VIEWPOINT SWITCHING IN MULTIVIEW VIDEOS USING SP-FRAMES

Ki-Kit Lai, Yui-Lam Chan, Chang-Hong Fu, and Wan-Chi Siu

Centre for Signal Processing
Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
{kikit.lai, enylchan, enwcsiu}@polyu.edu.hk

ABSTRACT

The distinguishing feature of multiview video lies in the interactivity, which allows users to select their favourite viewpoint. It switches bitstream at a particular view when necessary instead of transmitting all the views. The new SP-frame in H.264 is originally developed for multiple bit-rate streaming with the support of seamless switching. The SP-frame can also be directly employed in the viewpoint switching of multiview videos. Notwithstanding the guarantee of seamless switching using SP-frames, the cost is the bulky size of secondary SP-frames. This induces a significant amount of additional space or bandwidth for storage or transmission, especially for the multiview scenario. For this reason, a new motion estimation and compensation technique operating in the quantized transform (QDCT) domain is designed for coding secondary SP-frame in this paper. Our proposed work aims at keeping the secondary SP-frames as small as possible without affecting the size of primary SP-frames by incorporating QDCT-domain motion estimation and compensation in the secondary SP-frame coding. Simulation results show that the size of secondary SP-frames can be reduced remarkably in viewpoint switching.

Index Terms— Multiview, viewpoint switching, SP-frame, QDCT-domain, motion estimation

1. INTRODUCTION

The rapid advancement of video coding techniques has enabled the expansion of conventional single-view videos into multiview videos. Recently, the MPEG committee has led the 3D audio-visual (3DAV) activity to explore the emergence of this new technology. Potential applications include free-viewpoint video (FVV) or free-viewpoint television (FTV), 3D television (3DTV), immersive teleconference and surveillance. To support these applications, video systems require capturing a scene from different viewpoints which result in generating several video sequences from different cameras. The key characteristic of multi-view video is interaction, which can give users the opportunity to choose their favourite viewpoint freely [1]. Instead of transmitting all the views, an efficient switching technique to seamlessly switch among different viewpoint bitstreams is thus necessary.

In the meantime, H.264 has introduced SP-frames which are used to facilitate error resilience, bitstream switching, splicing, random access, fast forward and fast backward [2]. This special SP-frame is composed of primary and secondary SP-frames. They both exploit temporal redundancy with predictive coding, but use different reference frames. Although different reference frames are used, it still allows identical reconstruction. This property can be applied to drift-free switching between compressed bitstreams of different viewpoints.

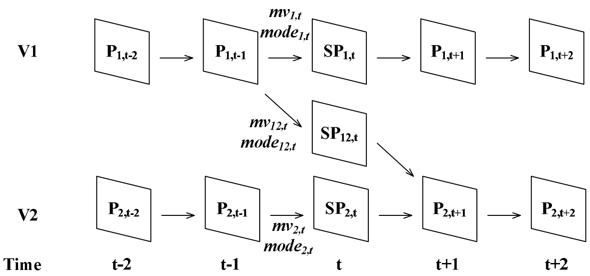

Figure 1. Switching between two views using SP-frames.

Figure 1 illustrates the idea of viewpoint switching in a multiview video using SP-frames. In this figure, two different views, which are captured by two cameras at the same time in the same scene, are encoded into two bitstreams (V1 and V2). V1 is an encoded sequence in viewpoint 1 while V2 is in viewpoint 2. Within each view, two primary SP-frames –$SP_{1,t}$ and $SP_{2,t}$ are placed at frame *t* (switching point). To allow seamless switching, a secondary SP-frame ($SP_{12,t}$) is produced, which has the same reconstructed values as $SP_{2,t}$ even different reference frames are used. When switching from V1 to V2 is needed at frame *t*, $SP_{12,t}$ instead of $SP_{2,t}$ is transmitted. After decoding $SP_{12,t}$, the decoder can obtain exactly the same reconstructed values as normally $SP_{2,t}$ decoded at frame *t*. Therefore it can continually decode V2 at frame *t+1* seamlessly.

In this scheme, extra storage for secondary SP-frames is inevitably required. In [3], an investigation has been conducted to evaluate the trade-off between the coding performance of primary SP-frames and the storage cost of secondary SP-frames. It has been found that a primary SP-frame with high quality results in a significantly high storage requirement for the secondary SP-frame. For a multiview video system, the bulky size of secondary SP-frames is more severe since an additional sequence of secondary SP-frames is required between two views. For instance, extra N-1 sequences of secondary SP-frames are stored in the multiview video system with N views. In this paper, we propose a novel

978-1-4244-1764-3/08/$25.00 ©2008 IEEE

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 30,2010 at 09:16:58 UTC from IEEE Xplore. Restrictions apply.

coding arrangement to reduce the size of secondary SP-frames in the multiview video system.

This paper is organized as follows. Section 2 gives a detailed description of SP-frame encoding structures between two views by using pixel-domain motion estimation and compensation. Section 3 presents an in-depth study of the problem on applying the traditional pixel-domain motion estimation technique into secondary SP-frame encoder. Analysis of using QDCT-domain motion estimation is also covered in this section. After the detailed investigation, a novel secondary SP-frame encoder is proposed. In Section 4, experimental results are shown, which focus on the performance of our proposed scheme over the conventional secondary SP-frame encoder in multiview video sequences. Concluding remarks are provided in Section 5.

2. ENCODING PROCESS OF SP-FRAMES

The way of encoding primary SP-frames of view 1 or view 2 is similar to that of encoding P-frames except additional quantization / dequantization steps with the quantization level *Qs* are applied to the transform coefficients of the primary SP-frame ($SP_{2,t}$ in Figure 2), as shown in Figure 2. Interested readers are encouraged to read the references [4-5]. These extra steps ensure that the quantized transform coefficients of $SP_{2,t}$ (denoted as $SP_{2,t}^{Q_s}$) can be quantized and de-quantized without loss at *Qs*, which is used in the encoding process of the secondary SP-frame, $SP_{12,t}$.

For coding $SP_{12,t}$, the reconstructed $P_{1,t-1}$($\hat{P}_{1,t-1}$) acts as the reference and its target is to reconstruct $SP_{2,t}^{Q_s}$ perfectly. By using the reference frame $\hat{P}_{1,t-1}$, its prediction is first transformed and quantized using *Qs* before generating the residue with $SP_{2,t}^{Q_s}$. Both the prediction and $SP_{2,t}^{Q_s}$ are thus synchronized to *Qs* and there is no further quantization from this point. This means that the decoder, with $\hat{P}_{1,t-1}$, *Qs*, and the residue available, can perfectly reconstruct $SP_{2,t}^{Q_s}$.

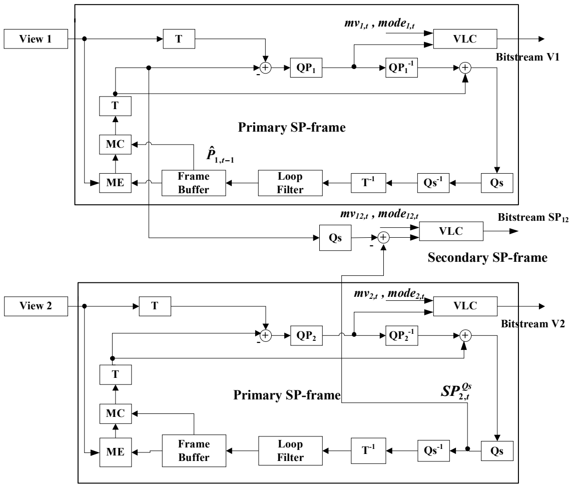

Figure 2. Simplified encoding block diagram of primary and secondary SP-frames [4].

Producing secondary SP-frames involves the processes of motion estimation and motion compensation. In H.264, it supports motion estimation using different block sizes such as 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4 [6]. To compute the coding modes and motion vectors for the secondary SP-frame, motion estimation is firstly performed for all modes and submodes independently by minimizing the Lagrangian cost function J_{motion}.

$$J_{motion}(mv_{12}, \lambda_{motion}) = SAD(s,r) + \lambda_{motion} \cdot R_{motion}(mv_{12} - pmv_{12}) \quad (1)$$

where mv_{12} is the motion vector used for prediction, λ_{motion} is the Lagrangian multiplier for motion estimation, $R_{motion}(mv_{12}$ - $pv_{12})$ is the estimated number of bits for coding mv_{12}, and SAD is sum of absolute differences between the original block *s* and its reference block *r* [6].

After motion estimation for each mode, a rate-distortion (RD) optimization technique is used to get the best mode and its general equation is given by

$$J_{mode}(s,c,mode_{12},\lambda_{mode}) = SSD(s,c,mode_{12}) + \lambda_{mode} \cdot R_{mode}(s,c,mode_{12}) \quad (2)$$

where λ_{mode} is the Lagrangian multiplier for mode decision, $mode_{12}$ is one of the candidate modes during motion estimation, SSD is sum of the squared differences between *s* and its reconstruction block *c*, and $R_{mode}(s,c,mode_{12})$ represents the number of coding bits associated with the chosen mode. To compute J_{mode}, forward and inverse integer transforms, and variable length coding are performed. In the implementation of H.264 codec such as JM11.0 [7], the motion estimation of the secondary SP-frame uses $\hat{P}_{1,t-1}$ and the original $SP_{1,t}$ as the reference and current frames respectively. This arrangement allows the reuse of coding modes ($mode_{1,t}$ in Figure 1) and motion vectors ($mv_{1,t}$ in Figure 1) during secondary SP-frame encoding.

However, the reuse of coding modes and motion vectors reduces the coding efficiency of a secondary SP-frame since the purpose of the secondary SP-frame is to reconstruct $SP_{2,t}$ instead of $SP_{1,t}$. In [8], a secondary SP-frame is encoded to match the exact target frame (reconstructed $SP_{2,t}$, $\hat{SP}_{2,t}$) based on the exact reference ($\hat{P}_{1,t-1}$), By using the correct target and reference frames, better compression performance of secondary SP-frames can be achieved. Note that the computational complexity inevitably increases without reusing coding modes and motion vectors. Nevertheless, secondary SP-frames are always generated in off-line for bitstream switching applications. Thus, complexity is not the major concern for coding secondary SP-frames.

3. SIZE REDUCTION OF SECONDARY SP-FRAMES AMONG VIEWS

3.1. QDCT-domain motion compensated prediction

However, the storage reduction in [8] is not so significant, which prevents it from using in viewpoint switching in multiview videos. In this section, we explore the drawback of using the conventional pixel-domain motion estimation and compensation processes for secondary SP-frames. Figure 3 shows the encoding steps of a block in a P-frame using pixel-domain motion estimation. In this example, most of the transform coefficients become zero after transformation and quantization. This property paves the way for entropy coding.

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 30,2010 at 09:16:58 UTC from IEEE Xplore. Restrictions apply.

$$\begin{array}{c}SP_{2,t}\\ \begin{array}{|c|c|c|c|}\hline 247&247&249&210\\\hline 254&254&248&200\\\hline 254&254&210&195\\\hline 254&254&222&184\\\hline\end{array}\end{array} - \begin{array}{c}MC(P_{1,t-1})\\ \begin{array}{|c|c|c|c|}\hline 224&248&255&192\\\hline 255&193&228&255\\\hline 193&255&213&251\\\hline 212&193&208&204\\\hline\end{array}\end{array} = \begin{array}{c}\text{Residue}\\ \begin{array}{|c|c|c|c|}\hline 23&-1&-6&18\\\hline -1&61&20&-55\\\hline 61&-1&-3&-56\\\hline 42&61&14&-20\\\hline\end{array}\end{array} \xrightarrow{\text{T+Qs}} \begin{array}{|c|c|c|c|}\hline 1&1&-1&0\\\hline 0&-1&0&0\\\hline 0&0&0&0\\\hline 0&0&1&0\\\hline\end{array}$$

Figure 3. Motion-compensated prediction using pixel-domain motion estimation in encoding a P-frame.

In contrast, in [8], secondary SP-frame encoding performs transformation and quantization of original $SP_{2,t}$ and $\hat{P}_{1,t-1}$ first. Then, quantized coefficients of the secondary SP-frame at t, $Qs[T(SP_{12,t})]$, can be obtained as,

$$Qs[T[SP_{12,t}]] = Qs[T[SP_{2,t}]] - Qs[T[MC(\hat{P}_{1,t-1})]] \qquad (3)$$

where MC() is the motion-compensation operator.

$$\begin{array}{ccccc}\begin{array}{c}SP_{2,t}\\ \begin{array}{|c|c|c|c|}\hline 247&247&249&210\\\hline 254&254&248&200\\\hline 254&254&210&195\\\hline 254&254&222&184\\\hline\end{array}\end{array} & & \begin{array}{c}MC(P_{1,t-1})\\ \begin{array}{|c|c|c|c|}\hline 224&248&255&192\\\hline 255&193&228&255\\\hline 193&255&213&251\\\hline 212&193&208&204\\\hline\end{array}\end{array} & & \\ \big\downarrow \text{T+Qs} & & \big\downarrow \text{T+Qs} & & \text{Residue}\\ \begin{array}{|c|c|c|c|}\hline 15&1&-1&0\\\hline 0&0&0&0\\\hline 0&0&0&0\\\hline 0&0&0&0\\\hline\end{array} & - & \begin{array}{|c|c|c|c|}\hline 15&0&0&0\\\hline 1&0&0&1\\\hline -1&0&-1&1\\\hline 0&0&-2&-1\\\hline\end{array} & = & \begin{array}{|c|c|c|c|}\hline 0&1&-1&0\\\hline -1&0&0&-1\\\hline 1&0&1&-1\\\hline 0&0&2&1\\\hline\end{array}\end{array}$$

Figure 4. Motion-compensated prediction using pixel-domain estimation in encoding a secondary.

Figure 4 uses the same example in Figure 3 again to show the residue of a secondary SP-frame in which a block is transformed and quantized prior to calculating the residue. In this case, their quantized coefficients are only near, but not equal, resulting in generating many non-zero residue, especially for a small *Qs*. Since there is no further quantization from this point, these coefficients should be encoded completely. In entropy coding, even only one high-frequency coefficient exists, significant demanding of bits is required. Therefore, size of secondary SP-frames becomes large, and this explains the situation in which the pixel-domain motion estimation is not suitable for coding secondary SP-frames. This case always happens in multiview video sequences since V1 and V2 are from different views. Therefore, motion estimation becomes critical of generation secondary SP-frames in multiview video systems. In this paper, we thus propose performing motion estimation and compensation in the quantized transform (QDCT) domain rather than the pixel domain to improve the coding efficiency of secondary SP-frames.

3.2. A novel scheme for encoding secondary SP-frames in viewpoint switching

In this section, we propose a quantized transform-domain motion estimation (QDCT-ME) technique that minimizes $Qs[T[SP_{2,t}]] - Qs[T[MC(P_{1,t-1})]]$ (quantized transform domain) instead of $SP_{2,t} - MC(P_{1,t-1})$ (pixel domain). From (1), SAD between pixels of the original block *s* and its reference block *r* is used to compute the distortion of J_{motion}. The investigation in section 3.1 reveals that pixel-domain distortion measure is not appropriate for coding secondary SP-frames. In the proposed QDCT-ME, the Lagrangian cost function J_{motion} in (1) needs to be rewritten as

$$J'_{motion}(mv_{12}, \lambda_{motion}) = k \bullet SAQTD(s,r) + \lambda_{motion} \cdot R_{motion}(mv_{12} - pmv_{12}) \qquad (4)$$

where *SAQTD(s,r)* is now the sum of absolute differences between the quantized transform coefficients of the original block *s* and the quantized transform coefficients of its reference block *r*, and it can be defined as

$$SAQTD(s,r) = \sum \left| Qs[T(s)] - Qs[T(r)] \right| \qquad (5)$$

k is a scaling factor to compensate for the energy loss of *SAQTD(s,r)* due to the quantization.

For coding a secondary SP-frame, this distortion measure can find a better motion vector and mode for minimizing the residue, $Qs[T[SP_{12,t}]]$, in (3). Note that SAQTD is computationally intensive since all the pixel blocks are necessary to be transformed and quantized to QDCT domain. However, the complexity is not the major concern for secondary SP-frame encoding since this frame type is always encoded off-line for multiview videos. On the other hand, the accuracy of distortion measure increases the coding efficiency of secondary SP-frames which results in the significant reduction of the storage requirement in the video server.

Figure 5 shows the block diagram of applying our new QDCT-domain motion estimation technique in the secondary SP-frame encoder. The reference and target frames in the QDCT domain are the inputs of QDCT-ME. After the motion vectors for each block are obtained, a corresponding QDCT-domain motion compensation (QDCT-MC) is used to compute the motion-compensated frame, $Qs[T[MC(\hat{P}_{1,t-1})]]$. With $Qs[T[MC(\hat{P}_{1,t-1})]]$ and $Qs[T[SP_{2,t}]]$, as depicted in Figure 5, the residue $Qs[T[SP_{12,t}]]$ can then be calculated.

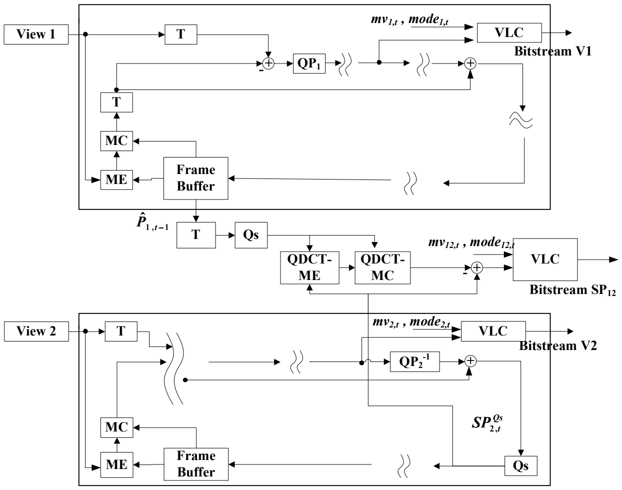

Figure 5. The proposed secondary SP-frame encoder in the QDCT-domain.

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 30,2010 at 09:16:58 UTC from IEEE Xplore. Restrictions apply.

4. EXPERIMENTAL RESULTS

To evaluate the performances of the proposed QDCT-ME and QDCT-MC scheme, three test sequences, "Ballroom" (VGA), "Exit" (VGA) and "Vassar" (VGA), were used in our experiments [9]. There are 8 views of testing sequences captured by 8 parallel-aligned cameras. For simplicity but without loss of generality, only view 0 and view 1 were selected to perform switching using SP-frames. The H.264 reference codec (JM11.0 [7]) was employed to encode primary SP-frames and secondary SP-frames with a frame rate of 30 fps. All test sequences have a length of 100 frames. These two different bitstreams from two different viewpoints were encoded with same sets of Q_P and Q_S. Q_P was varied from 16 to 28 whereas Q_S was set to $Q_P - 6$, which is the optimal setting according to [8]. Here, the scaling factor k of equation (4) was set to 3, which was found by experimental observations. To have fair comparisons between both schemes, every frame was encoded in turn as a SP-frame while non-switching frames were encoded as P-frames.

Figure 6 shows the average size reduction of secondary SP-frames with different Q_P. The values of the Y-axis mean our scheme has better size reduction of a secondary SP-frame in percentage difference over the scheme in [8]. From Figure 6, it is observed that the proposed scheme can substantially reduce the size of secondary SP-frames, up to 8%, 10% and 4% in "Ballroom", "Exit" and "Vassar", respectively. The significant improvement of the proposed scheme is due to the benefit of performing motion estimation and compensation in the QDCT domain. In [8], even though a proper target frame is selected for motion estimation, the performance is still not significant. It is due to the reason that only the conventional pixel-domain motion estimation technique is employed for coding secondary SP-frames. In encoding two different views, most of transformed coefficients become non-zero after transformation and quantization, as shown in Figure 4, which unfavour the use of entropy coding. Consequently, more bits are required to encode secondary SP-frames. On the other hand, our proposed scheme produces secondary SP-frames using motion estimation in the QDCT domain. The quantized and transformed coefficients are used to calculate the distortion in the Lagrangian cost function. The new SAQTD really finds the motion vector with more coefficients to be zero that benefits the entropy coding of secondary SP-frames. This provides the remarkable size reduction of our proposed scheme.

5. CONCLUSION

In this paper, an efficient scheme for switching bitstreams among different viewpoints has been proposed by using H.264 secondary SP-frames. The scheme can perform seamless switching even a different reference frame is used. Furthermore, the use of conventional pixel-domain motion estimation is not appropriate for a secondary SP-frame encoder, which incurs considerable size of secondary SP-frames in multiview videos. To alleviate this, we have incorporated the QDCT-domain motion estimation technique in the encoding process of secondary SP-frames. Experimental results show that the proposed scheme can significantly reduce the size of secondary SP-frames for the use of switching between two views. Besides, the proposed technique does not affect the coding efficiency of primary SP-frames.

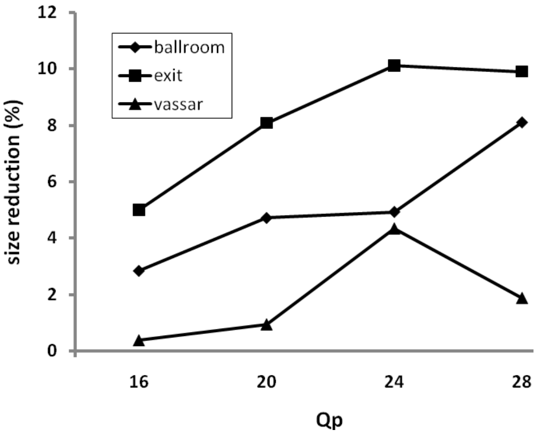

Figure 6. Size reduction of secondary SP-frames in percentage difference achieved by the proposed scheme over the scheme in [8].

6. ACKNOWLEDGEMENTS

The work described in this paper is partially supported by the Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (PolyU 5125/06E). Ki-Kit Lai acknowledges the research studentships provided by the University.

7. REFERENCES

[1] X. Guo, Y. Lu, W. Gao and Q. Huang, "Viewpoint Switching in Multiview Video Streaming," IEEE ISCAS, vol 4, pp. 3417-3474, China, May 2005.

[2] Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, "ITU-T Recommendation H.264 Advanced video coding for generic audiovisual services," 2005

[3] C.P. Chang and C.W. Lin, "R-D optimized quantization of H.264 SP-frames for bitstream switching under storage constraints," IEEE ISCAS, pp. 1241-1235, China, 2005.

[4] X. Sun, S. Li, F. Wu, K. Shen and W. Gao, "The improved SP frame coding technique for the JVT Standard," ICIP, III, vol 2, pp. 297-300, September 2003

[5] R. Kurceren and M. Karczewicz, "Synchronization-Predictive coding for video compression: The SP frames design for JVT/H.26L," ICIP, pp. 497-500, USA, 2002

[6] R. Suhring, T. Wiegand and H. Schwarz, "The emerging H.264/AVC standard," EBU Technical Review, 2003

[7] K. Suhring, H.264 Reference Software JM11.0, 2006, http://iphone.hhi.de/suehring/tml/

[8] W.T. Tan and B. Shen, "Methods to improve coding efficiency of SP frames," IEEE ICIP, pp. 1361-1364, Atlanta, USA, October 2006

[9] ISO/IEC JTC1/SC29/WG11 MPEG05/m12077, "Multiview Video Test Sequences from MERL," April 2005.

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 30,2010 at 09:16:58 UTC from IEEE Xplore. Restrictions apply.