Multi Trajectory Pose Adjustment for Life-long Mapping

Marc Sons[1], Henning Lategahn[2], Christoph G. Keller[3], Christoph Stiller[1]

***Abstract*— Current highly automated and self-driving vehicles heavily depend on detailed maps since they free the system from many otherwise complex onboard processing tasks. However, depending on the environment and the fineness of the map, the validity span of maps is often short and a periodic remapping of large areas with sensor-packed mapping vehicles is beyond any feasibility. Crowd base mapping approaches using low cost sensors appear more practicable.**

Herein we propose a general method to align several survey trajectories of the same area which is fundamental for any life-long mapping. Our algorithm requires previously acquired pose differences as input. These differences induce a pose graph which is aligned yielding a minimum least-squares residual. Therefore, our method is independent from the underlying sensor technology.

For evaluation purposes, we align pose graphs from simulated pose differences and compare it against the ground truth. Furthermore, stereo cameras are used to obtain pose difference estimates by common visual odometry methods. We present quantitative results of the robustness and accuracy of our method based on these pose differences. The results are compared against a high precision GPS receiver. Our approach clearly outperforms this costly reference sensor.

I. INTRODUCTION

Highly automated and self-driving vehicles will reach the mass market within the next decade. Many current systems heavily depend on detailed map data structures [1], [2], [3], [4]. The map based vehicle automation seems to be the most promising approach at the moment since storing relevant information within maps is extremely appealing and frees one from many otherwise complex onboard processing tasks. Moreover, recent re-localization approaches [1], [2], [4], [5] are in fact map dependent.

A static world assumption is in the nature of maps which is, however, violated in nearly every realistic scenario. The finer the details of the maps are, the shorter is their validity span. Periodically remapping of e.g. city-scale or larger areas with sensor-packed mapping vehicles is obviously beyond any feasibility. In contrast, crowd based mapping approaches using low-cost series sensors close to serial production appear more practicable.

Aligning several independent survey trajectories of the same area is a key step of crowd based mapping methods. Computing this alignment with great precision is fundamental for any life-long mapping.

[1] with the Institute of Measurement and Control, Karlsruhe Institute of Technology, Karlsruhe, Germany, www.mrt.kit.edu

[2] with the Atlatec UG, Karlsruhe, Germany, Vision based mapping and localization, www.atlatec.de

[3] with the Daimler AG, Research & Development, Sindelfingen, Germany, www.daimler.com

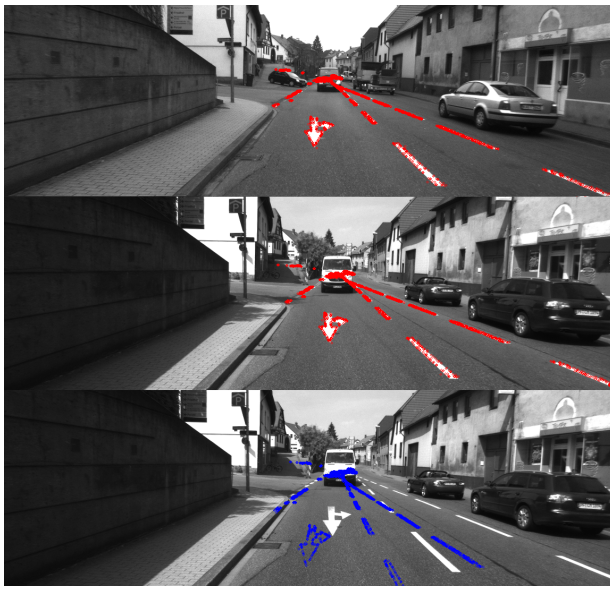

Fig. 1. Re-projection of extracted ground marking labels (red points) into different drives in an urban scenario. The image in the first row shows the drive from where the features are extracted from (ground truth). The image in the second row shows the re-projection of these features (red points) using the pose information of the resulting pose graph from our method. In comparison, the bottom image shows the re-projection of the same features with poses from a high precision GPS. Our method clearly outperforms DGPS.

Herein we present a novel algorithm to align several trajectories from different drives of the same area. The input to our method is a set of pose differences which establishes relations between poses within one trajectory and, additionally, between poses of different trajectories. The resulting pose graph forms a nonlinear optimization problem where all poses within are finally optimized yielding a minimum least-squares residual.

Our method is in fact independent from the underlying pose difference estimation method. Thus, any suitable sensor setup could be used. Moreover, our algorithm easily scales to vast environment as all optimizations are performed only on poses and circumvents the drastic limitations in scalability known from full bundle adjustment.

In order to complete the whole process and for evaluation purposes, we solely utilize stereo cameras and initially pair all images of a multitude of image recordings showing the same place. Thereafter, common visual odometry methods are used to estimate pair-wise pose differences. These estimates forms the input to the pose graph optimization problem.

We present a Monte Carlo simulation which demonstrates the potential of drift reduction of a multi trajectory alignment.

Thereafter we show results from real world experiments using stereo vision which demonstrate the robustness and high accuracy of our method. Therefore, salient 3D structures are projected from one drive into all interconnected ones merely by using pose information of the optimized pose graph. We compare these to a high precision GPS receiver (see Fig. 1). Our approach clearly outperforms this costly reference sensor in urban environments; many times by a significant margin.

Section II reviews related work. A general formulation of our algorithm is presented in Section III followed by a brief explanation of how pose differences can be estimated from stereo cameras. An experimental evaluation based on simulated and on real world data is given in Section IV. Finally, a summary of our contribution is presented in Section V.

II. RELATED WORK

The topics within this work are related to mapping in general [6], [7], [2], [5], [1], [8] and Simultaneous Localization and Mapping (SLAM) [9], [10], [11].

Mapping an unknown area and localize the ego position in this map at the same time is the main idea of SLAM. This approach has developed considerably in the past years and the preferred methodology changed from Kalman Filters to bundle adjustment [9], [11]. However, the computational complexity of SLAM scales poorly with the size of the map and becomes infeasible for large scale areas. Therefore, recent approaches decouple the mapping and localization problem [1], [2]. A map is precomputed by measurements from a survey drive. Thereafter a map relative relocalization with a sufficient accuracy is possible.

Mapping with methods of bundle adjustment constitutes an optimization problem where residuals of any type can be minimized. This allows the addition of pose difference measurements into the optimization problem [11].

Lategahn et al. [12] proposed a method to compute a map consisting of point features and 3D poses. These poses are obtained from an alignment of visual odometry and GPS measurements. Thereafter landmarks are added relatively to the precomputed poses in a second step. Schreiber et al. [8] presented a 2D mapping approach where local lane features are concatenated to a global representation. The concatenation bases on poses from a single survey drive. A continuous map update over the entire life span is not addressed. An iteratively addition of recent trajectories could keep these maps up to date.

A pioneering work on pose graphs was proposed from Lu and and Milios [10] where many horizontal range scans are aligned to each other in order to obtain a consistent world model. The pose graph was aligned on pose relations obtained from IMU measurements and pairwise ICP scan matchings. These matchings establishes loops closures and, therefore, improve the consistency of the world model significantly. In a broader sense our approach also close loops even though a particular loop was mapped at no time.

Olson et al. [6] presented a stochastic gradient descent method to optimize pose graphs. An advantage of this approach is the robustness against wrong initial guesses. Grisetti et al. [7] improved the convergence speed of this method significantly through a tree based parameterization.

Most of the related work concentrates on optimizing single trajectories. We found no currently existing work which explicitly align independent survey drives. Hence, our novel method provides an opportunity to an further improvement and to extend all these pose graph based approaches towards a life-long mapping by aligning several drives on different days and conditions.

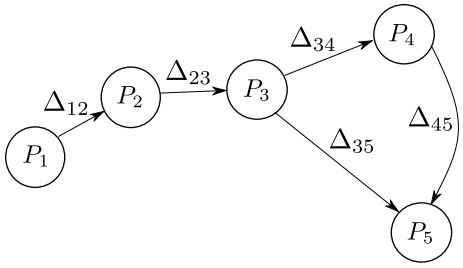

Fig. 2. Example pose graph with 5 poses (nodes) and and 5 pose differences (edges). A spanning tree is constructed e.g., by keeping all edges except of wether Δ_{34}, Δ_{35} or Δ_{45}.

III. COMPUTING THE POSE GRAPH

The first part (Section III-A) of this Section describes the general problem and explains how poses from several independent survey trajectories can be aligned to each other by a given set of estimates of pose differences. We assume that such a set is preexisting at this point since our algorithm abstracts from the underlying source of these estimates. However, we present one particular stereo vision based method to estimate pose differences in order to complete our process chain in the second part (Section III-B) of this section.

A. Pose Alignment

The input to our algorithm is a finite set of tuples $\left(\hat{\Delta}_{ij}, \Omega_{ij}\right)$ where $i \neq j$ and $i, j \in \{1, \ldots, N\}$. Thereby $\hat{\Delta}_{ij}$ denotes an estimate of the true pose difference

$$\Delta_{ij} = P_j \ominus P_i,$$

where $\ominus : SE(3) \times SE(3) \longrightarrow SE(3)$ denotes the pose difference operator which computes the difference from pose P_i to pose P_j. Furthermore, $\Omega_{ij} \in \mathbb{R}^{6\times 6}$ denotes a positive definite symmetric matrix which represents a weighting for the corresponding estimate $\hat{\Delta}_{ij}$. This input set induces a set of interconnected poses P_i.

The optimal set of pose deltas can be computed easily by integrating motion:

$$P_k^* = \sum_{i=1}^{k} \hat{\Delta}_{i,i+1} \oplus P_1, \;\; k \in \{2, \ldots, N\} \tag{1}$$

for an arbitrary start pose P_1 once every P_i is related to merely less than two estimates $\hat{\Delta}_{ij}$. The $\oplus$ denotes the

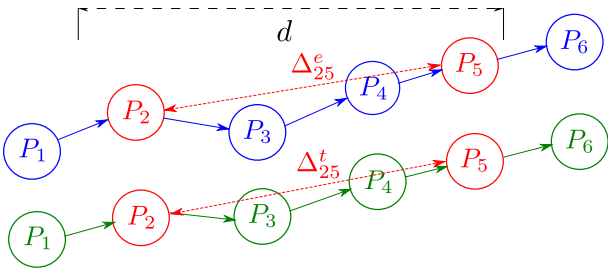

Fig. 3. Illustration of the drift error. The green poses denotes the ground truth whereas the blue ones denotes the estimated trajectory. Same indeces of the estimate and ground truth means related poses. The norm of the euclidean distance between pose 2 and 5 is assumed as approximately d in this example.

pose concatenation operator here. Equation (1) is known as odometry and forms the optimal set of poses in this case since the solution is unique. It is worth to mention that the weightings Ω_{ij} have no influence in equation (1) which means that worse estimates $\hat{\Delta}_{ij}$ with a low weighting have the same influence to the solution as good ones.

This is contrasted by the general case where the solution is overdetermined. Therefore, our approach uses a least squares optimization method to find an optimal solution. A 3D transformation is fully determined with 6 parameters which means that an unambiguously reversible depiction $\delta = \phi(P) \in \mathbb{R}^6$, $P \in SE(3)$ exists. Furthermore, the minimal representation of the identity pose $I \in SE(3)$ is $\phi(I) = \underline{0}$.

Therefore, given all pose difference estimates, the set of optimal poses P_i^* can be obtained through

$$\arg \min_{P_2,\dots,P_n} \sum_{ij}^{n} \phi\left(\Psi_{ij}\right)^T \Omega_{ij} \phi\left(\Psi_{ij}\right), \tag{2}$$

with $\Psi_{ij} = \left(P_i \oplus \hat{\Delta}_{ij}\right)^{-1} \oplus P_j$. This can be seen easily because for the true poses and differences holds

$$\begin{aligned}
&\left(P_i \oplus \Delta_{ij}\right)^{-1} \oplus P_j = \\
&\left(P_i \oplus \left(P_j \ominus P_i\right)\right)^{-1} \oplus P_j = \\
&P_j^{-1} \oplus P_j = I.
\end{aligned}$$

The pose P_1 is fixed to an arbitrary pose without the loss of generality.

Equation (2) constitutes a non-linear optimization problem due to the implicit 3D transformations and the conversion to a minimal representation. Therefore, the computation of the set of optimal poses is treated as a graph optimization problem, where every node represents a pose and every edge a pose difference estimate between two related nodes (see Fig. 2). This structure is called pose graph and enables the usage of common graph based optimization methods [11] to find a solution for problem (2). Furthermore, this representation reveals that a fully connected graph is assumed in order to obtain reasonable optimization results.

However, non-linear least squares methods require an appropriate initialization of the parameters to avoid local minima

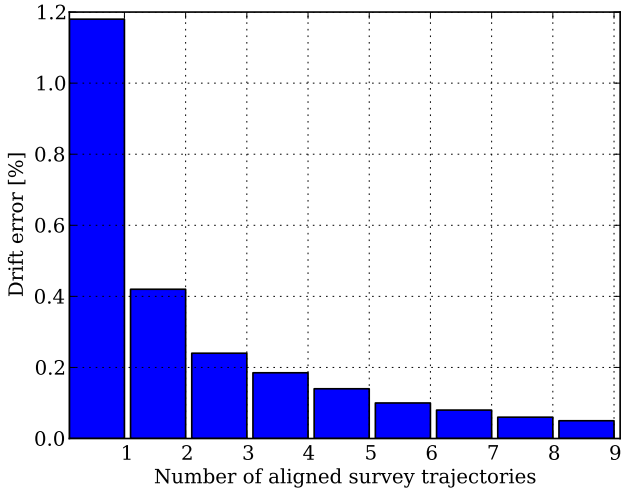

Fig. 4. Evolution of the mean drift error for an increasing number of aligned drives. Number of Monte Carlo iterations: 50, $d = 50$ meters, $\sigma = 1$ px, Probability of outlier occurrence: 0.5.

solutions. We compute an arbitrary spanning tree of the pose graph and integrate the poses over the corresponding deltas along all pathes of the tree for this purpose (see Fig. 2). Since a fully connected graph is assumed at least one spanning tree exists.

B. Vision based Pose Difference Estimation

Pose difference estimates can be obtained by several methods depending on the application and the environment. The estimation of ego-motion is part of many approaches and can be obtained by several sensors, e.g., cameras, GPS, IMU, or laserscanners. We obtain ego-motion from stereo visual odometry (SVO) [13], [14], [15] and treat it as pose difference estimates between two consecutive poses of a single trajectory within this work.

The first step of SVO is to compute circle matches [13] between the four images of two consecutive stereo image pairs. This requires a detection and description of salient points in all images followed by a descriptor comparison from which the final matches result. Thereafter two pixels of each match from the previous image pair are utilized to reconstruct a corresponding 3D landmark position. Finally, the 3D landmarks are re-mapped into the 2D space by the camera calibration with respect to a particular camera pose. The difference between the pixel positions of the matches in the current image pair and the pixel positions of the corresponding re-mapped landmarks is called back-projection error and is minimized by non-linear optimization methods which yields the pose of the camera rig at the time the current image pair was recorded.

In fact, a lot of workaround is required at every step of SVO to yield robust and accurate pose differences in practice [13]. The most basic assumption to SVO is that all images show the same scene. Otherwise no matches can be found. However, when the pose difference between the two consecutive camera rig poses is small enough this assumption is fulfilled implicitly.

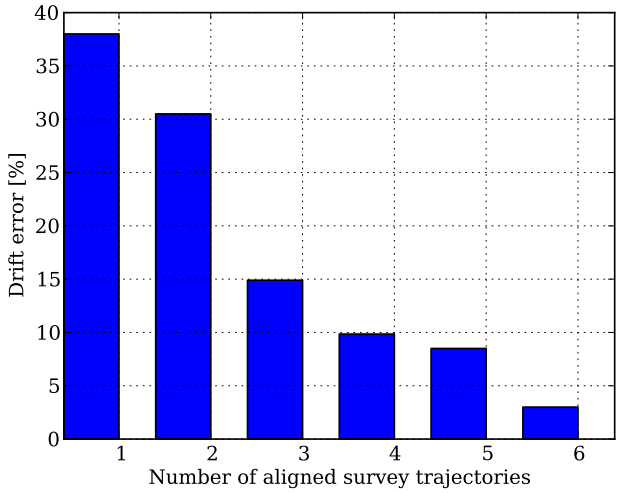

Fig. 5. Evolution of the mean drift error by an increasing number of aligned drives under bad simulation conditions. Number of simulation loops for each number of aligned drives: 50, Segment length $d = 50$ meters, Standard deviation of the Gaussian Pixel noise: 3 px, Probability of outlier occurrence: 0.7.

It remain the estimation of differences between poses of independent survey trajectories in order to align them. We also use SVO for this issue. This requires finding similar camera rig poses in the independent survey trajectories. This can be provided by a low cost GPS receiver or a loop closure detection [16]. SVO can be used straight forward once two corresponding image pairs of different drives are found.
Furthermore, covariance matrices for all pose difference estimates are obtained through a direct error propagation from prior defined covariances for all feature points. The inverse matrix of the propagated uncertainty is used as the weighting in equation (2).

IV. EXPERIMENTS

The experiments in the remainder of this Section mainly base on the visual odometry ego-motion estimation method. Some simulation experiments are presented in Section IV-A followed by a demonstration of the accuracy and robustness of our method in a real application in section in IV-B.

A. Simulation

The basic idea of the simulation is to generate distorted feature matches from a known ground truth. These simulated feature matches have the same structure as the matches obtained from the feature matching on real images. Thereafter pose differences based on these matches are estimated by the subsequent SVO processing steps and the attendant pose graph is aligned. This graph is compared against the known ground truth.
First of all, a ground truth is rendered along a particular reference trajectory. A fixed number of 3D landmarks are randomly placed along this trajectory. Furthermore, multiple survey trajectories are simulated by applying stochastic modifications to the poses of the reference trajectory. All trajectories and landmarks form the ground truth.

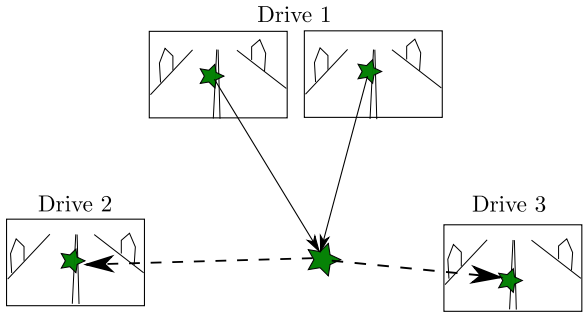

Fig. 6. Schema of the re-projection evaluation. Ground markings are detected and reconstructed in 3D using dense stereo information. The reconstructed landmarks are re-projected in the images of the other drives through the poses of the optimized pose graph.

Thereafter a virtual stereo camera pair is placed at all poses of all trajectories and the 3D landmarks within the aperture angle of the camera are projected. Occlusion effects are neglected at this point. The projected points represents the pixel of a circle match [13]. A Gaussian noise term is added to every pixel and, additionally, a uniformly-distributed random subset of all matches is heavily distorted to simulate outliers. Thereby, the number and distribution of the features in the virtual images is similar to our real experiments in urban scenarios.
The distorted virtual matches constitute the input for the remaining SVO computation steps from which the pose differences within the simulation are estimated. Thereafter the pose graph is initialized and aligned with our algorithm as described in Section III-A.
Finally. the quality of the estimated poses is evaluated against the known ground truth in terms of a trajectory drift error [17]. The drift error e between the ground truth and the estimated trajectory is computed by

$$e = \frac{1}{d \cdot N} \sum_{ij}^{N} \left|\left| \Delta_{ij}^{t} \ominus \Delta_{ij}^{e} \right|\right|, \tag{3}$$

where $||\cdot||$ denotes the norm of the Euclidean distance. Furthermore, Δ_{ij}^{t} denotes the pose difference between the i-th and j-th pose of the ground truth and Δ_{ij}^{e} the corresponding pose difference between the aligned poses of the pose graph. The i-th and j-th pose is thereby chosen so that the Euclidean distance between the poses is approximately d meters. The final drift error is computed by shifting a window of length d over the entire trajectory and averaging over all single errors which is illustrated in Fig. 3.
We perform this simulation in a Monte-Carlo scheme with different numbers of survey trajectories and varying simulation conditions. This means that a fixed number of survey trajectories and virtual matches are generated, the pose graph is aligned and evaluated against the ground truth. This step is repeated many times and the drift errors are averaged over all loops.
Fig. 4 shows the evolution of the drift error for an increasing number of aligned drives with zero-mean Gaussian noise with a standard deviation $\sigma = 1$ px and an outlier rate

Fig. 7. Comparison between our method (left column) and a high precision GPS measurement unit (right column). The lane markings are extracted from the first row images and reprojected into the images of the other drives (second and third row). The images show clearly the high accuracy of our estimated poses whereas the globally referenced GPS poses show high errors due multipath propagation.

of 50%. The length of the trajectories is approximately 1 km and the window length for the error evaluation is $d = 50$ m. The absolute drift percentage is quite small even with one drive because of the benign simulation conditions. However, the trend shows a converging decrease by aligning an increased number of drives.

Another simulation result is shown in Fig. 5 where the standard deviation of the pixel noise is $\sigma = 3$ px and the outlier rate is 70%. Furthermore, before the pose graph was aligned, several pose difference estimates from a randomly chosen subset of the aligned drives are post-distorted by a concatenation with a random noise pose whereas the weightings for these estimates remained unchanged. This results in high absolute drift errors. However, the drift can be decreased significantly whenever at least one relatively good survey trajectory is added. The result shows that a worse single trajectory estimate, e.g. because of bad weather or inappropriate lightning conditions can be improved clearly by an alignment with a better estimated trajectory.

B. Real World Experiments

First we present our experimental setup. Thereafter quantitative results of the accuracy and robustness of our method are shown.

Our vehicle is equipped with a stereo camera pair with a base width of 30 cm. The cameras field of view is approximately 80°. Image resolution is 1263×389 pixels after rectification. The image recording frequency is 10 Hz. Furthermore, the driven trajectory has a length of round 2 km passing through an inner city village and partly cross country. Every survey drive produces nearly 2000 stereo image pairs. Every image pair is related to one pose. We record several survey drives with our trial vehicle in order to evaluate our algorithm on real data.

Pose differences within one drive are estimated between two consecutive camera poses. Furthermore, we use a loop closure detection [16] to find similar images between independent drives. One camera pose is interconnected with at most one camera pose from another drive thereby. Pose difference estimates where the propagated uncertainty is higher than a upper border are leaved out at the final alignment.

For evaluation purposes, we detect edges of ground markings automatically in the images of one of the aligned drives. The detected pixels are reconstructed in 3D by using dense stereo information. These 3D reconstructed salient structures are then re-projected in the images from all other drives since we know the spatial relations of the cameras through our pose graph alignment and the camera calibration (see Fig. 6).

The left image in the first row of Fig. 7 is from the drive were the ground markings are extracted from. The left images in the second and the third row show the re-projected ground markings. The markings are projected exactly there where they expect to be. Along the entire trajectories nearly all reconstructed and re-projected markings are placed at their expected position which shows the high accuracy of the aligned pose graph.

Furthermore, we recorded high precision GPS data during all survey drives for a quantitative comparison. The same ground markings are re-projected with the recorded GPS poses which is possible since they are global referenced.

Again, the right image in the top row of Fig. 7 shows the

back-projected labels from the drive where they extracted from. The right images in the second and third row show the projection results (blue points) from the global referenced poses of our high precision GPS receiver. In contrast to the poses from the aligned pose graph, the re-projection with GPS poses show clearly poorer results which is observed over the entire trajectory. The resulting poses from our method outperforms the high precision GPS poses obviously. These results reveal that methods e.g, path planning which require a high re-localization accuracy quickly encounter their limits in urban scenarios when GPS is utilized. In contrast, our method shows precise and robust results in these areas and enables an opportunity for accurate and sustainable life-long mapping.

An unsatisfactory result is shown in Fig. 8. The reason for this is the surrounding environment which is poor in structure over a longer part of the trajectory. This is reflected in poor pose difference estimates since our applied SVO method requires structured area to work well. However, the poor results at this part of the trajectory are predictable by the propagated covariances. We observed a worsening of the covariance matrices in these areas.

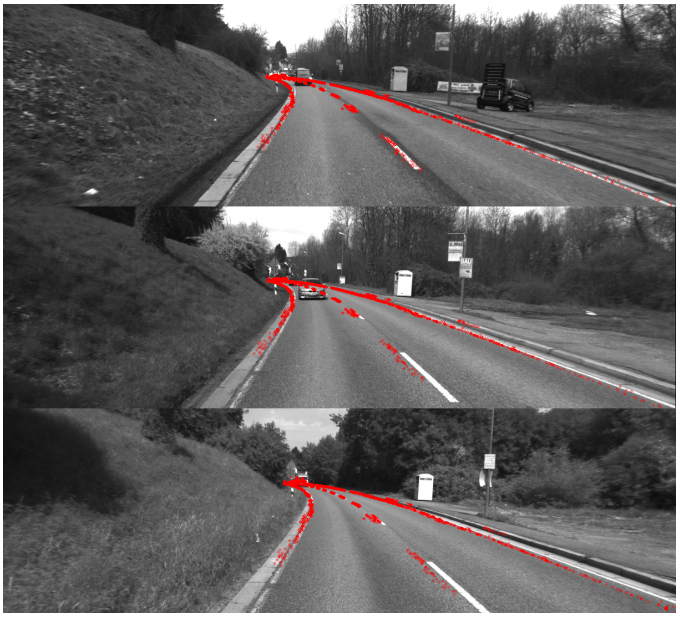

Fig. 8. Re-projection of extracted ground marking labels (red points) into different drives. The images show that the re-projection into the other drives is incorrect. The environment is poor in structure and hence, the pose difference estimates between the drives are worse.

V. CONCLUSIONS AND FUTURE WORK

Within this work, we presented a method to align several independent trajectories of the same area. A general method to optimize a set of related poses by a given set of pose differences were presented. Common non-linear graph based optimization methods were utilized for that purpose. In order to complete our process chain, a stereo vision method were used to estimate pose differences.

Furthermore, results from a simulation were presented which show the potential of drift reduction. This was extended by quantitative results from real vision based data which shows the accuracy and robustness of the method in urban areas. A comparison against high precision DGPS poses shows that our approach clearly outperforms this costly sensor in these areas.

The vision based approach reaches its limits whenever the trajectory passes long distances of rarely structured areas. Further improvements of the vision based frontend and combining several different sensors by our method open up possibilities for fruitful approaches which could overcome these flaws.

REFERENCES

[1] H. Lategahn, M. Schreiber, J. Ziegler, and C. Stiller, "Urban localization with camera and inertial measurement unit," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 719–724.

[2] J. Ziegler, H. Lategahn, M. Schreiber, C. G. Keller, C. Knoppel, J. Hipp, M. Haueis, and C. Stiller, "Video based localization for bertha," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*. IEEE, 2014, pp. 1231–1238.

[3] J. Levinson and S. Thrun, "Robust vehicle localization in urban environments using probabilistic maps," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 4372–4378.

[4] H. Badino, D. Huber, and T. Kanade, "Visual topometric localization," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE, 2011, pp. 794–799.

[5] A. Napier, G. Sibley, and P. Newman, "Real-time bounded-error pose estimation for road vehicles using vision," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE, 2010, pp. 1141–1146.

[6] E. Olson, J. Leonard, and S. Teller, "Fast iterative alignment of pose graphs with poor initial estimates," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 2262–2269.

[7] G. Grisetti, C. Stachniss, S. Grzonka, and W. Burgard, "A tree parameterization for efficiently computing maximum likelihood maps using gradient descent." in *Robotics: Science and Systems*, 2007.

[8] M. Schreiber, C. Knoppel, and U. Franke, "Laneloc: Lane marking based localization using highly accurate maps," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 449–454.

[9] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (slam): Part ii," *IEEE Robotics & Automation Magazine*, vol. 13, no. 3, pp. 108–117, 2006.

[10] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous robots*, vol. 4, no. 4, pp. 333–349, 1997.

[11] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g 2 o: A general framework for graph optimization," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 3607–3613.

[12] H. Lategahn and C. Stiller, "Vision-only localization," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 3, pp. 1246–1257, June 2014.

[13] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011.

[14] F. Fraundorfer and D. Scaramuzza, "Visual odometry: Part ii: Matching, robustness, optimization, and applications," *Robotics & Automation Magazine, IEEE*, vol. 19, no. 2, pp. 78–90, 2012.

[15] R. Hartley and A. Zisserman, *Multiple View Geometry*, 7th ed. Cambridge University Press, 2010.

[16] H. Lategahn, J. Beck, B. Kitt, and C. Stiller, "How to learn an illumination robust image feature for place recognition," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 285–291.

[17] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3354–3361.