Visual Odometry: Literature Survey

Alex Kreimer

January 12, 2016

Abstract

1. Stereo tracking and three-point/one-point algorithms-a robust approach in visual odometry [?]

The authors use Harris [6] features and KLT constrained by epipolar geometry as a front-end and LM re-projection minimization procedure as geometric estimation procedure. They peculiarity is that they present an initialization procedure for the rotation of the camera based on the infinite homography computation. H_∞ is computed by taking a patch in current image and minimize its difference from the warped template in the original image (the authors do not specify how the patch is chosen). The warping is performed by H_∞ that is parameterized by the rotation angles of the camera. After the rotation is determined, the translation is obtained from re-projection minimization over translation parameters only. They authors present results on their proprietary data without comparison to other methods or well established data sets. They argue that the decomposition of the motion problem leads to a more efficient (computation time) solution.

2. Vision meets robotics: The KITTI dataset [?]

This works describes the KITTI dataset. It provides a detailed information about the hardware setup, sensor calibration procedures, and their development kit. The setup has 2 pairs of stereo cameras (color and gray-level), laser scanner, inertial and GPS navigation system with RTK correction.

3. Real-Time Stereo Visual Odometry for Autonomous Ground Vehicles [8]

This work presents a real-time stereo odometry algorithm. Its estimation part is LM re-projection minimization (similar to StereoScan). The peculiarity of the algorithm is in its feature matching procedure (it uses Harris [6] or FAST [14] features). The pair of features is called consistent if the euclidean distance between them in frame t_1 is about the same as in frame t_2. The authors build an adjacency matrix when there is a link between feature pairs iff they are consistent. After this they search (greedy algorithm, since the problem of finding largest clique in graph is NP-complete) for the largest clique. Thus they do inlier detection, instead of usual outlier rejection using RANSAC.

4. An efficient solution to the five-point relative pose problem [12]

Five-point problem is to find possible camera poses between two calibrated views given 5 point correspondences. First, the authors use epipolar constraint to derive $E = xX + yY + zZ + wW$ where X, Y, Z, W are known vectors. The scalars x, y, z are solved for using ten cubic constraints of the Essential matrix (e.g., zero determinant and the constraint of equal eigenvalues). R, t are recovered from E using usual decomposition technique. The algorithm is compared to 8-pt, 6-pt, 7-pt algorithms. It is found to be superior for sideways motion and usually is outperformed by other methods for forward motion.

5. Visual Odometry [13]

The authors describe a complete visual odometry system and show its performance versus DGPS data. The system is real time, uses Harris corners, and corner matching as tracking. It works in both mono and stereo setups. In mono setting they solve relative orientation using the 5-point algorithm [12] as initial orientation with subsequent iterative refinement. In stereo setting 3-point algorithm [5] is used for initial hypothesis generation with subsequent re-projection error refinement. The authors show a number of shortcuts that make usually computationally expensive tasks run in real time (e.g., feature computation, matching, robust estimation). The lengths of sequences they test on is hundreds of meters. The experiments show that the system compares favorably to DGPS and INS sensors.

6. Parameterizing Homographies [1]

The motion of the plane may be described by a homography. This work compares different ways to parameterize homography for plane estimation. The authors deal with a number of cases (1) Unknown Relative Orientation (2) Fully Calibrated Case (3) A Moving Stereo Rig. The parameterization studied are 3×3 matrix parameterization, direct plane and 4 point parameterization. The authors show that in the (1) case 4 point parameterization is superior to the others.

7. Monocular Visual Odometry in Urban Environments Using an Omnidirectional Camera [16]

This work solves monocular odometry. The contribution of the paper is that they decouple the rotation estimation from the translation in order to estimate the pose of a new image. The reason why this works is that reconstruction of far away 3d points is very poor and thus these points are a “burden” on a re-projection error minimization methods. The authors make use of these features by computing epipolar geometry and recovering rotation out of it. This method significantly reduces drift over the state of the art methods (the algorithm is tested on a 2.5 km sequence). Translation scale/direction is still recovered using 3d based method. They use SIFT instead of Harris since it gave better results.

8. The Fundamental matrix: theory, algorithms, and stability analysis [9]

This work defines the Fundamental matrix as known today. Before it was common to use the Essential matrix to express the geometry of 2 views. The advantage of using F is that the camera does not have to be calibrated. The authors develop the formalism of F, propose estimation methods, analyze their stability and study degenerate configurations. Stability of the estimation is measured by error in the epipole (the authors argue that usually this is most commonly used computation result). The results show that linear (8-point) estimation method is highly sensitive to noise (especially if the epipole is in the image) and may always be improved by non-linear refinement. They study two different methods to parameterize F and a number of different optimization objectives (the best being Sampson error).

9. Real Time Localization and 3D Reconstruction [11]

This work estimates a motion of a calibrated camera (mono) set on an experimental vehicle. Interest points (Harris) are tracked and matched (ZNCC score). Robust estimates of the camera motion are computed in real-time, key-frames are selected and permit the features 3D reconstruction. The author introduce a procedure they call local bundle adjustment that ensures both good accuracy and consistency over long sequences. [12] is used to initialize the pose.

10. Refining essential matrix estimates from RANSAC [3]

The paper deals with the problem of estimating relative pose of two cameras from outlier contaminated feature correspondences. It is a common practice to use RANSAC with conjunction with a linear method to estimate a model (in this case the Essential matrix) and then to refine it using non-linear optimization method. They evaluate several refinement methods which minimize functions of Sampson’s error. All perform well on range sets of correspondences or sets with low outlier rates; but many perform poorly otherwise. The most accurate solution is give by minimizing a robust function (Blake-Zisserman) of a Sampson’s error. The rotations are parameterized as quaternions and the optimization is performed over the essential manifold, see [15]. The authors use IRLS combined with LM [10] method to optimize the objective.

11. Vision-based robot localization without explicit object models [4]

The authors propose a solution to localize a robot in an unknown 2d environment using visual input. The authors train neural net as a regression model N(I)=(x,y). As an input for the net they use image statistics (first and second moments of the edge distributions, mean edge orientations, densities of lines).

12. Visual Homing: Surfing on the Epipoles [2]

They propose a solution for visual homing. Using this method a robot can be sent to desired positions and orientations in 3D space specified by single images taken from these positions. The method is based on estimating epipolar geometry between a pair of views. A 3D model of the environment is not required. Using the epipolar geometry most of the parameters which specify the differences in position and orientation of the camera between the two images are recovered. From a pair of images translation may be recovered only up to a signed scale. In order to find out the real distance the robot make an extra step and takes an additional image. The authors prove that when the camera moves on a straight line the features move along the epipolar lines while their coordinates and the coordinate of the epipole obey the cross-ratio relation.

13. A way to parameterize rotations [15]

This work proposes a method to use quaternions in an unconstrained nonlinear optimization. Quaternions representing rotation have four elements but only three degrees of freedom, since they have to be of norm one. This constraint has to be taken into account when applying e.g., Levenberg-Marquardt algorithm. One of the ways to address this issue is to use appropriate parameterization (others are a projection step and Lagrange multipliers). Well known parameterizations are Euler angles and axis-angle representation.

The [7] call a parameterization fair if it does not introduce more numerical sensitivity than inherent to the problem itself. This is guaranteed, if any rigid transformation of the space to be parameterized results in an orthogonal transformation of the parameters. Both axis-angle and quaternion parameterizations are fair, while Euler angles is not.

Authors search for a parameteriation that:

(a) is minimal, i.e., uses only three parameters
(b) the three parameters may be changed arbitrarily by the optimization algorithm
(c) the resulting quaternion has always norm 1.

This new approach is based on the observation that all quaternions of norm-1 lie on the unit sphere in $\mathbb{R}^4$. The authors use the shortest connection between two points on a sphere, i.e., a great circle. For describing a movement on the sphere starting at $\mathbf{h_0}$ they use a vector v_4 lying in the tangential hyperplane that touches the sphere at $\mathbf{h_0}$. This hyper-plane is a subspace of $\mathbb{R}^4$, thus vectors in this plane may be represented as 3-vectors with respect to a plane-local coordinate frame.

Experiments are made on a synthetic (small) data-set. The authors perform bundle adjustment and compare their approach with axis-angle representation. The conclusion is that this representation performs better for rotations, for transnational motion both method are approximately equal.

References

[1] Simon Baker, Ankur Datta, and Takeo Kanade. Parameterizing homographies. *Robotics Institute, Carnegie Mellon University, Tech. Rep*, 2006.

[2] Ronen Basri, Ehud Rivlin, and Ilan Shimshoni. Visual homing: Surfing on the epipoles. *International Journal of Computer Vision*, 33(2):117–137, 1999.

[3] Tom Botterill, Steven Mills, and Richard Green. Refining essential matrix estimates from ransac.

[4] Gregory Dudek and Chi Zhang. Vision-based robot localization without explicit object models. In *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, volume 1, pages 76–82. IEEE, 1996.

[5] Robert M Haralick, Chung-nan Lee, K Ottenburg, and Michael Nölle. Analysis and solutions of the three point perspective pose estimation problem. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 592–598. IEEE, 1991.

[6] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Citeseer, 1988.

[7] Joachim Hornegger and Carlo Tomasi. Representation issues in the ml estimation of camera motion. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 640–647. IEEE, 1999.

[8] Andrew Howard. Real-time stereo visual odometry for autonomous ground vehicles. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 3946–3952. IEEE, 2008.

[9] Quan-Tuan Luong and Olivier D Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *International journal of computer vision*, 17(1):43–75, 1996.

[10] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441, 1963.

[11] Etienne Mouragnon, Maxime Lhuillier, Michel Dhome, Fabien Dekeyser, and Patrick Sayd. Real time localization and 3d reconstruction. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 363–370. IEEE, 2006.

[12] David Nistér. An efficient solution to the five-point relative pose problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(6):756–770, 2004.

[13] David Nistér, Oleg Naroditsky, and James Bergen. Visual odometry. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–652. IEEE, 2004.

[14] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *Computer Vision–ECCV 2006*, pages 430–443. Springer, 2006.

[15] Jochen Schmidt and Heinrich Niemann. Using quaternions for parametrizing 3-d rotations in unconstrained nonlinear optimization. In *VMV*, volume 1, pages 399–406, 2001.

[16] Jean-Philippe Tardif, Yanis Pavlidis, and Kostas Daniilidis. Monocular visual odometry in urban environments using an omnidirectional camera. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 2531–2538. IEEE, 2008.