Preface

The Workshop on Consumer Depth Cameras for Computer Vision (CDC4CV) brought together researchers from around the world, to explore the latest ideas about using a revolutionary consumer-priced depth camera, Microsoft's Kinect. This book is an outgrowth of that workshop, with the workshop organizers (Andrea Fossati, Juergen Gall, and Helmut Grabner from ETH Zürich; Xiaofeng Ren from Intel Labs Seattle; and Kurt Konolige from Willow Garage) inviting expanded contributions from the workshop participants.

The Kinect appeared as a consumer device in late 2010, and developer versions of the PrimeSense technology were available to some researchers up to a year earlier, Kurt Konolige at Willow Garage, and Xiaofeng Ren at Intel Labs in Seattle among them. Prior to this, Kurt Konolige had pursued a stereo system at Willow Garage that projected a texture onto a scene, and so allowed stereo to work on untextured objects. It worked great, and the limitation of "stereo dropouts", the lack of depth information on untextured surfaces, could finally be overcome. This system was incorporated into Willow Garage's PR2 robot platform, and was a primary enabler for many robot tasks. Along with colleagues Gary Bradski in vision, Sachin Chitta in manipulation, and many others both at Willow Garage and elsewhere, he implemented applications such as obstacle avoidance for robot arms, segmentation of planes, simple object recognition and pickup, discovery and modeling of articulated objects, and automatic plug-in to standard sockets, to name a few. With the advent of PrimeSense devices, however, he realized that the Willow Garage development of textured projection stereo could not compete in price, robustness, or convenience, and quickly abandoned it in favor of the former.

The Intel Seattle lab actively pursued the use of prototype PrimeSense devices as a major research thrust between 2010–2011, in collaboration with the University of Washington. The Intel-UW team promoted the name *RGB-D*, to emphasize the synergy of jointly using depth and color, and completed a number of successful projects on 3D mapping and modeling, object and gesture recognition, as well as interactive toy playing using projected display. Much of the RGB-D research transitioned to the newly founded Intel Science and Technology Center for Pervasive Computing.

Due to its low cost, once the sensor became publicly available, many research groups around the world began to exploit it. In the early 2011, the team at ETH Zürich started exploring several potential applications for RGB-D cameras, including action recognition, head pose estimation, physics-based tracking and human-computer interfaces. In fact it turned out that many very challenging vision tasks had great benefits from the addition of real-time reliable depth data.

Since that time, the field has progressed rapidly, to the point where it made sense to have a workshop in the autumn of 2011, to exchange ideas on the now widely available Kinect device, both its functioning and applications. The response to the call for participation was overwhelming—over 60 submissions, with 10 accepted oral presentations, and an equal number of posters. This book contains a selection of papers originating from the workshop, expanded and revised to full-length expositions, taking advantage of the book format to present results in more detail. It is divided into three main categories.

Part I, 3D Registration and Reconstruction, addresses the classic vision problem of multi-view geometry: how to correlate images from different viewpoints to simultaneously estimate camera poses and world points. The addition of depth information, as well as the constraint of real-time performance on a video stream, offer opportunities as well as difficulties that are addressed in these contributions.

Part 2, Human Body Analysis, offers a set of studies on human pose estimation, perhaps the most-researched area using video-rate depth images. The studies range from pose estimation for gaming and motion capture, through 3D human body scans for individuals, to recognition of hand pose and its application to sign language parsing.

Part 3, RGB-D Datasets, gives an overview of approaches to various recognition problems: category and instance learning of objects, human activity recognition. In computer vision, much progress has been made in these areas by having commonly available databases to test against; here we see the first steps towards the same approach using depth information as well.

As Jamie Shotton so aptly explained in the Forward, the Kinect, based on structured-light technology developed by PrimeSense, Ltd., has extraordinary depth quality, especially considering its cost. In the 1990s, Kurt Konolige witnessed and participated in amazing advances in mapping and navigation for robotics that were made possible by accurate depth information in a single plane from scanning laser rangefinders; we can only expect that PrimeSense-class devices, with precise depth over a full-field view, at video rates, will similarly help to advance applications in a much larger variety of fields.

We would like to thank all our collaborators at our respective institutions—at the Computer Vision Laboratory of the Swiss Federal Institute of Technology in Zürich, at the Computer Science and Engineering Department of the University of Washington, and at Willow Garage, Inc. in Menlo Park. Further thanks go to the heads of our groups for providing financial support and setting up constructive and fruitful working environments. To the workshop participants and presenters, as well as Gary Bradski and Jamie Shotton, the invited speakers, we extend our gratitude for their efforts at making it a great experience. Finally we would like to acknowledge

all the contributors of this book, especially Pushmeet Kohli and Jamie Shotton for preparing a very interesting invited chapter.

Zürich, Switzerland — Andrea Fossati
Juergen Gall
Helmut Grabner
Seattle, USA — Xiaofeng Ren
Menlo Park, USA — Kurt Konolige

Springer

http://www.springer.com/978-1-4471-4639-1

Consumer Depth Cameras for Computer Vision
Research Topics and Applications
Fossati, A.; Gall, J.; Grabner, H.; Ren, X.; Konolige, K. (Eds.)
2013, XVI, 210 p., Hardcover
ISBN: 978-1-4471-4639-1