

Aerial Video Encoding Optimization Based On x264

Fan Yang, Hongbing Ma

Department of Electronic Engineering, Tsinghua University, Beijing, China
Email: yf693036805@gmail.com, hbma@mail.tsinghua.edu.cn

Received 2012

ABSTRACT

x264 video codec uses lots of new video encoding technology based on H.264/AVC video encoding standard which enhances compression efficiency. However this results in so heavy computation that the x264 codec is not fit for real-time encoding application of high resolution video. This paper analyses the character of aerial video and then optimizes the inter-frame mode decision and motion estimation in x264 codec according to its character by reducing a lot of unnecessary computation. In the result, about 19% computation and encoding time is reduced with total bits and PSNR decreasing lightly.

Keywords: x264; Aerial Video Encoding; Mode Decision; Motion Estimation

1. Introduction

Aerial video real-time encoding and transmission is very important in application. With the definition of aerial video increasing to about 1080P, the big data brings the challenge for real-time encoding. Therefore, some scholars do lots of analysis and research about aerial video encoding which focus on two aspects mainly. (1) According to H.264/AVC encoding standard[1] and the character of aerial video, they use the global motion vector estimated by the speed of aircraft to substitute local motion vector to save the encoding time[2]. (2) By identifying the new area and object motion, they integrate the last frame into a new one[3]. This is a new encoding method which is independent of the traditional encoding standard. The two methods both decrease the complexity and both have locality. The first one is only used for the aerial video which is taken when the aircraft is parallel with the horizon. Otherwise, the method will lead to very low compression efficiency. The second method's challenge is in the recognition of the object motion. The error of recognition will have large bad impact in encoding.

H.264/AVC is a new generation video encoding standard which is proposed by ITU-T and ISO. H.264/AVC has high compression efficiency. x264[4] is the open-source release of H.264/AVC which is used widely. As x264 is not satisfying in the application of real-time encoding for aerial video, x264 must be optimized firstly before being used.

In x264, mode decision account for about 50%~70% in all the encoding time according to different command line parameter. For P frame, the candidate modes include P_skip, P16x16, P16x8, P8x16, P8x8, P4x4, P8x4, P4x8,

I16x16, I8x8, I4x4. For B frame, the candidate modes include B_skip, B16x16, B16x8, B8x16, B8x8, I16x16, I8x8, I4x4. We must calculate the encoding cost of all the modes to select the best one to gain the best compression efficiency. x264 use the Rate Distortion Cost[1] to measure the encoding cost. The process of calculating all the modes' encoding cost will cause so heavy computation. How to reduce the unnecessary computation is the focus of this paper.

Currently, some scholars propose improvement programs about the mode decision[5-8]. In one word, the main process is like this. Because of the correlation of adjacent frame in time and space for the aerial video, they predict the mode type of current macro block directly according to the analysis of the surrounding macro block's mode, motion vector and content. This method reduces lots of needless computation. But the prediction still needs lots of analysis and computation and may not be so accurate. It has limited optimization for the application.

The second part in this paper mainly analyses the character of the aerial video. The third part proposes the optimization algorithm in mode decision for the aerial video. The fourth part optimizes the motion estimation and the fifth part makes the simulation of the two algorithms. The last part gives the conclusion.

2. Character of Aerial Video

In aerial video, the change of adjacent frame is caused by the motion and the shake of the camera mainly. The motion of object in video will lead to so slight change that we can omit it. This means that the middle part of every

frame forms from the rotation and flat move of last frame and the edge is made of the new area.

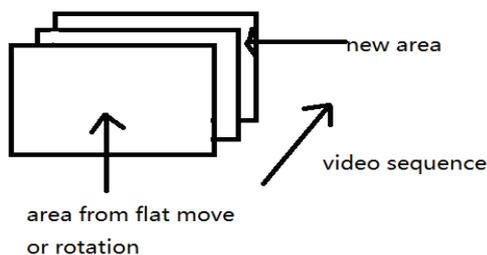


Figure 1. Video Sequence Change.

From the perspective of motion vector, the direction of motion vector is consistent for most macro blocks. But the magnitude of the motion vector has something to do with the angle between the camera and horizon. When the camera is parallel with the horizon, the magnitude is mostly the same which can be called the global motion vector. When the camera forms some angle with the horizon, the magnitude of the upper macro blocks are less. However, the motion vector is almost the same for the adjacent macro blocks in direction and magnitude.

In one word, the character of the aerial video is as follows:

- (1) The change of the adjacent frame mainly occurs in the edge district.
- (2) The motion vector of most adjacent macro blocks are the same in direction and magnitude.

3. Mode Decision Optimization

In the process of mode decision, the big macro blocks(16x16) are more suitable for the area of background or changing slowly because of the same motion vector. While the small macro blocks(4x4) are fitter for the area changing quickly. As in these areas, we can split the big macro block into the small macro blocks to do motion estimation respectively to gain better compression efficiency. Based on the analysis and the character of aerial video, we do some statistics about the mode decision on aerial video. We define that the edge area is the district whose distance is less than 5% to the boundary. We split the frame into two areas called the edge and the middle respectively. Then we do the statistics about the mode type for the two areas. Table 1 shows that the mode type ratio in the edge and the middle area.

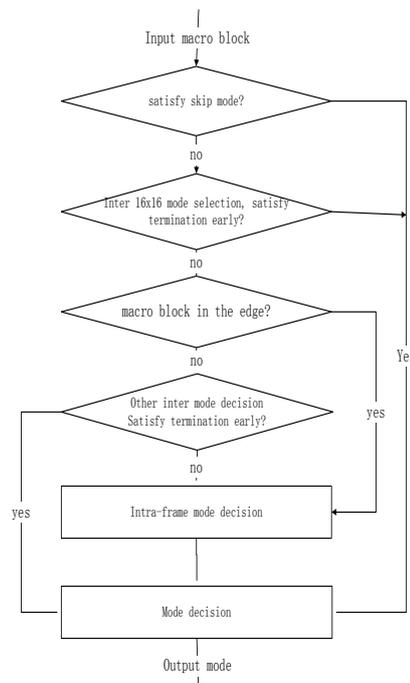
From Table 1, we can see that in the middle area, the skip mode and inter-frame 16x16 mode account for about 98% and the intra-frame mode about 1.5%, others about less than 0.5%. The current x264 encoder will seek the less than 0.5% inter-frame mode(except skip and Inter-frame 16x16 mode) in the cost of calculating all the inter-frame modes.

Table 1. Mode Type Ratio.

	Skip	Inter16x16	Inter16x8	Inter8x16
Edge	41.33%	52.78%	0.13%	3.81%
Mid	41.18%	56.93%	0.11%	0.25%
	Inter8x8	Intra16x16	Intra8x8	Intra4x4
Edge	0.08%	1.45%	0.39%	0.01%
Mid	0.07%	1.0%	0.44%	0.02%

This process will consume lots of encoding time and enhance little compression efficiency. Therefore, we propose an optimization algorithm. In the edge area, we compute all the intra-frame mode and inter-frame mode to find the best mode. While in the middle area, we only check the skip mode, inter-frame 16x16 mode and intra-frame mode, Omitting the check of the other inter-frame mode. The flow of mode decision after optimization is as Figure 2 shows.

Figure 2. Mode Decision Flow



4. Motion Estimation Optimization

Motion estimation is the process of searching the macro block which makes the motion compensation least. This is the most time-consuming process. x264 realizes 5 kinds of full-pixel motion estimation algorithms. They are dia(diamond search), hex(hexagon search), umh(UMH-exagon search), esa(exhaustive search) and

tesa(hadamard exhaustive search)[4]. The compression efficiency becomes better, the computation more complex, and encoding time longer. We do some statistics about the five motion estimation algorithms with the aerial video. Then we compare the change in encoding time, total bits and PSNR. The result shows in Table 2.

Table 2. Contrast of Motion Estimation Algorithms

	dia	hex	Umh
Coding time	14.00s	14.84s	18.13s
Total bits	424.9KB	424.3KB	424.2KB
PSNR	41.91	41.92	41.94
	esa	tesa	
Coding time	27.52s	33.35s	
Total bits	423.9KB	422.5KB	
PSNR	41.97	41.97	

In Table 2 we can see that the encoding time increase by 138% at most, while the total bits and PSNR change within 0.5% which can be omitted. According to the character of the aerial video, the motion vector of the adjacent macro block is familiar. x264 encoder can predict the motion vector accurately with some simple method such as median prediction. Based on the predicted motion vector, the encoder can find the macro block very fast which makes the current motion compensation least. Therefore, the best motion estimation method for aerial video is diamond search given compression efficiency and encoding time.

Diamond search algorithm is the process of iteratively matching macro blocks with diamond template. For the diamond search algorithm in aerial video, we do the statistic about the relationship between iterative counts and ratio. The result shows in Table 3.

Table 3. Diamond Search Algorithm Iterative Number Ratio.

1	2	3	4	5
65.7%	25.3%	5.60%	1.41%	0.62%
6	7	8	9	10
0.37%	0.24%	0.17%	0.12%	0.08%
11	12	13	14	15
0.07%	0.05%	0.03%	0.02%	0.02%

From Table 3, we can see something as follows:

(1) The macro blocks whose iterative number is 1 account for 65.7%. This kind of macro block's matching macro block is the same as the predicted one. Therefore, iterative computation is not useful for this kind macro blocks.

(2) The macro blocks whose iterative number are 2 or less than 2 account for about 96.6%. The macro blocks whose iterative number is small accounting for so high ratio shows that the prediction for the motion vector is very accurate. The matching macro block is almost near the predicted one.

According to the fact, this paper proposes an algorithm which is used to terminate the motion estimation earlier by self-adaptive threshold.

(1) Initialize the variable threshold = 0.

(2) Calculate the average cost of predicted matching macro block.

$$AC = PC / (H * W) \quad (1)$$

AC means average cost of predicted matching macro block. PC means the cost of predicted matching macro block. H and W are the height and the wide of current macro block.

If the average cost AC is less than (ratio * threshold), terminate the motion estimation, otherwise go to step 3. Ratio is the control constant which can be defined by ourselves.

(3) Conduct diamond search. If the count of the iteration equal to 1, calculate average value between the threshold and average cost AC computed in the step 2, assign the value to threshold. If the count of iteration is more than 1, finish the motion estimation directly.

Algorithm statement:

(1) Initialize the threshold at the first time we conduct the motion estimation.

(2) At the second step, the prediction cost is calculated according to the predicted vector which is realized in x264.

(3) Ratio is a constant used to control the filtering. For example, the ratio being set to 0 means filtering nothing, the same to no optimization. If the ratio is set to 100000, this means all the motion estimation will be omitted. According to the statistic, it is reasonable that the ratio is set to 0.8.

5. Simulation

This part simulates the optimization algorithms in the third and fourth part describe respectively. Then we combine the two algorithms to do the simulation together. The simulation result is as follows.

We take some aerial videos in the suburbs of Beijing whose contents include farmland, forest, factory school and moving cars. We name the video as video1, video2, video3, video4, video5.



Figure 3. Aerial Video Samples

Firstly, we explain some command line parameters in x264:

- (1) --no-psy, disable psychology optimization. PSNR will fail if the psychology optimization enabled.
- (2) --qp 30, qp is quantization parameter. 30 is a good threshold for the video encoding for our psychology.
- (3) --partitions all, select the best mode from all the mode type.
- (4) -me[*dia, umh, hex, esa, tesa*], which motion estimation method will be selected. In our experiment we choose *dia*.
- (5) --psnr, calculate the signal and noise ratio.

The negative data in the following tables means decrease after optimization. Positive data means increase after optimization.

We optimize the x264 encoder as the third part describes and Table 4 shows the change in the field of encoding time, total bits and PSNR after optimization.

Table 4. Change after Mode Decision Optimization.

	Coding time	Total bits	PSNR
video1	-15%	0.4%	-0.01
video2	-13%	0.3%	-0.01
video3	-22%	0.5%	-0.02
video4	-12%	0.9%	-0.02
video5	-14%	0.3%	0
average	-15%	0.5%	-0.012

We optimize the x264 encoder as the fourth part states and Table 5 shows the difference after optimization.

From the experiments we can reach some conclusions.

- 1) The two algorithms both decrease the needless computation to reduce the encoding time with the total bits and PSNR decreasing lightly.
- 2) From the first two experiments we can see that the mode decision optimization is better than motion estimation optimization.
- 3) The combined optimization is better than the separate one.
- 4) The two algorithms are simple to realize and have

stable effect on the real-time encoding for the aerial video.

Table 5. Change after Motion Estimation Optimization.

	Coding time	Total bits	PSNR
video1	-3%	-0.04%	-0.05
video2	-5%	-0.4%	-0.04
video3	-4%	0.8%	-0.1
video4	-8%	1.5%	-0.1
video5	-4%	0.2%	-0.09
average	-5%	0.4%	-0.08

We combine the two algorithms and optimize the x264 encoder. The difference is as Table 6 shows.

Table 6. Change after Mode Decision and Motion Estimation Optimization.

	Coding time	Total bits	PSNR
video1	-17%	0.6%	-0.05
video2	-16%	0.8%	-0.04
video3	-25%	0.7%	-0.11
video4	-18%	1.5%	-0.11
video5	-17%	0.3%	-0.09
average	-19%	0.8%	-0.08

6. Conclusion

This paper analyzes the aerial video and summarizes its characters. According to the characters, we optimize the x264 encoder from two perspectives. 1: We omit the needless mode decision to reduce the computation according to the position of the macro block. 2: We compute the threshold which is used to terminate the motion estimation earlier to reduce the encoding time. The two optimization save about 19% encoding time with the bit rates and PSNR decreasing slightly.

REFERENCES

- [1] Draft Recommendation and Final Draft International Standard of Joint Video Specification, Mar.2003, ITU-T Rec.H.264 and ISO/IEC 14 496-10 AVC, Joint Video Team
- [2] Malavika Bhaskaranand and Jerry D.Gibson. "Low-complexity Video Encoding For UAV Reconnaissance And Surveillance." The 2011 Military Communications Conference, Nov.2011.
- [3] Holger Meuel, Macro Mundeloh and Jörn Ostermann. "Low Bit Rate Based ROI Video Coding for HDTV Aerial Surveillance Video Sequence." 2011 IEEE Computer Society Conference On Computer Vision And Pattern Recognition Workshops, June 2011.
- [4] x264 Software(120711). Available from:

- <http://www.videolan.org/developers/x264.html>
- [5] Huanqiang Zeng, Canhui Cai and Kai-Kuang Ma. "Fast Mode Decision For H.264/AVC Based On Macroblock Motion Activity." IEEE Transactions On Circuits And Systems For Video Technology, Vol.19, No.4, pp.491-499, April 2009.
- [6] Tiesong Zhao, Hanli Wang, Sam Kwong and C.-C. Jay Kuo. "Fast Mode Decision Based on Mode Adaptation." IEEE Transactions On Circuits and Systems For Video Technology, Vol.20, No.5, pp.697-705, May 2010.
- [7] D.Wu, F.Pan, K.P.Lim, S.Wu, Z.G.Li, X.Lin, S.Rahardja, and C.C.Ko. "Fast Intermode Decision in H.264/AVC Video Coding." IEEE Transactions On Circuits And Systems For Video Technology. Vol.15, No.7, pp.953-958, July 2005.
- [8] Po-Hung Chen, Hung-Ming Chen, Mon-Chau Shie, Che-Hung Su, Wei-Lung Mao, Chia-Ke. Huang. "Adaptive Fast Block Mode Decision Algorithm for H.264/AVC." 2010 the 5th IEEE conference on Industrial Electronics and Applications, June 2010.