# Relocation using Laser and Vision

D. Ortín, J. Neira, J.M.M. Montiel

Computer Science Dept. of the University of Zaragoza,
María de Luna 1, 50018 Zaragoza (Spain).
E-mail: {dortin,jneira,josemari}@unizar.es

*Abstract*— **We present a method for solving the first location problem using 2D laser and vision. Our observation is a two-dimensional laser scan combined with its corresponding image. The observation is segmented in textured vertical planes; each vertical plane contains geometrical information about its location given by the laser scan, plus the gray level image obtained by the camera. The rich plane texture allows a safe plane recognition and the computer vision geometry allows to compute the relative camera motion once two planes are recognized as correspondent.**

**The proposed algorithm outperforms both laser-only or vision-only algorithms. This is shown in the experimental results where a map composed of 8 observations of a $20 \times 3$ meters corridor allows to successfully locate the robot(without any other prior) in 163 out of 192 initial robot locations.**

## I. Introduction.

The localization of a vehicle in a map is a problem of particular relevance for the mobile robotics and SLAM (simultaneous localization and mapping) communities. The first location problem consists on placing a vehicle in a map with no prior information about its current situation. This is a difficult problem, as it often involves determining whether the vehicle is actually in the map or not. However, it happens to be of practical interest, as its resolution has direct applications in typical SLAM bottlenecks, such as loop closing or the lost robot problem.

The first location problem is twofold: you must determine whether an observation corresponds to some part of a map, and if so, compute the sensor location. Broadly, the first location problem can be regarded as a membership classification in a map, followed by the recognition of corresponding features that can be used to compute the geometric location of the vehicle.

Several authors have reported solutions to the first location using geometrical information gathered by a 2D laser scanner [1],[2],[3]. Our contribution is to combine this laser geometrical information with vision because cameras gather rich photometric information that can solve geometrical ambiguities. In particular, our basic observation consist of one two-dimensional laser scan combined with its corresponding image. The observation is segmented in textured vertical planes; each vertical plane contains geometrical information about its location given by the laser scanner, plus the gray level texture obtained by the camera (see fig 1).

Continuous SLAM incrementally compute both the path followed by a vehicle and the elements of a map, typically modelled as a correlated stochastic vector [4],[5],[6]. Vision provides both geometrical and photometrical information that
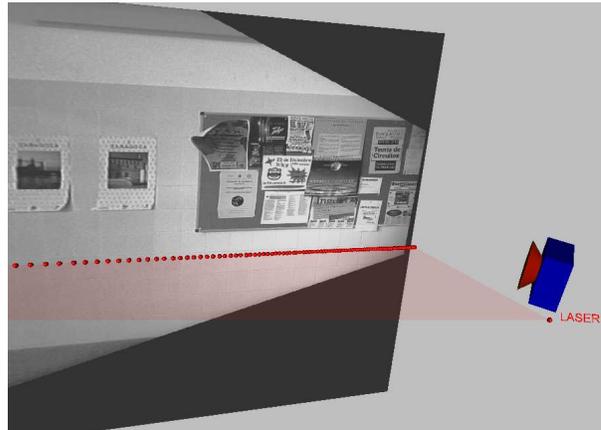


Fig. 1. The combination of complementary sensors allows the use of highly descriptive observations. The subsequent identification of corresponding features between the measurement and the map is therefore greatly simplified.

can be exploited in SLAM, either using a trinocular [7], stereo [8] or a single camera [9].

Viewpoint invariant features used for image matching show a promising venue to make the most of the photometric information contained in the images. Some impressive implementations can be found in [10],[11],[12] or [13].

An alternative solution for mapping an area is to build a topological description of the environment rather than a geometrical one [14],[15],[16]. In this case, the geometrical information becomes less significant for relocating the vehicle, and the relocation relies either in the information provided by the nodes of the map or in the existing relations among them.

In this work a combined use of laser and vision is used to generate multisensorial observations. Those observations are used to label uniquely the nodes of a topological map. Relocation is thus reduced to the determination of correspondences between the current observation and the nodes of the map.

Section II shows how the relative location between two observations can be derived. The use of this method to relocate the vehicle in a map is illustrated in section III. The reliability of this approach is experimentally validated in section IV.

## II. Relative Localization of Two Observations.

The problem that we want to solve is the following one: given two different observations taken at unknown locations, determine if they correspond to the same area or if they are unrelated. In the first case, compute their relative position.
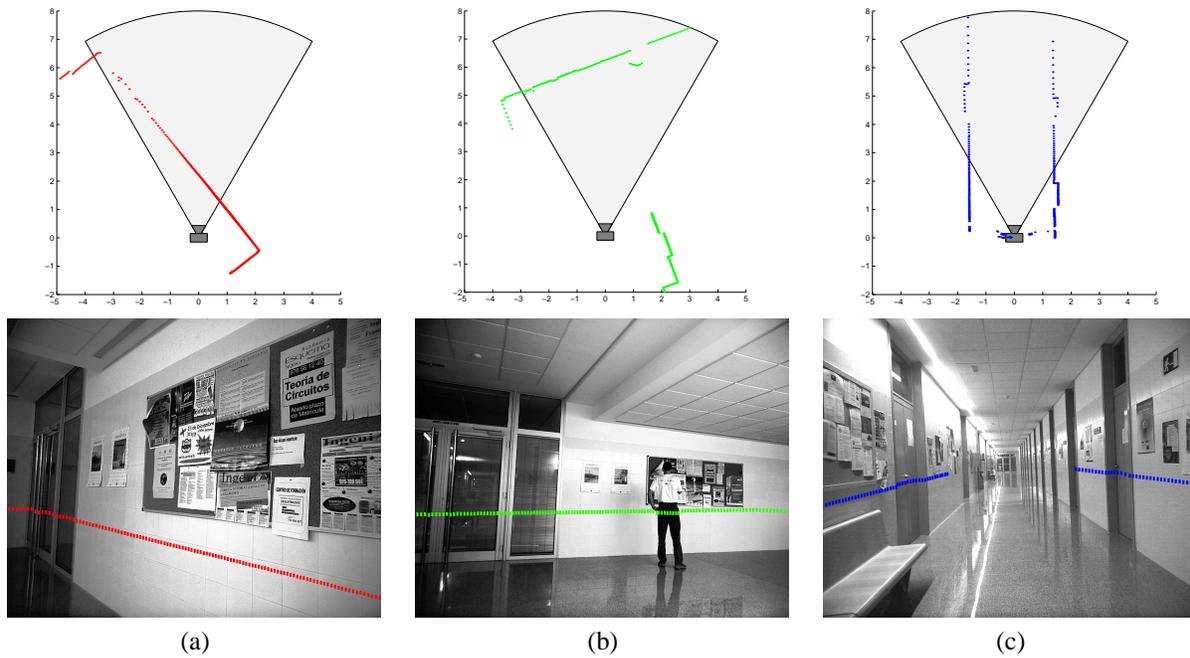
Fig. 2. Three different observations, each composed of a laser scan and its corresponding image. The laser measurements have been projected in the images just for representation purposes. Notice how the poor descriptive capability of the laser scans and the different viewpoints from where the pictures were taken complicate their individual matching. Our goal is to identify that (a) corresponds with (b), but not with (c). The relative location between (a) and (b) is also to be computed. Distances are in meters

The solution to this problem has direct applications in SLAM, in loop closing or in the lost robot problem, and topological mapping.

Consider the observations depicted in fig 2. We would like to conclude that (a) and (b) correspond to the same area seen from different viewpoints, whereas (c) corresponds to a different one. (The observant reader might notice that the glass door at the end of the corridor in (c) is in fact that depicted in (a) and (b); however, areas beyond the range of the laser scanner are not taken into consideration.) We would also like to compute the relative localization between the viewpoints in (a) and (b).

It is important to be able to relocalize the vehicle when the viewpoints are very different, as this provides the capability of mapping a large area with a reduced number of measurements. However, the perspective deformation in the images complicates the identification of corresponding elements, as illustrated in fig 2 (a) and (b). Standard correlation techniques cannot reliably identify corresponding points between these images.

In the following we propose an effective method for identifying corresponding planes and reliably computing the localization between the two viewpoints. The steps of the algorithm are sketched in table I.

### A. Correction of texture.

The image formation process consist in the projection of a three-dimensional world in one picture. The appearance of the imaged features is very dependent on the viewpoint from where the pictures were taken. This complicates the automatic

TABLE I

OUTLINE OF THE TWO OBSERVATIONS ALGORITHM.

| | |
|---|---|
| (1) | Segment the laser in straight segments. |
| (2) | Create one orthophotograph for every segment found. |
| (3) | Detect interest points. |
| (4) | Find corresponding points using normalized cross-correlation. |
| (5) | Fit the relative translation using robust statistics. |
| (6) | Reject the hypotheses with little support. |
| (7) | Otherwise, compute the relative localization of the cameras. |
| (8) | Fit a general homography using guided matching (optional). |
| (9) | Refine the motion using a bundle adjustment (optional). |

identification of point matches.

However, if we had some prior information about the three-dimensional scene we could solve the correspondence problem in the original domain of the data (where no perspective deformation occurs) rather than in its projections.

In this work, we make use of the two-dimensional laser scanner to detect possible vertical planes in the scene using a RANSAC based segmentation (step 1). Afterwards, the texture is mapped from the images to the planes by means of a homography, creating non-scaled orthophotographs of the planes. Fig 3 illustrates the corrected views for the wall detected in fig 2 (a) and (b) (step 2).

Without loss of generality, let us chose a reference system such that the selected plane is defined by $\{(X, Y, Z) \in \Re^3 | Z = 0\}$. In this case the coordinates in the image $(x, y)$ are related with those of the plane (up to a scale factor $\lambda$) by:

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \qquad (1)$$

<center>(a)                           (b)</center>
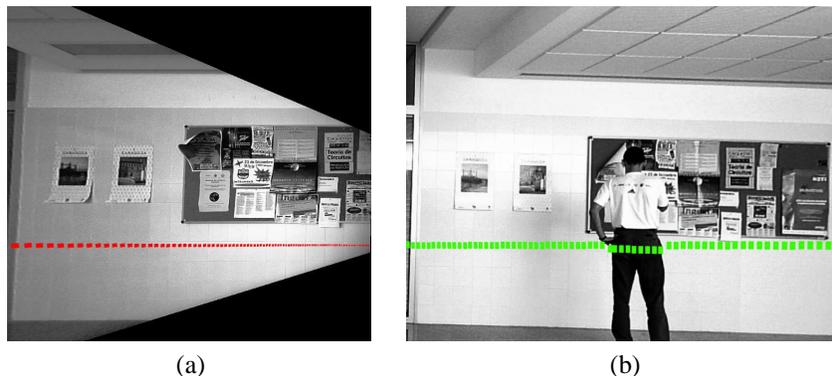
Fig. 3. Pictures (a) and (b) depict two orthophotos created from the sample images of figure 2 after the perspective correction of texture. Standard cross-correlation can be easily applied in these images. Both views can be roughly aligned by means of a two-parameter translation. However, if higher accuracy is required, a general eight-degree of freedom homography should be used to align the observations.

being $\mathbf{K}$ the inner calibration of the camera, $\mathbf{t}$ its location and $[\mathbf{r}_1 \ \mathbf{r}_2]$ the two first columns of its rotation matrix.

Corresponding orthophotographs look similar. This allows to easily identify homologous points detected as (in [17]) by means of normalized cross-correlation (steps 3 and 4). Larger (and therefore more discriminative) correlation windows can be used in the orthophotographs. Also, as the motion of the laser scanner is assumed to be roughly horizontal (although different heights are allowed) only a two-dimensional translation remains unknown between two corresponding planes.

### B. Alignment of two views.

Normalized cross-correlation will provide matches, a fraction of them spurious (outliers), even if we are working with orthophotographs. Given the high sensitivity of least-squares techniques to outliers, we make use of robust statistics to deal with the spurious matches.

We align two orthophotographs with a translation using a RANSAC-based algorithm [18][19]. In this probabilistic algorithm, given an acceptable probability of success $p$, $N$ attempts are carried out to determine the translation by randomly selecting the number of matches required to calculate a solution ($s$), and then determining the support of the solution among the rest of the matches. Given an outlier rate $\epsilon$, the number of attempts will be:

according to:

$$N = \frac{\log{(1-p)}}{\log{(1-(1-\epsilon)^s)}}. \tag{2}$$

For values of $p = 0.9$ and $\epsilon = 0.5$, only four attempts are required. Notice that the number of attemps does not depend on the number of matches. It can be further reduced when it is adaptively computed, as in [20].

A hypothesis is accepted when there is strong support (the number of matches is over a threshold) by the data. In this way we can determine whether two different observations correspond to the same area or not (step 6). If they do, the translation between both orthophotographs can be combined with the camera to plane transformation, and the non-scaled relative location of both cameras can be computed (step 7).

Last but not least, deviations on the assumptions of horizontal motion and vertical planar walls mean that a two-parameter translation might not be able to describe an accurate alignment for the planes. Therefore, after a rough bidimensional correction, a general homography should be found in the original images using guided matching (step 8), and the location of the cameras should be computed making use of standard photogrammetric techniques [21] (step 9). This avoids the artifacts introduced by the resampling of the texture and the non-gaussian noise present in the corrected orthophotos.

Fig 4 illustrates the aligned laser scans and pictures of the observations (a) and (b) depicted in fig 2.

### III. LOCALIZATION IN A MAP OF MULTIPLE OBSERVATIONS.

To use the relative localization process described in the previous section for vehicle relocation in an environment, assume that we have at our disposal a topological map composed of some nodes or *key observations*, whose relative locations are already known. Techniques to map these key observations could be conventional SLAM, odometry or a batch processing of our method, among others.

Fig 6 illustrates the topological map used in our experiments: a set of observations related by their corresponding locations. Notice that the individual observations are not combined into a single model. Only the location of the camera is referred to the global coordinate system, whereas the key observations remain independent of each other. Nevertheless, a 3D planar model of the environment can be generated by fusing the observations in the nodes of the map, whenever required.

The vehicle can be relocated by aligning the measured observation with the each of the key observations of the topological map, as explained in section II. Any prior information about the location of the vehicle could be directly used to reject some pairs before the testing is done.

During normal operation, multiple correspondences between the new observation and the nodes of the map are are detected, and hence, multiple hypotheses (a fraction of them spurious) for the location of the vehicle. This redundancy can be
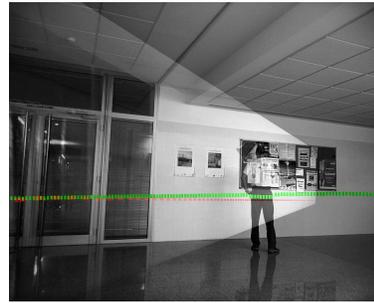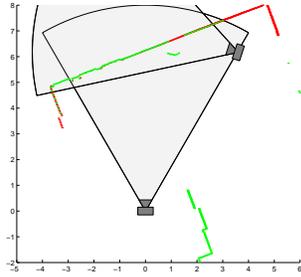
Fig. 4. Alignment of the observations (a) and (b) of fig 2. Strong evidence of correspondence (measured as the support of image point matches) is used to accept or reject a hypothesis. After the alignment, the three-dimensional location between both observations is readily available. Distances are in meters.

TABLE II
RELOCALIZATION RESULTS.

| Ref view | True positives | True negatives | False positives | False negatives |
|---|---|---|---|---|
| (a) | 31 | 138 | 0 | 23 |
| (b) | 24 | 128 | 0 | 40 |
| (c) | 32 | 128 | 0 | 32 |
| (d) | 19 | 138 | 2 | 33 |
| (e) | 47 | 132 | 0 | 13 |
| (f) | 30 | 128 | 0 | 34 |
| (g) | 54 | 108 | 0 | 30 |
| (h) | 35 | 141 | 0 | 16 |

exploited to increase the robustness and the accuracy: a) to detect spurious correspondences; b) to increase the precision by fusion of the inlier estimates; c) to detect (and possibly correct) any errors in the original map.

## IV. EXPERIMENTAL RESULTS.

In order to analyze our capability to correctly relocating the vehicle, 192 unordered horizontal observations were taken at different heights within a $20 \times 3$ meters corridor. A three dimensional motion in a corridor was chosen because it produces ambiguous configurations for the matching of the laser scans. Under these circumstances, processing the visual information becomes critical to robustly and accurately localize the vehicle.

Eight representative key observations were selected as the nodes of a topological map, sketched in fig 6. Each of the 192 observations were tested with every node of the map (a total of 1536 combinations), without making use of any prior information (no odometry is available).

The 1536 locations were computed as proposed in section II. The results were manually checked and are classified in table II.

The pairs were correctly located in 272 cases (*true positives*) and were correctly classified as non-corresponding in 1041 cases (*true negatives*).

There were 221 undetected corresponding pairs (*false negatives*) in which the vehicle was not relocated even when common areas were observed. The main reasons for failure were: a) little overlap between the views; b) poorly textured areas; c) occlusions the common area by non-planar elements; d) failed segmentation of the wall; d) specular reflections and severe illumination changes;

Only in two cases a pair of non-corresponding observations were erroneously reported as matches (*false positives*). In both of them the errors were due to the presence of strong symmetries in the corridor (see one of them in fig. 5). However, in both false positive cases the right location was also detected. When two inconsistent alternative locations for the vehicle are identified, heuristic criteria can be used to reject one or both of them.

Among the 192 test locations, in 2 the vehicle cannot be located because none of the nodes is visible. Both cases were successfully classified as true negatives. In 27 different cases the vehicle was not detected to be in the map, even when some of the nodes were covisible with its actual location.

Of the remaining 163 cases the vehicle was correctly located in 161: at least a true positive was detected without any false positive. In the 2 remaining two different locations, one false positive and one true positive were detected, although false positives were reported, their inconsistency can be detected because of the redundancy of the measurements.

Results for three typical observations (I), (II) and (III) are in fig 6, where examples of true positives, true negatives, false positives and false negatives are illustrated for particular cases. Please, see the caption of the figure for a more detailed explanation. Notice in these examples the important role played by the vision information to disambiguate matches and to reject hypotheses compatible if only laser information were considered.

## V. CONCLUSIONS.

This work addresses the fist location problem: identifying if a vehicle is revisiting an already known area and determining its location within a map without any prior estimation.

By exploiting multisensorial fusion, the individual capabilities of complementarity sensors (one 2D laser and one camera) are extended. This provides highly descriptive observations that, together with a topological representation of the map, allow a effective, robust and accurate localization of the vehicle.

The conducted experiments show the potential of the method to deal with real data. The experimental rate of successful relocations is above the $83\%$, and the false positive rate is about the $1\%$. Besides multiple hypotheses are produced for some of the locations, this redundancy can be exploited

(a)



(b)



(c)
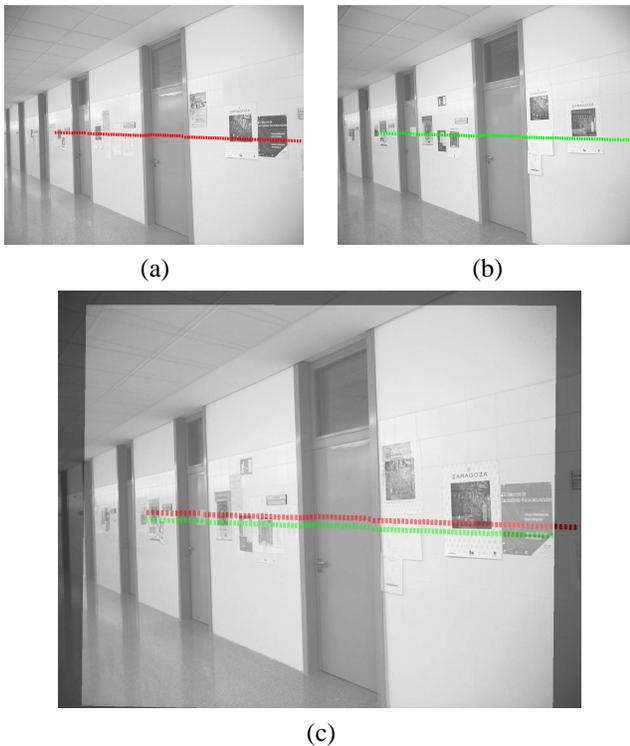
Fig. 5. One of the two false positive relocation errors. The key observation (a) and the test one (b) were mismatched due to symmetries in the scene. The good alignment achieved in (c) makes the posters be regarded as outliers, in the same way that the person appearing in fig 4.

to improve robustness and accuracy. Sample observations illustrate that the relocation of the vehicle was by no means trivial.

Whereas the method is currently constrained to indoor environments, it is our belief that the underlying ideas can be extended to work with more unrestricted ones. Future work is oriented in this direction.

## VI. DISCUSSION.

Despite many improvements in the last years, robust data association remains a critical issue for the robotics community. Regarded as a classification problem, it can be simplified if representative descriptors are selected to characterize the data.

The proposed techniques are well suited for topological map building because of their capability to locate with respect to the map nodes. This is a promising venue for our future work.

Determining correspondences is often based on statistical similarity criteria. This implies that unlimited data association is prone to fail, whatever the descriptors used. An approach that systematically integrates every new observation compromises the robustness of the relocation and even the reliability of the whole map in the long term.

## REFERENCES

[1] J. Neira, J. Tardós, and J. Castellanos, "Linear-time vehicle relocation in SLAM," in *IEEE Int. Conf. Robotics and Automation*, 2003, p. To appear.

[2] T. Bailey, E. M. Nebot, J. K. Rosenblatt, and H. F. Durrant-Whyte, "Data association for mobile robot navigation: A graph theoretic approach," in *Proc. IEEE Int. Conf. Robotics and Automation*, San Francisco, California, 2000, pp. 2512–2517.

[3] D. Fox, W. Burgard, F. Dellaert, and S. Thrun, "Monte carlo localization: Efficient position estimation for mobile robots," in *AAAI*, 1999, pp. 343–349.

[4] R. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty," *Intl. Journal of Robotics Research*, vol. 5, no. 4, pp. 56–68, 1986.

[5] J. Castellanos and J. Tardós, *Mobile Robot Localization and Map Building: A Multisensor Fusion Approach*. Boston. USA: Kluwer Academic Publishers, 1999.

[6] J. Leonard and H. Durrant-Whyte, "Simultaneous map building and localization for an autonomous mobile robot," in *Workshop on Intelligent Robots and Systems. IROS.*, Feb 2001, pp. 1442–1447.

[7] S. Se, D. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *International Journal of Robotics Research*, vol. 21, no. 8, pp. 735–758, 2002.

[8] A. Davison and D. Murray, "Simultaneous localization and map-building using active vision," *IEEE Trans. on PAMI*, vol. 24, no. 7, pp. 865 – 880, July 2002.

[9] A. Davison, "Real-time simultaneous localization and mapping with a single camera," in *Proc. International Conference on Computer Vision*, 2003.

[10] P. Pritchett and A. Zisserman, "Wide baseline stereo matching," in *6th ICCV, Bombay*, Jan 1998, pp. 754–760.

[11] T. Tuytellars and L. Van Gool, "Wide baseline stereo matching based on local, affinely invariant regions," in *Proceedings of the British Machine Vision Conference*, 2000.

[12] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets , or "How do I organize my holiday snaps?"," in *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, vol. 1. Springer-Verlag, 2002, pp. 414–431.

[13] P. H. S. Torr and C. Davidson, "IMPSAC: Synthesis of importance sampling and random sample consensus," *IEEE Trans. on PAMI*, vol. 25, no. 3, pp. 354 – 364, March 2003.

[14] B. Kuipers and Y. Byun, "A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations," *IEEE Journal of Robotics and Automatonomous System*, vol. 8, pp. 47–63, Apr-Jun 1991.

[15] J. Gaspar, N. Winters, and J. Santos-Victor, "Vision-based navigation and enviromental representations with an omnidirectional camera," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 6, pp. 890–898, december 2000.

[16] H. Choset and K. K. Nagatani, "Topological simultaneous localization and mapping (SLAM): Toward exact localization without explicit localization," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 2, pp. 125–137, Apr 2001.

[17] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, 1988, pp. 147–151.

[18] M. A. Fischler and R. C. Bolles, "Random sample consensus, a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381 – 395, 1981.

[19] P. H. S. Torr and A. Zisserman, "Robust computation and parametrization of multiple view relations," in *ICCV*, 1998, pp. 727–732.

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

[21] K. Kraus, P. Waldhausl, J. Jansa, and H. Kager, *Photogrammetry. Vol1: Fundaments and Standar Processes. Vol2: Advanced Methods and Applications*. Dümmler Bonn, 1997.
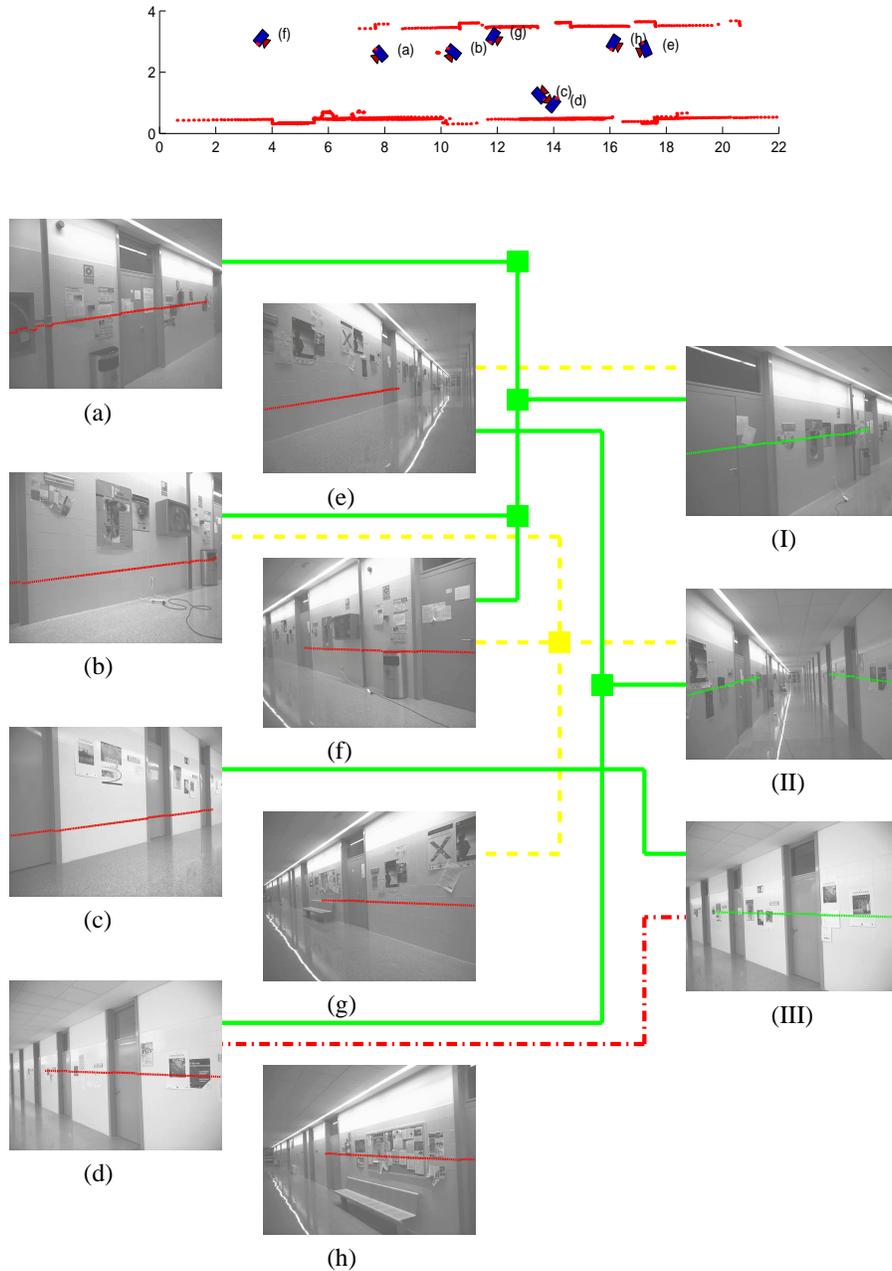
Fig. 6. Map used for the relocation. Eight unordered representative key observations, (a) to (h), are the nodes of our topological map, covering a length of about twenty meters in a corridor. Top view sketches the location of the cameras as well as their corresponding laser scans in the global reference.

Three different observations (I), (II) and (III) are evaluated with the nodes (a) to (h). Every detected pair provides an estimation for the three-dimensional location of the vehicle in the map. Typically, multiple estimations are found for each observation.

**True positive** correspondences (that actually match and are properly detected) are shown in continuous green. They correspond to the pairs (I)-(a), (I)-(b), (I)-(f), (II)-(d), (II)-(e) and (III)-(c). Notice that the correct location is found at least once for all the observations (I), (II) and (III), even when they were acquired at different heights and with very different viewpoints than those of the nodes of the map. In particular, notice that observation (II) is properly detected with respect to both sides of the corridor.

**False negative** correspondences (that match but are not detected) are depicted in dashed yellow (the pairs (I)-(e), (II)-(b), (II)-(f) and (II)-(g)). In can be observed that in all the cases they correspond to situations where very little overlap occurs between the nodes and observation (II). Identifying some of these situations was hard even for some of the authors.

The **false positive** correspondence (that does not match but are erroneously detected), only one (III)-(d), is printed in dash-dotted red. It is due to the symmetries in the corridor. However, the also available correct location of observation (III) (the pair (III)-(c)) is inconsistent with the previous one. This inconsistency could be automatically detected and exploited to resolve this relocation error.

**True negative** correspondences, those that do not match and are not detected, correspond to all the remaining non-linked possible pairs. For example, in observation (I), these are the pairs (I)-(c), (I)-(d), (I)-(g) and (I)-(h). This corresponds to the (non-trivial) capability of determining that the vehicle is *not* present in one part of the map.