

# Research proposal: Statistical Visual Odometry

Ehud Rivlin, Ilan Shimshoni, Alex Kreimer

March 24, 2015

## Abstract

Visual odometry is the process of estimating the ego-motion of the agent (e.g, human, vehicle) using only the input of a single or multiple cameras attached to it. In this work we propose a novel stereo visual odometry algorithm which uses motion priors, uncertainty modeling/propagation and maximum likelihood based inference procedure, that incorporates uncertainty measurements, to improve current results. In order to test our algorithm, we build a hardware setup that is capable to produce video sequences and ground truth suitable for stereo visual odometry algorithms.

## 1 Introduction

Visual Odometry has been successfully used in scientific and industrial environments (see [1] for an example on VO on Mars exploration rovers). Keeping track of a vehicle's location is one of the most challenging aspects of building an autonomous vehicle. The advantage of VO with respect to wheel odometry is that VO is not affected by wheel slip in uneven terrain or other adverse conditions. It has been demonstrated that compared to wheel odometry, VO provides more accurate trajectory estimates, and thus it matured from a "nice to have" capability into a critical autonomous vehicle system.

VO odometry system comprises a software for processing stereo pairs taken by the cameras. The system computes and updates the 6 Degree of Freedom vehicle pose (x, y, z, roll, pitch, yaw) by tracking the motion of image features.

In section 2 we do a brief summary of a related work in the field, in section 3 we state the main idea of our work, sections 4 and 5 go into the details of the proposed algorithm and the setup that we use.

## 2 Related work

Work on estimating motion with stereo cameras can be traced back to Moravec's work [2]. Following Moravec's work, Matthies et al. treated motion estimation as a statistical estimation problem and developed sequential methods for estimating the vehicle motion and updating the landmark models. This system

achieved an accuracy of 2% of distance over 5.5 meters and 55 stereo image pairs ([3];[4]) with a consistent level of accuracy reported more recently ([5]). Similar work has been reported elsewhere ([6], [7], [8]). Recently, Nister et al. have reported a successful real-time Visual Odometry implementation ([9],[10]). This implementation contains two motion estimation schemes: the stereo scheme, which is an iterative pose refinement scheme, and the monocular scheme, which was based on the framework of the 5-point algorithm [10] and the outlier rejection scheme [11] and good results have been reported on very long image sequences. Other approaches to the subject of Visual Odometry schemes have been reported as well. For example, McCarthy and Barnes have reported the performance of optical flow based motion estimation ([12]) and Vassalo and Gluckman and others have developed ego-motion estimation with omnidirectional images [13],[14],[15].

### 3 Main idea

In this work we seek to build an end to end VO system which comprises the following parts:

1. Implement a novel VO algorithm which exploits vehicle motion priors and uncertainty modeling/propagation to improve current results.
2. Hardware setup that is capable to capture the video sequences and to produce ground truth to evaluate a VO algorithm. The setup will be used to evaluate the performance of the algorithm along with an industry standard datasets (e.g. [16])

Below we summarize novel features of our algorithm:

1. Motion priors (e.g., [17]) we used in a mono VO algorithm, while we propose to use them in stereo VO
2. We propose to model uncertainty of the image features with the application to stereo VO algorithm, which was not done before
3. We propose to incorporate image features/3D points uncertainty into 3D-to-2D reprojection error minimization scheme, which was not done before (it was done in 3D-to-3D setting)

In the following sections we describe in more detail each part of our proposal.

## 4 Novel visual odometry algorithm

### 4.1 Motion priors

Motion priors for VO mean that we would like to exploit our knowledge of the problem (e.g., the vehicle can not jump) to improve existing results.

One of the main issues in VO algorithm is a presence of outliers in feature matching. Usually, a RANSAC procedure is employed to cope with the problem. Its drawback is that the number of required RANSAC iterations grows exponentially and thus a wrong hypothesis may be chosen. One of the ways to improve the situation is to build a “proposal distribution” (e.g., [17]) and use it to sample the data, instead of random RANSAC sampling. They show that this approach improves the runtime of the algorithm and yields more accurate results, which favor solutions that comply with vehicle motion constraints.

## 4.2 Statistical VO

We use [16] as the baseline algorithm. Our goal in this part of the research is to use uncertainty modeling and error propagation to improve the existing VO results. Previous studies suggest that the value of uncertainty modeling in vision algorithms may vary greatly depending on how close the model is to the true noise distribution (see [18], [19] for a discussion).

With regards to uncertainty modeling/propagation our algorithm may be divided into three parts:

1. Feature processing. We plan to use Harris corners as features, since they posses the required properties: fast, robust and provide a large number of features. These features are discussed in section 4.3
2. Triangulation procedure. We plan to evaluate empirically a number of algorithms. (Naive, DLT, Gold standard, see [20]). This will be discussed in 4.4
3. Motion parameters ( $\mathbf{r}, \mathbf{t}$ ) fitting procedure, which will be discussed in 4.5

## 4.3 Statistical characteristics of Harris corners

The incentive of modeling uncertainty (represented by a covariance matrix) for feature points is rather simple: not all corners are the same w.r.t. their tracking properties. We plan to evaluate a number different methods proposed in the literature to model the uncertainty of Harris corners (e.g., [21], [18]) and their influence on the VO. It is worth to say that previous studies explored uncertainty propagation mainly in (bi)-linear estimation problems, which is not the case for VO.

The results of this step is a pair of covariance matrices  $\Sigma_L/\Sigma_R$  for the features in the left and right images respectively.

## 4.4 Statistical characteristics of reconstructed 3D points

The next important step w.r.t. error propagation is a triangulation of image features into the corresponding 3D points. Given the covariance matrices  $\Sigma_L, \Sigma_R$  the covariance matrix of the 3D point is given by: (this is commonly referred

to as “sandwich covariance”):  $\Sigma_X = J \begin{pmatrix} \Sigma_L & 0 \\ 0 & \Sigma_R \end{pmatrix} J'$ , where  $J$  is the Jacobian of the triangulation algorithm (it may either be calculated analytically or computed numerically, depending on the triangulation algorithm). This method was successfully used in [22]. The work of [22] uses a similar up to this point procedure to model the uncertainty of the 3D points.

## 4.5 Robust parameter fitting

At this point our proposed algorithm and the algorithm proposed in the paper diverge. Their work proposes to reconstruct the features of a new frame into 3D and solve 3D-to-3D rigid motion problem. Previous research has shown (TODO: fill in the ref, still can’t find this paper) that 3D-to-2D reprojection error minimization is superior to the 3D-to-3D method and thus our proposition is to use the computed uncertainty in the 3D points to minimize reprojection error in the new frame directly. The maximum likelihood based solution leads to the following minimization problem:

$$[\mathbf{r}, \mathbf{t}] = \underset{\mathbf{r}, \mathbf{t}}{\operatorname{argmin}} \sum_i (x_i - \hat{x}_i(\mathbf{r}, \mathbf{t}))' \Sigma_{x_i}(\mathbf{r}, \mathbf{t}) (x_i - \hat{x}_i(\mathbf{r}, \mathbf{t}))$$

where  $x, \hat{x}$  are observed and predicted feature locations,  $\mathbf{r}, \mathbf{t}$  are camera motion parameters,  $\Sigma_{x_i}(\mathbf{r}, \mathbf{t})$  is uncertainty information derived from previous steps. Note that if covariance matrices are modeled as identity we obtain the minimization procedure used in [23].

## 5 Hardware setup

The setup is intended to recover the motion of a vehicle in an urban or rural environment by a sole use of stereo image sequences. It comprises a number of parts:

1. A calibrated and synchronized stereo rig. While it is relatively easy to build a calibrated stereo rig, installing it on a moving vehicle poses a new challenge: video stream synchronization, which is required by a stereo algorithm. We use hardware camera sync, which guarantees the desired result.
2. (D)GPS, IMU units to produce a ground truth. These units should provide data that is in sync with the cameras, which poses a challenge.

As stated before, the setup will be used to evaluate the performance of the algorithm along with the industry standard datasets (e.g. [16])

## References

- [1] M. Maimone, Y. Cheng, and L. Matthies, “Two years of visual odometry on the mars exploration rovers,” *Journal of Field Robotics, Special Issue on Space Robotics*, vol. 24, p. 2007, 2007.
- [2] H. P. Moravec, *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. PhD thesis, Stanford, CA, USA, 1980. AAI8024717.
- [3] L. Matthies and S. Shafer, “Error modeling in stereo navigation,” 1987.
- [4] L. H. Matthies, *Dynamic Stereo Vision*. PhD thesis, Pittsburgh, PA, USA, 1989. AAI9023429.
- [5] C. F. Olson, L. H. Matthies, M. Schoppers, and M. W. Maimone, “Rover navigation using stereo ego-motion,” *Robotics and Autonomous Systems*, vol. 43, no. 4, pp. 215–229, 2003.
- [6] Z. Zhang, O. D. Faugeras, and N. Ayache, “Analysis of a sequence of stereo scenes containing multiple moving objects using rigidity constraints,” in *Computer Vision., Second International Conference on*, pp. 177–186, IEEE, 1988.
- [7] S. Lacroix, A. Mallet, R. Chatila, and L. Gallo, “Rover self localization in planetary-like environments,” in *Artificial Intelligence, Robotics and Automation in Space*, vol. 440, p. 433, 1999.
- [8] D. Nistér, O. Naroditsky, and J. Bergen, “Visual odometry,” in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, pp. I–652, IEEE, 2004.
- [9] D. Nistér, O. Naroditsky, and J. Bergen, “Visual odometry for ground vehicle applications,” *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006.
- [10] D. Nistér, “An efficient solution to the five-point relative pose problem,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 6, pp. 756–770, 2004.
- [11] D. Nistér, “Preemptive ransac for live structure and motion estimation,” *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, 2005.
- [12] C. McCarthy and N. Bames, “Performance of optical flow techniques for indoor navigation with a mobile robot,” in *Robotics and Automation, 2004. Proceedings. ICRA’04. 2004 IEEE International Conference on*, vol. 5, pp. 5093–5098, IEEE, 2004.
- [13] R. F. Vassallo, J. Santos-Victor, and H. J. Schneebeli, “A general approach for egomotion estimation with omnidirectional images,” in *Omnidirectional Vision, 2002. Proceedings. Third Workshop on*, pp. 97–103, IEEE, 2002.

- [14] J. Gluckman and S. K. Nayar, “Ego-motion and omnidirectional cameras,” in *Computer Vision, 1998. Sixth International Conference on*, pp. 999–1005, IEEE, 1998.
- [15] P. Corke, D. Strelow, and S. Singh, “Omnidirectional visual odometry for a planetary rover,” in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 4, pp. 4007–4012, IEEE, 2004.
- [16] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [17] D. Scaramuzza, A. Censi, and K. Daniilidis, “Exploiting motion priors in visual odometry for vehicle-mounted cameras with non-holonomic constraints,” in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pp. 4469–4476, IEEE, 2011.
- [18] M. J. Brooks, W. Chojnacki, D. Gawley, and A. Van Den Hengel, “What value covariance information in estimating vision parameters?,” in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, pp. 302–308, IEEE, 2001.
- [19] Y. Kanazawa and K.-i. Kanatani, “Do we really have to consider covariance matrices for image features?,” in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2, pp. 301–306, IEEE, 2001.
- [20] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, “second” ed., “2004”.
- [21] U. Orguner and F. Gustafsson, “Statistical characteristics of harris corner detector,”
- [22] L. H. Matthies, “Dynamic stereo vision,” 1989.
- [23] A. Geiger, J. Ziegler, and C. Stiller, “Stereoscan: Dense 3d reconstruction in real-time,” in *Intelligent Vehicles Symposium (IV)*, 2011.