Information Sciences 320 (2015) 306–319

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A bundled-optimization model of multiview dense depth map synthesis for dynamic scene reconstruction

You Yang [a,b], Xu Wang [c], Qiong Liu [a,*], Mingliang Xu [d], Li Yu [a]

[a] Department of Electronics & Information Engineering, Huazhong University of Science & Technology, Wuhan, China
[b] Department of Automation, Tsinghua University, Beijing, China
[c] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong
[d] School of Information Engineering, Zhengzhou University, Zhengzhou, China

ARTICLE INFO

Article history:
Received 15 June 2014
Received in revised form 26 October 2014
Accepted 6 November 2014
Available online 20 November 2014

Keywords:
Three-dimensional video
Optimization method
Depth image processing
Energy function optimization

ABSTRACT

Depth map is the basic requirement for all three-dimensional (3D) applications, but facing sensor noises, low frame-rate and low resolution in the procedure of data acquisition, especially in multiview cases. These problems bring obstacles to high quality 3D applications. Among the existing approaches, depth propagation is one of the promise approaches, and it can be utilized in temporal or spatial manner. However, propagation based algorithms process one aspect of the mentioned problems to pursuit local optimal solution. Actually, the process chain of depth map is from capture to application, and the optimization should be coupled instead of mutually independent. In this paper, we proposed a bundled-optimization scheme to process the thorough chain from capture to multiview dense depth map generation for the 3D applications. In this scheme, sensor noises in the captured low-resolution depth map are first detected and removed through a frequency-counting based non-linear filter. The filter refrains from the noise amplification in the procedure of depth map up-sampling. Low-pass blurring effect around high frequency areas is the by-product in up-sampling, and it is hard to detect in the depth map. We therefore propose a Blocklet based depth map optimization method for this blurring effect, and the accuracy of the high resolution depth map is then improved. Temporal depth propagation is then utilized on the optimized depth maps through the optical flow field rectified by temporal and spatial constrains. After that, a multi-set graph cut model is proposed to synthesize the multiview dense depth map. The experimental results indicate that our scheme can achieve at least 13.2575% PSNR gains when comparing to the benchmark depth map synthesis methods, and suggest the effectiveness of the proposed bundled-optimization method.

1. Introduction

Recently, three-dimensional (3D) video has witnessed a rapid development in both academia and industry, including 3D imaging [30,16,19], 3D video compression and communication [32,41,2], 3D modeling and retrieval [34,8,26,36,37,35,11,28], remote 3D medical treatment [43], and other fields. In these applications, the reconstructed 3D natural scene can provide realistic viewing experiences, and it has become an important extension for the traditional 2D vision applications. Among

* Corresponding author. Tel.: +86 27 87542667.
E-mail address: q.liu@hust.edu.cn (Q. Liu).

http://dx.doi.org/10.1016/j.ins.2014.11.014
0020-0255/

all of the applications, depth map is critical because it records the 3D space coordinate information for every pixel corresponding to the captured color image. However, the generation of high resolution and rate depth map is not as convenient as color image, which can be easily captured by traditional cameras. These limitations bring challenges to 3D applications, for example, shutter glass based ultra-high Definition 3D movie with 120 Hz display frequency. So far, there is a trade-off between resolution and rate for depth map generation. As for high resolution, stereo matching methods were utilized to color image pairs. The requirement of heavy computation resources is the main hinder in high resolution depth image generation. On the other hand, as for high rate, RGB-D (e.g., Kinect) and ToF cameras are adopted for video-rate depth image capture but only low resolution (e.g., 320×240 pixels) can be obtained, and it is far away from practical usages [20,17,12,7]. Furthermore, multiview depth image capturing results in serious noise problem when multi-RGB-D or ToF cameras are utilized simultaneously [39]. In many applications, the inaccurate depth information is one of the limitations, such as multiple view analysis for 3D object retrieval [10,9,5]. High quality of multiview dense depth maps can lead to performance improvement in many applications. Therefore, multiview dense depth map synthesis methods were proposed to solve the resolution-rate problem.

Among the methods of dense depth map synthesis, propagation based schemes play an important role and they can be utilized in temporal or spatial manner. It is assumed in these schemes that depth and color images are different representations of the same scene, and thus they share the same temporal and spatial features [43,42,33,38,21,24]. Therefore in temporal schemes, motion vectors or optical flow among consequent color images which represents the temporal features is utilized for depth map synthesis. Specifically, the depth value for the region containing static object maintains un-changed in the consequent depth maps, while motion in consequent color images corresponds to variation of depth value in the corresponding region [33,38,42,21,24]. Therefore, the status (i.e., static or moved) of objects in consequent color images is utilized to identify the depth value variation in the depth map. Temporal depth propagation via motion vector or optical flow can solve the resolution-rate problem effectively in many cases, but the performance mainly depends on the fidelity of color images. For example, the performance drops down dramatically in the case of low bit-rate of 3D video communication [43]. In order to optimize the quality of depth map, side information such as motion vectors in compressed video bit-stream is utilized [43,23]. Besides that, the reconstructed depth map can be further rectified via localized quality improvement. For example, the block-partition parameter can be extracted from the bit-stream as the required side information [43]. The second type of depth propagation is in spatial manner, and constrains from inter-view affine transformation is the fundament of this scheme [22,18]. For example, cross-view iterative filtering is utilized for depth map spatial propagation at the decoder with the help of inter-lace sampling method adopted at the encoder [23]. However, spatial schemes are unable to solve the problem of rate. Therefore, up-sampling filter is usually accompanied with this scheme. So far, the problem of multiview dense depth map synthesis is still open for both temporal and spatial depth propagation.

In this paper, we propose a bundle-optimization scheme for multiview dense depth map synthesis. Noisy, low frame rate and low resolution depth maps are captured by depth sensors, and they should be optimized for better quality, higher resolution and higher frame rate. In this method therefore, we first solve the problem of low resolution for the key-time depth map. After that, flying pixel problem in the up-sampled depth map brings background–foreground ambiguous. A Blocklet based clustering method is therefore proposed for flying pixel removal in depth map. In this method, depth image is warped into 3D point cloud in the first. After that, a Blocklet region selection method is utilized to identify and rectify the flying pixels. Through the proposed scheme, the noises in original depth maps are removed, and flying pixel problem in up-sampled depth maps is cleared. Finally, in the procedure of multiview dense depth map synthesis, the multi-set graph cut model is built on the rectified candidates through optical flow. The experimental results show that our method can achieve at least 36.78% and 0.198 dB average gains on bad point ratio and PSNR for the test images, respectively. These results suggest the effectiveness of our bundled-optimization scheme in generating multiview dense depth maps.

The rest of this paper is organized as follows. The proposed scheme is discussed in details in Section 2. After that, experiments and discussions are presented in Section 3. Finally, we conclude our work in Section 4.

2. The proposed bundled-optimization scheme

As mentioned previously, we synthesize multiview dense depth maps for all non-key frames through the bundled-optimization methods. The processing procedure of the proposed scheme is given by Fig. 1. As described by this flowchart, we first solve the problem of low resolution for the key-time depth map. In this step, noises in the captured key-time depth map are detected and removed before up-sampling. After that, flying pixel problem is left in the up-sampled depth map and brings background–foreground ambiguous in the procedure of 3D content generation. In the second step therefore, the up-sampled depth map is optimized by Blocklet and clustering joint method to remove the flying pixels. Based on the high resolution key-time depth maps, we utilize a multi-set graph cut model to generate dense multiview depth maps for the non-key frames. We utilize the symbols and notations in Table 1 for following discussions.

2.1. Noise removal for original depth map

The noises in the captured depth map are the results of mis-detection over object convex surfaces or cross disturbance between depth sensors [14]. The noise distribution is content based and the stochastic model is usually unknown, and