TELKOMNIKA, Vol.11, No.3, September 2013, pp. 575~582
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v11i3.1049

A New Transcoding Scheme for Scalable Video Coding to H.264/AVC

Zhenyu Wu*[1], Daiying Zhou[1], Hong Hu[2]
[1]University of Electronic Science and Technology of China
Xiyuan Ave, Western High-tech, Chengdu, Sicuan, China
[2]Huawei Inc. Xiyuan Ave, Western High-tech, Chengdu, Sichuan, China
*Corresponding author, e-mail: zywu813cn@yahoo.com.cn

Abstract

Permintaan dari berbagai terminal video mendorong server video untuk melengkapi dengan skalabilitas untuk video terdistribusi dengan cara yang berbeda. Video Coding Scalable (SVC) sebagai perpanjangan dari standar H.264/AVC dapat menyediakan skalabilitas untuk server video dengan encoding video ke dalam satu lapisan dasar dan beberapa lapisan tambahan. Untuk mengaktifkan perangkat mobile tanpa skalabilitas menerima video pada batas terbaik mereka, mengubah bit-stream dari SVC ke H.264/AVC adalah menjadi teknik kunci. Bit-stream ulang adalah cara paling sederhana tanpa kehilangan kualitas. Namun, menulis ulang bukan skema transcoding nyata, karena kebutuhan untuk memodifikasi encoders SVC. Makalah ini mengusulkan sebuah pendekatan baru untuk mendukung transcoding skalabilitas spasial dengan meminimalkan distorsi yang dihasilkan dari proses re-encoding. Skema yang diusulkan terus memasukan informasi bit-stream secara maksimum dan mengadopsi metode hybrid upsampling untuk melakukan teknik scaling residu, yang dapat mengurangi distorsi transcoding ke minimalisasi. Hasil penelitian menunjukkan bahwa hilangnya tingkat-distorsi (RD) kinerja skema transcoding yang diusulkan lebih baik daripada metode Full Decoding Re-encoding (FDR), dimama hasil ini mendapatkan kualitas video tertinggi dalam arti umum, dengan mencapai hingga 0,9 dB Y-gain PSNR sambil menyimpan 95% ~ waktu proses 97%.

Kata kunci: *SVC, H.264/AVC, transcoding, menulis ulang*

Abstract

Requests from various video terminals push video servers to equip with scalability for video contents distribution in different ways. Scalable Video Coding (SVC) as the extension of H.264/AVC standard can provide the scalability for video servers by encoding videos into one base layer and several enhancement layers. To enable mobile devices without scalability receive videos at their best extent, converting bit-streams from SVC into H.264/AVC becomes the key technique. Bit-stream rewriting is the simplest way without quality loss. However, rewriting is not a real transcoding scheme, since it needs to modify SVC encoders. This paper proposes a novel transcoding approach to support spatial scalability by minimizing the distortions generated from re-encoding process. The proposed scheme keeps the input bit-streams' information at maximum and adopts the hybrid upsampling method to do residue scaling, which can reduce the transcoding distortion into minimization. Experimental results demonstrate that the loss of the rate-distortion (RD) performance of the proposed transcoding scheme is better than Full Decoding Re-encoding (FDR) which can get the highest video quality in general sense, by achieving up to 0.9 dB Y-PSNR gain while saving 95%~97% processing time.

Kata kunci: *SVC, H.264/AVC, transcoding, rewriting*

1. Introduction

H.264/AVC has been widely adopted for several years. Scalable Video Coding (SVC), as the scalable extension of H.264/AVC, has been recently approved to be much more complicated for both encoder and decoder devices. According to recent research [1] , scalable video coding is one of trends for video coding in the near future. However the number of decoder equipments is much larger than encoders. To consider the afford abilities of low-end users, the cost of decoders must be low enough. Therefore, it is foreseen that during a relative long time in the future, there will be much more H.264/AVC decoders in use than SVC decoders, especially for portable devices. It will lead to applications where both SVC and H.264/AVC are adopted. For example, in a live broadcasting or video conference, some

Received March 12, 2013; Revised June 13, 2013; Accepted July 24, 2013

receivers are equipped with SVC decoders, while others are with H.264/AVC ones. When the content provider sends SVC bit-streams, which can be with QVGA to WXGA spatial scalabilities for instance. Typical H.264/AVC receivers can only render QVGA services wherever how strong computation and display capability they have. To meet the requirement of those high-end users, a transcoder suggests being equipped at the media gateway (shown in 0) to convert the spatial scalable SVC bitstreams into H.264/AVC ones in WXGA or other higher spatial resolutions than base layer. Otherwise, the video server must send several H.264/AVC bitstreams with various spatial resolutions, which results in much higher network bandwidth , computation and storage requirements.

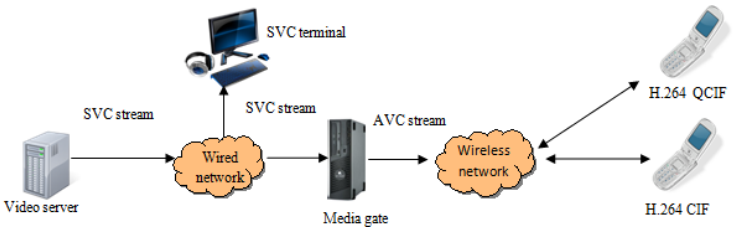

Figure1. Transcoding application

Straightforward approach to convert SVC bit-streams into H.264/AVC is FDR scheme, which fully decodes SVC streams and re-encodes into H.264/AVC one. As well known, although FDR scheme can provide highest video quality, it is typically not implementable due to high complexity. In addition to FDR, just a few researches have been made on SVC to H.264/AVC transcoding. Cascade [1] is one of them. It has reduced computation complexity quite a lot with probably sacrificing the quality of transcoded streams. Hannuksela[3] worked out a directed way to convert SVC into AVC by rewriting SVC headers, if SVC layers are independent. Segall [4] proposed a bit-stream rewriting method for CGS (Coarse Grain Scalability) case, wherein inter-layer prediction mechanism was modified and transform-domain intra-prediction was introduced as well as some other constraints.

Schemes mentioned above can transform SVC into AVC without drift. However they only focus on temporal or quality converting. Moreover, those rewriting schemes should modify the SVC encoding and decoding processes. So they cannot be real transcoders. This paper introduces a new spatial transcoding scheme from SVC to H.264/AVC. In the proposed scheme, we utilize the original information as much as possible to speed up transcoding process with quality improving at the same time. Furthermore, the hybrid up-sampling method [5] is adopted to scale the residues into expect substantial achievement of better RD performance than FDR and Cascade (which re-encodes video streams without motion estimation) schemes with inherently low computational complexity.

The rest of paper is organized as: a brief introduction of SVC is given in Section 2. Section 3 describes the proposed transcoding scheme for SVC to AVC in spatial classification. Section 4 presents the experimental results. Section 5 concludes the whole paper with a summary.

2. Brief introduction to SVC and spatial scalable transcoding

As the scalable extension of H.264/AVC, SVC inherits all the coding tools of AVC therein to guarantee the performance of SVC. Meanwhile, with the layered coding structure, various inter-layer prediction techniques such as inter-layer motion prediction, inter-layer residual prediction, and inter-layer intra prediction were designed for SVC to fully employ the inter-layer correlation so as to improve coding efficiency. SVC supports mainly three types of scalabilities (temporal, spatial and quality scalability) and implements them by layered coding structure. In the base layer, SVC encodes video sequences with the lowest frame rate, smallest frame size and largest quantization step in the same way as H.264/AVC. Meanwhile, SVC encodes higher levels of sequences in enhancement layers by inter-layer prediction mechanisms, which include inter-layer motion prediction, inter-layer residual prediction and

inter-layer intra-prediction. More details on SVC can be found in [7] .

To support spatial scalable coding, SVC follows multilayer coding. Each layer corresponds to a supported spatial resolution. In each spatial layer, motion-compensated prediction and intra-prediction are employed for single-layer coding. But in order to improve coding efficiency in comparison to simulcasting different spatial resolutions, so-called inter-layer prediction mechanisms, which include inter-layer motion prediction, inter-layer residual prediction and inter-layer intra-prediction, are incorporated additionally illustrated in 0. Model decision is one of the key techniques in spatial scalability of SVC. Kim[6] proposed a fast mode decision algorithm by classifying the MBs into three different categories, which can reduce encoding time up to 63 %. More details on spatial scalability of SVC are reported in [8] .

According to the SVC decoding process for inter-frame with the same spatial resolution between base layer and enhancement layer, the current frame can be reconstructed as (1):

$$f_n^i = T^{-1}(Q^{-1}(R_E)) + T^{-1}(Q^{-1}(R_B)) + f_k^j \quad (1)$$

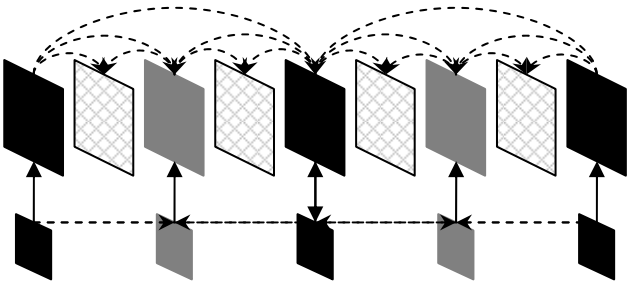

Figure2. Multilayer structure with additional inter-layer prediction for enabling spatial scalable coding

where f_k^j is taken as the pixel values of predicted macroblock of i[th] macroblock in n[th] frame in base layer, R_E as the residue of the corresponding macroblock in enhancement layer, and R_B as the residue of the reference macroblock in base layer in inter-layer prediction. $T^{-1}(\square)$ and $Q^{-1}(\square)$ denote the inverse transform and inverse quantize.

In [9] , an algorithm of rewriting SVC bit-stream for quality scalability was proposed. The decoding process can be formulated as (2):

$$f_n^i = T^{-1}(Q^{-1}(R_E + Scale(R_B))) + f_k^j \quad (2)$$

where $Scale(\square)$ denotes scaling the quantized coefficients, since the QPs of two layers are different.

If the quality enhancement layer is rewritable, the residue of H.264/AVC bit-stream can be obtained by (3).

$$R_{AVC} = R_E + Scale(R_B) \quad (3)$$

Then the reconstructed macroblock in enhancement layer can be got by (4).

$$f_n^i = T^{-1}(Q^{-1}(R_{AVC})) + f_k^j \quad (4)$$

According to (1)~(4), there is no drift introduced by the above rewriting operation. However, this method cannot be applied to spatial scalable case, because we cannot rewrite the corresponding residue of base layer into enhancement layers by $Scale(\square)$ operation. In the next section, we propose a new transcoding method for SVC to H.264/AVC in spatial scalability.

3. Proposed spatial transcoding scheme from SVC to H.264/AVC

Cascade and FDR are two typical structures of close-loop transcoding system which can get the highest RD performance in general sense, by fully decoding and then re-encoding with or without motion estimation and mode decision.

However re-encoding won't always get the highest transcoding quality, besides high computation consumption. According to the principle of video encoding, the decoded block must be "more look like" the coding block S, which determines model decision (*MD*), motion vector (*MV*) and residue. On the other hand, re-encoding will decrease the video quality even with the same QP values of input streams. We shall describe the above judgments in detail as follows.

Let us present the current reconstructed block in the input stream as $\hat{S}$, its reconstructed reference frame as C, motion vector as MV_0 and the corresponding residue is R. So $C(MV_0)$ represents the best matched block of current block in reference frame extracted from input stream. Then $\hat{S}$ and R can be got by (5) and (6):

$$\hat{S} = C(MV_0) + Q^{-1}\left(Q\left(R\right)\right) \quad (5)$$

$$R = S - C(MV_0) \quad (6)$$

Suppose FDR structure is taken to transcode and without considering accumulate errors, then the corresponding reference frame is still to be C. The re-encoded motion vector is denoted as MV_1, while the reconstructed best matched block is represented as $C(MV_1)$ and residue as R'. The reconstructed re-encoded block $\hat{S}'$ and residue R' can be written as (7) and (8).

$$\hat{S}' = C(MV_1) + Q^{-1}\left(Q\left(R'\right)\right) \quad (7)$$

$$R' = \hat{S} - C(MV_1) \quad (8)$$

Submitting（5）into（8）, we get R' as (9).

$$R' = C(MV_0) - C(MV_1) + Q^{-1}\left(Q\left(R\right)\right) \quad (9)$$

According to (9), the re-encoding residue must be probably larger than the input stream, when MV_1 is different from MV_0. Furthermore, if considering accumulative errors and re-quantization errors, FDR will introduce more quality decreasing, even with the same QP values as input stream.

In the other hand, if we re-use MV_0 of the input stream to transcode the necessary blocks. Let us denote residue as R_c which represented in (10), then the reconstructed block $\hat{S}_c$ can be got by (11).

$$R_c = \hat{S} - C(MV_0) \quad (10)$$

$$\hat{S}_c = C(MV_0) + Q^{-1}\left(Q\left(R_c\right)\right) \quad (11)$$

Submit（7）into（11）, we get:

$$R_c = C(MV_0) + Q^{-1}\left(Q\left(R\right)\right) - C(MV_0) = Q^{-1}\left(Q\left(R\right)\right) \quad (12)$$

Comparing (12) and (9), we can see clearly that the residue of transcoded block re-using side information (MVs) will be no larger than the one re-encoded by FDR. So $\hat{S}_c$ will be more look like the original un-coded block than $\hat{S}'$ without considering accumulate errors. That is why Cascade structure may get better RD performance than FDR in some cases, especially if re-encodes with fast motion estimation.

If consider accumulate errors, the corresponding reference frame becomes to be C'. The residue transcoded by FDR shall become:

$$R' = C(MV_0) - C'(MV_1) + Q^{-1}\left(Q\left(R\right)\right) \quad (13)$$

And the residue generated by re-using MVs will be:

$$R' = C(MV_0) - C'(MV_0) + Q^{-1}\left(Q\left(R\right)\right) \quad (14)$$

Comparing (9), (12), (14) with *R*, we can draw a conclusion that re-encoding will introduce extra errors into transcoded streams either re-using side information or re-encoding totally.

3.1 Proposed transcoding scheme

Based on the discussion above, we have kept the original blocks with no needs of transcoding in the input stream at maximum in the proposed scheme. The whole spatial transcoding scheme from SVC to H.264/AVC is illustrated in 0, which can be divided into "decoder" and "encoder" parts.

In the "decoder" part, the input SVC stream is first entropy decoded from the base layer to the enhancement layer whose spatial resolution is required, while higher layers will be skipped. Then side information such as model types, motion vectors is extracted. After that go through a transcoding judgment model, which is described in sub-section 3.2 to bypass those blocks needn't to be transcoded. Finally, inverse quantization Q^{-1} and inverse transform T^{-1} are used to get residues of referenced base layer and enhancement layers' blocks.

In the "encoder" part, residue up-sampling and fast MD/ME determination models are two key parts. Since residue upsampling is similar with image/video frame upsampling, methods for image/video frame upsampling can be adopted directly. Taking the tradeoff between computation complexity and quality, we implement the hybrid upsampling method to resize residue, which was described in our earlier work[5] . Fast MD/ME determination will be described in the subsection 3.3. Other encoding processes of the "encoder", including Transform (T), Quantization (Q), and Entropy Coding (*CABAC*) remain the same as H.264/AVC encoder.

3.2 Transcoding models' Judgement

In the proposed scheme, all MB modes are classified by judgment model into three types, and processed separately:

a) Normal modes with reference block unchanged: including all modes inherited from H.264/AVC, excluding skip and direct modes. The proposed scheme will output the current block with side information the same as input. For those blocks with reference blocks re-encoded, we should re-encode the current block through fast MD/ME determination model describes in sub-section 3.3.

b) Skip and direct modes: including all skip and direct modes inherited from H.264/AVC as well as the modified skip and direct modes in SVC. The proposed scheme will check whether their references MBs are re-encoded or not. If not, we will keep the current block with skip or direct modes, otherwise we should re-encode the current block through fast MD/ME model, for the reference MBs are not the exact ones.

c) Inter-layer prediction modes: including all modes for SVC special. There are two actions based on the principle described at the beginning of section 3. Firstly, if the current block is coded by inter-layer intra prediction (INTRA_BL), the proposed scheme will combine the upsampled base layer block with the residue in enhancement layer, through the residue upsampling method described in sub-section 3.3. While, if the current block is coded by inter-layer inter prediction (INTER_BL), we shall re-encode it by fast MD/ME determination.

3.3 Fast MD/ME (model decision and motion estimation) determination

In order to reuse the input information as much as possible, the proposed scheme will determine final MD/MV as following two steps.

i) The scaled MVs in referred base layer of input stream, and the MVs of neighbor blocks are all selected into candidate set. If the neighbor blocks have the same MVs, the proposed scheme will merge those blocks into a larger partition and put it inside the candidate set.

ii) etermine the final MD/MV from the candidate set got in step i) by minimizing Sum of Absolute Difference (SAD).

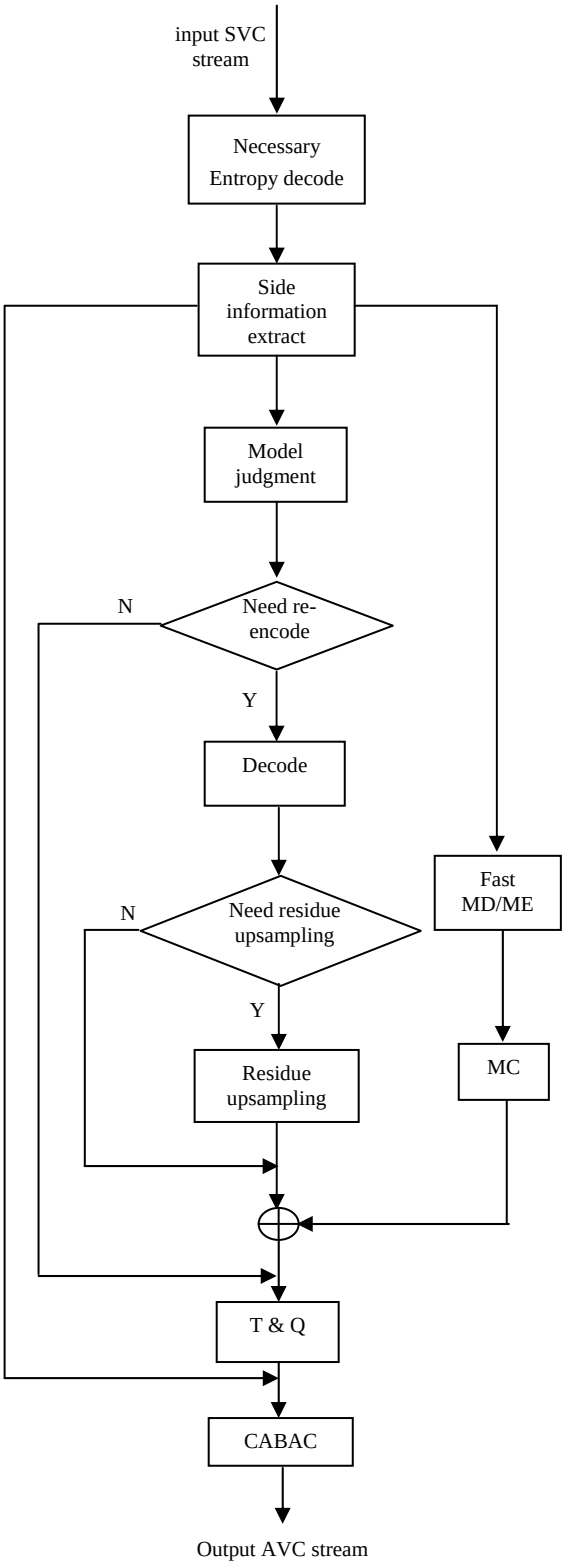

Figure 3. System diagram of the proposed transcoding scheme

4. Experimental results

The proposed spatial transcoding scheme from SVC to H.264/AVC was implemented with JSVM_9_17 and JM14.1[10] . Twenty standard video sequences have been tested. Test conditions for the proposed implementation are as follows:

- Two layers, base layer (layer 0) of QCIF@15Hz , one spatial enhancement layer (layer 1) of CIF@15Hz.
- Hierarchical B coding structure with GOP size 8.
- CABAC used
- Loop filter enabled
- 8x8 transform enabled
- The input QP values remained unchanged.

Since FDR is the transcoding scheme to get the highest quality traditionally and Cascade is another commonly used transcoding scheme, in the experiment part we only compare the proposed scheme with Cascade and FDR schemes using default fast ME in JM model (searching area defined as ±16 pixels).

Transcoding efficiency varies sequence by sequence. Due to lack of space, we only present rate-distortion (RD) curves for the best, the worst and middle cases of the proposed scheme in **Error! Reference source not found.**.

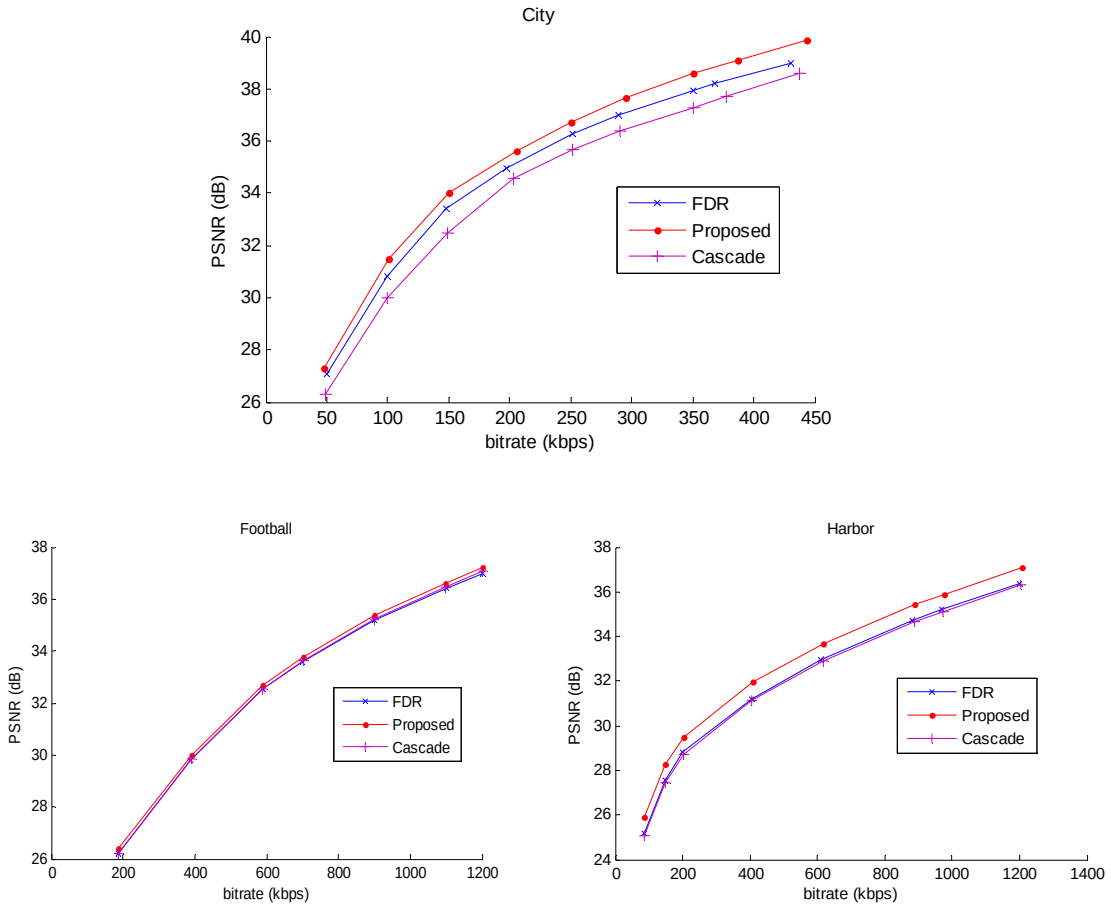

Figure 4. R-D comparison between FDR, Cascade and proposed scheme

The PSNRs are calculated between original un-coded video frames and transcoded ones. We can see that the proposed scheme can get 0.2dB~0.9dB Y-PSNR gains than FDR and 0.26~1.6dB than Cascade, which are meeting the analysis presented in section 3. The transcoding complexity of the transcoding schemes are shown in Table 1, by measuring the

running time consumed ratio in transcoding process. It is clearly that the proposed scheme is about 29 times in average faster than FDR and 1.17 times faster than Cascade.

Table 1. Time consuming ratio comparison for transcoding

sequence	***Ratio of transcoding time***		
	FDR	***Cascade***	***Proposed***
City	31.5	1.02	1
Football	23.1	1.31	1
Foreman	34.3	1.03	1
Mobile	28.7	1.15	1
Harbor	27.9	1.22	1
Soccer	28.6	1.32	1

5. Conclusion

A low-complexity scheme for transcoding spatial scalable SVC bit-streams into H.264/AVC was developed and presented in this paper. It keeps those non-interlayer prediction blocks without re-encoded reference blocks the same as input streams. And then adopts the proposed MD/MV determination method and hybrid residue upsampling approach to implement inter-layer prediction blocks' transcoding by reusing the original side information (MDs/MVs) as much as possible during the transcoding process. Experiment results have shown that the proposed scheme achieved up to 0.9dB gain in PSNR and around 29 times speedup than FDR scheme which shall get the best transcoding quality in general sense**.**

.

References

[1] Z. Shi, X. Sun, and F. Wu, "Spatially Scalable Video Coding For HEVC", IEEE Trans. On Circuits Syst. Video Technol., vol. 22, no. 12,pp. 1813-1826, Dec., 2012.
[2] A. Vetro, C. Christopoulos, and H. sun, "Video transcoding architectures and techniques: An overview", IEEE Signal Processing Magazine, vol. 20, pp. 18-29, Mar. 2003.
[3] M. M. Hannuksela, and Y.-K. Wang, Support for SVC Header Rewriting to AVC, Joint Video Team, Doc. JVT-W046, Apr. 2007.
[4] A. Segall, CE 8: SVC-to-AVC Bit-Stream Rewriting for Coarse Grain Scalability, Joint Video Team, Doc. JVT-V035, Jan. 2007.
[5] Z. Wu, H. Yu, and C. W. Chen, "A New Hybrid DCT-Wiener-Based Interpolation Scheme for Video Intra Frame Up-Sampling", IEEE Signal Process. Letters, vol. 17, no. 10, pp. 827-830, Oct. 2010.
[6] S. Kim, K. R. Konda, P. M.ah, and S. Ko,"Adaptive Mode Decision Algorithm for Inter Layer Coding in Scalable Video Coding", Image Processing, 2010. ICIP2010.
[7] MPEG, ITU-T IssUE Scalable Video Coding Standard-ISO/IEC 14496-10, Aug., 2007.
[8] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard", IEEE Trans on Circuits and Systems for Video Technology, vol. 17, No. 9, pp. 1103-1120, Sept. 2007.
[9] A. Segall and J. Zhao, "Bit-stream rewriting for SVC-to-AVC conversion", Image Processing, 2008. ICIP 2008.
[10] http://www.hhi.fraunhofer.de/fields-of-competence/image-processing/applications.html.