# A Visual Landmark Framework for Mobile Robot Navigation

J.B.Hayet, F. Lerasle, M. Devy

*LAAS-CNRS, 7 avenue Colonel Roche, 31077 Toulouse Cedex (France)*

---

## 1 Introduction

Vision has become a major element in mobile robot navigation and many strategies relying on images have already been proposed, based on environment representation either by image databases[10] or by visual landmarks. Classically, the latter are detected by the robot, mapped into the environment representation and recognized during the execution of a navigation task. In general, the robot's position estimate is computed mainly from the integration of outputs of odometers, which tends to accumulate small displacement errors and produces drift. When recognized, visual landmarks allow to make this drift vanish, so they play a key role in making navigation systems efficient.

The work presented here aims to be part of a navigation strategy relying on "natural" visual landmarks, i.e., salient objects a mobile robot detects/recognizes and from which it can either simply localize itself (if the map of the environment is known) or incrementally build a metric map integrating perceptual data and position estimation, according to the Simultaneous Localization And Mapping (SLAM) paradigm[17]. Our robot has several localization modalities, based either on laser segments learnt using a laser range finder and on visual landmarks detected from a single B&W camera: this paper is mainly devoted to the visual modality. The reader could refer to [1] for a description of these modalities.

Numerous techniques have been proposed to model landmarks for navigating in indoor environments. They all rely on two assumptions: (1) landmarks have to be easily detected in the image signal and (2) they can be locally characterized to distinguish them from others. In that scope, landmark-based navigation research has started by using remarkable characteristics of office-like environments (3D room corners, lights...) [3,11,9], or collections of simple edge segments[16]. Point sets can also serve as landmarks when combined to define projective invariants[2].

*Preprint submitted to Elsevier Science*

Most recent work make use of points to define landmarks[4,15], taking advantage of new, powerful interest point detection and characterization algorithms such as SIFT, which makes landmark-point recognition much easier[5].

In our work, planar quadrangular objects (posters, doors, cupboards...) are selected as landmarks, as they are one of the basic structures man-made environments are made of. Among research works similar to ours, we can quote [12], where the authors take advantage from genetic algorithms techniques to recognize 2D landmarks.

The paper is organized as follows. Section 2 details the landmarks detection process, with results on images acquired during robot navigation. Section 3 presents the landmarks recognition process; an evaluation of our recognition method, with respect to different acquisition criteria, proves its robustness. Navigation experiments are presented in section 4. Finally, section 5 sums up our approach and opens a discussion for our future work.

## 2 Landmarks detection

The landmark extraction is focused on planar, mostly quadrangular objects, e.g., doors, windows, posters, cupboards.... A natural way of extracting quadrilaterals relies upon perceptual grouping on edge segments.

### *2.1 Overview of the method*

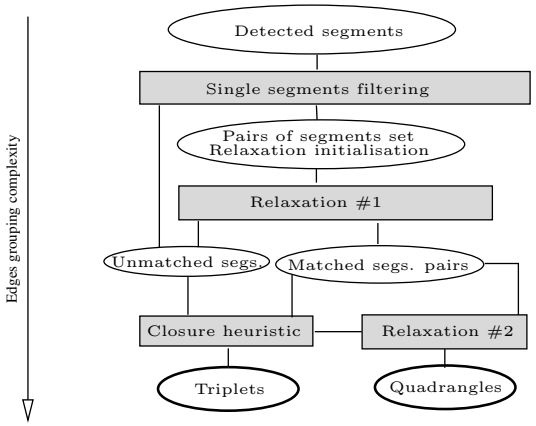


Fig. 1. The landmark detection scheme. Fig. 2. Segments in a typical indoor scene.

Let a set of $n_L$ edge segments set be $\mathscr{L} = \{l_i\}_{1 \leq i \leq n_L}$. A naive approach to test all possible 4-uples inside $\mathscr{L}$ does not make sense, as illustrated by Fig.2.

To reduce the problem complexity, we propose a two-step algorithm: first, mapping $\mathscr{L}$ to $\mathscr{L} \cup \{\emptyset\}$ so that each segment is matched with at most one segment; second, associating pairs of matched segments to form quadrangles. The whole process is described on Fig. 1.

**Extracting edge segments.** The output of a Canny-Deriche edge detector is first thinned and chained. The resulting edge chains are then recursively segmented to produce the set $\mathscr{L}$ of line segments as illustrated by Fig. 2. Before the matching process starts, small segments are filtered, altogether with segments that may correspond to repetitive patterns. Typically, segments corresponding to the floor tiling (as in the central image of Fig. 5) are found by an accumulator technique and are eliminated.

**Generating segment matches.** An initial set of matches is generated by looking for couples $(l_k, l_l)_{k \neq l}$ for which a similarity measure $s_{kl}$ is above a given level. Indices $k$ and $l$ are associated to individual segments. This measure combines several cues, as explained hereafter, so that segments corresponding to opposite sides of quadrangles have high values of $s_{kl}$.

Moreover, a set of geometric constraints on segment pairs denoted by $Q^1_{klmn}$ is used in a first relaxation scheme to validate pairs belonging to quadrangles, i.e., to generate a set of coherent potential landmarks. Again, indices $k, l, m, n$ represent individual segments.

**Generating potential quadrangular landmarks.** With constraints on pairs of detected quadrangles, a second relaxation process selects only the more consistent four-segment sets corresponding to landmarks; these constraints are denoted by $Q^2_{klmn}$. Three-segment sets are useful as they may correspond to occluded landmarks or doors, so a simple heuristic is used to combine two-segment sets rejected from the second relaxation process with single segments rejected from the first one by using constraints $T^2_{klm}$. All these constraints, specified in section 2.3 are applied through a relaxation scheme depicted hereafter.

### *2.2 Relaxation scheme*

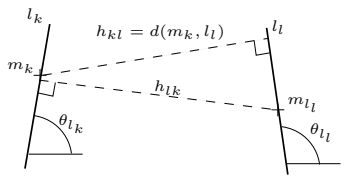


Fig. 3. Conventions for segment matching.

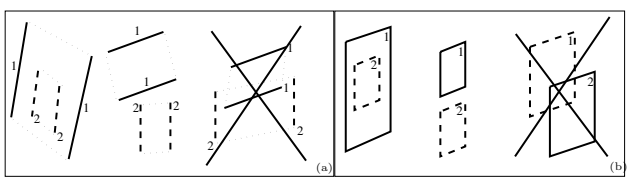


Fig. 4. Examples of accepted configurations for two segments pairs or two quadrangles

Given two sets $S_1$ ($n_1$ elements) and $S_2$ ($n_2$ elements), the principle of relax-

ation is to iteratively make all the probabilities $p_{kl}$ of associations between items $k$ (index for an element of $S_1$) and $l$ (index for an element of $S_2$) evolve towards 1 or 0, i.e., towards unambiguous match or mismatch. Let $A$ be the $n_1 \times n_2$ matrix such as $A_{kl} = p_{kl}$.

We define the variety $\mathscr{A} = \{A \in M_{n_1 \times n_2} \mid \forall (k,l)\ A_{kl} \geq 0 \text{ and } \forall k\ \sum_l A_{kl} = 1\}$.

The relaxation steps maximize iteratively a global consistency score using gradient ascent in $\mathscr{A}$. In our specific case, for each relaxation process $i \in \{1, 2\}$, we maximize a score $G^i(A)$ :

$$G^i(A) = \sum_{klmn} Q^i_{klmn} A_{kl} A_{mn}.$$

The terms $Q^1_{klmn}$ (resp. $Q^2_{klmn}$) represent a compatibility degree between pairs of segment pairs (resp. quadrangles) $(k,l)$ and $(m,n)$. It is derived from constraints detailed in section 2.3.

The gradient step $\alpha^{(p)}$ at iteration $p$ is adaptive and defined by $\alpha^{(p)} = \arg\min_{\alpha} G^i(A^{(p)} - \alpha \nabla G^{i,(p)})$. Regarding initialization, a priori probabilities are computed from similarity measures $s_{kl}$ only. If the measure $s_{kl}$ is below a threshold $s_{min}$, $p^{(0)}_{kl}$ is set to 0, otherwise it is estimated by:

$$p^{(0)}_{kl} = \frac{s_{kl}}{\sum\limits_{s_{kn} > s_{min}} s_{kn}}. \tag{1}$$

The next section describes the different criteria and constraints we use in the relaxation schemes.

### *2.3 Comparing sets of segments*

In this section, we make the way we use sets of segments more explicit. We first describe the similarity measure $s_{kl}$ between two segments $l_k$ and $l_l$ used to initialize probabilities $p_{kl}$. Then, we give details on the constraints $Q^1_{klmn}$ between pairs of segments, and $Q^2_{klmn}$ between quadrangles which are used in the two relaxation schemes.

#### *2.3.1 Segment similarity*

The measure $s_{kl}$ is defined by a weighted sum of the following geometric and luminance cues:

- segments length ratio $\frac{1}{2}(\frac{|l_l|}{|l_k|} + \frac{|l_k|}{|l_l|})$ in Fig. 3,
- angular difference $|\theta_{l_k} - \theta_{l_l}|$ in Fig. 3,
- a shape criteria giving favour to square-like shapes $\frac{1}{2}(\frac{|l_l|+|l_k|}{h_{kl}+h_{lk}} + \frac{h_{kl}+h_{lk}}{|l_l|+|l_k|})$ where $h_{kl}$ represents the distance defined in Fig 3,
- the overlapping rate between $l_k$ and $l_l$,
- presence of a third segment in the neighbourhood that forms a convex three-segments set with the given pair. Segments pairs $(l_k, l_l)$ without at least a third segment $l_m$ are discarded for the next.

As far as luminance criteria are concerned, an average grey-level profile is computed in the direction orthogonal to each segment, so that an association $(l_k, l_l)$ is characterized by the Zero Normalized Cross Correlation (ZNCC) score between the two segments profiles. In fact we assume here that the intensity in the background is uniform around a trustworthy quadrangle while its two opposite insides are supposed to include quite similar texture.

#### *2.3.2 Second degree constraints*

Here, uniqueness and convexity of potential matches among segments pairs are checked. Uniqueness constraint allows to reduce the relaxation algorithm complexity and enforces the assumption that landmarks are supposed to be locally unique. Convexity rule says that two segments pairs, correspond to opposite sides of two trustworthy quadrangles which must verify rules of full inclusion or no intersection as shown in the left part of Fig. 4.

From the constraints $Q^1_{klmn}$ described above, the first relaxation outputs a set of segments pairs. The next step is to match two segment pairs delimiting trustworthy quadrangles. Indexes $k, l, m$ and $n$ refer now to segments *pairs*.

#### *2.3.3 Third and fourth degree constraints*

The fourth degree constraints $Q^2_{klmn}$ ensue from accepted configurations for two quadrangles which are shown in the right part of Fig.4 and are applied throughout the second relaxation scheme.

From the previous steps, it is possible to extract 3-segment sets that can be helpful in robot navigation. These sets involve an unmatched segment pair $(k, l)$ coming from relaxation #2 and an unmatched segment $m$ coming from relaxation #1. The selection of these potential landmarks is based on uniqueness, on the resulting shape convexity and on vicinity relationships (constraints $T_{klm}$).

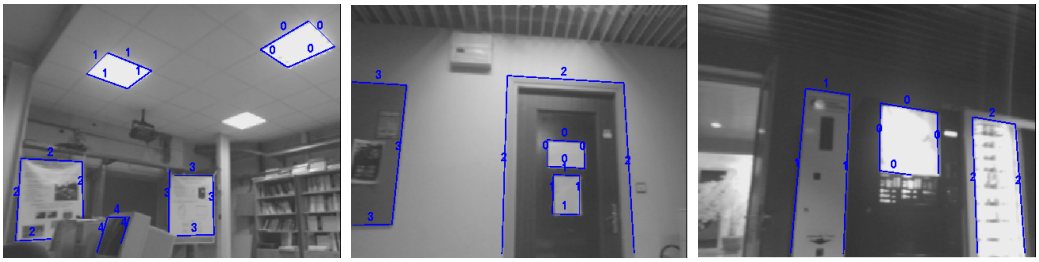

Fig. 5. Examples of landmarks detection: the numbers on the segments indicate the final tag associated to the detected landmark.

### *2.4 Detection results*

Experiments have been performed on a large database of about 300 images acquired from our robot navigating either in a corridor network or in cluttered open areas. The robot is a Nomadic XR4000, equipped with a SICK laser range finder and a CCD camera mounted on a pan-tilt platform.

Figure 5 shows examples of landmarks detected in an open cluttered environment. We note that both quadrangles and three-segment sets are extracted.

During the environment exploration, the robot executes two operations: (1) a SICK laser map is built by a classical SLAM procedure, and (2) visual landmarks are detected and combined with the laser segments. The resulted map is represented in Fig.6, with all laser segments and all detected landmarks: windows or posters in green (lateral walls or ceiling), doors by a grey icon. Their associated locations on the walls are triangulated from their perspective views and the planes defined by the laser segments assuming the multisensory system is fully calibrated. For every detected landmark, a visibility map (not shown here) is statically computed according to the environment model by an analytical method.

Detection rates are computed over the database of images taken by the robot. In this database, all quadrangular objects have been identified by a human operator; the landmarks detection module extracts 88% of existing landmarks, without any false detection.

During this environment exploration step, the robot could stop to perform both detection and recognition processes, so that only the representation of new discovered landmarks is learnt. Only quadrangular objects which are successfully detected from different view points (section 3.3) are considered as landmarks in the environment model. Later, when the robot navigates using the set of learnt landmarks, it must be able to achieve these tasks dynamically.

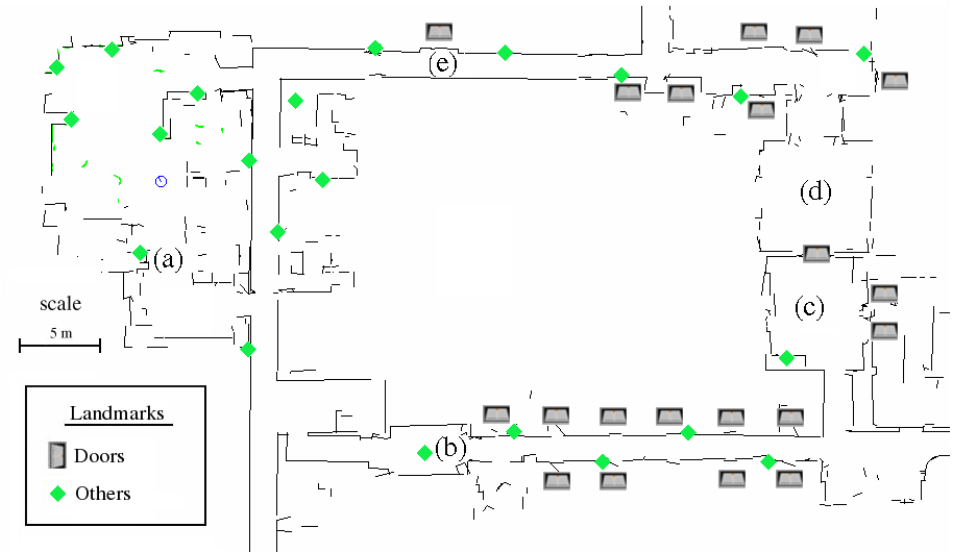


Fig. 6. Landmarks detection in office environment

## 3 Landmarks recognition

Once a landmark has been detected, an appearance model is built so that it can be recognized from different viewpoints. In section 3.1, we describe the landmark representation: boundaries of a detected landmark allow to rectify the observed pattern; and such a mapping provides an *invariant* representation under scale and perspective changes. We call it "icon".

In section 3.2, we propose distances to compare icons and perform recognition. Section 3.3 thereafter describes the landmark model. Based on this model, a confidence factor on the recognition process is proposed in section 3.4. In section 3.5, a correlation-based method is compared with an approach based on interest points extracted from icons.

### *3.1 Landmarks iconification*

Let us consider, (1) an extracted quadrangular landmark $Q$ from an image $I$, and (2) a fixed-size reference square $S$. The two shapes are related by a homography $H_{SQ}$ that maps points from $S$ to $Q$.

By using $H_{SQ}$, a new small-sized image $I'$ is built from the image $I$ by averaging pixels from $I$ into pixels in $I'$ (see Fig. 7). The computation of $H_{SQ}$ is straightforward as four point correspondences are available [14].

Averaging is performed in order to avoid too much information compression in the low-scale front view $I'$: the grey level value of a pixel $(a, b)$ in image $I'$ is

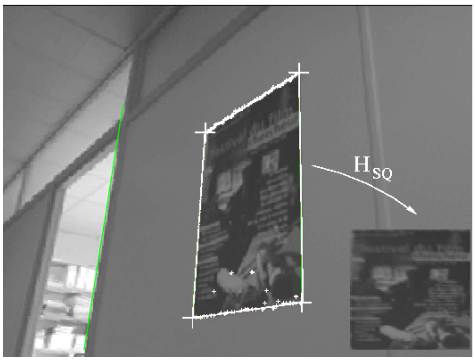


Fig. 7. Model construction: quadrilaterals are transformed into icons by the mean of $H_{SQ}$.

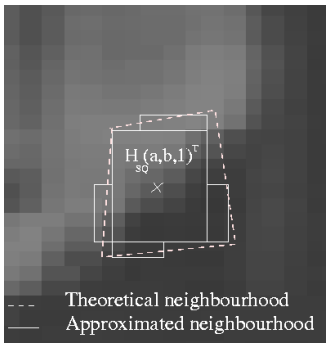


Fig. 8. Approximated averaging for iconification: zoom of the neighbourhood of the image of an icon pixel $(a, b)$.

determined by taking into account all pixels in image $I$ belonging to a certain neighbourhood of $H_{SQ}(a, b, 1)^T$, i.e., its image in $I$. This neighbourhood is computed by approximating the image of a pixel square with simple heuristics (see Fig. 8).

The icon $I'$ is processed by the Harris operator to get a set of $n$ interest points $\{X_i\}_{1\leq i\leq n}$ and a local descriptor[13] in $\mathbb{R}^7$, based on Gaussian derivatives, is associated to every interest point.

### *3.2 Metric on icons*

To perform the recognition between a set of learnt landmarks noted $\{C_l\}_{1\leq l\leq N}$ and a detected landmark $\mathcal{Q}$, metrics on icons are defined.

**A correlation-based distance.** The centered and normalized correlation score $\mathscr{C}$ provides a distance which is theoretically invariant to overall light changes.

To be less sensitive to local variations or occlusions, the icons $\mathcal{Q}$ (from the new landmark) and $C_l$ (from the reference landmark $l$) are divided into $5 \times 5$ buckets. Then, we define a robust correlation score $\mathscr{C}^r$ between two icons by using separated correlations $\mathscr{C}_{ij}(\mathcal{Q}, C_l)$ between buckets $i$ and $j$, and by choosing the $k^{th}$ greatest correlation score between buckets. The number $k$ is expressed as a ratio $r$ of admissible outliers among all the buckets. It allows to ignore the most important local differences between $\mathcal{Q}$ and $C_l$. From this new score, we derive the distance:

$$\mathscr{C}^r(\mathcal{Q}, C_l) = 1 - k^{th}_{1\leq i,j\leq 5}\mathscr{C}_{ij}(\mathcal{Q}, C_l)$$

**A local features-based distance.** Many popular appearance-based methods for object recognition are based on interest points matched thanks to their local descriptors[13,15]; these local features are remarkably stable under moderate rotation or light changes. We propose to use the partial Hausdorff distance[8] to compare sets of interest points $\{X_i\}_{1\leq i\leq n}$ extracted from icons.

Let be two sets of points $S_l = \{X_i^l\}_{1\leq i\leq n_l}$ associated with a known landmark $C_l$, and $S = \{X_j\}_{1\leq j\leq n}$, extracted from a new landmark $\mathcal{Q}$. To handle outliers, the Hausdorff distance between $S_l$ and $S$ is modified in the same way as $\mathscr{C}^r$, i.e., by considering only a fraction $r$ of all the points, $k = r\min(n_l, n)$:

$$\begin{cases} d_h^r(S_l, S) = \max(h^r(S_l, S), h^r(S, S_l)) \\ h^r(S_l, S) = \ k^{th}_{1\leq i\leq n_l} \min\limits_{1\leq j\leq n} d(X_i^l, X_j) \end{cases}$$

A threshold $\tau_l$ on the distance is set to recognize landmarks $\mathcal{Q}$ as instances of known landmarks $C_l$, as it will be described in section 3.3.2. An interpretation of this distance is that an object is recognized provided that for at least $k$ points of the second set, a similar point can be found in the first set, and reciprocally.

The partial Hausdorff distance between two sets of points depends on the *local distance* $d$ between points. We could simply use the Euclidean distance, but we would lose explicit local photogrammetric information. In order to take into account both spatial and photogrammetric similarities between points, we define a local distance noted $d_p$:

$$d(a, b) = d_\nu(a, b)\|a - b\|,$$

where $d_\nu(a, b)$ is the Mahalanobis distance between the descriptor vectors at points $a$ and $b$. The Hausdorff distance based on $d$ is denoted by $\mathscr{H}^r$.

### *3.3 Building appearance models*

For each landmark $C_l$, a model is built from a set of $N_l$ representative images $I_i$ at several viewpoints (typically $N_l = 50$), from which iconified views $I_i'$ are extracted.

### *3.3.1 Reducing landmark representation*

A Principal Component Analysis is first performed on the set of raw icons. We keep only three icons, denoted respectively by $\mathcal{Q}_l^1, \mathcal{Q}_l^2, \mathcal{Q}_l^3$. The first one $\mathcal{Q}_l^1$ corresponds to the mean icon of $I_i', 1 \leq i \leq N_l$, whereas $\mathcal{Q}_l^2$ and $\mathcal{Q}_l^3$ correspond to the more significant modes on this icon set.

For distance $\mathscr{H}^r$, such a process is followed by the extraction of Harris points and their characteristics in the $I_i'$ icons closest to the selected eigenvectors.

### *3.3.2 Determining recognition thresholds*

During the recognition step, a detected landmark is compared to each known landmark $C_l$, using a recognition threshold $\tau_l$ specific to it. During the modelling step, an optimal threshold is computed for each landmark $C_l$ by computing distances ($\mathscr{C}^r$ or $\mathscr{H}^r$) between extracted icons for this landmark, with either the $C_l$ model or all the other models noted $\neg C_l$.

The distance distributions on representative sets of icons from $C_l$ and $\neg C_l$ give us a good approximation of the probability densities on the distances, given the knowledge of $C_l$ or $\neg C_l$. To specify an optimal threshold $\tau_l$, we minimize:

$$S(\tau_l) = \lambda \underbrace{\int_0^{\tau_l} p(d|\neg C_l)\mathrm{d}d}_{\neg S_l(\tau_l)} + \mu \underbrace{\int_{\tau_l}^{+\infty} p(d|C_l)\mathrm{d}d}_{S_l(\tau_l)},$$

with $\lambda$ and $\mu$ being two weights for respectively false positive and false negative, noted $\neg S_l(\tau_l)$ and $S_l(\tau_l)$. The choice $\mu = \frac{1}{6}\lambda$ allows to give more importance to false positives than to false negatives. The security in the robot navigation being critical, the recognition of a landmark in a bad position cannot be accepted, i.e., false positive are more important to be avoided.

### *3.3.3 Validation gates*

For every landmark $C_l$, the modelling step ends with a verification of two criteria: (1) $C_l$ must be salient enough, and (2) the $N_l$ images from which the $C_l$ appearance model has been generated, must give a good approximation of all possible viewpoints on $C_l$.

The **saliency** criterion is verified from the *covariance* of the icons $I'$, and from the number of stable extracted interest points. The **visibility** criterion indicates how far from each other are the extreme positions at which the landmark has been detected during this learning step. For all couples $(i, j) \in$

$[1, N_l]^2$, an inter-image homography $H^{ij}$ maps corresponding vertices of the landmark in images $I_i$ and $I_j$. Let us consider the normalized homography $\hat{H}^{ij}$, such as $\hat{H}^{ij}_{33} = 1$, and where image coordinates have been centered and normalized. Then, we define a visibility confidence as: $v_c = max_{ij}\|\hat{H}^{ij} - I_{33}\|$.

$I_{33}$ is the 3x3 identity matrix. The greater is $v_c$, the more extended is the area on which the landmark has been perceived during the learning step. The value $v_c$ is clearly correlated to the planarity: planar landmarks are recognized in a larger area and under greater camera parameters changes than non-planar ones.

### *3.4 Confidence in the recognition result*

The recognition task requires to index and compare detected landmarks. For a set of N modelled landmarks $\{C_l\}_{1\leq l\leq N}$ and a detected landmark $\mathcal{Q}$, let us note $\mathscr{D}_l = \mathscr{D}(\mathcal{Q}, C_l)$, the distance between $\mathcal{Q}$ and each class $C_l$ ($\mathscr{D}$ being either $\mathscr{C}^r$ or $\mathscr{H}^r$). The probability $P(C_l|\mathcal{Q})$ of labeling $\mathcal{Q}$ to $C_l$, is defined by:

$$\begin{cases} P(C_\emptyset|\mathcal{Q}) = 1 \text{ and } \forall l\ P(C_l|\mathcal{Q}) = 0 \text{ when } \forall l\ \mathscr{D}_l > \tau_l \\ P(C_m|\mathcal{Q}) = 1 \text{ and } \forall l \neq m\ P(C_l|\mathcal{Q}) = 0 \text{ when } \exists! m\ \mathscr{D}_m < \tau_l \\ P(C_\emptyset|\mathcal{Q}) = 0 \text{ and } \forall l\ P(C_l|\mathcal{Q}) = \frac{h(\tau_l - \mathscr{D}_l)}{\sum_p h(\tau_l - \mathscr{D}_p)} \text{ otherwise} \end{cases}$$

where $C_\emptyset$ refers to the empty class and $h$ the Heaviside function: $h(x) = 1$ if $x > 0$, 0 otherwise. This allows us to use the entropy-based measure:

$$m(\mathcal{Q}, \{C_l\}) = 1 + \frac{1}{N+1} \sum_j P(C_j|\mathcal{Q}) \log P(C_j|\mathcal{Q}).$$

### *3.5 Recognition evaluation*

An important issue for our recognition process, is the way the algorithm behaves with light effects, scale/perspective changes and bad segmentation from the detection step. Other questions are related to the discriminating power of proposed distances. To investigate this robustness problem, a large test image database has been constituted both by:

(1) 270 real images of different landmarks acquired while the robot wandered around the lab (see Fig. 9) represented by the map of Fig. 6.
(2) synthetic images of 300 movie posters with different light, scale/perspective conditions and occlusions, these modalities remaining quite difficult to perform and quantify in real conditions (see Fig. 10).

Fig. 9. Examples of real images with variable scales, occlusions, brightness variations or specular reflexions

Fig. 10. Examples of synthetic images with variable scales, occlusions, brightness variations or specular reflexions. Movie posters were used as the basic texture.

### *3.5.1 Discriminating power*

Let us consider probability densities computed from the distribution of distances between a given landmark and other ones from the database of real images. A poster found in this database has been selected and learnt as a landmark, and Fig. 11 now represents distributions of distance values obtained (a) for the objects corresponding which are instances of this landmark (class $C_l$) and (b) for objects that are not (class $\neg C_l$). This distribution can be approximated by a Gaussian function, which center and variance depends on the Hausdorff fraction and on sets cardinals.

The overlapping surface under the two curves are relatively small for the two distances that have been investigated. By following the process described in section 3.3.3, we have rates of false positive around 1%, whereas false negative where about 30%, which reflects the high level of disturbances we put on synthetic data sets.

### *3.5.2 Behavior under viewpoint changes*

The graphs in the left part of Fig. 12 represent the evolution of the ratio $\frac{distance}{threshold}$ for distances $\mathscr{H}^r$ and $\mathscr{C}^r$ under scale change. This ratio has to remain below 1 to ensure recognition. Even for a scale factor about 3, values for both of the compared distances remain small w.r.t. their respective thresholds. However, as expected, results are degrading fast as soon as the apparent size of the extracted pattern is below the size of the square used for the iconic representation.

As far as perspective distortions are concerned, the evolution of the ratio $\frac{distance}{threshold}$ have been studied for distances $\mathscr{C}^r$ and $\mathscr{H}^r$ by performing a planar

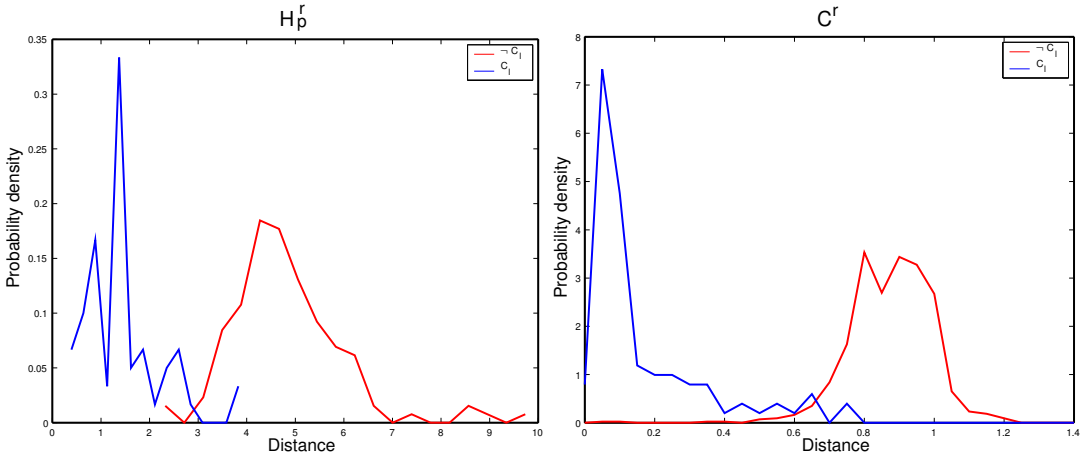


Fig. 11. Discriminating power: distribution of distances on classes $C_l$ and $\neg C_l$ for $\mathscr{H}^r$ (left) and $\mathscr{C}^r$ (right).

rotation in the horizontal plane of a landmark. Results on the right part of Fig. 12 show that the combination of invariants vectors and interest points is a powerful tool to achieve recognition of planar objects, as distances remain reliable up to $\pm 75°$ from the normal to the landmark plane, which is reasonable.

### *3.5.3 Behavior under light effects and occlusions*

The left graph in Fig. 13 shows that it is possible to have good recognition results for the two distances until local or global light saturations appear in the image.

Moreover, as it can be seen on the right part of Fig. 13, the representation is also robust to partial occlusions, which occurs for partially detected landmarks, that compose the majority of detected landmarks in indoor environment. With the distances $\mathscr{H}^r$ and $\mathscr{C}^r$, occlusions of the landmark up to 46% and 56% of its area do not prevent the landmark from being recognized.

## *3.6 Discussion: comparing the two metrics*

We have compared two different representations and associated metrics by applying tests w.r.t the main sources of image noise and variations. Both of the metrics have quite satisfactory results on ambiant brightness variations, scale or perspective changes, which makes our concept of quadrangles-landmarks a powerful tool for modelling environments. The $\mathscr{C}^r$ metric gives slightly better recognition results on all these tests, but it is limited by the size of data that have to be stored, i.e., all the icons have to be stored.

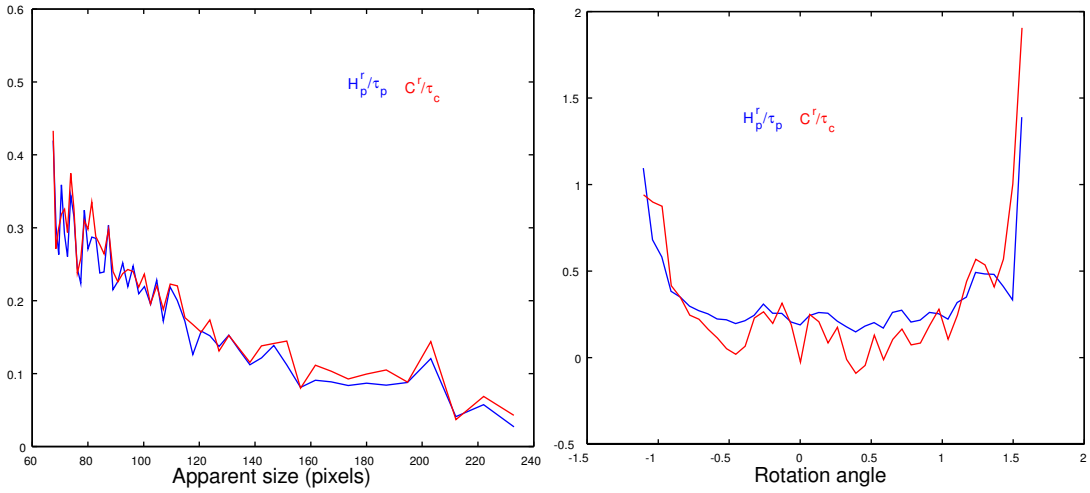


Fig. 12. Variations of the ratio $\frac{\text{distance}}{\text{threshold}}$ under scale changes (left) and rotations around the vertical axis (right) for distances $\mathscr{H}^r$ and $\mathscr{C}^r$.

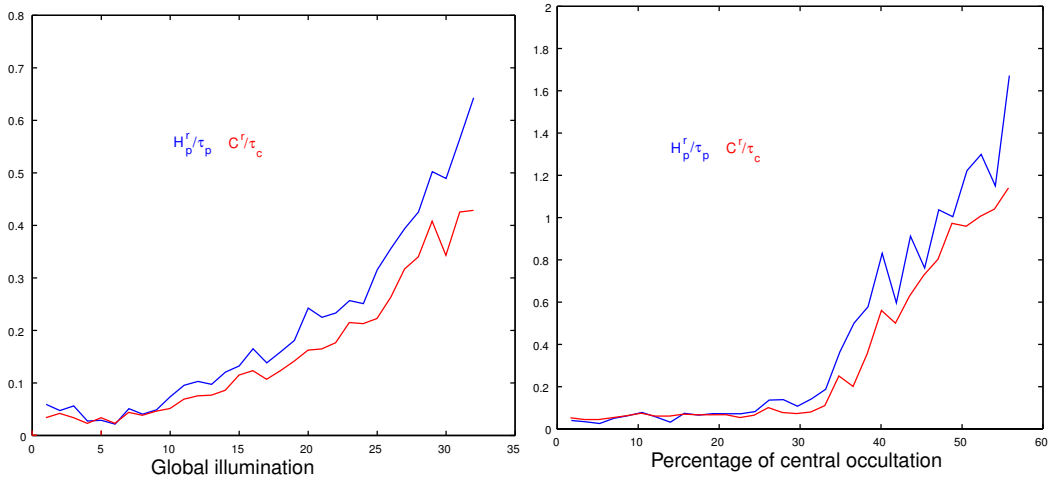


Fig. 13. Variations of the ratio $\frac{\text{distance}}{\text{threshold}}$ under light changes (left) and central occlusions (right) for distances $\mathscr{H}^r$ and $\mathscr{C}^r$.

That is why in practice $\mathscr{H}^r$ is preferred for our experimental work: this distance is compact and gives fairly good recognition results.

## 4 Application to robot navigation

Our landmark detection and recognition scheme has been integrated as a visual localization module in our Diligent Nomadic XR4000 robot, shown in Fig. 14. Subsection 4.1 describes the landmark localization with calibrated vision, and experiments showing our robot navigating in indoor environments are presented. Then, we introduce an extension we developed to handle unknown camera parameters.

*4.1 Localization with a calibrated camera*

Let us assume that our vision system is fully calibrated and that a 3D model of the quadrangle $\mathcal{Q}$ has been determined i.e. its four corners noted $\{P_i^n\}_{i=1..4}$ in the poster frame are *a priori* known. The landmark localization in the camera frame, i.e., the displacement $[R^{cn}, T^{cn}]$, is based on the decomposition of the homography $H^m$ relating four matches of image points $p_i$ and model points $P_i^n$. This matrix $H^m$ can be interpreted in terms of a displacement between the poster frame and the camera frame[14]:

$$[r_2^{cn}, r_3^{cn}, T^{cn}] = \lambda K^{-1}[h_1, h_2, h_3] \tag{2}$$

$[r_1^{cn}, r_2^c, r_3^{cn}]$ (resp. $[h_1, h_2, h_3]$) are columns of $R^{cn}$ (resp. $H^m$), $[t_x^{cn}, t_y^{cn}, t_z^{cn}]$ are the components of $T^{cn}$, $K$ the intrinsic parameters matrix and $\lambda$ the scale factor.

Fig. 14. The XR4000 "DILIGENT" robot we used in the experiments is equipped with a laser and BW cameras.

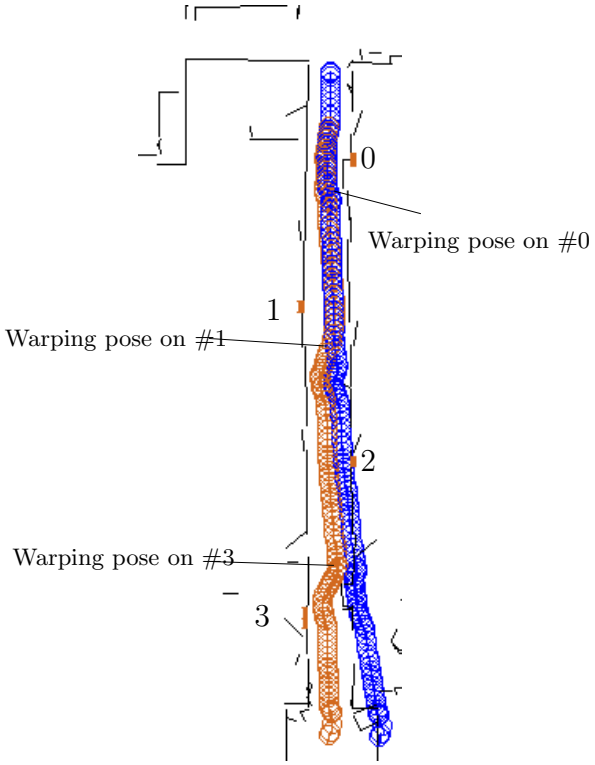


Fig. 15. Robot localization: recognizing known landmarks (marked 0, 1, 2, 3) allow to correct the robot's position.

Let us recall the robot is considered as a complete system, equipped not only by a camera, but also by a laser range finder and by odometry. A localization module is associated to every sensor: all computed positions are logically fused by a dedicated position manager module [1]. The localization strategy is based on a loose coupling of these modalities. During an off line statistical analysis,


Fig. 16. Robot localization: external view (bottom) and robot view (up).

the robot learns the better localization modality it must execute to locate itself in every area in the environment, according to their intrinsic performances and to local configurations of learnt landmarks or features. For example, the robot learns by itself, that: (1) in open space, it is relevant to fuse localizations computed by all modalities, even if they are computed at various frequencies, (2) in a given place, due to an uneven area on the ground, the odometry modality gives an important bias, (3) in a long corridor, vision modality is better than laser modality.

Fig. 15 and 16 illustrates navigation experiments in a 25 $m$ long corridor (annoted (b) in Fig 6) where laser localization is known to be inefficient as there is no identifiable beacon in the direction orthogonal to the corridor. In each left sub-figure, the blue trace corresponds to current odometry positions; without another modality, the robot would clearly bump against the left wall of the corridor. The red trace gives the current corrected position from the vision method, executed on four previously learnt posters annotated #0 to #3 (red color) on the laser map. Sub-figures show the robot respectively at corridor entry, at two positions close to posters and finally at corridor exit. Each upper right image shows the current robot perception while the bottom right image shows the robot in its environment. The robot perception is ensured by the camera mounted on a pan and tilt platform; in every place, the camera is pointed towards the best landmark, selected with respect to its visibility area and saliency coefficients estimated during the learning step. The number of positions corrections performed since the robot enters the corridor is displayed in the superimposed box on each sub-figure. The robot's position is corrected

13 times during its navigation. In such a corridor, the robot can be localized in the corridor direction, with an error lower than 20 *cm*.

### *4.2 Auto-calibration with quadrangles*

An extension of our work deals with active vision, which implies to re-estimate camera intrinsic parameters. We propose to do it online, from several views of a planar quadrangle. It is assumed here that these parameters are constant on these views. Using Eq. (2), we evaluate the image of the absolute conic $\omega = K^{-T}K^{-1}$, under the simplified form:

$$\omega = \begin{pmatrix} \omega_1 & 0 & \omega_2 \\ 0 & \omega_1 & \omega_3 \\ \omega_2 & \omega_3 & \omega_4 \end{pmatrix}$$

Let $\Omega = (\omega_1, \omega_2, \omega_3, \omega_4)^t$ be the vector to estimate. Such a parametrization allows to write linear constraints on intrinsic parameters[14]. First, constraints on planar homography deduced from Eq. 2 lead to:

$$\begin{cases} h_1^T \omega h_1 = h_2^T \omega h_2 \\ h_1^T \omega h_2 = 0 \end{cases} \qquad (C_1)$$

Secondly, assuming the roll angle of the camera platform to be neglected, makes the skyline and vertical vanishing point be known. This entails also:

$$(\,0\ 1\ 0\,)\ \omega h_2 = 0 \qquad (C_2)$$

In the same way, given Eq. (2), the constraint of planar robot motion can be written as follows:

$$-\,t_z^{nc}(h_1^T \omega h_1) = h_2^T \omega h_3 \qquad (C_3)$$

Combining these three constraints $(C_1)$, $(C_2)$ and $(C_3)$ allows to solve intrinsic parameters $K$.

Fig. 17 shows calibration results for different constraint combinations. Synthetic experiments (see Fig. 17, left) show that the first constraint $(C_1)$ seems to be sufficient. On the right part of Fig. 17, calibration results for real images are presented. The relative error to ground truth is inferior to 1% which

is suitable for active vision purposes. From five to ten views are required to recover intrinsic parameters with a good precision.

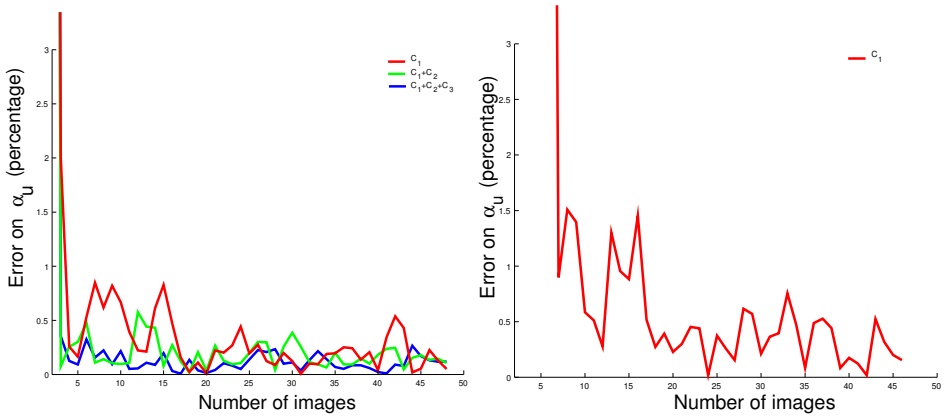


Fig. 17. On-line calibration: errors on $\alpha_u$ constraints for synthetic (left) or real (right) views vs. number of views

## 5 Conclusion and future works

We present an original framework to use quadrangular visual landmarks for robot navigation in indoor environment. A first contribution concerns a method for extracting quadrangles in open cluttered and corridor-like spaces. These quadrangles can correspond to planar objects (posters, doors, cupboards,...). A new representation and associated recognition method for such landmarks is presented. It has been verified that this method remains efficient despite ambiant brightness variations or viewing changes.

Navigation experiments have been performed; the extraction of visual landmarks is very efficient, as well as the landmark recognition method. During the environment exploration, about 90% of pertinent landmarks are extracted; then, when the robot goes along a path planned in the environment model, landmarks are actively searched and exploited for the robot localization. When only posters are considered as landmarks, the recognition rate is greater than 97%. Failures are due to unforeseen occlusions or specific ambiant brightness variations. Our method proposed to select the thresholds, allows to avoid false positive errors.

Two directions are currently studied regarding our visual landmarks based navigation system. Firstly, visual functions described here are exploited for topological navigation and qualitative localization purpose[7]. Considering ambiguous landmarks (doors,...), a markovian localization [6] will be implemented to handle multi-hypothesis on the robot position. Secondly, a more tied coupling strategy is studied to improve the explicit robot localization; the land-

mark model, will be learnt together with the laser map, using a SLAM approach to build a heterogeneous stochastic map.

## References


[1] B.Morisset and M.Ghallab. Synthesis of Supervision Policies for Robust Sensory-motor Behaviors. In *Proc. of Int. Conf. on Intelligent Autonomous Systems (IAS'02)*, 2002.

[2] A. Branca, E. Stella, and A. Distante. Landmark-based Navigation using Projective Invariants. In *Proc. of Int. Symp. on Robotics and Automation (ISRA'00)*, pages 569–574, 2000.

[3] H. Choset, K. Nagatani, and A. Rizzi. Sensor based Planning : Using a Honing Strategy and Local Map Method to Implement the Generalized Voronoi Graph. In *Graph. SPIE Mobile Robotics*, 1997.

[4] C.I. Colios and P.E. Trahanias. Landmark Identification based on Projective and Permutation Invariant Vectors. In *Proc. of Int. Conf. on Pattern Recognition (ICPR'00)*, pages 128–131, 2000.

[5] P. Elinas, R. Sim, , and J.J. Little. $\sigma$slam: Stereo vision slam using the rao-blackwellised particle filter and a novel mixture proposal distribution. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, 2006.

[6] D. Fox, W. Burgard, and S. Thrun. Markov Localization for mobile Robots in Dynamics Environments. *Journal of Artificial Intelligence Research*, 11:1265–1278, 1999.

[7] J.B. Hayet, F. Lerasle, and M. Devy. Environment Modeling for Topological Navigation using Visual Landmarks and Range Data. In *Proc. of Int. Conf. on Robotics and Automation (ICRA'03)*, 2003.

[8] D.P. Huttenlocher, A. Klanderman, and J. Rucklidge. Comparing Images Using the Hausdorff Distance. *IEEE Trans.. on Pattern Analysis and Machine Intelligence (PAMI)*, 15(9), 1993.

[9] G. Jang, S. Kim, W. Lee, and I. Kweon. Color Landmark-based Self-localization for Indoor Mobile Robots. In *Proc. of Int. Conf. on Robotics and Automation (ICRA'02)*, pages 1037–1042, 2002.

[10] J.Santos-Victor, R.Vassallo, and H.J. Schneebeli. Topological Maps for Visual Navigation. In *Proc. of Int. Conf. on Computer Vision Systems (ICVS'99)*, pages 1799–1803, 1999.

[11] F. Launay, A. Ohya, and S. Yuta. A Corridors Lights based Navigation System including Path Definition using a Topologically Corrected Map for Indoor Mobile Robots. In *Proc. of Int. Conf. on Robotics and Automation (ICRA'02)*, pages 3918–3923, 2002.

[12] M. Mata, J. M. Armingol, A. de la Escalera, and M.A. Salichs. A visual landmark recognition system for topological navigation of mobile robots. In *Proc. of Int. Conf. on Robotics and Automation (ICRA'01)*, 2001.

[13] K. Mikolajczyk and C. Schmid. A Performance Evaluation of Local Descriptors. In *Proc. of Int. Conf. on Computer Vision and Pattern Recognition (CVPR'03)*, 2003.

[14] R.Hartley and A.Zimmerman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, 2000.

[15] S. Se, D.G. Lowe, and J. Little. Global Localization using Distinctive Visual Features. In *Poc. of Int. Conf. on Intelligent Robots and Systems (IROS'02)*, pages 226–231, 2002.

[16] R. Sim and G. Dudek. Learning Visual Landmark for Pose Estimation. In *Proc. of Int. Conf. on Robotics and Automation (ICRA'99)*, pages 1972–1978, 1999.

[17] S.Thrun. *Robotic mapping: a survey.* G.Lakemeyer and N.Nebel, editors, Exploring Artificial Intelligence in the New Millenium, Morgan Kaufmann, 2002.