important, private-preserving block-level deduplication of multimedia content is still frequently disregarded [24]. The proposed frame level deduplication system improves deduplication efficiency by performing deduplication not only on the units of a single video, but also on the blocks of similar content videos [25].

The core video data can be used to organise deduplication processes. The reduction of the frame square, in which duplicate frame square copies are eliminated and record squares are isolated, is another option that can be considered for better deduplication process [26]. To eliminate duplicates, data is transferred to the farthest possible server in a server-side subtraction [27]. This tactic allows for the accumulation of more things without exceeding some kind of data transfer cap. Client-facing de-duplication relies on a detailed history of users movements [28]. Although the record prompts the user to check if the server is live, this approach keeps the restriction position and movement speed same. The degree of reduction, which guarantees video data reduction technique divided by the total number bytes of measurement of bytes, is

a common metric of subtraction efficiency. This is the gap between the fixed size of records moved by all customers and the total extra room utilised by administration, as provided by the distributed storage system. The deduplication rate, often known as the space decrease rate, is the ratio of deduplications created per unit of original videos.

Video compression has been the subject of a lot of research in the previous years, yet most methods still fail to faithfully recreate the original video in their output. It is proposed that feature reduction techniques and the wavelet transform be used in conjunction to create a new point selection approach for video compression [29]. The suggested technique can recognise scenes and break them down into discrete video views. The suggested method takes object maps from predefined stages and picks out features that change between scenes. A set of feature points representing the scene's foreground is generated for each video scene [30]. The original video can be recreated with the use of an object map that contains all the elements that were in the original [31]. By determining which items and features are shared by the backdrop and the foreground, users may narrow down the feature set. Only the unique images are copied across, while the originals are restored with the same characteristics. When compared to other methods, the compression ratio obtained by the suggested method is better. The proposed model framework is shown in Figure 3.

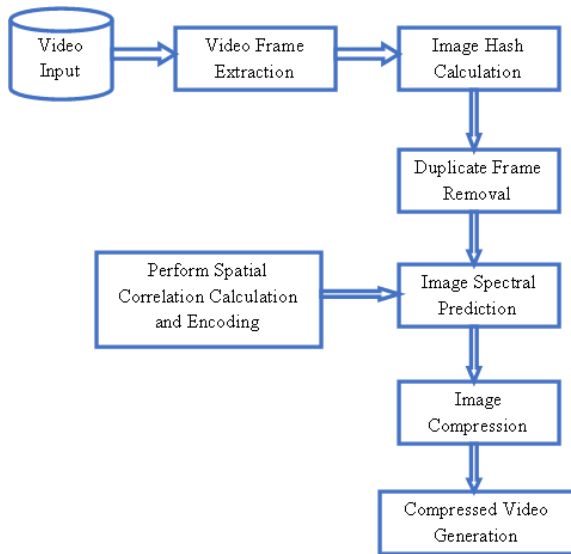


Figure 3. Proposed model framework

The quality of the video determines how many still images can be extracted from the input footage. In most cases, 40 still images are produced for every second of video time. As new images are taken, interest spots are calculated. To lessen the time complexity, the interest spots are created using an identically sized image window to those used in conventional model. To determine how much variation exists between frames of a video clip, the similarity measure is considered to compare sets of feature descriptors. If the time gap between two frames is large enough, the video will transition to the next scene. If the frame doesn't exist in the same video scene, the procedure still uses the feature descriptors from the original video scene.

Multi-level enhanced wavelet transform is performed on detected small video sequences. Enhanced Wavelet compression is applied to each still image once it has been

extracted from its corresponding video scene. Afterwards, the image that has been modified by wavelets is utilised. The original characteristics of the image are preserved by the multi level enhanced wavelet transform, allowing for their restoration at a later time with the same quality. In this research a Consistent Video Frame Duplication Removal with Precise Compression (CVFDR-PC) model for efficient memory handling is proposed.

Algorithm CVFDR-PC

{
Input: Videos Dataset {VDS}

Output: Deduplicated and Compressed Video

Step-1: Consider the video record from the dataset and the video will be divided into frames. Video is a collection of still images played rapidly enough to create the illusion of motion. Each of these pictures is referred to as a frame. The considered video is segmented into frames for further analysis. The video frame division is performed as:

$$VFrameSet[F] = \sum_{i=1}^F \frac{\delta(VD(i))}{size(VD(i))} * getFrame(i) + \frac{getfeyframe(VD(i))}{maxkeyframe(i)}$$

Step-2: The frames from the video are considered and the frame hash values are calculated to perform frame deduplication. Deduplication is the process of eliminating duplicate frames from the video frame set. The frame hash value calculation is performed as:

$$FHV(VFrameSet[F]) = \frac{N_y}{N_x} * \sum_{i=1}^F mean(VFrameSet_i, VFrameSet_{i+1}) + \frac{H(VFrameSet(i))}{size(VFrameSet(i))} + Th$$

Step-3: The duplicate frames will be removed in the frame set and then the remaining images spectral prediction is performed for performing frame compression, the process is performed as:

$$FCompress[F] = \sum_{i=1}^F \frac{H(VFrameSet(i + 1)) - FVH(VFrameSet(i))}{\frac{F}{4} * \sum_{i=1}^F (VFrameSet(i))}$$

Step-4: The Spatial Correlation Calculation and Encoding is performed on the deduplicated frame set for avoiding duplicate frames in a video. In most cases, samples of the image same colour component in an image will exhibit high vertical, horizontal, and spatial correlations. The proposed model can achieve modest compression when applied to bilevel or pseudocolor image stored data in scan-line order by taking advantage of horizontal correlation among pixels. The three stages of information processing are referred to as encoding, storage and retrieval, respectively.

The process of spatial correlation calculation and encoding

is performed as:

$$\begin{aligned}
 & SCorr(F) \\
 &= \frac{\sum_{i=1}^F [H(VFrameSet(i) - G(\alpha))]}{\sqrt{\sum_{i=1}^F \max(\lambda) + \min(\alpha) + Th}} \\
 & * \frac{[[std(\max(FCompress(i^2)) - \min(H(FHV(i)))_2)]]}{\sqrt{\sum_{i=1}^F \max(\lambda) + \min(\alpha) + Th}}
 \end{aligned}$$

Step-5: The image compression is performed on the final set and the video is generated from the deduplicated frameset that generates a final video with reduced size. The image compression is performed and the video is generated as:

$$\begin{aligned}
 VFCo[F] = \sum_{i=1}^F \Sigma_{i=1} I \begin{pmatrix} x \\ y \end{pmatrix} + pix(i + 1, i - 1) \\
 + \left\{ \frac{|F_y - SCorr(x)|}{|F_x - SCorr(y)|} \right\} * Th
 \end{aligned}$$

$$\begin{aligned}
 & VCompSet(F(x, y)) \\
 &= \sum_{i=0}^F \frac{pixset(x, y) + \max(\lambda) - \min(VFCo(i))}{\max(VFCo(i)) + size(SCorr(i))} \\
 & - \sum_{i=0}^F \frac{\sum_i^F pixset_{x,y} (y - x)^2}{\max(VFCo(i)) + size(SCorr(i))} + \alpha
 \end{aligned}$$

}

4. RESULTS

Distributed storage systems typically feature deduplication as a means to conserve bandwidth and storage space. However, deduplication is more difficult to achieve with encoded data since different key encryptions transform the same data into different arrangements. In this research, various approaches where deduplication algorithms are carried out on encrypted data in a vast warehousing facility is analyzed and a new deduplication and compression model is suggested. The majority of the methods are based on integrated encryption, a simple approach that guarantees flawless deduplication even with encrypted data. Security and data duplication across capacity regions are trade-offs that just cannot be made in today's information-rich world.

To facilitate progress in data hoarding without demanding user input on encryption method, data deduplication solutions in data collecting servers where open data is encrypted should be developed. By leveraging the information gleaned from the created item mappings and video frames, the image is compressed. Differences in the object map between scenes and frames are used to compress the image. The wavelet-transformed image from the first frame is sent to the receiver unchanged so they may use it as a reference, while the rest of the frames are filtered away using background subtraction and just the object map's distinctive features are sent. The received picture is utilised to generate a new object map, which is then compared to the current one by the receiver. Once a distinction is made, the original qualities are preserved.

In this research a Consistent Video Frame Duplication Removal with Precise Compression (CVFDR-PC) model for efficient memory handling is proposed. The proposed model effectively removes the duplicate frames in the video and also compression is applied for storage space optimized usage. The proposed model is compared with the traditional Synthetic video Texture dataset (BVI-SynTex) model. The frame is one photo in a video. Usually, a video will have 24 or 30 frames per second (FPS), where 1 second equals the number of frames displayed. Together, the image and the moment at which the photograph was taken make up the frame. The considered video will be divided into frames to check for duplicates. The video frame generation time levels of the proposed and existing models are shown in Figure 4.

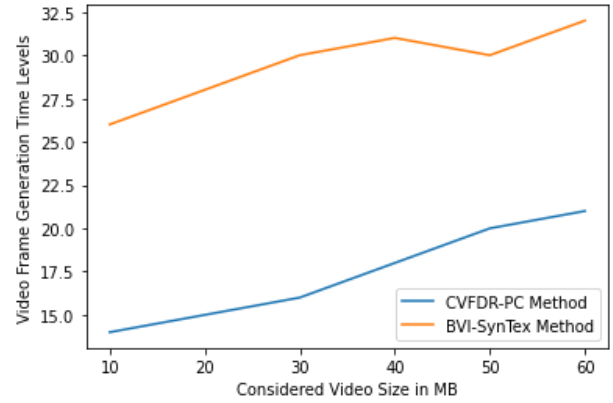


Figure 4. Video frame generation time levels

Video is one of the most sought-after forms of multimedia content since it can satisfy a wide variety of users' needs in a short amount of time. One of the difficult goals of video processing research is the development of methods to reduce computational burdens. The next step in video retrieval tasks involves classifying videos and extracting essential frames from them using vital protocols. The frames from the video is extracted completely and then they are analyzed for deduplication. The video frame generation accuracy levels are depicted in Figure 5.

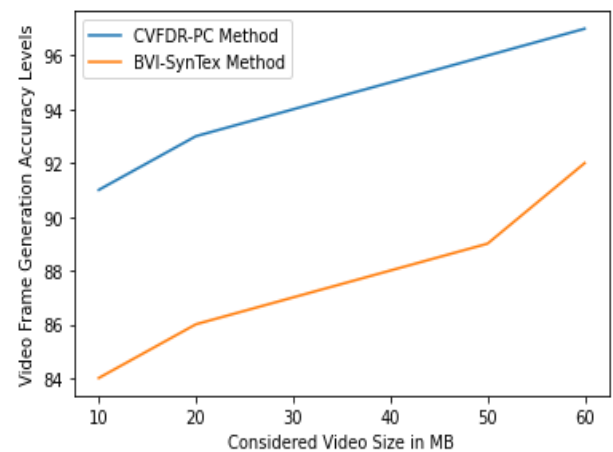


Figure 5. Video frame generation accuracy levels

The deduplication feature relies on a filter configuration that determines which incoming video, if any, is to be deduplicated. This includes video on a certain part or of a particular kind. Frames are compared using a correlation key

formula to identify duplicates, and the user can customise which parts of the frames are analyzed using the various protocol offsets and masking options at their disposal. The Figure 6 represents the frame deduplication detection time levels of the proposed and existing models.

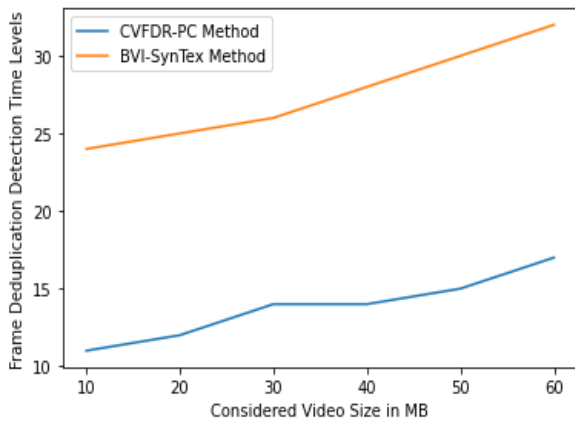


Figure 6. Frame deduplication detection time levels

A duplicate frame is an exact replica of an adjacent frame. 23.976 (24) fps. It was established that 24 frames per second (23.976 for "digital" film) was the bare minimum required to capture video while still keeping realistic motion, therefore it's the standard at which physical film footage is shot. The Figure 7 shows the frame deduplication detection accuracy levels of the traditional and proposed models.

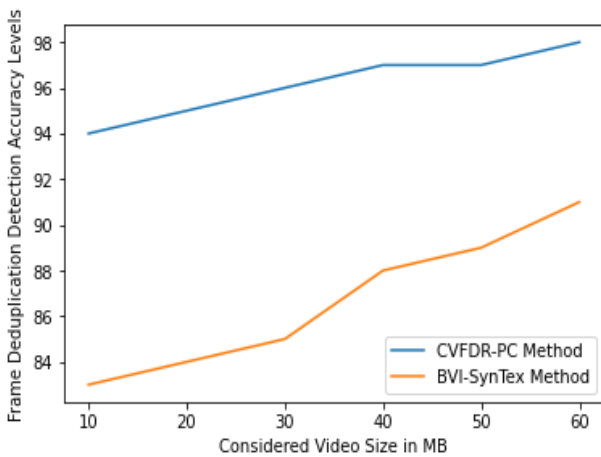


Figure 7. Frame deduplication detection accuracy levels

Compression techniques that take into account events occurring between frames are collectively referred to as inter-frame compression. Video file sizes can be reduced by taking advantage of the fact that there may be very few changes from one frame to the next. The image compression model effectively reduces the frame size while maintaining quality. The image compression accuracy levels of the proposed and existing models are shown in Figure 8.

The process of hashing an image involves running the image through an algorithm to get a unique identifier. Hash values for many copies of the image are identical. For the average hash algorithm, the first step in processing an image is to grayscale it, and then to scale it down. After that, the image's grayscale average is determined, and each pixel is

analysed in sequence from left to right. The Figure 9 represents the image hash calculation accuracy levels.

When it comes to video compression, a video frame can be compressed using a variety of algorithms, each with its own set of benefits and drawbacks, most of which revolve on the total amount of data compressed. Picture types, also known as frame types, refer to the varying algorithms used to create individual frames in a video. The proposed frame compression reduces the pixel size with same quality for storage. The image compression time levels of the proposed and existing models are shown in Figure 10.

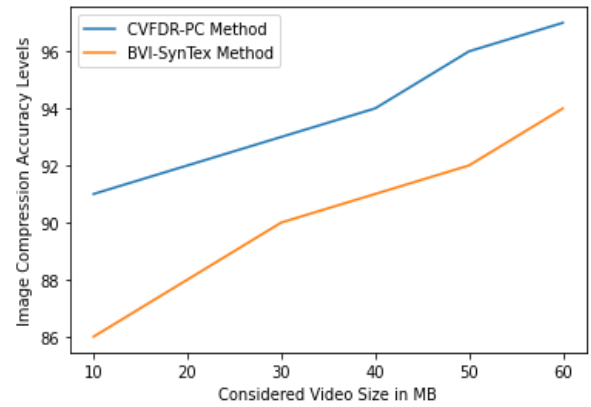


Figure 8. Image compression accuracy levels

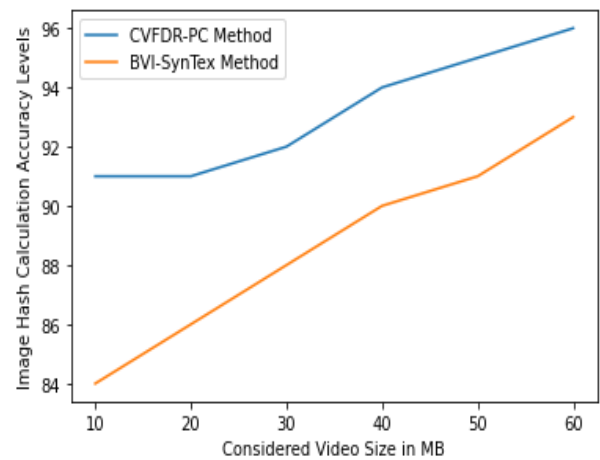


Figure 9. Image hash calculation accuracy levels

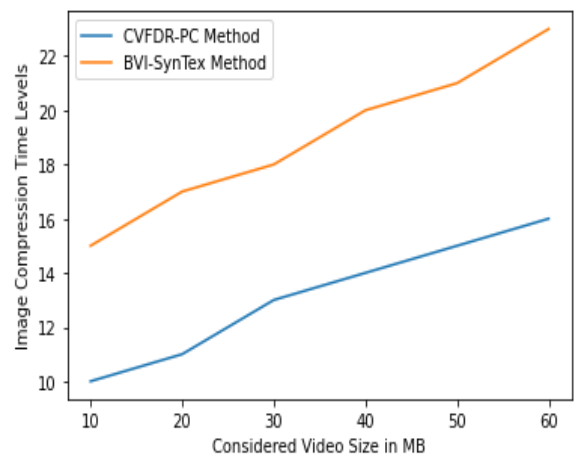


Figure 10. Image compression time levels

5. CONCLUSION

Intensity pattern change is used to identify scenes after the video is segmented into frames. The variation in successive images or frames within a scene is calculated based on previously detected scenes. A method is developed to identify and preserve the intensity-variable pixel while cancelling out the rest. Using only the unique pixels in this way decreases the frame size and the amount of room required to store them. The frames are then compressed further using an enhanced wavelet transform, which results in a higher compression ratio. Both the compression ratio and video quality have improved in the proposed method. The purpose of this study is to develop new methods for better video compression with duplicate frame removal. The input video is broken up into individual frames in the proposed technique employing enhanced wavelet transform for video compression. Sub-sampled pictures and feature descriptors are created from the resulting frames. The approach determines the scenes by their feature similarity and distortion rate, both of which are derived from the feature descriptor. Next, frames are picked and scenes are compressed one at a time using the multi-attribute feature estimation method. In this research a Consistent Video Frame Duplication Removal with Precise Compression model for efficient memory handling is proposed. With a time complexity of 4.5 seconds and a compression ratio of 95.8, the proposed technique achieves a higher duplicate frame removal accuracy of 98%. In future, the problem of fast, real-time duplicate recognition within a large video collection is explored via a variety of powerful algorithms. The features of the duplicate frames can be reduced for reduced time complexity and for better duplicate frame removal rate.

REFERENCES

- [1] Katsenou, A.V., Dimitrov, G., Ma, D., Bull, D.R. (2021). BVI-SynTex: A synthetic video texture dataset for video compression and quality assessment. *IEEE Transactions on Multimedia*, 23: 26-38. <https://doi.org/10.1109/TMM.2020.2976591>
- [2] Guan, T., Zeng, X., Seok, M. (2019). Recursive synaptic bit reuse: An efficient way to increase memory capacity in associative memory. *IEEE Transactions on Very Large-Scale Integration (VLSI) Systems*, 27(4): 757-768. <https://doi.org/10.1109/TVLSI.2018.2884250>
- [3] El-Moursy, A.A., Sibai, F.N., Khaled, H., Nassar S.M., Taher, M. (2020). Deployment and analysis of a hybrid shared/distributed-memory parallel visualization tool for 3-D oil reservoir grid on Openstack cloud computing. *IEEE Access*, 8: 212280-212297. <https://doi.org/10.1109/ACCESS.2020.3037731>
- [4] Tan Z., Liu, B., Chu, Q., Zhong, H.S., Wu, Y., Li, W.H., Yu, N.H., (2021). Real time video object segmentation in compressed domain. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(1): 175-188. <https://doi.org/10.1109/TCSVT.2020.2971641>
- [5] Yuan, Y., Wang, L., Wang, Y. (2021). CSANet for video semantic segmentation with inter-frame mutual learning. *IEEE Signal Processing Letters*, 28: 1675-1679. <https://doi.org/10.1109/LSP.2021.3103666>
- [6] Zhao, L., Wang, S., Wang, S., Ye, Y., Ma, S., Gao, W. (2022). Enhanced surveillance video compression with dual reference frames generation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3): 1592-1606. <https://doi.org/10.1109/TCSVT.2021.3073114>
- [7] Chen, Z., He, T., Jin, X., Wu, F. (2020). Learning for Video Compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(2): 566-576. <https://doi.org/10.1109/TCSVT.2019.2892608>
- [8] Ma, S., Zhang, X., Jia, C., Zhao, Z., Wang, S., Wang, S. (2020). Image and video compression with neural networks: A review. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(6): 1683-1698. <https://doi.org/10.1109/TCSVT.2019.2910119>
- [9] Choi, K., Chen, J., Park, M.W., Yang, H., Choi, W., Ikonin, S., Piao, Y., Esenlik, S., Park, M., Wang, Y.K., Choi, N., Zhao, Y., Jeong, S. et al. (2020). Video codec using flexible block partitioning and advanced prediction transform and loop filtering technologies. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(5): 1326-1345. <https://doi.org/10.1109/TCSVT.2020.2971268>
- [10] Zhuo, T., Cheng, Z., Zhang, P., Wong, Y., Kankanhalli, M. (2020). Unsupervised online video object segmentation with motion property understanding. *IEEE Transactions on Image Processing*, 29: 237-249. <https://doi.org/10.1109/TIP.2019.2930152>
- [11] Gui, Y., Tian, Y., Zeng, D.J., Xie, Z.F., Cai, Y.Y. (2020). Reliable and dynamic appearance modeling and label consistency enforcing for fast and coherent video object segmentation with the bilateral grid. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(12): 4781-4795. <https://doi.org/10.1109/TCSVT.2019.2961267>
- [12] Kim, M.J., Kim, J.G., Yoon, S.K., Kim, S.D. (2021). Functionality-based processing-in-memory accelerator for deep convolutional neural networks. *IEEE Access*, 9: 145098-145108. <https://doi.org/10.1109/ACCESS.2021.3122818>
- [13] Liu, W., Lin, G., Zhang, T., Liu, Z. (2021). Guided co-segmentation network for fast video object segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(4): 1607-1617. <https://doi.org/10.1109/TCSVT.2020.3010293>
- [14] Maninis, K.K. et al. (2019). Video object segmentation without temporal information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(6): 1515-1530. <https://doi.org/10.1109/TPAMI.2018.2838670>
- [15] Oh, S.W., Lee, J.Y., Xu, N., Kim, S.J. (2022). Space-time memory networks for video object segmentation with user guidance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1): 442-455. <https://doi.org/10.1109/TPAMI.2020.3008917>
- [16] Li, K., Tao, W., Liu, L. (2019). Online semantic object segmentation for vision robot collected video. *IEEE Access*, 7: 107602-107615. <https://doi.org/10.1109/ACCESS.2019.2933479>
- [17] Zhong, Y., Shu, M., Liu, Z., Lu, X. (2022). Spatio-temporal dual-branch network with predictive feature learning for satellite video object segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-15. <https://doi.org/10.1109/TGRS.2022.3148310>
- [18] Xuan, S., Li, S., Han, M., Wan, X., Xia, G.S. (2020). Object tracking in satellite videos by improved correlation filters with motion estimations. *IEEE*

- Transactions on Geoscience and Remote Sensing. 58(2): 1074-1086.
- [19] Hu, Z., Yang, D., Zhang, K., Chen, Z. (2020). Object tracking in satellite videos based on convolutional regression network with appearance and motion features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 13: 783-793. <https://doi.org/10.1109/JSTARS.2020.2971657>
- [20] Lu, X., Zhong, Y.F., Zheng, Z., Liu, Y.F., Zhao, J., Ma, A., Yang, J. (2019). Multi-scale and multi-task deep learning framework for automatic road extraction. *IEEE Transactions on Geoscience and Remote Sensing*. 57(11): 9362-9377. <https://doi.org/10.1109/TGRS.2019.2926397>
- [21] Xiao, H., Kang, B., Liu, Y., Zhang, M., Feng, J. (2020). Online meta adaptation for fast video object segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(5): 1205-1217. <https://doi.org/10.1109/TPAMI.2018.2890659>
- [22] Liang, H., Liu, L., Bo, Y., Zuo, C. (2021). Semi-supervised video object segmentation based on local and global consistency learning. *IEEE Access*, 9: 127293-127304. <https://doi.org/10.1109/ACCESS.2021.3112014>
- [23] Yin, Y., Xu, D., Wang, X., Zhang, L. (2021). Directional deep embedding and appearance learning for fast video object segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8). <https://doi.org/10.1109/TNNLS.2021.3054769>
- [24] Abdulhussain, S.H., Al-Haddad, S.A.R., Saripan, M.I., Mahmmod, B.M., Hussien, A. (2020). Fast temporal video segmentation based on Krawtchouk-Tchebichef moments. *IEEE Access*, 8: 72347-72359. <https://doi.org/10.1109/ACCESS.2020.2987870>
- [25] Vlagkoulis, V., Sari, A., Antonopoulos, G., Psarakis, M., Tavoularis, A., Furano, G., Boatella-Polo, C., Poivey, C., Ferlet-Cavrois, V., Kastriotou, M., Martinez, P.F., Alía, R.G. (2022). Configuration memory scrubbing of SRAM-Based FPGAs using a mixed 2-D coding technique. *IEEE Transactions on Nuclear Science*, 69(4): 871-882. <https://doi.org/10.1109/TNS.2022.3151977>
- [26] Choi, S.H., Park, E., Shin, H., Yoo, S. (2020). McDRAM v2: In-dynamic random access memory systolic array accelerator to address the large model problem in deep neural networks on the edge. *IEEE Access*, 8: 135223-135243. <https://doi.org/10.1109/ACCESS.2020.3011265>
- [27] Dong, Y., Yin, W., Wang, S., Zhang, L., Sun, L. (2019). Memory Leak Detection in IoT Program Based on an Abstract Memory Model SeqMM. *IEEE Access*, 7: 158904-158916. <https://doi.org/10.1109/ACCESS.2019.2951168>
- [28] Vogel, P., Marongiu, A., Benini, L. (2019). Exploring shared virtual memory for FPGA accelerators with a configurable IOMMU. *IEEE Transactions on Computers*, 68(4): 510-525. <https://doi.org/10.1109/TC.2018.2879080>
- [29] Wouafo, H., Chavet, C., Coussy, P. (2019). Clone-based encoded neural networks to design efficient associative memories. *IEEE Transactions on Neural Networks and Learning Systems*, 30(10): 3186-3199. <https://doi.org/10.1109/TNNLS.2018.2890658>
- [30] Jung, H., Ju, J., Kim, J. (2019). Randomized voting-based rigid-body motion segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(3): 698-713. <https://doi.org/10.1109/TCSVT.2018.2805838>
- [31] Xu, Z., Chen, D., Wei, K., Deng, C., Xue, H. (2022). HiSA: Hierarchically semantic associating for video temporal grounding. *IEEE Transactions on Image Processing*, 31: 5178-5188. <https://doi.org/10.1109/TIP.2022.3191841>