

Evaluation and Performance Comparison between JPEG2000 and SVC

Xin Lu¹ and Xuesong Jin^{1*}

¹*School of Electronics and Information Engineering, Harbin Institute of Technology, China*

²*School of Computer and Information Engineering, Harbin University of Commerce, China*

luxin5321@163.com, jinxuesong@163.com

Abstract

The scalable extension to the H.264/AVC has been established as a scalable video coding (SVC) standard, and JPEG2000 is a state of the art still image coding standard. The most distinctive difference between SVC and JPEG2000 is, in terms of spatial-frequency transformation, that SVC is based on the Discrete Cosine Transform (DCT) and JPEG2000 is based on Discrete Wavelet Transform (DWT). JPEG2000 outperforms its predecessor-JPEG which adopts DCT by nearly 30% in terms of compression performance. Despite the DWT has been successfully employed in still image coding, the latest video coding standard keeps on utilising DCT and the relevant techniques as its coding tools. In this paper, we investigate SVC for intra frame coding and JPEG2000 for still image coding, on video sequences with different resolutions and features. From the experimental evaluations, we attempt to gain deeper insight into the application scenarios of these two transform techniques. The experimental results demonstrate that at high bitrates, JPEG2000 outperforms SVC, and it is vice versa. We also note that the PSNR difference between SVC and JPEG2000 decreases with increasing the picture resolution. Therefore, JPEG2000 is more suited for high definition picture compression. The spatial scalability of SVC and JPEG2000 is also investigated in the evaluation, and the results show that SVC gains quite better performance than JPEG2000.

Keywords: *Scalable Video Coding(SVC), JPEG2000, Discrete Cosine Transform(DCT), Discrete Wavelet Transform(DWT), Still Image Coding*

1. Introduction

SVC is the scalable video coding extension based on H.264/AVC for supplying temporal, spatial and SNR scalability. It has been finalized as the video standard in July 2007 by the Joint Video Team (JVT) of the VCEG of ITU-T and the MPEG of ISO/IEC [1]. Compared with its predecessors, SVC achieves better compression performance in coding intra frame. In intra frame coding, SVC employs the spatial prediction within one frame, various block-size integer DCT transformation, and the CABAC for entropy coding [2]. Besides encoding tools mentioned above, the unique inter layer intra prediction technique is employed in SVC, which exploits the lower layer information for encoding enhancement layer signal [3]. These superior features improve the compression efficiency significantly. Intra frame coding in SVC makes use of the correlation in the current frame and its corresponding frame in the base layer without utilizing the information from the other reference frames. Therefore, it can be used for still image compression.

* Corresponding author

JPEG2000 is the first international still image compression standard based on wavelet transformation, which was developed by the Joint Photographic Experts Group (JPEG) committee [4]. At present, wavelets become powerful tools in signal analyzing, and has widely used in video and image coding fields. The predecessor of JPEG2000 is JPEG, which is based on 8×8 block DCT transform. Because of the defect of artificial blocking effect, DCT was discarded in the competition with DWT. The core encoding tools of JPEG2000 include DWT and entropy coding algorithm-Embedded Block Coding with Optimized Truncation (EBCOT). With these tools, JPEG2000 outperforms the JPEG in the term of compression performance by nearly 30% [5]. Besides the compression efficiency, JPEG2000 also provides several features e.g. intrinsic multi-resolution characteristic, which are useful in multimedia applications [6].

2. DCT in SVC Extension of H.264/AVC

Owe to the fine decorrelation and energy concentration capability of DCT, it has been adopted in the still image compression standard JPEG and the video compression standard such as MPEG-1/2, H.261/263, *et al.*, [7]. Similar to the previous video coding standards [8], SVC keeps on utilizing DCT as its transformation tools. In order to exploit the spatial correlation in intra frame coding, H.264 performs intra frame prediction in spatial domain by referring to the neighboring previously-coded blocks before DCT transformation. Subsequently, the residual signal is transformed by using the integer discrete cosine transformation which has the similar properties with DCT. However, different to the early video coding standard, SVC is based on the 4×4 block size but not 8×8 block size.

The integer DCT used in SVC has several features compared with DCT [9]:

- (1) Integer transformation (all the operations in the encoder and decoder can be completed in integer arithmetic, without any loss of decoding precision);
- (2) Without any mismatch between the transformation in the encoder and the inverse transformation in the decoder;
- (3) The core part of the transformation can be implemented using only additions and shifts operation;
- (4) Scaling multiplication is removed from the transformation and is incorporated into the quantizer, thus the number of multiplications can be reduced.

A 4×4 forward DCT transformation equation can be written as follows [10]:

$$\mathbf{Y} = (\mathbf{C}_f \mathbf{X} \mathbf{C}_f^t) \otimes \mathbf{E}_f = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{X} \\ \mathbf{X} \\ \mathbf{X} \end{bmatrix} \right) \otimes \begin{bmatrix} a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \\ a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \end{bmatrix} \quad (1)$$

Although, the above equation is an integer approximation of DCT, it can achieve almost the same compression performance with the DCT in the SVC Codec. Meanwhile, the integer DCT possesses many unique advantages. By observing the transform kernel, all of the elements in the matrix are equal to ±1 or ±2. It means that the multiplications can be implemented by using additions and left shifting operations. Since the scaling factor matrix \mathbf{E} scales the output of the kernel transformation, the scaling can be incorporated in the quantization procedure. Thus, there is no actual multiplication operation required in the transformation.

Therefore, the integer DCT adopted in the SVC standard achieves a great reduction in computing complexity for the transformation procedure.

3. DWT in JPEG2000

Compared with DCT, DWT has a better energy compaction capability and the feature of multi-resolution representation. Besides that, DWT is able to get rid of the critical

problem arisen in DCT-the blocking artificial. Therefore, DWT has been used in the latest still picture compression standard-JPEG2000. As the first wavelet-based image compression algorithm, JPEG2000 not only achieves superior compression performance, but also supports various kinds of features, e.g. multi-resolution representation, robustness to bit-errors, Region Of Interest (ROI) coding, and protective image security, et al. With the purpose of reducing the correlation between pixels, DWT decomposes the spatial image into a number of frequency subbands which represent the horizontal and vertical frequency components of the original image. Subsequently, the coefficients in each wavelet subband are quantized and coded independently with different coding strategies. Compared with DCT which is employed in JPEG codec, DWT provides a better time-frequency localization of a given image. Therefore, wavelet-based image coding has the superior compression performance to its predecessors.

Two kinds of coding schemes-lossy and lossless compression are supported in JPEG2000. The lossy compression utilize Daubechies 9/7 filter bank for irreversible transform, and the lossless compression uses Le Gall 5/3 filter bank for reversible transform.

The analysis and synthesis coefficients of Daubechies 9/7 and Le Gall 5/3 are listed in Table 1 and Table 2, respectively [Error! Bookmark not defined.].

Table 1. Daubechies 9/7 Analysis and Synthesis Lowpass and Highpass Filter Coefficients

n	Analysis filter coefficients		Synthesis filter coefficients	
	LPF	HPF	LPF	HPF
0	0.6029490182363579	1.11508705245699	1.115087052456994	0.6029490182363579
±1	0.2668641184428723	-0.5912717631142470	0.5912717631142470	-0.2668641184428723
±2	-	-	-0.05754352622849957	-
	0.07822326652898785	0.05754352622849957		0.07822326652898785
±3	-	0.09127176311424948	-0.09127176311424948	0.01686411844287495
	0.01686411844287495			
±4	0.02674875741080976			0.02674875741080976

Table 2. Le Gall 5/3 Analysis and Synthesis Lowpass and Highpass Filter Coefficients

n	Analysis filter coefficients		Synthesis filter coefficients	
	LPF	HPF	LPF	HPF
0	6/8	1	1	6/8
±1	2/8	-1/2	1/2	-2/8
±2	-1/8			-1/8

4. Core Coding Tools in SVC Intra-frame Coding

Intra frame coding denotes that the video pictures are compressed only by exploiting the redundant in spatial domain. SVC standard keeps on using the intra prediction coding to reduce the spatial redundant. Depending on the spatial correlation of the adjacent macroblocks, the coded macroblocks which are located to the above and to the left of the current macroblock are utilised to predict the current macroblock. Consequently, the DCT and the entropy coding are applied to the prediction error of the current macroblock and the prediction value. Mode types of intra 4×4 and intra 16×16 are supported in SVC. There are nine prediction patterns in intra 4×4 type and four prediction patterns in intra 16×16 type. The prediction value is predicted from the spatially adjacent macroblock and the base layer. The small block type intra 4×4 is intuitionally suitable for coding the pictures with significant detail; on the other hand, the big block type intra 16×16 is fit for coding very smooth regions in a picture.

For example, the 16 samples labeled as A~Q are reconstructed samples in neighboring blocks in the intra 4x4 prediction, which are used as prediction reference samples. The samples labeled as a~p denote the samples to be predicted. By using one of the nine prediction patterns, the prediction values can be obtained from the recent-coded neighboring samples. Figure 1 shows the relative position for A~Q and a~p. Nine prediction patterns of intra 4x4 type are shown in Figure 2.

Q	A	B	C	D	E	F	G	H
I	a	b	c	d				
J	e	f	g	h				
K	i	j	k	l				
L	m	n	o	p				

Figure 1. Relative Position for A~Q and a~p

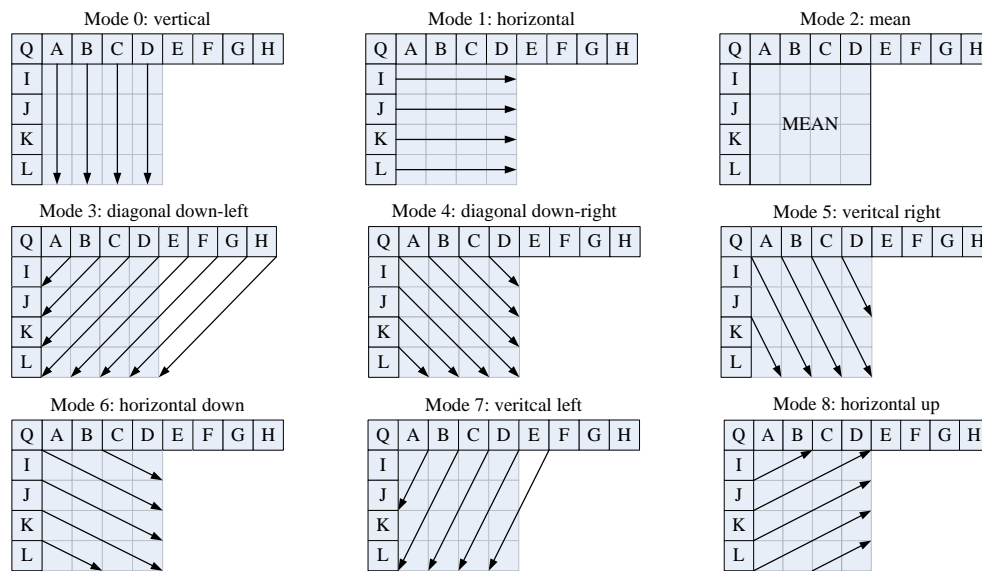


Figure 2. Nine Prediction Patterns of Intra 4x4 Type

In order to support the spatially scalable video coding, SVC continues to employ the multiple-layer coding scheme, which is used in conventional video coding standards including MPEG-2, H.263, and MPEG-4 visual. The SVC bitstream is composed of one base layer (BL) which is referred to the spatial layer with the lowest resolution and several enhancement layers (ELs) which are referred to the layers have higher resolutions compared with the base layer. Besides performing inter and intra prediction in each layer as in single layer coding, inter-layer prediction is also employed in SVC, which explores the reconstructed data of the lower layers [11]. Therefore, the prediction signal in enhancement layer intra coding is either formed by intra prediction signal or by upsampling the reconstructed lower layer signal.

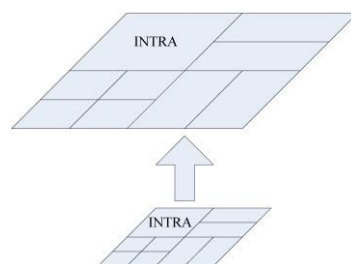


Figure 3. Inter Layer Prediction

In fact, SVC introduces three new prediction ways to reduce the inter-layer redundancy. In our experiment, the inter layer intra prediction is employed. Inter layer intra prediction is referred to when a sub-macroblock in the lower layer is intra-coded, the prediction signal for the enhancement layer macroblock can be obtained by up-sampling the corresponding reconstructed intra-signal of the lower layer, as shown in Figure 3.

5. Core Coding Tools in JPEG2000

JPEG2000 is mainly based on DWT, which possesses multi-resolution property. Furthermore, some other advanced image coding techniques such as EBCOT are involved in JPEG2000. The advanced coding tools make JPEG2000 outperform its predecessor JPEG by up to 30% in term of compression ratio. A brief coding procedure of JPEG2000 is described as follows [Error! Bookmark not defined.]:

- (1) Data pre-processing: prior to applying DWT, colour components transformation, DC level shifting, and tile partition are performed on the original image.
- (2) DWT: the wavelet transform is applied on each tile to obtain multi-level wavelet coefficients.
- (3) Quantisation of transform coefficients: the subbands of coefficients are quantised and collected into “code blocks”.
- (4) Entropy coding: The quantised coefficients are sent to the context-based binary arithmetic encoder for entropy coding.

6. Experiment Descriptions

In order to compare the coding efficiency of the two scalable coding schemes—SVC intra coding and JPEG2000, the Joint Scalable Video Model (JSVM) 9.18 is chosen as the evaluation model for SVC, and Kakadu 2.2.3 is selected for JPEG2000.

The performance in terms of PSNR and bitrates are evaluated in our experiments. SVC uses intra-prediction and inter-layer prediction for intra coding. CABAC is selected for entropy coding. JEPG2000 sets the DWT decomposition levels to 3 for CIF sequence, 4 for 4CIF, 5 for 720P, 6 for 1080P, 7 for 2160P video sequences.

In order to make a comprehensive comparison, we choose the video sequences with all the commonly used resolutions from CIF to 2160P (HD). In each resolution, 2~5 standard testing sequences with different picture features are chosen. The video pictures and resolutions used in the comparison are listed in Table 3.

Table 3. Video Sequences used in the Evaluations

Resolution	Video sequence
CIF (352×288)	bus(150f), foreman(300f), mobile(300f), mother-daughter(300f)
4CIF (704×576)	city(300f), crew(300f), harbor(300f), ice(240f), soccer(300f)
720P (1280×720)	duck(90f), park(90f), town(90f), tree(90f)
1080P (1920×1080)	sky(90f), rush(90f)
2160P (3840×2160)	crowd(90f), park(90f), town(90f)

Multi-resolution is an intrinsic property of DWT. SVC is also designed for supporting several spatial resolutions by using multi-layer coding structure. Therefore we intend to compare the spatial scalability in this paper.

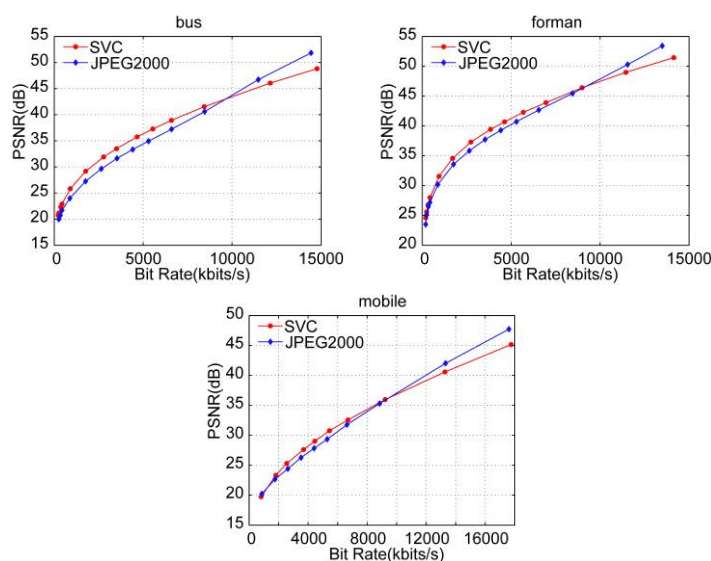
In kakadu2.2.3, the bits for per pixel can be directly set. However in SVC, to control the bitrates, we have to adjust the Quantisation Parameter (QP) to get the corresponding bitrates with kakadu2.2.3. As JPEG2000 is a still image coding standard, it has to process every single frame independently. To compress an image set, the bitrates of JPEG2000 can be calculated by the following equation:

$$bitrates = \left(\frac{bits}{pixel} \right) \times FrameWidth \times FrameHeight \times FrameRate \quad (2)$$

7. Experiment Results and Analysis

The performance is measured by PSNR of the Y-component according to various compressed bitrates. For the five sets of testing video sequences, the rate distortion curves are presented and analyzed.

For the same resolution reconstructed picture, when the bitrates is high, JPEG2000 outperforms SVC in terms of Y-PSNR, but when the bitrates is low, it is vice versa. This result can be found for all the video sequences. It can be seen from the figures, there is an intersection between the RD curves of SVC and JPEG2000. When the bitrates is less than this intersection, SVC gains better performance than JPEG2000. However, when the bitrates surpasses the intersection, JPEG2000 predominates in PSNR. The position of the intersection is closely linked with the resolution and the picture content. For the lower resolution pictures, the value of intersection is relative small, and it is vice versa. We also note that the PSNR difference between SVC and JPEG2000 decreases with increasing the picture resolution. For the CIF video sequences, the difference in PSNR can achieve by up to 3dB. While for the HD video the difference is less than 1 dB. We have investigated the internal reasons caused such phenomenon. DWT is applied to the entire picture, and the picture energy is concentrated into the low frequency subband. At low bitrates, almost all of the high frequency energy, which represents the significant detail information, and parts of the low frequency are lost, which leads to the overall PSNR drops dramatically. However, since DCT in SVC is based on 4×4 block, the information lost is restricted to the local block. Therefore, AVC gains better performance than JPEG2000 at low bitrates. At high bitrates, the information lost is minor, considering DWT is a global transformation, but DCT is a local transformation. DCT cannot avoid the boundary blocking effect, which is introduced by block-based transformation. Based on the above analysis, we can conclude that JPEG2000 is fit for high bitrates picture compression; SVC is more suited for low bitrates transmission. Therefore, JPEG2000 is suitable for high resolution picture compression, which is one of the reasons why digital cinema video selects JPEG2000 instead of AVC as its video coding standard.



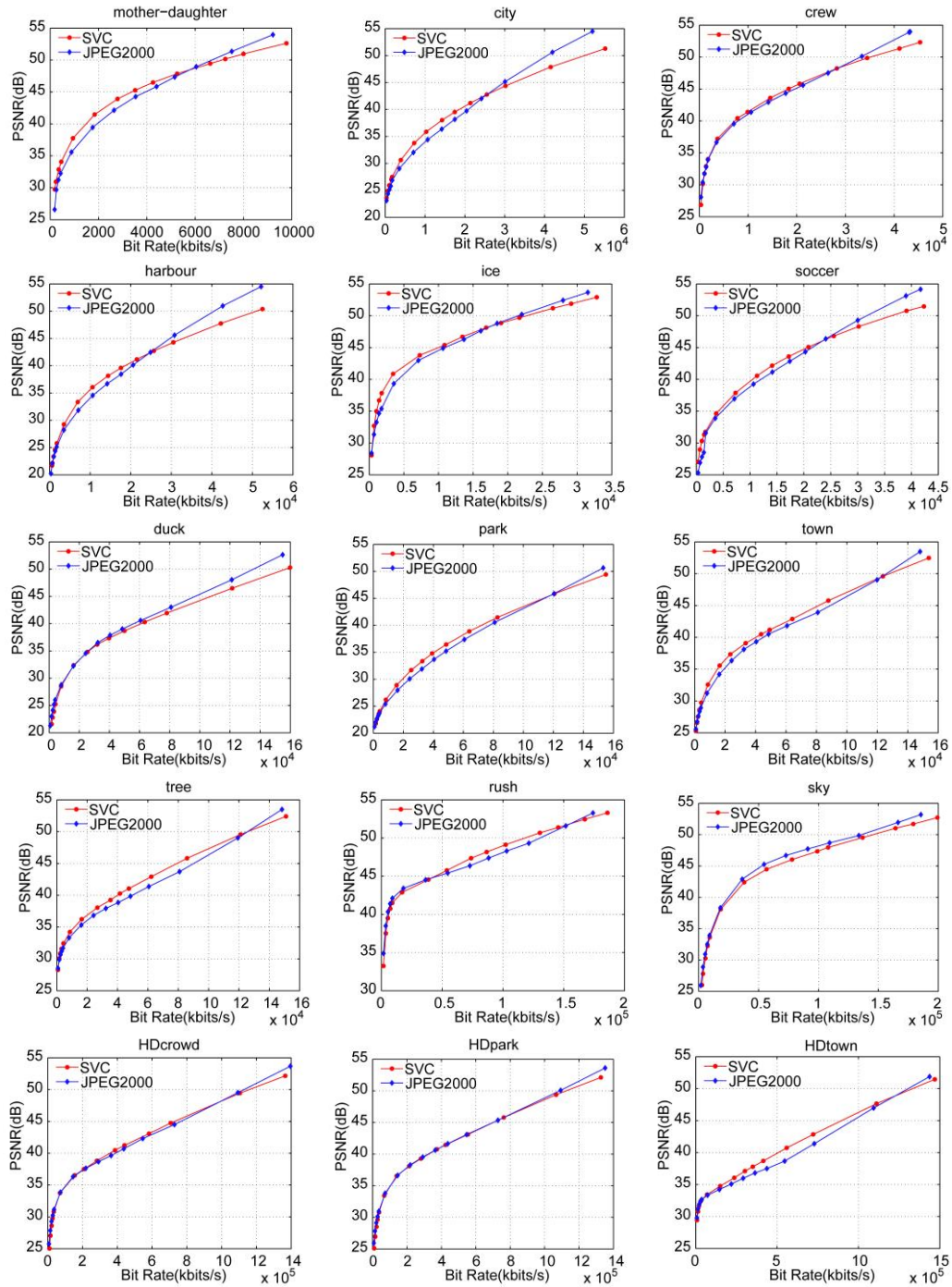
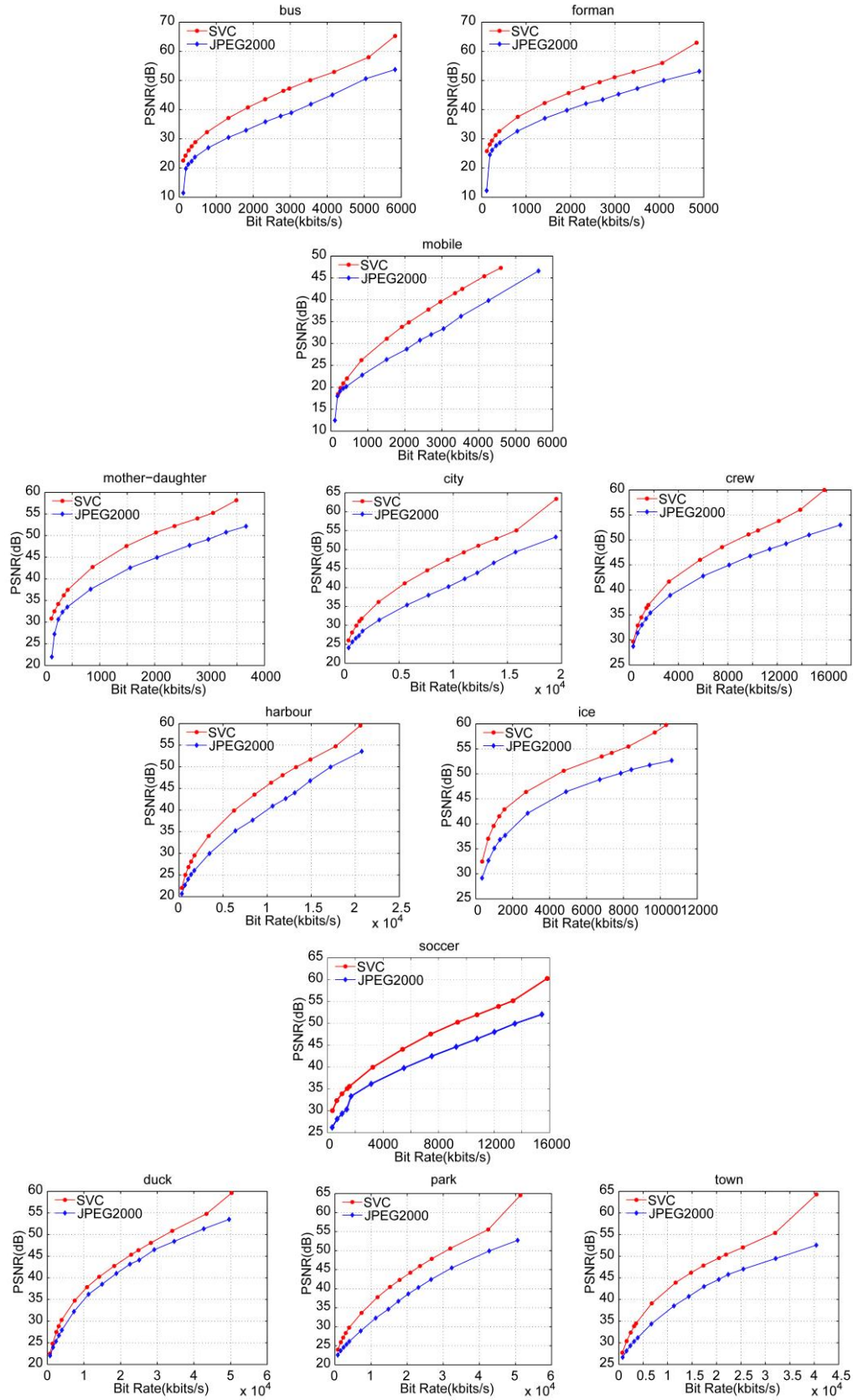


Figure 4. Rate Distortion Curves of the Full Construction for Various Video Sequences

Both JPEG2000 and SVC support spatial scalability. In order to evaluate the scalability performance, we extract the first level spatial picture from the compressed bitstream. From the results presented below, we note that SVC outperforms JPEG2000 over all the bitrates for all the video sequences.



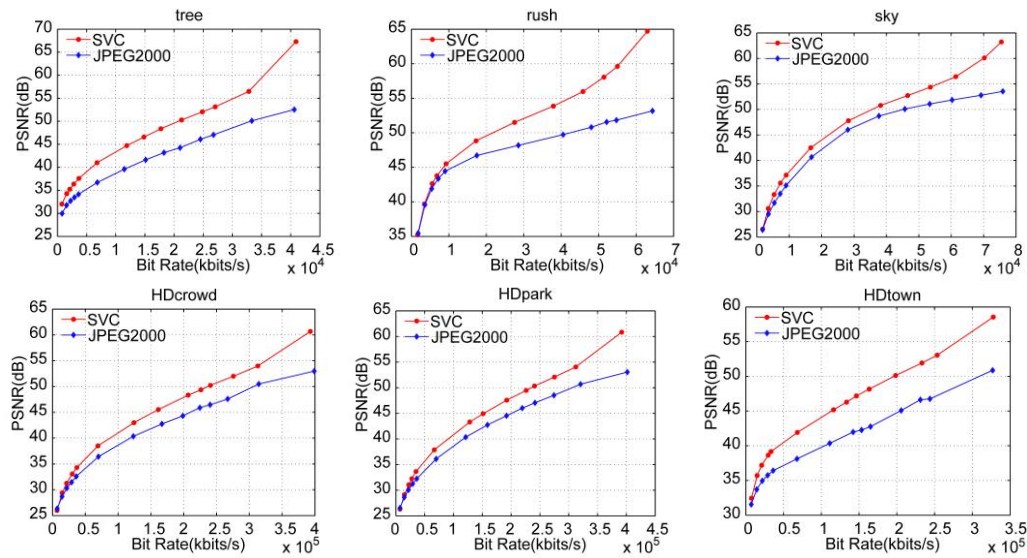
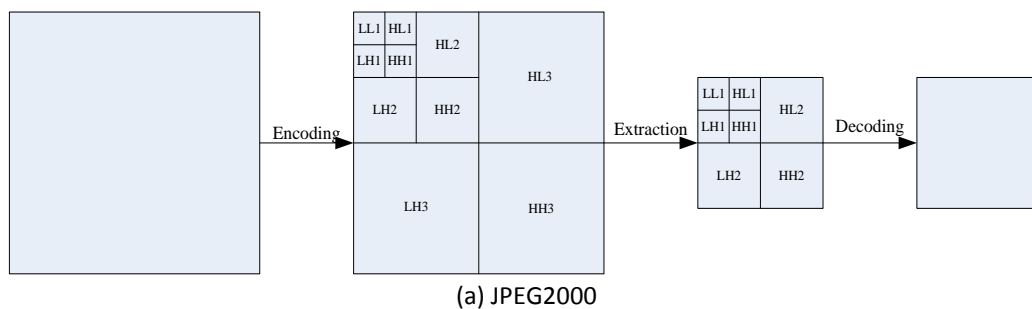


Figure 5. Rate Distortion Curves of the Spatial Extraction for Video Sequences

In SVC standard, the bottom-up layered structure is employed to fulfill the spatial and quality scalabilities. The video sequence is derived by downsampling the original video, and is coded using H.264/AVC compliant encoder. Thus, the coding performance of the SVC base layer is exactly the same with that of H.264/AVC. Different to SVC, JPEG2000 uses a top-down coding structure, picture is decomposed horizontally, and vertically to spatial subbands. The defect of top-down structure is that it is not suitable at lower resolutions, since all the decomposition is conducted at the full resolution. The reason is attributed to the way to obtain the reconstructed lower spatial resolution video. In SVC, the reduced resolution picture is obtained by reconstructing the downsampled picture as the way in H.264/AVC. However in JPEG2000, the reduced resolution picture is the first level wavelet decomposition picture. It is an approximation of the downsampled picture of the original one. Obviously, SVC can reconstruct the reduced resolution picture more precise. The approaches of SVC and JPEG2000 to get the reduced resolution picture are depicted as follows.



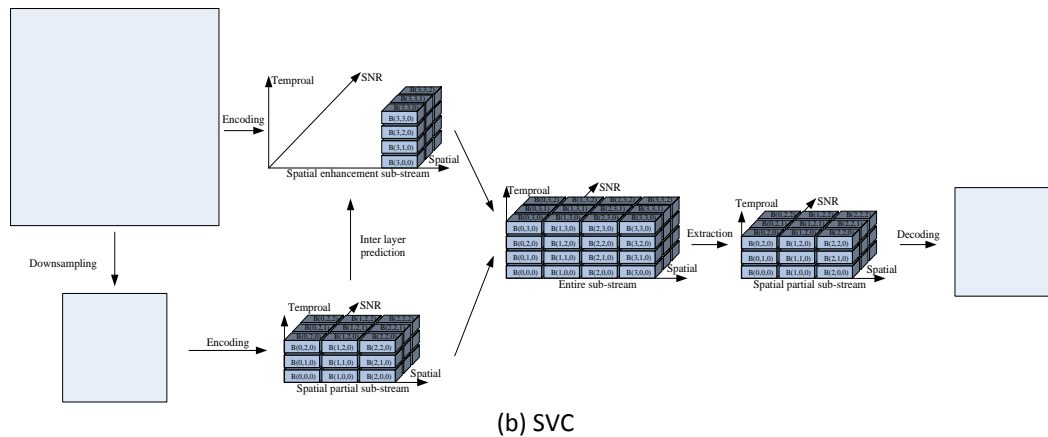


Figure 6. Approaches to Reconstruct the Reduced Spatial Resolution Picture

8. Conclusion

In this paper, the performance of the DCT-based SVC standard and the wavelet-based JPEG2000 were investigated thoroughly. We compared the coding efficiency using a variety of video pictures with resolutions from CIF to 2160P. The experimental results show that at lower bitrates, SVC can achieve better performance than JPEG2000. However when the bitrates is high, JPEG2000 outperforms SVC for all the pictures. We also noted that the PSNR difference between SVC and JPEG2000 decreases with increasing the picture resolution. Therefore, JPEG2000 is more suited for high definition picture compression. In addition, this is the first time in the literatures to evaluate the performance of spatial scalability. Observing the results, it shows at high bitrates SVC achieves 5 dB PSNR gains over JPEG2000 on average. Thus we draw a conclusion that SVC provides better performance than JPEG2000 in terms of spatial scalability.

Acknowledgements

This work has been supported by the National Natural Science Foundation of China (NSFC) under the projects of No. 61401123, and Harbin Science and Technology Bureau under project No. 2014RFQXJ166.

References

- [1] "Advanced Video Coding for Generic Audiovisual Services", ITU-T Rec. H.264 and ISO/IEC 14496-10 Advanced Video Coding, Joint Video Team (JVT) of ITU-T VCEG and ISO/IEC MPEG, (2010)
- [2] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard", IEEE Trans. on Circuits Syst. Video Technol., vol. 13, no. 7, (2003).
- [3] C. A. Segall and G. J. Sullivan, "Spatial scalability within the H.264/AVC scalable video coding extension", IEEE Trans. on Circuits Syst. Video Technol., vol. 17, no. 9, (2007).
- [4] A. Skodras, C. Christopoulos and T. Ebrahimi, "The JPEG 2000 still image compression standard", IEEE Signal Processing Magazine, vol. 18, no. 5, (2001).
- [5] C. Christopoulos, A. Skodras and T. Ebrahimi, "The JPEG2000 still image coding system: an overview", IEEE Trans. on Consumer Electronics, vol. 46, no. 4, (2000).
- [6] A. Samet, M. A. Ben Ayed, M. Loulou and N. Masmoudi, "Comparison between JPEG and JPEG2000 still image compression standard", Proc. Vis., Imag., Image Process, (2002) September 9-12, Marbella, Spain.
- [7] H. S. Malvar, A. Hallapuro, M. Karczewicz and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC", IEEE Trans. on Circuits Syst. Video Technol., vol. 13, no. 7, (2003).
- [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 Syst. Video Technol., vol. 17, no. 9, (2007).
- [9] A. Hallapuro and M. Karczewicz, "Low Complexity Transform and Quantization – Part I: Basic Implementation", JVT document, JVT-B038, Geneva, (2002) February.

- [10] I. E. G. Richardson, "H.264/MPEG-4 Part 10 White Paper: transform and quantization", www.vcodex.com, (2003) May.
- [11] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the scalable H.264/MPEG4 AVC extension", IEEE Int'l Conf. on Image Processing, (2006) October 161-164, Atlanta, USA.

Authors

Xin Lu, received the B.Sc. and M.Sc. degrees from the Harbin Institute of Technology, Harbin, China, in 2008 and 2010, respectively, and the Ph.D. degree in computer science from the University of Warwick, Coventry, U.K., in 2013.

He is currently a lecturer with the School of Electronics and Information Engineering, Harbin Institute of Technology. His current research interests include video coding standards, data compression, and image or video processing.

Xuesong Jin, received the M.Sc. and PhD. degrees from the Harbin Institute of Technology, Harbin, China, in 2004 and 2013, respectively.

He is currently an associate Professor with the School of Computer and Information Engineering, Harbin University of Commerce. His current research interests include image processing, video compression and pattern recognition.

