



THE UNIVERSITY

of ADELAIDE

Interactive 3D Reconstruction From Video

Benjamin Ward

School of Computer Science

The University of Adelaide

August 2012

Contents

List of Figures	xi
1 Introduction	1
1.1 Contribution	4
1.2 Outline	5
2 Background	7
2.1 Automated reconstruction methods	7
2.1.1 Feature-based reconstruction	8
2.1.1.1 Reconstruction from two views	8
2.1.1.2 Structure from motion	10
2.1.2 Visual hull methods	12
2.1.3 Shape from photoconsistency	15
2.1.4 Model based reconstruction	17
2.1.4.1 Generic primitives	18
2.1.4.2 Predefined models	19
2.1.4.3 Parameterised models	20
2.2 Interactive reconstruction methods	21
2.2.1 Image-based modelling	21
2.2.2 Sketch-based modelling	25
2.3 Applications of 3D reconstruction	26
2.3.1 Uses for architecture	26
2.3.2 Uses for visual effects	27
2.3.3 Uses for augmented reality	29

CONTENTS

3	Fitting models to images	31
3.1	Input to the method	34
3.2	Model definition	36
3.2.1	Model representation	36
3.3	Fitting process	37
3.3.1	3D likelihood functions	39
3.3.1.1	Evaluating distance to the model surface	41
3.3.1.2	3D template function	42
3.3.2	2D likelihood function	44
3.3.2.1	Photoconsistency likelihood	45
3.3.2.2	Evaluating photoconsistency likelihood	45
3.3.3	Interactive refinement	47
3.4	Fitting process	47
3.4.1	Initialisation	47
3.4.2	Sampling model parameters	49
3.4.3	3D likelihood evaluation	50
3.4.3.1	Optimising for 3D likelihood	51
3.4.4	2D likelihood evaluation	53
3.4.5	Hypothesis selection	53
3.4.6	Interactive refinement	54
3.5	Experiments	55
3.6	Summary	56
4	Modelling with User-Defined Geometry	61
4.1	Specifying 2D shape	64
4.1.1	Bounded plane model	66
4.2	From 2D sketching to 3D models	66
4.3	Single face fitting	67
4.3.1	Generating hypotheses	67
4.3.2	Selecting feature points	69
4.3.3	3D likelihood function	69
4.3.4	2D likelihood function	70
4.3.5	Frame selection	70

4.3.6	Appearance testing	72
4.3.7	Optimising the selected face	74
4.3.8	Manually refining fitting results	74
4.4	Experiments	75
4.5	Fitting for multiple faces	77
4.5.1	Generating hypotheses with multiple faces	78
4.6	Fitting with sparse points	79
4.6.1	Generating hypotheses with sparse features	79
4.6.2	Frame selection	83
4.7	2D modelling interface	86
4.7.1	Specifying vertices	86
4.7.2	Finding faces	86
4.7.3	Specifying shape with a sketch	87
4.7.3.1	Clustering vertices	87
4.7.3.2	Finding faces	89
4.7.3.3	Fitting edges to the image	89
4.7.4	Editing tools	89
4.7.5	Adding curves to faces	89
4.8	Results	91
4.9	Summary	91
5	Modelling With Geometric Constraints	95
5.1	Extrusion and mirroring	96
5.1.1	Defining the operations	97
5.1.2	Model definition	98
5.1.3	Generating hypothesised models	99
5.1.4	Likelihood functions	99
5.2	Modelling with primitive types	100
5.2.1	Primitive shape models	101
5.2.2	Interactively defining a primitive	101
5.2.3	Generating primitive hypotheses	102
5.2.4	Evaluating likelihood for primitives	102
5.2.5	Experiments	103

5.3	Defining faces with geometric constraints	104
5.3.1	Plane constraints	105
5.3.2	Boundary constraints	107
5.4	Summary	109
6	Modelling for Augmented Reality	111
6.1	Video tracking for AR	112
6.2	Modelling for AR environments	114
6.3	A method for in-camera modelling	116
6.3.1	Method overview	118
6.3.2	Graph cut segmentation	119
6.3.3	Segmentation for successive frames	120
6.3.4	Initial model generation	122
6.3.5	Generating the coarse model	123
6.3.6	Initialising the graph cut with camera interaction	124
6.3.7	Volume model generation	127
6.3.8	Results	128
6.4	Summary	128
7	Conclusions	133
7.1	Summary	133
7.2	Directions for future research	135
7.3	Conclusion	136
	Bibliography	139

Abstract

This thesis explores several approaches to 3D reconstruction from video sequences in which the reconstruction process is aided by information about the scene provided interactively by a user. This user-supplied information may describe, for example, the types of objects in the scene and their positions, the boundaries of faces of an object in one view, geometric properties of an object, or geometry which is not seen in the video sequence.

By incorporating this information into the modelling process, we can reconstruct scenes which would be difficult or impossible to reconstruct with a fully automated process due to elements of the scene having minimal texture, or translucent or reflective surfaces, or due to significant parts of the scene being occluded or poorly visible in all views. These interactive methods allow the user to reconstruct the desired parts of the scene to the desired level of detail, with automated processing minimising the required interaction.

We first propose a method in which individual objects in a video sequence are reconstructed by fitting pre-defined model types selected by a user. A novel fitting process is used to efficiently evaluate and optimise models sampled from the large space of possible models. Models are evaluated using a novel combination of user-supplied information, 2D image information, and 3D point cloud data recovered with a Structure from Motion process.

A method is also presented for polygonal modelling of objects in video sequences. This method allows the user to define the faces of the model in a single frame through an intuitive sketch-based interface. An automated process generates a 3D model from this set of 2D faces. Interactive techniques are also described for generating a complete model from a partial model

of an object, for fitting primitive shapes, and for incorporating geometric constraints into the modelling process.

We demonstrate the use of this polygonal modelling method for rapidly generating models in Augmented Reality environments. We then describe an additional method for Augmented Reality applications in which the camera is used as the input device. In this method interaction with the camera is used first to select an object in a scene, and then to provide sufficient views of the object for a complete reconstruction.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holders of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Benjamin Ward

August 14 2012

Publications

In carrying out the research which contributed to this thesis, a number of papers were published.

- A. van den Hengel, A. Dick, T. Thormählen, B. Ward, P. H. S. Torr. Fitting multiple models to multiple images with minimal user interaction. In *Proceedings of the International Workshop on the Representation and Use of Prior Knowledge in Vision (WRUPKV)*, May 2006
- A. van den Hengel, A. Dick, T. Thormählen, B. Ward, P. H. S. Torr. Hierarchical model fitting to 2D and 3D data. In *Proceedings of the 3rd International Conference on Computer Graphics, Imaging and Visualisation*, July 2006

- A. van den Hengel, A. Dick, T. Thormählen, B. Ward, P. H. S. Torr. Building Models of Regular Scenes from Structure-and-Motion. In *Proceedings of the British conference on Machine vision (BMVC '06)*, September 2006
- A. van den Hengel, A. Dick, T. Thormählen, B. Ward, P. H. S. Torr. Rapid Interactive Modelling from Video with Graph Cuts. In *Proceedings of Eurographics 2006*, September 2006
- A. van den Hengel, A. Dick, T. Thormählen, B. Ward, P. H. S. Torr. VideoTrace: rapid interactive scene modelling from video. *ACM Transactions on Graphics*, 26, July 2007
- A. van den Hengel, A. Dick, T. Thormählen, B. Ward, P. H. S. Torr. Interactive 3D Model Completion. In *Proceedings of Digital Image Computing: Techniques and Applications 2007*, December 2007
- A. van den Hengel, R. Hill, B. Ward, and A. Dick. In situ Image Based Modelling. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2009*, October 2009
- J. Bastian, B. Ward, R. Hill, A. van den Hengel, and A. Dick. Interactive Modelling for AR Applications. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2010*, October 2010
- B. Ward, S. B. Kang, and E. P. Bennett. Depth Director: A System for Adding Depth to Movies, *IEEE Computer Graphics and Applications*, 31, Jan-Feb 2011

Acknowledgements

First and foremost, I wish to thank my supervisors Anton van den Hengel and Anthony Dick for their excellent guidance throughout my candidature. My thanks as well to my colleagues Rhys Hill, John Bastian, Daniel Pooley, and Thorsten Thormählen for all of their help and advice. I am also extremely grateful to my family and friends for all the support they have given me over the years. I would especially like to thank my partner Kate Hoffmann for her endless patience, and for being a constant source of encouragement and inspiration.

List of Figures

2.1	Projective camera geometry	8
2.2	Stereo camera geometry	9
2.3	Reconstructing the visual hull	13
2.4	A 3D model in Google Earth	26
2.5	Frame of a video containing a synthetic copy of an object	27
3.1	Results of SfM reconstruction	35
3.2	Mesh model for a toy ambulance.	36
3.3	Distances from points to a cube model	40
3.4	Cube model histogram	43
3.5	Sample points generated for the cube model	46
3.6	Cube test sequence	48
3.7	Manually selecting an object	48
3.8	Positions and scales for hypothesised models	50
3.9	The top ≈ 100 results for the template likelihood on the cube sequence .	51
3.10	Optimising the 3D likelihood for the ambulance model	52
3.11	Optimising the 2D likelihood for the ambulance model	52
3.12	2D optimisation results for one section of the model	53
3.13	Manually refining the fit for the cylinder model	54
3.14	Test results for model fitting	58
3.15	Test results for stages of fitting the cube model	59
4.1	Specifying shape with a drawing interaction	65
4.2	Model face fitted to the images	68
4.3	A selection of the hypotheses generated for a model face	68

4.4	Frame selection	71
4.5	Manually refining a face	74
4.6	Test results for face fitting	76
4.7	Fitting for multiple faces	80
4.8	Selecting features outside of the face	81
4.9	Sampled planes for the selected features	81
4.10	Test results for fitting with sparse features	85
4.11	Specifying shape with freehand lines	87
4.12	Clustering vertices	88
4.13	Specifying a curve by fitting to a drawn curve	90
4.14	Specifying a curve by dragging an edge of the model	90
4.15	Models generated for a range of sequences	92
5.1	A car with symmetric structure	96
5.2	Initialising the mirroring operation for the car model	98
5.3	A selection of the hypotheses generated in mirroring	99
5.4	Mirroring the geometry of the car	100
5.5	Test results for fitting primitives	104
5.6	Selecting an image region to define an orientation constraint	105
5.7	Test results for fitting with boundary constraints	108
6.1	The PTAM interface	112
6.2	Modelling in an AR environment	113
6.3	A synthetic car model occluded by a real guitar	115
6.4	A synthetic car model casting a shadow onto a real ramp and toy ambulance	116
6.5	The graph cut region and resulting silhouette	121
6.6	Automatic model generation	124
6.7	Initialising the graph cut	125
6.8	Silhouettes for a tracked object in several views, and the resulting model	126
6.9	Timing results (in seconds) for the stages of the modelling process . . .	129
6.10	Replicated objects in an AR space	129
6.11	Replicating reflective objects	130
6.12	Real object occluding a synthetic object	130

Notation

S	Scalar
M	Matrix
\mathbf{X}	Vector
(X_1, \dots, X_n)	Elements of vector \mathbf{X}
\mathbf{x}	2D points and vectors
$\Pr(X)$	Probability of X
$\Pr(X Y)$	Conditional probability of X given Y
$\mathcal{L}(X, Y)$	Likelihood function