In 17th Portuguese Conference on Pattern Recognition (RecPad 2011)
Porto, Portugal, October 2011

Auto-Calibration of Pan-Tilt-Zoom Cameras: Estimating Intrinsic and Radial Distortion Parameters

Diogo Leite, Alexandre Bernardino, José Gaspar
leitediogo@live.com.pt,{alex,jag}@isr.ist.utl.pt

Institute of Systems and Robotics, IST / UTL
Lisboa, Portugal

Abstract

In this paper we propose a method for PTZ camera auto-calibration over the camera's zoom range. The method is based on the minimization of re-projection errors of feature points detected in images captured by the camera at different orientations and zoom levels. Experiments conducted on both synthetic and real data show the method achieves good results compared to methods that require higher computational costs or access to the camera.

1 Introduction

Pan-tilt-zoom (PTZ) surveillance cameras allow constructing background models of wide scenes and, consequently, detecting scene changes, provided one can accurately map pixels captured from the camera into 3D space. In other words, background modeling and scene-changes detection imply camera calibration. In this paper we present a calibration method for PTZ cameras that does not required physical access to the camera or the environment nor specific objects or structures in the image sequences used.

There are several documented methods for camera calibration requiring physical access to the camera such as [2] where intrinsic and radial distortion parameters are estimated by changing the orientations of a chess pattern in front of the camera. Past work on active camera calibration was essentially conducted for intrinsic parameter estimation only. Hartley [5] presented a self-calibration method for stationary (rotating) cameras and later Agapito *et al*. [1] introduced a self-calibration method for rotating and zooming cameras. These methods answered the problem of geometric calibration with methods that did not require physical access to the cameras to achieve good intrinsic parameters estimation. The method we propose is very much alike the one proposed by Sinha [7]. The main differences lye in the simplicity and reduced computational costs associated to ours and in the possibility ours offers of using any kind of image sequences for calibration provided there is overlapping between images. In addition, we estimate the radial distortion caused by the camera which is an effect that attenuates with camera zoom.

The method we propose is composed by a first step calibration procedure that uses homographies between acquired images to estimate the camera's intrinsic parameters at minimum zoom level. With the application of a cost function and the use of Levenberg-Marquardt optimization algorithm we are able to accurately estimate the radial distortion coefficients at this specific zoom level. The final step of our method enables the estimation of both the intrinsic parameters and the radial distortion coefficients for progressive zoom steps in a way simpler to that used in the calibration at minimum zoom level. We do a 3D reconstruction of pixel points acquired by the camera with a cube based model [3].

2 Camera Model

The pin-hole camera model for the perspective pan-tilt-zoom camera consists of a mapping from 3D projective space to 2D projective space. This is represented by a 3x4 rank-3 perspective matrix, $\mathbf{P}$. The mapping from 3D to the image plane takes a point $\mathbf{X} = [X\ Y\ Z\ 1]^T$ to a point $\mathbf{u} = \mathbf{PX}$ in homogeneous coordinates. The matrix $\mathbf{P}$ may be decomposed in

$$\mathbf{P} = \mathbf{K^z}[\mathbf{R} \mid \mathbf{t}] \tag{1}$$

where $\mathbf{t}$ is a 3x1 vector that represents the camera location, $\mathbf{R}$ is a 3x3 rotation matrix that represents the orientation of the camera with respect to an absolute coordinate frame and $\mathbf{K}^z$ is a 3x3 upper triangular matrix encompassing the intrinsic parameters of the camera:

$$\mathbf{K}^z = \begin{bmatrix} k_u & s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where k_u and k_v are the magnifications in the respective u and v directions, u_0 and v_0 are the coordinates of the principal point of the camera and s is a skew parameter (in this work we assume $s = 0$). The pan and tilt movements are included in $\mathbf{R}$. These movements are simply rotations about the projective center $\mathbf{O}$, which is usually chosen to be the world origin and thus $\mathbf{t} = [0\ 0\ 0]^T$.

Most cameras deviate from the pin-hole model due to radial distortion. This effect decreases with increasing focal length. Due to radial distortion a 3D point $\mathbf{X}$ is projected to a point $\delta\mathbf{x_d} = [\delta x_d\ \ \delta y_d]^T$. This point is deviated from the point $\mathbf{x} = [x\ y]^T$ according to the radial distortion function, $\Re^z$:

$$[x_d\ y_d]^T = \Re^z\left([x\ y]^T\right) = \mathbf{L}(r)[x\ y]^T = (1 + k_1 r^2 + k_2 r^4)[x\ y]^T \tag{3}$$

where $r = \sqrt{x^2 + y^2}$. This radial distortion model corresponds to a simplified two coefficient version of the one proposed by Heikkila [6] where r is the radial distance (distance from point $\mathbf{x}$ to the center of distortion (x_c, y_c)), $\mathbf{L}(r)$ is a radially symmetric distortion factor and k_1 and k_2 are the two radial distortion coefficients considered. For every zoom level z, $\Re^z$ is parameterized by $(x_c^z, y_c^z, k_1^z, k_2^z)$. In our model the principal point (u_0, v_0) is constrained to be the center of distortion.

Concluding, the goal of calibrating a PTZ camera involves estimating the unknown parameters of $\mathbf{K}^z$ and $\Re^z$ for any pan and tilt angles, while covering full range of optical zoom, from z_0 to z_{max}.

3 Calibration Method

The calibration method proposed consists of two steps. The intrinsics and radial distortion coefficients are first iteratively estimated at minimum zoom level and then computed for an increasing zoom sequence.

3.1 Calibration at minimum zoom level

The first step of our method is computing the intrinsics and radial distortion coefficients at minimum zoom level (z_0). To achieve this images are captured in a spherical grid. Every arc connecting two images is listed with informations such as identification of the two connected images, feature points detected in both and their correspondences. The feature locations and correspondences are determined with the SIFT algorithm followed by a match filtering. Homographies between every adjacent horizontal, $\mathbf{H_i}$ and vertical images, $\mathbf{V_i}$ are then robustly computed using RANSAC-based homography estimation and nonlinear minimization [4] (Chapter 3, p. 108). One of the images, $\mathbf{I_r}$ is chosen to be the reference one and homographies, $\mathbf{T_i}$ mapping points from the other images, $\mathbf{I_i}$ to $\mathbf{I_r}$ are computed through composition of the previous obtained $\mathbf{H_i}$ and $\mathbf{V_i}$ homographies.

Corresponding feature points, detected in two or more images, are back-projected to a unit sphere centered at the camera position using an estimate of the camera parameters. The corresponding 3D points, back-projected from different images, are merged into a single 3D mean point. Then to achieve global image alignment and an accurate geometric calibration a non-linear optimization of the re-projection error ($\mathbf{D}$) of all points ($\mathbf{u_i^j}$) is applied:

$$(\mathbf{K^{z_0}}^*, \Re^{\mathbf{z_0}*}) = \arg\min_{\mathbf{K^{z_0}}, \Re^{\mathbf{z_0}}} \sum_{i=1}^{n} \sum_{j=1}^{m(i)} \mathbf{D}(\mathbf{u_i^j},\ \mathbf{K^{z_0}} \Re^{\mathbf{z_0}}(\mathbf{R_i X^j}))^2 \tag{4}$$

where $\mathbf{R}_i$ are the rotation matrices for the respective images, $m(i)$ and n are the feature-count and image-count and $\mathbf{X}^j$ is a global feature list. This global minimization problem is solved using the Levenberg-Marquardt algorithm, implemented in the Matlab function `lsqnonlin`. The optimization algorithm is initialized with the matrix of intrinsic parameters $\mathbf{K^{z_0}}$ obtained using Agapito's *et al*. method [1] and null radial distortion coefficients ($k_1 = 0$ and $k_2 = 0$ initialize $\Re^{z_0}$, see Eq. 3).

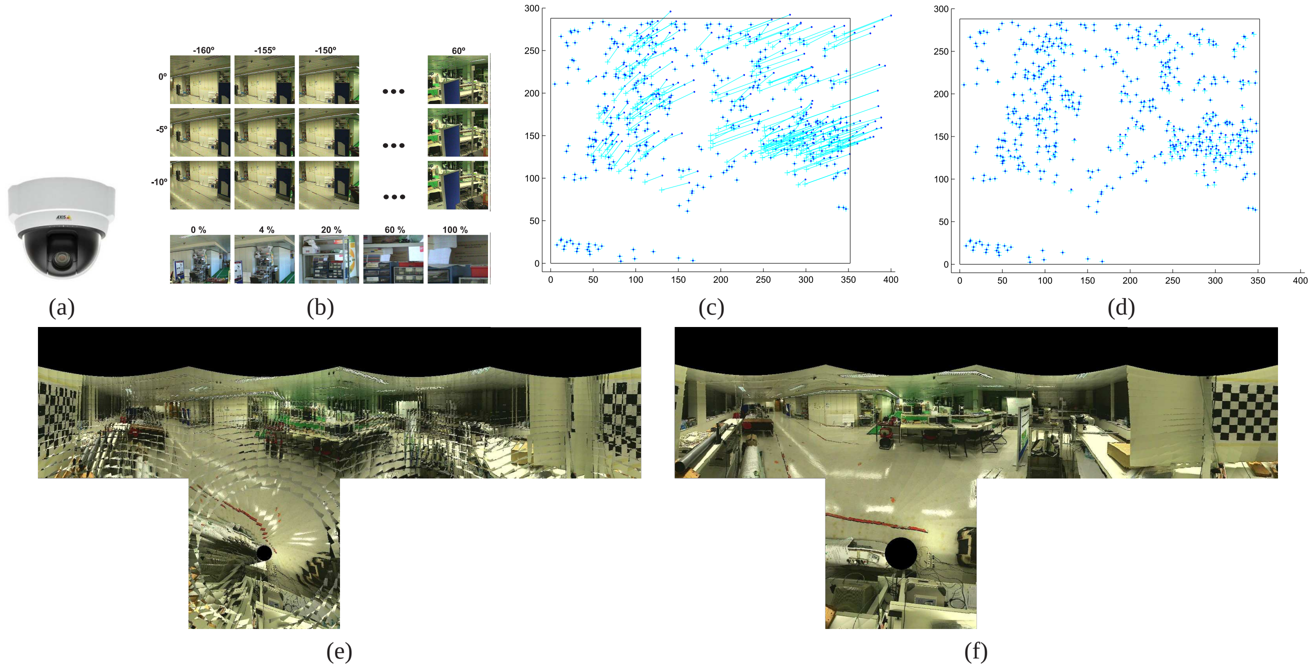

Figure 1: Display of different phases of the calibration method. (a) Axis 215 PTZ camera, (b) regular grid of images and zoom sequence used to calibrate camera,(c) re-projection errors at first iteration of the method, (d) re-projection errors at last (30) iteration of the method, (e) mosaic generated with intrinsic parameters estimated from Agapito's method and null radial distortion parameters and (f) mosaic generated with our estimation of intrinsic and radial distortion parameters.

3.2 Zoom sequence calibration

To achieve a zoom range calibration the camera is fixed at a certain orientation and images with progressively larger zoom levels are acquired. The computation steps that follow are similar to the ones applied in the minimum zoom calibration method (See Section 3.1). Parameters $\mathbf{K}^{z_i}$ and $\mathfrak{R}^{z_i}$ are iteratively estimated for every zoom level i with the starting estimates being the parameters found for zoom level $i-1$, $\mathbf{K}^{z_{i-1}}$ and $\mathfrak{R}^{z_{i-1}}$:

$$(\mathbf{K}^{\mathbf{z_i}*}, \mathfrak{R}^{\mathbf{z_i}*}) = \arg\min_{\mathbf{K}^{\mathbf{z_i}}, \mathfrak{R}^{\mathbf{z_i}}} \sum_{k=i-1}^{i} \sum_{j=1}^{m(k)} \mathbf{D}(\mathbf{u_k^j}, \mathbf{K}^{\mathbf{z_k}} \mathfrak{R}^{\mathbf{z_k}}(\mathbf{X^j}))^2. \qquad (5)$$

To initialize this second step of the calibration method the initial estimates for the parameters are the ones obtained as result of the first step calibration method, $\mathbf{K}^{z_0}$ and $\mathfrak{R}^{z_0}$. Then, the optimization problem, Eq. 5, is solved for every zoom level z_i, $i \in \{1, 2, \ldots, N\}$, where N denotes the maximum zoom level. The estimation of radial distortion parameters from a single zoom sequence has inherent ambiguities as the radial distortion at a particular zoom level may be compensated by the radial distortion at another level. We avoided this by keeping the parameters at zoom level z_0 fixed.

4 Results

The experiments conducted with our method encompassed synthetic and real data. First a camera with known parameters was simulated and 3D points were created and randomly placed in front of it. The method was tested in the presence of matching errors due to different noise levels applied. For real data testing an Axis 215 PTZ camera (Figure 1 (a)) was chosen and two sequences of images were used to calibrate the camera over its full zoom range. A regular grid of 164 images was acquired to estimate the parameters at minimum zoom level and then a progressive zoom sequence of 25 images was obtained with the camera at a fixed orientation to calibrate it over its full zoom range (Figure 1 (b)). The initial estimates at minimum zoom level result from the application of the self-calibration method proposed by Agapito *et al*. [1]. This estimates produce substantial alignment errors which are reproduced in the re-projection of matched feature points in the first image plane (Figure 1 (c)) and result in a poor mosaic construction (Figure 1 (e)). With our method these initial re-projection errors are rapidly minimized achieving sub-pixel errors after 30 iterations (Figure 1 (d)). The estimates obtained produce an improved mosaic as represented in Figure 1 (f).

5 Conclusions

In this paper we proposed an automatic calibration method for PTZ cameras. Intrinsic parameters and radial distortion coefficients are estimated over the full range of pan, tilt and zoom. These estimations are achieved by computing homographies between images and through minimization of re-projection errors. Future work will be focused in target and event tracking using active PTZ cameras. These cameras possess wide pan and tilt ranges and a large zooming capacity so potentially they have great tracking capacity. Given the calibration method proposed in this paper it is possible to generate high-definition mosaics of the whole environment and so it will be possible to accurately track events in the vast majority of the environment of the camera.

6 Acknowledgments

This work has been partially supported by the Portuguese Government - Fundação para a Ciência e Tecnologia (ISR/IST pluriannual funding) through the PIDDAC program funds and through the project DCCAL, PTDC / EEACRO / 105413 / 2008.

References

[1] L. Agapito, R. Hartley, and E. Hayman. Linear calibration of a rotating and zooming camera. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 15–21, 1999.

[2] Jean-Yves Bouguet. Camera calibration toolbox for matlab. `http://www.vision.caltech.edu/bouguetj`.

[3] Ricardo Galego, Alexandre Bernardino, and José Gaspar. Surveillance with pan-tilt cameras: Background modeling. 2008.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[5] Richard I. Hartley. Self-calibration of stationary cameras. In *International Journal of Computer Science*, pages 5–23, 1997.

[6] Janne Heikkila and Olli Silven. A four-step camera calibration procedure with implicit image correction. 1997.

[7] Sudipta N. Sinha and Marc Pollefeys. Pan-tilt-zoom camera calibration and high-definition mosaic generation. In *Computer Vision and Image Understanding 103*, pages 170–183, 2006.