Real-time pose estimation using stereo vision and mapping

PhD Student: Jakob Wilm[1,2]

Supervisors: Rasmus Larsen[1], Rasmus R. Paulsen[1], and Liselotte Højgaard[2]

1: Applied Mathematics and Computer Science, DTU 2: Dept. of Clinical Physiology, Nuclear Medicin & PET, Rigshospitalet

Outline

We approach the problem of estimating the 6DOF pose of a rigid object by means of a stereo camera setup. No prior knowledge of the object geometry is used — instead the object surface is reconstructed simultaneously with pose estimation. This is not unlike the SLAM (simultanous localisation and mapping) methodology used commonly in e.g. robot navigation.

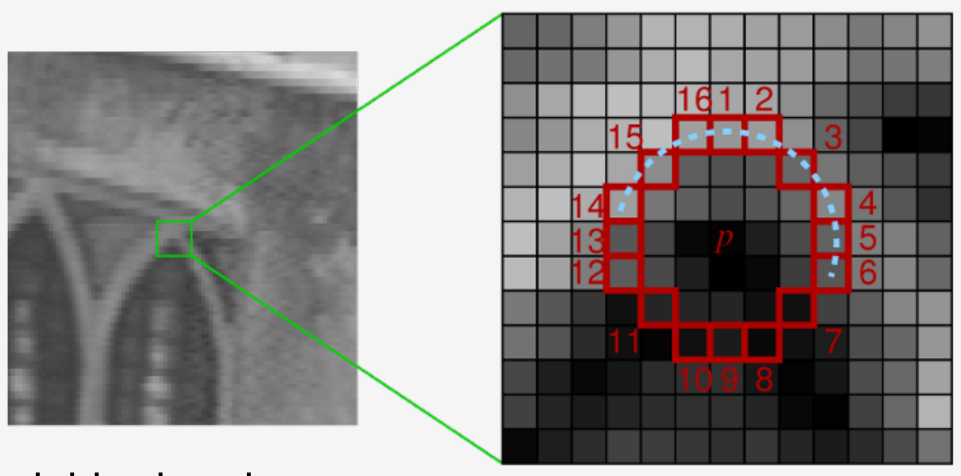
The motivation for this work is the problem of head motion during medical scanning. MRI and PET often require very long acquisition times during which patient motion is likely to occur. In MRI, these effects include Nyquist violations which are impossible to correct for post-aquistion.

Hence a real-time procedure for motion estimation is sought. The presented approach is similar to the work of Kyme et al. [1], but solves a n-perspective n-pose problem and makes use of binary features for real-time performance.

FAST and BRIEF features

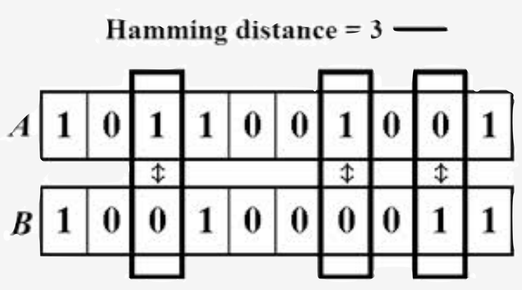
A high number of SIFT-like features locations are detected, described and matched, which comprises the computationally heaviest part of our algorithm. Consequently, we make use of a modern detector and binary features, which are considerably faster to compute and match.

Features locations are detected using a multi-scale adaptation of FAST (Features from Accelerated Segment Test) [2] over 4 scale levels.
The illustration shows how candidate features points are checked by determining the longest segment of higher brightness pixels in a ring-shaped neighborhood.

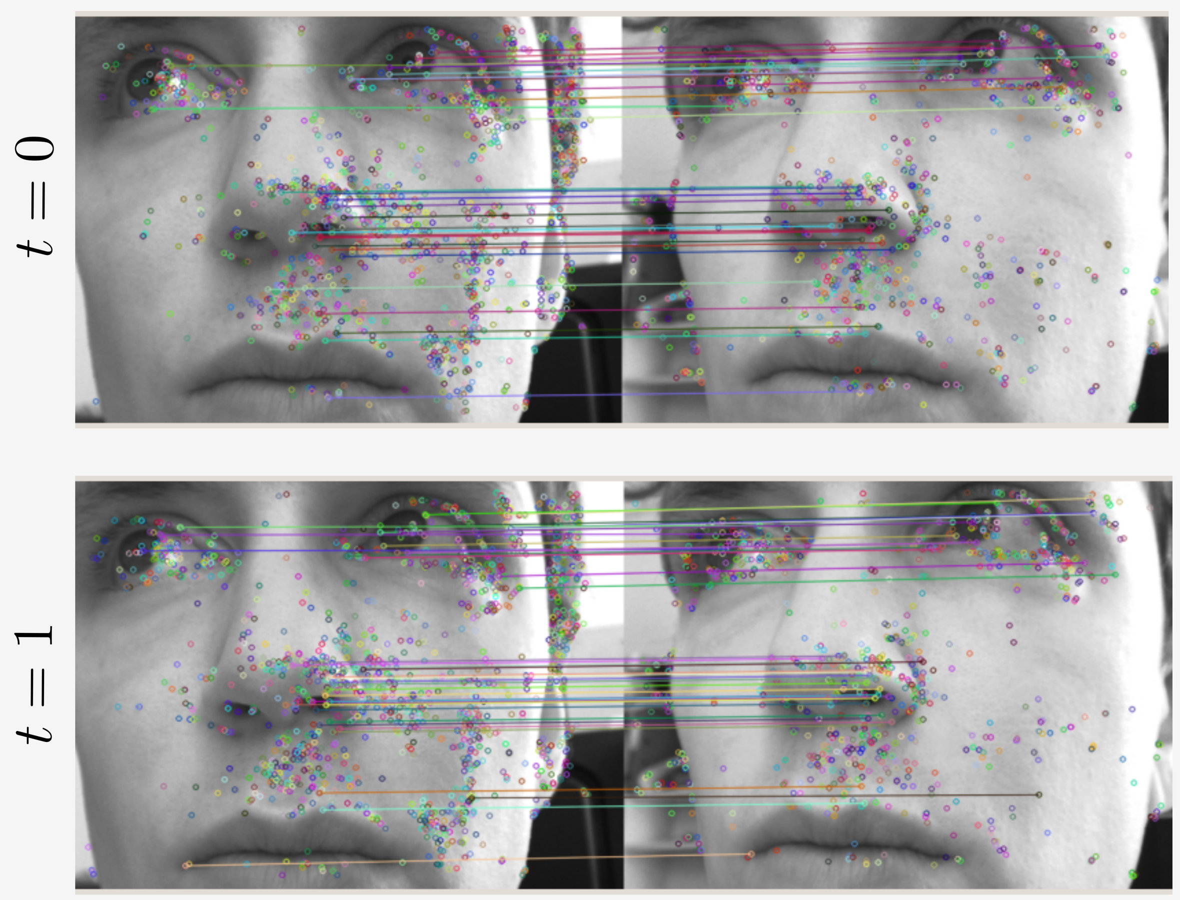


BRIEF is a fast binary feature descriptor with high robustness to geometric and photogrammetric transformations. The feature vector contains evaluations of a brightness comparison between the feature location and surrounding points in a predefined pattern. This comparison is done in a rotation/scale normalised space.

Feature matching comprises of a high number of distance calculations, which in the case of binary features can be done very efficiently by means of the Hamming distance as illustrated to the right.



Sample feature matches from two time instances are shown below.



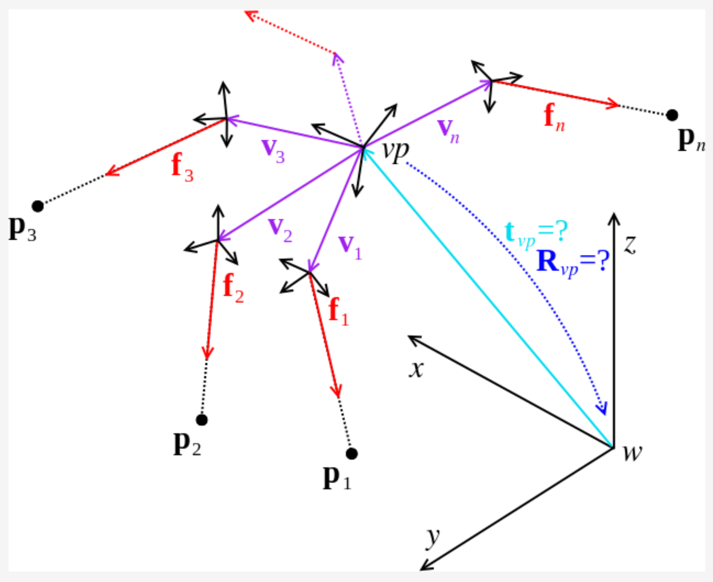
Left camera Right camera

N-perspective n-point problem (nPnP)

In the classical perspective n-point problem, the pose of a camera is estimated based on a series of 3D-2D point correspondences $(p_i \rightarrow f_i)$. The minimal case is three non-coaxial pairs.
Because we observe 3D points in two distinct, calibrated cameras, we solve the generalised n-perspective n-point problem using a linear algorithm [3] with sampling consensus, followed by a few iterations of a non-linear refinement of the estimate.

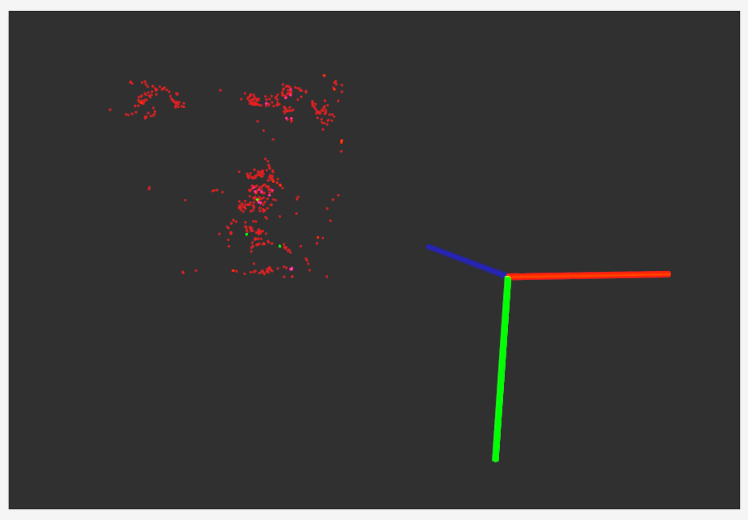
The illustration shows the n-perspective n-point problem as defined in OpenGV library[4]. vp denotes a "viewpoint", in this case the rigidly coupled stereo-setup. f_i are "bearing angles" corresponding the observed 2D projections of points p_i.

Solving for rotation R and t in this case provides us with the current object pose.



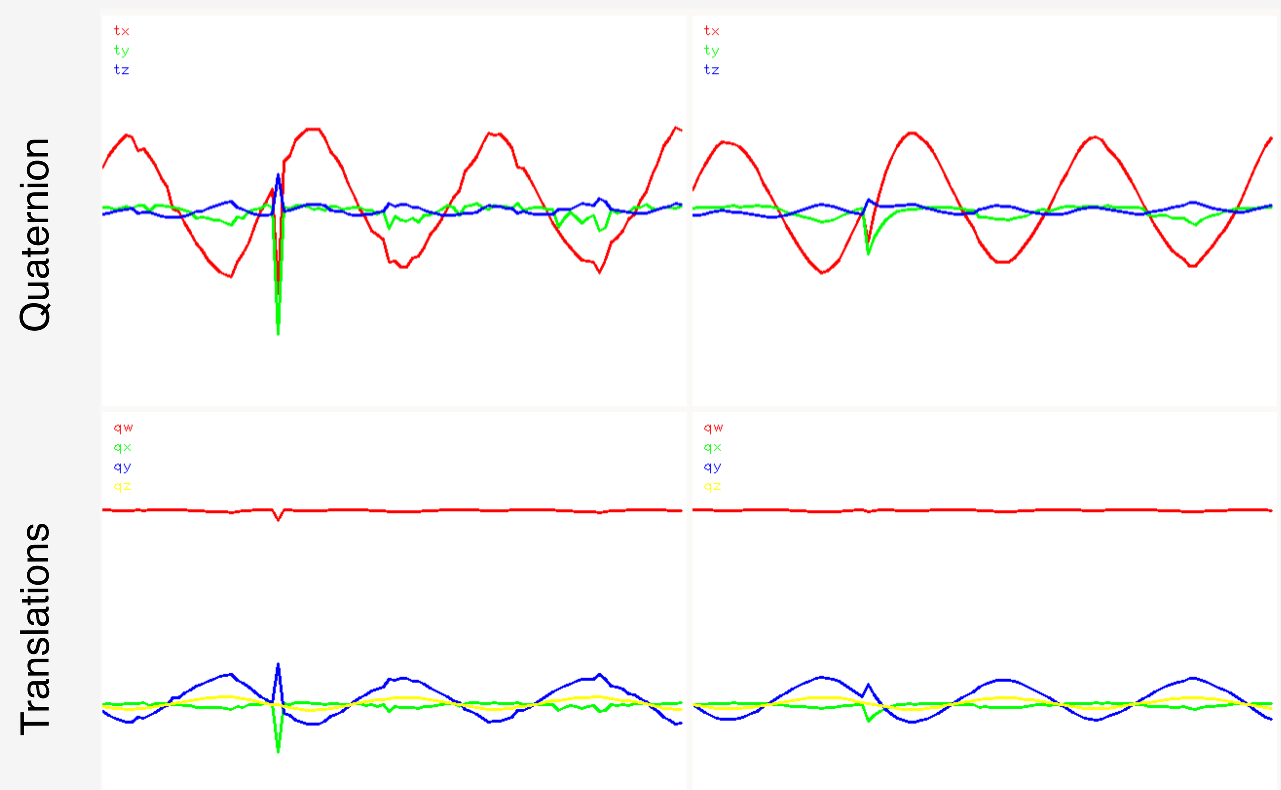
Mapping

Following the SLAM methodology, we match current features to a model of the object surface, while extending the model with unmatched features. This ensures a drift-free result, while providing ample opportunity to match.
Typical numbers for the "face" scene, shown throughout the poster, are 900 features detected in each image, with 120 stereo matches and 100 inliers after sampling consensus.



Kalman Filtering

Pose estimates are corrupted by noise. To mitigate this problem we filter the pose estimates as parameterised by quaternion and translation vector with a 6DOF scented Kalman filter similar to [5]. Qualitative results are shown below for the "face" scene used throughout this poster.



w/o filter with filter

Conclusion

Using binary feature descriptors and Kalman filtering, we have implemented a real-time system for 6DOF pose tracking. Qualitative results are promising, however, organic scenes are challenging to this method, and result in sporadic loss of tracking. Future work will address these challenges.

[1] Kyme, A., Se, S., Meikle, S., & Fulton, R. *Markerless Motion Tracking for Motion-Compensated Clinical Imaging.* In Proc. NSS/MIC (pp. 2–5). 2013.
[2] Rosten, E., & Drummond, T. *Machine learning for high-speed corner detection.* Vol. 3951 LNCS, pp. 430–443. doi:10.1007/11744023_34. 2006.
[3] Kneip, L., Furgale, P., & Siegwart, R. *Using multi-camera systems in robotics: Efficient solutions to the NPnP problem.* In Proc. IEEE ICRA (pp. 3770–3776). doi:10.1109/ICRA.2013.6631107. 2013.
[4] Kneip, L., & Furgale, P. (n.d.). *OpenGV: A Unified and Generalized Approach to Real-Time Calibrated Geometric Vision.* In Proc. IEEE ICRA. 2014.
[5] Goddard, J., & Abidi, M. *Pose and Motion Estimation using Dual Quaternion-Based Extended Kalman Filtering.* SPIE Phot. West, 3313 (January), 189–200. 1998.