Visual Control for Memory-Based Navigation using the Trifocal Tensor

H. M. Becerra and C. Sagues

Abstract— **In this paper, we present a control scheme for visual path-following of wheeled mobile robots based on a robust geometric constraint: the trifocal tensor (TT). The proposed control law only needs one element of the TT as feedback information, which is computed from the current and the target images along the sequence of the visual path. The scheme is valid for images captured by cameras having approximately a unique center of projection, e.g., conventional, central catadioptric and some fisheye cameras. The benefits of the proposed scheme are that explicit pose parameters decomposition is not required and the rotational velocity is smooth or eventually piece-wise constant avoiding discontinuities that generally appear when a new target image must be reached. Additionally, the translational velocity is adapted as required for the path. The validity and performance of the approach is shown through realistic simulations using synthetic images.**

I. Introduction

The locomotion of most of service robots is based on a wheeled platform, so that, the strategies to improve their navigation capabilities result of great interest for the robotics research community. In this sense, machine vision has shown good advantages as sensor for robot control (visual servoing) [1] and navigation [2]. In this paper, we present a control scheme for following a visual path, which is extracted from a *visual memory*. The proposed scheme uses the trifocal tensor (TT) as feedback information.

The concept of visual memory means that there is a learning stage in which a set of target images are stored. These key images define the path to be replayed in an autonomous stage. This strategy has been introduced for omnidirectional images in [3]. Recently, there are contributions toward the development of autonomous vehicles under this approach. Some of them are position-based approaches, in which, a 3D reconstruction is carried out either using an EKF-based SLAM [4] or a structure from motion algorithm through bundle adjustment [5]. A complete map building is avoided in [6] by relaxing to a local Euclidean reconstruction from the essential matrix. In visual control, image-based approaches generally offer a faster closed loop control with good performance. The work in [7] propose a qualitative visual navigation scheme that is based on some heuristic rules. An approach that uses the centroid of the abscissas of the feature points for feedback is presented in [8]. Most of the mentioned approaches suffer the problem of generating discontinuous rotational velocities when a new key image must be reached. This problem is tackled in [9] for conventional cameras, where the authors propose a varying reference instead of constant. The epipolar geometry has been used as feedback information for mobile robot navigation [10]; however, this geometric constraint is less robust than the TT and it presents the problem of short baseline.

The proposed control scheme uses the value of one element of the trifocal tensor as the only required feedback information. Thus, this approach does not require explicit pose parameters estimation unlike position-based schemes [4], [5]. In our proposal, the visual servoing problem is transformed in a reference tracking problem for the selected tensor element. It avoids the recurrent problem of discontinuous rotational velocity at key image switching of image-based schemes [6], [7], [8]. The use of the TT allows the gathering of many visual features into a single measurement, so that, undesired pseudoinverse of matrices is not needed. The TT as visual measurement improves the robustness of the control scheme against image noise [9] and it avoids the problem of short baseline [10]. As used in our approach, the TT gives the possibility of taking into account valuable a priori information that is available in the visual memory and that is not exploited in previous image-based approaches. We use this information to adapt the translational velocity and also achieve piece-wise constant rotational velocity according to the taught path. Additionally, the proposed scheme can be applied not only to conventional cameras but also to any vision systems having approximately a unique center of projection, which increase the applicability of the proposal.

The paper is organized as follows. Section II introduces the robot and camera model, and the TT for generic cameras. Section III details the proposed control strategy. Section IV presents the performance of the visual navigation via realistic simulations using synthetic images and finally, Section V provides the conclusions.

II. Mathematical Modeling

A. Robot Kinematics

Let $\chi = (x, y, \phi)^T$ be the state vector of a differential drive robot (Fig. 1(a)), where $x(t)$ and $y(t)$ are the robot position coordinates in the plane, and $\phi(t)$ is the robot orientation. The kinematic model of the robot expressed in state space can be written as follows:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} -\sin\phi & 0 \\ \cos\phi & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix} \tag{1}$$

being $v(t)$ and $\omega(t)$ the translational and angular input velocities, respectively.

This work was supported by project DPI 2009-08126.

H. M. Becerra is with Centro de Investigación en Matemáticas (CIMAT), Jalisco S/N, Col. Valenciana, 36240, Guanajuato, Gto, Mexico bhecerra@gmail.com

C. Sagues is with Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, C/ María de Luna 1, E-50018 Zaragoza, Spain csagues@unizar.es

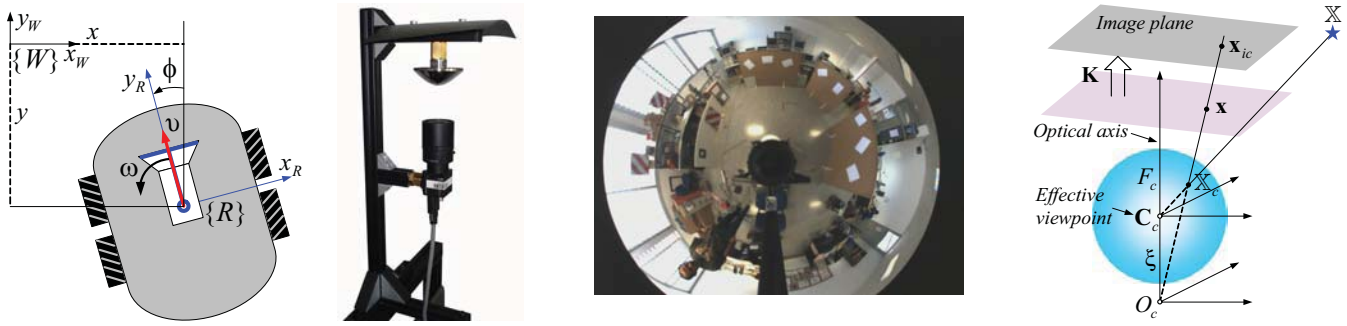


Fig. 1. Representation of the robot model and the camera model. (a) Robot frame definition. (b) Example of central catadioptric vision system. (c) Example of an image captured by a catadioptric system. (d) Generic camera model of central cameras.

B. The Trifocal Tensor (TT) for Generic Cameras

The constrained field of view of conventional cameras can be enhanced using wide field of view imaging systems such as fisheye cameras or full view omnidirectional cameras like the one shown in Fig. 1(b-c). It is known that the imaging process performed by conventional and catadioptric cameras can be modeled by a unique representation [11]. Such unified projection model works properly for imaging systems having approximately a single center of projection. This representation allows the computation of a geometric constraint, like the trifocal tensor (TT), in the same way for any central vision system.

The unified projection model describes the image formation as a composition of two central projections. The first is a central projection of a 3D point onto a virtual unitary sphere and the second is a perspective projection onto the image plane through a collineation $\mathbf{K}$. Let denote a 3D point as $\mathbb{X}$, and its corresponding coordinates as $\mathbf{X}$. Thus, point coordinates on the sphere $\mathbf{X}_c$ can be computed from point coordinates on the normalized image plane $\mathbf{x}$ (refer to Fig. 1(d)) and the sensor parameter ξ as follows

$$\begin{aligned} \mathbf{X}_c &= \left(\eta^{-1}+\xi\right)\bar{\mathbf{x}}, \\ \bar{\mathbf{x}} &= \left[\begin{array}{cc} \mathbf{x}^T & \frac{1}{1+\xi\eta} \end{array}\right]^T, \end{aligned} \tag{2}$$

where $\eta = \frac{-\gamma-\xi\left(x^2+y^2\right)}{\xi^2\left(x^2+y^2\right)-1}$, $\gamma = \sqrt{1+\left(1-\xi^2\right)\left(x^2+y^2\right)}$. In this work, we assume that the camera is calibrated, which allows us to exploit the representation of the points on the unit sphere.

The TT encapsulates all the geometric relations between three views independently of the structure of the scene and depending nonlinearly on the motion parameters among the three views. The TT has 27 elements and it can be encapsulated using three 3×3 matrices ($\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{T}_3$). There are, therefore, 26 independent ratios apart from the common overall scaling of the matrices. In this work, we focus on the use of points as image features. Consider three corresponding points projected on the unitary sphere $\mathbf{p}$, $\mathbf{p}'$ and $\mathbf{p}''$ and expressed in homogeneous coordinates (p^1,p^2,p^3). The incidence relation between these points is given as

$$[\mathbf{p}']_\times \left(\sum_i p^i \mathbf{T}_i\right) [\mathbf{p}'']_\times = \mathbf{0}_{3\times 3},$$

where $[\mathbf{p}]_\times$ is the common skew symmetric matrix. This expression provides a set of nine equations, however, only four of them are linearly independent. Thus, seven triplets of point correspondences are needed to compute the 27 elements of the tensor. This is done by solving a singular value decomposition (SVD) problem for the set of linear equations.

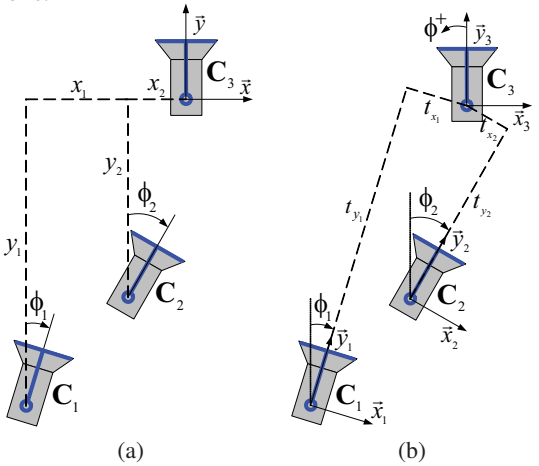


Fig. 2. Geometry between three camera locations in the plane. (a) Absolute locations with respect to a reference frame in $\mathbf{C}_3$. (b) Relative locations.

In the case in which the three cameras are located in the same plane, for instance, with the same vertical position from the ground, several elements of the tensor are zero and only 12 elements are in general non-null. Fig. 2 depicts the upper view of three cameras with global reference frame in the third view, in such a way that the corresponding locations are $\mathbf{C}_1 = (x_1, y_1, \phi_1)$, $\mathbf{C}_2 = (x_2, y_2, \phi_2)$ and $\mathbf{C}_3 = (0, 0, 0)$. Analytically, the TT can be deduced for this framework as done in [12], resulting in that the non-null elements are given as

$$\begin{aligned} T^m_{111} = -t_{x_1}\cos\phi_2 + t_{x_2}\cos\phi_1, T^m_{113} = t_{x_1}\sin\phi_2 + t_{y_2}\cos\phi_1, \\ T^m_{131} = -t_{y_1}\cos\phi_2 - t_{x_2}\sin\phi_1, T^m_{133} = t_{y_1}\sin\phi_2 - t_{y_2}\sin\phi_1, \\ T^m_{212} = -t_{x_1}, T^m_{221} = t_{x_2}, T^m_{223} = t_{y_2}, T^m_{232} = -t_{y_1}, \\ T^m_{311} = -t_{x_1}\sin\phi_2 + t_{x_2}\sin\phi_1, T^m_{313} = -t_{x_1}\cos\phi_2 + t_{y_2}\sin\phi_1, \\ T^m_{331} = -t_{y_1}\sin\phi_2 + t_{x_2}\cos\phi_1, T^m_{333} = -t_{y_1}\cos\phi_2 + t_{y_2}\cos\phi_1, \end{aligned} \tag{3}$$

where $t_{x_i} = -x_i \cos\phi_i - y_i \sin\phi_i$, $t_{y_i} = x_i \sin\phi_i - y_i \cos\phi_i$ for $i = 1, 2$ and the superscript m states that they are the tensor elements given by metric information. In practice, the estimated tensor has an unknown scale factor and this factor changes as the robot moves. We can fix a common scale during the navigation by normalizing each element of the tensor as follows:

$$T_{ijk} = \frac{\mathbf{T}^e_{ijk}}{T_N}, \tag{4}$$

where T^e_{ijk} are the estimated TT elements obtained from point matches, T_{ijk} are the normalized elements and T_N is a suitable normalizing factor, which must be different than zero. We can see from (3) that T_{212} and T_{232} are constant and non-null, assuming that the camera location $\mathbf{C}_1$ is different to $\mathbf{C}_3$. Therefore, any of these elements is good option as normalizing factor.

III. Control Strategy for Memory-Based Navigation

This section describes the proposed control strategy for wheeled mobile robot navigation based on the visual memory approach. First, we describe such approach briefly. Next, we present the way in which the TT is used in this approach and finally, the proposed control law is described in detail.

A. The Visual Memory

The navigation based on a visual memory consists of two stages. The first one is a learning stage where the visual memory is built. In this stage, the user guides the robot along the place where it is allowed to move. A sequence of images are stored from the onboard camera during this stage in order to get a representation of the environment. We assume that during learning, the translational velocity is never zero. From all the captured images, a reduced set is selected as key images by ensuring a minimum number of shared point features between two images. Thus, the visual memory defines a path to be replayed in the autonomous navigation stage. We assume that n key images are chosen and that these images are separated along the path in the Cartesian space by a minimum distance $d_{\min}$. For more details about the visual memory building and key images selection refer to [6].

B. The TT for Memory-Based Navigation

The TT has been exploited for the positioning of a mobile robot in [12] and [13]. In these works, both, the rotational and the translational velocities are computed from the elements of the tensor, which are driven to zero in order to accomplish the positioning task. The visual path-following problem only requires a rotational velocity to correct the deviation from the desired path. Consider that we have two images $I_1(\mathbf{K}, \mathbf{C}_1)$ and $I_3(\mathbf{K}, \mathbf{C}_3)$ belonging to the visual path and the current view of the onboard camera $I_2(\mathbf{K}, \mathbf{C}_2)$. As can be seen in Fig. 3, the element T_{221} of the TT (3) provides direct information of the lateral deviation of the current location $\mathbf{C}_2$ with respect to the target $\mathbf{C}_3$. It is worth emphasizing that the 1D TT does not provide this particular information of lateral error, so that, the 2D TT is used.

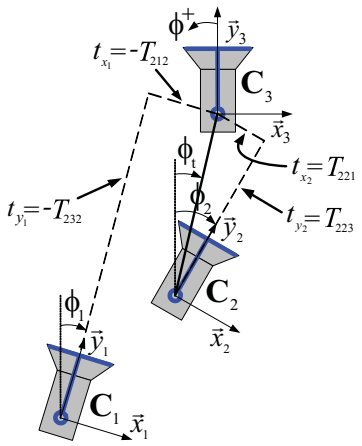


Fig. 3. The relative locations between cameras up to a scale are provided by the trifocal tensor.

Assuming that the center of projection coincides with the rotational axis of the robot, the element T_{221} of the tensor is related to the current location of the robot as follows:

$$T^m_{221} = t_{x_2} = -x_2 \cos\phi_2 - y_2 \sin\phi_2.$$

It can be seen that if $T^m_{221} = 0$,

$$\phi_2 = \phi_t = -\tan\left(\frac{x_2}{y_2}\right),$$

and consequently the current camera $\mathbf{C}_2$ is looking directly toward the target. Thus, we propose to compute the rotational velocity from feedback information given by the element T_{221}. The control goal is to drive this element with smooth evolution from its initial value to zero before reaching the next key image of the visual path. We can define a reference tracking control problem in order to avoid discontinuous rotational velocity in the switching of key images. It is also possible to exploit the a priori information provided by the visual path to compute an adequate translational velocity and a nominal rotational velocity according to the shape of the path.

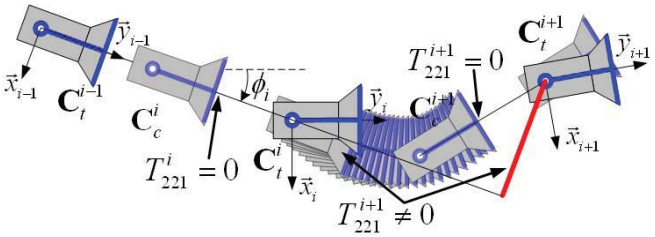


Fig. 4. Control strategy based on driving to zero the element of the trifocal tensor T_{221}.

C. Control Law for Autonomous Navigation

In this section, we describe the proposed control law that corrects the lateral deviation of the robot with respect to the taught visual path for each key image. As depicted in Fig. 4, the control law must take to zero the following one-dimensional function that represents the tracking error of the normalized tensor element T_{221} with respect to a desired reference $T^d_{221}(t)$:

$$\zeta = T_{221} - T^d_{221}(t). \tag{5}$$

The normalization of the TT is done as defined at the end of section II using $T_N = T_{232}$, which is non-null assuming that $\mathbf{C}_1 \neq \mathbf{C}_3$. The desired evolution of the tensor element is defined by the differentiable sinusoidal reference

$$\begin{aligned} T^d_{221}(t) &= \frac{T_{221}(0)}{2}\left(1+\cos\left(\frac{\pi}{\tau}t\right)\right), \quad 0 \leq t \leq \tau, \\ T^d_{221}(t) &= 0, \qquad\qquad\qquad\qquad\quad t > \tau, \end{aligned} \tag{6}$$

where $T_{221}(0)$ is the initial value of the normalized tensor element or the value at the time of key image switching, and τ is a suitable time in which the tensor element must reach zero, before the next switching of key image. Thus, the time is restarted at each instant when a change of key image occurs. This reference trajectory drives the tensor element to zero in a smooth way from its initial value.

The tracking error is computed using information extracted from the i^{th} key image as $I_3(\mathbf{K}, \mathbf{C}_3)$, the $(i-2)^{th}$ key image as $I_1(\mathbf{K}, \mathbf{C}_1)$ and the current image $I_2(\mathbf{K}, \mathbf{C}_2)$. According to the expressions of the trifocal tensor elements (3) and using the derivatives of the robot state as given by the model of the unicycle, we have that the time-derivative of T_{221} is

$$\begin{aligned} \dot{T}^m_{221} &= -\dot{x}_2\cos\phi_2 + x_2\dot{\phi}_2\sin\phi_2 - \dot{y}_2\sin\phi_2 - y_2\dot{\phi}_2\cos\phi_2, \\ &= (x_2\sin\phi_2 - y_2\cos\phi_2)\,\omega \\ &= T^m_{223}\omega. \end{aligned}$$

This time-derivative is also valid for normalized tensor elements and therefore, the differential equation relating the rate of change of the error with the reference tracking (RT) velocity is as follows:

$$\dot{\zeta} = T_{223}\omega_{rt} - \dot{T}^d_{221}. \tag{7}$$

Thus, the velocity ω_{rt} is worked out from the error dynamics (7). The following rotational velocity assigns a new dynamics through the auxiliary input δ_a:

$$\omega_{rt} = \frac{1}{T_{223}}\left(\dot{T}^d_{221} + \delta_a\right).$$

We define the auxiliary input as $\delta_a = -k_c\zeta$ to keep the tensor element T_{221} tracking the reference trajectory, where k_c is a control gain. Thus, the resulting rotational velocity is

$$\omega_{rt} = \frac{1}{T_{223}}\left(\dot{T}^d_{221} - k_c\zeta\right). \tag{8}$$

This velocity yields the error dynamics $\dot{\zeta} = -k_c\zeta$, which is exponentially stable for $k_c > 0$. This RT velocity is continuous with a sinusoidal behavior between key images. A nominal rotational velocity ($\bar{\omega}$) can be added in order to obtain an RT+ velocity that is maintained almost constant between key images, i.e., almost piece-wise constant rotational velocity during the navigation. So, the complete velocity can be eventually computed as

$$\omega = k_t\omega_{rt} + \bar{\omega}, \tag{9}$$

where $k_t > 0$ is a weighting factor on the reference tracking control ω_{rt}.

D. Exploiting Information from the Memory

Previous image-based approaches for navigation using a visual memory only exploit local information [4], [5], [6], i.e., the required rotational velocity is computed from the current and the nearest target images. We propose to exploit the visual memory in order to have an a priori information about the whole path without the need of a 3D reconstruction or representation of the path. A kind of qualitative map of the path can be obtained from the tensor element T_{221} using three consecutive key images of the memory. The value of this element, denoted as T^{ki}_{221}, shows qualitatively the orientation of the camera in the $(i-1)^{th}$ key image with respect to the i^{th} one and so, we can set an adequate translational velocity according to the curvature of the path as well as to compute the nominal rotational velocity that appears in (9). Recall that a priori, the tensor is computed between all consecutive triplets of key images with target in the i^{th} one. We propose the following smooth mapping $T^{ki}_{221} \rightarrow (\upsilon_{\min}, \upsilon_{\max})$ to modify the translational velocity between two limits accordingly:

$$\upsilon = \upsilon_{\max} + \upsilon_{\min} + \tfrac{\upsilon_{\max}-\upsilon_{\min}}{2}\tanh\left(1 - \frac{\left|T^{ki}_{221}/d_{\min}\right|}{\sigma}\right), \tag{10}$$

where σ is a positive parameter that determines the inflection point of the function. Once the translational velocity is set from the previous equation for each key image, its value can be used to compute the nominal velocity $\bar{\omega}$ proportional to the tensor elements T^{ki}_{221} as follows:

$$\bar{\omega} = \frac{k_m\upsilon}{d_{\min}}T^{ki}_{221}, \tag{11}$$

where $k_m < 0$ is a constant factor. This velocity by itself is able to drive the robot along the path, but correction is introduced in (9) through (8).

E. Timing Strategy and Key Image Switching

The proposed control method is based on taking to zero the tensor element T_{221} before reaching the next key image, which imposes a constraint for the time τ of the reference. Thus, a strategy to define this time is related to the minimum distance between key images ($d_{\min}$) and the translational velocity (υ) as follows:

$$\tau = \frac{d_{\min}}{\upsilon}.$$

We have found that a good approach to relate this time with the settling time γ of the tracking error is to consider $0.4\tau = 5/k_c$, from which $k_c = 12.5/\tau$.

By using the controller (8) with the reference (6), the time τ and the control gain k_c as described above, an intermediate location determined by $d_{\min}$ is reached. In the best case, when $d_{\min}$ coincides with the real distance between key images, the robot reaches the location where the corresponding key image was acquired. In order to achieve a good correction of the longitudinal position for each key image, the reference (6) is maintained to zero, which implies that $\omega = 0$, until the image error starts to increase. The image error is defined as the mean squared error between corresponding image points of the current image ($\mathbf{p}_{i,j}$) and points of the next closest target key image ($\mathbf{p}_j$), i.e.,

$$\epsilon = \frac{1}{r}\sum_{j=1}^{r} \|\mathbf{p}_j - \mathbf{p}_{i,j}\| \tag{12}$$

where r is the number of used corresponding points. As shown in [7], the image error decrease monotonically until the robot reaches each target view. In our case, the increment of the image error is the switching condition for the next key image, which is confirmed by using the current and previous difference of instantaneous values of the image error.

IV. Evaluation of the Proposed Control Scheme

In this section, we present simulations in Matlab of the proposed navigation scheme. We use the generic camera model [11] to generate synthetic key images from the 3D scene of Fig. 5(a) according to the robot motion on the predefined path shown in the same figure. This learned path starts in the location (5,-5,0^o) and finishes just before to close the loop of 54 m long. The vision system is hypercatadioptric with parameters $\alpha_x = 950$, $\alpha_y = 954$, $x_0 = 512$, $y_0 = 384$ all of them in pixels, $\xi = 0.9662$ and the size of the images is 1024×768 pixels. The TT is estimated using the typical 7-point algorithm as introduced in section II and using the projected points on the sphere. Fig. 5(b) shows an example of a triplet of catadioptric images projected onto the unitary sphere.

The performance of the proposed control scheme including image noise (2 pixels of standard deviation) is evaluated for a visual path of 36 key images distributed randomly along the learned path. The random distance between consecutive key images is between 1.42 m and 1.6 m, in such a way that a minimum distance $d_{\min} = 1.4$ m is assumed. The translational velocity is bounded between 0.2 m/s and 0.4 m/s. We present results for the two cases according to (9): 1) only reference tracking (RT) and 2) reference tracking + nominal velocity (RT+). We can see in Fig. 6(a) that the resultant path of the autonomous navigation stage is almost similar to the learned one in both cases; however, as expected, the performance is better for the RT control. The RT+ control decreases its performance in sharp curves, but it is better as long as the key images are closer. The first plot of Fig. 6(b) shows how the translational velocity effectively changes according to the shape of the path. For instance, between 55 s and 85 s the higher velocity is applied, which corresponds to the almost straight part of the path. The second plot of the same figure shows the behavior of the

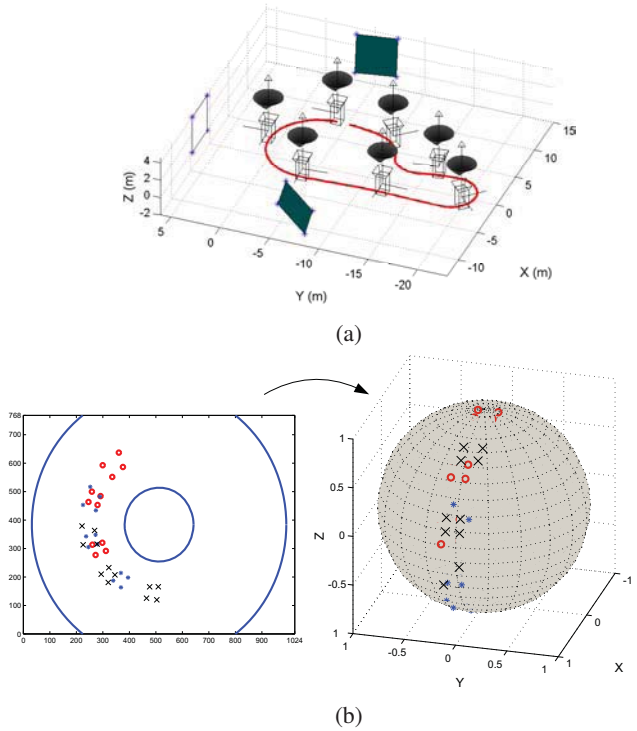
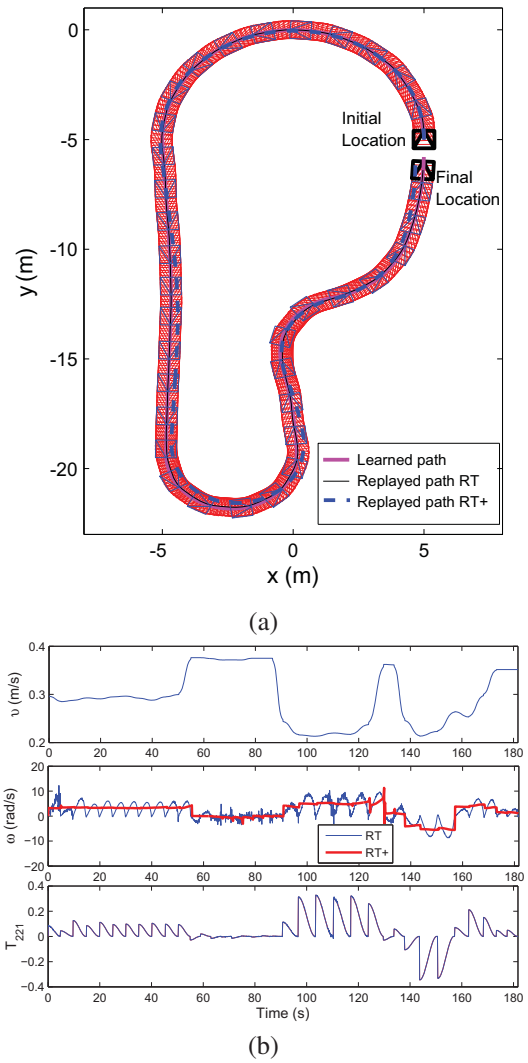


Fig. 5. Virtual scene and example of the synthetic images used. (a) Tridimensional scene and predefined path. (b) Example of a triplet of images projected onto the unitary sphere.

rotational velocity. On one hand, the velocity given by the RT control is smooth, performing cycles that start from zero and return to zero for each key image. On the other hand, the piece-wise constant velocity given by the RT+ control is more adequate. The third plot of the same figure shows the behavior of the reference tracking of the measurement T_{221}. It can be seen that the tensor element does not present unstable behavior when a key image is reached (T_{221} reaches zero), which means that the problem of short baseline is not present. Additionally, although the noise in the image points is significative, the use of the trifocal tensor reduces such effect and increases the robustness of the control scheme.

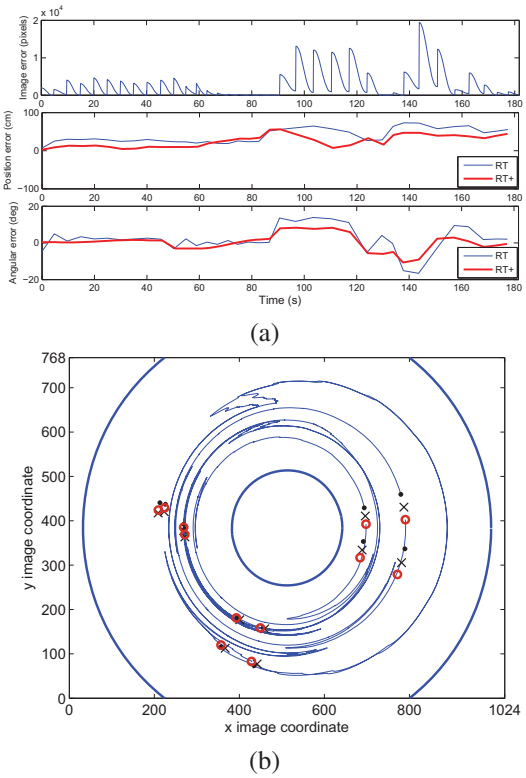
Fig. 7 presents the performance of the approach for the same experiment. The first plot of Fig. 7(a) shows the behavior of the image error for the RT case. It can be seen that the image error exhibits a monotonic decay before reaching each key image. The largest peaks in the image error correspond to the sharp curves in the path, which also causes the highest error in the path following, as can be seen in the second and third plot of the same figure. Nevertheless, the path following errors to reach each key image are small and comparable for both controllers. In order to show the behavior of the visual information, Fig. 7(b) presents the motion of the image points along the whole navigation. Although 12 points are used to compute the tensor, only the motion of 7 points is shown. It is appreciable the advantage of using a central catadioptric system looking upward, which is able to see the same scene during the navigation task.



(a)

(b)

Fig. 6. Simulation results for a navigation task. (a) Resultant paths and key images distribution. (b) Velocities and evolution of the element T_{221}.



(a)

(b)

Fig. 7. Performance of the navigation task for the results in Fig. 6. (a) Image error and path following errors. (b) Motion of the points in the images. Markers: "·" initial image, "O" final key image, "×" final reached location.

V. Conclusions

In this paper, we have proposed a control scheme for wheeled mobile robot navigation based on a visual memory, which exploits the advantages of the trifocal tensor. The value of one element of the tensor is the unique required information by the control law. Our proposed image-based approach does not need pose parameters decomposition. In this context, the scheme avoids discontinuous rotational velocity when a new target image must be reached providing piece-wise constant velocity if desired. The translational velocity is adapted according to the path and the approach is independent of its value. We exploit the advantages of wide field of view cameras, so that the scheme can be applied using any central camera, which increases its applicability. The control scheme has presented good performance according to simulation results using synthetic images.

References

[1] F. Chaumette and S. Hutchinson. Visual servo control part I: Basic approaches. *IEEE Robotics and Autom. Mag.*, 13(14):82–90, 2006.

[2] G.N. DeSouza and A.C. Kak. Vision for mobile robot navigation: A survey. *IEEE Trans. on Pat. Anal. and Mach. Intel.*, 24(2):237–267, 2002.

[3] Y. Matsumoto, K. Ikeda, M. Inaba, and H. Inoue. Visual navigation using using omnidirectional view sequence. In *IEEE Int. Conf. on Intelligent Robots and Systems*, pages 317–322, 1999.

[4] T. Goedeme, T. Tuytelaars, L. V. Gool, G. Vanacker, and M. Nuttin. Feature based omnidirectional sparse visual path following. In *IEEE/RSJ Int. Conf. on Intel. Rob. and Syst.*, pages 1806–1811, 2005.

[5] E. Royer, M. Lhuillier, M. Dhome, and J. M. Lavest. Monocular vision for mobile robot localization and autonomous navigation. *International Journal of Computer Vision*, 74(3):237–260, 2007.

[6] J. Courbon, Y. Mezouar, and P. Martinet. Autonomous navigation of vehicles from a visual memory using a generic camera model. *IEEE Trans. on Intel. Transportation Systems*, 10(3):392–402, 2009.

[7] Z. Chen and S. T. Birchfield. Qualitative vision-based mobile robot navigation. In *IEEE Int. Conf. on Robotics and Automation*, pages 2686–2692, 2006.

[8] A. Diosi, A. Remazeilles, S. Segvic, and F. Chaumette. Outdoor visual path following experiments. In *IEEE Int. Conf. on Intelligent Robots and Systems*, pages 4265–4270, 2007.

[9] A. Cherubuni and F. Chaumette. Visual navigation with a time-independent varying reference. In *IEEE Int. Conf. on Intelligent Robots and Systems*, pages 5968–5973, 2009.

[10] H. M. Becerra, J. Courbon, Y. Mezouar, and C. Sagüés. Wheeled mobile robots navigation from a visual memory using wide field of view cameras. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5693–5699, 2010.

[11] C. Geyer and K. Daniilidis. An unifying theory for central panoramic systems and practical implications. In *European Conf. on Computer Vision*, pages 445–461, 2000.

[12] G. López-Nicolás, J.J. Guerrero, and C. Sagüés. Visual control through the trifocal tensor for nonholonomic robots. *Robotics and Autonomous Systems*, 58(2):216–226, 2010.

[13] H. M. Becerra, G. López-Nicolás, and C. Sagüés. Omnidirectional visual control of mobile robots based on the 1D trifocal tensor. *Robotics and Autonomous Systems*, 58(6):796–808, 2010.