3D Modeling from Multiple Views in the Presence of Occlusion

Department of Computer Science University of North Carolina at Chapel Hill March 2007

The Challenge

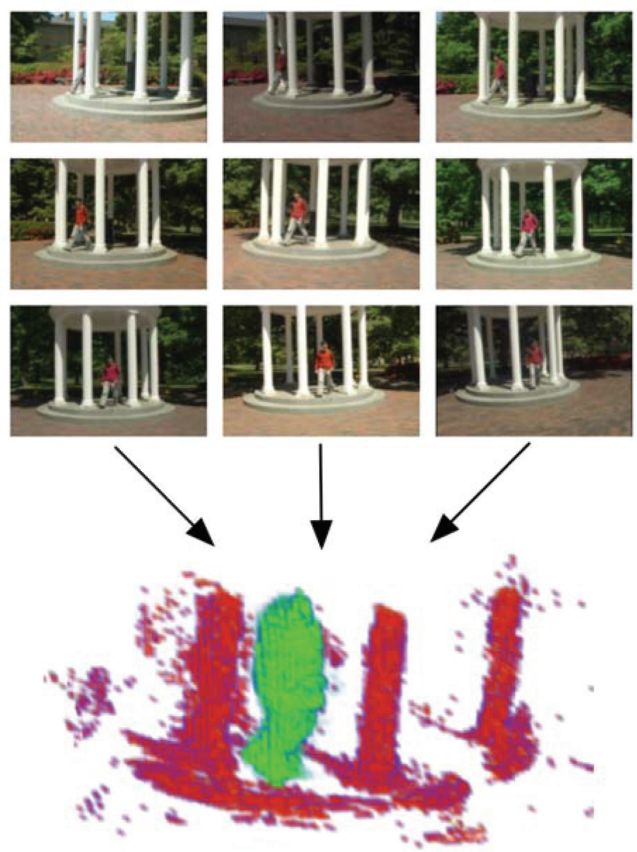
Building 3D models from multiple image sequences is a classical, but difficult computer vision problem. The problem has been well studied by researchers, and existing methods use various cues in images to deduce and construct a geometrical model. The main cues used currently are the segmentation of the silhouettes of objects of interest in images, and color consistency information used to pinpoint and triangulate a surface coherent with the conjunction of observed colors in images. Additional geometrical constraints such as boundary smoothness or prior knowledge about shapes are often used to help fill in the gaps where image data is inconclusive. All existing methods are generally very successful in controlled environments, where the lighting is constrained and the viewing conditions used to obtain images of objects made optimal. They however face substantial difficulties when brought outdoors or in generally unconstrained environments. This is because typical assumptions about lighting and color break in harsher conditions, making segmentation and color consistency reasoning much more difficult and error prone. In addition, dealing with uncontrolled setups means that choosing optimal, occlusionless viewpoints can prove impossible. Our aim is thus to propose 3D modeling methods that drastically improve robustness over existing approaches, and explicitly deal with the occlusion problem.

The Approach

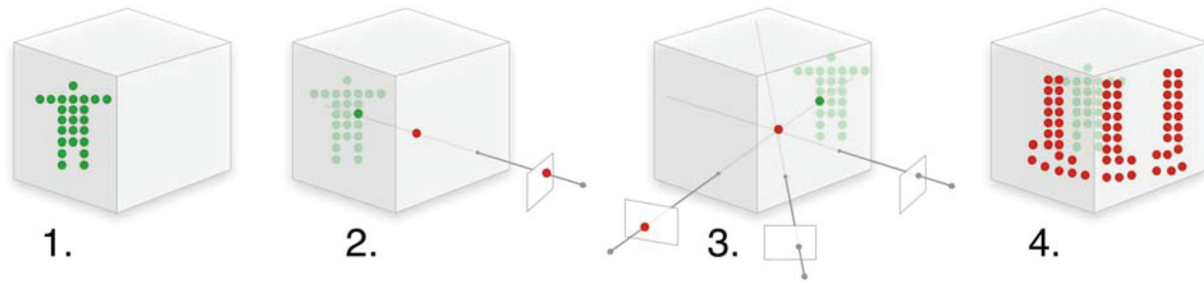
We have proposed approaches toward achieving these goals by focusing on silhouettes of objects of interest in the scene, and by using two ideas. First, we cast the 3D modeling problem into a sensor fusion problem. By using Bayesian probabilistic techniques, it is possible to infer probabilities of occupancy for each voxel on a regular grid, pinpointing the regions likely to be consistent with silhouettes. Second, we explicitly model and reconstruct the 3D shapes of static occluders that appear in the interaction space of objects of interest. We do so by using our best estimate of the shapes of objects of interest and compare it to the silhouettes images actually seen to detect discrepancies caused by occlusion. We then accumulate those occlusion cues in time, by processing all occlusion cues in a multi-view sequence. This knowledge about occluders can be used to simultaneously reconstruct occluders and objects of interest. The process is visually summarized on the following page.

Highlights

- **Propose a new method to take 3D modeling from silhouettes to difficult environments with static occluders.**
- **Use a Bayesian sensor fusion framework to model the uncertainty in the process and gain robustness.**
- **Simultaneously build the shape of objects and static occluders in the scene, by accumulating occlusion cues over time.**



Nine views from video sequences used to infer the 3D shape of an object of interest (person, green) and occluder (Old Well, red).



Technical Summary

1. Use an existing Bayesian sensor fusion method (ICCV 2005) to estimate the shape of dynamic objects of interest, in the form of an occupancy probability grid.
2. Knowledge about dynamic objects is used as prior to build a sensor model, describing how an image pixel forms given the state of the corresponding viewing line. By using Bayesian inference, this sensor model enables to make deductions about the occupancy of an occluder at the red voxel in space, when a dynamic object (green) is behind it. If the corresponding pixel in the image detected the dynamic object's silhouette, then we know the pixel observes the dynamic object and is unobstructed by an occluder. The red voxel's occluder occupancy probability decreases. Conversely if the image pixel did not detect a silhouette despite a dynamic object being on its viewing line, than it is likely that the viewing line is obstructed somewhere in front of the dynamic object by a static occluder: the red voxel's occluder occupancy increases. We label this information "occlusion cues" for this voxel.
3. Repeat this reasoning for all views, and all time instants in the sequence, to benefit from different positions of the dynamic object, which bring more independent verification to make deductions about our voxel. Bayesian sensor fusion enables soft integration of all cues observed by our model, inducing large robustness to noise and momentarily erroneous observations.
4. Do this for every voxel to obtain probabilistic occluder shape estimation. This can in turn be used to improve inference of dynamic objects, by accounting for occlusions.

Project Participants

Marc Pollefeys, associate professor
Jean-Sébastien Franco, postdoctoral research associate
Li Guan, graduate research assistant

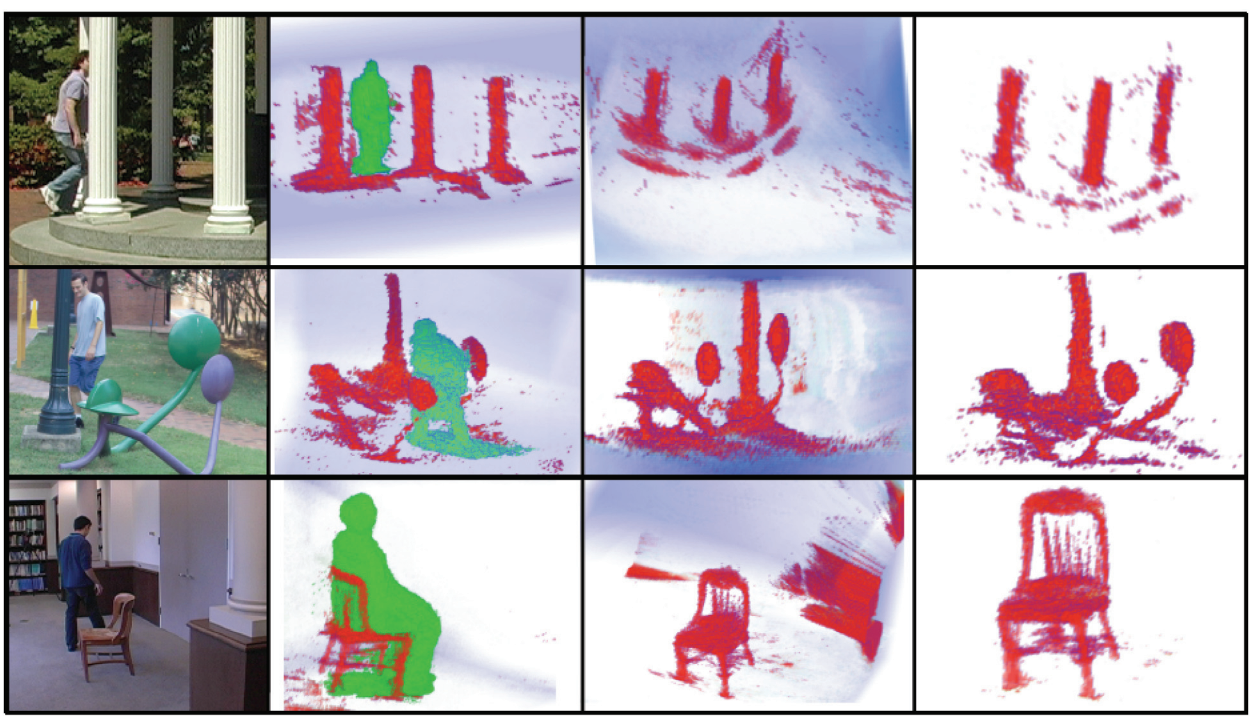
Research Sponsor

Packard Fellowship Grant

Selected Publications

L. Guan, JS. Franco, M. Pollefeys. 3D Occlusion Inference from Silhouette Cues. Submitted, *Computer Vision & Pattern Recognition*, 2007.

JS. Franco, and E. Boyer. Fusion of Multi-View Silhouette Cues Using a Space Occupancy Grid. *Proc Int'l Conf. on Computer Vision*, Oct. 2005 (2)1747–1753.



Resulting shapes obtained for 3 sequences, with object of interest in green and static occluder shape estimate in red: Old Well, Sculpture, and Chair. All models obtained with 9 calibrated views, acquired using standard DV Camera equipment.