IN-LOOP RADIAL DISTORTION COMPENSATION FOR LONG-TERM MOSAICING OF AERIAL VIDEOS

Holger Meuel, Stephan Ferenz, Marco Munderloh, Hanno Ackermann, Jörn Ostermann

Institut für Informationsverarbeitung, Leibniz Universität Hannover, Germany
Email: *{meuel, ferenz, munderl, ackermann, office}*@tnt.uni-hannover.de

ABSTRACT

For the generation of overview panoramic images from aerial surveillance videos, registered video frames are stitched together. Assuming a planar landscape, feature points can be detected and used to estimate a homography. However, if the features are affected by radial distortion, their mapping depends on their position within the frame and the resulting homography becomes inaccurate. As a result, the length of aerial panorama images is typically restricted to several hundred frames. To overcome this issue, we derive a model for the joint estimation of several homographies and one constant radial distortion. Due to the computational complexity of the solution, we propose a fast, iterative algorithm. Based on geometrical constraints, we regularize the projection of a jointly estimated picture group. We present panorama images from uncalibrated aerial videos with more than 1500 frames.

Index Terms— Long-term mosaicing, Panorama Image, Stitching, On-Board Image Distortion

1. INTRODUCTION

For aerial surveillance applications with a moving camera, *e. g.* when the camera is attached to a helicopter or an unmanned aerial vehicle (UAV), it is often desirable to have a scene overview in addition to the surveillance video. One common approach for visualization is the generation of a panorama image. To do so, single frames of a vertical aerial surveillance video sequence are projected in a common coordinate system. Assuming a planar surface of the ground as a world model, each frame of a video sequence can be registered into the panorama image by employing global motion compensation (GMC) techniques [1, 2, 3, 4, 5, 6]. These approaches work well for non-distorted cameras and if the video sequence characteristic perfectly matches the world model. Real world scenarios, however, rarely fulfill these constraints. Consequently, a lens distortion compensation or a full camera parameter estimation has to be performed. For unconstrained cameras with arbitrary motion and unknown camera parameters, automatic algorithms tend to estimate non-perfect parameters. Moreover, changing parameters over time are difficult to consider and to compensate. Feature points on non-planar structures like houses and trees may further impact an accurate estimation. Both effects finally result in inaccurate frame registrations. Consequently, existing mosaicing approaches, including those claiming long-term capabilities, are typically limited to only several hundred frames in one panorama image. To overcome this issue, it was proposed to mosaic only a few hundred frames and to concatenate the resulting panorama images afterwards [1, 4, 6]. Camera lens distortions like radial distortion are typically not taken into account when dealing with mosaicing approaches. In [7], the radial distortion was considered for the homography estimation for a pair of images. However, for noisy data, the estimation became challenging.

Our two contributions of this paper are as follows:

1. We present a model for the joint estimation of several homographies and one constant radial distortion. In contrast to [7], our model extents the estimation to an arbitrary number of homographies.
2. As fast approximative solution, we propose a fully automatic in-loop radial distortion compensation. In order to limit the change of the homographies' rotation angles, we restrict the change of shape and size of the projection from the camera-plane onto the mosaic.

The remainder of the paper is organized as follows: Section 2 introduces a mathematical model of radial distortion and gives an overview of related work in radial distortion estimation and compensation. In Section 3 our model for the joint estimation of several homographies with a common radial distortion is derived. Since the solution is tedious, we propose a faster, adaptive in-loop radial distortion compensating mosaicing system in Section 4. In Section 5 we present experimental results for artificially generated content as well as for real camera-captured surveillance videos before Section 6 concludes the paper.

2. RADIAL DISTORTION COMPENSATION

Radial distortion is one of the main distortions of optical lenses [8]. For a realistic radial distortion, we observed an error of about 0.5 pel between the undistorted and the distorted pixel projected into a panorama image for real aerial video sequences. This error can not be neglected and accumulates quickly to errors of several pel. Thus, radial distortion has to be considered for long-term mosaicing.

2.1. Related work

There has been a lot of research about radial distortion and radial distortion compensation [9, 10, 8, 11, 12]. Also in computer-vision, radial distortion has to be compensated depending on specific application requirements [13]. Most correction methods rely on any kind of test pattern to calibrate a lens at a given focal length. However, calibration pattern based methods like [11] can be applied only for known cameras. Even if the camera is known, it is challenging to calibrate lenses with long focal distances and the focus set to infinity as typical for aerial surveillance. Moreover, in aerial surveillance, the camera type and parameters are often unknown and thus have to be estimated from the video sequence. In [14], it was proposed to estimate the complete camera matrix including the radial distortion. The method is based on the estimation of projective homographies from corresponding image feature points, but is restricted to static scenes and limited degrees of freedom and thus not appropriate for aerial surveillance applications with a moving camera. In contrast to that, [8] proposed an approach to estimate the radial distortion based on edge detection and subsequent polygonal approximation in order to first detect straight lines and second to iteratively minimize the distortion error of different estimated radial distortion parameters while taking into account the straightness of detected lines in the image. However, in aerial surveillance applications, it cannot be guaranteed that straight lines *are* in the image and that those lines are indeed exactly *straight*. Consequently, a method not relying on specific image structures is more preferable. For an accurate homography estimation between two frames affected by unknown (and theoretically) different radial distortions, the radial distortion parameters have to be jointly estimated with the homography. Recently, a frame-to-frame-based approach was proposed and combined with *Random Sample Consensus* (RANSAC) for noise robustness for camera-captured signals [7]. However, for image sequences with more than two frames, a frame-to-frame-based method tends to estimate different radial distortions for different pairs of subsequent images, especially for noisy signals. In order to estimate constant radial distortions for a high number of subsequent frames, we propose to jointly estimate homographies for several frames with one common radial distortion.

2.2. Model of radial distortion

The mathematical model of radial distortion is well-known and will be summarized shortly based on [12]. The mapping of a real camera lens can be seen as a transformation modeling the deviations to the pinhole camera. Therefore the model projects each image point (x_d, y_d) of the distorted image to undistorted image coordinates (x_u, y_u). For further calculations we assume the center of the radial distortion to be equal to the principal point. We assume it to be in the center of the image; therefore the coordinates are shifted by half of the image width w and height h, each.

Referring to [8], the image distortion transformation can be described by a radial distortion function R. It can be expressed by an infinite series:

$$\begin{aligned} x_u &= x_d \left(1 + \kappa_{1_x} r_d^2 + \kappa_{2_x} r_d^4 + \dots\right) \\ y_u &= y_d \left(1 + \kappa_{1_y} r_d^2 + \kappa_{2_y} r_d^4 + \dots\right), \end{aligned} \tag{1}$$

with $r_d = \sqrt{\left(x_d^2 + y_d^2\right)}$ being the distance from the radial distortion's center and $\kappa_1, \kappa_2, \dots$ being the radial distortion parameters first, second, ... order. According to [9, 15], it is sufficient to only consider the first order radial distortion. Assuming radial symmetric lenses, the radial distortion parameters in horizontal and vertical direction κ_{1_x} and κ_{1_y}, respectively, are equal: $\kappa_1 := \kappa_{1_x} = \kappa_{1_y}$. With these assumptions we get:

$$\begin{aligned} x_u &= x_d \left(1 + r_d^2 \kappa_1\right) = x_d + x_d^3 \kappa_1 + x_d y_d^2 \kappa_1 \\ y_u &= y_d \left(1 + r_d^2 \kappa_1\right) = y_d + y_d^3 \kappa_1 + y_d x_d^2 \kappa_1. \end{aligned} \tag{2}$$

3. JOINT HOMOGRAPHIES AND RADIAL DISTORTION ESTIMATION

Since in surveillance scenarios the camera as well as its parameters (*e. g.* focal length) are constant over several frames, we propose to jointly estimate several homographies and one unknown, but constant radial distortion parameter κ_1. We optimize κ_1 in a way that the distortions introduced by mosaicing are kept small for all estimated frames, and thus to enable long-term mosaicing. Assuming a 3×3 projective transformation (homography) matrix $\mathbf{H}$, a point can be mapped from its homogeneous source coordinates $\vec{x_u} = (x_u, y_u, 1)^\top$ in one frame to its destination coordinates $\vec{x_u}' = (x'_u, y'_u, 1)^\top$ in the second frame:

$$\vec{x_u}' = \mathbf{H} \cdot \vec{x_u}. \tag{3}$$

A concatenation of several homographies leads to $\vec{x_u}'^n = (\mathbf{H_n} \cdot \dots \cdot \mathbf{H_1}) \cdot \vec{x_u}$. To map the undistorted feature coordinates $\vec{x_u}'$ back to the distorted ones $\vec{x_d}'$, we employ the inversion of Equation (2) with a given κ_1 and replace (x_u, y_u) by (x'_u, y'_u) and (x_d, y_d) by (x'_d, y'_d), and r_d by r'_d, respectively. The distorted radius is accordingly defined as $r'_d = \sqrt{\left({x'_d}^2 + {y'_d}^2\right)}$. To numerically solve this non-linear inversion, we use a Quasi-Newton method to determine the corresponding radially distorted point coordinates $\vec{x_d}'$ in dependence of the undistorted points coordinates $\vec{x_u}'$. We use the *Mean Squared Error* (MSE) as optimization criterion, denoting the mean squared Euclidean distance between all estimated feature points $\vec{x_d}'$ and their corresponding measured feature position point position $\vec{\hat{x}_d}'$.

Since an accurate estimation is highly computational complex and the run-time increases exponentially with the number of jointly estimated homographies (as we will show in the experimental section), we additionally propose a simplified, iterative, adaptive estimation approach.

4. IN-LOOP RADIAL DISTORTION COMPENSATION

We use the KLT- and RANSAC-based mosaicing framework from [5] and extend it by our proposed iterative in-loop radial distortion compensation approach in order to generate long panorama images. Since this approach works without any prior knowledge, it can adapt to any aerial video sequence which can be mapped into a plain panorama image. Assuming a (predominantly) planar ground, one frame $k-1$ can be mapped into frame k by applying a projective transformation as described above. The *Kanade-Lucas-Tomasi* (KLT) feature tracker [16] is employed to link Harris features [17] to correspondences between frames $k-1$ and k. Finally, *Random Sample Consensus* (RANSAC) [18] is performed to eliminate outliers and to determine the final projective transformation matrix $\mathbf{H}$ and thus the registration between both frames. After applying these steps for any pair of subsequent frames of a video sequence, all frames can be finally registered into one panorama image.

Whereas this frame-by-frame registration works well for undistorted video frames, it will fail for radial distortion affected frames due to accumulated registration errors, especially for longer image sequences with more than several hundred frames (example in Fig. 1a in the next section). Errors will occur, since the same object is contained in different frames at different positions and thus is affected differently by radial distortion. Moreover, also the accuracy of the homography will be influenced for distorted videos. Thus, we propose an adaptive, iterative, in-loop radial distortion compensation during the homography parameter estimation. For this, we use the radial distortion model from Subsection 2.2. We decompose the homography matrix $\mathbf{H}$ into the rotation matrix $\mathbf{R}$ and the translational vector $\vec{t}$ from one view to the other, the camera parameter matrices $\mathbf{K}$ and $\mathbf{K}'$ of the views, and the surface normal vector $\frac{\vec{n}}{d}$, with d being the distance between the camera center and the surface [19]:

$$\mathbf{H} = \mathbf{K}'\left(\mathbf{R} - \frac{\vec{t}\,\vec{n}^\top}{d}\right)\mathbf{K}^{-1}. \tag{4}$$

The rotation matrix is represented as

$$\mathbf{R} = \exp\left(\theta_x(t)\mathbf{W_x}\right) \cdot \exp\left(\theta_y(t)\mathbf{W_y}\right) \cdot \exp\left(\theta_z(t)\mathbf{W_z}\right), \tag{5}$$

with $\mathbf{W_x}$, $\mathbf{W_y}$, $\mathbf{W_y}$ being the skew-symmetric matrices induced by rotation around the x- (roll), y- (pitch) and z-axis (yaw), respectively. Assuming the surface geometry to be constant and the z-axis to be vertical to the earth surface, the change in size and shape of the projection of the camera target onto the mosaic only depends on $\theta_x(t)$ and $\theta_y(t)$. For a typical (and physically possible) motion of an aerial vehicle, we assume $\theta_x(t)$ and $\theta_y(t)$ to only change slowly. Thus, we optimize the radial distortion parameter in a gradient descent so that:

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\theta_x(t)\right| < c_x, \qquad \left|\frac{\mathrm{d}}{\mathrm{d}t}\theta_y(t)\right| < c_y. \tag{6}$$

For our iterative algorithm, the video is separated into *Picture Groups* (PGs) first, *e. g.* containing $n = 60$ frames each. Thus, our approach can be used on-the-fly for each PG and we don't have to wait until an entire sequence is recorded.

The initial radial distortion parameter for the first PG $l = 1$ is set to $\kappa_{1,l=1} = 0$, but can be arbitrary. As the radial distortion parameter usually does not change or changes only slowly over time (*e. g.* for changed intrinsic camera parameters such as focal length), $\kappa_{1,l}$ for all subsequent PGs is initialized with the previous PG's parameter $\kappa_{1,l-1}$. In an iterative loop, first, homographies for each frame of the current PG are estimated with the fixed, current $\kappa_{1,l_{\mathrm{cur}}}$. With these homographies, the mapping of each frame into a panorama image is evaluated. A plausibility check is performed, based on geometrical constraints: assuming a constant flight altitude and fixed camera parameters for the current PG, geometrical properties of the frame k have to be similar to these of the frame $k-1$, i. e. the frame should have similar shape and size. We tested different geometrical constraints for their suitability for optimization and found that the lengths of opposite sides as well as the size of the projected frame has to change only slowly. Since both of these measures have distinct minimums for the optimal radial distortion parameters, they are well-suited for optimization. Thus, we restrict these changes to be smaller than $c_{\mathrm{shape,max}}$ and $c_{\mathrm{size,max}}$, respectively. If these constraints are not fulfilled, another κ_1 is estimated using a bisection method and the loop is repeated until convergence or until a maximum number of iterations is reached. Since we aim at minimizing the distortion of projected frames into the panorama image, the estimated radial distortion will not necessarily be the real camera lens distortion but can also conceal (limited) model violations. Thus, our proposed algorithm enables long-term mosaicing over very long image sequences.

Finally, the entire PG is mapped into the panorama image, applying the optimal set of homographies.

5. EXPERIMENTS

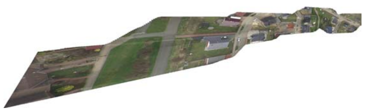
We present experimental results for the proposed joint estimation of several homographies and one radial distortion parameter in Section 5.1. Our results indicate that the accurate estimation is not practicable for larger picture groups due to run-time limitations. Thus, we present results for our proposed approximate solution in Section 5.2.

5.1. Joint homographies and radial distortion estimation

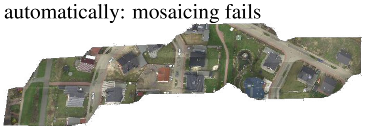
Simulations using Matlab by a Quasi-Newton descent proves our approach of a joint estimation of several homographies and one common radial distortion parameter κ_1 to be working. For noise-free, artificially generated point clouds with $N = 1000$ coordinates, using RANSAC to randomly sample $S = 30$ points for the homography estimation, we are able to achieve approximate solutions according to Table 1. Since the run-time increases exponentially, it is obvious that this approach is impractical for higher numbers of joint homographies estimations.

Table 1. Run-times and MSE of Matlabs (Quasi-Newton). No noise, initilization of solver with movement of distorted point cloud center, optimal value: κ_1=−3e−3, optimization criterion: MSE, (*): aborted, since convergence was not reached after $\gg 1000$ s.

# of homographies	MSE ≤ 1e−3		MSE ≤ 1e−6	
	$\kappa_{1_{est}}$	Time [s]	$\kappa_{1_{est}}$	Time [s]
2	−2.61e−3	7.8	−2.91e−3	10.9
3	−2.91e−3	17.3	−2.98e−3	42.0
4	−2.01e−3	32.7	−2.99e−3	222.2
5	−3.66e−3	57.1	(*)−2.97e−3	(*)1408.0
10	−3.48e−3	257.8	(*)−2.93e−3	(*)4268.7
10	−3.48e−3	257.8	−2.99e−3	561436.6



(a) Without any radial distortion compensation, neither manually nor automatically: mosaicing fails



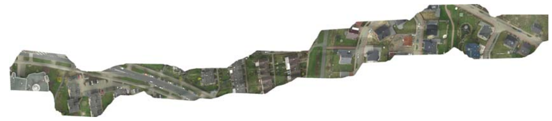
(b) With in-loop radial distortion compensation

Fig. 1. Panorama images after 330 frames (*350 m sequence*, TAVT [20] [21], scaled to fit column width)

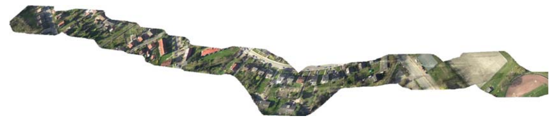
5.2. In-loop radial distortion compensation

For our experiments we used the unprocessed (not radially distortion compensated) test sequences from the *TNT Aerial Video Testset* (TAVT) [20] [21]. We used the iterative algorithm from Section 4, setting the maximum number of iterations to the empirically optimized value of $i = 14$ and the size of one picture group to $n = 60$. In order to limit the rotations as described by Equation (6), we use geometrical constraints as approximations instead. We allow for the empirically determined summarized changes of shape and size of $c_{\text{shape,max}} = 10\,\%$ and $c_{\text{size,max}} = 20\,\%$ per PG, respectively. Since for the test set the correct radial distortion parameter is not available, we use a manually corrected video instead as a reference. Matching the *350 m sequence* against Google Earth, we achieve a drift of only 4 pel per 909 pel, corresponding to 0.0044 pel/frame or 1 m per 230 m, respectively.

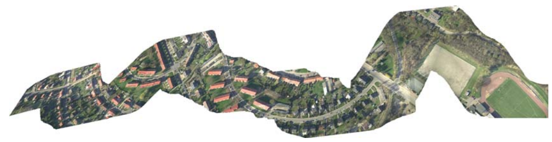
Whereas it is impossible to generate a panorama image from uncorrected content as recorded by the camera, our in-loop radial distortion compensation is able to fully automatically mosaic the same sequence (Fig. 1). On an Intel Xeon CPU E5-2670@2.60 GHz, our unoptimized C/C++ algorithm runs for the first PG of a sequence about 1000 ms/frame due to the initial estimation of a radial distortion parameter κ_1. This time can be decreased by providing a good initialization for κ_1, by reducing the picture group size, or by limiting the maximum allowed number of iterations; albeit the latter may reduce the accuracy of κ_1. For any other PG, the average run-time per frame is about 200 ms/frame, depending on the number of iterations necessary for fulfilling the optimization criteria. Entire panorama images for the test set are presented in Fig. 2. We would like to emphasize that common other mosaicing algorithms rely on either orthorectified images or use georeferenced images. Our algorithm only relies on the quasi-planar ground of the video sequence itself.



(a) *350 m sequence*, 821 frames



(b) *500 m sequence*, 1121 frames



(c) *1000 m sequence*, 1166 frames

(d) *1500 m sequence*, 1571 frames

Fig. 2. Panorama images of (non preprocessed, full HDTV) sequences (TAVT [20] [21], scaled to fit column width)

6. CONCLUSION

In this work we derive a model for the joint estimation of several homographies and one unknown, constant radial distortion parameter κ_1. Since the accurate and sophisticated estimation is impractical for higher numbers of jointly estimated homographies and tends to become instable, we propose a fast, fully-automatic in-loop radial distortion compensation algorithm instead. We use geometrical constraints to limit the change of shape and size of the projections in the panorama image. Thus, we get an optimized radial distortion parameter for projection, which does not necessarily match the correct radial distortion. By applying our algorithm to picture groups, we can guarantee a locally constant radial distortion compensation and achieve a low drift of only 0.0044 pel/frame. Furthermore, we are capable of adapting to changes in the scene or camera settings. Finally, we fully automatically generate long-term panorama images based on more than 1500 frames from camera captured surveillance videos without any manual radial distortion compensation or preprocessing.

7. REFERENCES

[1] S. Negahdaripour and P. Firoozfam, "Positioning and Photo-Mosaicking with Long Image Sequences; Comparison of Selected Methods," in *OCEANS, MTS/IEEE Conference and Exhibition*, 2001, vol. 4, pp. 2584–2592 vol.4.

[2] C.C. Dos Santos, S.A. Stoeter, P.E. Rybski, and N.P. Papanikolopoulos, "Mosaicking Images [panoramic imaging]," *Robotics Automation Magazine, IEEE*, vol. 11, no. 4, pp. 62–68, Dec 2004.

[3] D. Farin, M. Haller, A. Krutz, and T. Sikora, "Recent Developments in Panoramic Image Generation and Sprite Coding," in *IEEE 10th Workshop on Multimedia Signal Processing*, Oct 2008, pp. 64–69.

[4] Bo Yao, Xiaodong Cai, and Bizhong Wei, "Long-Term Background Reconstruction with Camera in Motion," in *2nd International Congress on Image and Signal Processing (CISP)*, Oct 2009, pp. 1–5.

[5] 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)*, Jun 2011, pp. 13 –20.

[6] S. Yahyanejad, M. Quaritsch, and B. Rinner, "Incremental, Orthorectified and Loop-Independent Mosaicking of Aerial Images Taken by Micro UAVs," in *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*, Sept 2011, pp. 137–142.

[7] Z. Kukelova, J. Heller, M. Bujnak, and T. Pajdla, "Radial Distortion Homography," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 639–647.

[8] Frédéric Devernay and Olivier Faugeras, "Straight lines have to be straight," in *Proceedings of SPIE*, 2001, vol. 2567.

[9] H.A. Beyer, "Accurate Calibration of CCD-Cameras," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 1992, pp. 96 –101.

[10] Thorsten Thormählen and Hellward Broszio, "Automatic line-based estimation of radial lens distortion," *Integrated Computer-Aided Engineering*, vol. 12, no. 2, pp. 177–190, Apr 2005.

[11] Tobias Elbrandt and Jörn Ostermann, *Enabling accurate measurement of camera distortions using dynamic continuous-tone patterns*, vol. 18, IOS Press, 2011.

[12] Holger Meuel, Marco Munderloh, and Jörn Ostermann, "Radial Distortion in Hybrid Video Coding," Technical report, Institut für Informationsverarbeitung, Leibniz Universität Hannover, Hannover, Germany, June 2012.

[13] R. Szeliski, *Computer Vision: Algorithms and Applications*, Texts in Computer Science. Springer, 2010.

[14] Thorsten Thormählen, Hellward Broszio, and Ingolf Wassermann, "Robust line-based calibration of lens distortion from a single view," *Proceedings of Mirage 2003 (Computer Vision/Computer Graphics Collaboration for Model-based Imaging, Rendering, Image Analysis and Graphical Special Effects), INRIA Rocquencourt, France, March, 10-12 2003, pp. 105-112*, 2003.

[15] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *Robotics and Automation, IEEE Journal of*, vol. 3, no. 4, pp. 323 –344, August 1987.

[16] Jianbo Shi and Carlo Tomasi, "Good Features to Track," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seattle, June 1994.

[17] C. Harris and M. Stephens, "A Combined Corner and Edge Detection," in *Proceedings of The Fourth Alvey Vision Conf.*, 1988, pp. 147–151.

[18] Martin A. Fischler and Robert C. Bolles, "Random Sample Consensus: a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[19] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, 2nd edition, 2004.

[20] 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/.

[21] Holger Meuel, Marco Munderloh, Matthias Reso, and Jörn Ostermann, "Mesh-based Piecewise Planar Motion Compensation and Optical Flow Clustering for ROI Coding," in *APSIPA Transactions on Signal and Information Processing*, 2015, vol. 4.