Calibrating an Outdoor Distributed Camera Network using Laser Range Finder Data

Agustin Ortega, Bruno Dias, Ernesto Teniente, Alexandre Bernardino, José Gaspar and Juan Andrade-Cetto

Abstract— Outdoor camera networks are becoming ubiquitous in critical urban areas of large cities around the world. Although current applications of camera networks are mostly limited to video surveillance, recent research projects are exploiting advances on outdoor robotics technology to develop systems that put together networks of cameras and mobile robots in people assisting tasks. Such systems require the creation of robot navigation systems in urban areas with a precise calibration of the distributed camera network. Despite camera calibration has been an extensively studied topic, the calibration (intrinsic and extrinsic) of large outdoor camera networks with no overlapping view fields, and likely to suffer frequent recalibration, poses novel challenges in the development of practical methods for user-assisted calibration that minimize intervention times and maximize precision. In this paper we propose the utilization of Laser Range Finder (LRF) data covering the area of the camera network to support the calibration process and develop a semi-automated methodology allowing quick and precise calibration of large camera networks. The proposed methods have been tested in a real urban environment and have been applied to create direct mappings (homographies) between image coordinates and world points in the ground plane (walking areas) to support person and robot detection and localization algorithms.
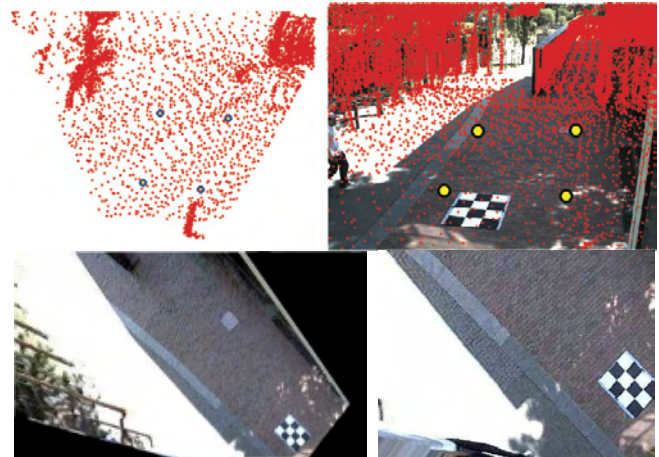
Fig. 1. Results of the proposed calibration system. The top row shows a graphical user interface to select planar regions and the registration of the laser range data on an image view. The bottom row shows recovered orthographic views of the ground plane. The chess pattern shown is not used for calibration, serves just to visually evaluate the quality of the ground-plane rectifying homography.

I. INTRODUCTION

Many urban areas and public buildings around the world have currently large camera networks. Applications have been focused mainly in security and surveillance but new trends in robotics are extending their usage to support the operation of mobile robotic devices in urban areas [1]. The camera network serves as a mean to detect, localize and map environmental information in a globally coherent frame of reference. Persons, robots and other targets must be localized in a unique coordinate system even though they are observed by distant cameras. This is a complex problem since camera networks have few or no overlapping fields of view. Additionally, being an outdoor system, it is constantly susceptible to weather conditions, such as rain and wind, and thus it is expected to have slight but visible positioning

A. Ortega, E. Teniente and J. Andrade-Cetto are with the Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain, {aortega, ehomar, cetto}@iri.upc.edu.

B. Dias, A. Bernardino and J. Gaspar are with the Institute for Systems and Robotics at Instituto Superior Técnico, Technical University of Lisbon, Portugal, {bdias, alex, jag}@isr.ist.utl.pt.

changes from time to time. The calibration methodology must therefore encompass simple self-adjusting mechanisms.

Recently, the development of powerful laser sensors combined with Simultaneous Location and Mapping (SLAM) methodologies [2], [3] allow the possibility to have available high precision Laser Range Finder (LRF) data registered over large areas. These large outdoor LRF datasets have started recently to be acquired also for the purpose of creating robot navigation systems in urban areas. The LRF data is acquired over the complete area of the network and, in particular, contains the areas corresponding to the fields of view of the cameras. This paper exploits this novel technologies proposing a methodology for calibrating an outdoor distributed camera network using LRF data.

The paper contributes in the use of LRF as external information to aid the calibration procedure of the distributed camera network. Whenever cameras have no overlapping view fields it is not possible to estimate the relative position between cameras unless external data is used to refer the camera calibration parameters to a global reference frame. Since calibration inevitably requires some user intervention, in large camera networks this can be a very tedious procedure if one does not develop practical and semiautomated methods that facilitate and speed up user input.

The idea of the approach is the following: in a first stage, the LRF map is registered to an aerial view of the site
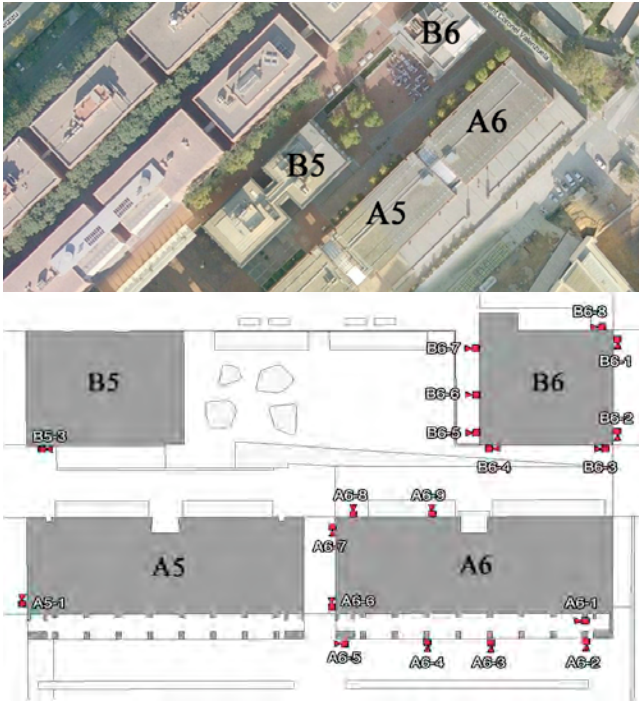
Fig. 2. An aerial view of our experimental site, the Barcelona Robot Lab (top), and the distribution of cameras in the network (bottom).

and the user sets up the position and nominal calibration parameters of the cameras in the network. This allows user selection of an initial camera field of view onto the LRF area of interest likely to be observed by each camera. In a second stage, lines extracted from the LRF area of interest are represented in the nominally calibrated camera coordinate system and are reprojected to the real-time cameras' acquired images. This allows the user to perceive the calibration errors and input information to a non-linear optimization procedure that refines both intrinsic and extrinsic calibration parameters. The optimization process matches 3D lines to image lines. The 3D lines are extracted by intersecting planes on the segmented LRF set. A novel approach to 3D range segmentation based on local variation is used [4]. To show the applicability of the calibration results, homographies of the walking areas are computed.

This work is associated with the European project Ubiquitous Networked Robotics in Urban Settings (URUS), that puts together networks of cameras and mobile robots in people assisting tasks. Fig. 2 shows an aerial view of our application scenario, the Barcelona Robot Lab. installed at the UPC Campus Nord, as well as a floor plan of the outdoor camera network.

The paper is organized as follows. First, related work in the calibration of camera networks is presented. Then, our method to extract 3D lines from available range data sets is explained, and the proposed method to refine the calibration parameters by matching these features with the same features on images is shown. Experiments on a real urban environment are depicted, and finally, conclusions and future work are discussed.

II. RELATED WORK

Different techniques have been proposed to calibrate cameras. Some require using patterns, either planar [5] or non-planar [6], with known metric structure. However, for large outdoor camera networks, calibration patterns of reasonable sizes often project on images with very small resolution, mainly because the cameras are located at a considerable height with respect to the floor; consequently making pattern segmentation difficult. In addition, a pattern-based independent calibration of each camera would require a secondary process to relate all camera coordinate systems to a global reference frame, but establishing this relation with small to null overlapping fields of view is nearly impossible. For planar scenarios, a Direct Linear Transformation (DLT) [7] suffices to estimate image to plane homographies [8]. Unfortunately, the planar scenario assumption is too restrictive, especially in situations with unparallel locally planar surfaces such as ramps and plazas, which often occur in real urban environments, as in our case.

An interesting technique to calibrate the camera network, without the need of a pattern, is with the aid of a bright moving spot [9]. The technique assumes overlapping fields of view to estimate the epipolar geometry of the camera network, to extract homographies, estimate depth, and finally compute the overall calibration of the camera network. In our case the cameras' fields of view seldom overlap, and the visibility of the bright spot does not always hold at sunlight. Another alternative is to place the led light on a moving robot and to have a secondary robot equipped with a laser sensor tracking the first one, relating their position estimates to the camera network [10]. Another system that relies on tracking a moving object to estimate the extrinsic parameters is [11], which assumes a constant velocity model for the target. In contrast to these approaches, we opt for a system that does not rely explicitly on a moving platform to calibrate the network.

For camera network systems that incorporate controlled camera orientation changes (pan and tilt), and motorized zoom, it is possible to use such motion capabilities to first estimate the intrinsic parameters rotating and zooming fitting parametric models to the optical flow, and then to estimate extrinsic parameters aligning landmarks to image features. Unfortunately, in our case, the cameras are not active.

We benefit instead from the availability of a dense LRF dataset acquired during a 3D laser-based SLAM session with our mobile robot mapping devices [12]. The set contains over 8 million points and maps the environment with accuracies that vary from 5 cm to 20 cm approximately. This data replaces the need for a tracked beam, a robot, or active capabilities of the camera network, and is used as external information to calibrate the camera network.

III. CALIBRATION METHODOLOGY

The calibration procedure, illustrated in Fig. 3, consists of two main steps. In the first step, a nominal calibration of the cameras is generated by registering the LRF data to an aerial image of the experimental site, showing both in a graphical
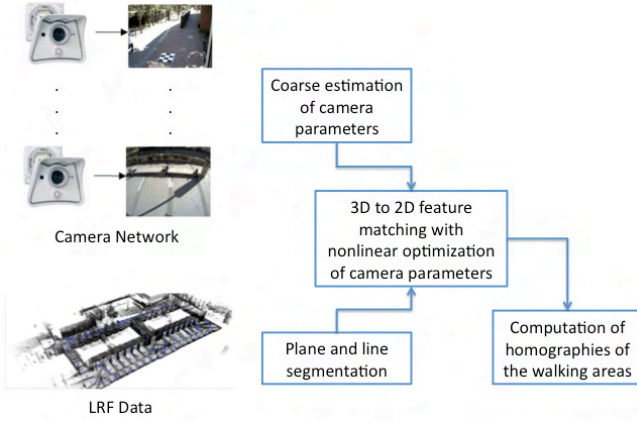
Fig. 3. Distributed camera network calibration methodology.

Fig. 4. Registration of aerial view with the LRF data and visual selection of camera location and orientation.

user interface, and prompting a user to coarsely specify the camera location, orientation, height and field of view. These initial parameters allow the cropping of the entire LRF into regions of interest compatible with the field of view of each camera as shown in Fig. 4.

The second stage aims at refining the cameras nominal calibration by matching, in a semi-automatic manner, 3D features to the corresponding 2D features in the cameras' images. The LRF data corresponding to each cameras' field of view is segmented into a set of best fitting planes with large support from the point clouds and then, straight lines are computed from the intersection of perpendicular planes from the set. The extracted 3D lines are then associated with 2D image lines and this information is fed to a non-linear optimization procedure that improves both intrinsic and extrinsic camera calibration parameters. Finally, homographies of the walking areas are computed by selecting planar regions in the LRF data. The final output of the whole calibration procedure consists in a) the extrinsic camera parameters (the relative position and orientation in the world frame), b) the intrinsic parameters (focal distance, image center and aspect ratio) and c) the homographies of the walking areas.

The first step of the calibration procedure needs to be performed only once, during the camera network installation, or when the network topology changes, i.e., cameras are added/moved. The second step can be executed as frequently as needed to keep the system calibrated despite small modification in camera orientation due to weather conditions and maintenance operations.

A. LRF Registration and Nominal Calibration

The registration of the LRF data with an aerial view of the environment is the first step in the calibration process. To that end, a graphical user interface is developed in which each camera region of interest in the LRF data is selected. Fig. 4 shows the interface where the user coarsely selects the position of a camera and its viewing direction (indicated by the magenta line in the zoomed area). The cameras are set with default intrinsic parameters. The LRF data corresponding to the field of view of each camera can
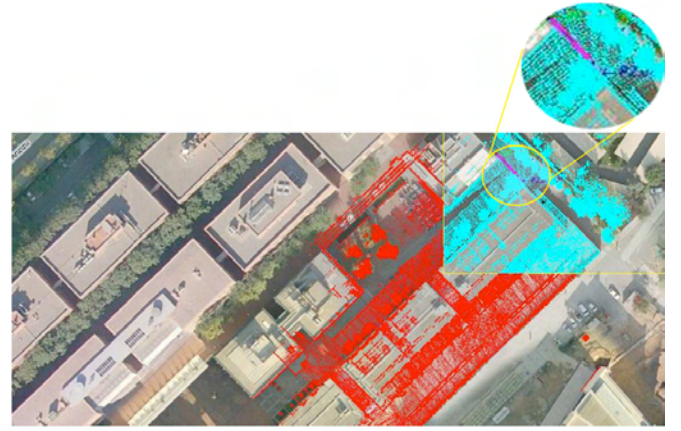
be visually adjusted by manually changing the intrinsic and extrinsic parameters, but this process is only required if the initialization is too erroneous.

The user interaction with the interface for nominal calibration consists of the following steps:

1) pointing the camera location, p_1 in the aerial view;
2) pointing a ground point, p_2 assumed to be in the field of view of the camera;
3) entering an elevation angle, θ;
4) entering an horizontal field of view, ϕ and the aspect ratio of the images.

Steps 3 and 4 will usually have default values, in order to make as simple as possible the task to the user. For instance, $\phi = 40^o$ corresponds to a common 8 mm lens in a 1/4 in CCD. Default values for θ depend on the location, but in our case many cameras are at the level of the second floor (about 6 meters high), imaging objects in the ground plane closer than 20 meters, and thus we have a typical value of $\theta = 17^o$.

With the parameters referred in steps 1 to 4, one can compute completely a pin-hole (perspective), projection model: p_1 and p_2 define the azimuth direction, the elevation is given by the user, and the roll of the camera is assumed null (these three parameters suffice to define the rotation matrix, R); p_1, p_2 and θ define the projection center of the camera, t in world coordinates; the field of view combined with the size of an image, and assuming the principal point equal to the image center, give the intrinsics matrix, K. Hence we obtain the perspective projection model:

$$P(\vartheta_j) = K[R|t] \tag{1}$$

where P denotes the projection matrix and ϑ_j represents a vector containing the listed parameters for camera j.

Note that radial distortion could be included also in the model but in this work we assume that it has been estimated independent from camera installation. That is, we assume that the knowledge of the radial distortion allowed correcting the images and thus obtaining novel images as if they were acquired by a non distorting optical system.

In the second stage of the calibration procedure, the parameters collected in ϑ_j will be refined by matching LRF (3D) and image (2D) straight line features.

B. Improving the Calibration

In order to improve camera calibration from the nominal parameters we could let an operator iteratively change the calibration parameters to find the best visual match between the LRF and its projection over the image. However this is a time consuming process and, in large camera networks it is cumbersome and tedious. Instead we propose a semi-automatic way where relevant 3D lines are automatically extracted from the LRF and the user just has to select points in the corresponding image lines. In practice the method works well with about half dozen lines selected for each camera image.

This procedure is expected to be conducted right after the nominal calibration, which gives just a rough approximation, and whenever the camera's position or orientation is changed, due to weather (wind, rain, etc.) or maintenance operations (repair, cleaning).

1) Extracting Lines: The computation of straight lines from the LRF data relies on identifying and intersecting planes. The method to segment planar regions is motivated by Felzenszwalb's algorithm to 2D image segmentation [13], and extended to deal with non-uniformly sampled 3D range data [4]. The algorithm sorts point to point distances for each point's k-nearest neighbors and then traverses the list of sorted distances in increasing order, growing planar patches by joining points that meet two matching criteria, i.e., distance constraints and orientation constraints. Thanks to the use of union by rank and path compression [14], the algorithm runs nearly in linear time with respect to the number of points in the LRF dataset.

To avoid the bottleneck of finding each point's nearest neighbors, an efficient library for the computation of approximate nearest neighbors is used instead [15]. Then, a plane is fit to each set of neighboring points [16] minimizing the sum

$$\varepsilon = \sum (p_i^T \hat{n} - d)^2$$

for all neighbors to that point p_i. The term $\hat{n}$ is the resulting surface patch normal of the best fit plane, given as the eigenvector associated to the smallest eigenvalue of

$$\left(Q - \frac{q\,q^T}{N^2} \right) \hat{n} - \lambda \hat{n} = 0 \qquad (2)$$

with

$$\varepsilon = \hat{n}^T \underbrace{\left(\sum p_i\, p_i^T \right)}_{Q} \hat{n} - 2d \underbrace{\left(\sum p_i^T \right)}_{q} \hat{n} + N^2 d^2$$

Once local surface normals and planar patches are computed for each point in the LRF dataset, segments are merged by growing a forest of trees based on curvature and mean distance. Curvature is computed from the angle between the normals of two neighboring regions, and for the regions to merge, they must be below a user selected threshold t_c,

$$|cos^{-1}(n_1^T n_2)| < t_c . \qquad (3)$$

For two segments passing the curvature criteria, they can be joined if their distance is below a user selected threshold t_d,

$$\frac{k_1 d_1 + k_2 d_2}{k_1 + k_2} < t_d$$

with

$$d_1 = (p_1 - p_2).n_2$$
$$d_2 = (p_2 - p_1).n_1$$

and k_1 and k_2 are the number of points each segment holds.

Once a set of segments is obtained, their intersecting lines are computed, and the ones with sufficient support from their generating planes, and with good orthogonality conditions are selected for projection onto the images.

2) Optimization Procedure: Given the nominal calibration, the 3D straight lines extracted from the LRF data can now be projected in the image and guide the user to select the corresponding 2D image lines. This 3D-2D association allows improving the nominal calibration by minimizing a cost function containing the camera projection matrix P.

Let $m_i = [u_i\ v_i]^T$ denote points that belong to an image line and $M_i = [X_i\ Y_i\ Z_i\ 1]^T\ i = 1,..,n$ denote the corresponding 3D points on the matching line in the LRF data . The cost function is defined as:

$$\hat{\vartheta}_j = \arg\min_{\vartheta_j} \sum_i \left\| m_i - h\left(P(\vartheta_j) \cdot M_i \right) \right\|^2 \qquad (4)$$

where h is a de-homogenization function, $P(\vartheta_j)$ is the projection matrix of the j-th camera as defined in (1), and ϑ_j are the calibration parameters, namely focal length and principal point, plus the extrinsic parameters for position and orientation. The optimization is solved using Levenberg-Marquardt nonlinear optimization.

C. Computing Homographies

Once the calibration parameters have been improved in each camera, this information is used to compute homographies of planar patches in the ground floor. The idea is to have a practical way to transfer 2D image to 3D world coordinates of targets detected in the images. The algorithm to compute the homographies is the Direct Linear Transformation (DLT) that associates LRF data points in the planes of interest to image points. The patches selected to be represented by homographies are the ones where it is likely to have people walking and where robots are expected to provide services to people. The user selects polygonal regions corresponding to the desired patches and the 3D LRF points inside these patches are used to compute the approximating 3D planes. Notice that this step is only possible having a sufficiently precise projection matrix P so that 3D patches are correctly associated to the selected image regions, otherwise erroneous planar approximations are likely to be computed.
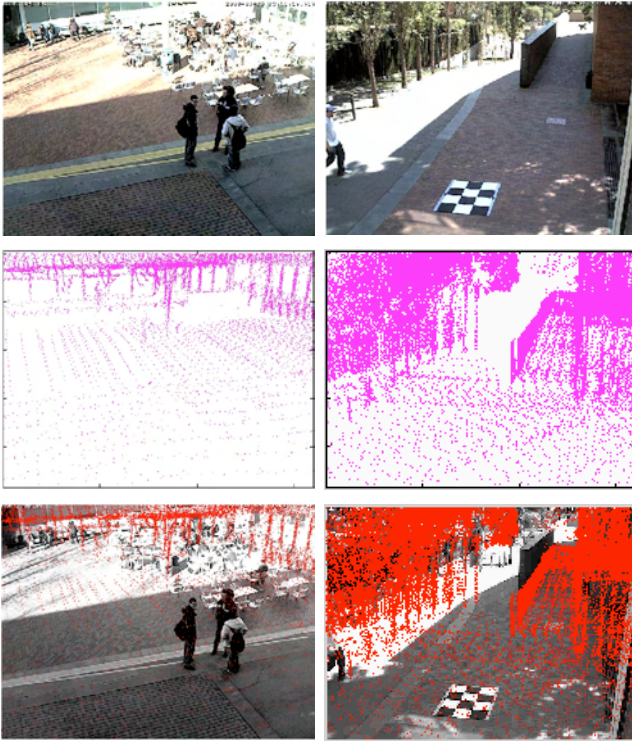
Fig. 5. Nominal calibration of cameras *A6-8* and *B6-2*. Original images (first row), projection of the LRF data to the image plane (second row) and superposition of the projected LRF data over the image (third row). Note the significant registration error in the second column (camera *B6-2*). This error will be corrected during the optimization process.

IV. EXPERIMENTS

In order to test the validity of the proposed calibration methodology we have performed tests at three levels: (i) LRF registration and nominal calibration, (ii) optimization of the calibration, based on improving the LRF data through image registration, (iii) application of the calibration to obtain an orthographic view of the ground plane. As described in the introduction, the experiments were performed in the outdoor non-overlapped camera network of the Barcelona Robot Lab. (see Fig. 2).

Using the registration of the global LRF data with the aerial view, one obtains a first calibration of a camera by pointing in the aerial view two points and using some nominal parameters. Fig. 5 shows two such calibrations, for cameras *A6-8* and *B6-2*. The figure shows the original image taken from each of the cameras, the LRF data in the field of view of the camera, and the 3D LRF data projected over the image. See in part 1 of the accompanying video a demonstration of the nominal calibration phase complemented with some manual improvement.

Given such initial calibration of the cameras, one can now run the optimization procedure described in Sec. III B. Fig. 6 a shows the segmented planes and lines out of the LRF data within the field of view of camera *B6-2*. Each color represents a segmented plane. The parameters that were used in this example to segment the data were: $k_n = 25$ neighbors, and distance and curvature thresholds of $t_d = 0.5$ and

$t_c = 0.8$, respectively. Furthermore, only lines in intersecting orthogonal planes in the interval $[\frac{\pi}{2} - 0.03, \frac{\pi}{2} + 0.03]$ were considered. Frames b and c, show the lines superimposed on the image, before and after optimization, and frame d shows the LRF data projected on the image. Parts 2.1 and 2.2 of the accompanying video show more views of the detected planes and lines, and a sequence of iterations of the optimization procedure.

Once we have the calibration of the cameras, we can relate 3D LRF date to its image counterpart, and vice-versa. One typical application is to observe orthographically the ground plane, i.e. obtaining a bird's eye view by computing an homography as discussed in Sec. III C. Fig. 1, shows the input data, points selected in the LRF (top-left) and their corresponding image points (top-right). The bottom row shows the resulting image and a zoom region on it. Note that as expected, the chess pattern placed in the floor for evaluating the results, is dewarped correctly (perspective effect removed). See also part 3 of the accompanying video detailing the process of selecting a region of interest of the ground plane and obtaining an orthographic view of the selected area.

V. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a methodology for calibrating outdoor distributed camera networks having small or inexistent overlapping fields of view between the cameras. The methodology is based on matching image data with LRF data acquired and registered along the complete area of the network using SLAM methodologies. In a first stage the user obtains the nominal calibration by using default intrinsic parameters for the cameras and indicating their positions and orientations on an aerial view aligned with the LRF data. Next, the calibration of each camera is improved by a semi-automatic optimization procedure detecting lines in the LRF and matching them with image lines. The lines are detected in the LRF by automatically segmenting out planar regions and finding such plane intersections. The optimization procedure minimizes the distances between points in the lines found in the LRF data and their corresponding points in image lines.

Experiments performed in a real outdoor camera network show that the methodology effectively allows calibrating camera networks. In particular the obtained calibration proved to have enough precision to allow the computation of dewarping homographies to observe orthographically the ground plane. In more general terms, the LRF data of the area mapped, actually acquired for robot navigation tasks, was shown to have as a by-product useful calibration information for the camera network.

Future work will focus on a deeper evaluation of the precision and accuracy of the proposed methodology. In addition, alternative primitives available both on the LRF and image data, will be explored to build not only geometric but also algebraic cost functionals, which may mitigate the complexity and further automate the complete calibration process.

(a) Plane intersections.

(b) Projection on images before optimization.

(c) Projection on images after optimization.

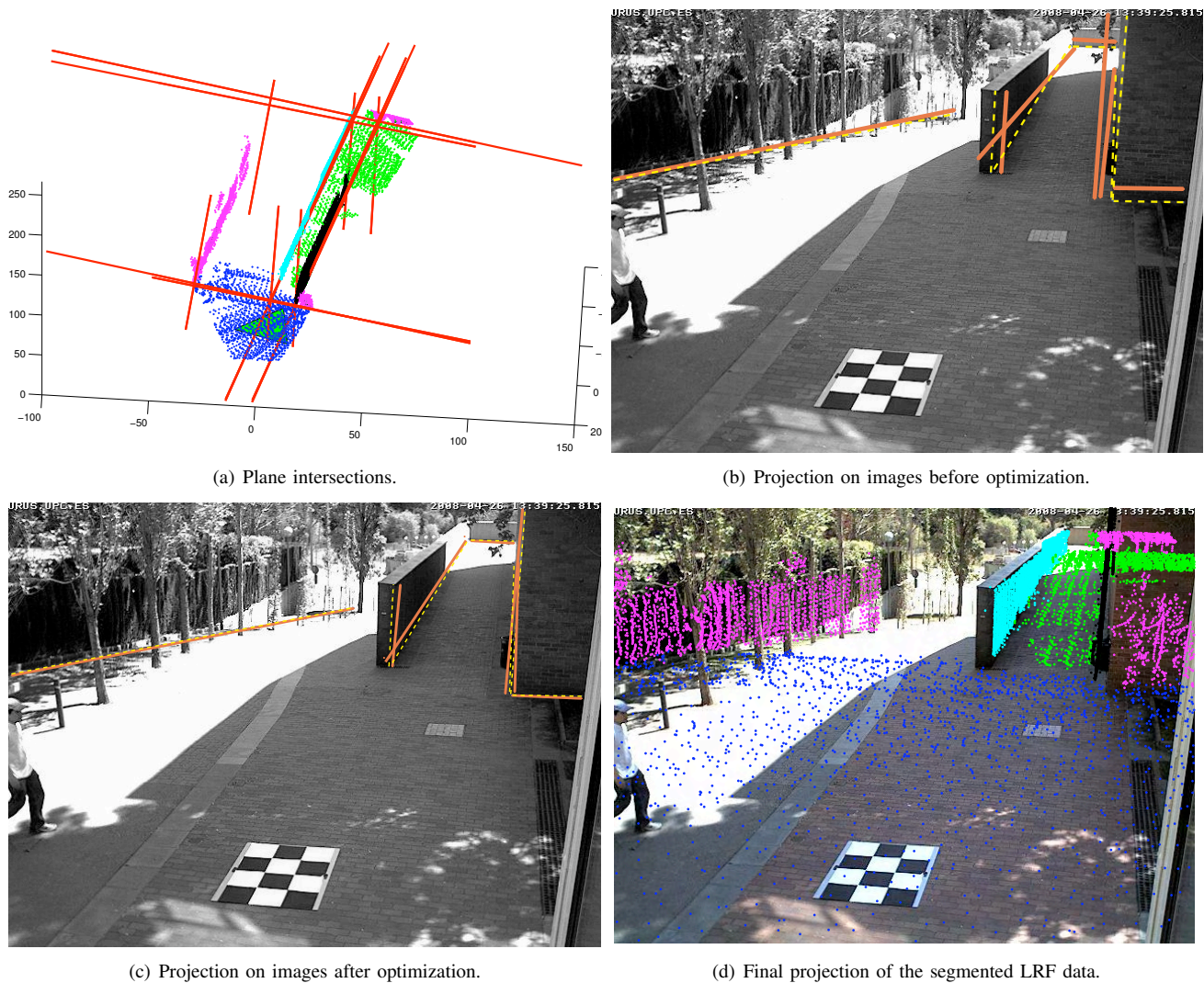(d) Final projection of the segmented LRF data.

Fig. 6. Improving the calibration. Input data is formed by planes and lines (a). The optimization approximates the projected laser lines (red) to the image lines (yellow) (b, c). Cloud of LRF data points projected in the image after the optimization (d).

REFERENCES

[1] A. Sanfeliu and J. Andrade-Cetto, "Ubiquitous networking robotics in urban settings," in *Proc. IEEE/RSJ IROS Workshop Network Robot Syst.*, Beijing, Oct. 2006, pp. 14–18.

[2] V. Ila, J. Andrade-Cetto, R. Valencia, and A. Sanfeliu, "Vision-based loop closing for delayed state robot mapping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, San Diego, Nov. 2007, pp. 3892–3897.

[3] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge: MIT Press, 2005.

[4] A. Ortega, I. Haddad, and J. Andrade-Cetto, "Graph-based segmentation of range data with applications to 3d urban mapping," in *Proc. European Conf. Mobile Robotics*, Dubrovnik, Sep. 2009.

[5] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.

[6] R. Tsai, "A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras," *IEEE J. Robot. Automat.*, vol. 3, no. 4, pp. 323–344, Aug. 1987.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge: Cambridge University Press, 2004.

[8] K. Okuma, J. Little, and D. G. Lowe, "Automatic rectification of long image sequences," in *Proc. Asian Conf. Comput. Vision*, Jeju Island, Jan. 2004.

[9] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multicamera self-calibration for virtual environments," *Presence: Teleop. Virt.*, vol. 14, no. 4, pp. 407–422, 2005.

[10] T. Yokoya, T. Hasegawa, and R. Kurazume, "Calibration of distributed vision network in unified coordinate system by mobile robots," in *Proc. IEEE Int. Conf. Robot. Automat.*, Pasadena, Apr. 2008, pp. 1412–1417.

[11] A. Rahimi, B. Dunagan, and T. Darrell, "Simultaneous calibration and tracking with a network of non-overlapping sensors," in *Proc. 18th IEEE Conf. Comput. Vision Pattern Recog.*, Washington, Jul. 2004, pp. 187–194.

[12] R. Valencia, E. Teniente, E. Trulls, and J. Andrade-Cetto, "3D mapping for urban serviece robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Saint Louis, Oct. 2009.

[13] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[14] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. Cambridge: MIT Press, 1992.

[15] D. M. Mount and S. Arya, "ANN: A library for approximate nearest neighbor searching," in *Proc. 2nd Fall Workshop on Computational and Combinatorial Geometry*, Durham, Oct. 1997.

[16] J. Andrade-Cetto and A. C. Kak, "Object recognition," in *Wiley Encyclopedia of Electrical and Electronics Engineering*, J. G. Webster, Ed. New York: John Wiley & Sons, 2000, supplement 1, pp. 449–470.