# Visual Odometry: Literature Survey

Alex Kreimer

April 18, 2016

**Abstract**

## 1. Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints [?]

This work deals with scale estimation in monocular visual odometry. The authors exploit "nonholonomicity" of the vehicle motion to recover the scale. It turns out that if the camera origin has an offset with the vehicle origin the fact that there is an instantaneous radius to vehicle motion may be exploited to obtain real motion scale.

## 2. Are we ready for autonomous driving? the kitti vision benchmark suite [?]

This work presents the KITTI data-set. The authors claim that at the moment when the work is published computer vision systems are rather rare in practical robotic systems. They take a step towards a wider use of vision in such systems by publishing a demanding data set that mimics real world requirements. The dataset includes data from four high resolution video cameras, a Velodyne scanner and state-of-the-art localization system. The benchmark includes 389 stereo and optical flow image pairs, 39.2 km of visual odometry data, and more than 200k 3D object annotations. The paper describes the setup, data-set organization and algorithm evaluation tools.

## 3. Stereoscan: Dense 3d reconstruction in real-time [?]

This method solves visual odometry. The authors use custom corner detector (primarily for run-time efficiency). Which are matched between pairs of stereo images as well as between pairs of consecutive images. Outliers are pruned by imposing rigidity constraint on a scene using Delaunay triangulation and also using circle match heuristic. Relative orientation is solved using non-linear Levenberg-Marquard reprojection minimization procedure wrapped into RANSAC. Despite the fact that LM may fall into local minima this method shows good results on the KITTI [?] dataset.

4. Visual odometry by multi-frame feature integration [?]

   This is an incremental work over the method proposed in [?]. The improvement proposed in this paper is a method to take into account the location of the feature in more than just a pair of consecutive stereo frames. Basically, the authors compute an average 3D location of the feature for a sequence of frames. When current relative orientation is solved by reprojection error minimization, projections of such averaged 3D points are taken into account. This method showed improved results on KITTI compared to [?].

5. Efficient computation of relative pose for multi-camera systems [?]

   This work generalizes the constraints presented in [?] to the generalized camera model presented in [?]. The problem is formulated as minimal eigenvalue minimization problem and solved using Levenberg-Marquardt optimization algorithm.

6. Direct optimization of frame-to-frame rotation [?]

   This work is a follow-up for [?]. They use the same constraint to solve relative orientation, but propose a different solution. The authors formulate the problem as a minimal eigen-value minimization scheme. The solutions proposed in this work is a non-linear optimization scheme based on Levenberg-Marquardt algorithm and a branch-and-bound algorithm. Both solution compare favorably to the state of the art on synthetic data.

7. Finding the exact rotation between two images independently of the translation [?]

   The authors propose constraints on the motion of the image features for pure rotation/translation of the camera. They state that for pure translation the normals to the epipolar planes are co-planar, while for pure rotation the feature displacement vectors must be co-planar. They also claim that the rotation (translation) is correctly estimated if the remaining motion adheres to the pure translation (rotation) constraint. The above constraints allow the authors to compose systems of polynomial equations that are solved with algebraic geometry methods.

8. Analysis and solutions of the three point perspective pose estimation problem [?]

   Three point perspective pose estimation problem refers to finding the pose of the camera relative to a known 3D triangle given the projections of its vertices on the camera plane. In our context this problem arises as an initial step of the odometry algorithms. The authors evaluate a number of methods to solve the problem mainly from numerical stability perspective. All algorithms have singularity points, but their probability is rather low. The paper shows that the order of substitutions that the solution

undertakes may influence result significantly. Numerical stability analysis is given for 6 different solutions.

## 9. An Efficient Minimal Solution for Multi-Camera Motion [?]

The authors propose an efficient method for estimating the motion of a multi-camera rig from a minimal set of feature correspondences. The authors develop a first order approximation to the generalized epipolar constraint developed in [**?**]. They propose and evaluate two solution methods to the posed problem: a method based on the Grobner basis and another methods based on the polynomial root finding using Sturm sequences.

## 10. Least-squares fitting of two 3-D point sets [?]

This paper deals with absolute orientation problem. It proposes a solution based on SVD which may be an alternative to [**?**]. The authors do not compare numerical stability of both solution, but do provide computation time comparison. For minimal case (3pt) quaternion based solution is faster, for larger point sets SVD is faster.

## 11. Closed-form solution of absolute orientation using unit quaternions [?]

Similar to three point perspective pose estimation problem discussed in [**?**] absolute orientation arises as initial step in some of the visual odometry algorithms. Absolute orientation problem is to determine relative orientation of the coordinate systems, given the coordinates of the same 3D points in each one of them. This paper presents a closed form solution to this problem based on unit quaternions.

## 12. Using many cameras as one [?]

The authors evaluate different camera configurations for camera ego-motion estimation. The tool used to analyze different camera system configurations is Fisher Information Matrix. The Fisher Information Matrix is a powerful analysis technique that can apply to any problem which involves searching for a parameter set that minimizes an error function. The author use Plucker line coordinates and a unified camera framework [**?**] to view a multi-camera system as a single camera. They derive a generalized epipolar constraint and generalized epipolar flow constraint and then analyze their sensitivity to noise using Fisher Information Matrix.

## 13. A General Imaging Model and a Method for Finding its Parameters [?]

Linear perspective camera model turned to be a defacto standard in computer vision. The authors come to conclusion that there are a number of imaging systems/sensors that were developed that do not fit the perspective camera model (e.g. cathadioptric, dioptric wide-angle, compound camera systems). The authors propose a generalized camera models that

is based on virtual imaging elements called raxels (ray pixels). Ray pixels sample the plenoptic function at some location in certain direction. The image surface is defined as caustic, a special case of a ray surface where the direction of the ray $q$ is a function of the location $p$ and the incoming rays are tangent to the surface. The authors propose methods to compute the parameters of the generalized imaging system.

## 14. Ego-motion and omni-directional cameras [?]

The authors combine wide-angle imaging systems with ego-motion algorithms designed for a usual perspective projection camera. Wide-angle imaging systems have advantages over a regular ones (e.g., it is possible to disambiguate rotation and translation effects). They show, on one hand, how to map wide-angle imaging system motion onto the unit sphere (i.e. use spherical projection model). On the other hand, they show how to adapt available perspective projections ego-motion estimation algorithms for the spherical motion field.

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

The authors use Harris [**?**] 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. $H_\infty$ 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 $H_\infty$ 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.

## 16. 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.

## 17. Real-Time Stereo Visual Odometry for Autonomous Ground Vehicles [?]

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 [?] or FAST [?] 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.

## 18. An efficient solution to the five-point relative pose problem [?]

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.

## 19. Visual Odometry [?]

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 [?] as initial orientation with subsequent iterative refinement. In stereo setting 3-point algorithm [?] 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.

## 20. Parameterizing Homographies [?]

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 $3 \times 3$ matrix parameterization, direct plane and 4 point parameterization. The authors show that in the (1) case 4 point parameterization is superior to the others.

## 21. Monocular Visual Odometry in Urban Environments Using an Omnidirectional Camera [?]

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.

## 22. The Fundamental matrix: theory, algorithms, and stability analysis [?]

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).

## 23. Real Time Localization and 3D Reconstruction [?]

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. [**?**] is used to initialize the pose.

## 24. Refining essential matrix estimates from RANSAC [?]

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 [?]. The authors use IRLS combined with LM [?] method to optimize the objective.

## 25. Vision-based robot localization without explicit object models [?]

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).

## 26. Beyond RANSAC: user independent robust regression [?]

This paper deals with shortcomings of RANSAC, e.g., user-set inlier threshold, which is hard to guess a-priory. The authors formulate usual regression problem in a projection pursuit framework (e.g., they derive a projection based M-estimator) for heteroscedastic (e.g., such that has different variability in it) data. The advantage for such formulation is that there exist rules for bandwidth selection for such estimators which are purely data driven.

## 27. Error modeling in stereo navigation [?]

This work proposes a method to model invariance in 3D point triangulation. They propose a way to estimate an error in a triangulated point given an assumption of an error in image features (basically, using Gauss-Markov error propagation theorem). The authors also propose a way to update landmark locations using the estimated error information. Using synthetic benchmarks they show that their method produces better results.

## 28. Visual Homing: Surfing on the Epipoles [?]

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.

## 29. A way to parameterize rotations [?]

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 [**?**] 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.

## 30. A simple and efficient rectification method for general motion [**?**]

This paper proposes a rectification method. It addresses issues present in most other rectification methods that sometimes produce very large images and sometimes can not rectify at all (usually when the epipoles are in the images). This method is based on a polar parameterization of the images around the epipole. No pixels are lost and images of minimal size are produced. One of the downsides of this method is that new cameras (rectified) are not available.

## 31. On measuring the accuracy of SLAM algorithms [**?**]

This paper attempts to create an objective benchmark to measure the accuracy of SLAM algorithms. The metric that this work proposes uses

only relative poses and does not rely on a global reference frame (e.g. benchmark dependence on time) . This work is a basis for KITTI [**?**] benchmarks. It allows to compare methods that use different estimation techniques and different sensor moralities.