Rwanda Journal of Engineering, Science, Technology and Environment, Volume 7, Issue 1, March 2025
eISSN: 2617-233X | print ISSN: 2617-2321

Mitigating RDO and MEC Complexity in H.26x Video Codecs using DWT

James Ntaganda[1,*], Richard Musabe[2]

[1]*Electrical and Electronics Engineering Department, School of Engineering , College of Science and Technology ,University of Rwanda, Kigali P.O.Box 3900.*
[2]*Computer and software Engineering Department, School of ICT, College of Science and Technology, University of Rwanda P.O.Box 3900, Rwanda*
*Corresponding author: j.ntaganda1@ur.ac.rw ; jamesntaganda10@gmail.com

Abstract

The current rise of audiovisual content demand needs dynamic approaches for content compression. In video compression, the weakness of human visual system (HVS) is exploited to reduce spatial and temporal redundancies, hence reducing inherent bit content in video frame sequences. Nevertheless, processing involved pose new challenges of computation cost in terms of time, power and system design complexity especially in real-time video streaming. Mitigating the trade-off between compression ratios, video quality and computation cost continues to be the core area of research in video coding. In H.26x codecs, motion estimation (ME) and motion compensation (MC) algorithms, sometimes simply called motion estimation and compensation (MEC) are used in inter-frame prediction. Conventionally, MEC involves time consuming searching algorithms in macroblock and sub-block matching. Also, H.26x series employ rate distortion optimisation (RDO) in intra prediction (IP) modes which is repetitive and exhaustive. In this paper, we present two approaches to mitigate MEC and RDO complexities in H.26x codecs: (a) we present a Discrete Wavelet Transform-assisted intra prediction (DWTIP). DWTIP avoids exhaustive and blind evaluation of all possible intra prediction modes by restricting specific modes to a specific macroblock and sub-block size predetermined based on pixel homogeneity levels. This leads RDO relaxation. (b) We also present a DWT-assisted motion estimation and compensation (DWTMEC) for inter-frame prediction. DWTMEC limits the search window to DWT approximation sub-band and reduces the search area, hence reducing MEC computation time. Objective video quality metric is used to compare conventional MEC with DWTMEC and conventional IP with and DWTIP. Results showed that employing DWTMEC along with DWTIP, video coding rates are improved with negligible degradation in video quality.

Keywords: Discrete Wavelet Transform, Rate Distortion Optimisation, Video Compression, H.26x Video Codecs, Video Quality Metrics

https://dx.doi.org/10.4314/rjeste.v7i1.2

1. Introduction

The 2015 media report (Hemant Joshi, 2015) estimated that video and audio content sources would generate 89% of consumer internet data traffic (Exabytes per month) by 2018. This trend continued to be influenced mainly by the tremendous growth of advertising video on demand (AVOD) and video streaming on demand (VSOD) as indicated in fig.1 (Statista Market Insights, 2024). Such a rise in demand is associated with technical challenges. Some of these challenges include: (a) Customers' demand for high quality audiovisual content delivered on thin and light platforms with high throughput continues to push designers to the edge (Bing, 2015). In responding to such a demand, video codec system designers have to look for optimal designs. They focus on less power consumption, high processing speed, sufficient memory, battery life, display size and less complex designs. These are the six major competing parameters in video codec designs, especially on portable platforms used in mobile environment. (b) Bandwidth as a scarce resource becomes constrained as more people get connected to the networks for audio-visual information, with inherent enormous bit content**.** Even when bandwidth is not constrained, the cost per content downloads needs to be addressed (Bing, 2015). Mitigating such challenges requires dynamic approaches as well as continuous improvement from academia, standardisation bodies and industry sector, especially in real time video streaming. Based on these competing parameters, a trade-off is always inevitable.

The quality is normally traded off to bandwidth and transmission cost. This is achieved by using video compression techniques to reduce enormous information content in a video sequence. The trade-off has to be done in such a way that visual quality of the video is not degraded below specific minimum quality of service (QoS), depending on field of application. Nevertheless, the algorithms involved pose others challenges of computation speed, power and system design complexity. Specifically, H.26x which is used to represent H.263, H.264 and H.265 codecs involve intensive computation operations such as spatial-frequency transformations, motion predictions (MP) and MEC (Battista et al., 2022).The overall goal is to reduce spatial redundancies from within intra predicted frames and temporal redundancies between successive frames. MEC only accounts for about 80 percent of total computation power during encoding process (Chatterjee & Chakrabarti, 2011). MEC coupled with RDO, leads to high computation costs in terms of time, power and system design complexity. Thus, in this paper, we attempt to reduce the overall time taken by RDO and MEC to improve encoding and decoding rates by using DWTIP and DWTMEC.

The rest of this paper is organised as follows: Section 2 introduces a brief review of intra prediction, inter predictions and related work. Section 3 highlights the platforms used. Section 4 explains proposed DWT approaches. Section 5 presents the results and in section 6, we draw conclusions. We also highlight further research work that we envisage.

2. A review on intra prediction, inter predictions and related work

All H.26x video codecs are coded by a group of pictures (GOP) comprised of intra predicted frames (I-frames) and inter predicted frames (P and B frames). This is depicted in fig.2. In I-frames, macroblocks (MBs) and sub-block (SB) are predicated without any reference frame. In P frames, some MBs and SBs are predicted from reference frame while others are not. In B frames, MBs and SBs are predicted using bi-directional modes (Richardson., 2003). Because of accumulated errors arising from inter prediction and block mismatches, I frames are inserted frequently along the video frames and are the anchor frames. I frames require no motion estimation to generate them. The number and frequency of I-frames insertion depends on a specific video sequence dynamics and settings. For example, the news

video sequence recoded in newsroom is encoded with less I frames compared with a moving bus video sequence. This is due to differences motion vectors involved.

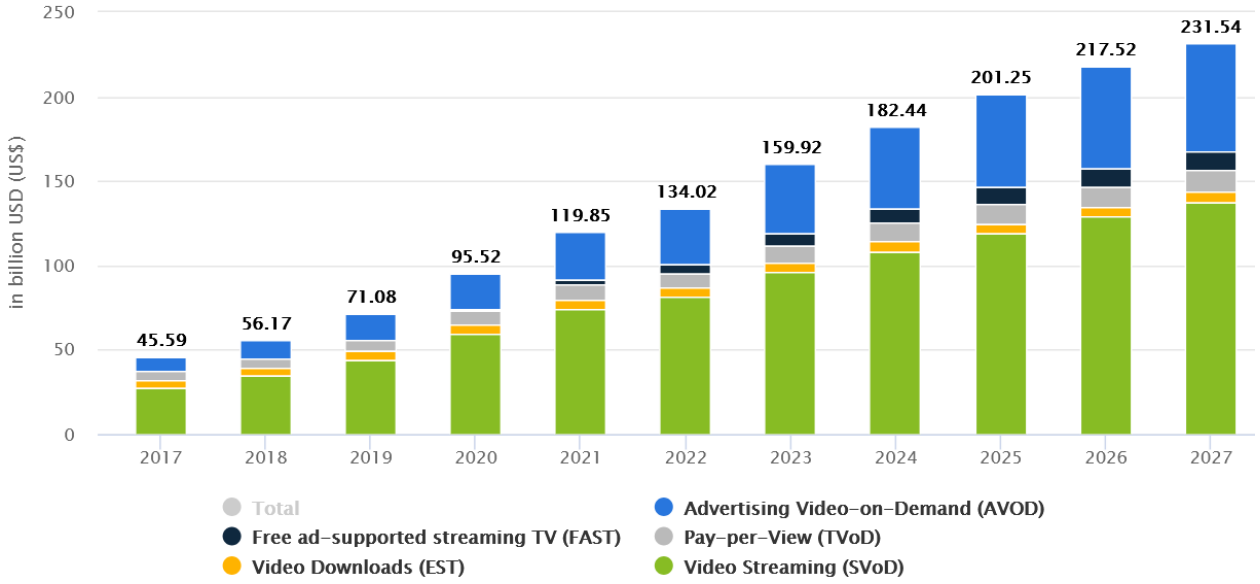

Fig. 1. Revenue by market from Video content (Statista Market Insights, 2024)

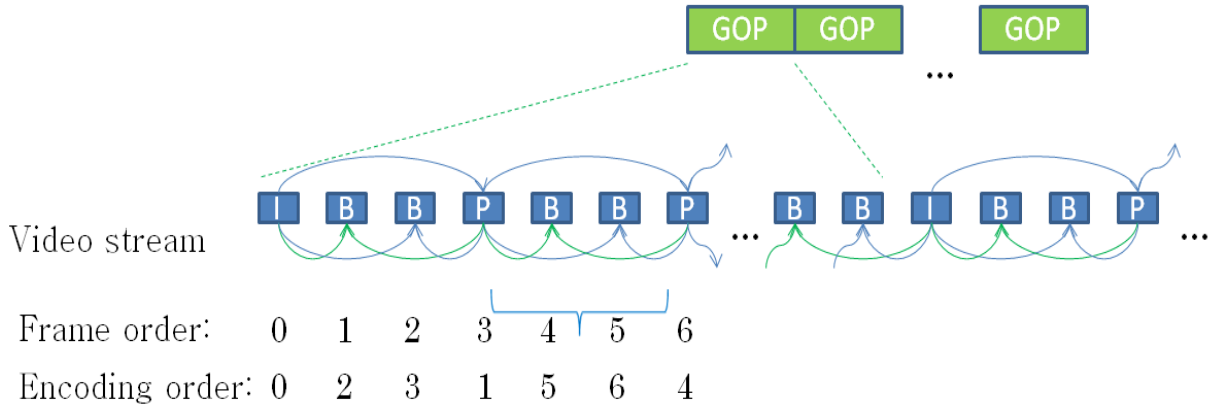

Fig. 2. A representation of GOP in H.26x video codecs (Nabeel & Al-Jammas, 2022)

2.1 Intra-prediction and RDO complexity

For proper analysis of RDO complexity, we based on Advanced Video Coding (AVC) which is still enjoyed and deployed in many areas such as multi view video coding (Jiang & Nooshabadi, 2016). Each frame of AVC is split into MBs of 16x16 pixels down to SBs of 4x4 pixels. In AVC, intra prediction is done by interpolation of neighbouring previously coded pixels (Richardson., 2003); AVC has a total of 17 intra prediction modes. These are 9 for 4x4 SBs, 4 for 8x8 SBs and 4 for 16x16 MBs (Richardson., 2003).The decision of the best prediction mode requires blind and exhaustive evaluation of all modes and leads to RDO complexity–challenge that we attempt to mitigate. In MB-based CODECs, RDO involved is computationally intensive and requires prior knowledge of distortion rate (D) and bit rate generated (R)

during intra prediction (Kwan-Jung O. and Yo-Sung H., 2005). To optimise such competing parameters, Lagrange optimisation cost function (*C*), depicted in equation 1 is used (Kwan-Jung O. and Yo-Sung H., 2005).

$$C(I_{MODE}) = D(I_{MODE}|QP, \lambda_{MODE}) + \lambda_{MODE} R(I_{MODE}|QP, \lambda_{MODE}) \quad (1)$$

Where *QP* is the quantisation parameter and λ is Lagrange multiplier, and I_{MODE} means I-frame coded with intra prediction mode at associated levels of *D*, *QP*, R and λ (Kwan-Jung O. and Yo-Sung H., 2005).If high profile of AVC is enabled in configuration file, RDO has to check all possible combinations (Lee et al., 2009). The number of mode combinations for luma and chroma components in an MB or SB is given by C8× (L4×16 + L16) (Richardson., 2003), (Lee et al., 2009).Where, C8, L4, and L16 represent the number of modes for chroma prediction, 4×4 luma prediction and 16×16 prediction, respectively. This implies that for a single MB or a SB, it has to perform 4× (9×16 + 4) = 592 different RDO computations in order to attain the best RDO mode (Jun Sung Park, 2006);(Miličević et al., 2012). If the 8×8 luma prediction of AVC is included, the number of mode combinations is C8× (L4×16 + L8×4 + L16) =4× (9×16 + 9×4 + 4) = 736 (Jun Sung Park, 2006);(Miličević et al., 2012), The best mode is the one having the minimum rate-distortion (RD) cost. In order to compute RD cost for each mode, the same operation of forward and inverse transform/quantization and entropy coding is repetitively performed. These exhaustive and repetitive computations explain the reason behind high complexity of conventional RDO (Kwan-Jung O. and Yo-Sung, 2005); (Sarwer & Po, 2007).

2.2 Inter-prediction and conventional MB and SB matching

A video sequence progressively has temporal redundancy. This is because two consecutive frames are often similar, especially in less moving objects and backgrounds where variations occur only due to object movement, illumination and camera movements. MEC techniques are used to reduce this redundancy –temporal redundancy. The well-known technique is Block Matching Algorithms (BMA) (Choudhury & Saikia, 2015); (Verma et al., 2013); (Casey, 2008).These techniques have been adopted by H.26x standard video codecs. Video frames coded from best matches using BMA are regarded as inter predicted frames, contrary to intra prediction where the coded frames are predicted from within the same frame. Inter prediction involves both ME and MC, and most of literature combine the two phenomena as MEC. ME of a MB involve finding a sample MB from a reference frame that closely matches the current MB (Richardson., 2003).An area in the reference frame centred on the current MB is searched and the 'best match' selected. The selected 'best' matching region in the reference frame is subtracted from the current macro block to produce a residual macro block (Richardson., 2003).All the predicted MBs are put together to reconstruct the final frame upon decoding. In fig.3(a) and fig.3(b), we present the two successive frames and the residue frame, after subtraction as well as associated motion vectors attained during the sample testing. The residue frame is encoded and sent jointly with a motion vector field relating the location of the best matching region relative to the current macro block position. It is essential to use a decoded residual frame to reconstruct the macroblock in order to make sure that encoder and decoder use same reference frame for MC.

In conventional MEC, a number of BMA have been in use for different codecs (Choudhury & Saikia, 2015); (Verma et al., 2013); (Casey, 2008). The most common ones are: Exhaustive Search (ES), Three Step Search (TSS), New Three Step Search (NTSS), Simple and Efficient Search (SES), Four Step Search (4SS), Diamond Search (DS), Adaptive Rood Pattern Search (ARPS), One-at-a-time Search Algorithm (OTA), cross search algorithm (CSA). Based on previous work by number of researchers, Adaptive Rood Pattern Search (ARPS), have most of the times shown less search points within search

window. We therefore base on these qualities, and compare ARPS termed as conventional MEC and DWTMEC using ARPS (Choudhury & Saikia, 2015); (Verma et al., 2013); (Casey, 2008).

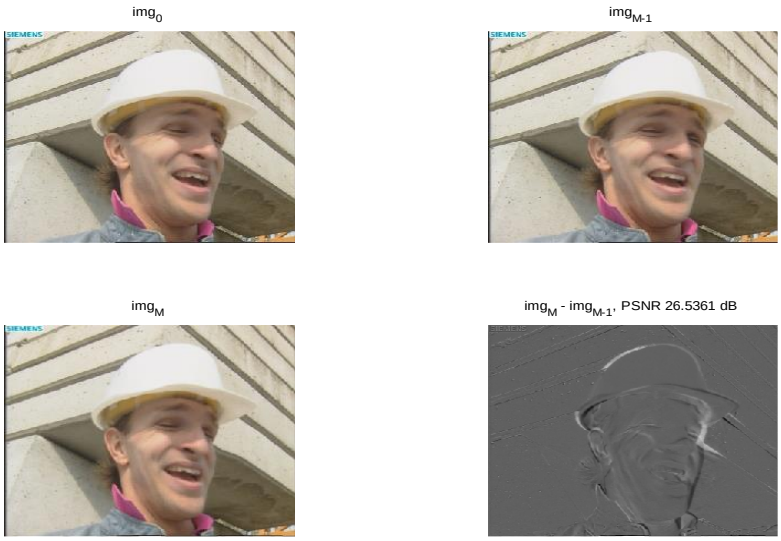

Fig. 3. (a) Motion compensation

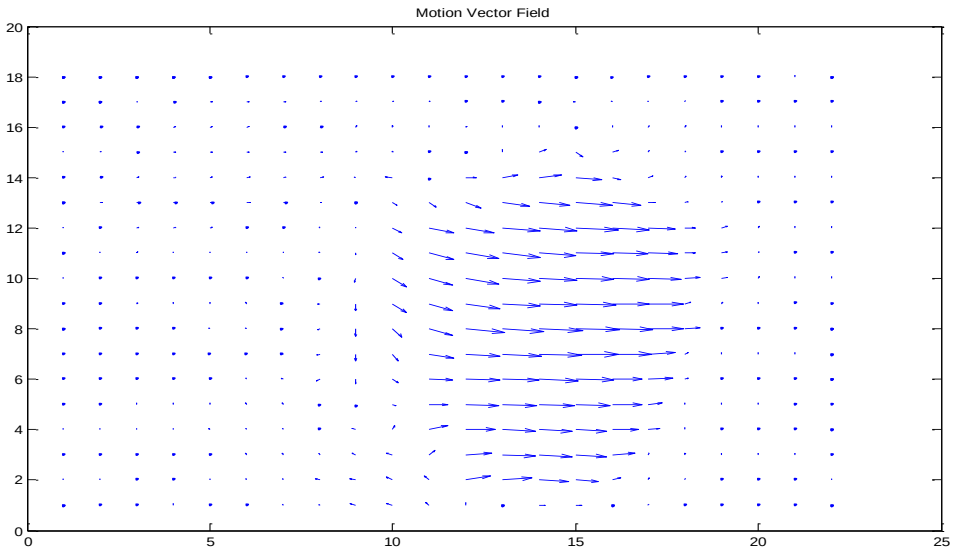

Fig. 3. (b) Motion vector field

3.Methodology

The JM16 software for Windows as downloaded, including associated configuration files. For consistency in results one video sequence was used. RDO complexity is first analysed using RDO-ON and RDO-OFF configuration modes. A DWT computation function is configured to be called with in the main software program. DWTIP and DWMEC output files are then compared with output files without using DWT. The comparison metrics used are the objective metric used in video coding.

3.1 Experiment, Tools and Hardware Platforms used

Experiments and analysis were based on JM16 reference software (Suehring, 2018). CodecVisa 4.38 (Codecian Co. Ltd, 2017) for Windows was used to analyse real-time video sequences approved to be decoded using H.26x principles (Arizona State universty, 2012);(Xiph, 2015). Carphone video sequence was chosen due its small size for experiments and analysis. Analysis was done on computer with Intel(R) core TM i3-2330M CPU@2.20GHz , with a 4.00GB installed RAM. The system also used the 64-bit operating system.

3.2 RDO and IP complexity analysis

In order to analyse RDO complexity that reveals quest for fast mode prediction techniques, JM16 reference codec software was used (Suehring, 2018).By opening a configuration file, RDO was enabled and disabled while recording encoding time. 300 frames of foreman video test sample in Quarter Common Intermediate Format (QCIF) format were encoded for different target bit rate (Suehring, 2018).Using recorded time taken and target bitrates, the two graphs for RDO-OFF and RDO-ON modes are plotted. Fig. 4 indicates that "RDO on" mode takes much longer than "RDO Off" mode during encoding process. From fig. 4, it is indicated that at an encoding rate of 250kbits/sec, when the RDO is on, time taken was more than 120 seconds. On the other hand, when RDO is OFF, the time taken was less than 20 seconds. But the reduction of time comes with big video quality degradation. The aim of our work is to reduce encoding time with negligible degradation in video quality.

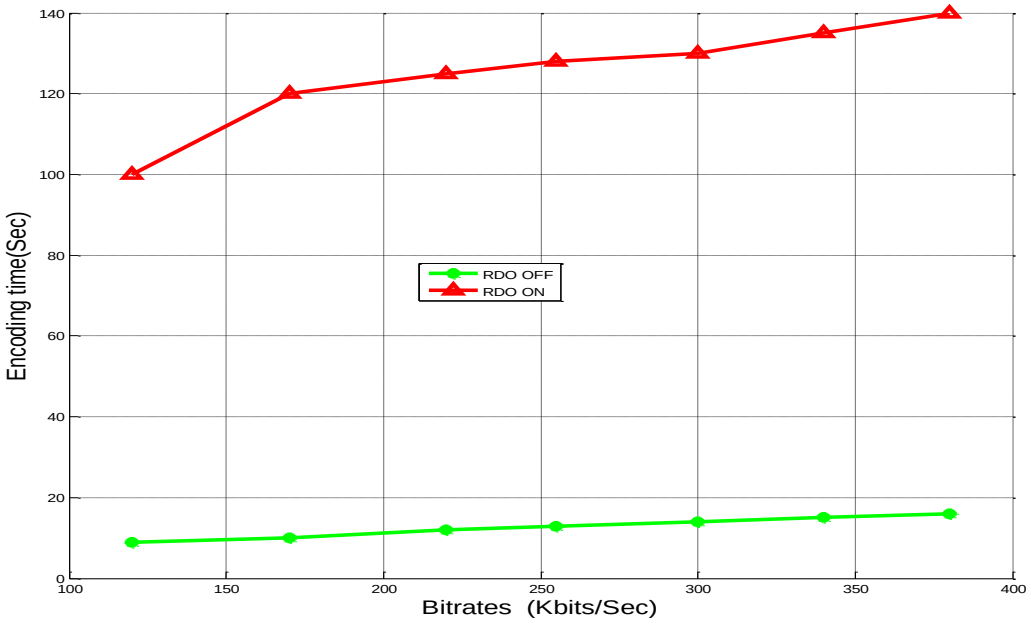

Fig. 4. RDO Complexity analysis using JM16 software

https://dx.doi.org/10.4314/rjeste.v7i1.2

3.3 Spatial Redundancy, MB Size and Intra Prediction mode decision

Using samples of video sequences and AVC codec, this section presents the relationship between the MB spatial homogeneity, size of MBs and prediction modes, hence laying a foundation and justification for proposed DWTIP approach. Fig. 5 (a) and fig. 5 (b) depict important features that form a basis for DWT as a tool for fast prediction modes. The left pane indicates bit content of the candidate MB and memory addresses required. The middle pane indicates the video frame and MB under analysis marked. The right pane, the upper part indicates general information of the MB while the lower part shows the prediction modes associated with the candidate MB in the current frame.

Using CodecVisa tools, frame 0 which is the first frame in the whole sequence is analysed after coding and decoding. It can be shown that MB0 at location (0, 0) has the largest "int bits" equivalent to 677. This means that it lies in region of pixels with less homogeneity levels and has many details to be transmitted compared to the rest of MBs. The details to be transmitted are indicated by the bit content in the left pane. Based on H.26x coding principles, it has to be coded using the smallest SB type of 4x4. It is also intra predicted using mode 0, mode 2, mode 3, mode 5 and mode 8 (Richardson., 2003). On the contrary, fig. 5 (b) shows MB88 located at position (0, 8) with "int bit" of 45. This means that it lies in the region of pixels with more homogeneity levels than MB0 (0, 0). It is therefore intra predicted by MB type of 16x16. It is the largest MB type in AVC, which used only mode 0 (vertical).

Analytically, it becomes clear that MBs or SBs in less spatially homogeneous regions of a video frame, as shown in fig. 5 (a) inherently has much bit content. They require more prediction modes within the smallest SBs. On the contrary, MBs and SBs in more homogeneous regions, as indicated in fig. 5 (b) shows less bit content and requires only DC (Mode 0) prediction mode with 16x16 MB, which is the largest in AVC Codec. This leads to a conclusion that regions of high spatial homogeneity levels have less bit content and are predicted using bigger micro blocks and vice versa. We based on this conclusion to propose DWTIP in the next section. The frame under test in this section is "frame 0", I-frame which is always the first frame in H.26x codecs. Thus, no motion vectors indicated.

4 Proposed DWT approaches

In H.26x, encoding techniques are not rigidly standardised. Normally, only the bit stream syntax and the decoding process are standardised (Suehring, 2018).Other components of a video transmission such as pre-processing, encoding, loss/error recovery, and post-processing are intentionally left out for designers (Ohm et al., 2012). This gives the designers ample degree of flexibility to look for new algorithms and rooms for improvements.

4.1 Proposed DWTIP algorithm for MB based CODECs

Based on the mentioned degree of flexibility in codec designs, our work proposes a DWTIP algorithm presented in fig. 6. (a) and fig. 6. (b) for a relaxed and optimised RDO. It aims at performing quick partitioning and obtaining optimum MB size prior to RDO, and by so doing, we restricted the known prediction modes to specific MB size in a given MB region. Unlike in conventional RDO, which evaluates every mode before partitioning the MBs, DWTIP avoids blind evaluation for different possible modes, hence RDO relaxation and fast prediction process. There are scholarly works available using DWT for video or image compression but with different approaches, objectives and application. Iwasokun and Olaoye (Iwasokun & Olaoye, 2021) used both DWT and Discrete Cosine Transform (DCT) for compression of but used a different method of testing results where the mean square error (MSE) is plotted

against the video frames (Iwasokun & Olaoye, 2021). In a very specific task of video compression, (Kumar et al., 2014), used the DWT for Motion Estimation(ME), as one task of entire video compression process. Due to computational intensiveness of such functions , hardware chips often called accelerators have been designed (Farghaly & Ismail, 2020). With such an approach, the hardware chip is called as subroutine to perform the DWT function instead of called a function code as subroutine in the main program. Based on the DWT capability in the spatial decomposition of video frame, the DWT used in this study DWT spatially transforms a given 2D image into four sub bands LL, LH, HL and HH at each transform level (Yang & Seo, 2023).These sub bands carry four (4) details about the transformed video frame. These are: approximation coefficients, horizontal details, vertical details and diagonal details respectively (Yang & Seo, 2023).Technically, 2D DWT spatially operates on a 2D video frame as a progressive low, high filter on both rows and columns respectively. By transforming LL sub band for "L" levels successively, quick MB partitioning can be achieved based on regional spatial similarity in a specific frame. This is contrary to conventional approach, where RDO is done to evaluate all possible prediction modes before obtaining the best prediction mode. The proposed algorithm in fig.6, first determines pixel similarly level *(η)* in a given block.

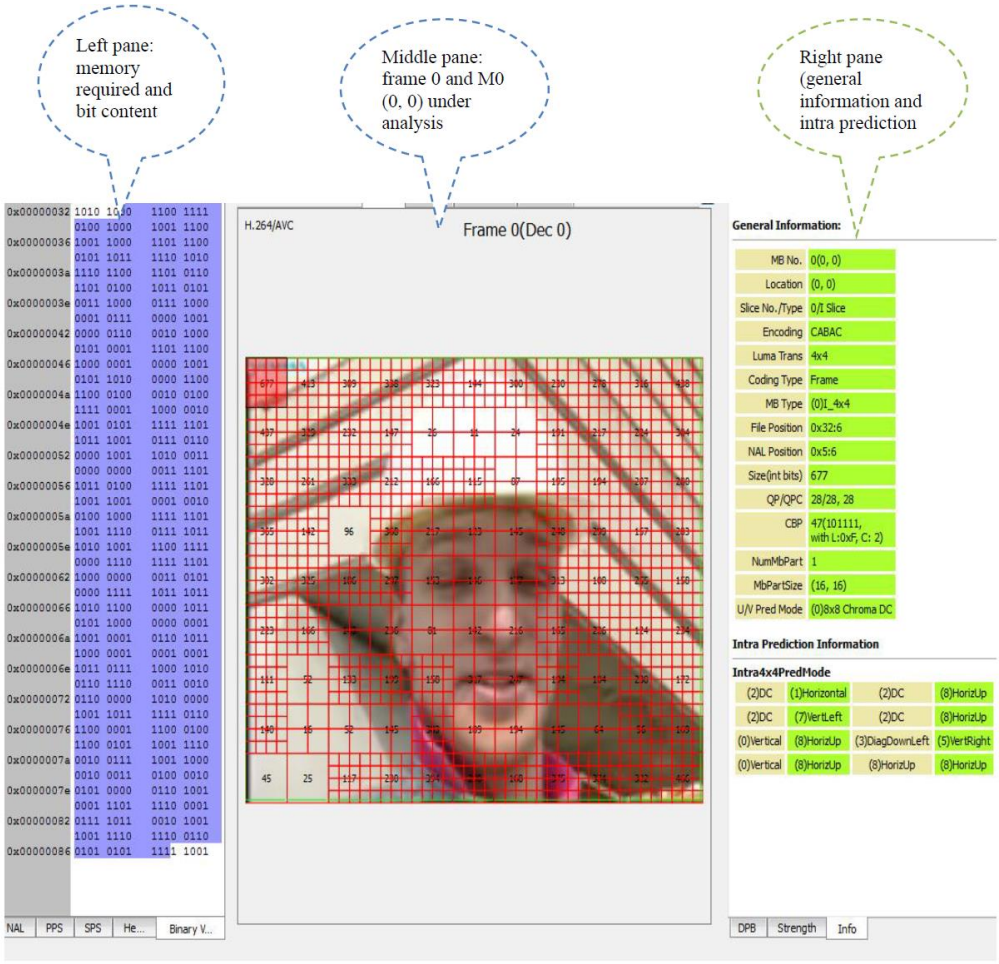

Fig. 5. (a) Less homogeneous region and associated smallest MB

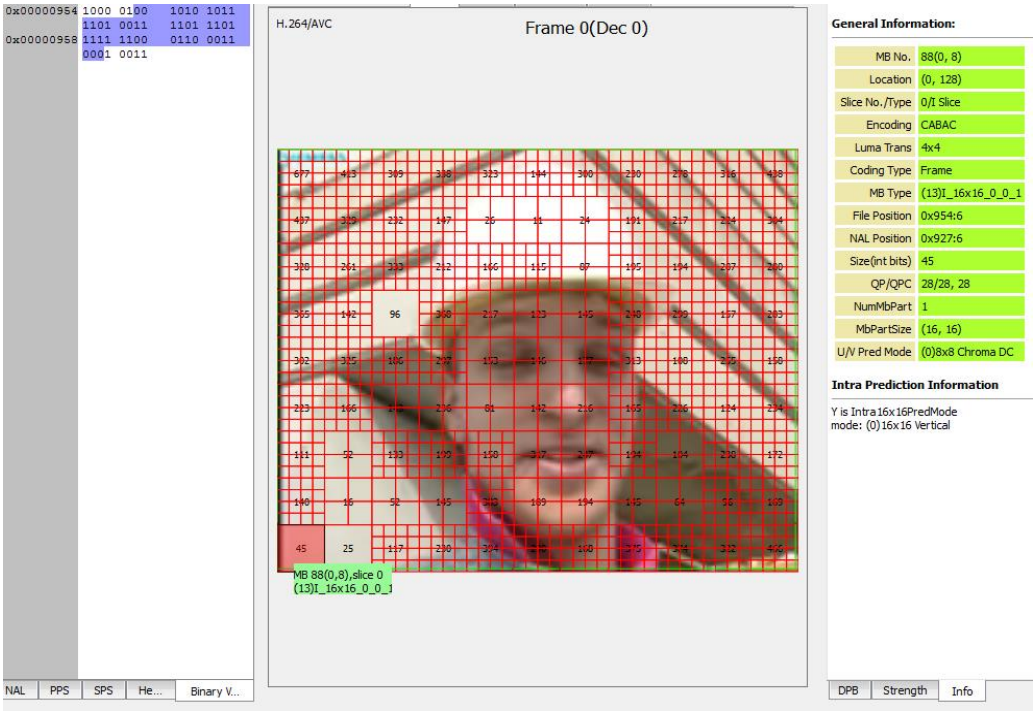

Fig. 5. (b) More homogeneous region and associated largest MB
Fig. 5. MB homogeneity analysis

We determined the size of an MB in a specific region of a video frame based on *'η'*. For each sub-band, the pixel similarity level *'η'* was obtained from two important parameters associated with DWT. These are DWT transformation coefficient energies and sub-band variance. After transforming a video frame (I_B) to be predicted by intra prediction mode and obtaining a *BxB* coefficient blocks, the statistical energy per block is obtained by equation 2. Where *B* is maximum block size detected based on level of regional homogeneity. For a 4x4 SB, both *i* and *j* maximum values becomes 4. If for example an 8x16 MB was to be encoded, the maximum values of *i* and *j* would be 8 and 16 respectively. In such an order, the upper limit value of *B* for a 16x16 MB is 16.

$$E_B = \frac{1}{B}\sum_{i}^{B}\sum_{j}^{B} C_B^2\,(i,j) \qquad (2)$$

Where C_B and E_B are the block's highest coefficient and block energy, respectively. Similarly, the sub-band variance is given by equation 3.

$$\delta_B^2 = \frac{1}{BxB}\sum\nolimits_{i}^{B}\sum\nolimits_{J}^{B}(I_B(i,j) - \mu_B)^2 \qquad (3)$$

Where μ_B is the mean of I_B(i, j), and given by equation 4.

$$\mu_B = \frac{1}{BxB}\sum\nolimits_{i}^{B}\sum\nolimits_{J}^{B} I_B(i,j) \qquad (4)$$

Hence, the required pixel similarity level metric, is obtained as quotient of E_B and δ_B^2.

https://dx.doi.org/10.4314/rjeste.v7i1.2

$$\eta = \frac{E_B}{\delta_B^2} \qquad (5)$$

Comparing the homogeneity levels, η , with the preset threshold levels, π , MB and SB size is predetermined, and partitioning is done prior to Intra prediction . This approach is depicted in fig.6.

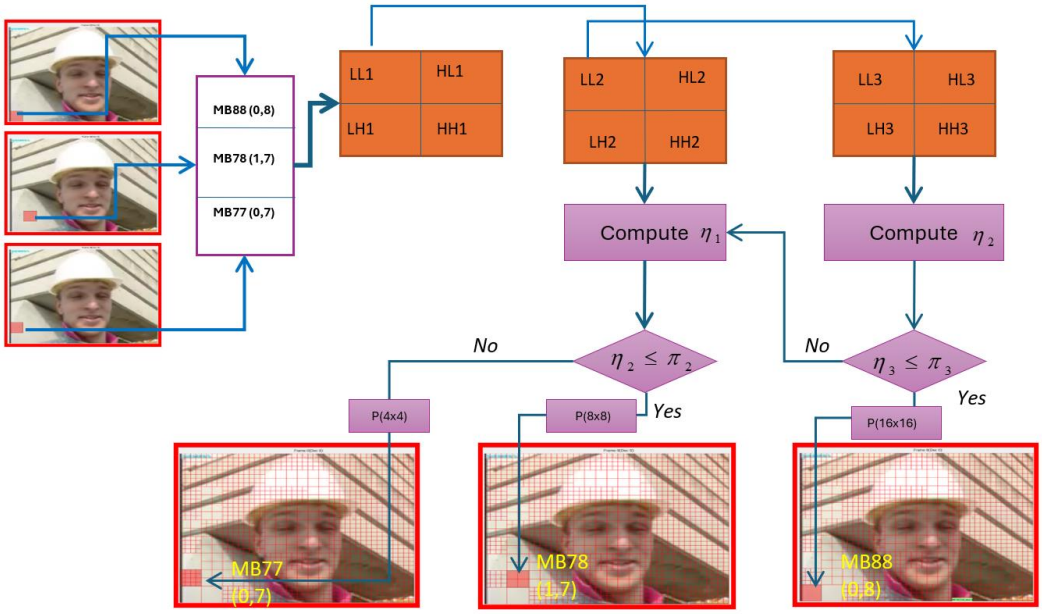

Fig. 6. (a) Sample of MBs partitioned into SBs and their location in entire video frames. MB77 has 16 SBs of 4x4 pixels, MB78 has 4 SBs of 8x8 pixels while MB88 has 1 MB with 16x16 pixels.

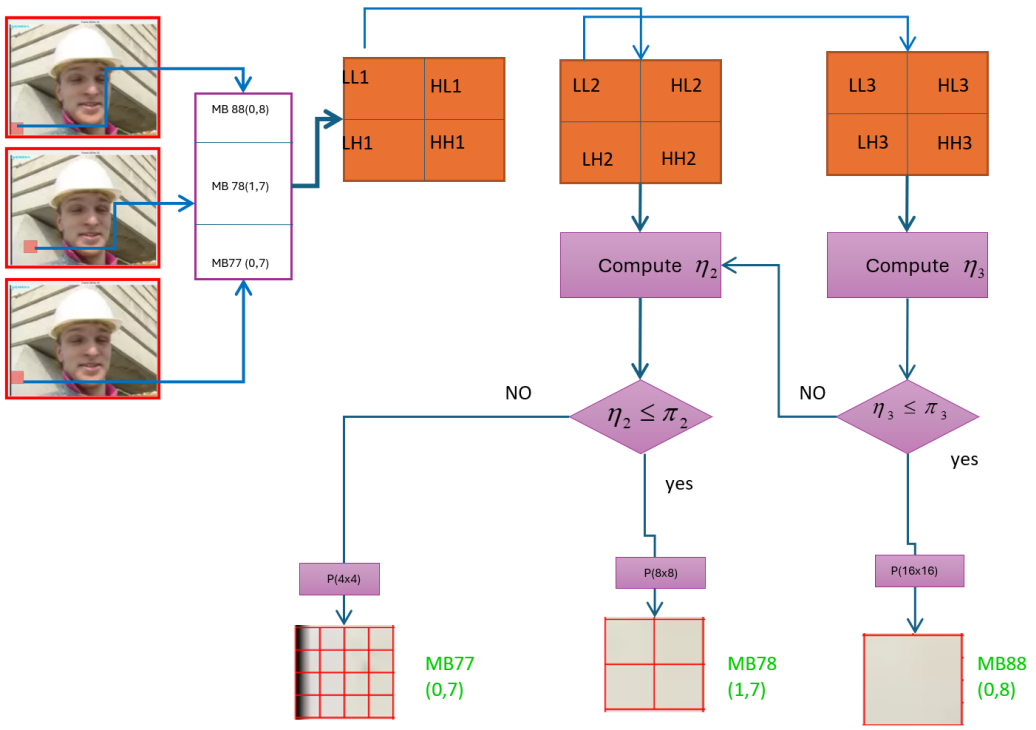

Fig. 6. (b) Samples of partitioned MBs and associated SBs enlarged for clarity.

Fig. 6. Proposed DWTIP algorithm

Fig. 6 shows three sampled MBs out of 99 MBs that makeup AVC video frame. MB77, MB78 and MB88 are located at (0, 7), (1, 7) and (0, 88) respectively. Using DWTIP approach, it was found that that the spatial similarity levels in these three MBs are different. As result, M88 exhibits more spatial homogeneity levels, followed by MB 78 and then MB77. Based on DWTIP, they are spilt into one 16x16, four 8x8 and sixteen 4x4 MBs respectively. Because a specific MB size is associated with fixed number of prediction modes (Richardson., 2003), the proposed approach avoids exhaustive and blind evaluation of all modes as mentioned earlier. It obtains fast and optimum partition modes by only evaluating specific modes associated with partitioned MB and SB size. Thus, RDO is considerably relaxed and encoding takes less time. The comparative results are presented in fig.10.

4.1 Proposed DWTMEC

In the proposed DWTMEC, DWT is performed on Reference frames prior to ME. Then ME is performed on DWT approximations sub bands of both frames. Because approximation sub band contains most of visual information to Human visual systems (HVS), by performing ME on wavelet coefficients, we are able to get estimation quickly and then motion vectors calculated directly using equations 6 and 7. By exploiting the hierarchical relationships between DWT sub bands, Motion vectors (MVs) in the rest of detail sub bands can be obtained from geometrical translations and the same hierarchical relationship. In this paper, we used three decomposition levels. We used (L=3), a three-levels DWT. Let n {1, 2, 3} be the number levels in L=3 and sub four associated with the transformation be represented by s= {1, 2, 3, 4}. From fig. 7, the Motion Vector (MV) from any sub band, s can be calculated by equation 6.

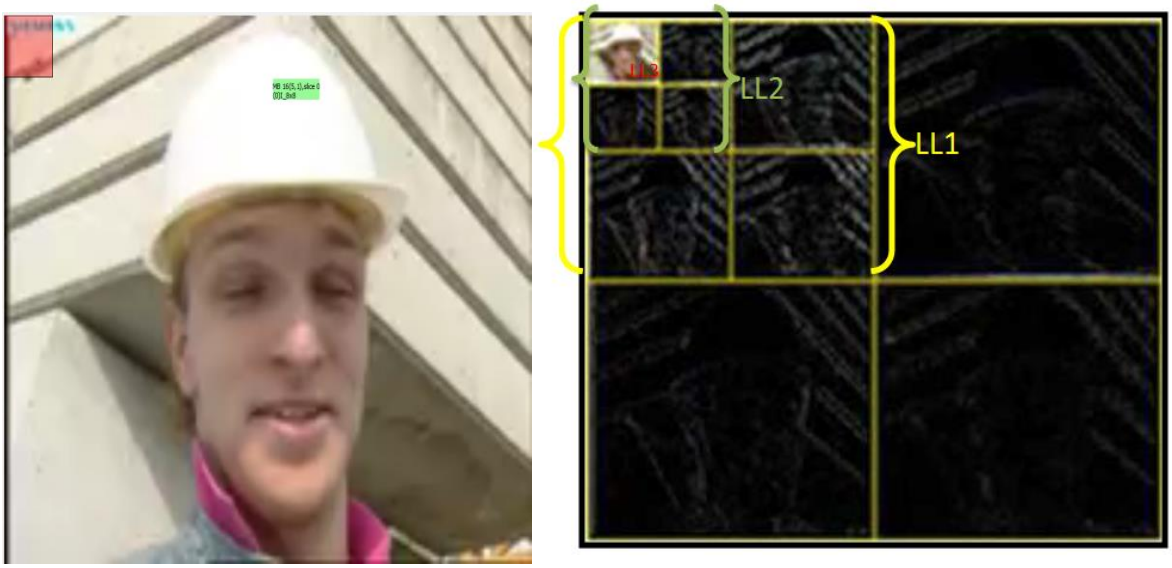

Fig.7. (a) Foreman video, frame 2 sampled prior to DWT **Fig.7. (b)** DWT transformation on foreman video, frame 2, L=3.

Fig. 7. DWT transformation and sub band details (frame 2 in foreman clip)

$$MV_{n,s} = 2^{l-n} * D_{L1}(i,j) + \delta_{n,s} \qquad (6)$$

https://dx.doi.org/10.4314/rjeste.v7i1.2

$$\begin{pmatrix} If\ n = L, \delta_{n,s} = 0, MV_{n,s} = D_{L1} \\ Else, MV_{n,s} = 2^{l-n} * D_{L1}(i,j) + \delta_{n,s} \end{pmatrix} \qquad (7)$$

The translational displacement in each sub band can be obtained by doubling the displacement of the matching sub band block in the lower DWT level. By adding correcting factor $\delta_{n,s}$, we correct the estimation error as given in equations 6 and 7 and presented in the fig. 8 and fig. 9.Estimating the motion and displacements in the approximation sub band equips us with three (3) major advantages:

1) Approximation sub band inherently contains the important information bit content and for HVS. This maintains PSNR values for decoded video sequence in acceptable range for most of application standards.
2) The search area is reduced compared to the original frame. This reduces the search points and computation time.
3) Compression ratios are improved. Thus, mitigating bandwidth and storage capacity constraints.

4.1.1 Evaluating MEC and Associated Errors and video quality using Video Quality Metrics

Matching and estimating motions in one macroblock DWT coefficients with others, depends much on cost function output. The MB with least cost is selected as the one that the closest to current block. In video codecs, there are different cost functions. The most popular and less computationally expensive is Mean Absolute Difference (MAD), and commonly used cost function is the Mean Squared Error (MSE) shown in equation 8 (Verma et al., 2013)

$$MSE = \frac{1}{NM}\sum_{i=o}^{N-1}\sum_{j=0}^{M-1}(C_{ij} - R_{ij})^2 \qquad (8)$$

For the case of Square MB sizes as we used in H.26x codecs, equation (8) is modified to Equation (9)

$$MSE = \frac{1}{N^2}\sum_{i=o}^{N-1}\sum_{j=0}^{N-1}(C_{ij} - R_{ij})^2 \qquad (9)$$

Where N is the size of the macroblock, C_{ij} and R_{ij} are the coefficients being compared in current MB and reference MB, respectively. The common objective metric of video quality after removal of both temporal and spatial redundancies (compression) is the Peak-Signal-to-Noise Ratio (PSNR) (Verma et al., 2013).

$$PSNR = 10log\left[\frac{(2^{np} - 1)}{MSE}\right] \qquad (10)$$

In this equation, "*np*" is the number of bits in a pixel and PSNR will be used in our results analysis and comparison as the measure of compressed video distortion. Fig.9 depicts a reconstructed final video frame from best MB searches and motion vectors involved during inter predictions. With close look at the left pane, it can be seen that information content and associated memory requirement have been tremendously reduced which is one of the main purposes of video codecs. Motion vectors involved in inter predictions are also depicted in the frame as well as their location in the rightmost pane. Another main concern is encoding/decoding time reduction. In the next section, we present results showing number of search points which to a great extent affects the overall encoding/decoding time. We also depict

the quality of the reconstructed video frames using the well-known objective video quality metric known as the peak signal-to-noise ratio PSNR.

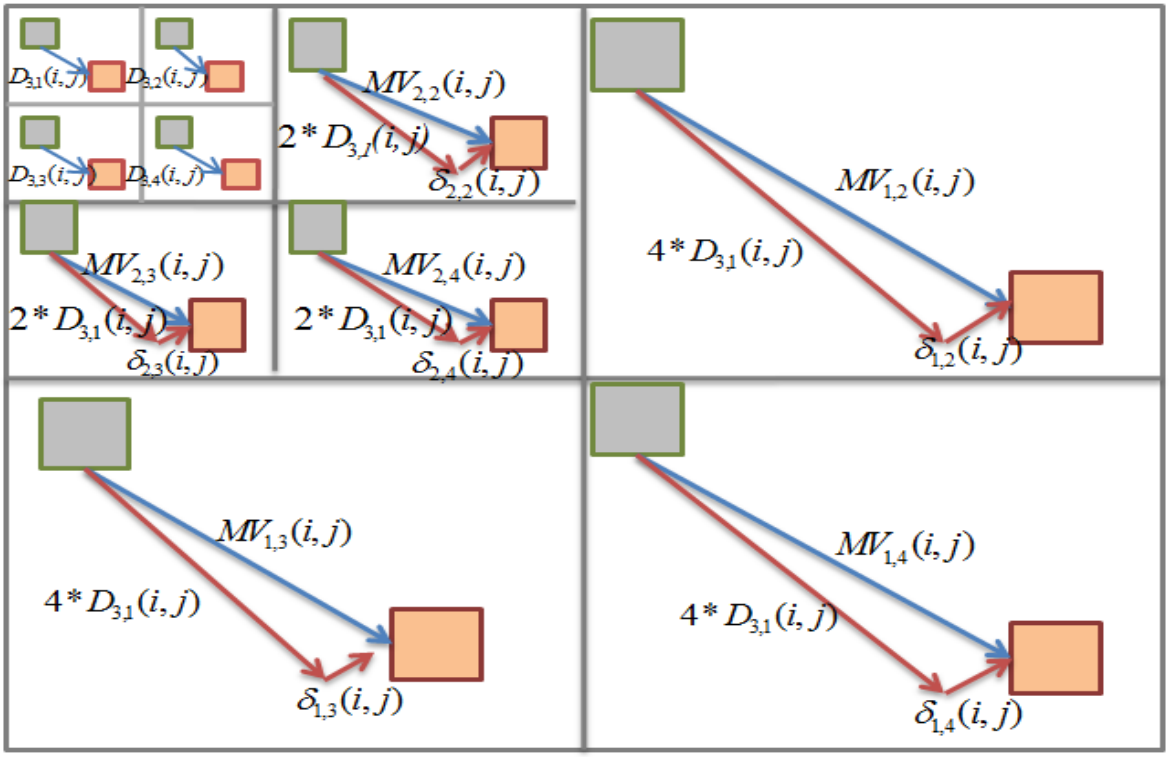

Fig.8. DWTMEC Illustration

Fig. 9. Final reconstructed frame and associated motion vectors

https://dx.doi.org/10.4314/rjeste.v7i1.2

5. Results Analysis

Fig.10 shows that RDO OFF mode consumes less time compared to RDO ON mode. But RDO Off mode affects PSNR to great extent. On the other hand, Using RDO ON with the proposed DWTIP approach mitigates RDO complexity as well as keeping PSNR in acceptable ranges.

Fig.11-14 show and compare the number of search points and associated PSNR using both conventional MEC (conventional ARPS) and DWTMEC (DWT assisted). It can be seen that DWTMEC reduces search points per frame without affecting much the PSNR. During experiment, it was also noted that video sequences that involved more movements, such as bus sequence, exhibited more PSNR degradation compared to sequences with less movements such as news sequences. This is because less similar consecutive frames in rapidly changing background are likely to cause more MB mismatches compared highly correlated consecutive frames.

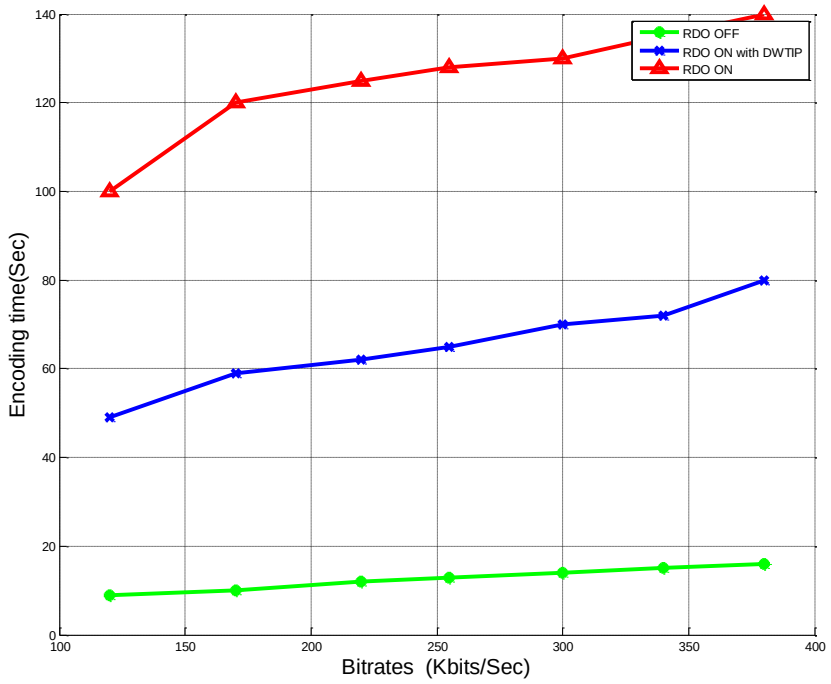

Fig.10. RDO complexity comparison

https://dx.doi.org/10.4314/rjeste.v7i1.2

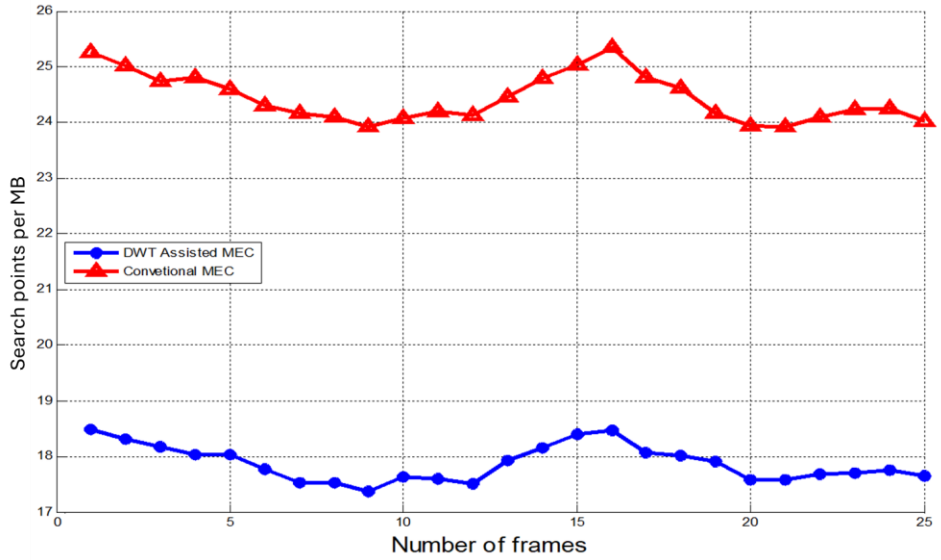

Fig. 11. (a) Search points per MB vs. progressive number of video frames

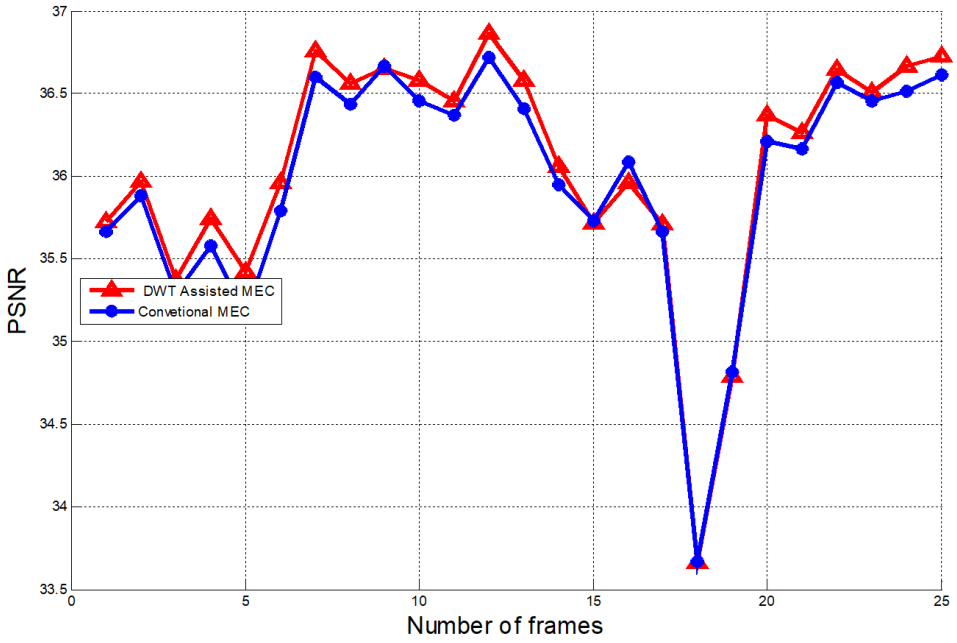

Fig.11. (b) PSNR comparisons

Fig.11.(b) shows that there is negligible video quality degradation using PSNR metric. On the other hand, Fig.11.(a) indicates that DWTMEC approach reduced search points per MB tremendously, thus an indication of improved coding and decoding rates.

https://dx.doi.org/10.4314/rjeste.v7i1.2

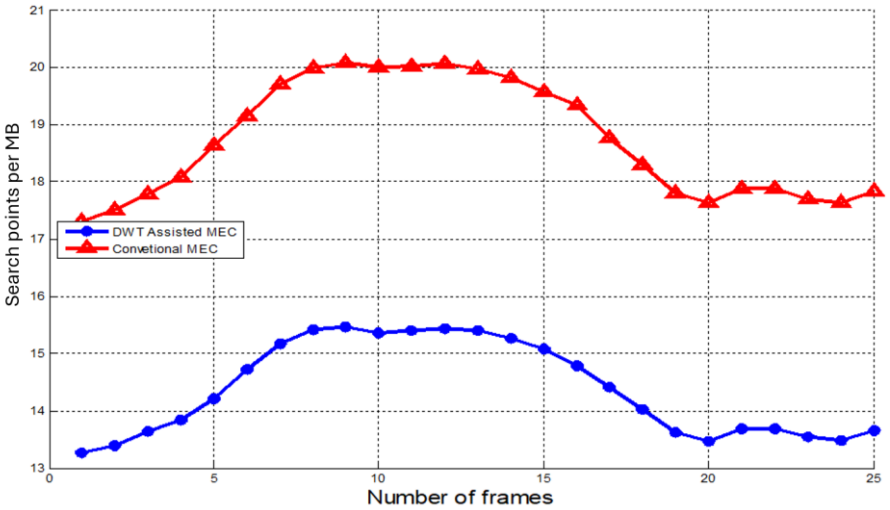

Fig.12. (a) Search points for foreman sequence

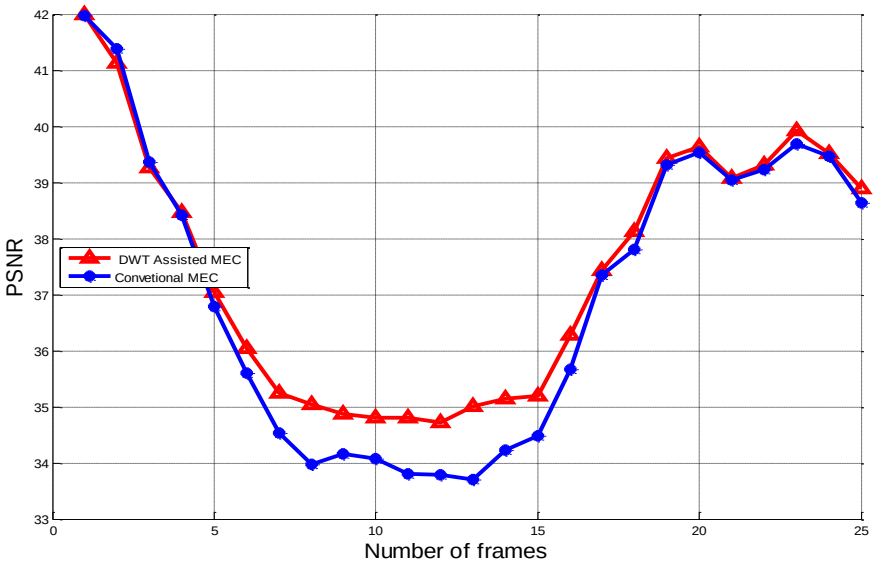

Fig.12. (b) PSNR foreman sequence

Fig.12.(b) shows that the video quality degradation between **frame 8 to frame 15** using PSNR metric. However, it is still tolerable given the advantage of improvement in search point reduction as indicated in fig.12(a). The degradation resulted from the sudden change in the background of foreman video sequence.

https://dx.doi.org/10.4314/rjeste.v7i1.2

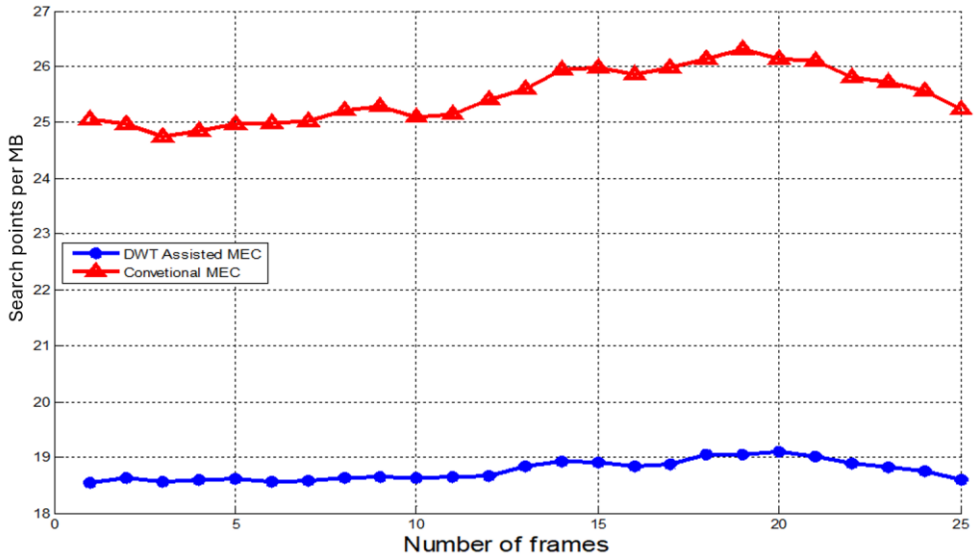

Fig.13.(a) Search points for bus sequence

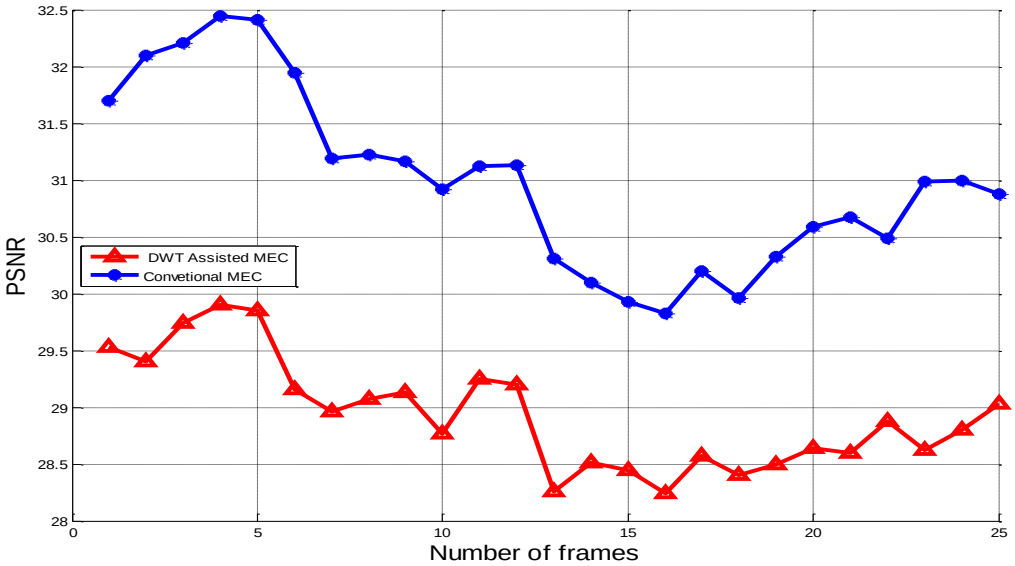

Fig.13. (b) PSNR

Fig .13. Search points and PSNR comparison for bus sequence

In fig.13. (b) the bus sequences depicted many mismatches between Micro blocks. This is because the bus direction and speed involved caused more rapid changes in the scene. But still the consistence in search point reduction holds as depicted in fig.13.(a).

https://dx.doi.org/10.4314/rjeste.v7i1.2

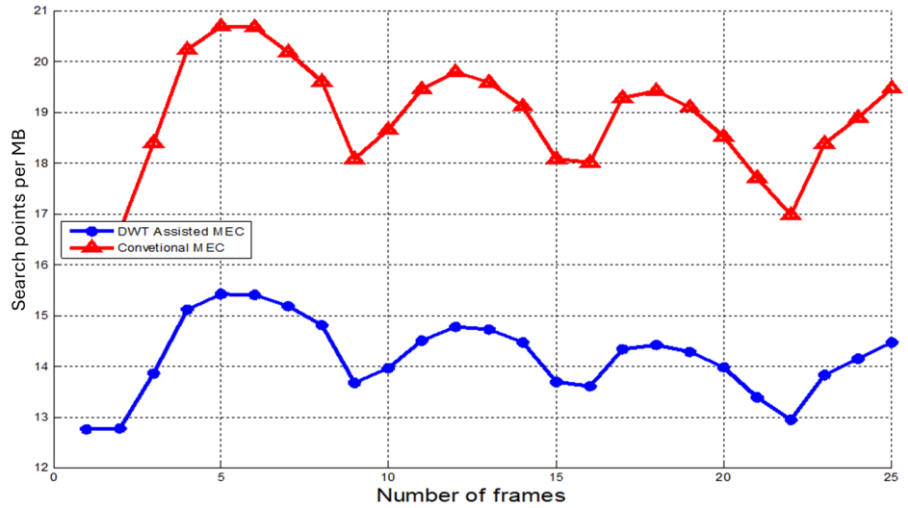

Fig.14.(a) Search points for car phone video sequence

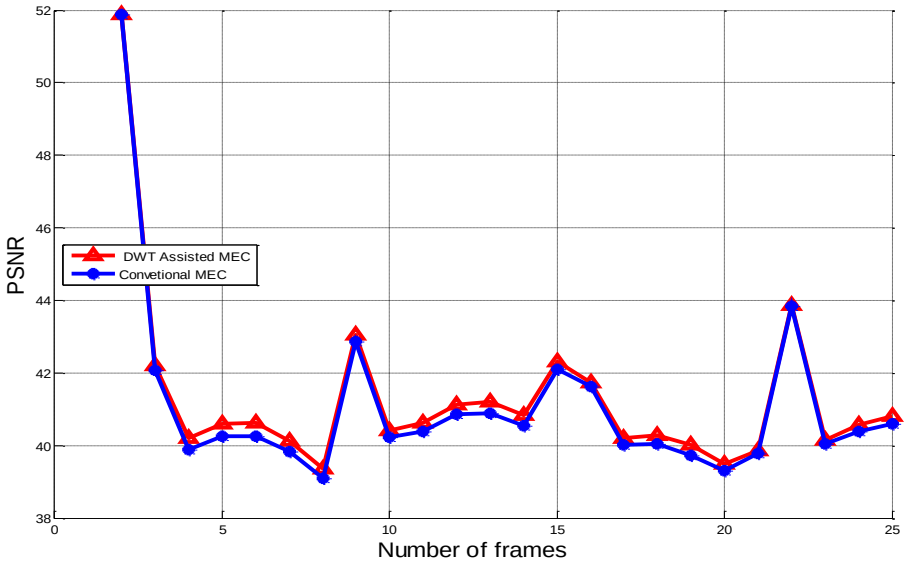

Fig.14. (b) PSNR

Fig.14. Search points and PSNR comparison for car phone video sequence

Fig.14.(b) shows a negligible deterioration in PSNR for car phone video sequence by objective metrics. On the other hand, fig.14.(a) shows a considerable reduction in search points per frame. It can be noticed that car phone video sequence shows some repetitive trends, this is due to repeating frames of the same features against its background.

6. Conclusion

The aim of this paper was to use DWT and its intrinsic features to mitigate both computational costs associated with RDO and MEC in H.26x video codecs. DWT assisted Intra prediction (DWTIP) led to faster selection of intra-prediction mode for a target region in video frame. By so doing, we avoided wastage of time for exhaustively testing all intra prediction modes. This was achieved by restricting specific modes for a given MB size based on their pixel similarity levels. It led to a more relaxed RDO without much degradation in video quality. DWT assisted MEC (DWTMEC) reduced searches per MBs

in entire video frame. This reduced time for MEC while maintaining PSNR in acceptable ranges. In fact, in some sequences where the scenes were not changing rapidly such as news and car phone video frame sequences, DWTMEC exhibited almost the same PSNR as conventional MEC. With both DWTIP and DWTMEC, the feedback loop (decoder) found in H.26x encoder can be optimised in terms of search points and time cost and be replicated to decoder end. Further research work envisages running the code on SOC FPGAs and designing an FPGA for DWT module to be called routinely by the main program instead of calling DWT subroutine software program. This is a modern approach for hardware-software codesign approach in system on chip (SoC) for computationally intensive designs.

7. References

Arizona State universty. (2012). *YUV Video Sequences*. http://trace.eas.asu.edu/yuv/

Battista, S., Meardi, G., Ferrara, S., Ciccarelli, L., Maurer, F., Conti, M., & Orcioni, S. (2022). Overview of the Low Complexity Enhancement Video Coding (LCEVC) Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(11), 7983–7995. https://doi.org/10.1109/TCSVT.2022.3182793

Bing, B. (2015). *Next-Generation Vide Coding and Streaming*. John Wiley & Sons, Ltd.

Casey, J. (2008). *An Investigation of Block Searching Algorithms for Video Frame An Investigation of Block Searching Algorithms for Video Frame Codecs* [Dublin Institute of Technology]. https://arrow.tudublin.ie/cgi/viewcontent.cgi?article=1000&context=schmuldistoth

Chatterjee, S. K., & Chakrabarti, I. (2011). A fast and low-power VLSI architecture for half-pixel motion estimation using two-step search algorithm for HDTV application. *IETE Journal of Research*, *57*(3), 263–270. https://doi.org/10.4103/0377-2063.83648

Choudhury, H. A., & Saikia, M. (2015). Block matching algorithms for motion estimation: A performance-based study. *Lecture Notes in Electrical Engineering*, *347*(October), 149–160. https://doi.org/10.1007/978-81-322-2464-8_12

Codecian Co. Ltd. (2017). *CodecVisa 4.38*. https://codecian.com/downloads.html

Farghaly, S. H., & Ismail, S. M. (2020). International Journal of Electronics and Communications (AEÜ) Floating-point discrete wavelet transform-based image compression on FPGA. *AEUE - International Journal of Electronics and Communications*, *124*, 153363. https://doi.org/10.1016/j.aeue.2020.153363

Hemant Joshi. (2015). *Digital media rise of on-demand content*. https://www2.deloitte.com/in/en/pages/technology-media-and-telecommunications/articles/digital-media-rise-of-on-demand-content.html

Iwasokun, G. B., & Olaoye, M. O. (2021). Discrete transformation technique for video compression. *Iran Journal of Computer Science*, *4*(4), 281–292. https://doi.org/10.1007/s42044-021-00085-3

Jiang, C., & Nooshabadi, S. (2016). Parallel multiview video coding exploiting group of pictures level parallelism. *IEEE Transactions on Parallel and Distributed Systems*, *27*(8), 2316–2328. https://doi.org/10.1109/TPDS.2015.2485993

Jun Sung Park, and H. J. S. (2006). Selective Intra Prediction Mode Decision for H.264/AVC Encoders. *World Academy of Science, Engineering and Technology*, *13*(May), 51–55.

Kumar, H., Rai, P., Sarma, D., & Thapa, G. (2014). Applicability of wavelet transform in Multi-Resolution Motion estimation technique. *International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2014*. https://doi.org/10.1109/ICRAIE.2014.6909211

Kwan-Jung O. and Yo-Sung H. (2005). Lecture Notes in Computer Science: Preface. In *Lecture Notes in Computer Science* (Vol. 3639).

Lee, S., Park, S., & Park, J. (2009). 270 MHz Full HD H . 264 / AVC High Profile Encoder with Shared Multibank Memory-Based Fast Motion Estimation. *ETRI Journal*, *31*(6), 784–794. https://doi.org/10.4218/etrij.09.1209.0007

Miličević, Z., Bojković, Z., & Rao, K. R. (2012). An approach to interactive multimedia systems through subjective video quality assessment in H.264/AVC standard. *WSEAS Transactions on Systems*, *11*(8), 305–314.

Nabeel, R., & Al-Jammas, M. H. (2022). The GOP Inter Prediction of H.264 AV\C. *Journal of King Saud University - Computer and Information Sciences*, *34*(1), 1345–1351. https://doi.org/10.1016/j.jksuci.2019.06.005

Ohm, J. R., Sullivan, G. J., Schwarz, H., Tan, T. K., & Wiegand, T. (2012). Comparison of the coding efficiency of video coding standards-including high efficiency video coding (HEVC). *IEEE Transactions on Circuits and Systems for Video Technology*, *22*(12), 1669–1684. https://doi.org/10.1109/TCSVT.2012.2221192

Richardson., I. E. G. (2003). *H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470869615

Sarwer, M. G., & Po, L. M. (2007). Bit rate estimation for cost function of 4×4 intra mode decision of H.264/AVC. *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, ICME 2007*, 1579–1582. https://doi.org/10.1109/icme.2007.4284966

Statista Market Insights. (2024). *Video-on-Demand - Worldwide*. https://www.statista.com/outlook/dmo/digital-media/video-on-demand/worldwide

Suehring, K. (2018). *H.264/AVC Software*. https://iphome.hhi.de/suehring/

Verma, V., Asst, S., & Ravi, P. (2013). Comparison And Implementation Of Block Matching Algorithms. *International Journal of Engineering Research and Applications*, *3*(4), 1202–1206.

Xiph. (2015). *Video Test Media*. https://media.xiph.org/video/derf/

Yang, D., & Seo, S. W. (2023). Discrete Wavelet Transform Meets Transformer: Unleashing the Full Potential of the Transformer for Visual Recognition. *IEEE Access*, *11*(September), 102430–102443. https://doi.org/10.1109/ACCESS.2023.3316144