

Preface

Depth cameras have been exploited in computer vision for several years, but the high price and poor quality of such devices have limited their applicability. With the invention of the low-cost depth sensors such as Kinect and Time-of-Flight (TOF) cameras, high-resolution depth and visual (RGB) sensing have become available for widespread use as an off-the-shelf technology. The complementary nature of the synchronized depth and RGB information opens up new opportunities to solve fundamental problems in computer vision, including human pose estimation, activity recognition, object and people tracking, 3D mapping and localization, etc. Furthermore, the robustness gained with depth cameras allow us to take computer vision out of the lab and into real environments (e.g., people's homes).

While it is beneficial to use RGB-D sensors in many vision applications such as object/scene recognition, human pose estimation, and gesture/activity recognition, machine learning plays an important role in bridging the gap between feature representations and decision making. For instance, the human pose recognition algorithm associated with Kinect adapts a well-known learning technique into a real-world task, where body parts are inferred using a randomized decision forest, learned from over 1 million training examples. The success of this algorithm bootstraps the investigation of applying intelligent machine learning techniques to this new type of sensor representation. Many systems have already demonstrated the possibility of making a better decision by learning useful information from a large set of RGB-D data.

This book brings together high-quality and recent research advances on RGB-D based computer vision. The targeted readers are researchers and practitioners working in the areas of computer vision, human-computer interaction and machine learning from both academia and industry. It can also be used as a reference book for graduate students studying computer vision, pattern recognition or multimedia.

We generally divide the chapters into four parts. In the first part, we have included two survey chapters, which overview the prior arts in this filed. In the second part, six chapters dedicate to the research of RGB-D based 3D

reconstruction, mapping, and synthesis. In the third part, there are two chapters describing novel techniques that employ depth data for object detection, segmentation, and tracking. The last part consists of four chapters, demonstrating accurate interpretations of human actions with the aid of the depth information. In the following we summarize all the chapters.

In Chap. 1, Kadambi et al. explain the theoretical difference between the first generation Kinect and the recent delivered second generation Kinect from depth creation perspective. This chapter also investigates the methods that can correct artifacts on depth maps.

In Chap. 2, Berger summarizes the developments that include two or more RGB-D sensors in one scene. The chapter focus on discussing effective means of mitigating interference errors between multiple sensors in different applications. In addition, the chapter lists the most prominent datasets, which are publicly available.

In Chap. 3, Zhang et al. deal with the problem of calibrating the color and the depth cameras. The algorithm is designed to enhance the traditional checkerboard based calibration scheme because it does not work properly in the particular application. The new algorithm proposes a maximum likelihood solution for the joint depth and color calibration based on the assumption that points on the checkerboard shall lie on a common plane.

In Chap. 4, Koschan and Abidi present a new method to reduce the noise on depth map. Here, a joint bilateral filter is carried out in order to enhance depth information with the aid of color information. The novelty lies in the fact that the joint bilateral filter is applied to a common distance transform map, which represents the degree of pixel-modal similarity between a depth pixel and its corresponding color pixel.

In Chap. 5, Liu et al. aim at capturing real performances of human actors, in which the core problem is to reconstruct 3D human performance. To do so, the algorithm automatically tracks the motion of the handheld cameras and aligns the surface and skeleton of each tracked performer to the captured RGB-D data. In order to solve more practical problems, like occlusions and human body orientation changes, the system synchronize the signals from three handheld Kinects.

In Chap. 6, Liu et al. introduce three interesting applications where analyzing the depth signal plays important roles. In the first application, fusing depth and RGB information helps to accurately reconstruct a real human and provide personalized avatars for users. In the second application, depth-based human skeleton tracking is used to evaluate energy consumption of users during a game playing. In the last application, authors show the possibility of obtaining more accurate human action classification by interpreting the depth map.

In Chap. 7, Feinen and Grzegorzek present a novel approach that is able to facilitate the RGB-D object retrieval application. The main idea is to match the curves derived by object contours. However, the traditional 2D curve matching is incapable of handling viewpoint changes. To overcome this problem, this chapter comes up with new ideas that match 3D curve configurations.

In Chap. 8, Berger et al. deal with the application of reconstructing the gas flow from three Kinect cameras. In this chapter, a new subpixel accurate flow detection algorithm is proposed, facilitating the sparse image data such as the Kinect spot pattern. Afterwards, they exploit the sparse spot detection algorithm to provide masks for a GPU-based visual hull reconstruction.

In Chap. 9, Tian develops a RGB-D based computer vision system to assist blind and visually impaired persons. The core technique contributed by the chapter is a stairs and pedestrian crosswalks detection algorithm based on RGB-D images. The system serves as a navigation aid that can help blind users to gain improved perception and better understanding of the environment changes.

In Chap. 10, Han et al. utilize Kinect in a smart environment, where sensing the location and identity of the users is essential. In this chapter, viewpoint invariant features are extracted from the object shape and used for detecting human. Moreover, human re-entry is identified by a boosting-based classification algorithm, in which depth information helps to select suitable positive samples.

In Chap. 11, Dominio et al. discuss the effective features that characterize static hand poses. Several depth-based descriptors have been presented and examined in the chapter. The final conclusion is that some features, such as curvatures, distance, and correlation features, have better performances. The combination of multiple features allows to obtain much better performance.

In Chap. 12, Liang and Yuan present a unified framework to enforce both the temporal and spatial constraints for hand parsing. In this work, a superpixel Markov Random Field is leveraged to efficiently remove the misclassified regions produced by per-pixel classification. Finally, a rotation-invariant hand gesture recognition algorithm is designed to recognize digit number gestures.

In Chap. 13, Krupka et al. describe a novel classifier architecture based on discriminative ferns ensemble. This classifier architecture optimizes both classification speed and accuracy, given a large training set. To speed up the algorithm, simple binary features and direct indexing are employed. In addition, a large capacity model and careful discriminative optimization help to gain the algorithm accuracy.

In Chap. 14, Yao et al. propose a new hand segmentation and gesture recognition system. The contribution is a 3D contour model from the classified pixels. Instead of matching between images, this 3D contour model can be coded into strings. Therefore, the correspondence sample gesture can be found by a fast nearest neighbor searching method.

Sheffield, UK
Eindhoven, The Netherlands
Cambridge, UK
Redmond, USA

Ling Shao
Jungong Han
Pushmeet Kohli
Zhengyou Zhang

Computer Vision and Machine Learning with RGB-D
Sensors

Shao, L.; Han, J.; Kohli, P.; Zhang, Z. (Eds.)

2014, X, 316 p. 163 illus., 148 illus. in color., Hardcover

ISBN: 978-3-319-08650-7