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