PhD Forum: Efficient Communication Scheme for Mobile Visual Sensor Networks Equipped with RGB-D Cameras

Xiaoqin Wang, Y. Ahmet Şekercioğlu and Tom Drummond

Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, Australia

Email: {Xiaoqin.Wang|Ahmet.Sekercioglu|Tom.Drummond}@monash.edu

***Abstract*—This work presents a new method to efficiently communicate the correlated visual and depth information captured by multiple mobile RGB-D camera sensors. Our proposed scheme adopting the special properties of the depth map can extract the uncorrelated information between two images without conducting computationally heavy motion estimation. Moreover, a novel entropy coding scheme is proposed to further compress the depth data.**

I. Introduction and Problem Statement

In a densely deployed Visual Sensor Network (VSN), there exists correlation among the visual information observed by cameras with overlapped field of views (FoVs) [1]. In this project, we consider a circumstance in which a mobile visual sensor network uses depth sensing to explore and color information to map an unknown environment. And each sensor has to send the captured RGB-D information to a common receiver. Significant correlation exists in the RGB-D information captured by two or more mobile sensors when they have overlapped FoVs. Our goal is to efficiently extract and encode the uncorrelated RGB-D information that can only be observed by each sensor and avoid transmitting the same surface geometry and color information repeatedly.

Conventional approach to solve this problem is to apply the block-based motion estimation [2] on two images. However, this approach is not feasible in our case. As each sensor does not have the knowledge of the images captured by the other sensors. And if every sensor exchanging complete images with other sensors to extract the uncorrelated information will waste the limited bandwidth of wireless links. In order to avoid transmitting highly correlated data, we propose a novel framework in which each mobile sensor can use the relative pose (location and orientation) information to determine the RGB-D information which can only be observed by itself while cannot be captured by the other sensors. Then each mobile sensor only requires to transmit the uncorrelated information to the common receiver. In order to further compress the RGB-D information before transmission, the uncorrelated color and depth information is coded using the conventional color image coding scheme and a novel differential Huffman coding scheme, respectively.

II. Data

A depth map is an image or image channel which contains information relating to the distance of the scene surfaces from a viewpoint. Depth map and corresponding color image can be directly provided by RGB-D camera sensor that are based on structured light, such as the Microsoft Kinect [3]. The default RGB video stream provided by Kinect uses 8-bit VGA resolution (640×480 pixels) and the monochrome depth video stream is also in VGA resolution with 11-bit depth, which provides 2,048 levels of sensitivity. We use the mobile sensor [4] with a Kinect mounted on the top which can captured both color and the depth information of the environments.

III. Methods

A. Forward Prediction and Backward Check

Consider a situation in which sensor a and b have overlapped FoVs and they need to transmit their observed RGB-D information to a common receiver. The poses and FoVs of mobile sensor a and b are illustrated in Fig. 1. The relative pose between two sensors are estimated by using the fast approach presented in ICDSC paper [4]. Let Z_a and Z_b denote a pair

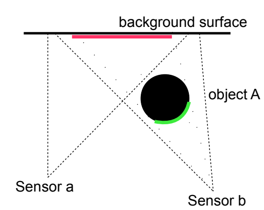

Fig. 1. Object A occludes the background surface in the view of sensor b.

of depth maps returned by sensor a and b. C_a and C_b are corresponding color images. With the accurate relative pose information between sensor a and b, sensor a can generate a depth map Z_b^* and a color image C_b^*, which are virtually captured at sensor b's viewpoint, by applying Depth Image Based Rendering (DIBR) technique [5] on Z_a and C_a. All of the frames are decomposed into 8×8 macro blocks. In the generated virtual frames, some blocks have no depth and color information. It is because none of the pixels in Z_a, C_a can be warped to these regions. The blocks with the same coordinates in Z_b and C_b contain the information that can only be observed by sensor b while cannot be seen by sensor a. Therefore, after sensor a sends these block coordinates to sensor b, sensor b will record these block coordinates as a set, $B_{forward}$, and only need to transmit the RGB-D information in these blocks of Z_b and C_b to the receiver node.

However, this process may fail to operate correctly in the situations when some points are occluded by the objects that can only be seen by the other sensor. According to Fig. 1, sensor a will incorrectly warp the RGB-D information of the red background surface onto the virtual frames as the information that can be observed by sensor b. In order to resolve this problem, we introduce a backward check mechanism. Similarly to the warping process from sensor a to b, in the backward check process sensor b can also generate virtual frames Z_a^* and C_a^*, which are virtually captured at sensor a's viewpoint. In the example shown in Fig. 1, the pixels representing the color and depth information of the green surface will move out of the image range and will not be shown in the virtual frames. Thus sensor b needs to determine the blocks including pixels in Z_b and C_b that move out of the image range in the backward check process. The set of these block coordinates is $B_{backward}$. Then, sensor b will derive the union of the block coordinates sets $B_{forward}$ and $B_{backward}$ as $B_{overall} = B_{forward} \cap B_{forward}$. The blocks of $B_{overall}$ in Z_b and C_b contain the information which can only be observed by sensor b. Therefore, sensor a will send the complete captured color image and depth map while sensor b will only send the RGB-D information in $B_{overall}$ to the common receiver.

B. Lossless Coding Scheme for Depth Maps

We propose a simple entropy coding scheme to further compress the depth information. As the depth maps usually include smooth regions representing the object surface, there is generally a high degree of correlation between neighboring pixels. The proposed scheme functions by comparing the current depth pixel value with a reference pixel and encodes the difference. Instead of using one dictionary, two different Huffman coding dictionaries are generated according to the various relations between the reference pixel and current pixel.

The flowchart of the proposed framework is illustrated in Fig. 2. For a network with more than two mobile sensors, the proposed scheme can be applied on every two sensors with overlapped FoVs to achieve efficient information communication.

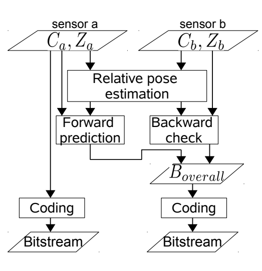

Fig. 2. Flowchart of the proposed framework.

IV. Experiments

The scheme proposed in III-A has been simulated on the datasets provided in [6]. An intuitive example is shown in Fig. 3. Depth maps are not shown as the depth maps are not for viewing purpose. (a) and (b) are the color images, C_a and C_b, captured by sensor a and b. (c) is the virtual image C_b^* generated from C_a. In (d), the black regions indicate the pixels in C_b that move out of the image range in the backward check mechanism. The union of the macro blocks in black regions in (c) and (d) is $B_{overall}$. Therefore, instead of transmitting the complete frames C_b and Z_b, only the blocks belonging to $B_{overall}$ in C_b and Z_b need to be transmitted by sensor b.

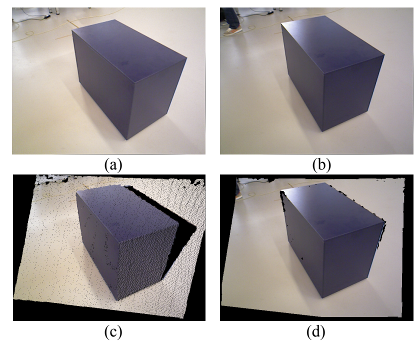

Fig. 3. Example of simulations.

The lossless coding scheme proposed in III-B for depth maps has been developed. The average length of a codeword is 1.532 bits. The uncompressed depth map returned by Kinect requires 2 bytes to store each pixel value. Therefore, the lossless compression ratio is 10.44.

V. Conclusion and Future Work

The main contribution of the proposed framework is that the uncorrelated information between two color images and depth maps can be determined efficiently without deriving the pixel difference. Our future work focuses on integrating III-A and III-B together and implementing the overall framework on the real mobile visual sensor network. Furthermore, the under-sampling [7] problem in the receiver side will also be addressed.

References

[1] X. Wang, A. Sekercioglu, and T. Drummond, "Multiview image compression and transmission techniques in wireless multimedia sensor networks: A survey," in *Distributed Smart Cameras (ICDSC), 2013 7th IEEE/ACM International Conference on*, October 2013.

[2] L.-K. Liu and E. Feig, "A block-based gradient descent search algorithm for block motion estimation in video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 6, no. 4, pp. 419–422, 1996.

[3] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, *Depth mapping using projected patterns*, 2008.

[4] X. Wang, A. Sekercioglu, and T. Drummond, "A real-time distributed relative pose estimation algorithm for rgb-d camera equipped visual sensor networks," in *Distributed Smart Cameras (ICDSC), 2013 7th IEEE/ACM International Conference on*, October 2013.

[5] C. Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," *Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems*, pp. 93–104, 2004.

[6] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Proc. of the International Conference on Intelligent Robot Systems (IROS)*, Oct. 2012.

[7] W. R. Mark, "Post-rendering 3d image warping: Visibility, reconstruction and performance for depth-image warping," Ph.D. dissertation, Univeristy of North Carolina at Chapel Hill, 1999.