

# PRECISION OF VISUAL LOCALIZATION USING DYNAMIC GROUND CONTROL POINTS

P. Trusheim,\* C. Heipke

Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany - (trusheim, heipke)@ipi.uni-hannover.de

Commission I, WG I/9

**KEY WORDS:** visual localization, dynamic ground control points, sequential bundle adjustment, simulation

## ABSTRACT:

Localization is one of the first steps in navigation. Especially due to the rapid development in automated driving, a precise and reliable localization becomes essential. In this paper, we report an investigation of the usage of dynamic ground control points (GCP) in visual localization in an automotive environment. Instead of having fixed positions, dynamic GCPs move together with the camera. As a measure of quality, we employ the precision of the bundle adjustment results. In our experiments, we simulate and investigate different realistic traffic scenarios. After investigating the role of tie points, we compare an approach using dynamic GCPs to an approach with static GCPs to answer the question how a comparable precision can be reached for visual localization. We show, that in our scenario, where two dynamic GCPs move together with a camera, similar results are indeed obtained to using a number of static GCPs distributed over the whole trajectory. In another experiment, we take a closer look at sliding window bundle adjustments. Sliding windows make it possible to work with an arbitrarily large number of images and to still obtain near real-time results. We investigate this approach in combination with dynamic GCPs and vary the no. of images per window.

## 1. INTRODUCTION

Localization is one of the basic tasks in navigation, a precise and reliable position is, e.g., a fundamental pre-condition for automated driving. Besides classical positioning sensors like global navigation satellite system (GNSS) receivers and inertial navigation systems (INS), more and more approaches also employ camera and/or laser scanner data to improve localization (e.g. (Garcia-Fernandez, Schön, 2019)). Both sensors can also be used in GNSS denied areas. In addition, cameras in particular have a relatively low cost and low weight, which can be a decisive advantage, for instance when deployed on unmanned aerial vehicles (UAV), although they need an external light source. Due to these reasons, researchers increasingly combine GNSS and INS sensors with image-based localization techniques. As an example, (Cavegn et al., 2016) shows the improvement of image-based georeferencing compared to direct georeferencing. The authors also demonstrate, that in challenging urban areas the uncertainty derived from only GNSS/INS can be too optimistic. Besides use in these challenging urban areas, image-based localization can also be employed in tunnels and for indoor navigation (e.g. (Cavegn et al., 2018)).

To solve the localization task state-space approaches based on Kalman or particle filters are often used. In photogrammetry and computer vision, network-based approaches using bundle adjustment are more common. (Colomina, Blázquez, 2004) compare both methods and point out the respective advantages and disadvantages: State-space approaches have a fixed and relatively small amount of state-parameters and are therefore typically chosen if real-time is a requirement. Traditional network approaches, on the other hand, suffer from a much larger size normal equation matrix. Also, they solve for all data in a simultaneous adjustment after data acquisition is finished. They thus achieve the most precise results, but real-time as such is not possible. Window-based sequential network methods, e.g.

\* Corresponding author

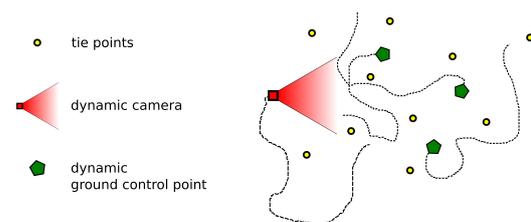


Figure 1. Basic scenario with one dynamic camera and dynamic GCPs moving between static tie points.

(Beder, Steffen, 2008), (Wilbers et al., 2019), offer a solution and can achieve near real-time results.

In this paper, we introduce a window-based sequential bundle adjustment for cooperative visual localization of moving cameras in an environment suited for autonomous driving. Following an idea of (Molina et al., 2017) along with static tie points, we introduce dynamic GCPs, i.e. GCPs which move in the scene. An example scenario is depicted in figure 1, showing one dynamic camera and multiple dynamic GCPs as well as a set of static tie points. Using dynamic GCPs brings several positive effects. On the one side it is more flexible because GCPs do not have to be placed on the trajectory in advance. In addition, in an automotive scenario if, for instance, a car coming from a GNSS denied area observes some other cars which accurately know and can communicate their position (i.e. act as dynamic GCPs), the first car can use them for self-localization.

Our main contribution is first to show, that based on simulations in a cooperative setting where vehicles pass on the information about their own position, the ego-motion can be calculated by using these vehicles as dynamic GCPs in a bundle adjustment with similar precision as achievable by using static GCPs. Second, we demonstrate how our approach can be extended to a sliding window variant. We determine the precision based on the variance-covariance matrix of the unknowns of the bundle

adjustment with simulated data. As we want to show if such an approach can be used in the automotive field, we investigate typical traffic scenarios.

The paper is organized as follows: After introducing related work in section 2, we show in section 3 the methods we are using. We first introduce our simulation scenario, which consists of dynamic cameras, dynamic GCPs and static tie points. We then present the functional and stochastic models we use, and also take a closer look at how we obtain the results, before introducing the sliding window approach. Section 4 contains our experiments. First, the simulation process is described and validated in a setting with regularly spaced tie points and no GCPs, followed by the comparison between static GCPs and dynamic GCP results. Finally, we report the investigations of the sliding window approach. In section 5 we recapitulate the results and discuss steps for future work.

## 2. RELATED WORK

Visual localization is a broad topic with many applications and lots of ongoing research. Three examples are (Wolcott, Eustice, 2014), (Cavegn et al., 2016) and (Cavegn et al., 2018). The authors investigate the use of information derived from images for georeferencing in environments, which are challenging for GNSS. (Wolcott, Eustice, 2014) use a monoscopic camera together with a 3D-map obtained by light detection and ranging (LIDAR). They compare the localization accuracy of the global positioning system (GPS), a mono camera and the LIDAR map-based localization, and show that both, the camera and the LIDAR approach yield better results than GPS. Although the camera is significantly cheaper than the LIDAR the obtained errors are in a similar order of magnitude. (Cavegn et al., 2016) deal with the challenge of georeferencing image sequences. As in urban canyons direct georeferencing is not suitable due to poor GNSS coverage, the authors add image-based georeferencing by using bundle adjustment. With this method, they are able to reduce the residuals at the checkpoints from approx. 40 cm to 4 cm. (Cavegn et al., 2018) use a multi-stereo system in their work and obtain georeferencing by combining simultaneous localization and mapping (SLAM) with highly redundant image sequences in a bundle adjustment. They test their work in urban environments with poor GNSS coverage and also indoors, achieving a root mean square error (RMSE) at checkpoints on the cm level. These results support the idea to use visual localization in autonomous driving for determining ego-motion. A difficulty here is, however, that the cited approaches need GCPs or current and accurate maps to work, which is not always available. One possible solution is to use a cooperative setting where some participants know their position and can therefore be used to determine the position of other participants.

In the work of (Stoven-Dubois et al., 2018), the authors introduce a UAV tandem system for surveying objects in GNSS denied areas. The surveying UAV flies next to the object to be captured and takes images while being tracked in an image and georeferenced by another UAV that flies at a higher altitude with a good GNSS signal. Also MapKITE (Molina et al., 2017) and (Nahon et al., 2019) use a tandem system. Here, the authors combine a terrestrial mobile mapping van with a UAV, so they can make use of both types of measurements. The van has a much higher payload which can be used for heavier but more accurate GNSS sensors. Therefore, the vehicle can be used as

a dynamic ground control point. For accurate automatic positioning, a circular target is placed on the vehicle roof. This cooperation significantly reduces the effort for having to place multiple static GCPs in the scene. We use this idea for localization in a network of cooperating vehicles.

While classical network approaches in photogrammetry typically assume a static environment, approaches using dynamic GCPs like (Molina et al., 2017) and (Nahon et al., 2019) have to introduce time-varying parameters for the object scene. (Colomina, Blázquez, 2004) describes a model that can handle time-dependent parameters including, for example, the trajectory and the sensor orientation. To compute these parameters, they compare a state space and a network approach and point out the respective advantages and disadvantages. Inspired by this work we decided to use a bundle adjustment with dynamic GCPs. As we do not assume the observation of the GCP positions to be synchronized to the image data capture, we interpolate the epoch of the latter from the former. Similar problems are discussed in (Cucci et al., 2017a) and (Cucci et al., 2017b) regarding raw observations from inertial measurement units (IMUs) in dynamic networks.

One issue of using bundle adjustment is the computing time, in particular for larger blocks. This is especially a problem in traffic situations where results need to be available in real-time. For image sequences, by using a window with a fixed number of images, the size of the equation system can be bounded so the computing time is bounded as well, see (Beder, Steffen, 2008) and (Wilbers et al., 2019). Along those lines (Beder, Steffen, 2008) introduce a sequential bundle adjustment approach with recursive estimation for speeding up the computation.

In our approach, we combine these ideas and transfer them to an automotive setting: we develop an incremental bundle adjustment using dynamic GCPs for localization in realistic traffic scenarios.

## 3. PRECISION DETERMINATION BY BUNDLE ADJUSTMENT

In the following, we describe the functional and the stochastic model we use for introducing dynamic GCPs into bundle adjustment. For the sake of completeness, we also describe how we obtain the precision of the unknowns. Finally, we present the sliding window approach we use.

### 3.1 Functional and stochastic model

We discriminate three different types of objects.

1. Dynamic cameras: The cameras we use are moving in the scene. Therefore, the parameters of the exterior orientation are functions of time.
2. Dynamic GCPs: These GCPs are points moving in the depicted scene. They can measure their 3D position in the global coordinate system. The time of the measurement is not necessarily synchronized with that of the image data capture, therefore the latter have to be interpolated.
3. Static tie points: Tie points have a stable position in space and time. Obviously, we could also use dynamic tie points, but in this paper, we restrict ourselves to the static variant.

Figure 1 shows such a scenario with one dynamic camera and three dynamic GCPs as well as a set of tie points.

As GCP coordinates are not necessarily measured at the same time as image data capture occurs, we map the object coordinates to the camera frame according to equation 1, where  ${}^{Cam}\mathbf{X}$  are the 3D camera frame coordinates,  $R$  and  $\mathbf{X}_0$  represent the exterior orientation of the camera and  $\mathbf{X}_{GCP}$  the GCP coordinates.  $t$  is the time at which those coordinates were determined, and  $f$  is an appropriate interpolation function.

$${}^{Cam}\mathbf{X} = R^T \cdot (f(\mathbf{X}_{GCP}, t) - \mathbf{X}_0) \quad (1)$$

In our set-up the optical axis is horizontal and the camera looks in driving direction, which is the X-axis of our camera frame. As a consequence the image plane is a vertical plane. The Z-axis of the camera frame is thus chosen to be vertical as well, and the Y-axis completes the system to be right-handed, leading to equation 2 for the relation between image and camera frame coordinates ( $x_l$  and  $y_l$  are the image coordinates,  $c$ ,  $x_0$  and  $y_0$  represent the elements of interior orientation of the camera).

$$\begin{aligned} x_l &= x_0 - c \cdot \frac{{}^{Cam}Y}{{}^{Cam}X} \\ y_l &= y_0 - c \cdot \frac{{}^{Cam}Z}{{}^{Cam}X} \end{aligned} \quad (2)$$

We assume that the position of the dynamic GCPs is updated relatively frequently so we use linear interpolation for  $f(\mathbf{X}_{GCP}, t)$  considering the two closest positions of the observed point concerning the time the image is taken (equation 3).

$$f(\mathbf{X}_{GCP}, t) = \mathbf{X}_{GCP,t_0} \cdot \frac{t_1 - t}{t_1 - t_0} + \mathbf{X}_{GCP,t_1} \cdot \frac{t - t_0}{t_1 - t_0} \quad (3)$$

Here  $t_n$  is the time when the position of the GCP  $\mathbf{X}_{t_n}$  was observed,  $t$  describes the time at which the image was taken.

As in our model the position of the tie points is independent of time we can use equation 4 instead of equation 1 to transform the tie points into the camera frame, where  $\mathbf{X}$  is the 3D position of the tie point in the global frame.

$${}^{Cam}\mathbf{X} = R^T \cdot (\mathbf{X} - \mathbf{X}_0) \quad (4)$$

The 3D position of the dynamic GCP is observed in the global frame at the time  $t_n$ , giving rise to a direct observation  $\mathbf{X}_{(GCP,t_n),l}$  for the unknown global position  $\mathbf{X}_{GCP,t_n}$  to be able to introduce uncertainty for the GCP position.

$$\mathbf{X}_{(GCP,t_n),l} = \mathbf{X}_{GCP,t_n} \quad (5)$$

Finally, in some experiments we also use direct observations for the image orientation which are introduced in a similar way as shown in equation 5.

Besides the functional model, we also need a stochastic model of the observations (equation 6),

$$\Sigma_{ll} = \sigma_0 \cdot \mathbf{Q}_{ll} \quad (6)$$

where  $\Sigma_{ll}$  represents the covariance matrix of the observations,  $\mathbf{Q}_{ll}$  the cofactor matrix of the observations and  $\sigma_0$  the variance factor. In our work, we assume uncorrelated observations for all

observations. As all groups of observations are introduced with their corresponding standard deviation we choose  $\sigma_0 = 1$ .

### 3.2 Precision

The precision of visual localization is obtained by standard variance propagation. According to the standard formulae of least squares adjustment the cofactor matrix of the unknowns  $Q_{xx}$  is determined as follows.

$$A = \left. \frac{\delta f}{\delta x} \right|_{x_0} \quad (7)$$

$$P = Q_{ll}^{-1} \quad (8)$$

$$N = A^T P A \quad (9)$$

$$Q_{xx} = N^{-1} \quad (10)$$

Equation 10 shows that  $Q_{xx}$  depends on one hand on the structure of the system given by the Jacobi matrix of the functional model  $A$  and on the other hand on the stochastic model of the observations  $Q_{ll}$ . The diagonal of  $Q_{xx}$  contains the variances of the unknowns. To obtain the precision of the exterior orientation of the camera for every image we use the two corresponding  $3 \times 3$  sub-matrices of  $Q_{xx}$  regarding the position and the rotation angles of the projection centre. As the unknowns are defined in the global frame we use the rotation matrix of the exterior orientation  $R$  to transform the cofactor sub-matrices into the camera frame. The diagonal elements of  $Q_{xx}$  then contain the variances of the unknowns regarding the driving direction and the direction of the horizontal Y and vertical Z axis of the image plane. We use the square root of these elements, i.e. the standard deviations, as the measure of precision.

### 3.3 Sliding window

To achieve near real-time behaviour when using bundle adjustment, sequential window-based approaches are often used (Beder, Steffen, 2008). In this paper, we make use of such sliding windows. A window contains a certain number of images (size)  $W_{sz}$  and overlaps with the next window by a certain number of images  $W_{ol}$ . The bundle adjustment then uses all images in one window, whereby the six exterior orientation parameters of all  $W_{ol}$  images overlapping with the previous window (and thus having been computed in the previous adjustment) are used as additional direct observations. In the stochastic model, the entries of the  $Q_{xx}$  matrix of the predecessor window are used in  $Q_{ll}$  to describe the variance and covariance of these direct observations.

## 4. EXPERIMENTS

In our experiments, we first discuss the influence of the tie points on the stability of the image block. Then, we compare the precision of the camera exterior orientation using dynamic GCPs to the case with static GCPs. Finally, we demonstrate the effects of using sliding windows with a dynamic GCP setup. All investigations are based on simulations.

### 4.1 Simulation environment

In our simulation setup, the camera moves on a pre-defined trajectory through the scene with constant velocity and with a horizontal viewing direction parallel to the driving direction. Static tie points are placed throughout the whole scene. The distribution of the static and dynamic GCPs differs in the different

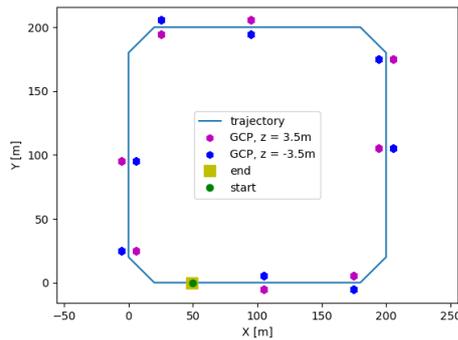


Figure 2. The camera trajectory and the placement of the GCPs for testing the static setup.

experiments. Images are taken at regular intervals. For every image exposure, the time is recorded and the global 3D coordinates of all GCPs and all tie points are mapped into image space. The 3D positions of the dynamic GCPs are recorded at different time steps, which leads to the time offset we consider in equation 3.

The camera we use has a resolution of  $1936 \times 1216$  px with a pixel size of  $5.86 \mu\text{m} \times 5.86 \mu\text{m}$  and a focal length of 5 mm. These are typical values for a camera used in the automotive field, so our simulation can be used in such an application.

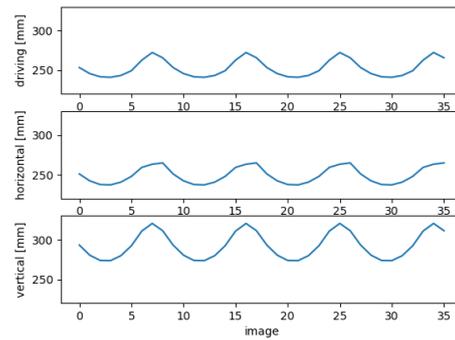
In the stochastic model we use  $\sigma_X = \sigma_Y = 0.1$  m and  $\sigma_Z = 0.2$  m for the dynamic GCP coordinate observations and  $\sigma_X = \sigma_Y = 0.05$  m and  $\sigma_Z = 0.1$  m for static GCP observations to take into account that static GCPs can typically be measured with a higher accuracy, and that GCP coordinates (typically measured with GNSS/IMU these days) in general have a higher uncertainty in Z direction. The standard deviation of the image coordinate observations is set to  $\sigma_x = \sigma_y = 0.5$  px.

#### 4.2 Scenario

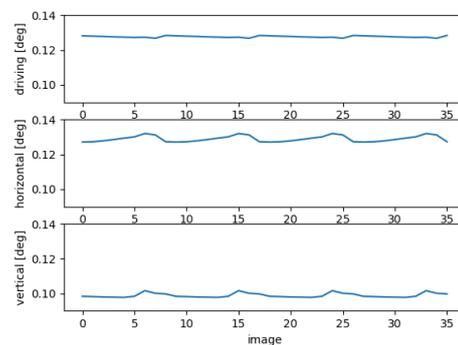
For our experiments, we selected a symmetric trajectory, so the interpretation of the results becomes more meaningful (figure 2). The shape of the trajectory is square with corners cut in 45-degree angles. The length of the trajectory is a little more than 700 m. Except for the experiments using window-based bundle adjustment, images are taken every 20 m, resulting in 35 images altogether. The viewing distance for the camera is restricted to 100 m to reduce effects caused by the camera seeing more tie points at the beginning of a long straight part of the trajectory than towards the end. For simplicity, the height of the projection centre is set to 0 m for the whole trajectory under the assumption of a flat environment. In the first experiments the tie points are arranged in a regular 3D grid with a grid size of 20 m, and all tie points lying in the viewing cone of the camera are assumed to be visible in the images. In further experiments the tie points are randomly placed within each 3D grid mesh, and for the window-based approach a grid size of 15 m is used.

#### 4.3 Block stability without GCPs

For all experiments we made sure to have enough tie points, adequately distributed in image space. Despite the fact that moving and viewing direction are parallel, we obtain a somewhat stable photogrammetric block in this way. To demonstrate this stability, we first investigate the precision of exterior orientation



(a) Precision of the position of the projection centre in the vertical and horizontal axes of the image and in driving direction.



(b) Precision of the angles around the vertical and horizontal axes of the image and in driving direction.

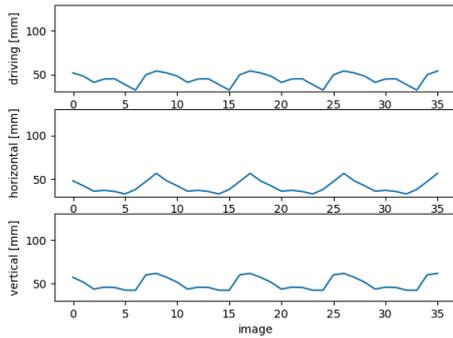
Figure 3. Results of visual localization with regularly spaced tie points, but without GCPs.

without GCPs. The block datum is defined by direct observations of the camera projection centres with  $\sigma_{X_0} = \sigma_{Y_0} = \sigma_{Z_0} = 1$  m, which are realistic values for the kind of GNSS receivers used in the automotive field. The results can be seen in figure 3, where the precision of the exterior orientation is depicted as a function of image number (or time).

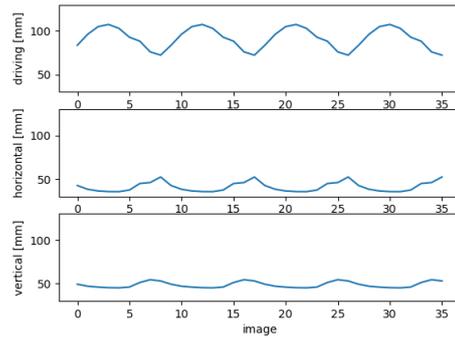
Due to the symmetric course, also the obtained precisions show a symmetric behaviour. The precision of the projection centre improves by about a factor of 3 compared to the precision of the direct observations, which demonstrates that the bundle based on the tie points is indeed relatively stable. It is also visible, that in the curves the precision decreases, although not by a large amount. This effect is explained by the fact that images in these positions of the trajectory are connected to fewer images than those along the straight lines. The angular precisions show a similar behaviour. When comparing the different directions it can be seen, that the precision of the vertical coordinate is lower than that of the horizontal coordinate perpendicular to the driving direction, which is a consequence of the rectangular image format. The precision of the angles around the horizontal and the vertical axis again show a similar behaviour.

#### 4.4 Comparison between static and dynamic GCPs

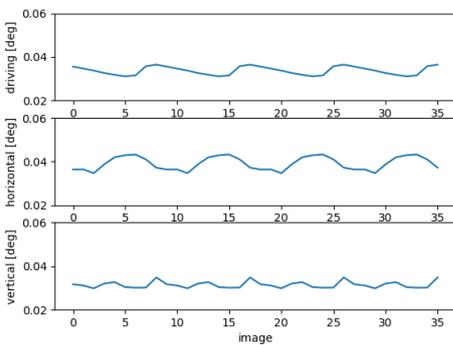
In the next step, we investigate the question, under which conditions dynamic GCPs can reach similar precision as static GCPs in a realistic traffic scenario. We choose a convoy formation



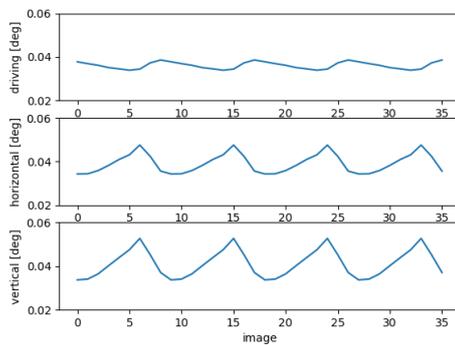
(a) Precision of the position of the projection centre in the vertical and horizontal axes of the image and in driving direction.



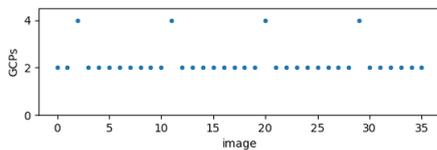
(a) Precision of the position of the projection centre in the vertical and horizontal axes of the image and in driving direction.



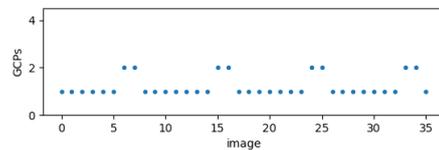
(b) Precision of the angles around the vertical and horizontal axes of the image and in driving direction.



(b) Precision of the angles around the vertical and horizontal axes of the image and in driving direction.



(c) Number of seen GCPs in every image.



(c) Number of seen GCPs in every image.

Figure 4. Results of visual localization with regularly spaced tie points and static GCPs.

Figure 5. Results of visual localization with regularly spaced tie points and dynamic GCPs.

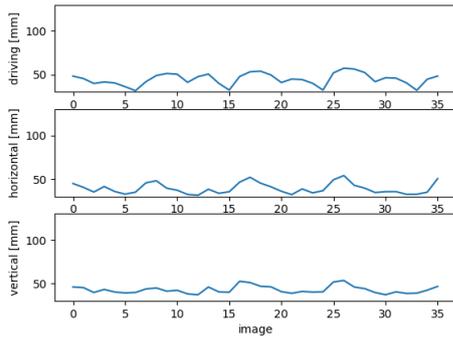
where the camera follows two dynamic GCPs. Convoy situations are typical for traffic and enable the possibility to use the same dynamic GCPs over a long distance. In our case, we use two GCPs in front of the camera. The first GCP drives 15 m in front of the camera and the second one 30 m, which corresponds to the recommended distance between vehicles at a speed of 50 km/h. The camera as well as the two dynamic GCPs have a height of 0 m. As a consequence, the camera sees the 2nd GCP only in the curves, otherwise, it is occluded. In addition, on the straight parts of the trajectory, the dynamic GCP is depicted in the centre of the camera image.

We placed the GCPs in tuples all around the trajectory (see figure 2), so the whole path can be consistently connected to the global coordinate system. As we want to have GCPs in most images to further increase the block stability, we placed two GCPs in the middle and two GCPs at the end of every straight part of the trajectory. Heights of 3.5 or -3.5 m guarantee that the image coordinates move towards the respective image corners, when the camera approaches the GCP tuples. By alternating the heights the GCP position in the images is evenly distributed

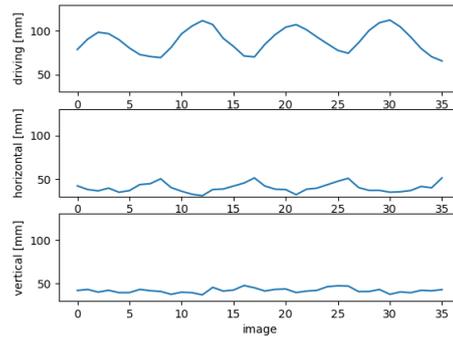
among all four corners.

First, we analyse the influence of the static GCPs (figure 4) on the precision by comparing the results to those without GCPs. The static GCPs improve the precision significantly: the position of the projection centre is improved by a factor of 5, those of the angles by approximately a factor of 3. When inspecting the projection centre results it can be seen, that - not surprisingly - when four GCPs per image are visible instead of two, the precision improves. This effect adds two local minima per straight trajectory part to the plots, as two tuples per straight part are used. In the precision of the angles, we obtain the same effect around the vertical and horizontal axis, although the angle around the horizontal axis is not so strongly influenced. The angle around the driving axis is improved using the GCPs, but the variations are very similar to those without GCPs.

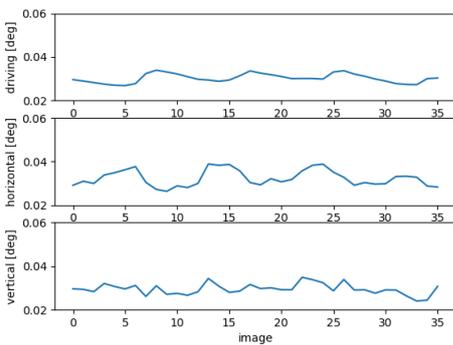
When investigating the precision reached with the dynamic GCPs especially the driving direction is noticeable (figure 5). Here we see an opposite behaviour as in the two previous cases. The precision decreases while driving on the straight line and



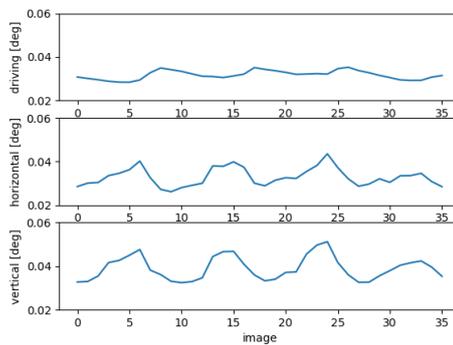
(a) Precision of the position of the projection centre in the vertical and horizontal axes of the image and in driving direction.



(a) Precision of the position of the projection centre in the vertical and horizontal axes of the image and in driving direction.



(b) Precision of the angles around the vertical and horizontal axes of the image and in driving direction.



(b) Precision of the angles around the vertical and horizontal axes of the image and in driving direction.

Figure 6. Results of visual localization by using static GCPs in a random tie point cloud.

Figure 7. Results of visual localization by using dynamic GCPs in a random tie point cloud.

risers in the curve. This effect is caused by the convoy configuration, as on the straight line only one GCP is visible and is depicted in the image centre. If either the 2nd GCP or the camera turn around the corner, two GCPs become visible and the precision improves. The horizontal and the vertical position on the other hand are dominated by the block stability provided by the tie points, because, except the overall level, the plots of figures 4a and 5a are rather similar to figure 3a; the same is true for the angles.

The results reported so far were obtained using a regular 3D tie point grid, for which the results are easier to interpret, but which is less realistic. Next, we present results obtained with tie points randomly placed within this grid. The point density is one tie point per cube with edge length 20 m. By using randomly scattered tie points (see figures 6 and 7) the symmetry of the previous plots is of course somewhat disturbed, but the main patterns can be still seen (compare figures 4 vs. 6 and 5 vs. 7). This influence can be especially seen in the angles. All in all, the results show that by a comparison between static and dynamic GCPs mainly the precision in driving direction differs, which is due to the selected convoy scenario and the resulting occlusions. In the other parameters, a similar level of precision was reached, and the higher precision of the static GCP coordinates is compensated by the larger amount of dynamic GCPs measurements.

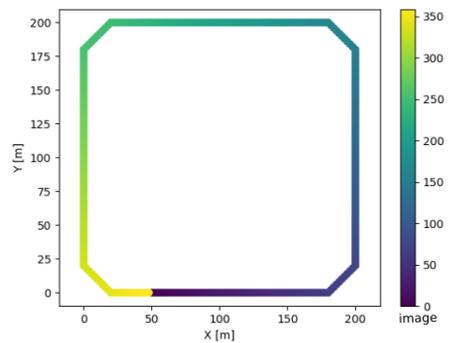
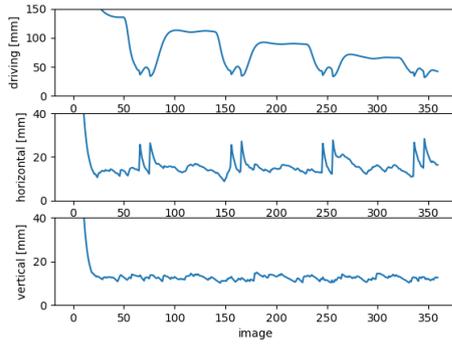


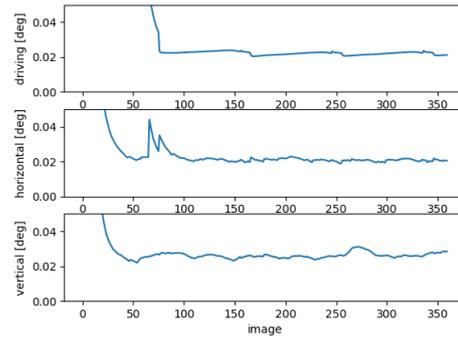
Figure 8. Distribution of images for the window-based approach. The image number is color coded.

#### 4.5 Sliding window in combination with dynamic GCPs

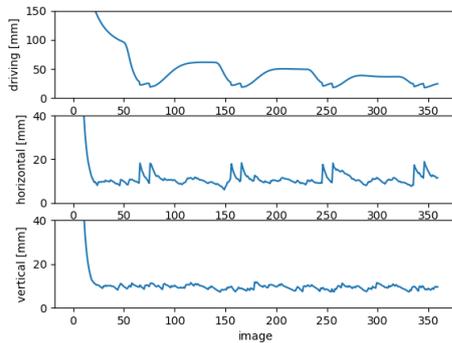
In the last experiment, we investigate the influence of a sliding window on the precision of the results when using dynamic GCPs. As mentioned before, the background is that sliding windows allow for near real-time results, which are important in the automotive field. In general, the sliding window has two independent parameters, namely window size and overlap. As we want to obtain results of each image as fast as possible, we select an overlap of  $(W_{sz} - 1)$  images, where  $W_{sz}$  is the window size. The choice of  $W_{sz}$  itself is a compromise between



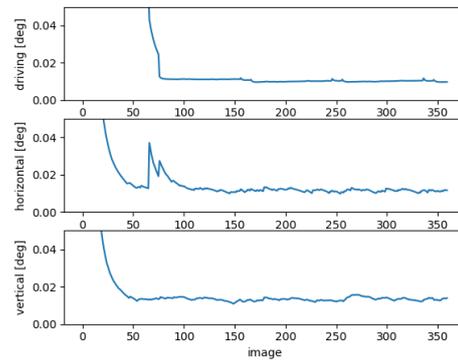
(a)  $W_{sz} = 5$ .



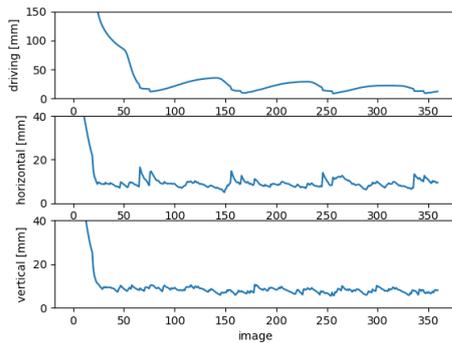
(a)  $W_{sz} = 5$ .



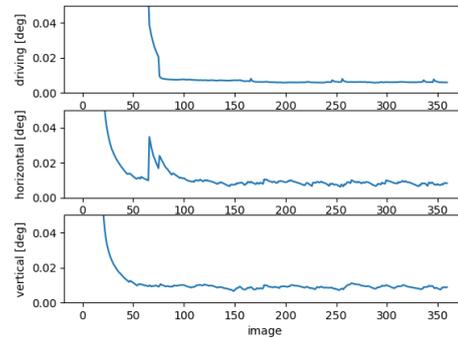
(b)  $W_{sz} = 10$ .



(b)  $W_{sz} = 10$ .



(c)  $W_{sz} = 20$ .



(c)  $W_{sz} = 20$ .

Figure 9. Precision of the position in the vertical and the horizontal image axes and the driving direction with  $W_{sz}$  10, 20 and 30. Note that the scale of the driving direction is adapted to better visualize the results.

Figure 10. Precision of the angle around the vertical and the horizontal image axes and the driving direction at  $W_{sz}$  10, 20 and 30.

the necessary precision, which is better for larger windows, and the required computational speed, as smaller windows can be processed faster.

A problem occurs when using dynamic GCPs in combination with the sliding window in our scenario: for small windows only part of the straight course is included, leading to only one visible GCP, located in the image centre, which results in numerically instable solutions. Therefore, in this experiment we move the second dynamic GCPs to a second lane to the left of the original trajectory, where it is not occluded any longer. Thus, we now simulate two-lane traffic rather than a convoy as before.

In our experiment we compare three different window sizes, namely  $W_{sz} = [5; 10; 20]$ . In order to have a denser dataset, we rise the rate of image capture to one image every 2 meters (figure 8) which is equivalent to 7 fps at our assumed speed of 50 km/h. This rate can easily be reached by typical automotive cameras. We use randomly scattered tie points with a point density of 1 tie point per cube with a grid size of 15 m.

The results (see figures 9 and 10) show that when using a sliding window with dynamic GCPs the precision is rather poor in the first part of the first straight line of the trajectory, even though the two-lane traffic scenario is used, in particular for the position in driving direction and for the angles. Afterwards, similar effects over the trajectory are obtained as by using the

full bundle (compare to figure 7). The four corners have a clear effect and, again, the precision of the position in driving direction is the lowest (note the different scale of the axis in figure 9). Also, as in the previous case the precision improves in the curves.

Regarding the window size, it can be seen, that as expected larger windows produce better precision. For the other two coordinates of the projection centre, the precision is similar to that for the full bundle (which, however, has a different number of tie points and of images), in the horizontal direction the curves are well visible. Due to the higher rate of image capture the smaller 45-degree parts of the trajectory are also visible. The precision of the angles achieve expected results. Stable values are reached after the first corner, then the values are more or less constant. Again, larger windows lead to better results. In summary, our results show that after a starting phase the precision obtained with the sliding window approach reaches a stable level. It can also be seen that there is a trade-off between the reachable level of precision and the needed processing time represented by the window size. Based on our results the sliding window approach seems to offer a good opportunity to reach near real-time behaviour for the bundle adjustment with dynamic GCPs.

## 5. CONCLUSION

Based on simulations we showed in this paper, to which extent dynamic GCPs can be used for visual localization in realistic traffic scenarios using the obtained precision for the elements of exterior orientation as criterion. We also presented the precisions obtained for a sliding window approach. While in case of a convoy formation problems can occur for the driving direction, the situation is improved for two-lane traffic.

In further research, we will study the effect of introducing additional GNSS and IMU measurements for the elements of exterior orientation. On the more methodological side we will study ways to deal with points at infinity (see (Förstner, Wrobel, 2016) for possible ways to do so). Another topic is the introduction of dynamic tie points and the question of how to then differentiate between the two types. Also, a combination of multiple cameras is a topic of interest for us. Finally, our results should be verified with real data.

## ACKNOWLEDGEMENTS

This work was supported by the German Research Foundation (DFG) as a part of the Research Training Group i.c.sens [GRK2159].

## REFERENCES

Beder, C., Steffen, R., 2008. Incremental estimation without specifying a-priori covariance matrices for the novel parameters. *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE.

Cavegn, S., Blaser, S., Nebiker, S., Haala, N., 2018. robust and accurate image-based georeferencing exploiting relative orientation constraints. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2, 57–64.

Cavegn, S., Nebiker, S., Haala, N., 2016. A systematic comparison of direct and image-based georeferencing in challenging urban areas. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B1, 529–536.

Colomina, I., Blázquez, M., 2004. A unified approach to static and dynamic modeling in photogrammetry and remote sensing. *Proceedings of the XXth ISPRS Congress, Istanbul, Turkey*, Citeseer, 178–183.

Cucci, D., Clausen, P., Skaloud, J., Matteucci, M., 2017a. A general approach to time-varying parameters in pose-graph optimization. *2017 European Navigation Conference (ENC)*, IEEE.

Cucci, D., Rehak, M., Skaloud, J., 2017b. Bundle adjustment with raw inertial observations in UAV applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 1–12.

Förstner, W., Wrobel, B. P., 2016. *Photogrammetric Computer Vision*. Springer International Publishing.

Garcia-Fernandez, N., Schön, S., 2019. Optimizing Sensor Combinations and Processing Parameters in Dynamic Sensor Networks. *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*, 10(2), 2048–2062.

Molina, P., Blázquez, M., Cucci, D., Colomina, I., 2017. First Results of a Tandem Terrestrial-Unmanned Aerial mapKITE System with Kinematic Ground Control Points for Corridor Mapping. *Remote Sensing*, 9(1), 60.

Nahon, A., Molina, P., Blázquez, M., Simeon, J., Capo, S., Ferrero, C., 2019. Corridor Mapping of Sandy Coastal Fore-dunes with UAS Photogrammetry and Mobile Laser Scanning. *Remote Sensing*, 11(11), 1352.

Stoven-Dubois, A., Jospin, L., Cucci, D., 2018. Cooperative navigation for an uav tandem in gnss denied environments. *Proceedings of the 31st International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2018)*, Miami, FL, USA, 24–28.

Wilbers, D., Merfels, C., Stachniss, C., 2019. A comparison of particle filter and graph-based optimization for localization with landmarks in automated vehicles. *2019 Third IEEE International Conference on Robotic Computing (IRC)*, IEEE.

Wolcott, R. W., Eustice, R. M., 2014. Visual localization within LIDAR maps for automated urban driving. *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE.