FAST CALIBRATION METHOD FOR ACTIVE CAMERAS

Piero Donaggio and Stefano Ghidoni
Videotec S.p.A., via Friuli, 6 – Schio (VI), Italy

Keywords: Camera calibration, lens distortion, camera networks, camera patrolling

Abstract: In this paper a model for active cameras that considers complex camera dynamics and lens distortion is presented. This model is particularly suited for real-time applications, thanks to the low computational load required when the active camera is moved. In addition, a simple technique for interpolating calibration parameters is described, resulting in very accurate calibration over the full range of focal lengths. The proposed system can be employed to enhance the patrolling activity performed by a network of active cameras that supervise large areas. Experiments are also presented, showing the improvement provided over traditional pin-hole camera models.

1 INTRODUCTION

In recent years, intelligent video surveillance systems based on camera networks have gathered increasing interest by both research and industry, since they are able to keep under control large regions, and to see a scene from multiple viewpoints, thus easily coping with occlusions, that often limit image processing systems based on a single video stream.

Nodes in a camera network typically need to exchange data, including spatial information, e.g. a tracked target position, or patrolling waypoints defined in the video stream of one camera that should be displayed from another camera perspective. This requires an accurate camera calibration, in order to perform correct associations between different video streams.

In this paper, we present an accurate camera model that enhances the 2D-3D point mapping that can be employed for defining patrolling paths with high accuracy, achieved by considering: (i) distortion caused by the lens, and (ii) pixel aspect ratio, that are usually neglected, or not precisely modelled, in commercially available systems.

Another important feature that has guided the system development is efficiency: lens distortion has not been eliminated by applying image undistortion for working on undistorted images, but rather, it has been considered in order to correctly undistort only the path waypoints projected from the image to the world, and correctly *distort* them when a conversion from world to image is needed. This keeps the computational load low, and lets the human operator work on the images acquired by the camera, and not on the undistorted ones, that usually feel less natural.

The paper is organized as follows: in section 2 camera network systems for patrolling are discussed, together with methods for camera calibration and lens distortion correction, while in section 3 our approach is presented. Finally, in section 4 experimental results are presented, while some final remarks are drawn in section 5.

2 RELATED WORK

In video surveillance applications, such as perimeter patrolling, a high level of accuracy is often required. Several researchers addressed the problem of calibrating pan-tilt-zoom cameras in real environments (Hartley, 1994; Agapito et al., 2001; Del Bimbo and Pernici, 2009). Most of them use a fairly simple geometric model in which axes of rotation are aligned with the camera imaging optics. This assumption is often violated in commercially integrated pan-tilt cameras (Davis and Chen, 2003).

Lens distortion must also be taken into account for increasing calibration accuracy. Collins et. al. (Collins and Tsin, 1999) calibrated a pan-tilt-zoom active camera system in an outdoor environment, assuming constant radial distortion and modelling its variation using a magnification factor. A more precise estimation of lens distorsion as a function of zoom is introduced in (Sinha and Pollefeys, 2004). Despite a high level of accuracy, the whole calibration process is very expensive and requires a closed-loop system to re-estimate the calibration every time the camera moves.

In this paper, we address both extrinsic and intrinsic camera calibration. Similarly to (Raimondo et al., 2010), we introduce a more complete geometric model, in which pan and tilt axes do not necessarily pass through the origin of the system and the imaging plane and optics are modeled as a rigid element that rotates around each of these axes. We exploit a priori knowledge of camera mechanics and rely on fine-grained pan-tilt-zoom encoders for maintaining calibration of active cameras while zooming and rotating. We also extend the work of (Raimondo et al., 2010) by introducing lens distortion compensation, removing the assumption of square pixel aspect ratio, and employing a technique for estimating calibration data at any zoom level, once they are measured off-line at a set of given focal lengths.

The proposed calibration method is computationally inexpensive, and therefore suitable for real time operations in camera networks. Moreover, our lens distorsion compensation technique is simpler and faster to achieve than the one proposed in (Sinha and Pollefeys, 2004) yet producing very accurate results.

3 MAPPING 2D-3D

The cameras employed in the system are PTZ units characterized by a rather complex mechanics, and require a specific geometric model in order to increase reprojection precision. Thus, the classic pinhole model has been replaced by a more realistic one, that takes into account all mechanical parameters, camera self-rotation as well as distortion effects.

Intrinsic parameters have been evaluated exploiting the camera calibration toolbox provided by OpenCV (Bradski, 2000). In order to estimate intrinsic parameters, several shots with a chessboard need to be taken and provided to the calibration framework.

The model employed for describing the distortion effect is accurate also for short focal lengths, as it is the case of some camera modules that reach an horizontal aperture angle of 58°. The calibration procedure provides both the camera matrix and the distortion coefficients, that make it possible to compensate for lens distortion. The OpenCV library provides functions for undistorting images and points.

3.1 Extrinsic calibration

As shown in (Raimondo et al., 2010), the relationship between a point $p_{\text{oc}} = [x_{\text{oc}}, y_{\text{oc}}, z_{\text{oc}}]$ expressed in the camera coordinate system and the point $p_w = [x_w, y_w, z_w]$ expressed in the world reference system is given by:

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ H \end{bmatrix} + R_\theta \left(\begin{bmatrix} D \\ 0 \\ 0 \end{bmatrix} + R_\varphi \left(\begin{bmatrix} x_{\text{off}} \\ y_{\text{off}} \\ z_{\text{off}} \end{bmatrix} + \begin{bmatrix} x_{\text{oc}} \\ y_{\text{oc}} \\ z_{\text{oc}} \end{bmatrix} \right) \right), \tag{1}$$

where H is the camera height from the ground, D the tilt rotation axis offset w.r.t. the pan rotation axis and $x_{\text{off}}, y_{\text{off}}, z_{\text{off}}$ are the displacements of the camera inside its case. R_θ and R_φ are the pan and tilt rotation matrices, respectively.

The angles θ and φ are acquired from pan-tilt encoders. Parameters $H, D, x_{\text{off}}, y_{\text{off}}, z_{\text{off}}$ are unknown but can be measured directly when the unit is assembled.

3.2 Image-world projections

Once intrinsic and extrinsic parameters are available, it is possible to precisely map image points to world coordinates, assuming that such points lie on the ground plane.

In order to calculate the coordinates (x_w, y_w, z_w) of a point p_w that lies on the ground plane from the coordinates (u, v) of its projection on the image plane expressed in pixel coordinates, we use the procedure proposed by (Raimondo et al., 2010), but considering also lens distortion and rectangular pixel aspect ratio.

Image sensors are in fact matrices of sensing elements that are often considered to be square; however, this is an approximation, and a more realistic model should consider pixels to be rectangles. This has an important consequence on those models that express focal lengths as a function of the pixel size: such models should in fact consider two different focal lengths, f_x and f_y, to model a rectangular pixel aspect ratio. By applying the above consideration, it is possible to modify equations (1) to obtain:

$$\begin{aligned} x_{w_0} &= c_\varphi x_{\text{off}} + s_\varphi z_{\text{off}} + D + \\ &\qquad \frac{(s_\varphi x_{\text{off}} - c_\varphi z_{\text{off}} - H)(s_y s_\varphi (v - c_y) - s_x c_\varphi f_x)}{s_x s_\varphi f_x + s_y c_\varphi (v - c_y)}, \\ y_{w_0} &= y_{\text{off}} + \frac{(s_\varphi x_{\text{off}} - c_\varphi z_{\text{off}} - H)(s_x (u - c_x))}{s_x s_\varphi f_x + s_y c_\varphi (v - c_y)}, \\ z_{w_0} &= 0, \end{aligned} \tag{2}$$

in which a zero-pan system is considered, without loss of generality. In the above equations, $x_{w_0}, y_{w_0}, z_{w_0}$ are the coordinates of the point p_{w_0} expressed in the zero-pan system, $c_\varphi = \cos(\varphi)$ and $s_\varphi = \sin(\varphi)$.

Eventually, the 3D-point coordinates in the world reference system are computed by simply multiplying p_{w_0} by clockwise pan rotation matrix.

Lens distortion can be compensated by removing its effect by means of specific functions provided by the OpenCV library. It is then enough to remove distortion before projecting an image point to the world ground plane to get an accurate result. On the way back, the mapping from world to image coordinates is followed by a transformation that actually distorts points: this function, however, is unfortunately not available in the OpenCV library and it has been implemented from scratch.

3.3 Calibration data interpolation

The complete pinhole camera model includes the focal lengths f_x and f_y, and therefore depends on the zoom level; the same dependency involves also distortion coefficients, since distortion is strongly influenced by the focal length and by the lens used. This means that in theory a calibration process should be carried out at each zoom level at which the camera will acquire images. This is very difficult to achieve in practice, unless the system is restricted to operate at very few zoom levels, which would represent a very strong limitation.

To overcome these issues, a set of calibration points, at several zoom levels, is collected. An interpolation method is then employed to recover the parameter values at any desired focal length. Such method is a linear interpolation: given an arbitrary focal length, parameters are evaluated with a linear combination of the nearest upper and lower values. While this can be reasonable for some parameters, like f_x and f_y, a linearization represents a strong assumption for distortion parameters, that can hardly adhere to the actual variation laws. On the other hand, more realistic models would require a very deep knowledge of the lens, that is difficult to obtain from the manufacturer.

In order to make the linear model precise, an adequate sampling rate of zoom levels at which calibration is performed off-line has been chosen. In particular, distortion coefficients are highly non-linear at wide angles, but become much easier to predict at higher zoom levels.

4 RESULTS

System performance has been verified by measuring reprojection errors for a set of points taken as ground truth. Errors are measured with and without the effects of distortion removal and rectangular pixel aspect ratio, and at several zoom levels. This way it is possible to evaluate the accuracy of the proposed model, and to understand which is the distortion level of the system.

Table 1: Mean reprojection errors at several zoom levels, with and without distortion removal. Calibration data between 1x and 2x (excluded) are interpolated. Values are in centimeters.

Zoom level	Distorted	Undistorted
1x	6.65573	3.67862
1.25x	6.13914	3.49851
1.5x	5.08299	4.48573
1.75x	4.20172	3.57173
2x	3.38452	2

To assess the calibration parameters interpolation accuracy, the test described above has been performed at several zoom levels, going from 1x to 2x, that is the range in which distortion effects are higher. However, calibration parameters has not been evaluated for all zoom levels, but only at 1x, 1.5x and 2x: the case of 1.25x and 1.75x rely on interpolated data. Results are summarized in table 1.

Regarding calibration parameter interpolation, it has been observed that by sampling data every 0.5x, reprojection errors for interpolated zoom levels are similar to those obtained at zoom level for which the calibration procedure had been performed. This holds in the range between 1x and 2x, that is, the region where lens distortion is stronger. For higher zoom levels experiments have not been performed, but it is possible to argue that less sampling points will be required, thanks to the reduced distortion at such zoom levels.

In figure 1 an example of point reprojection is shown: when lens distortion is not compensated and square pixel aspect ratio assumption holds, the patrolling path moves on the world reference when the camera is moved, as it can be seen comparing (a) and (c). The problem is solved by our model: comparing (b) and (d), the red line is at the bottom of the wardrobe in both images, even if it appears close to the image border (b) and towards the image center (d). The same precision is obtained also when changing the camera pan angle, because the model takes into account the non-square pixel aspect ratio.

5 CONCLUSIONS

In this paper, an accurate model for active cameras has been described, taking into account the com-

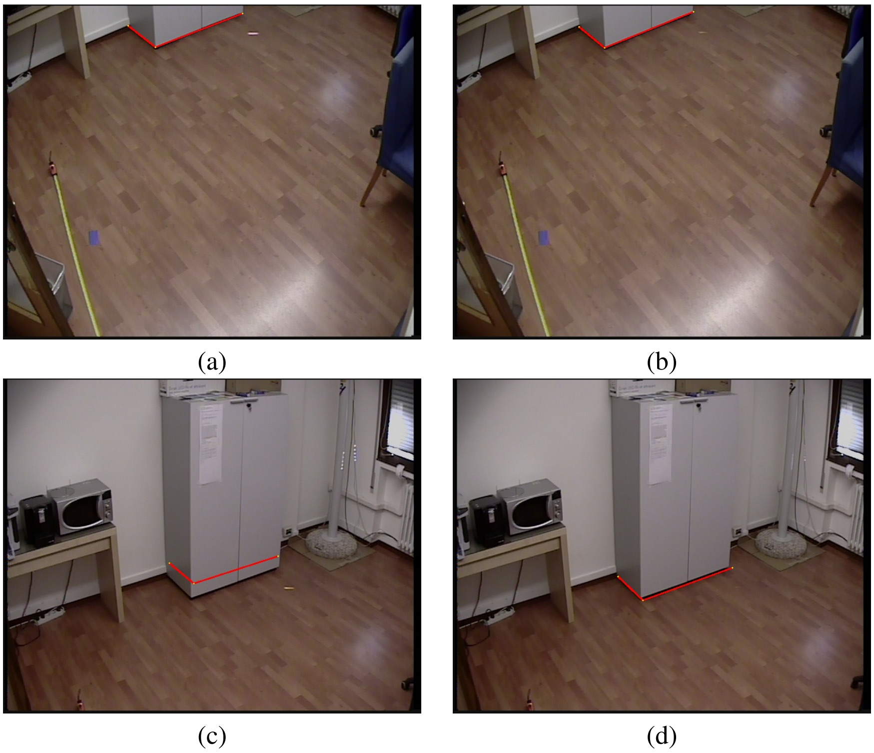

Figure 1: Effect of lens distortion when the PTZ camera is moved: the patrolling path is not correctly mapped during the movement between (a) and (c). Between (b) and (d) a similar movement happens, but lens distortion and pixel aspect ratio were taken into account when projecting waypoints to the ground plane.

plex mechanics of pan-tilt-zoom units as well as lens distortion while removing the assumption of square pixel aspect ratio. Using this model, a fast calibration procedure which exploits prior information on camera dynamics and interpolated lens distortion parameters has also been introduced.

Results show an increment in 2D-3D mapping accuracy over classical camera models, thus demonstrating the validity of our assumptions. The proposed system has proved to be suitable for real-time, accurate operations in a network of active cameras.

REFERENCES

Agapito, L., Hayman, E., and Reid, I. D. (2001). Self-calibration of rotating and zooming cameras. *International Journal of Computer Vision*, 45:107–127.

Aghajan, H. and Cavallaro, A. (2009). *Multi-camera networks*. Elsevier, 1st edition.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Bradsky, G. (2008). *Learning OpenCV*. O'Reilly, 1st edition.

Collins, R. T. and Tsin, Y. (1999). Calibration of an outdoor active camera system. In *In Proceedings of the 1999 Conference on Computer Vision and Pattern Recognition*, pages 528–534. IEEE Computer Society.

Davis, J. and Chen, X. (2003). Calibrating pan-tilt cameras in wide-area surveillance networks. In *Proc. of IEEE International Conference on Computer Vision*, pages 144–150. IEEE Computer Society.

Del Bimbo, A. and Pernici, F. (2009). Single view geometry and active camera networks made easy. In *Proceedings of the First ACM workshop on Multimedia in forensics*, MiFor '09, pages 55–60, New York, NY, USA. ACM.

Hartley, R. I. (1994). Self-calibration from multiple views with a rotating camera. pages 471–478. Springer-Verlag.

Raimondo, D., Gasparella, S., Sturzenegger, D., Lygeros, J., and Morari, M. (2010). A tracking algorithm for ptz cameras. In *2st IFAC Workshop on Distributed Estimation and Control in Networked Systems (NecSys 10), Annecy, France*.

Sinha, S. N. and Pollefeys, M. (2004). Towards calibrating a pan-tilt-zoom camera network. In *In Workshop on Omnidirectional Vision and Camera Networks at ECCV*.