Introduction

Goal:
Estimate 3D location and pose of objects, 3D location of points, and camera parameters from 2 or more images.

Motivation:
- Most 3D reconstruction methods do not provide semantic information.
- Most recognition methods do not provide geometry and camera pose.
- We propose to solve these two problems jointly.

Advantages:
- Improve camera pose estimation, compared to feature-point-based SFM.
- Improve object detections given multiple images, compared to independently detecting objects from each single image.
- Establish object correspondences across views.

Joint Likelihood Maximization

Main challenge:
High dimensionality of unknowns $=>$ Sample $P(q,u|O,C,Q)$ with MCMC

Parameter Initialization

- Use object detection scale and pose to initialize cameras relative poses
- Theorem: camera parameters can be estimated given:
  i) 3 objects with scale; ii) 2 objects with pose; iii) 1 object with scale and pose.

Monte Carlo Markov Chain

- Sampling starts from different initializations
- Proposal distribution $P(q,u|C,Q)$
- Combine all samples to identify the maximum

Results

Comparison Baselines
- Camera Pose Est.: Bundler [1]
- Object Detection: LSVM [2]

1. Car Dataset [3] (available online)
- Images and Dense Lidar Points
- ~500 testing images in 10 scenarios

2. Kinect Office Dataset (available online)
- Images and calibrated Kinect 3D range data
- Mouse, Monitor, and Keyboard
- 500 images in 10 scenarios

3. Person Dataset
- A pair of stereo cameras
- 400 image pairs in 10 scenarios

SSFM Problem Formulation

Measurements
- $q$: point features (e.g. DOG+SIFT)
- $u$: point matches (e.g. threshold test)
- $o$: 2D objects (e.g. [2])

Model Parameters (unknowns)
- $C$: camera (K is known)
- $Q$: 3D points (locations)
- $O$: 3D objects (locations, poses, categories)

Intuition:
In addition to point features, measurements of objects across views provide additional geometrical constraints that allow to relate cameras and scene parameters.

Model Overview

$\{O,Q,C\} = \arg \max P(q,u,o|C,O,Q)$

Point Likelihood $P(q,u|C,Q)$

Object Likelihood $P(o|C,O)$
- Estimate 3D object likelihood by 2D projection appearance:
  $P(o|C,O) \propto \prod_{t} \prod_{k} P(o|O_t,C_k)$

Assumption:
Given camera hypothesis, objects and points are independent

Reference


Acknowledgement
We acknowledge the support of NSF CAREER #1054127 and the Gigascale Systems Research Center. We thank Mohit Bagra for collecting the Kinect dataset and Min Sun for helpful feedback.