Visual navigation of indoor/outdoor nonholonomic robots with wide field-of-view cameras

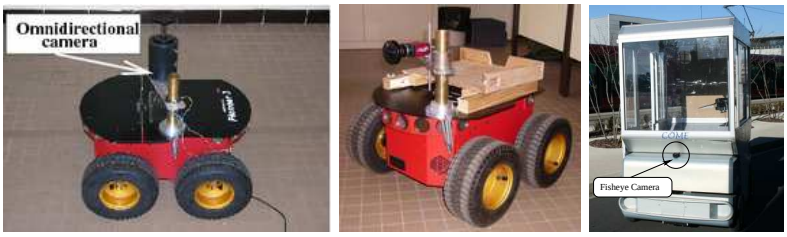
J. Courbon[1,2] and Y. Mezouar[1] and L. Eck[2] and P. Martinet[1]

[1] LASMEA, Aubiere F-63177, France
[2] CEA-List, Fontenay-Aux-Roses F-92265, France

Abstract. In this paper, we present results of a complete framework for non-holonomic robot navigation in indoor and outdoor environments using a single wide field-of-view camera. The proposed navigation framework for wheeled robots is based on a visual memory containing key images to reach. During a human-guided learning step, the robot performs paths which are sampled and stored as a set of ordered key images, acquired by an embedded camera. The set of these obtained visual paths is topologically organized and provides a visual memory of the environment. Given an image of one of the visual paths as a target, the robot navigation mission is defined as a visual route. When running autonomously, the control guides the robot along the reference visual route without explicitly planning any trajectory. The control consists in a control law adapted to the nonholonomic constraint and directly computed from visual points. The proposed framework has been designed for a generic class of cameras. In this paper, experiments have been carried on with catadioptric and fisheye cameras, in indoor environment with a AT3 Pioneer robot and in outdoor environment with an autonomous urban vehicle.

1 Introduction

Often used among more "traditional" embedded sensors - proprioceptive sensors like odometers as exteroceptive ones like sonars - vision sensor provides accurate localization methods. The authors of [1] accounts of twenty years of works at the meeting point of mobile robotics and computer vision communities. In many works, a map of the environment and the robot localization in this absolute frame are simultaneously updated. Both motion planning and robot control can then be designed in this space. The results obtained by the authors of [2] leave to be forcasted that such a framework will be reachable using a single camera. However, although an accurate global localization is unquestionably useful, our aim is to build a complete vision-based framework without recovering the position of the mobile robot with respect to a reference frame. In [1] this type of framework is ranked among qualitative approaches.
The principle of this approach is to represent the robot environment with a



(a) (b) (c)

Fig. 1. Three applications: navigation of a Pioneer AT3 equipped with a catadioptric camera (a) and with a fisheye camera (b) and navigation of an urban vehicle (c).

bounded quantity of images gathered in a set called visual memory. In the context of mobile robotics, [3] proposes to use a sequence of images, recorded during a human teleoperated motion, and called View-Sequenced Route Reference. This concept underlines the close link between a human-guided learning and the performed paths during an autonomous run. However, the automatic control of the robot in [3] is not formulated as a visual servoing task. In [4], *homing* strategy is used to control a wheelchair from a memory of omnidirectional images but the control of this holonomic robot is not part of the presented framework.
In the proposed framework, the control design directly takes into account the non-holonomic model of the robot and is computed from the feature matching. Panoramic views acquired by large field-of-view cameras are well adapted to this approach since they provide a large amount of visual features which can be exploited as well as for localization than for visual servoing.
The aim of this paper is to present different experimental validation. The concept of visual memory is briefly explained in Section 2 and more details can be found in [5]. The Section 3 deals with the vision-based control scheme designed to control the robot motions along a visual route using large field-of-view images. Finally, in Section 4, experiments on a Pioneer AT3 mobile robot using catadioptric and fisheye cameras (refer to Fig. 1 (a), (b)) and on an urban vehicle equipped with a fisheye camera (refer to Fig. 1 (c)) illustrate the proposed framework.

2 Vision-based memory navigation (VBMN) strategy

Our approach can be divided in three steps 1) visual memory building, 2) localization into the visual memory, 3) navigation into the visual memory (refer to Fig. 2).

2.1 Visual Memory Structure

The learning stage relies on the human experience. The user guides the robot along paths where the robot is authorized to go. Only some key views are kept

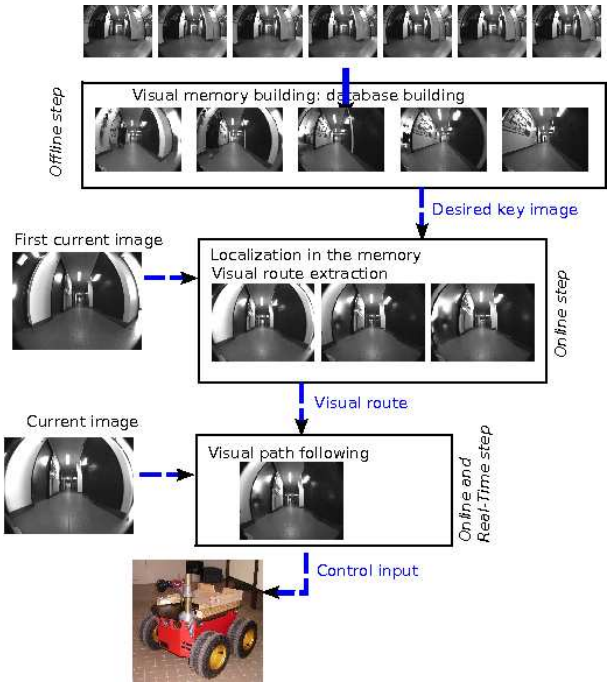


Fig. 2. Principle of our VMBN approach

and form the visual memory. The key image selection is done as detailled in the sequel. The considered visual features are points. As proposed in [2] and successfully applied for the metric localization of autonomous robots in outdoor environment, interest points are detected in each image with Harris corner detector and matched by computing a Zero Normalized Cross Correlation score. In our experiment, 500 points are detected in each image. The first image of the sequence acquired during the learning step is selected as the first key frame $\mathcal{I}_1$. A key frame $\mathcal{I}_{i+1}$ is then chosen so that there are as many frames as possible between $\mathcal{I}_i$ and $\mathcal{I}_{i+1}$ while there are at least $\mathcal{M}$ common interest points tracked between $\mathcal{I}_i$ and $\mathcal{I}_{i+1}$. From this selection, it results a visual path Ψ which is a directed graph composed of n successive key images (*vertices*): $\Psi = \{\mathcal{I}_i | i = \{1, 2, \ldots, n\}\}$. For control purpose, two hypothesis are supposed to be verified: 1) the authorized motions during the learning stage are assumed to be limited to those of a car-like robot, which only goes forward and 2) two successive key images $\mathcal{I}_i$ and $\mathcal{I}_{i+1}$ contain a set $\mathcal{P}_i$ of matched visual features, which can be observed along a path performed between ${}^R\mathcal{F}_i$ and ${}^R\mathcal{F}_{i+1}$ and which allows the computation of the control law.
Finally, the visual memory consits on a set of visual paths connected (multi-graph).

2.2 Visual route

A visual route describes the robot's mission in the sensor space. Given the images $\mathcal{I}_s^*$ and $\mathcal{I}_g$, a visual route is a set of key images which describes a path from $\mathcal{I}_s^*$ to $\mathcal{I}_g$. $\mathcal{I}_s^*$ is the closest key image to the current image of the robot determined during a localization step. This step consists in finding the image of the memory which best fits the current image acquired by the embedded camera. The last step of the framework is to follow this visual route.

3 Routes following using an omnidirectional camera

$\mathcal{I}_c$ is the current image and $\mathcal{I}_{i+1}$ is the next key image of the path to reach. The hand-eye parameters (*i.e.* the rigid transformation between $\mathcal{F}_c$ and the frame attached to the camera) are supposed to be known. The vehicle is supposed to locally navigate in a planar surface. Let us note $\mathcal{F}_{i+1} = (O_{i+1}, \mathbf{X_{i+1}}, \mathbf{Y_{i+1}}, \mathbf{Z_{i+1}})$ the frame attached to the robot when $\mathcal{I}_{i+1}$ was stored and $\mathcal{F}_c = (O_c, \mathbf{X_c}, \mathbf{Y_c}, \mathbf{Z_c})$ a frame attached to the robot in its current location (refer to Fig. 3). The origin O_c of $\mathcal{F}_c$ is the origin of the control frame of the robot.

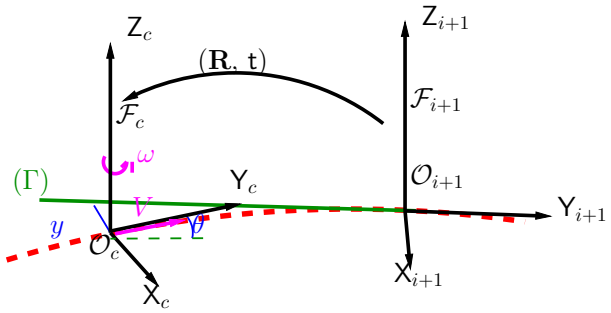


Fig. 3. The frame $\mathcal{F}_{i+1}$ along the trajectory (Γ) is the frame where the desired image $\mathcal{I}_{i+1}$ was acquired. The current image $\mathcal{I}_c$ is situated at the frame $\mathcal{F}_c$.

3.1 Principle and control law design

The control strategy consists in guiding $\mathcal{I}_c$ to $\mathcal{I}_{i+1}$ by regulating asymptotically the axle $\mathbf{Y}_c$ on the straight line $(\Gamma) = (O_{i+1}, \mathbf{Y}_{i+1})$ (refer to Fig. 3). The control objective is achieved if $\mathbf{Y}_c$ is regulated before the origin of $\mathcal{F}_c$ reaches the origin of $\mathcal{F}_{i+1}$. The longitudinal velocity (respectively the angular velocity) of the robot is V (respectively ω). Let y be the distance between O_c and (Γ) and θ the angular error between the current direction of the vehicle and the desired direction. As

proposed in [6], an asymptotically stable guidance control law based on the chain system approach can be designed to achieve this goal:

$$\omega(y,\theta) = -V\cos^3\theta K_p y - |V\cos^3\theta| K_d \tan\theta \tag{1}$$

As long as the robot longitudinal velocity V is non zero, the performances of path following can be determined in terms of settling distance [7] that is to say in terms of the distance to travel before reaching the desired position. K_p and K_d are two positive gains which set the performances of the control law depending on the settling distance. The lateral and angular deviations of $\mathcal{F}_c$ with respect to (Γ) can be obtained through partial Euclidean reconstructions as described in Section 3.2.
For robot relying on the Ackermann's model (bicycle model) as the RobuCab vehicle, the control law can be obtained using the same approach.

3.2 State estimation from the unified model of camera on the sphere

A classical model for central catadioptric cameras is the unified model on the sphere [8]. It has been shown in [9, 10] that this model is also suitable for fisheye cameras in robotic applications.
The point in the image plane corresponding to the 3D point $\mathcal{X}$ of coordinates $\mathsf{X} = [X\ Y\ Z\ ]^T$ is obtained after a projection on a virtual unit sphere, followed by a perspective projection on the normalized image plane $Z = 1-\xi$ and a plane-to-plane collineation [8] (refer to Figure 4) where the parameter ξ describes the type of sensor. The homogeneous coordinates $\underline{\mathsf{x}}_i$ of this image point is

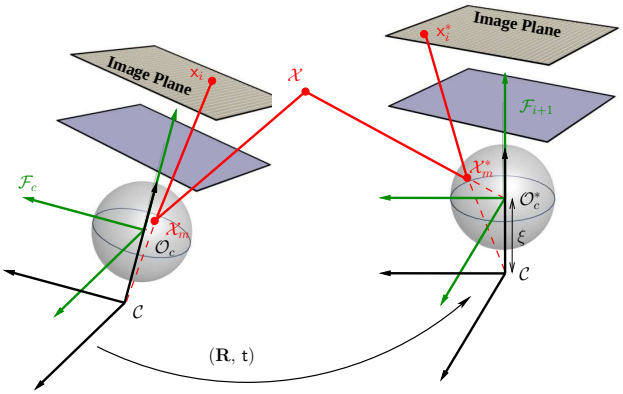


Fig. 4. Geometry of two views using the unified model on the sphere.

$$\underline{\mathsf{x}}_i = \mathbf{K}_c\mathbf{M}\begin{bmatrix}\dfrac{X}{Z+\xi\|\mathsf{X}\|} & \dfrac{Y}{Z+\xi\|\mathsf{X}\|} & 1\end{bmatrix} \tag{2}$$

$\mathbf{K}_c$ contains the usual intrinsic parameters of the perspective camera and $\mathbf{M}$ contains the parameters of the frames changes.

We notice that the coordinates X_m of the projection on the sphere can be computed as a function of the coordinates in the image x and the sensor parameter ξ:

$$\mathsf{X}_m = (\eta^{-1} + \xi)\bar{\mathsf{x}} \tag{3}$$

$$\bar{\mathsf{x}} = \begin{bmatrix} \mathsf{x}^T & \dfrac{1}{1+\xi\eta} \end{bmatrix}^T \text{ with: } \begin{cases} \eta = \dfrac{-\gamma - \xi(x^2+y^2)}{\xi^2(x^2+y^2)-1} \\ \gamma = \sqrt{1+(1-\xi^2)(x^2+y^2)} \end{cases}$$

Depending on the context, a scale euclidean reconstruction is computed using the homography matrix when 3D points are coplanar or using the essential matrix otherwise as detailled in the sequel.
Let $\mathcal{X}$ be a 3D point with coordinates $\mathsf{X_c} = [X_c\ Y_c\ Z_c\ ]^T$ in the current frame $\mathcal{F}_c$ and $\mathsf{X}^* = [X_{i+1}\ Y_{i+1}\ Z_{i+1}\ ]^T$ in the frame $\mathcal{F}_{i+1}$. This point is projected onto the unit spheres into the points $\mathcal{X}_m$ and $\mathcal{X}_m^*$ (refer to Fig. 4). We suppose that the camera is calibrated.

Scaled Euclidean reconstruction from planar 3D points Let consider that the point $\mathcal{X}$ belongs to a plane (π). After some algebric manipulation, we obtain:

$$\bar{\mathsf{x}} \propto \mathbf{H}\bar{\mathsf{x}}^* \tag{4}$$

where $\mathbf{H}$ is the Euclidean homography matrix relative to the plane (π), function of the camera displacement and of the plane coordinates in $\mathcal{F}_{i+1}$. As usual, the homography related to (π) can be estimated up to a scale factor with at least four couples of points. From the $\mathbf{H}$-matrix, the camera motions parameters (the rotation matrix $\mathbf{R}$ and the scaled translation $\mathsf{t_d}* = \frac{\mathsf{t}}{d^*}$) and the structure of the observed scene can be estimated (for more details refer to [11]).

Scaled Euclidean reconstruction from non planar points When considering non-planar 3D points, the epipolar geometry is used. The epipolar plane contains the projection centers O_c and O_{i+1} and the 3D point $\mathcal{X}$. The points of coordinates $\mathsf{X_m}$ and X_m^* clearly belong to this plane which is traduced by the relation:

$$\mathsf{X_m}^T\mathbf{R}(\mathsf{t} \times \mathsf{X_m^*}^T) = \mathsf{X_m}^T\mathbf{R}\,[\mathsf{t}]_\times\,\mathsf{X_m^*}^T = 0 \tag{5}$$

where $\mathbf{R}$ and t represent the rotational matrix and the translational vector between the current and the desired frames. Similarly to the case of pinhole model, the relation (5) can be written:

$$\mathsf{X_m}^T\mathbf{E}\mathsf{X_m^*}^T = 0 \tag{6}$$

where $\mathbf{E} = \mathbf{R}\,[\mathsf{t}]_\times$ is the essential matrix [12]. The essential matrix $\mathbf{E}$ between two images is estimated using five couples of matched points as proposed in [13].

From the essential matrix, the camera motion parameters (that is the rotation $\mathbf{R}$ and the translation $\mathtt{t}$ up to a scale) can be determined.

Finally, the estimation of the input of the control law (1), *i.e* the angular deviation θ and the lateral deviation y, are computed straightforwardly from $\mathbf{R}$ and $\mathtt{t}$.

4 Experimentations

For our experiments, cameras have been calibrated using the Matlab toolbox presented in [14]. The parameters of the rigid transformation between the camera and the robot control frames are roughly estimated. Grey level images are acquired at a rate of 15 fps. A learning stage has been conducted off-line and images have been memorized as proposed in Section 2.

4.1 Experimentation with a catadioptric camera and planar 3D points in indoor environment

The proposed framework is implemented on an external standard PC which wireless controls a Pioneer AT3 robot. A catadioptric camera is embedded on the robot and its principal axis is approximately confounded with the rotation axis of the robot (refer to Fig. 1 (a)). Three key views (refer to Fig. 5) have been selected to drive the robot from its initial configuration to the desired one. For this experiments, the positions of four planar points are memorized and then tracked. The control is realized using the homography matrix from the projection of the patterns onto the equivalence sphere. Note that distances from the optical center at the desired position to the reference plane have been overestimated and that the directions of the normals of the plane are roughly estimated. The results of the experimentation (refer to Fig. 6) show that the lateral and the angular errors are regulated to zero before reaching a key image.

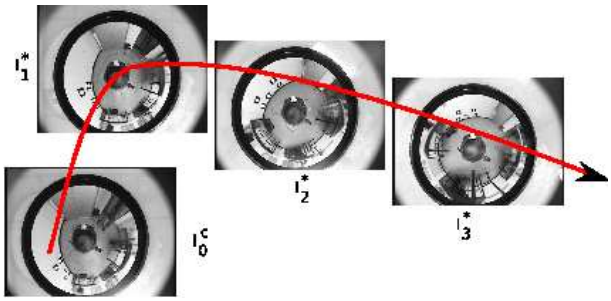


Fig. 5. Initial image I_c^0 and desired images the robot has to reach $I_j^*, j = 1:3$ (1st experimentation).

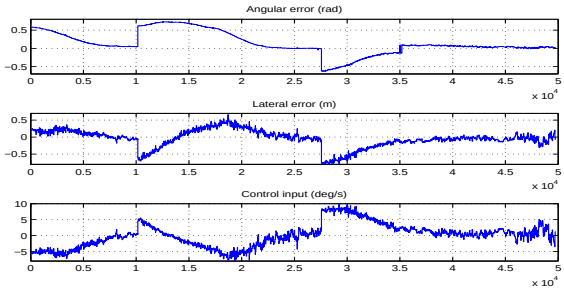


Fig. 6. Evolution of the lateral (in m) and the angular (in rad) errors and of the control input (angular speed in deg/s) during the experimentation (1st experimentation).

4.2 Experimentation with a fisheye camera in indoor environment

The Pioneer AT3 robot is now equiped with the Fujinon fisheye lens mounted onto a Marlin F131B camera . The camera providing a field of view of 185 deg and looking forward, is situated at approximately 30 cm from the ground. Several paths have been memorized (some of the images are shown in Fig. 7). The robot starts indoor and ends outdoor and the camera grabs images with natural landmarks. Given a goal image, a visual path has been extracted. At each frame, points are extracted from the current image and matched with the desired key image. A robust partial reconstruction is then applied using the current, desired and the former desired images of the memory. Angular and lateral errors are extracted and allow the computation of the control law (2). A key image is supposed to be reached when one of the "image errors" is smaller than a fixed threshold. In our experiment, we have considered two "image errors": the longer distance between an image point and its position in the desired key image (errImageMax) and the mean distance between those points (errPoints), expressed in pixels. The longitudinal velocity V has been fixed to $200\,mms^{-1}$. The gains K_p and K_d have been set in order that error presents a double pole located at value 0.3. For safety, the absolute value of the control input is bounded to 10 degrees by second. Lateral and angular errors as well as control input are represented in Fig. 8. Red crosses are plotted when key images change. As in the first experimentation, those errors are well regulated to zero for each key view. The image errors (expressed in pixels) are also decreasing before reaching the key views (refer to Fig. 9). Errors still remain different to zero because the current image do not reached exactly the desired image. As it can be noticed in Fig. 10, our method is robust to changes in the environment. A man was going in the direction of the robot (at the left) during the manually driven step whereas a man is walking at the right of the Pioneer AT3 robot during the autonomous navigation.

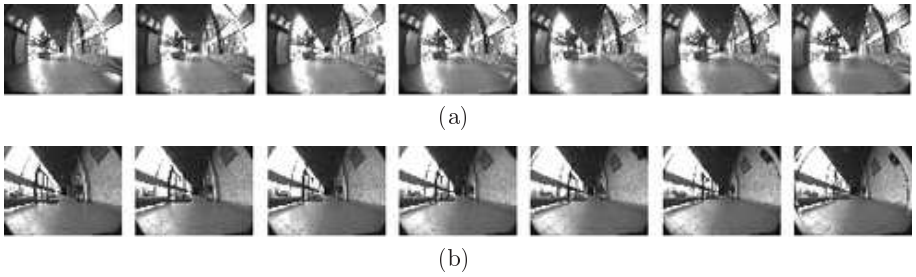


Fig. 7. Parts of the visual path to follow (2nd experimentation).

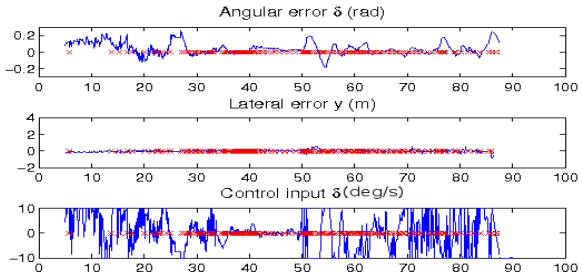


Fig. 8. Lateral y and angular θ errors and control input δ vs time (2nd experimentation)

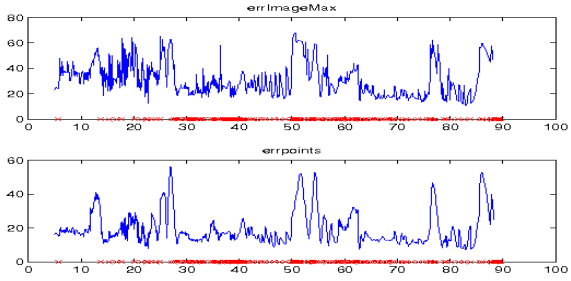


Fig. 9. Image errors: (errImageMax) and (errPoints) vs time (2nd experimentation)

4.3 Experimentation with a fisheye camera in outdoor environment

Our framework is now applied to the navigation of an urban electric vehicle, named RobuCab. The same fisheye camera as previously, looking forward, is situated at approximately 80 cm from the ground. This vehicle is manually driven along the 800-meter-long path shown in blue in Fig. 11. This path contains important turns as well as ways down and up and a come back.

After the selection step, 800 key images are kept and form the visual memory of the vehicle. The longitudinal velocity V is fixed between $1\,ms^{-1}$ and $0.4\,ms^{-1}$

(a) (b)

Fig. 10. The image (a) corresponds to the reached key image (b) of the visual memory (2nd experimentation).

depending on the position on the path to follow (straight lines or turns). The experiment lasts 13 minutes for a path of 754 meters which results to a mean velocity of $0.8\,ms^{-1}$. A mean of 123 robust matching for each frame has been found. The mean computational time during the online navigation was of 82 ms by image. The errors in the images decrease to zero until reaching a key image (refer to Fig. 12).

Lateral and angular errors as well as control input are represented in Fig. 13. As it can be noticed, those errors are well regulated to zero for each key view excepted when high turns occur. Our control law (line reaching) is not able to converge quickly in those cases. Significant errors are thus obtained during the large turns but errors are then decreasing. In future works, we plan to improve our control law to manage more efficiently the navigation in large turns.

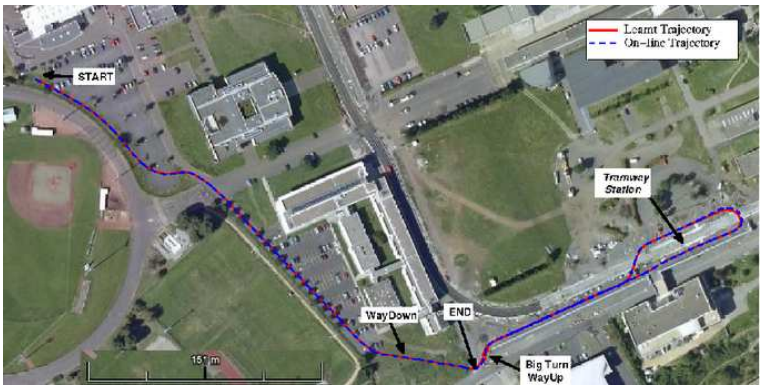


Fig. 11. Paths in the universitary campus executed during the memorization step (in red) and the autonomous step (in blue) (3rd experimentation).

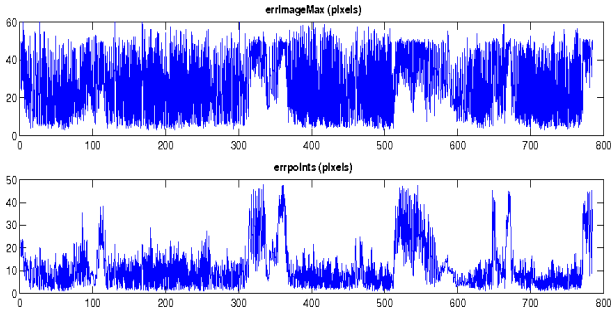


Fig. 12. Errors in the images vs time (3rd experimentation).

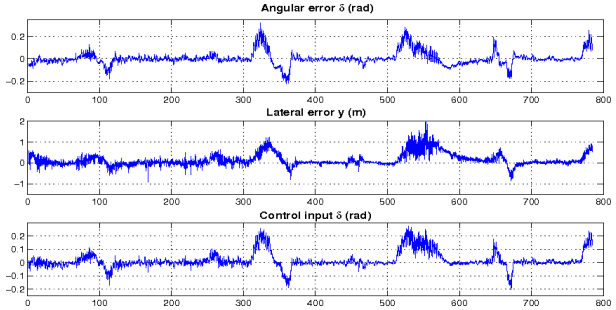


Fig. 13. Lateral y and angular θ errors and control input δ vs time (3rd experimentation).

5 Conclusion

In this paper an image-based navigation framework dedicated to nonholonomic mobile robots has been presented. The approach is illustrated in the context of indoor/outdoor environment using a single wide field of view camera and natural landmarks. We propose to learn the environment as a graph of visual paths, called visual memory. A visual route is made of a sequence of key images of the environment which describes, in the sensor space, an admissible path for the robot. This visual route can be performed thanks to a visual-servoing control law, which is adapted to the robot nonholonomy.

Future works will be devoted to relax the staticity constraint of the environment. We will try to analyse and to take into account environment modifications, which may occur between learning steps and autonomous runs, in both visual route building and following.

References

1. DeSouza, G.N., Kak, A.C.: Vision for mobile robot navigation: A survey. IEEE transactions on pattern analysis and machine intelligence **24**(2) (february 2002)

237–267
2. Royer, E., Lhuillier, M., Dhome, M., Lavest, J.M.: Monocular vision for mobile robot localization and autonomous navigation. International Journal of Computer Vision, special joint issue on vision and robotics **74** (2007) 237–260
3. Matsumoto, Y., Inaba, M., Inoue, H.: Visual navigation using view-sequenced route representation. In: IEEE International Conference on Robotics and Automation, ICRA'96. Volume 1., Minneapolis, Minnesota, USA (April 1996) 83–88
4. Goedemé, T., Tuytelaars, T., Vanacker, G., Nuttin, M., Gool, L.V., Gool, L.V.: Feature based omnidirectional sparse visual path following. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Canada (August 2005) 1806–1811
5. Courbon, J., Lequievre, L., Mezouar, Y., Eck, L.: Navigation of urban vehicle: An efficient visual memory management for large scale environments. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'08, Nice, France (September 2008)
6. Blanc, G., Mezouar, Y., Martinet, P.: Indoor navigation of a wheeled mobile robot along visual routes. In: IEEE International Conference on Robotics and Automation, ICRA'05, Barcelona, Spain (2005) 3365–3370
7. Thuilot, B., Bom, J., Marmoiton, F., Martinet, P.: Accurate automatic guidance of an urban electric vehicle relying on a kinematic GPS sensor. In: 5th IFAC Symposium on Intelligent Autonomous Vehicles, IAV'04, Instituto Superior Técnico, Lisbon, Portugal (July 5-7th 2004)
8. Geyer, C., Daniilidis, K.: Catadioptric projective geometry. In: International Journal of Computer Vision. Volume 43. (2001) 223–243
9. Barreto, J.: A unifying geometric representation for central projection systems. Computer Vision and Image Understanding **103** (2006) 208–217 Special issue on omnidirectional vision and camera networks.
10. Courbon, J., Mezouar, Y., Eck, L., Martinet, P.: A generic fisheye camera model for robotic applications. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'07, San Diego, CA, USA (29 October - 2 November 2007) 1683–1688
11. Faugeras, O., Lustman, F.: Motion and structure from motion in a piecewise planar environment. International Journal of Pattern Recognition and Artificial Intelligence **2**(3) (1988) 485–508
12. Svoboda, T., Pajdla, T.: Epipolar geometry for central catadioptric cameras. International Journal of Computer Vision **49**(1) (2002) 23–37
13. Nistér, D.: An efficient solution to the five-point relative pose problem. Transactions on Pattern Analysis and Machine Intelligence **26**(6) (2004) 756–770
14. Mei, C., Benhimane, S., Malis, E., Rives, P.: Constrained multiple planar template tracking for central catadioptric cameras. In: British Machine Vision Conference, BMVC'06, Edinburgh, UK (2006)