CS231A Project Proposal

Alex Fu, Lingjie Kong

April 21, 2017

1 Problem

We want to perform 3D reconstruction and recognition through mobile phone cameras in real-time. Specifically, we would like to produce a map representation of 3D reconstructed scenes on object level.

Ideally, recognition techniques can improve reconstruction accuracy and vice versa. A successful implementation can be useful for many downstream applications, such as helping disabled people moving around the street, etc.

2 Data

We will record videos of real life scenes by our own mobile phones cameras. We may also look at existing RGB-D SLAM Dataset and Benchmark for debugging and testing.

3 Method

We will first use open source 3D reconstruction algorithm from Dense Visual SLAM for RGB-D Cameras paper through minimizing both the photo-metric and the depth error over all pixels. We will also use open source feature-based methods to better exploit the pose accuracy as well as propose the open source entropy-based similarity measure for key-frame selection and loop closure detection.

Second, we will use deep learning especially convolutional neural network to train a model based on labeled data for different 3D object. Eventually, we should be able to use the 3D reconstructed model as well as the trained convolutional neural network to reconstructed the 3D scene as well as accurately classify the 3D objects in the scene.

Last, we are thinking about combining the 3D reconstruction as well as the classification into an Expectation Maximization (EM) algorithm to speed both reconstruction as well as classification by letting one boosting the result for the other.

4 Reading

3D reconstruction:

- Dense Visual SLAM for RGB-D Cameras
- ElasticFusion: Dense SLAM without A pose Graph

3D classification:

- Convolutional-Recursive Deep Learning for 3D Object Classification

EM for 3D reconstruction and classification:

- Accelerated Image Reconstruction Using Ordered Subsets of Projection Data

5 Evaluation

First, we will evaluate the performance of our 3D-reconstruction performance based on how many images do we need to clip from one video to fully reconstructed the 3D scene.

Second, we will evaluate the robustness of our 3D-reconsturction algorithm from different environments to see whether our algorithm and handle the variance.

Third, we will evaluate the accuracy of our classification based on a separated test set from training set and validation set to see how accuracy our convolutional neural network can predict the 3D-reconstructed object in one scene.

Last, we will also compare the result regarding training time as well as accuracy from two separated reconstruction and classification steps as well as one nested EM reconstruction and classification boosting algorithm eventually.

6 Progress

We would like to follow the timeline below:

- By May 5: Study and implement 3D reconstruction through SLAM.
- By May 12: Study and implement 3D recognition on top of previously reconstructed scenes.
- Midterm progress report: First pass implementation completed.
- By June 2: Improve reconstruction and recognition results, and implement it on mobile phones in real time.

References

[1] RGB-D SLAM Dataset and Benchmark,
`http://vision.in.tum.de/data/datasets/rgbd-dataset`

[2] Dense Visual SLAM for RGB-D Cameras,
`http://webdiis.unizar.es/~raulmur/orbslam/`

[3] ElasticFusion: Dense SLAM without A pose Graph,
http://www.roboticsproceedings.org/rss11/p01.pdf

[4] Convolutional-Recursive Deep Learning for 3D Object Classification,
https://papers.nips.cc/paper/4773-convolutional-recursive-deep-learning-for-3d-object-classification.pdf

[5] Accelerated Image Reconstruction Using Ordered Subsets of Projection Data,
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=363108