Retraction

Retraction: Survey on video compression techniques for efficient transmission (*J. Phys.: Conf. Ser.* **1916** 012211)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022
Survey on video compression techniques for efficient transmission

B SathiyaSivam¹, M G Sumithra², P Sreelatha³

¹Research Scholar, KPR Institute of Engineering and Technology, Coimbatore, TN, India
²Professor, Department of ECE, KPR Institute of Engineering and Technology, Coimbatore, TN, India
³Associate Professor, Department of BME, KPR Institute of Engineering and Technology, Coimbatore
shivasace@gmail.com

Abstract. Nowadays, due to the rapid growth of information technology, data representation has to be performed in several ways. Data file includes text, images, audio, video, and animations, which are large and require lots of space in the hard disk. The Video file consists of sequence of images to be framed in single entity. Image Compression is the effective way to reduce the storage space and speedup the transmission. The Video transmission incurs higher bandwidth requirements. It is necessary to transfer the high quality images to the user devices without loss and latency. This gives encouragement to the researchers to find acceptable compression algorithms. Even though many compression schemes exist, there will be the need for fast compression algorithms which produce acceptable quality images or video with minimum size. This paper presents a survey of various research articles about the image or video compression techniques.

Keywords: DCT, DWT, LWT, RDWT, CNN, RNN, ROI, Lossless Compression, PSNR, Compression Ratio.

1. INTRODUCTION
Video file occupies more space. Uncompressed full HD video requires 10.5 GB per minute. If a smart phone is used to shoot your video, 1080p video takes about 130 MB per minute, while 4K video takes up 375 MB per minute with standard 30 frames per second. Since video files require large space and limited bandwidth, video compression is used to reduce the file size.

Video compression techniques are used for reducing and removing the redundancy in video data. The compressed video file is smaller such that the video can be quickly transmitted over the network. The compression efficiency depends upon bit rate for a given resolution and frame rate. For lower bit rates, compression will be more efficient. In case the image quality is reduced, Video compression may be lossy. For lossy compression, efficient compression technique can be developed to provide good quality video. Even though the compressed video differs from the original, the differences are not clearly visible to the naked eye.
Video can be represented as a series of still frames. The sequence of frames contains both spatial and temporal redundancy. Many video compression algorithms use both spatial compression and temporal compression.

![Frame Splitting](frame_splitting.png)

Figure 1. Working of Video Compression Approach

Figure 1 shows the working approach of video compression. First, the input uncompressed video is split into number of frames of images. Second, selection of set of frames is made. Third, frames are reordered after compression and finally, video has been constructed to produce the compressed video. The deframing is done by generating sequence of images. The number of images generated for 1 second of video is between 12 to 24 images. Set of frames are selected and reordered to produce a compressed video. There are several methods available for the video compression.

The objective of video compression is to provide the video content with low bit-rate, maintaining the good quality picture. Compression can be carried by identifying redundancy and removing it. Mostly video compression demands high quality video for encoding and transmission. However, a good quality video requires more amount of storage and network bandwidth.

2. VIDEO COMPRESSION TECHNIQUES

2.1 Block Matching Techniques

Block Matching algorithm helps to find motion vector for each blocks within a search range and finds a best match that minimize error measure. Motion Vector is defined as the displacement of block location from current frame to the reference frame. Block Matching Techniques consists of three components namely block determination, search method and matching criteria. Block Determination specifies the size and position of Blocks in the current frame. The Search method used to indentify candidate block in the reference frames. Matching criteria is use to find the best match among the blocks in the reference frames. Motion Estimation [1] based temporal prediction is employed in MPEG video. In this technique, high quality video can be obtained which incurs slight cost for the coding.

2.2 DCT Based Video Compression

Discrete Cosine Transform [2] based video compression provides scalability. The compression mainly checks for the presence of redundancies among the several frames and correlation between them. Initially, Videos are converted into individual frames and DCT is applied to each frames. DCT converts the entire pixel values into frequency representation residing the high frequency pixels. The frames are compressed and made into sequence to form compressed video. This technique provides high compression ratio and gives better compressed quality video. This framework can be enhanced using video encoders to achieve higher quality video frames, high frame rate with high resolution.

2.3 DWT Based Video Compression
Discrete Wavelet Transform (DWT) is well known for its image processing and analysis techniques [3]. The wavelets time frequency representation performs sign handling applications. DWT results in better mutilation image quality over the existing techniques.

The DWT architectures available in lifting based, convolution-based and B-spline based designs. The lifting based architectures were mostly preferred since they require less multipliers and adders. B-spline-based architectures proposed along with DWT to reduce the number of multipliers utilizing B-spline factorization [4]. The redundant discrete wavelet transformation (RDWT) is used for motion estimation and compensation in video compression [5]. DWT based Video Compression provides high compression ratio, low mean square error and high peak signal to noise ratio.

2.4 Hybrid Wavelet Fractal based Video Compression

Hybrid Wavelet Fractal compression [6] is based on self similarity within the images. The important thing in fractal based compression is to find the range for each image blocks. This makes the encoding computationally very expensive. The decoding is much faster than encoding and it depends on number of iterations. The convergence usually takes no more than 4-8 iterations. Fractal Video compression partitions the video data and then exploiting the self-similarity present in the video to reduce redundancy. Wavelets are introduced in fractal based Video Compression to achieve high compression gain and fidelity error. Compression ratio decreases as the wavelet passes increases. This can be applicable for 2D fractal images. Unfortunately, 3D fractal compression algorithm is unrealistic.

2.5 Frame Difference based Video Compression

In [7], frame difference based video compression is presented by the author. In this approach, the method extracts the frames from the input video then extracts the set of features. This measures the difference between neighboring and adjacent frames. If the frame difference is higher, the frames are selected for video generation while the lower frame difference results in frame rejection. Three different methods are suggested for frame rejection. In Zero difference approach, the frames are eliminated when there is no difference between neighboring and adjacent frames occurs. In mean difference approach, the frame difference with less value than the mean value gets removed. In percentage difference approach, based on percentage value of difference, the compression is performed.

2.6 Deep Frame Interpolation based Video Compression

In [8], by providing the dense motion compensation, deep frame interpolation network for video compression solves the block based motion vector problem between the consecutive frames. Each frame’s are coded with bi-directional inter-prediction by adding one reference frame. The HEVC bit stream is modified to use reference frame. Deep interpolated frames are entropy coded to reduce syntax size. Deep frame interpolation manages several types of geometrical distortions by providing compact motion compensation. In [9], a codebook comprising of set of interpolation filters is proposed to attain rate-distortion.

2.7 Machine Learning based Video Compression

The boon of multimedia technological advancement gives rise to the significant advancement in video compression. Machine learning based compression [10] offers improved capabilities on complex data. This can be performed by automated analytical model with dynamic prediction. This type of machine learning approach fails in case of untrained videos since incapable of performing encoding operation.
In [11], a new approach is adopted for the problem of image compression with auto-encoders. The computational complexity of the proposed method has not been compared with the existing codec’s approximately. However, training the model on such dataset would require some high-end equipment and a considerable amount of time.

2.8 Deep Learning based Video Compression

With the latest advancement in Deep Learning techniques, Convolutional Neural Network (CNN) based video compression framework is proposed in [12]. Huffman Coding and Scalar Quantization techniques are used to encode feature maps into bit streams.

For 32x32 block, with H.264/AVC Deep Neural Network based video compression shows similar coding efficiency in compressing the video signal but it promises the great potential for future video coding.

There are several other deep learning frameworks are widely used for research such as TensorFlow [13], more recently PyTorch [14]. In the case of the Tensorflow framework, the networks are defined statically, where trainable weights and bias are defined as variable Tensors. These networks can be run inside a Session that initializes the necessary components of the framework to execute a neural network and registers the trainable variables that must be tracked to be optimized.

In the case of the PyTorch framework, the networks are defined as dynamic graphs in which user applied a set of methods during execution to create a neural network. By encapsulating the users can override the forward propagation of an empty network to define a custom architecture.

The training of this type of networks is possible due to the Auto-Grad feature (Automatic Gradient Differentiation) [15]. In comparison, Tensorflow provides a wider range of implemented operations and is established as a reference in deep learning for both research and production.

2.9 RNN based Video Compression

Recurrent Neural Network (RNN) is a hierarchical network in which the input needs hierarchical process in the form of a tree. Recurrent Neural Network can be extended overtime. In [16], the method trains the images and selects the frames for compression based on classification. The author proposed a new approach for pre and post processing using RNN. The framework consists of scalable parts which allow transmitting the frames and reconstructing with high accuracy. This algorithm works along with classic codec H.264. Recurrent Neural Network is used to analyze dynamic terrestrial behavior.

2.10 DNN based Video Compression

Conventional Video compression uses predictive coding. In [17], proposed End-to-end Deep Video Compression Framework to optimize the tradeoff between compression ratio and quality. Motion estimation and compression are not accomplished by deep models. Deep models provide accurate motion information at pixel-level, which can be optimized end-to-end. Correspondingly, deep models are faster than the Conventional Video compression. Based on this framework, new approaches for optical flow, bi-directional prediction and control rate can be analyzed. Deep Neural Network provides powerful non-linear representation.

2.11 Region of Interest based Video Compression
In this approach [18], ROI Detection Algorithm (RDA) is based on spatiotemporal coherent region detection from color and motion similarity analysis and posing the problem of grouping of these regions as an optimization problem. Color and positional similarity evaluation has been done using graph-cut in feature space; and motion similarity assessment has been done using phase-correlation based motion segmentation. By combining the color, motion and positional similarity, background regions are segregated and the initial ROI regions are generated. In the next step, the initial ROI regions are grouped to form the final Region(s) of Interest. In grouping model, predicate D is defined, for evaluation of two spatio-temporal regions. This is based upon dissimilarities among neighboring elements and adaptive with respect to the local characteristics of the data. This predicate determines the cost for the cluster. ROI region grouping algorithm aims at minimization of grouping cost of the entire scene.

2.12 Light Field based Video Compression

Light field video normally captured by arrays of cameras represents tens to hundreds of images at a instance. In [19], Light Field based Video Compression method reduces the storage requirements by more than 95% while maintaining the visual quality. Combining with ROI and Light Field technique, the compression method can protect 5% to 7% bitrates in compared with conventional Light Field video compression technique. In [20], Content based light field method provides good visual quality with improved compression efficiency. In [21], presented a new approach for Surface Light Field based on B-Spline wavelet thereby achieving competitive rate-distortion with low decoder complexity.

2.13 Object based Video Compression

In this type of video compression, different objects are matched with the frames for compression. In [22], video encoding structure achieves object detection, diminishes time based variations, provides rate-distortion and destroys redundant temporal fluctuations. In [23], hybrid object based video compression; a new object segmentation tool is incorporated for both objects and blocks results in high prediction accuracy. Along with spline-based representation, the scheme provides improved coding efficiency compared with the classical object-based coding.

2.14 Color Spaced based Video Compression

This approach extracts the color space from the initial image. The frame selection is done based on the similarity of color space measurements. In [24], the experiment was conducted for a set of 39 HDR video which utilizes cutting edge methodology. Based on this, compression efficiency, quality and complexity of the images gets derived.

2.15 Learning CCTV Compression

In CCTV, for digitizing signal 150mbps of digital stream can be achieved with lot of redundant information. Generally CCTV compression algorithms are lossy, because of higher compression ratio.

There are two categories of CCTV compression available namely Frame and Stream based compression. Besides Frame based compression, Stream based compression is widely used and provides more storage and simple transmission.

The following are the Stream based CCTV Compression Algorithms:

- JPEG2000
- MJPEG (Motion JPEG)
Motion JPEG is the most common compression algorithm for stream based CCTV compression. MJPEG is most suitable for video storage.

In terms of compression ratio, MPEG-4 is three times more efficient than MJPEG. However, it will be a wrong choice for the systems having 5-6 frames.

H.264 is 50-80% efficient than MPEG-4. For limited bandwidth, MPEG-4 and H.264 algorithms are suitable.

In JPEG2000, instead of DCT wavelet transform is used. JPEG2000 provides good quality image. Another advantage is that, it is good for motion detection algorithm since decompresses lower resolution. JPEG2000 requires high system performance thereby making it complex. Table 1 shows Comparison of Video Compression Techniques.

3. Comparative Study

Table 1. Comparison of Video Compression Techniques

| Compression Method                           | Pros                                      | Cons                             |
|---------------------------------------------|-------------------------------------------|----------------------------------|
| Block Matching Technique                    | Finds the Motion Vector                   | Need Codec incurs high cost      |
|                                             | Predicts the best match                   |                                  |
|                                             | Minimizes Error Measure                   |                                  |
| DCT                                         | Provides High Compression Ratio           | Need Enhanced Video Codec        |
|                                             | Good Quality Compressed Video             |                                  |
| DWT                                         | High Compression Ratio                    | Some loss of information         |
|                                             | Low MSE                                   |                                  |
|                                             | High PSNR                                 |                                  |
| Hybrid Wavelet Fractal based Video Compression | Reduce Redundancy                         | Complex Coding                   |
|                                             | High Compression Gain                     |                                  |
| Frame Difference based Video Compression    | Number of Frames selected is reduced      | Higher Compression Accuracy      |
|                                             |                                          |                                  |
| Deep Frame Interpolation based Video Compression | Frames are selected based upon            | Low Compression Ratio            |
|                                             | Interpolation Feature                     |                                  |
| Machine Learning based Video Compression    | Dynamic Prediction                       | Higher Time Complexity           |
|                                             | High Compression Ratio                    |                                  |
| Deep Learning based Video Compression       | Auto-Grad feature                         | Higher Time Complexity           |
|                                             | Dynamic Prediction                        |                                  |
|                                             | High Compression Ratio                    |                                  |
| RNN based Video Compression                | Dynamic Prediction                        | Hierarchy Process Time Complexity|
|                                             | Extended overtime                         |                                  |
| DNN based Video Compression                | Accurate Motion Information               | Hierarchy Process Time Complexity|
|                                             |                                          |                                  |
| ROI based Video Compression                | Selection of Frames based on ROI          | Time Complexity                  |
|                                             | High Compression Ratio                    |                                  |
| Light Field based Video Compression        | High Compression Ratio                    | Complexity                       |
Compression | High Quality Compressed Video
--- | ---
Object Based Video Compression | Selection of Frames based on objects. High Compression Ratio | Needs Cutting Edge Methodology
Color Spaced Based Video Compression | Selection of Frames based on colors. Reduces the frames | Illumination and Lightning affects.

There are some other works that were identified as closely related to the task of video Compression.

Compression of digital video is carried out with different bit rate. In [25], a new method for the retrieval of compressed video from disk to the video server is effectively proposed. Algorithms are derived for the allocation and reservation of resources. With this kind of method, the system can be utilized more effectively and efficiently.

In [26], a novel multiview video coding scheme is presented to improve the compression efficiency. In this video coding scheme, views are grouped as base and enhancement is made. Video encoder is used to code the estimated view based on dependent geometry. Rendering Techniques are used to generate the prediction images for the views. Further coding efficiency can be increased by adopting interpolation and depth estimation.

In [27], by the extraction of video latent representation, Recurrent Neural Network (RNN) can be used in spatio-temporal networks in order to improve video recognition tasks. RNN is used to identify the actions in the latent space. When reconstructing the video sequence from the latent inputs, these representations are more accurate and they provide some better results.

Adversarial network architecture [28], is proposed for real time up-sampling super-resolution videos. In this method, video quality gets improved. This architecture shows that adversarial training is important for video processing and with further exploration; this can be adopted for video compression.

In [29], presents a new architecture for predicting the frames from small video sequences. This architecture achieves considerable results in frame prediction windows by exploring temporal redundancy and capturing dependencies in the sequential frames. This can be employed further in prediction based architecture.

A Novel distributed video coding scheme in [30], is introduced for adaptive puncturing. The distributed coding sends more parity bits but with a lower bit rate. This scheme shows that the compression result gets improved by redirecting the parity bits.

In [31], a new framework is proposed for the wavelet transform with SPIHT algorithm. The wavelet transform provides motion selection instead of motion estimation technique. By combining SPIHT algorithm with DWT, certain improvements can be viewed in speed, memory, redundancy, error resilience, complexity, and compression ratio.

In [32], 3D SPIHT video coding scheme is introduced based on partitioning. In this scheme, even though there is no motion estimation and compensation involved, provides good visualization. This coding scheme provides greater scalability.

4. Conclusion
This paper presents a detailed survey on various methods of video compression for space reduction and efficient transmission over the network bandwidth. As part of analyzing the video compression techniques, we can understand different methods have different implications. Further studies may be carried out based on several hybrid compression models. This survey paper will be helpful for the researchers, academicians who are carrying their work on video compression and to identify the problem level.

References

[1] Avishek S, Jayanta M and Shamik S, A Neighbourhood Elimination Approach for Block Matching in Motion Estimation, *Signal Processing: Image Communication*, 26, 2011.

[2] Suresh G, Epsiba P, Rajaram M, Sivanandam S N, A Low Complex Scalable Spatial Adjacency ACC-DCT based Video Compression Method, *International Conference on Communication, Networking and Networking Technologies, IEEE*, 2010.

[3] Hage, Pankaj S, Pokle, Sanjay B, Gudur, Venkateshwarlu, Discrete Wavelet Transform based Video Signal Processing, *International Conference on Advances in Communication and Computing Technologies, IEEE*, 2014.

[4] Haldorai, A. Ramu, and S. Murugan, Social Aware Cognitive Radio Networks, Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi:10.4018/978-1-5225-5097-6.ch010

[5] R. Arulmurugan and H. Anandakumar, Region-based seed point cell segmentation and detection for biomedical image analysis, International Journal of Biomedical Engineering and Technology, vol. 27, no. 4, p. 273, 2018.

[6] Faiza M, Daoud B, Hybrid Wavelet-Fractal Image Coder Applied to Radiographic Images of Weld Defects, *Communications in Computer and Information Science*, 2011.

[7] Muzhir Shaban Al-Anil and Talal Ali Hammouri, Video Compression Algorithm Based on Frame Difference Approaches, *International Journal on Soft Computing*, 2, 2011.

[8] Begaint J, Galpin F, Guillotet P and Guilleminot C, Deep Frame Interpolation for Video Compression, *IEEE (DCC)*, 2019.

[9] Minoo K and Baylon D, On the Design of Optimal Sub-Pixel Motion Compensation Interpolation Filters for Video Compression, *IEEE (DCC)*, 2015.

[10] Yang Xin, Lingshuang K, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Z, Mingcheng Gao, Machine Learning and Deep Learning Methods for Cyber Security, *IEEE Open Access*, 2018.

[11] Lee J and Ebrahim T, Perceptual Video Compression: A Survey, *IEEE Journal of Selected Topics in Signal Processing*, 6, 2012.

[12] Chen T, DeepCoder: A Deep Neural Network based Video Compression, *IEEE (VCIP)*, 2017.

[13] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado G S, Davis A, TensorFlows Large-Scale Machine Learning on Heterogeneous Systems, 2015.

[14] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L and A. Lerer, Automatic Differentiation in Pytorch, 2017.

[15] Herrera A and Diou D, Image Quality Metrics: PSNR vs. SSIM, *International Conference on Pattern Recognition*, 2010.

[16] Abiodun O I, Aman Jantan, Omolara A E, Kemi Victoria, Mohamed N A, Humaira Arshad, State of the Art in Artificial Neural Network Applications: A Survey, *Heliyon*, 4, 2018.

[17] Liu W, Wang Z, Liu X, Zeng N, Liu Y, and Alsaaad F E, A Survey of Deep Neural Network Architectures and their Applications, *Neurocomputing*, 12, 2016.

[18] Rajandep Kaur, Rajneesh Rani, ROI and Non-ROI based Medical Image Compression Techniques: A Survey and Comparative Review, *IEEE First International Conference on Secure Cyber Computing and Communication*, 2019.
[19] Marco V, Armin M, Jean-Yves, Guillemaut, Adrian Hilton, Light Field Field Compression using Eigen Textures, *IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2019.

[20] Xiang Z, Philip A C, Ming Ting Sun, Mualong Tang, Shanshe Wang, Siwei Ma, Wen Gao A Framework for Surface Light Field Compression, *IEEE International Conference on Image Processing*, 2018.

[21] Kong L and Dai R, Object Detection based Video Compression for Wireless Surveillance Systems, *IEEE Multimedia*, 24, 2017.

[22] Nick A Tsoligkas, Donglai Xu and Ian French, Hybrid Object-Based Video Compression Scheme using a Novel Content based Automatic Segmentation Algorithm, *IEEE International Conference on Communications*, 2007.

[23] Aruna J, Jeyalakshmi C, Priyadarshini K and Revathi A, A Detailed Survey on Video Compression Techniques Towards Efficient Transmission, *International Journal of Advanced Science and Technology*, 29, 2020.

[24] Paek S, Chang S F, Video – Server Retrieval Scheduling and Resource Reservation for Variable bit rate Scalable Video, *IEEE Transactions on Circuits Systems and Video Technology*, 10, 2000.

[25] Shimizu S, Kitahara M, Kimata H, Kamikura K, Yashima Y, View Scalable Multiview Video Coding using 3-D Warping with Depth map, *IEEE Transaction on Circuits Systems and Video Technology*, 17, 2007.

[26] Srivastava N, Mansimov E and Salakhutdinov R, Unsupervised Learning of Video Representations using LSTMS, *International Conference on Machine Learning*, 2015.

[27] Shi W, Caballero J, Husz F, Totz J, Aitken A P, Bishop R, Rueckert D and Wang Z, Realtime Single Image and Video Super-Resolution using an Efficient Sub-Pixel Convolutional Neural Network, *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[28] Lotter W, Kreiman G and Cox D, Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning, *CoRR*, 1065, 2016.

[29] Mohamed H T, Turbo Code using Adaptive Puncturing for Pixel Domain Distributed Video Coding, *ICISP*, 2010.

[30] Chhabikirn Sao, SPIHT based Video Compression, *International Journal of Image Processing and Applications*, 2, 2011.

[31] Shang-Hsiu Tseng and Aldo Morales, A 3D SPIHT Low Memory Encoding Approach for Video Compression, *International Conference on Consumer Electronics*, 2009.