A VISUAL SERVOING APPROACH FOR ROBOTIC FRUIT HARVESTING IN THE PRESENCE OF PARAMETRIC UNCERTAINTIES[1]

Juan D. Gamba*, Pål J. From†, Antonio C. Leite*

*Pontifical Catholic University of Rio de Janeiro,
Department of Electrical Engineering,
Postal Code 22451-900, Rio de Janeiro RJ, Brazil.

†Norwegian University of Life Sciences,
Faculty of Science and Technology,
P.O. Box 5003, NO-1432 Ås, Norway.

juan212@ele.puc-rio.br, pal.johan.from@nmbu.no, antonio@ele.puc-rio.br

Abstract— In this paper, we address the robust visual servoing problem for a fixed target using an eye-in-hand camera configuration, when the camera calibration parameters are uncertain. The proposed solution combines the image-based visual servoing (IBVS) and the position-based visual servoing (PBVS) approaches to take advantage of both strengths to perform successful robotic fruit harvesting tasks. An image-based detection and recognition method, based on SURF and RANSAC algorithms, is used to extract the image feature of the fruit for the vision-based control algorithm. A kinematic control design, with robustness properties, is employed to cope with the uncertainties in the calibration parameters of the camera-robot system. To deal with possible singular configurations of the robot arm, that may arise during the task execution, we employ the Damped Least-Squares (DLS) inverse method. Experimental results, obtained with a Mitsubishi robot arm RV-2AJ carrying out strawberry harvesting tasks, are included to illustrate the performance and feasibility of the proposed methodology.

Keywords— Visual Servoing, Agricultural Robots, Automatic Fruit Harvesting, Kinematic Control.

Resumo— Neste trabalho, considera-se o problema de servovisão robusta para um alvo fixo usando uma configuração de câmera *eye-in-hand*, quando os parâmetros de calibração da câmera são incertos. A solução proposta combina as abordagens de servovisão baseada em imagem e servovisão baseada em posição utilizando suas principais vantagens para realizar tarefas bem-sucedidas de colheita robótica de frutas. Um método de detecção e reconhecimento de imagem, baseado nos algoritmos SURF e RANSAC, é usado para extrair a característica da imagem da fruta. Um algoritmo de controle cinemático, com propriedades de robustez, é empregado para lidar com as incertezas nos parâmetros de calibração do sistema câmera-robô. Para lidar com possíveis configurações singulares do braço robótico, que podem surgir durante a execução da tarefa, utiliza-se o método Damped Least-Squares (DLS) inverso. Resultados experimentais, obtidos com um robô manipulador Mitsubishi RV-2AJ executando tarefas de coleta de morangos, são incluídos para ilustrar o desempenho e a viabilidade da metodologia proposta.

Palavras-chave— Servovisão, Robôs Agrícolas, Coleta de Frutas Autômática, Controle Cinemático.

1 INTRODUCTION

In the last decades, the technological advances of sensors and communication systems have encouraged the agricultural industry to employ intelligent autonomous robots to carry out a number of repetitive and dull tasks for farmers (Edan et al., 2009). Fruit harvesting and picking, weed control, autonomous mowing, pruning, seeding and spraying, plant phenotyping, sorting and packing are just a few examples of how robots are taking over fields around the world. In general, these robots are equipped with navigation systems as well as specialized accessories and tools in order to accomplish all the tasks successfully, safely and efficiently (Bac et al., 2014). The agricultural environment introduces several challenges and difficulties, particularly, for robotic harvesting systems. Indeed, changes in seasons and weather conditions, crop growth and rotation, dense vegetation, different maturity levels of fruits, the existence of diseases and fungi in plants, all these factors create a dynamic and poorly structured environment. In this context, a new trend for agricultural robotics is to combine and integrate advanced control theory, computer vision algorithms and machine learning techniques into visual servoing approaches, allowing robots to operate with a high level of autonomy and decision-making (Kapach et al., 2012).

Visual servoing is the area of robotics that is concerned with the pose control of robot arms or mobile robots based on the feedback of visual information extracted from one or more image features, by using a single or multiple cameras (Chaumette and Hutchinson, 2006). Such cameras may have different models (e.g., monocular, stereo and RGB-D) and be mounted in different configurations: fixed in the robot workspace (i.e., eye-to-hand) or attached to the robot end-

[1]This research was supported by the UTFORSK Partnership Programme from The Norwegian Centre for International Cooperation in Education (SIU), project number UTF-2016-long-term/10097, and partially financed by CAPES - Higher Education Personnel Improvement Coordination within the Ministry of Education of Brazil.

effector (i.e., eye-in-hand). More advanced applications, however, can use different types and settings of cameras simultaneously. In general, the direct measurements provided by the vision system are related to the image feature parameters in the image space, while the robotic task is defined in the operating or task space in terms of the relative pose of the robot end effector with respect to the target object. In this context, vision-based controllers can be classified into two main groups: position-based visual servoing (PBVS), image-based visual servoing (IBVS) as well as hybrid visual servoing (HVS) that combines the benefits of both approaches. One of the main advantages of the IBVS approach over the PBVS approach is its higher robustness to camera calibration errors, which results in better positioning accuracy for the vision system (Chaumette and Hutchinson, 2007). Following this trend, a color-based object extraction method based on OHTA color space was developed by Wei et al. (2014) and used to carry out robotic strawberry picking tasks under complex agricultural background. Mehta et al. (2016) have designed a vision-based estimation and control system for robotic citrus harvesting based on the combination of large field-of-view of a fixed camera and the accuracy of a mobile camera. Barth et al. (2016) have proposed a visual servoing approach which uses the eye-to-hand camera configuration for sweet pepper harvesting in dense vegetation. A detection and localisation algorithm applied to robotic apple harvesting which detects red and bicoloured fruits on tree by using an RGB-D camera was introduced by Nguyen et al. (2016).

In this work, we address the soft-fruit harvesting problem by using a visual servoing approach based on the combination of computer vision, and control theory methodologies. The control design uses a kinematic control approach based on the Damped Least-Squares (DLS) inverse method to deal with the existence of singular configurations for the robot arm during the harvesting task. A robust vision-based control scheme which combines the image-based visual servoing (IBVS) and the position-based visual servoing (PBVS) approaches is designed to cope with parametric uncertainties in the camera-robot system. The color feature extraction method for fruit detection is based on OHTA color space, which is able to deal with color singularity problem and more suitable for real-time image processing compared to HSV and RGB color spaces (Wei et al., 2014). The fruit recognition method uses a combination of the Speeded-Up Robust Features (SURF) and Random Sample Consensus (RANSAC) algorithms (Bay et al., 2008), due to their well-known properties of robustness, fastness, and accuracy. Experimental results, obtained with a Mitsubishi RV-2AJ robot performing strawberry picking tasks, illustrate the performance and feasibility of the proposed visual servoing scheme.

2 PROBLEM FORMULATION

In this work, we address the robotic fruit harvesting problem using a visual servoing scheme with an RGB-D stereo camera mounted on the robot end effector (Fig. 1). Here, the following notation is considered: $p_{ij} \in \mathbb{R}^3$ and $R_{ij} \in \mathbb{SO}(3)$ denote respectively the position vector and orientation matrix of the frame $\mathcal{F}_j$ with respect to frame $\mathcal{F}_i$; $T_{ij} \in \mathbb{R}^{4\times4}$ is the homogeneous transformation matrix, which denotes the pose of the frame $\mathcal{F}_j$ with respect to frame $\mathcal{F}_i$. In this context, the pose of the camera frame $\mathcal{F}_c$ with respect to the base frame $\mathcal{F}_b$ is given by $T_{bc} = T_{be}\, T_{ec}$, say:

$$T_{bc} = \begin{bmatrix} R_{bc} & p_{bc} \\ 0^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} R_{be}\, R_{ec} & R_{be}\, p_{ec} + p_{be} \\ 0^{\mathsf{T}} & 1 \end{bmatrix} . \quad (1)$$

Here, we assume that (A1) T_{be} can be obtained from the *forward kinematics* map by using, for example, the Denavit-Hartenberg convention. In this case, $p_{be} = p_{be}(q)$ and $R_{be} = R_{be}(q)$. For simplicity, we also assume that (A2) the camera frame $\mathcal{F}_c$ and the end-effector frame $\mathcal{F}_e$ are aligned only with respect to z-axis, but the relative translation between their origins and the relative orientation of their z-axes, denoted by φ, may be uncertain. In this case, $R_{ec} = R_{ec}(\varphi)$.

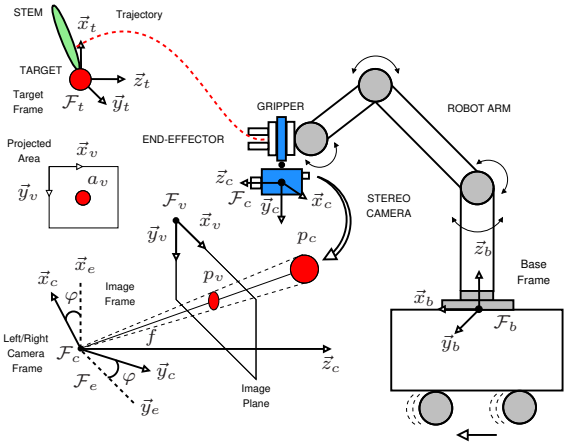

Figure 1: Visual servoing system for harvesting tasks.

The fruit harvesting task we consider consists of moving the robot arm to the vicinity of the fruit, cut the stem using a suitable device attached to the robot end-effector and store the fruit in a storage device. The first goal is to solve the image segmentation and interpretation problem, that is, to detect and recognize the target object located in the robot workspace by using an image-based detection and recognition (DR) algorithm. Once the target object is tracked by using a stereo vision system, the next step is to solve the correspondence and 3D reconstruction problem, that is, to

compute the 3D coordinates of the fruit with respect to the camera by using, for example, a simple triangulation technique (Siciliano et al., 2009). Thus, the pose of the target frame $\mathcal{F}_t$ with respect to the base frame $\mathcal{F}_b$ is given by $T_{bt} = T_{bc}\, T_{ct}$, say:

$$T_{bt} = \begin{bmatrix} R_{bt} & p_{bt} \\ 0^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} R_{bc}\, R_{ct} & R_{bc}\, p_{ct} + p_{bc} \\ 0^{\mathsf{T}} & 1 \end{bmatrix}, \quad (2)$$

where T_{ct} is pose of the target frame $\mathcal{F}_t$ with respect to the camera frame $\mathcal{F}_c$, which can be computed from the application of the DR algorithm and the triangulation technique. where the entries of T_{ct} can be computed from the application of the DR algorithm and the triangulation technique. Finally, since T_{bt} is computed, we can employ an *inverse kinematics-based algorithm* to transform the motion specifications, assigned to the robot end-effector in the task space, into the corresponding joint space motions, allowing for the successful execution of the desired motion.

3 VISUAL SERVOING APPROACH

The main idea of visual servoing is to use visual information obtained from a single or multiple cameras to control the pose of a robot arm with respect to a target object or a set of image features. In computer vision, an *image feature* may be any geometric characteristic that could be extracted from a given image. In general, an image feature can be obtained from the projection of a physical feature of the target object onto the camera image plane. Among several types of image features available in the literature, in this work, we adopt the *centroid coordinates* and the *area of the projected surface* of the fruit of interest, in order to carry out the harvesting task.

The main differences between position-based visual servoing (PBVS) and image-based visual servoing (IBVS) approaches are related to how the pose of the target object is obtained (e.g., pose estimate or feature measurements) and which coordinates space the output error is computed (e.g., task or image). The main advantages of using the visual information directly into a feedback control loop are twofold: (i) there is no need to estimate the pose of the target frame $\mathcal{F}_t$ with respect to the camera frame $\mathcal{F}_c$ (i.e., the object pose) in real-time; (ii) the lower sensitivity to calibration errors in the camera (Chaumette and Hutchinson, 2006).

3.1 Robot Kinematics

Here, we consider the kinematic model of a robot manipulator. In this framework, the *operational space* variables are related to the *joint space* variables by means of the following *forward kinematics* and *differential kinematics* mappings:

$$p = h(q)\,, \qquad \dot{p} = J_p(q)\,\dot{q}\,, \quad (3)$$

where $q, \dot{q} \in \mathbb{R}^n$ denote the position and velocity vectors of the robot joints, and $p, \dot{p} \in \mathbb{R}^m$ denote the position and linear velocity of the end-effector frame $\mathcal{F}_e$ with respect to the base frame $\mathcal{F}_b$. Notice that, $h(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^m$ is a nonlinear function and $J_p(q) = (\partial h / \partial q) \in \mathbb{R}^{m \times n}$ is the position part of the *analytical Jacobian* (Siciliano et al., 2009).

The orientation of the end-effector frame $\mathcal{F}_e$ with respect to the base frame $\mathcal{F}_b$ can be described by the *unit quaternion* representation, given by $\phi = \{\eta, \epsilon\} \in \mathbb{H}$, where $\eta \in \mathbb{R}$ is the scalar part and $\epsilon \in \mathbb{R}^3$ is the vector part, subject to the unit norm constraint $\eta^2 + \epsilon^2 = 1$. The so-called *quaternion propagation rule* relates the time-derivative of the unit quaternion $\dot{\phi} \in \mathbb{H}$ to the angular velocity of the end-effector frame $\mathcal{F}_e$ with respect to the base frame $\mathcal{F}_b$, denoted by $\omega \in \mathbb{R}^3$, as:

$$\dot{\phi} = \frac{1}{2} E(q)\,\omega\,, \qquad E(q) = \begin{bmatrix} \epsilon^{\mathsf{T}} \\ \eta I - Q(\epsilon) \end{bmatrix}, \quad (4)$$

where $Q(\cdot) : \mathbb{R}^3 \mapsto \mathrm{SO}(3)$ denotes the skew-symmetric operator and $J_r(\phi) = 2E^{\mathsf{T}}(\phi) \in \mathbb{R}^{3 \times 4}$ is the well-known *representation Jacobian*.

The *differential kinematics equation* provides the relationship between the joint velocities and the corresponding linear and angular velocities of the end-effector frame $\mathcal{F}_e$ with respect to the robot frame $\mathcal{F}_b$ as:

$$\mathbf{v} = \begin{bmatrix} \dot{p} \\ \omega \end{bmatrix} = \begin{bmatrix} J_p(q) \\ J_o(q) \end{bmatrix} = J(q)\,\dot{q}\,, \quad (5)$$

where $J(q) \in \mathbb{R}^{m' \times n}$ is the *geometric Jacobian* of the robot manipulator. Notice that, in general, the orientation of the robot end-effector is given in terms of the robot joint angles as $\phi = g(q)$, where $g(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^4$ is a nonlinear function. Thus, considering (5), implies that $J_o(q) = J_r(\phi)\,(\partial g(q)/\partial q)$, which is the so-called orientation part of the analytical Jacobian.

Now, we consider the kinematic control problem for a n-DoF robot manipulator. In this framework, the robot motion can be described by

$$\dot{q}_i \approx u_i\,, \qquad i = 1, \cdots, n\,, \quad (6)$$

where $\dot{q}_i$ is the velocity of the i-th robot joint and u_i is the velocity control signal applied to the i-th joint motor drive. Then, from the differential kinematic equation (5) and considering the kinematic control approach (6), we obtain the following control system:

$$\begin{bmatrix} \dot{p} \\ \omega \end{bmatrix} = J(q)\,u\,. \quad (7)$$

For the cases where the Jacobian matrix J is *non-square* ($m' < n$), the velocity control signal $u \in \mathbb{R}^n$ can be given simply by:

$$u(t) = J^{\dagger}(q)\,v = J^{\dagger}(q) \begin{bmatrix} v_p \\ v_o \end{bmatrix}, \quad (8)$$

where $J^{\dagger} = J^{\mathsf{T}}(JJ^{\mathsf{T}})^{-1}$ is the right pseudo-inverse of J and $v = [\, v_p \;\; v_o \,]^{\mathsf{T}}$ is a Cartesian control signal, for position and orientation, to be designed. Notice that, the control signal (8) locally minimizes the norm of the joint velocities, provided that (i) the robot kinematics is *fully known* and (ii) the Cartesian control signal $v(t)$ does not lead the robot to singular configurations. The failure of the last assumptions is a fairly open-problem in robotics area, and will be discussed later on.

3.2 PBVS control design

Consider the pose control problem for a n-DoF robot arm. Here, we assume that the task of interest is to move the robot end-effector towards a fixed target, located at the robot workspace (i.e., regulation task) using the eye-in-hand camera configuration. The control goal is to regulate the current pose of the robot end-effector $\mathbf{x} = [\, p \;\; \phi \,]^{\mathsf{T}}$ - estimated by the camera—to a constant desired pose $\mathbf{x}_d = [\, p_d \;\; \phi_d \,]^{\mathsf{T}}$, assumed to be bounded, that is:

$$\lim_{t \to \infty} \begin{bmatrix} p(t) \\ \phi(t) \end{bmatrix} = \begin{bmatrix} p_d \\ \phi_d \end{bmatrix} . \tag{9}$$

The position error e_p is defined as $e_p := p_d - p$, however the orientation error e_o should be defined in terms of the algebra of rotation groups, instead of the vector algebra (Siciliano et al., 2009). Thus, considering the *unit-quaternion representation*, it is usual to define the orientation error e_o as the vector part of the error quaternion:

$$e_o := \Delta\epsilon = \eta(q)\epsilon_d - \eta_d\epsilon(q) - Q(\epsilon_d)\epsilon(q) \,, \tag{10}$$

where the pair $\phi_d = \{\eta_d, \epsilon_d\}$ denotes respectively the scalar and vector parts of the desired quaternion. The error quaternion is defined as $\Delta\phi := \{\Delta\eta, \Delta\epsilon\} = \phi_d * \phi^{-1}$, where the symbol "$*$" denotes the quaternion product operator. Notice that, we have $\Delta\phi = \{1, 0^{\mathsf{T}}\}$ if and only if the quaternions ϕ and ϕ_d are coincident, which means that the corresponding coordinates frames are aligned.

Now, we are able to compute the velocity control signal u using a control algorithm based on the Jacobian pseudo-inverse:

$$u(t) := J^{\dagger}(q) \begin{bmatrix} v_p \\ v_o \end{bmatrix} = J^{\dagger}(q) \begin{bmatrix} \Lambda_p \, e_p \\ \Lambda_o \, e_o \end{bmatrix} , \tag{11}$$

where $\Lambda_p = \Lambda_p^{\mathsf{T}} > 0$ and $\Lambda_o = \Lambda_o^{\mathsf{T}} > 0$ are the position and orientation gain matrices. The stability and convergence analysis of the kinematic control approach is based on the Lyapunov stability theory and can be found in (Siciliano et al., 2009). Now, considering that during the execution of the fruit harvesting task the robot arm may move in the neighborhood of singular configurations, that is, those joint configurations where the Jacobian matrix has deficient rank, the key idea is to employ the well-known Damped Least-Squares (DLS) inverse method (Siciliano et al., 2009).

3.3 IBVS modeling

Here, we consider a visual servoing approach using an RGB-D stereo camera attached to the robot end-effector. Let $p_c = [\, x_c \;\; y_c \;\; z_c \,]^{\mathsf{T}}$ be the coordinates of a 3D point expressed in the camera frame $\mathcal{F}_c$. From the perspective projection model, the 3D point is projected in the image space as a 2D point with the coordinates $p_v = [\, x_v \;\; y_v]^{\mathsf{T}}$ expressed in pixels, say:

$$\begin{bmatrix} x_v \\ y_v \end{bmatrix} = \frac{f}{z_c} \begin{bmatrix} \alpha_x & 0 \\ 0 & \alpha_y \end{bmatrix} \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} x_{v0} \\ y_{v0} \end{bmatrix} , \tag{12}$$

where $\{x_{v0}, y_{v0}, f, \alpha_x, \alpha_y\}$ is the set of camera intrinsic parameters: x_{v0} and y_{v0} are the coordinates of the principal point; f is the focal length; α_x and α_y are the scaling factors in pixel per millimeter. The 3D point is projected in the image plane as a 2D point with normalized coordinates $p_p = [\, x_p \;\; y_p]^{\mathsf{T}}$ given by:

$$x_p = \frac{x_v - x_{v0}}{f\alpha_x} \,, \qquad y_p = \frac{y_v - y_{v0}}{f\alpha_y} \,. \tag{13}$$

Now, we suppose that the robot end-effector is moving with linear velocity $v_c \in \mathbb{R}^3$ and angular velocity $\omega_c \in \mathbb{R}^3$ both expressed with respect to the (instantaneous) camera frame $\mathcal{F}_c$. Then, using the well-known relationship of *velocity transformation* between the target frame $\mathcal{F}_t$ and the camera frame $\mathcal{F}_c$, we obtain the following motion equation (Chaumette and Hutchinson, 2006):

$$\dot{p}_{ct} = -v_c - Q(\omega_c)\, p_{ct} \,, \tag{14}$$

with $v_c = R_{bc}^{\mathsf{T}} \dot{p}_{bc}$ and $\omega_c = R_{bc}^{\mathsf{T}} Q({}^b\omega_{bc}) R_{bc}$. From the analysis of the camera-image coordinate transformation (12), we can observe that by using a *bidimensional image* may not be possible to obtain any explicit information on depth coordinate. Thus, we can not explicitly compute the depth between the target frame $\mathcal{F}_t$ and the camera frame $\mathcal{F}_c$ by using a single camera. This information, however, can be recovered (i) *indirectly*, using the image projected area of the object or (ii) *directly*, using a stereo vision system. In the indirect approach, the key idea is to use a target with spherical geometry so that the projected area in the image space becomes invariant with respect to the object rotations (Chaumette and Hutchinson, 2006). Let $a_v \in \mathbb{R}^+$ be the projected area of the target object expressed in the image frame $\mathcal{F}_v$. The dynamics of the *depth-to-area transformation* is given by:

$$\dot{a}_v = (-2a_v/z_c)\, \dot{z}_c \,. \tag{15}$$

Here, the following two assumptions can be considered: (A3) The image projected area a_v is

bounded and satisfies the inequality $0 < a_{min} < a_v(t) < a_{max}$ for all time t; (A4) The sign of z_c is assumed to be constant and known. Taking the time-derivative of (13) and using (14) yields:

$$\dot{s} = L_s \, \mathbf{v}_c \,, \qquad \begin{bmatrix} \dot{x}_p \\ \dot{y}_p \\ \dot{a}_v \end{bmatrix} = L_s \begin{bmatrix} v_c \\ \omega_c \end{bmatrix} , \qquad (16)$$

with

$$L_s = \begin{bmatrix} \frac{-1}{z_c} & 0 & \frac{x_p}{z_c} & x_p\, y_p & -(1+x_p^2) & y_p \\ 0 & \frac{-1}{z_c} & \frac{y_p}{z_c} & (1+y_p^2) & -x_p\, y_p & -x_p \\ 0 & 0 & \frac{2\,a_v}{z_c} & 2\,a_v\, y_p & -2\,a_v\, x_p & 0 \end{bmatrix}$$

where $L_s \in \mathbb{R}^{3\times 6}$ is the *interaction matrix* related to the state vector $s \in \mathbb{R}^3$, composed of the 3D point coordinates expressed in the image space. Notice that, since the target object is assumed to be fixed with respect to the base frame $\mathcal{F}_b$ the desired values for image features are assumed to be constant, and changes in s depends only on camera motion.

3.4 IBVS control design

Here, we assume that the control goal is to drive a set of features s to a desired set value s_d say:

$$s \to s_d \,, \qquad e_s := s_d - s \to 0 \,, \qquad (17)$$

where $e_s \in \mathbb{R}^3$ is the image feature error. From the differential kinematics equations (7) and (16), we obtain the following control system:

$$\dot{s} = L_s \, R_{bc}^{\mathsf{T}} \, J(q) \, u \,. \qquad (18)$$

From the time-derivative of (17), we define the velocity control signal u as:

$$u := J^{\dagger}(q) \, R_{bc} \, L_s^{\dagger} \, \Lambda_s \, e_s \,, \qquad \Lambda_s = \Lambda_s^{\mathsf{T}} > 0 \,, \quad (19)$$

where Λ_s is a positive definite gain matrix, which ensures the asymptotic convergence of the image feature error e_s to zero, that is, $\lim_{t\to\infty} e_s(t) = 0$, provided that (i) the robot arm is far from singular configurations and (ii) the calibration parameters of the camera-robot system are fully known. However, it is well-known that the knowledge of the camera calibration parameters and robot kinematics may not be satisfied from the practical point of view, particularly when the *eye-in-hand camera configuration* is used. In this context, the IBVS control system (18), in the presence of modeling errors or uncertainties, can be simply defined as:

$$\dot{s} = L_s(s) \, R_{bc}^{\mathsf{T}} \, J(q) \, u + \eta(s, q, \dot{q}, t) \,, \qquad (20)$$

where $\eta(\cdot)$ is a *state-dependent* nonlinear external disturbance which satisfies the linear growth bound

$$||\eta(s, q, \dot{q}, t)|| \leq \gamma ||e_s|| \,, \quad \gamma > 0 \,. \qquad (21)$$

Here, we also assume that $\eta(\cdot)$ is a nonlinear function that vanishes at origin, locally Lipschitz in $s, q, \dot{q}$ and uniformly in time t for all $t \geq 0$ (Khalil, 2002). Notice that, without loss of generality, we represent the perturbation term $\eta(\cdot)$ as an additive term on the right-hand side of the system equation (18). Therefore, an approximation or an estimation of the interaction matrix L_s, the rotation matrix R_{bc} as well as the Jacobian matrix $J(q)$ must be considered for the stability analysis (Chaumette and Hutchinson, 2006). In this case, the velocity control signal (19) takes the form:

$$u := \hat{J}^*(q) \, \hat{R}_{bc} \hat{L}_s^*(s) \, \Lambda_s \, e_s \,, \qquad (22)$$

where $\hat{R}_{bc} = R_{be}(q) \hat{R}_{ec}(\varphi)$, and $\varphi \in \mathbb{R}$ is the misalignment angle between the camera frame $\mathcal{F}_c$ and the end-effector frame $\mathcal{F}_e$ around the z-axes, assumed to be uncertain. Notice that, the estimated Jacobian matrix can be obtained from the DLS inverse method, say $\hat{J}^* = \hat{J}^{\mathsf{T}} (\hat{J} \hat{J}^{\mathsf{T}} + \alpha \, I)^{-1}$ where α is a damping factor that avoids the ill-conditioning problem of the Jacobian matrix.

Here, we also assume that (A3) $J(q)$ is bounded for all possibles values of q, that is $||J(q)|| \leq c_0$ for $\forall q \in [0, 2\pi]$ where $c_0 \in \mathbb{R}$ is a known positive constant, because $J(\cdot)$ depends on $q(t)$ as the argument of bounded trigonometric functions; (A4) $R_{bc} \in \mathbb{SO}(3)$ is bounded for all possible values of q and φ, that is $||R_{bc}|| \leq c_1$, where $c_1 \in \mathbb{R}$ is a known positive constant; (A5) $L_s(s)$ is bounded for all possible values of s, that is $||L_s(s)|| \leq c_2$, where $c_2 \in \mathbb{R}$ is a known positive constant, because we previously assumed that $\eta(\cdot)$ is state-dependent and vanishes at origin (Khalil, 2002).

To analyse the stability and convergence properties of the IBVS control scheme, we employ the Lyapunov stability formalism. Here, we consider the following positive-definite candidate Lyapunov function given by $V(e_s) = e_s^{\mathsf{T}} \, P \, e_s$, with $P = P^{\top} > 0$. The time-derivative of V along the system trajectories (17), (20) and (22), is given by

$$\begin{aligned} \dot{V}(e_s) \;=\; & -2\, e_s^{\top} \, P \, L_s \, R_{bc}^{\top} \, J \, \hat{J}^* \, \hat{R}_{bc} \, \hat{L}_s^* \, \Lambda_s \, e_s + \\ & 2\, e_s^{\top} \, P \, \eta(s, q, \dot{q}, t) \,. \end{aligned}$$

Then, for $P = (1/2)\Lambda_s$, the first term of the right-hand side of $\dot{V}(e_s)$ is negative definite. Since the second term, due to the effect of the perturbation, satisfies (21) we have:

$$\dot{V}(e_s) \leq c_3 ||e_s||^2 + c_4 \gamma ||e_s||^2 \,. \qquad (23)$$

If γ is small enough to satisfy the bound $\gamma < c_3/c_4$ implies that $\dot{V}(e_s) \leq -(c_3 - \gamma \, c_4) \, ||e_s||^2$ for $(c_3 - \gamma \, c_4) > 0$ and, therefore, the origin is an exponentially stable equilibrium point of the perturbed system (20). Notice that, the time-derivative of V is definite negative if the matrices $\hat{J}^*(q)$ and $\hat{L}_s^*$ are assumed to be full-rank, which can be guaranteed respectively if the robot arm has redundant

degrees of freedom and the IBVS control scheme uses one or more than two image features. The condition $\dot{V} < 0$ with $V > 0$ implies that the system trajectories uniformly converge to the origin, that is, the error system is asymptotically stable.

Moreover, since the robot kinematics and camera calibration intrinsic parameters are positive values, the presence of uncertainties in any or both matrices, $J^*(q)$ and $L_s^*(s)$, is not capable to violate the condition of negative definiteness (23). Accordingly, the misalignment angle φ between the camera and the end-effector frames needs to be less than $\frac{\pi}{2}$ rad, that is $|\varphi| < \frac{\pi}{2}$, in order to ensure the positive-definiteness property of the matrix $\hat{R}_{ec}(\varphi)$. Under these assumptions, it is possible to guarantee the asymptotic stability of the IBVS system for regulation tasks. Conversely, for tracking tasks, robust and adaptive control strategies could be used to overcome the restriction of the camera misalignment angle and deal with the existence of parametric uncertainties in the camera-robot system (Mehta et al., 2016; Mehta and Burks, 2016).

4 VERIFICATION AND VALIDATION

In this section, we present experimental results for a fruit harvesting task carried out in a strawberry farm (Sylling, Norway) in order to illustrate the effectiveness of the proposed visual servoing scheme (Fig. 2). We consider a RGB-D stereo camera attached to the end-effector of a 5-DoF robot arm, which is visually controlled at the velocity level by an image-based visual servoing approach in the neighborhood of singular configurations.

Numerical simulations were carried out on a Lenovo laptop with Intel Core i7-6500U Processor (4M Smart Cache, 2.5 GHz) 16 GB RAM, running Windows OS 64 bits. The control algorithm was implemented in MATLAB/Simulink (The MathWorks Inc.) Release 2017a and V-REP PRO EDU version 3.4.0 used as a robotic simulation platform. A set of scripts and function blocks were created to perform all necessary calculations and execute the control loop. Experiments were carried out using the Mitsubishi robot RV-2AJ and the controller Mitsubishi Melfa CR1. The communication channel between the laptop and the controller was established through RS-232C command protocol by means of an own-built cable, based on manufacturer manual. The camera attached to the end-effector was the ZED 2K Stereo Camera distributed by STEREOLABS and simple gripper was built ad-hoc using a pair of blades driven with a 24-VDC ABB motor. The parameters of the visual servoing system are $f = 10\ mm$, $\alpha_x = \alpha_y = 100\ pixel\, mm^{-1}$, $x_{v0} = y_{v0} = 500\ pixel$, $z_0 = 0.4640\ m$. The Denavit-Hartenberg parameters (Siciliano et al., 2009) of the robot arm are: $\alpha_1 = -\pi/2$, and $\alpha_4 = \pi/2$ and the offsets of joints two and four are $-\pi/2$ and $\pi/2$, where all angles are expressed in radians; $a_2 = 0.25$, $a_3 = 0.16$; $d_1 = 0.3$, and $d_5 = 0.072$, where all lengths are expressed in meters. The control gains were chosen as: $\Lambda_p = 7 \times 10^2\, I$, $\Lambda_o = 8 \times 10^4\, I$, $\Lambda_s = 3 \times 10^4\, I$. All other parameters are assumed to be zero.

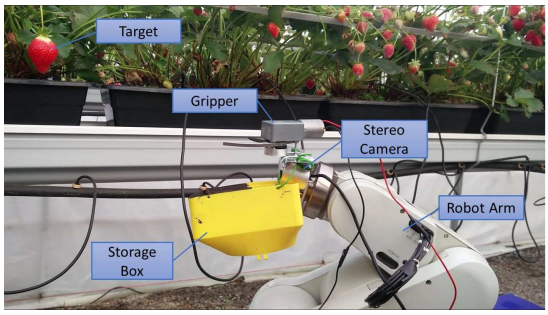

Figure 2: Experimental setup at the strawberry farm.

In this section, we describe some aspects of design and practical implementation of the proposed visual servoing approach. The flowchart of the fruit harvesting algorithm is shown in Fig. 3. From the calculation of the same image feature in both cameras of the stereo vision, we can compute the 3D point coordinates of the target in the operational space by using a triangulation technique (Siciliano et al., 2009). In computer vision, a well-known problem consists on recognizing matching points that belong to the same image feature in different scenes, along with object extraction or image segmentation (Forsyth and Ponce, 2012). In this context, the OHTA color space was chosen due to its ability for processing the fruit object extraction under complex agriculture background. In addition to dealing with the color singularity problem, others advantages of the OHTA space is its simplicity of calculation and low computational cost, which help us to reduce the problem of having a complex image processing during the execution of the vision-based control scheme (Wei et al., 2014).

After extracting the object of interest from the scene, we use the SURF algorithm (Bay et al., 2008) in order to obtain features from both images and, then, the RANSAC algorithm (Forsyth and Ponce, 2012) to identifying the corresponding

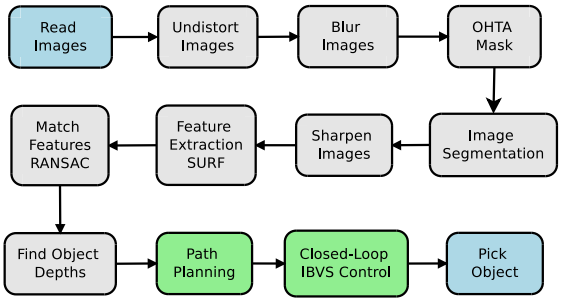

Figure 3: Fruit harvesting algorithm flowchart.

or matching points. It is worth noting that when performing the extraction of image features after object segmentation, a lower computational cost and more simple calculations are required than if the process is performed in the reverse order. On the other hand, it is well-known that RANSAC algorithm is a stochastic algorithm which does not guarantee the total recognition of the inliers features. In addition, by performing a triangulation with outliers features from a considerable distance will result in a target position which is very different from expected. Therefore, it is not always possible to ensure the system stability by performing an end-point open-loop control based on the PBVS approach, using as a target object the corresponding 3D point coordinates obtained from the visual system. Under this constraint, a closed-loop control system seems to be the more appropriate to ensure the reliability and safety of tasks performed by vision-based controllers. The proposed visual servo system involves different control techniques: (i) PBVS scheme for approaching the target object (approaching phase); (ii) IBVS scheme to maintain the object in the field of view of the stereo vision system (fine tuning phase); (iii) quaternion-based orientation control to ensure the proper pose of the robot end-effector for harvesting the selected fruit (picking phase). To deal with the minimum and maximum depth range presented in the stereo vision system as well as the depth needed for collecting the fruit properly, a depth estimation method based on the object projected area is used to calculate the object depth with respect to the camera. From a proper camera calibration, it is possible to compute the 3D point coordinates of the target using a single camera. After reaching the desired distance, a final image-based height adjustment is carried out for positioning the gripper close to the fruit stem, cut it and store the fruit in a storage box, completing the harvesting task successfully.

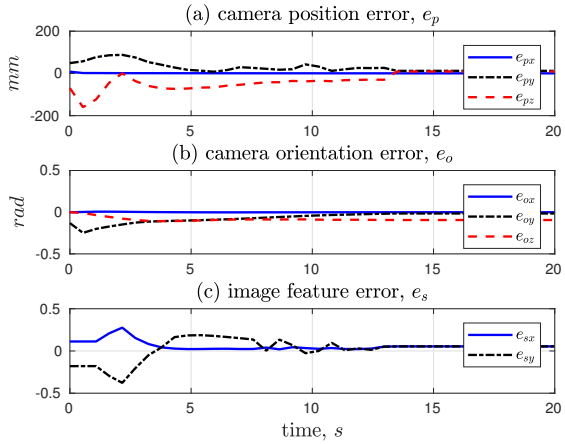

Figure 4: PBVS+IBVS: camera pose and image feature errors.

The *experimental results*[1] for a fruit picking task, performed with the Mitsubishi robot RV-2AJ at the strawberry farm in Sylling, Norway (Fig. 2) are presented in Figures 4-7. The behavior in time for position, orientation and image regulation errors can be observed in Fig. 4, where it is possible to see the asymptotic convergence of the errors. From Fig. 5, we can verify the time history of the position and orientation control signals provided by the PBVS scheme during the approaching phase. Conversely, the behavior in time of the position and orientation control signals provided by the IBVS scheme during the picking phase is shown in Fig. 6. The time history of the velocity control signals applied to the joints is shown in Fig. 7(a). We can observe the smooth behavior of the joint velocities, and the satisfactory performance of the orientation control scheme. The angular position of joint 5, q_5, is critical for the successful execution of the fruit harvesting task and the orientation controller ensures that the values assumed by q_5 remain close to zero as seen Fig. 7(b).

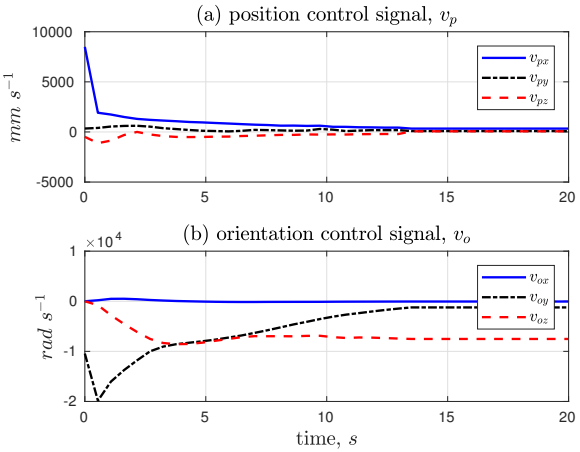

Figure 5: PBVS: pose control signals.

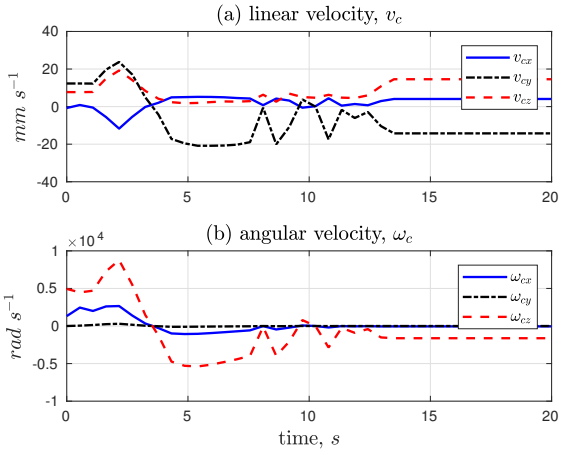

Figure 6: IBVS: linear and angular velocities.

[1]The preliminary experiments can be viewed in the accompanying video clip in: `https://youtu.be/Hi10oiuG0_I`

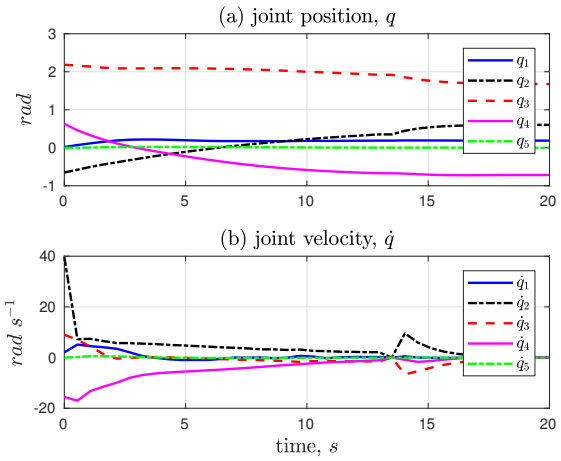

Figure 7: PBVS+IBVS: joint positions and velocities.

5 CONCLUDING REMARKS

In this work, we have developed a visual servoing control scheme based on PBVS and IBVS approaches for harvesting tasks in the presence of uncertainties. Combining the benefits of both approaches, we use the first one to perform the approach phase towards the target and, then, we use the second one to make fine adjustments in the movement. Practical results have shown the efficiency and feasibility of using robot arms to perform semi-autonomous harvesting tasks for soft fruits in orchards and poly-tunnels. We expected that in the near future other types of crops–that introduce different types of challenges - may be harvested by the robot arm and other agricultural tasks that use robots can also be implemented.

Some proposed topics for further investigation involve: (i) improve the fruit stem detection and segmentation algorithm to perform the fruit harvesting under different stems configurations, since by finding the top of the fruit and adding a small value to the height it is possible to pick only fruits with a stem in upright position; (ii) examine other methods for fruits recognition and image segmentation based on machine learning and deep learning approaches; (iii) analyse other vision-based control approaches, combining different camera models and mounting configurations, that are robust to parametric uncertainties and kinematic singularities.

References

Bac, C. W., van Henten, E. J., Hemming, J. and Edan, Y. (2014). Harvesting Robots for High-value Crops: State-of-the-art Review and Challenges Ahead, *Journal of Field Robotics* **31**(6): 888–911.

Barth, R., Hemming, J. and van Henten, E. J. (2016). Design of an Eye-in-hand Sensing and Servo Control Framework for Harvesting Robotics in Dense Vegetation, *Biosystems Engineering* **146**: 71–84.

Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L. (2008). Speeded-Up Robust Features (SURF), *Computer Vision and Image Understanding* **110**(3): 346–359.

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

Chaumette, F. and Hutchinson, S. (2007). Visual Servo Control - Part II: Advanced approaches, *IEEE Robotics and Automation Magazine* **14**(1): 109–118.

Edan, Y., Han, S. and Kondo, N. (2009). Automation in Agriculture, *in* S. Y. Nof (ed.), *Springer Handbook of Automation*, Springer-Verlag Berlin Heidelberg, pp. 1095–1128.

Forsyth, D. A. and Ponce, J. (2012). *Computer Vision - A Modern Approach*, 2nd edn, Pearson Inc.

Kapach, K., Barnea, E., Mairon, R. and Edan, Y. (2012). Computer Vision for Fruit Harvesting Robots - State of the Art and Challenges Ahead, *International Journal Computational Vision and Robotics* **3**(1/2): 4–34.

Khalil, H. K. (2002). *Nonlinear Systems*, Prentice-Hall Inc.

Mehta, S. and Burks, T. F. (2016). Adaptive Visual Servo Control of Robotic Harvesting Systems, *Proceedings of the 5th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture*, pp. 287–292.

Mehta, S. S., MacKunis, W. and Burks, T. F. (2016). Robust Visual Servo Control in the Presence of Fruit Motion for Robotic Citrus Harvesting, *Computers and Electronics in Agriculture* **123**: 362–375.

Nguyen, T. T., Vandevoorde, K., Wouters, N., Kayacan, E., Baerdemaeker, J. G. D. and Saeys, W. (2016). Detection of Red and Bicoloured Apples on Tree with an RGB-D Camera, *Biosystems Engineering* **146**: 33–44.

Siciliano, B., Sciavicco, L., Villani, L. and Oriolo, G. (2009). *Robotics: Modelling, Planning and Control*, Springer Publishing Company, Inc.

Wei, X., Jia, K., Lan, J., Li, Y., Zeng, Y. and Wang, C. (2014). Automatic Method of Fruit Object Extraction under Complex Agricultural Background for Vision System of Fruit Picking Robot, *Optik - International Journal for Light and Electron Optics* **125**(19): 5684–5689.