Hybrid Visual Servoing with Hierarchical Task Composition for Aerial Manipulation*

Vincenzo Lippiello[1], Jonathan Cacace[1], Angel Santamaria-Navarro[2], Juan Andrade-Cetto[2],
Miguel Ángel Trujillo[3], Yamnia Rodríguez Esteves[3], and Antidio Viguria[3]

Abstract—In this paper a hybrid visual servoing with a hierarchical task-composition control framework is described for aerial manipulation, *i.e.* for the control of an aerial vehicle endowed with a robot arm. The proposed approach suitably combines into a unique hybrid-control framework the main benefits of both image-based and position-based control schemes. Moreover, the underactuation of the aerial vehicle has been explicitly taken into account in a general formulation, together with a dynamic smooth activation mechanism. Both simulation case studies and experiments are presented to demonstrate the performance of the proposed technique.

Index Terms—Aerial Robotics; Visual Servoing.

I. INTRODUCTION

IN the last years new application domains in the field of aerial service robotics have been addressed by researchers from different disciplines, *e.g.* surveillance, inspection, agriculture, delivering, etc. Sophisticated prototypes have been developed with the capacity to physically interact with the environment [1]. The modeling and control of an unmanned aerial vehicle (UAV) able of interacting with the environment to accomplish simple robotic-manipulation tasks have been proposed in [2], which is based on a force-position control law designed through a feedback linearizing technique.

With the improvement of the batteries and the miniaturization of motors and servos, new high-performance UAV prototypes endowed with a robot arm —called unmanned aerial manipulators (UAMs)— have been designed. A control algorithm which is able to exploit all the degrees of freedom (DoFs) of a UAM is proposed in [3], where the execution of tasks with physical interaction with the environment has been achieved. However the employed UAM is completly actuated only along one direction and has no redundancy. In [4], [5] the dynamic model of a UAM and a Cartesian impedance control

have been designed providing a desired relationship between external wrench and the system motion. However, redundancy is exploited in a rigid way. Aerial manipulation tasks executed with a UAM endowed with a 2-DoFs robot arm have been presented in [6], where an adaptive sliding-mode controller has been adopted. A control solution considering valve turning with a dual-arm UAM has been proposed in [7]. In all these works no vision is employed for the task execution and there is no redundancy.

The use of vision for the execution of aerial robotic tasks is a widely adopted solution to cope with unknown environments. In [8], [9] new image-based control laws endowing a UAM with the capability of automatically positioning parts on target structures have been proposed, where the system redundancy and underactuation of the vehicle base have been explicitly taken into account. A task-oriented control law for aerial surveillance has been proposed in [10], where a camera is attached to the end-effector of the robot arm to perform visual servoing towards a desired target. However in these papers redundancy is employed in a rigid way and the interaction between dependent tasks is not considered.

In this paper a hybrid image- and position-based visual servoing via a hierarchical task-composition control is presented for the control of a UAM. The presence of redundancy in a UAM system allows combining a number of subtasks with the a new hierarchical-task formulation. Different subtasks can be designed both in the Cartesian space, *e.g.* obstacle avoidance, manipulation tasks, and in the image space of the camera, *e.g.* field-of-view constraints, as well as in the arm joint space, *e.g.* center-of-gravity balancing, joint-limits avoidance, manipulability, etc. Moreover, the underactuation of the aerial vehicle base has been systematically taken into account within a new recursive formulation. A number of practical tasks have been designed requiring only few DoFs to be accomplished, hence allowing an accurate profiling of the system behavior. The study of the task Jacobian singularity and a smooth task activation are also shown.

With respect to our previous work [11], a new advanced formulation is derived with the capability to guarantee decoupling of independent tasks (not only orthogonal as in the previous work), the stability analysis of the new proposed control law is discussed together with the derivation of all the task Jacobian matrices (in [11] the Jacobian matrices of the uncontrollable variables are missing), and finally both simulation and experimental results are provided to evaluate the effectiveness of the new proposed control law.
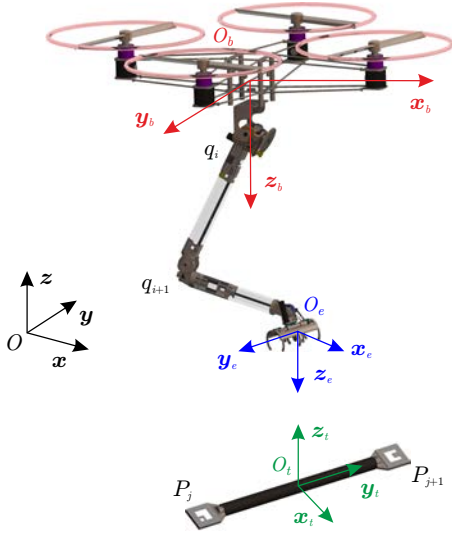
Fig. 1: Reference frames.

II. REFERENCE FRAMES AND CAMERA MODEL

Let $\boldsymbol{o}_b \in \mathbb{R}^3$ and $\boldsymbol{R}_b \in SO(3)$ be the position and rotation matrix, respectively, of the body reference frame $\{\mathcal{B} : O_b - x_b y_b z_b\}$, which is fixed with the UAV base, with respect to the inertial reference frame $\{\mathcal{I} : O - xyz\}$ (see Fig. 1). The triple $\boldsymbol{\phi} = (\varphi, \vartheta, \psi)$ of Euler roll-pitch-yaw angles is considered for the representation of the vehicle orientation, which in matrix form can be expressed as $\boldsymbol{R}_b(\boldsymbol{\phi})$.

A standard VToL UAV, *e.g.* a quadrotor, is an underactuated system with only 4 DoFs. In fact, the linear motion of the vehicle is generated by modifying the attitude, and thus the total propeller thrust generates a linear acceleration in the desired direction. Therefore, the roll and pitch rotations are constrained by imposing a desired linear motion.

A robot arm with ν DoFs attached to the UAV base is considered. Let $\boldsymbol{q} = \begin{bmatrix} q_1 & \dots & q_\nu \end{bmatrix}^\top$ be the arm joint vector describing the arm configuration, where q_i is the ith joint variable, with $i = 1, \dots, \nu$, and let $\{\mathcal{E} : O_e - x_e y_e z_e\}$ be the reference frame fixed with the arm end-effector (*e.g.* gripper). Notice that, by considering the DoFs of the vehicle base, even if a VToL UAV is an underactuated system, the addition of enough DoFs of the robot arm can generate a task-redundant system ($n = 4 + \nu$ total DoFs), *e.g.* with respect to the positioning of the gripper.

Finally, the vehicle base is endowed with a downward looking camera. Without loss of generality, the camera reference frame is considered coincident with $\mathcal{B}$. The pin-hole camera model is employed, *i.e.* by denoting with $\boldsymbol{p}^b = \begin{bmatrix} x^b & y^b & z^b \end{bmatrix}^\top$ the position of an observed point P expressed with respect to $\mathcal{B}$ (see Fig. 1), the optical projection of P onto the normalized image plane of the camera determines the so-called image feature vector $\boldsymbol{s}$:

$$\boldsymbol{s} = \begin{bmatrix} X \\ Y \end{bmatrix} = \frac{1}{z^b} \begin{bmatrix} x^b \\ y^b \end{bmatrix}. \tag{1}$$

III. OBJECT POSE ESTIMATION

We assume that m known visual markers are attached to the target object. Let $\{\mathcal{T} : O_t - x_t y_t z_t\}$ be a reference frame fixed with the object, and $\boldsymbol{o}_j^t$ and $\boldsymbol{R}_j^t$, with $j = 1, \dots, m$, be the known position and rotation matrix, respectively, of the reference frame attached to the jth marker in $\mathcal{T}$.

By using a visual tracker [12], [13], the pose of any visible marker can be measured with respect to the camera. Then, the pose of the target object can be reconstructed from the measurement of each visible marker as follows:

$$\boldsymbol{o}_t^b = \boldsymbol{o}_j^b - \boldsymbol{R}_j^{b\top} \boldsymbol{o}_j^t \tag{2}$$
$$\boldsymbol{R}_t^b = \boldsymbol{R}_j^b \boldsymbol{R}_j^{t\top}. \tag{3}$$

However, if more markers are simultaneously visible, a more robust and accurate solution can be achieved by combining the available information. Each marker contour is divided in several points (*e.g.* the corners and the middle points of the edges). Assuming the 3D model of the marker is known, we define the 3D coordinates of each contour point l relative to its marker frame as $\boldsymbol{p}_l^j$. Then, each 3D point of a marker is associated to the corresponding object frame $\mathcal{T}$ with

$$\boldsymbol{p}_l^t = \boldsymbol{R}_j^t \boldsymbol{p}_l^j + \boldsymbol{o}_j^t. \tag{4}$$

Once all contour correspondences are associated to the object frame, a Perspective-n-Point method [14] is used to obtain the camera pose with respect to the object. A RANSAC outlier rejection mechanism is integrated to remove point correspondences resulting from imaging artifacts that might be inconsistent with the computed transformation [15].

IV. DYNAMIC TASK PRIORITY CONTROL

The kinematic redundancy in the system allows control of the UAM end-effector by simultaneously achieving a number of secondary tasks. Indeed, redundancy means that the same gripper pose can be reached with several, even infinite, system configurations. Hence, the system can be suitably reconfigured by using internal motions, *i.e.* without affecting the gripper pose, to satisfy mechanical constraints (*e.g.* arm joint limits) and several subtasks (*e.g.* field of view, arm manipulability, robot-arm center of gravity control).

Let $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{o}_b^\top & \boldsymbol{\phi}^\top & \boldsymbol{q}^\top \end{bmatrix}^\top$ be the system state vector, and $\boldsymbol{\omega}_b = \begin{bmatrix} \omega_x & \omega_y & \omega_z \end{bmatrix}^\top$ the angular velocity of the body frame. To easily address the system underactuation we extract from $\boldsymbol{x}$ only the controlled variables in the new vector $\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{o}_b^\top & \psi & \boldsymbol{q}^\top \end{bmatrix}^\top$, and analogously for the corresponding velocity variables in $\boldsymbol{\upsilon} = \begin{bmatrix} \dot{\boldsymbol{o}}_b^\top & \omega_z & \dot{\boldsymbol{q}}^\top \end{bmatrix}^\top$, while $\boldsymbol{\varpi} = \begin{bmatrix} \omega_x & \omega_y \end{bmatrix}^\top$ represents the remaining vector of the uncontrolled angular velocities. Notice that the angular velocity $\boldsymbol{\varpi}$ can be measured with a standard onboard inertial measurement unit (IMU) under the assumption of a classical time-scale separation between the attitude controller (faster control loop, *e.g.* up to 1 kHz) and the velocity controller (slower control, in our case at camera frame rate of 25 Hz).

Moreover, let $\boldsymbol{\sigma}_0 = \boldsymbol{f}_0(\boldsymbol{x}) \in \mathbb{R}^{\mu_0}$ be the variables of a configuration-dependent main task; hence the following differential relationship holds:

$$\dot{\boldsymbol{\sigma}}_0 = \frac{\partial \boldsymbol{f}_0(\boldsymbol{x})}{\partial \boldsymbol{x}} \dot{\boldsymbol{x}} = \boldsymbol{J}_0(\boldsymbol{x}) \boldsymbol{\upsilon} + \overline{\boldsymbol{J}}_0(\boldsymbol{x}) \boldsymbol{\varpi}, \tag{5}$$

where $\boldsymbol{J}_0(\boldsymbol{x}) \in \mathbb{R}^{\mu_0 \times n}$ and $\overline{\boldsymbol{J}}_0(\boldsymbol{x}) \in \mathbb{R}^{\mu_0 \times 2}$ are the main task Jacobian matrices of the controlled and uncontrolled state variables, respectively. By inverting (5) and considering a regulation problem of $\boldsymbol{\sigma}_0$ to the desired value $\boldsymbol{\sigma}_0^*$, hence by defining $\widetilde{\boldsymbol{\sigma}}_0 = \boldsymbol{\sigma}_0^* - \boldsymbol{\sigma}_0$ as the main task error, the following velocity command can be considered:

$$\boldsymbol{v}^* = \boldsymbol{J}_0^\dagger(\boldsymbol{\Lambda}_0 \widetilde{\boldsymbol{\sigma}}_0 - \overline{\boldsymbol{J}}_0 \boldsymbol{\varpi}), \tag{6}$$

where $\boldsymbol{\Lambda}_0 \in \mathbb{R}^{\mu_0 \times \mu_0}$ is a positive-definite gain matrix, and $\boldsymbol{J}_0^\dagger$ is the generalized inverse of $\boldsymbol{J}_0$, which has been assumed to be full-rank. By substituting (6) into (5), the following exponentially stable error dynamics is achieved

$$\dot{\widetilde{\boldsymbol{\sigma}}}_0 = -\boldsymbol{\Lambda}_0 \widetilde{\boldsymbol{\sigma}}_0. \tag{7}$$

In case of $\mu_0 < n$ a second lower-priority subtask $\boldsymbol{\sigma}_1 = \boldsymbol{f}_1(\boldsymbol{x}) \in \mathbb{R}^{\mu_1}$ can be added with the following command:

$$\boldsymbol{v}^* = \boldsymbol{J}_0^\dagger \boldsymbol{\Lambda}_0 \widetilde{\boldsymbol{\sigma}}_0 + (\boldsymbol{J}_1 \boldsymbol{N}_0)^\dagger \boldsymbol{\Lambda}_1 \tilde{\boldsymbol{\sigma}}_1 - \overline{\boldsymbol{J}}_{0|1} \boldsymbol{\varpi}, \tag{8}$$

with $\boldsymbol{J}_1$ the Jacobian matrix of the second subtask, which is assumed to be full-rank, and

$$\overline{\boldsymbol{J}}_{0|1} = (\boldsymbol{J}_1 \boldsymbol{N}_0)^\dagger \overline{\boldsymbol{J}}_1 + (\boldsymbol{I}_n - (\boldsymbol{J}_1 \boldsymbol{N}_0)^\dagger \boldsymbol{J}_1)\overline{\boldsymbol{J}}_{0|0}, \tag{9}$$

where $\boldsymbol{N}_0 = \boldsymbol{I}_n - \boldsymbol{J}_0^\dagger \boldsymbol{J}_0$ is the projector onto the null space of $\boldsymbol{J}_0$, with $\boldsymbol{I}_n$ the n-dimension identity matrix, $\overline{\boldsymbol{J}}_{0|0} = \boldsymbol{J}_0^\dagger \overline{\boldsymbol{J}}_0$, $\boldsymbol{\Lambda}_1 \in \mathbb{R}^{\mu_1 \times \mu_1}$ is a positive definite gain matrix. The Jacobian matrix $\overline{\boldsymbol{J}}_{0|1}$ allows the compensation of the variation of the $\boldsymbol{\varpi}$. Notice that matrix $\boldsymbol{J}_1 \boldsymbol{N}_0$ is full-rank only if the two tasks are *orthogonal* ($\boldsymbol{J}_1 \boldsymbol{J}_0^\dagger = \boldsymbol{O}_{\mu_1 \times \mu_0}$) or *independent* (not orthogonal and $\mathrm{rank}(\boldsymbol{J}_0^\dagger) + \mathrm{rank}(\boldsymbol{J}_1^\dagger) = \mathrm{rank}([\boldsymbol{J}_0^\dagger \ \boldsymbol{J}_1^\dagger])$). See [16] for more details. Substituting (8) into (5) and by noticing that $\boldsymbol{N}_0$ is idempotent and Hermitian, hence $(\boldsymbol{J}_1 \boldsymbol{N}_0)^\dagger = \boldsymbol{N}_0(\boldsymbol{J}_1 \boldsymbol{N}_0)^\dagger$, the dynamics of the main task (7) is again achieved and so the exponential stability is proven. To study the behavior of the secondary task $\boldsymbol{\sigma}_1$ we can differentiate the subtask variables as follows

$$\dot{\boldsymbol{\sigma}}_1 = \frac{\partial \boldsymbol{f}_1(\boldsymbol{x})}{\partial \boldsymbol{x}} \dot{\boldsymbol{x}} = \boldsymbol{J}_1(\boldsymbol{x})\boldsymbol{v} + \overline{\boldsymbol{J}}_1(\boldsymbol{x})\boldsymbol{\varpi}. \tag{10}$$

Then, by substituting (8) and by assuming that the tasks are at least independent, the following error dynamics is achieved

$$\begin{aligned}
\dot{\tilde{\boldsymbol{\sigma}}}_1 &= -\boldsymbol{J}_1 \boldsymbol{J}_0^\dagger \boldsymbol{\Lambda}_0 \widetilde{\boldsymbol{\sigma}}_0 - \boldsymbol{J}_1(\boldsymbol{J}_1 \boldsymbol{N}_0)^\dagger \boldsymbol{\Lambda}_1 \tilde{\boldsymbol{\sigma}}_1 \\
&\quad + \left(\boldsymbol{J}_1 \boldsymbol{J}_0^\dagger \overline{\boldsymbol{J}}_0 + \boldsymbol{J}_1(\boldsymbol{J}_1 \boldsymbol{N}_0)^\dagger(\overline{\boldsymbol{J}}_1 - \boldsymbol{J}_1 \boldsymbol{J}_0^\dagger \overline{\boldsymbol{J}}_0) - \overline{\boldsymbol{J}}_1 \right) \boldsymbol{\varpi} \\
&= -\boldsymbol{J}_1 \boldsymbol{J}_0^\dagger \boldsymbol{\Lambda}_0 \widetilde{\boldsymbol{\sigma}}_0 - \boldsymbol{\Lambda}_1 \tilde{\boldsymbol{\sigma}}_1,
\end{aligned} \tag{11}$$

where we used the property $\boldsymbol{J}_1(\boldsymbol{J}_1 \boldsymbol{N}_0)^\dagger = \boldsymbol{I}_{\mu_1}$. Finally the dynamics of the error system can be written as follows

$$\begin{bmatrix} \dot{\widetilde{\boldsymbol{\sigma}}}_0 \\ \dot{\tilde{\boldsymbol{\sigma}}}_1 \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\Lambda}_0 & \boldsymbol{O}_{\mu_0 \times \mu_1} \\ -\boldsymbol{J}_1 \boldsymbol{J}_0^\dagger \boldsymbol{\Lambda}_0 & -\boldsymbol{\Lambda}_1 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{\sigma}}_0 \\ \tilde{\boldsymbol{\sigma}}_1 \end{bmatrix}, \tag{12}$$

that is characterized by a Hurwitz matrix, hence the exponential stability of the system is guaranteed. Moreover, we can notice a term representing the coupling effect of the main task on the secondary task. In case of orthogonal tasks this term is zero, *i.e.* $\dot{\tilde{\boldsymbol{\sigma}}}_1 = -\boldsymbol{\Lambda}_1 \tilde{\boldsymbol{\sigma}}_1$, and the behavior of the main and that of the secondary tasks are decoupled.

By generalizing (8) to the case of η prioritized subtasks, the following general velocity command can be formulated:

$$\boldsymbol{v}^* = \boldsymbol{J}_0^\dagger \boldsymbol{\Lambda}_0 \widetilde{\boldsymbol{\sigma}}_0 + \sum_{i=1}^{\eta} (\boldsymbol{J}_i \boldsymbol{N}_{0|\dots|i-1})^\dagger \boldsymbol{\Lambda}_i \widetilde{\boldsymbol{\sigma}}_i - \overline{\boldsymbol{J}}_{0|\dots|\eta} \boldsymbol{\varpi}, \tag{13}$$

with the recursively-defined compensating matrix

$$\begin{aligned}
\overline{\boldsymbol{J}}_{0|\dots|\eta} &= (\boldsymbol{J}_\eta \boldsymbol{N}_{0|\dots|\eta-1})^\dagger \overline{\boldsymbol{J}}_\eta \\
&\quad + (\boldsymbol{I}_n - (\boldsymbol{J}_\eta \boldsymbol{N}_{0|\dots|\eta-1})^\dagger \boldsymbol{J}_\eta)\overline{\boldsymbol{J}}_{0|\dots|\eta-1},
\end{aligned} \tag{14}$$

where $\boldsymbol{N}_{0|\dots|i}$ is the projector onto the null space of the *augmented Jacobian* $\boldsymbol{J}_{0|\dots|i}$ of the ith subtask, with $i = 0, \dots, \eta - 1$, which are respectively defined as follows

$$\boldsymbol{J}_{0|\dots|i} = \begin{bmatrix} \boldsymbol{J}_0^\top & \cdots & \boldsymbol{J}_i^\top \end{bmatrix}^\top \tag{15}$$

$$\boldsymbol{N}_{0|\dots|i} = (\boldsymbol{I}_n - \boldsymbol{J}_{0|\dots|i}^\dagger \boldsymbol{J}_{0|\dots|i}). \tag{16}$$

The previous stability analysis can be straightforwardly extended to the general case of η subtasks.

The hierarchical formulation (13) guarantees that the execution of all the higher-priority tasks from 0 (main task) to $i - 1$ will not be affected by the ith subtask and by the variation of the uncontrolled state variables. In other words, the execution of the ith task is subordinated to the execution of the higher priority tasks present in the task stack, *i.e.* it will be fulfilled only if suitable and enough DoFs are available, while the complete fulfillment of the main task, instead, is always guaranteed. However, with this new formulation for all reciprocally annihilating or independent tasks, a fully decoupling of the error dynamics is guaranteed.

The task composition and priority can be modified at runtime as needed, *i.e.* by activating or deactivating subtasks as well as by changing the priority order of the current active tasks already present in the task stack. However, in order to avoid discontinuity of the control input, a smooth transition between different task stacks has to be considered. This goal can be achieved by adopting a time-vanishing *smoothing term* when a new task stack is activated. Without loss of generality, we suppose that the transition phase starts at $t = 0$, *i.e.* the rth task stack has to be deactivated and substituted by the new one $(r + 1)$th. During the transition the velocity command is computed as follows:

$$\boldsymbol{v}^*(t) = \boldsymbol{v}_{r+1}^*(t) + e^{-\frac{t}{\tau}}(\boldsymbol{v}_r^*(0) - \boldsymbol{v}_{r+1}^*(0)), \tag{17}$$

where τ is a time constant determining the transition phase duration, and $\boldsymbol{v}_k^*$ is the velocity command corresponding to the kth task stack. When t becomes sufficiently greater than τ, the rth task stack is fully removed and a new transition can start. Notice that the smoothing term $e^{-\frac{t}{\tau}}(\boldsymbol{v}_r^*(0) - \boldsymbol{v}_{r+1}^*(0))$ is bounded and exponentially vanishing, hence it will not affect the stability of the proposed control law. The time constant τ has to be smaller than the inverse of the maximum eigenvalue of the gain matrices $\boldsymbol{\Lambda}_i$ to ensure a short transient time response in comparison with the nullifying time of the task errors $\widetilde{\boldsymbol{\sigma}}_i$.

V. Hybrid visual servoing and system behavior control for aerial manipulation tasks

In this section several elementary tasks useful for the composition of an aerial manipulation (grasping/plugging) task by exploiting visual measurements will be proposed. Besides tasks allowing the control of the gripper pose with respect to the observed objects, tasks able to guarantee the camera FoV constraint, to minimize the effect of the motion of the arm on the vehicle positional stability, and that addresses the issue of the joint mechanical limits are proposed. These tasks are not all orthogonal and/or independent but they are essential for the specific mission purposes. The priority of the proposed tasks can be arranged on the basis of the desired behaviour, even if some general constraints have to be considered. For example, the FoV constraint is essential because the loss of the observed object from the camera image will determine the failure of the whole mission (depending on the available camera optics, this problem could be less significant). On the other hand, the constraint on the joint limits is very invasive because it affects all the arm joints. For this reason it is advisable to move this task initially to the lower priority and increase it only when some joint limits are being approached.

A. Gripper position and orientation

Let o_e^* be the desired position of the gripper, *e.g.* suitable to perform the object grasping in the considered case study. This value can be computed by combining the pose measurement of the target object provided by the visual system, with the desired relative displacement of the gripper with respect to the target object in the grasping configuration. The corresponding position error is defined as $e_p = o_e^* - o_e$, and the task function is chosen equal to its square norm, yielding

$$\sigma_p = e_p^\top e_p, \tag{18}$$

where the desired task variable is $\sigma_p^* = 0$ (*i.e.* $\tilde{\sigma}_p = -\sigma_p$). The corresponding task Jacobian matrix is

$$J_p = 2e_p^\top \begin{bmatrix} I_3 & S(R_b o_e^b)\imath_z & J_{q,P} \end{bmatrix}, \tag{19}$$

where $S(\cdot)$ is the skew-symmetric matrix representing the vectorial product, $\imath_z = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\top$, and

$$J_q = \begin{bmatrix} J_{q,P} \\ J_{q,O} \end{bmatrix} = \begin{bmatrix} R_b & O_3 \\ O_3 & R_b \end{bmatrix} J_q^b, \tag{20}$$

with $J_q^b(q)$ the arm Jacobian matrix with respect to $\mathcal{B}$, with $J_{q,P}$ and $J_{q,O}$ $(3 \times \nu)$-matrices. The corresponding task Jacobian matrix of the uncontrolled state variables is

$$\overline{J}_p = 2e_p^\top S(R_b o_e^b) \begin{bmatrix} \imath_x & \imath_y \end{bmatrix}, \tag{21}$$

where $\imath_x = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$, $\imath_y = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top$.

Notice that if o_e is also measured by using visual markers attached to the gripper, the camera calibration error and the arm direct kinematics error will not affect the grasping accuracy in a similar way as in an image-based approach. Moreover, with the proposed choice of σ_p, only one DoF is required to execute this subtask, because only the norm of e_p will be nullified, *i.e.* the motion of the gripper during the

transient is constrained on a sphere of radius $\|e_p\|$. However, the corresponding task Jacobian J_p becomes singular when $e_p \to 0$. Nevertheless, in the task composition the generalized-inverse $J_p^\dagger$ is multiplied by σ_p. Hence, if J_p is full-rank, its determinant goes to zero only linearly when $e_p \to 0$, but σ_p goes to zero squarely.

Let $\{\eta_e, \epsilon_e\}$ and $\{\eta_e^*, \epsilon_e^*\}$ be the unit quaternions corresponding to R_e and to its desired value R_e^*, respectively. The corresponding orientation error can be expressed as

$$e_o = \eta_e \epsilon_e^* - \eta_e^* \epsilon_e - S(\epsilon_e^*)\epsilon_e. \tag{22}$$

The task function is chosen equal to

$$\sigma_o = e_o^\top e_o, \tag{23}$$

with the desired task variable $\sigma_o^* = 0$ (*i.e.* $\tilde{\sigma}_o = -\sigma_o$), while the corresponding task Jacobian matrix is

$$J_o = 2e_o^\top \begin{bmatrix} O_3 & \imath_z & J_{q,O} \end{bmatrix}. \tag{24}$$

The Jacobian matrix of the uncontrolled state variables is

$$\overline{J}_o = 2e_o^\top \begin{bmatrix} \imath_x & \imath_y \end{bmatrix}. \tag{25}$$

Remarks similar to the position case concerning the number of required DoFs, the singularity of the task Jacobian matrix, and the direct visual measurement of the gripper orientation can be repeated straightforwardly. Notice that these two sub-tasks are orthogonal.

B. Camera field of view

Let $s_c \in \mathbb{R}^2$ be the image feature vector of the projection of the observed markers centroid $o_c^b = \begin{bmatrix} x_c^b & y_c^b & z_c^b \end{bmatrix}^\top$ onto the normalized image plane, *i.e.*

$$s_c = \begin{bmatrix} X_c \\ Y_c \end{bmatrix} = \frac{1}{z_c^b} \begin{bmatrix} x_c^b \\ y_c^b \end{bmatrix}. \tag{26}$$

The FoV subtask consists in constraining s_c within a maximum distance with respect to a desired position s_c^* in the normalized image plane (*e.g.* the center of the image) by moving the vehicle base, *i.e.* the camera point of view (notice that we assumed the camera mounted on the vehicle base). Without loss of generality, any point of the observed target can be chosen to be controlled in the image. To achieve this goal, the following task function is considered:

$$\sigma_c = e_c^\top e_c, \tag{27}$$

where $e_c = s_c^* - s_c$, and the desired task variable is $\sigma_c^* = 0$ (*i.e.* $\tilde{\sigma}_c = -\sigma_c$), while the corresponding task Jacobian is

$$J_c = \begin{bmatrix} 2e_c^\top L_p R_b^\top & 2e_c^\top L_o R_b^\top \imath_z & O_{1\times\nu} \end{bmatrix}, \tag{28}$$

where

$$L_p = \frac{1}{z_c^b} \begin{bmatrix} -1 & 0 & X_c \\ 0 & -1 & Y_c \end{bmatrix}, \tag{29}$$

$$L_o = \begin{bmatrix} X_c Y_c & -(1 + X_c^2) & Y_c \\ 1 + Y_c^2 & -X_c Y_c & -X_c \end{bmatrix}. \tag{30}$$

Finally, the Jacobian of the uncontrolled state variables is

$$\overline{J}_c = 2e_c^\top L_o R_b^\top \begin{bmatrix} \imath_x & \imath_y \end{bmatrix}. \tag{31}$$

Notice that only one DoF is required to accomplish this subtask. In fact, the distance of o_c^b with respect to the desired optical ray corresponding to s_c^* is controlled. However, the corresponding task Jacobian matrix J_c is singular when $e_c \to 0$, but since it is not strictly required to accomplish the main mission that the target object is exactly in the desired position of the image (*e.g.* the center), this subtask can be activated only when σ_c exceeds a *safety threshold*.

C. Center of gravity

The weight of the robot arm can generate an undesired torque on the vehicle base depending on the configuration. In particular, the arm motion statically perturbs the system attitude and position when the center of gravity (CoG) of the arm p_g is not aligned with the CoG of the vehicle base along the gravitational line of action ι_z. Without loss of generality, the CoG of the vehicle base is assumed to be in O_b. Hence, by denoting with $e_g = \left((p_g - o_b)^\top \iota_z\right) \iota_z - (p_g - o_b)$ the error between the desired and the current position of the arm's CoG, the designed task function is defined as follows:

$$\sigma_g = e_g^\top e_g, \tag{32}$$

with the desired task variable $\sigma_g^* = 0$ (*i.e.* $\widetilde{\sigma}_g = -\sigma_g$). The corresponding task Jacobian matrix can be computed from the corresponding Jacobian represented. In fact, p_g^b is only a function of the arm joint configuration defined as

$$p_g^b = \frac{1}{m} \sum_{i=1}^{\nu} m_i p_{gi}^b, \tag{33}$$

where m_i and p_{gi}^b are the mass and the position of the CoG of the ith arm link, respectively, and $m = \sum_{i=1}^{\nu} m_i$.

The CoG of a partial chain of links can be represented, with respect to $\mathcal{B}$, from the link j to the end-effector, yielding

$$r_{gj}^b = \frac{1}{m} R_j^b \sum_{i=j}^{\nu} m_i p_{gi}^b, \tag{34}$$

where R_j^b is the rotation matrix between the jth arm link and $\mathcal{B}$. Finally, the differential relationship between p_g and the arm joint configuration is

$$\dot{p}_g^b = J_g^b \dot{q}, \tag{35}$$

where $J_g^b \in \mathbb{R}^{3 \times \nu}$ is the CoG Jacobian expressed in $\mathcal{B}$ and defined as follows

$$J_g^b = \frac{\partial p_g^b}{\partial q} = \begin{bmatrix} j_{g1}^b & \cdots & j_{g\nu}^b \end{bmatrix}, \tag{36}$$

with j_{gi}^b the ith joint Jacobian formulated from the partial CoG as follows

$$j_{gj}^b = \frac{\sum_{i=j}^{\nu} m_i}{m} S(R_j^b \iota_z) r_{gj}^b. \tag{37}$$

Finally, the corresponding task Jacobian is defined as

$$J_g = 2e_g^\top \begin{bmatrix} O_{1 \times 3} & S(R_b p_g^b) \iota_z & R_b J_g^b \end{bmatrix}, \tag{38}$$

while the uncontrolled state-variables Jacobian is

$$\overline{J}_g = 2e_g^\top S(R_b p_g^b) \begin{bmatrix} \iota_x & \iota_y \end{bmatrix}. \tag{39}$$

Notice that only one DoF is required and similar considerations as in the previous tasks definitions on the singularity of the Jacobian matrix when $e_g \to 0$ can be done.

D. Joint-limits avoidance constraint

A possible solution to avoid mechanical joint limits is to make attractive the central position of the joint ranges. Let $e_q = q^* - q$ be the corresponding error, where $q^* = q_L + \frac{1}{2}(q_H - q_L)$, with $q_L = \begin{bmatrix} q_{1L} & \cdots & q_{\nu L} \end{bmatrix}^\top$ and $q_H = \begin{bmatrix} q_{1H} & \cdots & q_{\nu H} \end{bmatrix}^\top$ the vectors of the low and high joint limits, respectively. The corresponding weighted square distance can be used as a task function to push the system to reach a safe configuration as follows

$$\sigma_l = e_q^\top \Lambda_l e_q, \tag{40}$$

where Λ_l is a following weighting matrix needed to normalize the control action with respect to the joint range

$$\Lambda_L = \text{diag}\{(q_{1H} - q_{1L})^{-2} \quad \cdots \quad (q_{\nu H} - q_{\nu L})^{-2}\}. \tag{41}$$

The desired task variable is $\sigma_l^* = 0$ (*i.e.* $\widetilde{\sigma}_l = -\sigma_l$), and the corresponding task Jacobian matrix is

$$J_l = \begin{bmatrix} O_{1 \times 4} & -2\Lambda_L e_q^\top \end{bmatrix}. \tag{42}$$

Notice that the uncontrolled state variables do not affect this subtask, hence the corresponding Jacobian is the null matrix.

Due to the higher priority tasks, some joint could reach anyway its limit. However, when a joint is approaching a mechanical limit, the corresponding component of the task function can be extracted from the previous subtask to form a new isolated subtask that can be activated on the top of the task stack. With this policy, if mechanically viable, the system will reconfigure its internal DoFs to achieve all the remaining subtasks until the dangerous condition will disappear and the original priority will be restored but starting from a different system configuration.

VI. SIMULATION RESULTS

The proposed approach has been tested in the simulator developed in the ARCAS project (www.arcas-project.eu), which is based on GAZEBO physics engine (http://gazebosim.org) (see Fig. 2(f)). A quadrotor with a weight of 5 kg and endowed with a downward looking camera at 25 Hz and a 6-DoFs robot arm plus a gripper has been employed. The camera has been positioned 50 cm ahead the vehicle base with an inclination of 30 deg with respect to the vertical axis in a way to observe the grasping manoeuvre without self-occlusion. The target object is a bar endowed with two visual markers at the ends. A UAM velocity control has also been employed [17], [18].

The assigned task is composed of two phases:

- *approaching phase* — the UAM starts from a distance of about 125 cm and has to move the gripper to a pre-grasping pose at 10 cm over the grasping pose;
- *grasping phase* — once the intermediate pose has been reached with an error less than a suitable threshold (2 cm for the position and 2 deg for the orientation), the target pose is moved towards the final grasping pose in 10 s;
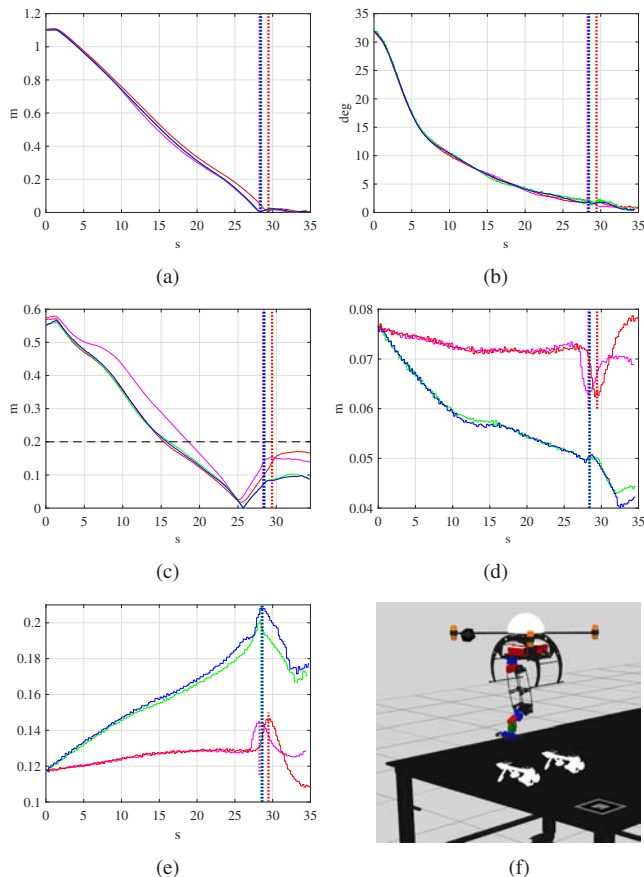
Fig. 2: Results of the simulated grasping task (case 1) in magenta, case 2) in red, case 3) in green, case 4) in blue): a) norm of the position error $\|e_p\|$; b) norm of the orientation error; c) norm of the FoV error $\|e_s\|$ with a 20 ± 2 cm activation/deactivation threshold; d) norm of the CoG error $\|e_g\|$; e) minimum distance from the joint limits normalized to the joint range; f) the GAZEBO simulator. The vertical lines indicate the conclusion of the approaching phase, while the end of each trajectory indicates the grasping time.

the closing of the gripper is then commanded when the final pose has been reached with a good accuracy (1 cm for the position and 1 deg for the orientation).

Four task-stack configurations have been simulated:

1) only main tasks for position and orientation control (Section V-A) active;
2) as case 1) plus FoV subtask (Section V-B);
3) as case 2) plus CoG subtask (Section V-C);
4) as case 3) plus joint limits subtask (Section V-D).

The achieved results are shown in Fig. 2 with different colors for the four considered control behaviors. A dashed vertical line is employed to highlight the end of the approaching phase and the starting of the grasping phase. Notice that the approaching and the grasping phases in all the considered case studies have different durations depending obviously on the selected control behavior, *i.e.* the active tasks stack.

Figure 2(a) shows the time history of the position error in norm during the task execution for each case study. For all cases, a smooth nullification of the pose error is observed. In particular, during the approaching phase the position error decreases almost linearly due to the saturation of the maximum

vehicle cruise velocity (10 cm/s in these case studies). Figure 2(b) shows the time history of the norm of the orientation error. The initial orientation is only a few degrees far from the final grasping pose, hence the error goes under the threshold in few seconds in all the case studies. Notice how the behavior of both the position and orientation errors are similar in all the cases coherently with the hierarchical task combination adopted in the proposed formulation, *i.e.* the activation of subtasks cannot affect significantly the behavior of the higher priority tasks.

Figure 2(c) shows the results achieved with the activation of the camera FoV subtask. In detail, this subtask is dynamically activated and deactivated by comparing the error norm with a double threshold, *i.e.* with a suitable hysteresis (20 ± 2 cm) to avoid chattering phenomena. By taking into account the camera pose with respect to $\mathcal{B}$, the desired position of the image features centroid has been chosen equal to $s_c^* = \begin{bmatrix} 0, -0.1 \end{bmatrix}^\top$. The achieved results show how except for case 1), *i.e.* when this subtask is activated, the FoV error is improved without affecting the movement of the gripper.

Figure 2(d) shows the time histories of the error in norm of the CoG subtask. For the chosen initial arm configuration the distance of the CoG with respect to the vehicle gravitational axis is 7.6 cm. In cases 1) and 2) this distance remains almost constant, while when the CoG subtask is active, *i.e.* for cases 3) and 4), the behavior is always improved without affecting the tasks with a higer priority in the stack.

Finally, in the last case study 4) also the joint-limits avoidance constraint is activated. Differently with respect to the previous cases, as it is shown in Fig. 2(e) (zero indicates the reaching of a joint limit, while 0.5 indicates that all joints are in the middle of the joint range), the task is not completely fulfilled, even if a clear increase of the distance with respect to the closest joint limit is guaranteed. This behavior is mainly due to the conflict with other subtasks that have a higher priority in the tasks stack. As described before, it is possible to increase the priority of this task in the stack when a joint limit is excessively close in a way to guarantee mechanical safety at the expense of other tasks.

VII. EXPERIMENTAL RESULTS

The UAM employed for the experimental tests has been developed in the ARCAS project. It is a multi-rotor aircraft with eight rotors in coaxial configuration with a 105 cm tip-to-tip wingspan, height of 50 cm, 13-inches propellers, and a total mass of 8.2 kg including batteries and the 6-DoFs robotic arm (see Fig. 3). The employed autopilot has been developed by CATEC (www.catec.aero) and allows also the control of the robot arm. A model-based design methodology [19] established on MATLAB/SIMULINK code generation tools has been adopted. The UAM has been endowed with an i7 ASCTEC MASTERMIND on-board for costly computing code, such as image processing. A motion capture system running at 100 Hz has been used as the positioning system, while the attitude is measured with the onboard IMU. A 6-DoFs manipulator [20] running at 50 Hz is attached below the vehicle base. The robotic manipulator direct kinematic model
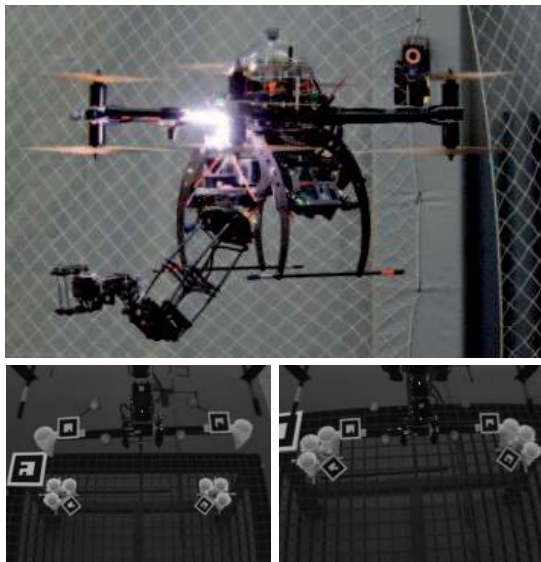
Fig. 3: Bonebraker UAM employed during the experiments (top), and images from the onboard camera during the approaching phase (bottom-left) and at the plugging instant (bottom-right).
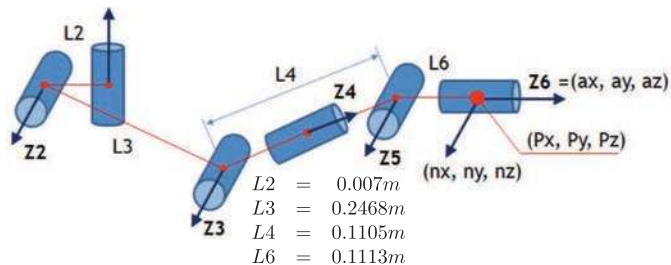

$$L2 = 0.007m$$
$$L3 = 0.2468m$$
$$L4 = 0.1105m$$
$$L6 = 0.1113m$$

Fig. 4: Reference frames following Denavit-Hartenberg convention.

TABLE I: Arm Denavit-Hartenberg parameters.

Link i	θ_i	d_i	a_{i-1}	α_{i-1}
1	θ_1	0	0	0
2	θ_2	0	-L2	$\pi/2$
3	$\theta_3 + \pi/2$	0	L3	0
4	θ_4	L4	0	$\pi/2$
5	θ_5	0	0	$-\pi/2$
6	θ_6	L6	0	$\pi/2$

is obtained by using the well known Denavit-Hartenberg convention (see Fig. 4 and Table I).

A high-definition camera running at 14 Hz has been positioned as in the simulation case study. The calibration of the vision system has been divided in two steps. First, the camera intrinsic parameters are obtained with several views of a calibration pattern (*i.e.* a chessboard). Secondly, the extrinsic parameters are obtained using the motion capture system to precisely localize the platform body frame ($\mathcal{B}$) and an object in the scene (which corresponds to a marker). By knowing the pose of the camera attached to the quadrotor body frame, we can trivially obtain the frame transformation between the camera and the object. However, the estimation of the error between the camera and the optical frames is also required. The marker detector is employed estimating the marker pose with respect to the optical frame. Then, a pose average of the difference between the camera and the optical frames is
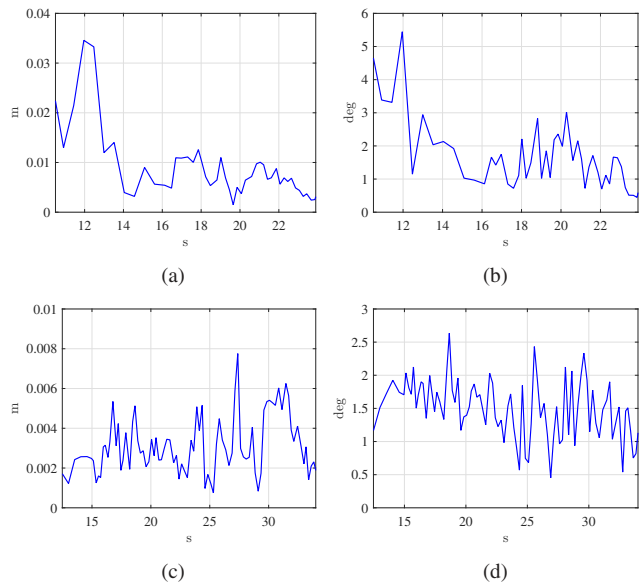

Fig. 5: Norm of the object position errors with respect to the ground-truth during grasping (a) and plugging (c) maneuvers. The corresponding orientation errors are shown in (b) and (d).

computed with respect to the object.

Figure 5 shows the error between the detected bar and the ground truth poses during grasping and plugging tasks.

The experimental task consists in plugging a bar endowed with two clipping systems at the ends into a fixed base[1], as shown in the bottom part of Fig. 3. As for the simulated case studies, the mission has been decomposed into two steps: the approaching phase, to move the bar over the plugging base at a distance of 5 cm, and the final plugging phase. During this latter phase the FoV task is turned off because the constraint is always satisfied by the system mechanical configuration and the adopted optics. The task requires high accuracy both in position and orientation, *i.e.* about 1 cm for the position and 1 deg for the orientation, that has to be guaranteed stable in time to avoid undesired collisions. To cope with this requirement, the bar has been endowed with visual markers as for the plugging base. Hence, the positioning error has been computed by using the measurement of the bar and of the base in a way to mitigate the effects of the calibration errors.

The achieved results are shown in Fig. 6. Figures 6(a) and 6(b) show the time history of the norm of the position and orientation errors, respectively. The vertical dashed line indicates the end of the approaching phase and the beginning of the plugging phase. The plugging instant corresponds with the end time of the plots. One can observe how the initial errors are quite high because the system starts from a distance of about 40 cm from the goal position, and with a significant orientation error too, however for both errors the target accuracy has been reached in a fast and stable way.

The time history of the norm of the FoV error $\|e_s\|$ is shown in Fig. 6(c), from which one can observe how this subtask is

---

[1]Several grasping and plugging experiments of different type of bars are shown in the video attached to this paper.
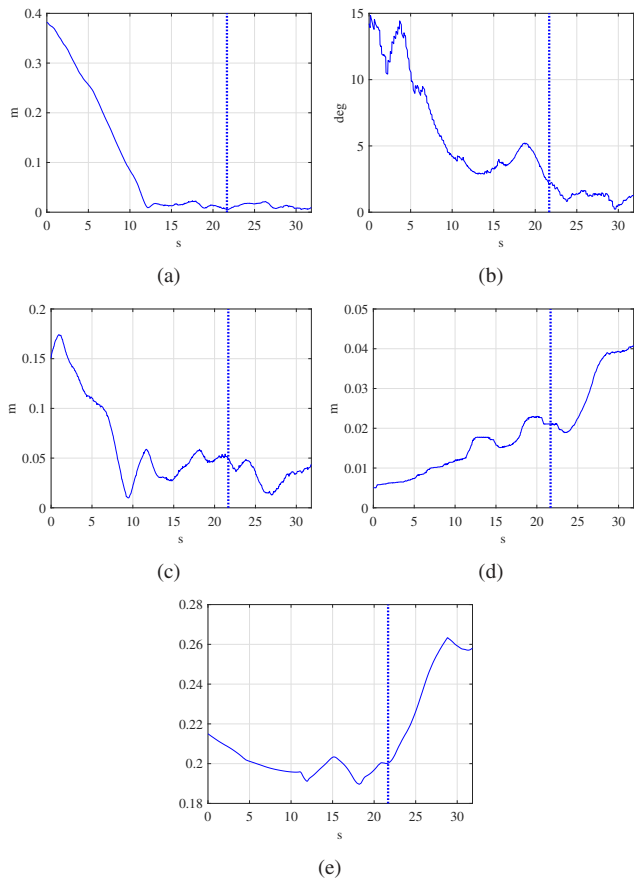
Fig. 6: Experimental results of the plugging task: a) norm of the position error $\|e_p\|$; b) norm of the orientation error; c) norm of the FoV error $\|e_s\|$; d) norm of the CoG error $\|e_g\|$ with a 15 ± 2 cm activation/deactivation threshold; e) minimum distance from the joint limits normalized to the joint range. The vertical dotted lines indicate the time instant when the approaching phase is concluded, while the each trajectory ends at the plugging time.

suitably executed, hence the system is able to prevent the loss of the visual markers from the camera FoV.

The CoG subtask has been employed with an activation/deactivation threshold of 15 ± 2cm. However, it is never activated because the high-priority FoV subtask determines arm configurations already compatible with the CoG subtask. In fact, the alignment error of the CoGs is lower than 4 cm.

Finally, Figure 6(e) shows the minimum distance computed over all the arm joints from the corresponding joint limits normalized to the joint range (zero indicates the reaching of a joint limit, while 0.5 indicates that all joints are in the middle of the joint range). Even if this is the lower priority task, a safety distance of more than 20% of the joint ranges, in the worst case, is always preserved.

VIII. CONCLUSIONS

A hybrid visual servoing with a hierarchical task-priority control framework suitable for UAM has been presented in this work. The underactuation of the VToL vehicle has been explicitly taken into account in a new general formulation that also guarantees the decoupling of independent tasks. Several subtasks have been proposed to reduce the error of the desired

gripper position and orientation, to maintain the target in the camera FoV, to vertically align the arm CoG with the quadrotor gravitational vector, and to avoid arm joint limits. Simulations and experiments validated the proposed solution.

REFERENCES

[1] L. Marconi, F. Basile, G. Caprari, R. Carloni, P. Chiacchio, C. Hurzeler, V. Lippiello, R. Naldi, J. Nikolic, B. Siciliano, S. Stramigioli, and E. Zwicker, "Aerial service robotics: The AIRobots perspective," in *2nd International Conference on Applied Robotics for the Power Industry*, Sept 2012, pp. 64–69.

[2] L. Marconi and R. Naldi, "Control of aerial robots: Hybrid force and position feedback for a ducted fan," *IEEE Control Systems*, vol. 32, no. 4, pp. 43–65, Aug 2012.

[3] F. Forte, R. Naldi, A. Macchelli, and L. Marconi, "On the control of an aerial manipulator interacting with the environment," in *IEEE International Conference on Robotics and Automation*, May 2014, pp. 4487–4492.

[4] V. Lippiello and F. Ruggiero, "Cartesian impedance control of a UAV with a robotic arm," in *10th IFAC Symposium on Robot Control*, Sep 2012, pp. 704–709.

[5] ——, "Exploiting redundancy in cartesian impedance control of uavs equipped with a robotic arm," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2012, pp. 3768–3773.

[6] S. Kim, S. Choi, and H. Kim, "Aerial manipulation using a quadrotor with a two dof robotic arm," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013, pp. 4990–4995.

[7] C. Korpela, M. Orsag, and P. Oh, "Towards valve turning using a dual-arm aerial manipulator," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2014, pp. 3411–3416.

[8] R. Mebarki, V. Lippiello, and B. Siciliano, "Exploiting image moments for aerial manipulation control," in *ASME Dynamic Systems and Control Conference*, Oct 2013.

[9] R. Mebarki and V. Lippiello, "Image-based control for aerial manipulation," *Asian Journal of Control*, vol. 16, no. 3, pp. 646–656, 2014.

[10] A. Santamaria-Navarro, V. Lippiello, and J. Andrade-Cetto, "Task priority control for aerial surveillance," in *12th IEEE International Symposium on Safety, Security, and Rescue Robotics*, Oct 2014, pp. 1–6.

[11] L. R. Buonocore, J. Cacace, and V. Lippiello, "Hybrid visual servoing for aerial grasping with hierarchical task priority control," in *23st Mediterranean Conference on Control Automation (MED)*, June 2015, pp. 651–657.

[12] M. Fiala, "ARTag, a fiducial marker system using digital techniques," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, June 2005, pp. 590–596.

[13] A. Amor-Martinez, A. Ruiz, F. Moreno-Noguer, and A. Sanfeliu, "On-board real-time pose estimation for UAVs using deformable visual contour registration," in *2014 IEEE International Conference on Robotics and Automation*, May 2014, pp. 2595–2601.

[14] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155–166, 2009.

[15] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[16] G. Antonelli, "Stability analysis for prioritized closed-loop inverse kinematic algorithms for redundant robotic systems," *IEEE Transactions on Robotics*, vol. 25, no. 5, pp. 985–994, 2009.

[17] R. Mebarki, V. Lippiello, and B. Siciliano, "Image-based control for dynamically cross-coupled aerial manipulation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2014, pp. 4827–4833.

[18] R. Mebarki and V. Lippiello, "Image moments-based velocity estimation of uavs in gps denied environments," in *IEEE International Symposium on Safety, Security, and Rescue Robotics*, Oct 2014, pp. 1–6.

[19] D. Santamaria, F. Alarcon, A. Jimenz, A. Viguria, M. Béjar, and A. Ollero, "Model-based design, development and validation for UAS critical software," *Journal of Intelligent & Robotic Systems*, vol. 65, no. 1–4, pp. 103–114, 2012.

[20] R. Cano, C. Pérez, F. Pruano, A. Ollero, and G. Heredia, "Mechanical design of a 6-DOF aerial manipulator for assembling bar structures using UAVs," in *2nd RED-UAS 2013 Workshop on Research, Education and Development of Unmanned Aerial Systems*, Compiegne, F, 2013.