Intra Scalable Video Coding With Pipelining

Snitha A

Pg Student, Vlsi & Embedded Systems, Ece Department
Tkm Institute Of Technology
Karuvelil P.O, Kollam, Kerala-691505, India

Gokul P G

Assistant Professor, Ece Department
Tkm Institute Of Technology
Karuvelil P.O, Kollam, Kerala-691505, India

Abstract— **Scalable video coding is an extension of H.264/advanced video coding (AVC) standard. Video coding today is used in a wide range of applications ranging from multimedia messaging, video telephony and video conferencing over mobile TV, wireless and Internet video streaming, to standard and High-Definition TV broadcasting. If this SVC is constructed on the fundamental H.264/AVC, the complexity of SVC is much higher than that of standard H.264/AVC. So a VLSI design for all-intra scalable video encoder is essential for efficient scalable video encoding. Inorder to achieve better memory bandwidth requirements an efficient coding method, called Frame level coding, is used. It achieves best tradeoff between internal memory usage and external memory access. Afterward, an all-intra SVC encoder combined with several advanced techniques, including fast intra prediction algorithm and three level pipeline is also employed to increase data throughput. The processing speed and memory utilization are also improved. These functionalities provide enhancements to transmission and storage applications. The coding is done in VHDL language, synthesized using Xilinx ISE 13.2 and simulated using Modelsim SE 6.5**

Keywords— *H.264/AVC, SVC, Frame level coding, fast intra prediction algorithm.*

I. INTRODUCTION

Scalable video coding (SVC) is an extension of the H.264 (MPEG-4 AVC) video compression standard for video encoding. The video codec allows video transmission to scale so that content is delivered without degradation between various endpoints, for example, between a laptop and a mobile device. The SVC codec translates bits from a network data stream into a picture and conversely translates camera video into a bit stream. It breaks up video bit streams into bit stream subsets that add layers of quality and resolution to video signals.

SVC codecs adapt to sub-par network connections by dropping these bit stream subsets or packets in order to reduce the frame rate, resolution or bandwidth consumption of a picture, which prevents the picture from breaking up. For example, a mobile phone would receive only the base layer or bit stream while a high-definition video conferencing console would receive both the base layer and bit stream subset or enhancement layer.

SVC is backwards compatible, so an SVC codec can communicate with an H.264 codec that is not SVC capable. The evolution of digital video technology and the continuous improvements in communication infrastructure is propelling a great number of interactive multimedia applications, such as

real-time video conference, web video streaming and mobile TV, among others. The new possibilities on interactive video usage have created an exigent market of consumers, which demands the best video quality wherever they are and whatever their network support. On this purpose, the transmitted video must match the receiver's characteristics such as the required bit rate, resolution and frame rate, thus aiming to provide the best quality subject to receiver's and network's limitations. Besides, the same link is often used to transmit to either restricted devices such as small cell phones, or to high-performance equipments, e.g. HDTV workstations. In addition, the stream should adapt to wireless loss networks (Ohm). Based on this reasoning, these heterogeneous and non-deterministic networks represent a great problem for traditional video encoders which do not allow for on-the-fly video streaming adaptation.

Scalable video coding involves generating a coded representation (bit-stream) that allows decoding of appropriate subsets to reconstruct complete pictures of resolution or quality with the proportion of the bit-stream decoded. The minimum bit-stream subset that can be decoded is called base layer. The remaining bits in the bit stream are called enhancement layer and by decoding the enhancement layer more details are obtained to get the video at higher resolution or quality as compared to base layer. The research on scalable video coding has been an active area for about 20 years. Many early standards, e.g. MPEG-2 Video/H.262 and MPEG-4 Visual, have included tools to provide several important scalabilities. However, the scalable profiles of these standards have rarely been used. One reason is due to the characteristics of traditional video transmission systems in which scalabilities is not really necessary. Another main cause for the situation is the fact that scalability always comes along with a significant loss in coding efficiency as well as a large increase in decoder complexity compared to the corresponding non-scalable profiles[4].

In July 2007, a scalable extension of H.264/MPEG-4 AVC (Advanced Video Coding) was jointly published by MPEG and ITU-T Video Coding Experts Group (VCEG), which makes the scalable extension to be the state-of-the-art scalable video codec. Several new coding techniques were developed in the scalable extension and the gap of coding efficiency has been reduced with state-of-the-art non-scalable codec while the complexity increase is reasonably maintained.

A video bit stream is called scalable when parts of the stream can be removed in a way that the resulting sub-stream forms another valid bit stream for some target decoder, and the

sub-stream represents the source content with a reconstruction quality that is less than that of the complete original bit stream but is high when considering the lower quantity of remaining data. Bit streams that do not provide this property are referred to as single-layer bit streams. The usual modes of scalability are temporal, spatial, and quality scalability. Spatial scalability and temporal scalability describe cases in which subsets of the bit stream represent the source content with a reduced picture size (spatial resolution) or frame rate (temporal resolution), respectively. With quality scalability, the sub-stream provides the same spatio-temporal resolution as the complete bit stream, but with a lower fidelity, where fidelity is often informally referred to as signal-to-noise ratio (SNR). Quality scalability is also commonly referred to as fidelity or SNR scalability.

The report is organized as follows. Chapter 2 includes the literature survey of the project. It gives the basic idea of different video compression techniques such as MPEG, H.264 and scalable video coding. It also specifies the intra frame prediction technique and different coding methods available for an SVC encoder. Chapter 3 presents the architecture. The simulation results are given in the Chapter 4. It includes the simulation result obtained from MATLAB and modelsim and finally Chapter 5 concludes this project and outlines the further researches followed by the references.

II. LITERATURE REVIEW.

A. Video Encoding

Video data may be represented as a series of still image frames. The sequence of frames contains spatial & temporal redundancy that video compression algorithms attempt to eliminate or code in a smaller size.Compression aims at lowering the total number of parameters required to represent the signal, while maintaining good quality. These parameters are then coded for transmission or storage. Result of compressing digital video is that it becomes available as computer data, ready to be transmitted over existing communication networks. Many video compression techniques are available today. Of these, the MPEG standards are the most widely used video coding standards.

1) MPEG: MPEG stands for Moving Picture Coding Exports Group . At the same time it describes a whole family of international standards for the compression of audio-visual digital data. The most known are MPEG-1, MPEG-2 and MPEG-4, which are also formally known as ISO/IEC-11172, ISO/IEC-13818 and ISO/IEC-14496.

The MPEG-1 Standard was published 1992. MPEG-1 was designed to allow a fast forward and backward search and a synchronisation of audio and video. In 1994 MPEG-2 was released, which allowed a higher quality with a slightly higher bandwidth. MPEG-2 is compatible to MPEG-1. MPEG-2 is more scalable than MPEG-1 and is able to play the same video in different resolutions and frame rates.

MPEG-4 was released 1998 and it provided lower bit rates (10Kb/s to 1Mb/s) with a good quality. It was a major development from MPEG-2 and was designed for the use in interactive environments, such as multimedia applications and video communication. This standard promises much higher compression than that possible with earlier standards. It allows coding of non-interlaced and interlaced video very efficiently,

and even at high bit rates provides more acceptable visual quality than earlier standards. Further, the standard supports flexibilities in coding as well as organization of coded data that can increase resilience to errors or losses. As might be expected, the increase in coding efficiency and coding flexibility comes at the expense of an increase in complexity with respect to earlier standard.

MPEG-4 Part 10/AVC for Advanced Video Coding is widely known as H.264 video compression standard. H.264 is an open, licensed standard that supports the most efficient video compression techniques available today. H.264 is the result of a joint project between the ITU-T's Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group (MPEG). ITU-T is the sector that coordinates telecommunications standards on behalf of the International Telecommunication Union. H.264 also has the flexibility to support a wide variety of applications with very different bit rate requirements. With H.264, a new and advanced intra prediction scheme is introduced for encoding frames. This scheme can greatly reduce the bit size of a frame and maintain a high quality by enabling the successive prediction of smaller blocks of pixels within each macroblock in a frame.

B. Overview Of H.264 Standard

H.264 Standard is based on the use of a block-based transform for spatial redundancy removal. H.264 uses an adaptive transform block size, 4x4 and 8x8 (High Profiles only), whereas previous video coding standards were using the 8x8 Discrete Cosine Transform (DCT). The smaller block size leads to a significant reduction in ringing artifacts. The 4x4 transform has the additional benefit of removing the need for multiplications. Transform convert spatial domain to frequency domain.

H.264 uses entropy coding method to for matching a simbol to code,based on context characteristics. All syntax codes except for the residual data are encoded by the Exp-Golomb codes. To read the residual data zigzag 1 scan(interlaced) or alternate scan(no interlaced) is used. For coding the residual data, a more sophisticated method called Context Adaptive Variable Length Coding(CAVLC) is employed. Context Adaptive Binary Arithmetic Coding(CABAC) is also employed in main and high profiles, CABAC has more coding efficiency but higher complexity compare to Context Adaptive Variable Length Coding(CAVLC). A coded H.264 stream or an H.264 file consists of a series of coded symbols.

A filter can be applied to every decoded macro block in order to reduce blocking distortion. The de-blocking filter is applied after the inverse transform in the encoder (before reconstructing and storing the macro block for future predictions) and in the decoder (before reconstructing and displaying the macro block).
The filter has two benefits:
(1) Block edges are smoothed, improving the appearance of decoded images (particularly at higher compression ratios)
(2) The filtered macro block is used for motioncompensated prediction of further frames in the encoder, resulting in a smaller residual after prediction.

C. Scalable Video Coding

SVC (Scalable Video Coding) is an extension to the H.264 codec standard that is used by most of the today's video conferencing devices. SVC video technology allows video conferencing devices to send and receive multi-layered video streams composed of a small base layer and optional additional layers that enhance resolution, frame rate and quality .

The term scalability refers to the removal of parts of the video bit stream in order to adapt it to the various needs or preferences of end users as well as to varying terminal capabilities or network conditions. Bit-stream scalability for video is a desirable feature for many multimedia applications. The need for scalability arises from graceful degradation transmission requirements, or adaptation needs for spatial formats, bit rates or power. To fulfill these requirements, it is beneficial that video is simultaneously transmitted or stored with a variety of spatial or temporal resolutions or qualities which is the purpose of video bit-stream scalability.

1) Modes of scalability **:** The usual modes of scalability are temporal, spatial, and quality scalability. Different scalabilities offer different tradeoffs, and in general some are more suitable for one set of applications while others are better suited for another set of applications.

Spatial scalability is also known as picture size scalability .In this, the video is coded at multiple spatial resolutions. The data and decoded samples of lower resolutions can be used to predict data or samples of higher resolutions in order to reduce the bit rate to code the higher resolutions.

Temporal scalability is also known as frame rate scalability. A video bit stream is called temporal scalable when parts of the stream can be removed in a way that the resulting substream forms another valid bit stream for some target decoder, and the substream represents the source content with a frame rate that is smaller than the frame rate of the complete original bit stream.

Quality scalability is also known as signal-to-noise ratio(SNR) scalability or fidelity scalability. A video bit stream is called quality scalable when parts of the stream can be removed in a way that the resulting substream forms another valid bit stream for some target decoder, and the substream represents the source content with a reconstruction quality that is less than that of the complete original bit stream.

D. Intra SVC Encoder

SVC supports three scalabilities: spatial, temporal, and quality scalability. To support spatial scalability, SVC adopts pyramid coding structure in which the frame resolution of each spatial layer is different from the other layers. As a result, such coding structure results in the high correlation between spatial layers, and thus motivating the adoption of interlayer predictions [1] to fully utilize the similarities between spatial layers. However, the adoption of interlayer prediction also complicates the hardware design. There are three coding methods available. They are frame level coding method,Row level coding method and MB level coding method.

Afterward, the coding method which achieves best tradeoff between internal memory usage and external memory access requirement will be selected as the best coding method in the SVC encoder design. The external memory requirements of all the coding methods is same. But the internal memory storage requirements may vary with different coding methods.

1) Frame level coding method: In this coding method, the spatial enhancement layer would be encoded after the entire reference layer is encoded. The next frame is taken only after completing the current frame.

The internal memory storage requirement of the frame level coding method can be calculated by using the formula,

Internal memory storage requirement

$$= 256 \times m \times (W_{mb,max}+1) \tag{1}$$

Where m is the number of quality layers and $W_{mb,max}$ is the the width of the largest frame in the unit of MB.

2) Row level coding method: The second method is called the row-level method. In this coding method, a row of MBs would be encoded after a corresponding row of MBs in the reference layer has been encoded.

The internal memory storage requirement of this coding method can be calculated by using the following formula,

Internal memory storage requirement

$$= 256 \times m \times \sum_{i=0}^{n-1} (W_{mb,i}+1) \tag{2}$$

Where m is the number of quality layers and $W_{mb,i}$ is the frame width of the ith spatial layer in the unit of MB.

3) MB level coding method: The third one is called the MB-level method. In this coding method, an MB in the reference layer would be encoded first, and then the corresponding MBs in the enhancement layer will be encoded. The internal storage requirements remain the same as in the row-level method.

In MB level and row level coding methods, some information such as the prediction data of neighboring MBs have to be stored temporary inside the internal memory due to spatial layer switching[1]. Although the prediction data of neighbouring MBs can be stored into external memory, the frequent external memory access will lead to the overall video coding system performance drop. Therefore, it is more reasonable to store such data inside internal memory. However,for the frame-level coding method, the problem of internal memory space caused by spatial layer switching would no longer exist and would thus result in less internal memory space requirements[7]. Thus frame level coding method has less internal memory storage requirement compared to the other coding methods. So it can be used in the encoder design for efficient video coding.

The frame level coding method indicates the way in which each macroblock is taken for processing. According to this coding method the first macroblock in the first frame is taken first for processing. Then the second macroblock of the first frame is taken for processing and so on. After completing all the macroblocks in the first frame, second frme is taken for

processing. The next frame is taken only after completing the current frame.

III. PROPOSED METHOD

To satisfy the video application diversities, an extension of H.264/advanced video coding (AVC), called scalable video coding (SVC), is designed to provide multiple demanded video data via a single video encoder. However, constructed on the fundamental of H.264/AVC, the complexity of SVC is much higher than that of H.264/AVC. So a new VLSI architecture is used for the design of SVC. Inorder to achieve better memory bandwidth requirements an efficient coding method,called Frame level coding, is used. A fast intra prediction algorithm[9] and three pipeline stages are employed to incresae the system throughput.

A. System architecture

The basic block diagram of the project work is shown in the Figure 3.1. It consists of three pipeline stages. First is the intra perdiction stage which takes video pixel inputs obtained from Matlab. Second pipeline stage itself consists of two stages; quality refine phase and the quantization and reconstruction phase.

Fig. 1. Basic Block Diagram

The overall scheduling of the SVC encoder adopts the interlaced frame parallel three-stage scheduling method as illustrated in Fig.1. Most of the computations are interlaced for two MBs from parallel frames to increase the hardware utilization. At the beginning, the two first MBs of the two frames will enter into Intra Prediction Phase for intra prediction. Afterward, the two second MBs of the two frames will be inputted into Intra Prediction Phase and the two first MBs go into Quantization and Reconstruction Phase.

However, due to the data dependency between the first stage and second stage, the NOP operations will be held in the first stage temporarily for the two second MBs in order to wait for the reconstruction of the two first MBs. Once the necessary pixels of the two first MBs have been successfully reconstructed, the reconstructed pixels, accompanied with a valid signal, will be forwarded to the first stage from the second stage to start up the intra prediction and transformation process for the two second MBs. Finally, the two first MBs go into the third stage to generate the SVC bitstream and remove the blocking effects. At the same time, the two second and third MBs of the two frames will, respectively, enter into the second stage and first stage for encoding.

1) *Three Pipeline Stage Architecture Design*: Fig.1 shows the proposed intra SVC encoder architecture design with three pipeline stages. In this architecture, all operations are grouped into four phases including Intra Prediction Phase, Quality Refine Phase, Quantization and Reconstruction Phase, and Entropy and Deblocking Phase. First, in Intra Prediction Phase, the input is given to the TraDED intra prediction. The function of quality scalability can be done in Quality Refine Phase. Then, the data are further quantized and reconstructed in the Quantization and Reconstruction Phase, along with the quality layer computation. Finally, the Entropy and Deblocking Phase executes the entropy coding for all coefficients that are needed to be encoded into bitstream and removes the blocking effect of reconstructed pixels.

2) *Architecture Design of First Stage*: It basically follows the TraDED algorithm to compute the intra prediction output. In TraDED design,pixels values at the top and left boundary values are used to predict all other pixel values within a macroblock. This is used for increasing the efficiency of the video encoder.

3) *Architecture Design of Second Stage*: In the second pipeline stage the operations like quantization, reconstruction, and quality scalability (Quality Refine Phase) are executed. The quantization module also supports quality refinement processes to share the hardware cost since most of the operations in this coding technique are very similar to the quantization process except the normalization and coefficient subtractions.

4) *Architecture Design of Third Stage*: In this pipeline stage, two modules called Deblocking and Entropy Coding are implemented. The Deblocking module is used to remove the blocking effects of reconstructed pixels for further prediction usage.

To achieve high system throughput entropy coding, an improved CAVLC design is adopted in this design[3]. Context-based adaptive variable length coding (CAVLC) is a form of entropy coding used in H.264/MPEG-4 AVC video encoding. It is an inherently lossless compression technique, like almost all entropy coders. It is an alternative to context-based adaptive binary arithmetic coding.

In the baseline profile of H.264/AVC, the context-based adaptive length coding (CAVLC) is used for coding quantized transform coefficients of the residual images[8]. In CAVLC, the reverse zigzag scanned run-length coding, and adaptive VLC tables are used to encode 4x4 or 2x2 block residual data.

IV. SIMULATION RESULTS

The design entry is modelled using VHDL in Xilinx ISE Design Suite 13.2 and the simulation of the design is performed using modelsim SE 6.5 from Xilinx ISE to validate the functionality of the design.

A. Video pixel values obtained from MATLAB

Video pixel values are obtained from MATLAB. For that a video found in the MATLAB library, 'xylophone.mpg', is used. Pixel values corresponding to each frame is obtained using MATLAB R2013b.
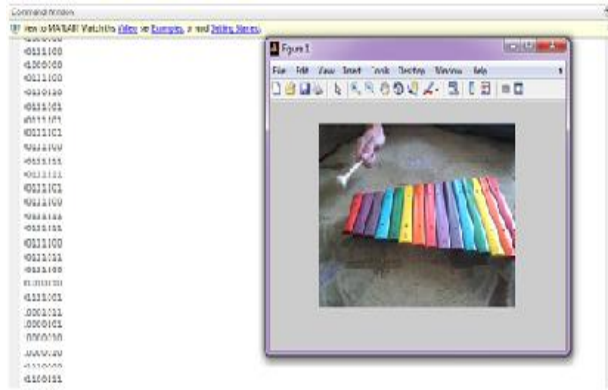
Fig. 2. Obtained pixel values from a video

Fig.2. shows the image frame pixel values of the video 'xylophone.mpg' found in the MATLAB video library. It also shows the 8th image frame of the video.
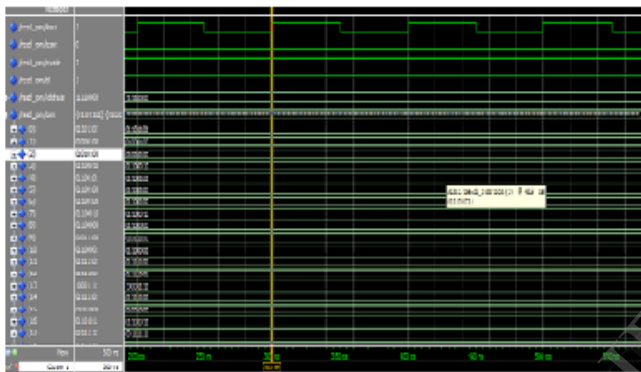
B. Video pixel values stored in a buffer

Fig.3. pixel values stored in the buffer

Fig.3. shows the image pixel values stored in the buffer. The pixel values obtained from the MATLAB are stored in a text file and the text file is read and simulated in Modelsim

C. Simulation of Intra prediction algorithm

Intra prediction algorithm is applied in the first pipeline stage of the SVC design. The intra prediction algorith is used for predicting the pixel values in a Macroblok if the neighbouring information are provided.
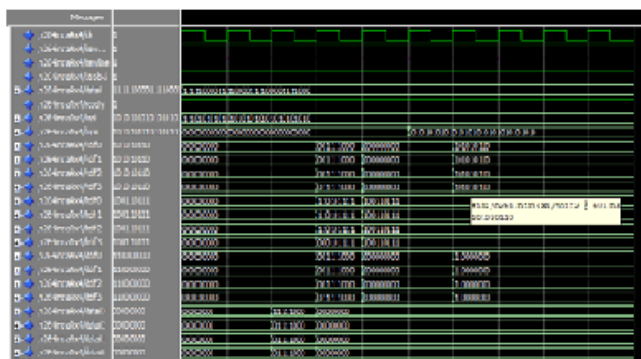
Fig. 4. Simulation result of intra prediction algorithm

Fig.4. shows the simulation result of intra prediction algorithm. It contains the output of four prediction modes; vertical,diagonal,horozontal and DC mode.

V. CONCLUSIONS

The project is intended to design a new VLSI architecture for scalable video encoder. Scalable video coding (SVC) is designed to provide multiple demanded video data via a single video encoder. Frame level coding method is used to achieve better performance. Pixel values are obtained using MATLAB and the values are stored in the buffer. The entire architecture is divided in to three pipeline stages to improve the system throughput.

REFERENCES

[1] Gwo-Long Li, Tzu-Yu Chen, Meng-Wei Shen, Meng-Hsun Wen, and Tian-Sheuan Chang" 135-MHz 258-K Gates VLSI Design for All-Intra H.264/AVC Scalable Video Encoder" *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 21, no. 4, pp.636-647,april 2013.

[2] H.-Y. Lin, K.-H. Wu, B.-D. Liu, and J.-F. Yang, "An efficient VLSI architecture for transform-based intra prediction in H.264/AVC," *IEEETrans. Circuits Syst. Video Technol.,* vol. 20, no. 6, pp. 894–904, Jun.2010.

[3] T.-H. Tsai, S.-P. Chang, and T.-L. Fang, "Highly efficient CAVLC encoder for MPEG-4 AVC/H.264," *IET Circuits Devices Syst.*, vol. 3,no. 3, pp. 116–124, Jun. 2009.

[4] Y.-K. Lin, C.-W. Ku, D.-W. Li, and T.-S. Chang, "A 140-MHz 94 K gates HD1080p 30-Frames/s intra-only profile H.264 encoder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 432–436, Mar. 2009.

[5] K. Xu and C.-S. Choy, "A five-stage pipeline, 204 cycles/MB, singleport SRAM-based deblocking filter for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 3, pp. 363–374, Mar.2008.

[6] F. Tobajas, G. M. Callico, P. A. Perez, V. de Armas, and R. Sarmiento,"An efficient double-filter hardware architecture for H.264/AVC deblocking filtering," *IEEE Trans. Consumer Electron.*, vol. 54, no. 1,pp. 131–139, Feb. 2008.

[7] C.-H. Hsia, J.-S. Chiang, Y.-H. Wang, and T.-Y. Teng, "Fast intra predictionmode decision algorithm for H.264/AVC video coding standard,"in *Proc. IEEE Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.,*vol. 2, pp. 535–538. . Nov. 2007.

[8] M.-C. Tsai and T.-S. Chang, "High performance context adaptive variable length coding encoder for MPEG-4 AVC/H.264 video coding,"in *Proc. IEEE Asia Pacific Conf. Circuits Syst.*, Dec. 2006, pp.586–589.

[9] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. Wu, and S. Wu,"Fast mode decision algorithm for intraprediction in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 15, no. 7, pp.813–822,Jul.2005