Regular Scene Reconstruction from Image Sequences

Thorsten Thormählen, Anthony Dick, Anton van den Hengel, Ben Ward
School of Computer Science
University of Adelaide
Australia
Email: thorsten@cs.adelaide.edu.au

Philip H. S. Torr
Department of Computing
Oxford Brookes University
UK
Email: philiptorr@brookes.ac.uk

Introduction

Most of us live our daily lives in man-made environments that contain obvious regularity and repetition. This video shows work in progress [1] that is specifically geared towards modelling such scenes from a set of images. The idea is to use computer vision techniques such as structure and motion estimation, and camera self-calibration, to obtain as much information as possible about the scene automatically. The system then has the facility for the user to interactively specify some high-level information about the scene, such as the type of objects it contains and their relations to each other. Information from the user is combined with that obtained from the vision techniques to generate an graphical parameterised model of the scene. The experience of using this system is rather like using a 3D modelling package that already has some idea of the scene to be modelled, and uses this knowledge to accelerate the modelling process. Visually convincing models of real world scenes can be created extremely rapidly and with far less user interaction than is possible in a modelling package that does not have this information available.

Description of the video

The video shows the system operating on two outdoor scenes that contain some regularity. In the first, a model is acquired of a set of stone bollards. This is quite challenging, as the bollards lack distinctive features and occlude each other. Structure and motion estimation recovers the camera parameters accurately but is not able to reconstruct many points in the foreground area, and the structure inherent in the scene is not obvious from the point based reconstruction.

The user can then select an area in an image as representing a particular kind of shape. In the video, the user is shown selecting a planar region: the ground plane. By selecting an area in the image, an area in the reconstruction is also selected. This creates a plane in the model that fits the data both in the image and the reconstruction. The user then selects a cuboid in an image (a bollard). Because there are almost no reconstructed points on the bollard, and it blends into the background in the image, the initial fit is poor. However, the user can also specify constraints between objects. A single button press denotes that the bollard abuts the ground plane. This extra information is enough to automatically pin down the correct position, size and orientation of the bollard.

The video next shows how repeated shapes can be modelled extremely rapidly, simply by drawing a line along the axis of repetition in an image. In this case the constraints on the shape, size and position of the objects is enough to automatically determine how many times the bollard is repeated, and the spacing between repetitions. The repeated bollards are constrained to lie on the ground plane, and in a straight line. Even the most distant bollard is modelled accurately, as information from the most visible shapes is effectively propagated to areas of the scene where less visual information is available.

As well as repeating individual shapes, an entire configuration of shapes can be repeated. This is shown by copying the modelled row of bollards across the image, to model an identical row to the right. Again this is achieved by a single mouse drag. After each interaction, an optimisation procedure ensures that the current model fits both the 2D and 3D data, so that user interaction does not need to be precise. The video then shows one application of the 3D model we have generated: it enables objects to be inserted into the scene and interact convincingly with the modelled structure.

The second example in the video shows the same process applied to an architectural scene. Interestingly, in this case the pedestals are not all the same size. They are still modelled accurately, due to the probabilistic nature of our problem formulation. This makes the system less brittle in the face of slight irregularity, allowing overall regularity to be captured while allowing slight variation in individual cases.

To recap, this video shows how a regular real world scene can be modelled extremely rapidly with the assistance of image data. As an example, a model of a forecourt is created using only one mouse selection for the ground plane, one mouse selection for a bollard, and 2 mouse drags to repeat the bollards. We note that the system is also applicable to scenes with little regularity, at the price of more user interaction.

References

[1] A. van den Hengel, A. Dick, T. Thormaehlen, P. H. S. Torr and B. Ward. Fitting Multiple Models to Multiple Images with Minimal User Interaction, Proceedings of the International Workshop on the Representation and use of Prior Knowledge in Vision (WRUPKV), in conjunction with ECCV'06, May 2006.