IEEE Int. Conf. on Computer Vision and Pattern Recognition (*CVPR'03*)
pp 313 – 318, Madison (U.S), June 2003.

Visual Landmarks Detection and Recognition for Mobile Robot Navigation

J.B.Hayet[†], F.Lerasle[†‡], M.Devy[†]
[†]LAAS-CNRS, 7 avenue Colonel Roche, 31077 Toulouse Cedex (France)
[‡] Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex (France)
{jbhayet, lerasle, michel}@laas.fr

August 25, 2006

Abstract

This article describes visual functions dedicated to the extraction and recognition of visual landmarks, here planar quadrangles detected from a single camera. The extraction of these landmarks is based on a relaxation scheme, used to verify constraints between image segments. Once extracted, such a landmark must be characterized by invariant attributes, and then recognized from any viewpoint, distance and orientation. The landmark is represented by an icon of fixed size, built using the homography between the current viewpoint and an invariable one. A landmark which is detected again, is recognized, relying on distances between icons: either a classical correlation score or the Hausdorff distance between sets of interest points extracted from the icons. An evaluation on actual and synthetic images, shows how invariant and discriminating, is our landmark representation, and allows to select the best distance for the recognition procedure. Results issued from experimentations of a mobile robot navigating indoor are finally presented.

1 Introduction

This paper presents visual functions integrated on a mobile robot equipped with a single camera, to extract, characterize and recognize landmarks while navigating indoor. For visual navigation, many strategies have already been proposed, based among others, either on the environment representation by an image data base, and indexation methods [5], or on landmarks, detected by the robot during a learning step, and recognized during the execution of a navigation task [6]. For indoor scenes, these landmarks could be simple features (vertical lines, interet points, vanishing points) or some characteristic objects, like doors, furnitures, posters...

Previous works [1, 2] have been devoted to our different strategies to deal with robot navigation. Landmarks detection was limited to quadrangular and planar posters lying on the lateral walls. This paper describes a more generic method suitable to detect and recognize visual landmarks not limited to such vertically-oriented quadrangles: neons on ceiling, groundsheet, doors, posters in any orientation and possibly partially occluded. The detection is based on a relaxation scheme to match image segments corresponding to landmarks boundaries. Each discovered landmark is characterized by an intrinsic representation, a fixed-size icon invariant with respect to illumination, scale changes and small occlusions. When it is perceived again from any viewpoint, this same icon will be built and compared to the ones learnt for every known landmark. Several distances are proposed to make more robust this recognition step.

This paper focuses only on the visual functions, dealing only with the extraction and the recognition of visual landmarks, without considering explicit localization, which could be required for several navigation strategies. Using only qualitative information, our method can be performed with an uncalibrated and active camera. Sections 2 and 3 detail respectively the detection and the recognition procedures. In section 4, a systematic evaluation of our recognition method, with respect to different criteria, shows how robust is this method, especially when variable light, scale and viewpoint conditions are considered. Experimental results from images acquired by a robot navigating indoor, are commented in section 5. Finally, section 6 sums up our approach and opens a discussion for our future works.

2 Landmarks detection

2.1 Overview of the method

The visual landmark extraction is focused on planar, mostly quadrangular objects (e.g. doors, windows, posters, cupboard,...). A natural way of extracting quadrilaterons has been selected, relying upon perceptual grouping on edges segments. Our approach requires an image segmentation in edge segments. Constraints at different levels of features complexity are then taken into account to label the segments $\mathscr{L} = \{l_i\}_{1 \leq i \leq n_L}$ on the classes $\mathscr{L} \cup \{\emptyset\}$. Figure 1 sums up all this process on a simplified way.

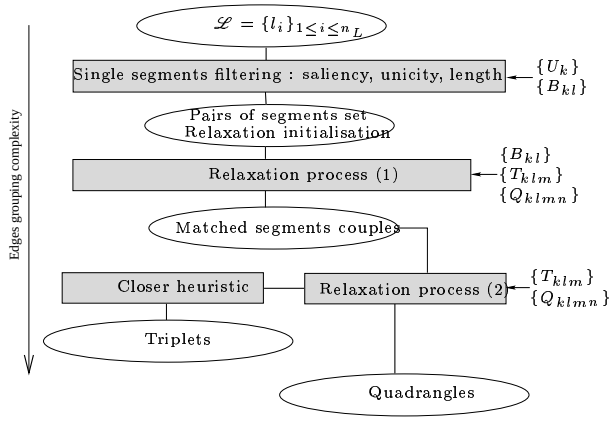


Figure 1: Edge grouping for landmarks detection

Sequential application of constraints between single segments l_k, noted U_k allows to filter the extracted segment set $\mathscr{L}$. Constraints, between couples $(l_k, l_l)_{k \neq l}$, noted B_{kl}, are then applied through a first relaxation scheme in order to define an initial set of couples of segments matchings. For the remaining matchings between segment couples, three-segment sets (resp. four-segment sets) are generated from last types of constraints noted T_{klm} (resp. Q_{klmn}). Finally, a second relaxation process based in particular on spatial configurations for one or two four-segments sets are proposed to select quadrangles corresponding to potential landmarks. Note that we do not deal only with quadrangles but with three-segment sets, that may also be a partially occluded door or quadrangle.

All these constraints, specified in sections 2.3 from 2.5, are applied through a continuous relaxation scheme depicted here after.

2.2 Relaxation formulation

Like Hummel in [3], we define a probability $p_{kl} \in [0, 1]$ for the event e_{kl} corresponding to the association between image segments k and l, given B_{kl}. This formulation can be transposed afterwards to other entities, for example couples of segments given Q_{klmn}.

$p_{kl} = 1$ if k is unambiguously associated with l, 0 if it is not the case. The relaxation process consists in making the ambiguous matchings ($p_{kl} \in [0, 1]$) evolve towards 0 or 1. The $n \times n$ matrix A such as $A_{kl} = p_{kl} = P(e_{kl}/B_{kl})$ and $\mathscr{A}$ the space defined as:

$$\mathscr{A} = \{A \in \mathbb{R}^{n_L} \times \mathbb{R}^{n_L} \mid \forall (k,l)\ A_{kl} \geq 0 \text{ et } \forall k \sum_l A_{kl} = 1\}$$

The iterative relaxation algorithm maximizes the global consistency score $G(A)$ using gradient ascent in the $\mathscr{A}$ space:

$$G(A) = \sum_{klmn} r_{klmn}.A_{kl}.A_{mn}$$

where the r_{klmn} represents the compatibility degree between associations (k, l) and (m, n). In opposition to [3], the gradient step α^i at iteration i is adaptative and defined by: $\alpha^i = \arg\min_{\alpha} G(A^i - \alpha.\nabla G^i)$

Regarding the relaxation process initialization, confidence measures s_{kl} for the event e_{kl} based on B_{kl} are computed. For measures inferior to a certain threshold s_{max}, we have $p^0_{kl} = 0$. For the others:

$$p^{(0)}_{kl} = \frac{s_{kl}}{\sum\limits_{n/s_{kn} > s_{max}} s_{kn}}$$

2.3 Constraints on single segment

This first class of constraints U_k filter the initial segment set $\mathscr{L}$ issued from the edge segmentation. The vector U_k for a given segment k is composed on: $|l_k|$ (resp. θ_k) the length (resp. the orientation) in the image, ul_k the local unicity which is a flag indicating its conservation or not for the next step. Typically, segments corresponding to the floor tiling will be labelled (by a accumulator technique) and so eliminated.

We have introduced a very simple but quite effective heuristic that reduces computer time by limiting the initial number of segments to be managed. From the estimation of the position of the robot, an expected average length of the segments corresponding to the landmarks edges, can be computed. This value is selected as the minimum segment lenght, with a confidence margin of 20% .

2.4 Constraints on segment couples

Segment couples corresponding to potential landmarks are formed according to geometric and luminance consistency criteria.

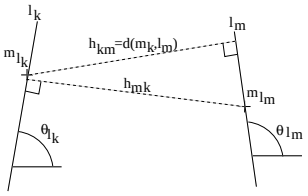


Figure 2: Geometric criteria for segments couples

With the notations in figure 2, the geometric criteria expressed by the vector B_{km} are as follows:

- the segments length ratio $\frac{1}{2}(\frac{|l_m|}{|l_k|} + \frac{|l_k|}{|l_m|})$;
- the overlapping rate in the segment orientation;
- the angular difference $|\theta_{l_k} - \theta_{l_m}|$;
- a shape criteria $\frac{1}{2}(\frac{|l_m|+|l_k|}{h_{km}+h_{mk}} + \frac{h_{km}+h_{mk}}{|l_m|+|l_k|})$.
- a third segment in the neighbourhood that forms a polygonal convex chain (three-segments set) with the given couple.

For the luminance criteria, an average grey-level profile is computed in the direction orthogonal to each extracted segment l_k. For each association (l_k, l_m), we can deduce a photogrammetric similarity score resulting from correlation between these profiles. All these criteria are taken into account in the computation of the elements $p_{kl}^{(0)}$.

2.5 Constraints between two segments couples

Uniqueness and convex rules are checked for the potential segments couples matchings. Uniqueness constraint allows to reduce the relaxation algorithm complexity. Convexity rule says that two segments couples define two quadrangles which must verify rules of full inclusion or no intersection.

The entities Q_{klmn} and r_{klmn} correspond to the verification of these two constraints. The entities T_{klm} useful in the heuristic for three-segments sets are based on the same criteria.

3 Landmarks recognition

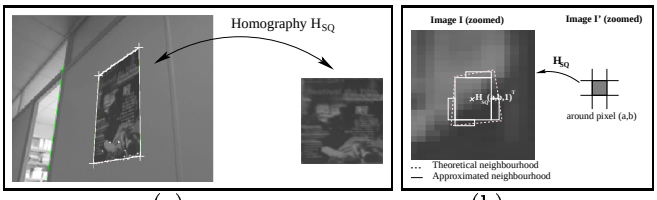
Once landmarks are detected, landmark models must be built in order to deal with their recognition from different viewpoints. At first, the boundaries of a detected quadrangular landmark allow to rectify the observed pattern. Using such a mapping for recognition provides an *invariant* representation under scale and perspective changes.

To perform recognition, the simplest methods could be based on correlation between these icons. In the section 4, a correlation-based method is compared with an approach based on interest points extracted from the icons. At first, we describe hereafter how appearance models of landmarks are learnt and compared.

3.1 Landmark iconification

Let us consider, (1) an extracted quadrangular landmark $Q = \{P_i\}_{1 \leq i \leq 4}$ from an image I, and (2) a fixed-size square S, corresponding to a $s \times s$ picture (s typically equal to 75), at lower scale than in image I. The four matchings between the S and I corners, allow to define an homography H_{SQ} mapping points from S to Q. Using H_{SQ}, a new small-sized image I' is built from the image I by averaging pixels from I into the pixels in I'. This process is illustrated on figure 3-(a). To handle perspective distortions, averaging is done in order to avoid too much information compression in the low-scale front view I' approximated from the view I. If we consider a pixel (a, b) in image I', its grey level value is determined by taking into account all the pixels in image I belonging to a certain neighbourhood of $H_{SQ}(a, b, 1)^T$, its image in I. This neighbourhood is computed by approximating with simple heuristics the image of a pixel square, i.e. a certain quadrilateron (figure 3-(b)).

This icon I' is processed by the Harris operator to get a set of n interest points $\{P_i\}_{1 \leq i \leq n}$. A local descriptor [8] in $\mathbb{R}^7$, based on Gaussian derivatives is then associated to every point.



(a) (b)

Figure 3: (a) landmark model construction, (b) averaging invariant with respect to perspective distortions

3.2 Distance between landmark models

To perform recognition between a set of N learnt landmarks noted $\{C_l\}_{1 \leq l \leq N}$ and a detected landmark $\mathcal{Q}$,

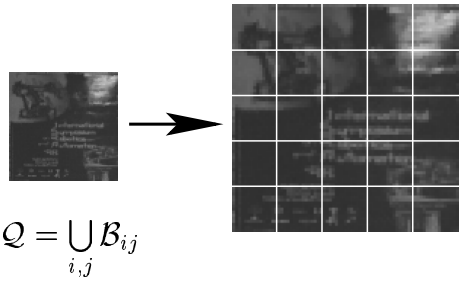


Figure 4: Icon divided in buckets

we must define for each C_l, a distance noted $\mathscr{D}_l = \mathscr{D}(\mathcal{Q}, C_l)$. From the icons built for the C_l and $\mathcal{Q}$ landmarks, the classical centered and normalized correlation score noted $\mathscr{C}^f$, provides a distance invariant to overall light changes. However two drawbacks arise: first, the lack of compactness of the data to be stored, as we would need to memorize the icons for each learnt landmark C_l, and, second, the difficulty to handle partial occlusions, or local light effects.

Using the ZNCC score makes this distance more robust against the illumination changes. To be less sensitive to local variations or occlusions, the correlation score between two icons $\mathcal{Q}$ and C_l, is based on the separated correlations $C_{ij}(\mathcal{Q}, C_l)$ between buckets i et j ($1 \leq i, j \leq n_B$, typically $n_B = 5$, figure 4). A robust correlation score is provided using the k^{th} greatest correlation score between buckets, k selected to ignore the more important local differences between $\mathcal{Q}$ and C_l,

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

Another popular appearance-based method [8] for object recognition, is based on interest points matched thanks to their local descriptors; the remarkable behavior of the Harris operator as far as scale, rotation or light changes are concerned, has been proven in [9].

Here, the partial Hausdorff distance, introduced by Huttenlocher[4], is used to compare sets of interest points. Let be two sets of points $S_l = \{P_i^l\}_{1 \leq i \leq n_l}$ extracted from a learnt landmark C_l, and $S = \{P_j\}_{1 \leq j \leq n}$, extracted from a detected landmark $\mathcal{Q}$. The Hausdorff distance between S_l and S is defined from a distance d between points:

$$\begin{cases} d_h(S_l, S) = \max(h(S_l, S), h(S, S_l)) \\ h(S_l, S) = \max_{1 \leq i \leq n_l} \min_{1 \leq j \leq n} d(P_i^l, P_j) \end{cases}$$

A natural way of avoiding outliers, consists in relaxing this definition, taking the k^{th} greatest minimum distance or, equivalently, a fraction r of $\min(n_l, n)$, to define:

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

A threshold τ on the computed distance must be selected to recognize the current landmark $\mathcal{Q}$. Physically speaking, an object is recognized provided that we can find a matched point in the first set for at least k points in the second, and vice versa. From this definition, we see that we can stop computation as soon as $\min(n_l, n) - k$ outliers have been detected, so recognition tests can be performed in relatively short time.

The Hausdorff partial distance between two sets of points, depends on the d distance. Comparing only the spatial configuration of the interest points in the compared icons, the euclidian distance d_2 could be used between two points a and b:$d_2(a, b) = \|a - b\|^2$. Comparing only photogrammetric attributes expressed on each point by the local descriptors, the Mahalanobis distance d_ν could be proposed. In order to take into account both spatial and photogrammetric similarities between points, these two distances are combined to define a distance noted d_p:$d_p(a, b) = d_\nu(a, b).d_2(a, b)$.

The Hausdorff distance based on the simple euclidean distance d_2, will be noted $\mathscr{H}_2^f$; the one based on the composite distance d_p will be noted $\mathscr{H}_p^f$.

3.3 Learning appearance models

The learning step is performed for each landmark C_l, from a set of N_l representative images I_i (typically 50

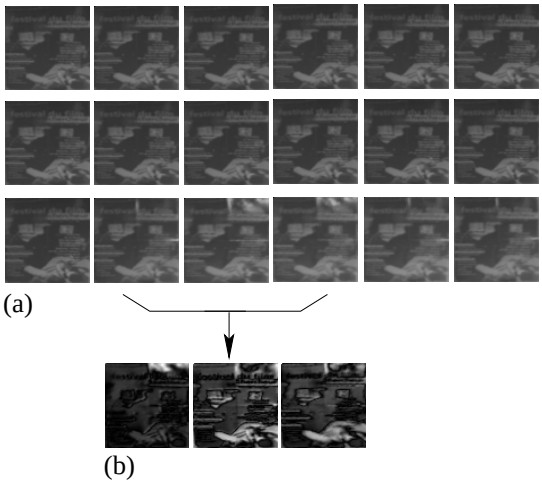


Figure 5: (a) Some icons generated for a landmark (b) the three more representative icons

images) from which iconified views I'_i are extracted. From these icons, stable interest points must be extracted, and knowing all landmark models, an optimal recognition threshold could be learnt.

3.3.1 Learning landmark representation

For the distance $\mathscr{C}^f$, a Principal Component Analysis on the raw data extracted from each landmark allows to keep only the three most representative icons I'_i built from the landmark images (figure 5), noted $\mathcal{Q}_l^1, \mathcal{Q}_l^2, \mathcal{Q}_l^3$: $\mathcal{Q}_l^1$ corresponds to the mean icon from $I'_i, 1 \leq i \leq N_l$; $\mathcal{Q}_l^2$ and $\mathcal{Q}_l^3$ correspond to the more significative variations on this icon set.

For the distances $\mathscr{H}_2^f$ and $\mathscr{H}_p^f$, interest points could be extracted from these three icons, but another method, based on a clustering technic, has been preferred in order to keep only the more salient points amongst all the icons I'_i. The figure 6 presents the histograms of the interest point positions extracted from all the icons shown on figure 5(a); the local extrema extracted from this histogram, provide the more invariant interest points.

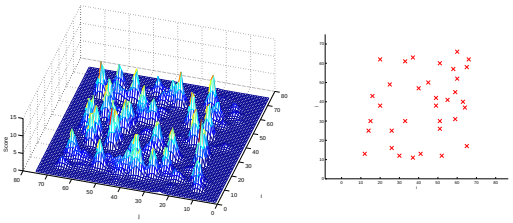


Figure 6: Clusters of the interest points.

3.3.2 Learning recognition thresholds

During the recognition step, a detected landmark will be successively compared to each learnt landmark C_l, using a recognition threshold τ. During the learning step, the distances between all C_l models and all the other models noted $\neg C_l$, are computed, using the distances $\mathscr{C}^f$, $\mathscr{H}_2^f$ or $\mathscr{H}_p^f$. Let us note $p(d|C_l)$ and $p(d|\neg C_l)$, the probability densities. These probabilities are approximated by their corresponding distance distributions (figure 9).

Given a threshold τ, we aim to minimize the sum:

$$S(\tau) = \lambda \underbrace{\int_0^{\tau} p(d|C_{\neg l})\mathrm{d}d}_{S_l(\tau)} + \mu \underbrace{\int_{\tau}^{+\infty} p(d|C_l)\mathrm{d}d}_{\neg S_l(\tau)} \quad (1)$$

with λ and μ two weights respectively for the false positive and false negative decisions, noted $\neg S_l(\tau)$ and $S_l(\tau)$ in figure 7. For our application, we fix $\mu = \frac{1}{6}\lambda$ to enforce the security in the robot navigation. From this

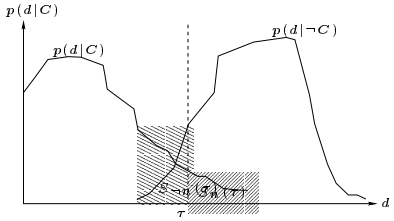


Figure 7: Confusion area for threshold determination

assumption and relation (1), the two graphs presented on figure 8, show the function $S(\tau)$ and the estimated thresholds for distances $\mathscr{C}^f$ and $\mathscr{H}_p^f$.

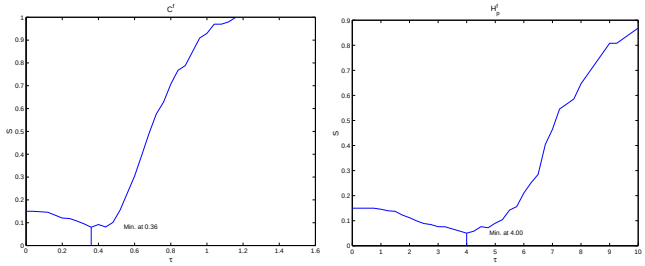


Figure 8: Graphs $S(\tau)$ for thresholds determination

3.3.3 Learning validation

For every landmark C_l, the learning step ends with a verification of two criteria: (1) C_l must be salient enough, and (2) the N_l learning images I_i 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 to be satisfied, is verified from the *covariance* of the iconified views I' and from the number of stable interest points extracted from I'.

The **visibility** criterion indicates how far from each other are the extreme positions at which the landmark is seen. For all the couples $(i,j) \in [1, N_l]^2$, an inter-image homography H^{ij} allows to map the vertices of the landmark in image I_i to its corresponding vertices in image I_j. Let us consider the normalized homography $\hat{H}^{ij}$, such as $\hat{H}^{ij}_{33} = 1$. Then we define the visibility confidence by: $v_c = max_{ij}\|\hat{H}^{ij} - I\|$.

I is the identity matrix. The greater is v_c, the more extended is the area on which the landmark has been perceived during the learning step.

3.4 Confidence in the recognition result

The recognition task requires the robot to index and compare the detected visual landmarks during its nav-

igation task. For a set of N learnt landmarks (classes) noted $\{C_l\}_{1\leq l\leq N}$ and a detected landmark $\mathcal{Q}$, we can define for each class C_l, a distance noted $\mathscr{D}_l = \mathscr{D}(\mathcal{Q}, C_l)$ and an a priori probability $P(C_l|\mathcal{Q})$ of labeling to C_l:

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

with $C_\emptyset$ refers the empty class and h the Heaviside function: $h(x) = 1$ if $x > 0$, 0 otherwise. Leray in [7] proposed an entropy based measure in order to minimize the probabilities on the losing classes:

$$m_e(\mathcal{Q}, \{C_l\}) = 1 + \frac{1}{N+1} \sum_P (C_l|\mathcal{Q}) \log P(C_l|\mathcal{Q})$$

The learning step, considering for each landmark, a set of representative images, makes possible also a bootstrapping method to estimate a variance on the a posteriori probabilities. Considering only the three representative $\mathcal{Q}_l^j$ icons with $j = 1 \leq j \leq 3$, a second confidence measure, related to the learning quality, can be computed by:

$$m_b(\mathcal{Q}, C_l) = \frac{\mathscr{D}(\mathcal{Q}, \mathcal{Q}_l^{\hat{j}})}{\mathscr{D}(\mathcal{Q}, \mathcal{Q}_l^{\check{j}})}$$

with $\hat{j}, \check{j}$ corresponding to the minimum and median of $\mathcal{Q}_l^j$ with $1 \leq j \leq 3$ (§ 3.3.1).

Finally, a global confidence measure on the recognition task, taking into account the learning quality, is given by:

$$m(\mathcal{Q}, \{C_l\}) = m_e(\mathcal{Q}, \{C_l\}).m_b(\mathcal{Q}, C_l)$$

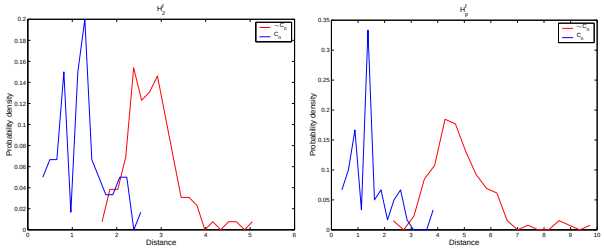
4 Method evaluation

An important issue we must care about our recognition process is the way the algorithm behaves with light effects (classic in indoor environment), scale/perspective changes and bad warping from the detection step. Especially, does our landmark representation is stable enough? Other questions are relative to the discriminating power, especially with partial occlusions of landmark areas. To investigate this robustness problem, a test image database is both constituted by:

1. real images of different landmarks $\mathcal{Q}_n$ acquired while the robot wandered around the lab.
2. synthetic images of 300 movie posters $\mathcal{Q}'_n$ with different light, scale/perspective conditions and occlusions, these modalities remaining quite difficult to perform and quantify in real conditions.

4.1 Discriminating power

First, the representation discriminating power is analyzed through the distribution of the distances we get between a given landmark and other ones from the database $\mathcal{Q}_n$. A poster we find in this database has been selected and learned as a landmark, and figure 9 now represents the distributions of the distance values we get (a) for the objects corresponding to this learned landmark (class C_l) and (b) for the objects not corresponding to it (class $\neg C_l$).



(a) (b)

Figure 9: Distances distribution on $C_l \cup \neg C_l$ ($\mathcal{Q}_n$) for: (a) distance $\mathscr{H}_2^f$, (b) distance $\mathscr{H}_p^f$

We note that classes are well separated even if the separability is partial, especially for distance $\mathscr{H}_2^f$ for which the confusion area is more important. This distance has been neglected during the remainder of the evaluation. Moreover, these distances distribution can be approximated by a Gaussian function, which center and variance depends on the Hausdorff fraction and on the sets cardinals.

4.2 Behavior under view point change

The graphs in figure 10 represent the evolution under scale change of the ratio $\frac{distance}{threshold}$ for the distances $\mathscr{C}^f$ and $\mathscr{H}_p^f$. For a scale factor about 3, the values for the two remain small comparatively to the thresholds τ_c and τ_p defined in § 3.3.2 and so scale changes do not really affect recognition results. Moreover and as expected, results are deteriorated as soon as the pattern apparent size is below the size of the square we use for representation, i.e. 75.

As far as perspective distortions are concerned, we have studied the evolution of the ratio $\frac{distance}{threshold}$ for the distances $\mathscr{C}^f$ and $\mathscr{H}_p^f$ when performing a planar rotation in the horizontal plane of a quadrangular landmark. The results (not illustrated here) 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 $\pm 75°$, a situation that may occur in corridor-like environments.

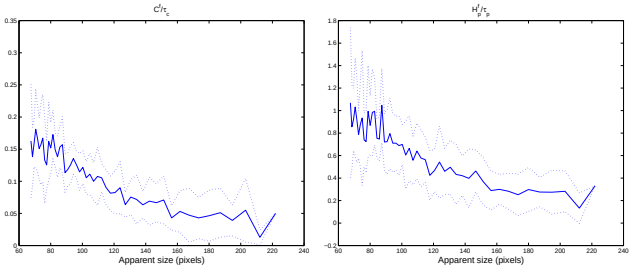


Figure 10: Variations and standard deviation under scale change of the ratio $\frac{\text{distance}}{\text{threshold}}$

4.3 Behavior under light effects and partial occlusions

The two graphs in figure 11 show that it is possible to have good recognition results for distances $\mathscr{C}^f$ and $\mathscr{H}_p^f$ until local or global light saturations appear in the image.

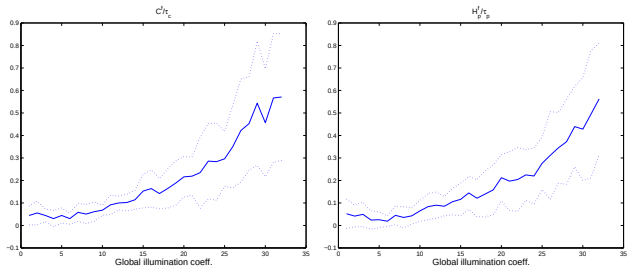


Figure 11: Variations and standard deviation of the ratio $\frac{\text{distance}}{\text{threshold}}$ under lightening changes

Other studies concern the robustness to the partial occlusions because partially detected landmarks compose the majority of detected landmarks in indoor environment. The aspect of the ratio $\frac{distance}{threshold}$ for the distances $\mathscr{H}_2^f$ and $\mathscr{H}_p^f$ when performing partial occlusions is close to one of figure 11. These partial occlusions are quantified by the percentage p_0 of the occlused surface in the two bottom graphs in figure 11. We have noted that we can consider partial occlusions of the landmark up to 46%, 51% and 56% of its area (maximum of p_0), respectively for distance $\mathscr{C}^f$, $\mathscr{H}_2^f$, $\mathscr{H}_p^f$.

4.4 Discussion

These robustness performances have been verified during robot navigation task. From all these experiments, we finally conclude that each of the two landmark representations and associated distances $\mathscr{C}^f$ and $\mathscr{H}_p^f$ for recognition task are robust with respect to the viewing conditions changes. This behavior can be extended to scale change and perspective distortion. Moreover the behavior under partial occlusions of the two metrics is quite interesting. All these characteristics prove that our approach seems to be suitable for mobile robot navigation in variable environment.

We have noted in § 4.1 that distances $\mathscr{C}^f$ and $\mathscr{H}_p^f$ are more discriminant than distance $\mathscr{H}_2^f$. Beside the well-known partial correlation technique, distance $\mathscr{H}_p^f$ gives some advantages in terms of information compacity since it guarantuees the same robustness. An other key that makes this distance more practice is that this representation is adapted to landmark pose refinement from interest points matchings. In this way, recognition process embelled on the mobile robot is carried out using this last representation.

5 Mobile robot navigation

Experiments in a large images database of variable environments have been realized: the environment can be either a corridor network or environment with complex background. The robot is a Nomadic $XR400$, equipped with a SICK laser range finder, two belts of ultrasonic sensors and a CCD camera mounted on a pan and tilt platform. Figure 12 shows examples of landmarks detection in an office-like environment represented in figure 13 by laser SICK data. Detected landmarks do not necessarily correspond to physically distinct objects.

All the detected landmarks have been superimposed on the figure: windows or posters in red color (lateral walls or ceiling), neons in blue color (noted Ne), doors represented by a grey icon.

During its navigation task, where the robot stopped to perform detection and recognition procedures, it is able to achieve these tasks dynamically. The detection rates are computed over the images database $\mathcal{Q}_n$ taken by the robot in the office environment. On this database, landmarks detection rates are 88% in the corridors and 89% in the open spaces. Only detected landmarks which are detected from different view points (§ 4) are considered for the environment modeling.

6 Conclusion

We have presented an original framework for the use of visual landmarks. A first contribution concerns the method proposed for extracting landmarks: a relaxation scheme allows us to extract potential quadrangles from the set of segments extracted from a single image. These quadrangles can correspond to planar objects (poster, doors, cupboard,...) located everywhere

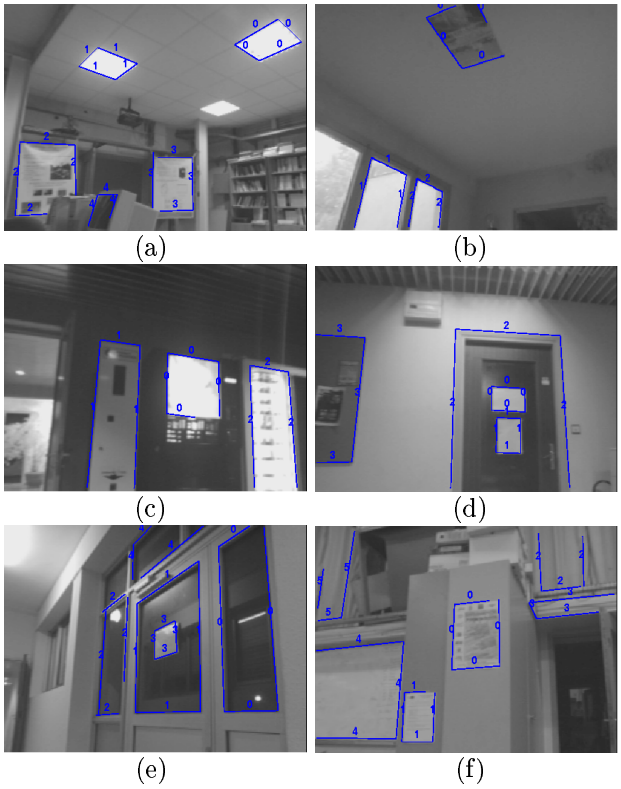


Figure 12: Landmarks detection

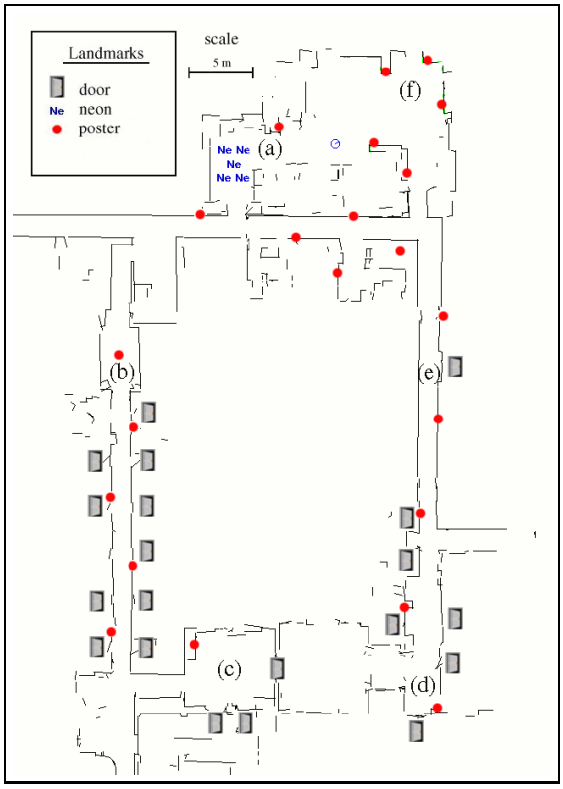


Figure 13: Landmarks detection in office environment

in the scene.

A new representation and associated recognition method for this kind of landmarks is presented and is dedicated to robot navigation. We have shown that this method remains efficient despite the lightening or viewing changes due to our application.

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 the pertinent landmarks are extracted; after, when the robot goes along a path planned in the environment model, these landmarks are actively searched and exploited for the robot localization: if we consider landmarks corresponding to posters, the recognition fails only in 3% of the situations, due to some unforeseen occlusions or specific lightening conditions.

In future works, our navigation system could be improved considering a more active visual system; up-to-now only pan and tilt rotations are generated in order to point the camera towards the searched landmarks. Zooming will be used to actively adapt the focal lenght with respect to the landmark-robot distance.

References

[1] J.B. Hayet, C. Esteves, M.Devy, and F.Lerasle. Qualitative Modeling of Indoor Environments from Visual Landmarks and Range Data. In *Proc. Int. Conf. on Intelligent Robots and Systems (IROS), Lausanne*, 2002.

[2] J.B. Hayet, F.Lerasle, and M.Devy. A Visual Landmark Framework for Indoor Mobile Robot Navigation. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), Washington*, volume 4, pages 3942–3947, 2002.

[3] R.A. Hummel and S.W. Zucker. On the Foundations of Relaxation Labeling Processes. *IEEE trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 5(3):267–287, 1983.

[4] 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.

[5] J.Santos-Victor, R.Vassallo, and H.J. Schneebeli. Topological maps for visual navigation. In *1st Int. Conf. on computer Vision Systems (ICVS'99)*, pages 1799–1803, jan 1999.

[6] D. Kortenkampf and T. Weymouth. Topological Mapping for Mobile Robots using a Combination of Sonar and Vision Sensing. In *National Conf. on Artificial Intelligence (AAAI)*, 1994.

[7] P. Leray, H. Zaragoza, and F. d'Alché Buc. Relevance of Confidence Measurement in Classification. In *Reconnaissance des Formes et Intelligences Artificielles (RFIA), Paris*, volume 1, pages 267–276, 2000.

[8] C. Schmid and R. Mohr. Local Gray-value Invariants for Image Retrieval. *IEEE trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 1(19):530–535, May 1997.

[9] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and Evaluating Interest Points. In *Int. Conf. on Vision System (ICCV), Bombay*, pages 313–320, 1998.