Expert Systems with Applications 41 (2014) 1532–1542

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Hierarchical camera auto-calibration for traffic surveillance systems

S. Álvarez *, D.F. Llorca, M.A. Sotelo

Computer Engineering Department, Polytechnic School, University of Alcalá, Madrid 28871, Spain

A R T I C L E I N F O

Keywords:
Auto-calibration
Pan-tilt-zoom cameras
Vanishing points
Intelligent transportation systems
Urban traffic infrastructures

A B S T R A C T

In this paper, a hierarchical monocular camera auto-calibration method is presented for applications in the framework of intelligent transportation systems (ITS). It is based on vanishing point extraction from common static elements present on the scene, and moving objects as pedestrians and vehicles. This process is very useful to recover metrics from images or applying information of 3D models to estimate 2D pose of targets, making a posterior object detection and tracking more robust to noise and occlusions. Moreover, the algorithm is independent of the position of the camera, and it is able to work with variable pan-tilt-zoom (PTZ) cameras in fully self-adaptive mode. The objective is to obtain the camera parameters without any restriction in terms of constraints or the need of prior knowledge, to deal with most traffic scenarios and possible configurations. The results achieved up to date in real traffic conditions are presented and discussed.

1. Introduction

Camera calibration is a fundamental stage in computer vision. The process is the determination of the relationship between a reference plane and the camera coordinate system (extrinsics), and between the camera and the image coordinate system (intrinsics). These parameters are very useful to recover metrics from images or applying prior information of 3D models to estimate 2D pose of targets, giving an idea of the size of the objects and making their detection and tracking more robust to noise and occlusions.

In a previous paper Álvarez et al. (2012), the authors presented a target detection system for transport infrastructures based on manual camera calibration through vanishing points. After that, the approach was improved, as described in Álvarez et al. (2013), with a preliminary automatic calibration method based on camera zooming and zebra-crossings. The current paper extends these works with a hierarchical camera auto-calibration system, which deals with most traffic scenarios and configurations with no restrictions. The work begins with the paper presented in Álvarez et al. (2011).

The standard method to calibrate a camera is based on a set of correspondences between 3D points and their projections on image plane as presented by Hartley and Zisserman (2000) and Tsai (1986). However, this method requires either prior information of the scene or calibrated templates, limiting the feasibility of surveillance algorithms in most possible scenarios. In addition, calibrated templates are not always available, they are not applicable for already-recorded videos and if the camera is placed very high their small projection can derive in poor accurate results. Finally, in case of having PTZ cameras, using a template each time the camera angles or zoom changes is not feasible. One novel method which solves the problem of the template is the orthogonal calibration proposed by Kim (2009). The system extracts the world coordinates from aerial pictures (on-line satellite images) or GPS devices to make the correspondences with the image captured. However this system is dependent on prior information from an external source and it does not work indoor. Therefore auto-calibration seems to be the more suitable way to recover camera parameters for surveillance applications.

One of the distinguished features of the perspective projection is that the image of an object that stretches off to infinity can have finite extent. For example, parallel world lines are imaged as converging lines, which image intersection point is called *vanishing point*. Caprile and Torre (1990), developed a new method for camera calibration using simple properties of vanishing points. In their work, the intrinsic parameters of the camera were recovered from a single image of a cube. In a second step, the extrinsic parameters of a pair of cameras were estimated from an image stereo pair of a suitable planar pattern. The technique was improved by Cipolla et al. (1999), who computed both intrinsic and extrinsic parameters from three vanishing points and two reference points from two views of an architectural scene. However these assumptions were incomplete, because as demonstrated by Hartley, Zisserman and Liebowitz in different publications, and summarized in Hartley and Zisserman (2000), it is possible to obtain all the parameters needed to calibrate a camera from three *orthogonal* vanishing points. From the mentioned works, a lot of research has been done to calibrate cameras in architectural environments (Rother, 2002; Tardif, 2009, etc....). All these methods are based on scenarios

* Corresponding author. Tel.: +34 91 885 6702.
E-mail address: sergio.alvarez@aut.uah.es (S. Álvarez).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.08.050

where the large number of orthogonal lines provide an easy way to obtain the three orthogonal vanishing points.

Nevertheless, in absence of so strong structures, as usual in the case of traffic scenes, the vanishing point-based calibration is not applicable. In this context, a different possibility is to make use of object motion. The complete camera calibration work using this idea was introduced by Lv et al. (2006). The method uses a tracking algorithm to obtain multiple observations of a person moving around the scene. Three orthogonal vanishing points are then computed by extracting head and feet positions in their leg-crossing phases. The approach requires accurate localization of these positions, which is a challenge in traffic surveillance videos. Furthermore, the localization step uses FFT based synchronization of a person's walk cycle, which requires constant velocity motion along a straight line. Finally it does not handle noise models in the data and assumes constant human height and planar human motion, so the approach is really limited. Based on this knowledge, in Junejo (2009) it is proposed a quite similar calibration approach for pedestrians walking on uneven terrains. There are no restrictions as with Lv's work, but the intrinsic parameters are estimated by obtaining the infinite homography from all the extracted points in multiple cameras.

To manage these inconveniences, the solution lies in computing the three vanishing points by studying three orthogonal components with parallel lines in the moving objects or their motion patterns. Zhang et al. (2013) presented a self-calibration method using the orientation of pedestrians and vehicles. The method seems to extract a vertical vanishing point from the main axis direction of the pedestrian trunk, perpendicular to the ground plane. Additionally, two horizontal vanishing points are extracted by analysing the histogram of oriented gradients of moving cars. The idea is interesting and it was initially implemented for this work. However, the straight vehicles used by Zhang differ from the modern ones, usually with more irregular and rounded shapes. Finally, the pedestrian detection step is not described and results are not depicted in the paper. Hodlmoser et al. (2010) present a different approach. They use zebra-crossings with known metrics to obtain the ground plane information, and pedestrians to obtain the vertical lines. The problem is the maximum distance that the camera can be from the ground and the necessity of knowing real distances from the scene.

In this paper, a self-calibration procedure based on vanishing points is presented. It is done through a hierarchical process which covers most of traffic scenarios and possible configurations. The objective is to obtain both intrinsic and extrinsic camera parameters without restrictions in terms of constraints (restrictions mentioned in previous paragraphs, vehicles driven in only one road direction (Hue et al., 2008), deprecated camera roll (Schoepflin and Dailey, 2003), etc.) or the need of prior information, except for the camera height.

After the present introduction, the remainder of the document is organized as follows. Section 2 describes the developed camera auto-calibration method, based on vanishing points, and the hierarchical system proposed. In Section 3, an application of this technique in the context of traffic surveillance is depicted with the developed segmentation and tracking algorithms. The results obtained are presented and discussed in Section 4 and finally Section 5 contains the conclusions and future work.

2. Camera autocalibration

The camera model used and the equations to obtain the calibration parameters from orthogonal vanishing points are described in the previous paper, Álvarez et al. (2013). In summary, the conclusion is that it is possible to calibrate a camera if the principal point and two orthogonal vanishing points are known; or by computing the principal point as the orthocentre of the triangle formed by three orthogonal vanishing points as vertices. The current work is focused on the way to extract these points from common elements of traffic scenarios. Fig. 1 summarizes the proposed camera calibration process.

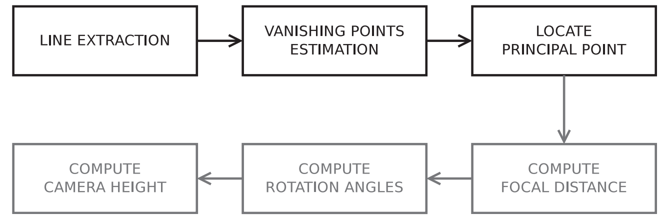

Fig. 1. Camera auto-calibration process.

2.1. Hierarchical auto-calibration

This section presents the proposed method to extract the vanishing points from the image through a hierarchical process. Depending on which elements appear in the scene and the chance of using camera zoom, 5 levels have been established to determine the hierarchy of each developed method and the priority of the solution adopted. Before presenting the hierarchical tree of Fig. 2, and to make its comprehension easier, the different options developed to obtain the vanishing points and optical center are described in the following paragraphs:

- *Zoom:* when zooming, if several features of the image are matched between frames they converge in a common point which corresponds to the optical center.
- *Crosswalk (cross):* the alternate white and gray stripes painted on the road surface provide a perfect environment to obtain two perpendicular sets of parallel lines. It means that two vanishing points of the ground plane can be extracted.
- *Pedestrians (ped):* humans are roughly vertical while they stand or walk. This characteristic makes them very useful to extract perpendicular lines to the ground.
- *Vehicle motion (vmot/vperp):* if one vanishing point from the ground plane is needed, it can be obtained from vehicles moving along the main motion direction (vmot). In case of a perpendicular intersection (in 3D coordinates), vehicles along the two main directions will provide perpendicular sets of parallel lines corresponding to the two ground plane vanishing points (vperp).
- *Structured scene (struct):* in case of scenes with a considerable number of architectural elements, the orthogonal vanishing point extraction can be done by brute force gradient analysis.
- *Optical center assumption (OC):* when it is not possible to obtain one of the three vanishing points, the optical cener can be assumed as the center of the image, although a small error is committed.

The different possible cases are also summarized in Fig. 2 and Table 1.

2.2. Principal point through camera zoom

When zooming, if several features of the image are matched between frames, the lines which join the previous and new feature positions converge in a common point which corresponds with the optical center. This effect is demonstrated in Álvarez et al. (2013) and represented in Fig. 3: an image was taken before and