LONG TERM BACKGROUND REFERENCE BASED SATELLITE VIDEO CODING

Xu Wang[1]*, Ruimin Hu*[1]*, Zhongyuan Wang*[2]*, Jing Xiao*[3]*, Yuhui Zhang*[1]

[1]National Engineering Research Center for Multimedia Software,
School of Computer Science, Wuhan University, China
[2]Hubei Key Laboratory of Multimedia and Network
Communication Engineering, Wuhan University, China
[3]Collaborative Innovation Center of Geospatial Technology, Wuhan, China

ABSTRACT

Video transmission from satellites to terrestrial devices usually requires a large amount of channel resources due to the huge amount of satellite video data. Subject to limited transmission bandwidth in space environment, the video encoder for video satellite calls for higher coding efficiency. In this paper, we propose a high efficiency satellite video coding method based on long term background reference (LTBR) to eliminate redundancy caused by periodical revisit. Firstly, data of Google Earth is used to provide prior information for establishing LTBR. Then a novel intra prediction method guided by pixels' cluster information from LTBR is introduced. Experiments demonstrate that our method outperforms HEVC and H.264 , in terms of rate-distortion, BD-PSNR and BD-Rate performance.

Index Terms— Video satellite, video coding, background reference, long-term redundancy.

1. INTRODUCTION

Along with the development of space information network, video satellite has been developed rapidly in recent years. More and more video satellites have been put into use, such as Chinese Jilin-1, Carbonite-2 of SSLT company and Skysat series, capturing remote videos of the earth in space day and night.

Generally, in contrast to terrestrial video, satellite video takes more gigantic volume due to ultra-high space resolution. For example, the Jilin-1 can capture 12000×5000 resolution frames at the rate of 30 fps. Regardless of the limited computational resource in satellite, due to the constraint of bandwidth between satellite and the ground, the demand for real-time transmission poses a great challenge for video compression. At present, most of video satellites follow traditional video encoders, like H.264 [1], HEVC [2], to compress observed earth video in space, but specific coding framework designed for satellite video has seldom been exploited yet.

The traditional video coding schemes, like the most advanced coding standard HEVC [2], make full use of intra and inter prediction to reduce the spatial or temporal redundancy. However, they only take intra and inter redundancy among frames in a short term into consideration. Few methods on satellite video coding are specifically proposed so far, but some research are conducted on compression for video by unmanned aerial vehicle (UAV) which shares similar characteristics with satellite video. These works can be roughly classified into two categories: complexity related and bitrate related. For the former, they use additional information from sensors of UAV to reduce computational complexity of the encoder, especially for motion estimation process. For example, global motion compensation (GMC) is the mostly used method [3, 4, 5, 6]. For the latter, region-of-interest (ROI) [7] based methods are used to reduce bitrate. Generally, according to pre-calculated ROI, these methods will allocate more bitrate for ROI and less bitrate for other regions to save the overall bitrate[8] and [9]. Although they can maintain subjective quality of visually attentive regions, the overall quality is undermined. Meanwhile, the definition of ROI largely depends on the task of video satellite rather than merely the human visual system, which makes ROI uncertain. In brief, the rectified approaches under the existing video coding framework can only make little progress in terms of rate-distortion performance.

Furthermore, some research consider the whole redundancy as the motionless background in satellite video occupies a large ratio. Similar to surveillance video, background redundancy exists through the entire satellite video. Long-term reference generated by the background modeling [10, 11] is widely exploited to reduce repeated background information within a surveillance video [12, 13, 14, 15, 16, 17]. However, such long-term reference is still subject to the scope of a single video.

In further examination, satellite video exhibits two unique

The research was supported by the National Nature Science Foundation of China (61671336, 91738302, 61671332, U1736206), the Hubei Province Technological Innovation Major Project (2017AAA123), and the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (17E03).

characteristics. First, due to the high imaging angle of view, the observed area can be seen as a two-dimensional plane where motionless background takes large ratio within a frame. Second, regular revisits in one area can cause a lot of periodical redundancies in the video. Therefore, background redundancy in satellite video can be further extended into scope of multiple videos, called long-term redundancy. Following this idea, we proposed a method which makes use of Google Earth data as reference for I frames in our previous work [18]. In addition, a definition and color correction step was introduced to keep consistency between reference and coding frame for higher reference efficiency in that work. However, due to long-term interval between reference and coding frames, the color and definition deviations cannot be eliminated totally. As a matter of fact, no matter how different the definition and color between the reference from Google Earth and coding frame, people can still recognize that they are the same place according to the structure and texture information. This fact means that the most common redundancy can be attributed to structure and texture information. Therefore, the long-term reference coding method independent of color and definition variations is worth exploiting.

In this work, we proposed a high efficiency satellite video coding method based on long- term background reference. Firstly, Google Earth data is considered to be a priori knowledge of satellite imaging scenario. Then, we use satellite's position information to roughly search the corresponding region of the captured frame in Google Earth and further perform accurate registration. Next, in order to eliminate long-term structure and texture redundancy to achieve high efficiency, a novel intra prediction method guided by pixels' cluster information from LTBR is proposed. Experimental results demonstrate that our method boosts the coding performance for satellite video, in terms of rate-distortion curves, BD-PSNR and BD-Rate [19, 20].

2. PROPOSED METHOD

The proposed long term background reference (LTBR) based satellite coding scheme can be divided into two steps. Firstly, we match current coding frames with data of Google Earth according to the coordinates of the captured area to establish LTBR. Meanwhile, the matching parameters are integrated into the coding stream to transmit for recovery of LTBR. Secondly, structure and texture prior information of LTBR are used to guide intra prediction.

2.1. Match LTBR with Google Earth

To obtain LTBR, the first thing is to search the best matched image from data of Google Earth which is indexed by geographic coordinates. Thus, the matching process is divided into follow two steps.

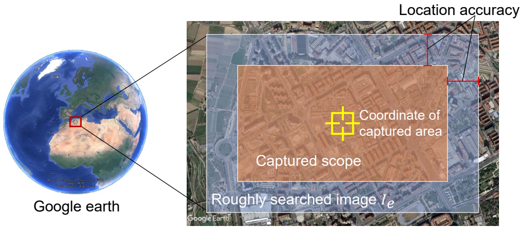

Fig. 1. Roughly search from data of Google Earth according to coordinate.

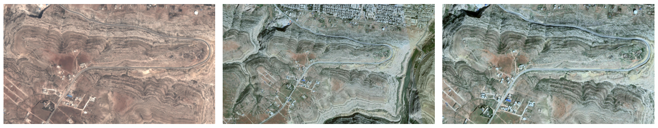

Fig. 2. The comparison of coding frame I_f, roughly searched image I_e and precisely matched image I_m. From left to right are I_f, I_e and I_m.

2.1.1. *Rough search according to coordinate*

Fig. 1 shows the roughly searching step from Google Earth. In this step, the longitude and latitude of the captured area are used to locate position of the captured area in Google Earth. And according to the scope of the camera lens, a general range is obtained, as the orange area in Fig. 1. Meanwhile, considering the accuracy of location, an additional part, the blue area in Fig. 1, is added to obtain the final roughly searched image from Google Earth. The image is donated by I_e.

2.1.2. *Precise registration by feature points*

In order to make the roughly searched image I_e can be referenced, a precise registration is proposed in this step. Firstly, the SIFT algorithm is used to detect key points of I_e and captured video frame I_f, and the key points are comprised by feature descriptors and positions which are donated by $f_i^{I_e}$, $p_i^{I_e}$ and $f_j^{I_f}$, $p_j^{I_f}$. Next, according to Euclidean distance of feature descriptor pairs between I_e and I_f, matched key point pairs are searched, and the matched position pairs are donated by $P_k^m = \left\{ p_i^{I_e}, p_j^{I_f} \right\}$. Then, the RANSAC algorithm is employed to obtain an affine transformation matrix H from I_e to I_f. Finally, the image I_e is affine transformed by matrix H to obtain a precisely similar image I_m to I_f. Fig. 2 shows the comparison of I_e, I_f and I_m.

2.2. LTBR based Intra Prediction

In this section, we introduce an intra prediction method guided by structure and texture prior information from LTBR. The outline of the method is shown in Fig. 3.

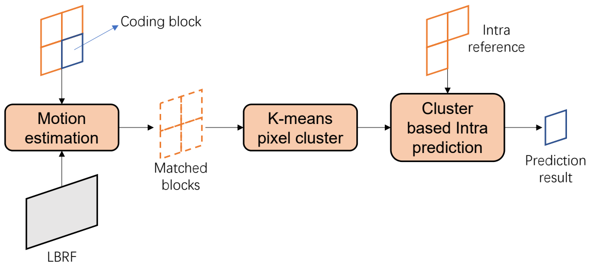

Fig. 3. Outline of LTBR based Intra Prediction.

2.2.1. *Structure and texture based motion estimation*

In order to make full use of structure and texture information from LTBR, we first estimate the best structure and texture matched block. As traditional motion estimation method may be inefficient because of color and resolution differences between LTBR and coding frame, we propose a novel motion estimation method in this step.

Firstly, we calculate the gradient of coding frame and LTBR by using Roberts operator. Then the x direction gradient and the y direction gradient of blocks are donated by G_x^{mb} and G_y^{mb}. Next the same search strategy with traditional coding scheme is employed to find the best matched block in LTBR. What different is that the error criterion is calculated as

$$\begin{aligned} J =& MSE(G_x^{current\_mb}, G_x^{LTBR\_mb}) \\ &+ MSE(G_y^{current\_mb}, G_y^{LTBR\_mb}), \end{aligned} \tag{1}$$

where the MSE represents mean-square error.

2.2.2. *K-means pixel cluster*

The intra prediction is based on the spatial correlation of pixels. Without structure and texture information, the closer the more relevant is the only principle to conduct intra prediction in the traditional scheme. Actually, pixels in the same category share stronger relevant relationship. Thus, we use K-means cluster method to extract structure and texture information in this step.

i. Randomly select n pixels from matched block as initial centroids of clusters in RGB color space, which are named as $(c_1...c_n)_{i=0}$ and initialize i variable, which represents current iteration number, with a zero.

ii. Compute Euclidian distances in RGB color space of pixels to each centroid.

iii. Classify pixels into n clusters according to their minimum Euclidian distances to the current n centroids, generating n clusters named: $(cl_1...cl_n)_i$.

iv. Obtain new centroids by calculating average of clusters in above step, named: $(c_1...c_n)_i$ and increase i variable by one.

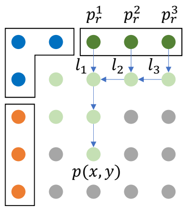

Fig. 4. Illustration of proposed intra prediction method.

v. Check if i is greater than maximum number of iterations, the current clusters are the final clusters. Otherwise go back to step ii.

2.2.3. *Cluster based Intra prediction*

After we obtained cluster information, a novel intra prediction method based on cluster information is introduced in this step. Firstly, cluster information from matched blocks is directly copied into current coding block and its adjacent reference blocks. Then, as mentioned above that pixels in the same category share stronger relevant relationship, we make intra prediction for each pixel by referring the same category pixels from adjacent reference blocks.

As Fig. 4 shows, we make these pixels as a graph, and each pixel is a node in graph. Only two pixels in the same category also adjacent are connected by an edge. Thus, the prediction of pixel in coding block is

$$p(x, y) = \frac{1}{L} \sum_{i=1}^{N} \frac{1}{l_i} p_r^i, \tag{2}$$

where l_i is the length of the shortest path from p_r^i to predicted pixel, and L is the sum of $\frac{1}{l_i}$

So far, the description of intra prediction method based on LTBR is almost completed. However, there is still a problem unsolved that in the K-means cluster step how many categories should be generated. For this problem, we employ the same strategy with mode selection in traditional intra prediction algorithm. The amount of category K is seen as a mode, and the optimal value is decided by the bit size of block's coding results. Also, range of K adopted in our experiments is integer from 1 to 5.

3. EXPERIMENTS EVALUATION AND DISCUSSION

3.1. Experimental Data and Method

In the experiment, we employ four satellite video sequences to make comparison. The sequences include two captured by Chinese Jilin-1 and two captured by Carbonite-2 of SSTL company. Each sequence has 300 frames, and the resolutions are all 1080p. The scenes of the sequences include mountain, city and airport. Example frames are illustrated in Fig. 5.

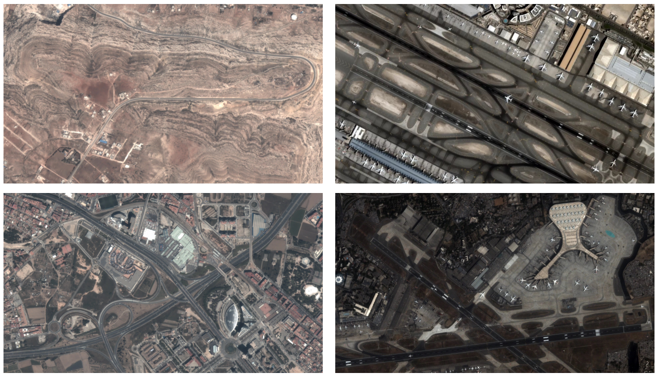

Fig. 5. Example frames of the satellite video sequences. The left two are captured by Jilin-1 and the right two are captured by Carbonite-2. The names of four sequences are Derna (DN), Valencia (VC), Dubai Airport (DA) and Mumbai Airport (MA) from the left column to the right column.

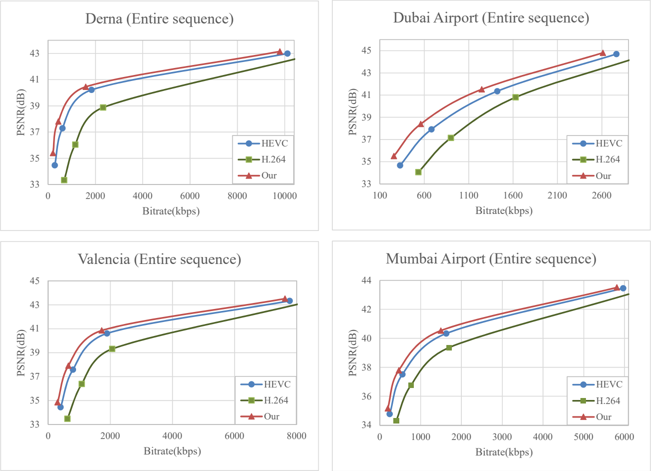

Fig. 6. Rate-distortion performance comparison.

Commonly, the rate-distortion curve, BD-Rate and BD-PSNR [19] are employed to evaluate the coding performance of our method. For comparison, we use IPPIPP structure to make compression and the interval of I frames is 30, the same as the frame rate. The coding results are obtained under the four equispaced qp, 22, 27, 32 and 37. Besides, our method is implemented by extending HM16.9, and the contrasted encoders are HM16.9 and JM19.0 which are implementations of HEVC and H.264. All the experiments are executed on Intel(R) Xeon(R) CPU E5-1620 v4 @ 3.50GHz and 32GB 2133MHz DDR4 Samsung memory.

3.2. Experimental Results

Firstly, we make a comparison from the view of rate distortion curves. As Fig. 6 shows that our method obviously improves the coding efficiency of four different sequences compared with HEVC and H.264.

Furthermore, it's easy to see that our method makes a great influence of I frames whose prediction types are all intra prediction. Thus, coding efficiency comparison of I frames is individually presented in the Table 1 by BD-Rate and BD-PSNR. As the table shown, at the same PSNR, our method decreases 25.58% and 55.98% bit-rate over HEVC and H.264 on average. Also, at the same bit-rate, our method enhances 1.49dB and 4.58dB PSNR over HEVC and H.264 on average. For the entire sequences, as shown in Table 2, our method decreases 23.1% and 55.56% bit-rate over HEVC and H.264 on average at the same PSNR. The PSNR of our method enhances 0.72dB and 2.23dB on average at the same bit-rate over HEVC and H.264.

Comparison of computational complexity is shown in Table 3. As can be observed, our encoding time is 115% of HEVC, and 164% of H.264 in average.

Table 1. The BD-Rate (%) and BD-PSNR (dB) of I frames compared with HEVC and H.264.

Vs.		DN	VC	DA	MA	Avg
HEVC	dB	1.94	1.47	1.53	1.00	1.49
	%	-34.81	-23.44	-23.48	-20.57	-25.58
H.264	dB	5.45	4.98	4.36	3.54	4.58
	%	-66.81	-56.69	-48.11	-52.31	-55.98

Table 2. The BD-Rate (%) and BD-PSNR (dB) of entire sequence compared with HEVC and H.264.

Vs.		DN	VC	DA	MA	Avg
HEVC	dB	0.72	0.64	1.02	0.51	0.72
	%	-31.01	-21.82	-20.89	-18.67	-23.10
H.264	dB	2.46	2.19	2.55	1.73	2.23
	%	-72.24	-56.67	-42.34	-50.97	-55.56

Table 3. The computational complexity compared with HEVC and H.264.

Vs.	DN	VC	DA	MA	Avg
HEVC	115%	118%	112%	114%	115%
H.264	164%	170%	158%	165%	164%

4. CONCLUSION

In this paper, we have proposed a LTBR based compression method for video satellite motivated by its acquisition characteristics. The data of Google Earth is first used as prior information to establish LTBR. A LTBR-guided intra prediction method is proposed to eliminate structure and texture redundancy, irrespective of differences of color and definition. The experiments verify that our method outperforms the state-of-the-art coding schemes in terms of rate-distortion curve, BD-PSNR and BD-Rate.

5. REFERENCES

[1] T Wiegand, G. J Sullivan, G Bjontegaard, and A Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[2] Gary J. Sullivan, Jens Rainer Ohm, Woo Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2013.

[3] Nirmala Ramakrishnan, Alok Prakash, and Thambipillai Srikanthan, "Low-complexity global motion estimation for aerial vehicles," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 402–410.

[4] Andres Rodriguez, Bryce Ready, and Clark Taylor, "Using telemetry data for video compression on unmanned air vehicles," in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 2006, p. 6468.

[5] Junbin Gong, Chenlin Zheng, Jinwen Tian, and Dingxue Wu, "An image-sequence compressing algorithm based on homography transformation for unmanned aerial vehicle," in *International Symposium on Intelligence Information Processing and Trusted Computing*, 2010, pp. 37–40.

[6] C. V. Angelino, L. Cicala, M. De Mizio, P. Leoncini, E. Baccaglini, M. Gavelli, N. Raimondo, and Roberto Scopigno, "Sensor aided h.264 video encoder for uav applications," in *Picture Coding Symposium*, 2013, pp. 173–176.

[7] Claudio M. Privitera and Lawrence W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *Pattern Analysis & Machine Intelligence IEEE Transactions on*, vol. 22, no. 9, pp. 970–982, 2000.

[8] Holger Meuel, Marco Munderloh, and Jörn Ostermann, "Low bit rate roi based video coding for hdtv aerial surveillance video sequences," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, 2011, pp. 13–20.

[9] Holger Meuel, Matthias Reso, Jorn Jachalsky, and Jorn Ostermann, "Superpixel-based segmentation of moving objects for low bitrate roi coding systems," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2013, pp. 395–400.

[10] Fangdong Chen, Houqiang Li, Li Li, Dong Liu, and Feng Wu, "Block-composed background reference for high efficiency video coding," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2016.

[11] Gang Wang, Bo Li, Yongfei Zhang, and Jinhui Yang, "Background modeling and referencing for moving cameras-captured surveillance video coding in hevc," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–1.

[12] Zhenfeng Shao, Jiajun Cai, and Zhongyuan Wang, "Smart monitoring cameras driven intelligent processing to big surveillance video data," *IEEE Transactions on Big Data*, vol. PP, no. 99, pp. 1–1, 2018.

[13] X. Zhang, T. Huang, Y. Tian, and W. Gao, "Background-modeling-based adaptive prediction for surveillance video coding.," *IEEE Trans Image Process*, vol. 23, no. 2, pp. 769–84, 2014.

[14] Thomas Wiegand and Bernd Girod, "Long-term memory motion-compensated prediction," *IEEE Trans.circuits Syst.video Technol*, vol. 9, no. 1, pp. 70–84, 1999.

[15] Tung Chien Chen, Chuan Yung Tsai, Yu Wen Huang, and Liang Gee Chen, "Single reference frame multiple current macroblocks scheme for multiple reference frame motion estimation in h.264/avc," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 17, no. 2, pp. 242–247, 2007.

[16] Mayank Tiwari and Pamela C. Cosman, "Selection of long-term reference frames in dual-frame video coding using simulated annealing," *IEEE Signal Processing Letters*, vol. 15, pp. 249–252, 2008.

[17] Manoranjan Paul, Weisi Lin, Chiew Tong Lau, and Bu Sung Lee, "Video coding using the most common frame in scene," in *IEEE International Conference on Acoustics Speech and Signal Processing*, 2010, pp. 734–737.

[18] Xu Wang, Ruimin Hu, Zhongyuan Wang, and Jing Xiao, "Virtual background reference frame based satellite video coding," *IEEE Signal Processing Letters*, vol. 25, no. 10, pp. 1445–1449, 2018.

[19] Gisle Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16/Q.6, Doc. VCEG-M033, Austin, TX, April 2001.

[20] Gisle Bjøntegaard, "Improvements of the BD-PSNR model," document ITU-T SC16/Q6, Doc.VCEG-AI11, 2008.