# Moving Object Tracking for Aerial Video Coding using Linear Motion Prediction and Block Matching

Holger Meuel, Luis Angerstein, Roberto Henschel, Bodo Rosenhahn, Jörn Ostermann
Institut für Informationsverarbeitung
Gottfried Wilhelm Leibniz Universität Hannover
Hannover, Germany
Email: *{meuel, angerste, henschel, rosenhahn, office}*@tnt.uni-hannover.de

***Abstract*—Region of Interest (ROI) coding is a common method for data reduction in scenarios where bandwidth is crucial like in aerial video surveillance from Unmanned Aerial Vehicles (UAVs). In order to save bits, non-ROI areas are typically reduced in quality or not transmitted at all and thus, an accurate ROI classification is mandatory. Moving objects (MOs) are often considered as ROIs and consequently have to be accurately detected on-board. However, common detection approaches either rely on computationally demanding processing which is not available at small UAVs with only limited energy, are model based or cannot provide a sufficient detection precision. While not detected MOs lead to a degraded representation at the decoder, erroneously detected MOs lead to an unnecessary high bit rate. We tackle all these issues utilizing an efficient object proposal computation.**

**Based on a dual-threshold strategy applied to image differences, we propose a linear prediction-supported block matcher. Compared to a simple thresholding approach, it shows superior performance and is robust to threshold tuning. By integrating superpixels into the framework, we further recover the complete shape of the MOs. Finally, an efficient tracking-by-detection system is employed to produce accurate detections from the proposals, thereby recovering missed MOs and denying wrong proposals, making the coding more efficient.**

**We achieve an improved detection precision of up to 76 % compared to a simple difference image-based approach. By using a general ROI coding framework we reduce the bit rate of our test set by 70 % compared to common HEVC.**

## I. Introduction

The high resolution *Pulse Code Modulation* (PCM) video data rate of 622 Mbit/s for a color video sequence with full *High Definition Television* (HDTV) resolution (1920×1080) can be typically compressed to 4–13 Mbit/s at a reasonable image quality [1], [2] using standardized hybrid video coding like *High Efficiency Video Coding* (HEVC) [3]. Since bandwidth is of crucial importance for some scenarios like aerial surveillance from Unmanned Aerial Vehicles (UAVs) and taking into account even higher camera resolutions or multi-camera setups, the bit rate often has to be much further reduced, *e. g.* by using *Region of Interest* (ROI) coding. It typically relies on quality reduction of non-ROI areas, resulting in a reduced bit rate for non-ROI areas and thus for the entire video. For aerial surveillance systems, moving objects (MOs) are typically of high interest and thus are considered as ROI. In order to achieve the optimal coding performance and to provide a high quality for all MOs, the latter have to be accurately detected on-board. Therefore, we propose to employ a model-less, three-stages moving object detector (Fig. 1). We use a dual-threshold, block matching supported (pel-wise) difference image first (stage 1) to select corresponding superpixels from an independent superpixel segmentation [4] (stage 2). To cope with non-perfect detections from the simple detectors, *e. g.* introduced by noise or motion parallax, a tracking-by-detection [5], [6] approach is used to eliminate outliers and also link single detections (stage 3). The resulting coding mask is used for the control of a highly efficient ROI-based coding framework for aerial videos [7].

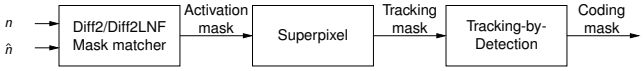



Figure 1. Block diagram of the entire 3-stages moving object detection system

### A. Related Work

Most ROI coding systems provide the highest image quality only for predefined regions whereas non-ROI areas are degraded in quality. For instance, blurring or a coarse quantization of a frame either as preprocessing prior to actual video encoding or within the video encoder itself can be applied [8]–[10], resulting in reduced bit rates. In contrast to that a video coding system relying on reconstruction of non-ROI areas by means of *Global Motion Compensation* (GMC) was proposed in [7], [11], [12]. By using GMC, a subjectively high image quality could be provided over the entire frame at very low bit rates of 0.8–2.5 Mbit/s for full HDTV resolution aerial video sequences [2]. The ROI coding systems defines two different ROIs: the first one contains only newly emerging areas for each frame (ROI-NA) which were not contained in the previous frame. These ROI-NAs are stitched together by means of GMC into a background panorama image at the decoder, using a projective transform homography $\mathbf{H_n}$ estimated on-board the UAV and transmitted as side-information. From the panorama image, the video frames are reconstructed [13]. In order to also reconstruct objects with a local motion (moving objects), moving objects are detected by a moving object detector which calculates the pel-wise luminance difference (difference image) between the motion compensated frame $\hat{n}$ and the current frame $n$. Spots of high energy in this difference image are additionally defined as ROI (ROI-MO). Like other simple approaches, this difference image-based moving object detector is highly computational efficient and provide

satisfactory detection results in several applications [14]–[17]. However, since the detection performance typically depends on single thresholds, *e. g.* for noise filtering, these approaches tend to be non-robust for different scenarios. As a consequence, they often lack accuracy for small objects with only very small amplitudes in the difference image as well as for objects with low contrast compared to the background. Thus, these detectors cannot fulfill high detection precision demands. More efficient detectors were proposed, which exploit parallax effects [18] or use block matching motion vectors [19] or an optical flow analysis in order to detect moving objects [20], [21]. [22] finally tracks clustered image features over several frames in order to improve the detection accuracy.
All these approaches have in common that the actual moving object detection is performed only based on features but do not consider the image content after the feature derivation anymore. In contrast to those approaches, specialized detectors may be utilized for the detection of certain objects, *e. g.* cars, by sophisticated classification or machine learning algorithms (SVM, HOG, SIFT, CNNs etc.). However, since in a surveillance system it is more important to detect *any* local motion instead of only predefinded objects, these methods cannot be used.

Thus we propose to combine a model-less dual-threshold difference image-based moving object detector combined with a (modified) block matcher (mask matcher) in order to find and track moving objects also in subsequent frames. The result of the difference image-based analysis is used as seed for a mask matcher. By this approach we find corresponding areas to any detection in subsequent frames and thereby can highly improve the precision of the detector. Since small and highly optimized, low energy block matchers are available from common video coding (*e. g.* in mobile phones), no additional efforts must be undertaken for the energy-efficient usage on-board of UAVs. For accurate shape retrieval especially for homogeneous, unstructured areas of moving objects, a superpixel segmentation is employed before a tracking-by-detection framework is used to link moving objects over long temporal distances. Using the enhanced moving object detections, a ROI-based coding is performed.

The remainder is organized as follows: In Section II our proposed block matching-assisted difference image-based moving object detector is described in detail. We present experimental results in Section III before Section IV concludes the paper.

## II. Block Matching Supported Moving Object Tracking

We use the ROI detection and coding system from [7] as a basis, since it provides a subjectively high image quality over the entire frame by encoding only newly emerging areas for each frame (ROI-NA) and locally moving objects (ROI-MO) es explained at the beginning of Section I-A. To overcome the weaknesses of its difference image-based detector, we propose to combine this simple method with block matching to exploit similarities of the shape of moving objects over neighboring frames. Basically, we apply a block matching between two video frames $n-1$ and $n$ for every spot of high energy in the difference image between both frames. If a proposal trajectory can be built up for one object over several frames, we assume it to be a moving object, otherwise the proposal is discarded. The proposals of the remaining trajectories are extended to the MOs full shape using superpixels as proposed in [23] in order to obtain accurate shape information. At the same time, this circumvents splitting of an object into multiple proposals.

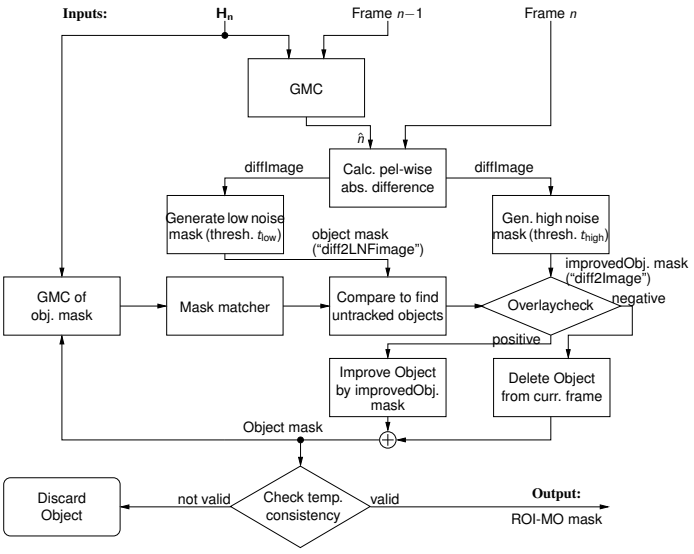



Figure 2. Flowchart of proposed moving object detector.

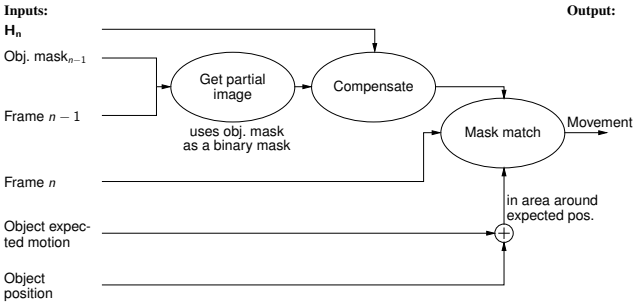



Figure 3. Mask matching process.

Finally an efficient tracking-by-detection framework [6] is employed to reliably link detections and to overcome the limits of the difference image-based tracker, *e. g.* missing detections as well as partial and long term occlusions (Fig. 1).

### A. Proposed Moving Object Detector

The block diagram of the proposed moving object detector is shown in Fig. 2. We use a dual-threshold strategy (also known as hysteresis thresholding) to filter as many false positive detections (static objects falsely detected as moving) on the one hand, and to retrieve the rough shape from the difference image on the other hand: Firstly, we calculate the pel-wise luminance differences between the global motion compensated (block "GMC" in Fig. 2) frame $\hat{n}$ and the camera-captured frame $n$ (diffImage output of block "Calc. pel-wise abs. difference" in Fig. 2). Secondly, we generate two different binarized versions (diff2image and diff2LNFimage) out of the same difference image (diffImage) with different noise filter settings, one with a high ($t_{\text{high}}$) and the other with a very low ($t_{\text{low}}$) noise filter threshold, respectively. For each pixel in the

diffImage we summarize the luminance values ("energy" in the difference image) in a $3 \times 3$ sliding window. If the summarized energy is above a threshold $t_{\text{high}}$, the current pixel is marked as moving object candidate in diff2image. Similarly, we generate the second binarized image diff2LNFimage using another threshold $t_{\text{low}}$. Diff2image (with $t_{\text{high}}$) basically controls, *where* a moving object is considered, whereas diff2LNFimage is used to improve the shape information which is passed to the mask matcher. Due to the combination of two different thresholds we become invariant against manual threshold tweaking in a wide range. By applying a region growing algorithm to both images, a set of object candidates (Objects) and object candidates with improved shapes (improvedObjects) are created. In the low-noise difference image, small and slow moving objects become more visible (see Fig. 4(b) in the next section) but also the false positive detection rate is increased, *e. g.* caused by noise or non-planar structures like houses or trees violating the implicit planarity assumption of the homography-based global motion compensation.

Starting with the known mask of all already tracked objects from the last frame, we firstly project the old mask into the current frame $n$ by applying the homography $\mathbf{H}_n$ ("GMC of object mask"). Next, a modified block matcher ("Mask matcher") is used to estimate the motion of the tracked object in the input video frames. For the search process, only pixels according to the mask of the object are considered instead of rectangular blocks, hence we call our matching module mask matcher (Fig. 3). As seed position for the mask matcher, the motion of the object from the last $m$ frames (we used $m = 3$) is linearly extrapolated into the current frame $n$, which is justified by the relatively high frame rates of typical video sequences of *e. g.* 25 or 30 fps. The resulting motion is applied to the objects mask.

In order to find new objects which are not tracked yet, the tracked objects are compared with the set of objects found in the difference image. A new Object is added for each untracked object. As the results of the mask matcher may not be the true motion of the object, a tracked object may be moved to a wrong position leading to non-ROI parts erroneously being marked as ROI. Therefore an overlaycheck is performed that compares all already tracked and all newly added Objects with the improvedObjects list. Every tracked object that does not match with a improvedObject will be marked as an virtual object. Based on the assumption that real moving objects (true positive detections) reappear at similar positions in subsequent frames, we can distinguish them from false detections: if a virtual Object is not recovered for $x$ frames, it is deleted. Otherwise, in the case of a successful overlaycheck, the estimated motion is confirmed and the mask of the Object is refined by merging its mask with the mask of the corresponding improvedObject. All pixels from a mask of a Object have a corresponding weight between 0 and 1, indicating the probability that it belongs to the Object. The weight of points in the mask of the Object that do not appear in the mask of the corresponding improvedObject mask is reduced in order to further refine the mask of the objects in subsequent frames and may further be used for detection and tracking confidence. The final ROI-MO activation mask (Fig. 2) is used for further processing.

### B. Accurate shape retrieval by Superpixel Segmentation

In order to retrieve accurate shape information of moving object proposals with unstructured texture, we employ the independently calculated superpixel segmentation of the input video frames [4]. Using the activation mask, a tracking mask (Fig. 2) is generated by inserting the areas of each superpixel which is covered by at least one marked pixel of the activation mask. By combining multiple detections of parts of the same moving object proposal to one proposal in the tracking mask, a correct processing of entire MOs including homogeneous parts is realized. The tracking mask finally is passed to the object tracker for a temporal linking of the single detections.

### C. Tracking-by-Detection

Using the proposed block matcher we obtain good initial objects proposals. Thereby superpixel help to improve the segmentation quality and to reduce the false positive rate by merging proposals belonging to the same object. However, the proposals are still prone to miss objects in cases of (partial) occlusion or deformation as they are based on image information. To recover objects also in cases where the appearance assumptions fail, we employ the robust and efficient multi object tracker [6] (MCA-Tracker), using only position information.

The hierarchical MCA-Tracker is a tracking-by-detection approach that seeks for a minimum cost arborescence in the detection association graph and obtains the optimal linking of the current association graph in linear time in the number of detections. It is thus suitable for the application in UAVs. Moreover, their proposed *tree tracklet* trajectory model allows it to reliably reject wrong detections and recover missing ones.

Finally, the computed detections are added as ROI to the coding mask.

## III. Experiments

We used the *350 m sequence* from the *TNT Aerial Video Testset* (TAVT) [2], [24] as well as the self-recorded aerial HD video sequence *Parking Lot* containing lots of moving objects on a parking lot and its nearby streets (overview frame in Fig. 4(e)). Whereas fast and large objects (*e. g.* the big white car in the *350 m sequence*) can be satisfactorily detected for a subsequent processing by the unmodified difference image-based moving object detector from [7], small and slow objects like the pedestrian in the *350 m sequence* or the cars in the *Parking Lot sequence* are still challenging. Since they may still be missing after tracking-by-detection was applied on the simple difference image (Table I), those objects may lead to erroneous results. Since our ROI coding system relies on background reconstruction (*i. e.* GMC), not detected MOs get lost and are replaced by corresponding background which is not meaningful for a surveillance system. To this end, a high detection recall rate in the first stage (Fig. 1) is desirable, which is achieved by our robust dual-threshold strategy. At

Table I

DETECTION RESULTS AGAINST MANUALLY CREATED GROUND TRUTH (30 FRAMES FOR *Parking Lot sequence*, 40 FRAMES FOR *350 m sequence*). *False positive* (FP) DETECTIONS DENOTE NON MOVING OBJECTS ERRONEOUSLY DETECTED AS MOVING (*e. g.* HOUSE EDGES), *false negative* (FN) DETECTIONS MOVING OBJECTS NOT RECOGNIZED AS MOVING (LNF: LOW NOISE FILTERING; SP: SUPERPIXEL; TBD: TRACKING-BY-DETECTION).

| Seq. | Input image | Recall | Prec. | FP | FN |
|---|---|---|---|---|---|
| Parking Lot | diff2Image | 76.3 | 26.8 | 985 | 112 |
| | diff2Image & SP | 93.4 | 37.0 | 753 | 31 |
| | diff2Image & SP & TbD | 94.7 | 38.9 | 704 | 25 |
| | diff2LNFimage | 95.3 | 15.9 | 2380 | 22 |
| | diff2 images &mask matcher | 86.7 | 65.6 | 215 | 63 |
| | diff2 images &mask mat.&SP | 86.0 | 90.8 | 41 | 66 |
| | **diff2 imgs. &mask mat.&SP&TbD** | 89.9 | 92.6 | 34 | 48 |
| 350 m | diff2Image | 100.0 | 7.8 | 377 | 0 |
| | diff2Image & SP | 96.9 | 12.7 | 214 | 1 |
| | diff2Image & SP & TbD | 96.9 | 12.5 | 217 | 1 |
| | diff2LNFimage | 96.9 | 5.9 | 492 | 1 |
| | diff2 images &mask matcher | 96.9 | 26.3 | 87 | 1 |
| | diff2 images &mask mat.&SP | 96.9 | 70.5 | 13 | 1 |
| | **diff2 imgs. &mask mat.&SP&TbD** | 96.9 | 83.8 | 6 | 1 |

the same time a high precision rate is required to keep the required bandwidth low. We accomplish this by the mask matcher at stage 1 together with the tracking-by-detection system. To obtain quantitative evidence of the effectiveness of our approach, we evaluate the system using the object based true positive (TP) and false positive (FP) rates and, respectively, and recall as well as precision measure as in [25] against manually created ground truth data:

$$\text{Recall} = \frac{\#\text{TP}}{\#\text{GT}} \qquad \text{Precision} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FP}}. \tag{1}$$

Table I shows that the precision of simple difference image-based moving object detectors is low. Whereas the integration of superpixels into the system improves the recall, the detection precision still remains low in absolute terms (below 37 %). By using our proposed mask matcher before the superpixel enhancement, we increase the precision to more than 70 %. With the improved detections, we were able to additionally increase the tracking accuracy of the tracking-by-detection framework [6] by up to 13 %.

In the *Parking Lot sequence* (Fig. 4) we observed that the detection of small and slow moving objects becomes possible only with our proposed system because they are often not contained in the common difference image (Fig. 4(a)). Using the low noise filter threshold alone leads to high false detections at the static background, *e. g.* due to motion parallax effects (Fig. 4(b), detected building structures top left). Since the latter detections do not have a directed movement but behave similar to noise, we can remove them with our proposed moving object detector without impairing the TP detections of the real moving objects (Fig. 4(c)). Accurate shape information can be retained by utilizing independently calculated superpixels (Fig. 4(d)). Fig. 4(e) finally shows the decoded image after reconstruction by means of GMC.

In order to demonstrate the robustness of our system we show that the detection results are mostly insensitive to the used threshold $t_{\text{high}}$. The noise filter threshold $t_{\text{high}}$ basically defines the number of objects which are processed by our mask matcher and can be set *e. g.* to values between 300 and 450

Table II

CODING RESULTS USING THE GENERAL ROI CODING FRAMEWORK FROM [7] WITH THE X265 VIDEO ENCODER [26] (BR: BIT RATE IN KBIT/S; ROI-PSNR: PSNR MEASURED IN ROI AREAS ONLY [2], [27]).

| | Parking Lot sequence | | 350 m sequence | |
|---|---|---|---|---|
| | BR [kbps] | ROI-PSNR [dB] | BR [kbps] | ROI-PSNR [dB] |
| HEVC | 20576 | 37.1 | 8575 | 40.8 |
| ROI HEVC | 4899 | 37.1 | 2597 | 41.0 |

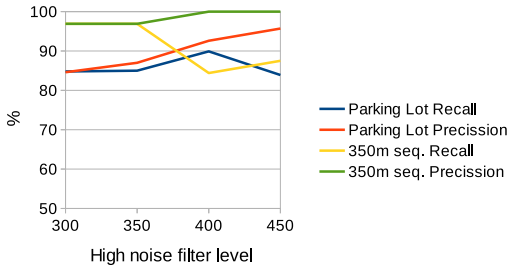



Figure 5. Robustness of the proposed system against parameter fine-tuning: the detection precission is mainly independent of the noise filter level $t_{\text{high}}$.

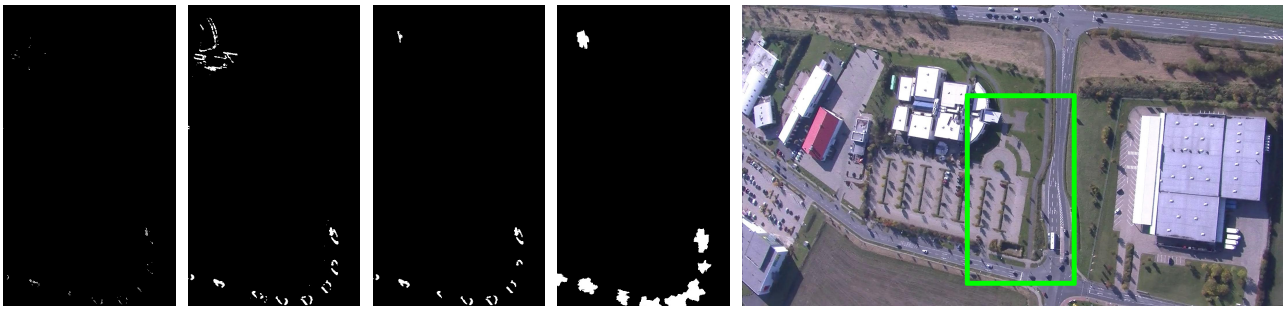
(we used $t_{\text{high}} = 350$) without impacting recall or precision much (Fig. 5). Since the low filter threshold ($t_{\text{low}}$) is only used to find a rough object shape as initialization for the mask matcher, the detection results are mainly unaffected as long as the threshold is set to a small value allowing a lot of energy in the diff2LNFimage (also including noise), *e. g.* $t_{\text{low}} = 200$.

Using the general ROI coding framework from [7] with the *x265* video encoder software (Lavc57.48.101 libx265) [26], we can reduce the bit rates by 76.2 % and 69.7 % for the *Parking Lot* and the *350 m sequence*, respectively, compared to common HEVC encoding at similar quality levels (37 dB and 41 dB ROI-PSNR [2], [27]). We would like to emphasize that a subjectively high image quality is provided over all entire frames due to the reconstruction of non-ROI by global motion compensation of previously transmitted ROI New Area.

## IV. CONCLUSION

We present a reliable detection system for small and slow moving objects in aerial video sequences which can be employed *e. g.* in ROI-based detection and coding systems onboard of (small) UAVs. By utilizing a dual-threshold strategy, our system becomes robust against parameter tweaking without impairing the detection precision much. After applying our proposed modified block matcher (mask matcher), we integrate independently calculated superpixels to cluster pelwise detections to compact objects and thus retrieve improved object shapes. Due to the decreased false positive and false negative detections, a tracking-by-detection framework can be utilized to efficiently eliminating remaining false detections, whereas a reliable tracking is not possible without our proposed preprocessing due to too many false detections.

We showed that the precision of the object-based detection is highly increased by the integration of the mask matcher from 12 % to more than 80 % in a fully automatic process. Employing the improved detections in a ROI coding framework we reduce the bit rates by about 70 % compared to common HEVC coding for the encoding of full HDTV resolution sequences (@30 fps) with a subjectively high image quality over



(a) Diff2image (b) Diff2LNFimage (c) After mask match. (d) W/ superpixels (e) Decoded & reconstructed

Figure 4. Detection precision improvement for self recorded parking lot sequence (a–d: magnifications); a) Diff2image: Binarized difference image from [7]: lots of MOs are not or not entirely detected in every frame; b) Diff2LNFimage: all MOs but also lots of false positive detections (*e. g.* building upper left) are detected; c) Diff2 images and applied mask matching: all MOs are detected, false positive detections are nearly entirely removed; d) Mask matching result (from c)) with superpixel enhancement: shapes of MOs are retrieved; e) Decoded and reconstructed entire video frame.

the entire frame. At the same bit rate, we retain more details with our ROI coding framework than with common HEVC.

## REFERENCES


[1] 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.

[2] 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 Sig. and Inform. Proc.*, vol. 4, 2015. [Online]. Available: http://journals.cambridge.org/article_S2048770315000128

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

[4] M. Reso, J. Jachalsky, B. Rosenhahn, and J. Ostermann, "Temporally Consistent Superpixels," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, Dec. 2013.

[5] L. Leal-Taixé, G. Pons-Moll, and B. Rosenhahn, "Everybody needs somebody: modeling social and grouping behavior on a linear programming multiple people tracker," *IEEE International Conference on Computer Vision Workshops (ICCVW). 1st Workshop on Modeling, Simulation and Visual Analysis of Large Crowds*, Nov. 2011.

[6] R. Henschel, L. Leal-Taixé, and B. Rosenhahn, "Efficient Multiple People Tracking Using Minimum Cost Arborescences," in *German Conf. on Pattern Recognition (GCPR)*, Sep. 2014.

[7] H. Meuel, M. Munderloh, F. Kluger, and J. Ostermann, "Codec Independent Region of Interest Video Coding using a Joint Pre- and Postprocessing Framework," in *Int. Conf. on Multim. & Expo (ICME)*, July 2016.

[8] 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 Trans. on Consumer Electr.*, vol. 49, no. 3, pp. 724–730, Aug 2003.

[9] 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.

[10] 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.

[11] 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)*, June 2011, pp. 13 –20.

[12] H. Meuel, F. Kluger, and J. Ostermann, "Illumination Change Robust, Codec Independent Low Bit Rate Coding of Stereo from Singleview Aerial Video," in *10th IEEE Int. 3DTV Conf.*, July 2016, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/document/7548961/

[13] 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. [Online]. Available: http://www.intechopen.com/books/advanced-video-coding-for-next-generation-multimedia-services/region-of-interest-coding-for-aerial-video-sequences-using-landscape-models

[14] R. Jones, B. Ristic, N. Redding, and D. Booth, "Moving Target Indication and Tracking from Moving Sensors," in *Proc.of Dig.Image Comput.: Techn.&Applicat.*, Dec 2005, pp. 46–46.

[15] A. Shastry and R. Schowengerdt, "Airborne Video Registration and Traffic-Flow Parameter Estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 4, pp. 391–405, Dec 2005.

[16] X. Cao, J. Lan, P. Yan, and X. Li, "KLT Feature Based Vehicle Detection and Tracking in Airborne Videos," in *Sixth International Conference on Image and Graphics (ICIG)*, Aug 2011, pp. 673–678.

[17] A. Ibrahim, P. W. Ching, G. Seet, W. Lau, and W. Czajewski, "Moving Objects Detection and Tracking Framework for UAV-based Surveillance," in *Fourth Pacific-Rim Symposium on Image and Video Technology (PSIVT)*, Nov 2010, pp. 456–461.

[18] J. Kang, I. Cohen, G. Medioni, and C. Yuan, "Detection and Tracking of Moving Objects from a Moving Platform in Presence of Strong Parallax," in *Proc. of the IEEE Int. Conf. on Computer Vision (ICCV)*, vol. 1, Oct 2005, pp. 10–17.

[19] M. K. Fard, M. Yazdi, and M. MasnadiShirazi, "A Block Matching Based Method for Moving Object Detection in Active Camera," in *5th Conf. on Inform. & Knowledge Technol. (IKT)*, May 2013, pp. 443–446.

[20] H. Yalcin, M. Hebert, R. Collins, and M. Black, "A Flow-Based Approach to Vehicle Detection and Background Mosaicking in Airborne Video," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, June 2005, p. 1202.

[21] M. Munderloh, *Detection of Moving Objects for Aerial Surveillance of Arbitrary Terrain*, ser. Fortschritt-Berichte. VDI Verlag, Mar. 2016, vol. 10, no. 847.

[22] M. Teutsch and W. Kruger, "Detection, Segmentation, and Tracking of Moving Objects in UAV Videos," in *Proceedings of the IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, Sept. 2012, pp. 313–318.

[23] H. Meuel, M. Reso, J. Jachalsky, and J. Ostermann, "Superpixel-based Segmentation of Moving Objects for Low-Complexity Surveillance Systems," in *Proc. of the Tenth IEEE Internat. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, Aug. 2013, pp. 395–400.

[24] Inst. f. Informationsverarb. (TNT), Leibniz Universität Hannover. (2010–2014) TNT Aerial Video Testset (TAVT). [Online]. Available: https://www.tnt.uni-hannover.de/project/TNT_Aerial_Video_Testset/

[25] K. Bernardin and R. Stiefelhagen, "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics," *J. Image Video Process.*, vol. 2008, pp. 1:1–1:10, Jan. 2008. [Online]. Available: http://dx.doi.org/10.1155/2008/246309

[26] VideoLAN Organization. (2014, Oct.) x265. V1.4. [Online]. Available: http://www.videolan.org/developers/x265.html

[27] P. Gorur and B. Amrutur, "Skip Decision and Reference Frame Selection for Low-Complexity H.264/AVC Surveillance Video Coding," *IEEE Trans.on Circ.&Syst. f. Vid.Techn.*, vol. 24, no. 7, pp. 1156–1169, July 2014.