Stereo Mosaicking and 3D-Video for Singleview HDTV Aerial Sequences using a Low Bit Rate ROI Coding Framework

Holger Meuel, Marco Munderloh, Jörn Ostermann
Institut für Informationsverarbeitung
Gottfried Wilhelm Leibniz Universität Hannover,
Germany
http://www.tnt.uni-hannover.de/~meuel/

Abstract

Low bit rate coding systems for the transmission of high quality aerial surveillance videos captured from UAVs are of high interest. One way to achieve high quality low bit rate video is to assume a planar surface of the earth, which is valid for sequences captured at high flight altitudes. Those systems only transmit the area of the current frame not contained in the previous frames (New Area) and reconstruct the already known areas by means of Global Motion Compensation (GMC) at decoder side. Although the bit rate can be reduced significantly compared to standardized video coders, no reconstruction of stereo video is possible at the decoder since each image pixel is transmitted only once and thus no motion parallax of objects can be observed in the reconstructed video. In this paper we present a coding system for stereo video reconstruction at very low bit rates. Onboard the UAV we employ the camera path estimated from the image data to create a second view of a virtual camera. We derive convenient baseline distances and demonstrate the resulting perceptively good stereo impression for different test sequences. Similar to the coding concept introduced above we transmit a second New Area 2 *in addition to the* New Area *already introduced. By doubling the bit rate to about 2 Mbit/s for a reasonable video quality of more than 38 dB, still saving more than 85 % BD-rate compared to common HEVC coding, we are able to reconstruct a full HDTV (30 fps) stereo video at the decoder.*

1. Introduction

In aerial surveillance, two constraints are of critical importance. On the one hand, the video bit rate has to be significantly reduced, taking into account the uncompressed bit rate of an 8 bit *Pulse Code Modulation* (PCM) encoded color video sequence with full *High Definition Television* (HDTV) resolution (1920×1080) of 622 Mbit/s. On the other hand, it is desirable to provide the best possible image quality as well as additional video evaluation methods (*e.g.* moving object detection) and visualization (*e.g.* stereo video, panorama images) [4, 6]. Especially stereo video, often also referred to as "3D" video in cinema and television, can provide additional information in several application scenarios like different height levels in disaster area monitoring.

978-1-4673-7632-7/15/$31.00 © 2015

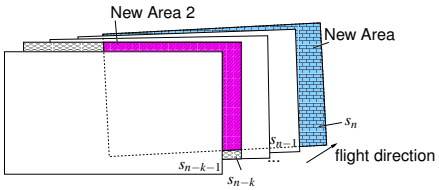


Figure 1. New Area Detection

1.1. Related Work

Modern hybrid video coders like *High Efficiency Video Coding* (HEVC) [11] are able to compress the huge amount of video data to bit rates of about 5–12.5 Mbit/s (HEVC), respectively, at a reasonable image quality [15, 18]. 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. 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 in favor of a reduced bit rate. 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 [3, 5, 13]. In [14], 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. However, these low bit rates are achieved at the cost of a lack of motion parallax for elevated objects. Basically only new emerging areas are transmitted for each frame (Figure 1, blue *New Area*)

whereas the remaining frame is reconstructed by means of global motion compensation at decoder side [16]. Common for the ROI coding techniques is that some important information contained in the video is lost during coding, *e.g.* parts of the image are only available in degraded quality at the decoder side or the changes in perspective – and thereby the information of the depth of the scene – are lost. This makes some further processing at the decoder impossible, *e.g.* to reconstruct a stereo video out of one monocular video sequence, which is possible in the case that entire images would have been encoded and transmitted in high quality.

This paper proposes to extend the GMC approach by transmitting additional areas out of the video sequence to the decoder in order to provide a second view of the scene while simultaneously holding the overall transmission rate small compared to competing approaches not using specially adopted video data reduction techniques. As the changes in perspective are preserved and the path of the camera is known, a real stereo video can be created from only one monocular video sequence recorded on-board the UAV. As a side benefit we also get a stereo panorama image for free without any additional computational cost which might be useful for a general overview of the observed landscape. Whereas the mosaicking process for stereo panorama image and stereo video generation itself is comparable to other monocular mosaic approaches like [4, 14, 16], our proposed system is capable of considerably reducing the bit rate and thus can transmit the video data over small bandwidth transmission channels. Hence, our system can provide a real-time view of the observed area and does not depend on the return of the UAV and a subsequent processing which might be highly interesting for surveillance missions. Moreover, in such a system, moving objects can be detected easily by a human observer due to a displacement not matching the static scene geometry. Since the flight path as well as the additional area needed for the second view are processed on-board the UAV and estimated from the video alone, no additional GPS/INS data is necessary. The remaining paper is organized as follows: Section 2 reviews the ROI coding system for aerial sequences from [14] which was used as a basis. We introduce the generation of the second view on-board the UAV in order to provide an accurate stereo video at the decoder. In Section 3 experimental results are presented and discussed for real video sequences recorded from different flight altitudes between 350 and 1500 m containing varying characteristics in terms of ground details. Section 4 finally concludes the paper.

2. ROI based Stereo Mosaic Coding System

We decided to use the reference coding system from [14] as a basis and extend it for stereo capability because

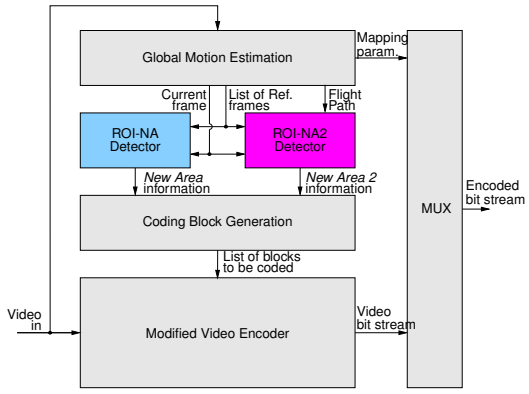


Figure 2. Block diagram of GME/GMC-based ROI coding system. Gray: unmodified (GME/GMC, external controlled video encoder), Blue: *ROI New Area* detector as in reference system, magenta: new proposed *ROI New Area 2* detector to allow for stereo video generation (based on [14]).

it is capable of retaining subjectively high quality over the entire image while simultaneously providing very low bit rates which is unique compared to other ROI coding systems. For a better understanding, we first shortly review this coding system with a focus on the detection, coding and transmission on encoder side (on-board the UAV) as well as the reconstruction at the decoder side. If moving objects should additionally considered, a thorough description of a highly accurate moving object detector can be found in [15]. The remaining section focuses on our extension for the integration of stereo video. The final block diagram is shown in Figure 2. It is worth noting that only the ROI-NA detector (blue) belongs to the original system (Subsection 2.1), whereas the ROI-NA 2 detector (magenta) is newly introduced in order to allow for stereo mosaicking and stereo ("3D") video generation (Subsection 2.2).

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 to be the main item in the scene, one frame $n-1$ is projected into the consecutive frame n by employing a projective transform using 8 parameters $\vec{a}_n = (a_{1,n}, a_{2,n}, \ldots, a_{8,n})^\top$. Hereby, the pixel coordinates from the preceding frame $\vec{p}_{n-1} = (x_{n-1}, y_{n-1})^\top$ are mapped to the position $\vec{p}_n = (x_n, y_n)^\top$ of the current one. Using homogeneous coordinates, this can be expressed in matrix notation:

$$\vec{p}_n = \mathbf{H}_n \vec{p}_{n-1}, \text{ with } \mathbf{H}_n = \begin{bmatrix} a_{1,n} & a_{2,n} & a_{3,n} \\ a_{4,n} & a_{5,n} & a_{6,n} \\ a_{7,n} & a_{8,n} & 1 \end{bmatrix}. \quad (1)$$

To determine $\vec{a}_n$, first, a global motion estimation is performed. For this purpose *Harris Corners* [10] are used to define a set of good-to-track feature points in the frame n.

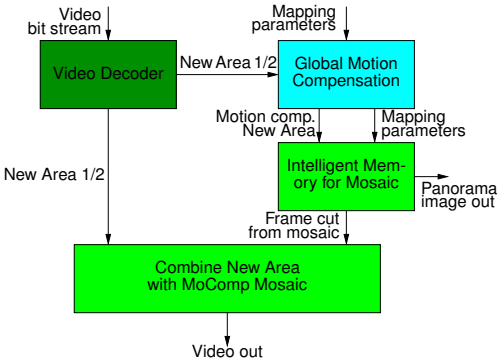


Figure 3. ROI decoder for stereo mosaicking and stereo video reconstruction (based on [14])

A *Kanade-Lucas-Tomasi* (KLT) [19] feature tracker is employed afterwards to relocate the feature positions in frame $n-1$ and thereby generate a sparse optical flow between the frames. Outliers such as false tracks are removed and the final mapping parameter set $\vec{a}_n$ is determined by *Random Sample Consensus* (RANSAC) [7]. This mapping parameter set is used for the *Global Motion Compensation* (GMC) as the first block in the block diagram of the coding system (Figure 2) by employing Equation (1). The mapping parameter set $\vec{a}_n$ is further employed to determine the *New Area* (NA) in the current frame n by the *ROI-NA Detector*. These determined ROI is passed to the *Coding Block Generation* block which basically assigns the pel-wise ROI to corresponding blocks for video coding. Any square block (*e.g.* a *Coding Tree Unit* in HEVC) containing at least one pel *New Area* is encoded as usual whereas any non-ROI block is forced to *skip mode* by an externally controlled modified video encoder, *e.g.* a modified HEVC encoder. By this external video coder control, the data rate is noticeable reduced compared to a not externally controlled video encoder while standard compliance of the bit stream is retained. The mapping parameter set $\vec{a}_n$ has to be transmitted in the data stream as well which could be realized by encapsulating the 8 parameters per frame in *Supplemental Enhancement Information* (SEI) messages.

To reconstruct the video from the transmitted *New Area*s, postprocessing is necessary after decoding of the bit stream to align ROIs from the current frame within the reconstructed background from the previous frames [16]. This *ROI decoder* is shown in Figure 3 (based on [14]). The *New Area* as well as the homography parameters are hereby fed into a *Global Motion Compensation* block and are mapped into a panorama image in the *Intelligent Memory for Mosaicking* block. "Intelligent" basically means the memory allocation scheme which allows streaming capability by dynamically allocating, releasing and realigning the amount of kept image content. Based on the homography parameters, video frames can be cut out from the mosaicked panorama image and concatenated as a video sequence at positions corresponding to the view which was originally recorded

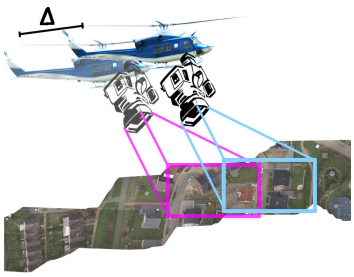


Figure 4. Principle of stereo ("3D") mosaic creation

on-board the UAV. We like to emphasize that any frame from the panorama image is generated just by *New Area* information. Due to the operating principle of the system, each point on the surface is only transmitted once, *i.e.* at the time when it is part of the *New Area*. Therefore, the generation of stereo views is not possible.

2.2. Stereo Mosaicking, Panorama Images Generation and Stereo Video Generation

To overcome the drawback of only having a single view for every part of the recorded scene, we suggest to integrate a stereo extension into the coding system.

Basically, for a real stereo representation, two views from different angles are needed for each ground object. Since no real second camera to generate the second view is feasible in a setup with small and medium UAVs, only one monocular video sequence is available. Thus, a second camera view has to be generated artificially out of the recorded video sequence by taking a second picture for the same ground area while the UAV has moved further (Figure 4). Whereas in the reference system only *New Area* – further referred to as *New Area 1* (NA 1) – is transmitted to keep the coding data rate as low as possible, we suggest to calculate and transmit an additional area for the second view – we call it *New Area 2* (NA 2) – within each frame.

The position of the *New Area 2* is calculated from the parallax that non-planar objects should achieve in the final video. According to [17], 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:

$$p = -\Delta C_x \frac{N_x}{s_x} \frac{f \cdot h}{(C_z - h) C_z}, \tag{2}$$

wherein f is the focal length of the camera. $\frac{N_x}{s_x}$ is a scaling factor and describes the size s_x of the camera sensor and the amount of pixel N_x it contains. We assume the x axis as the direction of flight. Then ΔC_x can be calculated from the speed of the aircraft divided by the number of frames recorded per second. It is applied to convert the unit of measurement into pel. By measuring the displacement of the camera as the displacement of the pixels of the ground

Figure 5. Two views from different time instances are necessary to generate stereo panoramas which serve as a basis for stereo video. The upper one is recorded, the lower one additionally needed.

plane Δx in pel between the frames, this can be simplified to:

$$p = -\Delta x \frac{h}{(C_z - h)}. \tag{3}$$

Moreover, this representation makes the solution independent of the speed of the aerial system, the camera lens or the sensor characteristics. Solving it for Δx retrieves the camera displacement in pel that has to be performed to achieve a given parallax p, whereby negative values of p result in objects sticking out of the screen.

As a simplification, and without loss of generality, we assume a constant flight altitude, speed, and direction (x direction) – *i.e.* a straight flight path of the UAV – as well as a camera looking straight downwards (nadir view). Thus, we get 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 several frames k later.

Experiments showed that a parallax of $-40 \leq p \leq 0$ pel gives a realistic impression of the height for aerial vertical videos. This corresponds to the recommendation given in [20]. Assuming a maximum object height of $h = 40$ m and a flight altitude of $C_z = 350$ m, Δx is approximated to 310 pel in this case. This camera displacement Δx is kept constant for height variations of the current flight path in order to create a consistent depth perception.

The actual position of the *New Area 2* is calculated by concatenating k homographies between the frames until the desired displacement Δx is reached. Given the homographies between $k+1$ preceding frames, which are available from the global motion estimation process, we calculate the projection of the current frame n into frame $n-k$:

$$\mathbf{H_k} = \prod_{i=n-k}^{n} \mathbf{H_i} \ . \tag{4}$$

Similarly we compute the projection $\mathbf{H_{k-1}}$ of the current frame n into frame $n-k-1$ for a running index $i = n-k-1, \ldots, n$. Given our constant baseline distance of k frames, we can align a virtual (second) camera (Figure 4, magenta) 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 1) have to be calculated and transmitted additionally.

Table 1. Parallax and subjective optimal baseline distances.

Sequence	*350 m*	*500 m*	*1000 m*	*1500 m*
max. parallax p	40	28	14	9
Frame offset k	10	15	20	30

Since the video encoding is block based, our calculation of *New Area 2* is based on the block raster of the current frame whereas all corners of the block are checked, if they belong to the current *New Area 2*. It is checked, if a corner pixel projected from the current frame n to the corresponding position (x_p, y_p) in frame $n-k$ is within frame $n-k$, but not in frame $n-k-1$. In other words, first, new areas between frames $n-k$ and $n-k-1$ are calculated (Figure 1, criss-crossed areas in frame $n-k$) and second, areas lying outside the current frame n are subtracted finally leading to the *New Area 2* as depicted in Figure 1 (magenta).

In order to generate two views for a stereo video sequence, we build two panorama images. Each of these is created by corresponding new area only, *i.e.* one panorama image is generated out of NA 1 only whereas the other one (for the virtual camera) is stitched from NA 2 only. Therefore, the ROI decoder (Figure 3) is applied twice, one time for NA 1 and one time for NA 2, resulting in two independent panorama images (Figure 5). For the mosaicking itself, each corresponding new area is registered into a global coordinate system [14]. A streaming capable mosaicking implementation can be found in *e.g.* [16]. Since at encoder side the NA 2 was calculated based on the flight path of the UAV, a constant baseline distance can be guaranteed for every pixel. However, if the UAV turned or just rolled, no corresponding NA 2 might be available for certain areas of NA 1 with the correct baseline distance and consequently, no stereo information will be available neither in the panorama image generated by NA 2 nor in the final stereo video sequence. These areas will appear without content in the second view.

Each view for the stereo video is generated by concatenating frames cut out of the corresponding mosaic, based on the flight path. In consequence the first view is generated from the first mosaic only whereas the second view is derived from the second mosaic only. The entire first frame of each view is inserted into the corresponding panorama image. This means that for the first mosaic the first frame of the recorded video sequence and for the second view the k-th frame of the video sequence are fully encoded (equal to "no ROI coding").

As a side benefit, local moving objects not matching the global motion of either of the views, will lead to an anomaly in the stereo video sequence and thus are easily to detect for a human observer.

3. Experiments

For the evaluation of our proposed approach we use the *TNT Aerial Video Testset* (TAVT) containing four high

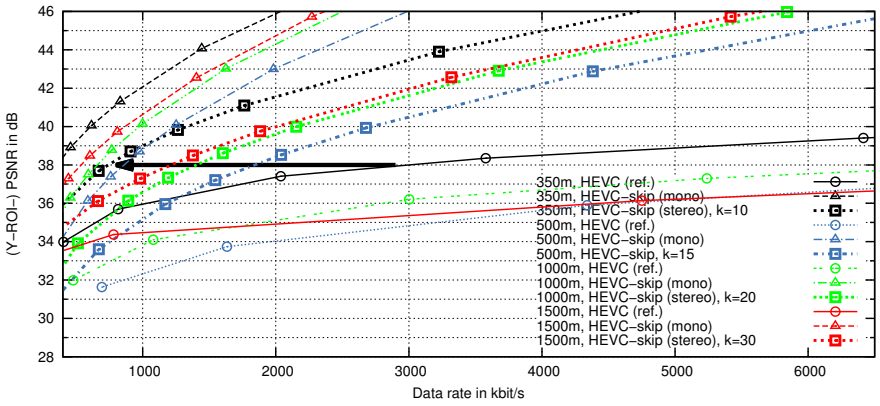
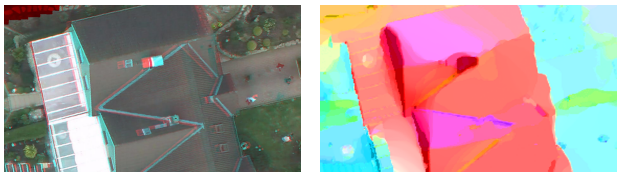


Figure 7. Rate-Distortion (RD) plot of stereo ROI ("3D") vs. mono ROI ("2D") coding efficiency, also compared to common HEVC.



(a) Anaglyph red-cyan (b) Depth map

Figure 6. Reconstructed frame (magnification) from *350 m Sequence*: (a) anaglyph red-cyan: red/cyan mark distinct 3D structures (*e.g.* roofs) or indicate disparity "errors" at moving objects (people right). For dark red areas (top-left), no 2nd view was present (Best viewed with red-cyan glasses.) (b) Depth map (generated after stereo ROI processing with algorithm of [22]).

resolution video sequences (full HDTV resolution, 30 fps), each between 821 and 1571 frames long with different image characteristics regarding ground resolution (43 pel/m–10 pel/m), noise and content (*e.g.* suburban houses, pitch, parking cars) [12]. We used a modified *x265* [21] HEVC video encoder and compared the common "single-view" coding with the bit rate of the proposed coding scheme. As described in Subsection 2.2, a baseline according to the distance of the human eyes leads to a negligible subjective stereo perception due to the distance between the camera and the ground. Thus, we tested different baseline distances corresponding to a frame offset of $k = 10 \ldots 30$ frames between both views. Subjective tests resulted in optimal baseline distances depending on the flight speed and height of the UAV according to Table 1. The processed video sequences are presented at [12].

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 [8, 9, 15]), 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 mosaics anyway.

Coding results are presented in a *Rate-Distortion* (RD) diagram in Figure 7. Bjøntegaard deltas [1, 2] (BD-rate, piecewise cubic interpolation, QP range: 10–38, 9 rate points) are provided in Table 2. Negative BD-rates represent

Table 2. Bjøntegaard delta (BD-rate) [1, 2], negative BD-rates represent coding gains (Y-ROI-PSNR).

Sequence	stereo ROI (proposed) vs. common HEVC BD-rate (in %)	stereo ROI (proposed) vs. mono ROI BD-rate (in %)
350 m Seq.	-84.08	120.40
500 m Seq.	-84.52	102.48
1000 m Seq.	-84.24	108.77
1500 m Seq.	-88.39	124.91
Mean	**-85.31**	**114.14**

coding gains compared to common HEVC encoding (*x265*, [21]), positive BD-rates represent additional bit rate consumption by the transmission of the *New Area 2*. It is obvious that the transmission of both stereo views consumes more bit rate compared to the transmission of only one single view ("mono"). Considering that *New Area 2* is not aligned with the coding block boundaries, it consumes 14 % more bit rate on average than *New Area 1*. Thus, in total the bit rate is increased by 114 % on average compared to the single view case. However, compared to a common HEVC encoder we can still 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. In contrast to other ROI coding approaches, also a subjectively very high quality is preserved over the entire image and thus over the entire stereo video sequence. A decoded frame from the reconstructed stereo video is shown in Figure 6 in anaglyph red-cyan coloration. Cyan represents the first view whereas red color indicates parallax of the second view. Dark red areas indicate areas where no second view was recorded by the UAV with the desired baseline distance.

Moving objects, which might be of high interest in an aerial surveillance system, can be easily identified by a human observer as disparity "errors" in the stereoscopic video (Figure 6(a)) as well as in a depth map (Figure 6(b)), generated out of the stereo video. By using the unmodified algorithm of Zach et al. [22] for the generation of the depth map, we demonstrate that our derived stereo views are suitable for subsequent computer vision tasks.

4. Conclusion

In this paper we propose a ROI based coding system for UAVs for the generation of stereo ("3D") aerial video sequences. By exploiting the planar characteristic of such sequences, which holds true also for test sets recorded in hilly terrain, only new emerging image content is transmitted for each frame. Based on the flight path – which is estimated from image data without external sensors like GPS/INS – corresponding *New Areas 2* are computed and additionally transmitted as a second (virtual camera) view, serving as another perspective from a different time instance for each ground object. We demonstrate that our approach has no negative impact on subsequent computer vision tasks like depth map generation. A convenient baseline distance was derived based on the maximally allowed parallax and experimentally validated. In order to enable stereo perception, the baseline distance was adjusted accordingly. As a side-benefit two panorama images generated from different views according to the baseline distance are created within the decoding process and thus are available without additional computational cost. Compared to a similar single-view ("mono") video compression, an additional bit rate of 114 % is consumed whereas the overall bit rate compared to common HEVC encoding can still be reduced by more than 85 %, leading to total bit rates of about 1–2 $\frac{\text{Mbit}}{\text{s}}$ for the transmission of full HDTV resolution (1920×1080, 30 Hz) stereo aerial video sequences at a reasonable subjective quality of 38–41 dB.

References

[1] G. Bjøntegaard. Calculation of Average PSNR Differences between RD Curves. In *ITU-T SG16/Q6 Output Document VCEG-M33*, Austin, Texas, April 2001.

[2] G. 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.

[3] M.-J. Chen, M.-C. Chi, C.-T. Hsu, and J.-W. Chen. ROI Video Coding Based on H.263+ with Robust Skin-Color Detection Technique. *IEEE Transactions on Consumer Electronics*, 49(3):724–730, Aug 2003.

[4] N. Dong, X. Ren, M. Sun, C. Jiang, and H. Zheng. Fast stereo aerial image construction and measurement for emergency rescue. In *Geo-Information Technologies for Natural Disaster Management (GiT4NDM), 2013 Fifth International Conference on*, pages 119–123, Oct 2013.

[5] N. Doulamis, A. Doulamis, D. Kalogeras, and S. Kollias. Low Bit-Rate Coding of Image Sequences using Adaptive Regions of Interest. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(8):928 –934, Dec 1998.

[6] F. Dufaux and F. Moscheni. Background Mosaicking for Low Bit Rate Video Coding. In *Image Processing, 1996. Proceedings., International Conference on*, volume 1, pages 673–676 vol.1, Sep 1996.

[7] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.

[8] P. Gorur and B. Amrutur. Skip Decision and Reference Frame Selection for Low-Complexity H.264/AVC Surveillance Video Coding. *IEEE Transact. on Circuits and Systems for Video Technology*, 24(7):1156–1169, July 2014.

[9] 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 International Conference on Image Processing (ICIP)*, pages 741–744, Sept. 2011.

[10] C. Harris and M. Stephens. A Combined Corner and Edge Detection. In *Proceedings of The Fourth Alvey Vision Conf.*, pages 147–151, 1988.

[11] HEVC. ITU-T Recommendation H.265/ ISO/IEC 23008-2:2013 MPEG-H Part 2/: High Efficiency Video Coding (HEVC), 2013.

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

[13] L. Karlsson, M. Sjöström, and R. Olsson. Spatio-Temporal Filter for ROI Video Coding. In *Proc. of the 14th European Signal Processing Conference (EUSIPCO)*, Sept. 2006.

[14] H. Meuel, M. Munderloh, and J. 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)*, pages 13 –20, June 2011.

[15] H. Meuel, M. Munderloh, M. Reso, and J. Ostermann. Mesh-based Piecewise Planar Motion Compensation and Optical Flow Clustering for ROI Coding. In *APSIPA Transact. on Signal and Informat. Processing, Submit. for Review*, 2015.

[16] H. Meuel, J. Schmidt, M. Munderloh, and J. Ostermann. *Advanced Video Coding for Next-Generation Multimedia Services – Chapter 3: Region of Interest Coding for Aerial Video Sequences Using Landscape Models*. Intech, Jan. 2013.

[17] M. Munderloh. *Detection of Moving Objects for Aerial Surveillance of Arbitrary Terrain*. PhD thesis, Leibniz Universität Hannover, Germany, 2015.

[18] 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*, 22(12):1669 –1684, Dec. 2012.

[19] J. Shi and C. Tomasi. Good Features to Track. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seattle, June 1994.

[20] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono. Stereoscopic 3d-tv: Visual comfort. *Broadcasting, IEEE Transactions on*, 57(2):335–346, June 2011.

[21] VideoLAN Organization. x265, Oct. 2014. v1.4.

[22] C. Zach, T. Pock, and H. Bischof. A Duality Based Approach for Realtime TV-L1 Optical Flow. In F. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 214–223. Springer Berlin Heidelberg, 2007.