Accurate self-calibration of two cameras from observation of a person moving

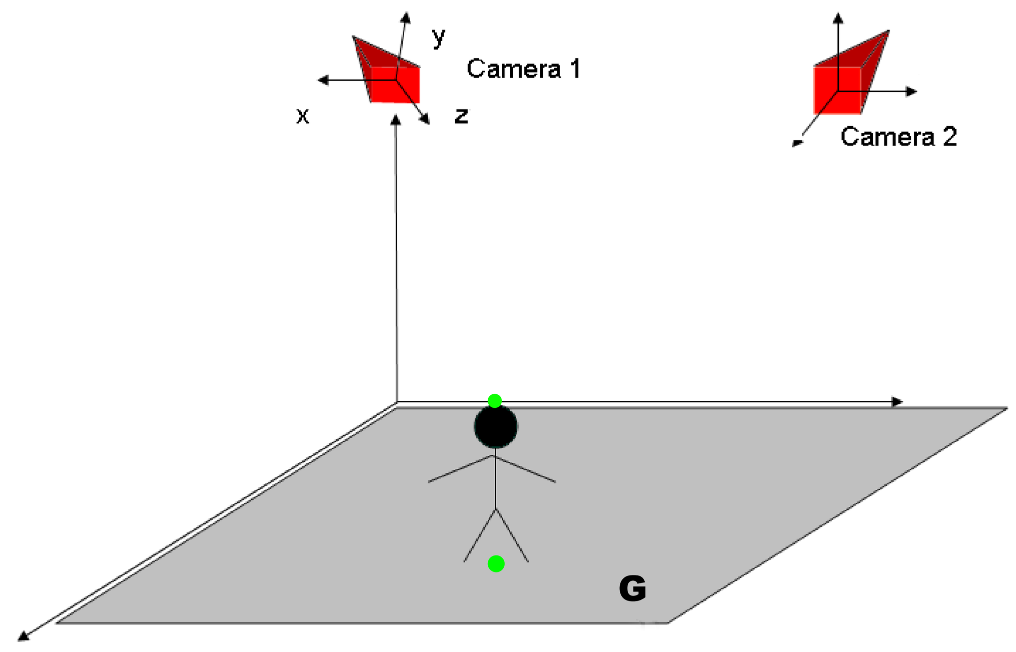
Tsuhan Chen
Advanced Multimedia Processing Lab
Carnegie Mellon University
Pittsburgh, U.S.A.

Alberto Del Bimbo, Federico Pernici and Giuseppe Serra
Media Integration and Communication Center
University of Florence, Italy

Abstract

A calibration algorithm of two cameras using observations of a moving person is presented. Similar methods have been proposed for self-calibration with a single camera, but internal parameter estimation is only limited to the focal length. Recently it has been demonstrated that principal point supposed in the center of the image causes inaccuracy of all estimated parameters. Our method exploits two cameras, using image points of head and foot locations of a moving person, to determine for both cameras the focal length and the principal point. In this poster we also describe a method to find the relative position and orientation of two cameras: the rotation matrix and the translation vector which describe the rigid motion between the coordinate frames fixed in two cameras. Results in synthetic and real scenes are presented to evaluate the performance of the proposed method.

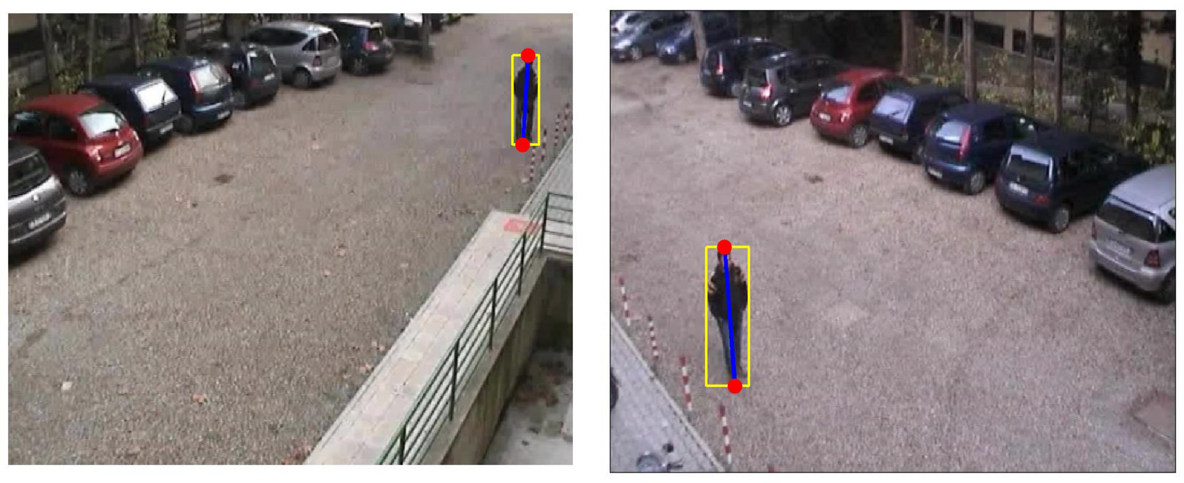
1 - The system



In this poster we present a method for calibration of two cameras based on features of a moving person in their common field of view. We use only the image of foot and head locations and we show how these points and their geometric relationship between cameras give enough information to find their relative position and orientation and the internal parameters of each camera, (i.e. the focal length and the principal point). We assume, like the previous methods, that the cameras have zero skew and unit aspect ratio.

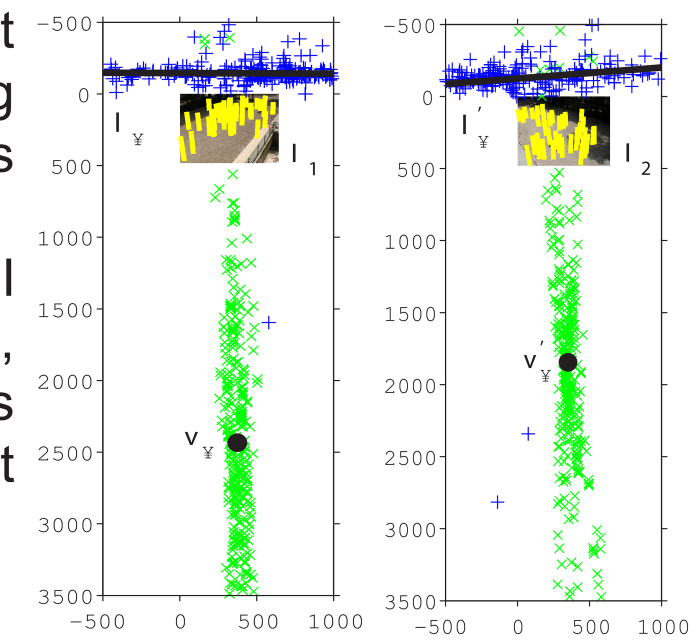
2 - Foot and head location

To determine the moving objects in the scene the Gaussian Mixture Model (GMM) for background subtraction is used. After localization, each moving blob is processed by the detector to determine if it is a person or not. Once the person is identified, the head and foot positions can be found by intersecting the smallest box around the person with its main axis (computed by the second order moment).



3 - Vanishing points and Vanishing lines

Considering a person as a vertical segment walking on a ground plane, the lines passing through the image head and the foot locations of each position identify vanishing point, $\mathbf{v}_\infty$. Since the height of a person is the same in all frames; given a set of different person positions, the line passing through the image head points and the line passing through the image foot points define vanishing line, $\mathbf{l}_\infty$.



4 - Infinite Homography

The Fundamental matrix can be computed using corresponding image points of foot and head (histogram is used to match person between cameras).
Once the vanishing points and vanishing lines of both cameras and the Fundamental Matrix are recovered, the Infinite Homography matrix can be computed:

$$\mathtt{H}_\infty = [\mathbf{l}'_\infty]_\times \mathtt{F} + \frac{(\mathbf{v}'_\infty \times \mathbf{e}')^T (\mathbf{v}'_\infty \times ((\mathtt{F}\mathbf{v}_\infty) \times \mathbf{l}'_\infty))}{\|\mathbf{v}'_\infty \times \mathbf{e}'\|^2 (\mathbf{l}_\infty^T \mathbf{v}_\infty)} \mathbf{e}' \mathbf{l}_\infty^T$$

5 - Calibration Cameras

Constraints on intrinsic camera parameters are enforced in terms of the absolute conic ω ($\omega = \mathtt{K}^{-T}\mathtt{K}^{-1}$, where $\mathtt{K}$ is the intrinsic parameters matrix). Once ω is known, $\mathtt{K}$ can be computed by Cholesky decomposition of ω. The matrix ω is symmetric, 3×3 matrix, defined up to scale, with five degrees of freedom. At least five constraints are needed to determine ω uniquely.

Zero skew assumption in the first camera	$\omega_{12} = \omega_{21} = 0$	One linear constraint
Square assumption in the first camera	$\omega_{11} = \omega_{22}$	One linear constraint
Pole-polar relationship in the first camera	$\mathbf{l}_\infty = \omega \mathbf{v}_\infty$	Two linear constraints
The zero skew assumption in the second camera	$\omega' = \mathtt{H}_\infty^{-T} \omega \mathtt{H}_\infty^{-1}$	One linear constraint

The same method is used to find the camera calibration matrix of the other camera.

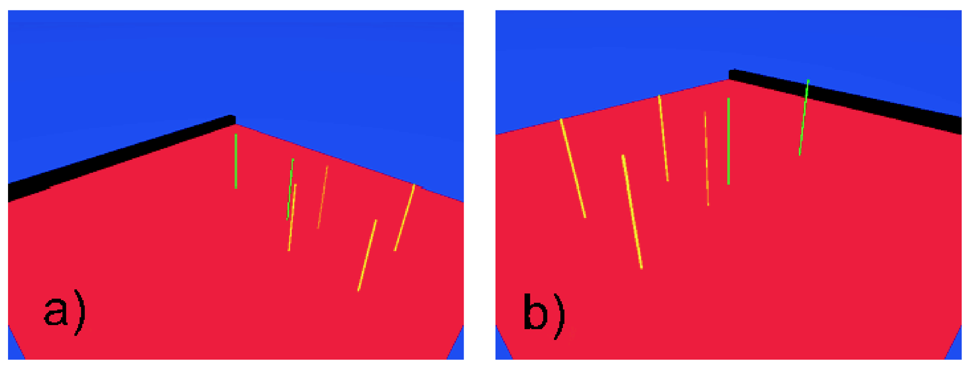
6 - Relative Rotation and translation of two cameras

The relative position and relative orientation are extracted from the essential matrix. In fact the Essential matrix, $\mathtt{E}$, is equal to: $\mathtt{E}=\mathtt{K}'^T\mathtt{F}\mathtt{K}$. It is known that this method find four solutions. The correct solution is obtained by testing a single image point matching between cameras. We used an image foot point correspondence.
Finally the translation scale is fixed by using the height values of the cameras from the ground. The height of a camera with respect to the ground is computed by

$$h_1 = \frac{1}{N} \sum_i \left[h / \left(1 - \frac{d(\mathbf{h}_i, \mathbf{c}_i) d(\mathbf{f}_i, \mathbf{v}_\infty)}{d(\mathbf{f}_i, \mathbf{c}_i) d(\mathbf{h}_i, \mathbf{v}_\infty)} \right) \right]$$

Experiments and Results

Synthetic experiments: To examine the performance of the proposed calibration algorithm we created a synthetic scene (fig. a and b). The first and the second camera were located respectively 3 and 2 meters above a ground plane with a tilt angle of π/6. Both cameras had a focal length of f = 480, unit aspect ratio, zero skew and principal point at (320,240). The image resolution was 640 × 480 pixels. In the scene randomly generated 1000 vertical segments of height 1.7 meters were inserted. The foot segments were posed on the ground plane. Gaussian noise with zero mean and different standard deviations were added to heads and feet separately to evalueated the performance of the proposed algorithm.

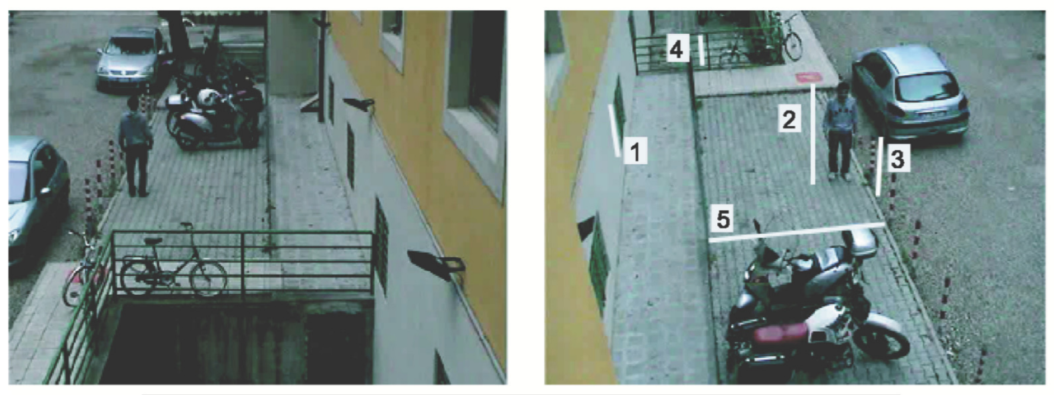


	Camera 1			Camera 2			Relative position					
							Center of the second camera			Rotation		
	f	u	v	f	u	v	x	y	z	α	β	γ
True value	480	320	240	480	320	240	4.243m	-1.255m	4.174m	-58.80°	-72.42°	88.391°
Std. Gaus. Error.												
0.5	1.276	2.838	3.52	2.108	0.368	3.394	0.0109m	0.0234m	0.0229m	0.39°	0.13°	0.63°
1	2.613	2.846	4.954	4.281	0.715	6.971	0.0223m	0.0475m	0.0466m	0.01°	0.79°	0.27°
2	5.396	2.935	9.219	9.497	1.453	16.072	0.0490m	0.0989m	0.0983m	0.01°	1.79°	0.53°
3	8.496	3.086	14.224	15.936	2.188	28.496	0.0816m	0.1552m	0.1574m	0,00°	3.06°	0.79°
5	12.174	3.308	20.431	24.165	2.936	46.273	0.1234m	0.2214m	0.2296m	0.00°	4.83°	1.05°

Real experiments: The proposed algorithm has been tested on multiple real sequences with a different walking person each. The image sequences had a resolution of 320 × 240 pixels and were captured at multiple locations and orientations. Here two examples are shown. To evaluate the performance of our algorithm we tested its ability to recover distances between objects or height of objects. For each experiment we captured about 70 measurements. The estimation errors are 0.04m and 0.05m respectively (fig. c and d).

Measurements	Estimated values	True values
1	1.923m	1.960m
2	1.864m	1.880m
3	0.311m	0.330m
4	3.521m	3.460m
5	0.919m	0.940m

c)



Measurements	Estimated values	True values
1	0.948m	0.980m
2	1.792m	1.800m
3	0.971m	0.940m
4	0.769m	0.800m
5	2.796m	2.820m

d)