

Cognitive Systems Monographs

Volume 5

Editors: Rüdiger Dillmann · Yoshihiko Nakamura · Stefan Schaal · David Vernon

Edilson de Aguiar

Animation and Performance Capture Using Digitized Models

Rüdiger Dillmann, University of Karlsruhe, Faculty of Informatics, Institute of Anthropomatics, Humanoids and Intelligence Systems Laboratories, Kaiserstr. 12, 76131 Karlsruhe, Germany

Yoshihiko Nakamura, Tokyo University Fac. Engineering, Dept. Mechano-Informatics, 7-3-1 Hongo, Bukyo-ku Tokyo, 113-8656, Japan

Stefan Schaal, University of Southern California, Department Computer Science, Computational Learning & Motor Control Lab., Los Angeles, CA 90089-2905, USA

David Vernon, Khalifa University Department of Computer Engineering, PO Box 573, Sharjah, United Arab Emirates

Author

Dr.-Ing. Edilson de Aguiar

Carnegie Mellon University

Disney Research Pittsburgh

4615 Forbes Avenue

Pittsburg, PA 15213

USA

E-mail: edilson@disneyresearch.com

ISBN 978-3-642-10315-5

e-ISBN 978-3-642-10316-2

DOI 10.1007/978-3-642-10316-2

Cognitive Systems Monographs

ISSN 1867-4925

Library of Congress Control Number: 2009940444

©2010 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable for prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

5 4 3 2 1 0

springer.com

To
Aparicio de Aguiar
and
Maria Auxiliadora de Aguiar

Foreword

Vision is our strongest sense. It enables us to quickly perceive and analyze our surroundings such that we can find our way around, recognize people and places, and avoid potential dangers. However, this rather functional way of looking at the human visual sense only grasps a fraction of the rich variety of sensual experiences that are channeled through optical stimuli. Visual perception is not only a tool for us but it can also induce great emotions, for instance if we are looking at a painting we like, or when we are intrigued by the visual effects of a feature film. The field of computer science that aims at algorithmically modeling these aspects of the human visual system is called visual computing. The field of visual computing subdivides in several more specific research areas, and i will briefly look at two of them in the following.

Researchers in computer vision and artificial intelligence are trying to equip computers and autonomous systems with visual analysis capabilities that match the ones of real humans. In recent years, this area of research has seen tremendous progress. Computers can nowadays perform optical recognition tasks, objects or people in image sequences can be tracked, and optical scene analysis can provide control inputs to steer autonomous vehicles. However, if we compare the performance, robustness and application range of even the best computer vision system today to the abilities of the human visual system, we have to humbly conclude that the field of computer vision is still in its infancy.

While computer vision focuses on the functional or reconstruction side of visual computing, computer graphics focuses on the synthesis or display aspect. In recent years algorithms for creating photo-realistic virtual imagery have greatly improved. Nowadays we can simulate entire virtual cities or imaginary foreign planets at a high visual fidelity, albeit at very high computational cost. Unfortunately, the same claim cannot be made for the rendering of virtual humans or virtual actors. Over millions of years the human visual system has developed the ability to quickly assess other humans, and it thus unmask in a glimpse of an eye even the slightest flaw in the appearance of a virtual actor. It is therefore of utmost importance that all aspects of a virtual human, including appearance and lighting effects at the surface, geometry and motion are modeled at the highest possible level of detail. It is no wonder that

achieving such a high level of quality comes at a high price which can be measured in several man months of work for animators.

Animation professionals can resort to a set of acquisition tools which help them to measure certain aspects of a virtual human, e.g. his motion or his shape from real world subjects. Laser triangulation scanners can acquire full-body geometry in a static pose. Marker-based motion capture systems can be employed to measure skeletal motion parameters, but are often cumbersome to use since they require the captured subject to wear optical markings on the body. Recently, in the field of computer vision we have also seen marker-less capturing systems which do not require such optical beacons and only expect multi-view video of an actor as input. However, most of these systems do not capture more than skeletal motion, require controlled recording conditions, and fail to reconstruct actors in normal everyday clothing (e.g. a skirt). It is thus fair to conclude that even state-of-the-art measurement technology only captures a small subset of the complexity of a moving human's appearance. Ideally, one would want to unify the process and measure much richer performance models, i.e. representations of detailed time-varying geometry and appearance using just one set of multi-view video recordings.

This book presents such joint reconstruction and display approaches that enable easier capturing of more complete scene models, as well as more faithful rendering of virtual humans. It describes a variety of significant leaps forward which the fields of human performance capture and animation have recently seen. In my opinion the book takes a refreshingly different perspective on cognitive system modeling in general and visual computing in particular. It does not focus on only the vision or graphics side but convincingly demonstrates that a joint investigation of both facets enables significant technological progress. I am convinced that also in other fields of computer science and cognitive system modeling such a more integrated approach to solving hard problems will prove highly successful.

Saarbruecken, Germany
September 2009

Christian Theobalt
MPI Informatik

Preface

In computer graphics, it is still challenging to authentically create virtual doubles of real-world actors. Although the interplay of all steps required by the traditional skeleton-based animation pipeline delivers realistic animations, the whole process is still very time-consuming. Current motion capture methods are not able to capture the time-varying dynamic geometry of the moving actors and need to be integrated with other special acquisition techniques. Furthermore, dealing with subjects wearing arbitrary apparel is still not possible. Another problem is that, even if it was possible to capture mesh animations, it would still be difficult to post-process or modify them. Not many papers in the literature have looked into this problem so far.

In this book, we propose algorithms to solve these problems: first, we describe two efficient techniques to simplify the overall animation process. Afterwards, we detail three algorithmic solutions to capture a spatio-temporally coherent dynamic scene representation even from subjects wearing loose and arbitrary everyday apparel. At the end, we also propose two novel algorithms to process mesh animations. By this means, real-world sequences can be accurately captured and transformed into fully-rigged virtual characters and become amenable to higher-level animation creation, e.g. by applying non-photorealistic rendering styles.

This book consists of four parts:

- Part I begins with the description of some general theoretical background information and elementary techniques shared by many projects in this book. Thereafter, the studio used to acquire the input data for the projects described in this book is presented.
- Part II reviews the steps involved in the traditional skeleton-based character animation paradigm and proposes two mesh-based alternatives to simplify the overhead of the conventional process. Both techniques can be directly integrated in the traditional pipeline and are able to generate character animations with realistic body deformations, as well as transfer motions between different subjects.
- Part III describes three algorithmic variants to passively capture the performance of human actors using a deformable model as underlying scene representation from multiple video streams. The algorithms jointly reconstruct spatio-temporally

coherent time-varying geometry, motion, and textural surface appearance even from subjects wearing everyday apparel, which is still challenging for related marker-based or marker-free systems. By using the acquired high-quality scene representations, we also developed a system to generate realistic 3D Videos.

- Part IV proposes two novel techniques to simplify the processing of mesh animations. First, an automatic method to bridge the gap between the mesh-based and the skeletal paradigms is presented. Thereafter, a method to automatically transform mesh animations into animation collages, a new non-photorealistic rendering style for animations, is proposed.

Although the methods described in this book are usually tailored to deal with human actors, their fundamental principles can also be applied to a larger class of subjects. Each method described here can be regarded as a solution to a particular problem or used as a building block for a larger application. All together, they exceed the capabilities of many related capture techniques and form a powerful system to accurately capture, manipulate, and realistically render real-world human performances.

Pittsburgh, PA
September 2009

Edilson de Aguiar

Acknowledgements

This book is adapted from my PhD dissertation and it would not have been possible without the help and support of many people. First of all, I would like to thank my supervisors Prof. Dr. Hans-Peter Seidel and Prof. Dr. Christian Theobalt. Prof. Seidel gave me the opportunity to work in an excellent and inspiring environment as the Max-Planck-Institut für Informatik (MPI) and supported me pursuing my own research interests.

I am also grateful to Prof. Theobalt, who always had time to discuss ongoing and future projects. I also thank him for the invaluable scientific guidance in my research. He has been working with me since the beginning of my PhD and we worked together on all projects described in this book.

Furthermore, I would like to thank Prof. Dr. Marcus Magnor for being my senior supervisor during the beginning of my PhD, Prof. Dr. Sebastian Thrun for hosting me in Stanford during my exchange research visit, and Prof. Dr. Jessica K. Hodgins who have agreed to be part of my graduation committee.

Special thanks go to all my former colleagues in the Computer Graphics Group at the MPI. Their cooperation, competence, creativity, and steady motivation makes the MPI the special place it is. In particular, I owe thanks to Naveed Ahmed, Christian Rössl, Carsten Stoll and Rhaleb Zayer, who were co-authors on some of my papers, Hitoshi Yamauchi for the support with the geometric modeling library, Andreas Pomi for helping with the studio.

I also thank Mayra Castro for the support during the production of this book and the people who kindly allowed me to record and scan them for my research and to many colleagues at the MPI for proofreading parts of my original dissertation. I am also grateful to the secretaries, the non-scientific employees of the institute, and the helpdesk team.

I also acknowledge the Max-Planck Center for Visual Computing and Communication, the International Max-Planck Research School for Computer Science, and the EU-Project 3DTV within FP6 under Grant 511568 for their partial financial support during my PhD studies.

Finally, I would like to thank my whole family and in particular my parents, Aparicio de Aguiar and Maria Auxiliadora de Aguiar, who always encouraged and supported me.

Contents

1	Introduction	1
1.1	Main Contributions and Organization of the Book	2
1.1.1	Part I - Background and Basic Definitions	3
1.1.2	Part II - Natural Animation of Digitized Models	3
1.1.3	Part III - Towards Performance Capture Using Deformable Mesh Tracking	4
1.1.4	Part IV - Processing Mesh Animations	4
	References	5
 Part I: Background and Basic Definitions		
2	Preliminary Techniques	9
2.1	The Camera Model	9
2.1.1	Mathematical Model	9
2.1.2	Camera Calibration	11
2.1.3	Geometry of Stereo Cameras	11
2.2	Modeling Humans	12
2.2.1	Modeling the Shape	12
2.2.2	Modeling the Appearance	13
2.2.3	Modeling the Kinematics	13
2.3	Computer Vision Algorithms	13
2.3.1	Background Subtraction	13
2.3.2	Optical Flow	14
2.3.3	Scene Flow	16
2.3.4	Image Features	17
	References	17

3	Interactive Shape Deformation and Editing Methods	19
3.1	Related Work	20
3.2	Mesh Editing Techniques	21
3.2.1	Guided Poisson-Based Method	21
3.2.2	Guided Laplacian-Based Method	23
3.3	Iterative Volumetric Laplacian Approach	24
	References	25
4	Recording Studio: Data Acquisition and Data Processing	29
4.1	Related Acquisition Facilities	29
4.2	Recording Studio	30
4.2.1	Studio Room	30
4.2.2	Camera System	30
4.2.3	Lighting System	32
4.2.4	Full Body Laser Scanner	32
4.3	Data Acquisition	33
4.3.1	Pre-recording	33
4.3.2	Recording	35
	References	35
 Part II: Natural Animation of Digitized Models		
5	Problem Statement and Preliminaries	39
5.1	Related Work	40
	References	41
6	Poisson-Based Skeleton-Less Character Animation	45
6.1	Overview	45
6.2	Prototype Interface	46
6.3	Animating Human Scans Using Motion Capture Data	48
6.3.1	Mesh-Based Character Animation	48
6.3.2	Video-Driven Animation	50
6.4	Results and Discussion	51
	References	54
7	Laplacian-Based Skeleton-Less Character Animation	55
7.1	Overview	56
7.2	Animating Human Scans with Motion Capture Data	57
7.2.1	Mesh-Based Character Animation	57
7.2.2	Video-Driven Animation	59
7.3	Results and Discussion	59
	References	60

Part III: Towards Performance Capture Using Deformable Mesh Tracking

8	Problem Statement and Preliminaries	65
8.1	Related Work	66
8.1.1	Human Motion Capture	66
8.1.2	Dynamic Scene Reconstruction	67
8.1.3	3D Video	68
	References	69
9	Video-Based Tracking of Scanned Humans	75
9.1	Framework	76
9.1.1	Acquisition and Initial Alignment	76
9.1.2	Step A: Purely Flow-Driven Tracking	77
9.1.3	Automatic Marker Selection	78
9.1.4	Step B: Flow-Driven Laplacian Tracking	80
9.2	Results and Discussion	80
	References	86
10	Feature Tracking for Mesh-Based Performance Capture	87
10.1	Overview	88
10.2	Hybrid 3D Point Tracking	89
10.3	Feature-Based Laplacian Mesh Tracking	91
10.4	Results and Discussion	93
	References	99
11	Video-Based Performance Capture	101
11.1	Overview	102
11.2	Capturing the Global Model Pose	103
11.2.1	Pose Initialization from Image Features	104
11.2.2	Refining the Pose Using Silhouette Rims	106
11.2.3	Optimizing Key Handle Positions	108
11.2.4	Practical Considerations	109
11.3	Capturing Surface Detail	109
11.3.1	Adaptation along Silhouette Contours	109
11.3.2	Model-Guided Multi-view Stereo	110
11.4	Results and Discussion	111
11.4.1	Validation and Discussion	112
	References	117
12	High-Quality 3D Videos	119
12.1	Creating 3D Videos	120
12.2	Results and Discussion	122
	References	124

Part IV: Processing Mesh Animations

13 Problem Statement and Preliminaries	127
13.1 Related Work	128
13.1.1 Motion-Driven Mesh Segmentation	128
13.1.2 Skeleton Reconstruction	128
13.1.3 Character Skinning	129
13.1.4 Editing Mesh Animations	129
13.1.5 Shape Matching	130
References	130
14 Reconstructing Fully-Rigged Characters	133
14.1 Overview	134
14.2 Motion-Driven Segmentation	135
14.3 Automatic Skeleton Extraction	137
14.4 Motion Parameters Estimation	139
14.5 Skinning Weight Computation	139
14.6 Results and Discussion	141
References	147
15 Designing Non-photorealistic Animation Collages	149
15.1 Overview	150
15.2 Rigid Body Segmentation	150
15.3 Building Animation Cells	152
15.4 Assembling the Collage	153
15.4.1 Shape Similarity Measure	154
15.4.2 Spatio-temporal Shape Fitting	154
15.5 Animating the 3D Collage	156
15.6 Results and Discussion	159
References	162
16 Conclusions	163
Index	167