J. Vis. Commun. Image R. 47 (2017) 48–61

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Video synthesis from stereo videos with iterative depth refinement ☆

Chen-Hao Wei [a], Shang-Hong Lai [a], Chen-Kuo Chiang [b,*]

[a] National Tsing Hua University, Taiwan
[b] National Chung Cheng University, Taiwan

ARTICLE INFO

Article history:
Received 15 April 2016
Revised 27 March 2017
Accepted 27 April 2017
Available online 20 May 2017

Keywords:
View synthesis
Depth refinement

ABSTRACT

We propose a novel depth maps refinement algorithm and generate multi-view video sequences from two-view video sequences for modern autostereoscopic display. In order to generate realistic contents for virtual views, high-quality depth maps are very critical to the view synthesis results. Therefore, refining the depth maps is the main challenging problem in the task. We propose an iterative depth refinement algorithm, including error detection and error correction, to correct errors in depth map. Error detection aims at two types of error: across-view color-depth-inconsistency error and local color-depth-inconsistency error. Then, error pixels are corrected based on sampling local candidates. A trilateral filter that considers intensity, spatial and temporal terms into the filter weighting is applied to enhance the spatial and temporal consistency across frames. So the virtual views can be better synthesized according to the refined depth maps. To combine both warped images, disparity-based view interpolation is introduced to alleviate the translucent artifacts. Finally, a directional filter is applied to reduce the aliasing around the object boundaries to generate multiple high-quality virtual views between the two views. We demonstrate the superior image quality of the synthesized virtual views by using the proposed algorithm over the state-of-the-art view synthesis methods through experiments on benchmarking image and video datasets.

1. Introduction

More and more 3D display applications can be found in high-tech products, including 3D LCD/LED displays, 3D cameras and mobile phone devices. People are used to wear 3D glasses to experience 3D scenes on stereo displays. Nowadays, advanced 3D display technologies such as autostereoscopic displays allow people to experience realistic 3D scenes without wearing 3D glasses. People can watch 3D contents from different viewpoints by changing the viewing directions toward screen. This requires multi-view contents rather than two views for stereoscopic displays. However, stereo cameras can only capture images or videos at two views. One way to obtain multi-view image content is to convert stereoscopic contents to multiple views. Another solution is to use a multi-view camera array to capture the images or videos which is quite expensive and usually restricted to static camera setting.

Given a stereo video pair, we aim to synthesize the virtual videos at different viewpoints located between the two viewpoints of the stereo video pair so that each synthesized video could be seen as one of multi-view contents for the modern autostereoscopic display. Estimating the depth maps from the two-view video pairs is an essential task to achieve the objective of multi-view synthesis. There are two primary approaches to acquiring depth images: using active sensors, such as Time-of-flight (ToF) camera or depth sensor, and computing depth maps from stereo images by stereo matching algorithm. Active device can capture depth maps for dynamic scenes in real time, while stereo matching algorithms usually take much time for computing from just a pair of stereoscopic images.

The depth maps obtained by the above approaches include error pixels and noises. Two common error pixels can be found in the estimated stereoscopic depth pair. One is across-view color-depth-inconsistency error and the other is local color-depth-inconsistency error. The across-view color-depth-inconsistency error occurs as the depth value of one pixel in the left image is not consistent with the depth value in the corresponding pixel in the right view. The local color-depth-inconsistency error denotes the pixels that their depth values are very different from those of pixels with similar color in a local neighborhood. Figs. 1 and 2 depict two types of depth errors. In this paper, a joint error detection and correction algorithm is proposed to refine depth maps

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author at: No. 168, Section 1, University Rd., Min-Hsiung Township, Chia-yi County 621, Taiwan.

E-mail addresses: ss780217@gmail.com (C.-H. Wei), lai@cs.nthu.edu.tw (S.-H. Lai), ckchiang@cs.ccu.edu.tw (C.-K. Chiang).

http://dx.doi.org/10.1016/j.jvcir.2017.04.008
1047-3203/

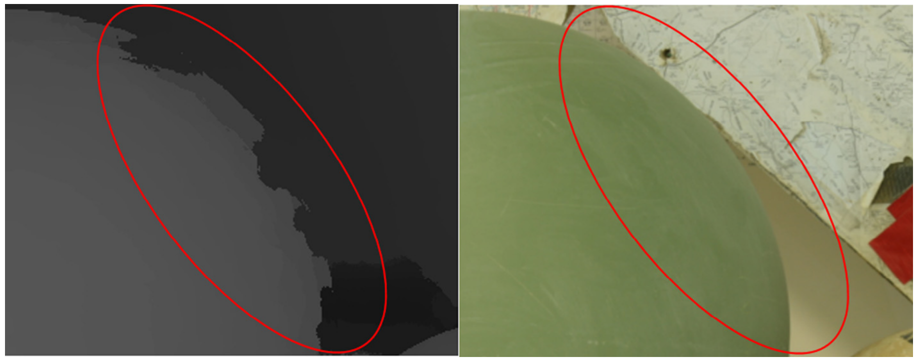

Fig. 1. The local inconsistent error in the depth map around object boundary compared to its color image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 2. Depth maps contain inconsistent errors across views. The depth values in the red windows are not consistent in both views. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that are estimated by any existing stereo matching technique or acquired from depth capturing devices.

After the errors are corrected, a spatio-temporal smoothing approach is introduced to smooth the depth maps and enhance the temporal consistency. Therefore, the noises, errors and visual artifacts across frames in the depth maps can be alleviated. After the depth maps are refined, view synthesis is applied to each stereo image pair of the stereo video with its corresponding refined depth map. First, the left and right views are warped to the desired viewpoint. A disparity-based view interpolation is applied to generate a virtual view using the warped left and right views. This method blends both warped images into a single virtual view effectively and reduce the error caused by occlusion when one pixel can be seen in the left view while occluded in the right view.

Next, we introduce a patch-based image inpainting to fill all holes. The quality of hole filling directly influences the view synthesis quality. However, some artifacts in the view synthesis are caused by image warping with inaccurate depth information around the object boundaries. A directional filter is proposed not only to reduce the artifacts but also to preserve the original edges of objects in the synthesized views. Thus, the virtual view for all frames is generated by using the above procedures.

The main contributions of this work are threefold.

- A joint error detection and correction algorithm is proposed to refine depth maps. It is effective for depth recovery and very general to any existing stereo matching technique or depth capturing devices.
- A complete view synthesis framework is proposed to synthesize virtual views between left and rightmost views by exploiting spatio-temporal smoothing approach, patch-based image inpainting and directional filters.
- We demonstrate the effectiveness of the proposed system on several publicly available sequences and apply our depth recovery method to the existing stereo matching algorithms [1,2]. Our method achieves better performance.

2. Related work

2.1. Depth refinement

Because the depth maps obtained by active sensing devices or stereo matching technique are usually far from perfect. There are two common active capturing devices, one is ToF camera and the other is Kinect. For ToF camera, auxiliary devices such as digital cameras are often exploited for depth map upsampling purpose because it cannot capture color image. In [3], depth maps which are captured from a low-resolution and noisy 3D-ToF camera can be refined into high quality depth maps with the help of corresponding color image by high-resolution digital camera. For Kinect, since there are occlusions in the captured depth maps, inpainting is required to fill in occluded regions. Lai et al. [4] applied a median filter on object construction to fill in missing depth values.

Once we have initial depth maps, we have to improve them because of several defects. In order to refine depth maps for images, advanced non-linear filters, such as non-local mean filters or bilateral filters, are quite effective for the purposes. Joint bilateral filtering and its variations are readily available tools for depth refinement using high quality auxiliary color images for cross filtering [5]. Yang et al. [6] applied bilateral filtering technique and interpolated the high-resolution depth values. Park et al. [3] used a NLM term to regularize depth maps and combined with a weighting scheme that involves edge, gradient, and segmentation information extracted from high quality color images.