Temporal Upsampling of Depth Maps Using a Hybrid Camera

Ming-Ze Yuan, Lin Gao*, Hongbo Fu, and Shihong Xia*

Abstract—In recent years, consumer-level depth cameras have been adopted for various applications. However, they often produce depth maps at only a moderately high frame rate (approximately 30 frames per second), preventing them from being used for applications such as digitizing human performance involving fast motion. On the other hand, low-cost, high-frame-rate video cameras are available. This motivates us to develop a hybrid camera that consists of a high-frame-rate video camera and a low-frame-rate depth camera and to allow temporal interpolation of depth maps with the help of auxiliary color images. To achieve this, we develop a novel algorithm that reconstructs intermediate depth maps and estimates scene flow simultaneously. We test our algorithm on various examples involving fast, non-rigid motions of single or multiple objects. Our experiments show that our scene flow estimation method is more precise than a tracking-based method and the state-of-the-art techniques.

Index Terms—Hybrid Camera, Scene Flow Estimation, Depth Upsampling

1 Introduction

In recent years, low-cost depth cameras such as Microsoft Kinect and Intel RealSense have been popular and employed for various computer graphics applications, including motion capture [1], scene reconstruction [2], and image-based rendering [3]. For such cameras, the resolution and speed of depth acquisition are sacrificed to achieve a low cost. For example, the latest Microsoft Kinect depth camera for Xbox One (Kinect V2) is able to capture depth frames with only 512 $\times$ 424 resolution at 30 frames per second (FPS). While such specifications might be sufficient for certain applications, they are not sufficient for applications involving fast motions and higher frame-rate video. On the other hand, with recent advancements in imaging sensors, high-resolution, high-frame-rate and low-cost video cameras such as GoPro have also opened up many possibilities in computer graphics, such as outdoor motion capture [4], structure from motion (SfM) and dynamic hair capture [5].

Video cameras have their advantages over depth cameras in terms of frame rate and resolution. Observing that high-resolution video cameras are cheap and available anywhere, several techniques (e.g., [6], [7]) have been proposed to use a hybrid camera, i.e., a high-resolution video camera and a low-resolution depth camera, to perform spatial upsampling of depth maps. Many applications, such as image-based rendering [8], [9] and image processing [10], can benefit from additional depth information. The Kinect V2 itself is already such a hybrid camera. However, the low-frame-rate capture problem of existing depth cameras is largely unexplored and thus is the focus of our work.

Motivated by the existing hybrid cameras for obtaining the spatial super-resolution of depth maps and the available high-frame-rate, low-cost video cameras, such as GoPro (with 240 FPS), we propose a hybrid camera to achieve temporal upsampling of depth maps (Fig. 1). Our hybrid camera consists of a low-frame-rate depth camera and a synchronized high-frame-rate video camera. The key challenge is to effectively extract fast motion information from color images using a high-frame-rate video camera and then use it to guide the interpolation of depth maps. A straightforward solution is to first compute the 2D optical flow [11] between consecutive images using the high-frame-rate camera and then employ the resulting motion flow to estimate intermediate depth maps between a pair of original depth maps. However, this simple solution works well only for translational motions.

Another possible solution is based on scene flow [12]. However, the traditional methods for scene flow estimation require both color images and depth maps acquired at roughly the same frame rate and thus cannot be directly used for temporal upsampling. To address this problem, we formulate an optimization to estimate the scene flow and intermediate depth maps jointly; the estimated scene flow is used to guide the interpolation of intermediate depth maps, which in turn help refine the scene flow estimation. We derive data constraints from the high-frame-rate color images and enforce spatiotemporal regularization based on the shortest motion path and the locally rigid deformation assumption.

We test our hybrid camera on various examples with quickly moving single or multiple objects and humans.

- *Corresponding Author: Lin Gao (gaolin@ict.ac.cn) and Shihong Xia (xsh@ict.ac.cn)*
- *M.-Z. Yuan is with the Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, and also with the University of Chinese Academy of Sciences, Beijing, 100190, China. E-mail: yuanmingze@ict.ac.cn.*
- *L. Gao is with the Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China. E-mail: gaolin@ict.ac.cn.*
- *H. Fu is with the School of Creative Media, City University of Hong Kong. E-mail: hongbofu@cityu.edu.hk*
- *S. Xia is with the Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China. E-mail: xsh@ict.ac.cn.*