

Foreword

Over the past few years the fields of computer vision and computer graphics, two well-established but distinct areas of computer science, have begun to investigate some common problems. As computer vision techniques have matured they have found applications to problems in many areas, including computer graphics. At the same time, the field of computer graphics has become more concerned with the use of image data for producing realistic, synthetic images.

The area of overlap between graphics and vision, now commonly termed image-based rendering, uses computer vision techniques to aid in synthesizing new views of scenes. Image-based rendering methods are having a substantial impact on the field of computer graphics, and also play an important role in the related field of multimedia systems, for applications such as teleconferencing, remote instruction and surgery, virtual reality and entertainment.

The recent interest in image-based rendering methods has brought with it a renewed investigation of some well-established computer vision techniques, in particular stereo vision and structure from motion. Using these computer vision techniques in the context of rendering new views creates new requirements that are just beginning to be understood. This book, which grew out of Daniel Scharstein's doctoral thesis, provides a rigorous introduction to some of these new requirements, and develops new computer vision techniques to address them.

In this book, Daniel Scharstein provides an introduction to the field of image-based rendering, including a broad survey of the state-of-the-art literature. Besides providing a well-written introduction to this area, this text makes several important research contributions. First, it develops a novel way of formalizing the view synthesis problem under the full perspective model, yielding a clean, linear warping equation. Second, it provides new techniques for dealing with visibility issues such as partial occlusion and "holes", problems that have received little attention in the literature. In addition, it provides a thorough re-evaluation of the requirements that view synthesis places on stereo algorithms. Finally, the book introduces two novel stereo algorithms specifically tailored to the application of view synthesis.

Image-based rendering has become an important research area only fairly recently. From the beginning, however, Daniel was firmly convinced that the new application area of image synthesis would result in a substantially different formulation of the stereo problem, which would then require new solutions. His insight was certainly right on. Moreover, Daniel sought out collaborators such as Richard Szeliski, who had established themselves as leaders at the boundary of computer vision and computer graphics. Students like Daniel really lead their faculty advisors, rather than the other way around.

December 1998

Daniel P. Huttenlocher
Associate Professor of Computer Science
and Weiss Presidential Fellow
Cornell University

Preface

The topic of this volume is an investigation of the use of stereo vision for the application of view synthesis. View synthesis – the problem of creating images of a scene as it would appear from novel viewpoints – has traditionally been approached using methods from computer graphics. These methods, however, suffer from low rendering speed, limited achievable realism, and, most severely, their dependence on a global scene model, which typically needs to be constructed manually.

Motivated by the shortcomings of traditional computer graphics methods, we present a new approach to view synthesis that avoids the above problems by synthesizing new views from existing images of a scene. Using an image-based representation of scene geometry computed by stereo vision methods, a global model can be avoided, and realistic new views can be synthesized quickly using image warping.

The first part of this book focuses on the view synthesis problem. Chapter 1 introduces and motivates the problem, and provides a brief review of stereo vision. Chapter 2 contains an in-depth survey of related work in image-based rendering and stereo vision. In Chapter 3, we formalize the view synthesis problem under the full perspective model and derive a linear warping equation using a special rectification step. We discuss how to resolve visibility, and how “holes” resulting from partially occluded areas can be filled. We also discuss how the view synthesis method can be used in a larger framework for image-based scene representations, and present experiments demonstrating that it is possible to efficiently synthesize realistic new views even from inaccurate and incomplete depth information.

The new application of stereo for view synthesis makes it necessary to re-evaluate the requirements on stereo algorithms. In Chapter 4, we compare view synthesis to several traditional applications of stereo and conclude that stereo vision is better suited for view synthesis than for applications requiring explicit 3D reconstruction. In particular, limited achievable depth resolution and matching ambiguities due to lack of texture are less of a problem for view synthesis. Other issues become more important, such as the correct recovery of depth discontinuities. We also discuss ways of dealing with partially occluded regions of unknown depth and with completely occluded regions of unknown texture.

The second part of the book presents several novel stereo algorithms that are motivated by the specific requirements imposed by view synthesis. In Chapter 5, we introduce a new evidence measure based on intensity gradients for establishing correspondences between images. This measure combines the notions of similarity and confidence, and allows stable matching and easy assigning of canonical depth interpretations in image regions of insufficient information. In Chapter 6, we present new diffusion-based stereo algorithms that are motivated by the need to correctly recover object boundaries. Our algorithms are based on iteratively diffusing support at different disparity hypotheses and locally controlling the amount of diffusion based on the current quality of the disparity estimate. In particular, we develop a novel Bayesian estimation technique that significantly outperforms area-based algorithms using fixed-sized windows. We provide experimental results for all algorithms on both synthetic and real images.

Chapter 7 concludes the volume with a summary and a discussion of possible directions for future work.

This book is based on my Ph.D. thesis, which was submitted to Cornell University in January of 1997. The original document has been revised and updated thoroughly; in particular, Chapter 2 now provides a comprehensive review of related work through December 1998. Some of the material in this volume is based on work that has been published previously. The view-synthesis method in Chapter 3 and some of the material in Chapter 4 was first presented in Scharstein [1996]. The stereo method in Chapter 5 is an extension of work described in Scharstein [1994a, b]. Finally, the material presented in Chapter 6 is based on joint work with Richard Szeliski [Scharstein and Szeliski, 1996; Scharstein and Szeliski, 1998].

Acknowledgements

I would first like to thank Dan Huttenlocher, who has been a great friend and advisor. He gave me a lot of support and freedom while I was trying to find my own path as a researcher, and did not mind when I picked a different area of research from his own. (And I did eventually manage to convince him that stereo is an interesting problem.) Many thanks to Sheila Hemami, Nick Trefethen, and Ramin Zabih, all of whom served on my Ph.D. committee and gave many valuable comments, suggestions, and corrections. Thanks especially to Ramin, who got me interested in the problem of synthesizing new views from real images in the first place. He also contributed many valuable ideas and insights during numerous interesting research discussions. I would also like to thank Amy Briggs for tirelessly reading and correcting countless drafts of this document, and for finding (hopefully) all the typos. (She also deserves the credit for the good English.)

I am indebted to Rick Szeliski, who has had a tremendous influence on my career as a researcher and has taught me a lot about effectively conducting

research. Thanks to Rick for a wonderful collaboration, which started with a week-long visit at DEC's Cambridge Research Lab in June, 1995, and which has contributed an important part to the work presented here.

Thanks to everyone who helped me during my last two years of working on my dissertation "remotely" from Vermont, including Karl Böhringer, Mark Hayden, and Nikos Pitsianis, and also the members of the great support staff at Cornell. I would also like to thank my colleagues in the Department of Mathematics and Computer Science at Middlebury College, for their support and friendship.

Finally, I would especially like to thank my family for their love, support, and confidence, and for putting up with my "emigration" in such a good-natured way. It is nice to know that someone is happy about me being happy.

And, most important of all, I thank my wife, Amy, who has made my life so wonderful. This book is dedicated to her.

Middlebury, January 1999

Daniel Scharstein