CODEC INDEPENDENT REGION OF INTEREST VIDEO CODING USING A JOINT PRE- AND POSTPROCESSING FRAMEWORK

Holger Meuel, Marco Munderloh, Florian Kluger and Jörn Ostermann

Institut für Informationsverarbeitung, Leibniz Universität Hannover, Germany
{meuel, munderl, kluger, office}@tnt.uni-hannover.de

ABSTRACT

For low bit rate scenarios (video conferencing, aerial surveillance), conventional video coding is unable to meet the small bit rate and high quality requirements. In contrast to that *Region of Interest* (ROI) coding provides an efficient compression by improving the quality of ROIs at the expense of non-ROIs. We also transmit ROI only, but reconstruct non-ROI from already transmitted content by means of global motion compensation in order to provide a high quality for the full frame. Previous ROI coding systems modified the video codec to control the coding of individual blocks. We propose a codec and ROI detector independent pre- and postprocessing framework instead. This enables the usage of off-the-shelf hard-/software and an easy adaption to the latest video coding technology. Maintaining the performance of subsequent computer vision tasks, we reduce the bit rate by 90–95 % to less than 1 Mbit/s using HEVC for full HDTV videos.

Index Terms— Region of Interest Video Coding, Preprocessing, AVC, HEVC, Global Motion Compensation (GMC)

1. INTRODUCTION

The spatial video resolution (HDTV, UHDTV) increases steadily for professional applications such as movie production, surveillance as well as for consumer products like hand-held video cameras. With high resolutions also the amount of data to be stored or transmitted increases drastically. For instance, a subjectively "good" representation of a full HDTV resolution sequence (1920 × 1080 pel, 30 fps) requires more than 5 Mbit/s with the most recent video coding standard *High Efficiency Video Coding* (HEVC) and even more than 10 Mbit/s with its often used predecessor, *Advanced Video Coding* (AVC) [1].

However, not all applications require high resolution and quality for the full frame. Also, bandwidth can be crucial in some application scenarios. *Region of Interest* (ROI)-based coding systems provide divergent compression for individual parts of the frame, usable for dedicated tasks like video conferencing or surveillance scenarios with areas of different importance. For instance, for mobile traffic monitoring from police helicopters, detailed information about moving objects is more important than for the (quasi static) background which can therefore be compressed more strongly. The same is applicable for fixed surveillance systems or video conferencing. Consequently, moving objects are often treated as ROI and are encoded in high quality at the cost of non-ROI areas (background). The resulting low quality of non-ROI areas strongly impacts computer vision applications, e. g. for background modeling or detection/classification of non-moving objects.

1.1. Related Work

In order to reduce the bit rate needed for the video coding of non-ROI areas of a frame, they can be blurred in a preprocessing step prior to the actual video encoding or they might be coarsely quantized during the encoding itself [2]. A modified and externally controllable block-based hybrid video coder like *Advanced Video Coding* (AVC) [3] or *High Efficiency Video Coding* (HEVC) [4] is employed in [5] and [6], respectively, in order to apply different quantization parameters – and thus "image quality levels" – for the coding of ROI and non-ROI blocks. However, the video encoders had to be modified for these systems and did therefore not allow the usage of off-the-shelf hardware. Moreover, it requires a costly adaption to each new coding standard and even to any alternative video encoder software.

To enable a fully-automatic processing, ROI and non-ROI have to be classified automatically. For our scenarios, a moving object detector can provide the desired information. Most of the recent work in the field of surveillance video processing, especially for automatic moving object (MO) detection, relies on a static (non-moving) camera, e. g. [7, 8] and does not deal with camera ego-motion. Common approaches for aerial surveillance rely on the global motion compensation of the background pixels due to the camera movement prior to calculating the pel-wise image differences (difference image) between two frames of the image sequence or between the current frame and a reconstructed background reference image [9]. More efficient detectors can also handle non-perfect conditions like parallax effects by employing the epipolar geometry [10]. [11] provides an extensive overview on recent work in aerial surveillance systems from moving cameras.

The drawback of the ROI detection and the coding approaches discussed above is the irreversible degradation of non-ROI areas (i. e. they cannot be reconstructed at full quality at the decoder). The ROI detection and coding system from [12, 13, 14] exploits the characteristic of aerial video sequences assuming a planar landscape to overcome this drawback. It maintains the full resolution and high video quality over the entire frame at low bit rates, enabling computer vision applications on the full frame, e. g. (static) object detection, classification or scene understanding. This ROI coding system relies on the transmission of new emerging image content (*New Areas*, NA) for each of the frames. These new areas are stitched together at the decoder in a post-processing step to reconstruct the static parts of the scene (background) by means of *Global Motion Compensation* (GMC) [13, 14]. In contrast to other approaches introduced here, this system is able to reconstruct the background in full quality at bit rates considerably lower than those of common state-of-the-art video encoder. It is used as reference system for our paper. Since common video encoders have no interface for an external coding control, intricate video encoder modifications are necessary for the video encoder in the reference system as well as in previously introduced ROI coding systems. With respect to economic and power consumption considerations for many real life applications, it is desirable to use off-the-shelf hardware or software video encoders instead of application specific encoders with either non standard conforming coding tools or special encoder control mechanisms.

In order to become independent of the video codec, we propose a preprocessing framework which enables an encoder-independent and reversible coding of regions of interest. Due to the codec independence, we refer to our framework as *general ROI coding* system. It is based on the modification of image blocks before the actual video encoding process. Consequently, the video encoder itself can be exchanged and coding efficiency improvements of upcoming video encoders can be inherited easily. Compared to other preprocessing approaches, our proposed method works non-destructive and introduces no extra image degradations or drawbacks. It can be combined with arbitrary ROI detectors as long as a list of pixels to be encoded as ROI is provided to our preprocessor. The versatility of the framework will be demonstrated for different video encoders by combining the moving camera GMC and ROI detection system from [13] with our proposed approach. Scenarios with a static camera like teleconferencing or building entrance observation are considered as special case with a global motion of zero. We will prove the benefit of our system also for those scenarios. Furthermore, we will demonstrate that our general ROI coding framework as well as the video encoding itself has no negative influence on subsequent computer vision tasks (e. g. people tracking).

Our contribution is a joint pre- and postprocessing ROI-based video coding framework which can be combined with arbitrary ROI detectors. The framework is able to achieve a similar coding performance like specifically adapted video codecs, but with off-the-shelf hardware or software video codecs (without any modifications).

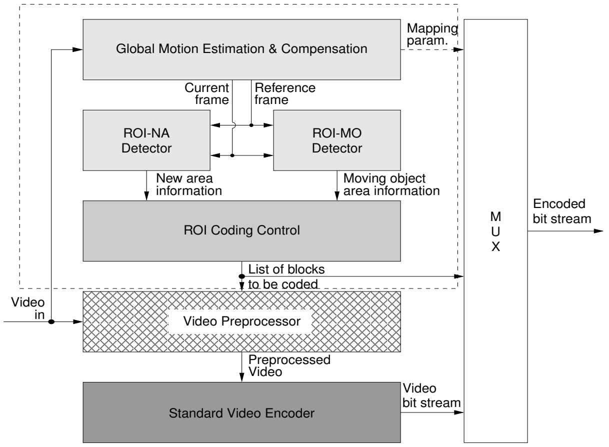


Fig. 1: Example block diagram of the *general ROI coding system*. The ROI detection system (dashed lined box, based on [13]) can be exchanged without loss of generality as long as a list of coding blocks is passed to the *Video Preprocessor* (crisscrossed).

The remainder of this paper is organized as follows: Section 2 reviews the GMC-based ROI detection and coding reference system and elaborates necessary modifications of the video encoder for this existing approach. Section 3 describes the proposed joint pre- and postprocessing framework in detail. Section 4 evaluates the experimental results before Section 5 concludes the paper.

2. ROI-BASED CODING EXTENSION FOR AVC

As a reference system we use the GMC ROI detection and coding systems from [13] because it is capable to retain subjectively high image quality over the entire image while simultaneously providing low bit rates which is unique compared to other ROI-based coding systems. In order to properly reconstruct local motion not covered by the global motion model, moving objects are additionally detected with a simple difference image-based detection approach. A thorough description of an highly accurate moving object detector for this system can be found in [12]. The unchanged blocks of the reference system are marked with a dashed lined box (Fig. 1).

Assuming a planar landscape, one frame $k-1$ can be projected into the consecutive frame k employing a projective transform with 8 parameters: $\vec{a}_k = (a_{1,k}, a_{2,k}, \ldots, a_{8,k})^\top$. The pixel coordinates from the preceding frame $\vec{p}_{k-1} = (x_{k-1}, y_{k-1})$ are mapped to the position $\vec{p}_k = (x_k, y_k)$ of the current one with the mapping parameter set $\vec{a}_k$:

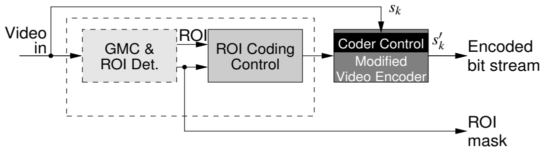
$$F(\vec{p}, \vec{a_k}) = \left(\frac{a_{1,k} \cdot x + a_{2,k} \cdot y + a_{3,k}}{a_{7,k} \cdot x + a_{8,k} \cdot y + 1}, \frac{a_{4,k} \cdot x + a_{5,k} \cdot y + a_{6,k}}{a_{7,k} \cdot x + a_{8,k} \cdot y + 1} \right)^\top . \quad (1)$$

To determine $\vec{a}_k$, first, a global motion estimation is performed: Harris Corners are used to define a set of good-to-track feature points in frame k. A *Kanade-Lucas-Tomasi* (KLT) [15] feature tracker is employed afterwards to relocate the feature positions in frame $k-1$ and thereby generate a sparse optical flow between the frames. Outliers such as false tracks are removed and the final mapping parameter set $\vec{a}_k$ is determined by *Random Sample Consensus* (RANSAC) [16]. This mapping parameter set is used for the global motion compensation in the first block in the block diagram of the coding system (Fig. 1). The mapping parameter set $\vec{a}_k$ is further employed to determine areas contained in the current frame k, but not in the previous one $k-1$ (new areas) by the *ROI-NA Detector*. These determined ROI is passed to the *ROI Coding Control* block which basically assigns the pel-wise ROI to corresponding blocks for video coding. Any square block (i. e. a *macroblock* in AVC or a *Coding Tree Unit* in HEVC, respectively) containing at least one pel new area is commonly encoded whereas any non-ROI block is not encoded (set to *skip mode*) by an externally controlled modified video encoder, e. g. a modified AVC [13] or a modified HEVC [12] encoder. As mentioned before, common video encoders have no interface to externally control or overwrite the coding mode, and thus intricate modifications are necessary in order to realize such control. By this external video coder control, the bit rate is significantly reduced compared to a not externally controlled video encoder while standard conformance of the bit stream can be retained. The mapping parameter set $\vec{a}_k$ has to be transmitted e. g. by encapsulating the 8 parameters per frame in *Supplemental Enhancement Information* (SEI) messages.

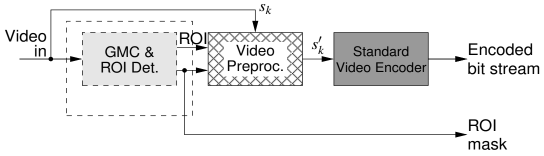
By applying the same global motion compensation in the decoder, the background is reconstructed by padding all new area stripes into a panorama image first (mosaicking). Afterwards it can cut frames according to their original positions and concatenate them as a video sequence. Finally, moving objects can be inserted in the video sequence (if available). We call this postprocessing method *quasi non-destructive* because the GMC background might show motion parallax effects for ground objects not matching the global plane [17].

3. PROPOSED GENERAL ROI CODING SCHEMES

The idea of our proposed *general ROI coding system* is to dedicate any coding decision to the highly optimized encoder internal *Rate Distortion Optimization*. Accordingly, we have to ensure that our input video stream contains only relevant information, needed at the decoder for reconstruction. Image areas, which are irrelevant for reconstruction (non-ROI) are replaced such that the encoder can encode the entire image as efficient as possible. While the ROI detection system stays unchanged, and can be exchanged in order to meet special application requirements, in common ROI coding systems the video codec itself has to be modified according to the above explanations. For a system employing the proposed joint pre- and postprocessing framework, off-the-shelf video encoders and decoders can be utilized (Fig. 1). In common video coding, the output is optimized to be as similar to the input as possible. In contrast to that, we accept the modifications of the input signal by the general ROI preprocessing by purpose.

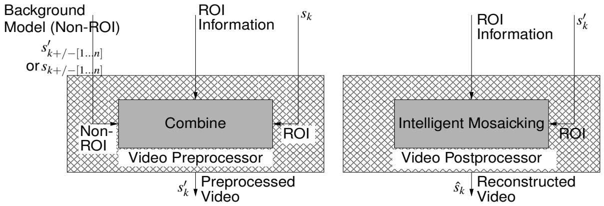


(a) Specifically adapted ROI Coding System



(b) Proposed General ROI Coding System

Fig. 2: Entire encoding scheme from the camera to the encoded bit stream, (a) with a modified video encoder, (b) with an unmodified video encoder (dashed white box: unchanged GMC & ROI detection from the reference system).



(a) Preprocessing (b) Postprocessing

Fig. 3: Image composition by pre- and postprocessing as introduced by the proposed general ROI coding. Subsequent image reconstructions have to be applied like required by the specific ROI detection and coding system (e. g. GMC).

In the reference systems described in the last section, every coding block, which is dedicated for being encoded in *skip mode* by a ROI detector, is forced to be skipped directly in the *modified* video encoder by external control (Fig. 2a). We propose to leave the video encoder itself unaltered (Fig. 2b) but replace non-ROI image blocks in a preprocessing (PP, Fig. 3a). ROI image blocks stay unmodified. We distinguish two operation modes, where each non-ROI block is replaced by:

Mode 1: the corresponding block from the (temporally) preceding frame. This mode aims at utilizing coding tools for unchanged content (e. g. skip mode).

Mode 2: a black block. This mode aims at utilizing coding tools for monochrome areas (skip, DC intra prediction).

Both modes are based on the fact that non-ROI areas are discarded in the postprocessing step anyway. To reconstruct the entire video sequence, for mode 2 it is mandatory to use a postprocessing after the actual video decoding using any standard conforming decoder. The postprocessing consists of the reconstruction of the decoded image blocks similar to the preprocessing (Fig. 3b). We would like to emphasize again that both, pre- and postprocessing are completely video codec independent. Consequently, no restrictions for special video codecs are introduced by the preprocessing nor by the postprocessing of our framework. However, since only the video codec was replaced by our framework, subsequent image reconstructions have to be applied like required by the specific ROI detection and coding system. For the reference system, the GMC parameters have to be taken into account for proper image reconstruction in a subsequent ROI decoder, i. e. the panorama image generation and video reconstruction [14]. This joint pre- and postprocessing framework offers several advantages compared to other ROI coding systems which rely on image degrading prefiltering [2] or usage of a modified video encoder as e. g. in [6]: No image degradation is introduced by our general ROI coding – beside the errors introduced by the video encoder itself, i. e. quantization errors for lossy coding modes. Additionally, any coding tool of any video codec can be used without restrictions (e. g. *skip mode* for AVC/HEVC).

Our proposed general preprocessing framework can also be used for application scenarios like surveillance operations with static cameras (e. g. entrance or parking lot observation), video conferencing or any other application where large parts of the frame are static and an arbitrary ROI detector distinguishes static and non-static parts of each frame.

3.1. Inherent Noise Removal of the General ROI Coding

Modern hybrid video encoders like AVC or HEVC already consist of very efficient coding modes also for encoding static parts of a frame (e. g. *direct mode* in AVC, *merge mode* in HEVC). These modes are most efficient for noise free use-signals s. However, camera captured signals contain additive white Gaussian noise (AWGN) n with the noise power $P(n)$. Thus, the superimposed signal $s+n$ is the input of the video encoder. Assuming a perfect motion compensation, the noise of the reference frame used for motion compensation n_{k-1} has to be removed and the noise n_k of the current frame has to be added to the prediction error signal. The resulting noise power accumulates to $2 \cdot P(n)$ in the prediction error signal, leading to higher bit rates than required for encoding s only.

For our proposed general ROI coding, the noise has to be encoded only once at the first occurrence of each coding block, since in all subsequent frames a pelwise copy is inserted from the temporally preceding frame (mode 1) or the block is entirely replaced by a black block (mode 2). Consequently, the resulting bit rate will decrease.

4. EXPERIMENTAL RESULTS

We evaluated the proposed general ROI coding framework by comparing its coding efficiency with common, non-modified video encoders as well as specifically adapted, modified encoders. For the latter we used the AVC-based implementation from Section 2, called AVC-skip [14], and a similar implementation based on the HEVC reference encoder *HEVC Test Model* (HM) 16.2, called HEVC-skip [12]. To evaluate our general ROI coding framework, we applied both proposed coding modes to the test sequences and fed the resulting videos in the unmodified video encoders *x264* v0.78 [18] (AVC), HM 16.2 (HEVC) and VP9 v1.3.0 [19] (WebM Project).

(a) Common AVC coding

(b) General ROI coding

Fig. 4: Encoded frame (a) with AVC and (b) after processing with reference GMC-based ROI detection system and general ROI video coding at the same bit rate (500 kbit/s). In (b) much more details are preserved, e. g. for the car trailer (red circle).

We used two HDTV resolution (1920 × 1080, 30 fps) aerial test sequences with a moving camera from the publicly available data set *TNT Aerial Video Testset* (TAVT) [12]: the efficiently to encode *350 m sequence* with a ground resolution of 43 pel/m and the highly noisy *1500 m sequence*, providing 10 pel/m. In order to demonstrate the benefits of the general ROI coding for scenarios with (quasi) static cameras, we additionally considered the publicly available *BIWI Walking Pedestrians dataset* (640 × 480, 25 fps) [20] and a self-recorded video conferencing sequence (1280 × 720, 10 fps).

The ROI detectors are the unmodified detectors from Section 2 (new area and moving object detector). Major parts of the frames of the aerial sequences were selected to be non-ROI by the ROI-detectors, depending on the camera movement relative to the ground and the amount of areas containing moving objects. By using the proposed general ROI preprocessing, those non-ROI parts are replaced by their corresponding areas from the preceding frame and stay unchanged (mode 1) or are replaced by black blocks (mode 2). For the static camera sequences, no global motion is present and thus only moving objects are detected. These preprocessed videos are fed into the video encoders. The resulting image quality using the proposed system is similar to the one achieved with the specifically adapted encoders for ROI and non-ROI areas. Compared to common video coding, the subjective quality stays high over the entire frame for all ROI-based systems even for low bit rates as more bits can be allocated to ROI areas (Fig. 4, car trailer in red circle) due to the bit savings in non-ROI regions.

For our evaluation, we adjusted all competitors to match the *Peak Signal-to-Noise Ratio* (PSNR) of the luminance (Y) channel provided by the modified AVC-skip encoder at a fixed quantization parameter (QP $= 25$) as close as possible. Only

Table 1: Gains of the proposed *general ROI preprocessing* (PP) with block insertion from previous frames (mode 1) or insertion of black (PP-black, mode 2) for non-ROI areas compared to a modified video coder (skip). AVC encoder: *x264*, v0.78 [18], HEVC encoder: HM 16.2 , LD profile, VP9 encoder: *WebM Project VP9 enc. v1.3.0 [19].* Negative numbers are gains compared to the reference (*Ref.*). The proposed framework provides similar (or better) coding performance than modified video coders (skip).

	350 m sequence, 43 pel/m, 821 frames, PSNR ≈ 38.9 dB [13]		1500 m sequence, 10 pel/m, 1571 frames, PSNR ≈ 37.6 dB [13]		BIWI Walking Pedestrians dataset, 750 frames, PSNR ≈ 42.2 dB [20]		Video conferencing sequence, 100 frames, PSNR ≈ 43.9 dB	
	Bit rate in kbit/s	Diff. in %	Bit rate in kbit/s	Diff. in %	Bit rate in kbit/s	Diff. in %	Bit rate in kbit/s	Diff. in %
AVC	9287	*Ref.*	13560	*Ref.*	270	*Ref.*	1690	*Ref.*
AVC-skip	943	−89.8	967	−92.9	161	−40.4	1229	−27.3
AVC-PP (mode 1)	785	−91.6	667	−95.1	182	−32.6	1080	−36.1
AVC-PP-black (mode 2)	709	−92.4	783	−94.2	222	−17.8	1134	−32.9
HEVC	5568	−40.0	11901	−12.2	210	−22.2	1397	−17.3
HEVC-skip	558	−94.0	614	−95.5	118	−56.3	874	−48.3
HEVC-PP (mode 1)	644	−93.1	686	−94.9	140	−48.2	911	−46.9
HEVC-PP-black (mode 2)	562	−93.9	618	−95.4	161	−40.4	930	−45.0
VP9	5387	−42.0	12969	−4.4	237	−38.9	1370	−18.9
VP9-PP (mode 1)	802	−91.4	829	−93.9	152	−43.7	670	−60.4
VP9-PP-black (mode 2)	577	−93.8	621	−95.4	165	−38.9	897	−46.9

the luminance values within ROI areas (ROI-Y-PSNR) are considered, similar to [21, 12]. Herby it is assumed that errors in non-ROI areas are irrelevant because non-ROI is reconstructed by external means as part of the postprocessing at the decoder anyway. If it was not possible to exactly match the ROI-Y-PSNR, we linearly interpolated in between the neighbored rate points, which is justified by a relatively linear curve between two neighbored rate points in a rate-distortion plot.

Table 1 shows the coding gains (negative numbers) relative to the AVC reference (*Ref.*) as marked in the columns. With the AVC-skip video encoder and for the HDTV aerial video sequences (with a moving camera) the bit rate is decreased by about 91 % on average. For the lower resolution test sequences the coding efficiency gain is smaller due to the proportionally higher amount of signaling data [22]. Using our proposed general ROI coding framework and the unmodified AVC video encoder, we achieve an even higher coding efficiency for the aerial sequences and the video conferencing sequence. Both proposed coding modes provide similar coding performances than the specifically adapted AVC-skip encoder. The optimal operation mode depends on the sequence characteristics.

Using our proposed general ROI coding framework, we achieve comparable bit rates to specifically adapted ROI encoders, without the need of encoder modifications. As expected, the HEVC-based systems outperform all AVC-based systems, especially for the high resolution video sequences. All HEVC-based ROI encoders achieve a bit rate of ~600 kbit/s for a perceptional "good quality" of about 38 dB and the full HDTV resolution sequences. This is very low compared to the bit rates of the unmodified HEVC codec between 5500–11900 kbit/s. Our proposed general ROI coding framework achieves comparable but slightly worse coding efficiencies of 12–14 % (or 644–686 kbit/s in absolute terms) using mode 1 (block copy) compared to the specifically adapted HEVC-skip implementation. In contrast to that we nearly reach the latter coding efficiency for general ROI coding in mode 2 (replace non-ROI blocks by black) with virtually no loss (0.7 % or 4 kbit/s for both aerial sequences, each). Whereas the bit rates of the unmodified VP9 encoder are within the same range than those of the unmodified HEVC encoder for our test set, HEVC-PP (mode 1) clearly outperforms VP9-PP (mode 1), except for the *Video conferencing sequence*. Depending on individual sequence characteristics, the VP9-PP-black (mode 2) bit rates are comparable with those of HEVC-PP-black.

To demonstrate that our general ROI coding framework does not negatively affect subsequent computer vision algorithms, we evaluated the performance degradation of a pedestrian detection and tracking algorithm on the differently coded *BIWI Walking Pedestrians* sequence. We achieved similar detection results using a simple *Gaussian Mixture Model* (GMM) based moving object detector and a similar people tracking accuracy in the tracking-by-detection framework from Henschel et al. [23]. We would like to point out that no additional errors are introduced by our proposed general ROI coding framework compared to a special ROI video encoder.

5. CONCLUSIONS

This paper aims towards a general *Region of Interest* video coding framework which is independent from any encoder modifications. Thus, it can be combined with arbitrary video encoders off-the-shelf and saves complicate and time-consuming encoder modifications. Using the proposed joint pre- and postprocessing framework, an improved coding efficiency of upcoming video codecs can be inherited easily. Based on the decision of arbitrary ROI detectors, non-ROI areas are replaced either by already known pixels or blocks, or by black pixels or blocks, leaving the optimal encoding strategy to the sophisticated video encoder internal rate-distortion optimization. Using our general ROI coding framework we retain the same image quality compared to a similar ROI detection and coding system employing a modified video codec. We demonstrate that the coding performance is similar to specifically adapted video codecs for different applications like (aerial) surveillance or video conferencing. Employing our general ROI coding, we achieve bit rates of less than 700 kbit/s for full HDTV resolution aerial video sequences at 30 fps at subjectively high quality with unmodified video codecs.

We encourage to combine our framework with highly optimized ROI detectors of your choice and provide a possibility to considerably reduce the bit rate without impairing subsequent computer vision applications.

6. 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),” *Circ.&Syst.f. Video Technology, IEEE Transact. on*, vol. 22, no. 12, pp. 1669 –1684, Dec. 2012.

[2] Linda Karlsson, Mårten Sjöström, and Roger Olsson, “Spatio-Temporal Filter for ROI Video Coding,” in *Proceedings of the 14th European Signal Processing Conference (EUSIPCO)*, Sept. 2006.

[3] AVC, *Rec. ITU-T H.264 & ISO/IEC 14496-10 MPEG-4 Part 10: Adv.Video Coding (AVC)-3rd Ed.*, ISO/IEC&ITU-T, Geneva, Switzerland, July 2004.

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

[5] Yang Liu, Zheng Guo Li, and Yeng Chai Soh, “Region-of-Interest Based Resource Allocation for Conversational Video Communication of H.264/AVC,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 134–139, Jan 2008.

[6] Peiyin Xing, Yonghong Tian, Tiejun Huang, and Wen Gao, “Surveillance Video Coding with Quadtree Partition based ROI Extraction,” in *Proc. of the IEEE Pict.Cod.Symp.(PCS)*, Dec. 2013, pp. 157–160.

[7] B. Dey and M.K. Kundu, “Robust Background Subtraction for Network Surveillance in H.264 Streaming Video,” *IEEE Transa. on Circ. and Syst. f. Video Technol.*, vol. 23, no. 10, pp. 1695–1703, Oct. 2013.

[8] H. Sakaino, “Video-Based Tracking, Learning, and Recognition Method for Multiple Moving Objects,” *IEEE Transact. on Circ. and Syst.f.Video Technol.*, vol. 23, no. 10, pp. 1661–1674, Oct. 2013.

[9] Xianbin Cao, Jinhe Lan, Pingkun Yan, and Xuelong Li, “KLT Feature Based Vehicle Detection and Tracking in Airborne Videos,” in *6th Int. Conf. on Image and Graphics (ICIG)*, Aug 2011, pp. 673–678.

[10] Jinman Kang, I. Cohen, G. Medioni, and Chang Yuan, “Detection and Tracking of Moving Objects from a Moving Platform in Presence of Strong Parallax,” in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2005, vol. 1, pp. 10–17.

[11] Michael Teutsch, *Moving Obj. Detect. and Segment. for Remote Aerial Video Surveillance*, Ph.D. thesis, Karlsruhe Inst. of Technology (KIT), Germany, 2014.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] VideoLAN Organization, “x264,” 2009.

[19] The WebM Project, “VP9,” 2013.

[20] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, “You’ll never walk alone: Modeling social behavior for multi-target tracking,” in *IEEE Int.Conf.onComp.Vis.*, Sept. 2009, pp. 261–268.

[21] 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.

[22] Sven Klomp, “Decoderseitige Bewegungsschätzung in der Videocodierung,” VDI Verlag GmbH, Düsseldorf, 2012, ISBN 978-3-18-382010-8, Fortschritt-Berichte VDI: Reihe 10, Informatik, Kommunikation ; 820.

[23] Roberto Henschel, Laura Leal-Taixé, and Bodo Rosenhahn, “Efficient Multiple People Tracking Using Minimum Cost Arborescences,” in *German Conf. on Pattern Recognition (GCPR)*, Sept. 2014.