Proteus – Semi-Automatic Interactive Structure-from-Motion

Malte Schwarzkopf and Christian Richardt

University of Cambridge Computer Laboratory
ms705@cam.ac.uk, christian.richardt@cl.cam.ac.uk

Abstract

Proteus is a semi-automatic structure-from-motion implementation that makes it straightforward – even for novice users – to obtain 3D models of real-world objects from digital video sequences. This is facilitated by the user tracing object outlines in key frames of the video. The process of doing so is supported by a novel guidance technique that interactively “snaps” the annotations to object outlines. As a result, Proteus can be used for rapid interactive capture of 3D models that can be used in virtual and augmented reality scenarios such as computer games, online worlds and urban modelling.

1 Introduction

Recently, user-created 3D content has become increasingly ubiquitous in virtual worlds such as Google Earth or SecondLife. Nonetheless, the process of building a 3D model of a real-world object generally still involves tedious manual labour.

Structure-from-motion is a promising approach to automate the model-building process, but existing research implementations suffer from a lack of accessibility to non-expert users. With *Proteus*, we present a novel approach that unifies automatic structure-from-motion and interactive user input, making use of the human perception of an object to improve results. At the same time, an intuitive front-end allows a user to merely trace the outlines of an object in order to generate a 3D model of it.

While previous implementations attempted to solve the correspondence as well as the shape reconstruction problem, *Proteus* takes an interdisciplinary approach and non-intrusively guides the user in providing the correspondence information.

Proteus also encompasses an efficient C++ implementation of the shape reconstruction algorithms, as well as various ways of exporting and displaying the reconstructed model.

2 Related Work

There is a breadth of existing literature and research on the subject of structure-from-motion, such as the extensive textbook by Hartley and Zisserman [2].

The closest related work is the VideoTrace project [4], which focuses on rapid interactive modelling of complex structures with a high degree of detail. It, however, differs from the approach we take with *Proteus* in that the structure-from-motion analysis is performed before interactive user input, and not based on the latter. Furthermore, VideoTrace does not support guidance techniques such as *Proteus*’ edge “snapping”.

3 Line Snapping

In the *Proteus* framework, features such as corners, edge joints, or distinctive points in the object texture, are identified and matched by the user using a simple GUI-based point-and-click mechanism.

One of the most difficult and tedious parts of the process of annotating a video frame is to place annotations with pixel accuracy.

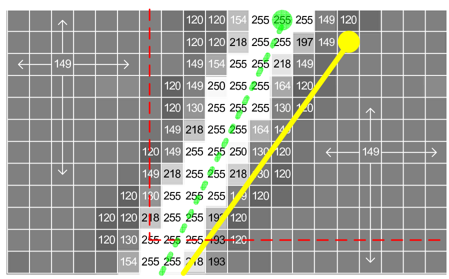
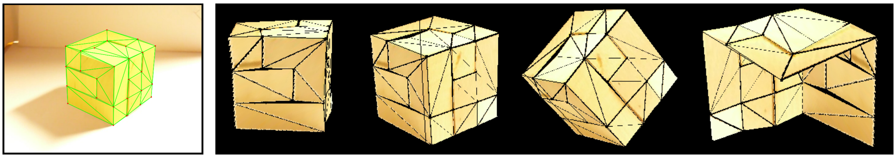


Figure 1: Magnified view of the snapping calculations. The dashed red line is the boundary of the 10×10 *snapping window, the green line is the line corresponding to the optimal snapping score, the yellow circle is where the user is pointing.*

VMV 2009 M. Magnor, B. Rosenhahn, H. Theisel (Editors)



(a) Annotated input frame (b) Metric reconstruction of the cube, various views

Figure 2: Example of a 3D model reconstructed using Proteus*.*

Hence a way of guiding the user in a non-intrusive, interactive fashion is required. We developed a novel mechanism for *Proteus* to "snap" lines to edges in the video frames, based on a technique by Agarwala [1].

Edges in the input images are detected using Sobel edge detectors and the results combined into a gradient magnitude ($\mathbf{g} = \sqrt{\Delta x^2 + \Delta y^2}$), so that a *gradient field* is obtained for the image. When placing annotations, an approximation to the line integral over the gradient field is computed for all pixels within a small window. This value is normalised by edge length and maximised over the window considered, yielding the coordinates of the line endpoint with the greatest average per-pixel gradient. Figure 1 illustrates this with a magnified example.

4 Reconstruction

Once correspondences between features in subsequent frames have been established, structure-from-motion is applied to infer 3D shape.

We implemented the *normalised 8-point algorithm* for finding the fundamental matrix – this requires a minimum of eight distinct feature correspondences in order to be able to find the fundamental matrix, subject to a singularity constraint.

The two camera models used in *Proteus* are *projective* and *metric* cameras, for the initial and the upgraded reconstructions respectively.

In order to "upgrade" the initial projective reconstruction, a small set of ground truth coordinates are specified by the user. These are then used for a *direct upgrade* to a metric reconstruction. Canonical approaches require a homography relating the ground truth to reconstructed points to be inferred. This is, however, overly complicated. We developed and implemented a new simple yet generic algorithm that determines the metric camera matrices directly, without explicitly solving for a homography $\mathtt{H}$.

Proteus offers both an internal OpenGL-based model viewer and a facility to export models into the PLY mesh format. For a more realistic reconstruction, textures are extracted from the video and imposed onto the generated 3D model. An example of a reconstructed model is shown in Figure 2.

5 Implementation

Proteus is implemented in C++, with a GUI built using the Qt toolkit. It is licensed under the GPL and the source code can be downloaded from the project website [3]. Future extensions will aim to look into implementing ways of performing the upgrade to a metric reconstruction which do not require ground truth information, as well as possibly adding further 3D primitives.

6 Evaluation

The evaluation that we undertook on *Proteus* is twofold: empirical evaluation to assess performance and accuracy, and a user study to investigate the subjective appeal and usability of the novel annotation guidance technique. We found that there is a significant user preference for interactive guidance.

References

[1] AGARWALA, A. SnakeToonz: A semi-automatic approach to creating cel animation from video. In *Proc. Intl. Symp. Non-Photorealistic Animation and Rendering* (2002), pp. 139–146.

[2] HARTLEY, R. I., AND ZISSERMAN, A. *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press, ISBN: 0521540518, 2004.

[3] SCHWARZKOPF, M. Proteus project website. `http://proteus.malteschwarzkopf.de`.

[4] VAN DEN HENGEL, A., DICK, A., THORMÄHLEN, T., WARD, B., AND TORR, P. H. S. VideoTrace: rapid interactive scene modelling from video. *ACM Transactions on Graphics 26*, 3 (2007), 86.