Collaborative Exploration with Dynamically Configurable Sensing Agents

Darius Burschka

Collaborative Exploration with Dynamically Configurable Sensing Agents

Darius Burschka

Computer Vision and Perception Group
Technische Universität München, Germany
burschka@cs.tum.edu

Abstract. We present a novel approach that provides the necessary localization and mapping capabilities to allow a dynamic placement of independent sensing agents. The presented approach allows to *build* complex sensor configurations using simple agents that carry basic sensors, like e.g. low-resolution cameras, which can be combined to high precision and high accuracy systems through fusion of their information and through optimized sensor placement.
We discuss the mathematical background and verify the resulting system on simulated and real world data. We compare our system to other similar systems in the field of collaborative localization and mapping.

1 Motivation

The size and complexity of high accuracy 3D sensors require large system dimensions and significant processing power. Especially in case of stereo systems, it is known that the accuracy of a binocular setup is scaled according to the basic stereo equation [8] that calculates the image disparity d to:

$$d = \frac{B \cdot f}{p_x} \cdot \frac{1}{Z} = K_s \cdot \frac{1}{Z} \tag{1}$$

with B representing the baseline between the cameras, f being the focal length and p_x representing the pixel-size in the image. Increasing the sensitivity of a stereo system means to increase the K_s factor in equation (1). A large field of view is essential for an exploration system. Therefore, an increase of the focal length f is a wrong solution. It reduces the viewing angle of the camera. An increase of the resolution of the camera by decreasing p_x improves the accuracy at a cost of increased processing power demands, because the number of pixels in the image increases quadratically with doubling of the image size. The only suitable parameter to achieve the required accuracy is the baseline B. There are physical limits to how much the cameras can be moved apart, which are usually given by the geometry of the vehicle. Additional problems are vibrations and distortions in the mechanical structure that reduce the final accuracy of the results.

Recent significant miniaturization of the devices provides a unique opportunity to deploy multiple small agents to explore a given environment. This

solution has the advantage of flexible task-dependent specification of their locations. These systems operate typically with low-resolution sensors and with very limited processing power.

Multiple groups work on flying agents with very small processing units on board. Most of them are equipped with video cameras. We treat these systems as *flying eyes*. The on-board systems are used only for navigation and stabilization tasks. An example of such a system is depicted in Fig. 1-left. The actual 3D reconstruction and model fusion cannot be done on-board due to limited resources. It needs to be moved to a ground station or a coordinating agent, like the blimp system in Fig. 1-right.

Fig. 1. Example of collaborating systems that explore an area.

In this paper, we propose a novel way of collaborative reconstruction of the environment with simultaneous localization of the independent cameras to each other. This approach is derived from the approach in astronomy where simple agents with limited resources (telescopes) spread around a large area are coupled together to powerful sensing systems that replace large and expensive single telescopes. Instead of mounting complex sensors on the agents, we define what sensor configurations are appropriate for a given task and we map them on distributed configurations of agents (robot swarms) to resemble the required perceptual capabilities. A good example for the gain achieved with our approach is an increase of the resolution of a sensor system by distributing the cameras on multiple agents that easily define arbitrary large baselines while moving in the environment. This way, the sensitivity of the system can be focused on a specific region in the world and a high information gain is expected from simple sensing units. This task can be performed by a single agent in a VSLAM approach [3, 6] but such a system tends to be very sensitive to motion in the environment. The motion is difficult to detect, because the images used for reconstruction are acquired at different points in time. A set of multiple agents performing collaborative localization and reconstruction avoids this problem.

Typically, navigation units operating in outdoor environments do not navigate with a sufficient accuracy to define the extrinsic parameters of the resulting stereo system (rotation matrix $\mathbf{R}$ and translation vector $\boldsymbol{T}$ describing the extrinsic motion parameters between the agents). Our approach presents a solution to this problem.

1.1 Related Work

The related work in this area can be subdivided in sensor network approaches and multi-agent localization. The sensor network approaches like [11] try to estimate the node geometry using distance and direction measurements between the nodes. They try to estimate the graph geometry from measurements about the neighbors. In general, design of a sensor network for localization and tracking involves both spatial sensor deployment and receiver pattern modulation. Deployment strategies were previously considered [14] assuming circular receiver pattern. Related work on sensor deployment and path identification used path exposure as a merit function to determine the quality of sensor deployment for target detection [5]. Other groups looked at recover pattern design in [12, 9, 15]. An excellent overview over network localization approaches based on graph localization is given in [1]. In opposite to these approaches, our system is designed to be used in dynamically changing configuration of flying systems. We use landmarks from the surrounding environment in our approach to simplify the reconstruction of the structure. A related approach is followed in work of C.J. Taylor [13], where localization of multiple agents is implemented based on bearing readings to the neighboring systems. Taylor's system requires 3 robots collaborating together, while our approach works already with two collaborating agents.

Our paper is structured as follows. In section 2.1, we present our approach to reconstruct 3D from motion stereo that allows a direct estimation of the resulting accuracy using the condition number of the matrix used for the reconstruction. Section 2.2 presents the way, how the extrinsic motion parameters of the cameras are estimated from the mutual observation of the image position of the participating cameras used in our collaborative exploration. We present additionally a solution based on standard perspective cameras that uses a simplified omnidirectional system based on a planar mirror and standard perspective cameras in surgical environments. This setup ensures single viewpoint geometry and allows simultaneous observation of other cameras and the reconstructed objects, which is essential for our approach. In section 2.3, we present a way how to keep the scale of the reconstructed data constant. Motion stereo does not recover an absolute scale for the reconstruction and additional processing is needed to keep the scale constant between consecutive frames of a video sequence. In the result section 3.2, we evaluate the sensitivity of the system to parameter errors showing how this analysis can be used to optimize sensor configuration for a given task. We compare our results with similar systems in section 3.3. We conclude the paper with some final evaluation of our system and discuss our future work in this area.

2 Approach

We present an algorithm performing relative localization of cameras undergoing arbitrary motion combined with simultaneous reconstruction of the environment surrounding them (Fig. 2). Our sensor model for the flying cameras is an omni-

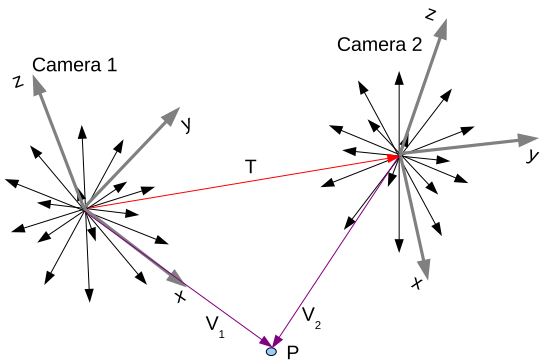

Fig. 2. Collaborative 3D reconstruction from 2 independently moving cameras.

directional system with a large field of view.

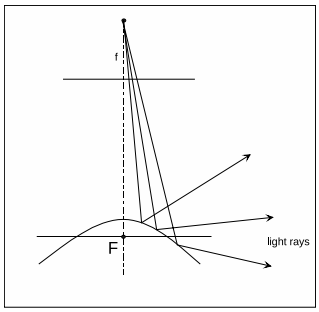

Fig. 3. Single viewpoint property.

We decided to use omnidirectional systems instead of fish-lens cameras, because their single viewpoint property [2] is essential for our combined localization and reconstruction approach (Fig. 3). This property allows an easy recovery of the viewing angle of the virtual camera with the focal point F (Fig. 3) directly from the image coordinates (u_i, ν_i). A standard perspective camera can be mapped on our generic model of an omnidirectional sensor. The only limitation of a standard perspective camera is the limited viewing angle. In case of a standard perspective camera with a focal point at F, we can estimate the direction vector $\boldsymbol{n_i}$ of the incoming rays from the uni-focal image coordinates (focal length of the camera f=1) (u_i, ν_i) to

$$\boldsymbol{n_i} = \frac{(u_i, \nu_i, 1)^T}{||(u_i, \nu_i, 1)^T||}. \tag{2}$$

We rely on the fact that each camera can see the partner and the area that it wants to reconstruct at the same time.

In our system, Camera 1 observes the position of the focal point F of Camera 2 along the vector $\boldsymbol{T}$, and the point P to be reconstructed along the vector $\boldsymbol{V_1}$ simultaneously (Fig. 2). The second camera (Camera 2) uses its own coordinate frame to reconstruct the same point P along the vector $\boldsymbol{V_2}$. The point P observed by this camera has modified coordinates [10]:

$$\boldsymbol{V_2} = \mathbf{R} * (\boldsymbol{V_1} + \boldsymbol{T}) \tag{3}$$

Since we cannot rely on any extrinsic calibration, we perform the calibration of the extrinsic parameters directly from the current observation. We need to find the transformation parameters $(\mathbf{R}, \boldsymbol{T})$ in (3) defining the transformation between the coordinate frames of the two cameras. Each camera defines its own coordinate frame.

2.1 3D Reconstruction from Motion Stereo

In our system, the cameras undergo an arbitrary motion $(\mathbf{R}, \boldsymbol{T})$ which results in two independent observations $(\boldsymbol{n_1}, \boldsymbol{n_2})$ of a point P. The equation (3) can be written using (2) as

$$\lambda_2 \boldsymbol{n_2} = \mathbf{R} * (\lambda_1 \boldsymbol{n_1} + \boldsymbol{T}). \tag{4}$$

We need to find the radial distances (λ_1, λ_2) along the incoming rays to estimate the 3D coordinates of the point. We can find it by re-writing (4) to

$$\begin{aligned} (-\mathbf{R}\boldsymbol{n_1},\ \boldsymbol{n_2}) \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} &= \mathbf{R} \cdot \boldsymbol{T} \\ \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = (-\mathbf{R}\boldsymbol{n_1},\ \boldsymbol{n_2})^{-*} \cdot \mathbf{R} \cdot \boldsymbol{T} &= \mathbf{D}^{-*} \cdot \mathbf{R} \cdot \boldsymbol{T} \end{aligned} \tag{5}$$

We use in (5) the pseudo inverse matrix D^{-*} to solve for the two unknown radial distances (λ_1, λ_2). A pseudo-inverse matrix to D can be calculated according to

$$\mathbf{D}^{-*} = (\mathbf{D}^{\mathbf{T}} \cdot \mathbf{D})^{-\mathbf{1}} \cdot \mathbf{D}^{\mathbf{T}}. \tag{6}$$

The pseudo-inverse operation finds a least square approximation satisfying the overdetermined set of three equations with two unknowns (λ_1, λ_2) in (5). Due to calibration and detection errors, the two lines V_1 and V_2 in Fig. 2 do not necessarily intersect. Equation (5) calculates the position of the point along each line closest to the other line.

We notice the similarity between the equations (1) and (5). Equation (1) can be written to solve for the unknown distance Z from the image plane of the coplanar binocular stereo system to:

$$Z = \frac{B \cdot f}{p_x} \cdot \frac{1}{d} = K_s \cdot \frac{1}{d} \tag{7}$$

We see that, as one should expect, the baseline B seems to correspond to the translation vector $\boldsymbol{T}$ which is divided by a disparity value d in binocular system. The "disparity" in our system appears to be the change in the view direction in both cameras $(n_1 \rightarrow n_2)$. The pseudo-inverse in (5) represents the division by d in (7). The rotation matrix $\mathbf{R}$ rotates the direction vector $\boldsymbol{n_1}$ into the coordinate frame of the second camera. We see that the radial distances (λ_1, λ_2) depend on the distance $\boldsymbol{T}$ between the two camera views and is reciprocal proportional to the change of view between them $(\mathbf{R}\boldsymbol{n_1} \rightarrow \boldsymbol{n_2})$. We know from previous work in stereo reconstruction [8] that the accuracy of the 3D reconstruction depends on the configuration between the cameras and the reconstructed point.

Cameras that are close together achieve poor reconstruction accuracy for the reconstructed points, which is reflected in decreasing disparity d values for distant points. It is difficult to talk about disparity for arbitrary motion of the cameras, as it is the case in our system. The sensitivity of the stereo system is encoded in the pseudo-inverse matrix in the equation (5).

The relative error in the solution caused by perturbations of parameters can be estimated from the condition number of the pseudo-inverse matrix $\mathbf{D}^{-*}$. The condition number is the ratio between the largest and the smallest singular value of this matrix [4].

The condition number estimates the sensitivity of the solution of a linear algebraic system to variations of parameters in matrix $\mathbf{D}^{-*}$ and in the measurement vector $\boldsymbol{T}$.

Consider the equation system with perturbations in matrix $\mathbf{D}^{-*}$ and vector $\boldsymbol{T}$:

$$(\mathbf{D}^{-*} + \epsilon\delta\mathbf{D}^{-*})\boldsymbol{T_b} = \lambda + \epsilon\delta\lambda \tag{8}$$

The relative error in the solution caused by perturbations of parameters can be estimated by the following inequality using the condition number κ calculated for $\mathbf{D}^{-*}$ (see [7]):

$$\frac{||T - T_b||}{||T||} \leq \kappa \left(\epsilon \frac{||\delta\mathbf{D}^{-*}||}{||\mathbf{D}^{-*}||} + \epsilon \frac{||\delta\lambda||}{||\lambda||} \right) + \mathcal{O}(\epsilon^2) \tag{9}$$

Therefore, the relative error in solution λ can be as large as condition number times the relative error in $\mathbf{D}^{-*}$ and $\boldsymbol{T}$. The condition number together with the singular values of the matrix $\mathbf{D}^{-*}$ describe the sensitivity of the system to changes in the input parameters. Small condition numbers describe systems with small sensitivity to errors compared to large condition numbers that suggest parallel direction of the measured direction vectors $(\mathbf{n_1}, \mathbf{n_2})$. A quantitative evaluation of this result can be found in Section 3.2. We use the condition number of the matrix D^{-*} as a metric to evaluate the quality of the resulting configuration of the two cameras. We can move the cameras for a given region of interest to a configuration with a small condition number ensuring high accuracy of the reconstructed coordinates.

If we know the extrinsic motion parameters $(\mathbf{R}, \boldsymbol{T})$, e.g. from a calibration process, then equation (5) allows to estimate directly the 3D position of an observed point for each camera using the two observations $(\mathbf{n_1}, \mathbf{n_2})$ from (2). Our system returns the 3D positions in local coordinate frames of each camera using the relation $\boldsymbol{V_i} = \lambda_i \boldsymbol{n_i}$. The sensor system is not responsible for any compensation of errors in the calibration and detection. These errors are compensated in higher level data fusion modules which are outside of the scope of this paper.

2.2 Estimation of Extrinsic Parameters

In our approach, we assume that we do not know the extrinsic motion parameters between the cameras, because each camera is mounted on independently moving agents. Therefore, we need to estimate the extrinsic motion parameters in parallel

to the 3D reconstruction. We assume, that the two cameras can see each other and the reconstructed point simultaneously. The coordinate frames of the two cameras can show significant rotation $\mathbf{R}$ to each other that need to be estimated for a correct estimation of the radial distances (λ_1, λ_2) in Section 2.1. We need to find a fixed coordinate frame that is defined based on the configuration of the cameras to each other and to the reconstructed point. Similar to a standard binocular system, we decided to define the x-axis of the Camera 1 frame along the vector $\boldsymbol{T}$ (Fig. 2). The direction vector $\boldsymbol{n_{c2}}$ in the local coordinate frame of Camera 1 pointing at the focal point F of Camera 2 defines the direction of this axis. This way, we fixed 2 rotation angles but our system can still rotate around the so defined x-axis. We need to define one more direction in space to fix the coordinate frame of our exploration system. We define the z-axis as a vector perpendicular to the x-axis and pointing towards the point P again to keep it similar to a binocular system. These definitions define now uniquely the coordinate frame $(\boldsymbol{e_x}, \boldsymbol{e_y}, \boldsymbol{e_z})$ of the exploration system in the frame of Camera 1 to:

$$\boldsymbol{e_x} = \boldsymbol{n_{c2}}, \quad \boldsymbol{e_y} = \frac{\boldsymbol{n_1} \times \boldsymbol{e_x}}{||\boldsymbol{n_1} \times \boldsymbol{e_x}||}, \quad \boldsymbol{e_z} = \boldsymbol{e_x} \times \boldsymbol{e_y} \tag{10}$$

with $\boldsymbol{n_{c2}}$ representing the direction to the focal point of the second camera in coordinate frame of the first and $\boldsymbol{n_1} = \boldsymbol{V_1}/||\boldsymbol{V_1}||$. The rotation matrix $\mathbf{R_1}$ between the coordinate frame of Camera 1 and the reference frame of the exploration system can be estimated to

$$\mathbf{R_1} = (\boldsymbol{e_x}\ \boldsymbol{e_y}\ \boldsymbol{e_z})\,. \tag{11}$$

Similarly, we calculate the position of reference coordinate frame in the coordinates of Camera 2 (f_x, f_y, f_z) to:

$$\boldsymbol{f_x} = -\boldsymbol{n_{c1}}, \quad \boldsymbol{f_y} = \frac{\boldsymbol{n_2} \times \boldsymbol{f_x}}{||\boldsymbol{n_2} \times \boldsymbol{f_x}||}, \quad \boldsymbol{f_z} = \boldsymbol{f_x} \times \boldsymbol{f_y} \Rightarrow \mathbf{R_2} = (\boldsymbol{f_x}\ \boldsymbol{f_y}\ \boldsymbol{f_z}) \tag{12}$$

In this case to preserve the direction of the x-axis in the exploration system, we have to invert the direction of the observation $\boldsymbol{n_{c1}}$ of the focal point of Camera 1. The vector $\boldsymbol{n_2}$ represents the direction to the reconstructed point P along $\boldsymbol{V_2}$ (Fig. 2).

Now, we are able to define the extrinsic motion parameters to

$$\boldsymbol{T} = \boldsymbol{n_{c2}}, \quad \mathbf{R} = \mathbf{R_2} \cdot \mathbf{R_1}^T \tag{13}$$

Typically for motion stereo, we are not able to find the magnitude of the translation, but we can define the translation $\boldsymbol{T}$ up to an unknown scale m. We will show in Section 2.3 how the scale can be kept at least constant between consecutive frames of a video stream. If the amount of translation is known between the frames then it can be used to scale the vector $\boldsymbol{T}$ close to the true value. Another way to recover the true translation is identification of known distances in the reconstructed data. The entire reconstruction is kept at the same scale, so any metric reference helps to recover the scale of the entire reconstruction.

We assumed above that the two cameras can see each other and the reconstructed object simultaneously. It can easily be achieved with an omnidirectional setup. In our proof of concept system, we used two standard perspective cameras observing the object in a mirror (Fig. 4).

Fig. 4. Test setup to verify the approach with standard cameras for surgerical applications.

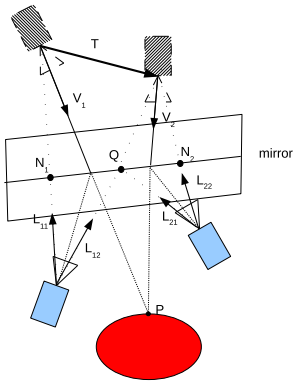

Fig. 5. Geometry of the setup in Fig. 4.

This setup represents a virtual system with the reconstructed object and the cameras themselves being on both sides of the mirror (Fig. 5). The limited field of view of each camera does not allow a direct observation of the other camera and the object simultaneously. According to [2] the simplest omnidirectional system with single viewpoint geometry consists of a planar mirror mounted in front of a perspective camera (Fig. 5). A combination of a planar mirror and a standard perspective camera preserves the single viewpoint geometry.

We can tell from Fig. 5 that although the cameras cannot see each other directly along the vector $\boldsymbol{T}$, the mirror allows to recover the orientation of this vector from the observations $(L_{11}, L_{12}, L_{21}, L_{22})$. The reflection properties of a planar mirror require that each camera can see itself with a direction vector that is perpendicular to the mirror surface. Therefore, L_{11} is the normal vector defined by Camera 1 and L_{22} is the normal vector to the mirror surface defined by Camera 2. The four vectors above intersect the plane of the mirror at points N_1, N_2, and Q. The vectors L_{12}, L_{21} are the directions at which the cameras can see each other. We define rotations of both cameras relative to a common system defined by the normal vector to the mirror as z-axis, x-axis along the points N_1, Q, N_2 and the

y-axis completing the right-hand coordinate frame. From this, we can establish the rotation matrices of the cameras relative to the mirror as

$$\begin{aligned} \boldsymbol{e_z} &= \boldsymbol{L}_{11}, \quad \boldsymbol{e_y} = \tfrac{\boldsymbol{e_z} \times \boldsymbol{L}_{12}}{||\boldsymbol{e_z} \times \boldsymbol{L}_{12}||}, \quad \boldsymbol{e_x} = \boldsymbol{e_y} \times \boldsymbol{e_z} \Rightarrow \mathbf{R_1} = (\boldsymbol{e_x}\, \boldsymbol{e_y}\, \boldsymbol{e_z}) \\ \boldsymbol{f_z} &= \boldsymbol{L}_{22}, \quad \boldsymbol{f_y} = \tfrac{\boldsymbol{f_z} \times (-\boldsymbol{L}_{21})}{||\boldsymbol{f_z} \times \boldsymbol{L}_{12}||}, \quad \boldsymbol{f_x} = \boldsymbol{f_y} \times \boldsymbol{f_z} \Rightarrow \mathbf{R_2} = (\boldsymbol{f_x}\, \boldsymbol{f_y}\, \boldsymbol{f_z}) \end{aligned} \tag{14}$$

The rotation matrices for single cameras relative to the mirror in (14) allow to find the absolute rotation matrix between the cameras observing the mirror to

$$\mathbf{R} = \mathbf{R_2} \cdot {\mathbf{R_1}}^T \tag{15}$$

We established already in Section 2.1 the relationships between the transformed points to

$$\begin{aligned} \lambda_{21} \boldsymbol{L}_{21} &= \mathbf{R}(\lambda_{11} \boldsymbol{L}_{11} + \boldsymbol{T}) \\ \lambda_{22} \boldsymbol{L}_{22} &= \mathbf{R}(\lambda_{12} \boldsymbol{L}_{12} + \boldsymbol{T}) \\ \lambda_{Q2} \boldsymbol{L}_{21} &= \mathbf{R}(\lambda_{Q1} \boldsymbol{L}_{12} + \boldsymbol{T}) \end{aligned} \tag{16}$$

Each condition above defines 3 equations for each of the three coordinates. From the reflection properties of the mirror, we can write additional conditions on the λ_x-values:

$$\lambda_{21} = \lambda_{Q2} + \lambda_{Q1}, \quad \lambda_{12} = \lambda_{Q2} + \lambda_{Q1} \tag{17}$$

The entire system of equations has the form $M\boldsymbol{\lambda} = 0$, which we solve with Singular Value Decomposition method from the matrix $(M^T M)$ as the vector corresponding to the smallest singular value. The underlying least square approximation compensates for calibration and detection errors in the system.

The solution can be multiplied with any constant value on both sides, therefore, the resulting translation vector $\boldsymbol{T}$ is known only up to a scale.

2.3 Constant Scale over Entire Sequence

Structure from Motion approaches are not able to recover the correct scale without any additional metric reference from the environment. In case of a binocular system, this reference is provided by the known calibrated baseline between the cameras. In our case, an a-priori extrinsic calibration is not possible. We mentioned in section 2.1 that our local coordinate system for a pair-wise camera reconstruction from dynamically configurable systems is defined by the vector connecting the cameras and a single reference point P observed in both cameras that allows to estimate the extrinsic parameters.

Knowing the extrinsic transformation $(\mathbf{R}, \boldsymbol{T})$ for a given moment in time, we are able to reconstruct 3D positions of all corresponding points in both camera images using equation (5). One way, how to keep the scale constant, is to track a reconstructed 3D structure in the world and to ensure a constant size of the distances between the reconstructed points. In this way, the new reconstruction can always be scaled up or down to the original scale used before. This approach requires a continues monitoring of the mutual position of the two flying agents. In our system, we added an additional mode developed by a group in [3] that allows

maintenance of the motion and scale data relative to a reference image without the necessity of continued relative monitoring of the mutual position of the agents. The drawback of that method is a cumbersome and difficult initialization method which can be replaced by our robust 3D estimate described in section 2.1. This mode is necessary in cases, where the required configuration of the cameras causes occlusions between the agents although the target is still in view of both cameras.

Our approach offers a robust initialization method for the system presented in [3]. The original approach relied on an essential method to initialize the 3D structure in the world. Our system gives a more robust initialization method minimizing the image error directly. The limited space of this paper does not allow a detailed description of this part of the system. The recursive approach from [3] is used to maintain the radial distance λ_x.

3 Results

Our flying systems use omnidirectional mirrors like the one depicted in Fig. 6

Fig. 6. Flying agent equipped with an omnidirectional sensor pointing upwards.

We tested the system on several indoor and outdoor sequences with two cameras observing the world through different sized planar mirrors (Fig. 4) using a Linux laptop computer with a 1.2 GHz Pentium Centrino processor. The system was equipped with 1GB RAM and was operating two Firewire cameras with standard PAL resolution of 768x576.

3.1 Accuracy of the Estimation of Extrinsic Parameters

We used the system to estimate the extrinsic motion parameters and achieved results comparable with the extrinsic camera calibration results. We verified the parameters by applying them to the 3D reconstruction process in (5) and achieved measurement accuracy below the resolution of our test system. This reconstruction was in the close range of the system which explains the high

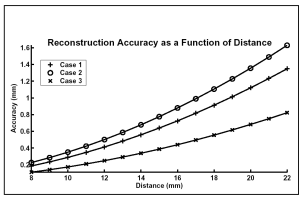

Fig. 7. Expected depth resolution for a standard video camera.

accuracy of the reconstruction (Fig. 7). The main difficulty is the definition of the focal point for the mutual observation of the cameras. We defined the position of the focal point as a center of the boundary circle of the lens. This was justified for the current setup with almost parallel cameras, but needs to be improved for the final submission of the paper.

3.2 Condition Number of the 3D Reconstruction

The collaborative exploration approach allows us to move the cameras to a desired position to establish the best sensitivity and the highest robustness to errors in the parameter estimation. We mentioned already in section 2.1 that the sensitivity of the reconstruction result (5) to parameter errors depends on the relative orientation of the reconstructing cameras (Fig. 8).

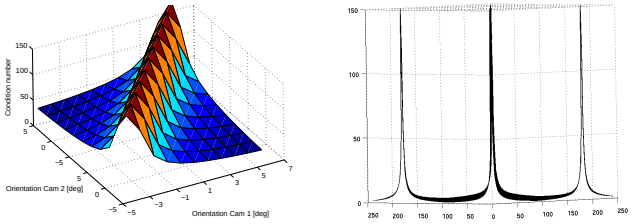

Fig. 8. (left) The condition number of the 3D reconstruction is dependent on the relative orientation of the corresponding direction vectors V_1, V_2 in Fig. 2 observing a point;(right) a minimal condition number corresponds to a difference in the viewing condition between the cameras of 90°.

As we already expected, the collaborative exploration system must avoid camera configurations resulting in similar viewing directions at a given 3D point P in the environment. It achieves best results with the cameras moving close to 90 degrees to each other, where the condition number reaches a minimum (Fig. 8 right). The natural limit is given by the occlusions in the scene that

do not allow to move the cameras too far from each other. From equation (9) we can see that increasing distance to the viewing point increases the absolute error in the estimation of the radial distances λ_i as well. It is a mathematical formulation of an empirical finding that the intersection angle of the rays originating from both cameras need to be close to perpendicular to ensure a highest insensitivity to parameter errors.

3.3 Comparison to Similar Systems

We work currently on better comparisons with similar systems. The system presented in [13] performed with an accuracy of approximately 10cm according to Fig. 9.

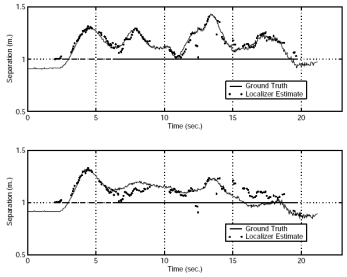

Fig. 9. Accuracy of the system presented in [13].

The accuracy of our system scales directly with the detection accuracy of the features and calibration accuracy of the camera system. We used the CalLab tool from the Matlab Toolbox website to obtain an exact calibration of the cameras. With these calibration values, we were able to verify the motion of the camera with a high accuracy KuKa manipulator with an accuracy below 1mm in an operating range to the object in the range 0.8-1.4m. The error is expected to increase with larger distances to the target, but we are currently looking for ways to verify the accuracy for larger distances.

4 Conclusions and Future Work

We presented a novel formulation for vision-based collaborative exploration systems based on multiple sensing agents with simple sensing capabilities. The limited space of this publication does not allow to explain the details how the configuration is chosen to obtain highest possible reconstruction accuracies with given sensor systems but we focused here on the co-localization and reconstruction part of the system. We showed how a modified 3D reconstruction algorithm

distributed between independent agents can help to build cheaper but powerful exploration systems. In the future work, we plan to work on better next-best view planning using the condition estimates derived in this approach.

Acknowledgments

This work was supported partially by EU IST-FP7 IP project GRASP and by the DFG Cluster "Cognition in Technical Systems" (CoTeSys).

References

1. J. Aspnes, W. Whiteley, and Y. Yang. A Theory of Network Localization. *IEEE Transactions on Mobile Computing*, 5:1663–1678, 2006.
2. S. Baker and S. Nayar. A Theory of Catadioptric Image Formation. *In Proc. of IEEE International Conference on Computer Vision*, 1998.
3. D. Burschka and G. D. Hager. V-GPS(SLAM): – Vision-Based Inertial System for Mobile Robots. In *Proc. of ICRA*, pages 409–415, April 2004.
4. S. Chandrasekaran and I. C. F. Ipsen. On the Sensitivity of Solution Components in Linear Systems of Equations. *SIAM Jounral on Matrix Analysis and Applications*, 16(1):93–112, 1992.
5. T. Clouqueur, V. Phipatanasuphorn, P. Ramanathan, and K. Saluja. Sensor deployment strategy for detection of targets traversing a region. *Mob. Netw. Appl.*, 8(4):453–461, 2003.
6. A. J. Davison. Real-Time Simultaneous Localisation and Mapping with a Single Camera. *Proc Internation Conference on Computer Vision*, 2, 2003.
7. J. Demmel. The componentwise distance to the nearest singular matrix. *SIAM Journal on Matrix Analysis and Applications*, 13:10–19, 1992.
8. O. D. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, 1993.
9. U. Gopinathan, D. Brady, and N. Pitsianis. Coded apertures for efficient pyroelectric motion tracking. *Optics Expressi*, 11(18):2142–2152, 2003.
10. Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An Invitation to 3D Vision: From Images to Geometric Models*. Springer Verlag, 2003.
11. D. Moore, J. Leonard, D. Rus, and S. Teller. Robust distributed network localization with noisy range measurements. In *SenSys '04: Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 50–61, New York, NY, USA, 2004. ACM.
12. P. Potuluri, U. Gopinathan, J. Adleman, and D. Brady. Lensless sensor system using a reference structure. *Optics Express*, 11(8):965–974, 2003.
13. J. Spletzer, A. Das, R. Fierro, C. Taylor, V. Kumar, and J. Ostrowski. Cooperative localization and control for multi-robot manipulation. *IROS*, pages 631–636 vol.2, 2001.
14. D. Tian and N. Georganas. A coverage-preserving node scheduling scheme for large wireless sensor networks. *ACM International Workshop on Wireless Sensor Networks and Applications*, 2002.
15. Y. Zheng, D. Brady, M. Sullivan, and B. Guenther. Fiber optical localization by geometric space coding with 2d gray code. *Appl. Optics*, 44(20):4306–4314, 2005.