Multi-view Video Summarization Algorithm for WMSN

Mei-yan Zhang
School of Electrical Engineering
Zhejiang University of Water Resources and Electric Power, Hangzhou, China
E-mail: Meiyan19831109@163.com

Wen-yu Cai*
School of Electronics & Information
Hangzhou Dianzi University
Hangzhou, China
E-mail: dreampp2000@163.com

***Abstract*—Wireless Multimedia Sensor Networks (WMSN) has been one challenge category of special wireless sensor networks. Different-view videos captured by cameras in various different locations are typical multi-view videos. In order to reduce transmission quantity and conserve nodes' energy, traditional multi-view video coding approaches are long suitable for WMSN. In this paper, we propose a multi-view video summarization algorithm for WMSN, where different-view videos are converted to one monocular video with consideration of spatio and temporal correlations of networks, so as to reduce the redundancy of multi-view video. Moreover, different-view videos of adjacent sensor nodes are coded with correlated arrangement of I frames, B frames and P frames. Finally, a few preliminary simulations are present.**

Keywords-Wireless Multimedia Sensor Networks (WMSN); Multi-view Video; Cooperative Summarization

I. INTRODUCTION

Wireless Multimedia Sensor Networks (WMSN) [1-2] belongs to a specific sensor networks equipped with many camera sensors to carry out video surveillance or object tracking applications. In recent years, the increased availability of inexpensive video cameras together with cheap memory and fast data transfer has led to a growing interest in multiple view video. In the scenes of wireless multimedia sensor networks, a large number of camera sensors are deployed with an extremely denser manner. This leads to the generation of video streams containing large areas of overlap, requiring efficient multi-view video coding schemes for storage and transmission. A multi-view video sequence can provide different views of the same scene and thus provide much more comprehensive. Fig.1 describes the framework of multi-view WMSN.

The researches on multi-view video have been an active research topic. It is assumed that multi-view video is captured from different viewpoints with multiple cameras with high correlation. Consequently, H.264/AVC standard has been amended to support Multi-View Coding (MVC) [3]. MVC combines temporal and inter-view predictions, where frames are not only predicted from their temporal neighbors but also from corresponding ones in adjacent views [4]. This approach is beneficial to process multi-view video with MVC, however, it still has too much redundancy for bandwidth-limited WMSN. Correspondence matching in multi-view video sequences facilitates many applications of WMSN. Since multi-view videos can be summarized into one mono-view with correspondence matching, significant data compression ratio can be achieved.

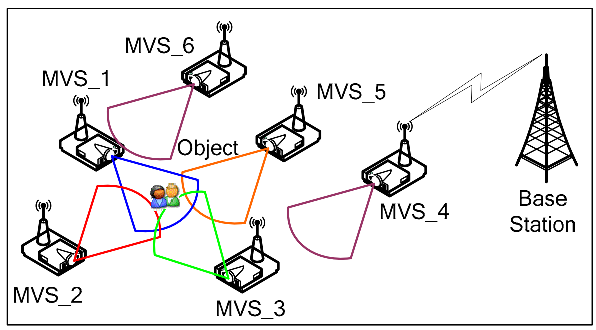


Figure 1. The framework of multi-view WMSN

The rest if this paper is organized as follows. Section II presents some related works on correspondence matching or synthesis methods of multi-view videos. Section III proposes the proposed cooperative summarization algorithm for multi-view video in WMSN. We evaluate our algorithm in Section IV and conclude the paper in Section V.

II. RELATED WORKS

Most state-of-the-art algorithms estimate correspondences by considering pairs of images. A comprehensive review of the state-of-the-art video summarization methods can be found in [5]. But there are a few new researches aiming to exploit inter-view video sequences correlations as well as spatio-temporal correlations in multi-view video sequences [6-9]. Fezza [6] et al proposed a preprocessing method using Histogram Matching (HM) algorithm for correcting these color discrepancies in multi-view video, so as to increase the matching robustness. Huo [7] et al presented the system architecture of WZ coding invoked for multi-view video. Then, they construct a novel mesh-structured pixel correlation model from the inter-view motion vectors and derive its decoding rules for joint source-channel decoding. Lee [8] et al proposed a correspondence matching algorithm for multi-view video sequences to provide reliable performance even when the multiple cameras with

978-1-4799-7091-9/14 $31.00 © 2014 IEEE
DOI 10.1109/WCSN.2014.50

significantly different viewing angles and positions. Sellent [9] et al provided a multi-video data structure to establish a confidence measure based on the consistency of the correspondences in a loop of three images.

Nevertheless, these above related works are mainly concerning to multi-view video coding or summarization, few researches starts from the view-point of WMSN. For energy-limited WMSN, traditional video process or coding approaches have many disadvantages to reduce data redundancy. Therefore, in this paper, we propose a multi-view video summarization framework in WMSN.

III. Multi-view Video Summarization Algorithm

Due to much more redundancy of multi-view video than that of monocular video, multi-view video summarization is much more comprehensive especially for multi-view WMSN. The overview of our method is shown in Fig.2. To the best of our knowledge, it is difficult to directly generate summarization, especially the video skims from multi-view videos. A common idea is to first parse the videos into shots. In this way, video summarization is transformed into a problem of selecting a set of representative shots.

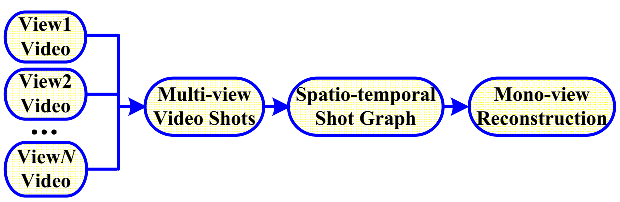


Figure 2. The principle of multi-view video summarization.

In the scene of multi-view video, the same scene is captured with multiple cameras from different viewpoints. For example in Fig.3, the point P is taken by three cameras. Te point P is projected onto three points Pm-1, Pm and Pm+1, which are expressed locally in the corresponding camera corresponding system respectively. The camera parameters and validation can be derived from [12].

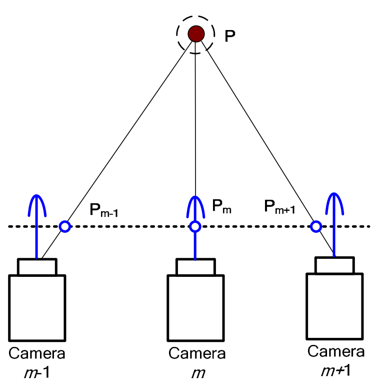


Figure 3. The view point relationship between sequential cameras.

It is supposed that there are m-views and we define Vt={V1,V2,…,Vm} as T=t video sequence. Then the detailed phases of our proposed multi-view summarization approach are described in below. The first step is to detect moving foreground objects using the background subtraction method in [10-11] and generate a binary shot finally. In the binary shot, pixels belonging to the foreground objects and the static background are assigned binary values of 1 and 0 respectively. The second and most important phase is how to reconstruction mono-view from different views according to temporal and spatial correlations. The principle of mono-view reconstruction process follows meanshift matching.

Finally, different-view videos of adjacent sensor nodes are coded with correlated arrangement of I frames, B frames and P frames, which is illustrated in Fig.4. The views in MVC structure can also be classified into three types, including I-view, P-view, and B-view, based on the type of the anchor frames. Among different views, inter-view prediction is employed to exploit spatial correlation. There are four views of same scene: View 1, View 2, View 3 and View 4. Considering correlation in spatial domain, only View 1 are coded into I frame, P frame and B frame, and View 2, View 3 and View 4 are coded only to P frame and B frame.

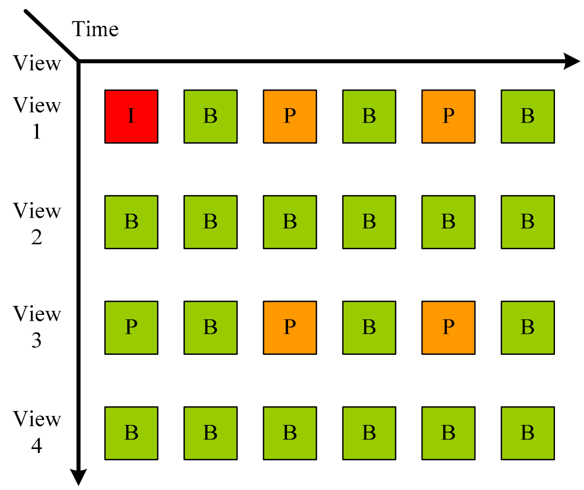


Figure 4. The sequence structure of frame type .

IV. Simulations

We use several test sequences of multi-view videos with 8-camera from MERL [13] and Fig.5 illustrates the layout of 8 cameras. Therefore, the difference between each view angle will be much smaller than arbitrary point of view.

The 178th frames from different views of Ballroom Multi-view videos are described in Fig.6, the shift difference values are very small and we use the proposed summarization algorithm to synthesize one video frame in Fig.7. The detailed summarization process will be described in our future papers, which considers the homogeneous parameters of used cameras, the distances and angles between objects with cameras, and also different synthesis methods.

Figure 5. Illustration of 8-camera setup

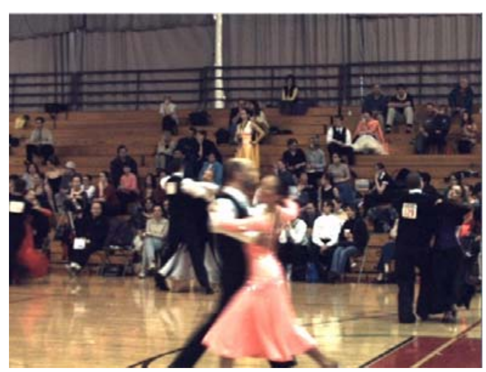
(a) View1

(b) View2

(c) View3

(d) View4

(e) View5



(f) View6

(g) View7

(h) View8

Figure 6. Ballroom Multi-view videos (178/250frames).

Figure 7. Summarization video frame.

The above summarization result is simple because we have the priori knowledge of camera sequence, therefore, its spatial relation can be revealed with some simple models. The other simulations such as shift matching result will be illustrated in our further works.

CONCLUSIONS

In this paper, we propose a video synthesis approach for multi-view WMSN to reduce transmission quantity and conserve nodes' energy. The proposed multi-view video summarization algorithm for WMSN converges different-view videos into one monocular video with consideration of spatial correlations of WMSN in order to reduce the redundancy of multi-view video. Meanwhile, a few crucial methods used in this summarization framework are illustrated in detail. The results of this paper are primary and need to be investigated in deep. Our future work is to take into account spatio-temporal correlation of multi-view videos and more simulation results will carry out to verify out proposed framework.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers whose insightful comments helped improve the presentation of this paper significantly. This research was partially supported by the Natural Science Foundation of Zhejiang Province (No.Y15F030066) and the National Natural Science Foundation of China (No.61102067).

REFERENCES

[1] I. F. Akyildiz, T. Melodia, K. R. Chowdhury, "A survey on wireless multimedia sensor networks", Computer Network, vol. 5, no. 1, 2007, pp.921-960.

[2] S. Soro and W. Heinzelman, "A survey of visual sensor networks", Adv. Multimedia Article ID 640386, vol. 2009, 2009, pp.1-21.

[3] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/AVC Standard", Proc. of the IEEE, 99(4), 2011, pp.626-642.

[4] C. Lipski, C. Linz, K. Berger, A. Sellent, M. Magnor, "Virtual video camera: image-based viewpoint navigation through space and time", Comput. Graphics Forum 29 (8) 2010, pp.2555-2568.

[5] B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," ACM Trans. Multimedia Comput., Commun., Appl., vol. 3, no. 1, 2007, pp.1-37.

[6] Sid Ahmed Fezza, Mohamed-Chaker Larabi, and Kamel Mohamed Faraoun, "Feature-based Color Correction of Multi-View Video for Coding and Rendering", IEEE Transactions on Circuits and Systems for Video Technology, 2013.

[7] Yongkai Huo, Tao Wang, Robert G. Maunder, and Lajos Hanzo, "Motion-Aware Mesh-Structured Trellis for Correlation Modeling Aided Distributed Multi-View Video Coding", IEEE Transactions on Image Processing, vol. 23, no. 1, 2014, pp.319-331.

[8] Soon-Young Lee, Jae-Young Sim, Chang-Su Kim, and Sang-Uk Lee, "Correspondence Matching of Multi-View Video Sequences Using Mutual Information Based Similarity Measure", IEEE Transactions on Multimedia, vol. 15, no. 8, 2013, pp.1719-1731.

[9] Anita Sellent, Kai Ruhl, Marcus Magnor, "A loop-consistency measure for dense correspondences in multi-view video", Image and Vision Computing, (30)2012, pp.641-654.

[10] E. B. Ermis, P. Calrot, P.-M. Jodoin, and V. Saligrama, "Activity based matching in distributed camera networks", IEEE Trans. Image Process., vol. 19, no. 10, 2010, pp.2595-2613.

[11] J. M. McHugh, J. Konrad, V. Saligrama, and P. M. Jodoin, "Foreground adaptive background subtraction", IEEE Signal Process. Lett., vol. 16, no. 5, 2009, pp.390-393.

[12] Anthony Vetro, Morgan McGuire, Wojciech Matusik, Alexander Behrens Jinho Lee, Hanspeter Pfister, "Multiview Video Test Sequences from MERL," ISO/IEC JTC1/SC29/WG11 m12077, Busan, Korea, April 2005.

[13] MERL. ftp://ftp.merl.com/pub/avetro/mvc-testseq, 2014..