

Dependency of Odometry and 3D Point Cloud Quality on Lighting Conditions

Hongmou Zhang¹, Ines Ernst¹, Sergey Zuev¹, Anko Börner¹, Martin Knoche², and Reinhard Klette³

¹ Institute of Optical Sensor Systems

German Aerospace Centre (DLR), Berlin, Germany

² Northland Innovation Centre (NIC), Whangarei, New Zealand

³ School of Engineering, Computer, and Mathematical Sciences

Auckland University of Technology (AUT), Auckland, New Zealand

Abstract—The paper discusses an evaluation of visual odometry accuracy with respect to available lighting. Very extensive test data (along one and the same road in New Zealand) have been recorded at day and at night. Used sensors are stereo cameras, inertial measurement unit (IMU), and GPS. The paper discusses odometry and 3D point cloud reconstruction results for the cases when using visual odometry only (i.e. based on the stereo camera data), or additionally also an IMU. Results show how trajectory and 3D point cloud recovery under low-light conditions benefit from the use of additional sensors.

I. INTRODUCTION

Odometry (estimation of trajectories, i.e. change of pose over time) and 3D point cloud reconstruction are two subjects relevant to modern vehicles, or to test field studies for evaluating modern vehicles. The *Northland Transport Technology Testbed* (N3T) aims at providing a test field for control and safety modules of modern vehicles, with a particular focus on the safety of trucks [1]. The research project “Digital Roads New Zealand” has been defined in connection with the needs of N3T, with participating academics at the *German Aerospace Centre* (DLR), see [2], and at *Auckland University of Technology* [3].

Visual odometry (VO) estimates trajectories by using mono- or multi-ocular camera systems; see [4] for an evaluation of various programs for visual odometry. Test sequences, provided by this benchmark website, are typically of a length of around 400 stereo frames. In this paper we use test data for a day-time and a night-time run along a road of about 6 km; during these runs we recorded 6,998 or 7,577 stereo frames (at 10 fps) for VO, respectively.

DLR developed an *integrated positioning system* (IPS) that may also be installed on a car for providing precise odometry of the vehicle on the road [5], [6], [7] while recording data by stereo vision and an *inertial measurement unit* (IMU). Additionally, the used test vehicle also operated a low-cost GPS. In this paper, we discuss odometry and 3D point cloud reconstruction results; at first odometry for VO only (using the ORB-SLAM method; see [8]), and then also 3D point cloud recovery for the IPS (using VO and IMU).

The studied road (see online maps for *Otaika Valley Road*) is one of the traffic-accident hotspots in New Zealand. Fig. 1

shows a measured trajectory for the night-time test run on this road; the red cross indicates the start position; the green cross is the end position (i.e. intended to be identical with the start position).

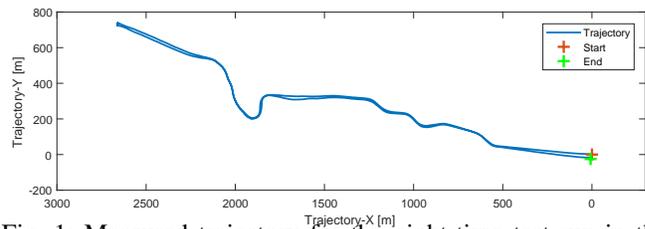


Fig. 1: Measured trajectory for the night-time test run in the local coordinate system

The paper is structured as follows. Section II discusses data quality and the exclusive use of visual odometry. A brief Section III informs about the used IPS (i.e. VO and IMU). Section IV discusses the quality of obtained clouds of 3D points. Section V concludes.

II. DATA QUALITY AND VISUAL ODOMETRY

Having exact calibration information about the given binocular camera system (a left and a right camera) at hand, a relative transform between stereo-frame pairs can be calculated with high accuracy. By integration of relative transforms, the motion trajectory of the camera system can be obtained.

First, visual features are selected in the left image of the stereo image pair at time k . Second, selected features are matched to the right image (known as stereo matching). Third, 3D points can be obtained from successfully matched features [9]. Fourth, those 3D points are back-projected into the subsequent stereo-image frame at time $k+1$ with an initial guess of a relative transformation T . Features from the image pair at time k are also matched with features in the image pair at time $k+1$ directly. An optimized relative transform can be obtained by minimizing the error distance between back-projected features and matched features. Eq. (1) defines the calculation:

$$\sum_n \left| p_{k+1}^{(n)} - K[R|t]P_k^{(n)} \right|^2 \quad (1)$$

where K is the camera calibration matrix. A matched feature with index n in image I_{k+1} is denoted by $p_{k+1}^{(n)}$. A reconstructed 3D object point at time t_k is denoted by $P_k^{(n)}$. Let R and t denote rotation and translation components of the relative transform T .

The above steps are repeated for each newly arriving image pair. By composition of all the optimized relative transforms, we obtain the sequence of subsequent positions and directions of the camera system (i.e. the trajectory).

Figure 2 shows images recorded by night (with no additional headlights on). For enhancing the performance of *semi-global matching* (SGM), selected for stereo matching, we pre-process night image data for contrast enhancement. Fig. 2, right column, shows pre-processing results.



Fig. 2: *Top left*: Original image. *Top right*: Pre-processed image with sufficient quality for SGM. *Bottom left*: Very dark original image. *Bottom right*: Pre-processed image with very poor quality for subsequent processing

For illustrating the varying quality of image data under low-light conditions, we analysed the relationship between the number of detected features and image contrast. Any input image is subdivided into 10×10 windows; we calculate the standard deviation of pixel intensities in these windows and the mean value of those standard deviations for the whole image. We also specify the number of detected features for each image. We apply a sliding mean (window size 51) to both resulting curves. Furthermore, we normalize both curves to zero mean and standard deviation equal to 1. The result for the night-time sequence is shown in Fig. 3. There are critical drops in both curves along the sequence. There is a high correlation between both curves.

We process the recorded day-time and night-time data by the visual odometry system ORB-SLAM.¹ Figure 4 shows ORB-SLAM results. In the figure, the first image frame is shown as the red triangle; each *keyframe*² is shown as a blue

¹ Position 29 on 27 June 2018 on [4]

² ORB-SLAM tracks features detected in a keyframe as far as it can; if the number of tracked features is below a threshold then a new keyframe is defined: Doing stereo matching, then track these new features.

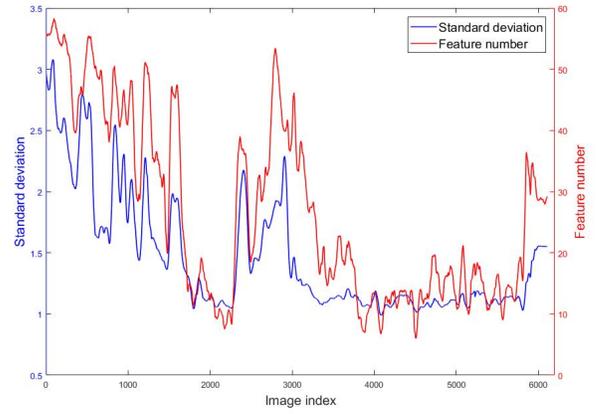


Fig. 3: Sliding means (window size 51; 25 on both sides of current time slot) for number of tracked features and for frame-by-frame mean of local standard deviation in recorded left images (for the whole night-time sequence)

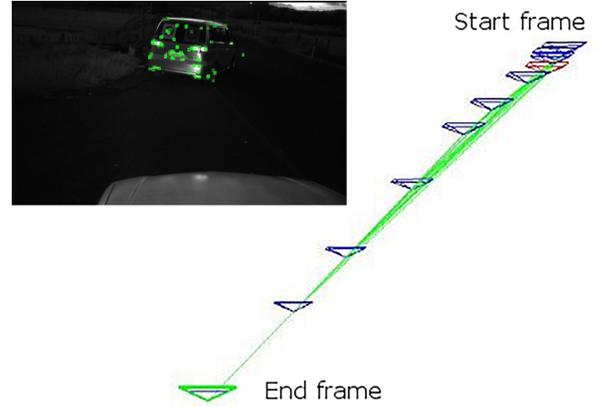


Fig. 4: *Top*: Screenshot for ORB-SLAM: Detected features in the night-time sequence are limited to a car driving in front, due to the strong reflections of headlights. *Bottom*: Measured trajectory by ORB-SLAM for the first 5 seconds

triangle; the green triangle indicates the current frame. During the test, ORB-SLAM fails feature tracking in the night-time sequences just after 50 frames (i.e. 5 seconds) hence just a short trajectory can be obtained. Moreover, the measured trajectory is completely wrong. In the first recorded frames, a car in front of the test vehicle moved forward while the test vehicle was still parked. However, ORB-SLAM wrongly generates a backward moving trajectory for these first frames due to features detected from the front car.

In general, VO alone performed badly under low-light conditions, but reasonably well under day-light conditions (as known from results such as on [4]; however, these benchmark sequences do not contain any night-time sequence). The quality of VO measurements highly relies on the number and distribution of features. Thus, the performance of a camera-only system is limited under low-light conditions where only a few features can be detected in each image.

III. VISUAL ODOMETRY AND IMU

In the IPS, features are detected by an extended AGAST feature detector [6]. To ensure a better balance between performance and processing time, the number of detected features for each image pair is dynamically adjusted, based on the number of successfully matched features in the next step [7].

In addition to the cameras, the IPS integrates an *inertial measurement unit* (IMU) as a self-contained sensor; the IMU measurement is independent of any external signal as well as of lighting conditions. The sampling frequency of the IMU is normally much higher than that of the camera system; thus, the IMU can provide a good reflection of the system's state. IPS's IMU contains a 3-axis gyroscope which provides angular velocity ω^b , and a 3-axis accelerometer which measures the acceleration a^b of the movement; superscript b stands for body-frame.

In the IPS, measurements from VO and IMU are fused by a Kalman filter to provide the most reliable trajectory measurement.

Both for day-time and night-time run, start and end point of the test vehicle with IPS are intended to be at exactly the same position. A measured trajectory was shown in Fig. 1. The distance error between measured start and end point is known as *closed-loop error*; a common measure to indicate the quality of odometry. For the night-time run, the test vehicle travelled 6,196 m, with a closed-loop error of 26.12 m, what is about 0.4% of the travelled distance. Due to the combination of VO with IMU, the IPS never stops calculating a trajectory.

Moreover, we measured GPS coordinates of some significant landmarks besides the road (e.g., traffic signs) by a *real-time kinematic* (RTK) GPS device. The average accuracy of landmark's coordinates measurement is 0.018 m in horizontal and 0.025 m in vertical direction. The measured landmarks are distributed along the testing road. They are used as ground truth to evaluate the quality of the measured trajectories. In order to compare with the landmarks, a measured trajectory is transformed and calculated by fusion with the VO-IMU-GPS data source in the global geographic coordinate system; details can be found in [15]. Figure 5 shows experimental results. From the latitude-longitude figure it can be seen that the trajectory accurately passes through all the landmarks; this demonstrates that the IPS measures a high-precision trajectory in the 2D map, which is normally important information for navigation tasks. Elevation-longitude shows us also a high-quality trajectory which confirmed a high consistence between trajectory and landmarks. Errors between landmarks and trajectory in the elevation-longitude figure are very small in comparison to the overall difference in height on the testing road. This experiment proves that the IPS system also provides an accurate trajectory in low-light environments.

IV. CLOUDS OF 3D POINTS

The IPS with its stereo camera approach generates two images at the same time, which can not only be used for reliable visual odometry estimation but also for the generation

of *high-density depth maps*. Subsequently, these point clouds from all image pairs can be merged into a high-density cloud and filtered into a voxel grid of an appropriate resolution and size. A large-scale 3D cloud of points for the entire observed area can be generated; see Fig. 6. Obviously, this approach only works if all contributing components (stereo matching for each pair of images and trajectory calculation) possess a very high accuracy.

A special challenge for the extended IPS was to allow high quality 3D point generation in real time. Our computationally expensive dense stereo matching is executed on a *graphics processor unit* (GPU). Based on earlier developments [16], we implemented SGM in OpenCL. From the various possible cost functions [17] with their pros and cons for different conditions we selected a census cost function for the data term; see, for example, [18] for a discussion of the influence of data and smoothness term in energy optimisation for stereo matching.

The overall frame rate for point cloud generation is dynamically adapted as follows: If the calculated IPS navigation solution shows a substantial difference in pose or time to the previously used image pair, a new local 3D point cloud is extracted from the depth maps and transformed into the global navigation frame. All steps, including trajectory estimation and dense depth map generation with a sufficient frame rate and disparity resolution can be carried out on a capable laptop PC in real time.

The global point cloud accuracy profits immediately from navigation precision. The cloud quality in detail is further strongly dependent on the stereo camera parameters, especially on pixel resolution and base length. For a fixed camera setup, the depth resolution and local accuracy of 3D points is

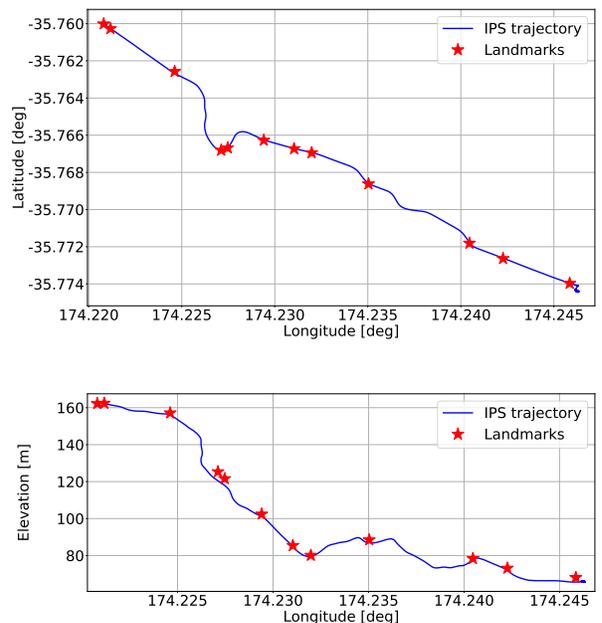


Fig. 5: Comparisons of measured IPS trajectory in a low-light environment with landmarks. *Top*: The trajectory and landmarks in latitude-longitude view. *Bottom*: The trajectory and landmarks in elevation-longitude view.

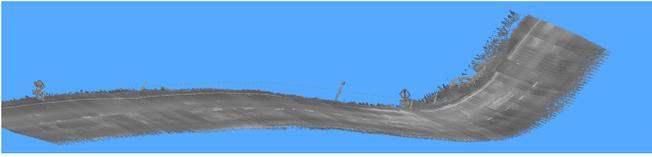


Fig. 6: Point cloud generated from a day-time run of the considered road segment is used as a reference (parallel projection)



Fig. 7: *Top*: 3D point cloud for the whole road segment (one direction) generated from pre-processed night-time images. *Middle*: Curve segment. *Bottom*: Details for two sections of the shown road segment.

mostly determined by the minimum distance of the shown objects to the camera while passing the object. In case of low light conditions, the extraction of depth images is a special challenge. SGM basically relies on the evaluation of a cost function for structure similarities in small image patches. This requires at least some texture and grey value differences in the images. A census cost function compares local intensity differences around the considered pixels in both images. Therefore it is very tolerant to overall brightness differences between the images but not to noise, which especially increases in homogeneous areas with low lighting conditions.

All point cloud segments presented in the paper are calculated from depth maps generated with full image resolution and filtered to a 2 cm sized voxel grid.

The 3D point cloud for the whole road in Fig. 7 shows image-quality dependent cloud densities; compare with Fig. 3. The top image in Fig. 7 gives an overview for the whole path (one direction), generated from pre-processed night-time images. The middle one provides a segment of this whole cloud. After light-condition changes and corresponding camera exposure control cycles (while driving into the forest), the

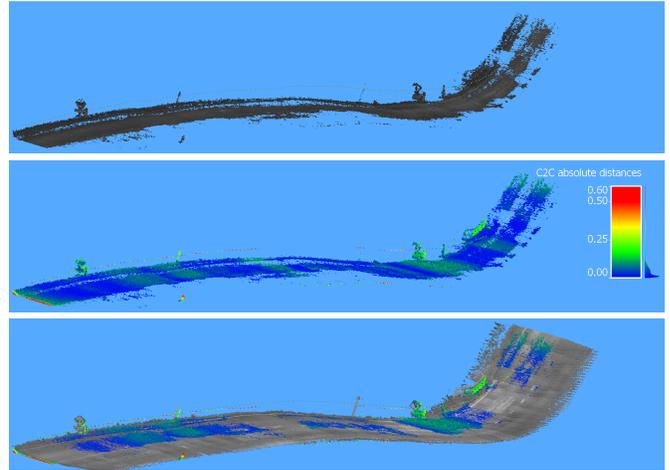


Fig. 8: Comparisons of cloud for original night-time sequence with day-time cloud (parallel projections)

generated point cloud density drops significantly due to an inadequate (despite pre-processing) image quality for SGM, up to almost only bright details like traffic posts or white road markings as shown between the yellow boxes (see bottom right of figure).

In case of dark and especially very noisy images the point cloud density suffers. In order to increase the number of valid 3D points, we executed image processing steps on the input images (with functions of scikit-image [19]) in an offline process. An adaptive histogram equalization followed by a total variance denoising proved favorable for SGM and the following point cloud generation. Figure 2 top row shows an image in original and sufficiently pre-processed form. However, extensive image processing may lead to inaccuracies of feature positions and therefore to a substantial negative influence on the visual odometry results. To avoid that, point clouds from pre-processed images pairs are fused with the help of a trajectory from original images to the final improved 3D point cloud.

Figure 8 shows point cloud results for original stereo images of the night-time run. The top of the figure shows the point

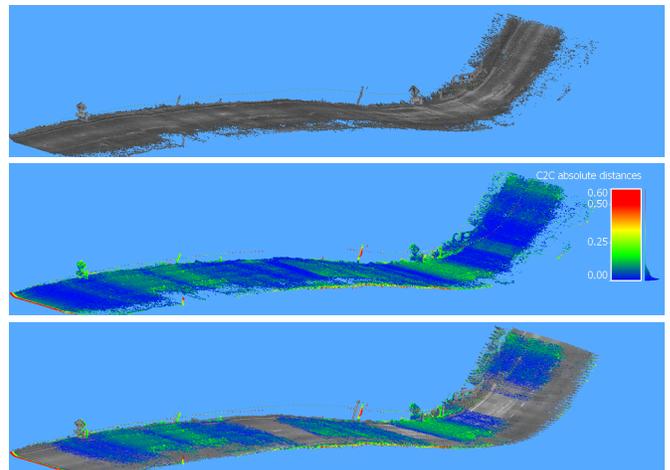


Fig. 9: Comparisons of cloud for pre-processed night-time sequence with day-time cloud (parallel projections)

cloud generated from original night-time data (compare with Fig. 6); the middle of the figure shows colour-coded point distances to the day-time cloud; the bottom of the figure shows the aligned clouds, the colour-coded night-time cloud and the grey-value-coded day-time cloud. Figure 9 shows point cloud results for pre-processed stereo images of the night-time run - in the same order as in Fig. 8 (i.e. point cloud, colour coded point distances to daylight cloud, and aligned colour coded cloud and grey value daylight cloud). With pre-processing, the point cloud contains significantly more 3D points in a visually better quality. If the images are too dark due to camera exposure control cycles induced by light condition changes (e.g. while driving into the forest) even using pre-processing gives poor results like in Fig. 2 bottom row; the 3D point cloud becomes sparse accordingly and contains almost only traffic posts and brighter road markings. Figure 7, bottom, shows quality differences which fundamentally depend on the remaining information in recorded images.

Generated point clouds for original and pre-processed images are given in Figs. 8 and 9; both rely on the same calculated trajectory from images recorded at night. That means that both clouds resemble each other in size and overall shape, but pre-processing led to significantly more and, maybe, slightly different points. A significant statement about quality and reproducibility of trajectories and point clouds based on low-light-condition data can be obtained by comparing 3D points with a ground truth model of the observed area. Such a model, unfortunately, is not yet available at sufficient resolution for our test area. The (outdoor) trajectory itself can be compared with high-accuracy GPS trajectories; see [BLINDED]. We recorded data on the test road several times under different conditions, hence we were able to generate a 3D point cloud for (“good”) day-time image data (taken just some hours before the selected night-time test run). This cloud serves as our point-cloud reference model. We compare the number and accuracy of generated 3D points, and even local (especially trajectory-dependent) shape differences of point clouds.

Figure 6 shows a uniquely identifiable cloud segment from the test area, generated under daylight conditions. Cloud segments of both of our clouds, calculated from original and pre-processed night-time images for the same region, have been aligned with the daylight cloud by an *iterative closest point* (ICP) algorithm. Subsequently, absolute point distances of both night-time clouds to the reference cloud have been calculated. These steps have been executed with the help of CloudCompare [20]; all presented point-cloud figures in this paper are also rendered with this software package. Figure 8 is for the original night-time cloud, and Fig. 9 for the optimized cloud; each figure shows a grey-value cloud, a colour-coded cloud of absolute point distances to the reference cloud, and an overlay of colour-coded cloud and grey-value reference cloud. The night-time clouds fit “very well” the daylight reference cloud; this means that the overall shape of the segment is reproduced from night-time images with high accuracy, and hence we can claim that the underlying daylight and night-

TABLE I: Parameters for absolute point differences to the reference cloud for original night-time cloud (Cloud 0), and the night-time cloud from pre-processed images (Cloud 1)

	Cloud 0	Cloud 1
Number of points	1,050,903	3,721,126
Mean distance to reference cloud	52 mm	50 mm
Standard deviation	124 mm	119 mm
Max. distance of 50%	34 mm	33 mm
Max. distance of 75%	62 mm	60 mm
Max. distance of 90%	88 mm	86 mm
Max. distance of 95%	105 mm	103 mm

time trajectories are of similar quality for the considered road segment.

Table I shows that the cloud from pre-processed images (Cloud 1) contains more than 3 times as many points as the cloud for original images (Cloud 0). Mean-absolute distances to the reference cloud, standard deviations, and percentiles are similar. Histograms of absolute point distances (see Fig. 10) illustrate that for both clouds, 50% of the points have a distance to the reference cloud smaller than 34 mm, and 90% smaller than 88 mm or 86 mm, respectively. (Blue histogram bars indicate the distance distribution for the original night-time cloud, green bars show the same for the cloud from pre-processed night-time images. Dashed lines mark the maximum distance for 50% of the points; solid lines show this for 90% of the points.) Pre-processing the images in the proposed way significantly improves the information contained in resulting point clouds, the additional points complete the original (day-time) cloud at a comparable accuracy.

Diagrams in Fig. 11 quantify point cloud qualities when comparing one of the clouds for night-time images with the daylight reference cloud. Both charts show (smoothed) numbers of VO features as red lines; the VO feature curve in the bottom chart is the left section of the overall plot given in Fig. 3. Green bars represent averaged numbers of 3D points generated from one image pair after voxel filtering in the final point clouds. For comparison, the vehicle speed, given by the trajectory, is plotted as a blue line. A green bar value is the quotient of the number of points (voxels) in a road-segment point cloud and the number of image pairs evaluated for that cloud. Assuming that image contents is only slightly changing while driving along a rural road, and the nearly constant vehicle velocity as shown, this ratio is a significant measure for point cloud densities under different conditions.

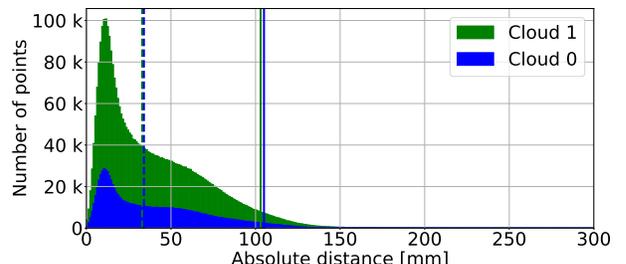


Fig. 10: Histograms of point distances to the daylight reference cloud

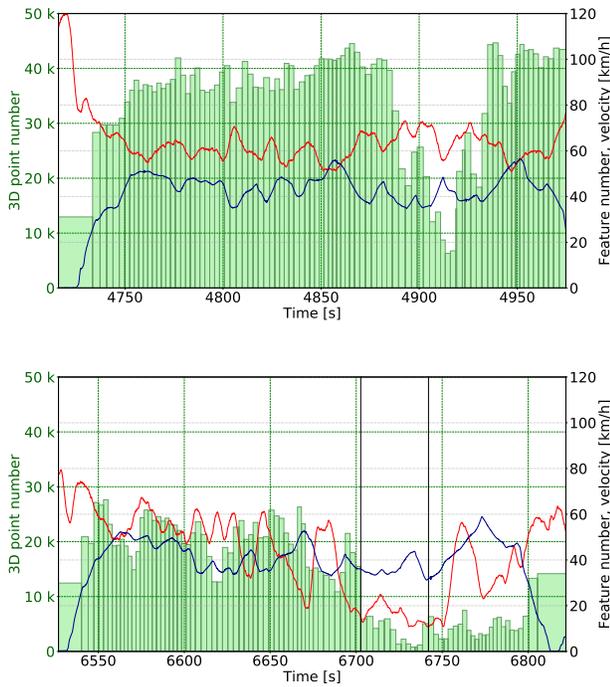


Fig. 11: Smoothed VO feature numbers (red lines) and averaged numbers of 3D points generated from one image pair in the final point clouds for one driving direction. A green bar value is the quotient of the number of points (voxels) in a road-segment point cloud and the number of image pairs evaluated for that cloud. *Top*: Diagram for the daylight (reference) point cloud. *Bottom*: Diagram for the night-time point cloud generated from pre-processed images

The upper part of Fig. 11 shows values for the daylight run used as a reference. The velocity is nearly constant over time, and the stable number of features found for VO calculation implies reliable VO information. The averaged number of 3D points per image pair is almost constant while driving, except for the section that leads through the forest. Due to darker images with lower contrast, this ratio drops to a value typically seen for night-time images, given in the lower diagram. For the night-time run with optimized input images this shows that the average number of 3D points that can be added per image falls to about half of the number for daylight reference values. The time period between the vertical black lines corresponds to the curved road segment between the yellow boxes in Fig. 7. Under these very-low-light conditions, both the number of VO features and the number of 3D points drop significantly. However, brighter objects like road markings and posts are still modelled in the point clouds as presented in the bottom part of Fig. 7 - they could be used, e.g., for an estimation of the slope angle for a continuation of a coarse road model.

V. CONCLUSIONS

The demonstrated combination of high-accuracy trajectory estimation (VO, IMU, and GPS) is stable over long time periods or trajectory intervals. It supports real-time 3D point

cloud generation and thus provides a promising technological basis for various applications, even for non-optimal (for the used camera system) low light conditions. In the case of very poor image quality, generated 3D point clouds contain only a few objects such as road markings and posts, but these are already sufficient for 3D matching with a given (partial) 3D road model. The proposed denoising step was prototypically implemented as an offline process, but it may be time-optimized or implemented on dedicated parallel hardware.

ACKNOWLEDGMENT

This is joint research undertaken at the German Aerospace Center's Institute of Optical Sensor Systems, Auckland University of Technology's Center for Robotics & Vision, and Northland Innovation Center, and authors appreciate the support by all the three institutions. We thank G. Russell, R. Brill and D. Williams from WSP-Opus, New Zealand, for surveying and providing the coordinates of the applied landmarks.

REFERENCES

- [1] The Northland Transport Technology Testbed. www.n3t.kiwi, 2018.
- [2] DLR Institute for Optical Sensor Systems. www.dlr.de/os/en/, 2018.
- [3] Centre for Robotics & Vision. cerv.aut.ac.nz, 2018.
- [4] Visual Odometry. The KITTI Vision Benchmark Suite, http://www.cvlibs.net/datasets/kitti/eval_odometry.php, June 2018.
- [5] A. Börner, D. Baumbach, M. Buder, A. Choinowski, I. Ernst, E. Funk, D. Griebßbach, A. Schischmanow, J. Wohlfeil, and S. Zuev, "IPS - A vision-aided navigation system," *Advanced Optical Technologies*, 6(2):121–129, 2017.
- [6] H. Zhang, J. Wohlfeil, and D. Griebßbach, "Extension and evaluation of the AGAST feature detector," *Proc. ISPRS Congress Annals*, vol. 3, pp. 133–137, 2016.
- [7] H. Zhang, J. Wohlfeil, D. Griebßbach, and A. Börner, "Eligible features segregation for real-time visual odometry," *Proc. 3D-NordOst*, in print, 2017.
- [8] R. Mur-Artal, J. Montiel, and J. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robotics*, 31(5):1147–1163, 2015.
- [9] D. Griebßbach, "Stereo-vision-aided inertial navigation," PhD thesis, Freie Universität, Berlin, 2014.
- [10] P. S. Maybeck, "Stochastic Models, Estimation and Control," *Mathematics in Science and Engineering*, Academic Press, 1979.
- [11] "Standard specification format guide and test procedure for single axis interferometric fiber optic gyros," *IEEE Standard 952*, 1997.
- [12] O. J. Woodman, "An introduction to inertial navigation," Univ. Cambridge, Computer Lab., UCAMCL-TR-696, 14, 2007.
- [13] J. Wendel, "Entwurfs- und Analysemethoden Integrierter Navigationssysteme," PhD thesis, Karlsruhe Institute of Technology, 2003.
- [14] D. Griebßbach, D. Baumbach, and S. Zuev, "Stereo-vision-aided inertial navigation for unknown indoor and outdoor environments," *Proc. Int. Conf. Indoor Positioning Indoor Navigation*, IEEE Xplore, 2014.
- [15] D. Baumbach, H. Zhang, S. Zuev, J. Wohlfeil, M. Knoche, and R. Klette, "GPS and IMU Require Visual Odometry for Elevation Accuracy," *Proc. IEEE Int. Conf. on Advanced Video and Signal-based Surveillance*, 2018.
- [16] I. Ernst and H. Hirschmüller, "Mutual information based semi-global stereo matching on the GPU," in *Int. Symp. Advances Visual Computing*, LNCS 5358, pp. 228–239, 2008.
- [17] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Trans. Pattern Analysis Machine Intelligence*, 31(9):1582–1599, 2009.
- [18] R. Klette, "Concise Computer Vision," Springer, London, 2014.
- [19] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Goullart, T. Yu, and the scikit-image contributors, "scikit-image: Image processing in Python," *PeerJ*, 2014.
- [20] GPL Software, "CloudCompare," version 2.8, retrieved from cloudcompare.org, 2017.