Appearance-Based SLAM Relying on a Hybrid Laser/Omnidirectional Sensor

Gabriela Gallegos[1], Maxime Meilland[1], Patrick Rives[1], Andrew I. Comport[2]

[1] INRIA Sophia Antipolis Méditerranée
2004 Route des Lucioles BP 93
Sophia Antipolis, France
{name.surname}@sophia.inria.fr

[2] CNRS, I3S Laboratory
2000 Route des Lucioles BP 121
Sophia Antipolis, France
comport@i3s.unice.fr

***Abstract*— This paper describes an efficient hybrid laser/vision appearance-based approach to provide a mobile robot with rich 3D information about its environment. By combining the information from an omnidirectional camera and a laser range finder, reliable 3D positioning and an accurate 3D representation of the environment is obtained subject to illumination changes even in the presence of occluding and moving objects. A scan matching technique is used to initialize the tracking algorithm in order to ensure rapid convergence and reduce computational cost. The proposed method is validated in an indoor environment using data taken from a mobile robot equipped with a 2D laser range finder and an omnidirectional camera.**

I. INTRODUCTION

Autonomous navigation in unknown environments has been the focus of attention in the mobile robotics community for the last three decades. When neither the location of the robot nor a map of the region are known, localization and mapping are two tasks that are highly inter-dependent and must be performed concurrently. This problem, known as "Simultaneous Localization and Mapping" (SLAM), was originally introduced by Smith and Cheeseman [22], [20].

Localization methods based only on proprioceptive sensors give bad results due to modeling approximations (rolling without slippage,...) which are not satisfied and dead reckoning drift. Mobile robots have to be equipped with a perception system that enables them to perform accurate localization and reconstruct a reliable and consistent representation of the environment.

Various techniques to solve the SLAM problem using laser range finders have been extensively studied. The information provided by laser range finders can be used not only to obtain a more accurate position estimate, but also to measure the distance to nearby objects. Localization schemes based on laser scan matching involve computing the most likely alignment between two sets of slightly displaced laser scans. Unfortunately, the most extensively used scan matching algorithm, "Iterative Closest Point" (ICP) [2], suffers from a high computational cost when dealing with large-scale environments due to its expensive point-to-point association rule. In order to reduce the computational resources spent on searching for these point associations and scale up to larger environments, Diosi and Kleeman proposed the Polar Scan Matching method [10], which simply matches points with the same bearing.

While laser-based schemes perform reasonably well in practice, the use of 2D laser alone limits SLAM to planar motion estimation and does not provide sufficiently rich information to reliably identify previously explored regions. Vision sensors are a natural alternative to laser range finders because they provide richer perceptual information and enable 6 degrees of freedom motion estimation. One of the first attempts to solve the SLAM problem using monocular vision was the work by Broida et al. [4]. Since then faster computers and ways of selecting sparse but distinct features have allowed new approaches to emerge.

Davison [8] proposes a real-time approach that attempts to minimise drift by detecting and mapping only long-term landmarks during the SLAM process. This approach, however, is not appropriate for long displacements because of algorithmic complexity and growing uncertainty. Another interesting approach better suited for outdoor environments and large displacements was proposed in [18].

Standard cameras only have a small field of view (typically between 40° and 50°) and are easily affected by occlusion. It was shown in [9] that a larger field of view (e.g. using fish-eye lens) makes it easier to find and follow salient landmarks. In contrast, omnidirectional cameras provide full 360° field of view, which makes easier to recognize previously observed places whatever the orientation of the robot is. Furthermore, in order to avoid the limitations due to planar projections, images captured by these cameras can be uniquely mapped to spherical images [13].

Many authors avoid the problems of monocular algorithms (i.e. scale factor, initialization, observability) by using multi-view constraints (see [14] for a survey). Others try to complement, merge or fuse the information of different sensors. In [23] a combination of ultrasonic and vision sensors is used to produce an occupancy grid representation of the environment. Fu et al. [11] extract environmental features from monocular vision and laser range finders to build metric maps and then fuse them using Kalman Filter to build a grid map simultaneously. Complementary information from cameras (panoramic or omnidirectional) with the depth information acquired by a laser range finder is not a new idea; see e.g. [6],

Gabriela Gallegos gratefully acnowledges her funding by a CONACYT (México) grant.

[5], [3].

In this paper, an appearance-based approach for indoor localization and mapping is proposed that merges data from an omnidirectional camera and a laser range finder to tackle the drawbacks of both sensors. The proposed framework adapts and extends the ideas presented in [7], [21], [17]. An initial guess of the position provided by scan matching is used to initialize the tracking algorithm to ensure rapid convergence and reduce computational cost.

II. Vertical line extraction and floor detection from omnidirectional images and laser scans

It is a well known fact that geometrical structures such as lines or planes characterize well a human-made environment. When using an omnidirectional camera, vertical lines in the scene (e.g. walls, facades, doors, windows) project as quasi-radial lines onto the image. A Hough transform was used to detect prominent lines from a binary edge image. Since the camera is calibrated, the image center (i.e, where all radial lines intersect) is available. It can also be approximated by the center of the mirror border (assumed to be a circle) by using a circle detector to improve accuracy. Finally, the lines detected by the Hough transform which intersect the image center are marked as radial lines. Thanks to the calibration between the laser and the camera, it is possible to project the laser trace onto the omnidirectional image. Subsequently, at each intersection point between the laser trace and a radial line, a depth measurement can be determined which then fully characterizes the vertical lines in the 3D scene [12].

Firstly, it is assumed that the distance between the laser frame and the floor is approximately known (which requires the plane to be horizontal). It is also assumed that the pose between the camera and laser frames is correctly estimated. Under these hypotheses, the laser scan can be shifted along the vertical lines and used to predict where a virtual laser trace, corresponding to the floor, should project onto the omnidirectional image. Due to calibration errors, the predicted trace does not exactly match the real boundary of the floor observed in the image. In practice, the neighborhood of the predicted trace is searched for the closest element of contour detected in the image. This match is finally taken as the intersection between the floor plane and the walls and is integrated into a partial 3D model. Figure 1(a) shows the predicted laser measurements projected onto the floor in blue. From the zoomed image in Fig. 1(c) it can be seen that in certain cases the prediction may be wrong as in the case of the fire-extinguisher which breaks the planarity hypothesis of the wall. Figure 1(b) shows in white the result of applying the Canny edge detector to the entire image and in yellow the detected edge points found by searching the neighborhood of the laser floor plan, therefore giving a more accurate floor estimation. A better view of this correction can be seen in Fig. 1(d) where overlapping traces are shown.

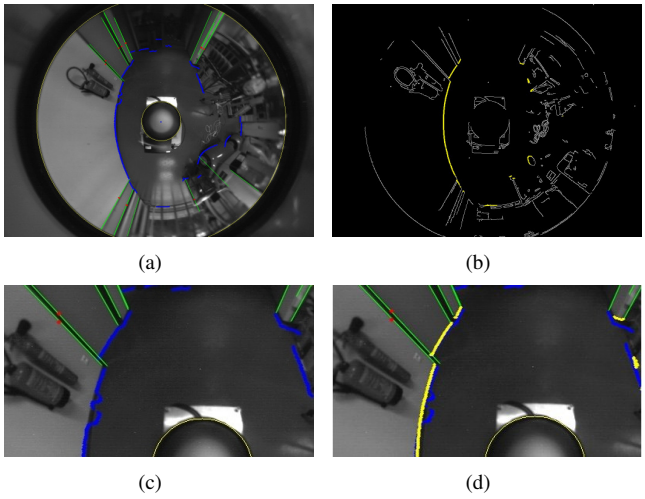


Fig. 1. Floor Detection 1(a) Line extraction and laser re-projection shifted at floor level. 1(b) Correct floor detection. 1(c) and 1(d) Zoom

III. Augmented Spherical View

In this section, a novel robot-centered representation is described that is well adapted to the appearance-based SLAM method. Central omnidirectional cameras can be modeled using two consecutive projections [13]: a spherical projection followed by a perspective one. An omnidirectional image can thus be mapped onto a sphere by means of an inverse projection. A point $\boldsymbol{P} \in \mathbb{R}^3$ is projected as a point $\mathbf{q}$ on the unit sphere S^2 and the projection is given by $\mathbf{q} = \frac{\boldsymbol{P}}{\|\boldsymbol{P}\|}$.

The coordinates of $\mathbf{q}$ can be expressed using standard spherical coordinates. The aim of this section is to show how this spherical representation can be constructed. Since the omnidirectional camera has been calibrated, the intrinsic parameters are known and the omnidirectional image plane $\boldsymbol{\mathcal{I}}_p(u, v)$ can be mapped onto the unit sphere image $\boldsymbol{\mathcal{I}}(\phi, \theta)$. This mapping is performed in four steps:

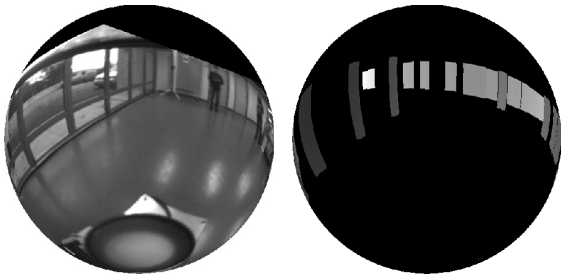
a) Sampling: First, the unit sphere is sampled at a constant angle in a spherical grid defined with the maximal radius of the omnidirectional image in ϕ and the laser points in θ, to respect both laser sampling and omnidirectional sampling.

b) To image plane: The spherical points $\mathbf{q}$ corresponding to couples (ϕ, θ) are mapped onto the image plane using the "Unified Projection Model" defined in [16], which is an extension of Geyer's [13] and Barreto's [19] models'.

c) Interpolation: The spherical image is obtained by interpolating the omnidirectional intensity image around the projected points as shown in Fig. 2(a).

d) Hybrid laser/image spherical view: Finally, the augmented spherical view is constructed using the depth information from the laser range finder and the floor plane, together with lines extracted from the omnidirectional image. In order to have a more dense estimation in the sphere, the laser trace is propagated down to the floor and upwards. The resulting spherical image is shown in Fig. 2(b).

In summary, the *current* augmented spherical view is denoted by $\boldsymbol{\mathcal{S}} = \{\boldsymbol{\mathcal{I}}, \boldsymbol{P}\}$ and the *reference* spherical view by $\boldsymbol{\mathcal{S}}^* = \{\boldsymbol{\mathcal{I}}^*, \boldsymbol{P}^*\}$. A superscript $*$ will be used throughout the paper to designate the reference view variables. $\boldsymbol{P}$ is initialized from the laser scan and the vertical 3D lines as



(a) (b)

Fig. 2. Augmented spherical view. 2(a) Grey levels on the unit sphere. 2(b) Depth spherical image.

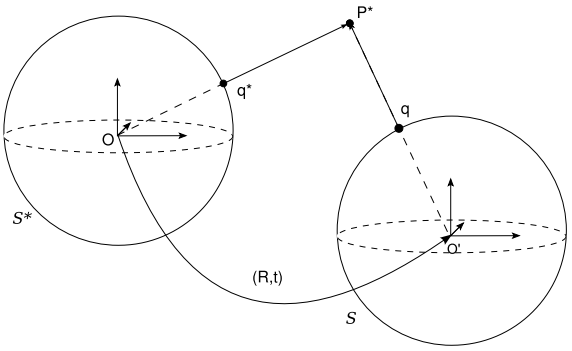


Fig. 3. Augmented spherical views: at each pixel on the unit sphere is associated with a grey level intensity and the corresponding depth of the 3D point.

shown in Fig. 3.

IV. Efficient hybrid laser/vision appearance-based localization

The main challenge of localization in the context of indoor environments is to obtain reliable odometry subject to illumination changes in the presence of occluding and moving objects. As expected for an odometry-based approach, the objective is to compute the trajectory of the robot (i.e. the laser/vision sensor) along a sequence by integrating elementary displacements estimated from the successive spherical views registered during motion.

An appearance-based localization method is proposed which minimizes a non-linear cost function directly built from the augmented spherical view defined above. As mentioned in the previous section, each pixel $\mathbf{q}$ of the spherical view $\boldsymbol{\mathcal{S}}$ is associated with a brightness function $\boldsymbol{\mathcal{I}}(\boldsymbol{\mathcal{P}})$ and is augmented with the depth of the associated 3D point (when data is available). In the following, the *reference template* is denoted $\boldsymbol{\mathcal{R}}^* = \{\mathbf{q}_1^*, \ldots, \mathbf{q}_n^*\}$, which defines the subset of the reference spherical view $\boldsymbol{\mathcal{S}}^*$ where both the grey level and the depth values are available, where n is the number of pixels .

A. Sphere-to-sphere mapping

Consider a 3D point $\boldsymbol{P}^* \in \mathbb{R}^3$ and its projection $\mathbf{q}^*$ onto the unit sphere. Using homogeneous coordinates, the spherical parameterization (θ, ϕ, ρ), gives:

$$\boldsymbol{P}^* = \begin{bmatrix} \rho\cos(\theta)\sin(\phi) \\ \rho\sin(\theta)\sin(\phi) \\ \rho\cos(\phi) \\ 1 \end{bmatrix} \tag{1}$$

and

$$\mathbf{q}^* = \frac{\boldsymbol{P}^*}{\|\boldsymbol{P}^*\|} = \begin{bmatrix} \cos(\theta)\sin(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\phi) \\ 1 \end{bmatrix}. \tag{2}$$

The motion of the sensor or objects within the scene induces a deformation of the reference template. Denote $\overline{\mathbf{T}} = (\overline{\mathbf{R}}, \overline{\mathbf{t}}) \in \mathbb{SE}(3)$ the true current sensor pose relative to the reference sensor pose (homogeneous transformation matrix). The function $w(\boldsymbol{P}^*, \overline{\mathbf{T}})$ which warps the current sphere onto the reference one is defined as

$$\boldsymbol{\mathcal{I}}^*(\boldsymbol{\mathcal{P}}^*) = \boldsymbol{\mathcal{I}}\left(w(\boldsymbol{P}^*, \overline{\mathbf{T}})\right), \quad \forall \boldsymbol{P}^* \in \boldsymbol{\mathcal{R}}^*. \tag{3}$$

In order to reduce computational time, the warping function is applied to the reference template $\boldsymbol{\mathcal{R}}^*$ only.

The warping function $w(\boldsymbol{P}^*, \overline{\mathbf{T}})$ defines a one-to-one mapping $\mathbf{q}^* \leftarrow \mathbf{q}^{cur}$ from the current sphere to the reference sphere such that

$$\mathbf{q}^{cur} = \frac{\boldsymbol{P}^{cur}}{\|\boldsymbol{P}^{cur}\|} = \frac{\overline{\mathbf{T}}\boldsymbol{P}^*}{\|\overline{\mathbf{T}}\boldsymbol{P}^*\|}. \tag{4}$$

The current image $\boldsymbol{\mathcal{I}}$ is then interpolated at points $\mathbf{q}^{cur}$ to obtain the corresponding intensities in spherical coordinates.

Considering that an initial estimation $\widehat{\mathbf{T}}$ of current image pose fully represents the pose of the current camera with respect to a reference sphere, the tracking problem is reduced to estimating the incremental pose $\mathbf{T}(\mathbf{x})$ assuming $\exists \tilde{\mathbf{x}}$: $\mathbf{T}(\tilde{\mathbf{x}})\widehat{\mathbf{T}} = \overline{\mathbf{T}}$. This estimate is updated by a homogeneous transformation $\widehat{\mathbf{T}} \leftarrow \mathbf{T}(\mathbf{x})\widehat{\mathbf{T}}$. The unknown parameters $\mathbf{x} \in \mathbb{R}^6$ are determined by the integral of a constant velocity twist that produces the pose $\mathbf{T}$ in 6 degrees of freedom:

$$\mathbf{x} = \int_0^1 (\boldsymbol{\omega}, \boldsymbol{v}) dt \in \mathfrak{se}(3). \tag{5}$$

The pose and the twist are related via the exponential map by $\mathbf{T} = e^{[\mathbf{x}]_\wedge}$, where the operator $[.]_\wedge$ is defined as

$$[x]_\wedge = \begin{bmatrix} [\boldsymbol{\omega}]_\times & \boldsymbol{v} \\ 0 & 0 \end{bmatrix} \tag{6}$$

and where $[.]_\times$ represents the skew symmetric matrix operator. Hence, the current camera pose can be estimated by minimizing a non-linear least squares cost function:

$$\boldsymbol{\mathcal{C}}(\mathbf{x}) = \sum_{\boldsymbol{P}^* \in \boldsymbol{\mathcal{R}}^*} \left(\boldsymbol{\mathcal{I}}\left(w\left(\boldsymbol{P}^*, \mathbf{T}(\mathbf{x})\widehat{\mathbf{T}}\right)\right) - \boldsymbol{\mathcal{I}}^*\left(\boldsymbol{P}^*\right)\right)^2. \tag{7}$$

B. Minimization of the cost function

The aim now is to minimize the difference in image intensity from the cost function (7) in an accurate and robust manner. Since this is a non-linear function of unknown parameters, an iterative minimization procedure is used. The minimization technique is quite similar to that used in [7], so only the broad lines of the method are outlined here. Rather than using a standard sum-of-squared differences (SSD) technique based on an L_2 norm, a robust M-estimator [15] is used in order to reject the outliers due to illumination changes, moving objects or occlusions in the scene. The objective function therefore becomes:

$$\boldsymbol{\mathcal{O}}(\mathbf{x}) = \rho\left(\sum_{\boldsymbol{P}^* \in \boldsymbol{\mathcal{R}}^*} \boldsymbol{\mathcal{I}}\left(w\left(\boldsymbol{P}^*, \mathbf{T}(\mathbf{x})\widehat{\mathbf{T}}\right)\right) - \boldsymbol{\mathcal{I}}^*(\boldsymbol{P}^*)\right), \tag{8}$$

where $\rho(u)$ is a robust weighting function (see [7], [15] for more details).

The robust objective function is minimized by $\nabla\boldsymbol{\mathcal{O}}(\mathbf{x})|_{\mathbf{x}=\tilde{\mathbf{x}}} = \mathbf{0}$, where ∇ is the gradient operator with respect to the unknown $\mathbf{x}$ from (5) and it is assumed that there is a global minimum within the convergence domain (which experimentally is quite large). The Jacobian of the objective function (8) can be decomposed in three parts:

$$\mathbf{J}(\mathbf{x})|_{\mathbf{x}=\tilde{\mathbf{x}}} = \mathbf{J}_{\boldsymbol{\mathcal{I}}^*}\mathbf{J}_w\mathbf{J}_{\mathbf{T}}. \tag{9}$$

Here $\mathbf{J}_{\boldsymbol{\mathcal{I}}^*}$ is the image gradient computed on the reference sphere with respect to spherical coordinates (θ, ϕ) of dimension $n \times 2n$, $\mathbf{J}_w$ is the derivative of spherical projection in (2) of dimension $2n \times 3n$, and $\mathbf{J}_{\mathbf{T}}$ depends on parametrization of $\mathbf{x}$ from (5) and has dimension $3n \times 6$. The objective function (8) is iteratively minimized by computing $\widehat{\mathbf{T}} \leftarrow \mathbf{T}(\mathbf{x})\widehat{\mathbf{T}}$ with the vector of unknown parameters $\mathbf{x}$ such that:

$$\mathbf{x} = -\lambda(\mathbf{D}\mathbf{J})^{+}\mathbf{D}(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{I}}^*), \tag{10}$$

where $(\mathbf{D}\mathbf{J})^{+}$ is the pseudo-inverse, $\mathbf{D}$ the diagonal matrix determinated from the robust function $\rho(u)$, and λ is a gain factor that ensures the exponential decay of the error.

C. Initialization step

It is a well known fact that direct iterative methods suffer from convergence problems when initialized far from the solution. This is also true for our method where an initialization sufficiently close to the solution is needed to ensure rapid convergence and reduce computational cost. This initial guess is obtained from the laser data using a 2D scan matching technique that is accurate enough to ensure fast convergence for the appearance-based method. Namely, the "Enhanced Polar Scan Matching" (EPSM) described in [12] was used for this initialization step. As any scan matching technique, it finds the pose of a laser scan with respect to a reference scan by performing a gradient descent search for the transformation that minimizes the square error between corresponding points. In contrast to other methods, it avoids an expensive search for corresponding points by matching points with the same bearing, therefore taking advantage of the natural representation of laser scans in polar coordinates. The method assumes the reference and current scans are given as sequences of range and bearing measurements of the form $\{r_{ri}, \phi_{ri}\}_{i=1}^{n}$ and $\{r_{ci}, \phi_{ci}\}_{i=1}^{n}$, respectively, and requires an initial estimate (x_c, y_c, θ_c) for the pose of the current scan in the reference scan coordinate frame. This estimate is obtained from the robot odometry.

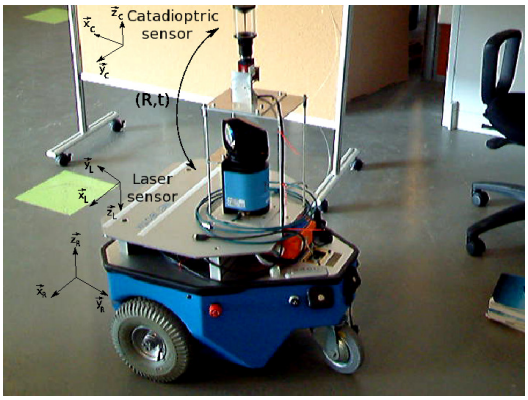


Fig. 4. Hannibal robot experimental testbed.

V. Experimental testbed

Figure 4 shows the robot used in the experiments, Hannibal, from Neobotix mobile platform (MP-S500). Hannibal is equipped with a Sick LD-LRS1000 laser, capable of collecting full 360° data. The laser head can revolve with a variable frequency ranging from 5Hz to 10Hz and the angular resolution can be adjusted up to 1.5° at multiples of 0.125°. The laser has a 30m range. To perform a 360° scan with a resolution of 0.25°, it was necessary to reduce the frequency of the rotor to 5Hz, thus obtaining 1,400 data points per scan. The perspective camera is a progressive-scan CCD camera (Marlin F-131B) equipped with a hyperbolic mirror HM-N15 from Accowle (Seiwapro) with a black needle at the apex of the mirror to avoid internal reflections of the glass cylinder. Careful calibration of the laser and the camera is required for merging image and laser data. The Matlab Omnidirectional Calibration Toolbox developed by Mei [1] was used to estimate the intrinsic parameters of the camera and the parameters of the hyperbolic mirror.

Odometry data arrives at a frequency of 50Hz, omnidirectional images at 15Hz and laser measurements at 5Hz. Since data from the different sensors arrive at different frequencies, a function to synchronize the data as it is acquired by the robot was implemented.

VI. Results

The method is validated using a sequence of 3,262 images and laser scans which were obtained by manually driving the robot in an indoor environment. The exploration trajectory constitutes a closed loop of about 40 meters across the robotic hall.

Figure 5(a) shows the map and the pose estimated by scan matching (in green) and the original odometry given by the

(a)

(b)

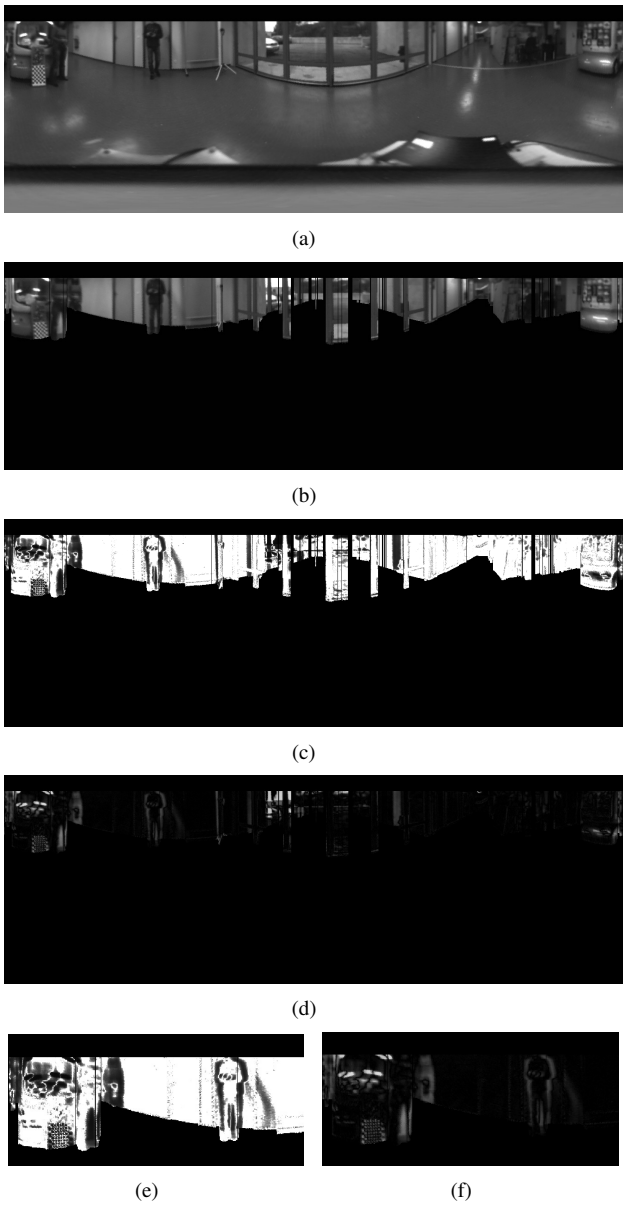
Fig. 5. 2D global maps obtained with the same laser data. 5(a) Map with EPSM pose estimation. 5(b) Map with spherical pose estimation.

robot encoders (in red). The shift that can be observed in the 2D map at the end of the loop is caused by *erroneous* laser measurements resulting in the failure of the scan matching process. Even in the presence of these errors it can be seen that the spherical tracking succeeds. This can be seen in Fig. 5(b) in blue. In this case the shift is corrected due to the redundancy between the laser and vision data leading to overall robustness and accuracy.

A representation of the images used for the pose estimation with the spherical tracking method is shown in Fig. 6. Observe visually that the final current warped image 6(b) (i.e. after convergence) is correctly matched with respect to the reference one in Fig. 6(a). Notice also that in Fig. 6(c), the moving pedestrian is rejected by the robust estimator function because his position in the current warped image differs too much from the reference one. In addition, the algorithm is capable of rejecting specular reflections on the ground and in the windows.

Under the assumption made in Sec. II that the walls are vertical, from Fig. 6(d) it is clear that the error is negligible for walls, while non vertical textured objects were not matched completely and were correctly rejected. As an example, notice the slanted calibration checkerboard on the left bottom of the images that is perfectly rejected in Fig. 6(e) and 6(f). Some parts of the static pedestrian on the left are partially matched because these parts are untextured and do not generate any matching errors, therefore estimation is not affected.

Figure 7 shows the 3D textured reconstruction with the



(a)

(b)

(c)

(d)

(e) (f)

Fig. 6. Images used for pose estimation. 6(a) Reference spherical image. 6(b) Current warped image. 6(c) Estimated rejection weights. 6(d) Final error. 6(e) 6(f) Weights and error zoom.

correctly matched 3D points obtained from the spherical tracking algorithm. Only the points that were not rejected by the robust estimation were used (i.e. with weight equal to 1). This leads to the rejection of moving pedestrians, that were plotted on the 2D maps in Fig. 5 as well as non-planar/vertical textured objects. The resulting 3D model was rendered in OpenGL and allows walk-throughs as well as bird eye views. The accompanying video[1] illustrates the incremental generation of a 2D map with both estimations of the robot trajectory, as well as the representations of the images used for the pose estimation with the spherical

[1] https://www-sop.inria.fr/arobas/videos/HybridLaserOmni_IROS10.mp4

tracking algorithm. A walk-through the 3D reconstructed model is shown at the end of the video.

In summary, the fusion of information from laser scan matching and an appearance-based method improves the robustness of localization and mapping. The robot trajectory is correctly estimated and the drift is minimized. The obtained 3D textured map represents the environment with a good level of precision as seen through the visually satisfying 3D model.

VII. Conclusion and Future work

The hybrid laser/vision appearance-based approach described in this paper has proved to be very efficient in obtaining reliable 3D odometry subject to illumination changes and in the presence of occluding and moving objects. A complete set of 3D points can be easily mapped to reconstruct a dense and consistent representation of the environment. As expected, the initialization of the tracking algorithm close to the solution using scan matching ensures fast exponential decrease of the error and avoids local minima.

Although the SLAM problem has been solved using many different approaches, some important problems need to be addressed that are often directly linked to the sensors used. Laser range finders cannot always help in evaluating the translation of a robot moving in a straight line in a corridor leading to potential observability problems. Mapping in dynamic environments is also hard using laser data only due to 2D measurements and slow acquisition rate. On the other hand, using exclusively visual sensors introduces issues such as propagating correctly the scale factor.

In this approach, the data is obtained by two full 360° field of view sensors: laser range finder and an omnidirectional camera. The experimental results are encouraging and provide a valuable insight into the possibilities offered by this hybrid approach.

In perspective, several research directions will be investigated including the extension of the formalism to deal with non-planar scenes. In this case the problem will be to formalise the uncertainties of different sensors so that the estimation can be decoupled into two separate minimization steps: 1) pose estimation, 2) depth refinement. Another direction will be to fuse the initialization, through EPSM scan matching, into the non-linear estimation scheme.

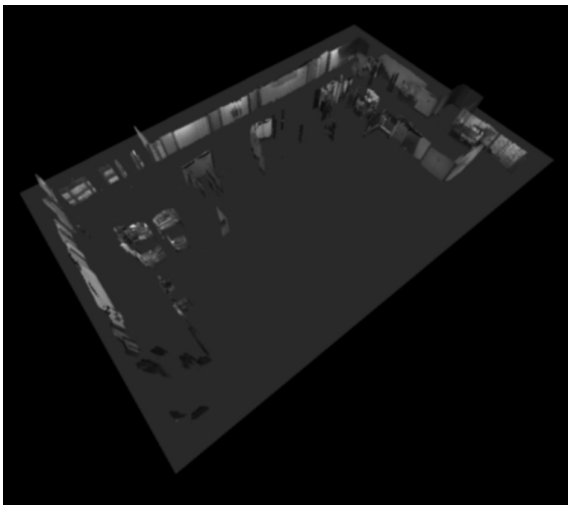


Fig. 7. 3D reconstruction.

References

[1] Calibration toolbox. `http://www.robots.ox.ac.uk/~cmei/Toolbox.html`.

[2] P.J. Besl and N.D. McKay. A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:239–256, 1992.

[3] P. Biber, H. Andreasson, T. Duckett, and A. Schilling. 3D modeling of indoor environments by a mobile robot with a laser scanner and panoramic camera. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3430–3435, 2004.

[4] T. J. Broida, S. Chandrashekhar, and R. Chellappa. Recursive 3D motion estimation from a monocular image sequence. *IEEE Transactions on Aerospace and Electronic Systems*, 26:639–656, 1990.

[5] A. Clerentin, L. Delahoche, C. Pégard, and E. Brassart. A localization method based on two omnidirectional perception systems cooperation. In *IEEE International Conference on Robotics and Automation*, pages 1219–1224, 2000.

[6] D. Cobzas, H. Zhang, and M. Jagersand. Image-based localization with depth-enhanced image map. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1570–1575, September 2003.

[7] A. Comport, E. Malis, and P. Rives. Accurate quadrifocal tracking for robust 3D visual odometry. In *IEEE International Conference on Robotics and Automation*, pages 40–45, Rome, Italy, April 2007.

[8] A.J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *IEEE International Conference on Computer Vision*, pages 1403–1410, Washington, DC, USA, 2003.

[9] A.J. Davison, Y. Gonzalez-Cid, and N. Kita. Real-time 3D SLAM with wide-angle vision. In *IFAC Symposium on Intelligent Autonomous Vehicles*, 2004.

[10] A. Diosi and L. Kleeman. Laser scan matching in polar coordinates with application to SLAM. In *IEEE International Conference on Robotics and Automation*, pages 3317–3322, 2005.

[11] Sheng Fu, Hui ying Liu, Lu fang Gao, and Yu xian Gai. SLAM for mobile robots using laser range finder and monocular vision. In *Mechatronics and Machine vision in Practice*, 2007.

[12] G. Gallegos and P. Rives. Indoor SLAM based on composite sensor mixing laser scans and omnidirectional images. In *IEEE International Conference on Robotics and Automation*, Anchorage, USA, May 2010.

[13] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical applications. In *European Conference on Computer Vision*, pages 445–461, 2000.

[14] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[15] P.-J. Huber. *Robust Statistics*. Wiler, New York, 1981.

[16] C. Mei. *Laser-Augmented Omnidirectional Vision for 3D Localisation and Mapping*. PhD thesis, INRIA Sophia Antipolis, France, 2007.

[17] C. Mei, S. Benhimane, E. Malis, and P. Rives. Efficient homography-based tracking and 3D reconstruction for single viewpoint sensors. *IEEE Transactions on Robotics*, 24(6):1352–1364, 2008. Special issue on visual servoing.

[18] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd. Monocular vision-based SLAM for mobile robots. In *International Conference on Pattern Recognition*, pages 1027–1031, 2006.

[19] J. Pedro and A. Barreto. *General central projection systems, modeling, calibration and visual servoing*. PhD thesis, 2003.

[20] M. Self R. Smith and P. Cheeseman. Estimating uncertain spatial relationships in robotics. *Proceedings of the Second Annual Conference on Uncertainty in Artificial Intelligence*, pages 435–461, 1986.

[21] G. Silveira, E. Malis, and P. Rives. An efficient direct approach to visual SLAM. *IEEE Transactions on Robotics (Special issue on visual SLAM)*, 24(5):969–980, 2008.

[22] R. Smith and P. Cheeseman. On the representation and estimation of spatial uncertainty. *International Journal of Robotics Research*, 5:56–68, 1986.

[23] Shin-Chieh Wei, Yasushi Yagi, and Masahiko Yachida. On-line map building based on ultrasonic and image sensor. In *IEEE Int. Conf. on Systems, Man and Cybernetics*, volume 2, pages 1601–1605, 1996.