



Institute of Internet and Intelligent Technologies
 Vilnius Gediminas Technical University
 Saulėtekio al. 11, 10223 Vilnius, Lithuania
<http://www.isarc2008.vgtu.lt/>

**The 25th International Symposium
 on Automation and Robotics in Construction**

June 26–29, 2008

ISARC-2008

**THE ROLE OF CAMERA NETWORKS IN
 CONSTRUCTION AUTOMATION**

Itai Katz

National Institute of Standards
 and Technology
 100 Bureau Drive,
 Gaithersburg, MD 20899-8611
itai.katz@nist.gov

Kamel Saidi

National Institute of Standards
 and Technology
 100 Bureau Drive,
 Gaithersburg, MD 20899-8611
kamel.saidi@nist.gov

Alan Lytle

National Institute of Standards
 and Technology
 100 Bureau Drive,
 Gaithersburg, MD 20899-8611
alan.lytle@nist.gov

ABSTRACT

Due to the complexity and variability of capital construction sites, real-time monitoring remains an elusive goal. As unit price continues to fall, camera networks are becoming increasingly viable sensing modalities. The benefits of combining multiple, overlapping image sources include the ability to produce 3D scene reconstruction as well as localize and track objects. In the construction domain, these benefits enhance situation awareness by facilitating tracking of construction equipment and material, generating progress reports, and providing automated alerts in the event of a safety hazard. In this paper we describe the use of a camera network with a discussion of practical considerations in calibration and tracking. We conclude with a demonstration of a usage example.

KEYWORDS

Construction automation, calibrated camera network, situation awareness

1. INTRODUCTION

1.1. The problem

Maintaining situation awareness on a construction site is a key component of safe and efficient operation. In other goods-producing industries, such as manufacturing, situation awareness can be achieved by systematically controlling the work environment. In contrast, the dynamic nature of a construction site does not easily lend itself to a similar degree of control. This lack of structure,

coupled with a limited information flow typical of construction sites results in a high probability of error. Particularly in the case of large capital projects, decision-making based on imprecise information can impose serious liabilities with respect to safety, scheduling, and cost. Critical information to aid the decision-making process may include:

- Locations of personnel and equipment on the site
- Locations of construction and manufactured components on the site

Level of completion for pending tasks

Areas having potential safety hazards

1.2. Camera networks as a solution

A number of manufacturers¹ produce video cameras designed specifically for construction site monitoring. When combined with digital networking, a collection of these devices allows for views from geographically disparate regions to be centralized. This feature could be of great benefit for large sites where continuous in-person monitoring is infeasible. One drawback of this system is that the effort required by the system operator increases with the number of video feeds to be monitored. Situation awareness may be hindered when large numbers of cameras are involved. While remote sensing increases the quantity of available information, it may not necessarily lead to better decision making. As visual clutter increases, the ability to discriminate interesting or anomalous events decreases [1].

In this paper we propose a solution to this problem by considering the use of *calibrated* camera networks. Rather than providing raw video feeds, calibrated camera networks provide a method to abstract the data into a form that is more readily usable by the system operator. While these systems have been studied extensively in the academic community, their use in the construction industry is novel.

1.3. Outline

In this paper we introduce the concept of the calibrated camera network and describe its potential role in the construction domain. Section 2 provides an overview of these systems and illustrates how they could potentially be applied to the construction domain. Section 3 discusses practical considerations

¹ Certain trade names and company products are mentioned in the text or identified in an illustration in order to adequately specify the experimental procedure and equipment used. In no case does such an identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products are necessarily the best available for the purpose.

when designing a camera network, including calibration and synchronization issues. Section 4 demonstrates the setup and testing of a camera network and reports on experimentally-derived error for 3D localization.

2. CALIBRATED CAMERA NETWORKS IN CONSTRUCTION

2.1. System description

A calibrated camera network is a collection of three or more vision sensors whose positions and orientations in 3D space are known. In contrast to a collection of ordinary cameras, calibrated camera networks have two key differences:

Each node in the network is a "smart" camera that combines an image sensor, on-board computing, and the capability to network with other nodes (typically wirelessly). This allows each node to independently carry out low-level image processing tasks (e.g., background subtraction) or more sophisticated vision algorithms (e.g., pedestrian detection). By communicating only the output of these algorithms, rather than the raw video streams, power and bandwidth are conserved.

Each camera is calibrated with respect to a common coordinate system (i.e., the pose and internal parameters of the cameras are predetermined). This step, performed before the nominal operating mode, allows the 2D image planes to be combined to recover 3D structures.

2.2. Benefits over existing sensors

An example of such a system is shown in Figure 1. In principle, a calibrated camera network operates as a 3D imaging system and can be used to detect and track objects in 3D, and to perform geometric reconstruction. Having multiple sensors operating collaboratively yields several advantages over single-sensor range measurement devices:

Robustness to noise: Individual localization errors can be averaged together. Localization error can be reduced by incorporating additional cameras.

Occlusion handling: Camera networks are inherently redundant. By placing cameras around the perimeter of the workspace, occluded objects can still be

tracked if they remain visible in two or more cameras.

Arbitrarily-sized workspace: Adding coverage in a camera network can be achieved by incorporating additional sensors.



Figure 1. Photo of a calibrated camera network of four cameras. The checkerboard pattern in the center is a calibration target.

2.3. Applications

Ultimately, the utility of any sensor on a construction site is determined by how well it aids project managers in the decision making process. The real-time 3D localization capabilities of calibrated camera networks can assist with a number of problems relevant to the construction industry:

- personnel and equipment tracking
- material tracking
- automatic project status tracking
- productivity assessment
- automated safety alerts (for safety and surveillance)

3. PRACTICAL CONSIDERATIONS

3.1. Design parameters

The primary design consideration in developing a camera network is the number of camera nodes to use. This number will depend on the desired resolution, the range of the cameras to the targets, the field of view of the lens, and the camera's position and orientation. Although two cameras are

sufficient to recover 3D points (through a process known as stereoscopy), additional cameras will improve precision and increase the size of the workspace. A discussion of these and other design decisions is described in [2].

3.2. Calibration

In order to link pixel coordinates to corresponding positions in 3D, calibration must be performed. In this phase, values describing each camera's internal characteristics ("intrinsic parameters") are defined, as well as the position and orientation of each camera ("extrinsic parameters"). Intrinsic calibration is performed by taking one or more images of a calibration target (commonly a checkerboard pattern) and applying one of several common algorithms. The resulting values include the camera focal length, CCD center, and lens distortion parameters. Extrinsic calibration is performed by finding points common to multiple images and comparing their projected pixel locations. Calibration is an extensively studied research topic. The interested reader is directed to [3] for a comprehensive list of references.

3.3. Temporal synchronization

Maintaining synchronization between cameras is of critical importance in order to ensure accuracy. Consider the case of tracking a piece of equipment (e.g., a vehicle) moving at a modest 10 m/s. At a camera framerate of 15 frames per second, a discrepancy of one frame between any two cameras corresponds to ≈ 66 ms, yielding a positional error of 0.66m. The required synchronization precision is a function of the speed of the object being tracked. In practice, accurate synchronization is difficult to maintain with software as clocks have a tendency to drift over time, necessitating frequent resynchronizations. A better solution is to use a common signal to provide a hardware trigger to each camera, although this negates the benefit of wireless networking.

4. EXPERIMENT

4.1. Procedure

Despite being an active area of research, to our knowledge there has been no systematic study on the performance of calibrated camera networks. While

comprehensive performance testing is beyond the scope of this paper, the experiment described in this section provides an order of magnitude estimate of localization error derived from empirical testing. This experiment also serves as a tutorial on the setup, calibration, and testing of calibrated camera networks.

Four cameras were arranged in a roughly 12 m x 6 meter rectangle. The camera resolution was set to 1024 pixels x 768 pixels, and the height, angle, and field of view were adjusted to provide a working volume of roughly 4 m x 4 m x 1.2 m. To provide ground truth for the position of the 3D points, a total station was used to measure 12 coplanar points at two different elevations (roughly 0 m and 1.2 m) for a total of 24 points. The model of total station used provides 0.2 mm spatial uncertainty, an estimated order of magnitude better performance than the camera network.

For each observation, a retro reflector was placed at the specified position and the 3D coordinates were obtained with the total station. Each camera recorded an image of the retro reflector in the same position. An example of a resulting image is shown in Figure 2. The pixel position of the retro reflector in each image was then manually determined by visual inspection.

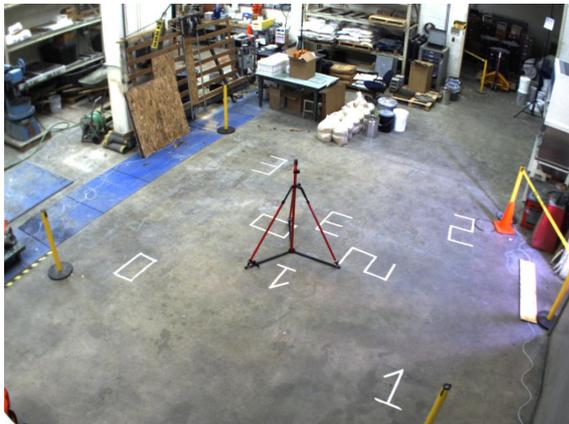


Figure 2. A typical view from one camera. The tripod in the center holds a 3.81 cm (1.5 in) spherically-mounted retroreflector

Intrinsic calibration was performed using the well-known Zhang method [4] as implemented in the MATLAB Camera Calibration Toolbox [5]. The mean focal length was found to be 913($\sigma=10.0$) pixel-lengths, and the average center was found to be 508($\sigma=37.2$) x 421($\sigma=17$) pixels. The POSIT algorithm [6] was used to calibrate the cameras. This algorithm takes as input a list of pixel coordinates corresponding to retro reflector positions along with the associated 3D points as measured with the total station. The output of the algorithm is the position and orientation of the camera.

To recover 3D points from the camera images, a least squares approach was used where a line is projected from the focal point of each camera through the pixel viewing the retro reflector. The reconstructed point is determined to be the point which minimizes the sum of the perpendicular distances to these lines.

To test performance, a subset of the 24 observed points were used for calibration and the remaining points were used to measure error. Two schemes were used for splitting the data between calibration and testing sets. In perimeter calibration, only the eight outermost points of the workspace were used for calibration, while the remaining 16 points were used for testing. In leave-one-out cross-validation (LOOCV), all but one point were used for calibration, and testing was performed on the remaining single point. This test is repeated for each 23/1 split until all observed points have been tested. LOOCV is commonly used to maximize calibration data, when the marginal cost of additional observations is expensive. The error metric used in both schemes is the average Euclidean distance between the reconstructed point and the point observed by the total station. The experiment was repeated three times to estimate statistical variance.

4.2. Results

The error analyses (see Tables 1 and 2) show that the calibrated camera network can resolve points to within 5 cm of ground truth, on average. While this level of performance may be unsuitable for surveying applications, sub-5 cm error is more than sufficient for tracking gross movement of

comparatively large objects such as personnel and equipment.

In addition to repeating the experiment with two calibration schemes, we also explored the impact of the number of cameras on the error. Several trends are evident from the results. As the number of cameras increases, mean error tends to decrease. LOOCV, which uses 23 points for calibration, produces lower error than perimeter calibration, which uses 8 points. This suggests that better performance can be achieved by increasing the number of calibration points, a phenomenon we leave for future study.

It should be noted that these two calibration schemes do not necessarily represent the best possible schemes, but are merely used to illustrate typical results. A synthetic reconstruction of the observed points, based on the calibration parameters, is shown in Figure 3.

Table 1. Error analysis using perimeter calibration

| Perimeter calibration | | | | | | |
|-----------------------|---------------|------------------|---------|----------|---------|----------|
| # cams | Trial 1 | | Trial 2 | | Trial 3 | |
| | μ (cm) | σ (cm) | μ | σ | μ | σ |
| 2 | 4.67 | 1.07 | 3.13 | 1.47 | 4.37 | 1.29 |
| 3 | 3.8 | 0.98 | 1.91 | 0.62 | 2.61 | 0.84 |
| 4 | 3.55 | 0.97 | 1.12 | 0.42 | 2.53 | 0.85 |

Table 2. Error analysis using leave-one-out cross-validation

| Leave-one-out cross-validation | | | | | | |
|--------------------------------|---------------|------------------|---------|----------|---------|----------|
| # cams | Trial 1 | | Trial 2 | | Trial 3 | |
| | μ (cm) | σ (cm) | μ | σ | μ | σ |
| 2 | 3.89 | 4.81 | 3.21 | 1.8 | 3.12 | 4.08 |
| 3 | 2.9 | 2.08 | 1.95 | 1.07 | 2.16 | 2.09 |
| 4 | 2.4 | 1.19 | 1.54 | 1.12 | 1.74 | 1.25 |

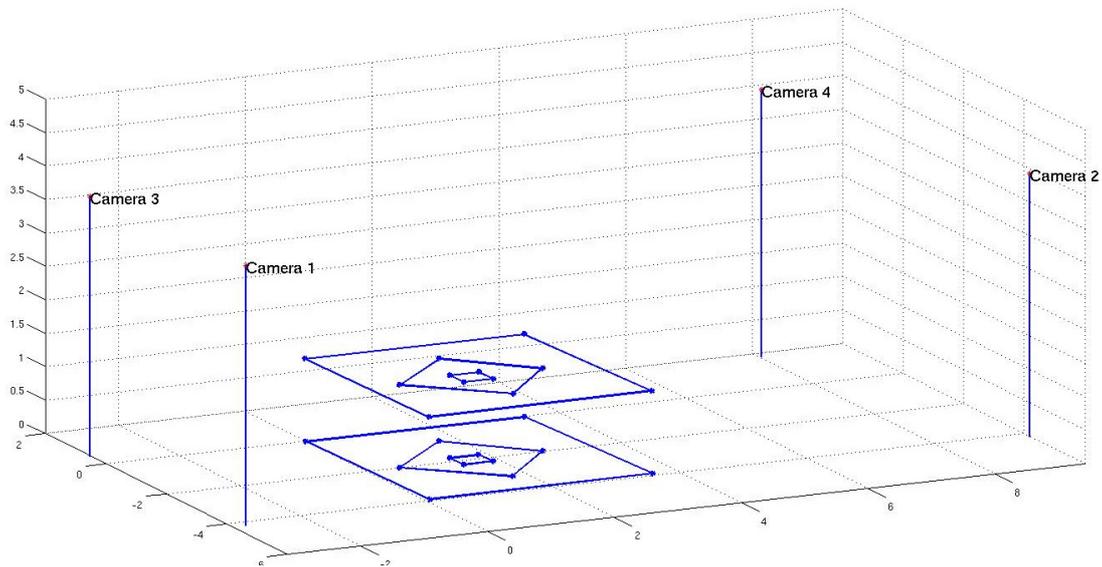


Figure 3. A reconstructed view of the workspace. The observed points are at the corners of the squares. Values are given in meters.

5. CONCLUSION

Situation awareness is a critical factor in maintaining a safe, efficient construction site. As the quantity of information increases, intelligent sensors will become increasingly important. Calibrated camera networks are well suited to this role by providing 3D imaging capabilities that are complementary to existing technologies. In this paper we described the characteristics and design considerations of camera networks, and empirically showed that errors of less than 5 cm are achievable.

As an in-depth experimental verification of calibrated camera networks has largely been unexplored, the experiment described in this paper can be considered a pilot study. Future experiments will characterize the effect of a number of variables on camera network error:

- number of calibration points
- calibration schemes
- number of cameras
- workspace volume
- lighting variations

6. ACKNOWLEDGEMENTS

The authors wish to warmly acknowledge Christoph Witzgall for deriving equations for the least mean squares 3D point reconstruction.

REFERENCES

- [1] Rosenholtz, R., Li, Y. and Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, Vol. 7, No. 2, 17, 1-22.
- [2] Carr, P., Thomas, P. and Hornsey, R. (2005). Performance benefits and limitations of a camera network, *Digital Wireless Communications VII and Space Communication Technologies*, Vol. 5819, No. 1, 328-339.
- [3] Trucco, E., and Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*. N.J.: Prentice Hall.
- [4] Zhang, Z. (2000) A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 22, No. 11, 1330-1334.
- [5] Bouguet, J.-Y. (2001) Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.
- [6] Dementhon, D. and Davis, L. (1995) Model-based object pose in 25 lines of code, *International Journal of Computer Vision*, Vol. 15, No. 1-2, 123-141.