Adaptive Pre-/Post-Filters for NRT-Based Stereoscopic Video Coding

Byung-Tak Lee, BongHo Lee, Haechul Choi, Jin-soo Kim, Kugjin Yun, Won-Sik Cheong, and Jae-Gon Kim

Non-real-time delivery of stereoscopic video has been considered as a service scenario for 3DTV to overcome the limited bandwidth in the terrestrial digital television system. A hybrid codec combining MPEG-2 and H.264/AVC has been suggested for the compression of stereoscopic video for 3DTV. In this paper, we propose a stereoscopic video coding scheme using adaptive pre-/post-filters (APPF) to improve the quality of 3D video while retaining compatibility with legacy video coding standards. The APPF are applied adaptively to blocks of various sizes determined by the macroblock coding mode and reference frame index. Experiment results show that the proposed method achieves up to 24.86% bit rate savings relative to a hybrid codec of MPEG-2 and H.264/AVC including the inter-view prediction.

Keywords: Stereoscopic video, adaptive loop filter (ALF), non-real-time 3DTV.

I. Introduction

Stereoscopic video has been considered as a possible approach to a 3D video service that could retain backward compatibility with an existing 2D video service. In the terrestrial digital television (DTV) system, a stereoscopic video codec should provide an additional video channel for 3D video alongside the base video channel while maintaining the existing channel capacity. Hence, there is a need for high coding efficiency. As a feasible solution, a hybrid codec consisting of MPEG-2 for a base video and H.264/AVC for an additional video has been suggested [1], [2]. The reason for the use of the MPEG-2 video codec is to maintain backward compatibility with existing 2D video services. Furthermore, non-real-time (NRT) delivery of stereoscopic video [1] is also considered as a service scenario for 3DTV services to overcome the limited bandwidth of the DTV system.

In stereoscopic video coding [3]-[6], an inter-view prediction scheme is introduced to remove view redundancy existing between the base video and the corresponding additional video. In our previous work [6], it was shown that the coding performance of the additional video can be significantly improved by allowing inter-view prediction. Furthermore, it was noted that the efficiency of the inter-view prediction relies on the quality of the reconstructed base-view picture. Hence, the reconstructed picture was used as a reference frame for the corresponding additional view in the inter-view prediction mode.

In this paper, a hybrid video codec combining MPEG-2 and H.264/AVC with inter-view prediction is adopted to compress the base video and the additional video, respectively. In addition to the hybrid video codec, we propose a stereoscopic...
video coding scheme using adaptive pre-/post-filters (APPF) to improve the coding efficiency of the additional video in NRT-based 3DTV services. The filters are designed in a similar way to the adaptive loop filter (ALF) adopted as one of the key coding tools in high efficiency video coding (HEVC) [7]. Specifically, to improve the quality of the reference picture for inter-view prediction, an adaptive filter is applied to the reconstructed base video frame as a pre-process and then the filtered picture is used as the reference. In addition, another adaptive filter is applied to the additional video picture as a post-process to directly enhance the quality of the additional video.

As mentioned previously, the design of the APPF, including such characteristics as the filter shape, filter tap length, and filter coefficient derivation, is based on the ALF used in HEVC [7]. The ALF is an adaptive Wiener filter which is a well-known optimal linear filter used to minimize the mean square error (MSE) between the original and reconstructed pictures. A picture filtered by ALF can be used as a reference picture for incoming pictures in decoding order. In general, the optimal filter coefficients of the Wiener filter are derived and signaled on a slice level. Chujoh and others [8] proposed a kind of ALF, called a block-based adaptive loop filter (BALF), and an extended version was adopted in the HEVC standard [9].

There are two main differences between the ALF and APPF approaches. One is that the APPF do not allow the filtered picture to be used as a reference picture for incoming pictures of the identical view in the inter prediction mode; this may result in the degradation of the coding efficiency obtained by the ALF. However, the constraint is intended to achieve compatibility with the legacy video coding standards MPEG-2 and H.264/AVC multiview video coding.

The second difference is the use of a coding scheme for filter information. In HEVC, the use of the ALF is switchable based on a block structure synchronized with the coding unit (CU) structure. On the other hand, in the proposed method, the switching of the APPF depends on a block structure determined by macroblock (MB) coding information, such as the MB mode and the reference frame index. These are available in H.264/AVC, whereas the CU structure is specified only in HEVC.

The adaptive filter, be it the ALF or the APPF, needs to signal side information, such as filter coefficients and the block-based filter on/off control information (control map), to the decoder side. For NRT delivery of stereoscopic video, side information for the APPF can be transmitted separately over an NRT channel. Furthermore, the additional delay required for APPF processing is likely to be feasible in NRT delivery.

This paper is structured as follows. Section II presents the delivery mechanism for NRT 3D video. Section III covers the proposed APPF-based stereoscopic video coding method. Section IV shows experiment results, followed by the conclusion in section V.

II. Delivery of NRT 3D Video

Figure 1 shows how one of the NRT 3DTV services considered by the ATSC (Advanced Television Systems Committee) 2.0 [1] can be practically realized. A file containing an additional video is transmitted through an NRT channel before a base video is broadcast in real-time over a legacy 2D channel. This NRT delivery scheme may be efficient when there is limited broadcasting channel capacity and a requirement for backward compatibility with a legacy 2D service.

In the NRT delivery scheme, stereoscopic video streams and side information produced by the proposed APPF can be transmitted through out-of-package or in-package delivery methods, as shown in Fig. 2.

In the case of out-of-package delivery, the base video is transmitted over a legacy 2D service real-time channel, while an additional video and the APPF side information are packed...
into a formatted file transmitted over an NRT channel. While this delivery method has the advantage of using exactly the same legacy 2D receiver, it does not require the transmission of a large amount of data (as a formatted file) over an NRT channel.

In the case of in-package delivery, both the base and additional videos are transmitted over the legacy real-time channel, and only ALF side information is transmitted over an NRT channel. The in-package delivery scheme can reduce the amount of data transmitted over an NRT channel at the cost of a smaller bit allocation to the base video. The proposed method will be evaluated for both delivery scenarios.

III. APPF-Based Stereoscopic Video Coding

The ALF is an adaptive Wiener filter that can process a picture corrupted by Gaussian noise, distortion, and blurring. The adaptive Wiener filter is applied to a reconstructed picture to minimize the MSE between the original and reconstructed pictures. The ALF adopted in HEVC [7] uses a set of adaptive filters, up to 16 diamond-shaped filters with a limited vertical size, and the filter selection is based on pixel activities. In particular, the tap length of a Wiener filter is chosen from 5×5, 7×7, and 9×9 filters on a slice-by-slice basis by rate-distortion (R-D) optimization. The ALF is applied not to an entire region within a picture, but to selected blocks. For this purpose, a picture is divided into various square block sizes ranging from 8×8 to 64×64, and then blocks to be filtered are optimally selected. To indicate whether a block is filtered or not, a flag is signaled on a block basis. Even though the set of Wiener filters can minimize the distortion, there is a large amount of ALF side information including a filter on/off control map, filter coefficients, the number of filters, and tap length of each filter. Transmitting such quantity of side information for each slice can degrade the coding efficiency.

To reduce the filter on/off control map, Chujoh and others proposed a quadtree-based adaptive loop filter (QALF) [9]. In HEVC, the filtering block split information to be signaled is synchronized to the CU structure. That is, block splits of the ALF follow the CU splits except for the minimum block size to be filtered, which is specified by the control depth of the ALF, as shown in Fig. 3. The control depth is optimally determined by the R-D cost. In this way, side information for block splits used in the QALF does not need to be transmitted.

In the proposed method, a hybrid video codec combining MPEG-2 and H.264/AVC is adopted. H.264/AVC has the MB structure instead of the CU structure. Thus, ALF information related to the CU structure cannot be directly used. It might be assumed that H.264/AVC is replaced with HEVC as a compression codec for an additional video; however, the standardization of HEVC is still in progress and HEVC is not yet a stable standard. H.264/AVC has already been adopted in various video applications. Therefore, we propose a method of applying adaptive Wiener filters (APPF) to H.264/AVC. As mentioned previously, H.264/AVC has an MB structure and the proposed method uses the MB-based coding information, such as the MB mode and reference frame index, to avoid the transmission overhead needed if block splits of the adaptive Wiener filter are sent. Specifically, if the adaptive Wiener filter is applied, it is applied to blocks of various sizes, such as 8×8, 16×16, 32×32, and 64×64. The procedure for selecting the block size is depicted in the flow chart of Fig. 4.

First, for a 64×64 block structure, if all MBs in the 64×64 block are coded as Inter 16×16 (P_16×16), Inter 16×8, Inter 8×16, or Intra 16×16 and have the same coding property (such as SKIP), intra prediction mode, or reference index for inter prediction, then the 64×64 block is not divided for the adaptive Wiener filter. Otherwise, the 64×64 block is divided into four 32×32 blocks. In the same way, each 32×32 block may or may not be divided into four 16×16 blocks. For a 16×16 block, the split is performed on the condition that the MB mode of the 16×16 block is a sub-MB (P_8×8), Intra 8×8, or Intra 4×4. The block boundary of the adaptive Wiener filter is aligned with the MB or sub-MB boundary. By using the proposed filter block split method, side information for the block split can be inferred on the decoder side without transmitting it explicitly. The motivation for using the MB properties is based on an observation that MBs with the same MB properties have similar error characteristics between the original and reconstructed pictures. The ALF_flag specifying the filter on/off for each block should still be signaled per a split block, as shown in Fig. 3.

1) For adaptive filtering, a picture is partitioned into various block sizes, which is referred to as filter block split in this paper. In principle, the filter block split is different from the prediction block partition of H.264/AVC.
In the proposed filter block split method, the number of homogenous blocks by which a block is split can be treated as a threshold. Through experiments where thresholds are applied differently, it was found that the relaxation of such a condition of homogeneity for the coding properties may improve the coding efficiency. Figure 5 shows an instance of the proposed block split determination with the relaxed condition: if 14 MBs of a 64×64 block are coded by the SKIP mode rather than 16 MBs, the 64×64 block is determined to be undivided.

Figure 6 shows the overall codec structure when adopting the APPF for stereoscopic video coding. A base video is coded by MPEG-2 and transmitted over a real-time channel, and an additional video is coded by H.264/AVC with inter-view prediction. The resulting bit stream is formatted into a file to be transmitted over an NRT channel prior to the base video. The adaptive pre-filter is applied to the reconstructed picture of the base video in a pre-processor to enhance the picture quality for use as the reference picture in inter-view prediction mode. In the post-processor, the adaptive post-filter is applied to the reconstructed additional video to enhance the picture quality directly.

The APPF side information, such as the filter coefficients and control information for filtering, is transmitted over an NRT channel. The 2D base video transmitted over the real-time channel is decoded and may be directly displayed, whereas 3D video can be provided by decoding the additional video transmitted over the NRT channel. In pre-/post-processing on the decoder side, the APPF are applied to the base and additional videos, respectively, in the same way as on the encoder side by using the delivered APPF side information.

IV. Experiment Results

To evaluate the proposed stereoscopic video coding scheme, some high-definition stereoscopic video sequences, namely, “Balloons,” “Kendo,” “Lovebird1,” and “Newspaper” [10], are used. The base video is encoded using MPEG-2 with a constant bit rate while applying the rate control method [11]. This corresponds with the general usage for legacy 2D
broadcasting, and the essential encoding parameters for the base video are listed in Table 1.

The additional video is encoded without rate control, and the proposed method is evaluated at several different quantization parameters (QPs) according to the Bjøntegaard delta bit rate (BDBR) and the Bjøntegaard delta peak signal-to-noise rate (BD-PSNR) [12] and according to the R-D performance. In addition, we evaluate the performance for the two delivery scenarios, namely, out-of-package and in-package. The additional video is encoded using a codec based on JMVM 8.0 [13], which is allowed for inter-view prediction, the encoding parameters of which are listed in Table 2.

In the following experiments, the BDBR and BD-PSNR are measured relative to an anchor that is coded using MPEG-2 for the base video and H.264/AVC including the inter-view prediction mode for the additional video. The anchor did not use the adaptive Wiener filter.

First, to evaluate the proposed adaptive filter, we code test sequences as single view encodings using H.264/AVC, H.264/AVC with a modified BALF, and H.264/AVC with the proposed adaptive post-filter. The modified BALF is the same as the BALF [8] except that a filtered picture is not allowed (for fairness) to be used as a reference picture for the inter-prediction mode. Figure 7 shows the R-D performance of the three coding conditions. It should be noted that the proposed adaptive post-filter outperforms the modified BALF as well as H.264/AVC alone. Unlike the adaptive post-filter, BALF uses a constant block size in a slice and the block size is determined as one of 8×8 to 128×128 on a slice basis. Therefore, this experiment result verifies that allowing various block sizes in

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution/frame rate</td>
<td>720p/30 fps</td>
</tr>
<tr>
<td>Number of frames</td>
<td>150</td>
</tr>
<tr>
<td>GOP structure</td>
<td>IBBP (intra period = 24)</td>
</tr>
<tr>
<td>Profile@level</td>
<td>Main profile@high level</td>
</tr>
<tr>
<td>Coding rate</td>
<td>In-package: 4.511 Mbps</td>
</tr>
<tr>
<td></td>
<td>Out-of-package: 6.447 Mbps</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy coding method</td>
<td>CABAC</td>
</tr>
<tr>
<td>Basis QP</td>
<td>{22, 27, 32, 37}</td>
</tr>
<tr>
<td>GOP structure</td>
<td>4 IBBBP (GOP size 4)</td>
</tr>
<tr>
<td>Intra period</td>
<td>24</td>
</tr>
<tr>
<td>Number of views minus one</td>
<td>1</td>
</tr>
<tr>
<td>View order</td>
<td>0-1</td>
</tr>
</tbody>
</table>

![Fig. 7. R-D performance of H.264/AVC, H.264/AVC with modified BALF [8], and H.264/AVC with proposed adaptive post-filter.](image-url)
the filter block split process to be dependent on local characteristics is more effective than using blocks of a constant size determined on a slice basis. Furthermore, the proposed method does not signal overhead bits for the block split.

For the case when both the proposed pre-filter and post-filter are applied for out-of-package delivery, the BDBR and BD-PSNR results are shown in Table 3. In these results, the rate includes the rate of filter side information as well as that of the additional video. It can be seen that applying the adaptive post-filter to the additional video achieves an average 10.47% bit savings. The reduction verifies that the adaptive post-filter improves the quality of reconstructed additional pictures by minimizing the error between the original and reconstructed pictures. On the other hand, applying only the adaptive pre-filter shows a reduction in coding efficiency. This result suggests that using only the pre-filter may increase bit rates for filter side information disproportionately to improvements in the quality of reference pictures for the inter-view prediction mode.

Table 4 shows the coding gain for the case of in-package delivery. Unlike the case of out-of-package delivery, the rate does not include the rate of ALF side information since it can be assumed that the side information is transmitted independently through an NRT channel. As shown in the results, there is up to a 24.86% bit rate improvement when applying the proposed APPF. This is a result of the quality improvement of the reference pictures for the inter-view prediction mode and the reconstructed additional pictures. It is worth noting that the quality of the base and additional videos transmitted through a legacy broadcasting channel could be enhanced significantly if the proposed filter information were freely transmitted through an NRT channel instead of the bandlimited broadcasting channel.

Figure 8 shows PSNR variations over 150 frames for sequences Balloons and Kendo when the pre-/post-filters are applied. The average coding gains for the cases are shown in Table 4 (the last column). The proposed pre-/post-filters consistently outperform the anchor over all frames for the sequences.

V. Conclusion

This paper proposed an efficient stereoscopic video coding scheme for NRT stereoscopic 3D services. The proposed method added the adaptive pre-/post-filters to a hybrid codec combining MPEG-2 and H.264/AVC. The coding efficiency of the additional video channel required for 3D was enhanced by using higher quality reference pictures for the inter-view prediction mode; these were generated by applying a pre-filter.
to the reconstructed base video. Furthermore, the quality of the additional video was significantly improved by applying a post-filter to the reconstructed additional video. To reduce the amount of side information for the filters, MB coding information was utilized to determine the block sizes to be filtered.

The proposed scheme showed up to a 24.86% coding gain improvement over an anchor coded using MPEG-2 for the base video and H.264/AVC including the inter-view prediction mode for the additional video. The result was obtained under the assumption that the side information for the proposed pre-/post-filters is transmitted through an NRT channel. The proposed scheme is expected to be well suited to NRT 3D services in which a separate NRT channel is available to deliver the side information for the pre-/post-filters or a combination of that side information and the additional video.

References


Byung-Tak Lee received his BS and MS in electronics engineering from Korea Aerospace University, Goyang, Gyeonggi-do, Rep. of Korea, in 2009 and 2011, respectively. He is currently an assistant engineer in Samsung Electronics. His research interests include video coding and video signal processing.

BongHo Lee received his MS in avionics engineering from Korea Aerospace University, Goyang, Gyeonggi-do, Rep. of Korea, in 1999. In 1999, he joined ETRI, Daejeon, Rep. of Korea, where he has been working on broadcasting system technologies. Currently, as a senior researcher, he is involved in the Realistic Broadcasting System Research Team focusing on the R&D of 3DTV technologies. His main research interests are mobile multimedia broadcasting systems, 3DTV, and digital hologram.

Haechul Choi received his BS in electronics engineering from Kyungpook National University, Daegu, Rep. of Korea, in 1997 and his MS and PhD in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 1999 and 2004, respectively. He is an assistant professor in the Department of Information and Communication Engineering at Hanbat National University, Daejeon, Rep. of Korea. From 2004 to 2010, he was a senior member of the research staff of the Broadcasting Media Research Group of ETRI, Daejeon, Rep. of Korea. His current research interests include image processing, video coding, and pattern recognition.

Jin-soo Kim received his BS in electronics engineering from Kyungpook National University, Daegu, Rep. of Korea, in 1991 and his MS and PhD in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 1993 and 1998, respectively. From 1998 to 2000, he was with the Business Division of System LSI at Samsung Electronics, where he was involved in the development of MCU chipsets. Since March 2000, he has been a faculty member in the Department of Information and Communication Engineering, Hanbat National University, Rep. of Korea, where he is a professor. His research interests include distributed video coding (DVC), high...
efficiency video coding, networked video rate shaping, and adaptation and media convergence.

Kugjin Yun received his BS and MS in computer engineering from Cheonbuk National University, Jeonju, Rep. of Korea, in 1999 and 2001, respectively. He joined ETRI, Daejeon Rep. Of Korea, in 2001, and he is currently a senior research member of the Realistic Broadcasting System Research Team. His current research interests include 3DTV and 3D mobile broadcasting.

Won-Sik Cheong received his BS, MS and PhD in electronic and electrical engineering from Kyungpook National University, Daegu, Rep. of Korea, in 1992, 1994, and 2000, respectively. He joined ETRI, Daejeon, Rep. of Korea, in 2000. Currently, he is the director of the Realistic Broadcasting Technology Research Team of the Broadcasting Systems Research Department. His research interests include 3DTV broadcasting systems, interactive multimedia broadcasting systems, video and image coding, and digital signal processing.

Jae-Gon Kim received his BS in electronics engineering from Kyungpook National University, Daegu, Rep. of Korea, in 1990, and his MS and PhD in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 1992 and 2005, respectively. From 1992 to 2007, he was with ETRI, Daejeon, Rep. of Korea, where he was involved in the development of digital broadcasting media services, MPEG-4/7/21 standards and related applications, and convergence media technologies. From 2001 to 2002, he was a staff associate in the Department of Electrical Engineering at Columbia University, New York, NY, USA. He is currently an associated professor in the School of Electronics, Telecommunication, and Computer Engineering at Korea Aerospace University, Goyang, Gyeonggi-do, Rep. of Korea. His research interests include video signal processing, video coding, video adaptation, and multimedia applications.