

# H.264 Sensor Aided Video Encoder for UAV BLOS Missions

Cesario Vincenzo Angelino<sup>1</sup>, Luca Cicala<sup>1</sup>, Marco De Mizio<sup>1</sup>, Paolo Leoncini<sup>1</sup>,  
E. Baccaglioni<sup>2</sup>, M. Gavelli<sup>2</sup>, N. Raimondo<sup>2</sup>, and R. Scopigno<sup>2</sup>

<sup>1</sup> CIRA, The Italian Aerospace Research Center, Capua, Italy  
{c.angelino,l.cicala,m.demizio,p.leoncini}@cira.it

<sup>2</sup> Istituto Superiore Mario Boella, Torino, Italy  
{baccaglioni,gavelli,raimondo,scopigno}@isbm.it

**Abstract.** This paper presents a new low-complexity H.264 encoder, based on x264 implementation, for Unmanned Aerial Vehicles (UAV) applications. The encoder employs a new motion estimation scheme which make use of the global motion information provided by the onboard navigation system. The results are relevant in low frame rate video coding, which is a typical scenario in UAV behind line-of-sight (BLOS) missions.

## 1 Sensor Aided Global Motion Estimation

Global motions in a video sequence are caused by camera motion, which can be modeled by parametric transforms. The process of estimating the transform parameters is called global motion estimation (GME) and compensation (GMC). A widely used global motion model is perspective model with 8 parameters. This model is suitable to represent the motion field for far backgrounds. In order to reduce the computational burden of GME, several works on the use of external motion sensors have been recently proposed. In [2], for example, the authors studied the application of GME techniques that process sensor data (gyroscopes, accelerometers and magnetometers) to video coding. The focus is on mobile devices and the complexity reduction of motion estimation reflects into increasing battery life.

Video coding systems currently employed in NATO military UAV applications meet the STANAG 4609 standard [5], which refers to MPEG4 and Motion JPEG2000 standards for motion imagery coding. However, these standards are for general purpose applications and do not rely on the particular features of aerial video sequences. Indeed, the latter present some peculiarities in terms of observed scene geometry and point of view. In particular, the observed scene is almost static (the camera movement is dominant) and generally easy to model (usually approximated with a plane). Moreover, the point of view (camera position and orientation) is far from the scene and approximately known since it can be obtained from the navigation system.

UAV are used typically for surveillance, for both military and civil applications such as the remote monitoring of hostile area, coasts and boundaries

surveillance, disaster management (earthquakes, floodings, fires, etc.). In such applications high defined images are needed. Furthermore, an UAV video coding system should meet both bandwidth (BLOS transmission) and computational (battery life) constraints.

Sensor aided video coding for UAV application gained attention in the last few years. Rodriguez et.al [3] use the available global motion information to simplify block ME in a MPEG-4 encoder. Although their approach reduces the complexity of a standard video encoder, transmitting the global motion information instead of the motion vectors obtained analyzing the video frames. M. Bhaskaranand and J. D. Gibson [4] proposed a global motion compensated low-complexity encoder with global parameters obtained from the camera mount system. Comparing the performance of the proposed encoder with that of a H.264, they showed that for videos typical of UAV reconnaissance, the proposed encoder achieves better R-D performance at lower bit rates. No particular attention has been payed in previous works to low overlap frame sequences and BLOS applications.

## 2 A Sensor Aided Encoder Based on x264

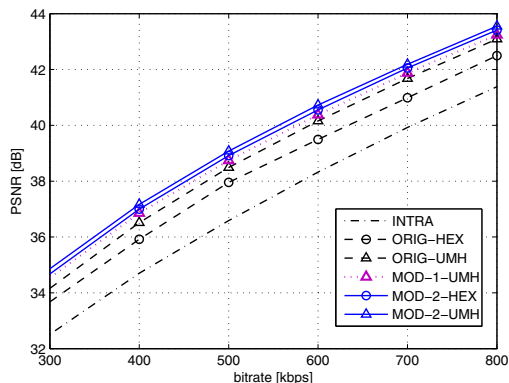
In this paper GME is performed using external position and orientation sensors. The observed scene is supposed to be approximated with a plane. Based on this simple hypothesis, a prediction of the motion field is obtained calculating the homography between the reference and the predicted frame. This model has been employed in the H.264 motion estimation system through its open source implementation x264<sup>1</sup>.

If motion vector predictors and refinement stage can receive additional information from sensors about the motion flow, it could be of great benefit for the whole encoding stage. The proposed work allows to choice the prediction method (inter/intra) for each macroblock and initializing the motion vectors by substituting the non-standard x264 predictors with predictors derived from external data. Furthermore, we also address the case of different refinement policies to improve the overall motion estimation process.

Especially for aerial sequences, a strong global motion flow is present in each frame and for each macroblock. The possibility of providing the encoder with some preferential motion directions and significant frames to be used as reference can thus leverage the encoding process. Our work mainly addresses two specific points: the decision on the frame type that has to be encoded (Intra, Predicted, Bi-Predicted) and the correspondent best frame reference and the initialization of the motion field to generate the predictors which will be adopted in the subsequent macroblock analysis.

The module developed to perform these operations is called initialization module. The proposed changes allow to interface this module with the x264 library. It receives as input two lists of metadata information. The first list contains metadata associated with the frames that can be used as a reference by a P frame in accordance with the number of reference frames which can be used.

<sup>1</sup> We used the version x264 0.128.30M 5d25de2.



**Fig. 1.** Quality (PSNR) vs output bitrate for the original x264 library and the proposed architecture

The second list contains metadata associated with the frame for which the decision has yet to be made. Wherever a frame is of P- or B- type, the motion field is initialized and MV predictors are provided for each macroblock of size  $4 \times 4$ . The preliminary stage of this proposed encoding system allows to initialize crucial parameters (frame type, MV, ...) with the available sensor information in order to speed up the whole encoding stage. In such way, a fully H.264/AVC-compliant video stream is generated and can be later decompressed and reconstructed by any compatible decoder. In fact, the proposed pre-encoding process exploits external data in the sense that it initializes the MV predictors and the frame type decision methods leaving unaltered the subsequent stages.

### 3 Performance Evaluation in BLOS Scenario

The focus of the tests is on low frame rate video coding, which is a typical scenario in BLOS missions. Test sequences have been generated by means of a computer graphic simulation tool for image acquisition over photorealistic and 3-dimensional scenarios. Acquisition camera is installed under the UAV in a looking-down setup. Simulated data reflects non-ideal characteristics of sensors in terms of sampling rate, delay, noise, available bandwidth. The sequence generated for encoding comparison in the current work is made at 0.5 fps, at a flying altitude of approx 200 meters over an urban landscape. The resulting frame overlap is lower than 60%.

In Fig.1 we report the results in terms of output quality (Peak Signal-to-Noise Ratio expressed in dB) versus the encoding bitrate (in kbps) and it is possible to highlight the benefits of the proposed approach. In the reported results, the curves labeled ORIG refer to the original x264 libraries in which the GOP structure is left to default optimization policies. Both UMH and HEX refinement, as implemented in x264, have been tested for the original encoder.

In general, for the tested sequences, short GOP are created so a significant number of Intra-coded picture are generated. The curve labeled INTRA refers to the intra-code mode i.e. all frames are I frames. Curves labeled MOD refer to the proposed libraries which exploit data from sensors to initialize motion fields.

The proposed algorithm outperforms traditional H.264/AVC coding because the motion estimation fails when the frame overlapping is very low. The GOP structure  $I, B_{ref}, B, I$  (MOD-1) outperforms traditional coding and the gain is even higher if the encoder adopts long sequences of P frames (MOD-2). Around 500 kbps, for instance, when using data received from sensors to initialize MV predictors, the gain can be approximately 1.4 dB when comparing the approach MOD-2 vs the original x264 with hexagonal motion refinement. Also in the case of UMH refinement (which is usually not adopted in real-time application) there is a the gain of 0.6 dB. These results are coherent with the fact that the subsequent frames overlapping is low and then the traditional x264 motion estimation algorithms fail when looking for similar MB in previous frames. For this reason, x264 rate-distortion policies decide to code frames in Intra-mode and this requires higher bandwidth. As a performance bound, in the figure we also report the encoder performance in Intra-mode (still image coding). As can be seen, the output quality is lower than the other schemes because the bandwidth required to code each frame as an I frame is significant.

Concerning encoding time, it is worth noting that motion estimation is significantly faster when it is possible to include in the video compressor system data acquired from sensors. For example, for a total output bitrate of 400 kbps, both for HEX and UMH refinement, a 25% speed-up in encoding time is achieved.

In conclusion, our experiment results show better performances of the proposed encoder w.r.t. standard H.264 encoder in low frame rate coding scenario, which is typical for UAV BLOS mission. Future works will address a more general model for the observed scene.

## References

1. Collin, R.: Video Encoding using Motion Data Nokia. International patent n. WO/2009/156965 (December 30, 2009)
2. Chen, X., Zhao, Z., Rahmati, A., Wang, Y., Zhong, L.: Sensor-Assisted Video Encoding for Mobile Devices in Real-World Environments. *IEEE Trans. on Circuits and Systems for Video Technology* 21(3), 335–349 (2011)
3. Rodríguez, A.F., Ready, B.B., Taylor, C.N.: Using telemetry data for video compression on unmanned air vehicles. In: *AIAA Guidance, Navigation, and Control Conference and Exhibit* (August 2006)
4. Bhaskaranand, M., Gibson, J.D.: Low-complexity Video Encoding for UAV Reconnaissance and Surveillance. In: *Proceedings of 2011 Military Communications Conference - Track 4 - Middleware Services and Application* (2011)
5. NATO STANAG 4609: Digital Motion Image Format, 3rd edn. (2009)