

Title

Pengju Jin, Siddhartha Srinivasa

Abstract—Abstract Goes here

I. INTRODUCTION

Introduction

II. RELATED WORK

Related work here

III. APPROACH

In order to obtain reasonable pose of the fiducial tag under noisy condition, we fused the depth information with RGB images of the kinect sensor. Our system uses the depth sensor data to provide an initial pose estimation of the Apriltag, and then further refine the pose using constrained optimization. In this section, we describe how we obtain the initial pose estimation of the tag using depth sensor data and fuse it with rgb information to refine the pose. Our key insight is that we can constrain the pose refinement process by weighting the sensor information with their uncertainty. In order words, we will weight the initial pose estimation more when the depth sensor is good otherwise we will allow the region of refinement to be bigger.

A. Depth Plane Fitting

Since the fiducial tags are planar, we can obtain the pose the tag using depth sensor by fitting a plane over the depth points. Although the corners of the fiducial tags found in the RGB images might be noisy, they are accurate enough to locate the rough region of the tag with at most a few pixels of offset. With a calibrated RGBD sensor, we can extract the patch of depth points containing the planar tag.

The range data retrieved from the Kinect sensor are usually far from perfect. Specifically, borders of the tag and occasionally dark regions of the tag produces highly reliable range data. Therefore, we first filter the data by removing points too far from the median before fitting the plane. In our system, we use a Bayesian plane fitting algorithm (describe the algorithm below) which assumes some noise model of the data and computes the mean and covariance of the plane. The noise model we used is a kinect specific noise model [reference], which increases with the axis rotation of the depth plane. Note that a higher covariance implies that we are more uncertain about our plane and our initial pose estimation.

B. Initial Pose Estimation

Once the plane is computed, we can estimate the pose of the tag. We project the noisy corners of the apriltag back onto the estimated plane to get their 3D coordinates in the camera frame. From here, the pose of the tag is the transformation from the two 3D points correspondence: The 3D projected points of the tag in the camera frame and the 3D points in the tag frame. Since the projected points are no longer going to be a perfect square, we want to find the optimal transformation that minimizes the reprojection error. This essentially becomes a 3-D rigid body transformation estimation problem. In our implementation, we

C. Pose Refinement