# Region-of-interest based rate control for UAV video coding

**ZHAO Chun-lei** (赵春蕾)[1,2]*, **DAI Ming** (戴明)[1], **and XIONG Jing-ying** (熊晶莹)[1,2]

*1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China*

*2. University of Chinese Academy of Sciences, Beijing 100049, China*

(Received 7 March 2016)



To meet the requirement of high-quality transmission of videos captured by unmanned aerial vehicles (UAV) with low bandwidth, a novel rate control (RC) scheme based on region-of-interest (ROI) is proposed. First, the ROI information is sent to the encoder with the latest high efficient video coding (HEVC) standard to generate an ROI map. Then, by using the ROI map, bit allocation methods are developed at frame level and large coding unit (LCU) level, to avoid inaccurate bit allocation produced by camera movement. At last, by using a better robustness $R$-$\lambda$ model, the quantization parameter ($QP$) for each LCU is calculated. The experimental results show that the proposed RC method can get a lower bitrate error and a higher quality for reconstructed video by choosing appropriate pixel weight on the HEVC platform.

**Document code:** A **Article ID:** 1673-1905(2016)03-0216-5

**DOI** 10.1007/s11801-016-6052-z

Unmanned aerial vehicles (UAVs) equipped with surveillance systems have been widely used in military missions and civilian domains[1,2]. The huge data has posed big challenges for the video compression systems on the UAVs. Rate control (RC) plays a major role in video coding. By using RC, better video quality can be achieved without violating the constraints imposed by the available channel bandwidth or memory requirement. Many valuable results about RC algorithms based on region-of-interest (ROI) have been published[3-6]. The methods in Refs.[3]—[6] are developed for different preceding compression standards, which are not suitable for the latest high efficiency video coding (HEVC) standard with the flexible picture partition and some other techniques[7-9]. Moreover, they are designed for low- resolution videos, which are not suitable for high-resolution and high-frame-rate video services, such as UAV videos. Although some RC methods[10,11] are proposed for HEVC at high resolutions, these methods compute the quantization parameter ($QP$) with a quadratic model, which may degrade the video quality due to the inaccurate bit allocation, and has been replaced by the $R$-$\lambda$ model with better robustness for HEVC[12,13]. ROI based RC approaches with $R$-$\lambda$ model[14,15] have already been applied in the video conferencing application. However, objects' diversity and uncertainty of an aerial video in various missions make ROI detection more difficult than the single type ROI in video conferencing. Furthermore, the ratio of bit allocation between every picture in one group of picture (GOP) is inflexible and predetermined in the previous schemes, which cannot meet the videos captured by high speed UAVs.

In this paper, we design a novel RC scheme which is appropriate for the high-resolution and high-frame-rate UAV video coding in HEVC. It uses ROI map for bit allocation at frame level and large coding unit (LCU) level. The scheme allocates more bits to the frames with a larger ROI at frame level. In this way, the larger ROI can get enough coding bits to improve the quality of ROI. At LCU level, different numbers of bits are dynamically allocated to LCUs with the ROI map. It can avoid inaccurate bit allocation from local object motions or global camera motions. Then the $QP$ for each LCU is calculated with the $R$-$\lambda$ model. Experimental results show that in HEVC, the proposed ROI based RC scheme can significantly improve the performance compared with state-of-art methods.

Fig.1 presents the overview of the proposed scheme. Firstly, the scheme generates an ROI map by using the objects detected in UAV videos. There are two classes of methods can be utilized to realize object detection. One is automatic detection based on computer vision, and the other is manual detection based on interactive information. After that, the RC module can make use of the generated ROI map to perform two efficient bit allocation methods at frame level and LCU level. Then $R$-$\lambda$ model is applied to calculate the $QP$.

* E-mail: ZhaoCL@hotmail.com

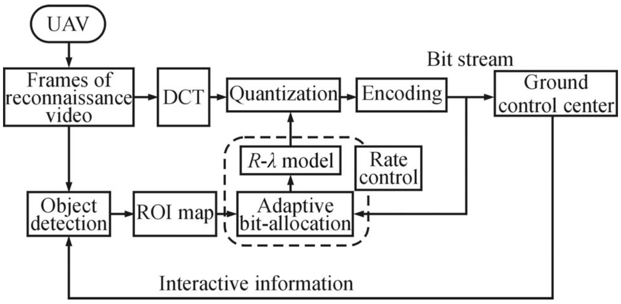



**Fig.1 The framework of the ROI based rate control**

The previous ROI based RC schemes are designed for the conversational applications, in which ROI detections are mostly implemented for face detection or motion detection. The object detection method for specific application limits the applied scope and makes the detected ROI fairly inaccurate for non-conversational scenarios. Different from the single type object in conversation, many regions may be the ROI (such as pedestrian, road and vehicles) in an aerial video. The diversity and uncertainty of objects in various missions has posed big challenges to the ROI detection in the UAV videos.

Fortunately, many valuable results about salient objects detection methods have been published. A comprehensive survey of salient object detection can be found in Ref.[16]. Cheng et al[17] designed a good generic objectness measure method with high object detection rate. Moreover, the development of the researches on target detection for UAV in different applications[18] brings convenience to multi-objective extraction for UAV videos. On the other hand, the interaction and the real-time performance are the two key characteristics for the UAV system. Targets can be detected with simple interaction with users or other equipments in the UAV system. The encoder obtains ROI with high quality by the manual detection to avoid a large amount of calculation and inaccuracy. After detecting ROI in one frame, pixels in ROI are assigned with weight value $k$, and pixels in non-ROI are assigned with a weight value 1, which can be expressed as

$$RM(w,h)=\begin{cases}k, & (w,h)\in \text{ROI}\\ 1, & (w,h)\in \text{non-ROI}\end{cases}, \tag{1}$$

where $RM$ is the ROI map at pixel level having the same size with the original frame, and ($w$, $h$) is the location of the pixel in the frame. Then, the ROI map is sent to HEVC encoder to guide an ROI based RC with an effective bit allocation method.

Since a human observer is the end-user of UAV video, enhancing the quality of ROI is an important task to improve the user satisfaction. As the overall coding and transmission resources are limited, allocating more bits on ROI while decreasing the quality of non-ROI is reasonable. Thus, the necessary bit stream can be reduced by allocating fewer bits for the less important parts. We introduce the proposed ROI based RC scheme with HEVC encoder for UAV videos. Bit allocation methods are applied in this scheme. At frame level bit allocation, we allocate bits to each frame in the GOP according to the frame weight measured by the ROI map at frame level. Similarly, RC scheme allocates bits to each LCU in the frame in accordance with their importance measured by the ROI map at LCU level.

In the traditional $R$-$\lambda$ method[12], the ratio of bit allocation between each picture in one GOP is determined by bit per pixel ($BPP$). Obviously, $BPP$ value ignores the characteristics of video content, since it only depends on the bandwidth, frame rate and resolution. Hence, the $BPP$ value used as a weight performs poorly at frame level bit allocation. Then, the problem of frame level bit allocation can be solved by the ratio between the Lagrangian multiplier ($\lambda$) values which are derived from $R$-$\lambda$ method[13]. However, the ratio of $\lambda$ within each GOP is fixed. It cannot adapt to the changeful aerial video well, especially when the UAV flies rapidly or works in zoom state. When working in short-focus mode for the target positioning, the ROI size is smaller. We can reduce the allocation bits to improve the compression ratio. On the contrary, when working in long-focus mode to identify and observe target details, the ROI size is larger. We should increase the allocation bits to improve the satisfaction of observer. Therefore, in order to meet the requirement of the UAV application, the frame level bit allocation problem can be solved by the ratio between the ROI maps at frame level in a GOP. The ROI map at frame level $F\omega[i]$ can be expressed as

$$F\omega[i]=\frac{\sum_{h=1}^{H}\sum_{w=1}^{W}RM(w,h)}{H\times W}=$$
$$\frac{N_{\text{ROI}}[i]\times k+(H\times W-N_{\text{ROI}}[i])}{H\times W}=$$
$$\frac{N_{\text{ROI}}[i]}{H\times W}(k-1)+1, \tag{2}$$

where $F\omega[i]$ represents the bit allocation weight of the $i$th frame in the current GOP, $H$ and $W$ are the height and width of one frame, and $N_{\text{ROI}}[i]$ is the number of pixels in the ROI of the current frame. Then RC scheme allocates bits adaptively for each frame in the GOP according to their importance measured by $F\omega[i]$. More bits are allocated to more important frames to improve the quality of video. Fewer bits are allocated to less important to improve the compression ratio. Then, the target bits for the $i$th frame in the current GOP can be achieved by

$$T_{\text{f}}[i]=\frac{F\omega[i]}{F\omega_{\text{rem}}}\times T_{\text{GOP,rem}}, \tag{3}$$

where $T_{\text{f}}[i]$ is the target bits to be allocated for the $i$th frame in the current GOP, $T_{\text{GOP,rem}}$ is the remaining bits used to encode the rest of frames in the current GOP, and $F\omega_{\text{rem}}$ is the sum of weight for the rest frames in the current GOP.

Mean absolute difference (*MAD*) of the LCU at the same position in the previously decoded frame is used as a weight at LCU level bit allocation in the typical RC algorithm[12]. However, *MAD* of the previous LCU may be quite different from the *MAD* of the current one with the high speed movement of UAV. Thus the *MAD* performs poorly at LCU level bit allocation. Gradient of the current LCU was used to allocate the bit at LCU level in Ref.[19], as it can be obtained directly from the current LCU. Instead of the gradient weight, we use ROI information of the current LCU to allocate bits at LCU level. The quality of ROI can be controlled flexibly by adjusting the pixel weight in the ROIs. The ROI map at LCU level $L\omega[i, j]$ can be expressed as

$$L\omega[i,j]=\frac{\sum_{h=r_j}^{r_j+H[j]-1}\sum_{w=c_j}^{c_j+W[j]-1}RM(w,h)}{H[j]\times W[j]}=$$

$$\frac{N_{\mathrm{ROI}}[i,j]\times k+(H[j]\times W[j]-N_{\mathrm{ROI}}[i,j])}{H[j]\times W[j]}=$$

$$\frac{N_{\mathrm{ROI}}[i,j]}{H[j]\times W[j]}(k-1)+1, \tag{4}$$

where $L\omega[i, j]$ represents the bit allocation weight of the $j$th LCU in the $i$th frame, $H[j]$ and $W[j]$ are the height and width of the $j$th LCU, $r_j$ and $c_j$ are row and column position of the top left corner of the $j$th LCU in $i$th frame, and $N_{\mathrm{ROI}}[i,j]$ is the number of pixels in the ROI of the $j$th LCU. The LCU with larger ROI means that it is more important and it needs more protections during video transmission.

Once the target bits for the $i$th frame are allocated, the target bits for LCU denoted by $T_{\mathrm{LCU}}[i,j]$ can be achieved with the ROI map at LCU level as

$$T_{\mathrm{LCU}}[i,j]=\frac{L\omega[i,j]}{L\omega_{\mathrm{rem}}}\times T_{\mathrm{f,rem}}, \tag{5}$$

where $L\omega_{\mathrm{rem}}$ is the weight of the rest LCU in the current frame, and $T_{\mathrm{f,rem}}$ is the remaining bits used to encode the rest of LCUs in the current frame. Clearly, the quality of ROI can be emphasized, since more target bits are allocated through Eq.(5) with high values of $L\omega[i, j]$ in ROI. We can flexibly control the quality of ROI by adjusting the value of $k$ in Eq.(1). In this way, more bits are allocated to ROI for ensuring better perceived quality.

$\lambda$ can be derived from $T_{\mathrm{LCU}}[i,j]$ by $R$-$\lambda$ model as

$$\lambda=\alpha\times\left(\frac{T_{\mathrm{LCU}}[i,j]}{H[j]\times W[j]}\right)^{\beta}, \tag{6}$$

where $\alpha$ and $\beta$ are initialized as 3.2003 and −1.367, respectively, and then updated by least square regression once a coded picture is available. From $\lambda$, $QP$ is finally obtained by

$$QP=4.200\ 5\times\ln(\lambda)+13.712\ 2\ . \tag{7}$$

Then, we evaluate the performance of the proposed ROI based RC scheme for UAV videos on HEVC platform. We utilize the HEVC reference software HM16.0[20] with its default $R$-$\lambda$ RC scheme as the reference scheme. Our ROI based RC scheme is then embedded into HM16.0 for comparison. We capture a raw uncompressed UAV video sequence in YUV format with 4:2:0 sampling as the test sequence. The test sequence contains 300 consecutive frames at 1 024×1 024 resolutions, and the frame rate is 30 frames per second. The encoder_lowdelay_P_main.cfg is used as the parameter setting with typical parameter settings of HM 16.0 as shown in Tab.1.

**Tab.1 The parameters for video coding**

| Section | GOP size | LCU size | Coding type |
|---|---|---|---|
| Contents | 4 frames | 64×64 pixels | IPPP…PPP |

The performance is measured in terms of the bitrate error from target bitrate, rate-distortion (RD) performance and mean peak signal-to-noise ratio (*PSNR*) between the proposed and the HM16.0 approaches. The bitrate error denoted by $\Delta R$ is expressed as

$$\Delta R=\frac{|R_{\mathrm{target}}-R_{\mathrm{actual}}|}{R_{\mathrm{target}}}\times 100\%\ . \tag{8}$$

where $R_{\mathrm{target}}$ and $R_{\mathrm{actual}}$ are the target and actual bitrates, respectively. In the evaluations, the sequence is encoded at different target bitrates from 100 kbit/s to 8 000 kbit/s, and the weight value $k$ in Eq.(1) is set to be 9. Tab.2 shows the comparisons of actual bitrate and the bitrate error for HM16.0 and our algorithm. It can be seen that the proposed algorithm generates smaller mismatch between target bitrates and actual output bitrates.

**Tab.2 Comparisons of bitrate and the bitrate error for HM16.0 and our algorithm**

| $R_{\mathrm{target}}$ (kbit/s) | $R_{\mathrm{actual}}$ (kbit/s) | | $\Delta R$ (%) | |
|---|---|---|---|---|
| | HM16.0 | Ours | HM16.0 | Ours |
| 100 | 117.802 | 103.756 | 17.802 | 3.756 |
| 200 | 202.847 | 198.572 | 1.424 | 0.714 |
| 500 | 512.334 | 497.141 | 2.467 | 0.572 |
| 1 000 | 992.813 | 997.845 | 0.719 | 0.216 |
| 2 000 | 1 992.447 | 1 996.148 | 0.378 | 0.193 |
| 4 000 | 3 994.872 | 3 992.625 | 0.128 | 0.184 |
| 6 000 | 5 995.639 | 5 997.598 | 0.073 | 0.040 |
| 8 000 | 8 996.817 | 8 995.613 | 0.040 | 0.055 |

Fig.2 shows the RD curves of HM16.0 and our algorithm in ROI, non-ROI and whole regions. It is observed that our approach outperforms HM 16.0 at various bitrates in terms of average *Y-PSNR* of the ROI. As for the cost, the RD performance of non-ROI may be degraded. However, thanks to the human vision system (HVS), people pay more attention to ROI, and the reduction of visual quality outside ROI is almost unnoticeable. Consequently, the degraded RD performance of non-ROI has few negative effects on the perceived visual quality of the entire frame.

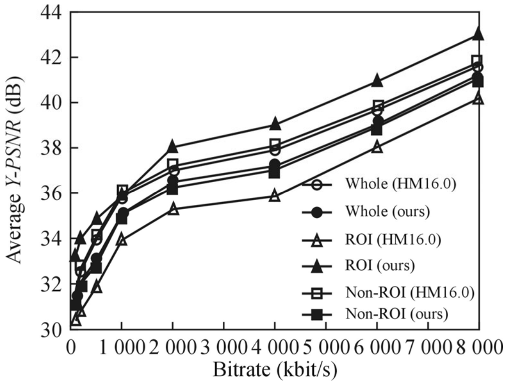
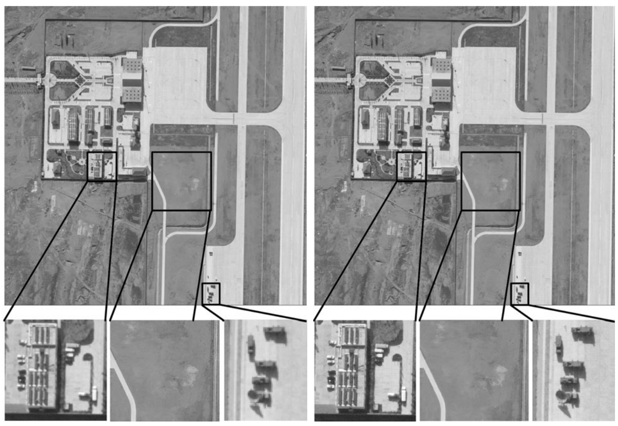



**Fig.2 RD performance comparison between the HM16.0 and our algorithm in ROI, non-ROI and whole regions**

Fig.3 demonstrates the reconstructed frame of the UAV video sequence compressed at 1 000 kbit/s obtained by HM16.0 and our algorithm. As expected, our approach can yield more favorable visual quality in the whole frame, since the visual quality in the ROI is improved and that in non-ROI is reduced almost unnoticeably, which can be seen in the amplified block in Fig.3.



(a) (b)

**Fig.3 Perceived visual quality comparison of encoded frame and amplified blocks between (a) HM16.0 and (b) our algorithm**

At last, we compare the results of our scheme with different $k$ and the gradient based $R$-$\lambda$ (GRL) method[19] in the aspect of $PSNR$ increase ($\Delta PSNR$) on HM16.0 in ROI. As shown in Fig.4, the larger $k$ leads to the bigger increase of $PSNR$ in ROI. This means that our scheme can flexibly control the quality of ROI by adjusting the $k$ value.

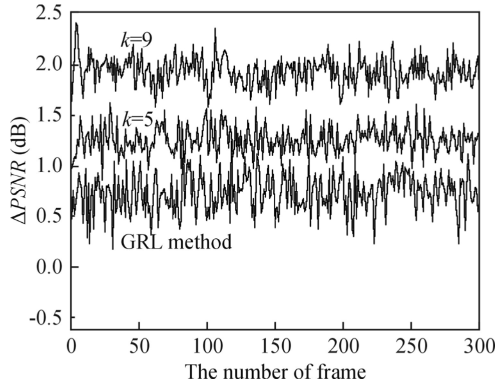



**Fig.4 Δ*PSNR* in ROI of our algorithm with different *k*=9 and 5 and GRL method**

In this paper, we put forward an ROI based RC scheme in HEVC standard. The scheme is developed for UAV applications. A more efficient and accurate ROI information can be obtained for real-time encoder through the salient object detection method and the interaction of UAV. More bits are assigned to the frame which contains larger ROI to adapt with the fast changing scene as the UAV flies at high speed. Bits are adaptively allocated with the ROI map at LCU level in order to improve the quality of the ROI. The quality of ROI can be flexibly controlled by using appropriate pixel weight in the ROI, and the $R$-$\lambda$ model with a better robustness is used to get a proper $QP$. The experimental results demonstrate that the reconstructed video quality is improved, and the bitrate error is reduced. Hence, the proposed RC scheme has potential value to be applied in UAV applications.

## References

[1] R. Schneiderman, IEEE Signal Processing Magazine **29**, 8 (2012).

[2] M. Bhaskaranand and J. D. Gibson, IEEE Journal of Selected Topics in Signal Processing **9**, 139 (2015).

[3] B.-R. Joan, S.-S. Joan and A.-L. Francesc, IEEE Signal Processing Letters **16**, 45 (2009).

[4] J. Y. Kim, C. H. Yi and T. Y. Kim, IEEE Transactions on Consumer Electronics **56**, 951 (2010).

[5] Y. Liu, Z. Li and Y. Soh, IEEE Transactions on Circuits and Systems for Video Technology **18**, 134 (2008).

[6] K. Sun and D. Wu, Journal of Visual Communication & Image Representation **30**, 234 (2015).

[7] Xu Sheng-yang, Yu Mei, Jiang Gang-yi, Fang Shu-qing and Shao Feng, Journal of Optoelectronics·Laser **26**, 2381 (2015). (in Chinese)

[8] Xu Jian, Wang Rang-ding, Huang Mei-ling, Li Qian and Xu Da-wen, Journal of Optoelectronics·Laser **26**, 1753 (2015). (in Chinese)

[9] G. J. Sullivan, J.-R. Ohm, W.-J. Han and T. Wiegand, IEEE Transactions on Circuits and Systems for Video Technology **22**, 1649 (2012).

[10] M. Xv, X. Deng, S. Li and Z. Wang, IEEE Journal of Selected Topics in Signal Processing **8**, 475 (2014).

[11] M. Simone, B. Riccardo and R. Roberto, A Saliency-Based Rate Control for People Detection in Video, IEEE International Conference on Acoustics, Speech and Signal Processing, 2016 (2013).

[12] B. Li, H. Li, L. Li and J. Zhang, Rate Control by R-lambda Model for HEVC, ITU-T SG16 Contribution, JCT-VC K0103, Shanghai, (2012).

[13] X. Wang and M. Karczewicz, Intra Frame Rate Control Based on SATD, ITU-T SG16 Contribution, JCT-VC M0257, Incheon, (2013).

[14] S. Li , M. Xu, X. Deng and Z. Wang, Signal Processing-Image Communication **38**, 127 (2015).

[15] M. Marwa, C. Marco and P.-P Beatrice, Region-of-Interest Based Rate Control Scheme for High Efficiency Video Coding, IEEE International Conference on Acoustics, Speech and Signal Processing, 7338 (2014).

[16] A. Borji, D. N. Sihite and L. Itti, IEEE Transactions on Image Processing **24**, 414 (2015).

[17] M.-M. Cheng, Z. Zhang, W.-Y. Lin and P. Torr, BING: Binarized Normed Gradients for Objectness Estimation at 300fps, IEEE Conference on Computer Vision and Pattern Recognition, 3286 (2014).

[18] H. Zhou, H. Kong, L. Wei, D. Creighton and S. Nahavandi, IEEE Transactions on Intelligent Transportation Systems **16**, 297 (2015).

[19] M. Wang, K. N. Ngan and H. Li, IEEE Signal Processing Letters **22**, 896 (2015).

[20] HEVC Reference Software, http://hevc.kw.bbc.co.uk/trac/ browser/jctvc-hm/tags, 2012.