COOPERATIVE LOCALISATION USING IMAGE SENSORS IN A DYNAMIC TRAFFIC SCENARIO

P. Trusheim*, Y. Chen, F. Rottensteiner, C. Heipke

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

Commission I, WG I/9

KEY WORDS: Visual Localisation, Dynamic Ground Control Points, Bundle Adjustment, Cooperative Localisation

ABSTRACT:

Localisation is one of the key elements in navigation. Especially due to the development in automated driving, precise and reliable localisation becomes essential. In this paper, we report on different cooperation approaches in visual localisation with two vehicles driving in a convoy formation. Each vehicle is equipped with a multi-sensor platform consisting of front-facing stereo cameras and a global navigation satellite system (GNSS) receiver. In the first approach, the GNSS signals are used as excentric observations for the projection centres of the cameras in a bundle adjustment, whereas the second approach uses markers on the front vehicle as dynamic ground control points (GCPs). As the platforms are moving and data acquisition is not synchronised, we use time dependent platform poses. These time dependent poses are represented by trajectories consisting of multiple 6 Degree of Freedom (DoF) anchor points between which linear interpolation takes place. In order to investigate the developed approach experimentally, in particular the potential of dynamic GCPs, we captured data using two platforms driving on a public road at normal speed. As a baseline, we determine the localisation parameters of one platform using only data of that platform. We then compute a solution based on image and GNSS data from both platforms. In a third scenario, the front platform is used as a dynamic GCP which can be related to the trailing platform by markers observed in the images acquired by the latter. We show that both cooperative approaches lead to significant improvements in the precision of the poses of the anchor points after bundle adjustment compared to the baseline. The improvement achieved due to the inclusion of dynamic GCPs is somewhat smaller than the one due to relating the platforms by tie points. Finally, we show that for an individual vehicle, the use of dynamic GCPs can compensate for the lack of GNSS data.

1. INTRODUCTION

In the field of automated driving, precise and reliable self-localisation is a fundamental pre-condition. Especially in densely built-up urban areas classical positioning sensors like global navigation satellite system (GNSS) receivers reach their limits, as occlusions and multipath effects can lead to systematic errors. In these areas, in particular, additional sensors such as laser scanners or cameras are used to improve the localisation, e.g. (Garcia-Fernandez and Schön, 2019). Although cameras require an external light source, they are lighter and cheaper than laser scanners, making them more flexible in use. Examples of the usage of cameras in challenging urban areas are (Cavegn et al., 2016) and (Cavegn, 2020), where the authors show an improvement of the accuracy of the object points by a factor of 10 when comparing image-based georeferencing using ground control points (GCPs) with direct georeferencing using GNSS/INS. They also demonstrate that the uncertainty of the localisation derived from GNSS/INS only can be too optimistic.

While typical visual localisation methods, for example visual simultaneous localisation and mapping (SLAM), use a single platform, sharing information among multiple platforms can lead to better reliability of the solution. Therefore, applications for cooperative visual SLAM with multiple agents have recently been developed (Zou et al., 2019). Due to the emergence of car-to-car and car-to-X communication, cooperation is also interesting for automated driving. Cooperation can be archived, for example, by sharing position, in which context

the cooperating vehicle can be used as a dynamic GCP (Molina et al., 2017), (Stoven-Dubois et al., 2018), (Trusheim and Heipke, 2020) or by sharing images or features derived from images (Zou and Tan, 2013). Especially in (Molina et al., 2017) and (Stoven-Dubois et al., 2018), where one of the participating platforms has worse possibilities for self-localisation, the additional information leads to better results.

In this paper, we investigate the potential improvement of localisation using image sensors in typical traffic scenarios due to cooperation. The paper extends our prior work (Trusheim and Heipke, 2020), in which only simulations were investigated, by presenting results achieved using real data. For this purpose, we recorded data in a typical traffic situation. Two vehicles equipped with multi-sensor platforms drive in a convoy formation through a road with normal traffic volume. Markers on the vehicles allow to identify them in the images (figure 1). Both multi-sensor platforms are equipped with a GNSS receiver and a stereo camera pair. Images are captured with GNSS timestamps. Therefore, all sensor data are available in a common time frame. The pose of all sensors and the markers relative to the platform is known from prior calibration.

Our main contribution is the demonstration of the advantages of cooperative visual localisation in urban traffic scenarios using bundle adjustment. For this purpose we qualitatively investigate the trajectory precision of the corresponding vehicles, using (a) common tie points and (b) so called dynamic GCPs.

The paper is structured as follows: After the introduction, we give a short overview of existing work in section 2. In section 3 we describe the employed bundle adjustment approach with
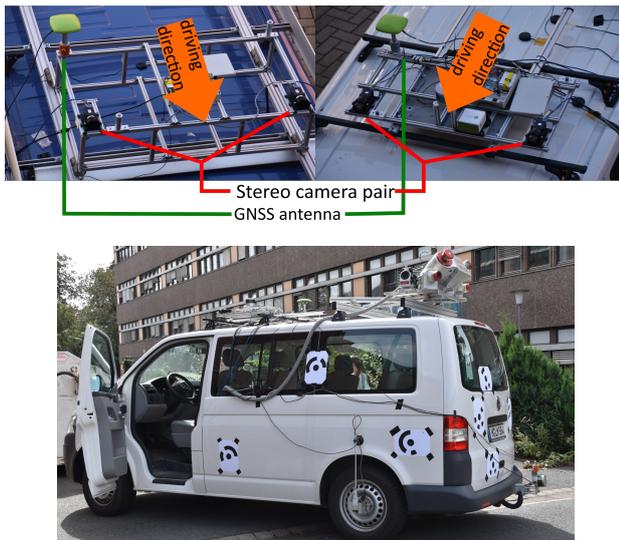
---
* Corresponding author

Figure 1. Top: multi-sensor platforms used for this paper. Left: platform of the trailing vehicle. Right: platform of the front vehicle. Both platforms are equipped with a GNSS receiver and a stereo camera pair looking into the direction of travel. Bottom: Front vehicle with attached markers.

a detailed description of the used functional models. Section 4 contains our experiments. We first introduce the different scenarios and give a short overview of the sensors and data used in the experiments. We then present and discuss the results of each scenario in a comparative way. In section 5 we recapitulate the results and discuss possible steps for future work.

2. RELATED WORK

Visual sensors provide important information for localisation in difficult environments. Cavegn et al. (2016) and Cavegn (2020) deal with the challenge of georeferencing image sequences. The authors employ image-based georeferencing by using bundle adjustment. With this method, they can reduce the residuals at checkpoints from approx. 40 cm, achieved with GNSS/IMU sensors only, to 4 cm. While these results show the potential of the idea to use visual localisation in autonomous driving for determining ego-motion, the authors use GCPs to obtain their results. In many applications, however, such GCPs are not available.

To increase the accuracy and reliability of localisation, a cooperation of several cameras is also used, especially in robotics; cf. the overview in (Zou et al., 2019). In CoSLAM (Zou and Tan, 2013), several moving cameras acquire data processed in a centralised adjustment. The authors state that they can scale their system to up to 12 cameras; image coordinates of points on moving objects are eliminated by outlier detection. Another possibility of cooperation is the recognition of cooperating platforms in the images. Stoven-Dubois et al. (2018) introduce an unmanned aerial vehicle (UAV) tandem system for surveying objects in GNSS denied areas. A so called surveying UAV flies next to the object to be surveyed 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. MapKITE (Molina et al., 2017; Nahon et al., 2019) also uses a tandem system. Here, the authors combine a terrestrial mobile mapping van with a UAV, so that they can make use of both types

of measurements. The van has a much higher payload and, thus, can be used for heavier and also more accurate equipment. In that approach, the vehicle is used as a dynamic GCP. For accurate automated positioning, a circular target is placed on the vehicle roof. In our previous work (Trusheim and Heipke, 2020), a comparison of static and dynamic GCPs in a traffic-related scenario based on simulated data was shown. Here, we apply this approach to real data.

Obviously, dynamic approaches must be able to handle time-dependent parameters. To this end, Colomina and Blázquez (2004) describe a model for trajectory and sensor orientation. The authors compare a state space and a network approach and point out the respective advantages and disadvantages. They show that, whereas state space approaches are faster, network approaches achieve higher precisions and allow for the integration of both time-dependent and time-independent models. For short trajectory sections, linear interpolation can be performed for time-dependent parameters; this is discussed in (Cucci et al., 2017a) and (Cucci et al., 2017b) regarding raw observations from inertial measurement units (IMUs) in dynamic networks.

3. COOPERATIVE VISUAL POSE ESTIMATION

In a dynamic environment, observations taken at different epochs are related to different states. In our case, we use observations of moving GNSS sensors to derive what we call dynamic GCPs which can be observed in images images acquired by cameras from another moving platform. In this case, the GNSS observations used to define the dynamic GCP do not refer to the 3D position that dynamic GCP was at when the images showing that GCP were captured. Thus, fusing observations of time-dependent processes typically requires interpolation of some entities, here platform poses, in a time-dependent model. We solve this problem by modelling the platform pose by a set of anchor points with linear interpolation in between. Each anchor point represents a 6 Degree of Freedom (DoF) pose of a platform at a specific point time. Image points in different images showing the same object point (i.e. potential tie points of the photogrammetric block) may also refer to different positions in 3D if these object points are not static. In this paper, we consider these non-static points as blunders, which are to be eliminated in a robust bundle adjustment.

In the bundle adjustment, we use three types of observations: (a) the position of the GNSS antenna of a platform in the global frame, (b) the image coordinates of conjugate points, which are used as tie points in space and time, and (c) the image coordinates of the marker points of a cooperating platform if these are visible. The (time-constant) 3D coordinates of the tie points in the global frame and the 6 DoF poses of the platform anchor points in the global frame are considered as unknowns; the platform poses at the time of image acquisition are interpolated from those of the anchor points by linear interpolation.

3.1 Functional models

We start the description of the functional model with the GNSS observations. Figure 2 shows the relationship between the position of the GNSS antenna in the platform frame and the position of the antenna observed in the global frame. We formulate this relationship as follows (the superscript indicates the frame; note that the global frame does not have a superscript. Also, a t in the subscript means "time-dependent"):

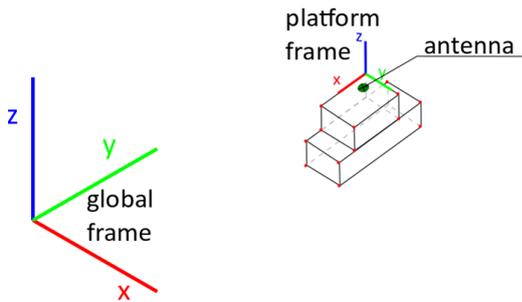$$X_{GNSS,t} = R_{plat,t} X_{GNSS}^{plat} + X_{plat,t}. \qquad (1)$$

Figure 2. Relationship between the position of the GNSS antenna in the platform frame and its position observed in the global frame.

Figure 3. Relationship between an object point in the global frame and its position observed in the image.

Here, $X_{GNSS,t,l}$ is the position of the GNSS antenna observed in the global frame at time t. $X_{plat,t}$ represents the position of the platform at time t, whereas $R_{plat,t}$ is a rotation matrix describing the rotation between the platform and the global systems at time t. It is a function of three time-dependent rotation angles $roll, pitch, yaw$ (r_t, p_t, y_t). Thus, $(R_{plat,t}, X_{plat,t})$ describe the unknown pose of the platform time t. X_{GNSS}^{plat} is the position of the antenna in the platform frame, which is constant and known from prior calibration.

Figure 3 illustrates the observation of a static object point which is used as a tie point. The functional model for the image coordinates of these tie points is described as follows:

$$X_{tp,t}^{plat} = R_{plat,t}^T (X_{tp} - X_{plat,t}) \tag{2}$$

$$X_{tp,t}^{cam} = R_{plat}^{cam\,T} \left(X_{tp,t}^{plat} - X_{plat}^{cam} \right) \tag{3}$$

$$x_{tp} = x_0 - c\frac{X_{X_{tp,t}^{cam}}}{Z_{X_{tp,t}^{cam}}} + \Delta x \tag{4}$$

$$y_{tp} = y_0 - c\frac{Y_{X_{tp,t}^{cam}}}{Z_{X_{tp,t}^{cam}}} + \Delta y.$$

In eq. 2 the coordinates of the tie point in the global frame X_{tp} and the pose $(R_{plat,t}, X_{plat,t})$ of the platform are unknown. The rotation matrix R_{plat}^{cam} and the shift X_{plat}^{cam} (eq. 3) represent the pose of the camera in the platform frame; this transformation is constant and given from prior calibration. Also, the interior orientation parameters x_0, y_0, c and the parameters of the distortion correction functions $\Delta x, \Delta y$ (eq. 4) are known from this calibration. Here, $X_{X_{tp,t}^{cam}}$ is the first component of the vector $X_{tp,t}^{cam}$, which represents the position of the tie point at time t in the camera frame, $Y_{X_{tp,t}^{cam}}$ is correspondingly the second component and $Z_{X_{tp,t}^{cam}}$ the third. x_{tp}, y_{tp} are the observed image coordinates of the tie point.

Finally, in figure 4 we show the case of an observed marker point attached to a cooperating vehicle. For such observations, we use the following model:

$$X_{mk,t}^{plat_k} = R_{plat_k,t}^T \left(R_{plat_j,t} X_{mk}^{plat_j} + X_{plat_j,t} - X_{plat_k,t} \right) \tag{5}$$

$$X_{mk,t}^{cam} = R_{plat_k}^{cam\,T} \left(X_{mk,t}^{plat_k} - X_{plat_k}^{cam} \right) \tag{6}$$

$$x_{mk} = x_0 - c\frac{X_{X_{mk,t}^{cam}}}{Z_{X_{mk,t}^{cam}}} + \Delta x \tag{7}$$

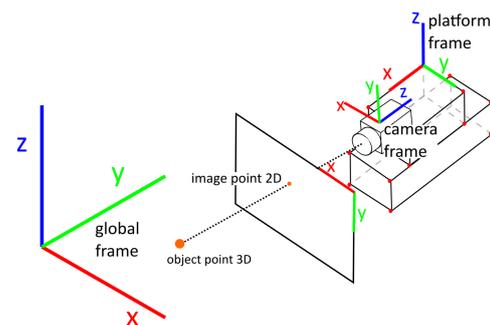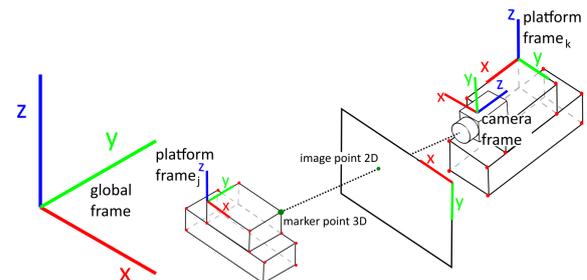$$y_{mk} = y_0 - c\frac{Y_{X_{mk,t}^{cam}}}{Z_{X_{mk,t}^{cam}}} + \Delta y.$$

Figure 4. Relationship between a marker point on a cooperating platform and its position observed in the image.

Here, platform j is the observed platform and platform k is the observing platform. We transform the position $X_{mk}^{plat_j}$ of the marker point on the observed platform j into the global frame using the pose $(R_{plat_j,t}, X_{plat_j,t})$ of platform j, where $X_{mk}^{plat_j}$ is determined in a prior calibration and assumed to be constant and the pose of the observed platform $(R_{plat_j,t}, X_{plat_j,t})$ is unknown (eq. 5). Using the pose $(R_{plat_k,t}, X_{plat_k,t})$ of the observing platform k, this point is transformed into the platform frame of k. Similar to eqs. 3 and 4, we then transform the point into image coordinates x_{mk}, y_{mk}, used as observations (see eqs. 6 and 7). Note that this type of observations contributes to the determination of the poses of both platforms.

3.2 Stochastic models

Each of the three different groups of observations is assumed to be of constant precision, and all correlations are neglected. As a consequence, the variance-covariance matrix of the observations is a diagonal matrix. For the observation of the GNSS antenna positions, we used a relatively conservative standard deviation of $\sigma_{XYZ_{GNSS}} = 0.5\ m$ for all components. The image coordinates are introduced with $\sigma_{xy_{tp}} = 0.8$ pixel for the tie points and $\sigma_{xy_{mk}} = 0.5$ pixel for the marker points. The accuracy for the marker points was chosen to be slightly smaller than the one for tie points because the markers were specially designed to be well identifiable in the images.

3.3 Platform trajectory modelling

As mentioned before, the trajectory of a platform is modelled by a series of anchor points A_{plat,t_i}, each consisting of a 3D vector X_{plat,t_i} representing the position of the platform at time t_i and another 3D vector $O_{plat,t_i} = (r_{t_i}, p_{t_i}, y_{t_i},)$ representing the time-dependent rotation angles forming the other component of the platform pose at time t_i. The components of the

anchor points are the actual unknowns in the adjustment. To determine the pose of the platform at the time t at which an image or a GNSS observation is acquired, we use linear interpolation between the neighbouring anchor points for both components:

$$X_{plat,t} = \frac{t - t_i}{t_{i+1} - t_i} \left(X_{plat,t_{i+1}} - X_{plat,t_i} \right) + X_{plat,t_i}$$

$$O_{plat,t} = \frac{t - t_i}{t_{i+1} - t_i} \left(O_{plat,t_{i+1}} - O_{plat,t_i} \right) + O_{plat,t_i} \quad (8)$$

where $X_{plat,t}$ and $O_{plat,t}$ represent the pose of the platform at time t, i.e. the time of observation, and the entities at times t_i and t_{i+1} are the corresponding entities of the anchor points. The time t_i is the time of the anchor point before the observation and t_{i+1} the one after the observation. The angles $O_{plat,t}$ are used to compute the rotation matrix $R_{plat,t}$ (e.g. eq. 1).

3.4 Precision

As a measure of quality of our investigation, we use the precision resulting from the bundle adjustment, contained in the covariance matrix of the unknowns. The inverse of the matrix of normal equations N contains information about the variances of the unknowns. In our case, we are interested in the 6 DoF poses of the anchor points of the trajectories and the 3D coordinates of the object points. The precision is then calculated by taking the square root of the corresponding entry of N^{-1}, multiplied by the a posteriori standard deviation of unit weight.

4. EXPERIMENTS

4.1 Scenarios

In this section, we consider five different scenarios involving two platforms to demonstrate the potential of cooperative localisation using bundle adjustment. The different combinations of observations used for the different scenarios are shown in table 1. In the following, the leading platform is named $platform$ $front$ and the trailing platform is named $platform$ $back$.

	platform back			platform front	
	TP	GNSS	MP	TP	GNSS
Sc. 1	✓	✓	✗	✗	✗
Sc. 2	✓	✓	✗	✓	✓
Sc. 3	✓	✓	✓	✗	✓
Sc. 4	✓	✓	✓	✓	✓
Sc. 5	✓	✗	✓	✗	✓

Table 1. Combination of observations used for the different scenarios (✓ = used, ✗ = not used). TP: image coordinates of tie points, GNSS: position of GNSS antenna, MP: image coordinates of marker points placed on $platform$ $front$. and observed by $platform$ $back$

Note that for some scenarios only image related observations from one platform are needed, and thus all computations (feature extraction, matching and bundle adjustment) can be executed locally on that vehicle: only the GNSS data of the other vehicle need to be communicated. For other scenarios, image observations from both vehicles are needed, thus locally extracted features and their description also need to be transferred to the vehicle which performs the bundle adjustment. However, a more detailed discussion of the communication aspects is beyond the scope of this paper. The five scenarios are defined as follows:

1. In the first scenario, we use only a single platform. The GNSS data and the image coordinates of tie points are used as observations in a bundle adjustment. This scenario provides a baseline regarding the obtainable precision using one multi-sensor platform and no cooperation.

2. In the second scenario the GNSS data and the image coordinates of the tie points of both platforms are used as observations in a common bundle adjustment. The inclusion of information from different platforms, including tie points observed from both platforms, should lead to a better result for the precision than in scenario 1.

3. In the third scenario, we use the image observations and the GNSS data of $platform$ $back$, in addition to the image coordinates of the observed marker points of $platform$ $front$ and its GNSS data, thus we use the markers to be able to use the observed platform as a dynamic GCP. Up to four marker points are visible in one image. Due to the use of the dynamic GCP, results should be better than in scenario 1, but not as precise as in scenario 2, because tie points observations are only used from $platform$ $back$.

4. The fourth scenario combines scenarios 2 and 3 by considering multiple platforms including a dynamic GCP. With this scenario, we want to check if the precision of scenario 2 can be further improved by using the additional cooperation strategy.

5. Finally, we show that the use of a dynamic GCP also makes it possible to calculate the ego-pose of the $platform$ $back$ even if no own GNSS data are available for that platform. For this purpose, we use the image coordinates of the tie points of $platform$ $back$ and the image coordinates of the observed marker points of $platform$ $front$, as well as the GNSS data of $platform$ $front$.

4.2 Data Acquisition

The data was recorded in a measurement campaign with several multi-sensor platforms. For this paper, a track is chosen where two platforms travel in tandem around a curve of approximately 90 degrees in an urban canyon. For the first half of the track, the vehicles travel in easterly direction and then turn turn south (figure 5). We use the stereo camera pairs and the GNSS receivers of both platforms shown in figure 1 in different combinations as described above. As the cameras look into the direction of travel, the distribution of tie points varies along the trajectory: for the part of the track after the turn there are significantly more points than for the first part, see figure 5.

The cameras we use are Grasshopper 3 USB cameras. They acquire images of 1920×1200 pixels at a frequency of $5\ Hz$ and have a focal length of $11.3\ mm$, equivalent to 1930 pixels. Image acquisition was initiated by an external trigger signal provided to both cameras. Based on this signal also the GNSS time is saved, therefore, all sensor data is given in the same time frame. The GNSS positions are captured using geodetic receivers Septentrio PolaRx5e SN 3061550 with a JAVRING-ANT G5T NONE, SN 06380 antenna at a frequency of $1\ Hz$. The images were taken on Aug. 25, 2020, at 5 pm, thus relatively late in the day. The sky was overcast, which led to difficult lighting conditions (figure 6).
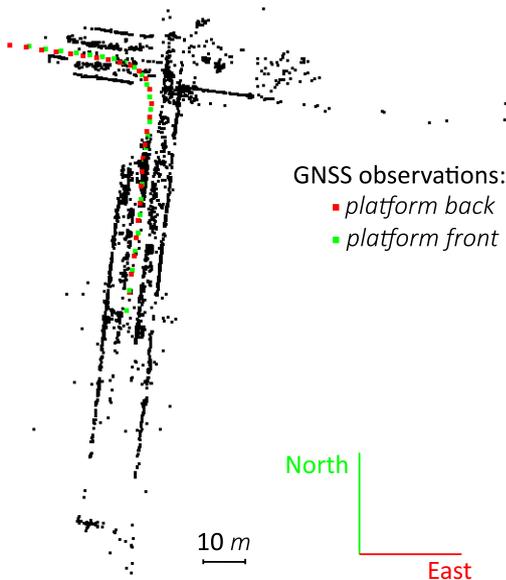
Figure 5. Data acquisition scenario. Black: sparse tie points; the GNSS observations of $platform\ back$ and $platform\ front$ are shown red and green, respectively.

Figure 6. Image taken by the trailing vehicle at a point in time between anchor point 12 and 13; the front vehicle can be seen in the centre of the image. Yellow: observed marker points. Blue: observed tie point. Red: tie points eliminated as outliers.

4.3 Experimental Setup

For the experiments, a section of the track covering about $25\ s$ was chosen. At a frequency of $1\ Hz$, this makes 25 GNSS observations for each platform, as well as 125 image pairs per stereo camera pair captured at $5\ Hz$ (500 images in total). To define the trajectory we place an anchor point every $0.25\ s$ ($4\ Hz$), yielding a total of 100 anchor points per trajectory. These anchor points were aligned with the GNSS observations such that the time stamp of every GNSS observation corresponds exactly to the time associated with one anchor point. Image coordinates of tie points were extracted using the software COLMAP (Schönberger et al., 2016; Schönberger and Frahm 2016), where we used SIFT-features and exhaustive matching.

In scenarios 3 and 5, where only GNSS observations are available for $platform\ front$, the anchor point density of this platform is reduced to $1\ Hz$ (25 anchor points), as there are no image observations which could support a denser selection. For the marker points, only observations in images closest in time to the GNSS observations of $platform\ front$ are used, which

is done to avoid interpolation errors. Thus, these marker point observations are only available at a frequency of $1\ Hz$.

4.4 Results

In all scenarios, the estimated standard deviation of the weight unit indicated a good fit of the stochastic model after convergence. The average back-projection error of the image coordinates amounts to approximately 0.65 pixels in the x-direction and 0.8 pixels in the y-direction. In general, it can be said that the height component of the positions of both the anchor and the object points can be determined with a lower precision than the planimetric components.

The number of tie points observed in the individual images is shown in figure 7. The reduction of the number of tie points at the beginning and the end of the right-hand turn of the vehicles (near anchor points 25 and 50, respectively, for $platform\ back$) is particularly noticeable. Figure 7 also shows that for $platform\ front$ there are significantly more tie points per image, which can be explained by a longer exposure time for the cameras on this platform.
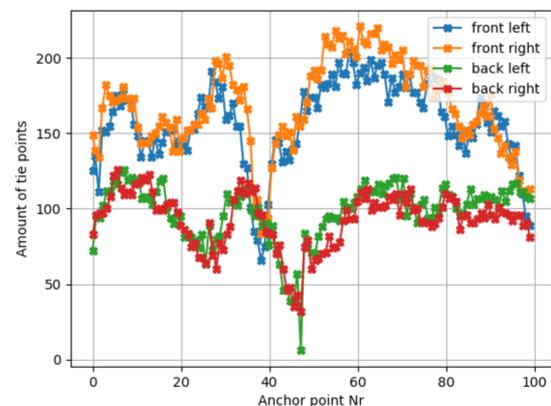

Figure 7. Number of tie points observed in each image.

An example for the distribution of these tie points can be seen in figure 6. The figure also shows the effects of using robust adjustment: in this image, some of the tie points were eliminated as blunders, and the majority of them lies on the moving $platform\ front$. We note, however, that not all these moving tie points were identified as blunders. In this image, four marker points can also be seen (marked in yellow). It has to be noted that they are not well distributed across the whole image, which somewhat weakens the solution.

Figure 8 shows the pose of the $platform\ back$ relative to its initial pose. The pose is represented by the positions and rotation angles of the individual anchor points. The upper part shows the positions. It can be seen that the platform first moves eastwards (increasing values of the blue curve) and then southwards (decreasing values of the orange curve), while the altitude remains largely constant (green curve). This 90° turn can also be seen in the yaw angle (green curve in the lower part of the figure). Roll and pitch (blue and orange curves), on the other hand, are nearly constant along the journey.

4.4.1 Scenario 1:
In this scenario, we only use $platform\ back$. Observations are the image coordinates observed by the stereo camera system and the positions of the GNSS antenna of the platform in the global frame. Figure 9 shows the precision for the 6 DoF poses of the anchor points in the global

frame. The overall precision of the position values (figure 9 top) is between 0.1 and 0.3 m. It is interesting to observe the impact of the right-hand turn of the platform on the precision. While driving along a straight trajectory, the σ_{Height} values decrease, whereas they increases during the right-hand turn of the platform before becoming smaller and increasing again towards the end of the trajectory. In general, the planimetric precision is better than the one in height. Due to the turning manoeuvre, the relationship between the two global planimetric position axes (East vs. North) and those of the platform (in vs. across driving direction) changes between anchor points 35 and 55. In the first part of the track, the orange curve (North/across-track) has behaviour a similar to the height, but is slightly more precise, partly due to the images having a wider horizontal than vertical extension. In the second half of the track, the blue curve (East/across-track) shows a corresponding behaviour. The precision in driving direction (first blue, then orange curve) is better than the one across driving direction.

The bottom part of figure 9 shows the estimated precisions of the rotation angles. The effect of the change of the driving direction on σ_{Roll} and σ_{Pitch} is similar to the one of σ_{East} and σ_{North}. Due to the spatial distribution of object points (figure 5), the precision of the rotation around the global East-axis (blue curve during the first part of the drive; orange curve in the second part) is better than the one around the global north-axis. The precision around the height axis (σ_{Yaw}) is the best one along the entire trajectory.

4.4.2 Scenario 2: In the second scenario, $platform\ front$ is introduced as a second platform to investigate cooperative localisation. In order to compare the results to scenario 1, again the results of $platform\ back$ are shown, see figure 10. It can be seen that the general shapes of the curves are similar to those in scenario 1, but in scenario 2 a better precision is achieved for all pose elements. This corresponds to our expectations according to the results of earlier simulations (Lenz, 2020).

4.4.3 Scenario 3: In the third scenario, the cooperative part is introduced by using the attached marker points (eqs. 5 and 7). This allows the $platform\ front$ to be used as a dynamic GCP. The observations used are the image coordinates of the marker points and the GNSS observations of $platform\ front$ in addition to those used in scenario 1.

In this scenario, a problem occurs: In the course of the trajectory there is a short time interval of about 5 s during the turn in which the marker points are not visible. Thus, we do not have any observations supporting the estimation of the related anchor points of $platform\ front$. This means that rotations of $platform\ front$ cannot be determined here. We find a workaround by regularising the solution; we introduce direct observation of the rotations of $platform\ front$ for each anchor point with relatively large standard deviations ($\sigma = 0.2\ rad\ (11.5^o)$ for roll and pitch and $\sigma = 0.5\ rad\ (28.6^o)$ for yaw), based on the assumption that the car only moves in the direction of travel and roll and pitch are rather constant over time. The yaw angle is estimated from two consecutive GNSS observations. In this way, numerical instabilities of the solutions are prevented.

If we compare the precision results of this scenario (figure 11) with those of the two previous ones, we notice that we achieve an improvement compared to scenario 1, but it is somewhat smaller than the one obtained in scenario 2. The reason is that the photogrammetric block is already geometrically rather stable in scenario 1, as enough well distributed tie points are

available, so the GNSS data for $platform\ back$ are sufficient to yield a rather precise solution, and the additional dynamic GCP does not have much effect. The precision of yaw is slightly decreased in the anchor points corresponding to the turn of the platform, when the $platform\ front$ is not visible.

4.4.4 Scenario 4: The fourth scenario combines scenarios 2 and 3. As the determination of the rotation of the $platform\ front$ is supported by the image coordinate observations of the tie points between the vehicles, the introduction of a regularisation as in scenario 3 is not necessary. The results achieved in this scenario are shown in figure 12. These are similar to the results obtained in scenario 2, which again shows that the introduction of the dynamic GCP does not have a significant effect if the photogrammetric block has a stable geometry stemming from a large enough number of well distributed tie points.

4.4.5 Scenario 5: In the fifth scenario, we consider the situation of localisation of a platform without GNSS observations taken by its own sensor, so that dynamic GCPs are the only information about the global frame. For this purpose, additional rotation observations of the $platform\ front$ are again needed to regularise the solution as described in section 4.4.3.

The results are shown in figure 13. The precision plots have a similar appearance than those for the other scenarios, but, as expected, compared to scenario 1, the solution is significantly less precise. The fact that during the right-hand turn the front vehicle is not visible has a further negative impact on these results.

4.4.6 Comparison Finally, we compare all scenarios based on the mean values for the precision of the pose parameters at the anchor points of the whole trajectory of $platform\ back$ (see table 2). We consider the first scenario as the baseline, as it does not contain any cooperation.

	$\overline{\sigma}_{East}\ [mm]$	$\overline{\sigma}_{North}\ [mm]$	$\overline{\sigma}_{Height}\ [mm]$
Scenario 1	131	109	168
Scenario 2	94	78	121
Scenario 3	98	80	126
Scenario 4	96	76	118
Scenario 5	159	126	218
	$\overline{\sigma}_{Roll}\ [^o]$	$\overline{\sigma}_{Pitch}\ [^o]$	$\overline{\sigma}_{Yaw}\ [^o]$
Scenario 1	0.61	0.52	0.29
Scenario 2	0.45	0.38	0.21
Scenario 3	0.48	0.42	0.22
Scenario 4	0.45	0.38	0.21
Scenario 5	0.93	0.77	0.37

Table 2. Mean precision of the 6 DoF pose over all anchor points of the different scenarios.

Scenario 2 yields an improvement in the precision of the position of 37 mm in East, 31 mm in North, 47 mm in height and an improvement in rotation precision of 0.16^o for roll, 0.14^o for pitch and 0.08^o for yaw. Overall, the average improvement over the precision of the 6 DoF pose is 27.5 % compared to the non-cooperative solution.

For scenario 3, an improvement in the precision of the 6 DoF poses of the anchor points compared to the baseline can also be observed. The improvements are 33 mm in East, 29 mm in North, 42 mm in height and 0.13^o in roll, 0.10^o in pitch and 0.07^o in yaw. This results in an average improvement of 24.0 % compared to the non-cooperative solution.
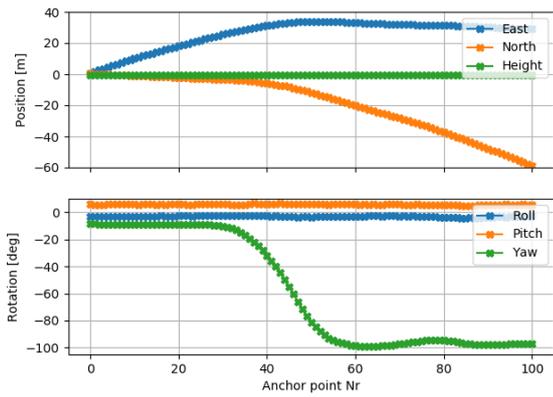
Figure 8. Trajectory of the *platform back* represented by the individual components of the anchor points. Top: Position of the anchor points in East, North and Height. Bottom: Rotations angles (roll, pitch, yaw).
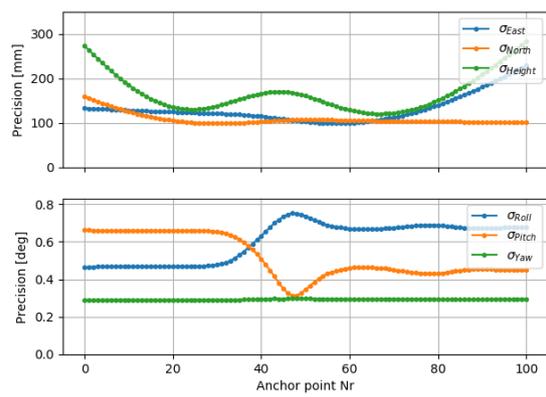

Figure 9. Results of *platform back*, scenario 1. Top: Precision of the anchor point positions in East, North and Height. Bottom: Precision of the anchor point rotations (roll, pitch, yaw).
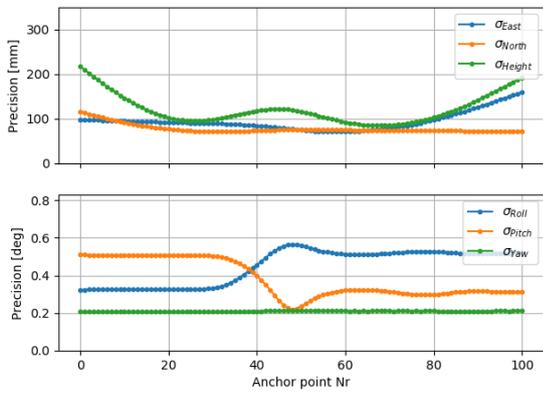

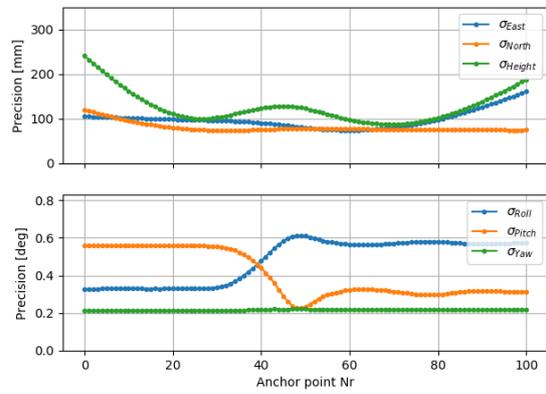Figure 10. Results of *platform back*, scenario 2 (details see figure 9).

Figure 11. Results of *platform back*, scenario 3 (details see figure 9).
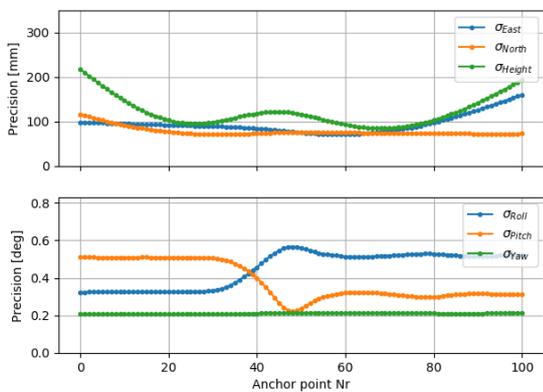

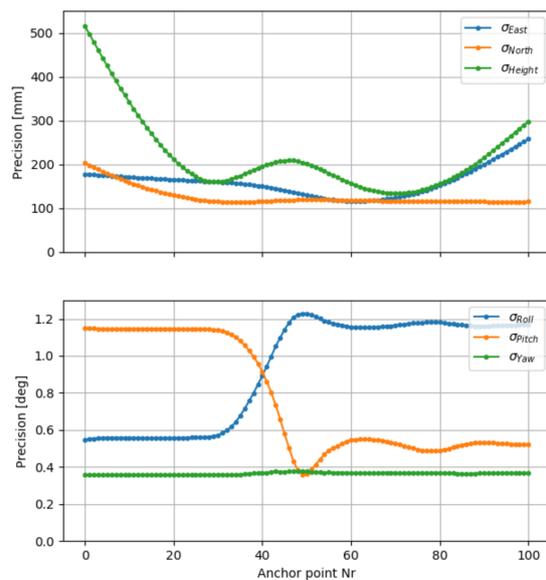Figure 12. Results of *platform back*, scenario 4 (details see figure 9).

Figure 13. Results of *platform back*, scenario 5 (details see figure 9).

For the combination of the two cooperation strategies (scenario 4), the individual, as well as the average improvement (27.8 %), is very similar to the one obtained in scenario 2.

Scenario 5 is interesting, as in this case cooperation is the only way to achieve a solution. Compared to scenario 1, the precision in position deteriorates by $28\ mm$ in East, $17\ mm$ in North and $50\ mm$ in height, a deterioration is also found in the precision of rotation of 0.32^o in roll, 0.25^o in pitch and 0.08^o in yaw. Overall, the precision of the 6 DoF pose deteriorates by 32.2 % on average.

5. CONCLUSION

In summary, we underline that a significant improvement in visual localisation can be obtained through cooperation. It was shown that both, cooperation by using other participants as dynamic GCPs as well as cooperation in a common bundle adjustment, lead to improvements in the precision of over 20 % with respect to the uncooperative approach.

The improvement obtained with a common bundle is slightly higher compared to using a single dynamic GCP, but a larger number and a better distribution of dynamic GCPs will improve the results obtained. Furthermore, the results show that cooperation can also compensate for the (temporal) absence of GNSS data. Such situations often happen in urban environments.

In further work, we will introduce more general interpolation schemes, which will allow us to be more flexible concerning defining anchor points. These will also be chosen as a function of driving mode (straight course, turn etc.), and we will select different distances between anchor points for different pose parameters. As a further improvement, additional sensor data could be introduced, such as IMU data, which are available in a higher measuring frequency and provide further information about the driving behaviour. These extensions should also be examined to avoid the regularisation for the rotation introduced in scenarios 3 and 5.

Another aspect is that in this work dynamic tie points are considered as blunders and are eliminated in a robust adjustment. In figure 6 it is shown that this is true for some points, but it is also visible that some points lying on the front vehicle are not eliminated. In future work, we will investigate possibilities to subdivide the tie points into dynamic and static before the adjustment. This should lead to a further improvement, also because trajectories can be defined for dynamic tie points also.

We conclude that our work shows that cooperative visual localisation in a real-world traffic environment leads to promising results with improved precision.

ACKNOWLEDGEMENTS

References

Cavegn, S., 2020. Integrated georeferencing for precise depth map generation exploiting multi-camera image sequences from mobile mapping. PhD thesis, Aerospace Engineering and Geodesy, University of Stuttgart, http://dx.doi.org/10.18419/opus-11210.

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

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.

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)*, 2048–2062.

Lenz, M., 2020. Untersuchung von visueller Lokalisierung mithilfe mehrerer kooperierender Kameras. Master's thesis, Institute of Photogrammetry and Geoinformation, Leibniz University Hannover, https://www.ipi.uni-hannover.de/fileadmin/ipi/abschlussarbeiten/master/2020/2020 _Thesis_Lenz.pdf.

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 foredunes with UAS photogrammetry and mobile laser scanning. *Remote Sensing*, 11(11), 1352.

Schönberger, J. L., Frahm, J.-M., 2016. Structure-from-Motion revisited. *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Schönberger, J. L., Zheng, E., Pollefeys, M., Frahm, J.-M., 2016. Pixelwise view selection for unstructured multi-view stereo. *European Conference on Computer Vision (ECCV)*.

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)*, 24–28.

Trusheim, P., Heipke, C., 2020. Precision of visual localization using dynamic ground control points. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2020, 363–370.

Zou, D., Tan, P., 2013. CoSLAM: Collaborative Visual SLAM in dynamic environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2), 354–366.

Zou, D., Tan, P., Yu, W., 2019. Collaborative visual SLAM for multiple agents: A brief survey. *Virtual Reality & Intelligent Hardware*, 1(5), 461 - 482.