THE UNIVERSITY OF ADELAIDE

School of Computer Science

Reconstructing 3D Geometry from Multiple Images via Inverse Rendering

John William Bastian

December, 2007

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE FACULTY OF ENGINEERING, COMPUTER & MATHEMATICAL SCIENCES

ABSTRACT

An image is a two-dimensional representation of the three-dimensional world. Recovering the information which is lost in the process of image formation is one of the fundamental problems in Computer Vision. One approach to this problem involves generating and evaluating a succession of surface hypotheses, with the best hypothesis selected as the final estimate. The fitness of each hypothesis can be evaluated by comparing the reference images against synthetic images of the hypothesised surface rendered with the reference cameras.

An infinite number of surfaces can recreate any set of reference images, so many approaches to the reconstruction problem recover the largest from this set of surfaces. In contrast, the approach we present here accommodates prior structural information about the scene, thereby reducing ambiguity and finding a reconstruction which reflects the requirements of the user. The user describes structural information by defining a set of primitives and relating them by parameterised transformations. The reconstruction problem then becomes one of estimating the parameter values that transform the primitives such that the hypothesised surface best recreates the reference images.

Two appearance-based likelihoods which measure the hypothesised surface against the reference images are described. The first likelihood compares each reference image against an image synthesised from the same viewpoint by rendering a projection of a second image onto the surface. The second likelihood finds the 'optimal' surface texture given the hypothesised scene configuration. Not only does this process maximise photo-consistency with respect to all reference images, but it prohibits incorrect reconstructions by allowing the use of prior information about occlusion. The second likelihood is able to reconstruct scenes in cases where the first is biased.

DECLARATION

This work does not contain material that has been accepted for the award of any other degree or diploma in any University or other tertiary institution. To the best of my knowledge and belief, it does not contain any material previously published or written by another person, except where due reference has been made in the text.

I give consent that this Dissertation to be made available by the University Library for loan and photo-copying.

I acknowledge that copyright of published works contained within this thesis (as listed below) resides with the copyright holder/s of those works.

Bastian, J. W. and van den Hengel, A. J. , Computing image-based re-projection error on graphics hardware, *Proceedings of the Seventh International Conference on Digital Image Computing: Techniques and Applications, (DICTA 2005)*, Cairns, Australia, 6-8 December 2005, IEEE Computer Society 2005

Bastian, J. W. and van den Hengel, A. J. , Computing image-based re-projection error on graphics hardware, *Proceedings of the Sixth International Conference on Digital Image Computing: Techniques and Applications, (DICTA 2003)*, Macquarie University, Sydney, Australia, 10-12 December 2003, CSIRO Publishing 2003

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ALGORITHMS

ACKNOWLEDGEMENTS

I would like to thank my supervisors Mike Brooks and Anton van den Hengel, and to Anthony Dick who offered advice and encouragement. My thanks also goes to Thorsten Thormählen, whose help with calibration was very much appreciated, and Darren Gawley for questioning the existence of zero-dimensions. I would also like to thank Rhys Hill, Daniel Pooley and Travis Olds for many interesting discussions and shared experiences.

This would not have been possible without the support and understanding of my family, and I thank them for their encouragement, empathy and quiet patience.

NOTATION

Probability

$\Pr(X)$	the probability of event X
$\Pr(X \mid Y)$	the conditional probability of event X given Y
$\mathscr{L}(X \mid Y)$	the likelihood that X given Y
$\mathcal{U}(a, b)$	uniform distribution in the range $[a, b]$
$\mathcal{N}(\mu, \sigma)$	normal distribution with mean μ and standard deviation σ

Images

$\mathsf{x}(\boldsymbol{p})$	the colour of the point $\boldsymbol{p}$ in the image x
$\ddot{\mathsf{x}}$	the RGB channels of image x
$\dot{\mathsf{x}}$	the α-channel of image x
r_i	the i^{th} reference image
$\mathcal{I}$	the set of reference images, $\mathcal{I} = \{\mathsf{r}_i\}$
$\mathsf{s}_{s,t}$	the synthetic image of the hypothesised surface rendered by camera t while using r_s as the surface texture.
v_i	the view-specific texture associated with r_i
t	the estimated surface texture
$\mathcal{C}$	the set of surface textures, $\mathcal{C} = \{t_i\}$

Graphics Pipe

$\boldsymbol{P}_{\mathcal{G}}$	OpenGL pipe-line's projection matrix
$\boldsymbol{M}_{\mathcal{G}}$	OpenGL pipe-line's modelview matrix
$\boldsymbol{T}_{\mathcal{G}}$	OpenGL pipe-line's texture matrix

Geometry

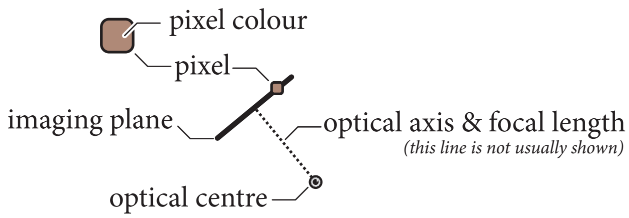
$\underline{\boldsymbol{p}}$	Euclidean 3D points and vectors
$\boldsymbol{M}$	matrix
$\boldsymbol{M}_n$	row n of matrix $\boldsymbol{M}$
$\mathcal{P}$	a geometric primitive
$\mathcal{S}$	the scene-graph
ρ	scene-graph parameter
Ψ	the scene-graph parameter vector, $\Psi = \{\rho_i\}$
$\mathcal{S}(\Psi)$	an instance of the scene-graph parameterised by the vector Ψ
$\mathfrak{t}$	triangle, $\mathfrak{t} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$

Projective Geometry

$\boldsymbol{p}$	homogeneous 3D point
$\boldsymbol{T}_{xyz}$	4×4 translation matrix; subscript indicates free axes
$\boldsymbol{R}_{xyz}$	4×4 rotation matrix; the subscript indicates free axes
$\boldsymbol{S}_{xyz}$	4×4 scale matrix; the subscript indicates free axes
$\boldsymbol{K}$	intrinsics matrix
$\boldsymbol{E}$	extrinsics matrix
$\boldsymbol{P}$	3×4 projection matrix to map 3D homogeneous points to 2D homogeneous points, $\boldsymbol{P} = \boldsymbol{KE}$
$\hat{\boldsymbol{K}}$	4×4 projection matrix to map 3D homogeneous points to eye co-ordinates
$\hat{\boldsymbol{P}}$	graphics projection matrix, $\hat{\boldsymbol{P}} = \hat{\boldsymbol{K}}\boldsymbol{E}$

Illustration

camera illustration



"The time has come," the Walrus said,
"To talk of many things."

Lewis Carroll
Through the Looking-Glass