Accordion-based SPIHT Coding for Video Compression using DCT/DWT

SRAVANTHI CHUTKE[1], N. M. NANDHITHA[2], PRAVEEN KUMAR. L[1]
[1]Electronics and Communication,
Sathyabama Institute of Science and Technology,
Chennai,
INDIA

[2]Electrical and Electronics
Sathyabama Institute of Science and Technology
Chennai
INDIA

Abstract: - Video signals tend to contain significant temporal compression because of the high degree of correlation between consecutive frames. This means that existing video compression technologies have not utilized all the information that is readily available in such signals. In particular, we will discuss a novel video compression technique that is capable of exploiting temporal redundancy to improve the details while processing requirements are kept low. In our case, the transformation is applied by converting the 3D video data into 2D space. This transformation allows for applying 2D transforms to search for temporal redundancy and makes motion compensation which is rather expensive computationally unnecessary. This transformation combines the spatial and temporal relations of the video signal and provides each group of pictures (GOP) with a single image with higher spatial relations. We then use SPIHT to increase the compression efficiency as a result of the properties of wavelet-transformed images. The use of discrete wavelet transform (DWT) provides a level of de-correlation that allows effective energy compaction to take place resulting in high video compression ratios. A wide range of experiments clearly indicates the potential of our technique especially when the bit rates are high and with slow-motion videos.

Key-Words: - DWT, DCT, SPIHT, GOP, MSE, PSNR, RLE Reduction, Code Word Reduction.

Received: May 29, 2024. Revised: November 16, 2024. Accepted: December 13, 2024. Published: February 19, 2025.

1 Introduction

The most important objective for most of us, when we place a video in a coding format, is to shrink the size of the video signal with no loss in visual quality so that we can store or transmit it. Optimal video performance relies on adequate quality, storage capacity, and communication bandwidth. With the increasing need for portable digital video products, Real Time Video Encoder and Decoder which are compact in nature, is of much importance and also, in demand. Nowadays frame prediction achieved through motion estimation when implemented with video definement of bulky complexity over the microchip level, proves to be the most adopted approach in video compression. However, the problem of motion estimation is always computationally complex so real-time implementation seems to be expensive and not effective, [1], [2].

On the other hand, there have emerged standards like MPEG (Moving Picture Experts Group), specifically designed for the purposes of stored video applications which usually consist of compressing powerful computers in online mode. 3D transform coders can reach video bit rates similar to those achieved by motion estimation-based coding but with much simpler processing requirements, [3], [4],[5]. However, the efficiency of 3D transforms may not be at their best when there are variations in the pixel values within a spatial or temporal region. In most cases, however, it has been observed that temporal redundancies are usually larger than spatial ones.

In order to take advantage of the temporal redundancies and thus enhance compression, the proposed method seeks to transfer the temporal redundancy of each picture group to the spatial domain. In this manner, temporal and spatial redundancies are combined within one entity that is highly spatially correlated, [6]. This resulting entity is then made to be encoded into a still format image by using a JPEG coder, thus facilitating better compression efficiency. The rest of the paper is organized as follows. In section 2, we focus on the specific details related to the compression ratio improvement techniques, which are the main focus and contribution of this work. Section 3 contains the results of research in which the proposed technique was compared against other existing video compression standards. Section 4 contains a discussion on the applicability of the suggested technique. Ultimately, section 5 wraps up the document in the form of a short conclusion based on our findings.

1.1 DCT-based Coding Techniques

In DCT-based 3D transform coding, the video signal is first partitioned into M× N×K blocks of pixels each of which is composed of a length M in the horizontal direction a length N in the vertical direction and a length K in the time, [7]. After that, a 3D DCT of each block is carried out, followed by its quantization and symbol encoding. One major advantage of this technique is that it completely eliminates the entire motion estimation step which is known to be time-consuming. Additionally, the 3D DCT-based video compression technique is known to have some types of side effects or distortions at low bit rates, including the presence of transparency artifacts that are caused by the 3D DCT, [8]. An example of this artifact appears in Figure 1, which shows a frame in a clip compressed with the use of MPEG (which employs the DCT).

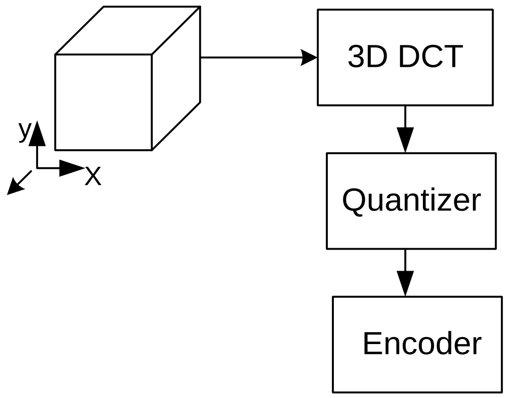

Fig. 1: Shows the 3D DCT application on a video cube

2 Proposed Methodology

The main concept of our constructed approach is video data representation in an almost one-dimensional shape utilizing video signal in the maximum possible amount by alleviating integration between temporal and spatial redundancies among the frames. A video cube is used as the input to our encoder and it contains a number of frames. This cube will first be split into temporal frames and then all the frames will be combined into video frames to form a single frame when viewed in 2D, [9]. Ultimately, the result frame is the one that gets coded after the previous process. This is because strategies that focus on mining for temporal redundancies and take advantage of the spatial redundancies created in the process are aimed at achieving better compression without incurring high processing costs, [10], [11], [12]. Further details on each step and the implementation specifics will be elaborated in the following sections as illustrated in Figure 2.

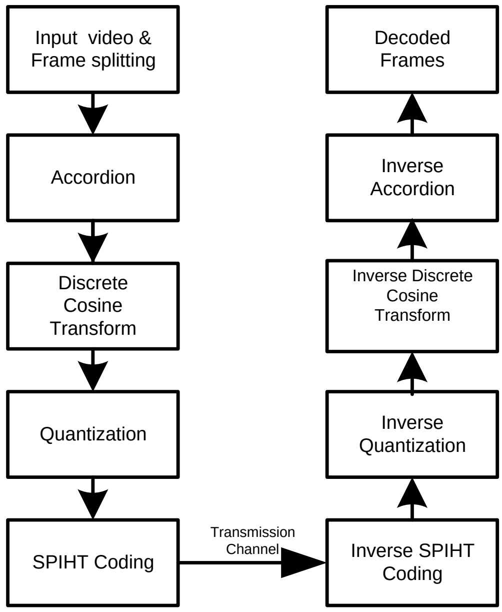

Fig. 2: Block Diagram of Proposed Methodology

2.1 Hypothesis

In 3D videos according to different studies, the temporal dimension is significantly less diverse than the spatial dimension. Thus, in the 3D video signal, pixels demonstrate stronger correlations in the time axis than in the space axis . Mathematically, we define this relationship as shown in equation (1):
for one reference pixel I(x,y,t) where:
I: pixel intensity value
x, y: space coordinate of the pixel
t: time (video instance)
We could have generally:

$$I(x,y,t) - I(x,y,t+1) < I(x,y,t) - I(x+1,y,t) \quad (1)$$

This assumption forms the foundation of our proposed technique, where we aim to arrange pixels with high temporal correlation in close spatial proximity. In such a scenario, the correlation within the resulting 2D frame will be increased thereby it can be assumed that the efficiency in compression will be increased, [13]. Coding efficiency and data redundancy in the final representation are enhanced by the spatial arrangement of temporally intercorrelated pixels. Work on selecting a specific segment of the project. Create a slideshow targeting a particular segment of the project "Accordion" Based Representation To reinforce such an assumption, we first notice a spatiotemporal factorization of the 3D video. This is both a spatial and temporal factorization of an 8x8x8 video cube. The retrieved frames were thus called temporal frames, [14], [15].

2.2 "Accordion" Model

The Accordion view model explained above is shown in Figure 3. It seeks to render pixels with identical co-ordinates of different coordinates of different coordinate images of the video cube adjacent. The transformation has the advantage of transforming the temporally correlated which exists in the original 3-dimensional source of video into high spatial correlation in the 2-dimensional representation called the IACC frame. The reason behind the flipping of the even temporal frames is a strategy that aims at further strengthening the spatial correlation at the edges of the cubes which are cut out from the spatial-temporal volume of the video signals. This modification of the gabled structure of the accordion representation made it possible to reduce the distances between the correlated pixels in the source file.

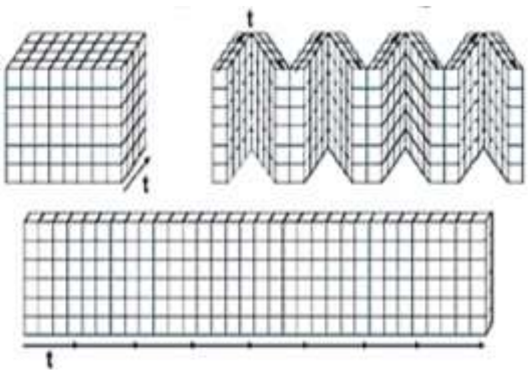

Fig. 3: Accordion view model

2.3 "Accordion" Analytic Framework

The transformation denoted the 'Accordion representation' is obtained from the process which takes the frames of pictures that belong to a group the sometimes-called GOP, (I 1-N) as input and generates the output frame known as IACC. Instead, the first process uses IACC as an input and produces frames There are Proceedings for a group of frames. The process of understanding these two leads to providing this scenario with instructions on how to operate as shown in Figure 4:

Let us note the following definitions for our "Accordion Representation":

1. LLL and HHH are respectively the length and height of the video source frames.
2. NR is the number of frames in a GOP (Group of Pictures/Frames).

We can represent the "Accordion Representation and its inverse" with I_{ACC} and in X_{ACC} as shown in equations (2) to (5) respectively:

$$I_{ACC} = I_n\left(\left(\frac{x}{N}\right), y\right) \quad (2)$$

where:

$$n = \left(\left(\frac{x}{N}\right) mod\ 2\right)(N-1) + 1 - 2\left(\left(\frac{x}{N}\right) mod\ 2\right)(x mod\ N) \quad (3)$$

ACC Inverse formulas:

$$I_n(x, y) = I_{ACC}(X_{ACC}, y) \quad (4)$$

where:

$$X_{ACC} = \left(\left(\frac{x}{N}\right) mod\ 2\right)(N-1) + n(1 - 2\left(\frac{x}{N}\right)\ mod\ 2 + x \quad (5)$$

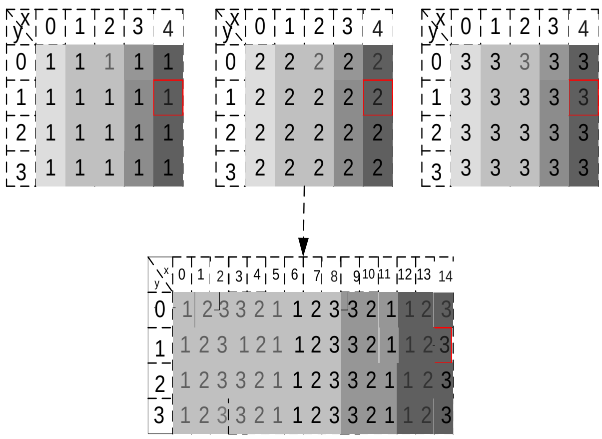

Fig. 4: Shows ACCORDION Framework example

2.4 SPIHT Coding

Set Partitioning in Hierarchical Trees (SPIHT) is a complex image compression technique based on wavelets which deserves special mention because of the following advantages:

1) It has been observed to achieve better quality images accompanied with a high Peak Signal-to-Noise Ratio (PSNR) in relation to colour images.
2) The method has been designed with progressive transmission in mind with images presented in different resolutions as data is being downloaded.
3) It produces a coded file that is completely embedded which makes its storage and transmission easier.
4) The quantization process is rather simple and this improves performance and the efficiency of the process.
5) Both the coding and the decoding processes are fast and almost symmetrical which makes it appropriate for real-time implementations.
6) SPIHT is highly flexible and can be applied to a variety of situations.
7) It offers lossless compression which is able to maintain the original image.
8) It is possible to get the required bit rate or distortion by adjusting the parameters of the technique.
9) Furthermore, the method works well with error protection schemes which improve data integrity during communication. This panoramic view makes it possible to understand why SPIHT is one of the most effective techniques for image compression.

3 Experiments

In this section, we present the results of the experiment carried out in order to evaluate the effectiveness of the proposed video coding technique on videos of various classes. The main aim was to establish whether this technology is worth it in terms of compression ratio, speed, and visual quality of the captured video. We present a (Table 1, Appendix) of results for the Proposed Method. The experiments are carried out using MATLAB software 2023a. Figure 5 shows the 2D frames of input and the compressed video.

The 2D version of the original Video

Compressed 2D version of original Video

Fig. 5: Original and Compressed 2D frames

4 Discussion

The Proposed method employing accordion-based methodologies, Discrete Cosine Transforms (DCT), quantization, and SPIHT coding has shown considerable improvement over the conventional compression techniques. The use of the accordion method improves frame representation by efficiently adjusting to different complexity of content and significantly improves data compression rates. When combined with DCT, which efficiently converts spatial domain data to frequency components, a more accurate quantization and fewer artifacts are produced. The use of the SPIHT coding processes also assists in the wastage elimination of bits and improves image quality at higher Peak Signal Noise Ratios (PSNR) and lower distortion. These new techniques in sequence do not only enhance efficiency in compression but also improve on factors such as flexibility and adaptability which makes them more attractive for use in compression of videos compared to the other techniques used.

5 Conclusion

In the end, video signals can also be described across the context of ‘time’ which most encoding techniques for videos don’t commonly utilize. This research proposed a video compression algorithm which has been based on time gaps that exist between successive frames. The scheme proposed not only simplifies the implementation of integrated video equipment due to low processing needs but also brings about new prospects for the enhancement of video encoding. The approach has demonstrated excellent performance in using temporal redundancy for compression, while the computational cost of using such an approach is relatively low. Due to features that are, for example, useful in video surveillance, this method has compression capabilities without a large sacrifice to processing capabilities, meaning that it is also suitable for real-time purposes. There are various topics that could be considered for further research. These include the use of different methods for representing videos and the combining of the "Accordion representation" with other transformations such as the wavelet transformation method, which would allow a global processing of the representation as opposed to the DWT block processing method.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of documentation work the authors used Grammarly in proper sentence formation and Grammar checks. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References:

[1] A. Arunadevi, Jaisiva. S, Mythile. C and Sudhapriya. K, "Reducing Computational Complexity Of 3D-DCT & IDCT in Video Coding Architecture," *2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT),* Trichirappalli, India, 2024, pp. 1-6. doi: 10.1109/ICEEICT61591.2024.10718370.

[2] Han Yang, Xiangyu Chen, Xi Chen, Bin Guo, "A 3D-DCT and Convolutional FEC Approach to Agile Video Streaming," *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS),* Denver, CO, USA, 2022, pp. 636-644. doi: 10.1109/MASS56207.2022.00094.

[3] Mouna Letaief, Anis Ammar, Hana Ben. Fredj and Chokri Souani, "Real time HEVC video compression using Hybrid DWT-DCT algorithm on an FPGA," *2021 International Conference on Control, Automation and Diagnosis (ICCAD)*, Grenoble, France, 2021, pp. 1-5. doi: 10.1109/ICCAD52417.2021.9638763.

[4] Pallavi Phadatare and Pratibha Chavan, "Implementation of DCT and Fractal Compression Technique," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2021, pp. 83-87. doi: 10.1109/ICECA52323.2021.9675895.

[5] Yeganeh Jalalpour, Li -Yun Wang, Wu -C. Feng and Feng Liu, "FID: Frame Interpolation and DCT-based Video Compression," 2020 IEEE International Symposium on Multimedia (ISM), Naples, Italy, 2020, pp. 218-221, doi: 10.1109/ISM.2020.00045.

[6] Maurizio Masera, Giulia. Fracastoro, Maurizio Martina and Enrico. Magli, "A Novel Framework for Designing Directional Linear Transforms with Application to Video Compression," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019, pp. 1812-1816. doi: 10.1109/ICASSP.2019.8683527.

[7] Radu Rădescu, "An Efficient Solution for Video Compression Using an Original Modified Algorithm Applied to Improve the 3D Discrete Cosine Transform," *2018 International Symposium on Fundamentals of Electrical Engineering (ISFEE),* Bucharest, Romania, 2018, pp. 1-5. doi: 10.1109/ISFEE.2018.8742460.

[8] Xianjian Xiao, Yanbin Zhuang, Zunzhi Wang and Xuewu Zhang, "A reconstruction algorithm based on 3D tree-structure Bayesian compressive sensing for underwater videos," *2015 IEEE International Conference on Information and Automation,* Lijiang, China, 2015, pp. 886-891. doi: 10.1109/ICInfA.2015.7279411.

[9] Afshan Mulla, Jaypal Baviskar, Amol Baviskar and Chirag Warty, "Image compression scheme based on zig-zag 3D-DCT and LDPC coding," *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Delhi, India, 2014, pp. 2380-2384. doi: 10.1109/ICACCI.2014.6968373.

[10] Hong Zhao, Hongxia Wang and Hafiz Malik, "Steganalysis of Youtube Compressed Video Using High-order Statistics in 3D DCT Domain," *2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing,* Piraeus-Athens, Greece, 2012, pp. 191-19. doi: 10.1109/IIH-MSP.2012.52.

[11] G. Suresh, P. Epsiba, M. Rajaram and S. N. Sivanandam, "A low complex scalable spatial adjacency ACC-DCT based video compression method," *2010 Second International conference on Computing, Communication and Networking*

Technologies, Karur, India, 2010, pp. 1-6. doi: 10.1109/ICCCNT.2010.5591628.

[12] Tomas Fryza, "Properties of Entropy Coding for 3D DCT Video Compression Method," *2007 17th International Conference Radioelektronika,* Brno, Czech Republic, 2007, pp. 1-4. doi: 10.1109/RADIOELEK.2007.371460.

[13] [13] Li Li and Zhengxin Hou, "Multiview video compression with 3D-DCT," *2007 ITI 5th International Conference on Information and Communications Technology,* Cairo, 2007, pp. 59-61. doi: 10.1109/ITICT.2007.4475617.

[14] V. Naga Prudhvi Raj and T. Venkateswarlu, "A Novel Approach to Medical Image Compression Using Sequential 3D DCT," *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007),* Sivakasi, India, 2007, pp. 146-152. doi: 10.1109/ICCIMA.2007.115.

[15] T. Fryza and S. Hanus, "Relation between character of real video sequences and 3D-DCT compression," Proceedings EC-VIP-MC 2003. *4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications (IEEE Cat. No.03EX667),* Zagreb, Croatia, 2003, pp. 107-112. doi: 10.1109/VIPMC.2003.1220447.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Sravanthi Chutke and Dr. N. M. Nandhitha identified Problem Definition and proposed a new methodology and carried out simulation.
- Praveen Kumar. L helped in writing the paper, carried out the optimization and responsible for the Statistics.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare.

APPENDIX

Table 1. This table compares the proposed system with other conventional methods

S no.	Parameter	Value			
		Proposed Technique	Existing Techniques		
			MRLE	Spiral scanning	DWT
1.	MSE	49.100	54.67	-	62.5
2.	PSNR	41.98 dB	40dB	27dB	39.8dB
3.	Compression%	90.16%	80-85%	90	80%
4.	RLE reduction	70%	60%	-	-
5.	Code word reduction	30%	25%	-	-