To appear in the *International Journal of Control*
Vol. 00, No. 00, Month 20XX, 1–24

Visual Servo Walking Control for Humanoids with Finite-time Convergence and Smooth Robot Velocities

Josafat Delfin[a*] , Hector M. Becerra[b] and Gustavo Arechavaleta[a]

[a] *Robotics and Advanced Manufacturing Group, Centro de Investigación y de Estudios Avanzados del IPN, Saltillo, Coah., Mexico, e-mail: (josafat.delfin, garechav)@cinvestav.edu.mx*
[b] *Centro de Investigación en Matemáticas (CIMAT), Guanajuato, Mexico, (e-mail: hector.becerra@cimat.mx)*

Word count: 6767 words

(*Received July 2015*)

In this paper, we address the problem of humanoid locomotion guided from information of a monocular camera. The goal of the robot is to reach a desired location defined in terms of a target image, i.e., a positioning task. The proposed approach allows us to introduce a desired time to complete the positioning task, which is advantageous in contrast to the classical exponential convergence. In particular, finite-time convergence is achieved while generating smooth robot velocities and considering the omnidirectional waking capability of the robot. In addition, we propose a hierarchical task-based control scheme, which can simultaneously handle the visual positioning and an obstacle avoidance tasks without affecting the desired time of convergence. The controller is able to activate or inactivate the obstacle avoidance task without generating discontinuous velocity references while the humanoid is walking. Stability of the closed-loop for the two task-based control is demonstrated theoretically even during the transitions between tasks. The proposed approach is generic in the sense that different visual control schemes are supported. We evaluate a homography-based visual servoing for position-based and image-based modalities, as well as for eye-in-hand and eye-to-hand configurations. The experimental evaluation is performed with the humanoid robot NAO.

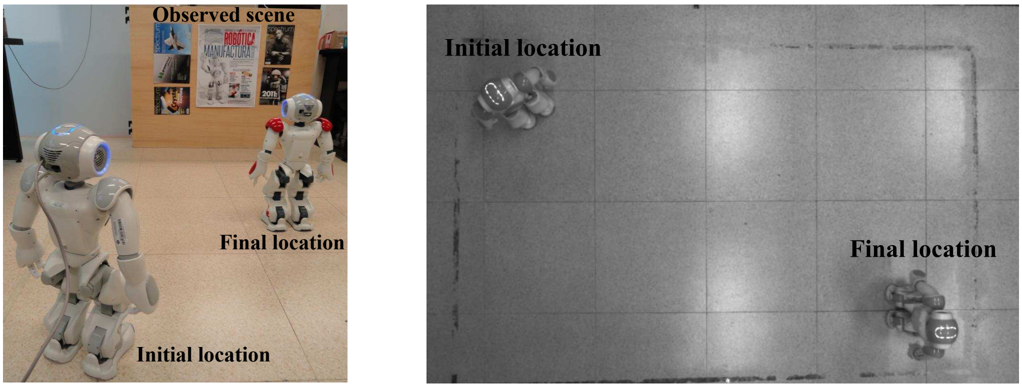
Keywords: Visual servoing; humanoid robots; finite-time convergence; obstacle avoidance; homography-based control

1. INTRODUCTION

The problem of humanoid locomotion has been widely studied in recent literature (Kajita et al. (2014)). The complexity of the problem is mainly related to maintain the dynamic equilibrium of the robot while walking. A pioneer walking pattern generator (WPG) (Kajita et al. (2003)) assumes predefined references for the zero moment point (ZMP) coming from the sequence of footsteps that the robot must follow during the coordination of its motion. In (Herdt et al. (2010)), the authors propose a scheme based on quadratic programming (QP), which generates the sequence of such steps in a reactive manner. In this approach, the input to the WPG is a reference in linear and angular velocities of the center of mass (CoM) of the humanoid. The outputs are the trajectories of the CoM that satisfy the criterion of the dynamic equilibrium and the trajectory of the foot that generates the next contact point with the ground.

The locomotion task assigned to the robot can be defined in terms of a visual target, i.e., the robot has to walk to a desired location defined from visual features. This positioning task is depicted in

[*] Corresponding author. E-mail: josafat.delfin@cinvestav.edu.mx



(a) (b)

Figure 1. **Positioning task of a humanoid robot.** The locomotion is guided by visual servo control. (a) Eye-in-hand configuration. (b) Eye-to-hand configuration.

Fig. 1(a) for an eye-in-hand and in Fig. 1(b) for an eye-to-hand configurations, respectively. In the first case the robot motion induces camera displacements and in the second case the camera is fixed, observing the motion of the robot from a stationary configuration (Chaumette and Hutchinson (2006)). Visual information has been used in a variety of humanoid robot navigation strategies. In (Ido et al. (2009)), a sequence of images obtained from a teaching phase together with a technique based on template correlation have been suggested to decide the actions that a humanoid has to perform for navigation purposes. The strategy proposed in (Faragasso et al. (2013)) copes with different corridor configurations by taking advantage of vanishing points easily computed in that particular environment. In (Oriolo et al. (2013)) a vision-based odometric localization method is used to close the loop for the task of tracking a desired trajectory. In both last works, the humanoid robot is assumed to behave as a non-holonomic system.

Humanoids are capable to perform several tasks simultaneously. In (Mansard et al. (2007)), a grasping task commanded by visual servoing is performed while the humanoid is walking. This strategy has been incorporated within a visually-guided locomotion planner in (Hayet et al. (2012)). In these cases, the locomotion task cannot be modified by an error function obtained from the visual task. The reactive walking controller introduced in (Herdt et al. (2010)) overcomes this limitation. Using this WPG, the sequence of footsteps of the humanoid can be computed from a set of visual features, as it is demonstrated in (Dune et al. (2010)). In that work, a classical visual servoing scheme provides a reference of velocity for the CoM that is introduced to the WPG. Such scheme is a decoupled approach in the sense that the visual controller is independent of the WPG. In contrast, a coupled approach suggested in (García et al. (2014)), introduces the visual constraints directly in the QP problem that is solved inside the WPG. Thus, this scheme is an extension of the WPG (Herdt et al. (2010)) including the constraints given by the visual task in the linear model predictive control scheme that solves the humanoid locomotion.

In this work, we focus on the locomotion guided from monocular vision while the robot is able to avoid unexpected obstacles that appear during its motion. In the literature, the obstacle avoidance task has been addressed from vision-based guidance of humanoid robots. In (Cupec et al. (2005)), a vision-guided walking in structured indoor scenarios is proposed. The strategy uses odometry and visual information to generate a local map of the environment where the robot is localized relative to the obstacles and then a reactive footstep planner drives the robot to avoid the obstacles. The authors of (Maier et al. (2013)) propose to detect obstacles from monocular images and sparse laser data. Transversability of the environment is estimated from that information to provide motion commands that drive the robot to avoid obstacles.

In this paper, we propose a visual servo control of humanoid robots that achieves finite-time convergence of a positioning task, even in the case that an obstacle is present in the path toward the target location. Finite-time convergence means that a task is finished in a predefined time interval. A vision-based control for humanoid robots able to manage reactively an obstacle avoidance task without jeopardizing the robot balance was not previously investigated in the literature. Both problems visual servoing and obstacle avoidance have been treated separately in previous works (Dune et al. (2010); García et al. (2014); Maier et al. (2013)). Additionally, ensuring finite-time termination of a task is a desirable property for robotics applications even if only one task is performed by the robot. Finite-time convergence improves the flexibility of the robotic system when more than one task must be performed, by allowing to have a controlled scheduling of the execution of each task (Jarquín et al. (2013); Ou et al. (2013)). The classical asymptotic stability of a visual task, presented for instance in (Dune et al. (2010); García et al. (2014)), is disadvantageous in humanoid robots control given that the small final velocities keep the robot oscillating around the goal without stopping, while a finite-time controller turns off the robot velocities when predefined. This might improve the final accuracy of the task. From asymtotic stability, a possibility to control the time of convergence might imply a no trivial tuning of control gains. We exploit the ideas reported in the context of movement neuroscience for studying human arm reaching tasks (Morasso et al. (1997)), where a time base generator (TBG) is proposed to achieve finite-time tracking of hand trajectories. The TBG as a terminal attractor has been used for kinematic control for handling postural task transitions of redundant robots in (Jarquín et al. (2013)). A different control scheme with finite-time convergence has been proposed in the context of visual servoing for reference tracking of multiple non-holonomic robots in (Ou et al. (2013)).

In the work herein, we propose a decoupled visual servo walking control that introduces a TBG, which also generates smooth robot velocities. The time to complete the positioning task is directly predefined in terms of the TBG. The contribution of the paper is then two-fold: first, the proposed visual servo walking control achieves finite-time convergence using smooth robot velocities and considering the omnidirectional waking capability of the robot; second, the proposed scheme corresponds to a hierarchical task-based control with smooth transitions that can deal with the avoidance of an obstacle detected in the path toward the desired robot location without affecting the desired time of convergence. The stability of the closed-loop for that couple of tasks is demonstrated even during the transitions between tasks. Additionally, the proposed scheme is shown to be generic in the sense that different visual control schemes can be used. In particular, we evaluate the approach using a homography-based visual control for position-based and image-based modalities, as well as for eye-in-hand and eye-to-hand configurations. The performance of the proposed approach is reported from online experiments using the NAO robot as experimental platform.

Regarding to our results, it will be shown that fixed-time convergence is indeed achieved, i.e., the same convergence time is maintained from different initial conditions. This nice property had not been explored before (Dune et al. (2010); García et al. (2014)). We consider that the time constraint to finish a visual task might be hard to be introduced in a coupled approach like the one in (García et al. (2014)). Exponential convergence of previous works (Dune et al. (2010); García et al. (2014)) gives an initial large control effort, which is not suitable for the dynamic balance of a humanoid robot. This issue is mitigated in our work thanks to the use of the TBG. Moreover, the TBG allow us to obtain smooth robot velocities even in the activation of an obstacle avoidance task while the visual servoing task is performing. As a consequence, the generated input references do not put at risk the walking stabilization of humanoid robots. It is also important to mention that the obstacle avoidance task is carried out without the need of an environment's map nor any previous training stage, in contrast to (Cupec et al. (2005); Maier et al. (2013)).

The paper is structured as follows. Section 2 presents a visual control scheme for humanoid robots that achieves exponential stability. In Section 3, a generic extension of the previous visual control to achieve finite-time convergence is proposed. In Section 4, we propose a generic visual control scheme that yields finite-time convergence while handling the avoidance of an obstacle and generating smooth robot velocities. Section 5 presents an experimental evaluation of the proposed visual control scheme with and without obstacles. Conclusions are given in Section 6.

2. VISUAL CONTROL SCHEME FOR HUMANOIDS

Visual control tasks can be formulated as a set of differential linear equations similar to other motion tasks that humanoids perform simultaneously. Here we describe how humanoid locomotion can be generated by only considering a linear model that captures the relation between the acceleration of the CoM and the ZMP during walking and a reference velocity representing the instantaneous output of a visual servo controller.

2.1 *Humanoid Locomotion*

The simplified discrete model of the ZMP-CoM that is commonly considered by a WPG is of the form (Kajita et al. (2003)):

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}u(k), \\ p(k) &= \boldsymbol{c}\boldsymbol{x}(k), \end{aligned} \tag{1}$$

where

$$\begin{aligned} \boldsymbol{x}(k) &= \left[x(kT)\ \dot{x}(kT)\ \ddot{x}(kT)\right]^T, \\ u(k) &= u_x(kT),\ p(k) = p_x(kT), \\ \boldsymbol{A} &= \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & T \end{bmatrix},\ \boldsymbol{B} = \begin{bmatrix} T^3/6 \\ T^2/2 \\ T \end{bmatrix}, \\ \boldsymbol{c} &= \begin{bmatrix} 1 & 0 & -h/g \end{bmatrix}, \end{aligned}$$

with T the sampling period, $\boldsymbol{x}$ represents the position of the CoM w.r.t. x-axis (the formulation is the same for y-axis). The jerk of CoM is given by $u_x = \dddot{x}$ while p_x stands for the position of ZMP. The height and gravity are h and g, respectively. The model (1) can be extended for a time horizon given by N steps. If the ZMP reference is known then a QP can be solved to find the ZMP-CoM solution. It is known that the QP formulation can be modified to consider automatic footstep placements and to incorporate additional inequality linear constraints (Herdt et al. (2010)). The WPG based on QP solves the following problem:

$$\min_{\bar{\mathbf{u}}(k)} \frac{\alpha}{2}\left\|\bar{\mathbf{u}}(k)\right\|^2 + \frac{\beta}{2}\left\|\dot{\boldsymbol{x}}(k+1) - \dot{\boldsymbol{x}}_r(k+1)\right\|^2 + \frac{\gamma}{2}\left\|\mathbf{p}(k+1) - \mathbf{p}_r(k+1)\right\|^2, \tag{2}$$

subject to $\boldsymbol{D}(k+1)\bar{\mathbf{u}}(k) \leq \boldsymbol{b}(k+1)$,
where $\dot{\boldsymbol{x}}(k+1) = [\dot{x}(k+1)\ \ldots\ \dot{x}(k+N)\ \dot{y}(k+1)\ \ldots\ \dot{y}(k+N)]^T$ is the velocity of CoM while $\dot{\boldsymbol{x}}_r$ represents its reference velocity, $\bar{\mathbf{u}}(k) = [u_x(k)\ x_f(k)\ u_y(k)\ y_f(k)]^T$, α and β are weights,

$$\mathbf{p}_r(k+1) = \mathbf{S}_c(k+1)\mathbf{x}_c(k) + \mathbf{S}(k+1)\boldsymbol{x}_f(k),$$

$\boldsymbol{D}(k+1)$ and $\boldsymbol{b}(k+1)$ contain the coefficients of the inequality constraints that define the feasible area to put the next step $\boldsymbol{x}_f(k)$, $\mathbf{S}_c$ and $\mathbf{S}$ play the role of selection matrices to handle the correspondence between the foot in contact $\boldsymbol{x}_c(k)$ and the foot for the next step. The outcome of

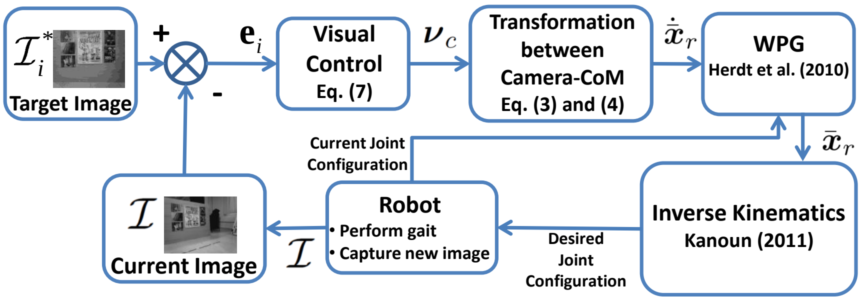


Figure 2. **Visual control scheme for humanoids.**

the pattern generator is used to compute the motion coordination by means of an efficient inverse kinematics method (Kanoun (2011)).

2.2 *Locomotion guided from visual control*

Let $\boldsymbol{\nu}_m$ be the velocity of the frame attached to the robot's CoM and $\boldsymbol{\nu}_c$ the velocity command given by a visual servo controller. If it is assumed that a constant transformation ${}^m\boldsymbol{T}_c$ exists between $\boldsymbol{\nu}_c$ and $\boldsymbol{\nu}_m$, then the following equation holds:

$$\boldsymbol{\nu}_m =^m \boldsymbol{T}_c \boldsymbol{\nu}_c, \tag{3}$$

where

$${}^m\boldsymbol{T}_c = \begin{pmatrix} {}^m\boldsymbol{R}_c^T & -{}^m\boldsymbol{R}_c^T[\boldsymbol{r}]_\times \\ 0 & {}^m\boldsymbol{R}_c^T \end{pmatrix} \in \mathbb{R}^{6\times 6},$$

with $[\boldsymbol{r}]_\times \in \mathbb{R}^{3\times 3}$ representing a screw-symmetric matrix and ${}^m\boldsymbol{R}_c \in SO(3)$. This constant transformation means that the robot's head is fixed w.r.t. the robot's body. The input for the WPG can be expressed as:

$$\dot{\bar{\boldsymbol{x}}}_r = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{\nu}_m. \tag{4}$$

Notice that the translational components of $\dot{\bar{\boldsymbol{x}}}_r$ are the input of (2) while the angular term is used to construct the linear inequality constraints that define the feasible region for the next footstep placement. The whole process for the visual control tailored to humanoids is shown in Fig. 2.

2.3 *Visual control strategies for eye-in-hand configuration*

In this section, we briefly recall the classical approach for visual servo control (Chaumette and Hutchinson (2006)). The aim is to drive a vector of m image features $\mathbf{s}$ to a desired value $\mathbf{s}^*$. Thus, we can define an error as follows:

$$\mathbf{e} = \mathbf{s} - \mathbf{s}^* \in \mathbb{R}^m. \tag{5}$$

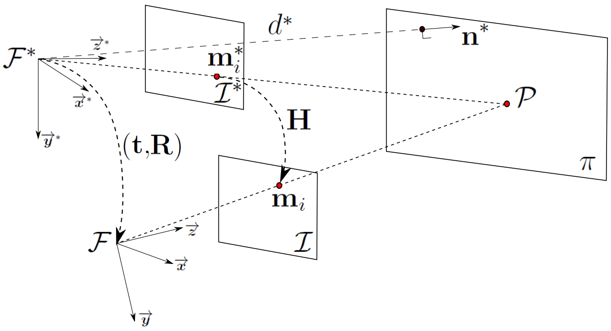


Figure 3. **Eye-in-hand configuration.** Reference frames definition and homography.

The relationship between the camera velocities and the change of the image features vector is given by $\dot{\mathbf{s}} = \mathbf{L}_s \boldsymbol{\nu}_c$, where $\mathbf{L}_s \in \mathbb{R}^{m\times 6}$ is an interaction matrix. In this paper, we concern for selecting an appropriate features vector of dimension 6, i.e. $m = 6$, in order to have a square interaction matrix. Besides, in our case, $\mathbf{s}^*$ is a vector of constant values and then $\dot{\mathbf{e}} = \dot{\mathbf{s}}$. Thus, the time derivative of the error is given by:

$$\dot{\mathbf{e}} = \mathbf{L}_s \boldsymbol{\nu}_c. \tag{6}$$

A typical choice of the camera velocity vector to impose an exponential decay of the error is as follows:

$$\boldsymbol{\nu}_c = -\lambda \hat{\mathbf{L}}_s^{-1} \mathbf{e}, \tag{7}$$

with λ a positive control gain. Notice that we will propose a six-dimensional features vector that allows us to have an invertible estimated interaction matrix $\hat{\mathbf{L}}_s$.

2.3.1 Position-based scheme (3D)

Let us denote by $\mathcal{F}$ and $\mathcal{F}^*$ the reference frames associated to the current camera pose (translation and rotation) and the target pose, respectively, as shown in Fig. 3. A vector $\mathbf{t}$ expressed in $\mathcal{F}^*$ represents the translation between the reference frames $\mathcal{F}$ and $\mathcal{F}^*$, and $\mathbf{R}$ is the rotation matrix between those frames. A features vector for a position-based control can be defined as follows:

$$\mathbf{s} = \begin{bmatrix} \mathbf{t} \\ \theta\mathbf{a} \end{bmatrix} \in \mathbb{R}^6, \tag{8}$$

where $\theta\mathbf{a}$ represents the angle/axis parametrization of the rotation matrix $\mathbf{R}$ (Rodrigues' formulation). Since the features vector encodes the translation and rotation with respect to the target frame, in this case $\mathbf{s}^* = 0$ and therefore $\mathbf{e} = \mathbf{s}$. The time-derivative of the error vector is given by (6) with the following square interaction matrix:

$$\mathbf{L}_s = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{\theta\mathbf{a}} \end{bmatrix},$$

where $\mathbf{I}$ is an identity matrix of 3×3 and $\mathbf{L}_{\theta\mathbf{a}} \in \mathbb{R}^{3\times 3}$ is such that $\mathbf{L}_{\theta\mathbf{a}}^{-1}\theta\mathbf{a} = \theta\mathbf{a}$ (Chaumette and

Hutchinson (2006)). The translational and rotational velocities result to be decoupled according to (7) and the velocity vector is given by:

$$\boldsymbol{\nu}_c = \left[\begin{array}{c} \boldsymbol{v}_c \\ \boldsymbol{\omega}_c \end{array} \right] = -\lambda \left[\begin{array}{c} \mathbf{t} \\ \theta \mathbf{a} \end{array} \right]. \tag{9}$$

Considering that the control gain accomplishes $\lambda > 0$ and that the camera pose is accurately estimated, the velocity vector (9) yields to an exponentially stable error dynamics of the form $\dot{\mathbf{e}} = -\lambda \mathbf{e}$.

There are several options to recover the relative pose between the current camera frame $\mathcal{F}$ and the target one $\mathcal{F}^*$. An option that does not require to know the 3D structure of the scene or a 3D model of an object in the scene is the homography matrix decomposition (Malis and Vargas (2007)). The Euclidean homography matrix $\mathbf{H}$ can be estimated using only the normalized images $\mathcal{I}$ and $\mathcal{I}^*$. In a normalized image, the calibration parameters encoded in $\mathbf{K}$ are removed using $\mathbf{K}^{-1}$. The homography encodes the rigid transformation that exists between the reference frames $\mathcal{F}$ and $\mathcal{F}^*$ as follows:

$$\mathbf{H} = \mathbf{R} + \frac{\mathbf{t}}{d^*}\mathbf{n}^{*T}, \tag{10}$$

where $\mathbf{R}$ and $\mathbf{t}$ are the rotation matrix and translation vector as defined above, $\mathbf{n}^*$ is the unitary vector expressed in $\mathcal{F}^*$ normal to a plane π, and d^* is the distance from π to $\mathcal{F}^*$. See Fig. 3 for details.

Thus, it is possible to decompose $\mathbf{H}$ according to (10) to obtain $\mathbf{R}$ and $\mathbf{t}$, which are necessary to implement the control law (9). An efficient algorithm to carry out a decomposition of $\mathbf{H}$ is proposed in (Triggs (1998)). It is well known that such kind of decomposition generates two geometrically valid solutions, however, only one of them is physically admissible. The correct solution can be selected taken the solution associated to the normal vector whose third component (n_z) will be the largest. We want to emphasize that although the homography model is valid for planar scenes, it is also possible to estimate a homography associated to a virtual plane for non-planar scenes using the algorithm proposed in (Malis et al. (2000)).

2.3.2 Image-based scheme (2D)

The homography model has also proved to be efficient for 2D visual servo control. A control scheme that formulates a control law directly in terms of the homography matrix is proposed in (Benhimane and Malis (2007)). Error vectors for translation and rotation are directly computed from the homography matrix. The core idea in that scheme is that the reference frames $\mathcal{F}$ and $\mathcal{F}^*$ coincide, if and only if the homography matrix $\mathbf{H}$ is equal to the identity matrix $\mathbf{I}$. Under that notion, a task function $\mathbf{s} = \mathbf{e} \in \mathbb{R}^6$ locally isomorphic to the camera's pose is defined. The task function $\mathbf{e} = \left[\mathbf{e}_v^T \; \mathbf{e}_\omega^T\right]^T$ is null if and only if the camera reaches the target pose and it is given by:

$$\begin{array}{rcl} \mathbf{e}_v & = & (\mathbf{H} - \mathbf{I})\mathbf{m}^*, \\ \left[\mathbf{e}_\omega\right]_\times & = & \mathbf{H} - \mathbf{H}^T, \end{array} \tag{11}$$

where $\mathbf{m}^*$ is any normalized point in the target image $\mathcal{I}^*$ that belongs to the plane (virtual plane for non-planar scenes) that defines $\mathbf{H}$, and $\left[\mathbf{e}_\omega\right]_\times$ represents the skew-symmetric matrix for the vector $\mathbf{e}_\omega = \left[e_{\omega x} \; e_{\omega y} \; e_{\omega z}\right]^T$.

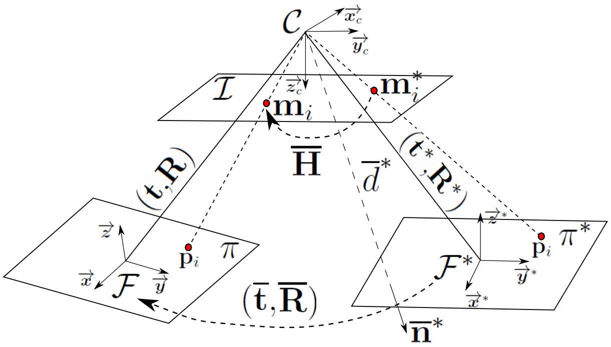


Figure 4. **Eye-to-hand configuration.** Reference frames definition and homography.

Thus, the task function $\mathbf{e}$ can be estimated using only the images $\mathcal{I}$ and $\mathcal{I}^*$, without obtaining the 3D structure of the target pose (d^*). It is only needed to estimate the homography $\mathbf{H}$. The time-derivative of the task function gives the linear system $\dot{\mathbf{e}} = \mathbf{L}\boldsymbol{\nu}_c$, where $\mathbf{L}$ is a square interaction matrix. In (Benhimane and Malis (2007)) it is shown that this control scheme is efficient even without using the interaction matrix. Thus, the linear control

$$\boldsymbol{\nu}_c = \begin{bmatrix} \boldsymbol{v}_c \\ \boldsymbol{\omega}_c \end{bmatrix} = -\lambda \begin{bmatrix} \mathbf{e}_v \\ \mathbf{e}_\omega \end{bmatrix}, \tag{12}$$

with $\lambda > 0$, yields an error dynamics locally stable. Therefore, the control law (12) only depends on the task function (11), which converges exponentially to zero. The local stability of the task function is guaranteed for any 3D structure and any point $\mathbf{m}^*$ (Benhimane and Malis (2007)).

2.4 *Extension of the control strategies for eye-to-hand configuration*

Let us consider the case of an eye-to-hand configuration as shown in Fig. 4. Under the same definition of the error $\mathbf{e} = \mathbf{s} - \mathbf{s}^*$ and assuming a static target, the time derivative of the error is well-known to be of opposite sign w.r.t. the case of the eye-in-hand configuration, i.e:

$$\dot{\mathbf{e}} = -\mathbf{L}_s \boldsymbol{\nu}_c. \tag{13}$$

Hence, the velocity command of the visual servo control is given by:

$$\boldsymbol{\nu}_c = \lambda \hat{\mathbf{L}}_s^{-1} \mathbf{e}, \tag{14}$$

being λ a positive control gain.

2.4.1 Position-based control

As depicted in Fig. 4, the Euclidean homography matrix that relates the frames $\mathcal{F}$ and $\mathcal{F}^*$ in the case of eye-to-hand configuration can be expressed as:

$$\overline{\mathbf{H}} = \overline{\mathbf{R}} + \frac{\overline{\mathbf{t}}}{\overline{d}^*} \overline{\mathbf{n}}^{*T}. \tag{15}$$

The homography $\overline{\mathbf{H}}$ is estimated from normalized point features of the image $\mathcal{I}$. The homography $\overline{\mathbf{H}}$ can be decomposed as described at the end of Section 2.3.1 to obtain $(\bar{\mathbf{t}}, \overline{\mathbf{R}})$. Similarly to the error defined in (8), here we define:

$$\mathbf{e} = \mathbf{s} = \left[\begin{array}{c} \bar{\mathbf{t}} \\ \overline{\theta \mathbf{a}} \end{array} \right] \in \mathbb{R}^6, \tag{16}$$

where $\overline{\theta \mathbf{a}}$ is the angle/axis parametrization of the matrix $\overline{\mathbf{R}}$. By applying (14), the velocity command of the visual servo control in the camera frame $\mathcal{C}$ is given by:

$$\overline{\boldsymbol{\nu}}_c = \left[\begin{array}{c} \overline{\boldsymbol{v}}_c \\ \overline{\boldsymbol{\omega}}_c \end{array} \right] = \lambda \left[\begin{array}{c} \bar{\mathbf{t}} \\ \overline{\theta \mathbf{a}} \end{array} \right]. \tag{17}$$

Assuming that the fixed desired robot pose $(\mathbf{t}^*, \mathbf{R}^*)$ with respect to the camera frame $\mathcal{C}$ is known, we can compute the pose of the frame $\mathcal{F}$ associated to the current robot pose as follows:

$$\mathbf{R} = \overline{\mathbf{R}}\mathbf{R}^* \text{ and } \mathbf{t} = \bar{\mathbf{t}} + \overline{\mathbf{R}}\mathbf{t}^*. \tag{18}$$

From the estimated current robot pose, a spatial motion transform, which transforms velocity commands expressed in the fixed frame $\mathcal{C}$ to the current frame $\mathcal{F}$, can be formed as follows:

$$^{\mathcal{F}}\mathbf{T}_{\mathcal{C}} = \left[\begin{array}{cc} \mathbf{R}^T & -\mathbf{R}^T[\mathbf{t}]_\times \\ \mathbf{0} & \mathbf{R}^T \end{array} \right].$$

Finally, the velocity command (17) expressed in the frame $\mathcal{F}$ is given by:

$$\boldsymbol{\nu}_c =^{\mathcal{F}} \mathbf{T}_{\mathcal{C}} \overline{\boldsymbol{\nu}}_c. \tag{19}$$

This vector of velocities feeds the WPG as defined in (3).

2.4.2 Image-based control

An image-based control scheme in an eye-to-hand configuration can be also formulated in terms of the homography $\overline{\mathbf{H}}$ defined in (15). Let us define similar errors to those in (11) as follows:

$$\begin{array}{rcl} \bar{\mathbf{e}}_v & = & (\overline{\mathbf{H}} - \mathbf{I})\mathbf{m}^*, \\ [\bar{\mathbf{e}}_\omega]_\times & = & \overline{\mathbf{H}} - \overline{\mathbf{H}}^T. \end{array} \tag{20}$$

Then, the velocity command of the visual servo control in the camera frame $\mathcal{C}$ is given by:

$$\overline{\boldsymbol{\nu}}_c = \left[\begin{array}{c} \overline{\boldsymbol{v}}_c \\ \overline{\boldsymbol{\omega}}_c \end{array} \right] = \lambda \left[\begin{array}{c} \bar{\mathbf{e}}_v \\ \bar{\mathbf{e}}_\omega \end{array} \right]. \tag{21}$$

It is also necessary to apply (19) to express the velocity vector in the current robot frame $\mathcal{F}$.

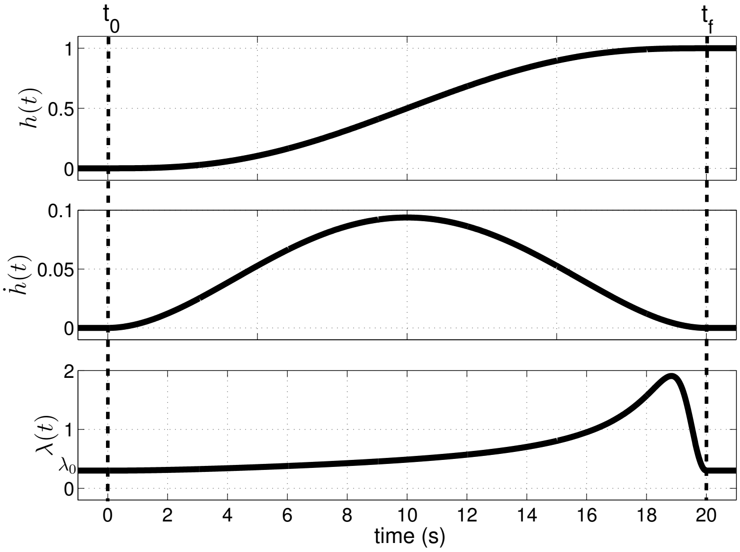


Figure 5. **Behavior of time-varying functions**. Top: profile of the terminal attractor $h(t)$. Middle: its time derivative $\dot{h}(t)$. Bottom: the resulting gain $\lambda(t)$. All functions are continuous and their final values are reached at a given final time $t_f = 20s$.

3. VISUAL CONTROL WITH FINITE-TIME CONVERGENCE

In visual servo control, the regulation of a task function is commonly performed by imposing an exponential decay as defined by (6) and (7). An important limitation of such regulation scheme is that the time evolution of the task errors cannot be directly controlled. In addition, discontinuous input signals appear at the activation time of the visual task as a by-product of the exponential behavior. Here, we introduce a terminal attractor, in particular a modified time-base generator (TBG) capable to ensure finite-time convergence while avoiding the undesired discontinuous effect occurred at the task activation time. It is of the form:

$$\lambda(t, t_0, t_f) = \lambda_0 + \frac{\dot{h}(t, t_0, t_f)}{1 - h(t, t_0, t_f) + \delta}, \tag{22}$$

where $0 < \delta < 0.1$, t_0 and $t_f > t_0$ are the initial and final times, respectively. The minimum value of (22) is reached at λ_0 which can be thought as a lower bound below the physical resolution of the robot's encoders. In the sequel, we will simplify the notation of the TBG as $h(t) := h(t, t_0, t_f)$ and the TBG gain as $\lambda(t) := \lambda(t, t_0, t_f)$. The function $h(t)$ can be implemented using the following smooth function whose range is the interval $0 \leq h(t) \leq 1$:

$$h(t) = \frac{1}{2}\left(1 - \cos\left(\frac{\pi(t-t_0)}{t_f - t_0}\right)\right), \; t_0 \leq t \leq t_f, \tag{23}$$

where t_0 and t_f are the initial and final transition times, respectively. The constant δ is useful to avoid the indetermination at $t \geq t_f$ when $\lambda(t \geq t_f) = 0/\delta$. By visual inspection of Fig. 5, we can verify how the involved time varying functions $h(t)$, $\dot{h}(t)$ and $\lambda(t)$ reach their final values at $t = t_f$.

In the following lemma, we introduce a visual controller that uses the described terminal attractor to replace the classical computation of the velocity command given in (7) and (14) for eye-in-hand and eye-to-hand configurations respectively. Since one of these two types of configurations can be used, in the sequel of the paper, the symbol ∓ or ± will appear to establish the sign of some expressions. The correct sign in each expression is the upper sign if we have an eye-in-hand configuration or the lower sign for an eye-to-hand configuration.

Lemma 1: *The time parametrized controller*

$$\boldsymbol{\nu}_c = \mp\lambda(t)\hat{\mathbf{L}}_s^{-1}\mathbf{e}, \tag{24}$$

where $\lambda(t)$ is given by (22), ensures finite-time convergence of the error dynamics (6) for eye-in-hand configuration, respectively (13) for eye-to-hand configuration, to the origin $\mathbf{e} = \mathbf{0} \in \mathbb{R}^6$.

Proof. Notice that due to the time-dependent gain $\lambda(t)$, the control law (24) makes the closed-loop system time-dependent. Introducing (24) in (6) or (13) with the correct sign depending if the configuration is eye-in-hand or eye-to-hand respectively, we have in both cases:

$$\dot{\mathbf{e}} = -\lambda\left(t\right)\mathbf{L}_s\hat{\mathbf{L}}_s^{-1}\mathbf{e}. \tag{25}$$

Hence, (25) is a non-autonomous system with equilibrium point $\mathbf{e} = \mathbf{0} \in \mathbb{R}^6$. An stability analysis for this system must consider the property of non-autonomy (Khalil and Grizzle (2002)). Let us define the following candidate Lyapunov function:

$$V\left(\mathbf{e}\right) = \frac{1}{2}\mathbf{e}^T\mathbf{e}, \tag{26}$$

which is continuously differentiable, positive definite and decrescent as required (Khalil and Grizzle (2002)). The last two properties mean that the inequality $W_1\left(\mathbf{e}_i\right) \leq V\left(\mathbf{e}_i\right) \leq W_2\left(\mathbf{e}\right)$ is accomplished for all $\mathbf{e} \in \mathbb{R}^6$. Because the candidate Lyapunov function (26) does not explicitly depend on time, $W_1\left(\mathbf{e}\right)$ and $W_2\left(\mathbf{e}\right)$ are trivially found to be $W_1\left(\mathbf{e}\right) = W_2\left(\mathbf{e}\right) = V\left(\mathbf{e}\right)$. The time-derivative of the candidate Lyapunov function is

$$\dot{V} = -\lambda\left(t\right)\mathbf{e}^T\mathbf{L}_s\hat{\mathbf{L}}_s^{-1}\mathbf{e}.$$

To show that the equilibrium point $\mathbf{e} = \mathbf{0}$ of the non-autonomous system (25) is asymptotically stable, a positive definite function $W_3\left(\mathbf{e}\right)$ must be found such that:

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial \mathbf{e}}\dot{\mathbf{e}} \leq -W_3\left(\mathbf{e}\right). \tag{27}$$

The term $\partial V/\partial t$ is zero because $V\left(\mathbf{e}\right)$ is not a function of time. Additionally, we have that:

$$\frac{\partial V}{\partial \mathbf{e}}\dot{\mathbf{e}} = \dot{V} = -\lambda\left(t\right)\mathbf{e}^T\mathbf{L}_s\hat{\mathbf{L}}_s^{-1}\mathbf{e}. \tag{28}$$

Since $\lambda(t)$ in (22) satisfies a lower bound condition:

$$\lambda(t) \geq \lambda_0 > 0.$$

Then, we have:

$$\frac{\partial V}{\partial \mathbf{e}}\dot{\mathbf{e}} = -\lambda\left(t\right)\mathbf{e}^T\mathbf{L}_s\hat{\mathbf{L}}_s^{-1}\mathbf{e} \leq -\lambda_0\mathbf{e}^T\mathbf{L}_s\hat{\mathbf{L}}_s^{-1}\mathbf{e}. \tag{29}$$

Therefore, we can set $W_3\left(\mathbf{e}\right) = \lambda_0\mathbf{e}^T\mathbf{L}_s\hat{\mathbf{L}}_s^{-1}\mathbf{e}$ and asymptotic stability is guaranteed if and only if $\mathbf{L}_s\hat{\mathbf{L}}_s^{-1} > 0$, i.e., the matrix $\mathbf{L}_s\hat{\mathbf{L}}_s^{-1}$ is positive definite. This is the typical stability condition for a visual servo controller, however, in this case the additional condition $\lambda(t) \geq \lambda_0 > 0$ on the time parametrized function must be satisfied. For the PBVS (9), the condition $\mathbf{L}_s\hat{\mathbf{L}}_s^{-1} > 0$ holds globally for a good estimation of the pose parameters. For the IBVS (12), the condition $\mathbf{L}_s\hat{\mathbf{L}}_s^{-1} > 0$ holds locally around $\mathbf{e} = \mathbf{0}$, as shown in (Benhimane and Malis (2007)).

Now, let us prove the finite-time convergence of the closed-loop error dynamics (25). Assuming that the error dynamics (25) are decoupled by $\hat{\mathbf{L}}_s^{-1}$ (which is feasible given a good estimation of the interaction matrix), we have that the individual dynamics are:

$$\dot{e}_i(t) = -\lambda(t)e_i(t) \text{ for } i = 1, ..., 6$$

with $\lambda(t)$ as in (22). By separation of variables and integration, the solution of each individual error dynamics using the TBG gain (22) become:

$$e_i(t) = c_0 e^{-\lambda_0(t-t_0)} \left(1 - h(t) + \delta\right), \tag{30}$$

where c_0 is found out evaluating the previous equation for $t = t_0$. Thus, $c_0 = \frac{e(t_0)}{1+\delta}$. Replacing c_0 in (30) and operating we have:

$$e_i(t) = e_i(t_0)e^{-\lambda_0(t-t_0)} \left(1 - \frac{h(t)}{1+\delta}\right). \tag{31}$$

Since λ_0 is a small gain, the exponential term is a slow dynamic. However $1 - \frac{h(t)}{1+\delta} \to 0$ as $t \to t_f$. Thus, the shape of $e(t)$ closely follow the profile of $1 - h(t)$ and the convergence time is controlled by $h(t)$. This means that $e(t)$ is a smooth function that asymptotically converges to zero at $t = t_f$ regardless of the initial condition $e(t_0)$ and finite-time convergence of the error is ensured. □

4. VISUAL CONTROL WITH OBSTACLE AVOIDANCE

The visual task considers the error between the current and target images to control the humanoid robot. However, if the camera detects a static obstacle while the robot walks toward the target location, then the robot should activate the obstacle avoidance as its primary task and, as a consequence, the degree of importance of the visual task decreases. This means that two tasks should be simultaneously solved within a time interval without causing conflicts in the locomotion behavior of the humanoid.

A widely used strategy to handle the compromise of active tasks consists of defining a strict hierarchy between them (Lee et al. (2012)). Following this scheme, the visual task is the single active task until a security distance (d_s) between the obstacle and the robot is violated. If this occurs, then the obstacle avoidance task becomes the primary active task and the solution of the visual task is projected onto the null-space of the primary task Jacobian. Otherwise, the obstacle avoidance task must remain inactive.

In particular, the task function to avoid an obstacle is defined as:

$$e_o = x_o - d_s \in \mathbb{R}, \tag{32}$$

such that:

$$x_o = ||\boldsymbol{x}_r - \boldsymbol{x}_o||, \tag{33}$$

where $\boldsymbol{x}_r = [x_r\ y_r]^T \in \mathbb{R}^2$ and $\boldsymbol{x}_o = [x_o\ y_o]^T \in \mathbb{R}^2$ corresponds to the relative robot and obstacle positions w.r.t. the current frame $\mathcal{F}$. By differentiating the task error (32) w.r.t. the vector $\boldsymbol{x}_r - \boldsymbol{x}_o$, we obtain the gradient of the task:

$$\boldsymbol{g}_o = \left(\frac{\boldsymbol{x}_r - \boldsymbol{x}_o}{x_o}\right)^T \in \mathbb{R}^{1\times 2}. \tag{34}$$

Then, assuming a common reference frame to express the obstacle avoidance task and the visual task, e.g. the target frame $\mathcal{F}^*$, we can express the kinematics of the obstacle avoidance task as follows:

$$\dot{e_o} = \boldsymbol{J}_o \boldsymbol{\nu}_c, \tag{35}$$

where $\boldsymbol{J}_o$ is the obstacle avoidance Jacobian given by:

$$\boldsymbol{J}_o = \begin{pmatrix} \boldsymbol{g}_o & \mathbf{0}_{1\times 4} \end{pmatrix} \in \mathbb{R}^{1\times 6} \tag{36}$$

and its corresponding null-space projector is given by:

$$\boldsymbol{Q}_o = \boldsymbol{I} - \boldsymbol{J}_o^T \boldsymbol{J}_o \in \mathbb{R}^{6\times 6}. \tag{37}$$

If the obstacle avoidance task is not active, i.e. if $e_o > 0$, the control law is obtained as in (24). Otherwise, if $e_0 \leq 0$, the control law is given by a hierarchical task-based scheme expressed as:

$$\begin{aligned} \boldsymbol{\nu}_c &= \boldsymbol{\nu}_o + \boldsymbol{\nu}_{\mathrm{v|o}}, \\ \boldsymbol{\nu}_o &= -\lambda_o \boldsymbol{J}_o^+ e_o, \\ \boldsymbol{\nu}_{\mathrm{v|o}} &= (\boldsymbol{J}_{\mathrm{v}} \boldsymbol{Q}_o)^+ \left(\mp\lambda_{\mathrm{v}}(t)\mathbf{e} - \boldsymbol{J}_{\mathrm{v}} \boldsymbol{\nu}_o\right), \end{aligned} \tag{38}$$

where $\boldsymbol{J}_o^+ = \boldsymbol{J}_o^T(\boldsymbol{J}_o \boldsymbol{J}_o^T)^{-1}$ is the Moore-Penrose pseudoinverse, $\boldsymbol{J}_{\mathrm{v}} = \hat{\mathbf{L}}_s$ is the corresponding interaction matrix, $\boldsymbol{\nu}_o$ is the velocity vector's solution of the obstacle avoidance task (high priority task) and $\boldsymbol{\nu}_{\mathrm{v|o}}$ is the velocity vector's solution for the visual task (secondary task) that takes into account the obstacle avoidance task.

An undesired effect of the instantaneous switching between the control laws (24) and (38) is an abrupt change in the input signals of the locomotion controller, i.e. the velocity profiles of the robot's CoM (Keith et al. (2011)). This is a well-known problem in the visual control community where some solutions have been suggested, e.g., (Mansard et al. (2009), Cherubini and Chaumette (2013)). In particular, we adopted the intermediate value strategy introduced in (Lee et al. (2012)). This scheme rewrites the control law (38) as:

$$\begin{aligned} \boldsymbol{\nu}'_c &= \boldsymbol{\nu}'_o + \boldsymbol{\nu}'_{\mathrm{v|o}}, \\ \boldsymbol{\nu}'_o &= \boldsymbol{J}_o^+ \dot{e}'_o, \\ \boldsymbol{\nu}'_{\mathrm{v|o}} &= (\boldsymbol{J}_{\mathrm{v}} \boldsymbol{Q}_o)^+ \left(\mp\lambda_{\mathrm{v}}(t)\mathbf{e} - \boldsymbol{J}_{\mathrm{v}} \boldsymbol{\nu}'_o\right), \end{aligned} \tag{39}$$

where

$$\dot{e}'_o \;=\; -\lambda_o h_{\mathrm{t}}(t) e_o \mp \left(1 - h_{\mathrm{t}}(t)\right) \boldsymbol{J}_o \boldsymbol{J}_{\mathrm{v}}^+ \lambda_{\mathrm{v}}(t)\mathbf{e} \tag{40}$$

encodes the intermediate desired values that can be interpreted as the weighted combination of the obstacle and visual tasks. The shape of the proposed transition function $h_{\mathrm{t}}(t)$ is the same as the TBG defined in (23) for an adequate transition time interval $t_{\mathrm{t},0} \leq t \leq t_{\mathrm{t},f}$.

The following theorem analyzes the convergence of the time-constrained hierarchical control law with smooth transitions (39).

Theorem 1: *Consider the stack of two tasks: obstacle avoidance and visual servo control, which can be represented by the following extended kinematics:*

$$\dot{\mathbf{e}}' = \begin{pmatrix} \boldsymbol{J}_o \\ \pm\boldsymbol{J}_v \end{pmatrix} \boldsymbol{\nu}_c, \tag{41}$$

where $\mathbf{e}' = \begin{pmatrix} e_o' & \mathbf{e}^T \end{pmatrix}^T$. *The control law (39) ensures finite-time convergence to the origin* $\mathbf{e}' = \mathbf{0}$.

Proof. A Lyapunov candidate function considering the two time-constrained hierarchical tasks in transition (39) is:

$$V(\mathbf{e}') = \frac{1}{2}\mathbf{e}'^T\mathbf{e}'. \tag{42}$$

The derivative w.r.t. time of (42) is:

$$\dot{V}(\mathbf{e}') = \mathbf{e}'^T\dot{\mathbf{e}}' = \mathbf{e}'^T \begin{pmatrix} \boldsymbol{J}_o \\ \pm\mathbf{L}_s \end{pmatrix} \boldsymbol{\nu}_c = \left(e_o\boldsymbol{J}_o \pm \mathbf{e}^T\mathbf{L}_s\right)\boldsymbol{\nu}_c. \tag{43}$$

By substituting (39) into (43), making $\boldsymbol{\nu}_c = \boldsymbol{\nu}_c'$ and organizing the terms in matrix form, we obtain:

$$\dot{V} = -\mathbf{e}'^T\mathbf{M}\mathbf{e}', \tag{44}$$

where

$$\mathbf{M} \;=\; \begin{pmatrix} \boldsymbol{M}_a & \boldsymbol{M}_b \\ \boldsymbol{M}_c & \boldsymbol{M}_d \end{pmatrix}, \tag{45}$$

with

$$\begin{aligned}
\boldsymbol{M}_a &= \lambda_o h_{\mathrm{t}}(t)\left(\boldsymbol{J}_o\boldsymbol{J}_o^+ - \boldsymbol{J}_o(\hat{\mathbf{L}}_s\boldsymbol{Q}_o)^+\hat{\mathbf{L}}_s\boldsymbol{J}_o^+\right), \qquad (46)\\
\boldsymbol{M}_b &= \lambda_{\mathrm{v}}(t)\boldsymbol{J}_o\left(\pm(\hat{\mathbf{L}}_s\boldsymbol{Q}_o)^+ + (1-h_{\mathrm{t}}(t))\left(\pm\boldsymbol{J}_o^+\boldsymbol{J}_o\hat{\mathbf{L}}_s^+ \mp (\hat{\mathbf{L}}_s\boldsymbol{Q}_o)^+\hat{\mathbf{L}}_s\boldsymbol{J}_o^+\boldsymbol{J}_o\hat{\mathbf{L}}_s^+\right)\right),\\
\boldsymbol{M}_c &= \lambda_o h_{\mathrm{t}}(t)\left(\pm\mathbf{L}_s\boldsymbol{J}_o^+ \mp \mathbf{L}_s(\hat{\mathbf{L}}_s\boldsymbol{Q}_o)^+\hat{\mathbf{L}}_s\boldsymbol{J}_o^+\right),\\
\boldsymbol{M}_d &= \lambda_{\mathrm{v}}(t)\mathbf{L}_s\left((\hat{\mathbf{L}}_s\boldsymbol{Q}_o)^+ + (1-h_{\mathrm{t}}(t))\left(\boldsymbol{J}_o^+\boldsymbol{J}_o\hat{\mathbf{L}}_s^+ - (\hat{\mathbf{L}}_s\boldsymbol{Q}_o)^+\hat{\mathbf{L}}_s\boldsymbol{J}_o^+\boldsymbol{J}_o\hat{\mathbf{L}}_s^+\right)\right),
\end{aligned}$$

By using the definition of the pseudoinverse, it can be verified that:

$$\begin{aligned}
\boldsymbol{J}_o(\hat{\mathbf{L}}_s\boldsymbol{Q}_o)^+ &= \boldsymbol{J}_o\boldsymbol{Q}_o\hat{\mathbf{L}}_s^{-1}, \qquad (47)\\
\mathbf{L}_s(\hat{\mathbf{L}}_s\boldsymbol{Q}_o)^+ &= \mathbf{L}_s\hat{\mathbf{L}}_s^{-1}. \qquad (48)
\end{aligned}$$

Additionally, we have the following properties from the Jacobian of the obstacle avoidance task:

$$\begin{aligned}
\boldsymbol{J}_o\boldsymbol{J}_o^+ &= \boldsymbol{J}_o\boldsymbol{J}_o^T = 1, \qquad (49)\\
\boldsymbol{J}_o\boldsymbol{Q}_o &= \mathbf{0}. \qquad (50)
\end{aligned}$$

Then, using the previous identities, the submatrices of $\mathbf{M}$ in (46) reduce as follows:

$$\begin{aligned} \boldsymbol{M}_a &= \lambda_o h_{\mathrm{t}}(t) \in \mathbb{R}, \\ \boldsymbol{M}_b &= \pm\lambda_{\mathrm{v}}(t)\left(1 - h_{\mathrm{t}}(t)\right) \boldsymbol{J}_o \hat{\mathbf{L}}_s^{+} \in \mathbb{R}^{1\times 6}, \\ \boldsymbol{M}_c &= \mathbf{0} \in \mathbb{R}^{6\times 1}, \\ \boldsymbol{M}_d &= \lambda_{\mathrm{v}}(t)\mathbf{L}_s\hat{\mathbf{L}}_s^{-1} \in \mathbb{R}^{6\times 6}. \end{aligned} \tag{51}$$

Therefore:

$$\mathbf{M} = \begin{pmatrix} \lambda_o h_{\mathrm{t}}(t) & \pm\lambda_{\mathrm{v}}(t)\left(1 - h_{\mathrm{t}}(t)\right) \boldsymbol{J}_o \hat{\mathbf{L}}_s^{+} \\ \mathbf{0}_{6\times 1} & \lambda_{\mathrm{v}}(t)\mathbf{L}_s\hat{\mathbf{L}}_s^{-1} \end{pmatrix}. \tag{52}$$

We have that $\lambda_o, \lambda_{\mathrm{v}}(t) > 0$ and $0 \leq h_{\mathrm{t}}(t) \leq 1$. Since the eigenvalues of matrix $\mathbf{M}$ only depend on $\lambda_o, \lambda_{\mathrm{v}}(t), h_{\mathrm{t}}(t)$ and assuming that $\mathbf{L}_s\hat{\mathbf{L}}_s^{-1} > 0$, from Lemma 1 we conclude that the error $\mathbf{e}'$ converges to the origin in finite-time. Hence, the convergence of the time-constrained hierarchical tasks in transition is ensured. □

The previous theorem allows us to claim that the visual servo control task is accomplished in a predefined time in spite of the occurrence of an obstacle avoidance task.

5. EXPERIMENTAL EVALUATION

We have tested the proposed approach using a NAO humanoid robot. For evaluation purposes, we decided to use the position-based scheme for experiments in eye-to-hand configuration and the image-based scheme for experiments in eye-in-hand configuration. In both cases, the robot was calibrated with respect to a fixed top camera to provide ground truth and measure the visual task performance.

5.1 *Visual control with TBG*

In this section, we report some results to evaluate the visual control approach using the TBG to achieve finite-time convergence only for the positioning task, i.e., without obstacles. We report three experiments with three different initial conditions using the eye-in-hand configuration. We set $t_0 = 0$ and $t_f = 39$ for $h(t)$ and $\delta = 0.001$. The performance of the positioning task is shown in Fig. 6. The convergence of the robot to the desired pose for the same time t_f from the three different initial locations can be appreciated from the plots in Fig. 7.

In the case of eye-in-hand configuration, we used the top camera mounted on the robot's head and the camera velocities are computed using (12). The images were obtained at a frame rate of 12Hz with a resolution of 640×480 pixels. The image features were acquired as follows: first, a corner detector based on (Shi and Tomasi (1994)) was used, which is implemented on the function `goodFeaturesToTrack` of the OpenCV library. Then, we assigned a SIFT descriptor (Lowe (1999)) to each detected point. A robust matcher based on RANSAC (Fischler and Bolles (1981)) matched all the points between the current image and the target image. Finally, we used the function `calcOpticalFlowPyrLK` from OpenCV, which tracks the points using an algorithm based on a sparse iterative version of the Lucas-Kanade optical flow in pyramids. The tracking of points from a camera mounted on a humanoid robot is not trivial due to the jerky camera movements generated by the robot's gait. Thus, the tracker was experimentally tuned to deal with this unavoidable behavior. We evaluated experimentally the performance of the tracker with the NAO robot. It was found that the tracker is able to work properly at the maxima forward, lateral and rotational

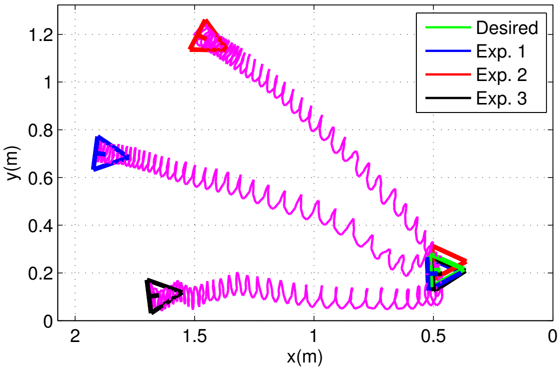


Figure 6. **Performance of the positioning task for TBG experiments.** The green triangle displays the desired pose of the robot and the blue, red and black triangles display the three experiments with different initial poses.

velocities whereas the maximum step sizes are 0.04m for forward direction, 0.14m for lateral and 0.349rad for rotation. The `DLT` function of ViSP library was used to compute the homography. It is important to say that it is possible to compute the homography for non-planar scenes using the implementation proposed in (Malis et al. (2000)), as we did in a previous work (Delfin et al. (2014)).

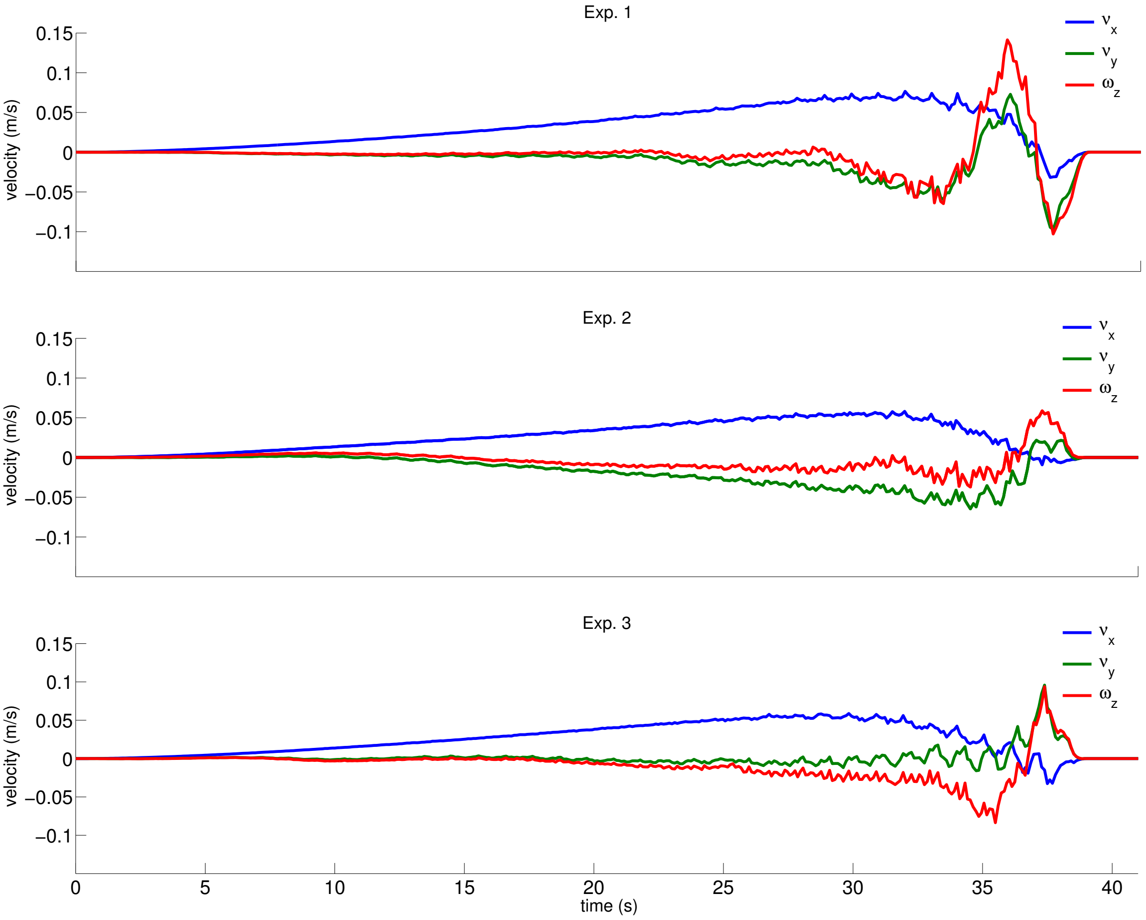


Figure 7. **Input velocities to the WPG for experiments with TBG.** From three different initial locations, the convergence is achieved at a given final time $t_f = 39$.

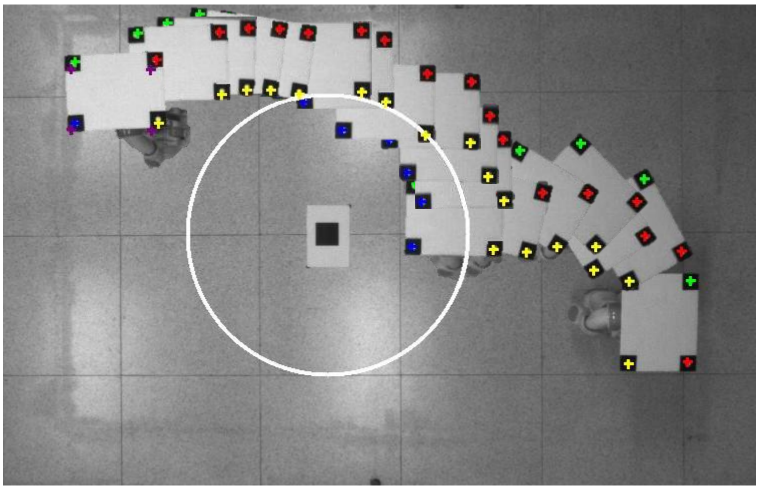


Figure 8. **Snapshots of the eye-to-hand configuration.** The purple crosses display the desired pose of the robot and the blue, red, yellow and green crosses display the current pose. We can see the convergence of the robot in the image to the desired position of the pixels.

5.2 *Visual control with TBG and obstacle avoidance*

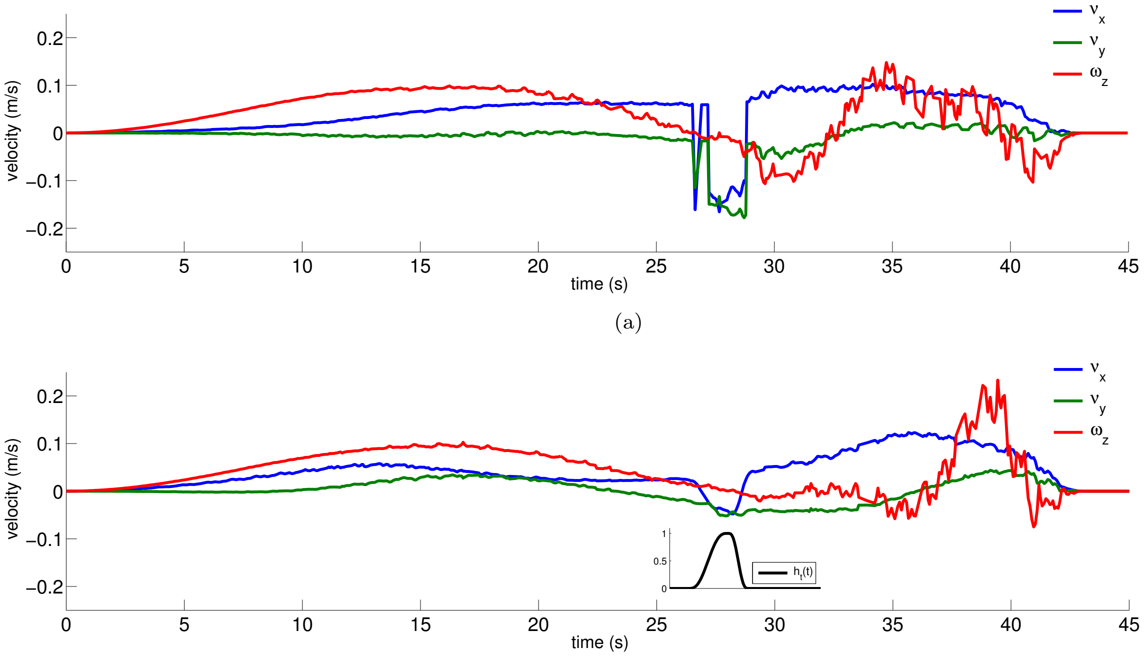
In this section, we present results of the proposed hierarchical task-based approach with smooth transitions, where the positioning task given in terms of a visual target must be finished in a desired time, in spite of the execution of an obstacle avoidance task when needed.

5.2.1 Eye-to-hand configuration

For the case of eye-to-hand configuration, an uEye 120SE-M-GL camera was mounted at 2.9m from the floor obtaining images at a frame rate of 16Hz with a resolution of 752×480 pixels. A pattern of points was placed on the NAO's head and the points were tracked, as seen in Fig. 8, and used to compute the required homography. In this case, the camera velocities are computed using (19). The initial selection of the points was done manually and they were tracked using the function `calcOpticalFlowPyrLK` from OpenCV as in the eye-in-hand experiments. The centroid of the four points was used to compute x_o, i.e. the distance between the robot and the obstacle.

For the reported experiments, the reference pose is displaced and rotated 1.67m and 86 degrees respectively from the initial pose. A landmark was used to simulate the obstacle and it was placed at 0.97m from the initial pose, we set $d_s = 140$ pixels $\simeq 0.39$m. The robot is controlled in order to achieve the reference pose while the obstacle is avoided using a switching between the control laws (24) and (38) for one experiment and (39) for another, both with $\lambda_o = 0.6$. For the TBG, we set $t_0 = 0$ and $t_f = 43$ for $h(t)$ and $\delta = 0.001$. We defined the transition time interval $t_{\mathrm{t},f} - t_{\mathrm{t},0} = 3$ for the obstacle avoidance using the transition function $h_{\mathrm{t}}(t)$ as defined in (23). During the experiments, we measured the Center of Pressure (CoP) of each foot, but only in the simple support phase of the gait. We recorded the lateral and frontal positions of the CoP using the pressure sensors on the robot's soles as seen in Figs. 10 and 11, where the dashed lines represent the edges of the robot's soles.

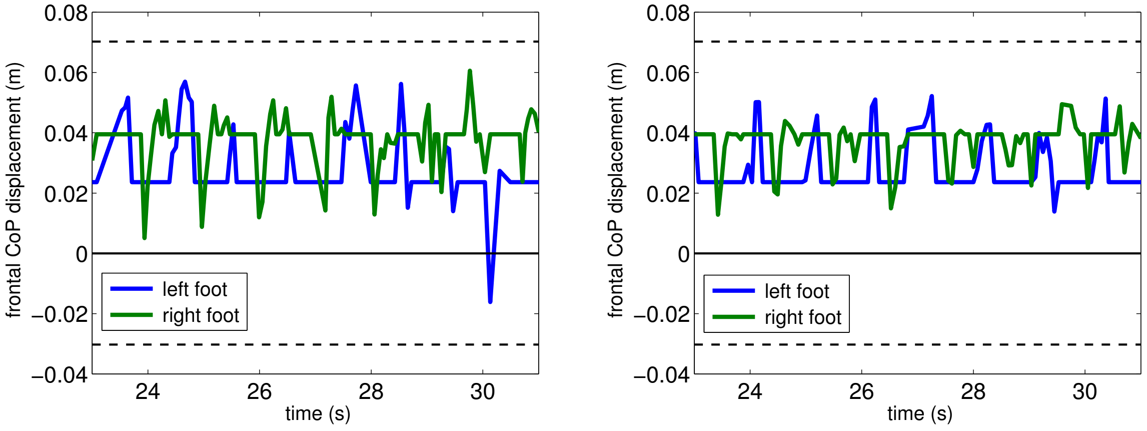
Fig. 9 shows the velocities for both experiments, without and with smooth transitions respectively. In Fig. 9(a) we can see the input velocities to the WPG using a switching between the control laws (24) and (38). It can be seen discontinuities caused by the switching between one and other control law between $t = 25$ and $t = 30$ when the security distance is violated. This discontinuity produces that the CoP gets close to the edge of the sole unbalancing the robot and consequently it might fall. This behavior can be observed in the Figs. 10(a) and 11(a), mainly in the frontal reading of the left foot where the CoP around the second 30 gets close to the edge. It can be seen



(a)

(b)

Figure 9. **Input velocities to the WPG for eye-to-hand configuration.** (a) Input velocities using a switching between the control laws (24) and (38). (b) Input velocities with smooth transitions using the control law (39). It can be seen the transition function $h_t(t)$ during the activation interval.



(a)

(b)

Figure 10. **Frontal displacement of the CoP for eye-to-hand configuration.** (a) Measurement of the CoP displacement during the interval of switching between the control laws (24) and (38). (b) Measurement of the CoP displacement during the activation of the transition function $h_t(t)$ using the control law (39).

that the reaction to the discontinuity is not instantaneous due to the humanoid locomotion.

When the transition function $h_t(t)$ is used, these discontinuities are alleviated. Fig. 9(b) shows the input velocities to the WPG using the control law (39). We can see the activation of the transition function $h_t(t)$ around the second 26. This smooth combination of the obstacle task and the visual task causes less displacement of the CoP as can be seen in Figs. 10(b) and 11(b). Thus, the control scheme with smooth transitions between tasks does not compromise the dynamic balance of the humanoid robot.

Fig. 12 shows the performance of the positioning task. The position of the obstacle is plotted as

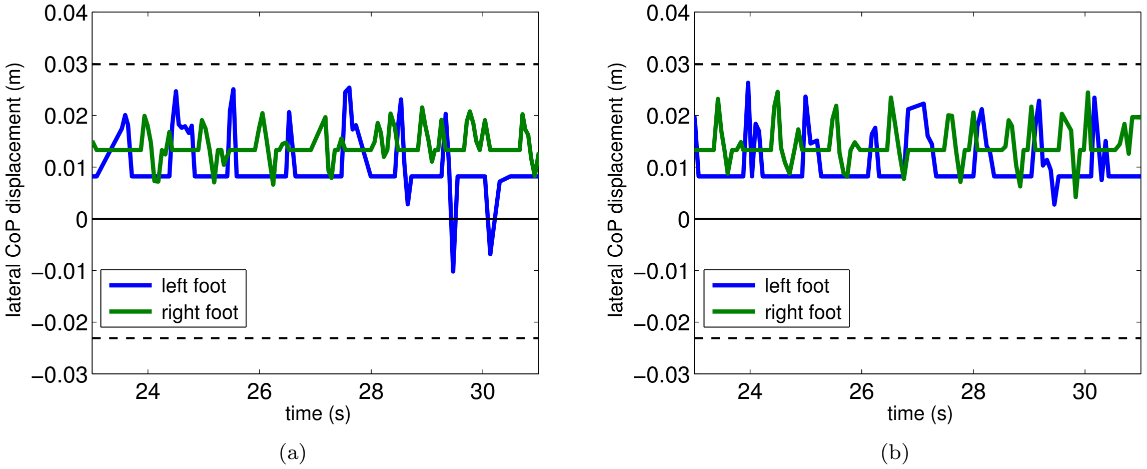


Figure 11. **Lateral displacement of the CoP for eye-to-hand configuration.** (a) Measurement of the CoP displacement during the interval of switching between the control laws (24) and (38). (b) Measurement of the CoP displacement during the activation of the transition function $h_t(t)$ using the control law (39).

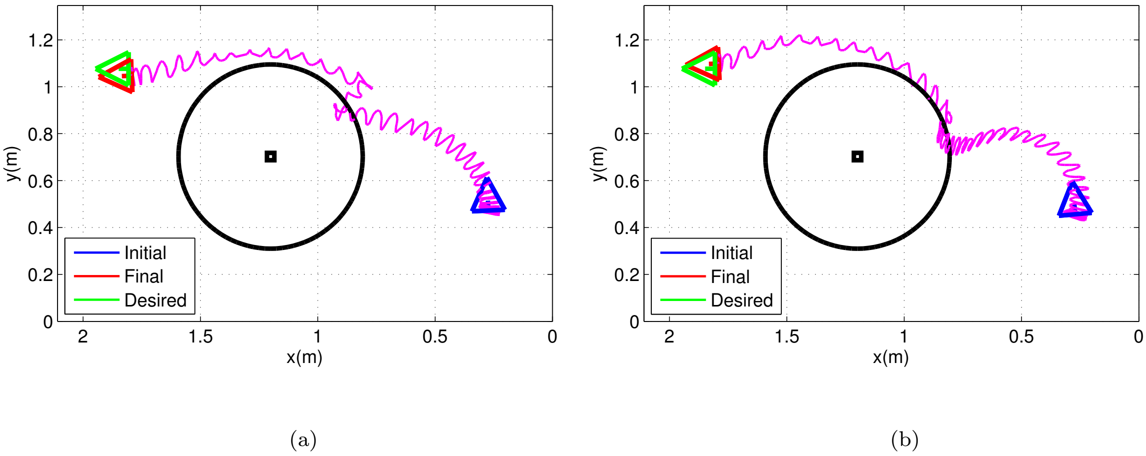


Figure 12. **Performance of the positioning task for eye-to-hand configuration.** (a) Path followed by the robot using a switching between the control laws (24) and (38). (b) Path followed by the robot with smooth transitions using the control law (39).

a black square and the security distance as the black circle. The switching of the control laws (24) and (38) is activated by e_o. In Fig. 12(a) this switching is clearly observed, i.e. when the robot invades to the security distance, it is ejected out of the circle. For the case of the control law (39), the transition function $h_t(t)$ is also activated by e_o, when the robot invades the security distance for the first time, as seen in Fig. 12(b). In this case the robot leaves smoothly the circle evading the obstacle. Some snapshots of the pattern of dots for this experiment are shown in Fig. 8, we can see the convergence of the pixels to their desired position. It is important to say that in all the tests done, the magnitude of the positioning error was less than 0.05m.

5.2.2 Eye-in-hand configuration

For the experiments with obstacle avoidance in the eye-in-hand configuration, an Asus Xtion Pro Live RGB-D sensor was used to measure the distance to the obstacle. It was mounted on the NAO's

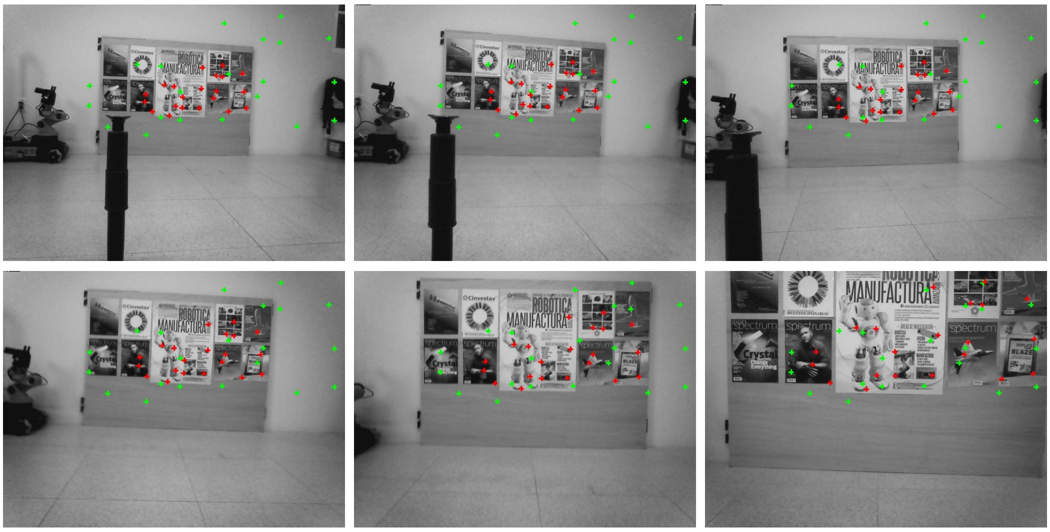


Figure 13. **Snapshots of the eye-in-hand configuration.** The green crosses display the desired position of the points in the image and the red crosses display the current position of points. We can see the convergence of the points in the image to their desired position.

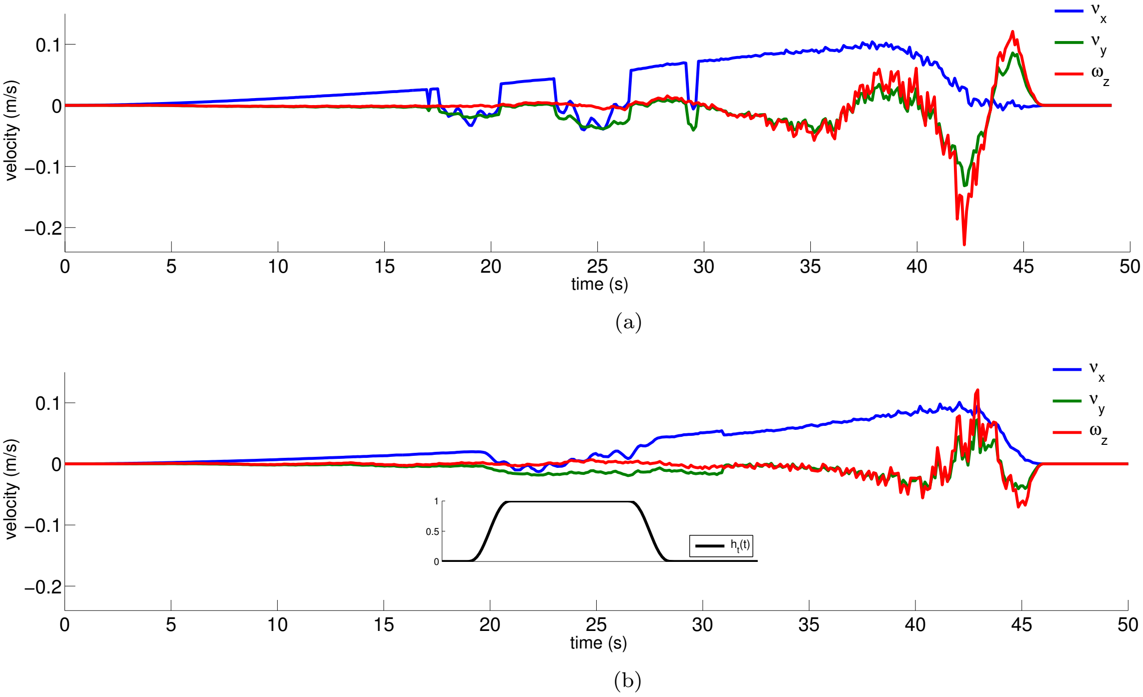


Figure 14. **Input velocities to the WPG for eye-in-hand configuration.** (a) Input velocities using a switching between the control laws (24) and (38). (b) Input velocities with smooth transitions using the control law (39). It can be seen the transition function $h_t(t)$ during the activation interval.

head and the detection of the obstacle was done using the OpenNI and OpenCV libraries. The reference image was captured at 1.55m and 10 degrees from the initial pose. A 0.44m long tube was used as obstacle and it was placed at 0.9m from the initial pose. Due to the operating range of the RGB-D sensor, we set $d_s = 0.50$m. As in the eye-to-hand configuration, the robot is controlled in order to achieve the reference pose using a switching between the control laws (24) and (38) for one experiment and (39) for another, both with $\lambda_o = 0.7$. In this case, the velocities associated to the visual task are computed using (12). We set $t_0 = 0$ and $t_f = 46$ for the TBG and $\delta = 0.001$. We

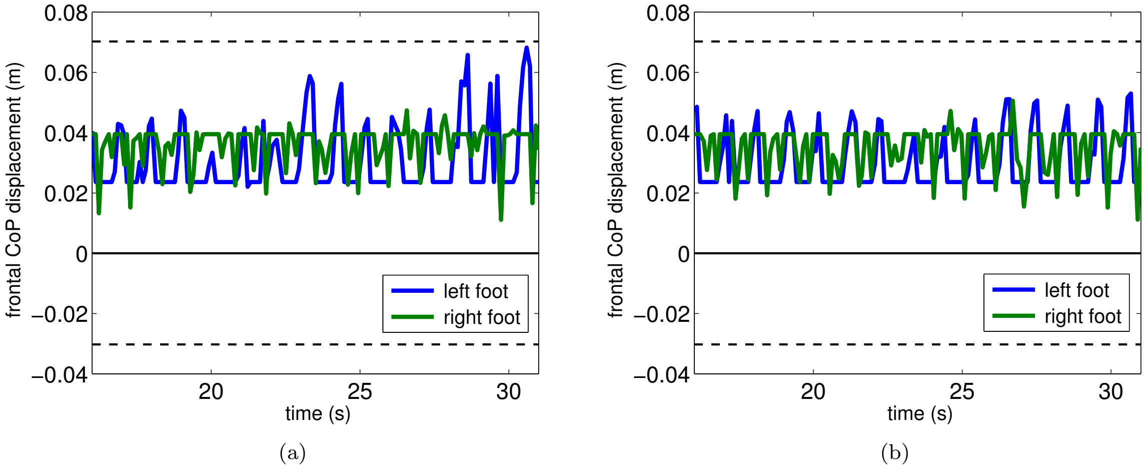


Figure 15. **Frontal displacement of the CoP for eye-in-hand configuration.** (a) Measurement of the CoP displacement during the interval of switching between the control laws (24) and (38). (b) Measurement of the CoP displacement during the activation of the transition function $h_t(t)$ using the control law (39).

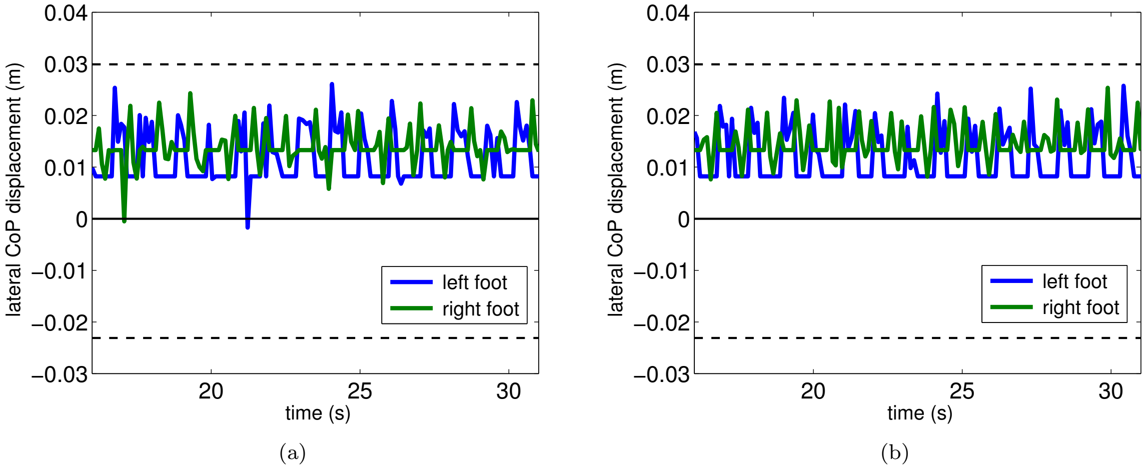


Figure 16. **Lateral displacement of the CoP for eye-in-hand configuration.** (a) Measurement of the CoP displacement during the interval of switching between the control laws (24) and (38). (b) Measurement of the CoP displacement during the activation of the transition function $h_t(t)$ using the control law (39).

defined for these experiments the transition time interval $t_{t,f} - t_{t,0} = 9$ for the obstacle avoidance. Also, we made a measurement of the CoP in each foot.

Fig. 13 shows some snapshots of the images captured by the NAO's camera during the experimental run using the control law (39). The points marked as red crosses are the tracked points in the current robot view. The green crosses are the corresponding points of the target image.

Fig. 14 displays the velocities for both experiments, and the Figs. 15 and 16 display the measurements of the robot's CoP. We can observe a similar behavior as in the eye-to-hand experiments regarding the discontinuities in the velocities and the displacement of the CoP caused by the switching of control laws without smooth transitions. The discontinuities are alleviated once the smooth transition is used.

The performance of the positioning task for these experiments is shown in Fig. 17. It can be seen that unlike the eye-to-hand experiments, the robot invades the security distance because the obstacle leaves the field of view of the sensor. This field of view is plotted in the Fig. 17 as the cyan cones for three different robot's poses: at the initial pose, at the pose when the obstacle leaves

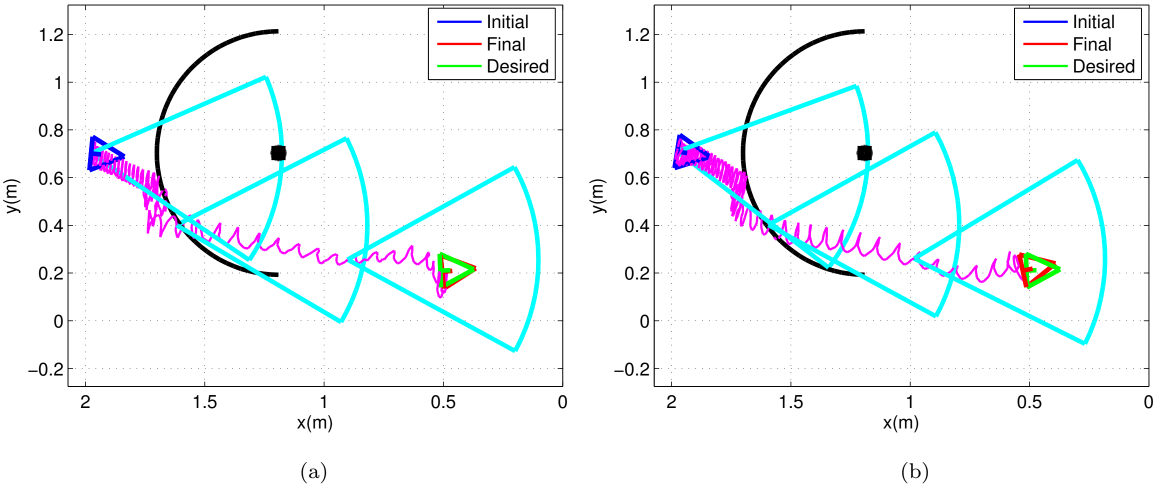


Figure 17. **Performance of the positioning task for eye-in-hand configuration.** The cyan cones depicts the field of view of the RGB-D sensor for three samples poses. (a) Path followed by the robot using a switching between the control laws (24) and (38). (b) Path followed by the robot with smooth transitions using the control law (39).

the RGB-D field of view and at the pose when the robot leaves the security distance. Notice that although the robot violates the security distance, the obstacle is avoided properly. This effect is due to the constrained field of view of the distance sensor and can be overcome by using another distance sensor with larger measurement range and field of view.

6. CONCLUSIONS

In this work, we have proposed a vision-based humanoid locomotion approach that uses monocular images to solve the problem of driving a humanoid robot to a desired location. The proposed approach is a decoupled visual servo walking control, where a reference velocity computed from a visual servoing scheme is given as input to a walking pattern generator. This approach facilitates the introduction of a time constraint to complete the positioning task, which is advantageous in contrast to the classical exponential convergence. We verified that the proposed approach achieves finite-time convergence using smooth robot velocities and considering the omnidirectional waking capability of the robot. Also, the proposed approach is able to handle simultaneously the visual positioning task while avoiding an unexpected obstacle without altering the desired time of convergence. Moreover, the continuity of the velocity profiles are maintained when task transitions occur. The stability analysis has been done to demonstrate the convergence of the two task-based control. We have shown theoretically and experimentally that the proposed approach can be applied using different visual servoing schemes and camera-robot configurations, which shows the generality of the approach. The online experiments were conducted with a NAO robot for evaluation purposes.

For future work, we consider extending the approach to handle more than two tasks in the hierarchical scheme, dealing with different types of obstacles: obstacles that can be surrounded or obstacles that can be passed over. In this case, the robot must identify the type of obstacle and decide how to carry out the obstacle avoidance task while keeping the smoothness of robot velocities. Additionally, in order to enlarge the workspace of the visual task, which is limited by the matching of visual features between the current and target images, we want to extend the proposed approach to follow a path that is defined in terms of a sequence of reference images as done in (Delfin et al. (2014)). The sequence of reference images is previously acquired in a supervised teaching phase. Thus, autonomous navigation can be performed by replaying the visual path but taken into account the avoidance of obstacles not present during the teaching phase.

References

Benhimane, S. and Malis, E. (2007). Homography-based 2D visual tracking and servoing. *Int. Journal of Robotics Research*, 26(7):661–676.

Chaumette, F. and Hutchinson, S. (2006). Visual servo control. Part I: Basic approaches. *IEEE Robotics and Automation Magazine*, 13(4):82–90.

Cherubini, A. and Chaumette, F. (2013). Visual navigation of a mobile robot with laser-based collision avoidance. *Int. Journal of Robotics Research*, 32(2):189–205.

Cupec, R., Schmidt, G., and Lorch, O. (2005). Vision-guided walking in a structured indoor scenario. *Automatika*, 46(1-2):49–57.

Delfin, J., Becerra, H. M., and Arechavaleta, G. (2014). Visual path following using a sequence of target images and smooth robot velocities for humanoid navigation. In *IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, pages 354–359.

Dune, C., Herdt, A., Stasse, O., Wieber, P.-B., Yokoi, K., and Yoshida, E. (2010). Cancelling the sway motion of dynamic walking in visual servoing. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 3175–3180.

Faragasso, A., Oriolo, G., Paolillo, A., and Vendittelli, M. (2013). Vision-based corridor navigation for humanoid robots. In *IEEE Int. Conf. on Robotics and Automation*, pages 3175–3180.

Fischler, M. A. and Bolles, R. C. (1981). Ransac random sampling concensus: A paradigm for model fitting with applications to image analysis and automated cartography. In *Communications of ACM*, pages 181–395.

García, M., Stasse, O., Hayet, J. B., Dune, C., Esteves, C., and Laumond, J. P. (2014). Vision-guided motion primitives for humanoid reactive walking: decoupled versus coupled approaches. *Int. Journal of Robotics Research*, pages 1–18.

Hayet, J.-H., Esteves, C., Arechavaleta, G., Stasse, O., and Yoshida, E. (2012). Humanoid locomotion planning for visually-guided tasks. *Int. Journal of Humanoid Robots*, 9(2):26.

Herdt, A., Diedam, H., Wieber, P.-B., Dimitrov, D., Mombaur, K., and Diehl, M. (2010). Online walking motion generation with automatic footstep placement. *Advanced Robotics*, 24(5-6):719–737.

Ido, J., Shimizu, Y., Matsumoto, Y., and Ogasawara, T. (2009). Indoor navigation for a humanoid robot using a view sequence. *Int. Journal of Robotics Research*, 28(2):315–325.

Jarquín, G., Arechavaleta, G., and Parra-Vega, V. (2013). Continuous kinematic control with terminal attractors for handling task transitions of redundant robots. In *IEEE Int. Conf. on Robotics and Automation*, pages 1976–1981.

Kajita, S., Hirukawa, H., Harada, K., and Yokoi, K. (2014). *Introduction to Humanoid Robotics.* Book series: Springer Tracts in Advanced Robotics 101. Springer Berlin Heidelberg.

Kajita, S., Kanehiro, F., Kaneko, K., Fujiwara, K., Harada, K., Yokoi, K., and Hirukawa, H. (2003). Biped walking pattern generation by using preview control of zero-moment point. In *IEEE Int. Conf. on Robotics and Automation*, pages 1620–1626.

Kanoun, O. (2011). Real-time prioritized kinematic control under inequality constraints for redundant manipulators. In *Robotics: Science and Systems VII*, Los Angeles, CA, USA.

Keith, F., Wieber, P.-B., Mansard, N., and Kheddar, A. (2011). Analysis of the discontinuities in prioritized task-space control under discrete task scheduling operations. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3887–3892, San Francisco, CA, USA.

Khalil, H. K. and Grizzle, J. W. (2002). *Nonlinear Systems.* Prentice Hall, Upper Saddle River.

Lee, J., Mansard, N., and Park, J. (2012). Intermediate desired value approach for task transition of robots in kinematic control. *IEEE Transactions on Robotics*, 28(6):1260–1277.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *IEEE Int. Conf. on Computer Vision*, volume 2, pages 1150–1157.

Maier, D., Stachniss, C., and Bennewitz, M. (2013). Vision-based humanoid navigation using self-supervised obstacle detection. *Int. Journal of Humanoid Robotics*, 10(2):1–28.

Malis, E., Chaumette, F., and Boudet, S. (2000). 2 1/2 Visual servoing with respect to unknown objects through a new estimation scheme of camera displacement. *Int. Journal of Computer Vision*, 37(1):79–97.

Malis, E. and Vargas, M. (2007). Deeper understanding of the homography decomposition for vision-based control. Technical report, INRIA, Sophia Antipolis.

Mansard, N., Remazeilles, A., and Chaumette, F. (2009). Continuity of varying-feature-set control laws. *IEEE Transanctions on Automatic Control*, 54(11):2493–2505.

Mansard, N., Stasse, O., Chaumette, F., and Yokoi, K. (2007). Visually-guided grasping while walking on a humanoid robot. In *IEEE Int. Conf. on Robotics and Automation*, pages 3041–3047.

Morasso, P., Sanguineti, V., and Spada, G. (1997). A computational theory of targeting movements based on force fields and topology representing networks. *Neurocomputing*, 15(3-4):411–434.

Oriolo, G., Paolillo, A., Rosa, L., and Vendittelli, M. (2013). Vision-based trajectory control for humanoid navigation. In *IEEE-RAS Int. Conf. on Humanoid Robots*, pages 118–123.

Ou, M., Li, S., and Wang, C. (2013). Finite-time tracking control for multiple non-holonomic mobile robots based on visual servoing. *Int. Journal of Control*, 86(12):2175–2188.

Shi, J. and Tomasi, C. (1994). Good features to track. In *IEEE Computer Vision and Pattern Recognition*, pages 593–600.

Triggs, B. (1998). Autocalibration from planar scenes. In Burkhardt, H. and Neumann, B., editors, *Computer Vision - ECCV'98*, volume 1406 of *LNCS*, pages 89–105. Springer Berlin Heidelberg.