ILLUMINATION CHANGE ROBUST, CODEC INDEPENDENT LOW BIT RATE CODING OF STEREO FROM SINGLEVIEW AERIAL VIDEO

Holger Meuel, Florian Kluger, Jörn Ostermann

Institut für Informationsverarbeitung
Gottfried Wilhelm Leibniz Universität Hannover, Germany
http://www.tnt.uni-hannover.de/~meuel/

ABSTRACT

Low bit rate transmission of HD video captured from UAVs is highly interesting. Assuming a planar surface, areas contained in the current frame but not in the previous frames (*New Area*) can be reconstructed using Global Motion Compensation (GMC). Aiming at stereo reconstruction from monocular video by using motion parallax, a second view of each image pixel has to be additionally transmitted. Whereas the bit rate can be considerably reduced compared to standardized video coding to about 1–2 Mbit/s, artifacts at the boundaries between new areas and GMC reconstructed areas may occur, *e. g.* due to illumination changes. We propose a gradient correction of the new areas to adjust the luminance. Furthermore, we utilize a general ROI coding framework to become independent of any encoder modifications. We achieve a subjectively higher video quality while saving 2 % BD-rate compared to a specifically adapted encoder by exploiting latest encoder optimizations of *x265*.

***Index Terms* —** Stereo from singleview video, general ROI coding, HEVC coding, HDTV low bit rate coding, luminance gradient correction filtering

1. INTRODUCTION

For aerial surveillance the *Pulse Code Modulation* (PCM) data rate of 622 Mbit/s for a color video sequence with full *High Definition Television* (HDTV) resolution (1920×1080) has to be significantly reduced for small bandwidth transmissions. Modern hybrid video coders like *High Efficiency Video Coding* (HEVC) [1] can compress such videos to about 5–12.5 Mbit/s at a reasonable image quality [2, 3]. But for small mobile platforms like *Unmanned Aerial Vehicles* (UAV), *e. g. Micro Air Vehicles* (MAV), with a very limited channel capacity of only a few Mbit/s, the bit rate has to be further reduced.

1.1. Related Work

One common solution is *Region of Interest* (ROI) -based video coding. Most ROI coding systems provide the best possible image quality only for predefined regions in an image and degrade the image quality of non-ROI areas. For instance, non-ROI areas of a frame could be blurred or coarsely quantized either in a preprocessing step prior to actual video encoding or within the video encoder itself [4, 5, 6]. In [7], a video coding system retaining subjectively high image quality over the entire image was presented. This system achieves very low bit rates of 0.8–2.5 Mbit/s for the transmission of full HDTV resolution aerial video sequences by reconstructing already known static parts of the image by means of *Global Motion Compensation* (GMC) [8]. Hence, no motion parallax can be observed for elevated objects since each background ground pixel is transmitted only once. This renders some further processing at the decoder impossible, *e. g.* to reconstruct a stereo video out of one monocular video sequence. In order to preserve depth information for stereo video reconstruction, it was proposed to additionally transmit a second view for each ground pixel in [9]. However, the latter approach did not address global illumination changes on the one hand (Figure 1), leading to disturbing artifacts in the reconstructed stereo video. On the other hand, extensive modifications were introduced in a HEVC encoder to externally control the different coding of ROI and non-ROI areas [10, 11].

Figure 1: Global illumination change in 2 seconds: upper half (car, lighter street) from frame 145, bottom half from frame 202 (magnifications from *350 m sequence* from TAVT data set [2, 12]).

This paper addresses both issues: First, in order to exploit latest encoder optimizations, we use the general coding framework from [13] and achieve additional coding gains by using a more optimized HEVC encoder. Second, we propose to consider illumination changes by the integration of a luminance correction filter in the reconstruction process of both views of the stereo video. By using local gradients, we conceal global illumination changes, resulting in a subjectively improved quality.

The remaining paper is organized as follows: Section 2 reviews the stereo ROI coding system from [9] which was used as a basis. In Section 3 we describe the combination of the coding system with the codec independent video coding framework from [13]. Our proposed illumination correction filter is described in Section 4. We present experimental results in Section 5 before Section 6 concludes the paper.

2. ROI-BASED STEREO VIDEO CODING SYSTEM

We decided to use the reference system from [9] as a basis since it is capable of generating a very low bit rate stereo video from a monocular aerial video while retaining subjectively high quality over the entire image which is unique compared to other ROI coding systems. We first review the basic ROI coding system and focus on the stereo extension in Section 2.2. If moving objects should additionally be considered, a thorough description of a highly accurate moving object detector can be found in [2].

2.1. ROI-based Coding System

The idea of data reduction with this system is to exploit the special characteristic of the *planar* landscape which is observed in aerial surveillance video. Assuming a planar landscape, one frame $n-1$ is projected into frame n by employing a projective transform using the homography $\mathbf{H}_n$. For the global motion estimation (first block in Figure 2), the homography $\mathbf{H}_n$ is determined using

978-1-5090-3313-3/16/$31.00

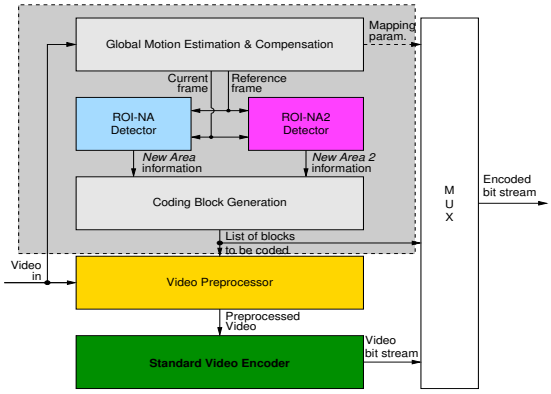
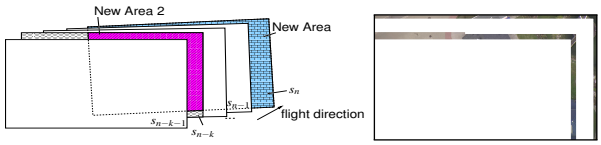


Figure 2: Block diagram of the *general ROI stereo coding system*. The ROI detection system (dark gray/dashed box) is the same as in [9] which is combined with the general ROI coding framework from [13], enabling the usage of off-the-shelf video encoders.



(a) Principle (b) Example (*350 m seq.*)

Figure 3: New Area Detection [9]

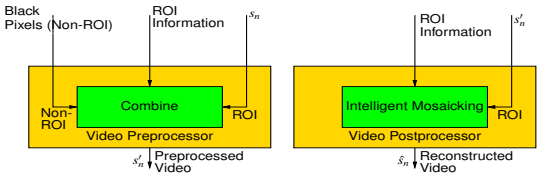
a *Kanade-Lucas-Tomasi* (KLT) feature tracker and *Random Sample Consensus* (RANSAC). This homography is used to determine the *New Area* (NA) in the current frame n by the *ROI-NA Detector* (Figure 2, blue block). This determined ROI is passed to the *Coding Block Generation* block which basically assigns the pel-wise ROI to corresponding blocks, *e. g.* of size 16 pels×16 pels. Whereas this block assignment was used for block-based video coding, it can theoretically be removed for general ROI coding (see Section 3) and the pel-wise ROI mask might be used instead. In order to leave the detection system as well as the reconstruction at the decoder unchanged, we also use the block-based classification of ROI and non-ROI. Finally, ROI and non-ROI have to be encoded differently, which is typically realized by an externally controlled, modified video encoder. We propose an alternative in Section 3.

To reconstruct the video from the transmitted *New Area*s, post-processing is necessary after the video decoding to align ROIs from the current frame within a reconstructed background panorama image from the previous frames [8, 9]. Based on the homography parameters, video frames can be cut out from the panorama and concatenated as a video sequence at positions corresponding to the view which was originally recorded on-board the UAV.

2.2. Stereo Panorama Images and Stereo Video Generation

For a real stereo representation, two views from different angles are needed for each ground object. Since no real second camera is feasible in a setup with MAVs, only one monocular video sequence is available. Thus, a second camera view has to be artificially generated out of the recorded video sequence by taking a second picture for the same ground area while the UAV has moved further. Similar to the *New Area* – further referred to as *New Area 1* (NA 1) – a second "New Area" (*New Area 2*, NA 2) is calculated and additionally transmitted for each frame (Figure 2, magenta block). The position of the *New Area 2* is calculated from the parallax that non-planar objects should have in the final video. According to [14], the resulting motion parallax p of overflown objects can be calculated from the displacement of the camera ΔC_x, the flight altitude C_z and the height h of the non-planar object as follows:

$$p = -\Delta C_x \frac{N_x}{s_x} \frac{f \cdot h}{(C_z - h)\, C_z}, \qquad (1)$$



(a) Preprocessing (b) Postprocessing

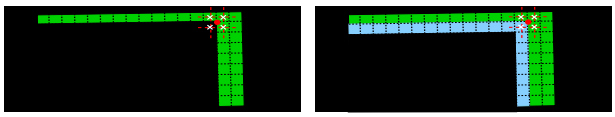
Figure 4: Image composition by pre- and postprocessing of general ROI coding. Subsequent processing has to be applied like required by the specific ROI coding system (*e. g.* GMC) (based on [9]).

wherein f is the focal length of the camera. N_x/s_x is a scaling factor and describes the size s_x of the camera sensor and the amount of pixels N_x it contains. Without loss of generality, a constant flight altitude, speed, and direction (x direction) is assumed – i. e. a straight flight path of the UAV – as well as a camera looking straight downwards (nadir view). This leads to a constant displacement Δx of the pixels on the ground plane, i. e. a constant translation in flight direction within the recorded vertical aerial video sequence. Given these assumptions, one object emerging in NA 1 will pass NA 2 k frames later. Experiments showed that a parallax of $-40 \leq p \leq 0$ pel gives a realistic impression of the height for aerial vertical videos, which corresponds to the recommendation given in [15]. The actual position of the *New Area 2* is derived by concatenating k homographies between the frames until the desired displacement Δx is reached. By concatenating homographies instead of predefining a constant frame offset, also aberrations of the camera (*e. g.* in y direction) can be correctly considered. Given the homographies between $k+1$ preceding frames, which are available from the global motion estimation, the projection of the current frame n into frame $n-k$ is computed: $\mathbf{H}_k = \prod_{i=n-k}^{n} \mathbf{H}_i$. Given a constant baseline distance of k frames, the virtual (second) camera is aligned based on the flight parameters with the recorded video sequence. As on decoder side this second view has to be available, areas emerging in the view-field of the second camera (*New Area 2*, Figure 3) have to be calculated and transmitted additionally. The calculation of *New Area 2* is based on the block raster of the current frame. First, new areas between frames $n-k$ and $n-k-1$ are calculated (Figure 3a, criss-crossed areas in frame $n-k$) and second, areas lying outside of the current frame n are subtracted finally leading to the *New Area 2* as depicted in Figure 3a (magenta). A real example is shown in Figure 3b. In order to generate two views for a stereo video sequence, two panorama images are reconstructed from NA 1 and NA 2, respectively. Each panorama is used for the generation of one view.

3. GENERAL ROI CODING

Whereas most ROI coding systems rely on externally controlled, modified video encoders, we propose to combine the above stereo ROI-detection system with the codec independent general ROI coding framework from [13]. Thereby, error-prone and time consuming encoder modifications can be avoided. Since an arbitrary, unmodified video encoder can be used, the change of the video codec becomes easy. Ongoing encoder optimizations can be used without any change in the system but only by replacing the video encoder. Thus, systems relying on the general ROI coding framework become more future-proof. As a side benefit, the standard compliance of the bit stream is provided by any standard video encoder. No efforts have to be made to enable external encoder control.

The general ROI coding framework supports two different operation modes: using the first one, non-ROI blocks are replaced by their corresponding regions from the preceding frame in a preprocessing step prior to the actual video coding. For the second one, non-ROI regions are similarly replaced by black pixels (Figure 4). Since only ROI areas are considered for background reconstruc-



(a) New area filtering without known pixels from panorama

(b) Proposed improved filtering using known areas from panorama

Figure 5: Proposed mosaicing: before current NA (green, (a), black dashed local coordinate system) is inserted in the panorama (red dashed world coord. system), neighboring pixels are transformed into the current NAs coordinate system (blue, (b)) to avoid filter artifacts at boundaries between NAs from different frames.

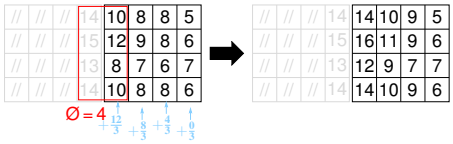


Figure 6: Example for luminance correction, integer rounded values after luminance correction (right).

tion in our ROI coding system and non-ROI areas are discarded, their content is irrelevant anyway. As the latter mode (mode 2) appears to be slightly more efficient for HDTV resolution aerial sequences in [13], we use mode 2. In the block diagram of the proposed system (Figure 2), we integrated the *Video Preprocessor* block and a *Standard Video Encoder*.

4. IMPROVED FILTERING

Our second contribution aims at the correction of luminance changes in the panorama images and consequently also in the reconstructed videos. Due to the operation principle of the ROI coding system, global illumination changes appear in the case that a ROI is inserted into the panorama image next to global motion compensated content (non-ROI) and the global illumination has changed, *e. g.* due to the change from sunny to cloudy sky. Our proposed filtering reduces filter artifacts introduced by non-optimal filtering of the new areas during the mosaicing process, *e. g.* an incorrect luminance may occur at ROI boundaries (Figure 7b). The trivial approach to add a new area to the panorama image would be to apply the global motion compensation to the current new area. Since only the new areas from one frame can be used for background generation by the *Postprocessor* (Figure 4b), no neighboring pixels are available for proper filtering, *e. g.* using a bilinear filter (Figure 5a). Since the coordinate system of the current new area c (dashed black) is not coherent with the "world" coordinate system C of the panorama image (dashed red), filter artifacts will necessarily occur. Typically, mirroring or replication of the boundary pixels is applied, which leads to imperfect results.

On the other hand, neighboring pixels are available for already known content from preceding frames, i. e. for the areas below and left of the NA in Figure 5a. Thus, we propose to utilize this information to avoid filter artifacts by transforming neighboring pixels from the panorama image to the current new areas coordinate system first, using replication of the boundary pixels (Figure 5). In the second step, we apply the filtering as described above to any ROI block in any direction, i. e. for the new area in the example. In contrast to the trivial approach, with our proposal real pixel information is used for filtering (Figure 5b, blue). The boundary problem of course occurs also during the transformation of pixels from C to the current new areas coordinate system c, but can be neglected, since the affected pixels are only intermediately used and discarded after the final filtering.

In order to get rid of illumination changes, for each block we compute the mean gradient of the luminance (Y channel in YCbCr representation) across the boundaries between current ROI and their adjacent pixels in the panoramic image, considering every horizontal and vertical neighboring pixels. This gradient is spread over the entire block width or height, respectively, by a linearly changing luminance: The luminance of each pixel of the current ROI block is corrected by $\frac{\text{averaged gradient}}{\text{blockwidth}-1}$. Thus, the luminance of adjacent pixels is adjusted whereas the pixels of the opposite column or row of the block remain untouched. As an example assume a mean gradient of 4 (*e. g.* caused by global illumination changes) between the new area and its left-side neighboring pixels in a 4×4 block (Figure 6). Then this gradient is distributed equally over all pixels in the block in horizontal direction, resulting in luminance corrections for each column as indicated with blue numbers in the figure. The same is applied in vertical direction, if applicable. In the special case that only a diagonal but no horizontal or vertical non-ROI neighboring block exists, the luminance correction is applied similarly but only with half of the gradient in each direction – which turned out to be subjectively the best. In order to compensate luminance changes between adjacent ROI blocks, an additional median filtering is applied over the gradients of several neighbored ROI blocks. While other illumination correction algorithms exploit global image characteristics, which are not available since only new areas are present for every frame, our method relies on local filtering of new areas and neighboring pixels from the panoramic image only.

5. EXPERIMENTS

For the evaluation of our proposed approach we used the *TNT Aerial Video Testset* (TAVT) [2, 12] containing four high resolution video sequences (full HDTV resolution, 30 fps), each between 821 and 1571 frames long with different image characteristics. We used the modified *x265* (v1.4) [16] HEVC video encoder from [9] as reference and compared the coding efficiency with our proposed general ROI coding framework, using the latest unmodified *x265* v1.9 (*x265*, preset *placebo* – representing the most efficient coding settings of *x265* – and PSNR tuning). In order to realize a convenient stereo impression we also used the proposed baseline distances according to [9] which are listed in Table 1. To evaluate the objective quality of the reconstructed videos (both views), we only considered luminance values (Y component in YCbCr video format) within ROI areas (*e. g.* similar to [17, 2]), assuming errors in non-ROI areas, *e. g.* introduced by the GMC due to parallax, to be irrelevant as the background is reconstructed from the panorama images anyway.

Bjøntegaard deltas [18] (BD-rate, piecewise cubic interpolation, QP range: 10–50, 9 rate points) are presented in Table 1. Negative BD-rates represent coding gains compared to the modified *x265* from [9]. From the results we see that the unmodified video encoder can slightly outperform the specifically adapted encoder by 1.97 % BD-rate on average. This result can be explained by encoder optimizations in the meantime. Compared to a conventional HEVC encoding, we achieve a bit rate saving of about 85 %, corresponding to a total bit rate of 1–2 Mbit/s for a subjectively good video quality of 38–41 dB (PSNR). In contrast to other ROI coding approaches and independent of the encoder used (modified or unmodified), we preserve a subjectively very high quality over the entire image and thus over the entire stereo video.

Subjective results of our improved filtering are presented in Figure 7 for challenging scenarios: The first example (a–c) shows the superior filtering for homogeneous areas: Whereas adjacent new areas can clearly be distinguished without our improved filtering in (b), our result (c) is similar to the conventionally encoded result but at only about one fifth of the bit rate. The second example (d–f) is much more challenging, since the global illumination changes are very high between the background generation

(a) Conventional coding(b) W/o improved filt. (c) W/ improved filt. (d) Conventional cod. (e) W/o improved filt. (f) W/ improved filt.

Figure 7: Results before (b,e) and after (c,f) the proposed luminance correction filtering for challenging scenarios. In addition, the conventionally coded magnifications are presented (a,d). We would like to emphasize that the bit rate of the conventional coding is about 5000 kbit/s instead of 1000 kbit/s with the proposed general ROI coding at a similar Y-ROI-PSNR of about 39 dB (*350 m sequence* [2, 12]).

Table 1: Frame offsets (from [9]) and resulting Bjøntegaard delta (BD-rate) [18] of general ROI coding using an unmodified *x265* (proposed) vs. a modified *x265* encoder from [9] (stereo ROI, HEVC, negative BD-rates represent coding gains, Y-ROI-PSNR, *x265* settings: preset *placebo* and *--tune psnr*).

Sequence	**frame offset** k	**BD-rate (in %)**
350 m sequence	10	−3.37
500 m sequence	15	−0.50
1000 m sequence	20	−1.67
1500 m sequence	30	−2.34
Mean		−1.97

and the emerging moving person. Although our luminance correction filter cannot entirely conceal the luminance differences, informal subjective tests certify a highly improved image quality, especially in the reconstructed video. By applying our luminance correction algorithm during the generation of both panorama images, the subjective image quality is also increased for stereo vision.

6. CONCLUSION

In this paper we use a ROI-based coding system for UAVs for the generation of stereo ("3D") aerial video sequences from monocular video. Our two contributions are: First, we combined the stereo ROI coding system with a codec independent general ROI coding framework. Exploiting newly emerged encoder optimizations, we outperform a specifically adapted video encoder by ~2 % BD-rate on average. By using an off-the-shelf video encoder, the usage of new or optimized video encoders is facilitated and the coding system becomes future-proof. Our second contribution is an online illumination change correction filtering during the (GMC-based) background reconstruction at decoder side, which highly improves the subjective quality without impacting the encoder – and thus the transmission bit rate. We achieve total bit rates of about 1–2 Mbit/s with an unmodified x265 software encoder for full HDTV resolution (30 Hz) stereo aerial video sequences at 38–41 dB.

7. REFERENCES

[1] HEVC, "ITU-T Rec. H.265/ ISO/IEC 23008-2:2013 MPEG-H Part 2/: High Eff. Video Coding (HEVC)," 2013.

[2] Holger Meuel, Marco Munderloh, Matthias Reso, and Jörn Ostermann, "Mesh-based Piecewise Planar Motion Compensation and Optical Flow Clustering for ROI Coding," in *APSIPA Transact. on Sig. and Inform. Proc.*, 2015, vol. 4.

[3] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards - Including High Efficiency Video Coding (HEVC)," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1669 –1684, Dec. 2012.

[4] Mei-Juan Chen, Ming-Chieh Chi, Ching-Ting Hsu, and Jeng-Wei Chen, "ROI Video Coding Based on H.263+ with Robust Skin-Color Detection Technique," *IEEE Trans. on Consumer Electr.*, vol. 49, no. 3, pp. 724–730, Aug 2003.

[5] N. Doulamis, A. Doulamis, D. Kalogeras, and S. Kollias, "Low Bit-Rate Coding of Image Sequences using Adaptive Regions of Interest," *IEEE Trans. on Circ. and Systems for Video Technol.*, vol. 8, no. 8, pp. 928 –934, Dec 1998.

[6] Linda Karlsson, Mårten Sjöström, and Roger Olsson, "Spatio-Temporal Filter for ROI Video Coding," in *Proc. of the 14th Europ. Sig. Proc. Conf. (EUSIPCO)*, Sept. 2006.

[7] Holger Meuel, Marco Munderloh, and Jörn Ostermann, "Low Bit Rate ROI Based Video Coding for HDTV Aerial Surveillance Video Sequences," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition - Workshops (CVPRW)*, June 2011, pp. 13 –20.

[8] Holger Meuel, Julia Schmidt, Marco Munderloh, and Jörn Ostermann, *Adv. Vid. Cod. for Next-Generation Multimed. Services – Chpt. 3: ROI Coding for Aerial Video Seq. Using Landscape Models*, Intech, Jan. 2013.

[9] Holger Meuel, Marco Munderloh, and Jörn Ostermann, "Stereo Mosaicking and 3D-Video for Singleview HDTV Aerial Sequences using a Low Bit Rate ROI Ccoding Framework," in *IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS), 2015*, Aug 2015, pp. 1–6.

[10] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "Region-of-Interest Based Rate Control Scheme for High Efficiency Video Coding," in *IEEE Int. Conf. on Acoustics, Speech &Signal Proc. (ICASSP)*, May 2014, pp. 7338–7342.

[11] Peiyin Xing, Yonghong Tian, Tiejun Huang, and Wen Gao, "Surveillance Video Coding with Quadtree Partition based ROI Extraction," in *Proceedings of the IEEE Picture Coding Symposium (PCS)*, Dec 2013, pp. 157–160.

[12] Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover, "TNT Aerial Video Testset (TAVT)," 2010–2014, https://www.tnt.uni-hannover.de/project/TNT_Aerial_Video_Testset/.

[13] Holger Meuel, Marco Munderloh, Florian Kluger, and Jörn Ostermann, "Codec Independent Region of Interest Video Coding using a Joint Pre- and Postprocessing Framework," in *Int. Conf. on Multim. & Expo (ICME)*, July 2016.

[14] Marco Munderloh, *Detection of Moving Objects for Aerial Surveillance of Arbitrary Terrain*, vol. 10 of *Fortschritt-Berichte*, VDI Verlag, Mar. 2016.

[15] Wa James Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: Visual Comfort," *IEEE Trans. on Broadcast.*, vol. 57, no. 2, pp. 335–346, June 2011.

[16] VideoLAN Organization, "x265," 2014–2016, v1.4–v1.9.

[17] D. Grois and O. Hadar, "Complexity-Aware Adaptive Spatial Pre-Processing for ROI Scalable Video Coding with Dynamic Transition Region," in *Proc. of the 18th IEEE Int. Conf. on Image Proc. (ICIP)*, Sept. 2011, pp. 741–744.

[18] Gisle Bjøntegaard, "AI11: Improvements of the BD-PSNR model. ITU-T Study Group 16 Question 6. 35th Meeting," in *ITU-T SG16 Q*, Berlin, Germany, 2008.