Camera Calibration for Uneven Terrains by Observing Pedestrians

Imran N. Junejo
IRISA/INRIA Rennes, France
imran.junejo@inria.fr

Abstract

A calibrated camera is essential for computer vision systems. The prime reason being that such a camera acts as an angle measuring device. Once the camera is calibrated, applications like 3D reconstruction or metrology or other applications requiring real world information from the video sequences can be envisioned. Motivated by this, we address the problem of calibrating multiple cameras, with an overlapping field of view (FoV), observing pedestrians in a scene walking on an uneven terrain. This problem of calibration from an uneven terrain has so far not been addressed in the vision community. We automatically estimated the infinite homography between the cameras by using the special geometric information obtained from observing pedestrians. This homography provides constraints on the intrinsic (or interior) camera parameters while also enabling us to estimate the extrinsic (or exterior) camera parameters. We test the proposed method on real as well as synthetic data; encouraging results demonstrate the applicability of the proposed method.

1 Introduction

Due to an exponential increase in the computational power of the present day computers, real-time applications of computer vision algorithms has not only been possible, but has acquired a great deal of interest from governments, commercial companies, security agencies and even general public. The area that has attracted the most attention is camera surveillance. Most of the video surveillance systems involves monitoring people (or pedestrians) in a scene. The system can be monitoring, for instance, a building entrance, an airport lobby, stairways, mall escalators or an embassy) using stationary or rotating cameras. The goal for such a system can be to model the behavior of objects (e.g. cars or pedestrians, depending on the situation), event reconstruction, or action recognition etc. In this paper, we present a novel method to calibrate each camera in a system of multiple cameras monitoring a particular area of interest by observing only pedestrians in the scene.

Camera calibration has now become an essential step for any meaningful computer vision based system. For a surveillance system, it is known that due to perspective projection the measurements made from the images do not represent metric data. This is evident from a simple observation: the objects grow larger and move faster as they approach the camera center, or two objects moving in parallel direction seem to converge at a point in the image. The projective camera thus makes it difficult to characterize objects - in terms of their sizes, motion characteristics, length ratios and so on - unless more information is available about the camera being used. Another application for such a surveillance system is making measurements in the image plane i.e. *metrology* [3]. Measurements like a person's height or true walking speed can be estimated easily once the camera is calibrated [2].

Often in real world scenarios, people do not walk on flat surfaces but rather on uneven terrains. For example, people on escalators at the airports or at the shopping malls, people climbing stairs, or walking from road to a sidewalk or on the grass etc. Recently, some researchers have addressed the problem of calibration from observing pedestrians walking on a flat surface (i.e. ground plane), but they fail to handle this general behavior of the objects. However, we do not restrict pedestrian movements to any particular direction or even require a constant velocity.

Due to space limitations, it is not possible to review all auto-camera calibration methods (see [4]). Therefore, we limit ourselves to related work on camera auto-calibration from observing pedestrians. The first work to deal with camera calibration from pedestrians was that of Lv et al. [10]. They recovered the horizon line and the vanishing points from observed walking humans. However, their formulation does not handle robustness issues. Similarly, Krahnstoever and Mendonça [8, 9] proposed a Bayesian approach for auto-calibration by observing pedestrians. Foot-to-head homology is decomposed to extract the vanishing point

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

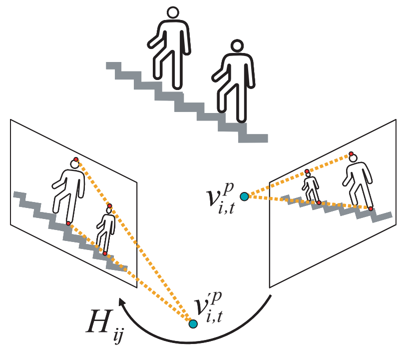


Figure 1. Setup: Top-Bottom locations of two different instances of a walking pedestrians provide a vanishing point in each observing camera. There also exists a homography $\mathbf{H}_{i,j}$ that maps vanishing points between the different cameras.

and the horizon line for calibration. They also incorporate measurement uncertainties and outlier models. However, their method requires prior knowledge about some unknown calibration parameters and prior knowledge about the location of people; and their algorithm is also non-linear. Recently, Junejo and Foroosh [7, 6] presented two methods for camera calibration from pedestrians. They decompose the fundamental matrix induced by different instances of a walking pedestrians to impose constraints on camera intrinsics [7]. However, their method only woks for cases when the pedestrians are walking on a ground plane.

In this paper, we propose a novel solution to the problem of camera calibration when the pedestrians are walking on an uneven terrain. The setup includes multiple cameras looking at the area of interest. We do *not* restrict pedestrians to walk in a certain manner or with a constant velocity. See Fig. 1 for an example of the scenario. The detected top (head) and bottom (feet) locations on a person, over at least two instances, are used to estimate the vanishing points in each view. These vanishing points are then used to estimate infinite homography matrix between the cameras. We estimate three camera parameters i.e. the focal length (f), and the principle point (u_o, v_o). The noise in data points is minimized by using all the vanishing points obtained from all detected people and all the frames. We demonstrate results on synthetic as well as on the real data.

In the next Section, we provide a brief introduction to the concepts related to a pinhole camera. Main method is presented in Section 3, followed by results and conclusion.

2 Background

For a pinhole camera model used in this paper, a 3D point $\mathbf{M} = [X\ Y\ Z\ 1]^T$ and its corresponding image projection $\mathbf{m} = [u\ v\ 1]^T$ are related via a 3×4 matrix $\mathbf{P}$ by

$$\mathbf{m} \sim \underbrace{\mathbf{K}[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3\ \mathbf{t}]}_{\mathbf{P}}\mathbf{M}, \quad \mathbf{K} = \begin{bmatrix} \lambda f & \gamma & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

where $\sim$ indicates equality up to multiplication by a non-zero scale factor, $\mathbf{r}_i$ are the columns of the rotation matrix $\mathbf{R}$, $\mathbf{t}$ is the translation vector, and $\mathbf{K}$ is a nonsingular 3×3 upper triangular matrix known as the camera calibration matrix including five parameters, i.e. the focal length f, the skew γ, the aspect ratio λ and the principal point at $(u_0,\ v_0)$.

The aim of camera calibration is to determine the calibration matrix $\mathbf{K}$. Instead of directly determining $\mathbf{K}$, it is common practice [4] to compute the symmetric matrix $\boldsymbol{\omega} = \mathbf{K}^{-\mathbf{T}}\mathbf{K}^{-1}$ referred to as Image of the Absolute Conic(**IAC**). **IAC** is then decomposed uniquely using the Cholesky Decomposition [12] to obtain $\mathbf{K}$.

3 Method

The scenario that we address in this work is when multiple cameras, having overlapping FoV observe a particular scene, indoors or outdoors, with pedestrians walking in that area of interest. Although the method applies to n number of cameras, for sake of simplicity, we consider the case of two cameras i and j from here on. The various steps involved in the proposed method are: (1) foreground object extraction, (2) estimation of the top and bottom locations, (3) estimation of vanishing points obtained from observing a pedestrian, (4) estimating the infinite homography matrix between different cameras, and (5) camera calibration, i.e. estimating the intrinsic and extrinsic camera parameters. We address these issues likewise.

3.1 Estimating Top and Bottom Locations

A pedestrian needs to be detected and tracked in the video sequence. Note that we are not solving the background subtraction and tracking problem, and therefore use one of the well know methods for performing this task [5]. We do require that top and bottom locations be correctly detected from the tracked pedestrians. In this regard, we adapt the approach proposed by [7] and [8]. Whereas Lv et al. [10] performs eigendecomposition of the detected blob to extract top and bottom locations, we calculate the center of mass and the second order moment of the lower and the upper portion of the bounding box of the foreground region to detect these points for each instance of a walking person seen from each camera (cf. Fig. 1). For a camera i at time instance t, we denote the detected top point for a pedestrian p as $\mathcal{T}^p_{i,t}$ and similarly the bottom point as $\mathcal{B}^p_{i,t}$.

3.2 Estimating The Infinite Homography

Under projective transformation, parallel lines in the word intersect at a point on the image plane, called the vanishing point. While observing a pedestrian of height h at different time instances in a camera, and assuming that h does not change between different instances, the line joining the top-top position of a detected pedestrian at time instance t and $t+1$ respectively, is parallel in the world to the line joining bottom-bottom at the same time instances. This is shown in Fig. 1. Let the intersection of these parallel lines be projected in camera i as: $\boldsymbol{v}^p_{i,t} = \overline{\mathcal{T}^p_{i,t}\mathcal{T}^p_{i,t+1}} \times \overline{\mathcal{B}^p_{i,t}\mathcal{B}^p_{i,t+1}}$, where $t = 1, 2, \ldots,$ and $p = 1, 2, \ldots, k$ persons in a scene. Similarly the top-bottom lines for a pedestrian at two different instances intersect at another vanishing point lying on the plane at infinity, as: $\boldsymbol{v}'^p_{i,t} = \overline{\mathcal{T}^p_{i,t}\mathcal{B}^p_{i,t}} \times \overline{\mathcal{T}^p_{i,t+1}\mathcal{B}^p_{i,t+1}}$. These vanishing points lie on the plane at infinity.

In the situation where people are walking only on a flat surface, as in [7, 8] or [10], the obtained vanishing points $\boldsymbol{v}'$ and $\boldsymbol{v}$ are orthogonal to each other and satisfy the pole-polar relationship with respect to $\boldsymbol{\omega}$. However, this is not the case when the pedestrians are walking on an uneven terrain and hence we cannot use their methods to obtain any constraints on the unknown parameters. We overcome this problem by estimating the *infinite homography* between different cameras viewing the scene. Consider two cameras i and j observing walking pedestrians. The mapping of points from camera i to camera j over the plane at infinity π_∞ is given by

$$\mathbf{H}_{i,j} = \mathbf{K}_j \mathbf{R}_{i,j} \mathbf{K}_i^{-1}, \tag{2}$$

where $\mathbf{R}_{i,j}$ is the relative rotation between the cameras and $\mathbf{H}_{i,j}$, the infinite homography, maps the points lying on the plane at infinity in camera i to their corresponding points lying on the plane at infinity in camera j. In practice, $\mathbf{H}_{i,j}$ is often estimated between two images by matching points. Therefore, a minimum of four point correspondences are necessary to compute $\mathbf{H}_{i,j}$ [4]. However, for a robust solution to $\mathbf{H}_{i,j}$, we use all the estimated vanishing points from the observed pedestrians:

$$\left[\begin{array}{ccc} \boldsymbol{v}^p_{j,t} & \boldsymbol{v}'^p_{j,t} & \ldots \end{array}\right] = \mathbf{H}_{i,j} \left[\begin{array}{ccc} \boldsymbol{v}^p_{i,t} & \boldsymbol{v}'^p_{i,t} & \ldots \end{array}\right] \tag{3}$$

where $\mathbf{H}_{i,j}$ is estimated by the DLT (Direct Linear Transform) algorithm [4].

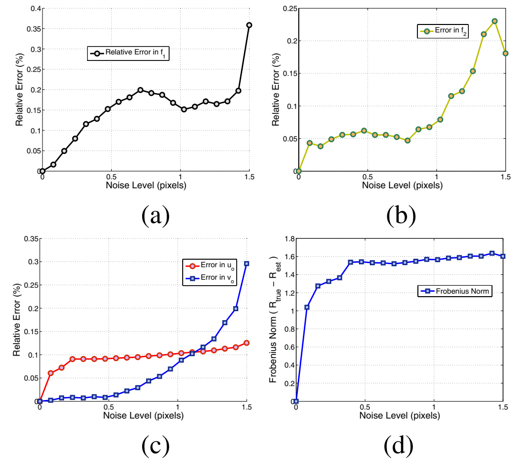
3.3 Camera Calibration

Using the property $\mathbf{R}_{i,j} = \mathbf{R}_{i,j}^{-T}$ we transform (2) to:

$$\boldsymbol{\omega}_j = \mathbf{H}_{i,j}^{-T} \boldsymbol{\omega}_i \mathbf{H}_{i,j}{}^{-1} \tag{4}$$

where $\boldsymbol{\omega}_j$ is the IAC for camera j. Noting the fact that $\boldsymbol{\omega}_j$ is a symmetric 3×3 matrix, (4) can be simply represented by:

$$\left[\begin{array}{ccc} \omega_{1,1} & \omega_{1,2} & \omega_{1,3} \\ * & \omega_{2,2} & \omega_{2,3} \\ * & * & 1 \end{array}\right]_j = \left[\begin{array}{ccc} c_{1,1} & c_{1,2} & c_{1,3} \\ * & c_{2,2} & c_{2,3} \\ * & * & 1 \end{array}\right]_i \tag{5}$$



(a) (b) (c) (d)

Figure 2. Performance of auto-calibration method VS. Noise level in pixels.

where '*' indicate symmetric values, and $c_{a,b}$ contain the unknown parameters of the RHS. Due to advancements in the camera manufacturing techniques, it is safe to assume that the aspect ratio $\lambda = 1$ and also the skew $\gamma = 0$ [1, 11, 10, 8, 7]. This gives us two linear constraints on the unknown parameters:

$$c_{1,2} = 0 \tag{6}$$

$$c_{1,1} = c_{2,2} \tag{7}$$

corresponding to $\omega_{1,2} = 0$ (from $\gamma = 0$) and $\omega_{1,1} = \omega_{2,2}$ (from $\lambda = 1$), respectively. Using (6) and (7) we express $\omega_{1,3}$ and $\omega_{2,3}$ in terms of $\omega_{1,1}$. Here we introduce a cost function on the algebraic distance of the principle point from the center of the image (I_x, I_y), which gives an extra weak constraint on $\boldsymbol{\omega}$:

$$\left[\begin{array}{cc} u_o & v_o \end{array}\right] = \arg\min \sum \left(\frac{\omega_{1,3}}{\omega_{1,1}} + I_x\right)^2 + \left(\frac{\omega_{2,3}}{\omega_{2,2}} + I_y\right)^2 \tag{8}$$

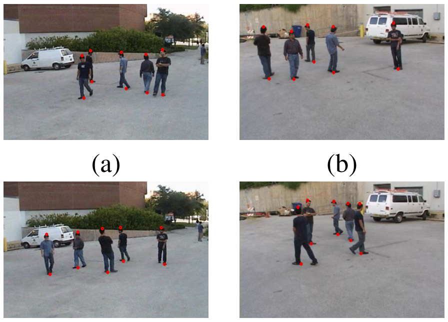
We solve this constraint by an application of Levenberg-Marquardt algorithm [12]. By substituting $\omega_{1,3}$ and $\omega_{2,3}$ obtained from solving (6) and (7) into (8) and minimizing it, $\omega_{1,1}$ can be estimated, which in turn determines $\omega_{1,3}$ and $\omega_{2,3}$ (note that $\omega_{1,1} = \omega_{2,2}$).

The relative rotation between the two camera is obtained by $\mathbf{R}_{i,j} = \mathbf{K}_j^{-1} \mathbf{H}_{i,j} \mathbf{K}_i$.

4 Results

We rigourously test the proposed method on synthetic data as well as on the real data. Due to space limitations, we show results for real data on only one sequence.

Synthetic data: We estimate for the unknown f, u_o and v_o. We generated 50 vertical synthetic objects on an uneven terrain i.e. having non-zero z-axis values. We



(a) (b)

(C) (d)

Figure 3. Instances from real data sequences: A column represents a unique camera.

project the data in two different views, having the principle points at $(320, 240)$, $\gamma = 0, \lambda = 1, f_1 = 1000$ and $f_2 = 1200$.

We gradually add a Gaussian noise with $\mu = 0$ and $\sigma \leq 1.5$ pixels to the data-points making up the vertical lines. Taking two vertical lines at a time, the vanishing points $\boldsymbol{v}$ and $\boldsymbol{v}'$ are estimated and $\mathbf{H}_{i,j}$ is recovered. The results of the estimated calibration parameters are shown in Fig. 2. For a maximum noise level of 1.5 pixels [13], the error rate for f_1 and f_2 is less than 1%. Similarly the error rate for the estimated principle is less than 0.4%. We also estimated the relative rotation matrix and in order to measure the error in estimation, we compute the Frobenius norm, as shown in Fig. 2d. In the curves shown, the error rate increases linearly. Note that the downward trend in the error curves is due to minimizing (8) in the vicinity of the image center.

Real data: The data was captured from cameras having a resolution of 720×480. We employed two such cameras looking at a scene with a number of people walking on a uneven ground. Fig. 3 shows two synchronized instances from the dataset consisting of more than 6000 frames. The top and the bottom points are marked with a red circle in the figure. The homography was extracting after estimating the vanishing points from the moving pedestrians, as described in Section 3. The extracted intrinsic parameter for camera one are:

$$\mathbf{K}_1 = \begin{bmatrix} 935.98 & 0 & 356 \\ 0 & 935.97 & 238 \\ 0 & 0 & 1 \end{bmatrix}$$

where as the focal length for the second camera is found to be $f_2 = 961.58$. Note that the estimated principle point is very close to the center of the image. Also note that f_1 and f_2 are very close to each other - indicating, qualitatively, the correctness of the estimated parameters.

5 Conclusion

Real world scenarios involves people walking on uneven terrains. We propose a novel method to obtain camera intrinsic and extrinsic parameters for such a scenarios when multiple cameras are looking at the area of interest. To the best of our knowledge, no work exists that deals with camera calibration for this specific scenario. Therefore, this method can be very useful for many of the existing multi-camera video surveillance systems that observe people in doors or outdoors. In the proposed method, tracking each walking pedestrian between the frames enables us to obtain vanishing points in each camera view. These points are then used to estimate the infinite homography that exists between any two cameras viewing the scene. We present novel constraints to solve for three of the camera intrinsic parameters. We show encouraging results on synthetic data, in the presence of large noise, as well as on real data.

References

[1] L. D. Agapito, E. Hayman, and I. Reid. Self-calibration of rotating and zooming cameras. *Int. J. Comput. Vision*, 45(2):107–127, 2001.

[2] X. Cao and H. Foroosh. Metrology from vertical objects. *Proc. of BMVC*, 2004.

[3] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *Int. J. Comput. Vision*, 40(2):123–148, 2000.

[4] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[5] O. Javed and M. Shah. Tracking and object classification for automated surveillance. *the seventh European Conference on Computer Vision*, 2002.

[6] I. Junejo and H. Foroosh. Euclidean path modeling for video surveillance. *Elsevier Journal of Image and Vision Computing (IVC)*, pages 512–528.

[7] I. Junejo and H. Foroosh. Trajectory rectification and path modeling for video surveillance. In *Eleventh IEEE International Conference on Computer Vision (ICCV)*, 2007.

[8] N. Krahnstoever and P. R. S. Mendonca. Bayesian autocalibration for surveillance. *Tenth IEEE International Conference on Computer Vision*, 2005.

[9] N. Krahnstoever and P. R. S. Mendonca. Autocalibration from tracks of walking people. In *British Machine Vision Conference*, 2006.

[10] F. Lv, T. Zhao, and R. Nevatia. Self-calibration of a camera from video of a walking human. *IEEE International Conference of Pattern Recognition*, 2002.

[11] M. Pollefeys, R. Koch, and L. V. Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *Int. J. Comput. Vision*, 32(1):7–25, 1999.

[12] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[13] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, 2000.