Addendum to "Robust direct vision-based pose tracking using normalized mutual information"

Hang Luo*[a], Christian Pape [a], Eduard Reithmeier [a]
[a] Institute of Measurement and Automatic Control, Leibniz University Hannover, Nienburger Str.17, D-30167 Hannover

ADDENDUM

Recently it came to the authors' attention, that there is a published work by Kumar Shaurya Shankar and Nathan Michael (see [1]) which relates closely to their above mentioned publication (see also [2]).

Both publications are concerned with robust pose estimation methods for the task of direct visual odometry. In that context, in both publications a hybrid-tracking scheme composed of Sum of Squared Difference (SSD) based tracking and an information-based tracking algorithm on a Gaussian pyramid is employed.

However, one main difference is the structure of the hybrid tracking schemes. Shankar and Michael use a mutual information (MI)-based tracking method on higher pyramid levels (low resolution) for a coarse estimate, which is refined by a standard SSD-based tracker on lower levels (high resolution). We use a re-weighting SSD-based tracker to get a coarse pose estimate on lower resolution levels. That pose is more likely located in the narrower convergence domain of normalized mutual information (NMI) optimizer. Afterwards, a NMI maximization on high-resolution levels is performed to optimize the coarse estimate.

The implementation of the information-based optimizer contains another difference. Combined with a line-search strategy, Shankar and Michael use a MI-based Newton descent method for direct pose estimation and compute the derivatives of those pixels with high gradient magnitude to save computational time. They use the SSD tracking for a fast refinement. Our work focuses on the implementation of a NMI metric. For this implementation, we directly evaluate the second-order analytical NMI derivatives and incorporate it into a generic Newton-type framework. In order to speed up the computation of the NMI metric with its derivatives on lower levels, we set and reuse proper intermediate variables in the computation process, and employ the coarse SSD optimizer to select a subset of pixels with high confidence for computation.

Besides, Shankar and Michael compare the estimated relative transformations between nearby poses to the actual relative transformations (from ground truth) on the Tsukuba dataset and their own-built datasets. We carry out simulations on different RGB-D datasets with different evaluation criteria. We firstly use the pixel distances to test the convergence rate and then run our method on popular RGB-D datasets (TUM RGB-D and ETH-ICL datasets) for a quantitative measure of the global pose deviation with respect to ground truth.

Finally, both papers mention that the use of information metrics (MI and NMI) increases the tracking robustness to scene variations. Shankar and Michael argue that their method uses the SSD to speed up the refinement process at high-resolution levels. We claimed to exploit SSD to increase the convergence percentage of the optimization for the complete tracking scheme in order to handle larger initial displacements.

[1] K. S. Shankar and N. Michael, "Robust direct visual odometry using mutual information," in *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR),* IEEE, 9-14(2016).

[2] H. Luo, C. Pape, and E. Reithmeier, "Robust direct vision-based pose tracking using normalized mutual information", in *Optical Metrology and Inspection for Industrial Applications V*, *Proc. SPIE* 10819, 108190T (2018).

Optical Metrology and Inspection for Industrial Applications V, edited by Sen Han,
Toru Yoshizawa, Song Zhang, Proc. of SPIE Vol. 10819, 1081920 · © 2018 SPIE
CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2534586