SETUP AND CALIBRATION OF A DISTRIBUTED CAMERA SYSTEM FOR SURVEILLANCE OF LABORATORY SPACE[1]

M. Eggers[2], V. Dikov[3], C. Steger[3], B. Radig[2]

[2] Intelligent Autonomous Systems Group (IAS), Technische Universität München (TUM), {eggers,radig}@in.tum.de
[3] MVTec Software GmbH, {dikov, steger}@mvtec.com

This paper describes the setup and realization of a distributed camera system designed to survey a laboratory area where humans and mobile manipulators collaborate jointly. The system consists of 40 industrial grade cameras surveying a 10m by 10m area from a top-down perspective, connected via Gigabit Ethernet (GigE) to a cluster of 40 computers for distributed image processing. Cameras were fully calibrated, achieving an average reprojection error of 0.13 pixels for the complete system, which exceeds state-of-the art accuracy. Current long-term testing has the system running with at least 99.994% availability for up to two weeks. Successful application tests of the system were conducted, where it was used to track the movements of robots and humans across the surveyed area.

Introduction

In areas where they can be installed, global sensor arrays can greatly facilitate joint collaboration between humans and mobile robots capable of manipulation, since they serve to extend the perception of the robot beyond the limitations of its own platform. For this purpose, a camera system capable of supporting long-term collaboration experiments with human and robot participants was developed at the CoTeSys Central Robotics Laboratory (CCRL) in Munich. This paper focuses on the technical setup and calibration of the system and elucidates the decisions taken during system design based on the requirements and initial conditions.

Initial Conditions

Fig. 1 depicts the experimental area the described camera system aims to survey. The area is 10m × 10m wide and 4m high, divided halfway by a wall 2.5m high. A metal scaffolding has been mounted on the ceiling at a height of 3.2m above the floor, to attach the cameras as well as several other sensors.

The main purpose of the assembled camera system is the continuous real-time surveillance of the aforementioned experimental area over extended periods of time. Most importantly, positions of humans and robots operating in the experimental area must be tracked in order to enable robust and safe navigation for the robots in the presence of humans.

Fig. 1. Surveyed experimental area

Camera Installation

A number of decisions had to be taken regarding the number and type of cameras to be used, as well as the positioning of the cameras. Since one of the main objectives of the system is to survey the entire experimental area without any gaps, it was decided to set up the cameras in a way to minimize occlusion by having them face top-down at the experimental area.

The number of required cameras for this kind of setup depends on the field of view of the cameras and lenses used, and the positioning of the cameras. With the maximum height of the cameras h_c given as 3.2m above the floor by initial constraints, the relationship between the camera angle α and the covered floor distance in the primary direction d_x in is as follows:

[1] Research supported by DFG as CoTeSys Project 410/SP5

$$d_x = 2 \times \tan\frac{\alpha}{2} \times h_c \qquad (1)$$

To reliably survey humans and robots, a complete (and redundant) coverage of the scene in 1.7m height h_0 is required. This height corresponds to the average height of an adult person [7]. This yields the new formula for the covered distance at h_0:

$$d_x = 2 \times \tan\frac{\alpha}{2} \times (h_c - h_0) \qquad (2)$$

This formula yields a requirement of 5 × 7 cameras to cover the 10m distance in the respective directions. Ultimately, it was decided to use an array of 5 × 8 cameras to cover a slightly larger area in the secondary direction. Fig. 2 depicts the positions of the cameras and their FOVs at h_0 = 1.7m.

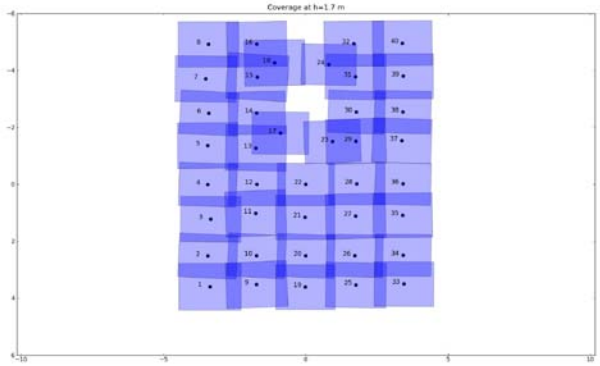

Fig. 2. Camera positions and fields of view.

Network Installation

Real-time processing of the images from a large number of cameras requires a lot of computing power, rendering it desirable for the image processing to be distributed to multiple computers rather than centralized on a single machine. Since a large set of computers requires space and dedicated cooling, it was deemed feasible to set up a server room containing the 40 processing client computers. Since this requires the cameras to be distant from the processing computers, it was opted to use Gigabit-Ethernet (GigE) [1] connected cameras, that can be connected to computers up to 100m away from the sensors themselves.

The network architecture for the camera system is divided into client network and the camera network. The client network interconnects all of the processing clients and the server PC, which handles all centralized processing tasks, via an 48-port GigE switch. Data rates required on this network are not critical, since images do not have to be streamed continuously at high frame rates.

The camera network, on the other hand, connects the processing clients to the cameras themselves, and is essentially not a single network but a set of uniform networks, consisting of only two nodes each. Here, an important factor is the load the network can handle to support continuous streaming of images from cameras to processing clients.

Since GigE-Vision uses User Datagram Protocol (UDP) packets on top of the GigE-Vision Streaming Protocol (GVSP), a data overhead of 46 bytes is created for each data packet, consisting of Ethernet header (14 bytes), IP header (20 bytes), UDP header (8 bytes), GVSP Header (8 bytes) and Ethernet trailer (4 bytes). Consequently, the gross data rate R for a camera can be computed as follows:

$$R = {n_x}^2 \times \frac{1}{A} \times BPP \times f \times \frac{MTU + 18B}{MTU - 36B} \qquad (3)$$

where n_x denotes the number of pixels in the primary image direction, A denotes the aspect ratio, BPP denotes the number of bits used to encode each pixel, f denotes the image frequency (number of images per second) and MTU denotes the size of the maximum transmission unit.

In the described setup, n_x = 1024, A = 4/3, f = 30 Hz and MTU = 9000 bytes (i.e. jumbo frames).

Camera Calibration

Calibration of the camera system was performed using the HALCON [9] software suite, which offers built-in support for multi-camera calibration.

In HALCON, the i-th camera from a setup of N cameras is specified by two sets of parameters: *external* parameters $E_i = (\mathbf{R}_i, \mathbf{T}_i)$ representing the pose of the camera relative to the world coordinate system and *internal* parameters $I_i = (\Pi_i, D_i)$ modeling the projection of 3D points from the camera coordinate system into camera image, where $i = 1 \ldots N$. Π_i describe a standard linear pin-hole camera pro-

[1] Research supported by DFG as CoTeSys Project 410/SP5

jection, whereas D_i define a non-linear radial and decentering distortion using a divisional distortion model [6]. A 3D point $\mathbf{X}$ is transformed into the camera coordinate system $\mathbf{X}_c = E_i \circ \mathbf{X} = \mathbf{R}_i \mathbf{X} + \mathbf{T}_i$ and then is projected in the camera image $\mathbf{p} = \pi(\mathbf{X}_c, I_i)$ (cf. [7,10]).

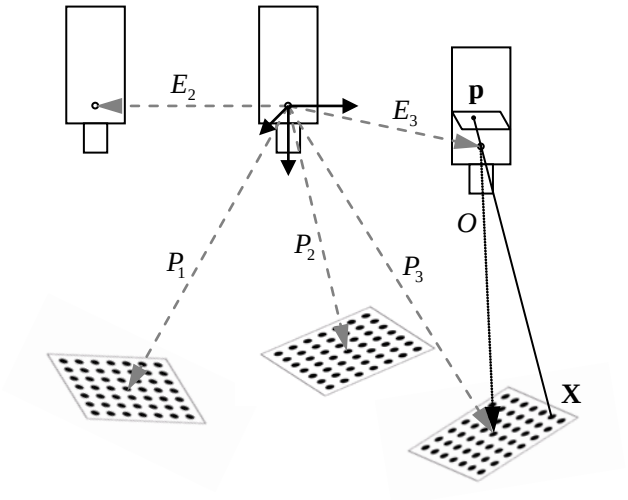

Fig. 3. Multiple camera setup projection model.

A *reference camera* is selected and its coordinate system is the world coordinate system (cf. Fig. 3.). To calibrate the setup, a known calibration object with M control points (marks) is used. Each mark has known coordinates $\mathbf{x}_m$, $m = 1 \ldots M$ in the local coordinates system of the calibration object. The object is exposed in K different poses P_k, $k = 1 \ldots K$, in front of the cameras. Thus the calibration marks define KM control points $\mathbf{X}_{km} = P_k \circ \mathbf{x}_m$ in the world. For each P_k, all cameras simultaneously take an image. Images in which the calibration object is not fully visible are ignored.

In the presented setup, the reference camera is 22 (cf. Fig. 2). To obtain calibration data, a quadratic calibration object with circular marks, measuring 0.425m, was slowly moved across the entire experimental area, while varying height, pitch and roll. A total of K = 9071 synchronized images were recorded, with the calibration plate being observed by up to five cameras at once.

The calibration of multiple cameras is formulated as a minimization problem:

$$d_{ikm} = \left\| \mathbf{p}_{ikm} - \pi(E_i \circ P_k \circ \mathbf{x}_m, I_i) \right\|^2 \qquad (4)$$

$$d = \sum_{i=1}^{N} \sum_{k=1}^{K} v_{ik} \left(\sum_{m=1}^{M} d_{ikm} \right) \rightarrow \min$$

Equation (4) is the *reprojection error* for control point m in pose k into i-th camera image and v_{ik} is 1 if pose k is visible from camera i and 0, otherwise. This is a typical bundle-adjustment problem formulation, in which an estimation for both the camera parameters and the calibration object pose is found.

Since there is no direct solution to this problem, an iterative numerical method is used, which requires initial values for I_i, E_i and P_k. Each I_i is initialized from the product specifications of the camera. Then a pose, in which each camera is observing the calibration object in its own coordinate system, can be estimated. Finally, through a chain of shared observations from different cameras on overlapping calibration object poses, the poses of cameras are transformed into the reference coordinate system and used as initial values for E_i. All poses of the calibration object P_k are similarly transformed.

The optimization is implemented by a general sparse Levenberg-Marquardt (LM) algorithm as described in [3], which scales linearly with the size of the camera setup.

The circular calibration features of the calibration plate projected onto camera images deform to ellipses, whose centers define the corresponding image points $\mathbf{p}_{ikm}$. Note that ellipse centers do not represent precisely the projection of the circular center due to perspective and radial distortions. The distortion of the extracted marks is corrected with the calibrated D_i, their centers $\mathbf{p}_{ikm}$ are re-estimated and perspectively corrected (as proposed in [4]) with the calibrated parameters Π_i. Subsequently, the calibration is performed again with the corrected $\mathbf{p}_{ikm}$ and the calibrated setup parameters as initial values.

The calibration procedure reports the RMS of d as average error:

$$e = \sqrt{\frac{1}{M \sum_{i=1}^{N} \sum_{k=1}^{K} v_{ik}} d} \qquad (5)$$

[1] Research supported by DFG as CoTeSys Project 410/SP5

Calibration Accuracy Comparison

For the calibration performed on the described camera setup, an average reprojection error of 0.13 pixels was achieved, which compares favourably with results achieved for multi-camera calibration by other researchers (cf. Table 1 for details).

Table 1. Calibration Accuracy Comparison

Researcher	*e* in px	Cameras
Pollefeys et al. [8]	0.11 – 0.26	25 – 4
Svoboda et al. [10]	0.2	16
Devarajan et al. [2]	0.59	60 (simul.)
Kurillo et al. [5]	0.153	4
Described system	0.13	40

Several factors contribute to the high accuracy achieved by HALCON. Using centers of circular features as control points provides a robust and accurate method for extracting them in the camera images. Then the adopted non-linear distortion models, both division and polynomial, correct the projection errors efficiently. In particular, re-estimating $\mathbf{p}_{ikm}$ with the calibrated parameters corrects both projective and distortion bias and further improves the information extracted from the projected marks. Finally, defining the calibration as a bundle-adjustment problem yields a geometrically optimal calibration for the entire camera setup, which scales well with respect to the setup size because of the sparse LM optimization algorithm.

Long-Term Testing

For long-term surveillance tasks, the stability and robustness of the system are paramount. So far, the system has been successfully tested operating for periods of up to two weeks.

Table 2. Results of Long-Term Testing

Component	Availability	Downtime (2 weeks)
Server	100%	0s
Client	100%	0s
Camera	99.996%	3.46s
Capturing (SW)	100%	0s
Tracking (SW)	99.994%	5.19s

Table 2 lists the availability of the important system components during the testing period, obtained by sampling the data in 30s intervals. Note that the test period of two weeks well exceeds the usual demands on the system for continued surveillance of human-robot experiments.

Conclusion

In this paper, a camera system tailored for surveillance tasks in a laboratory area was presented, focusing on system setup and calibration. It was demonstrated that the system compares favorably to the state of the art in terms of calibration accuracy and works robustly for extended periods of time.

References

1. AIA. *GigE Vision*: *Camera Interface Standard for Machine Vision*, Version 1.2, January 2010.
2. D. Devarajan, Z. Cheng, and R.J. Radke. *Calibrating distributed camera networks*. Proceedings of the IEEE, 96(10):1625–1639, 2008.
3. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, 2000.
4. J. Heikkilä. *Geometric camera calibration using circular control points.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, Issue: 10, 1066 – 1077, 2000.
5. G. Kurillo, Z. Li, and R. Bajcsy. *Framework for hierarchical calibration of multi-camera systems for teleimmersion*. In: Proceedings of the 2nd International Conference on Immersive Telecommunications, pages 1–6. ICST, 2009.
6. S. Lanser, C. Zierl, and R. Beutlhauser. *Multibildkalibrierung einer CCD-Kamera.* In Mustererkennung 1995, 17. DAGM-Symposium. Springer-Verlag, pp. 481-491, 1995.
7. C.L. Ogden, C.D. Fryar, M.D. Carroll, and K.M. Flegal. *Mean body weight, height, and body mass index, united states 1960–2002. Advance data from vital and health statistics,* (347), 2004.
8. M. Pollefeys, S.N. Sinha, L. Guan, and J.S. Franco. *Multiview Calibration Synchronization and Dynamic Scene Reconstruction.* Multi-camera networks: principles and applications, page 29, 2009.
9. C. Steger, M. Ulrich, and C. Wiedemann, *Machine Vision Algorithms and Applications*, Weinheim: Wiley-VCH, 2008
10. T. Svoboda, D. Martinec, and T. Pajdla. A *convenient multicamera self-calibration for virtual environments.* Presence: Teleoperators & Virtual Environments, 14(4):407–422, 2005.

[1] Research supported by DFG as CoTeSys Project 410/SP5