Structure from Motion without Correspondence

Frank Dellaert, Steve Seitz, Chuck Thorpe, Sebastian Thrun

Problem: How can we recover the 3D structure of a scene from sparse image features without knowing the correspondence between the features in different images ?

Impact: Reconstructing 3D scene geometry and camera motion from a set of images of a static scene is a primary objective of computer vision. There are countless applications where recovering 3D structure is a key component, from creating virtual worlds for gaming, surveying environments at a distance, to robot navigation in complex environments. We expect our method to significantly expand the conditions under which images can be taken, as we do not require features to be tracked (see below).

State of the Art: The current state of the art provides solutions that apply only under special conditions. Specifically, existing shape recovery techniques are strongly limited by their reliance on error-prone correspondence techniques, or require that features be tracked over time. In the latter case, problems arise when an image sequence turns back on itself, and there is now a correspondence problem between feature tracks. We feel the solution is to solve for the structure and correspondence simultaneously.

Approach: In this project we address the structure from motion problem (SFM) *without* prior knowledge of point correspondence. We frame the problem as finding the maximum likelihood estimate of structure and motion given only the measurements, integrating over all possible assignmen ts of 3D features to 2D measurements. While the full computation of this likelihood function is generally intractable, we propose to use the Expectation-Maximization algorithm (EM) as a practical method for finding its maxima.

The outline of our method is as follows: instead of solving for structure and motion given the original image measurements, we solve a new SFM problem using newly synthesized 'virtual' measurements, computed using our current knowledge about the correspondences. This knowledge comes in the form of a probability distribution, obtained using the actual image data and an initial guess for the structure and motion. By solving this new SFM problem we obtain a new estimate for the structure and motion. This basic step is iterated until convergence. A key step in our method is in computing the probability distribution over correspondences, which is hard to obtain analytically. To circumvent this problem, we use Markov Chain Monte Carlo methods to sample from the distribution, which can be done efficiently.

Future Work: So far we have only worked with sequences of images where all features were hand-picked, and we have not shown results on sequences with occlusions or spurious features. Clearly, in order to be more widely applicable, the performance of the algorithm under these more general conditions needs to be evaluated. This is a practical rather than a theoretical hurdle, as in theory occlusions and spurious features are easily incorporated in the formulation above. However, it does introduce the issue of how many features need to be instantiated, since now this will not be known a priori.

Selected References:

[1] F. Dellaert. Addressing the correspondence problem: A Markov chain Monte Carlo approach. Technical Report CMU-RI-TR-00-11, Robotics Institute, School of Computer Science, Carnegie Mellon University, January 2000.

[2] F. Dellaert, S.M. Seitz, C.E. Thorpe, and S. Thrun. Structure from motion without correspondence. Technical Report CMU-RI-TR-99-44, Robotics Institute, School of Computer Science, Carnegie Mellon University, December 1999.

[3] F. Dellaert, S.M. Seitz, C.E. Thorpe, and S. Thrun. Structure from motion without correspondence. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2000.

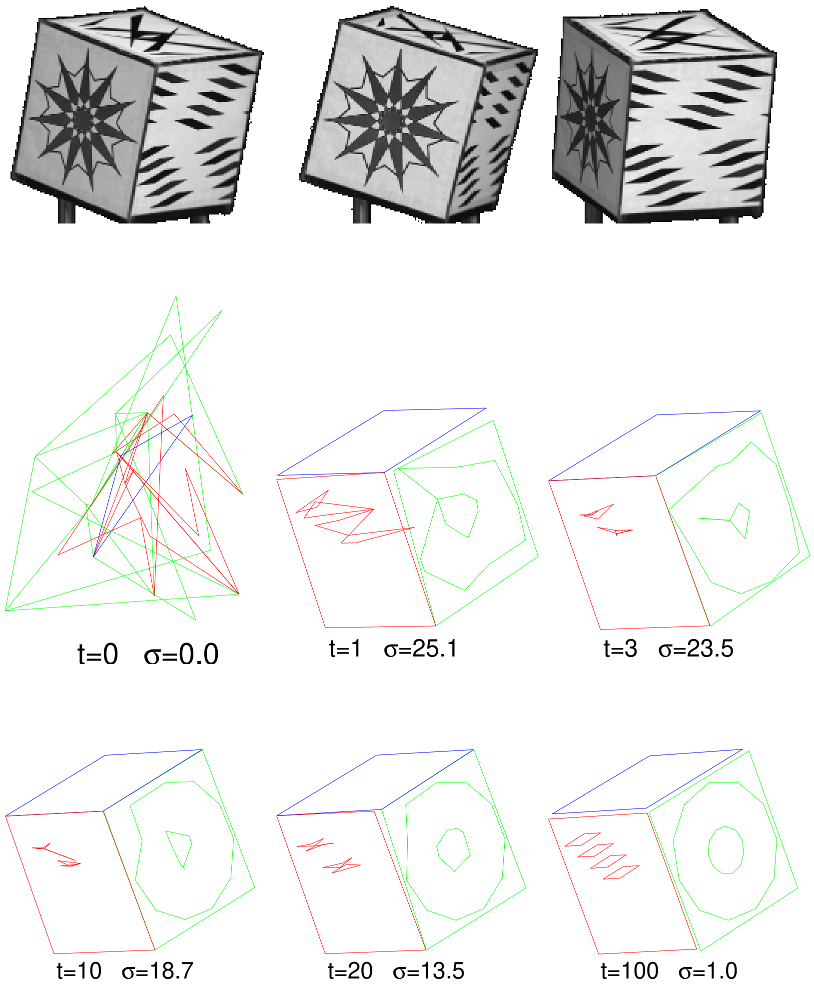

Figure 1: Top: three images of a cube, out of 11 images taken at the Calibrated Imaging Laboratory (CIL) at CMU. Bottom: starting from a random guess for the structure (t=0) we recover gross 3D structure in the very first iteration (t=1). As the annelaing parameter σ is gradually decreased, succesively finer details are resolved (iterations 3,10,20, and 100 are shown).