ELectronics EXpress

Real-time and smooth scalable video streaming system with bitstream extractor intellectual property implementation

Liang-Hung Wang[1a)], **Yi-Mao Hsiao**[2], **Rong-Shan Wei**[3], **and San-Fu Wang**[4]

[1] *College of Physics and Information Engineering, Fuzhou University, Fujian*
[2] *Infor. and Comm. Res. Lab., Industrial Technology Research Institute, Taiwan*
[3] *College of Physics and Information Engineering, Fuzhou University, Fujian*
[4] *Department of Electronic Engineering, Ming Chi University of Tech., Taiwan*
a) *eetommy@gmail.com*

Abstract: Network congestion causes the users cannot perform a real-time streaming service. The paper proposes a real-time and smooth scalable video coding (SVC) streaming system with a bitstream extractor intellectual property (BEIP) design. The system parses the source video to obtain the video information, moreover, adjusts adaptively streaming resolution for limited network bandwidth. The BEIP with very large scale integration (VLSI) implementation, the speed, power consumption, and chip area are 243 MHz, 25 mW and $1.073 \times 1.028\,mm^2$, respectively. The IP performs high system throughput of 7.8 Gbps to deal with the network traffic.

Keywords: H.264/SVC, congestion control, bandwidth adaptive
Classification: Integrated circuits

References

[1] H. Schwarz, D. Marpe and T. Wiegand: IEEE Trans. Circuits Syst. Video Technol. **17** [9] (2007) 1103.
[2] Y. Wang, L. P. Chau and K. H. Yap: Proc. IEEE Int. Conf. on Multimedia and Expo (2008) 577.
[3] D. T. Nguyen and J. Ostermann: IEEE J. Sel. Topics Signal Process. **1** [2] (2007) 246.
[4] J. S. Youn and J. R. Choi: IEICE Electron. Express **8** [17] (2011) 1380.
[5] S. W Park and S. U. Shin: IEICE Trans. Inf. & Syst. **E92-D** [5] (2009) 851.

DOI: 10.1587/elex.11.20140090
Received January 29, 2014
Accepted January 31, 2014
Publicized February 17, 2014
Copyedited March 10, 2014

1 Introduction

With the popularity of the network, wireless products with multimedia video service become more and more popular. The scalable video coding (SVC) is proposed as a new video standard [1] that not only extends the high compression ratio of H.264/AVC (Advance Video Coding) characteristic but also owns the scalability of layer coding. The SVC video streaming can be extracted into several sub streams for different requirements of client equipment by spatial, temporal and quality scalability, especially, in complex heterogeneous network environment. The video quality is obviously affected by bandwidth fluctuation in wireless network environment. If the channel bandwidth is small, it will cause the data loss, thereby affecting the video frame delay and fragment. However, the same video streaming in high bandwidth network, the excess bandwidth is wasted.

Wang proposes an algorithm to improve perceptual video quality [2] that is used to calculate the spatial resolution of the sub-stream based on the bandwidth constraint. Nguyen presents a SVC streaming system with a congestion control algorithm which is supported by channel bandwidth estimation [3]. The algorithm implements packet retransmission from the base layer to limit the burden in congested network. The paper is presented the bitstream extraction, encoding and decoding schemes to analyze the bitstream to get the video information, furthermore, it enables dynamic extraction for real-time transmission.

System-on-a-chip (SoC) is a mainstream device in which entire circuits, including digital, memory and mixed-mode signals are integrated into a single piece of silicon. The SoC concept with VLSI design could be widely used in microelectronics, especially in multimedia technology [4]. The application- specific integrated circuits (ASICs) solution of VLSI design is adopted to implement the video compression fast and effectively.

The proposed SVC streaming system with a bitstream extractor IP can be employed to against the bandwidth fluctuation. Based on terminal capacity and display resolution, the system uses spatial scalability to generate several resolution formats for the requirement of clients. In addition, if the network congestion occurred, the system changes different frame rate for the available bandwidth (ABW) with temporal scalability, which the ABW is calculated by bandwidth estimator of streaming system. According to the terminal capacity and the bandwidth information, the bitstream extractor selects the adaptive layer to perform real-time and smooth service for the users. The IP design of bitstream extractor is implemented to be integrated easily with the video codec and network multimedia to a multi-function SoC for real-time and smooth streaming system.

2 System architecture

The frame rate, resolution, and group of picture (GOP) of the CITY, SOCCER and HARBOUR test videos are employed to verify the SVC streaming system. The resolutions of 4CIF, CIF, and QCIF are 704×576, 352×288 and 176×144, and the related frame rates are 7.5 fps, 15 fps and 30 fps. When the resolution is changed from CIF to 4CIF, QCIF to CIF, and QCIF to 4CIF, the requirement of bandwidth are more than 3.85, 5.77, and 20.43 times, respectively. Nine layers (i.e. Fig. 3 (b)) including all resolutions and frame rates are defined to transmit the streaming data. As network congestion occurred, the frame rate is adjusted from 30 fps to 15 fps and 7.5 fps. Based on the information of spatial and temporal scalability, we propose a SVC system to against the network bandwidth

DOI: 10.1587/elex.11.20140090
Received January 29, 2014
Accepted January 31, 2014
Publicized February 17, 2014
Copyedited March 10, 2014

violent changing and provide a smooth streaming service. The basic concept of bitstream extractor is: if the bandwidth is not enough to transmit the video streaming, the system selects down to the lower layer to maintain the video quality. Otherwise, the system raises the layer up for the sufficient bandwidth requirement. Generally, the video streaming without bitstream extractor (i.e. original video streaming), the receiving data is altered immediately and the video displaying is not smooth.

Fig. 1 (a) shows the proposed system architecture including server and client sides. In server side, the down-convert converts the video source (e.g. YUV format, 4CIF, 30 fps) into several resolutions (i.e. from 4CIF to CIF and QCIF). The SVC encoder behind the down-convert encodes the source video into H.264 streaming. The types of network abstraction layer (NAL) packet, supplemental enhancement information (SEI), sequence parameter set (SPS), subset sequence parameter set (SSPS), and picture parameter set (PPS) are analyzed by the NAL parser. The frame rate, layer, resolution, and video related information are stored in SEI type. The decoding sequence is stored in SPS, SSPS, and PPS types. The bandwidth estimator calculates the network bandwidth, from the client side to the server side, and sends the calculated data to the bandwidth budget. As the bandwidth estimator detects the changing of network bandwidth, the time control and bandwidth budget controls the video traffic to overcome the network congestion.

In the client side, the SVC decoder decodes the H.264 streaming into YUV video. The display quality is decided by user information (i.e.

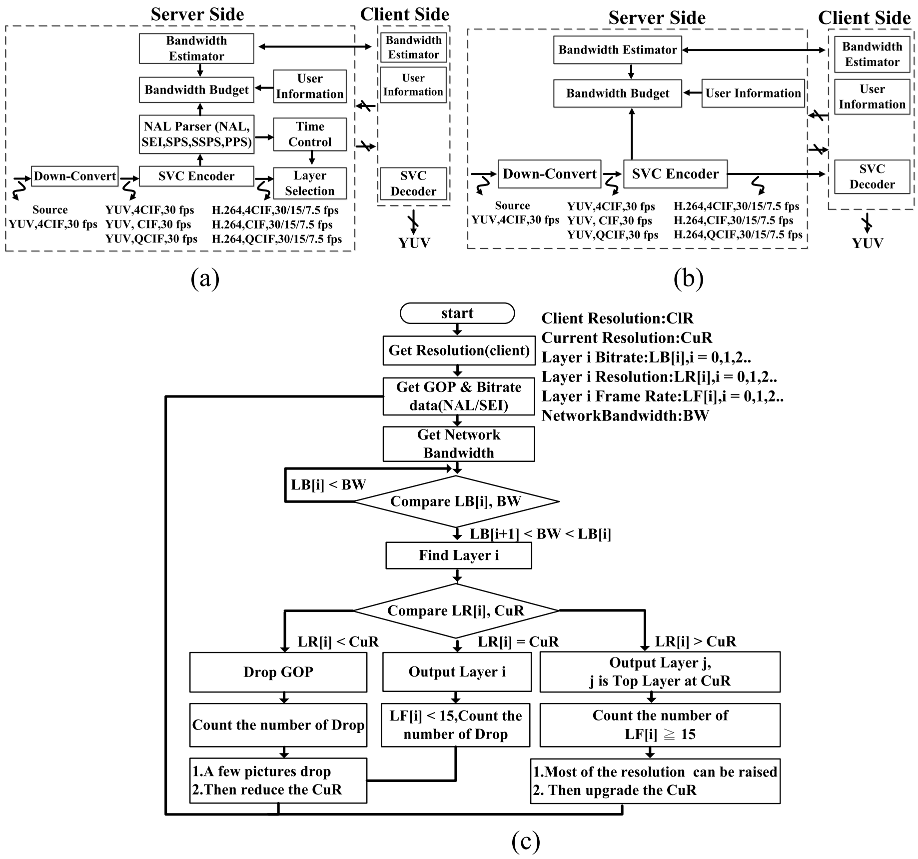


Fig. 1. (a) System architecture of the proposed system.
(b) System architecture of the original system.
(c) The flow chart of layer selection

DOI: 10.1587/elex.11.20140090
Received January 29, 2014
Accepted January 31, 2014
Publicized February 17, 2014
Copyedited March 10, 2014

IEICE Electronics Express, Vol.11, No.5, 1–6

displayer resolution), moreover, the bandwidth information is replied synchronously to server side by bandwidth estimator. In the original system, the streaming data is sent to the client side directly, the system diagram is illustrated in Fig. 1 (b). However, in the proposed system, the functionality of NAL parser, time control, and layer selection are adopted additionally to implement a real-time and smooth bitstream extractor.

The layer selection with time control is proposed to perform the smooth bitstream extractor. The data flow chart of layer selection is shown in Fig. 1 (c). The server gets the displayer resolution of client side, the video SEI/NAL packet, the video GOP and the bit rate of each layer. The server calculates the maximal available bandwidth, which the bandwidth is composed of the video bit rate and the estimated bandwidth. Compared the bit rate of nine layers (i.e. Fig. 3 (b)) and ABW, in addition, the server identifies the closest layer and stores the bandwidth value in bandwidth budget (i.e. Fig. 1 (a)). If the layer resolution (LR) is smaller than the video current resolution (CuR), the video GOP is dropped and the current resolution cannot be displayed, moreover, counts the dropped frame rate and recording. If the LR is equal to the CuR, the selected frame is output to the client. Besides, if the selected frame rate is less than 15 fps, the system still records the dropped data and continually reduces the current resolution. Finally, if the selected LR is bigger than the CuR, the server outputs the maximal resolution frame to the client. If the frame rate is higher than 15 fps, the system records the current bandwidth is enough. Furthermore, when the transmitted frame rate is enough for a period, the system upgrades to higher level video streaming to transmit the higher performance video.

3 Implementation the bitstream extractor IP

The SVC bitstream extractor is implemented as an IP design. Fig. 2 (a) shows the hardware architecture, including Bitstream_Buffer, Bitstream_Control, Layer_Selector and NAL_Porcessor. The Bitstrseam_Buffer with 128 bytes is employed to store the H.264 streaming data from the bus. If the data is less than 64 bytes, the register allows the data entry, and the input data is analyzed by Bitstream_Control. In SEI state, the system reads the payload of SEI information. If the video bit-rate is read, the bit-rate information is stored into the registers for Layer_Selector function. The SPS, SSPS and PPS states are similar to the SEI state that passes the NAL information to the NAL_Processor. Several types of NAL packets are decoded into two ways, Ext_Golomb_Decoder and Fixed_Len_Decoder ways. The executed functions of the types are: 1.) Finding the start code; 2.) Checking the NAL header to identify the types. If the type is in SEI, the system analyzes the payload. Otherwise, the data are sent to NAL processor; 3.) Analyzing the SEI payload. The Layer_Selector implements the layer selection strategy, moreover, the display information of client side is also stored in the SEI state. The Layer_Selector choices the adaptive layer from the bandwidth estimator and the bit-rate of current video. The NAL_Processor analyzes the NAL packets. If the packets are SPS, SSPS and PPS, the system outputs the video streaming of the selected layer. If the NAL types are 14 and 20, the NAL_Processor outputs the original NAL data and the system outputs the NAL packet header. The implementation

DOI: 10.1587/elex.11.20140090
Received January 29, 2014
Accepted January 31, 2014
Publicized February 17, 2014
Copyedited March 10, 2014

IEICE Electronics Express, Vol.11, No.5, 1–6

of the IP is carried out in very-high-speed integrated circuit hardware description language (VHDL) and synthesized by Synopsys design compiler with TSMC 0.18-μm standard CMOS process. Fig. 2 (b) shows the related chip area of $1.073 \times 1.028\,\text{mm}^2$.

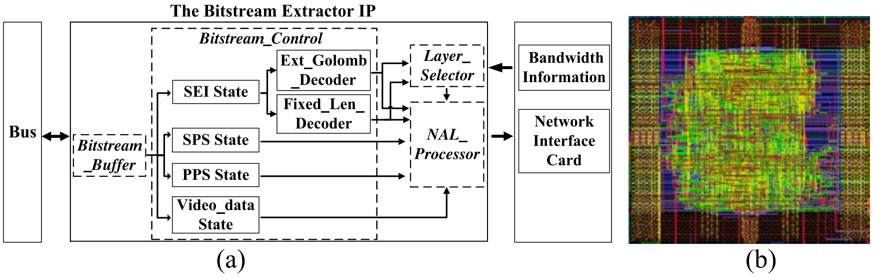


Fig. 2. (a) Hardware architecture of the bitstream extractor IP. (b) The layout of the IP.

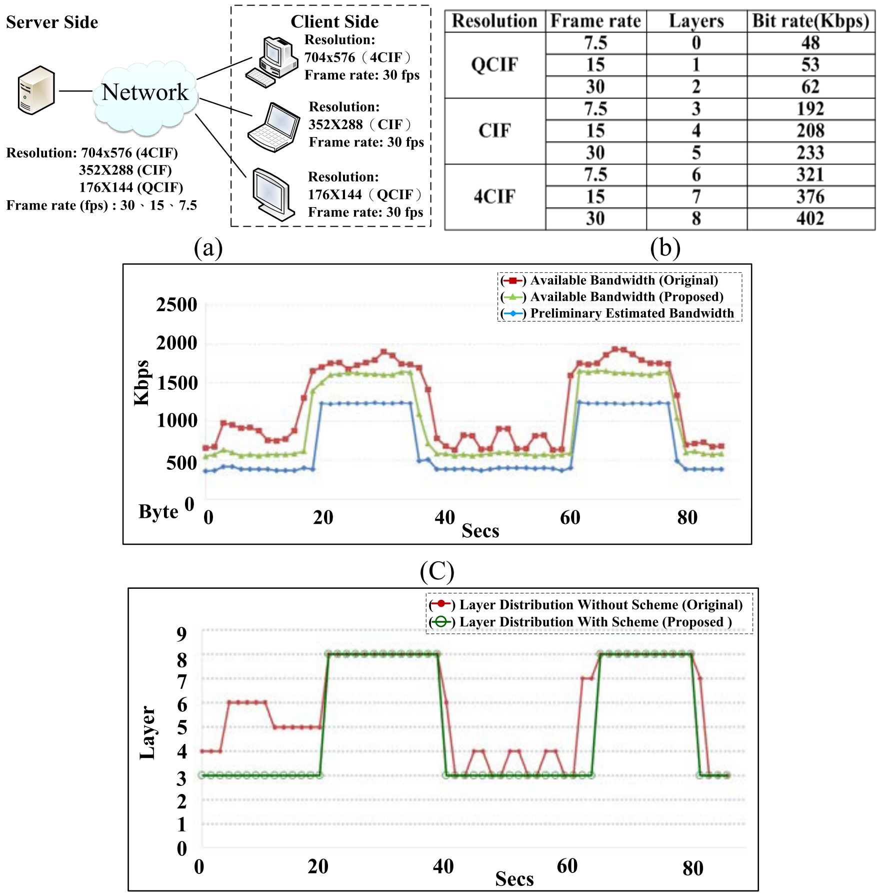
4 Simulation result and performance analysis

The Linux network simulator and the H.264/SVC streaming [5] encoded by JSVM software is used to verify the bitstream extractor IP, the simulation environment is illustrated in Fig. 3 (a). The bitrate setting of the proposed SVC encoder is defined by variable-bit-rate (VBR), therefore, the resolution and frame rate of the SVC sequence streaming are adjusted during the transmission when network bandwidth was changing. According to the available bandwidth, from the bandwidth estimator and the video bit rate, the BEIP selects a suitable video streaming transmission to the client sides. The video SOCCER is adopted as an experimental film. The bit rate of video and the estimated bandwidth are changing from 48 Kbps to 402 Kbps and 400 kbps to 1200 kbps, respectively, during video transmission. Based on the setting of VBR streaming, we define the nine layers with the dedicated resolutions and frame rates that are shown in Fig. 3 (b). Fig. 3 (c) shows the available bandwidth summed of the estimated bandwidth and the bit rate of current video.

The post-layout simulation result of layer selection is shown in Fig. 3 (d). In the original system, because of the streaming data is transmitted directly without layer selection, the layer is decided by the instantly available bandwidth. Therefore, the layer distribution of original system without proposed scheme shows that the layers are fluctuation. According to the proposed scheme, the layer 0 to 2, the layer 3 to 5 and the layer 6 to 8 are with QCIF, CIF and 4CIF resolution, respectively. If the available bandwidth is changed immediately, the system adjusts to adaptive layer to keep the video quality. Undoubtedly, the system with proposed scheme performs smooth streaming service than the original system. The operating frequency and the power consumption of the IP are 243 MHz and 25 mW, respectively. Eq. (1) is employed to compute the system throughput, which the high throughput 7.8 Gbps of 32 bits SVC bitstream extractor IP can deal with the requirement of network traffic.

$$T = clock \times bits \tag{1}$$

DOI: 10.1587/elex.11.20140090
Received January 29, 2014
Accepted January 31, 2014
Publicized February 17, 2014
Copyedited March 10, 2014



Resolution	Frame rate	Layers	Bit rate(Kbps)
QCIF	7.5	0	48
	15	1	53
	30	2	62
CIF	7.5	3	192
	15	4	208
	30	5	233
4CIF	7.5	6	321
	15	7	376
	30	8	402

Fig. 3. The simulation results of the SOCCER streaming for bandwidth changing. (a) The experimental environment. (b) The video bit rate of nine layers. (c) The available bandwidth. (d) Comparison with original and proposed layers.

5 Conclusion

Traditional video system just supports fixed resolution and frame rate in network transmission. If the channel bandwidth is fluctuation and network congestion occurred, the delay and loss of video frame causes the video performance decrease. We propose a real-time and smooth video streaming system with the bitsream exactor that uses scalability of SVC to select the adaptive video streaming for the client equipment. The simulation result shows that the proposed bitsream exactor IP performs more smoothly video to against the available bandwidth fluctuation for the heterogeneous network. We implemented the IP which is easily integrated with video codec and network protocol accelerator. According to the post-layout simulation results, the speed, power dissipation and chip area of the proposed BEIP are 243 MHz, 25 mW, and 1.103 mm^2, respectively, that performs 7.8 Gbps system throughput. The proposed bitstream extractor IP will be fabricated in TSMC 0.18-μm standard CMOS process to realize a multimedia SoC in the future.

DOI: 10.1587/elex.11.20140090
Received January 29, 2014
Accepted January 31, 2014
Publicized February 17, 2014
Copyedited March 10, 2014