J. Vis. Commun. Image R. 39 (2016) 65–81

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Efficient multiple-example based super-resolution for symmetric mixed resolution stereoscopic video coding ☆

Kuo-Liang Chung [a,1], Yong-Huai Huang [b,*,2]

[a] Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Road, Taipei 10672, Taiwan, ROC
[b] Department of Electronic Engineering, Jinwen University of Science and Technology, No. 99, AnChung Rd., Xindian Dist., New Taipei City 23154, Taiwan, ROC

ARTICLE INFO

Article history:
Received 3 January 2016
Revised 13 May 2016
Accepted 15 May 2016
Available online 19 May 2016

Keywords:
Asymmetric resolution stereoscopic video coding (ARSVC)
Bitrate reduction
Multiple-example
Stereoscopic quality
Super-resolution
Symmetric mixed resolution stereoscopic video coding (SMRSVC)
3D TV

ABSTRACT

In this paper, we first propose a new symmetric mixed resolution stereoscopic video coding (SMRSVC) model which can provide clear bitrate-reduction and visual merits. Based on the newly proposed SMRSVC model, we then propose a quality-efficient multiple-example based super-resolution method. In the proposed super-resolution method, the four block examples selected from the forward and backward key-frames, the reference super-resolved frame, and the interview super-resolved frame are referred so as to effectively fuse the high frequency component of the super-resolved current block of the downsampled non-key-frame, and then an enhanced super-resolved non-key-frame is followed. Based on six test stereoscopic video sequences, the experimental results demonstrate that besides the bitrate-saving effect, the proposed super-resolution method for the proposed SMRSVC model also has better quality performance in terms of six well-known quality metrics when compared with several state-of-the-art methods for the previous asymmetric resolution stereoscopic video coding model and the SMRSVC model.

1. Introduction

With the advance of coding and network technology, the stereoscopic video system has received growing attention in the three-dimensional television (3-D TV) market. By synthesizing stereoscopic video sequences, the stereoscopic video system can provide consumers with more realistic 3D scenes. For a stereoscopic video sequence, since it consists of one left-view video sequence and one synchronized right-view video sequence, large amount of data makes it necessary to be compressed for storage saving and transmission over the internet. However, encoding the left-view and right-view video sequences independently results in double storage space and transmission bandwidth requirements. Constrained by the limited storage and bandwidth, designing different stereoscopic video coding models with the goals of low bitrate and good quality is therefore crucial [1,2].

The suppression theory of the binocular vision system [3] indicates that one view frame in a stereoscopic image pair, say the right-view frame, could be encoded at a lower bitrate than the other, i.e. the left-view frame, without causing obvious visual quality degradation. To realize the above bitrate reduction suggestion for stereoscopic video sequences, several researchers have proposed different variants of the asymmetric stereoscopic video coding (ASVC) model. Among these variants, some methods [4–6] have been presented to encode stereoscopic video sequences with different quality levels. In [4,5], Shao et al. encoded right-view frames with a larger quantization step, which is determined by a specified just noticeable distortion threshold, than left-view frames so as to reduce the bitrate requirement and maintain the perceptual quality. Fezza et al. [6] reduced the bitrate required in encoding each right-view frame by blurring the frame with the disk filter. Different from the above methods, the methods in [7–12] aimed to encode stereoscopic video sequences with different resolution levels, where each right-view frame is usually downsampled to be a quarter the size of the corresponding left-view frame. In [7], Fehn et al. applied the disparity compensated prediction [29] to encode the downsampled right-view frame by referring the corresponding left-view frame. Chen et al. [8,9] proposed the least mean square method to improve the disparity compensated prediction accuracy when encoding the downsampled right-view frames. In [10], Park

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: klchung01@gmail.com (K.-L. Chung), yonghuai@ms28.hinet.net (Y.-H. Huang).

[1] Supported by the Ministry of Science and Technology of ROC under the Contract MOST 102-2221-E-011-055-MY3.
[2] Supported by the Ministry of Science and Technology of ROC under the Contract MOST 104-2221-E228-006.

http://dx.doi.org/10.1016/j.jvcir.2016.05.007
1047-3203/

and Sim developed a variant in which each stereoscopic video sequence is composed of a left-view sequence and the horizontally downsampled right-view sequence. For 2D display service, the developed variant directly uses the left-view sequence as the output; for 3D display service, each left-view sequence is horizontally downsampled to be paired with the right-view sequence so as to generate a symmetric stereoscopic video sequence. Without using the disparity compensated prediction, Aflakt et al. [11] quantized the sampled luma values in each downsampled right-view frame and encode the left-view video sequence and the downsampled right-view video sequence independently in H.264/AVC. In [12], for each group of pictures, Yu et al. adaptively used the horizontal or vertical sampling strategies to downsample the right-view frames for improving the rate-distortion performance.

Besides the bitrate performance, the quality performance of the right-view video sequence upsampled at the decoder side is also an important issue for the ARSVC model. Traditionally, the signal image-based super-resolution methods, such as the 6-tap-filter (6TF) method [19], the Wiener Filter-based method [20], the soft-decision adaptive interpolation (SAI) method [21], the interpolation-dependent image downsampling with the edge-directed interpolation method [22], and learning-based super resolution methods [23,24], are utilized to estimate the missing pixels in the right-view frame. Besides the above single image-based super-resolution methods, the single-view video super-resolution methods, such as the hybrid super-resolution method [32], the example-based super-resolution method [15], and the texture synthesis-based method [33], can be applied to reconstruct the right-view frames. Considering the strong correlation between the left-view sequence and right-view sequence, Chung et al. [13] developed a Wiener filter-based super-resolution method with interview prediction and error compensation, which is called the WFIP-EC method throughout the paper, to improve the quality of the upsampled right-view frame. The WFIP-EC method improved the quality of the upsampled right-view frame mainly by compensating the prediction errors in the regions with heavy irregular textures. Empirical results have showed that under the ARSVC model, the WFIP-EC method for middle and high bitrate cases has better quality performance when compared with the signal image-based super-resolution methods, but is less competitive for low bitrate cases. Unfortunately, in the ARSVC model, because the quality of the decoded left-view video sequences is always better than the decoded right-view video sequences, a quality-imbalance problem may occur and degrades the perceived visual quality when viewing the synthesized 3D video sequences [14] over a period of time.

In this paper, our contributions are twofold: (1) we first propose a new model, called the symmetric mixed resolution stereoscopic video coding (SMRSVC) model, to alleviate the quality-imbalance problem occurred in the previous ARSVC model, and (2) we then propose a new quality-efficient multiple-example based super-resolution method for the proposed SMRSVC model. At the encoder side of the proposed model, the left-view and right-view video sequences of the input stereoscopic video sequence are independently encoded into two mixed resolution video sequences. At the decoder side, in order to achieve better quality of the super-resolved non-key-frame, the proposed super-resolution method utilizes the blocks from the bi-directional key-frames, the reference super-resolved frame, and the inter-view super-resolved frame as four examples to fuse the high-frequency component of the super-resolved non-key-frame. Based on six typical test stereoscopic video sequences, in terms of six well-known metrics, the experimental results confirm the quality-efficient, bitrate-saving, and perceived visual advantages of our proposed super-resolution method for the proposed SMRSVC model when compared with the SAI method [21] and Hung et al.s method [15] for the SMRSVC model; the SAI method and the WFIP-EC method [13] for the ARSVC model.

The rest of this paper is organized as follows. In Section 2, the proposed bitrate-saving SMRSVC model is presented. In Section 3, the proposed quality-efficient multiple-example based super-resolution method is presented for the SMRSVC model. The experimental results and performance comparison are provided in Section 4. Some concluding remarks are addressed in Section 5.

2. Proposed symmetric mixed resolution stereoscopic video coding (SMRSVC) model

This section consists of two subsections. In the first subsection, we briefly introduce the previous ARSVC model for coding the stereoscopic video sequence and then point out its own quality-imbalance problem, which will bring about a visual discomfort side-effect when viewing the synthesized 3D video sequences over a period of time. In the second subsection, to resolve the mentioned quality-imbalance problem, we propose a new stereoscopic video coding model called the symmetric mixed resolution stereoscopic video coding (SMRSVC) model.

2.1. Quality-imbalance problem occurred in the ARSVC model

In Fig. 1, the downsampled $\frac{W}{2} \times \frac{H}{2}$ right-view sequence and the original $W \times H$ left-view sequence constitute the ARSVC sequence. After receiving the encoded ARSVC sequence by the decoder, the super-resolution process is needed to upsample each decoded right-view frame to the $W \times H$ one. Finally, the reconstructed stereoscopic video sequence is conveyed into the 3D TV to provide viewers with 3D scenes.

The ARSVC model has clear bitrate-saving merit due to the encoding of the downsampled right-view sequence; however, the reconstructed stereoscopic video sequence suffers from a quality-imbalance problem. We now take examples to explain this problem in more detail. After running the state-of-the-art WFIP-EC method [13] on the Alt Moabit and Butterfly test stereoscopic video sequences, Fig. 2 shows the plots of peak signal-to-noise ratio (PSNR) vs. quantization parameters (QPs) for the reconstructed left-view and right-view sequences of the two test stereoscopic video sequences, respectively. In Fig. 2(a) and (b), the blue solid curve and the blue dashed curve denote, respectively, the PSNR performance of the reconstructed left-view and right-view sequences. It is clear that the PSNR difference between the reconstructed left-view and right-view sequences statistically ranges from 2 dB to 14 dB, indicating the quality-imbalance problem indeed occurred in the ARSVC model.

The above real quality-imbalance problem has been studied in quality assessment research [4,30,31,27,28]. In the subjective assessment study, Aflaki et al. [31] concluded that although the ARSVC model can provide satisfied 3D viewing experience, the quality-imbalance problem occurred in the ARSVC model may degrade the perceived visual quality when compared with the traditional symmetric stereoscopic coding model. In [30], Saygili et al. expressed the quality-imbalance problem by using the just-noticeable level of asymmetry in terms of PSNR threshold. Their subjective assessment showed that if the PSNR of the reconstructed right-view sequence is less than the specified PSNR threshold, the asymmetrically coded sequences could result in worse perceived visual quality than the symmetrically coded sequences. In [4], Shao et al. indicated that when the PSNR difference between the reconstructed left-view and right-view sequences is larger than 2 dBs, the viewers could feel the quality degradation phenomenon. As shown in Fig. 2, the PSNR difference for the ARSVC model may range from 2 dB to 14 dB, and it assures the quality degradation possibility. The two newest objective assessment metrics proposed by Silva et al. [27] and Lin and Wu

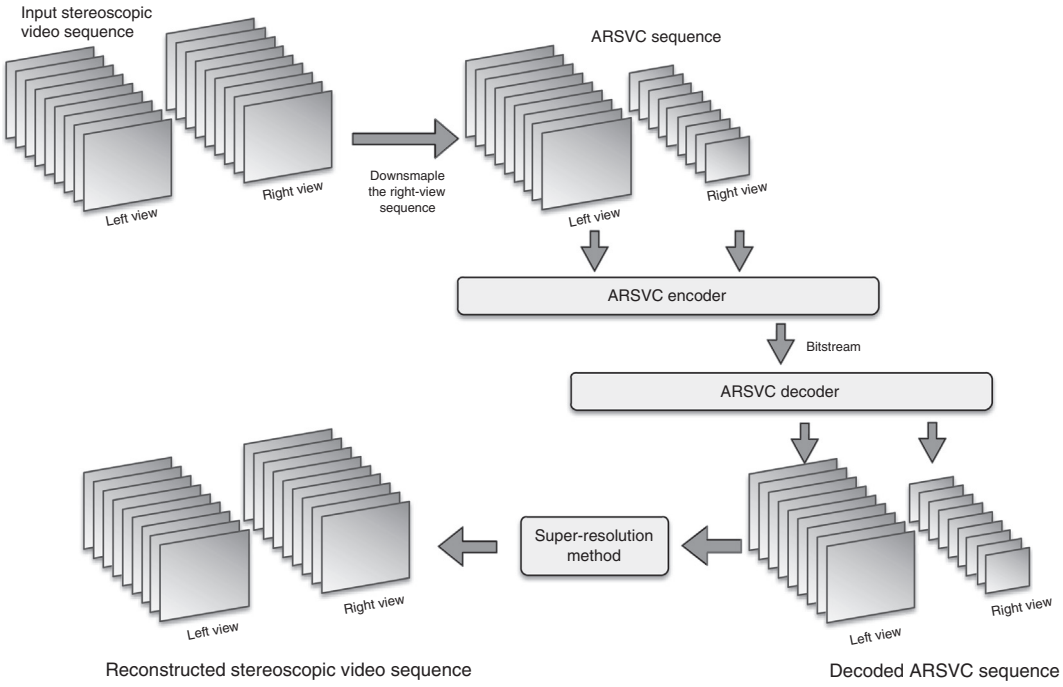

Fig. 1. The ARSVC model.

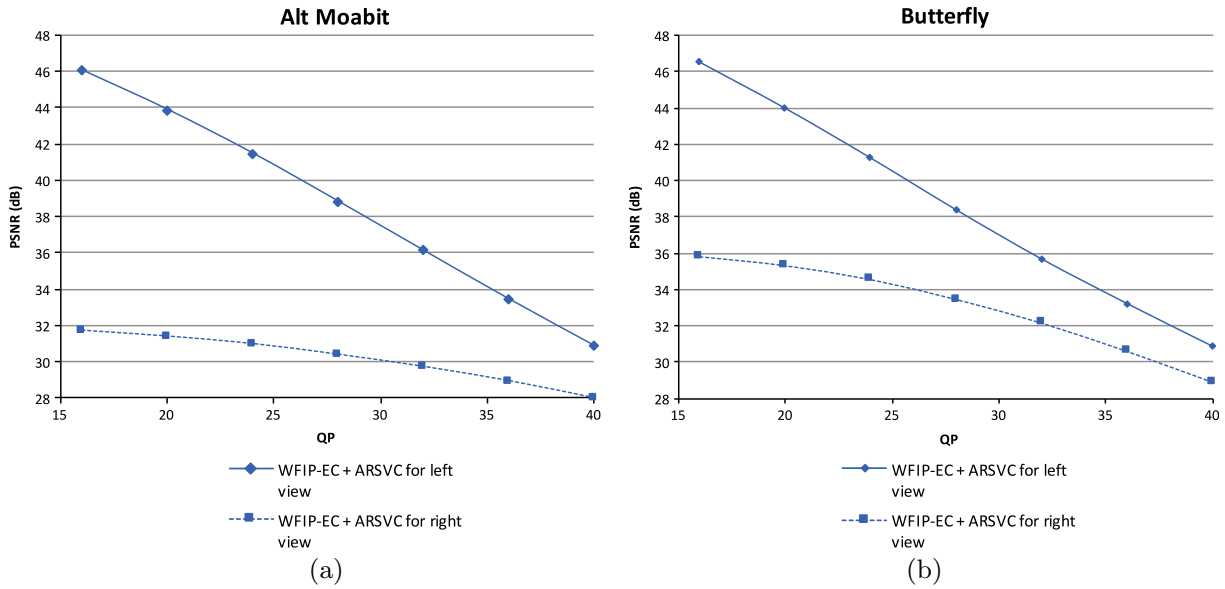

Fig. 2. The PSNR curves using the WFIP-EC super-resolution method for the ARSVC model. (a) Alt Moabit. (b) Butterfly stereoscopic video sequences.

[28] showed that the stereoscopic videos with imbalanced quality could lead to a significant difference in perceived visual quality when compared to that with balanced quality. The above discussions concluded that although the ARSVC model provides bitrate-saving merit, the existing quality-imbalance problem may degrade the perceived visual quality when compared with the symmetric stereoscopic video coding model.

2.2. Proposed SMRSVC model for alleviating the quality-imbalance problem in the ARSVC model

To resolve the quality-imbalance problem occurred in the previous ARSVC model and achieve more bitrate-saving merit, we propose a new model called the SMRSVC model, as shown in Fig. 3. When compared with the ARSVC model, the novelties of the proposed SMRSVC model associated with the proposed quality-efficient super-resolution method are listed below:

(1) Instead of the downsampled right-view sequence in which each frame is of size $\frac{W}{2} \times \frac{H}{2}$ in the ARSVC model, the right-view sequence in the proposed SMRSVC model alternately contains one key-frame and several non-key-frames in which each key-frame is the same size as the original image frame and each non-key-frame is downsampled to be of size $\frac{W}{2} \times \frac{H}{2}$.
(2) Instead of the left-view sequence in which each frame is of size $W \times H$ in the ARSVC model, in terms of the frame size

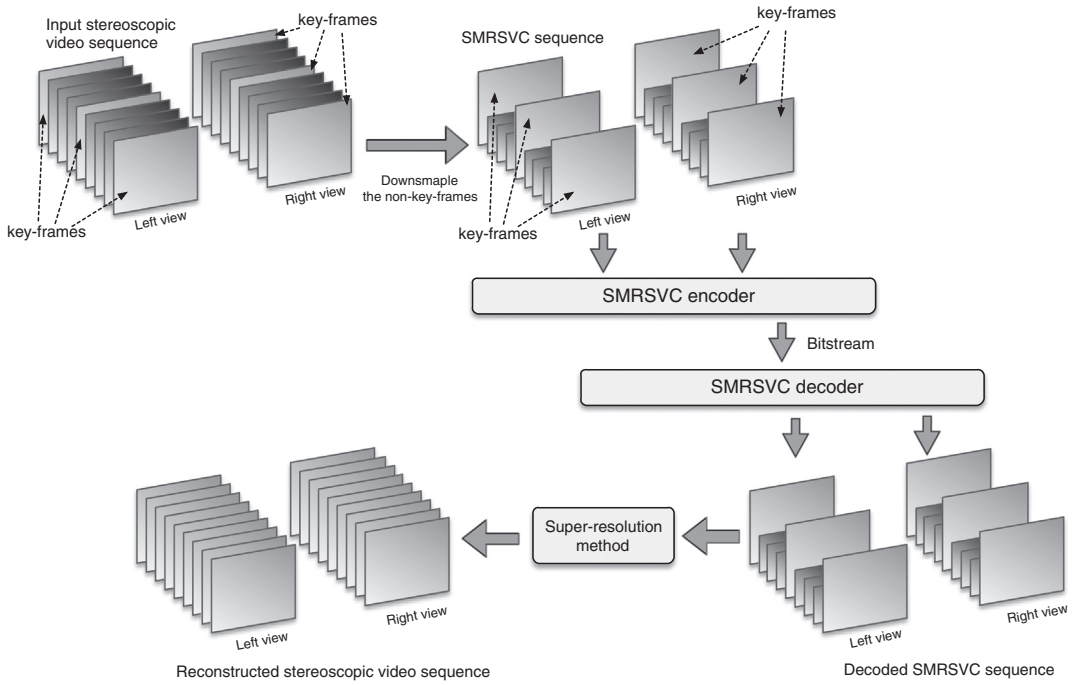

Fig. 3. The proposed bitrate-saving SMRSVC model.

configuration, the left-view sequence in the proposed SMRSVC model is the same as its right-view sequence.

(3) Instead of adopting the previous WFIP-EC method, which is originally designed to super-resolve one downsampled right-view sequence in the ARSVC model, we propose a new quality-efficient super-resolution method, which will be presented in Section 3, for super-resolving the downsampled right/left-view sequences in the SMRSVC model.

Suppose that in the input stereoscopic video sequence, both the left-view sequence and the synchronized right-view sequence contain N image frames. Let L_i and $R_i, 0 \leqslant i \leqslant N-1$, denote the i-th left-view frame and right-view frame, respectively, each with size $W \times H$. In the proposed SMRSVC model, the image frames of the left-view/right-view sequence are alternately transformed into the setting of a key-frame and several non-key-frames, in which each key-frame being with the original size regularly appears every P frames and each non-key-frame is downsampled to be of size $\frac{W}{2} \times \frac{H}{2}$. Let L_k and $L_{k'}, k \in \{0, P, 2P, \ldots, (M-1)P\}$ where $M = \frac{N-1}{P}$ and $k' = k + P$, denote two consecutive key-frames. Between two key-frames, we have $P-1$ downsampled non-key-frames and each one is denoted by $L_n, k < n < k'$. The key-frames and non-key-frames in the right-view sequence are denoted by the similar way, but we only replace the notation 'L' with 'R'.

Each $L_n, k < n < k'$, is downsampled to produce a quarter sized frame ℓ_n. Similarly, for the right-view sequence, each non-key-frame R_n is also downsampled to produce r_n. In fact, the encoder of the existing multi-view video coding standard, such as H.264/MVC [16] or MV-HEVC [17], can be slightly modified as the encoder for coding the SMRSVC sequence. Fig. 4 illustrates an example of the prediction structure used in the SMRSVC encoder/decoder. Here, the key-frame period is set to $P = 16$. Among these 34 frames in Fig. 4, the two key-frames in the left-view sequence are first encoded as the I-frames, and then the 15 downsampled non-key-frames in the left-view sequence are encoded as the B-frames. The right-view sequence is encoded in a similar way, but its two key-frames are encoded as the P-frames. Besides, each right-view frame can additionally refer to the synchronized left-view frame so as to further improve the bitrate performance. In order to remove more redundancy between the two views in a stereoscopic video sequence, the disparity compensated prediction [29] is considered in the encoding framework to predict the current downsampled right-view frame by referring to either the downsampled left-view frame or the previous downsampled right-view frame.

The experimental results show that the PSNR difference between the reconstructed left-view and right-view frames in the proposed SMRSVC model is rather small, less than 1 dB, and the low PSNR difference property is not surprising due to the symmetric structure of the proposed SMRSVC model. It comes to a conclusion that the quality-imbalance problem existed in the previous ARSVC model can be resolved by the proposed SMRSVC model. In addition, downsampling the non-key-frames in the left-view/right-view sequences leads to a better bitrate-saving effect. The experimental results in Section 4 will confirm the low bitrate, high perceived visual quality, and balanced visual experience merits of the proposed SMRSVC model. At the decoder side, after decoding the SMRSVC sequence, the proposed quality-efficient super-resolution method follows to effectively reconstruct the downsampled non-key-frames of the decoded SMRSVC sequence, which will be presented in the next section.

3. Proposed quality-efficient multiple-example based super-resolution method for the SMRSVC model

In our proposed super-resolution method, each non-key-frame is first upsampled to the one with the low-frequency component and then its high-frequency component is compensated by fusing the two referred key-frames, the super-resolved reference frame, and the super-resolved inter-view frame. In the following three subsections, we firstly present the basic idea behind the proposed super-resolution method; secondly, we present a two-stage method to realize the proposed method; finally, the difference among the previous methods and the proposed method two-stage is provided.

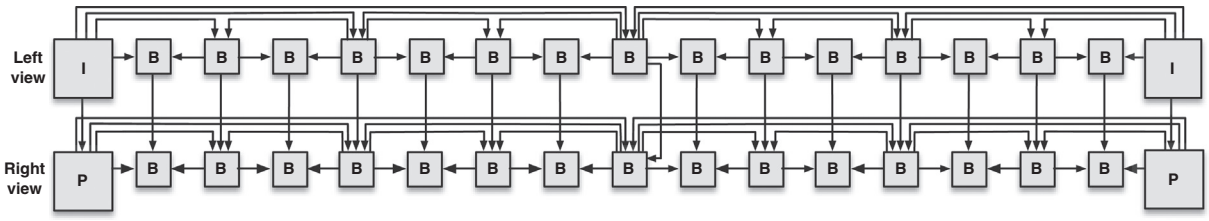

Fig. 4. The prediction structure for the SMRSVC.

3.1. Basic idea behind the proposed super-resolution method

After decomposing the original non-key-frame $L_n, k \leqslant n \leqslant k'$, into the high-frequency and low-frequency components, L_n can be represented by

$$L_n = L_n^{LF} + L_n^{HF}, \tag{1}$$

where L_n^{LF} and L_n^{HF} denote the low-frequency component and high-frequency component of L_n, respectively. The low-frequency component provides the coarse information of L_n and can be obtained by

$$L_n^{LF} = f_u(f_d(L_n)) = f_u(\ell_n), \tag{2}$$

where $f_u(x)$ and $f_d(x)$ denote the upsampling operator and the downsampling operator, respectively, on the image x using the Lanczos filter [18]; ℓ_n denotes the downsampled frame of L_n. The high-frequency component contains the textural information of L_n and can be obtained by

$$L_n^{HF} = L_n - L_n^{LF}. \tag{3}$$

Taking a car non-key-sub-frame as illustration, Fig. 5(a) shows the original image frame, while Fig. 5(b) and (c) display, after the decomposition, the corresponding low-frequency component and high-frequency component.

Let $\widehat{L}_n$ denote the super-resolved frame of the downsampled non-key-frame ℓ_n. Because L_n^{HF} cannot be directly calculated by Eq. (1) at the decoder side, we calculate the estimation of L_n^{HF}, denoted by $\widehat{L}_n^{HF}$, instead. Accordingly, the super-resolved non-key-frame can be determined by

$$\widehat{L}_n = L_n^{LF} + \widehat{L}_n^{HF} = f_u(\ell_n) + \widehat{L}_n^{HF}. \tag{4}$$

Since the non-key-frames in the right-view sequence are processed by the same way, we omit the related details.

To estimate the high-frequency component of the current non-key-frame, i.e. $\widehat{L}_n^{HF}$ in Eq. (4), the proposed method selects a set of Q example-frames, say $\mathcal{L}_n = \{\mathcal{L}_{n,0}, \mathcal{L}_{n,1}, \ldots,$ and $\mathcal{L}_{n,Q-1}\}$, in which each example-frame can be selected from either the left-view or the right-view sequence. In our proposed super-resolution method, empirically we set $Q = 4$ and the four selected example-frames are the forward and backward key-frames L_k and $L_{k'}$, the super-resolved reference non-key-frame which is previous to the current non-key-frame, and the inter-view frame at the same time slot in the right-view sequence. Based on the four selected example-frames, in the next section, a two-stage block-based method is presented to fuse the high-frequency component of the current non-key-frame for improving the quality of the super-resolved frame.

3.2. Two-stage multiple-example and block based super-resolution for SMRSVC

The proposed super-resolution method for the SMRSVC model consists of two stages, the intra-view stage and the inter-view stage. In the intra-view stage, each current downsampled non-key-frame in the left-view and right-view video sequences is super-resolved independently by referring to the triple-example in its own video sequence. Based on the super-resolved non-key-frames obtained from the intra-view stage, in the inter-view stage, we further enhance the quality of each current super-resolved frame by referring to the quadruple-example: the three same examples used in the intra-view stage and the inter-view example frame which is in the right-view sequence. The detail of the proposed two-stage super-resolution method is described as follows.

The flowchart of the intra-view stage for the left-view sequence is depicted in Fig. 6. The intra-view stage for the right-view sequence can be processed in the same way. After decoding the SMRSVC sequence, between the two adjacent key-frames, L_k and $L_{k'}$, these $P-1$ decoded downsampled non-key-frames will be super-resolved in the order: ℓ_{k+1}, ℓ_{k+2}, …, and $\ell_{k'-1}$. In order to obtain better super-resolved frame $\widehat{L}_n, k < n < k'$, as mentioned before, the multiple-example set $\mathcal{L}_n$ is used to estimate the high-frequency component $\widehat{L}_n^{HF}$ and for $k+1 < n < k'$, it yields

$$\mathcal{L}_n = \{\mathcal{L}_{n,0}, \mathcal{L}_{n,1}, \mathcal{L}_{n,2}\} = \{L_k, L_{k'}, \widehat{L}_{n-1}\}; \tag{5}$$

for $n = k+1$, it yields

$$\mathcal{L}_n = \{\mathcal{L}_{n,0}, \mathcal{L}_{n,1}\} = \{L_k, L_{k'}\}. \tag{6}$$

In order to estimate more accurate $\widehat{L}_n^{HF}$ from $\mathcal{L}_n$, a block based approach is presented. Let $B(\widehat{L}_n)$ denote the current 8×8 block in $\widehat{L}_n$. In the intra-view stage, as shown in Fig. 6, given $B(\widehat{L}_n)$, the block matching process is applied to find its best matched block $B^*(\mathcal{L}_{n,j})$ from each example-frame and the three (or two for $n = k+1$) found best matched blocks will be used to fuse the high-frequency component of $B(\widehat{L}_n)$, denoted by $B(\widehat{L}_n^{HF})$. Since we only have the low-frequency component of the current non-key-frame, the block matching process is performed on $\widehat{L}_n^{LF}$ and $\mathcal{L}_{n,j}^{LF}$. To minimize the blocking effect, the 2-pixel-wide overlapping block motion compensation technique recommended in [15] is adopted to implement the above block matching process.

According to the best matched block $B^*(\mathcal{L}_{n,j})$, the high-frequency block $B^*(\mathcal{L}_{n,j}^{HF})$ is obtained by

$$B^*(\mathcal{L}_{n,j}^{HF}) = B^*(\mathcal{L}_{n,j}) - B^*(\mathcal{L}_{n,j}^{LF}). \tag{7}$$

After determining the three high-frequency blocks, $B^*(\mathcal{L}_{n,0}^{HF}), B^*(\mathcal{L}_{n,1}^{HF})$, and $B^*(\mathcal{L}_{n,2}^{HF})$, based on the three example-frames, the high-frequency block $B(\widehat{L}_n^{HF})$ is fused by

$$B(\widehat{L}_n^{HF}) = \sum_{j=0}^{Q-1} \alpha_j B^*(\mathcal{L}_{n,j}^{HF}), \tag{8}$$

where $Q = 3$ and the weight α_j is given by

$$\alpha_j = \frac{1}{\epsilon_j} \left(\sum_{m=0}^{Q-1} \frac{1}{\epsilon_m} \right)^{-1} \tag{9}$$

Fig. 5. Decomposition of a car sub-non-key-frame into a low-frequency component and a high-frequency component. (a) The original image frame. (b) Low-frequency component of (a). (c) High-frequency component of (a).

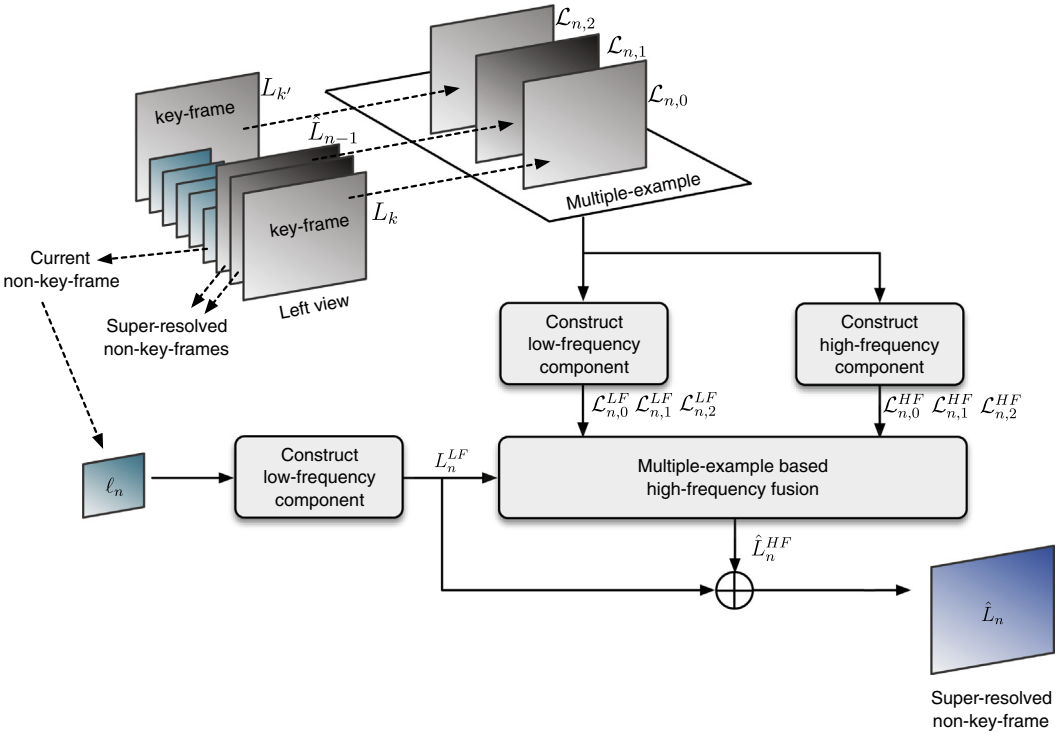

Fig. 6. The flowchart of the intra-view super-resolution stage for the left-view sequence.

and

$$\epsilon_j = \left\| B(\widehat{L}_n^{LF}) - B^*(\mathcal{L}_{n,j}^{LF}) \right\|^2. \tag{10}$$

Note that for $n = k + 1$, only two high-frequency blocks, $B^*(\mathcal{L}_{n,0}^{LF})$ and $B^*(\mathcal{L}_{n,1}^{LF})$, are used to fuse $B(\widehat{L}_n^{HF})$. After all the 8×8 high-frequency blocks of the current non-key-frame have been fused by Eqs. (8)–(10), the high-frequency component $\widehat{L}_n^{HF}$ can be constructed and then Eq. (8) is applied to obtain the super-resolved frame $\widehat{L}_n$. After performing the intra-view stage for all the non-key-frames in the left-view sequence, the similar intra-view super-resolution stage is applied to the right-view sequence.

The inter-view stage, as shown in Fig. 7, is used to refine the quality of the super-resolved non-key-frames obtained in the intra-view stage. In this stage, the used quadruple-example $\mathcal{L}_n$ is expressed as

$$\mathcal{L}_n = \{\mathcal{L}_{n,0}, \mathcal{L}_{n,1}, \mathcal{L}_{n,2}, \mathcal{L}_{n,3}\} = \{L_k, L_{k'}, \widehat{L}_{n-1}, \widehat{R}_n\}, \tag{11}$$

where $\widehat{R}_n$ denotes the super-resolved frame in the right-view sequence. Based on $\mathcal{L}_n$, the similar block matching process as in the intra-view stage is applied to obtain the four 8×8 high-frequency blocks $B^*(\mathcal{L}_{n,0}^{LF}), B^*(\mathcal{L}_{n,1}^{LF}), B^*(\mathcal{L}_{n,2}^{LF})$, and $B^*(\mathcal{L}_{n,3}^{LF})$. Next, the high-frequency block $B(\widehat{L}_n^{HF})$ can be refined in a similar way as in Eqs. (8)–(10) for $Q = 4$. After all the high-frequency blocks have been fused, the whole high-frequency component $\widehat{L}_n^{HF}$ follows. We then refine the super-resolved non-key-frame $\widehat{L}_n$ by Eq. (4). Using the above refinement process, the quality of the non-key-frames in the right-view sequence can be enhanced too. Finally, the resultant high quality stereoscopic video sequence is fed into the 3D TV to provide viewers with better perceived visual 3D quality.

3.3. Comparison with the previous two methods

In this subsection, we want to compare the difference among the proposed super-resolution method and the two state-of-the-art super-resolution methods, the WFIP-EC method [13], and Hung et al.'s method [15]. In the WFIP-EC method, each downsampled right-view frame is first upsampled by using the Wiener filter-based method. Next, each missing pixel in the upsampled right-view frame is interpolated by fusing the two prediction results obtained from the following two prediction processes. The first prediction process collects the pixels neighboring to the current missing pixel into a set, and then based on the collected pixels, a least square estimation is used to obtain the first prediction result. The second prediction process applies the disparity compensated

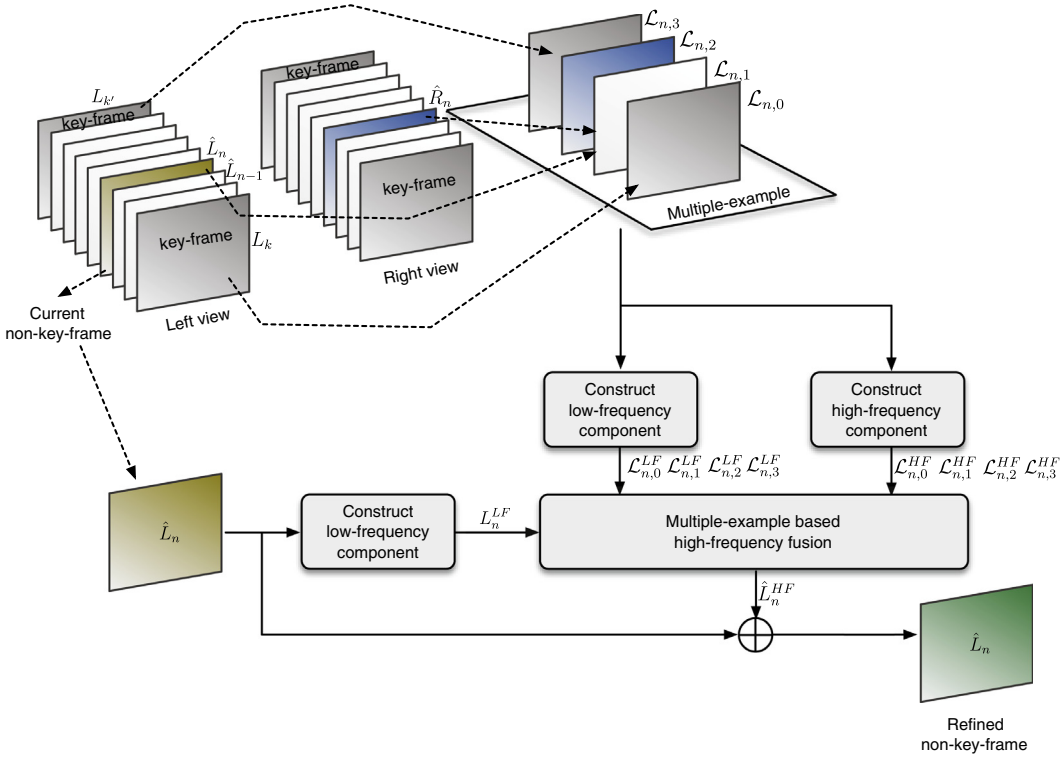

Fig. 7. The flowchart of the inter-view super-resolution stage for refining the left-view sequence.

Table 1
Comparison of the proposed and the two previous super-resolution methods.

Method	Estimation kernel	Inter-view prediction	Intra-view prediction	Video coding model
WFIP-EC	Wiener filter-based approach	Disparity compensated prediction	Spatial prediction	ARSVC model
Hung et al.	Estimation for low- and high-frequency components	N/A	Temporal prediction using key-frames	Mixed resolution 2D video coding model
Proposed	Estimation for low- and high-frequency components	Disparity compensated prediction	Temporal prediction using key- and non-key-frames	SMRSVC model

prediction to find a best matched block from the left-view frame, and then based on the found block, a least square estimation is used to obtain the second prediction result. Hung et al.'s super-resolution method is designed for the mixed resolution 2D video coding model, in which the mixed resolution video sequence is equal to that of the single-view case in the proposed SMRSVC model. In Hung et al.'s method, each downsampled non-key-frame is first upsampled by Lanczos filter to obtain the low-frequency component. Then, the high-frequency component is estimated by only using the nearest two key-frames.

After sketching all the differences among the proposed super-resolution method and the two previous super-resolution methods, their comparison is listed in Table 1. The WFIP-EC method uses both intra-view and inter-view predictions to assist the super-resolution process. Hung et al.'s method only uses the intra-view prediction to construct the low-frequency and high-frequency components of the super-resolved frames. Note that the intra-view predictions used in the two methods are different, the WFIP-EC method utilizes the spatial prediction and the other utilizes the temporal prediction. Due to the use of depth information, the view projection process can obtain better prediction result than the disparity compensated prediction used in WFIP-EC. The proposed super-resolution method uses both intra-view and inter-view predictions. The proposed intra-view prediction considers the temporal correlation of video to assist the super-resolution process. Different from the Hung et al.'s method, the proposed intra-view prediction uses not only the nearest two key-frames but also the super-resolved non-key-frame nearest to the current non-key-frame for constructing the high-frequency component. The strong temporal correlation between the current non-key-frame and the nearest one can improve the accuracy of the intra-view prediction. The proposed inter-view prediction is based on the disparity compensated prediction to construct the high-frequency component. The view projection process is not considered in the proposed inter-view prediction because including the depth maps in the proposed SMRSVC model will degrade the bitrate-saving merit.

4. Experimental results

In this section, six test stereoscopic video sequences downloaded from the website [25] are used to confirm the quality-efficient, bitrate-saving, and visual perception merits of our proposed super-resolution method for the proposed SMRSVC model when compared with several state-of-the-art super-resolution methods for the ARSVC and SMRSVC models. Here, we adopt two variants of the ARSVC model, one proposed by Fehn et al. [7] and the other proposed by Aflakt et al. [11], for performance comparison

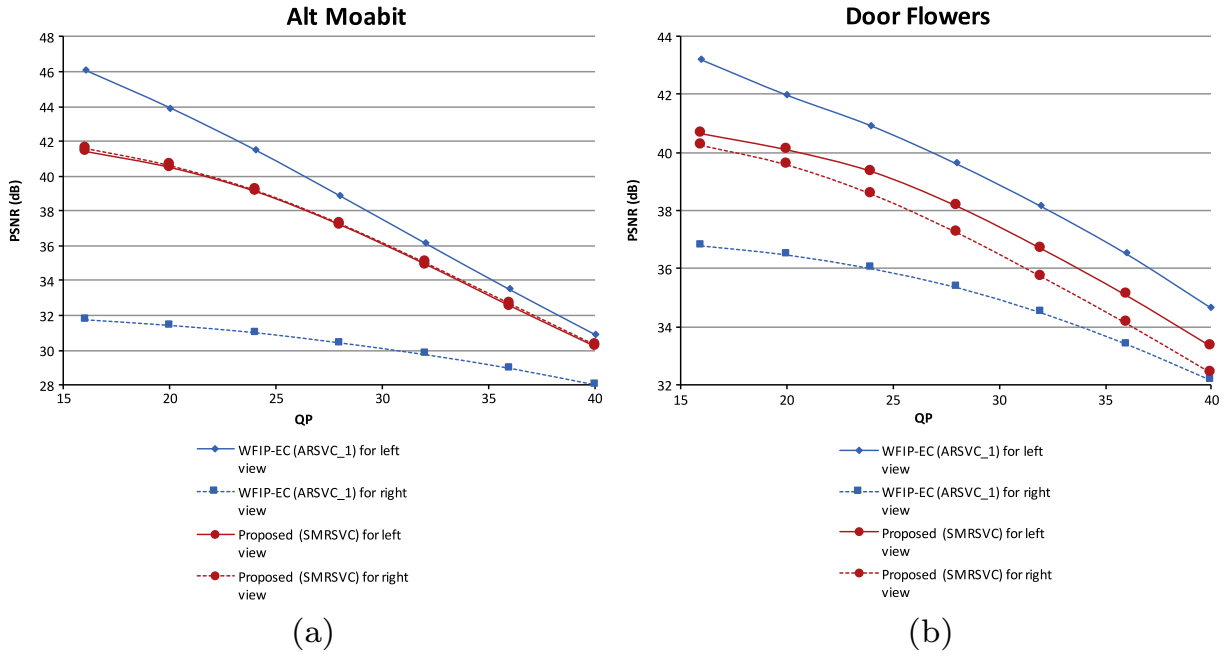

Fig. 8. The PSNR curves using the WFIP-EC super-resolution method for the ARSVC_1 model and the proposed super-resolution method for the proposed SMRSVC model. (a) Alt Moabit. (b) Door Flowers stereoscopic video sequences.

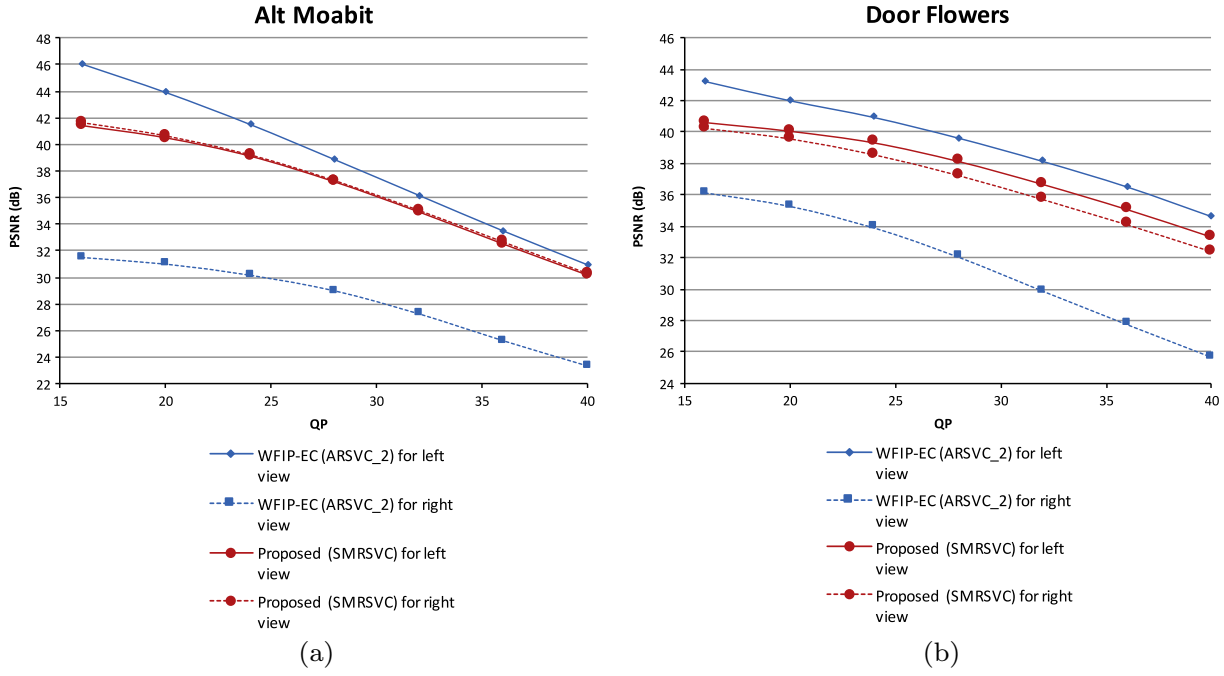

Fig. 9. The PSNR curves using the WFIP-EC super-resolution method for the ARSVC_2 model and the proposed super-resolution method for the proposed SMRSVC model. (a) Alt Moabit. (b) Door Flowers stereoscopic video sequences.

and they are denoted by ARSVC_1 and ARSVC_2, respectively. The six test video sequences are the Alt Moabit sequence, each frame with size 512 × 384, the Butterfly sequence, each frame with size 432 × 240, the Book Arrival and Door Flowers sequences, each frame with size 1024 × 768, and the Poznan Street and Poznan Hall2 sequences, each frame with size 1920 × 1088. The performance comparison among the concerned methods is evaluated by the visual comfort, the compression performance, and the quality performance. According to the quality-imbalance discussion in SubSection 2.1, here the perceived visual quality is measured by PSNR difference between the left-view frame and the synchronized right-view frame. The compression performance is measured by the two metrics, bitrate and Bjøntegaard delta bitrate (BD-BR) [26]. The quality performance is measured by PSNR, Bjøntegaard delta peak signal-to-noise ratio (BD-PSNR), stereoscopic structural distortion (StSD) metric [27], and Frequency-Integrated metrics (FI-metrics).

The concerned super-resolution methods for the ARSVC_1, ARSVC_2, and SMRSVC models were implemented on an IBM

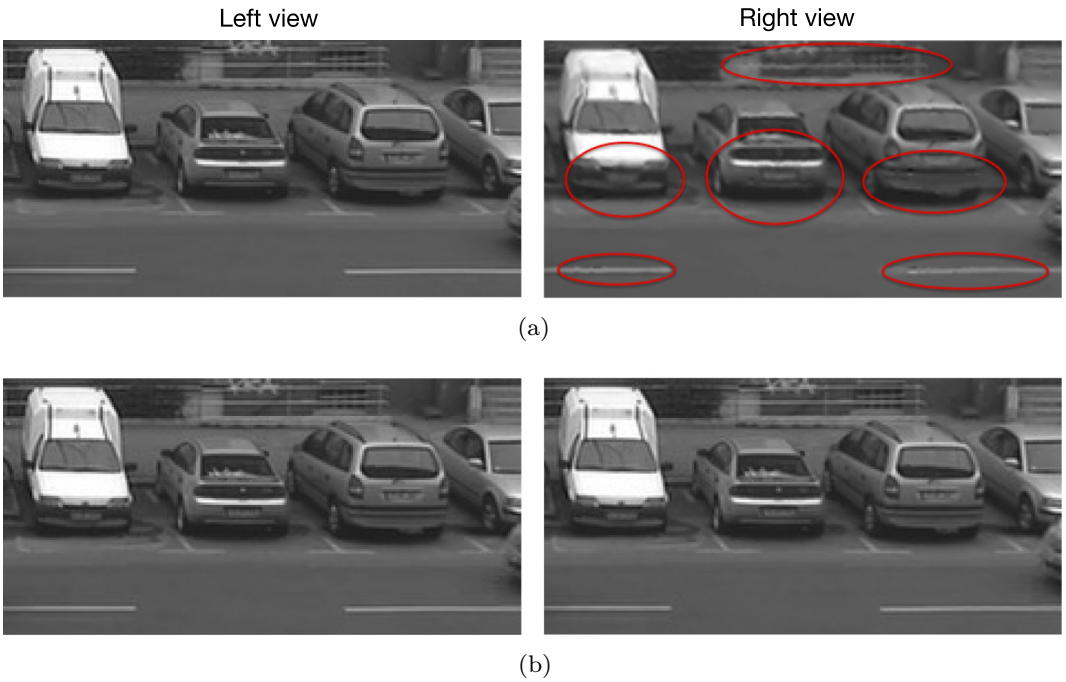

Fig. 10. The visual quality comparison for the Alt Moabit stereoscopic video sequence. (a) Amplified sub-images cut off from the second super-resolved ARSVC_1 left/right-view frames by the WFIP-EC method. (b) Amplified sub-images cut off from the second super-resolved SMRSVC left/right-view frames by our proposed method.

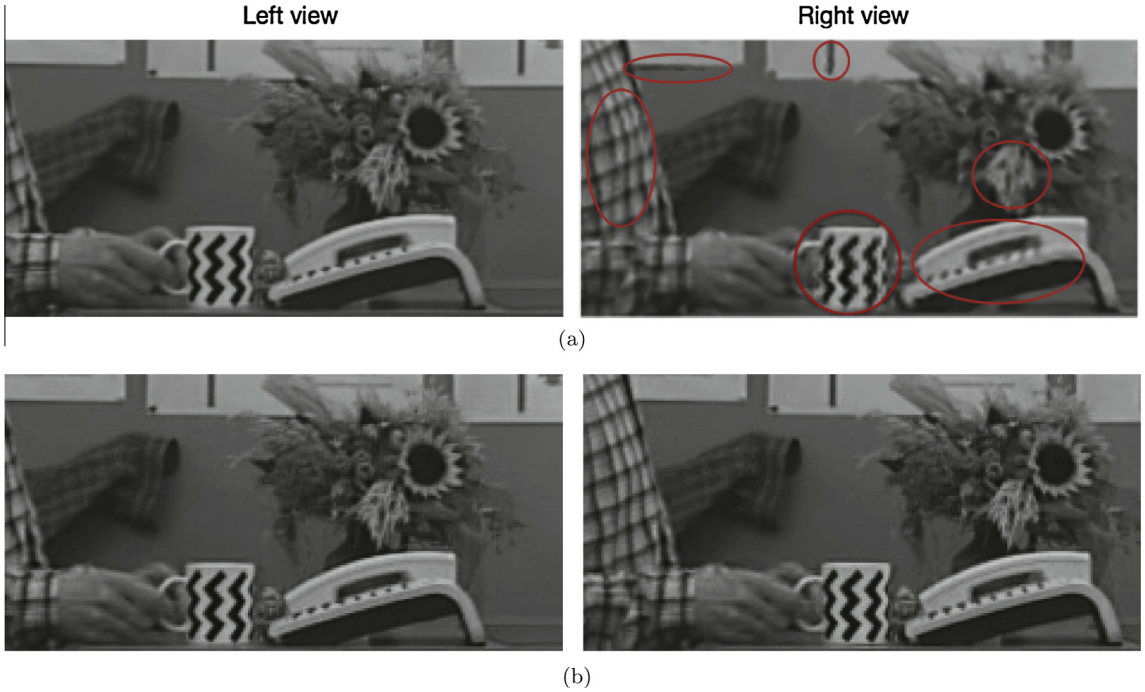

Fig. 11. The visual quality comparison for the Book Arrival stereoscopic video sequence. (a) Amplified sub-images cut off from the second super-resolved ARSVC_1 left/right-view frames by the WFIP-EC method. (b) Amplified sub-images cut off from the second super-resolved SMRSVC left/right-view frames by our proposed method.

compatible computer with an Intel Core i7-3770 CPU 3.40 GHz and 4 GB RAM. The operating system was Microsoft Windows 7. The program development environment was Visual C++ 2010 and the implementation platform was MV-HEVC [17]. In our implementation, the key-frame period was set to 40 empirically. In what follows, three kinds of experiments are carried out to demonstrate the merits of our proposed super-resolution method for the proposed SMRSVC model.

4.1. Quality-balance comparison

The first experiment is designed to show that in terms of the PSNR difference between the reconstructed left view and right view sequences, the proposed SMRSVC model has better quality-balance merit than the previous ARSVC models. Based on the two test stereoscopic video sequences, Alt Moabit and Door Flowers, the proposed SMRSVC model together with the proposed

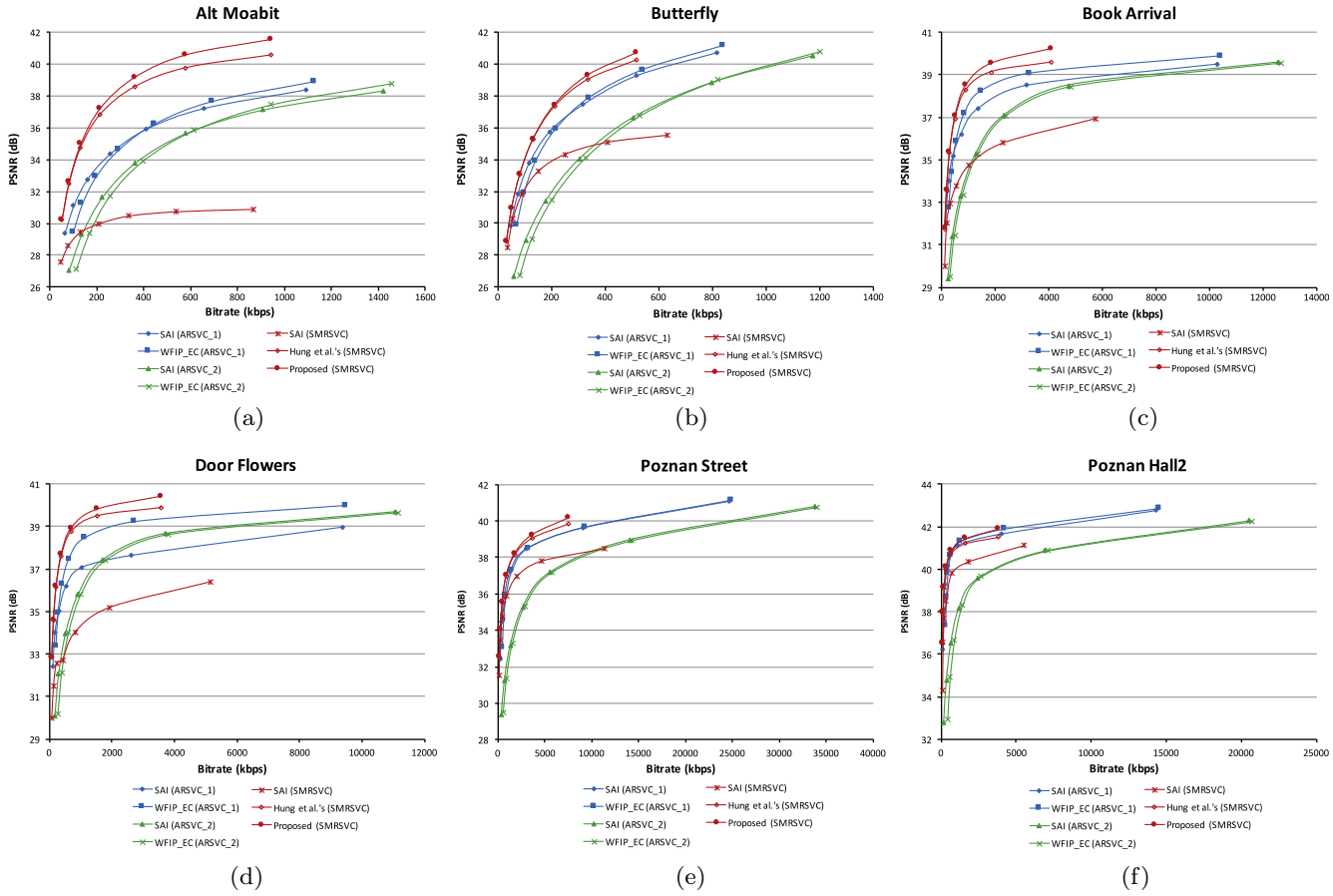

Fig. 12. In terms of PSNR, the RD curves of all concerned methods for (a) the Alt Moabit sequence, (b) the Butterfly sequence, (c) the Book Arrival sequence, (d) the Door Flowers sequence, (e) the Poznan Street sequence, and (f) the Poznan Hall2 sequence.

Table 2
BD-PSNR (in dB) of the proposed method over the Hung et al.'s method for the SMRSVC model and the WFIP-EC method for the ARSVC_1 and ARSVC_2 models.

Video sequence	BD-PSNR of ours over WFIP-EC for ARSVC_1	BD-PSNR of ours over WFIP-EC for ARSVC_2	BD-PSNR of ours over Hung et al.'s for SMRSVC
Alt Moabit	3.66	5.78	0.43
Butterfly	1.65	5.54	0.07
Book Arrival	1.34	4.23	0.23
Door Flowers	1.48	3.70	0.17
Poznan Street	0.85	0.06	0.07
Poznan Hall2	0.72	4.28	0.16
Average	1.62	3.93	0.19

Table 3
BD-BR (in %) of the proposed method over the Hung et al.'s method for the SMRSVC model and the WFIP-EC method for the ARSVC_1 and ARSVC_2 models.

Video sequence	BD-BR of ours over WFIP-EC for ARSVC_1	BD-BR of ours over WFIP-EC for ARSVC_2	BD-BR of ours over Hung et al.'s for SMRSVC
Alt Moabit	−59.40	−74.30	−9.90
Butterfly	−31.34	−67.05	−1.58
Book Arrival	−47.20	−79.94	−10.01
Door Flowers	−67.64	−87.07	−9.35
Poznan Street	−37.08	−82.83	−3.88
Poznan Hall2	−44.26	−89.47	−15.26
Average	−47.82	−80.11	−8.33

super-resolution method is carried out to compare with the ARSVC_1 and ARSVC_2 models, respectively, together with the WFIP-EC method [13]. The plots of PSNR vs. QPs of the proposed SMRSVC model and the ARSVC_1 model for the Alt Moabit and Door Flowers sequences are shown in Fig. 8, while the analog of the proposed SMRSVC model and the ARSVC_2 model is given in Fig. 9. In Figs. 8 and 9, the red curves are the PSNR performance of the reconstructed left-view and right-view sequences of the proposed SMRSVC model, while the blue curves are that of either the ARSVC_1 model or the ARSVC_2 model. For the ARSVC_1 model and the ARSVC_2 model, it is clear that their PSNR differences between the reconstructed left-view and the right-view sequences in Figs. 8 and 9 still range from at least 2 dB to at most 14 dB and such large PSNR differences confirm again the existing quality-imbalance problem in the ARSVC models. In contrast, all the PSNR differences between the reconstructed left-view and the right-view sequences of the proposed SMRSVC model are less than 1 dB, indicating that the proposed SMRSVC model does have the quality-balance superiority when compared with the two concerned ARSVC models. Similar conclusions also appear in the other four test stereoscopic sequences.

Besides the quality-balance merit, we also provide the visual quality comparison to demonstrate the visual perception merit of the proposed SMRSVC model. For the case of $QP = 16$, Fig. 10(a)

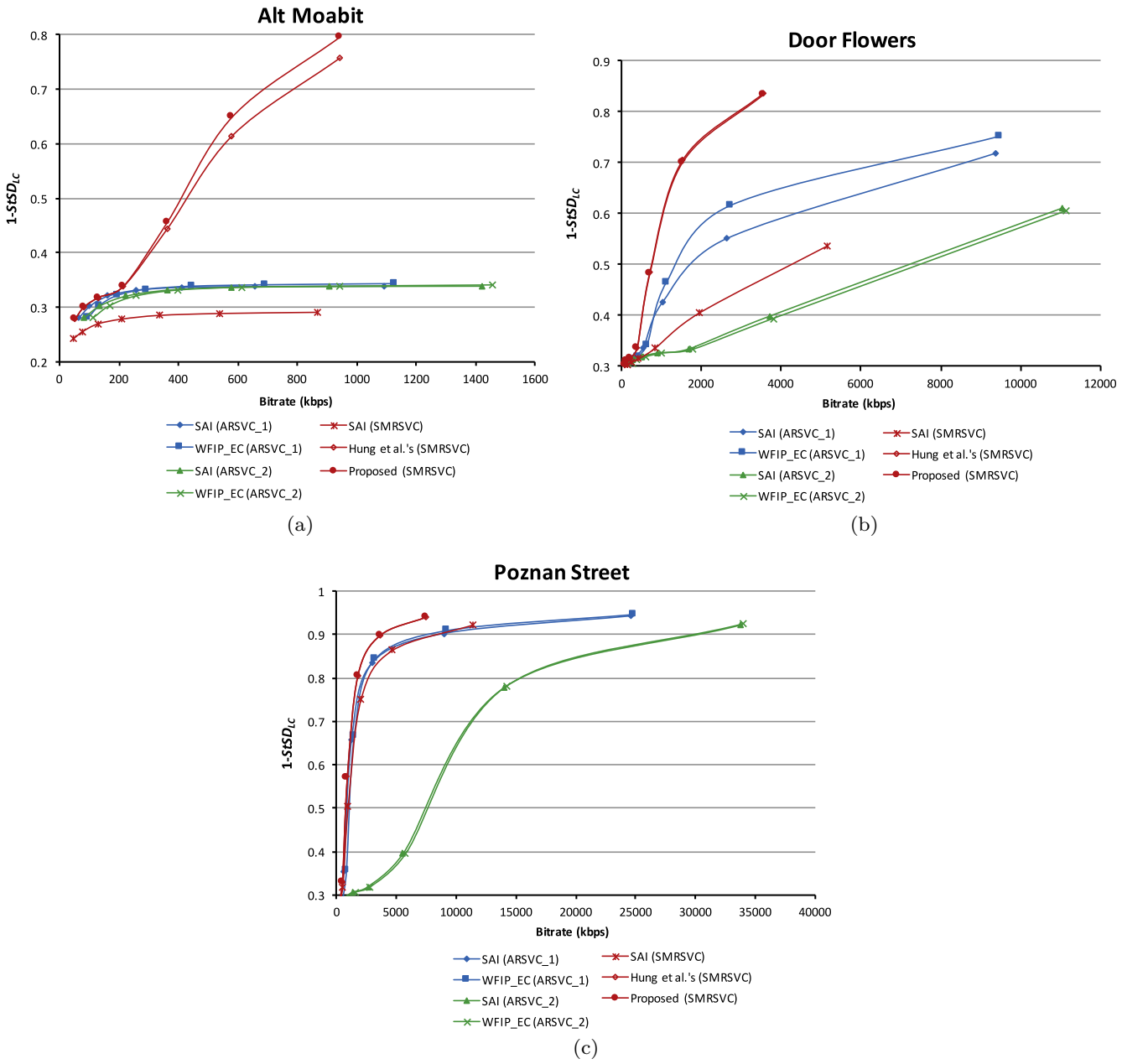

Fig. 13. In terms of the metric 1−StSDLC, the RD curves of all concerned methods for (a) the Alt Moabit sequence, (b) the Door Flowers sequence, and (c) the Poznan Street sequence.

shows the left and right amplified sub-images which are cut off from, respectively, the second left-view and right-view image frames in the reconstructed super-resolved Alt Moabit sequence generated by the WFIP-EC method for the ARSVC_1 model, while Fig. 10(b) shows the analog generated by the proposed super-resolution method for the SMRSVC model. It is clear that in the right amplified sub-image of Fig. 10(a), the boundaries of the objects highlighted by the red ellipses demonstrate obvious distortion when compared with the left amplified sub-image of Fig. 10 (a). In contrast, the two amplified sub-images corresponding to the proposed super-resolution method for the SMRSVC model in Fig. 10 demonstrate good and similar visual quality and no obvious visual degradation appears in the right amplified sub-image. For the stereoscopic Book Arrival sequence, Fig. 11 shows the analog and then similar conclusions are observed again, indicating the visual merit of our proposed SMRSVC model when compared with the ARSVC_1 model. The visual merit of the proposed SMRSVC model is kept when compared with the ARSVC_2 model.

4.2. Bitrate and distortion tradeoff comparison

In terms of rate-distortion (RD) curve, BD-PSNR, and BD-BR, the second experiment is conducted to show the bitrate and distortion tradeoff superiority of the proposed super-resolution method for the SMRSVC model when compared with the soft-decision adaptive interpolation (SAI) method [21] and Hung et al.'s method [15] for the SMRSVC model; the SAI method and the WFIP-EC method for the ARSVC_1 and ARSVC_2 models. In general, lower distortion of the reconstructed video sequence often accompanies higher bitrate requirement. For plotting the RD curves, seven different QPs, 16, 20, 24, 28, 32, 36, and 40, are used to obtain seven (PSNR, bitrate)-pairs. Fig. 12 shows the RD curves of the two concerned

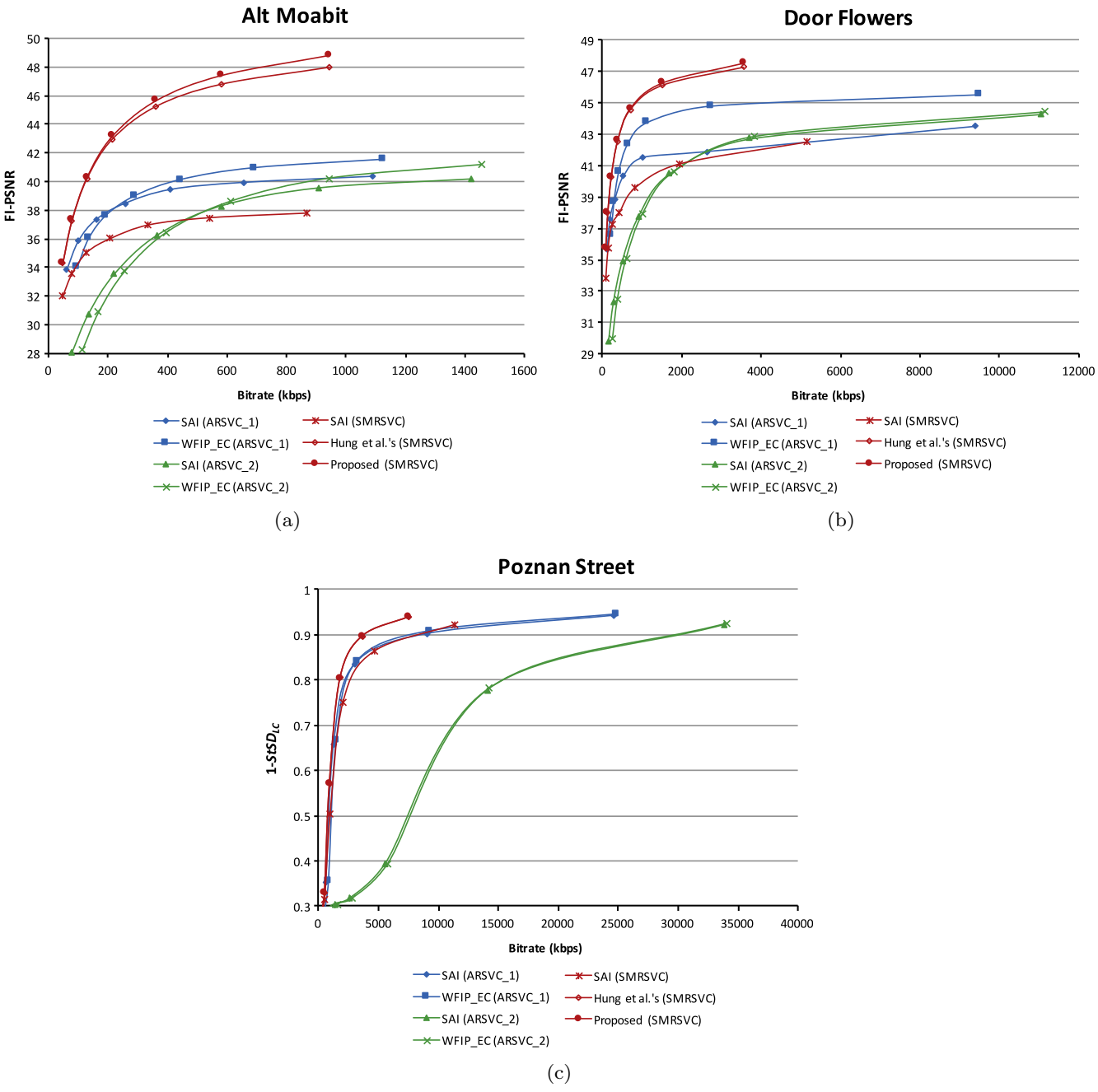

Fig. 14. In terms of the metric FI-PSNR, the RD curves of all concerned methods for (a) the Alt Moabit sequence, (b) the Door Flowers sequence, and (c) the Poznan Street sequence.

Table 4
Encoding/decoding time (in second) and bitrate (in kbps) comparison between the proposed SMRSVC model and the traditional one in MV-HEVC.

Sequence	MV-HEVC			SRMSVC		
	Encoding time	Decoding time	Bitrate	Encoding time	Decoding time	Bitrate
Alt Moabit	600	0.8	387	215	0.3	316
Butterfly	308	3.3	313	90	1.1	232
Book Arrival	2511	0.5	2673	1009	0.2	1448
Door Flowers	2446	3.7	2323	1058	1.5	1263
Poznan Street	6517	10.9	8541	3016	5.1	2843
Poznan Hall2	6690	10.1	4388	2943	4.6	1257
Average	3179	4.9	3104	1389	2.1	1226

Table 5
Execution-time (in second) comparison between Hung et al.'s super-resolution method and the proposed super-resolution method.

Sequence	Hung et al.'s	Proposed
Alt Moabit	142	267
Butterfly	69	135
Book Arrival	667	1224
Door Flowers	629	1184
Poznan Street	1637	3124
Poznan Hall2	2070	3761
Average	869	1615

super-resolution methods, respectively, for ARSVC_1 and ARSVC_2 and the three concerned super-resolution methods for the SMRSVC model. From Fig. 12, our proposed super-resolution method for the proposed SMRSVC model achieves the best bitrate and distortion tradeoff among all the concerned methods. In addition, the experimental results also demonstrated that integrating the intra-view stage and inter-view stage, which are included in the proposed super-resolution method, can improve the PSNR up to 0.2 dB when compared with using the intra-view stage only. Further, the inter-view prediction demonstrates better performance gain when QP is less or equal to 24. Therefore, the inter-view stage can be turned off for the case when QP is greater than 24.

The BD-PSNR and BD-BR essentially measure, respectively, the average differences in PSNR and bitrate of the two RD curves. Here, positive values of BD-PSNR and negative values of BD-BR imply that the proposed super-resolution method for SMRSVC has better RD performance. As shown in Tables 2 and 3, all the values of the BD-PSNR are positive and all the values of the BD-BR are negative, confirming the quality and bitrate superiority of the proposed method over the Hung et al.'s method for the SMRSVC model and the WFIP-EC method for the ARSVC_1 and ARSVC_2 models.

4.3. Stereoscopic quality comparison

In the third experiment, we adopt the two metrics, the StSD [27] and the FI-metrics [28], to evaluate the perceptual quality of the

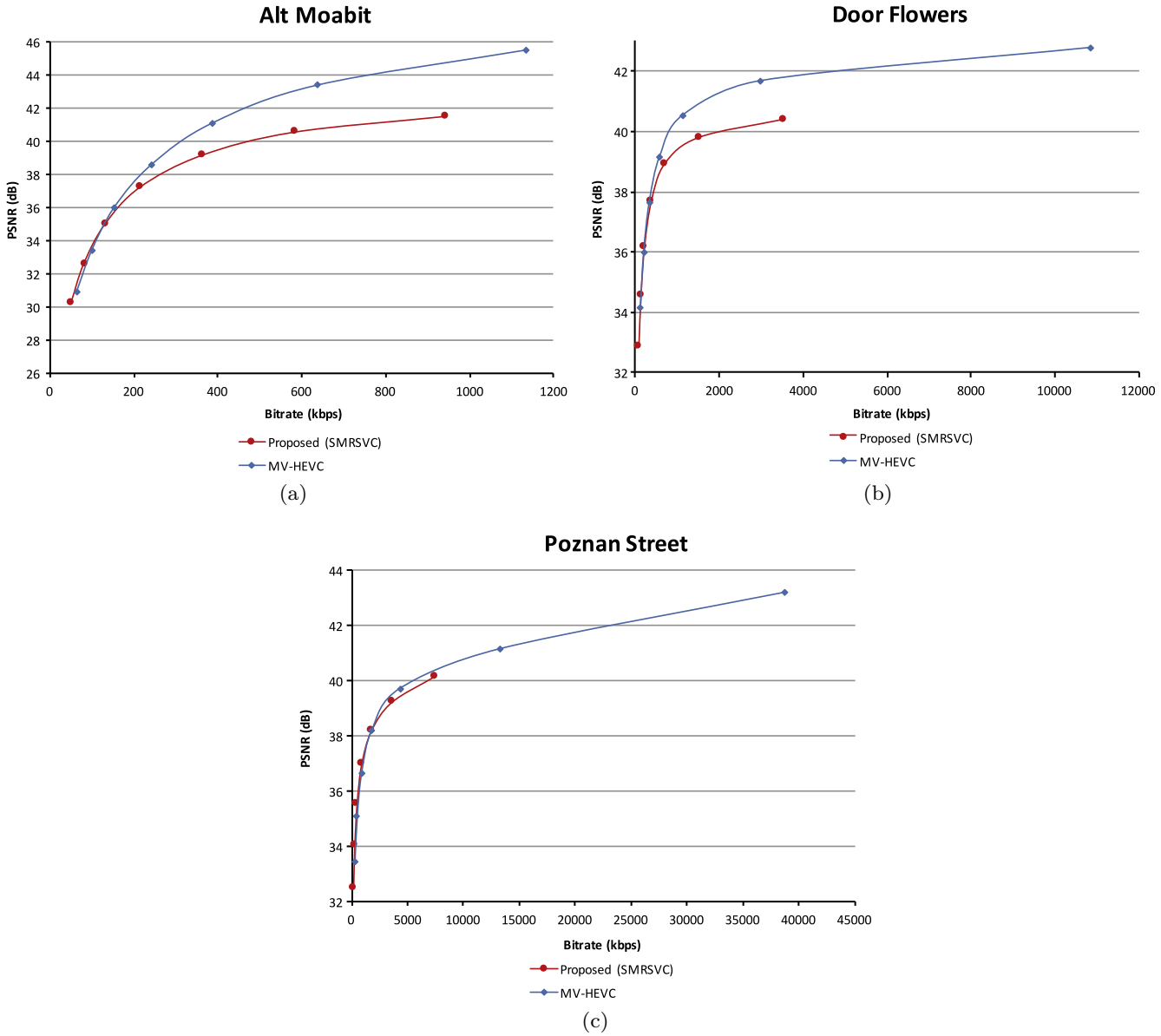

Fig. 15. In terms of PSNR, the RD curves of the traditional one in MV-HEVC and the proposed one in SMRSVC for (a) the Alt Moabit sequence, (b) the Door Flowers sequence, and (c) the Poznan Street sequence.

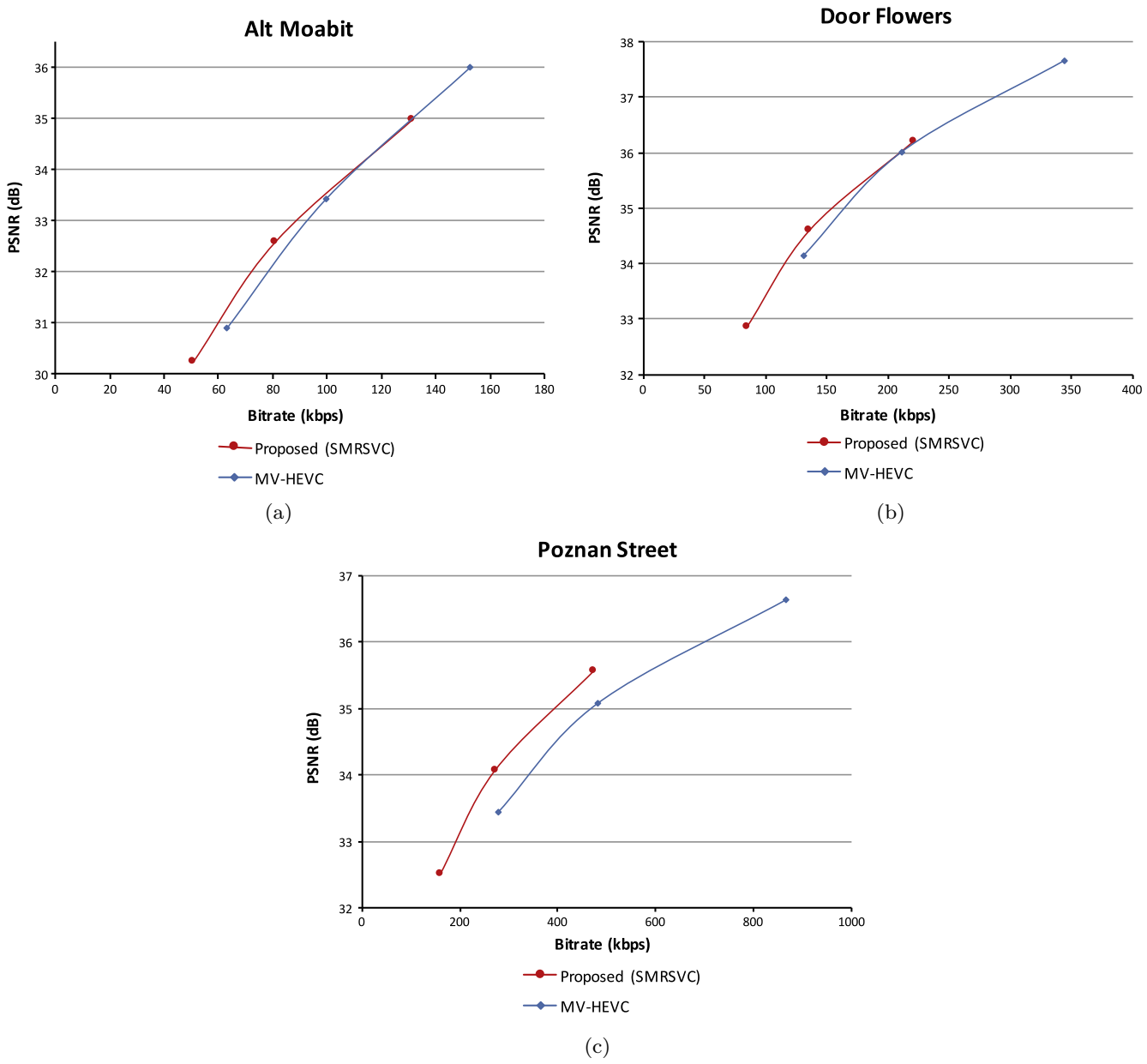

Fig. 16. In terms of PSNR for low bitrate case, the RD curves of the traditional one in MV-HEVC and the proposed one in SMRSVC for (a) the Alt Moabit sequence, (b) the Door Flowers sequence, and (c) the Poznan Street sequence.

stereoscopic video sequences. The strong consistency between the two stereoscopic quality metrics and the subjective quality assessment has been demonstrated in the experiment results in [27,28]. The metric StSD consists of three terms, one to measure the structural distortions, another to measure the blur measurement, and the other to measure the content complexity. A simplified version of StSD, called the StSDLC, has been provided by Silva et al. [25] to economically compute the value of StSDLC for measuring the stereoscopic quality, and it has the similar quality functionality of StSD. It is known that the higher the perceptual quality of the reconstructed stereoscopic video sequence is, the smaller the value of StSDLC ($0 < \text{StSDLC} < 1$) is. Here, we take (1−StSDLC) instead of StSDLC as the measurement of the perceptual quality of the stereoscopic video sequences. In addition, 'FI-metrics' which incorporates the binocular integration behaviors into the 2D quality metrics is another considered metric to measure the stereoscopic quality. Here, for reflecting more aspects of the stereoscopic quality, we combine PSNR and FI-metrics to form the metric 'FI-PSNR'.

For saving the context space, we only provide the RD curves for the three stereoscopic video sequences, Alt Moabit, Door Flowers, and Poznan Street. The RD curves in Figs. 13 and 14 demonstrate that in terms of (1−StSDLC) and FI-PSNR, the proposed method for the SMRSVC model remains the best when compared with the two concerned methods, respectively, for the two variants of the ARSVC model and the two concerned methods for the SMRSVC model. The perceptual quality merit of the proposed SRMSVC model and the proposed super-resolution method is kept for the other three stereoscopic video sequences.

In summary, based on the six test stereoscopic video sequences, in terms of several kinds of metrics, such as the quality balance, bitrate and distortion tradeoff, BD-PSNR, BD-BR, and two stereoscopic quality metrics, (1−StSDLC) and FI-PSNR, the results in the above three experimental sets have demonstrated that our proposed super-resolution method for the proposed SMRSVC model has the best performance when compared with the concerned super-resolution methods for the two variants of the ARSVC

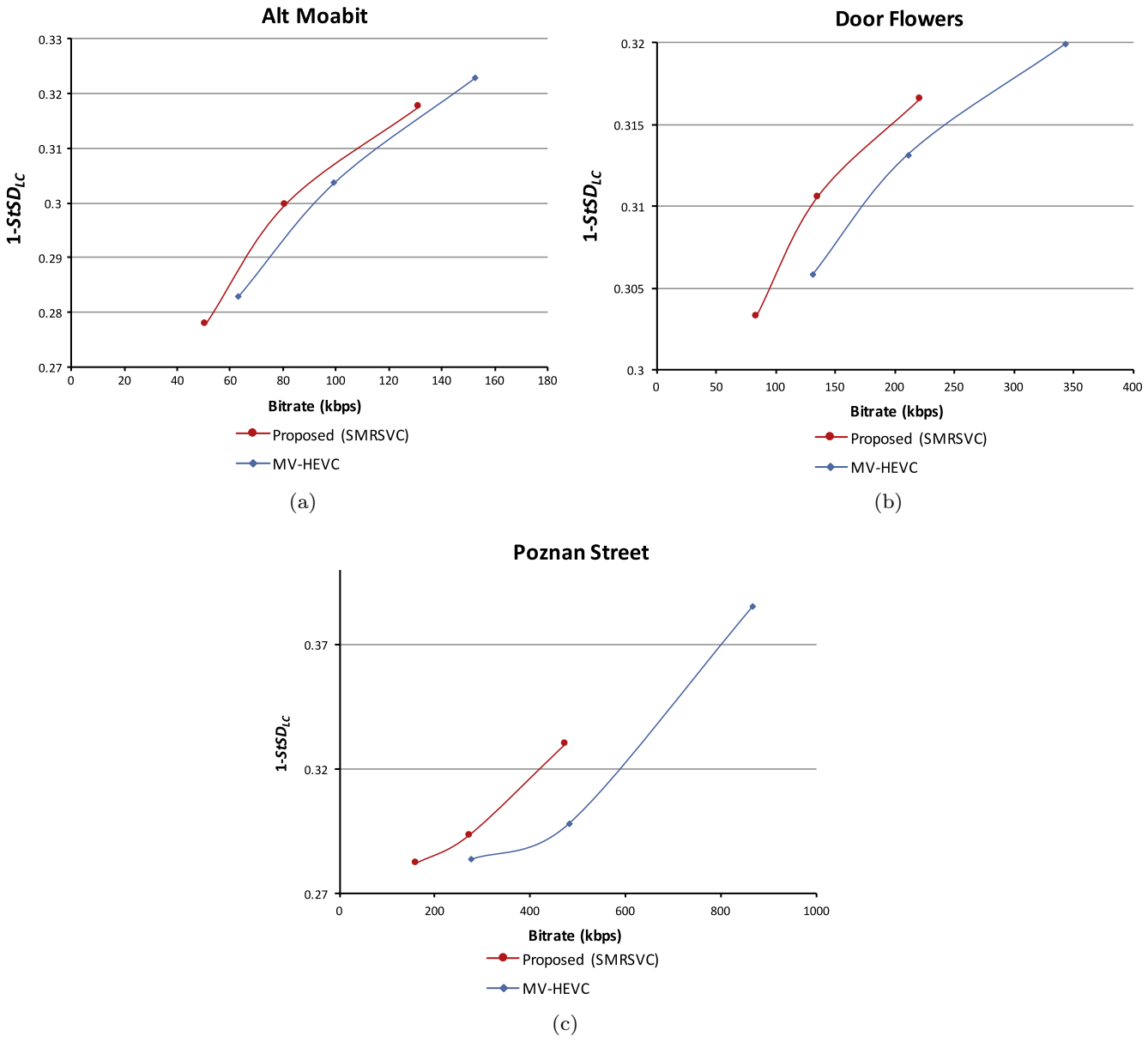

Fig. 17. In terms of 1−StSDLC for low bitrate case, the RD curves of the traditional one in MV-HEVC and the proposed one in SMRSVC for (a) the Alt Moabit sequence, (b) the Door Flowers sequence, and (c) the Poznan Street sequence.

models and the concerned super-resolution methods for the SMRSVC model.

4.4. Comparison with the traditional stereoscopic video coding model

Finally, we compare the performance of the proposed SMRSVC model with that of the traditional stereoscopic video coding model. Since in the traditional stereoscopic video coding model, each image frame in the video sequence still maintains the same size of the original image frame, the input stereoscopic video sequence can be directly encoded using MV-HEVC. Table 4 shows the encoding/decoding time and bitrate comparison between the proposed SMRSVC model and the traditional one in MV-HEVC. Note that each instance of the encoding/decoding time in Table 4 is an average result that is calculated over six QPs. From Table 4, it is clear that the proposed SMRSVC model only requires about 40% encoding/decoding time and bitrate of the traditional one in MV-HEVC on average.

Furthermore, Table 5 tabulates the execution-time of the proposed and Hung et al.'s super-resolution methods for the SMRSVC model, in which each instance of the execution-time is also an average result calculated over six QPs. The results in Table 5 show that the execution-time required in the proposed method is approximately double of that of Hung et al.'s method. The proposed method requires more execution-time mainly for the following three reasons. Firstly, the proposed method consists of the intra-view super-resolution stage and the inter-view super-resolution stage, whereas Hung et al.'s method only has the intra-view super-resolution stage. Secondly, the intra-view stage of the proposed method adopts two key-frames and one reference non-key-frame as example-frames for achieving better quality, whereas that of Hung et al.'s method only uses two key-frames as example-frames. Thirdly, unlike Hung et al.'s method, the proposed method needs to carry out the inter-view stage which adopts four example-frames, yielding an additional execution-time increase. Since more example-frames used in our proposed

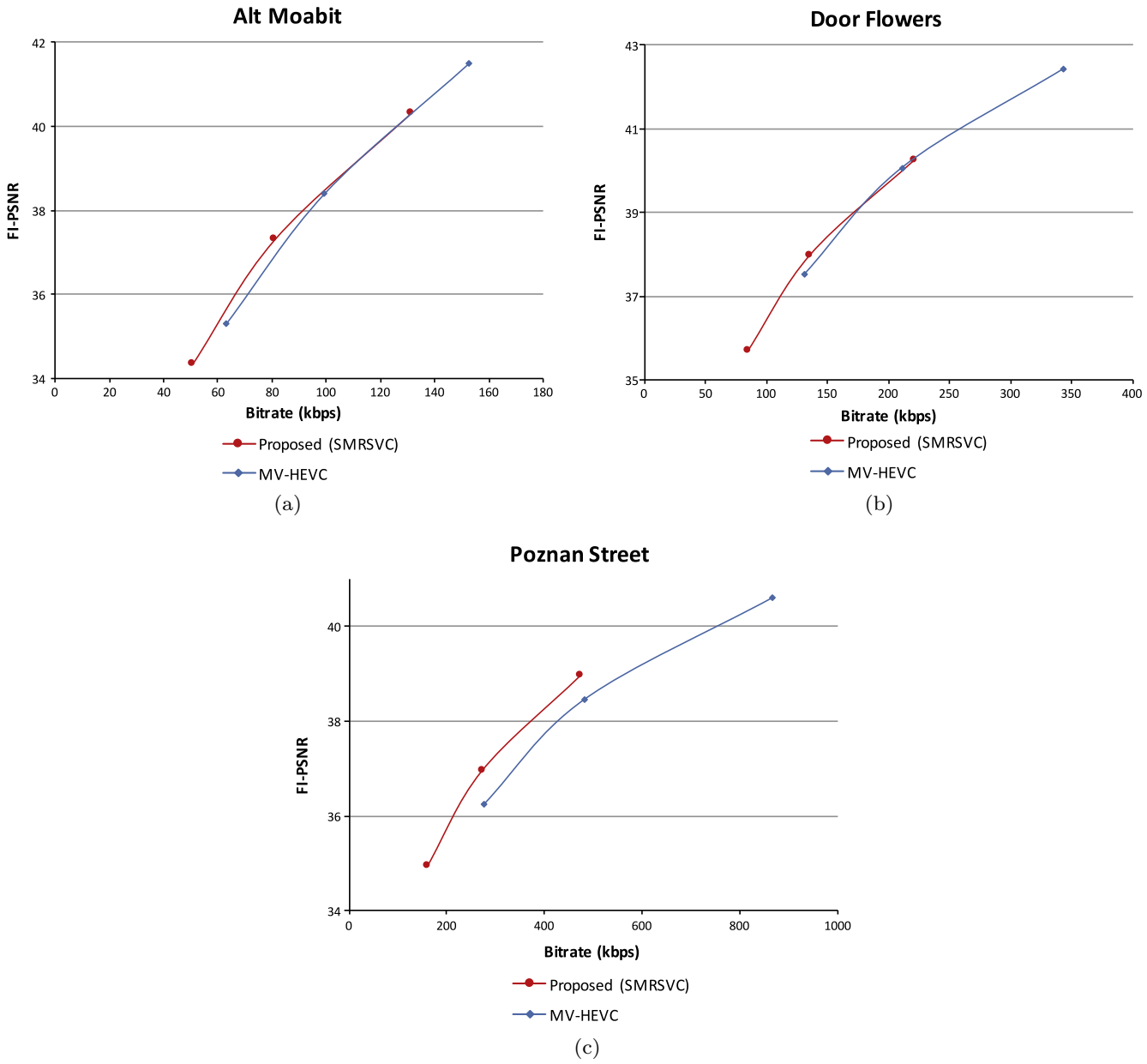

Fig. 18. In terms of FI-PSNR for low bitrate case, the RD curves of the traditional one in MV-HEVC and the proposed one in SMRSVC for (a) the Alt Moabit sequence, (b) the Door Flowers sequence, and (c) the Poznan Street sequence.

method lead to more execution-time required in the block matching process, in order to substantially reduce the execution-time in the block matching process for the proposed method, the motion vector of each block can be directly inherited from that included in the encoded SMRSVC sequence.

Fig. 15 shows the RD curves of the proposed super-resolution method for the SMRSVC model and the traditional one in MV-HEVC. Here, only the RD curves of the three stereoscopic video sequences, Alt Moabit, Door Flowers, and Poznan Street, are provided since those of the other three sequences also demonstrate similar behavior. Furthermore, in Fig. 15, due to the difficulty in examining the difference between the two concerned methods for the low bitrate case, we additionally enlarge the low bitrate case of the RD curves and show them in Fig. 16. From Figs. 15 and 16, it is clear that the proposed method for the SMRSVC model has better RD performance for low bitrate case, while the traditional one in MV-HEVC has better RD performance for the high bitrate case. Meanwhile, Figs. 17 and 18 display the RD curves for the lower bitrate case in terms of (1−StSDLC) and FI-PSNR. The curves in Figs. 17 and 18 indicate the perceptual quality merit of the proposed SRMSVC model and the proposed super-resolution method for the low bitrate case. From the above performance comparison between the proposed SMRSVC model and the traditional stereoscopic video coding model, it draws the conclusion that the proposed SMRSVC model can effectively reduce the encoding time required in the traditional stereoscopic video coding model and is more suitable for the applications in the low bitrate coding environment.

5. Conclusion

In this paper, we have presented a new model, called the symmetric mixed resolution stereoscopic video coding (SMRSVC) model, and a quality-efficient super-resolution method for the

proposed SMRSVC model. Because one stereoscopic video sequence in the proposed SMRSVC model consists of two identical mixed resolution video sequences, it could alleviate the quality-imbalance problem existed in the previous asymmetric mixed resolution stereoscopic video coding (ARSVC) model while still preserving the bitrate reduction merit. Instead of adopting one forward key-frame and one backward key-frame as example-frames in the super-resolution method by Hung et al., our proposed multiple-example based super-resolution method for the SMRSVC model fuses the high-frequency component by the intra-view stage which refers to the bi-directional key-frames, reference super-resolved non-key-frame, and then refines each super-resolved frame by the inter-view stage to increase the quality of the super-resolved videos. Based on the six test stereoscopic video sequences, in terms of the six typical metrics, the PSNR difference for measuring the quality-balance issue, the bitrate and BD-BR for measuring the compression performance, the PSNR and BD-PSNR for measuring the quality, and the (1−StSDLC) and FI-PSNR for measuring the quality of the reconstructed stereoscopic videos, the experimental results confirm the quality, visual perception, bitrate and distortion tradeoff, and stereoscopic quality advantages of our proposed super-resolution method for the proposed SMRSVC model when compared with several state-of-the-art super-resolution methods for the SMRSVC and ARSVC models.

Acknowledgment

The authors appreciate the programming help of Mr. W.C. Liu and the proofreading help of Ms. C. Harrington.

References

[1] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, A. Koz, Coding algorithms for 3DTV – a survey, IEEE Trans. Circ. Syst. Video Technol. 17 (11) (2007) 1606–1621.
[2] A. Aksay, G.B. Akar, Evaluation of stereo video coding schemes for mobile devices, in: Proc. True Vision Capture, Transmission Display of 3D Video, Potsdam, Germany, 2009, pp. 1–4. May.
[3] P. Seuntins, L. Meesters, W. Ijsselsteijn, Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation, ACM Trans. Appl. Percept. 3 (2) (2006) 95–109.
[4] F. Shao, G. Jiang, M. Yu, K. Chen, Y.S. Ho, Stereoscopic video coding with asymmetric luminance and chrominance qualities, IEEE Trans. Consumer Electron. 56 (4) (2010) 2460–2468.
[5] F. Shao, G. Jiang, M. Yu, K. Chen, Y.S. Ho, JND-based asymmetric coding of stereoscopic video for mobile 3DTV applications, in: Proc. Int. Conf. Image and Signal Processing, Shanghai, China, 2011, pp. 186–189. October.
[6] S.A. Fezza, M.C. Larabi, K.M. Faraoun, Asymmetric coding of stereoscopic 3D based on perceptual significance, in: Proc. IEEE Int. Conf. Image Processing, Paris, France, 2014, pp. 5656–5660. October.
[7] C. Fehn, P. Kauff, S. Cho, H. Kwon, N. Hur, J. Kim, Asymmetric coding of stereoscopic video for transmission over T-DMB, in: Proc. 3DTV Conf., Kos Island, Greece, 2007, pp. 1–4. May.
[8] Y. Chen, Y.K. Wang, M.M. Hannuksela, M. Gabbouj, Picture-level adaptive filter for asymmetric stereoscopic video, in: Proc. IEEE Int. Conf. Image Processing, San Diego, California, USA, 2008, pp. 1944–1947. October.
[9] Y. Chen, Y.K. Wang, M. Gabbouj, M.M. Hannuksela, Regionally adaptive filtering for asymmetric stereoscopic video coding, in: IEEE Int. Symp. Circuits and Systems, Turkey, 2009, pp. 2585–2588. May.
[10] S.N. Park, D.G. Sim, View-dependency video coding for asymmetric resolution stereoscopic views, Opt. Eng. 48 (7) (2009). 077009 (1–8).
[11] P. Aflakt, M.M. Hannuksela, J. Hakala, J. Hakkinen, M. Gabbouj, Joint adaptation of spatial resolution and sample value quantization for asymmetric stereoscopic video compression: a subjective study, in: Proc. Int. Symp. Image and Signal Processing and Analysis, Dubrovnik, Croatia, 2011, pp. 396–401. September.
[12] M. Yu, H. Yang, S. Fu, F. Li, R. Fu, New sampling strategy in asymmetric stereoscopic video coding for mobile devices, in: Proc. Int Conf. E-Product E-Service and E-Entertainment, Wuhan, China, 2010, pp. 1–4. December.
[13] K.L. Chung, Y.H. Huang, W.C. Liu, Quality-efficient upsampling method for asymmetric resolution stereoscopic video coding with inter-view motion compensation and error compensation, IEEE Trans. Circ. Syst. Video Technol. 24 (3) (2014) 430–442.
[14] W. Tam, Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV, JVT-W094, San Jose, CA, April 2007.
[15] E.M. Hung, R.L. de Queiroz, F. Brandi, K.F. Oliveira, D. Mukherjee, Video super-resolution using codebooks derived from key-frames, IEEE Trans. Circ. Syst. Video Technol. 22 (9) (2012) 1321–1331.
[16] ISO/IEC JTC1/SC29/WG11, Text of ISO/IEC 14496-10:200X/FDAM 1 Multiview Video Coding, Doc. N9978, Hannover, Germany, July 2008.
[17] G.J. Sullivan, J.M. Boyce, Y. Chen, J.-R. Ohm, C.A. Segall, A. Vetro, Standardized extensions of high efficiency video coding (HEVC), IEEE J. Select. Top. Signal Process. 7 (6) (2013) 1001–1016.
[18] C.E. Duchon, Lanczos filtering in one and two dimensions, J. Appl. Meteorol. 18 (8) (1979).
[19] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, Adaptive Basic Unit Layer Rate Control for JVT, Doc. JVT-G012, Pattaya, Thailand, 2003.
[20] X. Li, M.T. Orchard, New edge-directed interpolation, IEEE Trans. Image Process. 10 (10) (2001) 1521–1527.
[21] X. Zhang, Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation, IEEE Trans. Image Process. 17 (6) (2008) 887–896.
[22] Y. Zhang, D. Zhao, J. Zhang, R. Xiong, W. Gao, Interpolation-dependent image downsampling, IEEE Trans. Image Process. 20 (11) (2011) 3291–3296.
[23] L.W. Kang, B.C. Chuang, C.C. Hsu, C.W. Lin, C.H. Yeh, Self-learning-based low-quality single image super resolution, in: Proc. IEEE Int. Workshop on Multimedia Signal Processing, Pula, Croatia, 2013, pp. 224–229. September.
[24] L.W. Kang, C.C. Hsu, C.W. Lin, C.H. Yeh, Learning-based joint super-resolution and deblocking for a highly compressed image, IEEE Trans. Multimedia 17 (July) (2015) 921–934.
[25] http://sp.cs.tut.fi/mobile3dtv/stereo-video.
[26] G. Bjøntegaard, Calculation of average PSNR differences between RD curves, VCEG Contribution VCEG-M33, Austin, April 2001.
[27] V. De Silva, H. Arachchi, E. Ekmekcioglu, A. Kondoz, Toward an impairment metric for stereoscopic video: a full-reference video quality metric to assess compressed stereoscopic video, IEEE Trans. Image Process. 22 (9) (2013) 3392–3404.
[28] Y.-H. Lin, J.-L. Wu, Quality assessment of stereoscopic 3D image compression by binocular integration behaviors, IEEE Trans. Image Process. 23 (4) (2014) 1527–1542.
[29] M.G. Perkins, Data compression of stereopairs, IEEE Trans. Commun. 40 (4) (1992) 684696.
[30] G. Saygili, C.G. Gurler, A.M. Tekalp, Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming, IEEE Trans. Broadcast. 57 (2) (2011) 593–601.
[31] P. Aflaki, M.M. Hannuksela, J. Hakkinen, P. Lindroos, M. Gabbouj, Subjective study on compressed asymmetric stereoscopic video, in: Int. Conf. Image Processing, Hong Kong, China, Springer, 2010, pp. 4021–4024. September.
[32] B.C. Song, S.C. Jeong, Y. Choi, Video super-resolution algorithm using bi-directional overlapped block motion compensation and on-the-fly dictionary training, IEEE Trans. Circ. Syst. Video Technol. 21 (3) (2011) 274–285.
[33] C.C. Hsu, L.W. Kang, C.W. Lin, Temporally coherent super-resolution of textured video via dynamic texture synthesis, IEEE Trans. Image Process. 24 (3) (2015) 919–931.