Progress Report

Bardia Mojra

October 3, 2020

Robotic Vision Lab
University of Texas at Arlington

1 Progress

Following items are listed in order of priority:

- Pose Estimation: This week I read [1], it is a comprehensive road map for implementing camera pose estimation. It begins P3P as a classical solution for camera pose estimation where it uses minimal set of point correspondences between 2D measurement in the image to 3D features of the model. Later PnP introduced a non-linear implementation for N point correspondences, in authors' opinion, "the gold standard" solution to PnP problem consist of estimating six parameters of the transformation matrix. The solution is obtained by minimizing the norm of projection error using Gaussian-Newton of a Levenberg-Marquardt technique (not sure if this statement makes sense in theory - these are two separate optimization methods). It is noted that when Gaussian noise is assumed to be present in the measurements, projection error minimization results in Maximum Likelihood estimate. In section 3, they briefly describe what algorithms work best in their view for various applications. They consider RANSAC and M-estimators as two classical approaches for making robust estimation. Moreover, they look at numerous keypoint matching processes and markerless tracking proposals for camera relative pose estimation and motion tracking. When pose estimate of some object is known, a 3D contour of the object is projected onto the image. Object contour is sampled and a search is performed along the edge normal to the contour to find strong gradients in the next frame. For unknown object and environments, monocular VSLAM solutions are considered a viable approach by the authors as they are used in ARToolKit and other commercial solutions. State of the art VSLAM solutions such as FastSLAM, PTAM, ORB-SLAM take advantage of incremental new data integration techniques (mostly use EKF or a particle filter) and update error probability density on the go, as well as performing bundle adjustment (BA) offline to further increase mapping accuracy. Another advantage of using VSLAM end to end solutions is the integrated image registration directly in 3D space which is done by using Gaussian-Newton or Levenberg-Marquardt approaches for known matching points, and iterative closest point (ICP) for unknown ones. Binary feature descriptors such as ORB (orientation invariant), BRISK (scale and orientation invariant), FREAK, and BRIEF make it very efficient to track and recognize objects and scenery which translates into tracking object and camera

poses.

- OCRTOC: I setup [2] on my computer and began studying the code and I will re-implement it for learning purposes. I reinstalled Ubuntu on my lab computer for model training purposes. Next I need to setup MoreFusion or PoseNet on it. I think we can come up with some improved algorithm for object pose estimation if we thoroughly study state of the art VSLAM solutions.

- Reading list: [3] and [4].

- MoreFusion [2]: Still need to write a literature review on this.

- PoseNet [5]: I read this paper today, it seems like a simple approach but I am quite skeptical it would work in an inverse setting where we estimate the pose of the objects in the scene rather than the camera. I will set it up and train a data set on it and see how it performs.

- UR5e:

- Quaternions:

- TensorFlow:

- Fellowship: I attended NSF Grant webinar and I know I can write a good personal statement but I have no idea if my research is worthy of anything at all at this point. I am confident if I do good research I will produce great results but at this point I can not make any judgment on my work as I am still learning and absorbing at a high rate. Usually when my learning rate declines I know I am in a good position with respect to the material. This is my self-feedback.

- Machine Learning: Working on Keras tutorial.

2 Plans

Following items are listed in order of priority:

- (On pause) Continue with ROS Industrial tutorials and documentation.

- (On pause) Resume Robotic Perception course as soon as possible.

- (On pause) Read Digital Image Processing by Gonzalez and Woods.

References

[1] E. Marchand, H. Uchiyama, and F. Spindler, "Pose estimation for augmented reality: a hands-on survey," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 12, pp. 2633–2651, 2015.

[2] K. Wada, E. Sucar, S. James, D. Lenton, and A. J. Davison, "Morefusion: Multi-object reasoning for 6d pose estimation from volumetric fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14540–14549, 2020.

[3] J. Lampinen and A. Vehtari, "Bayesian approach for neural networksreview and case studies," *Neural networks*, vol. 14, no. 3, pp. 257–274, 2001.

[4] R. Li, Z. Liu, and J. Tan, "A survey on 3d hand pose estimation: Cameras, methods, and datasets," *Pattern Recognition*, vol. 93, pp. 251–272, 2019.

[5] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in *Proceedings of the IEEE international conference on computer vision*, pp. 2938–2946, 2015.