Multiple Camera Calibration with Bundled Optimization using Silhouette Geometry Constraints

Hirotake Yamazoe Akira Utsumi Shinji Abe
ATR Intelligent Robotics and Communication Laboratories
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan
{*yamazoe, utsumi, sabe*}*@atr.jp*

Abstract

We propose a method of calibrating multiple camera systems that operates by adjusting the camera parameters and the 3D shape of objects onto silhouette observations. Our method employs frontier points, which are geometrically meaningful points on object surfaces, to determine the geometrical relations among multiple cameras. In contrast to conventional methods, both camera parameters and the 3D positions of the frontier points are jointly estimated by minimizing the 2D projection errors between the 2D projected positions of the frontier points and observed silhouette contours for all cameras. This method makes it possible to obtain accurate calibration results without using any special instruments. Experimental results using real image data demonstrate the effectiveness of our method.

1. Introduction

Compared with systems based on one or two viewpoints, vision systems based on multiple cameras have many advantages, such as the ability to observe broad spaces, fewer occlusion problems, and higher tracking/3D reconstruction accuracy, etc. Therefore, many systems employing multiple cameras have been proposed, including surveillance systems, 3D modeling systems, and human-computer interfaces [1, 3–5, 10, 13].

To receive the maximum benefit from the multiple camera system, some researchers employ a very large number of cameras. For example, Kanade et al. proposed a 3D modeling system comprising 49 synchronous observations [5]. Matsuyama et al. [7] and Tomiyama et al. [10] also proposed a 3D modeling system using a "visual hull" observed with many cameras.

In such systems, however, it is difficult and time-consuming to calibrate the cameras, especially when the number of viewpoints (number of cameras) is large. Accordingly, automatic camera calibration must be a key technology for multiple camera systems to be practicable.

In open-air conditions, the number of valid (stable) image features that can be used for automatic calibration is very limited. From this perspective, for instance, Lee et al. have proposed a method to determine geometrical relations among multiple cameras [6] that requires coplanar image features (pedestrian trajectories) and a nonlinear algorithm for the calculations.

On the other hand, regarding indoor scenes (especially in a video studio), we can extract more features from a target object image. Some researchers have proposed a camera calibration method that uses silhouette images [2, 8, 9]. For example, Sinha et al. proposed a RANSAC-based algorithm [9] in which the first epipolar geometry (fundamental matrices and frontier points) is estimated for all camera pairs. These results are then integrated and optimized by bundle adjustment. In that method, however, frontier points themselves are not the targets of the optimization. Boyer proposed a calibration algorithm based on pairwise tangency constraints [2]. This method considers constraints on entire silhouette contours and minimizes them to calibrate multiple cameras. However, in silhouette contours, only the frontier points are geometrically meaningful points; therefore, this method runs the risk of becoming unstable during the optimization process.

In this paper, we propose a camera calibration method in which the camera parameters and the 3D shapes of targets are simultaneously adjusted to the observed silhouette shapes. Here, the 3D shapes are represented by frontier points. The camera parameters of all cameras can be estimated by directly minimizing the 2D projection errors between the projected frontier points and observed silhouette contours. The concept of the proposed method is based on bundle adjustment [11], thus it is different from other silhouette-based algorithms because it directly adjusts 3D object shapes to observed silhouette shapes. Our method makes it easy to calibrate a multiple camera system.

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006

The next section contains an explanation of the geometrical constraints used in our method. Section 3 describes our algorithm, which calibrates a multiple camera system based on silhouette images. In Section 4 we present the experimental results and conclude the paper in Section 5.

2. Silhouette Geometry Constraints

2.1. Silhouette and Frontier Points

First, we describe frontier points and their characteristics as employed in our method.

In 3D reconstruction based on a visual hull, the reconstructed shape is generally larger than the actual one; the reconstructed shape only matches the actual shape at the frontier points. Therefore, we employ frontier points as the representation of the 3D shape and adjust the geometrical relations between 3D shapes and observed silhouettes by using the properties of the frontier points

Here, let us consider the situation where a target is observed with cameras C_k and C_l. Frontier points are hypothetical locations on the target surface projected onto the silhouette contour in C_k and C_l.

Figure 1 shows these relations. Frontier point $\boldsymbol{P}_f$ is projected onto silhouette contours $\boldsymbol{S}_k$ and $\boldsymbol{S}_l$ at $\boldsymbol{x}_i^{(k)}$ and $\boldsymbol{x}_i^{(l)}$.

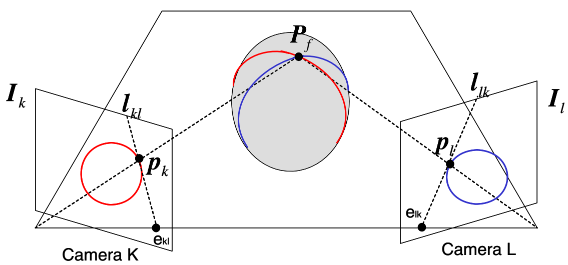

Figure 1. Silhouette and frontier points

2.2 Constraints among Cameras

Next, based on the relations described above, we explain the geometrical constraints among multiple cameras used in the proposed algorithm.

Let us consider a situation where the target is observed with three cameras C_k, C_l, and C_m. Let S_k, S_l, and S_m be the observed silhouette contours (Fig. 2).

We can obtain the following frontier points: $P_{kl,i}$, the frontier points with cameras C_k and C_l; $P_{km,i}$, the frontier points with cameras C_k and C_m; and $P_{lm,i}$, the frontier points with cameras C_l and C_m.

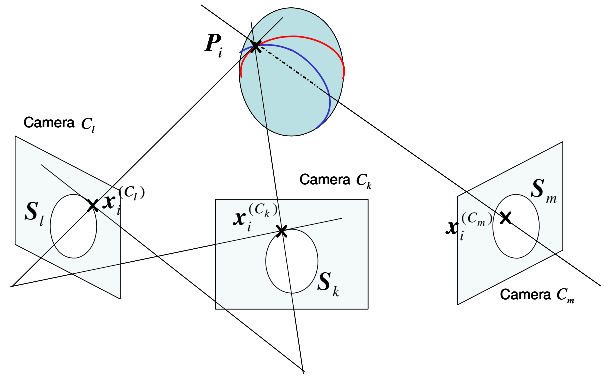

Figure 2. Constraint for frontier points among multiple cameras

Here, let us consider the observations at camera C_k. We categorize the obtained frontier points into two types: the frontier points concerning camera C_k $P_{km,i}$ and $P_{kl,i}$, and other frontier points $P_{lm,i}$.

According to the properties of frontier points, both $P_{kl,i}$ and $P_{km,i}$ should be observed on silhouette contour S_k.

On the other hand, $P_{lm,i}$ is not observed on silhouette contour S_k. However, since $P_{lm,i}$ is the point on the object surface, $P_{lm,i}$ should be visible inside the silhouette.

The proposed method optimizes 3D shape (frontier points) and camera parameters by satisfying these constraints among all cameras. The constraints resemble the those used in [2]. That method considers constraints on entire silhouette contours and minimizes them to calibrate multiple cameras. However, in silhouette contours, only the frontier points are geometrically meaningful points for representing the relationship among cameras. Therefore, the proposed method is straightforward and more efficient.

Finally, we define the 2D distance between silhouette S_k and 2D frontier point $\boldsymbol{x}_i^{(k)}$. First, we calculate the epipole of camera C_l on C_k $\boldsymbol{e}_{kl}$. Then, we determine tangential lines to S_k through epipole $\boldsymbol{e}_{kl}$ and locate a tangent point on the silhouette boundary. The 2D distance $d(\boldsymbol{x}_i^{(k)}, S_k)$ can be calculated as shown in Fig. 3.

The next section explains the details of the proposed method.

3. Bundle Camera Calibration that Minimizes Frontier-Point Projection Errors

3.1. Initialization

This subsection describes the initialization process. First, we calculate the initial values of all camera param-

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006

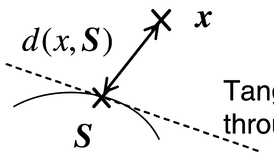

Figure 3. Distance between Silhouette and 2D Frontier Point

eters by using [12], enabling us to estimate the fundamental matrices and the epipoles among all cameras. Candidates for the initial frontier point are obtained by estimating tangential points from epipoles onto the silhouette contours. Then, we can choose initial frontier points that satisfy the property of the frontier points.

3.2. Minimizing Projection Errors

Next, we estimate camera parameters of multiple cameras by minimizing 2D projection errors.

Here, let us consider a situation where the target is observed with N cameras C_1,...,C_N. S_1,...,S_N are obtained silhouettes on cameras C_1,...,C_N, respectively.

We assume M frontier points $\boldsymbol{P}_1, ..., \boldsymbol{P}_M$ are obtained. Here, the frontier point $\boldsymbol{P}_i (i = 1, ..., M)$ is observed by two cameras C_{a_i} and C_{b_i}. Let $\boldsymbol{x}_i^{(C_j)}$ be the 2D observation positions of $\boldsymbol{P}_i$ on cameras C_j.

$$t \begin{bmatrix} x_i^{(C_j)} \\ y_i^{(C_j)} \\ 1 \end{bmatrix} = \boldsymbol{A}_j \left[\boldsymbol{R}_j \ \ \boldsymbol{T}_j\right] \begin{bmatrix} \boldsymbol{P}_i \\ 1 \end{bmatrix}. \quad (1)$$

$\boldsymbol{R}_j$, $\boldsymbol{T}_j$ and $\boldsymbol{A}_j$ are the rotation matrix, the position, and the intrinsic matrix of camera C_j, respectively. $\left(\boldsymbol{A}_j = \begin{bmatrix} f_j & 0 & u \\ 0 & f_j & v \\ 0 & 0 & 1 \end{bmatrix}\right)$.

Here, let us consider the relations between a camera C_j and frontier points $\boldsymbol{P}_1, ..., \boldsymbol{P}_N$.

If $\boldsymbol{P}_i$ is the frontier point directly observed by C_j, 2D distance $D_j(\boldsymbol{x}_i^{(C_j)})$ between the projected frontier point $\boldsymbol{x}_i^{(C_j)}$ and the silhouette S_j can be defined as $d(\boldsymbol{x}, S)$ described in Section 2.2.

$$D_i^{(C_j)} = d(\boldsymbol{x}_i^{(C_j)}, S_j). \quad (2)$$

On the other hand, if $\boldsymbol{P}_i$ is the frontier point defined by other cameras bur C_j, the 2D distance $D_j(\boldsymbol{x}_i^{(C_j)})$ can be defined as follows: when $\boldsymbol{x}_i^{(C_j)}$ is inside the silhouette S_j, we set $D_j(\boldsymbol{x}_i^{(C_j)})$ as zero. Otherwise, $D_j(\boldsymbol{x}_i^{(C_j)})$ is the minimum Euclidian distance $D_{min}(\boldsymbol{x}_i^{(C_j)}, S_j)$ between $\boldsymbol{x}_i^{(C_j)}$ and S_j.

$$D_i^{(C_j)} = \begin{cases} 0 & \boldsymbol{x}_i^{(C_j)} \in S_j \\ D_{min}(\boldsymbol{x}_i^{(C_j)}, S_j) & otherwise \end{cases} \quad (3)$$

We then estimate the optimal camera parameters by minimizing constraints regarding all cameras as follows:

$$D_{all} = \sum_{i=1}^{M} \sum_{j=1}^{N} D_i^{(C_j)} \rightarrow \min . \quad (4)$$

$$(\boldsymbol{P}, \boldsymbol{A}, \boldsymbol{R}, \boldsymbol{T}) = \arg \min_{\boldsymbol{P},\boldsymbol{A},\boldsymbol{R},\boldsymbol{T}} D_{all} \quad (5)$$

4. Experiments

To evaluate the proposed method's effectiveness, we performed experiments using real images.

Four cameras (cameras 1-4) were arranged in the scene. They were all roughly calibrated by [12] in advance, and initial frontier points were extracted by using initial calibration results. Figure 4 shows the experimental environment. All cameras captured ten images for about one second and performed silhouette extraction by color. Figure 5 shows the input images.

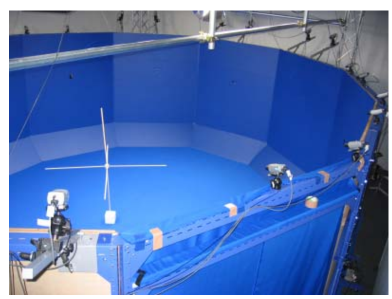

Figure 4. Experimental environment

All cameras were subsequently calibrated using our method. In this experiment, we estimated the positions and poses $\boldsymbol{T}$ and $\boldsymbol{R}$ as well as the focal length f of all cameras. Figure 6 illustrates the calibration results (red points are the frontier points, which should be visible on the silhouette contours; the green ones are other frontier points.) Figure 7 shows the summation of the projection errors of all cameras.

As the figure shows, projection errors decrease with an increase of iterations, indicating that our technique, makes it easy to calibrate systems based on multiple cameras.

0-7695-2521-0/06 $20.00

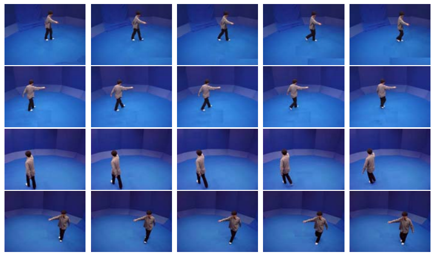

Figure 5. Input images

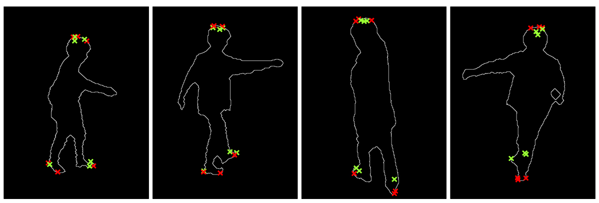

Figure 6. Silhouette images and estimated frontier points (Red: frontier points, which should be visible on the silhouette contours. Green: other frontier points.)

5. Conclusion

In this paper, we proposed a camera calibration method that operates by adjusting the camera parameters and 3D shapes of targets onto the observed silhouette shapes. In our method, we employed frontier points and their characteristics for calibration, and optimal camera parameters were estimated by minimizing the projection errors between the 2D positions of the frontier points and the observed silhouette shapes. Experimental results using real images revealed the effectiveness of our method. The proposed method makes it possible to obtain accurate calibration results without any special calibration instruments.

Future works will involve evaluating the system's robustness and integrating the initialization process.

Acknowledgement

This research was supported in part by the Ministry of Education, Culture, Sports, Science and Technology. under the "Development of fundamental software technologies for digital archives" project.

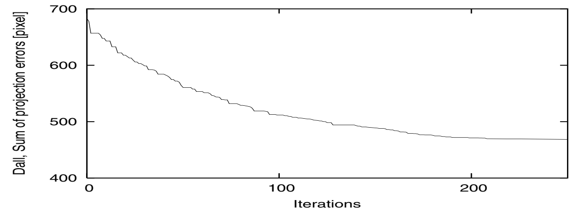

Figure 7. Projection errors

References

[1] D. Arita, R. Taniguchi, S. Yonemoto, and Y. Hamada. A real-time multi-view image processing system on PC cluster. In *Proc. of ACCV2000*, pages 270–275, 2000.

[2] E. Boyer. Camera calibration using silhouettes. Technical Report RR-5559, INRIA, 2005.

[3] D. M. Gavrila and L. S. Davis. 3-d model-based tracking of humans in action: a multi-view approach. In *Proc. of CVPR*, pages 73–80, 1996.

[4] H.Kato and M.Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proc. of the 2nd International Workshop on Augmented Reality(IWAR 99)*, 1999.

[5] T. Kanade, H. Saito, and S. Vedula. The 3D room: Digitizing time-varying 3D events by synchronized multiple video streams. *CMU-RI-TR-98-34*, 1998.

[6] L. Lee, R. Romano, and G. Stein. Monitoring activities from multiple video streams: Establishing a common coordinate frame. *IEEE Pattern Anal. Machine Intell.*, 22(8):758–767, 2000.

[7] T. Matsuyama, X. Wu, T. Takai, and S. Nobuhara. Real-time 3d shape reconstruction, dynamic 3d mesh deformation, and high fidelity visualization for 3d video. *Computer Vision and Image Understanding*, 96:393–434, 2004.

[8] P. Ramanathan, E. Steinbach, and B. Girod. Silhouette-based multiple-view camera calibration. In *Proc. of Vision, Modeling and Visualization 2000*, volume 1, pages 3–10, 2000.

[9] S. N. Sinha, M. Pollefeys, and L. McMillan. Camera network calibration from dynamic silhouettes. In *Proc. of CVPR2004*, volume 1, pages 195–202, 2004.

[10] K. Tomiyama, Y. Orihara, M. Katayama, and Y. Iwadate. Algorithm for dynamic 3D object generation from multi-viewpoint images. *Proc. of SPIE*, 5599, 2004.

[11] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle Adjustment – A Modern Synthesis. In *Vision Algorithms: Theory & Practice, LNCS 1883*, pages 298–372, 2000.

[12] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the shelf tv cameras and lenses. *IEEE J. Robotics and Automation*, 3(4):323–344, 1987.

[13] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Pattern Anal. Machine Intell.*, 19:780–785, 1997.