Robust Mapless Outdoor Vision-Based Navigation

Christophe Giovannangeli
CNRS UMR8051 ETIS
Neurocybernétic Team
Université de Cergy-Pontoise - ENSEA
2, avenue Adolphe-Chauvin
95312 Cergy-Pontoise, France
Email: giovannangeli@ensea.fr

Philippe Gaussier
CNRS UMR8051 ETIS
Neurocybernétic Team
Université de Cergy-Pontoise - ENSEA
2, avenue Adolphe-Chauvin
95312 Cergy-Pontoise, France
Member of the Institut Universitaire de France

Gaël Désilles
DGA-LRBA
Laboratoire de Navigation Autonome
Vernon, France

Abstract— This article presents an efficient and mature vision-based navigation algorithm based on a sensory-motor learning. Neither Cartesian nor topological map is required, but a set of biologically inspired place cells. Each place cell defines a location by a spatial constellation of online learned landmarks. Their activity provides an internal measure of localization. A simple set of place-action associations enable a robot to go back to a learned location or to follow an arbitrary visual path. The system is able to achieve sensory-motor tasks in indoor as well as in large outdoor environments with similar computation load. The behavior is robust to kidnapping, object and landmark addition or removal, presence of mobile obstacles and severe visual field occlusions.

I. INTRODUCTION

This paper describes our ongoing work to test the feasibility of building an autonomous sentinel robot inspired from neurobiological data for patrolling and exploration missions in an *a priori* unknow environment (with the constraint of using only local visual information and no global positionning system). Since the discovery of neurons called *place cells* (PCs) in the hippocampus of the rodents which activity were highly correlated with the position of the rat in its environment [1], our understanding of the cognitive navigation mechanisms has widely increased. Visual information is involved in taxon navigation (going back to a particular landmark), or place recognition from distant landmarks. Biological models of vision-based navigation use the azimuth of the landmarks [2], [3], or more rarely, their identity or a conjunction of both [4], [5], [6]. In parallel, robotics has largely taken advantages of the computer improvement, enabling to visually generate a 3D model of hundreds of meters of a road [7]. In contrast, autonomous navigation in various environments is still a challenge. Constraints in indoor and outdoor environments are so different that navigation algorithms have always been subdivided in two branches: indoor navigation vs. outdoor navigation. Biomimetic navigation seems to offers a relevant approach to reconcile outdoor and indoor navigation as it will be illustrated, since the different biomimetic models are segregated by the cognitive complexity of the task and not by their field of application [8], [9].

The next section briefly reviews the major navigation approaches and discusses on their inherent environmental depen-

dence. The third section presents a model of pre-hippocampal PCs providing a robust level of localization decreasing regularly over a large area, without the need of any explicit Cartesian or topological map. Section IV focuses on robustness enhancement for outdoor environment. The sensory-motor learning enabling the robot to go back to a location or to follow a trajectory is presented in section V. The robustness of our algorithm to kidnappings, object additions and removals, presence of mobile obstacles and severe visual field occlusions will be pointed out. The conclusion will emphasize the interest of our map-less sensory-motor approach and expose future orientations.

II. STATE OF THE ART IN AUTONOMOUS NAVIGATION

Three main approaches dominate indoor navigation: map-based navigation that requires a pre-defined model of the world, map-building-based navigation that includes classical SLAM and SPLAM methods, and map-less navigation [10]. Map-based navigation refers to occupancy maps, the VFF (Vector Field Forces) or S-MAP (S refers to Squeezing 3D world into 2D map) representations. The processing stream is the following: capturing the available sensorial information to detect landmarks and then establishing correspondences between extracted and expected landmarks in order to identify the current position. Those methods, widely used today in many navigation researches, are strongly dependent to a prior model of the world. Therefore autonomy of the robot depends on the quality of this prior knowledge, hence it is not robust to unknown environmental disturbances such as landmarks displacement or occlusion. To avoid this dependency, self building map algorithms have been introduced. Two major contributions in this field are occupancy-grid and topological approaches [11], [12]. Whereas the prior method implies the use of a 2D grid representing a plane world in which each cluster is affected of a computed probability to be occupied by the robot, the next tries to describe the world as the graph in which nodes and links can have various significations (for example, locations and directions of the adjacent locations). In [13], five major issues in indoor mapping are listed. 1) The nature of noise on the physical measurements is generally context-dependent in mobile robot navigation. 2) The dimensionality of the problem increases drastically with

the size of the environment and with the complexity of the internal representations. 3) The algorithms must deal with the *correspondence problem* (or data association problem) to determine if two sensorial measurements taken at different instants correspond to the same physical position in space. 4) Environmental changes may occur throughout the robot life. And finally, 5) the robot has to select the action during the mapping process (path planning in partially known environments). Moreover, designers are confronted to a trade-off between speed and reliability: the speed and the accuracy of the incremental algorithms (building a consistent map in one pass) strongly depend on the precision of the measurements, as opposed to the algorithms requiring several passes to build a consistent map which are slower but more reliable. However, the map-building-based methods exhibit impressive robustness in long term navigation when coupled with vision-based localization systems to deal with the correspondence problem [14]. Less advertised, mapless approaches allow navigation without prior information about the environment (as for instance a stabilized Cartesian map, a stabilized topological map, a world description ...). The biomimetic approach is prevalent in this field (see for instance [8], [9] for a non exhaustive review). Beside bio-inspired navigation systems, major contributions to mapless navigation include optical-flow-based and appearance-based navigation [10]. Even if most of the models have been tested only in indoor environments, they do not rely on their specific features.

Concerning outdoor navigation, successful results are less abundant due to harder constraints such as roughness of the terrain, size of the environment or absence of structures (*i.e.* walls). Relevant works in structured environments mainly focus on visual road-following and unmanned vehicle guidance in urban environments [7], requiring heavy image processing and/or a prior model of the road. Navigation in less structured outdoor environments still remains a great challenge [15], [Darpa challenge] since classical techniques for indoor mapping are rarely pertinent. Researches mostly focus on external referencing, Cartesian Elevation Map formation [16], or global positioning. As compared to indoor navigation, accuracy and robustness of navigation systems in unstructured outdoor environments are still very weak, as regard to the requirements of the missions. Moreover most approaches strive to keep a constant precision all along the mission, whereas very accurate positionning is needed only on critical phases.

III. A MODEL OF VISUAL PLACE CELLS

This section describes a model of pre-hippocampal visual PCs, tested on miscellaneous platforms[1]. The same neural network has been implemented on these robots to perform missions in open indoor and outdoor environments. Fig. 1 summarizes our visual processing chain. A place is defined by a spatial constellation of online learned visual features

[1]Koala, K-Team and Labo3,AAI and Pionner 3 AT, ActivMedia

(here a set of triplets *landmark-azimuth-elevation*). Coupled with a sensory-motor system, the place recognition algorithm has already been proved to be sufficient to generate a robust navigation behavior with a trivial landmark extraction method in structured indoor environments [17], [6]. The place recognition has been recently optimized for complex open outdoor environments [18].
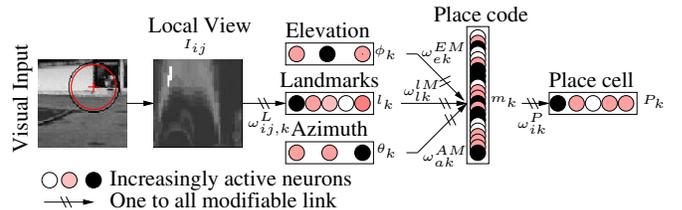

Fig. 1. Block diagram architecture of the place recognition: Our architecture is composed of a visual system that focuses on points of interest and extracts small images in log-polar coordinates (called local views), a merging layer that compresses *what* and *where* information, and a place recognition layer.

The first layer of the architecture is a visual system autonomously extracting landmarks from a panoramic image [17], [19], which may be either re-built from a set of classical images, covering the whole 360° of the available visual field, or captured by means of a conic mirror (Vstone VS-C42N-TR) to speed up the experimentations [18]. For sake of robustness to daylight intensity variation, the gradient image is computed from the CCD input. This gradient image is convolved with a DoG (Difference Of Gaussian) filter to detect robust focus feature points at low resolution. A competition between the feature points enables the system to primarily focus on most intesive points (based on a contrast and edge curvature criterium). A small image of a given circular area around each focus point in log-polar coordinates, is computed to enhance the pattern recognition when small rotations and scale variations occur [20]. In the following, these small log-polar images will be called *local views* since they contain visual information in the neighborhood of the focus point. Figure 2 illustrates the selected focus points and the size of the circular area converted in log-polar coordinates as a landmark view.

During the learning of a location, each local view is learned as a landmark for the system. Neurons are recruited to encode these local views. They are called *landmark neurons*[2]. Each landmark neuron expresses a confidence level granted to the recognition of the encoded local view. The kth landmark recognition level $l_k(t)$ of the current local view is:

$$l_k(t) = f^{RT}\left(\frac{1}{X_I.Y_I}\sum_{i,j=1}^{X_I,Y_I}\|\omega_{ij,k}^L(t) - I_{ij}(t)\|\right) \quad (1)$$

[2]To avoid biological discussions and simplify the explanations, figurative names will be used instead of their neurobiological counterpart. Details about the biological structures that may be associated to these processing can be found in [18], [21].

with X_I and Y_I the number of columns and rows of the local views. $\omega^L_{ij,k}(t)$ is the weight of the connection between the pixel i,j of the current local view and the kth landmark neuron. $I_{ij}(t)$ is the value of the ij^{th} point of the current local view. $f^{RT} = \frac{1}{1-RT}[x - RT]^+$ is an activation function that increases the dynamics of the responses (RT: Recognition Threshold), with $[x]^+ = x$ if $x \geq 0$ and 0 otherwise. The learning of a local view is performed in *one-shot* on a new recruited landmark neuron k according to the following rule (the value $\omega^L_{ij,k}$ are initially null): $\Delta\omega^L_{ij,k} = I_{ij}(t).R^L_k$. They adapt during the *one-shot* learning (at the recruitment, $R^L_k = 1$), and do not change anymore ($R^L_k = 0$ otherwise). The system also provides the azimuth and the elevation of the focus points by means of two groups: the *azimuth neuron* group and the *elevation neuron* group. The azimuth (absolute direction) is obtained with a compass or any simulation of a vestibular system [17] or inertial systems[3]. Each azimuth and elevation neuron has a preferred firing direction and expresses how near is the current extracted local view from its favorite direction. The activity $\theta_i(t)$ and $\phi_i(t)$ for the ith azimuth and elevation neuron are given by a strictly monotonous function $g^\rho()$ decreasing from 1 to 0 with the angular distance Δ_θ between the favorite direction and the position of the current local view (ρ_{Az} and ρ_{El} are the extent of the lateral diffusion).

θ = 37° θ = 49° θ = 87° θ = 106° θ = 109° θ = 113° θ = 121° θ = 128°

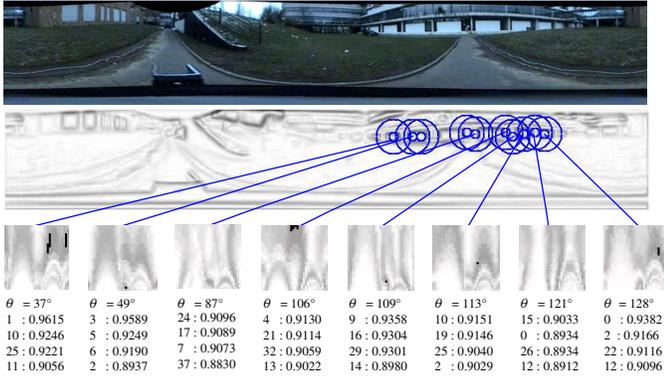| θ = 37°	θ = 49°	θ = 87°	θ = 106°	θ = 109°	θ = 113°	θ = 121°	θ = 128°
1 : 0.9615	3 : 0.9589	24 : 0.9096	4 : 0.9130	9 : 0.9358	10 : 0.9151	15 : 0.9033	0 : 0.9382
10 : 0.9246	5 : 0.9249	17 : 0.9089	21 : 0.9114	16 : 0.9304	19 : 0.9146	0 : 0.8934	2 : 0.9166
25 : 0.9221	6 : 0.9190	7 : 0.9073	32 : 0.9059	29 : 0.9301	25 : 0.9040	26 : 0.8934	22 : 0.9116
11 : 0.9056	2 : 0.8937	37 : 0.8830	13 : 0.9022	14 : 0.8980	2 : 0.9029	12 : 0.8912	12 : 0.9096

Fig. 2. Illustration of landmark extraction mechanism: the gradient of a panoramic image is convolved with a DoG filter. The local maxima of the filtered image correspond to points of interest (center of the circles). Here, the eight first focus points are displayed. The system focuses on these points to extract local views in log-polar coordinates corresponding to landmarks. The system also provides the bearing of the focus points by means of a magnetic compass.

Merging of the *what* and *where* information (the landmark recognition and its spatial localization in the visual field) is performed in a product space (*i.e.* a third-order tensor compressed in a vector of pi-sigma units which the elements $m_k(t)$ are called *merging neurons*, defining a tensorial place code $M(t)$. The recruited merging neurons characterize the learned

<hr>

[3]Interestingly, an azimuth sensor is not strictly necessary. Simulations has shown that a local reference (such as the bearing of a distant landmark) was a sufficient estimation of the robot orientation to infer correct the place recognition. Moreover, a visual compass has already been studied with exactly the same landmark extraction architecture [17].

landmark-azimuth-elevation triplets. The activity $m_k(t)$ of the kth merging neuron (*i.e.* the kth learned triplet) is:

$$m_k(t) = \max(\sum_{l=1}^{N_L} L_{lk}(t) \cdot \sum_{a=1}^{N_A} A_{ak}(t) \cdot \sum_{e=1}^{N_E} E_{ek}(t)$$
$$, [m_k(t - d_t) - r_k(t)]^+) \qquad (2)$$

with $L_{lk}(t) = \omega^{LM}_{lk}(t).l_l(t)$, $A_{ak}(t) = \omega^{AM}_{ak}(t).\theta_a(t)$ and $E_{ek}(t) = \omega^{EM}_{ek}(t).\phi_e(t)$. A binary reset signal $r_k(t)$ enabling to forget the previous activity. It fires at the beginning of a visual exploration of a panoramic image. $\omega^{LM}_{ak}(t)$, $\omega^{AM}_{bk}(t)$ and $\omega^{EM}_{ck}(t)$ are respectively the weight of the synapses of the kth merging neuron coming from the ath landmark, b^{th} azimuth and c^{th} elevation neuron and are initially null. N_L, N_A and N_E are the number of recruited landmark, azimuth and elevation neurons.

More precisely, the learning of a location triggers the learning of all the extracted landmarks based on the current panoramic image, inducing the build up of new triplets in the merging tensor to define a new place code. Each recruited merging neuron is linked to the new recruited landmark neuron and to the most active azimuth neuron and elevation neuron (giving the current landmark direction), defining the new *landmark-azimuth-elevation* triplet. Hence, the learning rule of a merging neuron k can be summarized as follow (the ω^{LM}_{jk} are initially null): $\Delta\omega^{LM}_{jk} = H_1(l_j(t)) \cdot R^M_k$, (idem with A (azimuth) and E (elevation) instead of L (landmark), and with θ and ϕ instead of l). The afferent synapses adapt during the *one-shot* learning (at the recruitment, $R^M_k = 1$), and do not change anymore ($R^M_k = 0$ otherwise). $H_y(x) = 1$ if $x > y$ and 0 otherwise (Heaviside function). Finally, the activity of a PC results from the computation of the distance between the learned place code and the current place code. Thus, the activity $P_k(t)$ of the kth PC can be expressed as follow:

$$P_k(t) = \frac{1}{W_k}\left(\sum_{i=1}^{N_M} \omega^P_{ik}(t).m_i(t)\right) \qquad (3)$$

where $\omega^P_{ik}(t) = 1$ expresses the fact that the triplet i (*i.e.* the ith merging neuron which activity is $m_i(t)$) has been used to encode the PC k. The number of triplets used by the kth PC is given by $W_k = \sum_{i=1}^{N_M} \omega_{ik}$ with $\omega_{ik} \in \{0, 1\}$, and with N_M the number of recruited neurons in the *what* and *where* tensor. The *one-shot* learning rule of a PC k is computed as follow (the ω^P_{ik} are initially null): $\Delta\omega^P_{ik} = H_1(m_i(t)) \cdot R^P_k$. They adapt during the *one-shot* learning (at the recruitment, $R^P_k = 1$), and do not change anymore ($R^P_k = 0$ otherwise).

A remarkable property of the system relies in its built-in generalization capability: a PC coding for the location A responds when the robot is precisely in A but also to a lesser degree in the neighborhood of A, creating a continuous and large place field around A. Experiments in the next section will illustrate several examples of place fields (fig. 3). The learning of several locations creates overlapping place fields and also leads to the paving of the space when the learning of

new locations is triggered by the detection of low PC activities (according to a given threshold). A predictable mathematical consequence of the *what* and *where* merging is the following: the shape of the place field is homothetic with the shape of the environment [22] (*i.e* the place fields extend with the distance to the landmarks). The prediction is verified in the experiment of the fig. 4 (large outdoor environment). On fig. 4, we can see that in an outdoor environment, the place fields are really larger than in an indoor environment. For instance, in fig. 4, the place fields have a useful diameter of about 25 m., which is almost the size of the environment.

As regard to the problem of the size of the world representation introduced in section II, our system exhibits a real interest. The system builds its own metrics based on the parallax and the recognition of the landmarks. Hence, the dimensionality of the internal representation is not given by the metric size of the explored area but rather by its visual regularity (*i.e.*: if the distance to the landmark is infinite, the world will be describe by means of a single PC).

IV. ROBUSTNESS OF THE PLACE RECOGNITION.

There are two important points to guaranty the robustness of our place recognition algorithm: 1) the addition of a soft competition allowing several interpretations of an extracted local view, 2) the use of a short term memory to select over a given temporal period the most pertinent *landmark-azimuth-elevation* occurrences. The built-in generalization capabilities of our place recognition system is largely improved.

A. Soft Competition at the level of the landmark recognition.

A first approach to classify an extracted local view is to suppose that the closest learned landmark is recognized. However, in two adjacent learned places, the same focus point can be the center of two different learned landmarks. Hence, the extracted local view has two correct interpretations and both should be accepted. Allowing a single interpretation for each local view is equivalent to impose the worth spatial localization error (equivalent to occlusion of the landmarks) to all the other interpretations even if they can be more or less valid. This dilemma is unsolvable if the landmark neuron group is a simple WTA (Winner Takes All) allowing a single interpretation for each local view. Moreover, the distance between learned prototypes shrinks with the number of encoded landmarks. Accordingly, the identification errors become more and more frequent with a strict competition (classical problem of clustering). We can wonder about the necessity to assign a single label to each local view. Due to the inherent ambiguity of the sensorial information, its interpretation ought to be fuzzy. As only the global behavior of the system matters [23], [24], instead of trying to perform an impossible choice, allowing multiple interpretations of the same local view seems to bring a lot of advantages.

In order to allow several interpretations of the same local view, fixing a recognition threshold is not a satisfying solution since an optimal value does not exist. However, a safety recognition threshold can be estimated to reduce the noise

effects (RT in equation 1). Another simple solution is to fix a maximum number of interpretations over this safety threshold RT. As it is impossible to assign a single label to a local view, the system is allowed to produce some mistakes if it provides at least a given number of valid interpretations. This kind of coding is well managed in a competitive network when the noise induced by the mistakes equally disturbs all the PCs. Fig. 3 illustrates the effect of a strict competition vs. a soft competition on the level of the landmark recognition. The robot had learned five aligned places in a working room of 5.4×7.2 m and went over the line to compute PC activities at each position. Generalization is largely increased by the soft competition. Using a strict competition prevents the place fields from overlapping (see fig. 3a). A fuzzy interpretation of the visual cues allows place fields to extent very far from the center of the learned places, without reducing the precision in their centers (fig. 3b).
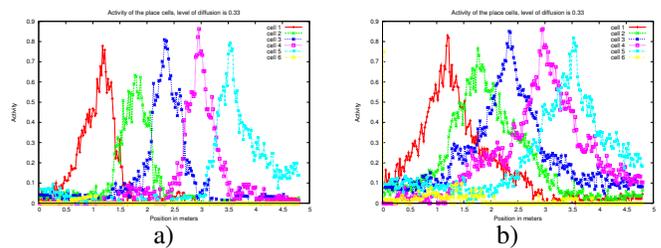

a)   b)

Fig. 3. Place fields induced by a strict (a) or a soft (b) competition (indoor env.). PC activities are computed every 2 cm over a line of 4.8 m long. Places have been learned every 60 cm. Fig. a shows place fields induced by a strict competition. Fig. b shows place fields induced by a soft competition. A strict competition at the level of the landmark recognition does not allow a good generalization and place fields to overlap.

B. Short Term Memory.

As opposed to our previous results in structured indoor environments, outdoor place field formation led to highly unstable results (see fig. 4b). The mechanism to control the focus of the attention is unable to guaranty the focus on a stable number of learned landmarks (low probability to retrieve them and therefore to recognize them). Step after step, variability of pertinent focus points was so high that even at the learned position, the PC activity could be very low. Interpolating the maxima of the curves or averaging the curves should allow to recover the shape of the traditional place fields. Even though the selection of the analyzed local views is very unstable due to the strict competition between a large number of focus points, the visual system seems able to extract pertinent landmarks. Since the extraction frequency of the pertinent landmarks is unknown, allowing the system to remind their occurrences for a while should help to promote the stability of the place fields. Practically, rather than resetting the whole place code at the beginning of each visual exploration, reset could be delayed until a given number of images. During, this period, if a more pertinent triplet is extracted, the value of the triplet should be

updated. Hence the following computation of the reset signal is proposed:

$$r_k^{Q_{max}}(t) = [1 - H_1(Q_k^{Q_{max}}(t))]^+$$

with

$$Q_k^{Q_{max}}(t) = max(\quad [Q_k(t-dt) - T(t)]^+ \\ , Q_{max}.H_0\left(m_k(t) - m_k(t-dt)\right) \quad)$$

This reset signal depends on each triplet *landmark-azimuth-elevation*, and occurs if the neuron has not been over-activated since Q_{max} visual explorations. $T(t)$ is a binary signal firing at the beginning of the visual exploration of a new panorama. $T(t)$ was previously the reset signal. Now, the computation of a PC relies on the number of pertinent visual cues that has been extracted during Q_{max} visual explorations. Despite a temporal hysteresis when $Q_{max} > 1$, the use of an extended short term memory (STM) allows place fields to become more stable and to exhibit a wide spatial generalization (see fig. 4c). Hence, the proposed place recognition system enables the robot to localize itself providing a robust measure of localization decreasing regularly over a large extent.
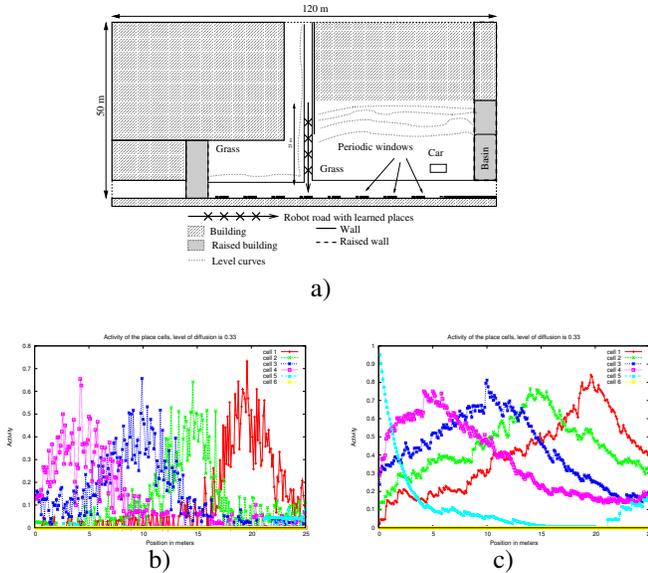

a)

b)   c)

Fig. 4. Place fields without (b) and with (c) extended STM (outdoor env. of fig. a). PC activities are computed every 10 cm over a line of 25 m long. Places have been learned every 5 m. Soft competition is used at the level of the landmark recognition. Fig. b shows highly noisy place fields. Fig. c illustrates the interest of a STM. In outdoor environment, the useful diameter of the place fields is about 25 m.

V. ROBUST SENSORY-MOTOR NAVIGATION

Several approaches may be adopted to go back to a location. Models inspired by the insects like the bees or the desert ant cataglyphis suggest to minimise the parallax between a learned location and the current location [2], [3]. Unless a cognitive map is added [25], this approach is difficult to use in a real dynamic environment. Another approach could be the use a hill-climbing algorithm on the place recognition
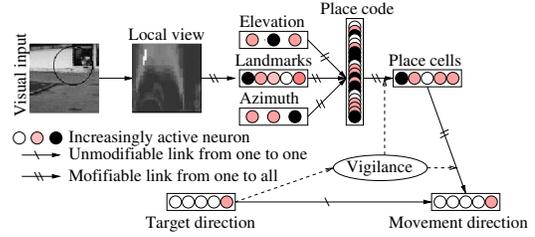

Fig. 5. PerAc architecture is able to create an attractive behavior to the goal. The Groups *target direction* and *movement direction* are WTA. Each new learned place is associated with a direction of movement which is triggered when the robot recognize the place

level of a goal cell (a particular PC). Unfortunately, even if the robot could keep a direction as long as the recognition level increases, a strong initialization problem occurs each time a new action has to be chosen. Finally, the noise on the place recognition level (our PC activity) can induce local maxima, and the duration of each movement represents a critical parameter for the convergence of such an algorithm. These limitations can be overcome by incorporating our PCs in a PerAc (Perception-Action) architecture [5] as illustrated in figure 5. Each PC is associated with a movement to trigger when being recognized. If the PCs and the actions are defined in the frame of a competitive structure, a minimum of three place-action associations around a goal creates an attraction basin, enabling the robot to return to the goal from each place in the attraction area (see fig. 6). Learning is equivalent to excavate this basin [24]. Step after step, the robot reacts according to the learned sensory-motor associations, as a ball rolling deeper and deeper in a valley. An arbitrary traject without cross points (a loop for instance) can also be learned (see fig. 8).
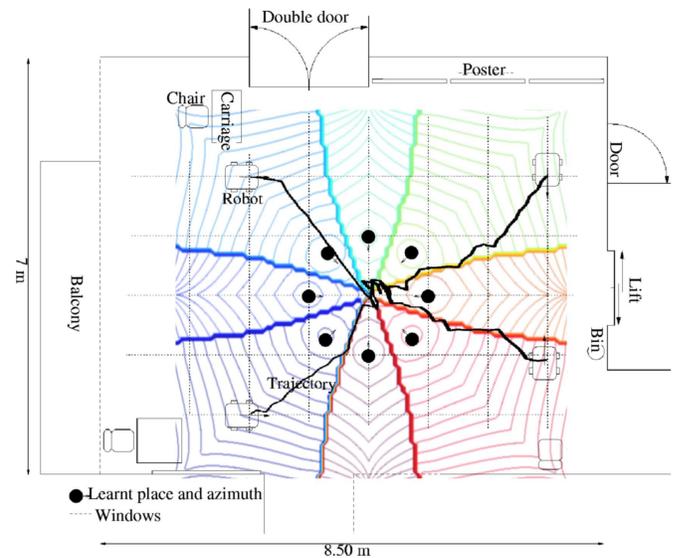

Fig. 6. Real trajectories of homing in indoor environment with an omni-directional camera. 8 places (black circles) are learnt at 1 m from the goal (size of the square on the floor). The theoretical place fields are superposed with the plan and the trajectories.

A. Robust navigation in changing environments

Problems of robustness to environmental changes are a major obstacle to the autonomy of navigation mechanisms like SPLAM and vision-based algorithms. Most of the technics are based on a very strong assumption: the world is static. Even incremental methods should be put at fault by environmental changes. A capital challenge for the navigation systems is to exhibits the same performance in static as well as in dynamical environments (people may hide a part of the visual field and objects can be added or removed). Accordingly, a robust multiple-landmarks-based navigation algorithm should not be designed with assumptions about the identity or the number of landmarks that can be retrieved from a place. Our previous experiments have put forward the robustness of the place recognition to environmental disturbances that equally penalize all the PCs [6]. As long as the environmental disturbances do not change the rank of the PCs in the competition, the behavior remains correct. Occlusions or displacements of few landmarks usually change the aspect and the spatial localization of the landmarks, penalizing all the PCs in the same way. Yet, since the spatial distribution of the landmarks in the different learned constellation can be non-uniform, larges occlusions of the visual field may hide a different number of pertinent landmarks for each cell. This problem has to be corrected.

We introduce here a more robust computation of the PC activities. As no assumption can be made about the number of pertinent landmarks in the current place code, distance should not take into account all the points of the learned place code but rather a given ratio. As compared to the less activated triplets in the current *landmark-azimuth-elevation* constellation, the most activated triplets in the constellation have a higher probability to be pertinent. Hence, the first maxima terms of the sum of the equation 3 should be sufficient (and more pertinent) to define a PC robust to severe inhomogeneous displacements or occlusions of landmarks. PC activities are now given by the following equation:

$$P_k(t) = \frac{\sum_{m=1}^{W_k \cdot \rho_L} \max_{i=1..N_M}^m \left(\omega_{ik}^P . m_i(t)\right)}{W_k \cdot \rho_L} \quad (4)$$

with ρ_L the ratio of pertinent landmarks necessary for the PC computation (25% of the landmarks in the following experiment because 50% of the visual field can be occluded), and with $\max_{i \in I}^m (x_i)$ computing the mth maximum in $\{x_i, \forall i \in I\}$. W_k is the number of learned triplets in the location k.

In the experiment of fig. 7, the robot learns 4 aligned places and goes along the line formed by the places (as in the previous experiments, fig. 4a). During learning, a whole panoramic image is used. In order to evaluate the performance of the place recognition algorithm using this new equation, half the panoramic image is occluded after the robot has reached the intermediate position. With the previous PCs model ($\rho_L = 100\%$), the occlusions of the visual field induce the place fields to collapse (see fig. 7a). The use of eq. 4 enables the place fields to maintain a reliable activity as long as the $\rho_L\%$ of

the triplets belonging to the learned constellation of a PC are visible (see fig. 7b).

B. Outdoor navigation without panoramic image building

As regard to the outdoor navigation task, it seems really difficult to obtain a perfect panoramic image. Indeed, if the capture plane is not horizontal, the conic mirror will induce deformations that can disturb an azimuth-based algorithm. Moreover, omni-directional camera must be calibrated. Hence, the use of a classical pan-tilt camera seems to be more accurate for outdoor navigation (a classical CCD camera with 60° of visual field mounted on a servo-motor allowing 240° of field of view). Thanks to the robustness of our new place recognition, the static reconstruction of a panoramic view is no more necessary. Images can be analyzed online since the STM of the merging layer is able to store occurrences of pertinent *landmark-azimuth-elevation* triplets which can be scattered in the time. The robot moves continuously without stopping to build a panoramic image. Hence the result of the merging of the different *what* and *where* information suffers from the azimuthal drift introduced by the permanent move of the robot. This drift has a negative impact on all the PCs. However their rank in the competition remain the same as with a static panorama, allowing the robot to select the correct sensory-motor association, making it fall in the attraction basin. To further enhance the autonomy of the robot, a simple dynamic neural field [26] merges in a fuzzy manner the obstacle avoidance behavior and the homing (or path following) behavior coming from the sensory-motor learning. The system has been intensively evaluated in dynamical indoor environments. Its robustness to kidnapping, to presence of mobile obstacles, to part of the visual field occlusion and to landmark addition or removal has been successfully confirmed. New experiments show our algorithm is efficient without any modification in complex outdoor environments (see fig. 8, sururban areas). These experiments were achieved at the beginning of september. The learning was performed around
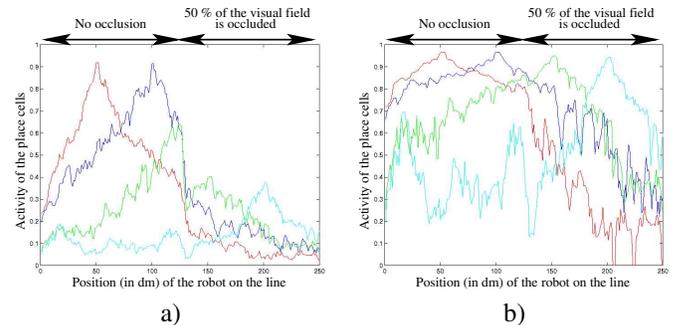

Fig. 7. Place fields (outdoor env.) with a visual occlusion of half the panorama from the intermediate position to the end. PC activities are computed every 10 cm over a line of 25 m long (see fig. 4a). Places have been learned every 5 m. Soft competition at the level of the landmark recognition, and STM are used. Fig. a shows collapsed place fields as soon as the visual field is occluded. Fig. b illustrates the interest of a computation that take into account the possibility of landmarks occlusion. The ratio of landmarks to be extracted (here, $\rho_L = 25\%$), controls the robustness to the occlusions.

Fig. 8. Outdoor environment and looped sensory-motor trajectory. Arrows represent the learned positions and the associated movements. The robot closes the loop of about 100 m in 20 mn.

5.00 pm and the performances evaluated at 10.00 am the following day. The sky was shiny and some shadows was projected on part of the environment. The robot closed the loop of about 100 m in 20 mn with a Intel Xeon 2.4 Ghz (dual processor). The system was slow because the compass measurements, the visual algorithm, the control of the robot head, and the control of the robot wheels were performed by the same sequential program. Hence, each step took about 10 s. In the new developments, these different tasks are computed in parallel by different programs which can be executed on several processors. The dynamic neural field integrates orders of obstacle avoidance every 0.05 s. The online active computation of succesive partial views of the whole panorama offers the same level of performances as the static analysis of a panorama and allows a faster update of the robot trajectory (after the analysis of each images). A elementary visual decision is made every 0.25 s. The same kind of outdoor sensory-motor learning was also achieved in a narrow suburban street with a 'S' shaping. The performances were successfully evaluated on several days of the week with a single learning. Another advantage to our sensory-motor navigation model is that the correspondence problem does not exist. Since our system is based on the assumption that the visual data define the places, the question is not to know if the current data correspond to a previously learned place but rather if the current movement is appropriate at the current recognized location. Unless the complexity of the environment induces a PC to have a multiple peak activity (local maxima of the activity), the correspondence problem while closing a loop is eluded.

VI. DISCUSSION AND CONCLUSION

This article has highlighted the interest of our map-less vision-based navigation system and the important points to guaranty the robustness for indoor as well as outdoor environments. Several biomimetic navigation systems have been developed. The *snapshot* model of insect-inspired visual navigation [2] have been intensively evaluated. The main conclusion are the following: the navigation behavior based on the parallax minimization can be simplified in a simple vectorial difference [3], that can even be learned (as opposed to a wired architecture) [27], and the use of a topoligical map seems necessary to generate a robust autonomous behavior in various environments [25]. Unfortunately, it seems the robustness to environmental changes has never been studied. Most of the models of the hippocampal system focus on the merging of idiothetic and allothetic information to providing a reliable PC coding [28], [29], highlighting the importance of the head direction cells. [28]'s model has been merged with a classical mapping algorithm (RatSLAM) and evaluated in large indoor environments. Recently, the system has been proved to generate a consistent map of a large outdoor environment with an omnidirectional visual system [30]. Map-building was performed offline due to the computation load that depends on the size of the environment. Another major contribution is the Psykarpax project [31], aiming at building an artificial rat. Even if the system could reliably navigate with a SLAM-like algorithm [32], the integration and the unification of the differents elements seems to remain a problem. As compared to the existing systems, our vision-based sensory-motor navigation system exhibits a weak computation load allowing a robot to reliably navigate in real time, in indoor [6] as well as in outdoor environments. Robustness to kidnapping, to part of the visual field occlusions, to landmark addition or removal and to presence of mobile obstacles has been shown. However, the computation and memory load linearly depends on the number of learned landmarks and the number of extracted local views. For a long term use, it will be primordial to predict the pertinent landmarks in order to select a bounded number of neurons in the competition process. Moreover, a simple sensory-motor architecture does not enable a robot to plan a trajectory according to its internal motivations. In our complete hippocampal model, neurons in the hippocampus proper (CA1/CA3 regions) learn and predict transitions between successive multimodal states. A cognitive map computes a latent learning of the spatial topology of the environment [33] and can be used to plan a sequence of actions to reach an arbitrary goal [29]. Future experiments will focus on this extended model in dynamical indoor and outdoor environments. The robustness of our visual PCs will be very important, in order to simplify the merging of allothetic and idiothetic information needed to learn and recognize complex multimodal states involved in the planning.

(movies available on http://www-etis.ensea.fr/Members/pgaussier)

APPENDIX

The upper parameters of the tab. I depend on the input image size (expect the gradient parameter) as opposed to the lower parameters. $Pano$: size of the panoramic image. $Deriche$: parameter for the Deriche gradient computation. σ_1 and σ_2

Param	Fig. 3b	Fig. 4c	Fig. 7b	Fig. 8
$Pano.$	1500.240	1500.240	750.120	9×192.144
$Deriche$	0.8	0.8	1.0	0.8
DoG: σ_1	3 px	3 px	1 px	2
DoG: σ_2	4 px	4 px	2 px	4
$DoG\ size$	15 px	15 px	5 px	10 px
$X_I \times Y_I$	32×32	32×32	16×16	16×16
$Radius$	40 px	40 px	20 px	35 px
RT	0.8	0.8	0.8	0.8
$Nb.\ Interp.$	4	4	4	3
ρ_{Az}	0.33	0.33	0.2	0.33
ρ_{El}	∞	∞	1	0.33
g^ρ	linear	linear	gaussian	gaussian
Q_{max}	0	4	4	9×2
ρ_L	1	1	0.25	0.33

TABLE I

PARAMETER DEFINITION FOR THE EXPERIMENT OF FIG. 3B, 4C, 7B, 8.

are the deviations defining the DoG filter. $DoG\ size$: size of the filtering window in pixels. $Radius$: the radius in pixels of the extent of the log polar transformation (radius of the circle of fig. 2). $Nb\ Interp.$: the number of allowed interpretations in the competition process between the landmark neurons. The other parameters are defined in the paper.

REFERENCES

[1] J. O'Keefe and N. Nadel, *The hippocampus as a cognitive map*. Oxford: Clarendon Press, 1978.

[2] B. Cartwright and T. Collett, "Landmark learning in bees," *Journal Comp. Physiology*, vol. 151, pp. 521–543, 1983.

[3] D. Lambrinos, R. Moller, T. Labhart, R. Pfeifer, and R. Wehner, "A mobile robot employing insect strategies for navigation," *Robotics and Autonomous Systems*, vol. 30, pp. 39–64, 2000.

[4] I. Bachelder and A. Waxman, "Mobile robot visual mapping and localization: A view-based neurocomputational architecture that emulates hippocampal place learning," *Neural Networks*, vol. 7, no. 6/7, pp. 1083–1099, 1994.

[5] P. Gaussier and S. Zrehen, "Perac: A neural architecture to control artificial animals," *Robotics and Autonomous System*, vol. 16, no. 2-4, pp. 291–320, December 1995.

[6] P. Gaussier, C. Joulain, J. Banquet, S. Leprêtre, and A. Revel, "The visual homing problem: an example of robotics/biology cross fertilization," *Robotics and autonomous system*, vol. 30, pp. 155–180, 2000.

[7] E. Royer, J. Bom, M. Dhome, B. Thuillot, M. Lhuillier, and F. Marmoiton, "Outdoor autonomous navigation using monocular vision. ieee/rsj iros2005," in *IEEE/RSJ IROS2005, International Conference on Intelligent Robots and Systems*, 2005, pp. 3395–3400.

[8] O. Trullier, S. Wiener, A. Berthoz, and J. Meyer, "Biologically-based artificial navigation systems: Review and prospects," *Accepted for publication in Progress in Neurobiology*, 1996.

[9] M. Franz and H. Mallot, "Biomimetic robot navigation," *Robotics and Autonomous System*, vol. 30, pp. 133–153, 2000.

[10] G. DeSouza and A. Kak, "Vision for mobile robot: a survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237–267, 2002.

[11] S. Thrun, "Learning metric-topological maps for indoor mobile robot navigation," *Artificial Intelligence*, vol. 99, no. 1, pp. 21–71, 1998.

[12] T. Duckett and U. Nehmzow, "Exploration of unknown environments using a compass, topological map and neural network," in *Proceedings of the 1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA'99)*, Monterey, CA, 1999.

[13] S. Thrun, "Robotic mapping: A survey," in *Exploring Artificial Intelligence in the New Millenium*, G. Lakemeyer and e. Nebel, B., Eds. Morgan Kaufmann, 2002.

[14] H. Andreasson, A. Treptow, and T. Duckett, "Localization for mobile robots using panoramic vision, local features and particle filter," in *IEEE International Conference on Robotics and Automation (ICRA 2005)*, Barcelona, Spain, 2005.

[15] R. Chatila, "Deliberation and reactivity in autonomous mobile robots," *Robotics and Autonomous System*, vol. 16, no. 2-4, pp. 197–211, December 1995.

[16] C. Ye and J. Borenstein, "A method for mobile robot navigation on rough terrain," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2004, pp. 3863–3869.

[17] P. Gaussier, C. Joulain, S. Zrehen, J. Banquet, and A. Revel, "Visual navigation in an open environement without map," in *International Conference on Intelligent Robots and Systems - IROS'97*. Grenoble, France: IEEE/RSJ, September 1997, pp. 545–550.

[18] C. Giovannangeli, P. Gaussier, and J. Banquet, "Robot as a tool to study the robustness of visual place cells," in *I3M'2005: International Conference on Conceptual Modeling and Simulation (CMS 2005)*, Marseille, 2005, pp. 97–104.

[19] S. Leprêtre, P. Gaussier, and J. Cocquerez, "From navigation to active object recognition," in *The Sixth International Conference on Simulation for Adaptive Behavior SAB'2000*. Paris: MIT Press, 2000, pp. 266–275.

[20] L. Schwartz, "Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding," *Vision Res.*, vol. 20, pp. 645–669, 1980.

[21] J. Banquet, P. Gaussier, M. Quoy, A. Revel, and Y. Burnod, "A hierarchy of association in hippocampo-cortical systems: cognitive maps and navigation strategies," *Neural Computation*, vol. 17, pp. 1339–1384, 2005.

[22] P. Gaussier, A. Revel, J. Banquet, and V. Babeau, "From view cells and place cells to cognitive map learning: processing stages of the hippocampal system," *Biological Cybernetics*, vol. 86, pp. 15–28, 2002.

[23] P. Gaussier, J. Baccon, K. Prepin, J. Nadel, and L. Hafemeister, "Formalization of recognition, affordances and learning in isolated or interacting animats," in *From Animals to Animats: SAB'04*. Cambridge, MA: MIT Press, 2004, pp. 57–66.

[24] M. Maillard, O. Gapenne, P. Gaussier, and L. Hafemeister, "Perception as a dynamical sensori-motor attraction basin," in *Advances in Artificial Life (8th European Conference, ECAL)*, ser. Lecture Note in ArtificialIntelligence, C. et al., Ed., vol. LNAI 3630. Springer, sep 2005, pp. 37–46. [Online]. Available: http://publi-etis.ensea.fr/2005/MGGH05

[25] V. V. Hafner, "Cognitive maps in rats and robots," *Adaptive Behavior*, vol. 13, no. 2, pp. 87–96, 2005.

[26] S. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, pp. 77–87, 1977.

[27] V. V. Hafner and R. Moller, "Learning of visual navigation strategies," in *Proceedings of the European Workshop on Learning Robots (EWLR-9)*, P. G. M. Quoy and J. W. (Eds.), Eds., 2000, pp. 47–56.

[28] A. Arleo and W. Gerstner, "Spatial cognition and neuro-mimetic navigation: A model of hippocampal place cell activity," *Biol. Cybern.*, vol. 83, no. 3, pp. 287–299, 2000.

[29] N. Cuperlier, M. Quoy, P. Gaussier, and C. Giovannangeli, "Navigation and planning in an unknown environment using vision and a cognitive map," in *IJCAI-05 Workshop: Reasoning with Uncertainty in Robotics (RUR-05)*, 2005, pp. 48–53.

[30] G. W. D. Prasser, M. J. Milford, "Outdoor simultaneous localisation and mapping using ratslam," in *International Conference on Field and Service Robots*, 2005.

[31] D. Filliat, B. Girard, A. Guillot, M. Khamassi, L. Lachze, and J. Meyer, "State of the artificial rat psikharpax," in *From Animals to Animats 8: Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior*, S. S., A. Ijspeert, A.and Billard, S. Vijayakumar, J. Hallam, and e. Meyer, J.-A., Eds. MIT Press, 2000, pp. 435–444.

[32] D. Filliat and J. Meyer, "Active perception and map learning for robot navigation," in *In from Animals to Animats 6, Proceedings of Sixth Conference on Simulation of Adaptive Behavior*. MIT Press, 2000.

[33] E. Tolman, "Cognitive maps in rats and men," *The Psychological Review*, vol. 55, no. 4, pp. 189–208, 1948.