CODING AND INTERMEDIATE VIEW SYNTHESIS OF MULTIVIEW VIDEO PLUS DEPTH

Karsten Müller1, Aljoscha Smolic1, Kristina Dix1, Philipp Merkle1, and Thomas Wiegand1,2

1 Image Processing Department
Fraunhofer Institute for Telecommunications - Heinrich Hertz Institute (HHI)
Einsteinufer 37, 10587 Berlin, Germany
{kmueller/smolic/dix/merkle/wiegand}@hhi.de

2 Image Communication Chair
Department of Telecommunication Systems
School of Electrical Engineering and Computer Sciences
Technical University of Berlin
Einsteinufer 17, 10587 Berlin, Germany

ABSTRACT
For advanced 3D Video (3DV) applications, efficient data representations are investigated, which only transmit a subset of the views that are required for 3D visualization. From this subset, all intermediate views are synthesized from sample-dense color and depth data. In this paper, the method for reliability-based view synthesis from compressed multi-view + depth data (MVD) is investigated and corresponding results are shown. The initial problem in such 3DV systems is the interdependency between view capturing, coding and view synthesis. For evaluating each component separately, we first generate results from the coding stage only, where color and depth coding is carried out separately. In the next step, we add the view synthesis stage with reliability-based view synthesis and show, how the separate coding results influence the view synthesis quality and what type of artifacts are produced. Efficient bit rate distribution between color and depth is investigated by objective as well as subjective evaluations. Furthermore, quality characteristics across the viewing range for different bit rate distributions are analyzed. Finally, the robustness of the reliability-based view synthesis to coding artifacts is presented.

Index Terms— View synthesis, MVD, 3D video, video coding, MVC.

1. INTRODUCTION
Autostereoscopic multi-view displays require capturing and transmission of multi-view data. Typical displays require N (e.g. N=9) views [1] at the receiver side. For efficient transmission, coding of multiview data was investigated, where dependencies between the views are exploited in addition to temporal dependencies within each sequence [2]. Although a reduction in bit rate was achieved in comparison to simulcast coding, the overall bit rate remained to be proportional to the number of views and with that too high for transmission.

Therefore, a multi-view representation format was investigated, that only includes a small number of views (e.g. 2 or 3) together with associated per-sample depth data [3]. The latter allows the reconstruction of intermediate views allowing the input views for multi-view displays to be synthesized. High-quality intermediate view synthesis is crucial for the applicability of a multi-view + depth (MVD) format to 3DV systems in order to provide constant quality over the viewing range. Until recently, synthesis techniques still showed quality degradation for synthesized views between original camera positions [3]. Accordingly, an improved synthesis algorithm was introduced [4] and results for uncompressed data were shown.

This paper combines coding and view synthesis by investigating efficient coding of MVD data for high-quality intermediate view synthesis. In section 2, the effects of coding artifacts in MVD data on view synthesis are analyzed. The reliability-based synthesis approach is presented in section 3. Detailed results for the coding techniques are shown in section 4. These results are used for combined coding and view synthesis with reliability-based layer processing in section 5.

2. GEOMETRY DISTORTION IN DEPTH-BASED APPLICATIONS
One problem for the design of 3DV systems is the interdependency between view capturing, coding, and view synthesis. The quality of the system is judged for all views including the synthesized views. All stages of the MVD system strongly influence these results. While the original camera setting influences the capability to generate good-quality views at any viewpoint, coding introduces compression artifacts. In case of an MVD representation, artifacts in rendered views are introduced directly because of coding of color data as well as indirectly through coding of depth data.

For the influence of coding artifacts, we look into the view synthesis, derived in [5]. Consider an arbitrary sample $s_k$ in an intermediate view at position $(u_k, v_k)$, which is interpolated from appropriate original samples $s_0$ and $s_1$ at associated positions:

$$s_k[x, y] = (1 - \kappa) \cdot s_0[x, y] + \kappa \cdot s_1[x, y].$$

Here, $\kappa \in [0..1]$ represents the intermediate position parameter, which specifies the intermediate position between views $s_0$ and $s_1$. For instance, a value of $\kappa = \frac{1}{2}$ specifies the middle position between both original views. For each original sample, depth values $z_0$ and $z_1$ exist, but for simplicity, we assume a strictly parallel camera setting, where depth can be converted into horizontal disparity shifts $d_0$ and $d_1$ (otherwise, a disparity shift in horizontal and vertical directions occurs). Thus, the above equation can be modified to relate the horizontal sample position in the intermediate view $u_k$ with the positions in the original views $u_0$ and $u_1$, using the $\kappa$-scaled disparity shifts:

$$s_k[x, y] = (1 - \kappa) \cdot s_0[x, y - (1 - \kappa) \cdot d_0, y] + \kappa \cdot s_1[x, y - d_1, y].$$

During coding, color and depth values are subject to coding errors, such that their reconstructed values differ from the original. While color errors only change the interpolation value, depth errors cause sample shifts $\Delta d_0$ and $\Delta d_1$ such that erroneous neighboring samples from original views are used for the individual color contributions:
\[ \hat{s}_c(x_c, y_c) = (1 - \kappa) \cdot \hat{s}_i(x_i, y_i) - (1 - \kappa) \cdot (d_0 + \Delta d_0), y_c \] 

\[ + \kappa \cdot \hat{s}_i(x_i - \kappa \cdot (d_i + \Delta d_i), y_c) \]  

(3)

Formula (3) implies that coding errors at depth edges translate into erroneous sample shifts. In combination with color edges, completely different color values could be used for interpolation and strong sample scattering and color bleeding is exhibited in the synthesized view. Therefore, the applied view synthesis algorithm was specifically designed to cope with such data.

3. VIEW SYNTHESIS APPROACH

For the creation of intermediate views at arbitrary positions, view synthesis with reliability-based processing is used, as described in [4]. Intermediate views are synthesized between any adjacent pair of original camera views by 3D projection in the most general case. Samples from original views are projected into the intermediate view, according to the intermediate position parameter \( \kappa \), described above. This parameter also controls the blending-weights between the samples of both original views.

In addition to this general approach, our reliability-based synthesis first detects unreliable areas along depth discontinuities. These areas are known to produce visual artifacts in the projection process and are therefore processed separately. Reliable areas are projected first, as described above. Then, the unreliable boundary areas are split into foreground and background data. Here, foreground areas are projected first and merged with the reliable data. Afterwards, the background data are projected and also merged. The important difference between foreground and background handling is the merging process: The foreground data is merged with the reliable data in a front most-sample-approach to preserve most of the important information. In contrast, background information is only used to fill remaining uncovered areas. Finally, different view enhancement algorithms are applied, including outlier removal, hole filling and natural edge smoothing. This synthesis approach is based on [6]. A more detailed description can be found in [4].

4. CODING RESULTS

First, individual coding results for color and depth data are analyzed in order to obtain initial bit rates and to investigate, which bit rate ratio between color and depth data is obtained at certain qualities. We used multi-view video coding (MVC extension of H.264/MPEG4-AVC) to separately compress color as well as depth data [7]. For this, we used the JMVM 7.0.1 reference software and a GOP-size of 16.

Fig. 1 shows the R-D results of individual color and depth coding, using 3 original views for each sequence.

These results show the typical PSNR curves for color and depth data. The important part here is the relation between color and depth data: For both sequences, Ballet and Breakdancers, the comparison between color and depth results highlights that with MVC coding, the depth data require relatively high bit rates for achieving the same PSNR values as the color data. A rough estimate gives an overall color-to-depth-ratio in terms of bit rate for same PSNR value of 3:1 for Breakdancers and even 1:1 for the Ballet sequence. These ratios are relatively low - the depth data for these sequences require a considerable portion of the overall bit rate. For other sequences, ratios of 10:1 are reported [8]. The bit rate distribution between color and depth however is highly dependent on the characteristics of the multi-view sequence. For sequences, like Ballet and Breakdancers, depth structures are very detailed and thus require more bit rate. However, the quality of decoded depth data in terms of PSNR has only limited practical meaning, because depth is never displayed. Only the influence on the quality of the synthesized views as evaluated in the next section is relevant.

5. VIEW SYNTHESIS RESULTS

Systems and algorithms for 3DV can only be evaluated by the quality of the synthesized views, which is being investigated below. These views are influenced by the individual color and depth coding results.

5.1. Combined Coding and Intermediate View Synthesis

After analyzing the coding behavior for color and depth data separately, the step of intermediate view synthesis is added. These intermediate views are taken for the overall judgment of the 3DV-system with all interconnected components. For the intermediate views, no original reference is available for objective PSNR measures. Hence a reference signal is synthesized from uncompressed color and depth data. Since the original data may already contain errors, the finally obtained objective PSNR values have to be verified by subjective evaluations.

For the objective quality evaluation, reference views at intermediate positions in the middle between each two original views were synthesized first, using the algorithm described in section 3 with view parameter \( \kappa = \frac{1}{2} \). Then, color and depth data with different coding qualities were combined for intermediate view synthesis. To analyze the overall behavior of combined coding and view synthesis, a full set of color-depth-pairs with 4 different coding qualities were recorded. The results for the Ballet sequence are shown in Fig. 2.

Here, the grid is shown, which is formed by the 4 curves with constant color fidelity and varying depth fidelity (solid line) and 4 curves with constant depth fidelity and varying color fidelity (dotted lines). The numbers in Fig. 2 indicate the quantization
parameter (QP) setting, used in the coder. A low QP value results in a higher quality or PSNR value at the expense of a higher bit rate. Following the solid curves with constant color fidelity, an increase in depth fidelity only moderately increases the intermediate view quality. For the dashed curves with constant depth fidelity, the behavior can be categorized as follows: First, at low color qualities, the curves are steep. Here, a small increase in color quality increases the intermediate view quality significantly. Then, at high color qualities, the curves are rather plane and a further increase in color quality only slightly increases the intermediate view quality. This second behavior was also verified by initial MVD coding results in [3].

It can be seen that the envelope of the rate distortion functions in Fig. 2 for the Ballet sequence roughly connects pairs of color and depth fidelity with approximately equal quality for both.

For the Breakdancers sequence, the results are shown in Fig. 3. Here, the solid curves for constant color fidelity and varying depth fidelity as well as the curves for constant depth fidelity and varying color fidelity are similar to those for the Ballet sequence. The envelope also roughly connects points with approximately equal quality for color and depth.

![Fig. 3: Results of view synthesis for intermediate positions for Breakdancers.](image)

For our limited number of experiments it seems that a good setting is found for the MVD codec when color and depth are coded at approximately the same fidelity (measured as MSE value). We will refer to this bit allocation as “equal MSE bit allocation”.

### 5.2. Synthesized View Quality Characteristic

The final goal of the presented 3DV system is the provision of high-quality views for N-view displays. For this, the optimal coding, and combined coding and view synthesis results have been shown in the previous subsection for one intermediate viewing position at $\kappa = \frac{1}{2}$. For further investigation, the quality characteristics across the viewing range of a 3DV system are investigated. For this, intermediate views are generated at various positions between each pair of original views by varying the intermediate position parameter $\kappa$.

For each sequence, the equal MSE bit allocation at constant overall bit rate is compared against a different bit allocation, which either assigns a lower rate to color or depth signals. The results are shown in Fig. 4. Here, the dark red curves show a parabolic characteristic between each pair of original views. The intermediate views have lower PSNR values, since they are influenced by geometric distortions due to depth coding errors. In contrast, PSNR maxima occur at original positions, where only color coding artifacts influence quality.

![Fig. 4: Equal MSE bit allocation (red curve) and different bit allocations (blue curve) between color and depth (top: Ballet, bottom Breakdancers).](image)

The light blue curves in Fig. 4 show the result for the different bit allocations. For the Ballet sequence in Fig. 4 top, the bit rate for color was increased at the expense of depth bit rate. This leads to better quality at original camera positions, where no depth data are required. For intermediate positions, the quality is lower due to higher geometric distortion caused by lower depth quality. Since a good quality is required across the full 3DV viewing range, the dark red curve is preferred. For the Breakdancers sequence, the opposite was applied for obtaining the light blue curve. The bit rate for color was decreased and more bit rate was assigned to depth. As a consequence, the light blue curve in Fig. 4 bottom shows a relatively constant characteristic, but at a very low quality. Here, the low color quality dominates all views, such that the depth data has no significant influence. Therefore, also the curve showing the equal MSE bit allocation is preferred.

Overall, we have found that the efficient bit allocation to color and depth through assigning equal MSE to both as described above gives very good quality characteristics across the viewing range. A change in bit rate allocation in either direction leads to worse overall results in our experiments. For visual comparison, detailed results for equal MSE bit allocation and the different bit allocations are shown in Fig. 5 for both sequences. For the Ballet sequence, the face texturing is sharper in the non-optimal case (left, bottom) in Fig. 5; however it has more annoying boundary distortion and background artifacts, in comparison to the optimal case (left, top).

For the Breakdancers sequence, the face texturing is worse in the non-optimal case (right, bottom) in Fig. 5, which dominates already the overall quality, including boundary distortions caused by depth map coding.
5.3. Reliability-based Layer Processing
The applied view synthesis algorithm uses reliability-based layer processing, to reduce visual artifacts. Therefore, Fig. 6 shows details from coding results.

Here, intermediate views were synthesized from uncompressed as well as compressed data. To show the capability of our view synthesis, reliability-based layer projection was switched off for the results in Fig. 6, top. For uncompressed data, the performance was already proven in [4] and is shown in Fig. 6 for comparison.

The quality improvement from reliability-based layer projection is also achieved for compressed data. The results in Fig. 6 show the visual artifacts along the object boundaries (top) as well as the reduction of the artifacts with reliability-based processing (bottom).

Both compressed views show the typical blurring artifacts caused by color compression, as well as different levels of color scattering, caused by depth compression. In the layer-based projection, areas along depth discontinuities are treated separately. The edge detection for depth discontinuity analysis is robust enough to detect smooth edges in the coded depth data. Then, the integrated filtering operators partially remove the scattered points, visible in Fig. 6, top. Thus, the reliability-based approach from [4] considerably improves the quality of synthesized views in 3DV systems also for compressed data.

5. SUMMARY AND CONCLUSIONS
The work, presented in this paper, combines multi-view video + depth coding with reliability-based view synthesis. It was found that areas with concurrent depth and color edges are especially vulnerable to coding errors, which cause visible artifacts in the view synthesis. To investigate the interdependency between coding and view synthesis, results are shown for individual stages, as well as for the complete 3DV chain. First, results for the coding stages are evaluated separately in order to obtain bit rate ratios for color and depth data. Then, the intermediate view synthesis is added and overall results are shown. Here, the major finding indicated a bit allocation that yields equal MSE for color and depth data for best intermediate view synthesis. Finally, the view synthesis evaluation showed that the applied reliability-based approach also suppresses visual artifacts in compressed data and is therefore suitable for a 3DV system.

6. ACKNOWLEDGMENT
We would like to thank the Interactive Visual Media Group of Microsoft Research for providing the Ballet and Breakdancers data sets.

7. REFERENCES