2015 IEEE Intelligent Vehicles Symposium (IV)
June 28 - July 1, 2015. COEX, Seoul, Korea

Robust Stereo Visual Odometry from Monocular Techniques

Mikael Persson[1], Tommaso Piccini[1], Michael Felsberg [1], Rudolf Mester [1,2]

***Abstract*— Visual odometry is one of the most active topics in computer vision. The automotive industry is particularly interested in this field due to the appeal of achieving a high degree of accuracy with inexpensive sensors such as cameras. The best results on this task are currently achieved by systems based on a calibrated stereo camera rig, whereas monocular systems are generally lagging behind in terms of performance. We hypothesise that this is due to stereo visual odometry being an inherently easier problem, rather than than due to higher quality of the state of the art stereo based algorithms. Under this hypothesis, techniques developed for monocular visual odometry systems would be, in general, more refined and robust since they have to deal with an intrinsically more difficult problem.**

In this work we present a novel stereo visual odometry system for automotive applications based on advanced monocular techniques. We show that the generalization of these techniques to the stereo case result in a significant improvement of the robustness and accuracy of stereo based visual odometry. We support our claims by the system results on the well known KITTI benchmark, achieving the top rank for visual only systems*.

I. INTRODUCTION

The main distinction we make when discussing visual odometry systems is that between stereo and monocular systems. Due to the difficulties imposed by a single camera setting, monocular visual odometry systems are still hard pressed to compete with stereo systems despite significant improvements in the last few years.

Monocular challenges such as scale estimation: Whether through ground-plane estimation or object recognition is a tough task with inherent uncertainties, requiring long tracks and long optimization windows which can propagate the scale information through the translations over longer time spans. Moving objects present a different challenge by causing outliers which cannot always be identified immediately and necessitate careful maintenance of tracks. Similarly noisy measurements and poor pose estimates further require a robustification of the triangulation and estimation processes. Finally since monocular systems must rely only on the quality of its existing tracks, the initialization of a monocular system must be performed with great care.

Stereo systems by comparison, have a fixed known baseline that can be easily exploited to recover a global metric scale. Though in rare cases using only local data may leave the system vulnerable to distant scenes. Moving objects are, in general, easier to identify for the same reason but suffer from a similar problem. Noisy tracks cause less problems since they can be readily triangulated at every new couple of frames for verification and new tracks can be added with estimated depth from a single stereo pair. Furthermore, redundancy in the visual information due to the overlapping field of view helps stabilize tracks observed in both images. In essence the problems monocular systems face in every frame for every track are reduced to rare occurrences for the stereo system.

The hypothesis motivating the design of our system is that the methods used in monocular visual odometry are more robust by necessity and would perform well if generalized to the stereo case. In particular we use motion model predicted tracking by matching not dissimilar to the scheme proposed by Song et al [1], delayed outlier identification [2], long optimization windows and robust iterative triangulation.

This work describes the building blocks of our system as well as the general outline. We present a novel way of combining and tuning state of the art components to obtain a new stereo visual odometry system that outperforms the state of the art. Since our target is visual odometry for automotive applications, the many driving sequences in the KITTI odometry benchmark [3] serve as a strong indicator of practical performance.

II. RELATED WORK

Visual odometry is undoubtedly one of the most active topics in computer vision in the last years. Since the seminal work of Moravec [4] more and more research groups became interested in the topic of self-localization of a system relying solely or mainly on visual data. Real time monocular SLAM using commodity components was first achieved by Davison et al with MonoSLAM [5] and the filtering approach dominated the field for four years. But it was PTAM [6] by Klein et al, a bundle adjustment based system, which finally broke the barrier to direct augmented reality applications and piqued the interest of the public in 2007. These two works are still considered among the most important in the field because they showed what was possible to do by relying just on inexpensive cameras for self-localization and mapping. The obvious applications of successful systems attracted the attention of military contractors and many leading industries, especially in the automotive field. This, in turn, sparked an explosion of high quality works on the topic in the last few years. An extensive literature review on the topic is beyond the scope of this paper and we will focus exclusively on a brief overview the most relevant works for automotive applications that are of specific interest for this paper. The

[1]Computer Vision Laboratory, Linköping University, Sweden, name.lastname@liu.se
[2]Visual Sensorics & Inf. Proc. Lab (VSI), C.S.Dept., Frankfurt University, Germany
*At the time of submission 2015-01-30

interested reader can, however, find more complete reviews of the topic in review papers such as [7].

Visual odometry is an extremely complex and difficult problem. Numerous approaches have been proposed and virtually uncountable variants and assumptions have been tried to make the problem easier to treat. We like to cluster the existing systems using four different aspects:

- Feature-based versus direct systems
- Global (loop-closing) versus local systems
- Filter based versus bundle adjustment based systems
- Monocular versus stereo systems

Feature based systems extract keypoints and track or match them in subsequent frames. These matches are then used to compute the egomotion by the means of essential matrix estimation [8], PNP [9], [1] or the trifocal tensor [10]. Direct methods on the other hand operate on the whole images thus producing a dense or semi-dense motion field [11], [12], [13], [14].

Global systems keep track of the complete map built over time so that at any time instant it is possible to identify a previously visited location, thus allowing to correct the current estimate of the trajectory [6], [12]. As argued by Song et al. this is in general not practical for automotive applications [1]. Alternative approaches keeping track solely of recent map information have become more popular in recent years [1], [9], [15], [15], [16].

Many approaches make use of some form of filter to obtain a smooth trajectory for the visual odometry and to reduce the negative effect of noisy measurements [10], [17], [18]. This approach simplifies the fusion of ancillary sensors such as IMUs, but complicates the process to revert associations retroactively [19], [20], [21], [22]. The main alternative to this approach is to instead rely on a bundle adjustment step to refine the map and camera poses by minimizing the total reprojection error [23], [9], [8], [15].

Monocular approaches make use solely of the information provided by a single moving camera, thus information on the scene and the motion can only be recovered up to a global scaling factor without assumptions on the scene [1], [24]. Stereo systems, on the other hand, can rely on the known transformation between the two cameras to extract the metric scale and on the redundancy of the visual data to improve stability and robustness [9], [24], [16], [10], [15], [11].

Comparison between different methods has been made easier recently thanks to a number of free datasets that have been made available [3], [25], [26], [27]. The KITTI dataset in particular is currently the most popular one, partly thanks to the benchmark test made available by the authors [3]. The top performing visual-only systems to date are all stereo based papers [9], [16], [11], [15]. An exception to this rule is given by the work of Song et al. which is among the best scoring systems despite relying on monocular data [1]. This system is a feature based, local method using bundle-adjustment. The system we propose has many affinities with this work, but also exploits the advantages given by a calibrated stereo setup.

III. Visual Odometry

This section outlines our visual odometry system and its components. We denote the chain of measurements in a sequence of images and the associated $3D$ point as a 'track'.

The system processes the stereo sequence as follows:

Algorithm 1 Main Loop

```
for each stereo pair do
    Features = Extract Features(FAST,BRIEF)
    Track and estimate pose
    Local bundle adjustment
    Add new tracks
```

Initialization of the system differs only in that the motion model is not applied during the first four stereo-pairs of tracking.

A. Feature extraction

A scale pyramid is built for each image in the stereo pair. For each level of the pyramid the FAST [28] corners are extracted and subjected to a corner response, Adaptive Non-Maxima Suppression (ANMS) filter with a fixed radius of five. The pyramid consists of the full size image and a $2/3$ subsampling. BRIEF descriptors are computed for each FAST corner at the corresponding level [29].

The ANMS filter both reduces the computational cost and improves estimation by improving the spatial distribution of the tracks [30]. The BRIEF descriptor was chosen due to a small advantage in training set accuracy over the FREAK [31] and ORB [32] descriptors.

B. Tracking

The latest BRIEF descriptor of every recently measured track is searched after in a large window (radius 25 pixels) around its predicted position in the new left image by BRIEF descriptor matching. The track position is predicted using the pose estimate from the motion model and the estimated $3D$ position of the corresponding point.

Since descriptor matching often results in several potential candidates, we use a two step approach. First, the minimum Hamming distance $2D - 3D$ correspondences are used to estimate the pose via RANSAC-PNP. Second, the estimated pose is used to guide the matching, selecting the best candidate taking both appearance and re-projection error, thresholded to 3 pixels, into account. Tracks found in the left image are matched into the right in the same way with the same criteria, however we do not require that a right image match is found. Finally every inlier is matched to the right image using descriptor matching, the measurement added to the track if it passes a test on the re-projection error given the known stereo configuration.

If fewer than fifty tracks are found, the result is considered untrustworthy and the system uses the motion model to predict the pose of the current frame and reinitializes.

C. Track Replenishment

We choose to keep ≈ 500 tracks at all times, adding new tracks as needed by the following method. The FAST corners of the left image are ANMS filtered with a fixed radius of 18 pixels removing every corner too close to an existing track or another corner with a higher response strength. The descriptors of the remaining features are then matched to the right image, triangulated and filtered by their re-projection error. Matches are added as candidate tracks. Candidate tracks are searched after in the next stereo pair images and added as proper tracks if they are found and pass the re-projection test and no more than 1000 tracks are already present. Excess tracks shorter than four stereo-pairs are discarded.

D. Pose prediction

We use a constant acceleration model in world coordinates to predict the pose with the added constraint that the car is driving forward i.e. along the optical axis of the camera.
Let:

- The position at time t: p_t
- The velocity at time t: v_t
- The acceleration at time t: a_t
- The velocity in the camera coordinates at time t: $v_{c,t}$
- A noise component $n_{x,t} \sim N(0,\sigma^2)$

Model:

$$p_{t+1} = p_t + vt_\delta + a\frac{t_\delta^2}{2} + n_{p,t} \tag{1}$$

$$v_{t+1} = v_t + a * t_\delta + n_{v,t} \tag{2}$$

$$a_{t+1} = a_t + n_{a,t} \tag{3}$$

$$(1,1,0)v_{c,t} = n_{vc,t} \tag{4}$$

The predictive states i.e. v_t,a_t are found by minimizing $\sum(n_{p,t}^2+n_{v,t}^2+n_{a,t}^2+n_{vc,t}^2)$ over a window of twenty vehicle poses.

E. RANSAC-PNP

We use the P3P of Kneip et al [33] wrapped in the MLESAC [34] loop followed by optimization of the re-projection errors of the inlier set using the *Ceres Solver*. We perform $[250-1000]$ RANSAC iterations depending on data.

F. Local Bundle adjustment

The total track re-projection errors are minimized by the *Ceres Solver* [35] over a local window over the track $3D$ positions and the vehicle poses. The cost contains every measurement of every recently measured track with an average re-projection error below 3 pixels.

Every track failing this test is re-triangulated independently, iteratively removing the worst measurement until the average re-projection error reaches the threshold. We found experimentally that this reduces the total computational cost compared to the use of robust cost functions.

The vehicle poses outside the optimization window which enter the cost are set constant approximating the proper transfer of the information.

- Let x_i be a $3D$ point feature.
- Let $\wp : \wp(x_i) = \frac{1}{x_{i2}}\binom{x_{i0}}{x_{i1}}$
- Let P_c transform from vehicle to camera c coordinates.
- Let P_t transform from world to vehicle at time t.
- Let y_{cti} be a pinhole normalized measurement of x_i by camera c at time t.

We minimize an equivalent to:

$$\sum(y_{cti} - \wp(P_cP_tx_i))^2 \forall c,t,i \in window \tag{5}$$

The rotations are parametrized as unit-quaternions. Normalization is maintained by backprojection to the unit sphere.

G. Robust Triangulation

New tracks or tracks which may have been corrupted by outliers need to be triangulated before becoming part of the BA optimization cost function. In the former case to ensure a good initalization and in the latter to improve convergence speed.

A track is triangulated by minimizing $\sum(y_{ct} - \wp(x_{ct}))^2$ where $x_{ct} = P_cP_tx$ The optimization is initialized by the midway method applied to the two measurements taken the furthest apart, a track which is found behind either camera is moved in front of both. We use iterative minimization of the re-projection error due to its superior accuracy and performance over polynomial methods [36].

IV. EXPERIMENTS

For our tests, we use the publicly available KITTI dataset. This dataset provides ground truth egomotion for the 11 training sequences, furthermore a benchmark is available to test methods on a separate set of test sequences. This benchmark is used to test the methods for the KITTI odometry challenge. The average change in pose per meter is used as error metric. The distribution of the per frame error is computed using the training data ground truth.Since automotive applications require low latency the errors are computed on the estimates directly after the tracking step for each stereo-pair frame.

Our method is currently the top ranked stereo system in the KITTI odometry benchmark [†] under the name *cv4xv1-sc*.

A. Benchmark Results

In this section we present some sample results our system achieved on the KITTI benchmark test sequences.

Figures 1 and 2 show the path reconstructed from our visual odometry system compared with the ground truth data on two test sequences of the KITTI benchmark.

Further test results are available on the KITTI odometry website.

Figures: 3 and 4 Show that our method(CV4X) outperforms the state of the art visual odometry systems. Including the previously top ranked MFI [9], TLBBA [15] and 2FOCC aswell as MLM-SFM [1] and even DEMO [37] a monocular system supported by a high end laser depth sensor. The

[†]http://www.cvlibs.net/datasets/kitti/eval_odometry.php (accessed on 30 Jan 2015)

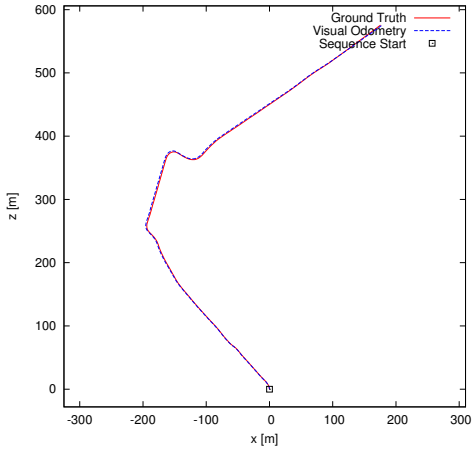

Fig. 1. Reconstructed path for test sequence 11. Ground truth and our method.

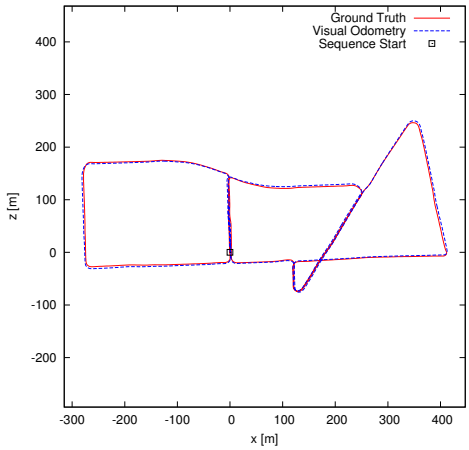

Fig. 2. Reconstructed path for test sequence 13. Ground truth and our method.

results show a slight improvement on the rotation estimate and a significant improvement on the translation estimate. Figures 1 and 2 show the trajectory of the vehicle as estimated by our system.

B. Results on the training data

We also provide some results our method achieved on the training set of the KITTI benchmark. Ground truth is provided for these sequences, we therefore evaluated our method on them using the same metrics used for the test set by the KITTI benchmark.

Figures 5 and 6 respectively show the rotation and translation error plots for all the 11 sequences in the KITTI training set.

In figures 7 and 8, we show the distribution of the errors in the training sequences, for rotation and translation respectively.

Its worth noting that the systems recovery method is not tested as no tracking failure occurs in any of the KITTI sequences.

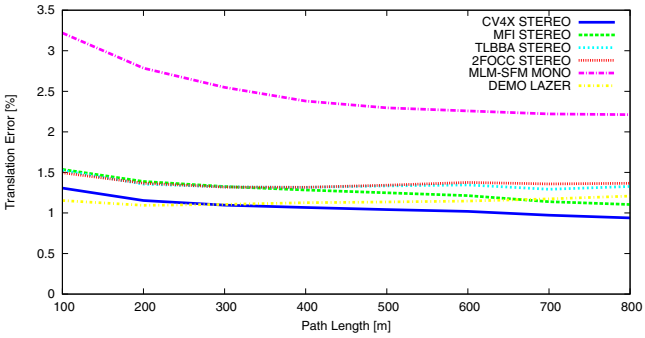

Fig. 3. Comparison of the average KITTI testing translation results

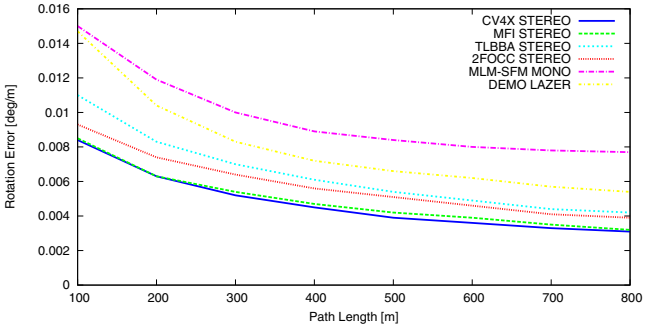

Fig. 4. Comparison of the average KITTI testing rotation results

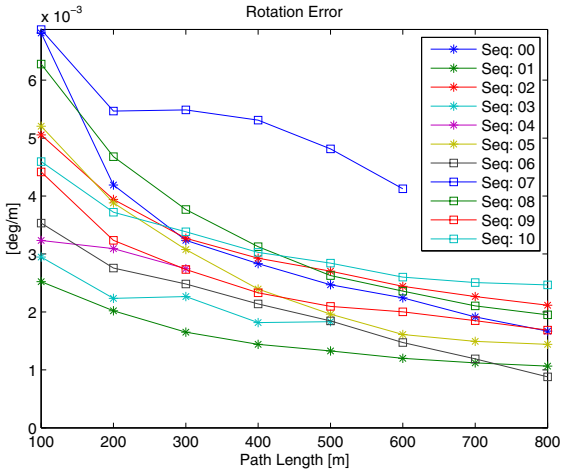

Fig. 5. Per training sequence rotation errors

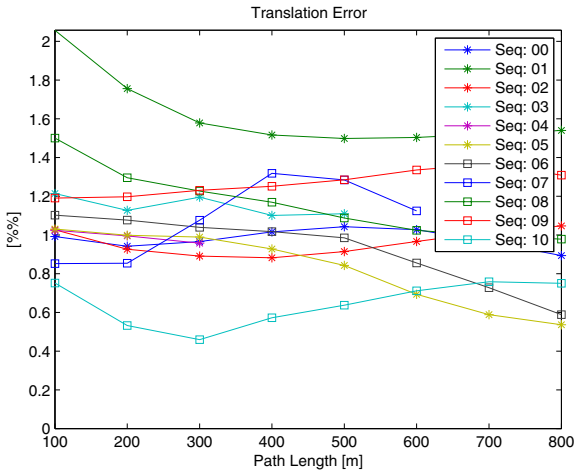

Fig. 6. Per training sequence translation errors

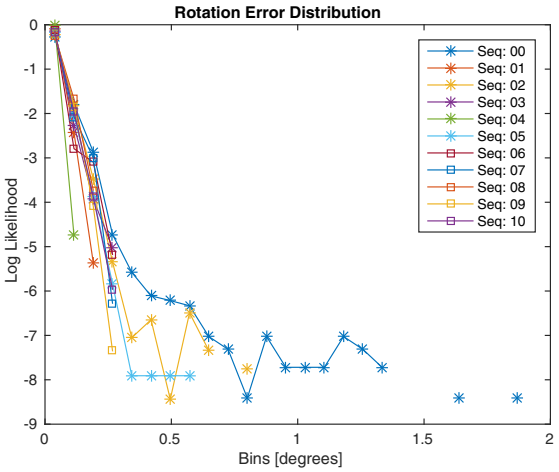

Fig. 7. Rotation error distribution

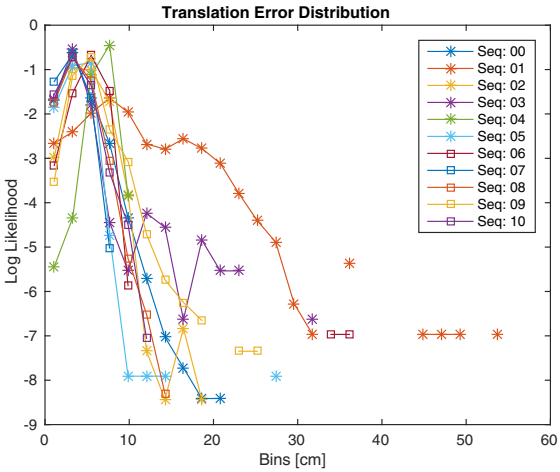

Fig. 8. Absolute Translation error distribution

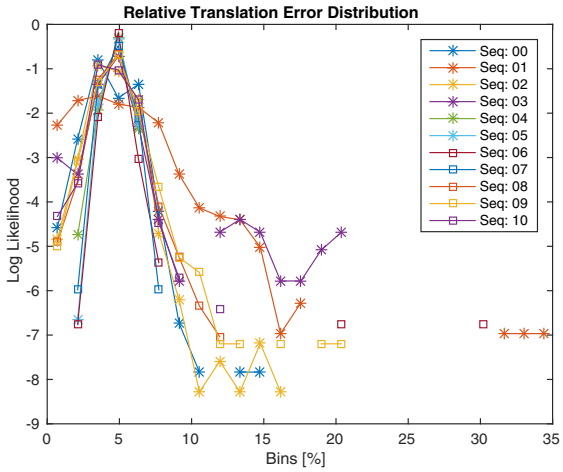

Fig. 9. Relative Translation error distribution for speeds $> 30km/h$

C. Time Performance

Similar to the elegant two thread PTAM system, our system parallelizes tracking and mapping. We also use OpenMP and cuda for further parallelization subtasks.

The bundle adjustment iterates for up to 90 ms and the tracking takes on average 140 ms. We believe that further CUDA paralellization should allow the system to run in the full 10 fps of the KITTI dataset video source. We performed the experiments on a 4Ghz core I7x4 and a GTX 770 graphics card.

V. Analysis

The analysis of the results shows that the translation errors vary significantly among the sequences in the training set. In particular for sequence 01, a highway scene with largely distant image areas driving at high speed, the translation estimates are poor. This is most likely caused by the very long distances of the majority of feature points from the cameras. This sequence also has severe perceptual aliasing during a few frames, which cause failures if the motion model tracking constraint is disabled, indicating that the motion prediction improves robustness.

The translation error distribution graphs show a significantly higher noise than the long term drift, this indicates that an error in one frame will be compensated for by opposed error in the next. Further the error is also reduced by the BA optimization. Figure 9 also shows that the error is proportional to speed as expected.

It is also interesting that the error distributions show how the system is able to recover from large PNP errors, up to 2 degrees or nearly 30% presumably surrounded by frames with opposed errors. This may also explain the rather counterintuitive observation that the system is less accurate and robust with a small re-projection threshold such as 1 pixel rather than 3 when evaluated on the training sequences. Though the errors do not cause a tracking failure, they are the main cause of the odd behaviour of the KITTI rotation

error for training sequence 00. The error occurs during rapid rotations which cause the loss of all longer older tracks.

Additionally, our results also show that, though there is a great deal of work aimed at achieving sub-pixel tracking accuracy, this is not necessary for visual odometry when the poses are constrained by hundreds of tracks.

The motion model provides predictions of sufficient quality for its purpose, given the smoothness of the trajectory of a car. Direct inclusion of the motion model in the main cost function, as is common in filtering approaches, holds the promise of further accuracy improvements.

VI. ACKNOWLEDGEMENTS

Research funded in part by: Daimler, Vinnova through grant iQmatic, and ELLIIT & EMC^2, funded by the Swedish Research Council.

VII. CONCLUSIONS

We have developed a stereo visual odometry system which incorporates several techniques developed for monocular systems and show that it achieves state of the art performace on the KITTI odometry benchmark, outperforming all other published, KITTI ranked, stereo odometry methods.

REFERENCES

[1] S. Song, M. Chandraker, and C. C. Guest, "Parallel, real-time monocular visual odometry," in *ICRA*, Karlsruhe, Germany, May 6-10, 2013.

[2] M. Persson, "Online Monocular SLAM: Rittums," Lith-ISY-EX–13/4741-SE, Linköping University, Sweden, 2014.

[3] 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.

[4] H. P. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover." DTIC Document, Tech. Rep., 1980.

[5] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, Oct 2003, pp. 1403–1410 vol.2.

[6] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE, 2007, pp. 225–234.

[7] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *Robotics & Automation Magazine, IEEE*, vol. 18, no. 4, pp. 80–92, 2011.

[8] J. Hedborg, P.-E. Forssén, and M. Felsberg, "Fast and accurate structure and motion estimation," in *Advances in Visual Computing*. Springer, 2009, pp. 211–222.

[9] H. Badino, A. Yamamoto, and T. Kanade, "Visual odometry by multi-frame feature integration," in *First International Workshop on Computer Vision for Autonomous Driving at ICCV*, December 2013.

[10] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*. IEEE, 2010, pp. 486–492.

[11] C. Forster, M. Pizzoli, and D. Scaramuzza, "Svo: Fast semi-direct monocular visual odometry," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, 2014.

[12] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "Dtam: Dense tracking and mapping in real-time," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2320–2327.

[13] S. Lovegrove, A. J. Davison, and J. Ibanez-Guzmán, "Accurate visual odometry from a rear parking camera," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE, 2011, pp. 788–793.

[14] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 834–849.

[15] W. Lu, Z. Xiang, and J. Liu, "High-performance visual odometry with two- stage local binocular ba and gpu," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 1107–1112.

[16] F. Bellavia, M. Fanfani, F. Pazzaglia, and C. Colombo, "Robust selective stereo slam without loop closure and bundle adjustment." in *ICIAP (1)*, ser. Lecture Notes in Computer Science, vol. 8156. Springer, 2013, pp. 462–471.

[17] J. Montiel, J. Civera, and A. J. Davison, "Unified inverse depth parametrization for monocular slam," *analysis*, vol. 9, p. 1, 2006.

[18] M. Kaess, A. Ranganathan, and F. Dellaert, "isam: Incremental smoothing and mapping," *Robotics, IEEE Transactions on*, vol. 24, no. 6, pp. 1365–1378, 2008.

[19] K. Konolige, M. Agrawal, and J. Sola, "Large-scale visual odometry for rough terrain," in *Robotics Research*. Springer, 2011, pp. 201–212.

[20] 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.

[21] E. S. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach," *The International Journal of Robotics Research*, vol. 30, no. 4, pp. 407–430, 2011.

[22] M. Pollefeys, D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, *et al.*, "Detailed real-time urban 3d reconstruction from video," *International Journal of Computer Vision*, vol. 78, no. 2-3, pp. 143–167, 2008.

[23] H. Strasdat, J. Montiel, and A. J. Davison, "Real-time monocular slam: Why filter?" in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 2657–2664.

[24] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *IV*, Baden-Baden, Germany, June 2011.

[25] G. Pandey, J. R. McBride, and R. M. Eustice, "Ford campus vision and lidar data set," *International Journal of Robotics Research*, vol. 30, no. 13, pp. 1543–1552, November 2011.

[26] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman, "The new college vision and laser data set," *The International Journal of Robotics Research*, vol. 28, no. 5, pp. 595–599, May 2009. [Online]. Available: http://www.robots.ox.ac.uk/NewCollegeData/

[27] P. Koschorrek, T. Piccini, P. Oberg, M. Felsberg, L. Nielsen, and R. Mester, "A multi-sensor traffic scene dataset with omnidirectional video," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. IEEE, 2013, pp. 727–734.

[28] E. Rosten and T. Drummond", ""machine learning for high-speed corner detection"," in *"European Conference on Computer Vision"*, vol. "1", "May" "2006", pp. "430–443". [Online]. Available: "http://edwardrosten.com/work/rosten_2006_machine.pdf"

[29] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "BRIEF: Computing a Local Binary Descriptor Very Fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1281–1298, 2012.

[30] S. Gauglitz, L. Foschini, M. Turk, and T. Hollerer, "Efficiently selecting spatially distributed keypoints for visual tracking," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, Sept 2011, pp. 1869–1872.

[31] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 510–517.

[32] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, Nov 2011, pp. 2564–2571.

[33] L. Kneip, D. Scaramuzza, and R. Siegwart, "A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation," in *Proc. of The IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, USA, June 2011.

[34] P. H. S. Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry," *Computer Vision and Image Understanding*, vol. 78, pp. 138–156, 2000.

[35] S. Agarwal, K. Mierle, and Others, "Ceres solver," http://ceres-solver.org.

[36] J. Hedborg, A. Robinson, and M. Felsberg, "Robust Three-View Triangulation Done Fast," in *Proceedings: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2014*, 2014, pp. 152–157.

[37] J. Zhang, M. Kaess, and S. Singh, "Real-time depth enhanced monocular odometry," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Chicago, IL, Sept. 2014.