Improved Motion Compensation Temporal Filtering (MCTF) for Scalable Video Coding (H.264/SVC) Standards

Ankur Vora, Krishna Chandramouli
School of Information Technology and Engineering
Vellore Institute of Technology, Vellore, India
ankuronthenet@yahoo.com, krishna.c@vit.ac.in

ABSTRACT: The newer of the ITU/MPEG standardization effort of H.264/AVC scalable extension, Joint Scalable Video Model (JSVM) has been made for efficient Scalable Video Coding (SVC). The Motion Estimation (ME) and Motion Compensation (MC) techniques are applied to exploit the redundancy between two consecutive frames. This paper deals with application of Multi-resolutional Discrete Wavelet Transform (DWT) to improve channel bandwidth aspects and the visual quality of the video in error prone network. A gradient based approach is used to recover the lost pixels via finding the direction of maximum amount of correctly received pixels and concoles the lost pixels. The algorithm applied in a temporal domain on the spatial frequency of the Motion Vectors (MV). To evaluate the proposed algorithm an open source reference software for SVC called Joint Scalable Video Model (JSVM) developed by ITU is used. In the results shows the proposed algorithm gets the significant improvement in visual quality for error prone networks with reduction in bandwidth & improves PSNR.

KEYWORDS: Motion Estimation (ME), Motion Compensation (MC), Group Of Picture (GOP), JSVM (Joint Scalable Video Model), Dyadic partition, Lifting Based Approach, Fine Granular Scalability (FGS), Low Pass component (LP), High Pass Component (HP).

1 INTRODUCTION

In order to serve the single encoded best possible quality video to a variety of end-users with different limitations in terms of internet bandwidth, different networks and processing powers; has become a major challenge. The answer is offered by the SVC (Scalable Video Codec) extension of the H.264/AVC video compression format. The H.264 scalable extension, developed by ITU JVT committee represents the most recent development of SVC [1]. The codec encodes the conventional single-stream video into multiple video streams or layers such that the resulting bit stream is still decodeable and can extract the highest layer most as per the user end capabilities [1]. To realize this in a coding way [5], a combination of intrinsically scalable technique and layered approach is used in Joint Scalable Video Model (JSVM). For quality of video in scalability an Embedded Quantization Approach for Coarse Granularity Scalability and Fine Granularity Scalability (FGS) is used [13] [3] [1] [13].

The whole paper is categorized in to 6 sections. The section 2 describes the overview of the DWT & Multi-resolutional DWT over a raw video sequence. The section 3 describes the existing algorithm in JSVM proposed by ITU/MPEG. The section 4 describes the proposed framework. The section 5 shows the implementation and results. The section 6 involves the conclusions and future work. While in the last part the references are added.

2 DISCRETE WAVELET TRANSFORM (DWT) & MULTI-RESOLUTIONAL DWT IN VIDEO

A layered based video compression ME & MC are done in recursive manner in the GOP, the first I frame are coded, and then the P & B frames are coded [4]. In between if any error occurs, these errors will get accumulated and for the entire process at the decoder. These errors are sufficiently large in a time span and produce significant disparities in visual quality [2] [3] as shown in figure1.

![Figure 1: ME/MC at GOP Level](image)

To overcome this problem normally some transformation techniques on the video signals are applied. The Figure 1 addressed the recursive loop prediction error picturably [2]. The popular transforming coding techniques like DCT, FFT, and DWT & their expansions of transformation techniques used in video compression. For representation of higher quality of video with least bandwidth we are using shapiros’s DWT transform which is adopted by ITU/MPEG [6]. In a GOP based video coding, the spatio temporal signal resolution to be represented by the base layer (i.e., layer 0) is first generated by decimation. The base-layer reconstruction is an approximation of all the higher layer resolution levels; it’s kind of lifting based approach described in [7]. Video analysis on raw videos where Y: Cb: Cr component is individually treated and transformed into a 2-D DWT. The filter coefficients used for both of the lifting-based implementations with dyadic partition is presented in Figure2.

The HP & LP functions can be used in a tree based approach as shown in Figure 3. The framework applies on an analyzer filter bank by decimation at the encoder side and the
synthesizer filter bank by interpolation at the decoder side by MC and IMC operation respectively as shown in Fig 3. So due to this bank of filters it will gives us the better results for prediction of HP & LP components of the signals, reduction in bandwidth with improving the visual quality in terms of PSNR. The two levels of filter banks applied on the signal at encoder and decoder side respectively to achieve the filtering action with LL, LH, HL an HH via dyadic partition respectively on rows n columns of that particular frame as shown in Figure 2. [8]. After applying the a bank of filters, rather to consider a Full Scale Granularity (FSG) of the spatial resolution we are achieving more and more detailed Coarse Granularity [13]. As shown in Figure 3 due to decimation at analysis filter each stage achieves decimation by n/4 signals, hence it reduces the bandwidth. Due to the more detailed Coarse Scale Granularity of the signal by applying filters, we can eventually apply a more n better prediction for the spatial component. Hence this will improve the visual quality as shown in result in section 4.

2.1 GOP based Multiresolutional Analysis

In a GOP each frame consist the essential amount of MV’s through the ME/MC process in order to sufficiently represent the video data. Consider for signal X[n] for N-dimensional vector space, the value of the N is the size of GOP 8, 16, 32 or 64 as per the encoding configuration. The number of frames will be represented as X_0, X_3, ..., X_N. The weight of MV’s associates with each frame is defines as some unique scalars like W_0, W_2, W_5, ..., W_N. For each frame in the GOP with N sample space the signal X[n] will be have like

\[ X = W_0X_0 + W_2X_2 + \ldots + W_NX_N \]

The set of all such linear combination is also called as vector space. It is of dimension N-1 and every vector in it is also lies in lies in vector subspace V_N. Here we consider an N-1 sample space because for the i^{th}-frame. For 1^{st}-frame firstly assigns a manual scaling by 1 in order to make half of its amplitude and via shifting function we can easily predict other signals call this as \( V_{N-1} \). So, subsequently the whole vector space from a nested sequence of subspaces is of like as shown below.

\[ V_1 \subset V_2 \subset V_3 \ldots \subset V_N \]

As we can seen in Figure 4 a multi-resolution approach using a set theory one can say that in recursive way each \( V_0, V_1, V_2 \) and V3 component of the above diagram are the subset of each other. So in a place where some errors or some high frequency components are presents, we can predict the value of those particular frames, by estimate the approx MV’s. So in fast motion video streaming like foot ball match or cricket match, we can approximate the packet loss or the sudden scene caused high energy, may be well predicted by this approach [9][6][4][2].

3 Existing Algorithm in JSVM for H.264 SVC

The motion updating method implemented in JSVM is in divided in two steps: first is to derive the motion information used in the update step from the IMC and then, generate the update component by MC with adaptive weight control using lifting based approach [10]. Lets see the working and demerits of the existing algorithm via one example as shown in Figure 5.
As shown in figure, a block B to be updated in the LP frame and the two blocks of A1 and A2, in HP frame are predicted from this block during the prediction step. Let’s take (X1, Y1) & (X1, Y2) be the MVs of the block A1 & A2 respectively & assumes that they are totally different from each other. While updating, one of the IMC of A1 & A2, which having largest nos. of connected pixels predicted from B, is set to be the update MV for B. Here, the update MV for B has 9 connected pixels with B and (X2, Y2) has only 3. However, among the 9 connected pixels through (X1, Y1) of A1, 2 of them are also connected through (X2, Y2) and 6 pixels of B are not connected by either of the two IMVs from A1 & A2. Consequently, the derived update MV for B, (X1, Y1), may not accurately represent the true motion of the whole block since some of its pixels are connected to different MVs, and others are not connected at all as per courtesy to[11][12].

As describe the JSVM does not keep any track of the updated MV with it. It only updates the MV to the current frame of the GOP by applying DWT & IMC to the current frames as per the order they arrive. So in high motion videos, this packet loss can not be retrieved back. Re-sending the lost packets through Resource Management (RM) cells in a network again it takes some time and ultimately video can’t be buffered smoothly. As we discuss in Section 2, that the ME and MC is in e recursive loops in nature. As a result generally we saw some kind of square blocks or other disparities on the screen.

4 PROPOSED ALGORITHM IN JSVM FOR H.264 SVC

In order to overcome the problem discussed in section 3 the proposed algorithm used multi-resolution DWT approach on the MCTF and find the gradient of the correctly received pixels. Where we keep a track of each frame’s MV of the GOP and based on that energy & MV. In a GOP for i-frame assigns a scaling function in order to get its half shifts. In this article it first assign some scaling function to the i-th-frame by assigning a lower QP value in order to keep the quality higher n obtain better prediction for P & B pictures of GOP. After scaling the i-th-frame of GOP, we did the ME & MC, at encoder. By obtaining the IMC, we get the MV’s of the i+1 frame, and we keeps’ the record of the energy in the framefor entire GOP as shown in Figure 6.

![Figure 6](image)

Figure 6: Proposed MV Scheme for JSVM

So, now whenever, a packet loss, high energy problem or any There may be probability that more nos. of received pixels is damaged, or consecutive frames MVs may lost recursively. So we are finding the gradient for the lost pixels. Gradient is calculated as absolute difference for each neighbouring pixel in x (Gx) as well as y (Gy) directions. The pixel edge direction is calculated using arc tan (Gx / Gy).

![Figure 9](image)

Figure 9: Gradient Based Direction Finding for Lost Pixels

The edge orientation is divided into 8 directions with an interval of 22.50 in between them. Concealment is performed in the direction which has the maximum magnitude as shown in Figure2. Directional interpolation based EC techniques rely on the fact that the information is heavily correlated. Here use of DWT and dyadic partition based HP & LP filter banks on spatial frequency components with respect to temporal domain.

5 IMPLEMENTATION & EVALUATION

In order to evaluate proposed algorithm the experiment is perform on open source called JSVM. The algorithm is tested on fast and slow motion raw format videos. For the evaluation first encodes a video through JSVM encoder and then apply some rtp loss on encoded video sequence manually to consider it as an error affected video. For error generation in rtp loss standard checker board pattern type error generations are available and apply the same. At the decoder side check that the proposed algorithm conceals these errors effectively or not. Firstly take the simple akiyo.yuv and claris.yuv sequences. These sequences are slow motion video, and try to obtain how much amount of bandwidth and PSNR improvement is observed, due to multi-dimensional DWT applied on it. Figure 7 shows the performance evaluation on these sequences with different GOP values and at different temporal frequencies. As per the result shown, the significant amount of bitrate & PSNR values are improved. The figure 8 gives the graphical comparsion between the existisng and proposed algorithm in terms of PSNR and bit-rates.

For error concealment evaluation, test is applied on two standard high motion raw videos named as foreman.yuv and tennis.yuv. Apply standard checker board pattern on it, as rtp loss. And compare the existing and proposed algorithm by visual quality as shown in Figure8. The Figure8 shows that our proposed algorithm consoles the errors more efficiently with compare to the existing algorithm and prevents the degradation of visual quality in a video sequence.

5 CONCLUSION

As per the result shown above, our proposed algorithm works well with existing JSVM codes, improves the visual quality in terms of PSNR values & the reduction in the bitrate for normal motion with error video. And also for higher motion
video, it consoles’ the error significantly as we can visualize and compare with existing results. In future we can apply more efficient operator to find the direction of maximum received pixels in a frame and can get better error concealments in video. We can also apply some rate control mechanisms with it to serve best possible quality video in error prone networks.

REFERENCES


