Computer Aided Geometric Design 64 (2018) 15–26

Contents lists available at ScienceDirect

Computer Aided Geometric Design

www.elsevier.com/locate/cagd

Reconstructing non-rigid object with large movement using a single depth camera ☆

Feixiang Lu *, Bin Zhou, Feng Lu, Yu Zhang, Xiaowu Chen, Qinping Zhao

State Key Laboratory of Virtual Reality Technology & Systems, Beihang University, China

ARTICLE INFO

Article history:
Received 1 March 2018
Received in revised form 18 April 2018
Accepted 11 June 2018
Available online 25 June 2018

Keywords:
Non-rigid object reconstruction
Large movement
A single depth camera
Canonical frame identification

ABSTRACT

Non-rigid detailed 3D reconstruction of real world scenes has witnessed great success in recent years. However, most existing methods take the first frame as canonical model and the topological structure of the input scenes are fixed during the reconstruction process, which is an assumption that may not hold in practice for highly non-rigid scenes. Regarding this issue, this work proposes a novel approach to reconstruct non-rigid object with large movement which often results in topological structure change. In this paper, we firstly introduce an adaptive strategy that can effectively identify the most fine-grained scene topology as the canonical model. Such model is then deformed to each depth map, constrained by robust inter-frame correspondences established from object contour and scene flows. After deformation, we further fuse the depth map to the canonical model via a novel adaptive selection scheme, so as to remove spurious noise without smoothing model details. Experimental results show that the proposed approach can effectively handle various input scenes with large movement and generate models with high-fidelity details.

1. Introduction

3D reconstruction of real-world scenes from depth cameras is a widely studied problem in the fields of computer vision and computer graphics. After long-term efforts, the 3D model of a scene can be now accurately built by fusing its depth maps captured in multiple views, as long as the scene is static (e.g. KinectFusion (Newcombe et al., 2011; Izadi et al., 2011)). However, reconstructing non-rigid scenes with a single depth camera is still largely unsolved due to a number of challenges, such as non-rigid deformation, incomplete scans, and large movement which might cause the inconsistency of topological structures of the scene.

In recent years, the challenges of handling non-rigid deformation and incomplete scans have been well studied and addressed by various previous works (Sumner et al., 2007; Xu et al., 2007; Li et al., 2009; Liao et al., 2009; Zhou et al., 2010; Oikonomidis et al., 2011; Taylor et al., 2012; Li et al., 2013; Yang et al., 2013; Zollhöfer et al., 2014; Dou et al., 2015; Zhang et al., 2015a; Yang et al., 2015; Dou et al., 2016). However, these methods rely on strong priors based on pre-designed templates, user direct manipulation, multiple depth sensors, or pre-learned statistical models. Moreover, some techniques need seconds to minutes to compute a single frame which is a waste of time for reconstruction. Newcombe et al. (2015) proposed the first system for densely reconstructing general dynamic scenes, which can generate high-quality results from a single camera in real-time. Although significant successes were made by these approaches, however, most of them do not

☆ This paper has been recommended for acceptance by Ligang Liu.
* Corresponding author.
E-mail address: flylu@buaa.edu.cn (F. Lu).

https://doi.org/10.1016/j.cagd.2018.06.002
0167-8396/

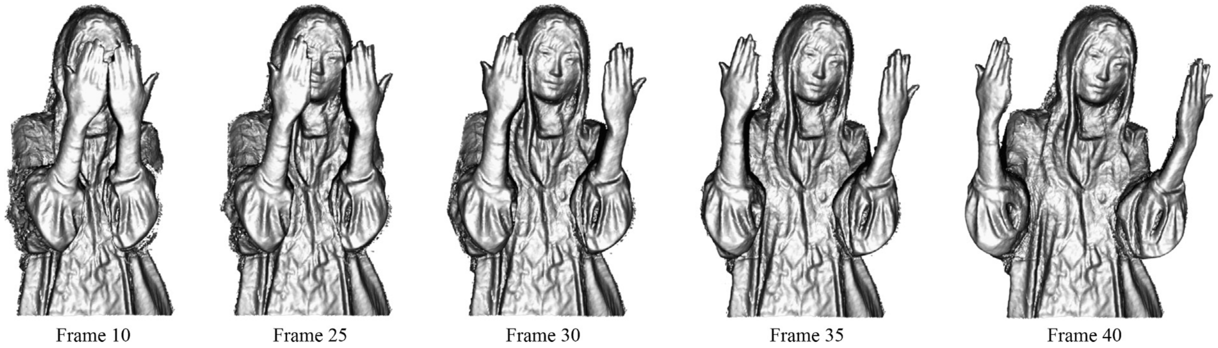

Fig. 1. We present a new approach to reconstruct non-rigid object with large movement. Our method only requires single depth camera (e.g. Kinect v2.0) to capture the depth maps and generate the plausible results rapidly (≈0.1 s/frame).

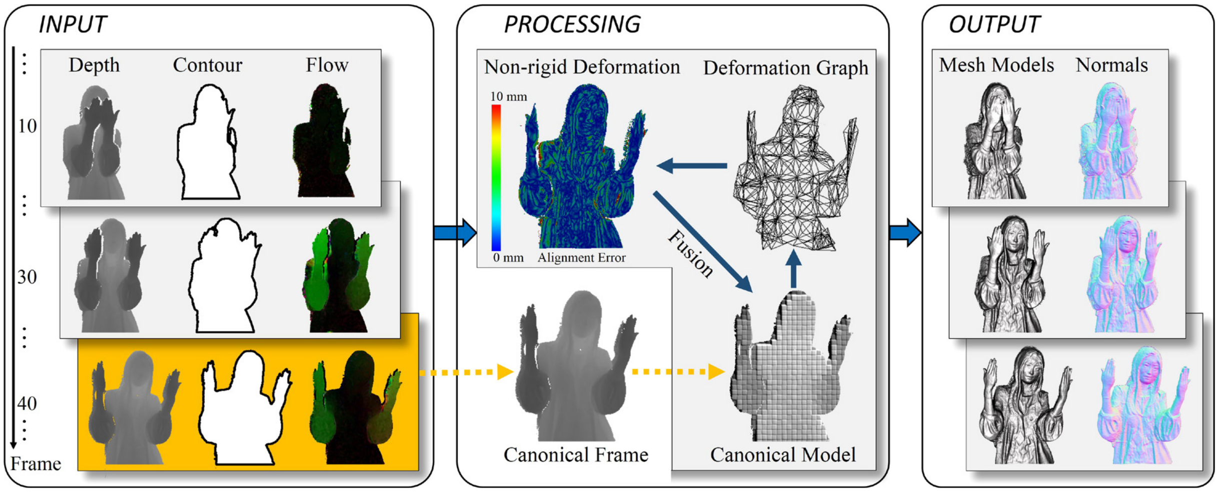

Fig. 2. An overview of the system pipeline. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

explicitly consider the challenging problem of large movement (i.e. potential topological change) of the input scene, which frequently happens for non-rigid objects. As illustrated in Fig. 1, with the arms of the person stretching away from the body, the topology of the person becomes inconsistent. In this case, the previous methods using fixed topological structure cannot reconstruct the input scene consistently. Slavcheva et al. (2017) have made the attempt towards this problem by a level-set evolution approach. However, the reconstruction without correspondences makes the results somewhat shaking in appearance and loose essential details.

To address this problem, this paper presents a novel approach to reconstructing non-rigid scenes with large movement from a single depth camera. As summarized in Fig. 2, the proposed approach takes the depth sequence captured by Kinect v2.0 sensor as input, and incrementally fuses the depth maps to generate a canonical model that can best fit the scene on each frame under certain deformations. To this end, we propose a novel adaptive strategy to identify the most fine-grained scene topology as the canonical model by analyzing the topological structure. Given the canonical model, we then deform it to each depth map constrained by robust inter-frame correspondences established from object contours and scene flows. Finally, we fuse the depth maps onto the deformed canonical models through a novel scheme that can adaptively select the appropriate interval of frames for fusion, which can generate high-quality reconstruction results without over-smoothing model details. Experimental results demonstrate that our approach can effectively handle various input scenes with topological structure change due to large movement.

The contributions of this paper are summarized as follows: 1) we present a novel approach via identifying the canonical frame to reconstruct the non-rigid scenes with large movement; 2) we efficiently deform the canonical model to fit each depth map using contour and scene flow cues; 3) we propose an adaptive fusion algorithm which can largely suppress the noise during fusion and preserve the model details.

2. Related work

There were various previous works on 3D scene reconstruction based on consumer-level depth cameras. While a large group of them focused on static scenes (Newcombe et al., 2011; Izadi et al., 2011; Roth and Vona, 2012; Whelan et al., 2012;