LOCAL MAP BUILDING FOR MOBILE ROBOTS BY FUSING LASER RANGEFINDER AND MONOCULAR VIDEO IMAGES

Geovany Araújo Borges[*]
gaborges@ene.unb.br

Marie-José Aldon[†]
aldon@lirmm.fr

[*]Laboratório de Controle e Visão por Computador (LCVC)
Departamento de Engenharia Elétrica (ENE) - Universidade de Brasília (UnB)
Caixa Postal 04591 - Asa Norte - Brasilia - CEP 70910-900 - Brazil

[†]Département Robotique - LIRMM
UMR CNRS/Université Montpellier II, n^o. C55060
161, rue ADA. 34392 - Montpellier - Cedex 5 - France

RESUMO

Esse artigo apresenta uma técnica de cartografia local baseada na fusão de características extraídas de imagens fornecidas por *laser rangefinder* e visão monocular. O método proposto é parte de um sistema de cartografia e localização simultâneos implantado em um robô móvel real. A partir de imagens do *rangefinder*, algumas estruturas do ambiente local são observadas sob a forma de características geométricas, as quais são usadas para construir uma primeira instância do mapa local. Características da imagem de vídeo são usadas com observações complementares das características da primeira instância de mapa local, que é atualizado levando a uma maior precisão na representação espacial de estruturas do ambiente. Incertezas nos parâmetros das características são também estimadas por meio de ferramentas estatísticas. Exemplos obtidos a partir de dados reais são apresentados. O desempenho satisfatório dessa abordagem tem contribuido para a obtenção de resultados consistentes do um sistema de cartografia e localização simultâneos.

PALAVRAS-CHAVE: Cartografia local de ambientes, fusão de dados multisensoriais, extração de características, modelamento de incertezas, robôs móveis.

ABSTRACT

This paper presents a local mapping technique based on the fusion of laser rangefinder and monocular vision images. This system is currently used as part of a concurrent localization and environment mapping system in a real mobile robot. From laser rangefinder images, some local map structures are observed as geometric features and a first instance of local map is built. Video image features are used as complimentary observations of the previous local map to improve its accuracy. Feature uncertainty is also estimated using statistical techniques. The paper presents examples obtained from real data only. The satisfactory performance of this approach has lead to consistent results as part of a simultaneous localization and map building system.

KEYWORKS: Local environment mapping, multisensor data fusion, feature extraction, uncertainty modeling, mobile robotics.

1 INTRODUCTION

Concurrent Mapping and Localization (*CML*) is an important and difficult research topic in robotics. In *CML*, exteroceptive sensor data gathered at a given time is used to build a local environment representation, the local map, relative to the robot sensory system reference frame. In the context of geometrical maps (Estrada et al., 2005), the local map is used for robot pose estimation and global map updating. In this process, any error in local mapping is propagated to pose estimation and map updating (Borges, 2002). Thus, the construction of local environment maps is a primary task for simultaneous localization and environment mapping systems in mobile robotics. For instance, Arras *et al.* (Arras et al., 2001) utilize environmental geometrical features cap-

tured by a 2D laser rangefinder and a video camera for pose estimation of a mobile robot. A similar approach has been followed by Neira *et al.* (Neira et al., 1999), but using a 3D rangefinder. However, when exploring complimentary or multiple informations about environment structures captured by different sensors, more accuracy in robot positioning can be obtained. In (Lallement et al., 1998), the authors have followed this direction, and applied data fusion techniques in a robot equipped with 2D laser rangefinder and video camera for detection of local environment corners.

This paper describes a geometrical local mapping strategy based on the fusion of 2D laser rangefinder and video camera images. This system is the basis of the local mapping module of a successful simultaneous localization and global environment mapping strategy presented previously in (Borges and Aldon, 2004a)(Borges and Aldon, 2002), and is described in more detail in this paper. It differs from most approaches by the use of more than one map structures. Map structures are more complex than low-level sensor features, such that used in (Arras et al., 2001), being useful in representing most structures found in indoor environments. Its reduced computing time has made this strategy a strong component in the navigation system of a mobile platform, the Omni robot.

The paper is organized as follows. Local map structures are presented in section 2. The multisensory system of Omni robot is presented in section 3, where calibration details are described. Section 4 described the procedures used for local map building, followed by experimental results in section 5. The conclusions are presented in section 6.

2 REPRESENTATION OF LOCAL MAPS

In most works on indoor environment modeling, map structures are given by 2-D geometrical primitives: infinite lines, represented in polar parametric form $\mathbf{l} = (\rho, \alpha)^T$, and points, given by their Cartesian coordinates $\mathbf{p} = (x, y)$. These primitives are supposed to be perturbed by zero mean Gaussian noise, with covariance matrices $\Lambda_{\mathbf{l}}$ and $\Lambda_{\mathbf{p}}$ associated to $\mathbf{l}$ and $\mathbf{p}$, respectively. In this work, extended map structures using other attributes make them more complex and suitable for indoor environment representation. The map structures, shown in Fig. 1, give higher-level identity to map components. These structures are:

- *Semiplanes* correspond to walls and other planar obstacles. They are represented by (i) an infinite line $\mathbf{l}$, (ii) two end-points and (iii) one flag indicating the visibility side of the structure;
- *Edges* correspond to extremities of walls. They are represented by a point $\mathbf{p}$, and the associated semiplane index is kept for record;
- *Corners* correspond to the intersection of two walls or two consecutive planar faces. They are represented by a point $\mathbf{p}$, and the associated semiplanes indexes are kept for record;
- *Photometric edges* correspond to artifacts observed as vertical lines in video images, and that correspond neither to *edges* nor to *corners*. They are represented by a point $\mathbf{p}$.

The above structures allow a rich representation for indoor environments, and the following relations apply to them: (i) every *edge* is associated to a *semiplane*, *i.e.,* the point primitive of the *edge* is over the line primitive of the *semiplane*; (ii) every *corner* is associated to two adjacent *semiplanes*, *i.e.,* the point primitive of the *corner* is on the intersection of the lines of the associated *semiplanes*. Such relations are used as prior information to constrain map updating. Furthermore, given the map structures uncertainty, constraints should be satisfied in the stochastic sense: the uncertainty of the constrained features should also be taken into account. For instance, an *edge* might not be exactly over the line of the associated *semiplane*. It may be close to the *semiplane*, but at a distance which should compatible with the uncertainties of both structures. The propagation of such constraints during environment mapping process has been shown to minimize map divergence (Borges and Aldon, 2002).

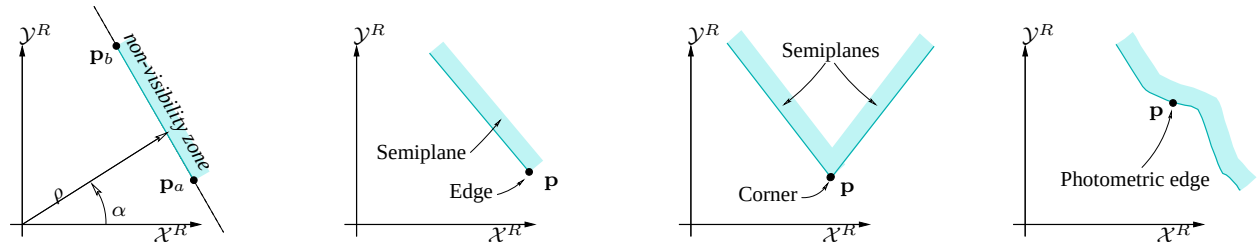
The symbols used for the structures in the maps are as follows: (—) for semiplanes, (⊗) for edges, (⊙) for corners, (△) for photometric edges.

3 MULTISENSOR SYSTEM

3.1 Description and geometry

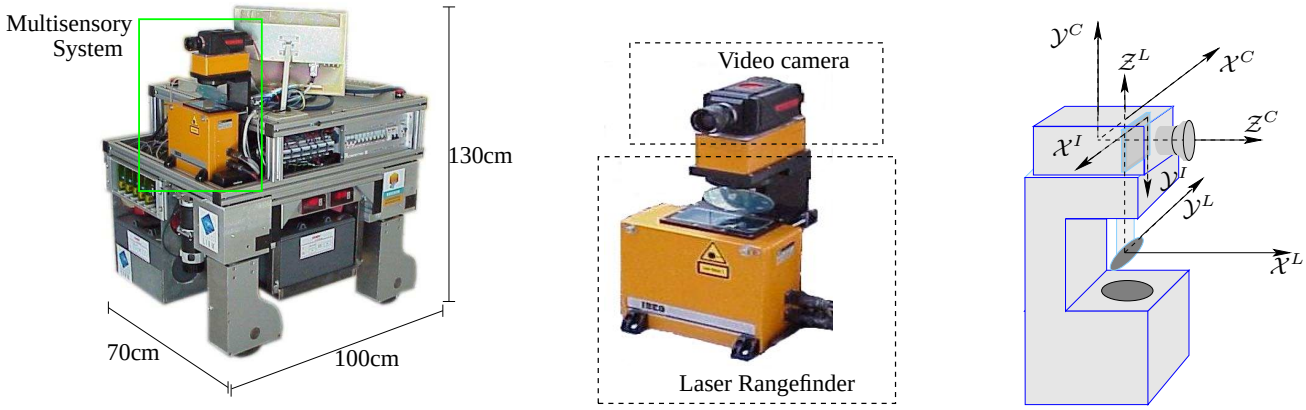
Omni is a mobile platform on which a localization and environment mapping architecture has been implemented, allowing it to navigate in indoor environments whilst a map is built. It is an omnidirectional robot with six motion axis, each one equipped with absolute and incremental encoders. A laser gyroscope completes its proprioceptive sensory system. Environment data is captured from two exteroceptive sensors, a 2D laser rangefinder and a gray-scale video camera, as shown in Figures 2(a)-(b). Figure 2(c) depicts the reference frames of the multisensory system, with $\mathcal{R}^L$: $\mathcal{X}^L \times \mathcal{Y}^L \times \mathcal{Z}^L$ being that of laser rangefinder and $\mathcal{R}^C$: $\mathcal{X}^C \times \mathcal{Y}^C \times \mathcal{Z}^C$ refers to the camera.

The rangefinder is a 2D30 model manufactured by *IBEO Lasertechnik GmbH*. It performs eight scans per second in the $\mathcal{X}^L \times Y^L$ plane and reaching objects placed up to *30m* from the origin of $\mathcal{R}^L$. The measurement principle is based on the time-of-flight of laser impulses. The laser beam carrying the impulses is generated inside the rangefinder and reflected by a 45^o rotating mirror in the $\mathcal{Z}^L$ axle. The mirror performs eight rotations per second in the counterwise sense, allowing the laser beam to cover a view of 270^o ($\pm 135^o$). The range measurements start every time the laser



(a) Semiplane (b) Edge (c) Corner (d) Photometric edge

Figura 1: Map structures.



(a) (b) (c)

Figura 2: (a) The Omni mobile robot, its (b) exteroceptive multisensory system and (c) local reference frames.

beam is at $\phi_{\min} = -135^o$ with respect to $\mathcal{X}^L$. From this direction, one measurement is done every $\Delta\phi = 0.6^o$, until the beam direction reaches $\phi_{\max} = 135^o$. Let ϕ_n denote the laser beam direction angle of the *n-th* measurement. The range measurement taken at ϕ_n is denoted as r_n. The entire set of measurements is referred as a 2D range image:

$$\mathcal{L} = \{(r_n, \phi_n) | n = 1, \ldots, N\}. \quad (1)$$

Each (r_n, ϕ_n) correspond to the polar coordinates of the intersection point between the laser beam and an obstacle at direction ϕ_n. The great accuracy of direction ϕ_n is guaranteed by an absolute encoder. However, each measurement r_n is assumed to be contaminated by an additive zero mean Gaussian noise with variance σ_r^2, denoted as $\mathcal{N}(0, \sigma_r^2)$. A previous characterization procedure of the rangefinder detected bias in the range measurements. Such a bias (moving average) is compensated in software after each scan, allowing the zero mean noise hypothesis to be assumed. The measurement standard deviation is $\sigma_r = 3\ cm$.

The video camera acquires gray-scale images represented by

$$\mathcal{I} = \{p(u, v) | u = 1, \ldots, M; v = 1, \ldots, N)\}. \quad (2)$$

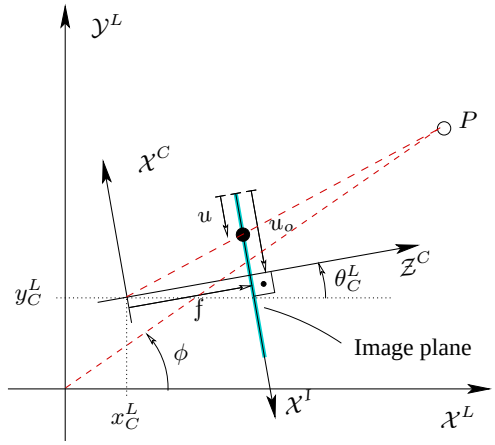
with $p(u, v)$ being the intensity of each pixel at coordinates (u, v) in the *image plane*, attached at $\mathcal{R}^I : \mathcal{X}^I \times \mathcal{Y}^I$ reference frame. Considering the simplified pin-hole geometrical model for the camera (Jähne and Hauβecker, 2000), the optical axis is orthogonal to the image plane at metric coordinate (x_o^I, y_o^I) of $\mathcal{R}^I$. The origin of the camera reference frame $\mathcal{R}^C$ is at distance f of (x_o^I, y_o^I), where f is the focal distance. The $\mathcal{X}^C$ and $\mathcal{X}^I$ are parallel and in opposite directions, such as $\mathcal{Y}^C$ and $\mathcal{Y}^I$. $\mathcal{Z}^L$ and $\mathcal{Y}^C$ are parallel and share the same direction.

3.2 Calibration

The positioning between camera and rangefinder allows to easily relate vertical lines observed in video images and their observation angle in the rangefinder reference frame. This is used to relate video and range image features during data fusion in the local map building process. In order to to this, the geometry between these sensors should be described in more detail, as well as calibration is necessary. Firstly, we suppose $\mathcal{Z}^L$ and $\mathcal{Y}^C$ are parallel. This results in a simplified superimposed view of $\mathcal{X}^C \times \mathcal{Z}^C$ and $\mathcal{X}^L \times \mathcal{Y}^L$ planes, as shown in Figure 3. In order to perform information fusion between the images of the multisensory system, we are interested in obtaining a relationship

$$\phi = g(u), \quad (3)$$

with u being the column index of a vertical line observed in the video image $\mathcal{I}$, and ϕ is its observation angle from the



(a)

Figura 3: Simplified superposition of laser rangefinder and video camera reference frames.

rangefinder.

Considering the rangefinder reference frame $\mathcal{X}^L \times \mathcal{Y}^L$, let $\mathbf{t}_C^L = (x_C^L, y_C^L)$ denote the projection point of the origin of $\mathcal{X}^C \times \mathcal{Z}^C$, as well as θ_C^L is the projected angle between $\mathcal{Z}^C$ and $\mathcal{X}^L$. Thus, $(x_C^L, y_C^L, \theta_C^L)$ characterize misalignment between the two sensors in the projected rangefinder reference frame. Let P be a range image point with coordinates $\mathbf{p}^L = (x^L, y^L)^T$ observed at ϕ direction. Its projection in the image plane is given by u. Further, let $\mathbf{p}^C = (z^C, x^C)^T$ be the coordinates of P in $\mathcal{X}^C \times \mathcal{Z}^C$. Such parameters are related by

$$\mathbf{p}^C = \mathbf{R}(\theta_C^L) \cdot \left(\mathbf{p}^L - \mathbf{t}_C^L\right), \tag{4}$$

with

$$\mathbf{R}(\theta_C^L) = \begin{bmatrix} \cos(\theta_C^L) & \sin(\theta_C^L) \\ -\sin(\theta_C^L) & \cos(\theta_C^L) \end{bmatrix}. \tag{5}$$

From projective geometry, P is observed at image column

$$u = u_o - \frac{x^C}{z^C} \cdot \frac{f}{h_u}, \tag{6}$$

with u_o being the column index of x_o^I, h_u is the metric width of a pixel. Since $x^L = r\cos(\phi)$ and $y^L = r\sin(\phi)$, with r being the range measurement of P, an analytical form of equation 3 can be derived from eqs. (4)-(6). It can be verified that $g(u)$ can be approximated by an affine function when there is no sensor misalignment, *i.e.*, $\theta_C^L \to 0$ and $\left\|\mathbf{t}_C^L\right\| \to 0$, which lead to $\arctan$ be approximately linear to $u - u_o$. Hence, in order to avoid nonlinear least-squares estimation techniques for calibration, sensor misalignment has been done manually as low as possible. In this way, a reasonable form of $g(u)$ is a first order polynomial (affine function).

The calibration procedure consisted of acquiring 50 pairs of images ($\mathcal{L}$,$\mathcal{I}$) of different positions of the robot with respect to two targets. The targets are vertical wood boards, whose both extremities were marked by black tires. In this way, such extremities are easily identified in video images. In range images, these extremities appear as large discontinuities in the scans. Figure 4(a) shows a video image, with four columns well identified as corresponding to the targets. The observation directions in the corresponding range image are shown in Figure 4(b). From the entire set of images, 152 pairs of corresponding rangefinder observation angle ϕ_m and camera vertical indexes u were used in a linear least-squares regression procedure. The ranges r were limited to $0.5\ m \leq r \leq 10.5\ m$. The estimated model is given by:

$$\hat{g}(u) = \hat{a} \cdot u + \hat{b}, \tag{7}$$

with $\hat{a} = -0.00162942$ and $\hat{b} = 0.32413517$. The estimation residuals $\phi_m - \hat{g}(u)$ with respect to u are shown in Figure 4(c). It should be pointed out that, despite the nonlinear characteristic of the geometric model, the linear approximation seems to be satisfactory, since the residual error variance is constant for the experiment. In this way, the stochastic model

$$\hat{\phi} = \hat{a} \cdot u + \hat{b} + \varepsilon, \tag{8}$$

is used henceforth, with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and $\sigma_\varepsilon = 0.412°$. This model allows to estimate the variance $\sigma_{\hat{\phi}}^2$ of $\hat{\phi}$ from the variance σ_u^2 related to u and the model error variance as

$$\sigma_{\hat{\phi}}^2 = \hat{a}^2 \cdot \sigma_u^2 + \hat{\sigma}_\varepsilon^2. \tag{9}$$

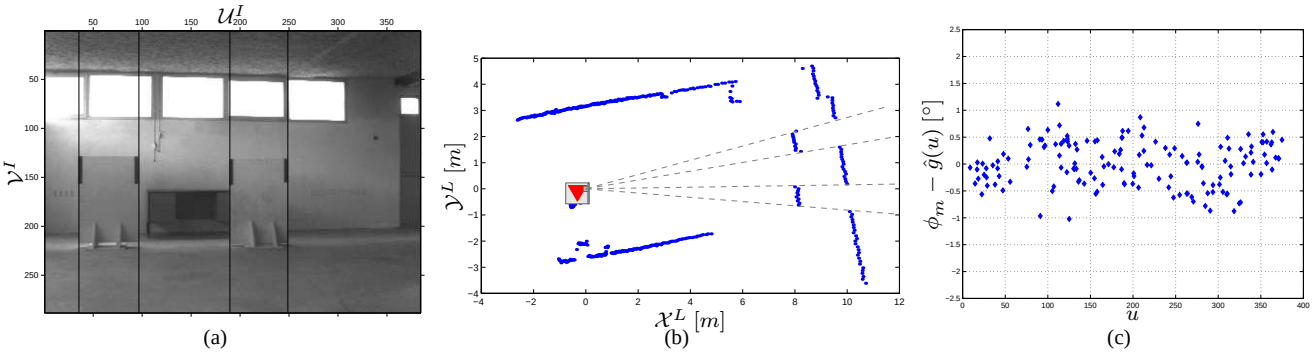
4 LOCAL MAP BUILDING

The local map $\mathcal{M}^R$ is obtained from the pair of synchronized exteroceptive sensor images ($\mathcal{L}, \mathcal{I}$). Multisensor data fusion consists in using vertical edges extracted from the video image to improve a range image map. In order to accomplish this, the following steps are followed: feature extraction, rangefinder map building and sensor data fusion. These procedures are described in this section.

4.1 Feature extraction

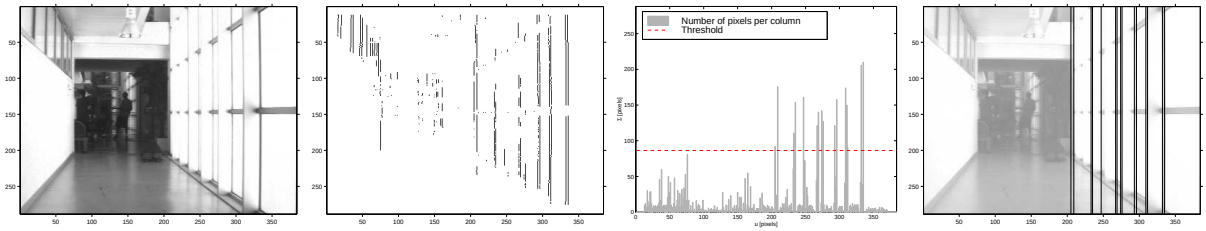
The feature extraction process issues two lists of features :

- $\Omega^U = \{U_k \mid k = 1, \ldots, N_{\Omega^U}\}$ is a list of indexes $U_k = \{u_k, \boldsymbol{\sigma}_{u_k}^2\}$ extracted from the video image $\mathcal{I}$. For the *k-th* feature, u_k is the index of the column related to the *k-th* vertical edge, and $\boldsymbol{\sigma}_{u_k}^2$ is its variance;
- $\Omega^S = \{S_k \mid k = 1, \ldots, N_{\Omega^S}\}$ is a list of line segments $S_k = \{\mathbf{l}_k, \mathbf{\Lambda}_{\mathbf{l}_k}, \mathbf{p}_{a_k}, \kappa_{a_k}^+, \mathbf{p}_{b_k}, \kappa_{b_k}^+\}$ extracted from the rangefinder image $\mathcal{L}$. For the *k-th* feature, $\mathbf{l}_k = (\rho_k, \alpha_k)^T$ represents the polar parameters of a infinite line, and $\mathbf{\Lambda}_{\mathbf{l}_k}$ is its covariance matrix. $\mathbf{p}_{a_k}$ and $\mathbf{p}_{b_k}$ are points corresponding to the extremities of the line segment. $\kappa_{a_k}^+$ and $\kappa_{b_k}^+$ are flags associated to $\mathbf{p}_{a_k}$ and $\mathbf{p}_{b_k}$, respectively, to indicate whether these extremities are strong breakpoints (discussed below);



(a) (b) (c)

Figura 4: Example used for calibration: (a) Vertical lines in video image and (b) projected observation angles in range image. In (c) obtained calibration residuals for the entire set of image pairs.



(a) (b) (c) (d)

Figura 5: Video image feature extraction: (a) raw image, (b) binary image with enhanced vertical edges, (c) accumulator and (d) extracted vertical lines.

Feature extraction of video image is based on well known techniques. Figure 5 gives an example of the application of these procedures. Firstly, a Canny filter (Canny, 1986) is applied for edge enhancement in both vertical and horizontal directions of $\mathcal{I}$ (Figure 5(a)), followed by local maxima selection. Pixels belonging to vertical edges are selected by eliminating pixels with gradient direction greater than 10° with respect to vertical. A binary image is obtained by applying hysteresis thresholding based on 10% and 60% histogram percentiles of the gradient module, as shown in Figure 5(b). The main drawback of these procedures is that very high (or low) levels of image brightness lead to no detection at all. Other more robust techniques may be employed (Jolion et al., 1991), at the cost of computational complexity. From the binary image, a specialized Hough accumulator is applied. The accumulator counts the number of active pixels of each column, as shown in Figure 5(c). A column is considered as a valid vertical edges if the number of active pixels is greater than $30\ \%$ of the number of pixels per column. This fixed threshold is justified, since normalization procedures have been applied in the previous steps. For the example, the resulting edges are shown in Figure 5(d).

Since each vertical edge can be composed of more than one column, starting from u_d up to u_f, these indexes are also considered for determining the parameters $U = \{u, \mathbf{\Lambda}_u\}$. In a first sight, the following statistics can be used:

$$u = \frac{\sum_{x=u_d}^{u_f} x \cdot A(x)}{\sum_{x=u_d}^{u_f} A(x)}, \qquad \boldsymbol{\sigma}_u^2 = \frac{\sum_{x=u_d}^{u_f} (u-x)^2 \cdot A(x)}{\sum_{x=u_d}^{u_f} A(x)}, \tag{10}$$

where $A(x)$ is the contents of the Hough accumulator at x column. However, $\boldsymbol{\sigma}_u^2$ becomes zero for the case of $u_d = u_f$, resulting in an inconsistent measure of variance given that the pixel width is not considered. Thus, we propose to use a 3σ reasoning about uncertainty in the interval $[(u_d - 0,5)\,, (u_f + 0,5)]$:

$$\boldsymbol{\sigma}_u^2 = \left[\max(u_f - u + \frac{1}{2}, u - u_d + \frac{1}{2}) \cdot \frac{1}{3}\right]^2. \tag{11}$$

From the range image $\mathcal{L}$, Ω^S is extracted by using the fast split-and-merge fuzzy algorithm (Borges and Aldon, 2004b). This algorithm uses classic fuzzy clustering for line finding, but in a split-and-merge manner. It is well known that most clustering approaches should use a guess of the number of clusters exist in a set of support points, which in this case means to know in advance the number of lines in a range image. In a split framework, this problem is solved by calling recursively the clustering algorithm to find two clusters every time. If one of the clusters cannot be validated as a line feature, the clustering algorithm is applied again to this cluster until having only valid line clusters composed of at least 10 support points. In the merge phase, very close clusters are merged. Laser scan breakpoints are detected using

an extended Kalman filter-based approach (Castellanos and Tardós, 1996). Breakpoints correspond to important discontinuities in the range image sequence, and may indicate the extremities of local environment surfaces. Breakpoints used in determining line extremities [1] $\mathbf{p}_{a_k}$ or $\mathbf{p}_{b_k}$ have their corresponding $\kappa^+_{a_k}$ or $\kappa^+_{b_k}$ flags set.

4.2 Local rangefinder map building

In this procedure, a first instance of the local map is built using only features extracted from the range image. This map can be composed by semiplanes, edges and corners. Figure 6(a) gives an example of line features extracted from a range image. In order to determine each map structure, the following steps are followed:

- *Semiplane structures*: every line segment extracted from the range finder image is converted to a semiplane. The visibility side of the semiplane is identified by a flag. A simple test procedure allows to identify whether the coordinates of a point in the local map is on the visibility side of a semi plane. The visibility side is a important feature of a semiplane, since it allows to correctly solve matching problems. The default visibility side of a semiplane is set to the side where the robot observed it. In the example, see Figure 6(b);

- *Edge structures*: all line segments extremities marked as strong breakpoint are candidate to be considered as edge structures. However, only those with small associated uncertainty are accepted as edges. The edge uncertainty is obtained considering the uncertainties associated to the scan points used in the determination of the line segment extremity, as well as the line segment uncertainty (Borges, 2002). In the example, see Figure 6(c);

- *Corner structures*: all line segments having extremities which are two consecutive range scan points are candidates to be used for corner determination. These extremities should not be classified as strong breakpoints. The crossing point between the line candidates is computed as well as its covariance matrix by uncertainty propagation from line parameters covariance matrices. Such a point is accepted as corner whether the angle between the supporting line features is greater than 60°. In the example, see Figure 6(d) shows only one corner found;

4.3 Sensor data fusion

The last procedure in local map building uses the vertical lines extracted from the video image to update the initial local map structures obtained from range image. The sensor fusion is performed for (i) reducing edges uncertainties and (ii) extracting photometric edges. This procedure is well illustrated in Figure 7.

Edges found in the local map could be observed by the camera as vertical line features U_1 and U_2 (see Figure 7(a)). Thus, the classical Extended Kalman Filter (EKF) is used to update the point coordinates (x_i, y_i) of all edges whose observation angle $\phi_i = \arctan(y_i/x_i)$ in the range image local map is in correspondence with the observation angle $\hat{\phi}_j = \hat{g}(u_j)$ of vertical edge U_j (eq. (7)). Such correspondences are verified using the classical χ^2-hypothesis test. For instance, Figure 7(b) shows the estimated lines of sight $\hat{O}_1$ and $\hat{O}_2$ in the local map associated to vertical features U_1 and U_2, respectively. From these features, only U_2 is in correspondence with an edge, as can be seen in more detail in Figure 7(c). Using the EKF formalism, the edge parameters (x_i, y_i) obtained from range features represent the predicted coordinates of the edge, with associated covariance matrix. The angle of view $\hat{\phi}_j$ is used to update such coordinates as an observation in the EKF measurement model. The resulting edge is shown in Figure 7(d). Even it is not visible in this figure, the updated edge is closer to the line of sight of feature U_2, and its associated uncertainty is reduced.

For last, photometric edges are estimated from all vertical lines which did not have correspondences in the initial local map. Since these structures are composed of a point, and the camera can only capture their observation angle $\hat{\phi}$, they are estimated using 100 bootstrap samples (Efron, 1979), obtained from the distribution of all scan points which lie, according to the χ^2-hypothesis test, in the vertical line angle of view $\hat{\phi}$. Photometric edges presenting large covariance matrices are discarded. Such a phenomena may arrive if the local region is very cluttered. In this way, we obtain reliable photometric edges. In in Figure 7(d), the vertical edge U_1 has originated a photometric edge in the line of sight $\hat{O}_1$ of the local map. It can be seen in Figure 7(a) that such a feature corresponds to the border in the right of a board fixed on a wall. Such features can only be detected using both range and video images. They are useful for robot localization, and should be represented in local maps. In this case, this is a photometric edge.

5 EXPERIMENTAL EVALUATIONS

The previous sections have shown some examples using real experimental data acquired with Omni. Thus, the experimental evaluations shown here illustrate some further properties of the proposed approach. In Figures 8(a)-(b) laser and video image features are used to build the local map of Figure 8(c) (Borges and Aldon, 2004a). In such example, the robot is in a narrow corridor, where features are more difficult to be detected.

In Figure 9(a), a sequence of range images are superimposed to illustrate the need to differentiate between edges, corners

[1] the extremities of a line segment are obtained by projecting the first and the last (in order of acquisition) support points in the line

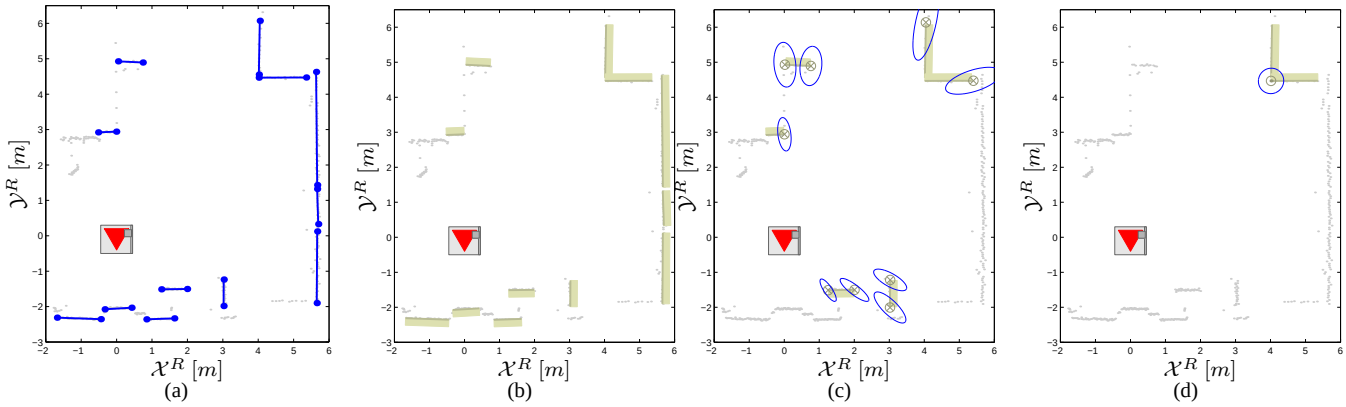


Figura 6: Local map obtained from a range image: (a) features obtained from a range image, (b) semiplanes, (c) edges and (d) corners. Uncertainty related to edges and corners are represented by ellipses, which were intentionaly increased in order to be visible.

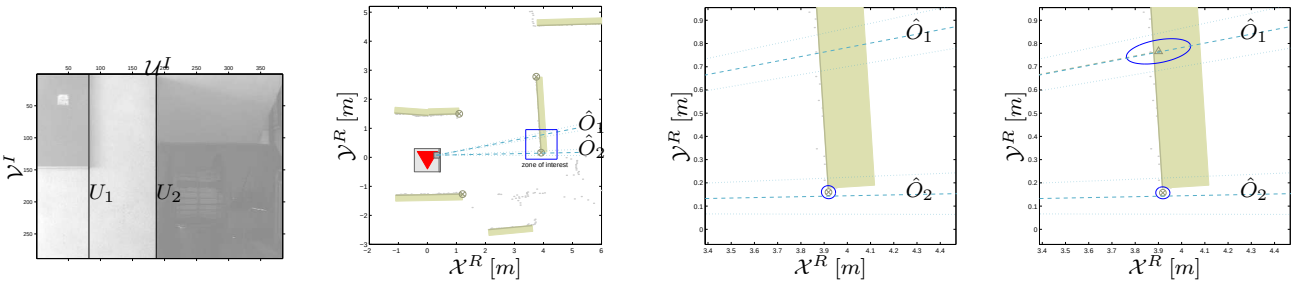


Figura 7: Example of sensor fusion: (a) vertical edges in video image, (b) lines of sight of these features in the range image map, interest zone before (c) and after (d) fusion.

and photometric edges. The features indicated by letters A, B and C are also observed in some video images, such as Figure 9(b) shows. In this sequence, these features are sometimes detected in the local maps as edges, corners or photometric edges. In means that the semantics related to each map structure can change according to its local observation. This rich variety of geometric features represented by points in the local map has lead to improve convergence of our previous work on environment mapping (Borges and Aldon, 2002). The computing time of the proposed method has been of at most $350\ ms$ in the Omni platform machine, based on a Pentium II 333 MHz CPU.

6 CONCLUSIONS

This paper presented the local mapping system used in Omni mobile robot. This system applies data fusion strategies of local environment information acquired by a video camera and a laser rangefinder. It is the the basis of the local mapping module of a successful simultaneous localization and environment mapping strategy presented previously in (Borges and Aldon, 2004a)(Borges and Aldon, 2002). In order to allow 3D environment modeling, the current work is based on the replacement of the monocular camera by a stereo system.

ACKNOWLEDGEMENTS

This work has been partially supported by Capes, Brasilia - Brasil. The Omni robot has been kindly borrowed to Universidade de Brasilia (Brasil) by *Université Montpellier II* (France), after agreement between the two institutions. The authors wish to thank the anonymous reviewers for their constructive comments.

REFERÊNCIAS

Arras, K. O., Tomatis, N., Jensen, B. T. and Siegwart, R. (2001). Multisensor on-the-fly localization: Precision and reliability for applications, *Robotics and Autonomous Systems* **34**: 131–143.

Borges, G. A. (2002). *Cartographie de l'environnement et localisation robuste pour la navigation de robots mobiles*, PhD thesis, Université Montpellier II, LIRMM, 161 rue ADA, 34392, Montpellier, Cedex 5, France. *One of the recipients of the 2001/2002 Club EEA prize for the best french thesis in Automatic Control*.

Borges, G. A. and Aldon, M.-J. (2002). A decoupled approach for simultaneous stochastic mapping and mobile robot localization, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 558–563.

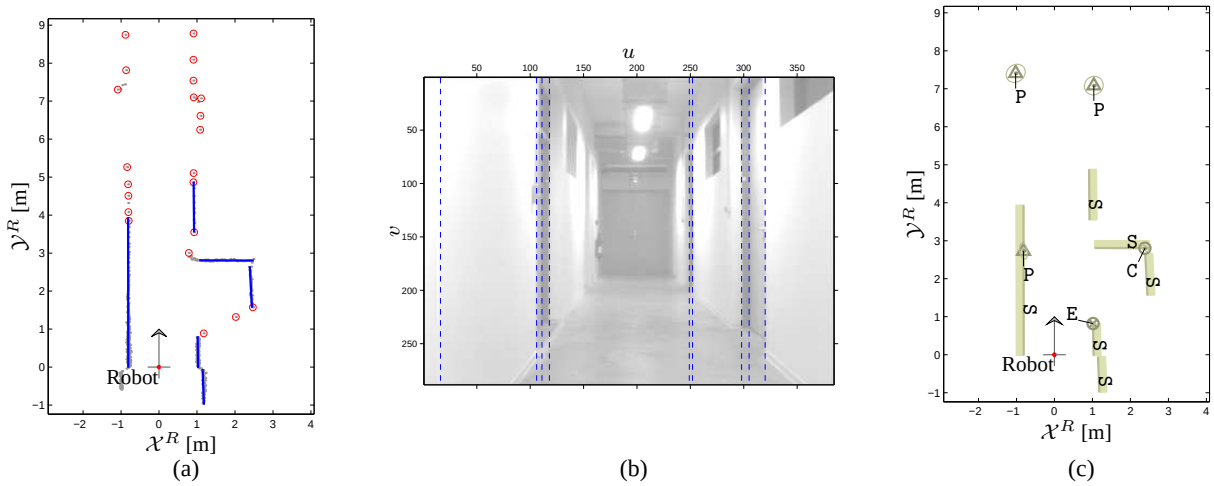


Figura 8: Local map building example . (a) Range image and features (line segments and breakpoints). (b) Video image and features (vertical lines). (c) Local map (S: Semiplanes, E: Edges, C: Corners, and P: Photometric edges

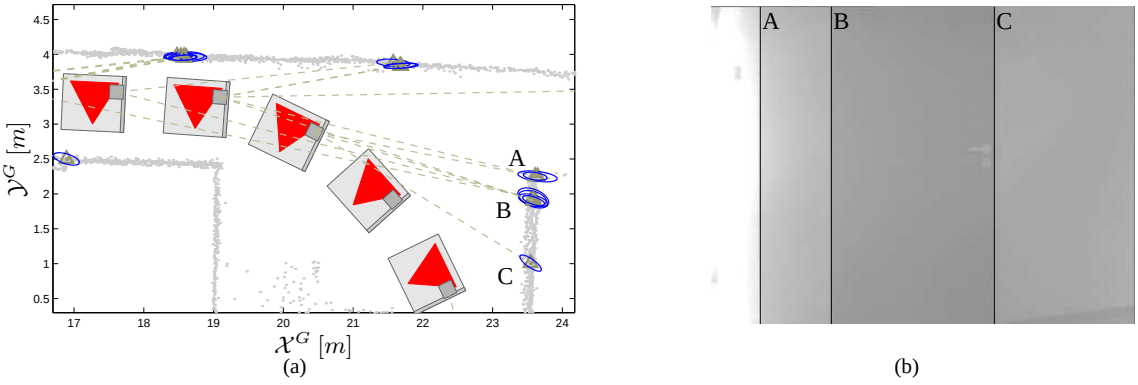


Figura 9: (a) Sequence of experimental data ilustrating coherence in consecutive detection of edges, corners and photometric edges. (b) video image showing these features observed as vertical lines.

Borges, G. A. and Aldon, M.-J. (2004a). Environment mapping and robust localization for mobile robots navigation in indoor environments, *Congresso Brasileiro de Automática, Gramado, Brasil*.

Borges, G. A. and Aldon, M.-J. (2004b). Line extraction in 2D range images for mobile robotics, *Journal of Intelligent and Robotic Systems* **40**(3): 267–297.

Canny, J. (1986). A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(6): 679–698.

Castellanos, J. A. and Tardós, J. D. (1996). Laser-based segmentation and localization for a mobile robot, *in* F. P. M. Jamshidi and P. Dauchez (eds), *Robotics and Manufacturing: Recent Trends in Research and Applications*, Vol. 6, ASME Press, pp. 101–109.

Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Journal of Ann. Stat.* **7**: 1–26.

Estrada, C., Neira, J. and Tardós, J. D. (2005). Hierarchical SLAM: Real-time accurate mapping of large environments, *IEEE Transactions on Robotics* **21**(4): 588–596.

Jähne, B. and Hauβecker, H. (2000). *Computer vision and Applications: a Guide for Students and Practitioners*, Academic Press, Londom, UK.

Jolion, J. M., Meer, P. and Bataouche, S. (1991). Robust clustering with applications in computer vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(8): 791–802.

Lallement, A., Siadat, A., Dufaut, M. and Husson, R. (1998). Laser-vision cooperation for map building and landmarks recognition, *IEEE ISIC/CIRA/ISAS Joint Conference*, pp. 387–392.

Neira, J., Tardós, J. D., Horn, J. and Schmidt, G. (1999). Fusing range and intensity images for mobile robot localization, *IEEE Transactions on Robotics and Automation* **15**(1): 76–84.