Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, October 6-9, 2013

Camera auto-calibration using zooming and zebra-crossing for traffic monitoring applications

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

Abstract— **This paper describes a camera auto-calibration system, based on monocular vision, for applications in the framework of Intelligent Transportation Systems (ITS). Using camera zoom and a very common element of urban traffic infrastructures as it is a zebra crossing, a principal point and vanishing point extraction is proposed to obtain an automatic calibration of the camera, without any prior knowledge of the scene. This calibration is very useful to recover metrics from images or apply 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* cameras in fully self-adaptive mode. In the paper, the results achieved up to date in real traffic conditions are presented and discussed.**

Index Terms— **Camera auto-calibration, pan-tilt-zoom cameras, vanishing points, urban traffic infrastructures.**

I. INTRODUCTION

Recently, a lot of research has been carried out on ITS to detect vehicles and pedestrians using vision from traffic infrastructures. Nevertheless very few address the problems of complex urban environments, the adaptability to every condition or the chance to vary the position, angle or zoom of the camera in order to make the system as versatile as possible. Before starting to program a computer vision algorithm, one of the first questions to make is related to the size of the targets. The fact is how far the camera from the objects is, because the size depends on the distance. In case of traffic applications, the position of the camera is totally random, and it is different from one infrastructure to another. Therefore, if the goal is to develop a "plug&play" system, the approximate dimensions of the objects are needed, and that is possible trough a camera calibration.

Camera calibration, is a fundamental stage in computer vision, essential for many applications. The process is the determination of the relationship between a reference plane and the camera coordinate system (extrinsic parameters), and between the camera and the image coordinate system (intrinsic parameters). These parameters are very useful to recover metrics from images or apply prior information of 3D models to estimate 2D pose of targets, making object detection and tracking more robust to noise and occlusions.

In previous papers [1], [2], the authors presented a target detection system for transport infrastructures based on manual camera calibration through vanishing points. The main goal of the current work is to extend the camera calibration method proposed in [1] for target detection in traffic monitoring applications by means of an automatic calibration process based on two main restrictions. First, camera zooming has to be applied as an initialization step to compute the camera optical center. Second, we need the presence of at least one zebra-crossing in the scene to automatically detect two vanishing points. Thus, both intrinsic and extrinsic camera parameters can be computed. The proposed approach does not need the presence of architectural elements. No prior knowledge of the scene or targets is needed. Furthermore, the algorithm is independent of the camera position, and it is able to work with variable pan-tilt-zoom cameras in fully self-adaptive mode.

S. Álvarez, D. F. Llorca, M. A. Sotelo are with the Computer Engineering Department, Polytechnic School, University of Alcalá, Madrid, Spain. email: sergio.alvarez,llorca, sotelo@aut.uah.es.

II. RELATED WORK

The standard method to calibrate a camera is based on a set of correspondences between 3D points and their projections on image plane [3], [4]. 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 changes its angles or zoom is not feasible. One novel method which solves this problem is the orthogonal calibration proposed in [5]. 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 approach depends 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. Since most of these applications make use of only one static camera, auto-calibration cannot be achieved from camera motion, but from inherent structures or flow patterns of the scene. One of the distinguished features of 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*. In [6] a new method for camera calibration using simple properties of vanishing points was presented. In their work the intrinsics were recovered from a single image of a cube. In a second step, the extrinsics of a pair of cameras were estimated from an image stereo pair of a suitable planar pattern. The technique was improved in [7], computing both intrinsic and

978-1-4799-2914-613/$31.00 ©2013 IEEE

extrinsics from three vanishing points and two reference points from two views of an architectural scene. However these assumptions were incomplete, because as demonstrated in [3], it is possible to obtain all the parameters needed to calibrate a camera from three *orthogonal* vanishing points.

From the works mentioned before, a lot of research has been done to calibrate cameras in architectural environments [8], [9]. All these methods are based on scenarios where the large number of orthogonal lines provide an easy way to obtain the three orthogonal vanishing points, just taking the three main directions of parallel lines. 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 in [10]. The method uses a tracking algorithm to obtain multiple observations of a person moving around the scene; computing the three orthogonal vanishing points 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, [11] proposed a quite similar calibration approach for pedestrians walking on uneven terrain. Although there are no restrictions, the intrinsics are estimated by obtaining the infinite homography from all the extracted points in multiple cameras.

To manage such 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. In [12] a self-calibration method using the orientation of pedestrians and vehicles was presented. The method extracts a vertical vanishing point from the main axis direction of the pedestrian trunk. Additionally, two horizontal vanishing points are extracted by analysing the histogram of oriented gradients of moving cars. However, the straight lines of the vehicles used in [12] differ from the modern ones, usually with more irregular and rounded shapes. Finally, the pedestrian detection step is not described and results are not presented in the paper.

III. CAMERA AUTO-CALIBRATION

A. Camera calibration from vanishing points

For a pin-hole camera, and with the common assumption of zero skew and unit aspect ratio, perspective projection from the 3D world to an image can be represented in homogeneous coordinates by the following expression:

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

where (u,v) and (X,Y,Z) are the respective pixel and world coordinates of a point, f is the focal length of the camera, (u_0,v_0) are the pixel coordinates of the principal point, R_{jk} are the elements of the Rotation Matrix and T_i is the Translation Vector.

To compute the intrinsics and the rotation angles, the origin of the world coordinate system (WCS) is placed on the ground plane, and it is initially aligned with the camera coordinate system (CCS). Then, it is translated to T, followed by a rotation around the Y-axis by angle *yaw* (α), a rotation around the X-axis by angle *pitch* (β), and finally, a rotation around the Z-axis by angle *roll* (γ). Therefore, as there are four unknown variables: the focal length f and the rotation angles α, β and γ; four expressions are needed.

A vanishing point V_x is defined at infinity, in homogeneous 3D coordinates, as $[1,0,0,0]^T$. Applied to Equation (1) with the CCS aligned to the WCS ($T=0$), it is possible to obtain useful relationships to find the value of the searched variables:

$$\begin{cases} u_{v_x} &= f\dfrac{R_{11}}{R_{31}} + u_0 \\ v_{v_x} &= f\dfrac{R_{21}}{R_{31}} + v_0 \end{cases} \quad (2)$$

In a similar way a vanishing point V_y is defined at infinity, in homogeneous 3D coordinates, as $[0,1,0,0]^T$. Following the same previous steps an analogous equation is obtained:

$$\begin{cases} u_{v_y} &= f\dfrac{R_{12}}{R_{32}} + u_0 \\ v_{v_y} &= f\dfrac{R_{22}}{R_{32}} + v_0 \end{cases} \quad (3)$$

Combining Equations (2) and (3) the necessary expressions are obtained:

$$\begin{cases} u_{v_x} &= f\dfrac{\cos\gamma\cot\alpha}{\cos\beta} + f\sin\gamma\tan\beta + u_0 \\ v_{v_x} &= -f\dfrac{\sin\gamma\cot\alpha}{\cos\beta} + f\cos\gamma\tan\beta + v_0 \\ u_{v_y} &= -f\sin\gamma\cot\beta + u_0 \\ v_{v_y} &= -f\cos\gamma\cot\beta + v_0 \end{cases} \quad (4)$$

The variable isolation is not a complicated task but a little bit laborious. Hence, for the sake of clarity it is summarized into the final expressions:

$$roll = \gamma = \tan^{-1}\left(\frac{u_{v_y} - u_0}{v_{v_y} - v_0}\right) \quad (5)$$

$$f = \sqrt{(\sin\gamma(u_{v_x} - u_0) + \cos\gamma(v_{v_x} - v_0))(\sin\gamma(u_0 - u_{v_y}) + \cos\gamma(v_0 - v_{v_y}))} \quad (6)$$

$$pitch = \beta = tan^{-1}\left(-\frac{f\sin\gamma}{u_{v_y} - u_0}\right) \quad (7)$$

$$yaw = \alpha = tan^{-1}\left(\frac{f\cos\gamma}{(u_{v_x} - u_0)\cos\beta - f\sin\gamma\sin\beta}\right) \quad (8)$$

978-1-4799-2914-613/$31.00 ©2013 IEEE

Although in theory the sign of the term under square root in Equation (6) should be always positive, it can be negative in practice. That is a good indicator of a wrong vanishing point estimation, to repeat the extraction process.

The goal is therefore to extract two orthogonal vanishing points and the principal point of the image (u_0, v_0).

B. Principal point estimation through camera zoom

Usually the objective of auto calibration approaches is to find three orthogonal vanishing points and compute the principal point as the orthocenter of the triangle formed by the three of them. However, if the equations are analysed, after this step only two points are required. Therefore, if it is possible to find the principal point, only two additional vanishing points are necessary.

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. To demonstrate this phenomenon the situation of Figure 1 is outlined.

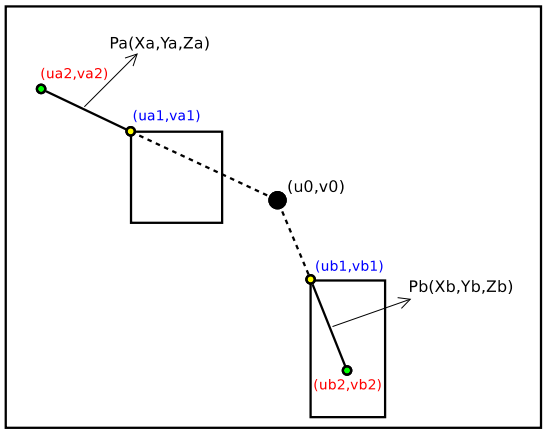


Fig. 1. Situation to analyse the relation between zoom and optical flow.

The objective is to find if the segments which join (u_{a2}, v_{a2}) to (u_{a1}, v_{a1}) and (u_{b2}, v_{b2}) to (u_{b1}, v_{b1}) have a common point corresponding to the optical center. For this purpose it is necessary to use the pin-hole camera model to obtain a geometric relationship between the 3D point, which does not change with zoom, and the point in the image which change with the focal length ($f1 \rightarrow f2$):

$$\begin{cases} u &= f\frac{X}{Z} + u_0 \\ v &= f\frac{Y}{Z} + v_0 \end{cases} \tag{9}$$

With simple geometric line analysis it is known that the lines which pass through (u_{a1}, v_{a1}) and (u_{b1}, v_{b1}) are:

$$\begin{cases} v - v_{a_1} &= m_a(u - u_{a_1}) \\ v - v_{b_1} &= m_b(u - u_{b_1}) \end{cases} \tag{10}$$

where m_i is the slope of the lines with the form:

$$\begin{cases} m_a = \dfrac{v_{a_2} - v_{a_1}}{u_{a_2} - u_{a_1}} = \dfrac{(f_2\frac{Y_a}{Z_a} + v_0) - (f_1\frac{Y_a}{Z_a} + v_0)}{(f_2\frac{X_a}{Z_a} + u_0) - (f_1\frac{X_a}{Z_a} + u_0)} = \dfrac{Y_a}{X_a} \\ m_b = \dfrac{v_{b_2} - v_{b_1}}{u_{b_2} - u_{b_1}} = \dfrac{(f_2\frac{Y_b}{Z_b} + v_0) - (f_1\frac{Y_b}{Z_b} + v_0)}{(f_2\frac{X_b}{Z_b} + u_0) - (f_1\frac{X_b}{Z_b} + u_0)} = \dfrac{Y_b}{X_b} \end{cases} \tag{11}$$

Therefore isolating a point (u, v), the following expression is derived:

$$v = \frac{Y_a}{X_a}(u - u_0) + v_0 \tag{12}$$

And finally if $u = u_0 \rightarrow v = v_0$.

To detect when the camera is zooming and compute the principal point, the motion of static feature points of the image is captured. These features are extracted and matched with SURF [13], between the current image and the background model extracted by the background subtraction algorithm presented in the previous author's work [1]. After that, the neighbourhood of each point is represented by a feature vector and matched between the images, based on Euclidean distance. If the detected motion is bigger than a simple shaking (experimentally established with a threshold) and the motion vectors are concurrent, the movement is considered as zoom and the principal point extracted as the intersection point. The computation time of this process depends on the zoom velocity and the number of SURF features matched. Usually between 5-10 frames at a rate of 15 frames per second.

C. Zebra crossing vanishing point extraction

A common intersection scenario usually has zebra crossings like the one presented in Figure 2.

Fig. 2. Example of zebra crossing.

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 the two vanishing points from the ground plane can be obtained.

To detect if there are crosswalks in the image for a posterior analysis, the following steps are done.

- **Background model estimation:** by the background subtraction algorithm mentioned before, the background model is extracted to look for crosswalk candidates without moving objects that can occlude them, or sudden illumination changes.
- **Thresholding:** as the typical zebra crossing has a strong white component, a thresholding step is done in order to highlight the white stripes.

978-1-4799-2914-613/$31.00 ©2013 IEEE

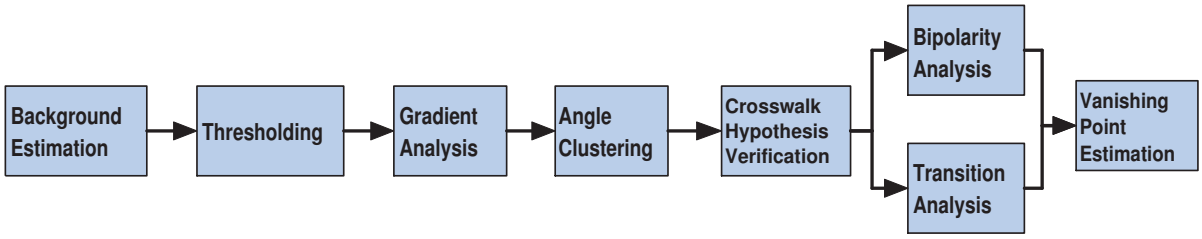


Fig. 3. Crosswalk detection process.

- **Gradient analysis:** the line extraction algorithm explained in another work published by the author's [14] is used in order to obtain the straight lines of the scene, necessary for the vanishing point estimation.
- **Angle clustering:** all the lines extracted are initially grouped by angle in order to distinguish between different kind of candidates. To separate lines with close angles but from different crosswalk candidates a RANSAC filter is applied. The input of the algorithm is the distance from each line to the rest of the cluster. Segments that do not belong to the neighbourhood are included in a different cluster or discarded.
- **Verify crosswalk hypothesis:** a confidence factor of each candidate is taken in order to decide if whether or not it can be consider as a zebra crossing. In the case of more than one valid candidate, the system chooses the one with the highest confidence factor, based on:

  1) *Bimodal analysis*. A gray color based histogram is constructed to analyse the bimodal component of a crosswalk. In case of a zebra crossing, this histogram should have two representative gaussian components, as shown in Figure 4(b).
  2) *Transition analysis*. The b/w transitions (in the binary image) are analysed in order to measure the number of changes and how constant the width of the stripes is. This process is done through a transitions binary pattern constructed by the values of the line which best represents the direction of the crosswalk. This line is obtained fitting by RANSAC the center of the gradient lines extracted for each zebra crossing.

  The corresponding gradients (in yellow), representing line (in red), bimodal histogram and transition pattern of the crosswalk of Figure 2 are represented in Figure 4.
- **First vanishing point estimation:** The vanishing point corresponding to the main direction of the crosswalk stripes is computed as also explained in [1], with the gradients extracted previously.
- **Second vanishing point estimation:** Due to the small size and the irregularity of the perpendicular segments of the stripes, the gradient analysis is not accurate enough to obtain the desired set of parallel lines. To solve this problem, the centroid of each segment is computed as the intersection of the central line of the stripe with the end of the stripe. All the points obtained are fitted to a line by RANSAC and the intersection between the upper and lower lane is consider the second vanishing point. The process is represented in Figure 5.

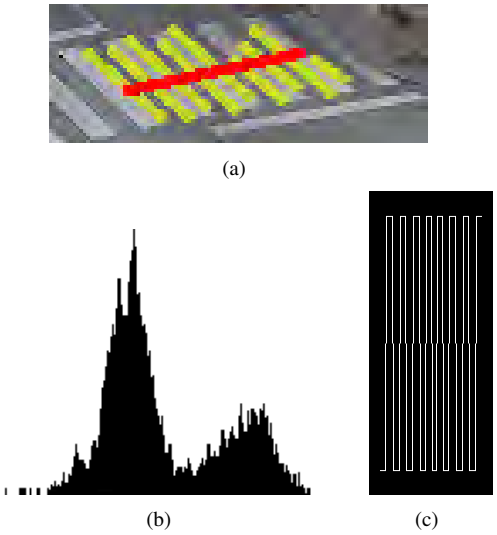


Fig. 4. Confidence factor indicators of a crosswalk. (a) Gradients and fitted representing line. (b) Bimodal histogram. (c) Transitions binary pattern.

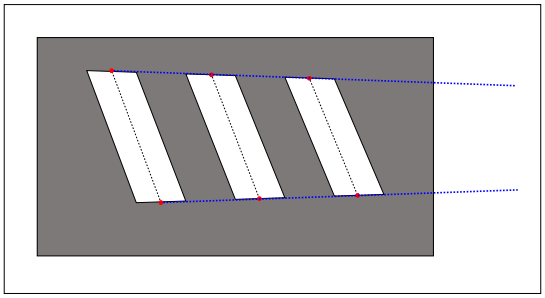


Fig. 5. Extraction of the second vanishing point from a crosswalk.

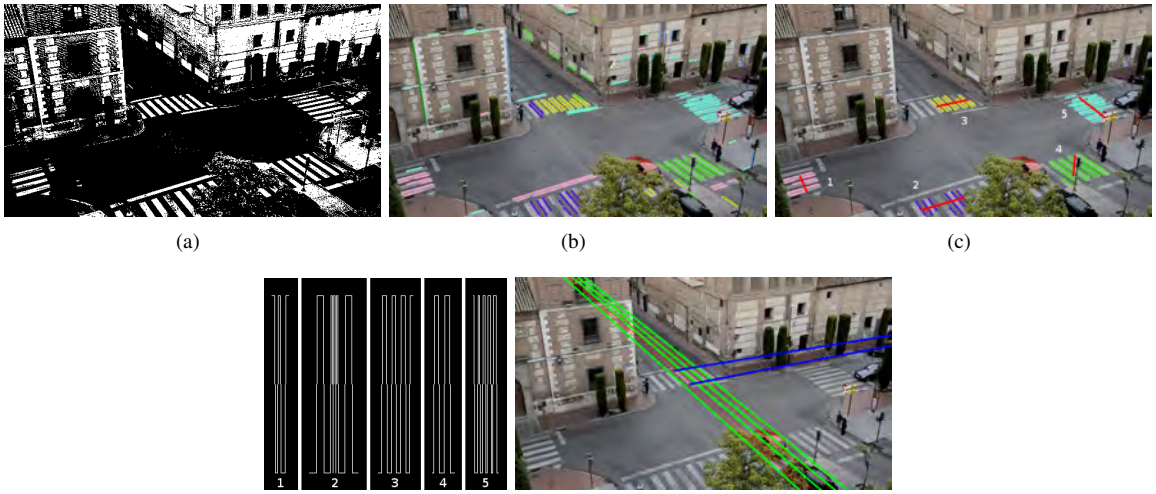
IV. EXPERIMENTS

The proposed approach is evaluated using the calibration based on manual vanishing point extraction presented in [1] as the groundtruth. Firstly, the two steps of the algorithm (principal point computation and vanishing points extraction) are depicted in Figures 6 and 7 with their application in one scenario. After that, a second scenario is presented to show the performance of the method in both experiments.

978-1-4799-2914-613/$31.00 ©2013 IEEE

(a) (b) (c)

Fig. 6. Principal point computation through camera zoom. (a) Image before zooming and extracted features. (b) Image after zooming and extracted features. (c) Feature matching. The common point corresponds to the optical center.



(a) (b) (c)

(d) (e)

Fig. 7. Crosswalk detection example. (a) Binarized background model. (b) Line extraction. (c) Grouped candidates with testing lines in red. (d) Transition pattern of candidates 1 to 5. (e) Parallel lines to compute the vanishing points.

Figure 6 shows an example of the principal point computation: an image was taken before and after zooming and the matched features converge to the principal point. To compute the intersection point, a RANSAC-based algorithm has been developed to delete wrong lines (outliers in red). The crosswalk detection method is illustrated in the Figure 7. Firstly, the background model image is binarized, and the lines are extracted by gradient analysis and grouped by angle. After that, a RANSAC-based filter is applied to get the final candidates. The red line is the one which best fits the candidate. Bimodal and transition analysis is then done in order to obtain the confidence factor shown in Table I.

TABLE I

CONFIDENCE FACTOR FOR EACH CROSSWALK.

Candidate	Confidence	Description
1	0.10	Irregular pattern
2	0.14	White stripe with black holes
3	0.96	Chosen candidate
4	0.40	Traffic light occlusion
5	0.77	Acceptable but irregular

Two representative scenes have been selected to show the performance of the approach. For the sake of clarity, the description of the variables used is presented:

- *OC*: computed principal point of the image.
- *Focal, pitch and roll:* values of the computed intrinsic and extrinsic camera parameters. Yaw is not consider because its variation does not modify the ground plane and does not have impact into the 3D projection.
- $dist_i$*:* 3D depth distance from the camera to three points of the image. The distance, computed with the equations of the pin-hole camera model (assuming the points in the ground plane), is compared to the one obtained by the Google Maps tool [15]. Figure 8(b) shows an example of how it is extracted from the website.
- vol_i*:* volumes of the projected prisms over three vehicles, assuming a fixed 3D standard size.

The tests were performed in two sequences recorded from the top of a tower (Figure 8(a)), and the obtained results are illustrated in Figure 9 with the projected prisms of three vehicles. The resolutions of the images is 640×480.

To analyse the results obtained in the test, Table II summarizes all the values extracted and computed by the system, compared with the groundtruth of the semi-automatic approach of [1]. The performance of the method has been described through the results obtained by two selected videos. 8 more sequences from different scenarios and conditions have been used to test the developed auto-calibration method. As a result, the following average errors on the camera parameters are extracted: $f = 3.85\%$, $pitch = 2.08^\circ$ and $roll = 0.52^\circ$.

978-1-4799-2914-613/$31.00 ©2013 IEEE

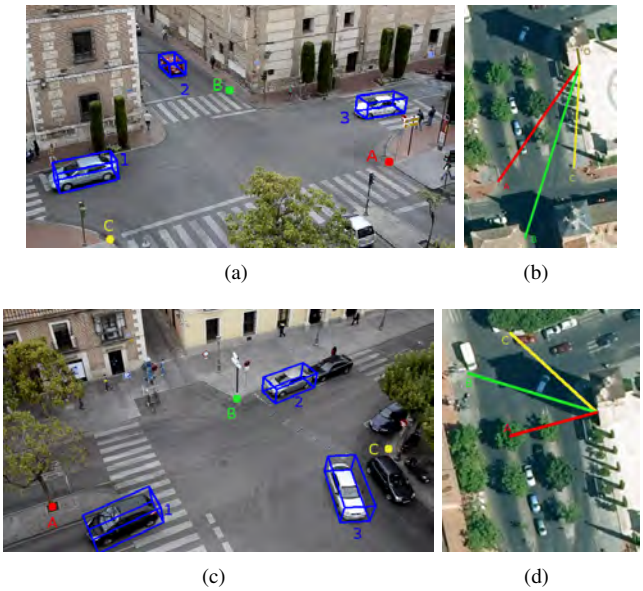
TABLE II

AUTO-CALIBRATION RESULTS FOR SCENARIOS 1 AND 2.

Scene	OC	FOCAL	PITCH	ROLL	distA	distB	distC	vol1	vol2	vol3
Groundtruth 1	(320.27,180.10)	685.35	-24.89	0.28	39	50	29	41271	8190	22142
Scene 1	(325.43,187.64)	700.52	-23.61	0.03	39.79	51.30	30.77	39296	6419	18706
Groundtruth 2	(321.67,182.49)	578.84	-39.23	0.13	24	33	29	67291	22220	66644
Scene 2	(322.25,184.02)	592.58	-41.57	0.38	26.38	33.05	30.04	82171	31655	80796

(a) (b)

Fig. 8. (a) Torre de Santa María, where the camera was located. (b) Example of distance extraction from the tower, with Google Maps.



(a) (b)

(c) (d)

Fig. 9. Scenarios used for the experiments and graphic results of the approach. (a) Scenario 1, selected points and projected volumes. (b) Measured distances from Google Maps. (c) Scenario 2, selected points and projected volumes. (d) Measured distances from Google Maps.

V. CONCLUSION

In this paper, a camera auto-calibration approach based on vanishing points has been presented. The objective is to extend the work proposed by the author's in [1] with an automatic calibration process based on camera zoom and crosswalk detection.

The performance of the method has been described through the results obtained by two selected videos, although 8 more sequences have been used to test the system. The obtained results are really satisfactory: the low error of the 3D prisms projections and distance measurements proves the strength of the method. Furthermore, the system is able to adapt the calibration parameters in case of PTZ camera displacements without manual supervision.

Future work will include an hierarchical procedure to calibrate the camera in presence of different elements of the scene, to create a robust multi-lever camera calibration which can provide high versatility to cover most of the possible traffic scenarios and configurations without any restriction in terms of constraints or the need of prior knowledge.

VI. ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science and Innovation under Research Grant ONDA-FP TRA2011-27712-C02-02.

REFERENCES

[1] S. Álvarez, D. F. Llorca, M. A. Sotelo, and A. G. Lorente, "Monocular target detection on transport infrastructures with dynamic and variable environments," in *IEEE Intelligent Transportation Systems Conference*, 2012.

[2] S. Álvarez, M. A. Sotelo, D. F. Llorca, and R. Quintero, "Monocular vision-based target detection on dynamic transport infraestructures," in *Lecture Notes in Computer Science*, 2011, pp. 576–583.

[3] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2000.

[4] R. Tsai, "An efficient and accurate camera calibration technique for 3d machine vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1986.

[5] Z. Kim, "Camera calibration from orthogonally projected coordinates with noisy-ransac," in *Proceedings of the IEEE Workshop on Application of Computer Vision*, 2009.

[6] B. Caprile and V. Torre, "Using vanishing points for camera calibration," *International Journal of Computer Vision*, vol. 4, pp. 127–140, 1990.

[7] R. Cipolla, T. Drummond, and D. Robertson, "Camera calibration from vanishing points in images of architectural scenes," 1999.

[8] C. Rother, "A new approach to vanishing point detection in architectural environments," *Image and Vision Computing*, vol. 20, pp. 647–655, 2002.

[9] J. P. Tardif, "Non-iterative approach for fast and accurate vanishing point detection," in *Proceedings of the IEEE Conference on Computer Vision*, 2009.

[10] F. Lv, T. Zhao, and R. Nevatia, "Camera calibration from video of a walking human," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1513–1518, 2006.

[11] I. N. Junejo, "Using pedestrians walking on uneven terrains for camera calibration," in *Machine Vision and Applications*, vol. 22, 2009, pp. 137–144.

[12] Z. Zhang, M. Li, K. Huang, and T. Tan, "Camera auto-calibration using pedestrians and zebra-crossings," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2011, pp. 1697–1704.

[13] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008.

[14] D. F. Llorca, S. Álvarez, and M. A. Sotelo, "Vision-based parking assistance system for leaving perpendicular and angle parking lots," in *IEEE Intelligent Vehicle Symposium*, 2013.

[15] G. Google Maps, "https://maps.google.es/," 2013.