Autocalibration from Tracks of Walking People

Nils Krahnstoever   Paulo R. S. Mendonça

GE Global Research, One Research Circle, Niskayuna, NY
E-Mail: {krahnsto|mendonca}@crd.ge.com

Abstract

It has been shown that under a small number of assumptions, observations of people can be used to obtain metric calibration information of a camera, which is particularly useful for surveillance applications. However, previous work had to exclude the common criticial configuration of the camera's principal point falling on the horizon line and very long focal lengths, both of which occur commonly in practise. Due to noise, the quality of the calibration quickly degrades at and in the vicinity of these configurations. This paper provides a robust solution to this problem by incorporating information about the motion of people into the estimation process. It is shown that under the assumption that people walk with a constant velocity, calibration performance can be improved significantly. In addition to solving the above problem, the incorporation of temporal data also helps to take correlations between subsequent detections into consideration, which leads to an up-front reduction of the noise in the measurements and an overall improvement in auto-calibration performance.

1 Introduction

The usefulness of surveillance systems can be greatly improved if metric information can be extracted from a scene, such as the velocity or size of tracked targets or their proximity to points of interest. To obtain such information, a necessary step is that of *camera calibration*, which can be performed under somewhat controlled conditions and using knowledge of the geometry of the scene [3, 18], or by trading this knowledge for assumptions on the camera motion and the rigidity of the scene [13, 14], in an approach generally known as *camera autocalibration*.

In surveillance systems targeted at tracking people one of the most conspicuous features available in the images are, hopefully, the tracked people themselves. Previous works have demonstrated that the use of people as calibration objects yields, under practical conditions, useful results [11, 12]. A common theme among these works is the modeling of people as vertical sticks walking along a planar surface and observed by a pinhole camera. Under such assumptions the problem of calibration from images of people can be mapped into that of *calibration from vanishing points* [4, 5], a technique which, despite its geometric elegance, is notoriously sensitive to noise. This issue has been tackled by the work in [11], but some difficulties remain. One problem is that of critical configurations for calibration from vanishing points, which is akin to that of critical camera motions

BMVC 2006 doi:10.5244/C.20.12

in autocalibration. A second problem is the use of information provided by the tracking system as a set of independent detections, neglecting the continuity of the tracks. This work demonstrates how to incorporate continuity constraints on human motion for the latter problem, which also yields a solution to the former.

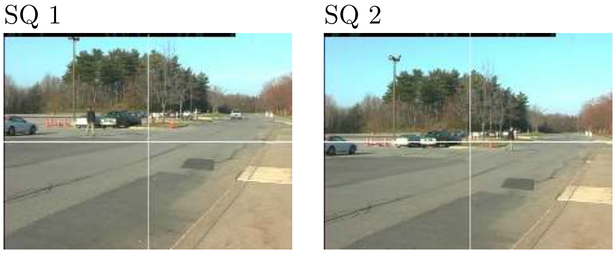

Figure 1: **Nearly Horizontal Camera Views.** *Two camera views that approach zero tilt angles, a critical configuration for autocalibration from people detections. The left view has a tilt angle of about* 4.5 ± 0.7 *degrees. The right camera has a tilt angle of about* 1.6 ± 1.1 *degrees.*

The difficulty with calibration from vanishing points is due to its reliance on the intersection of lines that are, in practice, nearly parallel. For a camera with zero skew and unit aspect ratio, the principal point will fall on the orthocenter of the triangle formed by vanishing points corresponding to three orthogonal directions [4]. Furthermore, the square of the focal length will be given by the distance between such orthocenter and an arbitrary side of the triangle multiplied by the distance between the orthocenter and the vertex opposite to the chosen side [4]. For a pinhole camera viewing a ground plane, a possible choice for this arbitrary side is the *horizon* of the ground plane, which is the image of the line at infinity contained in that plane, and the opposite vertex will be the vanishing point corresponding to the image of the point at infinity in the direction orthogonal to the ground plane. If the viewing direction of the camera is parallel to the ground plane (e.g., Fig. 1), the distance between the principal point and the horizon will be zero, and the distance between the principal point and the opposite vertex will be infinite; the focal length will be, therefore, undetermined. For illustration consider the calibration in Figure 2, which is close to a critical configuration. Standard autocalibration significantly overestimates the focal length of the scene, leading to an elongation of trajectories away from the camera, see Figure 2 (left).

This problem cannot be overcome unless further information about the scene or the camera is provided, which, in this work, comes in the form of the assumption of constant velocity for the people being tracked. The approach presented in this paper enables to deal with data as in Figure 1 (right) to obtain the result in Figure 2 (right). An additional benefit of considering tracks of person detections is that filtering can be applied to the observations, resulting in less noisy data. In order to appropriately handle the remaining uncertainty, a Bayesian framework is used to perform the estimation of the relevant camera parameters, taking the noise of the measurements into consideration in a principled way and allowing for the handling of nonlinearities in the motion as well as the observation models.

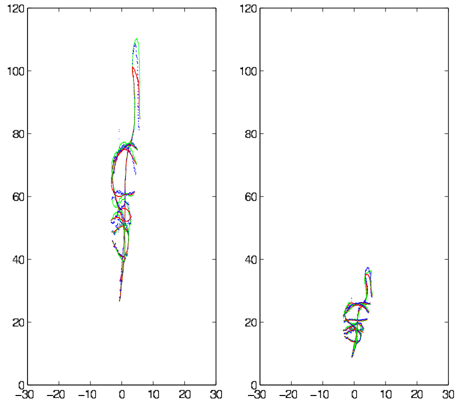

Figure 2: **Challenging Scene.** *Calibration of scene with horizon near the principal point as in Figure 1 (right). Left: Calibration without the use of motion information. The scale in the y-direction is incorrect due to poor estimation of the focal length. Right: Calibration using motion of the target as described in this work. The estimation of the focal length is much more accurate and the scale in y-direction more plausible. Blue dots denote foot detections projected into the ground plane, green tracks denote forward estimated trajectories. Red tracks denote backward filtered MAP trajectories. See section 3 for details.*

2 Related Work

A number of previous works have made use of camera or scene motion models to solve the structure-from-motion (SFM) problem, as well as for camera calibration. A particularly relevant example is the work by Han and Kanade [7], which uses a second-order motion model in conjunction with a projective factorization method to simultaneously recover structure, camera motion and intrinsic parameters in a scene with multiple constant-speed rectilinear motions. Kahl and Heyden [10] demonstrate how the incorporation of a smoothness constraint on the camera motion allows for stable solutions for SFM even in the presence of critical camera motions [17, 19] while avoiding problems caused by local minima and coupling of camera intrinsic and extrinsic parameters, such as focal length and forward motion. The motion model used in that work was simply a first-order Markov process, but, as the authors pointed out, the procedure could be straightforwardly extended to incorporate higher-order models. In [1] the problem is attacked in a recursive manner with the use of an extended Kalman filter and a first-order Markov process. A Bayesian approach with, again, a first-order Markov process, was used in [15].

None of the approaches above is applicable to the case of a static camera, which is the problem tackled in the present work. In this case scene constraints must be used, such as those adopted in [4, 5], but these works cannot cope with the levels on noise present in our data, although [5] does employ a simple covariance propagation technique. The work by Lv et al [12] deals with the particular problem of calibrating a static camera by tracking people, but it is based on an ad-hoc approach that fails to account for the uncertainty in the data. Recently, a novel camera calibration technique based on tracking people was introduced [11]. The method deals with realistic noise levels and provides error bounds for the estimates of the relevant camera parameters. However, it does not make full use of the tracks, which are dealt with as if they were independent detections, and, as with [12], it fails when the viewing direction is parallel to the ground plane.

3 Approach

The goal of this paper is to estimate a metric camera $\mathtt{P}$ from observations of people. To facilitate calibration, people are viewed as vertical calibration sticks of constant height h with a known distribution standing on a flat ground plane. Using standard assumptions on the camera intrinsic parameters (i.e., zero skew and unit aspect ratio) the estimation of the focal length f, the tilt angle θ, the roll angle ρ and the height above the ground plane z yields metric camera calibration. The estimation of $\mathrm{m} = [f, \theta, \rho, z]$ can be obtained from detections of the foot and head locations $\mathrm{y}_j = (\mathrm{y}_j^d, \mathrm{y}_j^u)$ of people in the image, since they allow for the computation of vanishing points [12].

The estimation is cast into a Bayesian framework, by which expected values and covariances from the posterior probability density function (pdf) $p(\mathrm{m}|D,M)$, with $D = \{\mathrm{y}_j, j = 0,\ldots,K-1\}$, could be obtained. Due to the ambiguities that arise for small tilt angles, we would like to incorporate assumptions about the *motion* of people that generated observations D. More specifically, we assume that people tend to either walk at a constant velocity with known distribution or stand still, switching between the two states at will. We denote this assumption as M, completing the description of the terms in the pdf $p(\mathrm{m}|D,M)$. For putative D and M, the 3D trajectories $\mathrm{X} = \{\mathrm{X}_i, i = 0,\ldots,N\}$ of detected people can be obtained. Here, each trajectory consists of 3D body center locations $\{\mathrm{X}_{i,k}\}, k = 0,\ldots,N_i$. We have now that

$$p(\mathrm{m}|D,M) = \int_{\mathrm{X}} p(\mathrm{X},\mathrm{m}|D,M)d\mathrm{X} = \frac{p(\mathrm{m})}{p(D)} \int_{\mathrm{X}} p(\mathrm{X},D|\mathrm{m},M)d\mathrm{X}. \tag{1}$$

Under the assumption that the motion model M is a first order Markov model, the joint probability of the data and the trajectories decomposes into a product of data likelihood and motion terms as follows:

$$p(\mathrm{X},D|\mathrm{m},M) = \prod_{i=0}^{N-1} \left[p(\mathrm{X}_{i,0}|\mathrm{m},M) \prod_{k=1}^{N_i-1} p(\mathrm{X}_{i,k}|\mathrm{X}_{i,k-1},\mathrm{m},M) \prod_{k=0}^{N_i-1} p(\mathrm{z}_{i,k}|\mathrm{X}_{i,k},\mathrm{m},M) \right]. \tag{2}$$

Here, the $\mathrm{z}_{i,k}$ represent the detections y_j after data association and filtering in image space, to be described in section 3.4. The joint probability (2) ties subsequent states together. This unfortunately prevents (1) to decomposes into a product of independent terms making this problem computationally much more challenging than autocalibration approaches based on independent person detections. Ideally, we would want to obtain the estimation of $E[\mathrm{m}]$ under the pdf in (1) or at least it's MAP estimate. Although Monte Carlo Markov Chain (MCMC) integration and sampling / optimization of the marginal is possible, this approach is in practice prohibitively expensive. We hence settle for the joint MAP estimate of $p(\mathrm{X},\mathrm{m}|D,M)$ and take the solution to our problem as

$$(\mathrm{m}^*,\mathrm{X}^*) = \arg\max_{\mathrm{m},\mathrm{X}} p(\mathrm{X},\mathrm{m}|D,M) = \arg\max_{\mathrm{m},\mathrm{X}} p(\mathrm{m})p(\mathrm{X},D|\mathrm{m},M). \tag{3}$$

In the following we discuss the data likelihood terms and motion models of (2).

3.1 Data Likelihood

The likelihood $p(\mathrm{z}_{i,k}|\mathrm{X}_{i,k},\mathrm{m},M) = p(\mathrm{z}_{i,k}|\mathrm{X}_{i,k},\mathrm{m})$ of an observation $\mathrm{z}_{i,k} = (\mathrm{z}^d_{i,k},\mathrm{z}^u_{i,k})$ given a state $\mathrm{X}_{i,k}$, representing a 3D body center, and the parameters m of the camera are obtained as follows. First $\tilde{\mathrm{X}}_{i,k}$, the homogeneous representation of $\mathrm{X}_{i,k}$, is projected into the image as $\tilde{\mathrm{x}}_{i,k} = \mathrm{P}(\mathrm{m})\tilde{\mathrm{X}}_{i,k}$, with Cartesian representation given by $\mathrm{x}_{i,k}$ and $\mathrm{P}(m)$ the projective matrix parameterized by m. The result is then mapped from the 2D projection $\mathrm{x}_{i,k}$ of the body center to the image locations of the foot and head, by center-to-foot and center-to-head *homologies* $\mathrm{H}^d_c(\mathrm{m})$ and $\mathrm{H}^u_c(\mathrm{m})$ [11, 16], $\tilde{\mathrm{x}}^d_{i,k} = \mathrm{H}^d_c(\mathrm{m})\tilde{\mathrm{x}}_{i,k}$ and $\tilde{\mathrm{x}}^u_{i,k} = \mathrm{H}^u_c(\mathrm{m})\tilde{\mathrm{x}}_{i,k}$ where $\tilde{\mathrm{x}}^d_{i,k}$ and $\tilde{\mathrm{x}}^u_{i,k}$ are the image locations of the feet and head of a detected person in homogeneous coordinates, with Cartesian representation given by $\mathrm{x}^d_{i,k}$ and $\mathrm{x}^u_{i,k}$, respectively. These projections are then compared against foot and head locations obtained through filtering and data association (i.e., tracking) of image measurements represented by locations y^d_j and y^u_j with uncertainty estimates $\boldsymbol{\Omega}^d_j$ and $\boldsymbol{\Omega}^u_j$. These filtered observations and their uncertainties (which are on average reduced during the tracking process) are be denoted as $(\mathrm{z}^d_j,\mathrm{z}^u_j,\Sigma^d_j,\Sigma^u_j)$. Note that the correspondence between the raw measurement indices j and the trajectory, time indices i,k is obtained during the data association and tracking step.

To account for effect of shadows, for missed or split detections, and for deviations from the assumptions about people being vertical calibration objects, observations are assumed to arise from three separate sources. The first is the *inlier model*, which is simply the distribution of filtered observations described in section 3.3. The second, occurring with probability P_o, is an *outlier model* corresponding to a wide distribution in the vicinity of the inliers. Finally, a uniform background distribution is selected with probability P_b, with $P_b < P_o \ll 1$. Omitting the corresponding subscripts, the overall model for the foot location is then defined as

$$p(\mathrm{z}^d|\mathrm{x}^d) = (1-P_o-P_b)N(\mathrm{z}^d;\mathrm{x}^d,\Sigma^d) + P_o N(\mathrm{z}^d;\mathrm{x}^d,\Sigma^d_o) + P_b\frac{1}{wh}, \tag{4}$$

where w and h are the width and height of the image and $\boldsymbol{\Sigma}^d_o$ is the variance of the outlier model; the model for head locations is defined analogously. The covariances of the original detections have eigenvectors aligned with the foot to head direction. The covariance $\boldsymbol{\Sigma}^d_o$ is assumed to be aligned with $\boldsymbol{\Sigma}^d$, but it is set to be three times as wide and twice as high as that, whereas $\boldsymbol{\Sigma}^u_o$ is set to be twice as wide and three times as high as $\boldsymbol{\Sigma}^u$, to which it is, in turn, aligned. This definition models the fact that foot location outliers are often caused by horizontal shadows, whereas head location outliers are caused by predominantly horizontal splits in the detections. These parameters are chosen to accommodate the particular person detector used in this work, and will vary for different detectors. Note that the observation model (4) is a mixture of Gaussian plus background pdf, and hence non-Gaussian.

3.2 Motion Model

Our model assumptions M about the motion of detected people have to be encoded into the dynamical model $p(\mathrm{X}_{i,k}|\mathrm{X}_{i,k-1},\mathrm{m},M)$. This is not trivial, since the model has to favour smoothly changing locations and velocities and at the same time favour a specifc velocity magnitude (i.e., the average walk velocity) in the long term. Toward this goal, we represent the state X_k of a particular trajectory point using its location (x_k,y_k) on the ground plane, its heading angle ϕ_k in the ground plane, and the magnitude v_k of the velocity of the tracked subject, $\mathrm{X}_k=(x_k,y_k,\phi_k,v_k)$. For the deterministic component of the model we use a standard constant velocity model $\mathrm{X}_{k|k-1}=(x_{k-1}+v_{k-1}\cos(\phi_{k-1})\Delta t_k, y_{k-1}+v_{k-1}\sin(\phi_{k-1})\Delta t_k, \phi_{k-1}, v_{k-1})$, where $\Delta t_k=t_k-t_{k-1}$. The stochastic component follows a standard white noise process for location and heading. For the velocities, two components representing a walking person and a standing person, govern the stochastic model. The walking component $p^w(v_k|v_{k-1})$ takes the form

$$p^w(v_k|v_{k-1})=cN(v_k;v_{k-1},\sigma_v^2)e^{-\frac{(v_k-\bar{v})^2}{2\sigma_{\bar{v}}^2}}, \tag{5}$$

where the first term is a white noise process governing local changes in speed (with variance σ_v) and the second term penalizes boundless deviations from a finite velocity $\bar{v}$ consistent with human motion. The above equation can be rewritten as

$$p^w(v_k|v_{k-1})=N(v_k;\hat{v}_{k-1},\hat{\sigma}_{k-1}), \text{ where } \hat{v}_{k-1}=\frac{\sigma_{\bar{v}}^2 v_{k-1}+\sigma_v^2\bar{v}}{\sigma_{\bar{v}}^2+\sigma_v^2}, \hat{\sigma}_{k-1}=\frac{\sigma_{\bar{v}}^2\sigma_v^2}{\sigma_{\bar{v}}^2+\sigma_v^2}. \tag{6}$$

One sees that as v_{k-1} deviates from $\bar{v}$, the center of the pdf $p^w(v_k|v_{k-1})$ shifts toward $\bar{v}$. Essentially, velocities are encouraged to diffuse toward the average speed $\bar{v}$.

A second component models the fact that people might stop or have lower speeds during turns. The stopping model has the same form as (5) except that $\bar{v}=0$ and $\sigma_v^2=\sigma_0^2$. The final stochastic model is given by

$$p(v_k|v_{k-1},M)=P_w p^w(v_k|v_{k-1})+(1-P_w)p^s(v_k|v_{k-1}), \tag{7}$$

where P_w is the probability that a given person is walking as opposed to standing still. Our model M is defined by the above structure and the parameters $(\sigma_v^2,\bar{v},\sigma_{\bar{v}}^2,\sigma_0^2,P_w)$. Note that the model allows sampling as well as evaluation of likelihoods, an important property for backward smoothing [9].

3.3 Optimization of Joint PDF

As described above, we need to maximize the joint pdf (3) with respect to the calibration parameters m and the trajectories X. This is a difficult task with respect to X since it in general requires the optimization over all the variables $\mathrm{X}_{i,k}$ simultaneously. In order to make the problem tractable, we perform an iterative coordinate descent with respect to m and X independently, given by

$$\mathrm{X}^{[l]} = \arg\max_{\mathrm{X}} p(\mathrm{X}, D|\mathrm{m}^{[l]}, M), \text{ and} \tag{8}$$

$$\mathrm{m}^{[l+1]} = \arg\max_{\mathrm{m}} p(\mathrm{X}^{[l]}, D|\mathrm{m}, M)p(\mathrm{m}). \tag{9}$$

Calibration Parameters: Optimization over the calibration parameters m under fixed trajectories X, as in (9), has much lower computational cost than an optimization over all variables simultaneously. However, convergence can still be slow, since a change in m will change all of the $\mathrm{z}_{i,k}$, due to $\tilde{\mathrm{x}}_{i,k} = \mathrm{P}\tilde{\mathrm{X}}_{i,k}$, (4) and (2). Therefore an intermediate step is taken between (8) and (9), by which, for the current $\mathrm{m}^{[l]}$, the projections $\mathrm{x}^{[l]}$ of the trajectories $\mathrm{X}^{[l]}$ are obtained and kept fixed during the computation of (9). We then perform the optimization in (9) with fixed *image-based* trajectories, which allows for changes in the actual 3D trajectories during optimization. To obtain the foot and head locations for varying m, we utilize homologies that map from the body center (i.e., the 3D plane at height $h/2$) to the foot and head plane respectively.

Trajectories: The optimization in (8) over the trajectories X under fixed camera geometry m is performed by sequential Monte Carlo filtering [6, 8] on (2), which obtains particle representations of the pdfs (with the superscripts $[l]$ omitted for convenience)

$$p(\mathrm{X}_{i,k}|\mathrm{z}_{i,k:0}, M, \mathrm{m}) = \sum_{n}^{N_s} \pi_{i,k}^{n} \delta(\mathrm{X}_{i,k} - \mathrm{X}_{i,k}^{n}). \tag{10}$$

We take particles representations $\mathrm{X}_{i,k}^{n}$ as discrete set of possible states and use the Viterbi algorithm to obtain the maximum (logarithm) of (2).

3.4 Data Association and Filtering

The tracking stage of the system performs standard track formation and data association, and filtering in image space [2]. Tracked targets are represented by eight dimensional state vectors for the foot and head locations and velocity components $\mathrm{z}_j = (\mathrm{z}_j^d, \mathrm{v}_j^d, \mathrm{z}_j^u, \mathrm{v}_j^d)$. Due to the lack of 3D geometric information, the initial tracking process is performed entirely in 2D image coordinates. The dynamics is modeled by a 2D white noise constant velocity linear Gaussian system driving the evolution of the foot locations, of the head locations, and of the velocities, all independently. However, at each time step the system noise in the dynamics is rescaled according to the height of the target in image coordinates. The observation likelihood is given by the terms in (4). We use particle filters in order to deal with the non-normality of our observation model; however, unlike traditional computer vision uses of particles filters, we perform explicit gating and data association to deal with the presence of multiple targets. Also, the observation likelihood does not need to be evaluated for any image data, but rather only operates on the previously collected foot and head location detections. Alternatively, an extended or unscented Kalman filter could have been utilized for tracking.

The particle filter yields posterior pdfs $p(\mathrm{z}_j|\mathrm{y}_{0:j})$ represented as particle sets. We utilize the mean and variances of these distributions projected into the image and their association to tracks as the filtered detections $\mathrm{z}_{i,k}^d, \mathrm{z}_{i,k}^u, \Sigma_{i,k}^d, \Sigma_{i,k}^u$ for the

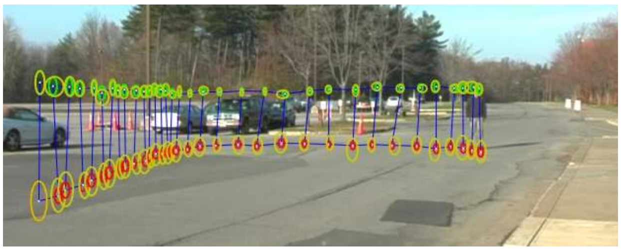

Figure 3: **Foot and Head Tracking.** *An image based particle filter is used to track and filter the foot and head detections (yellow). The tracker yields tracks (blue) of filtered observations (red and green). Only every 20th observation is shown.*

subsequent stages. See Figure 3 for an illustration of the original detections, a corresponding track and the filtered detections.

4 Results and Discussion

We first quantitatively demonstrate the challenge of calibrating scenes with degenerate configurations. Figure 1 shows two views with the principal point close to the horizon. When applying autocalibration on isolated people detections, the focal length estimates become unstable due to the uncertainties in the image measurement.

Table 1 shows the expected calibration parameters, together with their estimated variances for the two scenes in Figure 1, estimated using the approach described in [11]. As one can see, the focal length for SQ1 is already significantly uncertain, while for sequence SQ2 it becomes essentially undetermined. Note that all that was changed between SQ1 and SQ2 is the tilt angle of the camera.

To further quantify the calibration errors, a pattern was drawn on the ground plane to guide the trajectories of the subjects as they walked during video capture. By comparing the estimated trajectories with the location and dimension of the pattern, statements about the calibration accuracy can be made.

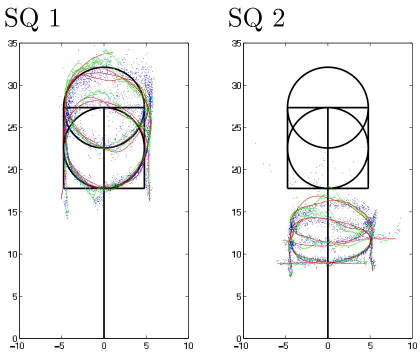

Figure 4: **Autocalibration Without Motion Information.** *This figure shows autocalibration similar to the one in [11], on the camera views in Figure 1. The tracks were visualized for the MAP estimate of the calibration.*

For the sequences SQ1 and SQ2 and the MAP calibration parameters estimated without the use of motion information, we obtain the ground plane projections in Figure 4. One clearly sees that the calibration of SQ1 is quite accurate, while the focal length for sequence SQ2 was underestimated, resulting on a reconstruction of the tracks too close to the camera.

	f	θ/deg	ρ/deg	z/m
SQ 1 Expect.	912 ± 142	4.7 ± 0.67	-0.20 ± 0.12	1.77 ± 0.03
SQ1 MAP	874	4.6	0.54	1.71
SQ2 Expect.	9699 ± 21838	1.6 ± 1.0	0.08 ± 0.15	1.81 ± 0.04
SQ2 MAP	443	3.98	0.59	1.80

Table 1: **Calibration Results Without Motion Information.** *This tables shows the auto-calibration results of the camera views in Figure 1. The first entries for each sequence denote the expectation and variance of the posterior distribution of the calibration parameters.*

By performing calibration with the use of motion information we obtain the results shown in Figure 5. One clearly sees that information about the motion of people has aided the calibration to the extent that good results were obtained even in the degenerate case of SQ2. The optimization of the joint PDF converges after eight iterations. The final calibration parameters for SQ2 are taken to be $\mathbf{m}^* = [923.4 \text{ pixel}, 1.822 \text{ deg}, 0.01511 \text{ deg}, 1.722 \text{ m}]$, which are consistent with the estimates obtained for SQ1 and non-motion based calibration, including the change in the tilt angle.

We further quantify the performance by fitting ellipses to the projected tracks for SQ2 in Figure 4 and Figure 5 and comparing them to the ground truth circles, which had radii of $4.78 \text{ m} \pm 0.05 \text{ m}$. The fitting was performed using a robust least squares approach. Without the use of motion information the ellipses fitted to the estimated trajectories have minor and major radii 5.1 m and 2.3 m, corresponding to an average tarjectory shape error of 28.9%. While the scale along one axis is accurately recovered without the use of motion information, the scale along the other is compressed due to the inaccurate focal length as already observed in Fig. 2. In contrast, the method presented in this work obtains radii 4.4 m and 4.8 m, corresponding to an error of 4.2%.

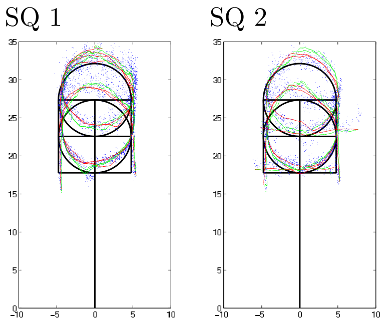

Figure 5: **Autocalibration With Motion Information.** *This figure shows auto-calibration on the camera views in Figure 1 using the approach presented in this paper.*

5 Conclusion

We have incorporated tracking information into the Bayesian estimation of camera calibration parameters from people observation. We have shown that information about the motion of people can be utilized to obtain calibration for scenes that are otherwise intractable for approaches that consider isolated detections of people. The overall calibration approach can deal with a wide variety of scene configurations and can gracefully handle large amount of measurement noise and outliers.

References

[1] A. Azarbayejani and A. Pentland. Recursive estimation of motion, structure and focal length. *IEEE Trans. Pattern Analysis and Machine Intell.*, 17(6):562–575, June 1995.

[2] S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems.* Artech House Publishers, 1999.

[3] D. C. Brown. Close-range camera calibration. *Photogrammetric Eng. and Remote Sensing*, 37(8):855–866, Aug. 1971.

[4] B. Caprile and V. Torre. Vanishing points for camera calibration. *Int. Journal of Computer Vision*, 4:127–139, 1990.

[5] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. In *Proc. 7th Int. Conf. on Computer Vision*, volume I, pages 434–441, Corfu, Greece, Sept. 1999.

[6] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo methods in practice.* Springer, 2001.

[7] M. Han and T. Kanade. Multiple motion scene reconstruction from uncalibrated cameras. In *IEEE Trans. Pattern Analysis and Machine Intell.*, volume 25, pages 884–894, July 2003.

[8] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 1(29):5–28, 1998.

[9] M. Isard and A. Blake. A smoothing filter for condensation. In *Proc 5th European Conf. Computer Vision*, volume 1, pages 767–781, 1998.

[10] F. Kahl and A. Heyden. Euclidean reconstruction and auto-calibration from continuous motion. In *Proc. 8th Int. Conf. on Computer Vision*, volume II, pages 572–577, Vancouver, Canada, July 2001.

[11] N. Krahnstoever and P. Mendonca. Bayesian autocalibration for surveillance. In *Proc. of IEEE International Conference on Computer Vision (ICCV'05), Beijing, China*, October 2005.

[12] F. Lv, Z. Tao, and R. Nevatia. Self-calibration of a camera from video of a walking human. In *Proceedings of International Conference on Pattern Recognition, Quebec City, Quebec, Canada*, August 2002.

[13] S. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. *Int. Journal of Computer Vision*, 8(2):123–151, Aug. 1992.

[14] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown intrinsic camera parameters. *Int. Journal of Computer Vision*, 32(1):7–25, August 1999.

[15] G. Qian and R. Chellappa. Bayesian self-calibration of a moving camera. *Computer Vision and Image Understanding*, 95(3):287–316, Sept. 2004.

[16] J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry.* Oxford Classic Texts in the Physical Sciences. Clarendon Press, Oxford, UK, 1998. Originally published in 1952.

[17] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1100–1105, San Juan, Puerto Rico, June 1997.

[18] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Analysis and Machine Intell.*, 22(11):1330–1334, Nov. 2000.

[19] A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. *Phil. Trans. Royal Soc. London A*, 356(1740):1193–1211, May 1998.