Vision-based Navigation, Environmental Representations and Imaging Geometries

José Santos-Victor[1] and Alexandre Bernardino[1]

Instituto Superior Técnico, Instituto de Sistemas e Robótica, Lisbon, Portugal,
{jasv,alex}@isr.ist.utl.pt – http://www.isr.ist.utl.pt/vislab

Abstract. We discuss the role of spatial representations and visual geometries in vision-based navigation. To a large extent, these choices determine the complexity and robustness of a given navigation strategy. For instance, navigation systems relying on a geometric representation of the environment, use most of the available computational resources for localization rather than for "progressing" towards the final destination. In most cases, however, the localization requirements can be alleviated and different (e.g. topological) representations used. In addition, these representations should be adapted to the robot's perceptual capabilities.

Another aspect that strongly influences the success/complexity of a navigation system is the geometry of the visual system itself. Biological vision systems display alternative ocular geometries that proved successful in different (and yet demanding and challenging) navigation tasks. The compound eyes of insects or the human foveated retina are clear examples. Similarly, the choice of the particular geometry of the vision system and image sampling scheme, are important design options when building a navigation system.

We provide a number of examples in vision based navigation, where special spatial representations and visual geometries have been taken in consideration, resulting in added simplicity and robustness of the resulting system.

1 Introduction

Most of the research on vision-based navigation has been centered on the problem of building full or partial 3D representations of the environment, which are then used to drive an autonomous robot. Instead of concentrating the available resources to progress towards the goal, the emphasis is often put on the process of building (or using) these 3D maps. This explains why many existing systems require large computational power, but still lack the robustness needed for many real-world applications. In contrast, examples of efficiency can be drawn from biology. Insects, for instance, can solve very large and complex navigation problems in real-time [1], in spite of their limited sensory and computational resources.

One striking observation in biology is the diversity of "ocular" geometries. Many animals eyes point laterally, which may be more suitable for navigation purposes. The majority of insects and arthropods benefit from a wide field of view and their eyes have a space-variant resolution. To some extent, the performance of these animals is related to their specially adapted eye-geometries.

Similarly, one possibility to explore the advantages of having large fields of view in robotics is to use *omni-directional cameras*.

Studies of animal navigation suggest that most species utilize a very parsimonious combination of perceptual, action and representational strategies that lead to very efficient solutions when compared to those of today's robots.

Both robustness and an efficient usage of computational and sensory resources can be achieved by using visual information in closed loop to accomplish specific navigation tasks or behaviors [2,3]. However, this approach alone cannot deal with global tasks or coordinate systems (e.g. going to a distant goal), because it lacks adequate representations of the environment. Hence, a challenging problem is that of extending these local behaviors, without having to build complex 3D representations of the environment.

At this point, it is worth discussing the nature of the navigation requirements when covering long distances, as compared to those for short paths. Many animals, for instance, make alternate use of landmark-based navigation and (approximate) route integration methods [1]. For example, to walk along a city avenue, position accuracy to within one block is sufficient. However, entering our hall door would require much more precise movements.

This *path distance/accuracy* tradeoff between long-distance/low-precision and short-distance/high-accuracy mission segments plays an important role in finding efficient solutions for robot navigation.

In the following sections we discuss how different imaging geometries and environment representations can be used for improving the navigation capabilities of an autonomous system.

2 Imaging geometries

In this section we discuss two aspects of the imaging geometry. Firstly, we consider the case of omni-directional cameras whose enlarged fields of view can be advantageous for navigation. Then, we detail the log-polar mapping which is a space-variant image sampling scheme, similar to those found in natural seeing systems. Finally, a combination of both camera and image sensor design is introduced.

2.1 Omni-directional Vision

Omni-directional cameras provide a 360° view of the robot's environment and have been applied to autonomous navigation, video conferencing and surveillance [4] -[7], among others. Omni-directional images are usually obtained with a combination of cameras and convex mirrors. Mirror shapes can be conic, spherical, parabolic or hyperbolic [7].

Visual landmarks are easier to find with omni-directional images, since they remain in the field of view much longer, than with a conventional camera. The imaging geometry has various properties that can be exploited for

navigation or recognition. For example, vertical lines in the environment are viewed as radial image lines (see Fig.3).

Our omnidirectional system [8] combines a camera and a spherical mirror, mounted on top of a mobile platform, as shown in Fig. 1.

Fig. 1. Left: omni-directional camera. Center: camera mounted on the mobile robot. Right: camera (spherical mirror) projection geometry - symmetry about the z-axis simplifies the geometry.

The geometry of image formation is obtained by equaling the incidence and reflection angles on the mirror surface. The resulting mapping relates the coordinates of a 3D point, $\mathbf{P}$, to the coordinates of its projection on the mirror surface, $\mathbf{P_m}$, and finally, to its image projection $\mathbf{p}$, as in Fig. 1.

2.2 Space variant (log-polar) sampling

Foveated active visual systems are widely present in animal life. A representation of the environment with high-resolution and a wide field of view is provided through the existence of the space-variant ocular geometry and the ability to move the eyes.

The most common space-variant image representation is the log-polar mapping, introduced in [9], due to its similarity to the retinal resolution and organization on the visual cortex of primates. The log-polar transformation is a conformal mapping from points on the *cartesian* plane $\mathbf{x} = (x, y)$ to points in the *cortical* plane $\mathbf{z} = (\xi, \eta)$ [9], as shown in Fig. 2. The log-polar mapping is described by :

$$[\xi, \eta]^t = \left[\log(\sqrt{x^2 + y^2}), \arctan\frac{y}{x}\right]^t \qquad [x, y]^t = \left[\mathrm{e}^{\xi}\cos\eta, \mathrm{e}^{\xi}\sin\eta\right]^t$$

The application of the log-polar mapping to artificial vision was first motivated by its perceptually based data compression capabilities. When compared to cartesian images, log-polar images allow faster sampling rates without reducing the size of the field of view and the resolution on the central part of the retina (fovea). In addition to rotation and scale invariance [10], the

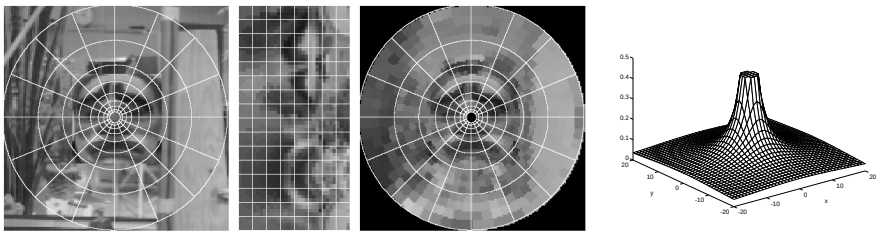

Fig. 2. The log–polar transformation maps points in the cartesian (far left) to the cortical planes (middle left). The effective image resolution becomes coarser in the periphery, as shown in the retinal plane (middle right). The log-polar mapping implements a focus of attention in the center of the field of view, equivalent to a weighting window in cartesian coordinates (far right).

log-polar geometry provides additional algorithmic benefits: easy computation of time-to-contact [3,11], increased stereo resolution on verging systems and good disparity selectivity for vergence control [12,13].

2.3 Omnidirectional vision and Space variant sampling

Both omnidirectional cameras and the log-polar sensor have a rotational symmetry, which suggests the combination of both. As a result, rather than getting the usual omni-directional images, a so-called panorama can be directly obtained by reading out the image pixels. As an additional benefit, the angular resolution is constant, as the log-polar geometry is based upon circular rings with a constant number of pixels (see Fig. 3). The joint mirror profile

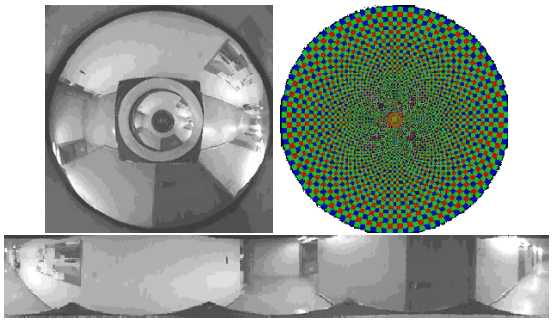

Fig. 3. Combination of omnidirectional images and a log-polar sensor (top) yields directly a constant resolution panorama (bottom).

and sensor layout design was addressed in the EU project Omniviews.

3 Environmental representations

In this section we discuss various environmental representations for navigation. We will focus on *alternatives* to the traditional geometric maps. First, we discuss the use of topological maps. Then, we shall see how to use various forms of image-based (local) representations. Finally, we mention visual servoing as an implicit local representation of the environment.

3.1 Topological Maps

Topological Maps [14,15,8] can be used to travel long distances in the environment, without demanding accurate control of the robot position along a path. The environment is represented by a graph. *Nodes* correspond to recognizable *landmarks*, where specific actions may be elicited, such as entering a door or turning left. *Links* are associated with regions where some environmental structure can be used to control the robot (see Section 3.3).

Landmarks are directly represented by *images* and a map is thus a collection of inter-connected images (Fig. 4). Precise metric information is not necessarily required to go from one particular locale to another. For example, to get from the city center, *Rossio*, to *Saldanha*, we may *go forward* until we reach the statue in *Rotunda*, *turn right* in the direction of *Picoas* and carry on until we finally reach *Saldanha* Square. The navigation problem is decom-

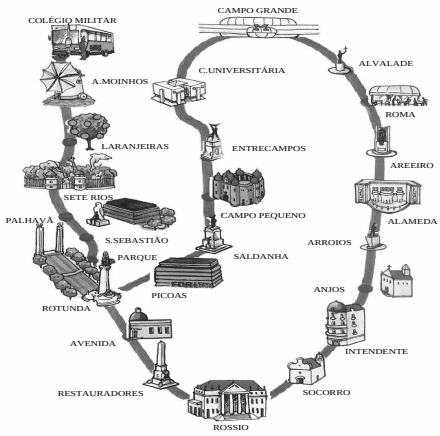

Fig. 4. A topological map of touristic landmarks in Lisbon, Portugal.

posed into a succession of sub-goals, identified by recognizable landmarks. The required navigation skills consist of following roads, making turns and recognizing landmarks.

3.2 Panoramas, Bird's eye views and mosaics

Images acquired with an omni-directional camera are distorted, when compared to those of a perspective camera. For instance, a corridor appears as an image band of variable width. However, the image formation model can be used to correct some distortions, yielding Panoramic images or Bird's Eye Views, which may serve as local image-based environment representations that facilitate tracking or feature extraction.

Scan lines of panoramic images contain the projections of all visible points at constant angles of elevation. Hence, the unwarping consists of mapping concentric circles to lines [16]. The horizon line is actually transformed to a scan line and vertical 3D lines are mapped as vertical image lines.

Bird's eye views are obtained by radial correction around the image center[1], corresponding to a scaled orthographic projection of the ground plane. For example, corridors appear as image bands of constant width, simplifying the navigation system. Image panoramas and bird's eye views are illustrated in Fig. 5.

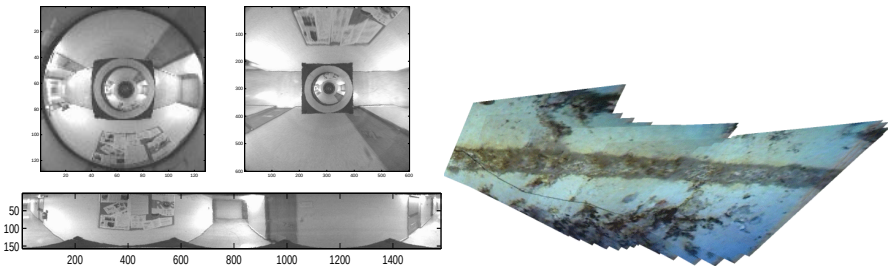

Fig. 5. Left: omni-directional image, the corresponding bird's eye view and the panoramic image. Right: Video mosaic of the sea bottom.

When the camera motion undergoes pure rotation or when the observed scene is approximately planar Video Mosaics constitute interesting representations. Video mosaics can be built by accurately registering images acquired during the camera motion. They offer high resolution and large fields of view, and can serve as visual maps for navigation [18]. Figure 5 shows a video mosaic of the sea-bottom[2].

3.3 Local structure (servoing)

Visual servoing can also be interpreted as yet another form of visual representation. The goal of (image-based) visual servoing consists in reaching

[1] Hicks [17] obtained ground plane unwarped images directly, with the use of a custom-shaped mirror.

[2] Work developed in the EU ESPRIT-LTR Project 30185 NARVAL

desired configurations of image features. As such configurations, (uniquely ?) constrain the camera pose with respect to the work space, they can be considered as an implicit camera-environment representation, rather than describing their (world) coordinates explicitly.

4 Examples of Navigation and Vision based Control

In this section we give various examples of visual based navigation and control. All the different examples explore certain camera/image geometries and specific representations of the environment.

4.1 Topological maps and image eigenspaces

When using a *topological map* to describe the robot's *global* environment, a mission can be specified as: "*go to the third office on the left-hand side of the second corridor*".

The topological map consists of a large set of reference (omni-directional) images acquired at pre-determined positions (landmarks), connected by links in a graph. During operation, the reference image that best matches the current view indicates the robot's *qualitative* position.

Reference images can be interpreted as points in a high-dimensional space, each indicating a possible reference position of the robot. As the number of images required to represent the environment can be very large, we build a lower-dimensional linear subspace approximation using Principal Component Analysis (PCA), [19].

Figure 6 shows the first 3 principal components (eigenimages) computed from 50 omni-directional images in one corridor, shown in descending order in accordance with their eigenvalues.

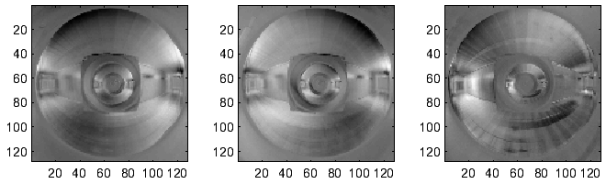

Fig. 6. The first 3 eigenimages obtained with the omni-directional vision system.

A "distance", d_k, between the current view and the set of reference images can be computed in real-time using the projections in the eigenspace. The position of the robot is that associated with the reference image, $\mathbf{I}_k$ having the lowest distance, d_k.

Omni-directional images help dealing with relatively *dynamic* environments, where people partially occlude the robot's view. Even when a person

is very close to the robot, the occlusion is not sufficiently large so as to cause the robot to misinterpret its topological position.

We have built a topological map from omni-directional images, acquired every 50 cm, along corridors. Reference positions were ordered according to the direction of motion, thus maintaining a causality constraint.

We acquired a set of prior images, P, and ran the robot in the corridor to acquire a different set of run-time images, R. Figure 7 shows the distance d_k, between the prior and run-time images.

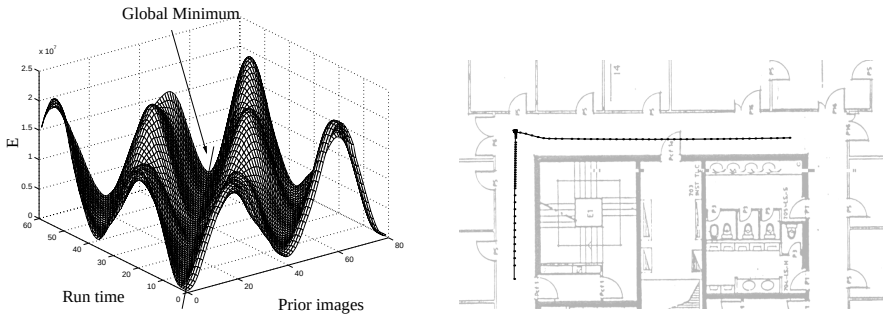

Fig. 7. Left: A 3D plot of the error (d_k) between images acquired at run time, R versus those acquired a priori, P. This plot represents the traversal of a single corridor. The global minimum is the estimate of the robot's topological position. Right: One of the paths traveled by the robot

The error surface presents a *global* minimum, corresponding to the correct estimate of the robot's topological position, and degrades in a piecewise smooth manner. Spurious local minima are due to distant areas of the corridor that may look similar to the robot's current position and can be avoided by restricting the search space to a neighborhood of the current position estimate. Figure 7 shows results obtained when driving the robot along a corridor. The distance traveled was approximately 21 meters. Odometry was used to display the path graphically.

4.2 Servoing on local structure

To navigate along the topological graph, we have to define a suitable vision-based behavior for corridor following (*links* in the map). In different environments, knowledge about the scene geometry can be used to define other behaviors. In this section we provide examples on how to explore local image structure for servoing the robot.

Centering Behavior: The first visually guided behavior is the *centering reflex*, described in [20] to explain the behavior of honeybees flying within

two parallel "walls". The qualitative visual measure used is the difference between the image velocities computed over a lateral portion of the left and right visual fields, [2]. The ocular geometry (*Divergent Stereo*) was an early attempt to use wide fields of view images for navigation [2].

The robot control system involves two main loops. The *Navigation loop* governs the robot heading in order to balance the bilateral flow fields, hence maintaining the robot at similar distances from structures on the right or left sides. The *Velocity loop* controls the robot forward speed as a function of the amplitude of the lateral flow fields. The robot accelerates in wide spaces and slows down when the environments becomes narrower.

Additionally, a *sustaining mechanism* is embodied in the control loops to avoid erratic behaviors of the robot, in the absence of (localized) flow information. It allows the use of the robot in rom-like environments or when the "walls" are not uniformly textured. Figure 8 shows the robot trajectories (from odometry) superimposed on the experimental setup, for various real-time experiments.

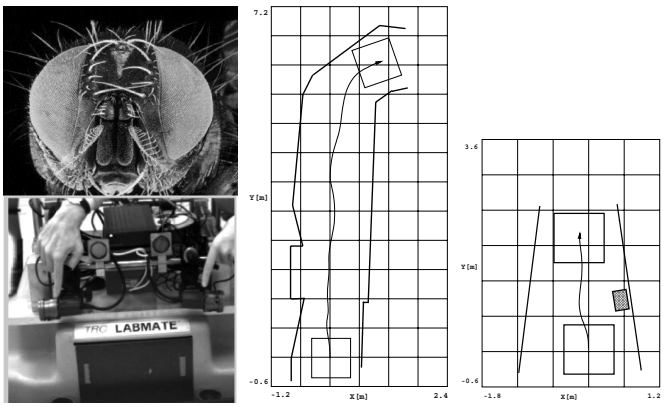

Fig. 8. Left to right: Compound eyes of insects and the divergent stereo configuration; results of the centering reflex obtained with the Divergent Stereo approach in closed loop operation for different scene layouts.

To test the velocity control, we considered the funneled corridor with varying width. As the corridor narrows down, the average flow increases and the velocity control mode forces the robot to slow down, enabling the robot to make a softer, safer maneuver.

Corridor Following with Bird's eye views: In another example, the parallelism of the corridor guidelines is used to control the robot heading direction. To simplify the servoing task, the visual feedback is provided by *Bird's eye views* of the floor, computed from omni-directional images.

Tracking the corridor guidelines is done with *bird's eye* (orthographic) views of the ground plane (see Fig. 9). Projective-planar transformations, computed from differential odometric data are used to predict the position of points and lines from one image to the next.

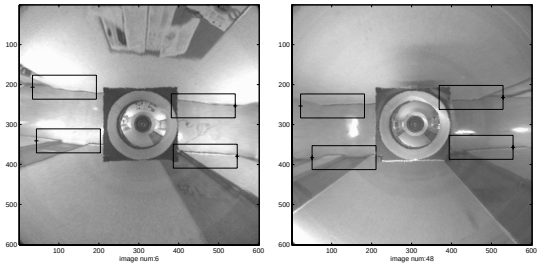

Fig. 9. Bird's eye views during tracking of the corridor guidelines.

The use of bird's eye views of the ground plane simplifies both the extraction of the corridor guidelines (the corridor has a constant width) and the computation of the robot position and orientation errors, with respect to the corridor's central path, which are the inputs of a closed loop controller.

Visual Path Following: Topological navigation can be used to travel between distant places, without accurate localization along the path. For local tasks, we rely on *Visual Path Following* when the robot must follow a reference trajectory accurately for e.g. door traversal, docking and navigation in cluttered environments.

Bird's eye views are used to track environmental features, estimate the robot's position/orientation and drive the robot along a pre-specified trajectory. Again, this geometry simplifies the tracking and localization problems.

The features used are corner points defined by the intersection of edge segments, tracked with a robust fitting procedure. Vertical lines project as radial (or vertical) lines, in the bird's eye view (or panoramic) images. Tracking is simplified by using bird's eye (orthographic) views of the ground plane, thus preserving angular measurements and uniformly scaling distances.

Figure 10 illustrates tracking and localization while traversing a door into a room. The tracked features (shown as black circles) are defined by vertical and ground-plane segments, in bird's eye view images. The robot position and orientation (in the image) are estimated with an Extended Kalman filter and used to control the robot's angular velocity [8] for trajectory following.

Figure 10 shows tracking and localization while following a reference trajectory, relative to a visual landmark composed of two rectangles. The figure shows the mobile robot at the final position after completion of the task. The processing time is about 0.4 sec/image.

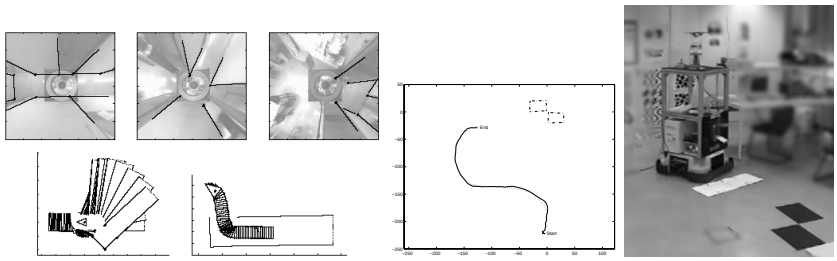

Fig. 10. Far left (clockwise) Feature tracking at three instants (black circles); estimated scene model and self-localization results. Far right: Visual Path Following results: Dash-dotted line shows the landmark. The dotted line is the reference trajectory, specified in image coordinates, and the solid line shows the filtered position estimates; robot at the final position

Tracking with log-polar images: The *Medusa* binocular head is an active tracking system shown in Fig. 11, running at video rate (25 Hz) without any special processing hardware. The mapping from 128x128 cartesian to 32x64 log-polar images takes about 3 ms, which is highly compensated by the reduction achieved (8 times) in the remaining computations.

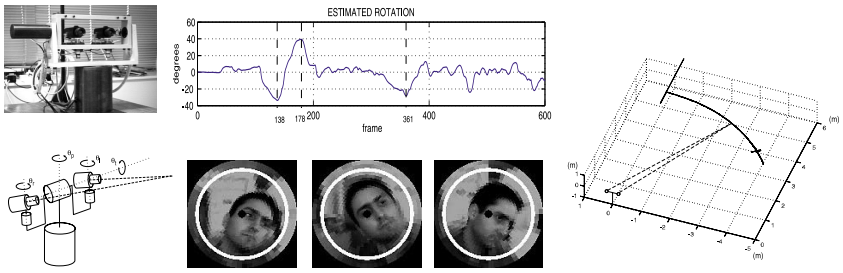

Fig. 11. Left: the Medusa stereo head with four joints (camera vergence, pan and tilt). Middle: estimated rotation in a tracking sequence (top). The frames shown below correspond to the notable points signaled in the plot, associated to local maxima in target motion. Right: Target trajectory measured with stereo

We developed tracking algorithms for objects undergoing general parametric 2D motions, using log-polar images [13]. Figure 11 shows results during a face tracking experiment with a model containing both translation and rotation. Similar results are obtained for more complex motions.

Currently, the binocular system is placed on a static platform and is able not only to track object motion but also to measure its distance and orientation relative to the camera system (see Fig. 11). In future work we intend to place the robotic head on a mobile platform where, by tracking objects, navigation behaviours like following or avoiding objects would be

possible. Also fixating static objects in the environment can be important for several navigation tasks [21], like ego-motion estimation or path planning.

4.3 Topological and local navigation

The following experiment integrates global and local navigation tasks, combining *Topological Navigation* and *Visual Path Following*. Figure 12 shows an image sequence of the robot during the entire run.

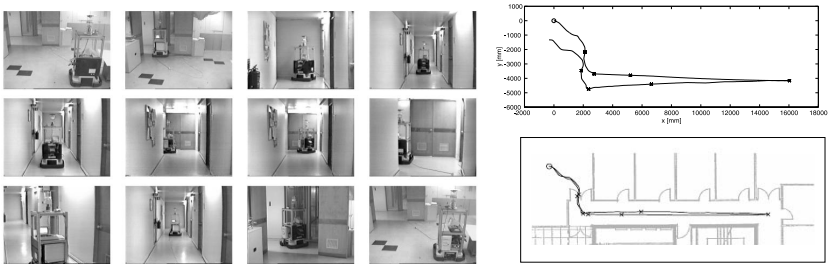

Fig. 12. Left: sequence of images of an experiment combining visual path following for door traversal and topological navigation for corridor following. Right: same type of experiment with showing the trajectory estimated from odometry (top) and the true one (bottom).

The mission starts in the VisLab. Visual Path Following is used to navigate inside the room, traverse the door and drive the robot out into the corridor. Then, control is transferred to the topological navigation module, which drives the robot all the way to the end of the corridor. At this position a new behavior is launched, consisting of the robot executing a 180^o turn, after which the topological navigation mode drives the robot back to the lab. entry point. Once the robot is approximately located at the lab. entrance, control is passed to the Visual Path Following module, which locates appropriate visual landmarks and drives the robot through the door. It follows a pre-specified path until the final goal position, well inside the lab., is reached.

Figure 12 shows the robot trajectory during one experiment, and its estimate using odometry. When returning to the laboratory, the uncertainty in odometry is approximately 0.5m. Thus, door traversal would not be possible without the use of visual control.

4.4 Mosaic Servoing

In another example we have used video mosaics as a map for the navigation of an underwater vehicle [18]. The vehicle pose is estimated with respect to the mosaic, thus allowing us to control its trajectory towards a desired configuration. Figure 13 shows estimates of the camera pose over time when the underwater vehicle is swimming over an area covered by the mosaic.

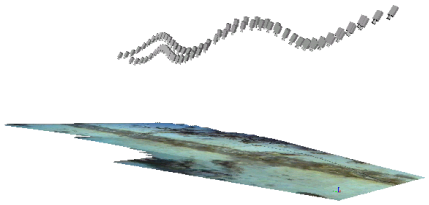

Fig. 13. Using video mosaics for navigation.

5 Conclusions

We discussed the fundamental issues of environmental representations and ocular/image geometries, when addressing the problem of visual based navigation.

In spite of all the recent progress in autonomous (visual) navigation, the performance of today's robots is still far from reaching the efficiency, robustness and flexibility that we can find in biology (e.g insect vision).

In biology, the diversity of ocular/image geometries seem to be extremely well adapted to the main tasks to handle by each species. Additionally, experiments show that spatial representations seem to be remarkably efficient.

Similarly, these aspects are determinant to the success of artificial visual navigation systems. Rather than being defined in advance, spatial representations and the geometry of the vision system should be at the core of the navigation system design. Different navigation methods, eye geometries and environmental representations should be used for different problems, with distinct requirements in terms of processing, accuracy, goals, etc.

In terms of eye geometries, we discussed omni-directional cameras and space variant image sampling. We gave examples of different environmental representations, including topological maps, image-based descriptions (panoramas, bird's eye views, mosaics) and local servoing structures.

Examples of vision based navigation and control were presented to illustrate the importance of the choice of eye geometry and spatial representation. In our opinion, studying eye/image geometries and spatial representations in artificial systems is an essential step, both for understanding biological vision systems and for designing truly flexible and robust autonomous systems.

Acknowledgments The authors would like to acknowledge the contribution of José Gaspar, Niall Winters and Nuno Gracias for the research described in this paper.

References

1. R. Wehner and S. Wehner, "Insect navigation: use of maps or ariadne's thread?," *Ethology, Ecology, Evolution*, vol. 2, pp. 27–48, 1990.

2. J. Santos-Victor, G. Sandini, F. Curotto, and S. Garibaldi, "Divergent stereo in autonomous navigation : From bees to robots," *Int. J. Computer Vision*, vol. 14, no. 2, pp. 159–177, 1995.
3. J. Santos-Victor and G. Sandini, "Visual behaviors for docking," *CVIU*, vol. 67, no. 3, September 1997.
4. Y. Yagi, Y. Nishizawa, and M. Yachida, "Map-based navigation for mobile robot with omnidirectional image sensor COPIS," *IEEE Trans. Robotics and Automation*, vol. 11, no. 5, pp. 634–648, 1995.
5. L. J. Lin, T. R. Hancock, and J. S. Judd, "A robust landmark-based system for vehicle location using low-bandwidth vision," *Robotics and Autonomous Systems*, vol. 25, pp. 19–32, 1998.
6. K. Kato, S. Tsuji, and H. Ishiguro, "Representing environment through target-guided navigation," in *Proc. ICPR*, 1998, pp. 1794–1798.
7. S. Baker and S. K. Nayar, "A theory of catadioptric image formation," in *Proc. ICCV*, 1998, pp. 35–42.
8. J. Gaspar, N. Winters, and J. Santos-Victor, "Vision-based navigation and environmental representations with an omni-directional camera," *IEEE Trans. on Robotics and Automation*, vol. 16, no. 6, pp. 890–898, Dec.r 2000.
9. E. Schwartz, "Spatial mapping in the primate sensory projection : Analytic structure and relevance to perception," *Biol. Cyb.*, vol. 25, pp. 181–194, 1977.
10. C. Weiman and G. Chaikin, "Logarithmic spiral grids for image processing and display," *Comp Graphics and Image Proc*, vol. 11, pp. 197–226, 1979.
11. C. Capurro, F. Panerai, and G. Sandini, "Dynamic vergence using log-polar images," *IJCV*, vol. 24, no. 1, pp. 79–94, August 1997.
12. A. Bernardino and J. Santos-Victor, "Binocular visual tracking: Integration of perception and control," *IEEE Trans. Rob. Automation*, vol. 15(6), Dec. 99.
13. A. Bernardino, J. Santos-Victor, and Giulio Sandini, "Foveated active tracking with redundant 2d motion parameters," *Robotics and Autonomous Systems*, June 2002.
14. B. Kuipers, "Modeling spatial knowledge," *Cognitive Science*, vol. 2, pp. 129–153, 1978.
15. R. Brooks, "Visual map making for a mobile robot," in *IEEE Conf. Robotics and Automation*, 1985.
16. J. S. Chahl and M. V. Srinivasan, "Reflective surfaces for panoramic imaging," *Applied Optics*, vol. 36, no. 31, pp. 8275–8285, 1997.
17. A. Hicks and R. Bajcsy, "Reflective surfaces as computational sensors," in *IEEE Workshop on Perception for Mobile Agents, CVPR 99*, 1999, pp. 82–86.
18. N. Gracias and J. Santos-Victor, "Underwater video mosaics as visual navigation maps," *CVIU*, vol. 79, no. 1, pp. 66–91, July 2000.
19. H. Murase and S. K. Nayar, "Visual learning and recognition of 3D objects from appearance," *Int. J. Computer Vision*, vol. 14, no. 1, pp. 5–24, 1995.
20. M.V. Srinivasan, M. Lehrer, W.H. Kirchner, and S.W. Zhang, "Range perception through apparent image speed in freely flying honeybees," *Visual Neuroscience*, vol. 6, pp. 519–535, 1991.
21. C. Fermüller and Y. Aloimonos, "The role of fixation in visual motion analysis," *IJCV*, vol. 11, no. 2, pp. 165–186, October 1993.