MIXED RESOLUTION CODING WITH INTER VIEW PREDICTION FOR MOBILE 3DTV

Heribert Brust\textsuperscript{(1)}, Gerhard Tech\textsuperscript{(1)}, Karsten Mueller\textsuperscript{(1)}, Thomas Wiegand\textsuperscript{(1)\textsuperscript{(2)}}

\textsuperscript{(1)} Fraunhofer Institute for Telecommunications Heinrich-Hertz-Institut Image Processing Department
Einsteinufer 37, 10587 Berlin, Germany heribert.brust@hhi.fraunhofer.de

\textsuperscript{(2)} Technical University of Berlin Image Communication, Faculty of EE & CS Einsteinufer 17, 10587 Berlin, Germany

ABSTRACT

This paper presents results on mixed resolution stereo video coding including inter view prediction. The objective tests show that coding one view of a stereo pair at a lower resolution outperforms coding of a low pass filtered view at base resolution. For both coding methods inter view prediction produces additional coding gains. Two different prediction approaches have been compared: One uses the lower resolution view as base view, which is used to predict the second view, the other uses the higher resolution view as base view. The prediction is realized by up-sampling or down-sampling the decoded base view before predicting. It turned out that equal results are achieved for low bit rates, while predicting from high resolution to low resolution view performs better for medium and high bit rates. During coding, different OP combinations for left and right view have been compared to achieve the maximum combined PSNR.

Index Terms — Mixed resolution, stereo coding, H.264, MVC, mobile services

1. INTRODUCTION

The interest in mobile entertainment devices and 3DTV increased over the past years. The combination of both techniques is an important research topic. The requirements of mobile 3DTV techniques are low bit rate, high quality stereoscopic views and low complexity. While the transmission of video and depth, which is described in [1], requires relatively complex view rendering at the receiver side, the transmission of two color views requires only decoding and displaying of the transmitted views. There are several techniques for stereo video coding like simulcast, H.264/AVC with stereo SEI message, H.264/MVC and mixed resolution (MR) coding [2][3]. Mixed resolution coding uses the binocular suppression theory for a reduction of the bit rate [4]. This theory states that the perceived visual quality of stereoscopic video with different sharpness of left and right view is rated close to the sharper view [5][6]. This means that a stereo video with one spatial sub-sampled view can be transmitted at lower bit rates in comparison to the full resolution (FR) case and yield the same subjective quality. It was observed in [7] that the subjective impairment, caused by different sharpness of left and right view, is reduced with increasing resolution and viewing distance and decreasing display size. Subjective tests showed that MR coding enhanced the overall quality on a small screen with a typical size used on mobile phones. The optimization of the bit rate distribution to left and right view was conducted in [7] by maximizing the mean PSNR of both views. It turned out that the maximum PSNR is obtained at a distribution of 30% to 35% of the total bit rate for the lower resolution view, applying sub-sampling by a factor of two in both directions. While the tests in [7] did not use any inter view (IV) prediction, it has been shown in [8] that a coding gain can be achieved by combining MR coding with IV prediction.

This paper compares different methods of MR coding with inter view prediction. Section 2 compares the rate distortion (RD) curves of coding a low pass filtered view to coding of a down-sampled view. It was further investigated which quality should be chosen to code left and right view when IV prediction is used. This was tested for the prediction from the low resolution view to the high resolution view (section 3) as well as from the high resolution view to the low resolution view (section 4). The best combinations of both prediction methods are compared in section 5 and the conclusion is given in section 6.

2. LOW PASS FILTERING AND SUB-SAMPLING

For the evaluation in this paper, the stereo videos Hands, Snail, Horse and Car, produced by KUK Filmproduktion GmbH, were used in this test. They were coded using the H.264/MVC JMVM 7.0 reference software [9] with typical coder settings for mobile applications: i.e. a GOP size of 2 with Intra Period 16 and CAVLC Symbol Mode. The quantization parameter (QP) varied from 20 to 44 with a step size of two.

The right view was coded as the base view with full resolution (480x272 pixels). For coding the left view different methods were compared. One method applied a low pass filter in row wise and column wise. For this, the non-normative dyadic filter in (1) from the JSVM reference software was used [10].

\[
\begin{pmatrix}
2 & 0 & -4 & -3 & 5 & 19 & 26 & 19 & 5 & -3 & -4 & 0 & 2
\end{pmatrix}/64
\] (1)

The low pass filtered left view was coded at full resolution without IV prediction as well as with IV prediction from the right view. For objective evaluation, the mean squared error (MSE) was calculated by comparing the decoded view to the low pass filtered uncoded view.

Another coding method which was compared in this test was applying the low pass filter from (1) and sub-sampling by a factor of two in both directions. After coding at a low resolution of 240x136 pixels, the decoded view was up-sampled again to...
the original resolution with the normative dyadic filter from (2) used in the JSVM reference software.

\[
(1 \ 0 \ -5 \ 0 \ 20 \ 32 \ 20 \ 0 \ -5 \ 0 \ 1)/32 \quad (2)
\]

Coding of the sub-sampled view was carried out without IV prediction, as well as with IV prediction. IV prediction in this case was realized by applying the low-pass filter and sub-sampling process to the decoded right view and using this low resolution view for prediction. In this case the MSE was calculated by comparing the decoded up-sampled view with the uncoded down- and up-sampled view according to [7]. The MSE was used to calculate the PSNR separately for each view and coding method.

In both cases, i.e. coding of the low pass filtered view and coding of the sub-sampled view, the mean squared error only reflects impairments caused by coding, while low pass filtering is not considered. This approach follows the binocular suppression theory when the binocular quality of a MR sequence is objectively evaluated.

Figure 1 shows the rate distortion (RD) curves of the tested sequences. Coding the right (full resolution) view requires the highest bit rate for a given PSNR value due to the high spatial frequencies. Coding the left, low pass (LP) filtered view requires a lower bit rate and more coding gain is achieved, when IV prediction from the right view is applied. The bit rate reduction caused by IV prediction increases for higher rates. For all bit rates and tested sequences the coding performance is better for the down-sampled (DS) view than for the low pass filtered view. This means that coding a sequence with high resolution requires more bit rate than coding a lower resolution sequence if both have similar spatial frequencies. The combination of down-sampling and IV prediction yields the highest coding gain. The following section describes further optimization of this coding method.

3. PREDICTION FROM HIGH RESOLUTION TO LOW RESOLUTION VIEW

Coding the views of a MR stereo pair with different quality for left and right view can result in higher mean PSNR values as shown in [7]. The mean PSNR is calculated by considering all pixels of both views. This was done by averaging the MSEs and then calculating a total PSNR from that.

Different quality of both views is set by choosing different QP values. The test, described in section 2, used equal QP values for left and right view, when IV prediction was used.

Next, we investigated how the overall quality can be improved when left and right views are coded with different QP values. In order to find the best QP combination, the full resolution right view was coded with the JMVM 7.0 reference software, using the same coding parameter as in section 2 but with a QP step size of 1 (ranging from 20 to 44). The decoded right view was low-pass filtered, sub-sampled by a factor of two in both direction and used for IV prediction of the low resolution left view. For low-pass filtering the coefficients shown in (1) were applied row wise and column wise. This means that the base view was coded at a resolution of 480x272 pixels, while the second IV predicted view was coded with a resolution of 240x136 pixels. The low resolution left view was coded with a QP step size of 4 (20 to 44).

After coding and decoding, the low resolution left view was up-sampled again to the base resolution of 480x272 pixels row wise and column wise with the filter coefficients from (2). Finally the total PSNR of both views was calculated by using the above described method. Figure 2 shows the total PSNR versus the total bit rate of both views. Each RD point belongs to one QP combination of left and right view. RD points with equal QP value for the low resolution left view are connected to one curve. The shown number pairs in Figure 2 are the QP values for the low resolution left view (first number) and the high resolution right view (second number) for some RD points. These points are closest to the envelop curve of best RD performance, and the QP combinations give the highest total PSNR for the particular bit rates.

It can be seen that for all sequences the selected QP for the left view is lower than the QP for the right view. This effect has already been observed in [7] for coding with different QP values without interview prediction. The reason for this effect is the perceptual suppression of artifacts introduced by low pass filtering. Low pass filtering, in contrast to coding, is not considered by the total PSNR. As for the left view, fewer pixels need to be coded, but the MSE has the same influence on the total PSNR. Therefore, this view is coded with higher quality.

---

**Figure 1.** PSNR vs. bit rate for right view with base resolution and low pass (LP) filtered or down-sampled (DS) left view, with inter view (IV) prediction and without inter view prediction - Sequences: Hands, Snail, Horse and Car

**Figure 2.** PSNR vs. bit rate for low resolution left view, coded with different QP values - Sequences: Hands, Snail, Horse and Car
4. PREDICTION FROM LOW RESOLUTION TO HIGH RESOLUTION VIEW

This section describes the coding of MR sequences with IV prediction, in which the low resolution view is coded as base view. It was investigated which QP combination results in the highest total PSNR when IV prediction from the low resolution left view to the high resolution right view is used.

The left view was coded at a resolution of 240x136 pixels with the above described coder settings and a QP step size of one (5 to 44). The decoded left view was up-sampled and used for IV prediction of the right view. A QP step size of four (20 to 44) was used for coding the right view.

Figure 3 shows the total PSNR versus total bit rate. Rate Distortion points with equal QP values for the right view are connected. The QP combinations for the low resolution left view (first number) and high resolution right view (second number) are shown for rate distortion points, which are close to the enveloping curve. It can be seen that with this prediction structure the left view is coded with a lower QP value, which was also observed in section 3.

5. COMPARISON OF BOTH PREDICTION DIRECTIONS

The QP value for the low resolution left view is below the QP value for the right view for both prediction directions. As described earlier this results from the lower number of pixels to encode for this view. The difference between both QP values is higher for the prediction from low resolution view to high resolution view than vice versa. The optimization with IV prediction allocates more rate for the base view which means that both views benefit from this. Figure 4 shows the comparison of the RD curves of both prediction directions with the optimized QP combinations of Figures 2 and 3. It can be seen that for low bit rates both prediction methods perform equally, while for medium and high bit rates the prediction from the low resolution view to the high resolution view achieves higher total PSNR values.
In comparison to MR coding without IV prediction and equal QP values for left and right view, both described prediction structures give a high coding gain, as shown in Figure 4.

The bit rate distributions for the chosen RD points of Figure 2 are between 14% and 42% for the low resolution left view compared to the total bit rate.

6. SUMMARY AND CONCLUSIONS

Coding of low pass filtered and down-sampled views have been investigated both with and without IV prediction. The combination of down-sampling and IV prediction yields the highest coding gains. Thus, lower bit rates are required for coding the low resolution view compared to the full resolution view. Two directions of IV prediction have been compared: Prediction from the high resolution view to the low resolution view and vice versa. For low bit rates both prediction directions perform equally, while for medium and high bit rates the prediction structure from high resolution to low resolution performs better.

Future work on this topic will include subjective evaluation of the MR coding with IV prediction.

7. ACKNOWLEDGEMENTS

This work was supported by EC within FP7 (Grant 216503 with the acronym MOBILE3DTV). We would like to thank KUK Filmproduktion GmbH, Munich, Germany, for providing the Hands, Snail, Horse and Car stereo video data (http://www.kuk-film.de/).

8. REFERENCES