Proceedings of the 2004 IEEE
International Conference on Robotics & Automation
New Orleans, LA • April 2004

A Global 3D Map-Building Approach Using Stereo Vision

Juan Manuel Sáez and Francisco Escolano
Robot Vision Group
Departamento de Ciencia de la Computación e Inteligencia Artificial
Universidad de Alicante, Ap.99, E-03080, Alicante, Spain
Email: {jmsaez, sco}@dccia.ua.es

Abstract— **In this paper we present a stereo-based approach for building 3D maps. First, the best local alignment between successive point clouds is computed by a fast ego-motion/action-estimation algorithm which relies on an incremental matches filtering process followed by energy minimization. Then, a quasi-random updating algorithm, a kind of multi-view ICP, minimizes the global inconsistency of the map. Such an inconsistency is defined in terms of the sum of local inconsistencies and an additional entropy-based regularization term which is effective in plane-parallel environments. For the sake of efficiency, we assume a flat floor and a fixed stereo camera mounted on the robot. We have successfully tested the approach by performing several indoor mapping experiments.**

I. Introduction

Recently, there has been a great interest in using mobile robots to build 3D maps of the environment. Existing approaches using laser range finders sensors, have evolved towards obtaining more and more compact 3D models made up of planar surfaces [18][8][11][5]. In such approaches, the problem of simultaneously computing the map and robot poses is formulated in terms of maximizing a log-likelihood function and then using an EM-like algorithm to obtain, at least, a local maximum. In other approaches, range finders are used in combination with cameras, following the idea of complementing sensor capabilities. For instance, in [6] a range finder is combined with eight CCD cameras to compute a mesh model of the environment from a large number of overlapped 3D images. In [4][1], a range scanner is combined with a single camera through fusing the results of 2D and 3D matching and planar segmentation processes, and a high-level planning system obtains the best next view for acquiring the model of the environment.

In this paper we focus on a third group of approaches that only use stereo cameras as 3D sensors. In [13], 3D information is condensed in 2D maps which are represented with the 2D occupancy grid model, because its 3D generalization [12] is neither computational feasible nor scalable. Thus, practical stereo approaches must perform some sort of feature extraction. For instance, in [7] stereo data is used to compute the Hough transform of the 3D planes in the environment and such planes are extracted through a voting scheme, although some manual guidance is used when there is a lack of sensor input. In [9], stereo vision is fused with inertial information in order to recover 3D segments. On the other hand, in [15][16] 3D landmarks based on scale-invariant image features are used to compute the map. Such a computation relies on estimating the ego-motion of the robot, tracking the landmarks using the odometry for prediction, and finally superposing the landmarks to obtain the map. Such an approach is not globally consistent because it only relies on local estimations. This is why it is extended in [17], although the global consistency proposed exploits the closing-the-loop constraint for performing backward corrections.

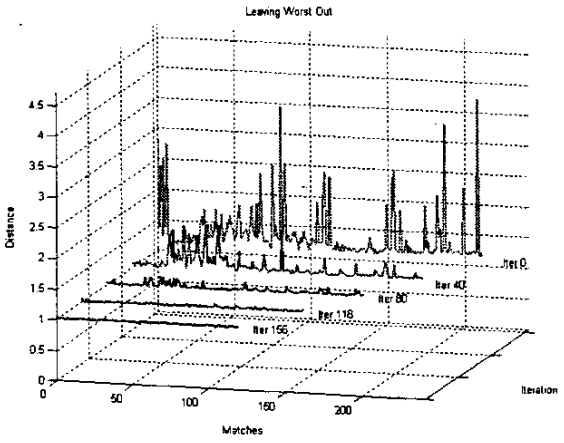

Fig. 1. Leaving-the-worst-out. Evolution of the averaged-ratio curves.

We propose an stereo approach that uses a subset of the 3D point clouds (those whose 2D projections are strict local maxima of the gradient image) and the local appearances of their projections, to feed an ego-motion/action-estimation process. Such a process performs local 3D alignment through filtering matches potentially associated to outliers, followed by energy minimization. Once we have a sequence of poses and actions/transformations we build an initial map by aggregating all point clouds with respect to a common reference system. Given such a map, we proceed to maximize a global-consistency criterion through minimizing the sum of local energies associated to each of the point clouds of the map. Each local energy consists of the averaged distance between the 3D points of its cloud and their closest point in the rest of the clouds. Then, we proceed to find the set of ac-

tions/transformations that minimize the sum of local energies (global energy). In order to do so, we enter a quasi-random updating process, which can be seen as a multi-view version of the classical ICP algorithm [2][10]. In each iteration, we simultaneously change all actions/transformations, being the amplitude of such a change proportional to their local energy. In this regard, our approach is close to the one presented in [3], where a genetic algorithm explores the space of actions/transformations looking for the sequence yielding the most consistent map. However, for the sake of computational efficiency we assume, both in the local action estimation and in the global rectification processes, a flat floor and a fixed camera. Thus, the robot and the camera are confined to the XZ horizontal-frontal plane and only rotation with respect to the vertical Y axis is allowed. Moreover, if we also assume a plane-parallel environment, in which all planes are either parallel or orthogonal, we can take advantage of a regularization term. Such a term relies on the entropies of the map projections both on the X and Z axes, and it allows us to rectify curved corridors that were originally straight. These latter assumptions are may be proper specially when working in indoor environments. On the other hand, we make no additional assumptions regarding cyclic robot trajectories.

The technical details of the approach are described in Section II. In Section III we present some indoor mapping experiments performed by tele-operating a mobile robot equipped with a stereo camera that implements our approach. Finally, in Section IV we present our conclusions and future work.

II. Approach Description

A. Action Estimation

Let $\mathbf{p}_t = [x_t, z_t, \theta_t]^T$ the $t-$th pose of the robot, where the two first components are the coordinates in the ZX horizontal-frontal plane, and the third one is the orientation with respect to the vertical axis Y. Such a pose is obtained from $\mathbf{p}_{t-1}$ through the incremental action $\mathbf{a}_t = [\delta x_t, \delta z_t, \delta\theta_t]^T$ which is supposed to be unknown. In order to estimate such an action we exploit both the 3D and appearance information contained in observations $\mathbf{O}_t$ and $\mathbf{O}_{t-1}$ through the following steps:

1) *Feature Extraction.* Let $\mathbf{C}_t(x,y,z)$ the 3D point cloud observed from the $t-$th pose, and let $\mathbf{I}_t(u,v)$ the right intensity image of the $t-$th stereo pair (reference image). For the sake of both efficiency and robustness, instead of considering all points $\mathbf{M}_t = [x_t, y_t, z_t]^T \in \mathbf{C}_t$ we retain only those points whose projections $\mathbf{m}_t = [u_t, v_t]^T \in \mathbf{I}_t$ are associated to strict local maxima of $|\nabla \mathbf{I}_t|$. These points define the constrained cloud $\tilde{\mathbf{C}}_t(x,y,z)$. The same holds for $\mathbf{C}_{t-1}(x,y,z)$ and $\mathbf{I}_{t-1}(u,v)$.
2) *Feature Matching.* Under the assumptions of flat floor and fixed camera, a point $\mathbf{M}_t = [x_t, y_t, z_t]$ should match another point $\mathbf{M}_{t-1} = [x_{t-1}, y_{t-1}, z_{t-1}]$ lying in the plane $Y = y_t$, that is, satisfying $|y_t - y_{t-1}| \leq \delta$, with δ sufficiently small. This reduces considerably the complexity of our matching process. However, in order to find such a match we will measure the similarity between the local appearances in the neighborhoods of their respective projections $\mathbf{m}_t$ and $\mathbf{m}_{t-1}$. Furthermore, in order to provide certain degree of invariance to change of texture appearance, such a matching relies on the Pearson correlation ρ (illumination invariance) between the log-polar transforms $\mathbf{LP}$ (local orientation invariance) of the windows $\mathbf{W}_{\mathbf{m}_t}$ and $\mathbf{W}_{\mathbf{m}_{t-1}}$ centered on both points, that is, $\mathbf{M}_{t-1}$ must maximize the score

$$S(\mathbf{M}_t, \mathbf{M}_{t-1}) = |\rho(\mathbf{Z}_t, \mathbf{Z}_{t-1})|, \qquad (1)$$

being $\rho(\mathbf{Z}_t, \mathbf{Z}_{t-1}) \in [-1, 1]$ the correlation coefficient of the random variables associated to the grey intensities of the log-polar mappings:

$$\mathbf{Z}_t = \mathbf{LP}(\mathbf{W}_{\mathbf{m}_t}), \mathbf{Z}_{t-1} = \mathbf{LP}(\mathbf{W}_{\mathbf{m}_{t-1}}). \qquad (2)$$

In order to ensure the quality of the match, we reject candidates with: (i) low strength $S(\mathbf{M}_t, \mathbf{M}_{t-1}) \leq S_{min}$; (ii) low distinctiveness, that is, exists also $\tilde{\mathbf{M}}_t$ satisfying $S(\mathbf{M}_t, \mathbf{M}_{t-1})/S(\mathbf{M}_t, \tilde{\mathbf{M}}_{t-1}) = R_{min} \approx 1$; and (iii) unidirectionality, that is, for $\mathbf{M}_t$, the match maximizing $S(\mathbf{M}_t, \mathbf{M}_{t-1})$ is $\tilde{\mathbf{M}}_{t-1}$, but for this latter one, the match maximizing $S(\mathbf{M}_{t-1}, \mathbf{M}_t)$ is $\tilde{\mathbf{M}}_t \neq \mathbf{M}_t$.
3) *Matching Refinement.* Despite considering the three later conditions, the matching process is prone to outliers. Thus, after computing the best matches for all points, we proceed to identify and remove potential outliers. Suppose that the $i-$th point $\mathbf{M}_t^i$ of $\mathbf{C}_t$ matches the $j-$th point $\mathbf{M}_{t-1}^j$ of $\mathbf{C}_{t-1}$, and similarly $\mathbf{M}_t^k$ matches $\mathbf{M}_{t-1}^l$. Let $d_{ik} = ||\mathbf{M}_t^i - \mathbf{M}_t^k||$ and $d_{jl} = ||\mathbf{M}_{t-1}^j - \mathbf{M}_{t-1}^l||$. Let also D_{ikjl} be the maximum between the ratios d_{ik}/d_{jl} and d_{jl}/d_{ik}. Then, in order to preserve structural coherence it is better to retain matches where D_{ikjl} is close to the unit and remove the others. More globally, in order to consider whether $\mathbf{M}_t^i$ matches $\mathbf{M}_{t-1}^j$ or not, we evaluate the quantity

$$D_{ij} = \frac{\sum_k \sum_l D_{ikjl}}{|\mathcal{M}|}, \qquad (3)$$

where $\mathcal{M}$ is the current set of matches, that is, for testing whether a given match should be removed or not, we consider the averaged sum of its maxima.
Leaving-the-worst-out is an iterative process in which we remove the match in $\mathcal{M}$, and their associated points, with higher D_{ij} and then proceed to re-compute, in the next iteration, the maxima for the rest of matches. If we plot the D_{ij} for all matches each iteration (see curves in Fig. 1) we can see that as we iterate we tend to a flat curve, that is, a curve whose standard deviation tends to zero. Thus, we stop the process when either such a deviation reaches σ_{min}, being σ_{min} sufficiently small, or a minimum number of matches $|\mathcal{M}|_{\min}$ is reached.
4) *Action/Transformation Estimation.* The purpose of the leaving-the-worst-out process is to provide a set of good-quality matches in order to face action estimation directly. The idea is to perform both the refinement and action estimation once, that is, to avoid an interleaved

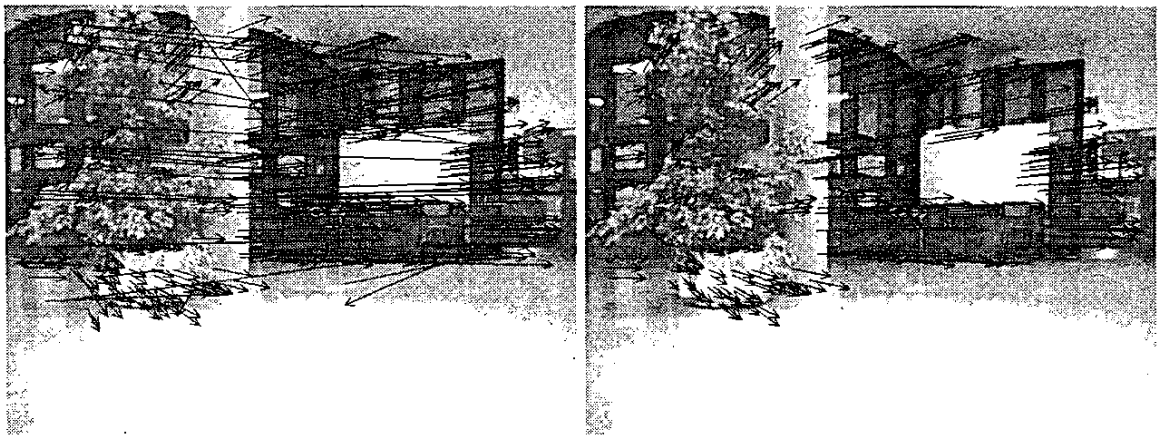

Fig. 2. Projections of the initial matches before refinement (left) and after refinement (right).

EM-like estimation process. Then, let $\mathbf{R}_t$ and $\mathbf{t}_t$ the 3×3 rotation matrix and 3×1 translation vector, respectively, associated to action $\mathbf{a}_t$. Paying attention to the constrained 3D point clouds $\tilde{\mathbf{C}}_t(x,y,z)$ and $\tilde{\mathbf{C}}_{t-1}(x,y,z)$, and given that each point $\mathbf{M}_t^i$ in the first cloud matches point $\mathbf{M}_{t-1}^j$ in the second one, the optimal action is the one yielding the transformation (rotation and translation) that maximizes the degree of alignment between both clouds, that is, the one that minimizes the usual quadratic energy function

$$E(\mathbf{B}) = \sum_i \sum_j \mathbf{B}_{ij} ||\mathbf{M}_{t-1}^j - (\mathbf{R}_t \mathbf{M}_t^i + \mathbf{t}_t)||^2, \quad (4)$$

being $\mathbf{B}_{ij}$ binary matching variables (1 when $\mathbf{M}_t^i$ matches $\mathbf{M}_{t-1}^j$ and 0 otherwise). In order to minimize the latter function we perform a conjugate gradient descent with an adaptive step through the space of incremental actions.

In the experiment showed in Fig. 2, there are 423 initial matches that are reduced to 156 after 156 iterations of the leaving-the-worst out process. Then, the obtained action is: $\delta x_t = -0.108$ meters, $\delta z_t = 0.757$ meters and $\delta\theta_t = 5.85$ degrees.

B. Map Building and Rectification

Let $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{N-1}$, be a robot trajectory, i.e. a time-indexed sequence of poses, of size N, and $\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_{N-1}$ be the estimated actions sequence of length N, where $\mathbf{a}_0$ a synthetic action that brings the robot from the origin of the global reference system to $\mathbf{p}_0$, that is, $\mathbf{a}_0 = \mathbf{p}_0$. Then, an initial approximation of the 3D map comes from superposing all the point clouds with respect to the referential pose $\mathbf{p}_0$. Recursively we compute,

$$\mathbf{A}_{N-2} = \mathbf{C}_{N-2} \cup \{\mathbf{R}_{N-1}\mathbf{M}_{N-1}^i + \mathbf{t}_{N-1} \,|\mathbf{M}_{N-1}^i \in \mathbf{C}_{N-1}\} \quad (5)$$

and follow the sequence $\mathbf{A}_{N-3}, \mathbf{A}_{N-4}, \ldots, \mathbf{A}_0$, until the last one becomes the current map, say $\mathbf{A}$,i.e. the aggregation of all observations in a common reference system. However, as this map accumulates the errors produced at local action estimations (errors due to the latter algorithm, or even to the absence of 3D cues when non-textured parts of the environment are observed), it is desirable to provide a consistency criterion and an updating strategy that exploits it in order to obtain a globally-consistent map.

1) *Consistency Criterion.* Let $\tilde{\mathbf{A}}$ the constrained version of the current map. We assume that maximizing the consistency of such a map is equivalent to minimizing the following energy function

$$E(\tilde{\mathbf{A}}) = \sum_t E_t, \quad (6)$$

that is, the sum of local energies corresponding to each constrained cloud. Each E_t is defined by the averaged sum of distances between each of the points $\mathbf{M}_t^i = [x_t^i, y_t^i, z_t^i]^T \in \tilde{\mathbf{C}}_t$ and their closest point in the rest of constrained point clouds which lie in the plane $Y = y_t^i$ (same height):

$$E_t = \frac{\sum_i ||\mathbf{M}_t^i - \mathbf{M}^{closest}||^2}{|\tilde{\mathbf{C}}_t|}, \quad (7)$$

where $\mathbf{M}^{closest}$ minimizes $||\mathbf{M}_t^i - \mathbf{M}_v^k||$ among all points $\mathbf{M}_v^k = [x_v^k, y_v^k, z_v^k]^T \in \tilde{\mathbf{C}}_v$, being $v \in \{0, 1, t-1, t+1, \ldots, N\}$ and $|y_t^i - y_v^k| \leq \delta$. Thus, we exploit the fact that there is no translation with respect to the Y axis.

2) *Quasi-random Update.* The underlying idea of our map-updating strategy is to modify all actions $\mathbf{a}_t$ simultaneously (including $\mathbf{a}_0$ because $\mathbf{p}_0$ could share information with $\mathbf{p}_{N-1}$, for instance, when the trajectory is cyclic) in order to obtain a new constrained map $\tilde{\mathbf{C}}^{new}$, but the amplitude of such a modification must be proportional to the contribution of $\tilde{\mathbf{C}}_t$ to the global energy. Such an amplitude is defined by the relative energy

$$e_t = \frac{E_t}{\sum_v E_v}. \quad (8)$$

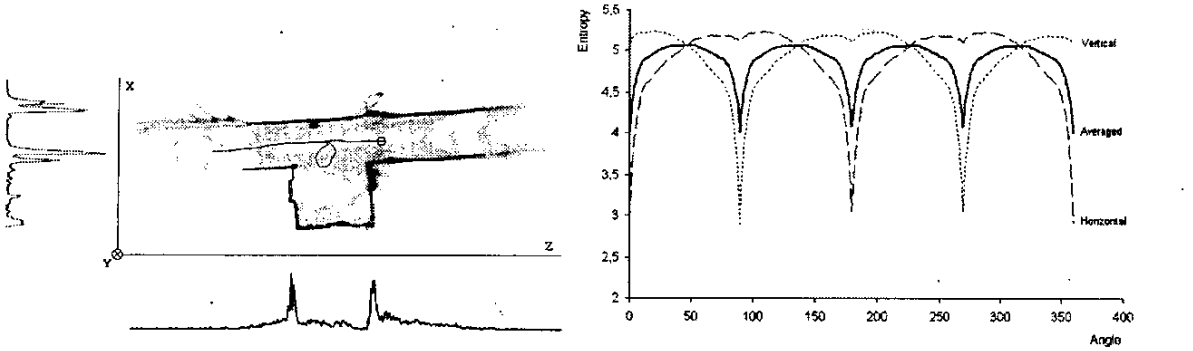

Fig. 3. Alignment term. A fragment of a map of $N = 36$ views with distributions $\mathbf{q}_X$ and $\mathbf{q}_Z$, being $H(\mathbf{q}_Z) = 5.1475$ and $H(\mathbf{q}_X) = 3.7985$ (left). Individual and global entropies as the map is rotated around Y (right).

Therefore, the new action $\mathbf{a}_t^{new}$ is a random variable following the Gaussian distribution $N(\mathbf{a}_t, e_t\mathbf{S}_t)$, where $\mathbf{S}_t$ is diagonal 3×3 matrix defined by the variances $\sigma_{\delta x}$, $\sigma_{\delta z}$ and $\sigma_{\delta\theta}$, that must be carefully specified, that is, a proper scale must be chosen.

Considering simultaneously $\mathbf{a}_0^{new}, \mathbf{a}_1^{new}, \ldots, \mathbf{a}_{N-1}^{new}$ implies considering new poses $\mathbf{p}_0^{new}, \mathbf{p}_1^{new}, \ldots, \mathbf{p}_{N-1}^{new}$, and consequently a new constrained map $\tilde{\mathbf{A}}^{new}$. Such a map is accepted if it yields an energy decrease, that is, if $E(\tilde{\mathbf{A}}^{new}) \leq E(\tilde{\mathbf{A}})$. Otherwise, none of the new actions are applied, we retain $\tilde{\mathbf{A}}$, and a new iteration of the algorithm begins. We finalize when the stabilization of the global energy is detected.

In practice, we only modify simultaneously $K < N$ actions, being these actions selected randomly. This is done for two reasons: (i) the complexity of the optimization problem is obviously reduced; and (ii) constrained clouds corresponding to extreme observations ($\tilde{\mathbf{C}}_0$ and $\tilde{\mathbf{C}}_{N-1}$) provide, in the general case where the robot trajectory is not necessarily cyclic, a higher contribution to the global energy (they share few points with the rest of the clouds). This leads to higher distortion amplitudes both for $\mathbf{a}_0^{new}$ and $\mathbf{a}_{N-1}^{new}$. Thus, a $K-$th order selection does not necessarily consider these actions.

3) *Alignment Term*: In plane-parallel environments, that is, in environments where the main planes (walls, doors, floor, ceiling, and so on) are either parallel or orthogonal, the global consistency criterion in Eq. 6, may be completed by an alignment term that exploits such an assumption. This allows, for instance, to correct a typical straight corridor that appears slightly curved if we only minimize the sum of local energies. However, as we do not have access to a set of planes, but to a huge point cloud, we analyze the distribution of the projections of such a cloud on both the X and Z axes (the XZ plane contains a zenithal view the map). Let $q(x)$, with $x \in X$, the sum of all points $\mathbf{M}^i = [x^i, y^i, z^i]$ in the constrained map $\tilde{\mathbf{A}}$ with $x = x^i$, and $q(z)$, with $z \in Z$ is defined in a similar way. Let $\mathbf{q}_X$ the probability distribution associated to normalizing $q(x)$ after performing a proper discretization of the X axis, that is, $\mathbf{q}_X(x) = q(x)/\sum_x q(x)$, and $\mathbf{q}_Z$ is defined similarly. Then, the alignment term relies on the scaled sum of the entropies associated to the latter distributions:

$$E^{align} = -\mu(H(\mathbf{q}_X) + H(\mathbf{q}_Z)), \tag{9}$$

being $\mu \geq 0$ the scaling factor. Intuitively, if a corridor is perfectly orthogonal to the X axis, we will observate two well-defined peaks in the projection and $H(\mathbf{q}_X)$ will be maximal with respect to any other rotation, whereas $H(\mathbf{q}_Z)$ will be minimal (see Fig. 3). Furthermore, the sum $H(\mathbf{q}_X) + H(\mathbf{q}_Z)$ is maximal when either the corridor is orthogonal to X or to Z. Therefore, E^{align} prefers maps yielding peaked projections and this is why curved corridors may be rectified by using the complete global-consistency criterion

$$E(\tilde{\mathbf{A}}) = \sum_t E_t + E^{align}, \tag{10}$$

as we can see in Fig. 4. In the experiment showed in such a figure, we have taken a trajectory of $N = 200$ views. From top to bottom, we show partial results corresponding to: the initial map ($E = 0.197024$), 20-th iteration ($E = 0.193770$) and 100-th iteration ($E = 0.185515$). In order to represent the final map we take the complete point clouds (the same for the rest of the experiments).

III. Experimental Results

For our mapping experiments we use a RWI Magellan Pro robot equipped with a trinocular Digiclops stereo system, a laptop with a Pentium IV/4.5GHz processor, and a wireless ethernet system for tele-operation (see Fig. 5). We obtain 320×240 stereo images producing clouds of 10.000 points on average which are typically reduced to 500 points in the constrained clouds. Given our previous experimental evaluations of the 3D estimation error for the Digiclops system, our maximum range is 8 meters, being the averaged error associated to such a distance below 0.75 meters.

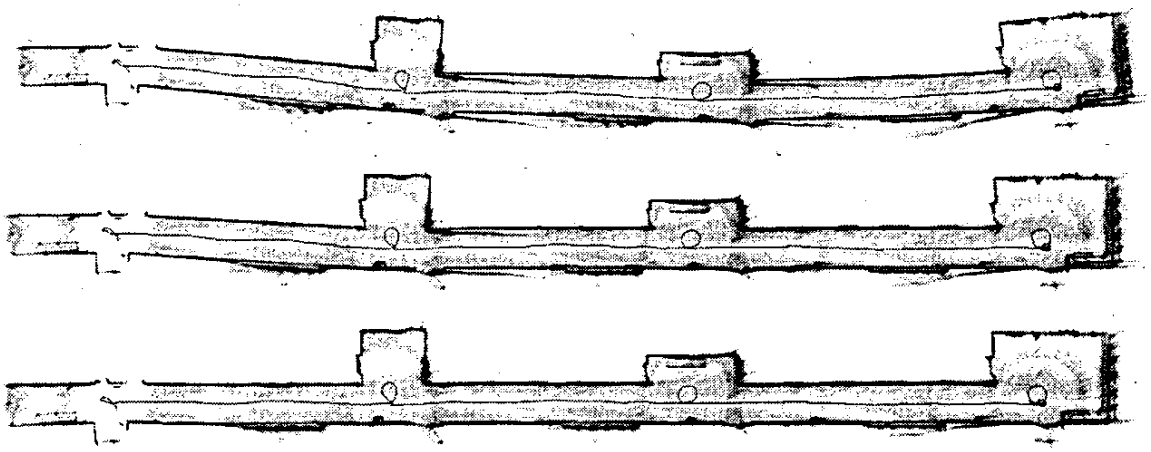

Fig. 4. Example of map rectification using the complete global-consistency criterion. From top to bottom: several iterations (0, 20, and 100).

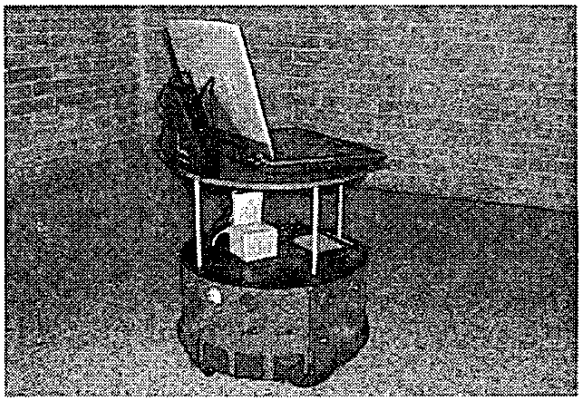

Fig. 5. Our mobile robot equipped with the stereo system.

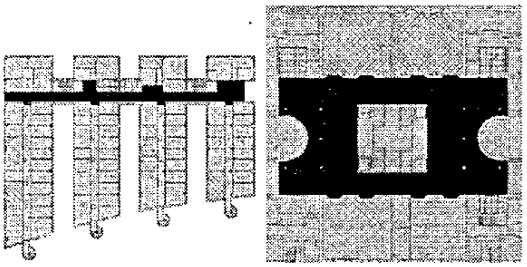

Fig. 6. Planes of the two indoor environments used in the experiments. Polytechnical corridor (left) and Economics Hall (right).

The parameters for action estimation are: $|\mathbf{W}| = 7 \times 7$ (the size of the appearance windows), $\delta = 0.05$ meters (range for considering that two points have similar height), $S_{min} = 0.8$ (minimal strength of a match), $R_{min} = 0.95$ (minimal distinctiveness ratio), $\sigma_{min} = 0.005$ and $|\mathcal{M}|_{\text{min}} = 10$ (variance limit and minimal number of matches for stopping the leaving-the-worst-out process).

The parameters for map building and rectification are: $\sigma_{\delta x} = \sigma_{\delta z} = 0.016$ meters, $\sigma_{\delta\theta} = 2.86$ degrees. The fraction K/N of simultaneously selected actions in the quasi-random update process is between 0.1 and 0.15 (then for typically registrations with N from $100 - 300$ views, K is in the range $15 - 45$). Regarding the regularization term E^{align}, the constant μ for weighting the sum of entropies is set to 0.001. We have found that such a setting of μ tends to slightly enforce global alignment in the early stages of the minimization process.

Given the latter settings and specifications, we have teleoperated the robot into two indoor environments (see Fig. 6). The Polytechnical corridor has a length of 56 meters and it has been recovered from $N = 200$ views (see Fig. 4).The averaged time for action estimation was 323 ms. The averaged time per rectification iteration was 2.54 secs. During the first five iterations, such a time is very high ($15-20$ secs) and after this point, it falls to $2-3$ secs. Such a behavior is very similar to the one observed in the mapping of the Economics Hall (see Fig. 6). In this latter case, the robot performed a double-cycle trajectory of $N = 308$ views. The averaged time for action estimation was 343 ms, and the averaged rectification time was 2.56 secs. Again the individual time per iteration falls from $20 - 30$ secs to $2 - 3$ secs.

IV. Conclusions and Future Work

We have proposed an approach to find globally consistent 3D maps using stereo vision, and we have successfully tested it in two indoor environments. However, there many aspects to solve, and our current and future research includes: (i) extending the approach for more unconstrained egomotion and environment conditions, (ii) investigate an on-line strategy for solving the SLAM problem with stereo vision, and (iii) extract more compact maps (we have performed some initial experiments for plane detection in [14] and the results were promising).

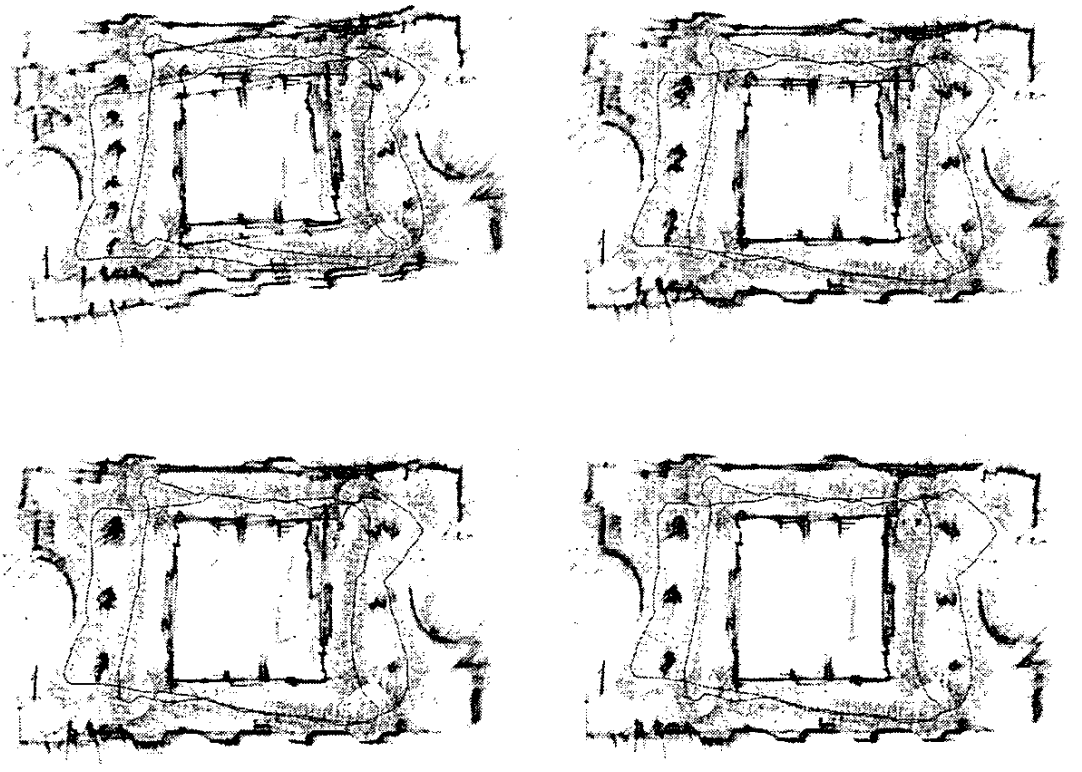

Fig. 7. Four iterations of the rectification process in the Economics Hall. From top to bottom and left to right, iterations: 0, 5, 10, and 600.

ACKNOWLEDGMENT

This research is partially funded by the project TIC2002-02792 of the Spanish Government.

REFERENCES

[1] P. Allen, I. Stamos, A. Gueorguiev, E. Gold, and P. Blaer. *AVENUE: Automated site modeling in urban environments*. In Proceedings of 3DIM2001: International Conference on 3D Digital Imaging and Modeling (2001).

[2] P. J. Besl and N. D. McKay. *A Method for Registration of 3D Shapes*. IEEE Transactions on Pattern Analysis and Machine Intellitence 14(2) (1992) 239-256.

[3] T. Duckett. *A genetic algorithm for simultaneous localization and mapping*. In Proceedings of ICRA'03: IEEE International Conference on Robotics and Automation (2003).

[4] A. Gueorguiev, P. Allen, E. Gold, and P. Blaer. *Design, architecture, and control of a mobile site modeling robot*. In Proceedings of ICRA'00: IEEE International Conference on Robotics and Automation (2000).

[5] D. Hähnel, W. Burgard, and S. Thrun. *Learning compact 3D models of indoor and outdoor environments with a mobile robot*. Robotics and Autonomous Systems, 44 (2003) 15-27.

[6] S.F. El-Hakim, P. Boulanger, F. Blais, and J.-A. Beraldin. *Sensor-based creation of indoor virtual environment models*. In Proceedings of VSMM'97: IEEE Confernce on Virtual Systems and Multimedia (1997).

[7] M.B.L. Iocchi and K. Konolige. *Visually realistic mapping of a planar environment with stereo*. In Proceedings of ISER'00: International Conference on Experimental Robotics (2000).

[8] Y. Liu, R. Emery, D. Chakrabarti, W. Bungard, and S. Thrun. *Using EM tlo learn 3D models with mobile robots*. In Proceedings of ICML'01: International Conference on Machine Learning (2001).

[9] J. Lobo, C. Queiroz, and J. Dias. *World feature detection using stereo vision and inertial sensors*. Robotics and Autonomous Systems 44 (2003) 69-81.

[10] J. Luck, C. Little, and W. Hoff. *Registration of range data using a hybrid simulated annealing amd Iterative Closest Point*. In Proceediings of ICRA'00: IEEE International Conference on Robotics and Automation (2000).

[11] C. Martin and S. Thrun. *Real-time acquisition of compact volumetric maps with mobile robots*. In Proceedings of ICRA'02: IEEE International Conference on Robotics and Automation (2002).

[12] H. Moravec.*Robot spatial perception by stereoscopic vision and 3D grids*. Technical Report, The Robotic Institute, CMU (1996).

[13] D. Murray and J. Little. *Using real-time stereo vision for mobile robot navigation*. Autonomous Robots, Vol. 8, N. 2 (2000) 161-171.

[14] J.M. Sáez and F. Escolano. *Compact mapping in plane-parallel environments using stereo vision*. In Proceedings of CIARP'03: Iberoamerican Congress on Pattern Recognition. LNCS, Vol. 2905 (in press) (2003).

[15] S. Se, D. Lowe, J. Little. *Vision-based mobile robot localization and mapping using scale-invariant features*. In Proceedings of ICRA'01: IEEE International Conference on Robotics and Automation (2001).

[16] S. Se, D. Lowe, J. Little. *Mobile robot localization and mapping with uncertainty using scale-invariant landmarks*. International Journal of Robotics Research, Vol. 21, No. 8 (2002) 735-758.

[17] S. Se, D. Lowe, J. Little. *Vision-based mapping with backward correction*. In Proceedings of IROS'02: IEEE/RSJ International Conference on Intelligent Robots and Systems (2002).

[18] S. Thrun, W. Burgard, D. Fox. *A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping*. In Proceedigns of ICRA'00: IEEE International Conference on Robotics and Automation (2000).