Variable Elimination for 3D from 2D

By Ji Zhang

Accurately reconstructing the 3D geometry of a scene or object observed on 2D images is a difficult problem: there are many unknowns involved (camera pose, scene structure, depth factors) and solving for all these unknowns simultaneously is computationally intensive and suffers from numerical instability. Here we algebraically decouple some of the unknowns so that they can be solved for independently. Decoupling the pose from the other variables has been previously discussed in the literature. Unfortunately, pose estimation is an ill-conditioned problem. In this paper, we algebraically eliminate all the camera pose parameters (i.e., position and orientation) from the structure-from-motion equations for an internally calibrated camera.
We then also fully eliminate the structure coordinates from the equations.
This yields a very simple set of homogeneous polynomial equations of low degree involving only the depths of the observed points. When considering a small number of tracked points and pictures (e.g., five points on two pictures), these equations can be solved using the sparse resultant method.