A Testing-environment for a Mobile Collaborative Stereo Configuration with a Dynamic Baseline

Andreas Sutorma[a], Matthias Domnik[b] and Jörg Thiem
Department of Information Technology, University of Applied Sciences and Arts Dortmund, Sonnenstraße 96, 44139 Dortmund, Germany

Keywords: Collaborative, Stereo, UAV, Mapping, Dynamic, Baseline, Visual, VICON, Environment, Mobile.

Abstract: This contribution deals with the construction of a testing-environment for the development of a camera based collaborative stereo configuration with a dynamic baseline. The use of a variable baseline for the stereo configuration allows to perform a more accurate depth calculation of the environment. For the development of such a collaborative stereo configuration it's necessary to compare the results with ground-truth data. A VICON systems is a very capable solution for UAV and Robotic studies because of the high accuracy and low latency. This external localization system is intended to determine the dynamic stereo baseline at the first step of development. At a later progress this task will be taken over by another calibrated stereo camera that belongs to the mobile collaborative stereo configuration.

1 INTRODUCTION

The development of autonomous UAVs/UGVs (Unmanned Air/Ground Vehicle) for the exploration of large areas is a current research task. In most cases the exploration is done with a RGB camera on a single UAV. This can be a mono camera but also a calibrated stereo camera. In the case of a mono camera, the poses between the individual shots are needed to determine the depth information. This is usually solved with the use of an IMU (Inertial Measuring Unit) or GPS, if this service is available. The IMU can only measure relative movements, this leads to a steadily increasing error in pose estimation over the time (drift). For this reason, it is necessary to cyclically support the measured data with other information, for example with visual data from cameras (visual odometry). Using visual odometry has the advantage of reducing the error of position estimation by renewing already measured points when they are recognized (loop closure). That is possible when the features of the points are stored in a generated map (mapping). This leads to the established Visual-SLAM (Simultaneous Localization and Mapping) method. The biggest disadvantage of the Visual-SLAM arises when the camera system degrades to a monocular case. This happens when the distance of both cameras (baseline) is much smaller than the distance of the object (feature). In this case a larger baseline within the calibrated stereo camera configuration can help. But a single UAV is limited in size and portability what makes it hard to realize a large baseline. This leads to a stereo configuration with two UAVs that are equipped with one mono camera to form a variable baseline. The difficulty lies in determining the exact length of the baseline. There are different approaches to solve this problem.

2 RELATED WORK

The term "Collaborative Stereo" is used for mobile robot applications that use a large stereo baseline to increase the measured 3D reconstruction (Achtelik et al., 2011; Boulekchour, 2015). This approach is used on small mobile robots like UAVs, because they are limited in their weight and size. The idea is to use at least two UAVs equipped with a mono camera to realize a stereo setting with a variable baseline. You need additional sensor information if you want to use a collaborative stereo configuration with just two UAVs equipped with mono cameras. Using only feature correspondences in the overlapping field of view from two monocular cameras, we estimate the relative 6 DoF transformation between the robots poses up to

[a] https://orcid.org/0000-0002-7965-2495
[b] https://orcid.org/0000-0001-6501-2246

Sutorma, A., Domnik, M. and Thiem, J.
A Testing-environment for a Mobile Collaborative Stereo Configuration with a Dynamic Baseline.
DOI: 10.5220/0007951104680474
In *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2019)*, pages 468-474
ISBN: 978-989-758-380-3

a scalar factor in translation. The knowledge of the correct transformation is important for the accuracy of a three-dimensional reconstruction. Some recent approaches address this problem of determining the true translation in this collaborative stereo configuration.

GNNS (Global Navigation Satellite System) services can be used to determine the relative pose of the mono cameras. The accuracy depends on the GPS localization (Dias et al., 2013). A high accuracy can not be assumed permanently and in some environment the availability can not be assumed all over the time.

Another work (Achtelik et al., 2011) estimates the relative pose of two UAVs by merging the poses of downward oriented mono cameras with IMU sensor data to solve the problem of the scalar factor in translation. But this approach needs a continuous relative movement between the UAVs to converge and this takes up to 8 seconds. Nonetheless, this approach finds application in swarm-based research projects.

The use of ultra-wideband technology shown in (Guo et al., 2017) can also determine the distance between two UAVs. The Jet Propulsion Laboratory (JPL) are working on similar approaches with two small (Roland Brockers, 2015). These UAVs are using a tandem formation during the flight to create depth maps of the terrain. The distance between the UAVs is determined by an ”antenna monitoring system”. In Addition, some solutions exist (Kwon et al., 2014) where two UAVs are tracking a visible target (fiducial marker) on a ground vehicle. However, this case has the disadvantage that the target must be very large, or the distance to the target must be very small to achieve accurate measurements.

Another possible solution is to use an external motion capture system to localize the UAVs (Ahmad et al., 2016). This is only applicable in proper equipped indoor rooms. Such a motion capture system is used in this work to realize a test environment.

3 COLLABORATIVE STEREO WITH MASTER-SATELLITE CONFIGURATION

This section outlines a novel method to realize a dynamic baseline for mobile robotic applications as presented in (Sutorma, Andreas and Thiem, Jörg, 2018). A flexible formation of at least three optically measuring systems (e.g. UAVs with cameras) makes it possible (see Fig.1). A master (calibrated stereo camera with fixed baseline) is located behind the so-called satellites (mono cameras with variable baseline). This master-satellites stereo (MSS) configuration allows the master to estimate the relative pose of the two mono cameras (satellites), whose positions may also vary over time. The satellites must be equipped with markers to estimate the translation and rotation with respect to each other (extrinsic stereo parameters). Furthermore, the master may control the baseline of the satellites configuration to optimize the triangulation and resolution for different (near vs. far) scenarios. The variable baseline increases the accuracy over standard, fixed-baseline stereo methods.

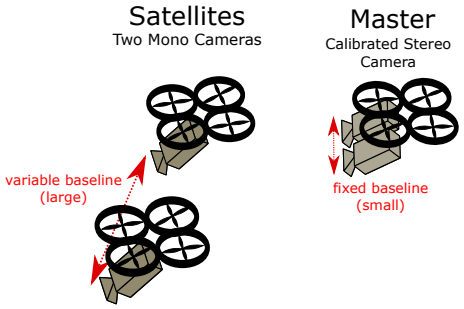

Figure 1: Variable baseline stereo configuration with two mono cameras and one calibrated stereo camera system.

4 ADVANTAGE

The advantage of collaborative stereo and therefore also our proposed MSS configuration is the realization of a flexible and large baseline. This makes it possible to achieve a high accurate depth estimation. However, as in all approaches for collaborative stereo, the relative pose of the two mono cameras have to be estimated in realtime with a proper uncertainty. This aspect has to be analyzed in theory and to be concidered in practical testing. This section therefore will show the theory of the resulting depth error within different baselines. The reconstructed depth coordinate Z of an object point $P = (X,Y,Z)^T$ in a rectified stereo image pair is calculated by the well-known relation Eq.1

$$Z = \frac{b \cdot f_{px}}{d} \tag{1}$$

with the stereo baseline b, the normalized focal length f_{px} and the disparity d. In our experimental environment the camera has a focal length of f_{px} = 1389 px. For the further calculation we consider the focal length as exact and expect a disparity error of 1 pixel.

$$\Delta Z_d = \left|\frac{\partial Z}{\partial d}\right| \cdot \Delta d = \left|-\frac{b \cdot f_{px}}{d^2}\right| \cdot \Delta d \tag{2}$$

Replacing the disparity d with the expression:

$$d = \frac{b \cdot f_{px}}{Z} \tag{3}$$

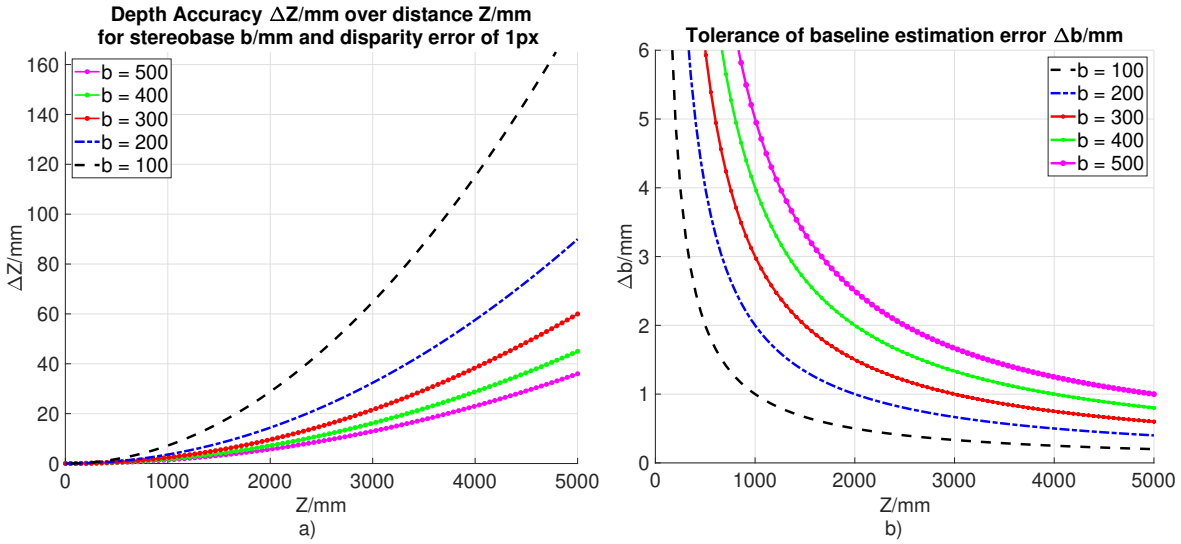

Figure 2: Depth accuracy over distance for different stereo baseline configurations (a) and the maximum error tolerance for the determination of the baseline to aquire a maximum error of 10mm (b).

leads to the simplified depth error ΔZ_d in Eq.4:

$$\Delta Z_d = \frac{Z^2}{b \cdot f_{px}} \cdot \Delta d \tag{4}$$

As we can see in Fig.2a), the expected depth error becomes less with a larger stereo baseline.

In this context, a larger stereo base in the application allows a greater tolerance for determining the baseline to achieve the same resolution as with a camera system with a smaller stereo baseline. To guarantee a certain resolution in the 3D map, an error limit for the determination of the baseline can be calculated. Same as in Eq.4, this time a derivation to the baseline b is performed. In addition, an inequality is used to determine an error limit Z_{max} (Resolution).

$$\Delta Z_b = \left|\frac{\partial Z}{\partial b}\right| \cdot \Delta b = \frac{f_{px}}{d} \cdot \Delta b = \frac{Z}{b} \cdot \Delta b < \Delta Z_{max} \tag{5}$$

This leads to the Eq.6 for the allowed error of baseline determination.

$$\Delta b < \frac{\Delta Z_{max}}{Z} \cdot b \tag{6}$$

The Fig.2b) shows the result of the baseline estimation error tolerance Δb. The bigger the baseline, the bigger is the tolerance for the baseline estimation error.

5 METHODOLOGY

The following method is used to evaluate multi-baseline stereo systems as well as collaborative stereo configurations. A special focus is laid on the fact that both static measurements, which allow a maximum repeatability of the experiments, as well as dynamic experiments are to be realized. Particularly with regard to the use of the test setup with UAVs, a great dynamic is to be depicted in future tests.

To determine the camera pose in the task space for each frame an optical reference system is used (see section 6). In the case of the Master-Satellite Stereo Configuration the reference system may be the calibrated stereo camera of the master. Of course, the measurements of the reference system contain inherent errors. This reference system is considered as error-free in the first step. However, the expected error is assumed to be sufficiently small to allow a first evaluation for the presented method.

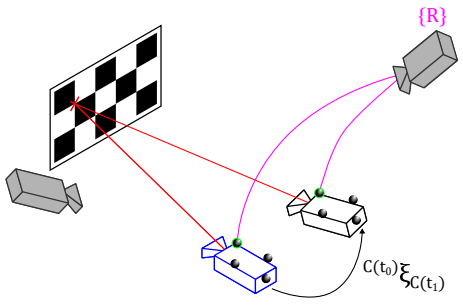

Figure 3: Schematic drawing of the setup for the indirect error estimation method presented here. The camera with the blue outlines represents the scene at time t_0 and the camera with the black outlines at time t_1. The gray cameras denote the optical reference system (VICON).

The basic setup is shown in Fig.3. A frame,

$I(t_0)$ respectively $I(t_1)$, is taken with a pre-calibrated monocular camera at time t_0 (camera with blue outlines) and at time t_1 (black outlines). The pose between frame $I(t_0)$ and $I(t_1)$ is determined with the given relation ${}^{C(t_0)}\xi_{C(t_1)}$ by the optical reference system (gray cameras in Fig.3). For our method, a reference frame is initially defined for each test cycle, to provide the corresponding stereo image in consecutive frames. Typically, the first frame $I(t_0)$ is taken for this purpose, but this is not mandatory. The inherent error of the optical reference system is, in a first approximation, constant over the entire volume and does not accumulate for consecutive measurements so that any frame can serve as a reference.

In the first step, a checkerboard is used to evaluate the point-by-point 3D reconstruction of the stereo images. The a priori known width and height of the checkerboard squares (see Fig.4) make it possible to carry out a measurement of the square widths and square heights after the calculation of the three-dimensional corner points (green crosses) of the individual fields.

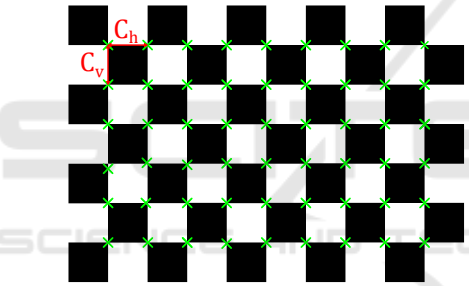

Figure 4: Layout of the checkerboard with a field with of C_W and field height of C_H. The green crosses mark the corners we used for the measurement process.

Due to the relationship $C_H = C_W = const$, only the square width is mentioned below. The three-dimensional coordinates of the checkerboard corner points, relative to the reference frame, are determined by triangulation. The calculation of the square width is based on the L_2-norm. We used a custom-made checkerboard of high precision. The width of its squares is considered as ground truth value. The calculated field widths provide a indirect error estimation for each stereo setting on the 3D reconstruction performed with the baseline b and the vergence angle Θ given by ${}^{C(t_0)}\xi_{C(t_1)}$.

6 EXPERIMENTAL SETUP

The used optical reference system is the VICON Vero 2.2 motion capture system, which scans a volume of approximately 3 m × 6 m × 3 m (WxDxH) at 300 Hz. For tracking objects in space, the motion capture system requires reflective markers, as shown in Fig.5.

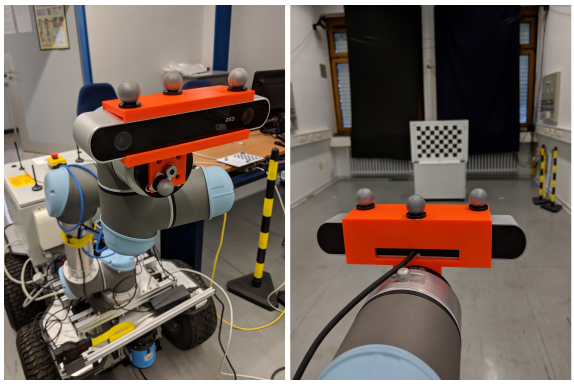

Figure 5: Our measuring setup consists of a camera system (ZED Stereolabs) with reflective markers mounted on a mobile robot platform. This camera is aligned to a checkerboard around 4 meters away.

For the first experiments, a mobile robot platform with a built-in universal robot UR5 was used, to accurately move a camera simulating the satellite camera system. This allows us to capture multiple frames at one pose. The frames are taken by the calibrated stereo camera ZED from stereolabs. However, we mainly use only the left ZED camera for our first step. This camera system has a fixed stereo baseline and provides a datastream through ROS nodes but also can be used directly with USB 3.0. This calibrated stereo camera with a fixed stereo baseline is used exclusively for the previously described calibration process. Further, the stereo camera is attached to a Nvidia Jetson TX2 development kit, that runs a ROS node for distributing the frames to the network. The frames are taken with a resolution of 1920 × 1080 (1080p) in single shot mode. The checkerboard has 10 squares in the horizontal direction and 7 squares in the vertical direction with a total size of 655.0 mm × 458.5 mm. A distance of approximately 4.5 m is provided between the checkerboard and the camera.

Fig.6 shows an exemplary plot of the camera positions (green identifies the reference frame) measured by the optical reference system and the corresponding checkerboard position. The increment between the consecutive camera positions for the variable baseline configuration is approximately 100 mm.

7 EXPERIMENTAL RESULTS

The results from the experimental setup show that the theoretical basics in this setup are fulfilled. Fig.8 shows the results of a series of measurements in

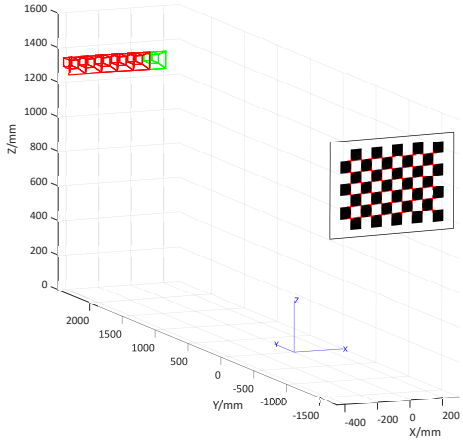

Figure 6: Exemplary plot of the camera positions in the experimental setup with the reference camera frame (green). This positions are estimated by the optical reference system.

which 30 images were taken for each stereo baseline configuration. A larger stereo baseline increases the accuracy of the measurement as it is expected from Eq.4 in the case of measuring Z. The standard deviation also decreases as the stereo baseline increases. Note, this is an indirect measurement of the depth accuracy ΔZ. So, we measure the error of the checkerboard square width ΔC and not the depth accuracy ΔZ directly.

8 FUTURE WORK

The proposed method serves a direct error estimation of the regarding measurement. The next step is to extend the system to provide quantitative depth information. An extension of the previous method can be done by determining all poses of the measurement setup, that the determined depth Z by the stereo camera system becomes comparable to a quantitative value Z_{ref}. The poses ${}^{R}\xi_{CB}$, ${}^{R}\xi_{C(t_0)}$, and ${}^{R}\xi_{C(t_1)}$ of the current setup (see Fig.7) are given with respect to a global reference point in the testing-environment. However, since the relative poses are crucial, the global reference point is not significant in this case.

The future measurement setup should work with absolute poses. This will be necessary to perform a continuous calibration of the stereo camera system (consisting of the satellites) during the flight in the previously presented formation of three UAVs (see Fig.1). For this reason, it is necessary to identify the unknown poses ${}^{CB_{MP}}\xi_{CB}$ and ${}^{C_{MP}}\xi_{C(t_0)}$. The marker point on the checkerboard $\{CB_{MP}\}$ can be placed very precisely by knowing the accurate geometric dimensions of the checkerboard. So, the error in the pose ${}^{CB_{MP}}\xi_{CB}$, between the reference point and the origin

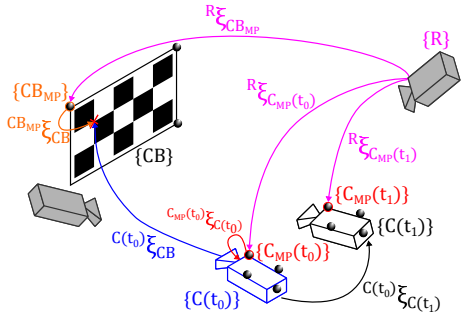

Figure 7: Schematic representation of the relations between the checkerboard, the camera over time and the optical reference system.

of the checkerboard can be disregarded. The pose ${}^{C_{MP}(t_0)}\xi_{C_{(t_0)}}$, between the marker point on the camera and the origin of the camera coordinate system, is not directly measurable. For determination, the known transformations are concatenated:

$$ {}^{C_{MP}}\xi_{C} = \ominus {}^{R}\xi_{C_{MP}} \oplus {}^{R}\xi_{CB_{MP}} \oplus {}^{CB_{MP}}\xi_{CB} \ominus {}^{C}\xi_{CB} \quad (7) $$

The temporal dependence of the poses can be neglected for this consideration, since the relative pose ${}^{C_{MP}}\xi_{C}$ does not change over time. The knowledge of all poses makes it possible to determine the reference distance to the origin of the checkerboard and thus to check the measurements of the stereo system at any time. The largest error is contributed by the pose ${}^{C(t_0)}\xi_{CB}$. Therefore, this pose can be checked in the following calibration. For initial error estimation, therefore, a stereo camera is used whose extrinsic parameters are known by the calibration.

For the initial calibration process, a frame is recorded from a static position. Subsequently, the three-dimensional reconstruction takes place with reference to the left as well as the right camera image. Using the resulting poses, considering Eq.7, the poses between C_{MP} and C_L, respectively C_R, are determined (see Fig.9). The following relationship is derived with the known pose of the extrinsic camera parameters ${}^{CR}\xi_{CL}$:

$$ {}^{C_{MP}}\xi'_{C_L} = {}^{C_{MP}}\xi_{C_R} \oplus {}^{C_R}\xi_{C_L} \quad (8) $$

The concatenation of ${}^{C_{MP}}\xi'_{C_L}$ and ${}^{C_{MP}}\xi_{C_L}$ corresponds to the identity matrix:

$$ {}^{C_{MP}}\xi'_{C_L} \ominus {}^{C_{MP}}\xi_{C_L} = \underline{I} \quad (9) $$

If this is not the case, the pose ${}^{C_{MP}}\xi_{C_L}$ is highly susceptible to errors and the calibration must be repeated. For practical implementation, suitable error limits with regard to rotation and translation must be defined. Note, that the errors in the calibration process of the stereo system are assumed to be negligibly small.

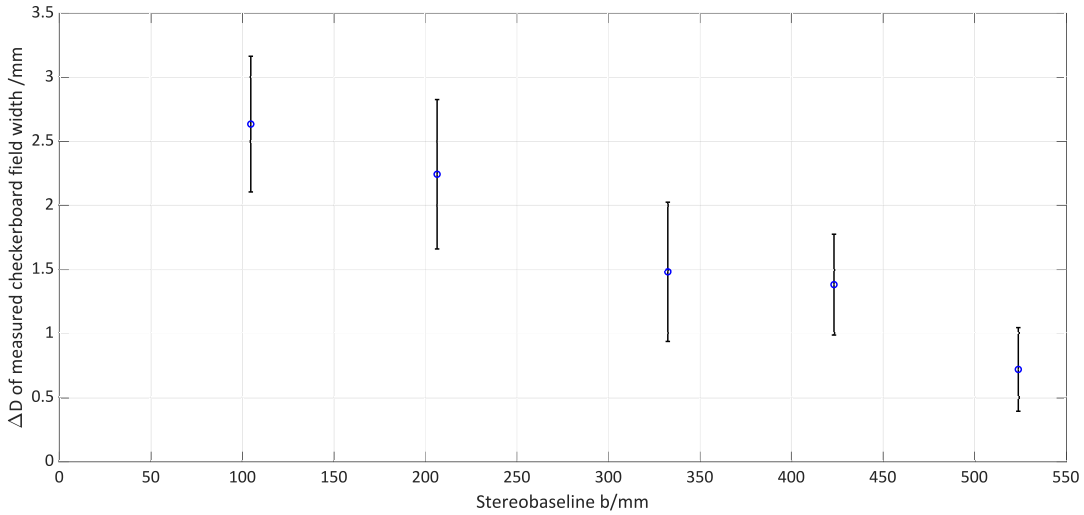

Figure 8: The mean value with the standard deviation of measured field width ΔC with different stereo baseline configurations.

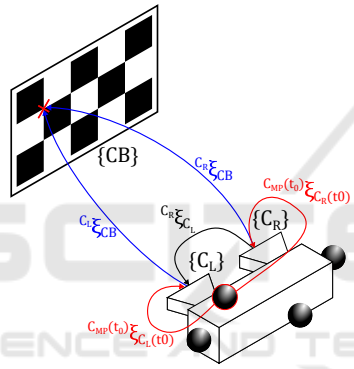

Figure 9: Theoretical verification approach for calculated pose between the marker on the camera and the camera itself.

In the upcoming development, this calibration process will take place on real UAVs, where vibrations during the flight and the synchronization of the stereo image pairs must be considered.

9 CONCLUSION

In this paper, an approach for a testing environment for a camera based collaborative stereo configuration was presented. Our method shows a first step towards a reliable and flexible testing-environment for multi-baseline systems with a great dynamic, like UAVs. The results show a comprehensible behavior regarding the theoretical part in section 4. Our presented method does not establish a direct relation between the depth accuracy ΔZ and the error of the checkerboard field width ΔC. Therefore, we have to further investigate the proposed concept in section 8, that allows to measure the depth accuracy ΔZ directly. However, our first effort shows promising results towards a reliable setup for testing multi-baseline stereo systems as well as collaborative stereo configurations.

The experimental setup presented here will be used in the further development and evaluation of the master-satellites stereo method. This is necessary to check the calibration process of the satellites, that will be done by the master UAV. The VICON camera system provides the Ground-Truth data and enables a flexible and reconstructible test design.

The proposed MSS method can always be of great importance in applications where constant high accuracy in a dynamic environment is required even at large object distances. This approach enables the usage when a high-resolution accuracy of a 3D depth map is required, e.g. such as surveying engineering offices for the planning of new construction projects. The survey method provides access to significantly more accurate 3D maps. Evidence protection tasks such as the progress of the construction, deviations from the construction plans or changes in the building structure could be examined even more precisely and digitized for eternity. In addition, the security sector can also make use of this collaborative stereo setup on UAVs. Environmental disasters have increased rapidly because of climate change. As a result, floods, storms and forest fires have increased in recent years. Disaster situations are typically chaotic, confusing, stressful and dangerous. UAVs offers the ability to scan a large area quickly and accurately, reducing search time and making deployment planning more reliable. Areas where earthquakes and floods occur in frequency, this 3D measurement tasks will benefit from a collaborative stereo approach. It will

be possible to locate particularly damaged areas (so-called hotspots) from the high-resolution 3D maps and thus to initiate appropriate relief measures in a targeted manner. In addition to environmental disasters, there are other scenarios for the security sector from safety-critical infrastructures, which need to be monitored on a regular basis. These include, for example, power plant and energy systems, road and railway routes and high-voltage roads.

REFERENCES

Achtelik, M. W., Weiss, S., Chli, M., Dellaerty, F., and Siegwart, R. (2011). Collaborative stereo. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2242–2248. IEEE.

Ahmad, A., Ruff, E., and Bülthoff, H. H. (2016). Dynamic baseline stereo vision-based cooperative target tracking. In *2016 19th International Conference on Information Fusion (FUSION)*, pages 1728–1734.

Boulekchour, M. (July 2015). *Robust Convex Optimisation Techniques for Autonomous Vehicle Vision-based Navigation*. Phd thesis, Cranfield University.

Dias, A., Almeida, J., Silva, E., and Lima, P. (2013). Multi-robot cooperative stereo for outdoor scenarios. In Calado, J. M. F., editor, *2013 13th International Conference on Autonomous Robot Systems (Robotica)*, pages 1–6, Piscataway, NJ. IEEE.

Guo, K., Qiu, Z., Meng, W., Xie, L., and Teo, R. (2017). Ultra-wideband based cooperative relative localization algorithm and experiments for multiple unmanned aerial vehicles in gps denied environments. *International Journal of Micro Air Vehicles*, page 1756829317695564.

Kwon, J.-W., Seo, J., and Kim, J. H. (2014). Multi-uav-based stereo vision system without gps for ground obstacle mapping to assist path planning of ugv. *Electronics Letters*, 50(20):1431–1432.

Roland Brockers (2015). Adaptive resolution stereo-vision (ares-v), https://www-robotics.jpl.nasa.gov/tasks/completed.cfm.

Sutorma, Andreas and Thiem, Jörg (2018). Collaborative stereo configuration of master and satellite unmanned air vehicles for a dynamic baseline. Indoor Positioning and Navigation Conference.