# Toward Autonomous Avian-Inspired Grasping for Micro Aerial Vehicles

Justin Thomas, Giuseppe Loianno, Joseph Polin, Koushil Sreenath, and Vijay Kumar‡

Abstract. Micro Aerial Vehicles, particularly quadrotors, have been used in a wide range of applications. However, the literature on aerial manipulation and grasping is limited and the work is based on quasi-static models. In this paper, we draw inspiration from agile, fast-moving birds such as raptors, that are able to capture moving prey on the ground or in water, and develop similar capabilities for quadrotors. We address dynamic grasping, an approach to prehensile grasping in which the dynamics of the robot and its gripper are significant and must be explicitly modeled and controlled for successful execution. Dynamic grasping is relevant for fast pick-and-place operations, transportation and delivery of objects, and placing or retrieving sensors. We show how this capability can be realized (a) using a motion capture system and (b) without external sensors relying only on onboard sensors. In both cases we describe the dynamic model, and trajectory planning and control algorithms. In particular, we present a methodology for flying and grasping a cylindrical object using feedback from a monocular camera and an Inertial Measurement Unit onboard the aerial robot. This is accomplished by mapping the dynamics of the quadrotor to a level virtual image plane, which in turn enables dynamically-feasible trajectory planning for image features in the image space, and a vision-based controller with guaranteed convergence properties. We also present experimental results obtained with a quadrotor equipped with an articulated gripper to illustrate both approaches.

# 1. Introduction

Nature provides many examples of graceful and yet purposeful flight that inspire us to create aerial robots. We are particularly interested in raptors that successfully hunt fast moving and elusive prey. Although birds differ from aerial robots in a number of fundamental ways, we still have the opportunity to incorporate avian behaviors, such as dynamic high-speed grasping, in aerial robots.

The capabilities of Micro Aerial Vehicles (MAVs) are rapidly expanding to include surveillance [3], construction [4], manipulation of slung loads [5], collaborative transportation [6, 7], and mapping of unknown environments using aerodynamic effects [8]. However, aerial vehicles' flight durations are limited by the energy density of batteries and the speed of aerial manipulation is restricted to quasi-static interactions with the environment. An aerial vehicle endowed with capabilities traditionally ascribed to raptors would be instrumental towards speeding-up interaction with the environment and mitigating the energy-density restriction. Acquiring, transporting and deploying payloads while maintaining a significant velocity is important since it would result in minimization of the required flight time, thereby improving time and energy efficiency. Beyond energy-efficiency, high-speed grasping would be particularly useful if a MAV was needed to quickly acquire or deploy sensors, materials, or robots.

High-speed grasping would be extremely difficult without an articulated appendage capable of interacting with the environment (*e.g.* see Figure 1). Previous research includes an adaptive hand [9], a passively actuated gripper for perching [10], a servodriven claw for construction [4], and a gripper which pierced its targets [11]. These grippers vary in method and application, but they suffer from a common limitation: to be effective, the vehicle must approach the target perpendicular to the plane of the target and the approach velocity must be near zero. The ingressive gripper in [3] was able to perch with more aggressive trajectories by triggering a spring-loaded claw that would engage upon contact, but still required a velocity normal to the target surface. Perching has been demonstrated using fixed-wing Unmanned Aerial Vehicles (UAVs) but focused either on the perching mechanism [12, 13] or trajectory planning and nonlinear control [14]. In the later, an external motion capture system was required.



Figure 1: (a) A Red Kite swoops down and uses visual feedback to approach, grasp, and retrieve food on the ground [15]. (b) A bald eagle uses a similar strategy to hunt prey in the water [16].

It is natural to look to nature for inspiration when approaching such design challenges. We can observe from video footage that a raptor sweeps its legs and claws backwards while capturing prey, thereby reducing the relative velocity between the claws and the prey [16]. This allows the bird, without stopping, to have a near-zero relative velocity between the claw and the prey. We draw inspiration from this hunting methodology to enable high-speed aerial grasping and manipulation for MAVs.

To enable autonomous grasping, the robot must be able to detect the object of interest and use visual feedback to control the robot's motion. However, to maintain agility, the robot must have low inertia (*i.e.* minimal sensor payload) and consider the dynamics of the system. Another limitation is our poor understanding of the perception-action loops required for agile flight and manipulation. One can observe that visual feedback is used to close the control loop while dynamically grasping prey (see Figure 1). In scenarios like this, a monocular camera is an ideal sensor, especially when combined with an Inertial Measurement Unit (IMU) [17, 18], and motivates either Position Based Visual Servoing (PBVS) or Image Based Visual Servoing (IBVS) [19]. PBVS requires an explicit estimation of the pose of the robot in the inertial frame while IBVS acts directly using feedback from the image coordinates. In particular, a single monocular camera is sufficient for visual servoing when there is some known geometry or structure in the environment.

There are many excellent tutorials on visual servoing [20, 19, 21, 22]; however, most approaches assume first-order or fully-actuated systems. For example, [23] demonstrated robustness to camera calibration, but only considered a first-order system. Stability was proven for second order systems, but assumed full actuation [24]. More recently, [25] and [26] leveraged a spherical camera model and utilized backstepping to design nonlinear controllers for a specific class of underactuated second-order systems. As is typical in backstepping, however, it is necessary to assume that the inner control loops are significantly faster than the outer ones. There have been some preliminary efforts towards autonomous landing, but an estimate of velocity in the inertial frame is obtained using an external motion capture system [27]. We are not aware of any work that addresses grasping and perching at high speeds using only onboard sensors.

Therefore, our goal is to develop and demonstrate an approach to grasping at significant velocities, requiring explicit modeling and control of the inertia and momentum of the flying robot and its appendages. In particular, we consider a quadrotor, which is appealing because of its mechanical simplicity, its agility, its ability to hover, and its well-understood dynamics [5]. Quadrotors are similar to helicopters, but have four rotors with parallel axes of rotation, and therefore, parallel thrust vectors [28]. The system is underactuated; however, it is possible to design controllers that guarantee convergence from almost any point on SE(3), the Euclidean motion group in three dimensions [29]. Similar controllers have been derived for a quadrotor carrying a cable-suspended payload [5]. However, both of these approaches require full knowledge of the state. In order to achieve these goals, we will design and model the system with an articulated appendage and consider the dynamics of the system directly in the image plane (rather than in the Cartesian space) to develop an IBVS controller based on visual features of a cylinder.

In this article, we describe a proof-of-concept robot with an articulated gripper, demonstrate dynamic grasping using a motion capture system, and make a step towards enabling grasping maneuvers using vision. The paper is organized as follows: we begin with a list of assumptions in Section 2 and develop the dynamic model of a quadrotor with a swinging gripper in Section 3. In Section 4, the dynamics of the robot are mapped to the image plane of an on-board camera, while Section 5 leverages the differential flatness property to enable trajectory generation in the Euclidean and the image spaces. Section 6 presents a controller for the coupled robot-gripper system to stabilize a MAV performing dynamic maneuvers with our gripper. Additionally in this section, we build upon existing vision and geometric-control literature to develop a vision-based controller. Following this, Section 7 describes the hardware used in experiments, particularly the actuated gripper and the camera system. We present experimental validation in Section 8, which includes high-speed dynamic grasping and vision-based control. The dynamic grasping results with a motion capture system validate (a) the dynamical model, (b) the proposed dynamic trajectory generation method, and (c) the feasibility of dynamic grasping maneuvers with our MAVs. Further, the vision-based results validate (a) the mapping of the dynamic model to the image space, (b) the feasibility of the trajectory generation method in the image space, (c) the stability of the proposed nonlinear controller, and (d) the feasibility of grasping maneuvers using vision feedback on our MAVs. Finally, Section 9 provides concluding remarks and comments on opportunities for future work.

## 2. Assumptions

In this paper, there are several assumptions that we make in the development of the dynamical model, control, and vision feedback. For ease of reading, we will enumerate a list of assumptions here. In particular, there will be two sets of assumptions, one for an inertial controller that uses a motion capture system, and the other for a vision based controller.

For developing a dynamical model, control design, and trajectory planning in the inertial frame, we make the following assumptions:

- A1) The dominant motion for the dynamic grasping maneuver occurs in the vertical plane of the inertial frame.
- A2) The location of the target payload is known. In particular, this is resolved by tracking the target with the motion capture system.
- A3) The target payload has no mass, which follows from the fact that the coupled mass of the quadrotor and gripper (658 g) is much larger than the payload's (24 g).

In addition to A1 and A3, the following assumptions are made for the development of the vision based feedback control mode:

- A4) The gripper is considered to have no mass, thereby reducing a degree of freedom from the dynamic system.
- A5) The cylinder is orientated such that its axis is perpendicular to the vertical plane.
- A6) The radius of the cylinder is known a priori. Though other sensors exist for measuring distance, we prefer to leverage sensors that are already mounted on the robot.

# 3. Dynamic Model

Following the list of assumptions, this section will develop the dynamical model of the quadrotor with a gripper in the vertical plane. The list of symbols used in this section and the rest of the paper are tabulated in Table 1.

# 3.1. Kinematics

We formulate the problem in the sagittal (*i.e.* vertical) plane of the robot based on Assumption A1, which has enabled high-speed dynamic grasping for aerial robots [1]. A

 $f \in \mathbb{R}$ Sum of forces produced by all 4 propellers  $M \in \mathbb{R}$ Moment generated by  $f_1$  and  $f_3$  about body's y-axis  $\mathbf{r}_q \in \mathbb{R}^2$ Position of robot's COM without gripper, or  $[x_q, z_q]^T$ Position of gripper's COM, or  $[x_q, z_q]^T$  $\mathbf{r}_g \in \mathbb{R}^2$  $\theta \in \mathbb{R}$ Pitch of robot  $\beta \in \mathbb{R}$ Angle of gripper relative to horizontal plane in inertial space  $R \in SO(2)$ Rotation matrix that rotates from robot to inertial frame by angle  $\theta$  $L_g \in \mathbb{R}$ Distance between the robot and gripper COM's  $m_q \in \mathbb{R}$ Mass of robot without gripper  $m_a \in \mathbb{R}$ Mass of gripper  $J_a \in \mathbb{R}$ Rotational inertia of robot about its COM  $J_q \in \mathbb{R}$ Rotational of inertia of gripper about its COM  $\omega_a \in \mathbb{R}$ Angular velocity of robot  $\omega_q \in \mathbb{R}$ Angular velocity of gripper  $g \in \mathbb{R}$ Gravitational acceleration  $R_t \in \mathbb{R}$ Radius of target cylinder  $i^{th}$  ray from center of camera to tangent point on cylinder, or  $[x_{t,i}, z_{t,i}]$  $\mathbf{r}_{t,i} \in \mathbb{R}^2$  $\mathbf{v} \in \mathbb{R}^2$ Feature coordinates in virtual image plane, or  $[v_1, v_2]^T$  $J \in \mathbb{R}^{2 \times 2}$ Image Jacobian  $\Gamma:S\to V$ Map from inertial frame, S, to the visual coordinate frame, V $\mathbf{y} \in \mathbb{R}^3$ Set of flat outputs, or  $[\mathbf{r}_a^T, \beta]^T$ 

Table 1: Glossary of important symbols.



Figure 2: (a) The quadrotor has a total thrust,  $f = \sum_i f_i$ , in the direction of the vertical body frame axis ( $Re_2$ ). A pitch moment, M, results from the difference between  $f_3$  and  $f_1$ . (b) We assume that the target is located at the origin and the quadrotor is located at  $(x_q, z_q)$ . The focal length of the camera,  $f_x$ , defines the location of the image plane relative to the quadrotor and the image coordinates are given by  $v_1$  and  $v_2$ . The optical ray tangent to the target intersects the target at  $(x_t, z_t)$ . The coordinate system of the camera is indicated by  $\mathbf{x}_c$  and  $\mathbf{z}_c$ .

rotating articulated appendage is attached to the robot like the legs of a raptor rotating about the hip.

The position of the robot's Center of Mass (COM) is given by  $\mathbf{r}_q = [x_q, z_q]^T$ and the gripper's COM is  $\mathbf{r}_g = [x_g, z_g]^T$ . Let the pitch of the robot be  $\theta$  and the angle of the gripper relative to the horizontal be  $\beta$  as displayed in Figure 2a so that  $R = R(\theta) \in SO(2)$  is the rotation matrix defining the pitch of the robot and  $R_\beta = R(\beta) \in SO(2)$  is the rotation matrix defining the rotation of the gripper. Then, the position of the gripper is entirely determined from the position of the quadrotor and the angle of the gripper through

$$\mathbf{r}_g = \mathbf{r}_q + L_g R_\beta \mathbf{e}_1 \tag{1}$$

where  $L_g$  is the constant distance between the quadrotor and gripper's center of masses and  $\mathbf{e}_1 = [1, 0]^T$ . Furthermore, higher-order derivatives of the gripper position can be expressed as functions of the position of the quadrotor, the angle of the gripper, and their higher-order derivatives.

#### 3.2. Dynamics in the Inertial Frame

The dynamics of the planar coupled system are determined using Lagrangian mechanics where the potential energy is

$$U = m_q g \left( \mathbf{r}_q \cdot \mathbf{e}_2 \right) + m_g g \left( \mathbf{r}_g \cdot \mathbf{e}_2 \right), \tag{2}$$

and the kinetic energy is

$$T = \frac{1}{2} \left( m_g \| \dot{\mathbf{r}}_g \|_2^2 + m_q \| \dot{\mathbf{r}}_q \|_2^2 + J_g \omega_g^2 + J_q \omega_q^2 \right)$$
(3)

where g is gravity,  $m_q$  is the mass of the quadrotor,  $m_g$  is the mass of the gripper,  $\mathbf{e}_2 = [0, 1]^T$ ,  $J_q$  is the angular inertia of the quadrotor,  $J_g$  is the angular inertia of the gripper (about its COM),  $\omega_g$  is the angular velocity of the gripper, and  $\omega_q$  is the angular velocity of the quadrotor. Then, the vector of generalized coordinates (**q**) and the vector of generalized forces and moments (**F**) are given by

$$\mathbf{q} = \begin{bmatrix} \mathbf{r}_q \\ \theta \\ \beta \end{bmatrix} \quad \text{and} \quad \mathbf{F} = \begin{bmatrix} fR\mathbf{e}_2 \\ M - \tau \\ \tau \end{bmatrix}, \quad (4)$$

where f is the total thrust,  $\tau$  is the actuator torque on the gripper arm, and M is the moment produced by the thrust differential between the front and rear rotors as portrayed in Figure 2a. The dynamics are determined using the Euler-Lagrange equations so that

$$\ddot{\mathbf{q}} = D^{-1} \left( \mathbf{F} - C \dot{\mathbf{q}} - \mathbf{G} \right) \tag{5}$$

where the matrices D, C, and **G** are

$$D = \begin{bmatrix} m_g + m_q & 0 & 0 & -L_g m_g \sin(\beta) \\ 0 & m_g + m_q & 0 & -L_g m_g \cos(\beta) \\ 0 & 0 & J_q & 0 \\ -L_g m_g \sin(\beta) & -L_g m_g \cos(\beta) & 0 & J_g + L_g^2 m_g \end{bmatrix},$$
 (6)

$$C = \begin{bmatrix} 0 & 0 & 0 & -L_g m_g \cos(\beta)\dot{\beta} \\ 0 & 0 & 0 & L_g m_g \sin(\beta)\dot{\beta} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{ and } \mathbf{G} = \begin{bmatrix} 0 \\ g(m_g + m_q) \\ 0 \\ -gL_g m_g \cos(\beta) \end{bmatrix}.$$
(7)

See [30] for the complete 3-D dynamic model of the specific case when  $\tau = 0$  and  $J_g = 0$ .

Having presented the dynamical model of the quadrotor with a gripper in the inertial frame, we will next present a model for vision to enable mapping the dynamical model in the inertial frame onto a camera image plane.

# 4. Vision

In this section, we present an overview of the vision system, outline the camera model, derive the geometric constraints on the cylinder detection in the image plane, and map the previously computed dynamics into the image plane. We will use the following nomenclature.

Let  $T \in SE(3)$  be the homogeneous transformation matrix from the camera frame to the world frame,  $f_{\alpha}$  denote a focal length in the  $\alpha$  direction,  $c_{\alpha}$  be the center image pixel in the  $\alpha$  direction, and  $\lambda$  be an arbitrary scaling factor.

#### 4.1. Camera model

The camera is modeled using a standard pinhole perspective camera model so that a generic point in the world,  $[X, Y, Z, 1]^T$ , is projected onto the image plane,  $[x', y', 1]^T$ , according to [31] such that

$$\lambda \begin{bmatrix} x'\\y'\\1 \end{bmatrix} = KP_0T^{-1} \begin{bmatrix} X\\Y\\Z\\1 \end{bmatrix}, \quad K = \begin{bmatrix} f_x & 0 & c_x\\0 & f_y & c_y\\0 & 0 & 1 \end{bmatrix}, \quad P_0 = \begin{bmatrix} \mathcal{I}_{3\times3} & \mathbf{0}_{3\times1} \end{bmatrix}.$$
(8)

From here on, we will use the calibrated image coordinates in the camera frame, (x, y),

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = P_0 T^{-1} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \implies \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K^{-1} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}, \qquad (9)$$

which are equivalent to the transformation and projection of points in the world to an image plane with unity focal length and a centered image coordinate system.

We will next present the geometry of the problem where the camera is fixed to the quadrotor and develop the mapping between the pose of the quadrotor and the location of the image features in the camera image plane.

#### 4.2. Geometry

Let the image features be the points whose rays are tangent to the cylinder and lie in the vertical plane. In contrast to typical visual servoing approaches, these points are now a function of the position of the robot. Therefore, we cannot use the standard image Jacobian, which assumes the target points are stationary in the inertial frame [21].

In order to formulate the mapping between the image plane and the robot pose, let the target cylinder be centered at the origin,  $R_t$  denote the radius of the target cylinder, and  $\mathbf{r}_t$  be a tangent point on it as shown in Figure 2b. With the camera at the same position as the quadrotor, there are two geometric constraints in the inertial frame,

$$\left\|\mathbf{r}_{t}\right\|_{2} = R_{t} \tag{10}$$

$$\|\mathbf{r}_{q}\|_{2}^{2} = \|\mathbf{r}_{q} - \mathbf{r}_{t}\|_{2}^{2} + R_{t}^{2}$$
(11)

where  $\|\cdot\|_2$  is the 2-norm in the Euclidean space. These equations have two solutions which represent the two tangent points,

$$\mathbf{r}_{t,i} = \frac{R_t^2}{\|\mathbf{r}_q\|^2} \left( \begin{bmatrix} x_q \\ z_q \end{bmatrix} \pm \begin{bmatrix} -z_q \\ x_q \end{bmatrix} \sqrt{\frac{\|\mathbf{r}_q\|^2}{R_t^2} - 1} \right).$$
(12)

Unfortunately, the features in the image plane are coupled with the attitude. Thus, the image features would not allow for the necessary attitude-decoupled mapping between the position of the robot and the image features as required for the features to be flat outputs as outlined in section 5.1.2. Similarly to [32], the calibrated image coordinates are mapped to coordinates on a level virtual image plane by rotating the camera coordinate system to a virtual frame where  $\theta = 0$ .

Then, the virtual calibrated coordinates of the features can be computed using the position of the quadrotor, (12), and

$$\lambda \begin{bmatrix} v_i \\ 0 \\ 1 \end{bmatrix} = P_0 T^{-1} \begin{vmatrix} x_{t,i} \\ 0 \\ z_{t,i} \\ 1 \end{vmatrix}$$
(13)

with the appropriate transformation, T, and independent of the pitch,  $\theta$ . The virtual coordinates,  $\mathbf{v} = [v_1, v_2]^T$ , in (13) provide two equations which can be solved to determine the robot and camera position as a function of the virtual image coordinates.

We also define the space  $S = {\mathbf{r}_q \in \mathbb{R}^2 | 2R_t \leq ||\mathbf{r}_q|| \leq B_r, z_q > R_t}$ , such that the quadrotor's position is bounded below by  $2R_t$  and bounded above by  $B_r$ , and the quadrotor is always above the cylinder. Then, there exists  $V \subset \mathbb{R}^2$  and a smooth global diffeomorphism  $\Gamma : S \longrightarrow V$  such that

$$\mathbf{v} = \frac{f_x}{z_q^2 - R_t^2} \begin{bmatrix} x_q z_q + R_t^2 \sqrt{\frac{\|\mathbf{r}_q\|^2}{R_t^2} - 1} \\ x_q z_q - R_t^2 \sqrt{\frac{\|\mathbf{r}_q\|^2}{R_t^2} - 1} \end{bmatrix} \equiv \Gamma(\mathbf{r}_q),$$
(14)

$$\dot{\mathbf{v}} = \frac{d\Gamma\left(\mathbf{r}_{q}\right)}{dt} = \frac{\partial}{\partial \dot{\mathbf{r}}_{q}} \left(\frac{d\Gamma\left(\mathbf{r}_{q}\right)}{dt}\right) \dot{\mathbf{r}}_{q} \equiv J \dot{\mathbf{r}}_{q},\tag{15}$$

where J is the image Jacobian [33]. Note that J can be expressed as a function of either the image coordinates or the position of the robot by using (14) and the fact that  $\Gamma$  is invertible. Having established a mapping between the Cartesian coordinates and the image coordinates, we will next develop a dynamic model of the quadrotor system directly in the image coordinates.

#### 4.3. Dynamics in the Image Plane

For simplicity in the visual system, we now assume that the gripper is massless (*i.e.*  $m_g = 0$  and  $J_g = 0$  according to Assumption A4) and leave its incorporation for future work since we have not yet found flat outputs for the coupled system with vision and a gripper. Then, D is diagonal (see (6)), the centripetal and Coriolis terms, C, are zero (see (7)), the gravitational term,  $\mathbf{G}$ , is zero except for the second element (also (7)), and  $\tau = 0$ . Since the system has three degrees of freedom (we no longer have  $\beta$ ), given by  $\mathbf{q}$ , and only two control inputs that appear in  $\mathbf{F}$ , the system is underactuated.

Now,  $\dot{\mathbf{r}}_q$  and  $\ddot{\mathbf{r}}_q$  can be expressed as functions of the image coordinates using the inverse of the image Jacobian, J. Then, the dynamics in (5) can be expressed in terms

of the image coordinates using

$$\dot{\mathbf{r}}_q = J^{-1} \dot{\mathbf{v}} \tag{16}$$

$$\ddot{\mathbf{r}}_q = J^{-1} \ddot{\mathbf{v}} - J^{-1} \dot{J} J^{-1} \dot{\mathbf{v}}$$
(17)

so that the dynamics in the image coordinates are:

$$\ddot{\mathbf{v}} = \frac{1}{m} J \left[ f R \mathbf{e}_2 - \mathbf{G}_{1:2} \right] + \dot{J} J^{-1} \dot{\mathbf{v}}$$
(18)

$$J_q \ddot{\theta} = M \tag{19}$$

where  $\mathbf{G}_{1:2}$  denotes the first two elements of  $\mathbf{G}$ . Equation (18) presents the translational dynamics directly in the image coordinates. In the next section, we will demonstrate that  $\mathbf{v}$  forms a set of flat outputs for the system, enabling trajectory design directly in the image space.

# 5. Dynamically Feasible Trajectories

#### 5.1. Differential Flatness

A system is differentially flat if there exists a change of coordinates which allows the state,  $(\mathbf{q}, \dot{\mathbf{q}})$ , and control inputs,  $\mathbf{u}$ , to be written as functions of the flat outputs and their derivatives  $(y_i, \dot{y}_i, \ddot{y}_i, ...)$  [34]. If the change of coordinates is a diffeomorphism, we can plan trajectories using the flat outputs and their derivatives in the flat space since there is a unique mapping to the full state space of the dynamic system.

5.1.1. Flat Outputs in the Inertial Frame The coupled system comprising of the quadrotor and the actuated gripper, whose dynamics is given by (5), is differentially flat with a set of flat outputs given by (see [1] for details)

$$\mathbf{y} = \left[ \begin{array}{cc} x_q & z_q & \beta \end{array} \right]^T \tag{20}$$

and we remind the reader that  $\beta$  is the angle of the gripper arm relative to the horizontal axis. Consequently, any sufficiently smooth trajectory in the space of flat outputs is automatically guaranteed to satisfy the equations of motion. Further, we see that the control inputs to the system are functions of the snap  $(\mathbf{y}^{(4)})$  of the trajectories. Thus, trajectories planned in the flat space are required to be smooth in position  $(\mathbf{y})$ , velocity  $(\dot{\mathbf{y}})$ , acceleration  $(\ddot{\mathbf{y}})$ , and jerk  $(\ddot{\mathbf{y}})$ .

5.1.2. Flat Outputs in the Image Space To simplify the planning for the vision system, we assume that the gripper is massless (Assumption A4), which reduces the degrees of freedom of the system. A proposed set of flat outputs in the image space are the image coordinates,  $\mathbf{v}$ . These would be convenient since planning dynamically feasible trajectories in the image space, V, would be as simple as planning a sufficiently smooth

trajectory in the image coordinates. Assuming that the radius,  $R_t$ , of the target cylinder is known (Assumption A6), there exists a diffeomorphism between the image coordinates and the position of the robot, namely  $\Gamma$  as defined in (14). From (18), we have

$$fR\mathbf{e}_2 = m_q J^{-1} \left( \ddot{\mathbf{v}} - \dot{J} J^{-1} \dot{\mathbf{v}} \right) + \mathbf{G}_{1:2}$$

$$\tag{21}$$

and defining

$$\mathbf{F}_{1:2} = m_q J^{-1} \left( \ddot{\mathbf{v}} - \dot{J} J^{-1} \dot{\mathbf{v}} \right) + \mathbf{G}_{1:2}, \tag{22}$$

we conclude that

$$f = \|\mathbf{F}_{1:2}\|, \quad \theta = \arctan\left(\frac{F_1}{F_2}\right).$$
 (23)

Now, recall the choice to use the virtual image coordinates and observe that solving for  $\theta$  in equation (23) would have been much more difficult (if possible at all) if the Jacobian, J, was dependent upon  $\theta$ . The derivative of (21) reveals that

$$\dot{f} = \mathbf{e}_2^T R^T \dot{\mathbf{F}}_{1:2} \tag{24}$$

and

$$\dot{\theta} = \frac{1}{f} \mathbf{e}_1^T R^T \dot{\mathbf{F}}_{1:2}.$$
(25)

The next derivative provides

$$\ddot{\theta} = \frac{1}{f} \left( \mathbf{e}_1^T R^T \ddot{\mathbf{F}}_{1:2} - 2\dot{f}\dot{\theta} \right)$$
(26)

and, using (19), the pitch moment is

$$M = J_q \frac{1}{f} \left( \mathbf{e}_1^T R^T \ddot{\mathbf{F}}_{1:2} - 2\dot{f}\dot{\theta} \right).$$
(27)

Upon inspection, we see that the 4<sup>th</sup> derivative of the image coordinates appears in (27) through the  $\ddot{\mathbf{F}}_{1:2}$  term, which means that trajectories in the image plane must be at least 4 times differentiable, or  $\mathcal{C}^4$ .

## 5.2. Trajectory Generation

The differential flatness analysis in the Euclidean space and further examination of the control inputs reveals that the snap (4<sup>th</sup> derivative) of the position of the quadrotor appears in the M term through  $\ddot{\theta}$ . In addition,  $\beta^{(4)}$  appears in M through the  $\mathbf{r}_s^{(4)}$  term in  $\ddot{\theta}$ . In the image plane case, the snap of the image coordinates appears in M.

Then, to minimize the norm of the input vector, it is appealing to minimize the following cost functional constructed from the snap of the trajectory:

$$\mathcal{J}_{i} = \int_{t_{0}}^{t_{f}} \left\| y_{i}^{(4)}(t) \right\|^{2} dt \quad \forall i$$

$$(28)$$

where  $y_i$  denotes the  $i^{th}$  flat output. Accordingly, we consider minimum-snap trajectories in both the Euclidean space as well as in the image space. The minimization problem can be solved by choosing a finite dimensional basis for the trajectories and numerically solving a quadratic program (QP) [35]. If only equality constraints are needed, the QP can be solved by a single matrix inversion, and in practice, even the inequality case can be solved fast enough for real-time integration. In our implementation, we precompute the trajectories and control the robot (using Vicon) to start at the appropriate starting point in the trajectory. The choice for this approach was motivated by ease-of-implementation and the fact that this allows the same trajectory to be flown numerous times.

The boundary conditions on the trajectories are the same as the observed boundary conditions of the trajectories of the raptors. In particular, we define a start and finish location, and we let the position at pickup be defined by the target's location. In the gripper-equipped case, the position at pickup is constrained such that the gripper is oriented vertically when grasping the target, but the velocity, acceleration, and jerk of the quadrotor are free and required to be continuous. In the vision-based case, we use pre-recorded measurements at a position in which the robot will capture the target to define the position constraints. Similarly, the higher-order derivatives at the pickup time are free.

See Figure 5a and Figure 5b for a desired and experimental trajectory of the position in the gripper-equipped case of the quadrotor and the gripper angle, respectively. See Figure 6b for the inertial-frame trajectories that result from planning in the image space.

Having shown that the system is differentially flat with two sets of flat outputs in the inertial and the image spaces respectively, and having used the differential flatness property to generate dynamically feasible trajectories, we now develop two controllers, one that uses the motion capture system for tracking in the inertial frame, and another that uses vision to track features in the image space.

#### 6. Control

#### 6.1. Control in The Inertial Frame

Now, we briefly present the controller that drives the quadrotor and gripper system along the desired trajectory. The quadrotor controller has an outer position control loop running at 100 Hz which generates desired attitudes and feedforward control inputs. The commanded thrust is

$$f = k_{pz} \left( z_{q,d} - z_q \right) + k_{dz} \left( \dot{z}_{q,d} - \dot{z}_q \right) + f_d \tag{29}$$

where  $k_{pz}$  and  $k_{dz}$  are proportional and derivative gains, respectively. The desired values of various variables, denoted with a subscript "d", are computed using the flatness property. A 1 kHz inner-loop attitude controller on-board the quadrotor is used to drive the robot to the desired attitude. The control moment is

$$M = k_{p\theta} \left(\theta_c - \theta\right) + k_{d\theta} \left(\dot{\theta}_d - \dot{\theta}\right) + M_d \tag{30}$$

where  $\dot{\theta}_d$  is the nominal angular rate,  $M_d$  is the feed-forward moment,  $k_{p\theta}$  is a proportional gain, and  $k_{d\theta}$  is a derivative gain. Finally,  $\theta_c$  is the command from the outer loop determined by

$$\theta_c = \sin^{-1} \left( k_{px} \left( x_{q,d} - x_q \right) + k_{dx} \left( \dot{x}_{q,d} - \dot{x}_q \right) \right) + \theta_d \tag{31}$$

where  $k_{px}$  is a proportional gain and  $k_{dx}$  is a derivative gain. The control design is similar to the quadrotor hover controller in [36], and the feedforward control input serves to compensate for the motion of the gripper. The state of the quadrotor is observed using Vicon [37] and feedforward control inputs are supplied to the control loops as displayed in Figure 3a. Further, we assume that the object being grasped is significantly lighter than the combined mass of the robot and gripper (Assumption A3), and therefore do not consider it in the control system.



Figure 3: (a) A block diagram of the controller used for experiments. A subscript "d" denotes a desired or nominal value (computed using the flatness property). (b) The gripper arm in motion from right to left as the claw is grasping. The shaded projections demonstrate the motion as the arm swings about the axis pointed into the page.

#### 6.2. Control in the Visual Space

6.2.1. Attitude Controller First, let  $R_d \in SO(2)$  denote the desired rotation matrix defined by a desired attitude,  $\theta_d$ , and recall that R is the rotation matrix defining the current attitude. The angular rate of the robot is  $\Omega$ , which, in the planar case, is equivalent to  $\dot{\theta}$ , and the desired angular rate is  $\Omega_d$ , or  $\dot{\theta}_d$ . Then, we define attitude errors

$$e_R = \frac{1}{2} \left( R_d^T R - R^T R_d \right)^{\vee} = \sin(\theta - \theta_d)$$
(32)

$$e_{\Omega} = \Omega - R^T R_d \Omega_d = \dot{\theta} - \dot{\theta}_d. \tag{33}$$

where  $\vee$  is the "vee" map as defined in [29]. These errors are similar to [29] but simplified for the planar case. Also, the configuration error function is defined as

$$\Psi(R, R_d) = \frac{1}{2} \operatorname{tr} \left[ I - R_d^T R \right].$$
(34)

The attitude controller is then given as below.

**Proposition 1.** [29, Prop. 1] (Exponential Stability of Attitude Controlled Flight Mode) Consider the control moment defined as

$$M = -K_R e_R - K_\Omega e_\Omega + J_q \ddot{\theta}_d, \tag{35}$$

where  $K_R$  and  $K_{\Omega}$  are positive scalars. Further, suppose the initial conditions satisfy

$$\Psi(R(0), R_d(0)) < 2 \tag{36}$$

$$\|e_{\Omega}(0)\|^{2} < \frac{2}{J_{q}}k_{R}\left(2 - \Psi(R(0), R_{d}(0))\right).$$
(37)

Then,  $(e_R, e_\Omega) = (0, 0)$  is exponentially stable for the closed-loop system.

*Proof.* Follows from [29, Prop. 1]. See Appendix A for more details.  $\Box$ 

6.2.2. Position Control Let errors in the image plane be defined by

$$\mathbf{e}_v = \mathbf{v} - \mathbf{v}_d \tag{38}$$

and we remind the reader that  $\mathbf{v}$  is a vector of the image feature coordinates. Similarly,  $\mathbf{v}_d$  is a vector of the desired image feature coordinates. Then, using (18), the image space error dynamics are

$$m_q \ddot{\mathbf{e}}_v = f J R \mathbf{e}_2 - J \mathbf{G}_{1:2} + m_q \dot{J} J^{-1} \dot{\mathbf{v}} - m_q \ddot{\mathbf{v}}_d.$$
<sup>(39)</sup>

where J is the image Jacobian and  $\mathbf{G}_{1:2}$  is the first two components of  $\mathbf{G}$ . The visual servoing controller is then given as below.

**Proposition 2.** (Exponential Stability of Visual Feature Controlled Flight Mode) Consider the total thrust component along the current body frame vertical axis defined by

$$f = \mathbf{A} \cdot R\mathbf{e}_2. \tag{40}$$

where

$$\mathbf{A} = \mathbf{G}_{1:2} + m_q J^{-1} \left[ -K_p \mathbf{e}_v - K_d \dot{\mathbf{e}}_v + \ddot{\mathbf{v}}_d \right], \tag{41}$$

 $K_p > 0, K_d > 0$ , and the commanded attitude is given by

$$R_c \mathbf{e}_2 = \frac{\mathbf{A}}{\|\mathbf{A}\|}.\tag{42}$$

Finally, if we meet the assumptions stated in Appendix B, then the zero equilibrium  $(\mathbf{e}_v, \dot{\mathbf{e}}_v, e_R, e_\Omega) = (\mathbf{0}, \mathbf{0}, 0, 0)$  is locally exponentially stable.

Proof. See Appendix B.

#### 7. Experimental Testbed

Having developed two sets of controllers for tracking dynamically feasible grasping trajectories, we now briefly describe the experimental testbed that will be used for experimentally validating our proposed methods of generating and tracking these trajectories.

# 7.1. An Avian Inspired Gripper

A capable gripper is critical for high-speed aerial manipulation. In particular, the gripper must enable payload capture at high relative velocities, reliably cage and secure payloads regardless of shape or size, and facilitate stable perching on arbitrary sites. As mentioned in section 1, review of slow-motion video footage revealed that a predatory bird swept its legs backwards during the capture phase of hunting, thereby reducing the relative velocity between its claws and its prey [16].

In order to ascribe the same functionalities to the quadrotor platform, we designed a gripper that consists of a 10.5 cm rotating arm cut from Acrylonitrile Butadiene Styrene (ABS). This linkage is analogous to the bird's leg and is intended to swing the gripping mechanism backwards during payload grasping to increase the time window of grasping. We attempt to reduce energy consumption, maintain agility, and be versatile by requiring the grasping mechanism to be lightweight and compliant with arbitrary shapes. Leveraging a design similar to [38], we were able to create an adaptive and underactuated gripper capable of grasping arbitrary shapes. Figure 3b provides a timelapse visualization of the servo motor rotating this arm.

## 7.2. Visual Feedback

The quadrotor is equipped with a global shutter Caspa<sup>TM</sup> VL camera and Computer on Module from Gumstix [39]. The automatic detection and tracking of the cylinder runs onboard the robot, is based on contour detection using Freeman chain coding, and is obtained using the C++ Visp library [40]. When the object is in the image and  $\mathbf{r}_q \in S$ , the measured image points from the camera are mapped to the virtual image plane using feedback from the IMU and the transformation shown in Figure 4a, which is mathematically equivalent to

$$v_i = \tan\left(\arctan\left(v_{i,m}\right) + \theta\right) \tag{43}$$

where  $v_{i,m}$  is the boundary of the cylinder as measured in the calibrated image.

The points in the virtual plane are filtered to improve the estimate of the image features and their derivatives to compute J and  $\dot{J}$ . A block diagram of the system is shown in Figure 4b. Since our visual controller is only designed for motion in the vertical plane, in experimentation, an external motion capture system is used as feedback to stabilize the yaw and out of plane motion. Note that the vision based controller stabilizes motion in the vertical plane as designed.



Figure 4: (a) The measured image feature points,  $v_{i,m}$ , which are affected by  $\theta$ , are mapped onto a virtual level image plane to decouple the motion from the attitude of the robot and determine the coordinates  $v_i$ . (b) A camera captures images of the cylinder, which are sent to the Gumstix Overo Computer on Module (COM) and processed at 65 Hz using blob tracking. The boundaries of the cylinder are undistorted, calibrated, and sent back to a ground station along with the pitch as measured from the IMU. Then, the ground station maps the points to the virtual plane and computes desired control inputs using the IBVS controller. Simultaneously, Vicon feedback is used to close the loop on the roll and yaw of the robot. Then, the desired attitude is sent to the onboard controller, which uses the IMU to control the attitude at 1 kHz.

Having briefly described the experimental platform that's being used, we will next present experimental results to validate our proposed methods of trajectory generation and tracking to achieve dynamic grasping.

# 8. Results

#### 8.1. High-speed grasping with control in the inertial space

We demonstrate experimental results on a 500 gm Asctec Hummingbird quadrotor [28] equipped with a 158 gm gripper. The experiments utilize the GRASP Multiple Micro UAV Testbed [41] and leverage a motion capture system to accurately determine the state of the quadrotor [37]. A 27 gm cylindrical target was tracked using Vicon [37]. Thus, compared to the combined mass of the vehicle and gripper, the mass of the target is quite small, which justifies Assumption A3.

The controller in the inertial space, which combines feedforward control inputs and a simple feedback controller on the quadrotor, was used to grasp the target while moving at 2 m/s with a success rate of 100% out of 5 attempts. Desired trajectories and the experimental results can be seen in Figure 5a and Figure 5b. Position errors for those trajectories are presented in Figure 5c and 5d. The quadrotor was able to successfully grasp the target at speeds up to 3 m/s, or 9 body lengths / second (Figure 7c). To see a video of sample experiments, see the supplementary video.



(a) Desired quadrotor position trajectories overlayed with experimental results.





(b) Desired  $\beta$  and  $\theta$  trajectories overlayed with experimental results.



(c) The x components of the normal error between the experimental and desired gripper positions for 5 consecutive trials.

(d) The z components of the normal error between the experimental and desired gripper positions for 5 consecutive trials.

Figure 5: Experimental results from planning, control, and grasping in the inertial frame. The pickup time is represented by a vertical dashed line.

# 8.2. Vision-Based Control

The stability of the proposed visual controller is demonstrated through several different experiments including hovering, vertical trajectories, "swooping" trajectories, and hovering above a moving cylinder. Here we present a sample "swooping" trajectory, which includes components from several of the previously mentioned trajectories. See Figure 6a for the planned and actual trajectories in the virtual image plane, Figure 6b for the corresponding estimated and actual position in the inertial frame, Figure 6c for a sequence of still images from a sample experiment, and the supplementary video for footage of sample trajectories.

# 8.3. Discussion

Avian-Robot Comparison: In assessing the success of our results, it is appropriate to use the eagle's performance as a standard of comparison. This is complicated by the fact that length and time scales can not be extracted from the video footage accurately. However, in order to facilitate a quantitative comparison between the trajectories, we



(a) Experimental results of the feature coordinates in the virtual plane for a "swooping" trajectory. The feature coordinates are denoted by  $v_i$  and the desired trajectory is given by  $v_{i,d}$ .

(b) Positions in the inertial frame for the experiment in Figure 6a. The vision estimates of the position (using  $\Gamma$ ) are denoted by the "v" subscript. The ground truth only has the "q" subscript.



(c) Still images from a sample "swooping" trajectory using the vision-based controller developed in this paper. Note: the background has been washed out to improve visibility of the robot.

Figure 6: Results from Vision Based Control Experiments

nondimensionalize the trajectories using the following relationships:

$$x^{\star} = \frac{x}{L}, \quad z^{\star} = \frac{z}{L}, \quad t^{\star} = \frac{tv_p}{L} \tag{44}$$

where  $v_p$  is the body velocity at pickup (pixels/frame or meters/second) and L is the length from the axis of rotation to the gripping surface of the claw (pixels or meters). Results using this approach are presented in Figures 7a and 7b. It can be seen that the horizontal position of the gripper closely matches that of the eagle's claw, while the vertical positions differ significantly, potentially due to the limited range of motion of the gripper arm of the quadrotor compared to that of the eagle.

Visual Controller: The results of the vision based control are shown in Figures 6a and 6b. In these, a "swooping" trajectory is executed with a variation of 1 m in the z direction and 50 cm in the x direction. The system is stable, and it is possible to notice that the swooping trajectory in the Cartesian space, as shown in Figure 6b, corresponds to a desired planned and executed trajectory in the image space Figure 6a. This is an





(a) A comparison of the nondimensionalized x positions of the quadrotor claw and the eagle claw.

(b) A comparison of the nondimensionalized z positions of the quadrotor and the eagle claw.



(c) A still image comparison between the eagle (extracted from [16]) and the quadrotor for a trajectory with the quadrotor moving at 3 m/s (9 body lengths / second) at pickup. See [42] for a video of the grasping.

Figure 7: Comparison with the Eagle Trajectory

experimental demonstration of the success of the proposed theoretical approach. We also notice that, in the Cartesian space, the error is quite small in the z direction, which presents a larger spatial change compared to x direction. Moreover, the z direction is the most challenging from a vision control point of view since the only source of information to recover the scale is the cylinder size.

*Limitations and Future Challenges:* A current limitation with the quadrotor-gripper system with control in the inertial frame is that the problem is formulated in the vertical plane. In our future work, we will attempt to generalize this to the full 3d problem by determining flat outputs for the coupled system.

It is also important to recognize that the experimental trajectories for the visionbased control (Figures 6a and 6b) are not as fast as the trajectories with control feedback in the inertial space, which demonstrated aggressive grasping at speeds up to 3 m/s. There are several reasons: the feedback is only from sensors onboard the robot (in contrast with an external motion capture system), the rate of feedback is nearly half in the vision-based case since we use a space, weight, and power constrained camera and computer, the position feedback loop is now closed using the onboard IMU, and the camera has a limited field of view. Although high speed visual control has been demonstrated earlier [43], it has not been achieved on space, weight, and computationally-constrained platforms. Thus, it is natural to expect trajectories that are not as aggressive. Our main goal is to show the feasibility of the proposed approach with a minimal sensor suite. Further, we observe that trajectory tracking is not perfect and attribute this to modeling errors such as distortion from the camera lens and external disturbances such as ground effect and the disturbed aerodynamics after the target is captured. Future work and the advancement of technology will help to reduce the limitations with the goal of eventually achieving similar performance to the experiments in a structured environment.

In the vision-based case, we are currently reliant upon a motion capture system to stabilize the lateral dynamics. This is mainly because the lateral velocity is not observable from the features selected. In future work, we will attempt to augment the feedback with optical flow for velocity estimates, and perhaps extend the feature points to be tangent lines (parallel to the axis of the cylinder), which would help provide an estimate of the roll of the robot.

We would also like to point out that the current vision approach requires the radius of the cylinder to be known a priori. In many cases, however, proper identification of the cylinder may lead to a good estimate of the size. For example, there are many common cylinders of similar or standard size such as railings and pipes. Additionally, once there is one successful grasp, the desired location of image features can be recorded to enable future grasping without needing to determine the size of the cylinder. Thus, we believe that this approach will not be difficult to generalize to grasping of unknown cylinders.

*Perching:* As researchers continue to develop quadrotors, the added ability to perchive will be critical in extending mission time. Unlike grasping and perching using fixed wing vehicles, the two tasks are very similar for quadrotors. The only difference for a quadrotor is that the planned trajectory would stop at the bottom of the swooping behavior in order to perch. Using the proposed trajectory methods and control schemes, this task would be a simple extension of the current work.

# 9. Conclusion

In this paper, we explored the challenges of high-speed aerial grasping using a quadrotor MAV. A novel appendage design, inspired by the articulation of a raptor's legs, was shown to enable a high rate of success while grasping objects at high velocities. The dynamic model of the quadrotor and gripper system was shown to be differentially flat in the inertial space, and dynamically feasible trajectories were generated for dynamic grasping. Experimental results were presented for quadrotor velocities of 2 m/s and 3 m/s (6 - 9 body lengths / second). A comparison of a nondimensionalized quadrotor trajectory with a sample avian trajectory reveals that the trajectories are similar although the curvature of the raptor's path is higher. We developed a non-

linear vision-based controller for trajectory tracking in the image space, which does not require an external motion capture system in the primary dimensions. We presented a proof of stability and provided validation of the controller through experimentation using a quadrotor equipped with a monocular camera system. In particular, we formulated the dynamics of the underactuated system directly in a virtual image plane and demonstrated that the system is differentially flat, with the image coordinates being the set of flat outputs. The proposed trajectory generation method in the image guarantees dynamic feasibility and enables incorporating visual constraints as linear constraints.

In summary, we have demonstrated a first step towards autonomous and dynamic grasping and manipulation for MAVs in unstructured environments. Future prototypes of the gripper will leverage better actuation, Shape Deposition Manufacturing (SDM) for lighter fingers, and a cam mechanism for actuating the arm, which will improve the performance and speed at which we can grasp objects. Finally, we will generalize the vision-based control law to the full three dimensions and incorporate a manipulator to enable truly autonomous dynamic grasping.

# Acknowledgments

The authors would like to thank Kartik Mohta for his help, insight, and patience during the development of the vision system used in our experiments.

# Appendix A

For the attitude controller, the Lyapunov candidate is

$$\mathcal{V}_R = \frac{J_q}{2} e_\Omega \cdot e_\Omega + K_R \Psi(R, R_d) + c_2 e_R \cdot e_\Omega, \tag{45}$$

with  $c_2$  being a positive scalar, such that,

$$\mathbf{z}_{\theta}^{T} M_{\theta} \mathbf{z}_{\theta} \le \mathcal{V}_{R} \le \mathbf{z}_{\theta}^{T} M_{\Theta} \mathbf{z}_{\theta}, \tag{46}$$

$$\dot{\mathcal{V}}_R \le -\mathbf{z}_{\theta}^T W_{\theta} \mathbf{z}_{\theta},\tag{47}$$

where  $\mathbf{z}_{\theta} = [\|e_R\|, \|e_{\Omega}\|]^T$ , and  $M_{\theta}, M_{\Theta}$ , and  $W_{\theta}$  are positive definite.

# Appendix B

# Stability of Translational Dynamics in the Image Coordinates

We take an approach very similar to [29] to show that the controller is exponentially stable. First, define  $K'_p, K'_d, B, \alpha \in \mathbb{R}$  as,

$$K'_{p} = m_{q} \|J\| \|J^{-1}\| K_{p}$$
(48)

Toward Autonomous Avian-Inspired Grasping for Micro Aerial Vehicles

$$K'_{d} = \|J\| \left( m_{q} \|J^{-1}\| K_{d} + \|J^{-1}\| \right)$$
(49)

$$B = \|J\| \left( \|G_A\| + m_q \|J^{-1}\| \|\ddot{\mathbf{v}}^d\| + \|J^{-1}\| \|\dot{\mathbf{v}}^d\| \right)$$
(50)

$$\alpha = \|e_R\| \tag{51}$$

and define  $W_{v_1}, W_{v_2}, W_{v\theta}, W_v \in \mathbb{R}^{2 \times 2}$  as

$$W_{v_1} = \begin{bmatrix} \frac{c_1 K_p}{m_q} & \frac{c_1 K_d}{2m_q} \\ \frac{c_1 K_d}{2m_q} & K_d - c_1 \end{bmatrix}, W_{v\theta} = \begin{bmatrix} \frac{c_1}{m_q} B & 0 \\ B & 0 \end{bmatrix}$$
(52)

$$W_{v_2} = \begin{bmatrix} \frac{c_1 \alpha K'_p}{m_q} & \frac{\alpha}{2} \left( \frac{c_1}{m_q} K'_d + K'_p \right) \\ \frac{\alpha}{2} \left( \frac{c_1}{m_q} K'_d + K'_p \right) & \alpha K'_d \end{bmatrix}$$
(53)

$$W_v = W_{v_1} - W_{v_2}.$$
 (54)

Suppose we choose positive constants  $c_1, K_p, K_d, K_R, K_\Omega$  such that,

$$K_p > \frac{c_1^2}{m_q} \tag{55}$$

$$\lambda_{\min}(W_{\theta}) > \frac{4 \left\| W_{v\theta} \right\|^2}{\lambda_{\min}(W_v)} \tag{56}$$

Then, there exists positive constants  $\gamma_1, \gamma_2, \gamma_3$ , such that  $||J|| \leq \gamma_1, ||J^{-1}|| \leq \gamma_2,$  $||\dot{J}^{-1}|| \leq \gamma_3$ , and if initial conditions and the desired trajectory satisfy

$$\alpha < \frac{1}{m_q \gamma_1 \gamma_2},\tag{57}$$

$$dist(\mathbf{v}_d(t), V^c) < \left\| \mathbf{e}_v(0) \right\|, \tag{58}$$

where  $V^c$  is the complement of V, and  $dist(\mathbf{v}_d(t), V^c) = \inf_{t \in [0,\infty), w \in V^c} \|\mathbf{v}_d(t) - w\|$ is the smallest distance between a trajectory and a set, then the zero equilibrium  $(\mathbf{e}_v, \dot{\mathbf{e}}_v, e_R, e_\Omega) = (\mathbf{0}, \mathbf{0}, 0, 0)$  is locally exponentially stable.

*Proof.* Using (18), we can determine the image errors

$$\ddot{\mathbf{e}}_{v} = \ddot{\mathbf{v}} - \ddot{\mathbf{v}}_{d} = \frac{1}{m} J \left[ f R \mathbf{e}_{2} - \mathbf{G}_{A} \right] + \dot{J} J^{-1} \dot{\mathbf{v}} - \ddot{\mathbf{v}}_{d}$$
(59)

so that

$$m\ddot{\mathbf{e}}_v = fJR\mathbf{e}_2 - J\mathbf{G}_A + m\dot{J}J^{-1}\dot{\mathbf{v}} - m\ddot{\mathbf{v}}_d.$$
(60)

Defining

$$\mathbf{X} = J \frac{f}{\mathbf{e}_2^T R_c^T R \mathbf{e}_2} \left( \left( \mathbf{e}_2^T R_c^T R \mathbf{e}_2 \right) R \mathbf{e}_2 - R_c \mathbf{e}_2 \right),$$
(61)

the error dynamics become

$$m\ddot{\mathbf{e}}_{v} = J\left(\frac{f}{\mathbf{e}_{2}^{T}R_{c}^{T}R\mathbf{e}_{2}}R_{c}\mathbf{e}_{2}\right) + \mathbf{X} - J\mathbf{G}_{A} + m\dot{J}J^{-1}\dot{\mathbf{v}} - m\ddot{\mathbf{v}}_{d}.$$
(62)

22

Next, let

$$f = \mathbf{A} \cdot R\mathbf{e}_2 \tag{63}$$

and the commanded attitude be defined by

$$R_c \mathbf{e}_2 = \frac{\mathbf{A}}{\|\mathbf{A}\|}.\tag{64}$$

Then, from the previous two equations, we have

$$f = \|\mathbf{A}\| \mathbf{e}_2^T R_c^T R \mathbf{e}_2.$$
(65)

Substituting this into (62) and using A, we have

$$m\ddot{\mathbf{e}}_{v} = J\left(\frac{\|\mathbf{A}\|\,\mathbf{e}_{2}^{T}R_{c}^{T}R\mathbf{e}_{2}}{\mathbf{e}_{2}^{T}R_{c}^{T}R\mathbf{e}_{2}}R_{c}\mathbf{e}_{2}\right) + \mathbf{X} - J\mathbf{G}_{A} + m\dot{J}J^{-1}\dot{\mathbf{v}} - m\ddot{\mathbf{v}}_{d}$$
(66)

$$= J\left(\left\|\mathbf{A}\right\| R_c \mathbf{e}_2\right) + \mathbf{X} - J\mathbf{G}_A + m\dot{J}J^{-1}\dot{\mathbf{v}} - m\ddot{\mathbf{v}}_d$$
(67)

$$= J\mathbf{A} + \mathbf{X} - J\mathbf{G}_A + m\dot{J}J^{-1}\dot{\mathbf{v}} - m\ddot{\mathbf{v}}_d$$
(68)

$$= -K_p \mathbf{e}_v - K_d \dot{\mathbf{e}}_v + \mathbf{X} \tag{69}$$

which has the same form as (83) in [29]. We use the same Lyapunov candidate, but in our image coordinates,

$$\mathcal{V}_{v} = \frac{1}{2} K_{p} \left\| \mathbf{e}_{v} \right\|^{2} + \frac{1}{2} m \left\| \dot{\mathbf{e}}_{v} \right\|^{2} + c_{1} \mathbf{e}_{v} \cdot \dot{\mathbf{e}}_{v}.$$
(70)

Now, let  $\mathbf{z}_{v} = \begin{bmatrix} \|\mathbf{e}_{v}\|, \|\dot{\mathbf{e}}_{v}\| \end{bmatrix}^{T}$ , then it follows that the Lyapunov function  $\mathcal{V}_{v}$  is bounded as

$$\mathbf{z}_{v}^{T} M_{v} \mathbf{z}_{v} \leq \mathcal{V}_{v} \leq \mathbf{z}_{v}^{T} M_{V} \mathbf{z}_{v}, \tag{71}$$

where  $M_v, M_V \in \mathbb{R}^{2 \times 2}$  are defined as,

$$M_{v} = \frac{1}{2} \begin{bmatrix} K_{p} & -c_{1} \\ -c_{1} & m \end{bmatrix}, \quad M_{V} = \frac{1}{2} \begin{bmatrix} K_{p} & c_{1} \\ c_{1} & m \end{bmatrix}.$$
 (72)

Then,

$$\dot{\mathcal{V}}_{v} = K_{p} \left( \dot{\mathbf{e}}_{v} \cdot \mathbf{e}_{v} \right) + m \left( \ddot{\mathbf{e}}_{v} \cdot \dot{\mathbf{e}}_{v} \right) + c_{1} \left( \mathbf{e}_{v} \cdot \ddot{\mathbf{e}}_{v} + \dot{\mathbf{e}}_{v} \cdot \dot{\mathbf{e}}_{v} \right),$$
(73)

and incorporating (69),

$$\dot{\mathcal{V}}_{v} = -\frac{c_{1}K_{p}}{m} \left\| \mathbf{e}_{v} \right\|^{2} - (K_{d} - c_{1}) \left\| \dot{\mathbf{e}}_{v} \right\|^{2} -c_{1}\frac{K_{d}}{m} \left( \mathbf{e}_{v} \cdot \dot{\mathbf{e}}_{v} \right) + \mathbf{X} \cdot \left( \frac{c_{1}}{m} \mathbf{e}_{v} + \dot{\mathbf{e}}_{v} \right).$$
(74)

23

Toward Autonomous Avian-Inspired Grasping for Micro Aerial Vehicles

Now, we establish a bound on  $\mathbf{X}$ . From (61),

$$\mathbf{X} = J \frac{f}{\mathbf{e}_2^T R_c^T R \mathbf{e}_2} \left( \left( \mathbf{e}_2^T R_c^T R \mathbf{e}_2 \right) R \mathbf{e}_2 - R_c \mathbf{e}_2 \right)$$
(75)

$$\|\mathbf{X}\| \le \|J\| \left\| \frac{\|A\| R_c \mathbf{e}_2 \cdot R \mathbf{e}_2}{R_c \mathbf{e}_2 \cdot R \mathbf{e}_2} \right\| \|e_R\|$$
(76)

$$\leq \|J\| \|\mathbf{A}\| \|e_R\| \tag{77}$$

$$\leq \|J\| \left\| \mathbf{G}_A + mJ^{-1} \left[ -K_p \mathbf{e}_v - K_d \dot{\mathbf{e}}_v + \ddot{\mathbf{v}}_d \right] + J^{-1} \left[ \dot{\mathbf{e}}_v + \dot{\mathbf{v}}_d \right] \right\| \|e_R\|$$
(78)

$$\leq \left(K_p' \left\| \mathbf{e}_v \right\| + K_d' \left\| \dot{\mathbf{e}}_v \right\| + B\right) \left\| e_R \right\| \tag{79}$$

where  $K'_{p}, K'_{d}, B$  are as defined in (48)-(50), and from [29],  $0 \le ||e_{R}|| \le 1$ .

Next we will show that there exists positive constants  $\gamma_1, \gamma_2, \gamma_3$  s.t.,  $||J|| \leq \gamma_1, ||J^{-1}|| \leq \gamma_2$ , and  $||J^{-1}|| \leq \gamma_3$ . Since  $\Gamma$  is smooth (we only require  $C^2$  here), J is smooth on the closed set S. This implies J is bounded on S, i.e.,  $\exists \gamma_1 > 0$ , s.t.  $||J|| < \gamma_1$ . Next, since J is smooth and nonsingular on S, the inverse is well defined and is smooth on S, which implies  $J^{-1}$  is bounded on S, i.e.,  $\exists \gamma_2 > 0$ , s.t.  $||J^{-1}|| < \gamma_2$ . Next, observe that  $\frac{d}{dt}J^{-1}(\mathbf{r}_q) = \frac{\partial}{\partial \mathbf{r}_q}J^{-1}(\mathbf{r}_q)\dot{\mathbf{r}}_q$  is a composition of smooth functions on S, implying that it is bounded on S, i.e.,  $\exists \gamma_3 > 0$ , s.t.  $||J^{-1}|| < \gamma_3$ .

Then, similar to [5], we can express  $\dot{\mathcal{V}}_v$  as

$$\dot{\mathcal{V}}_{v} = -\begin{bmatrix} \mathbf{e}_{v}^{T} & \dot{\mathbf{e}}_{v}^{T} \end{bmatrix} W_{v_{1}} \begin{bmatrix} \mathbf{e}_{v} \\ \dot{\mathbf{e}}_{v} \end{bmatrix} + \mathbf{X} \cdot \left(\frac{c_{1}}{m} \mathbf{e}_{v} + \dot{\mathbf{e}}_{v}\right)$$

$$\leq -\begin{bmatrix} \mathbf{e}_{v}^{T} & \dot{\mathbf{e}}_{v}^{T} \end{bmatrix} W_{v_{1}} \begin{bmatrix} \mathbf{e}_{v} \\ \dot{\mathbf{e}}_{v} \end{bmatrix}$$

$$+ K'_{p} \|\mathbf{e}_{v}\| \|e_{R}\| \left(\frac{c_{1}}{m} \|\mathbf{e}_{v}\| + \|\dot{\mathbf{e}}_{v}\|\right)$$

$$+ K'_{d} \|\dot{\mathbf{e}}_{v}\| \|e_{R}\| \left(\frac{c_{1}}{m} \|\mathbf{e}_{v}\| + \|\dot{\mathbf{e}}_{v}\|\right)$$

$$+ B \|e_{R}\| \left(\frac{c_{1}}{m} \|\mathbf{e}_{v}\| + \|\dot{\mathbf{e}}_{v}\|\right).$$

$$(80)$$

$$(81)$$

This can be written as,

$$\dot{\mathcal{V}}_{v} \leq -\mathbf{z}_{v}^{T} W_{v} \mathbf{z}_{v} + \mathbf{z}_{v}^{T} W_{v\theta} \mathbf{z}_{\theta}$$

$$\tag{82}$$

where  $W_{v\theta}, W_v$  are as defined in (52), (54). Since  $W_v = (W_v)^T$  and  $W_v \in \mathbb{R}^{2\times 2}$ , it is sufficient to show that  $\det(W_v) > 0$  and  $W_v(1,1) > 0$  in order to claim that  $W_v > 0$ . Then, from the assumption on  $\alpha$  in (57), we have  $w_{11} > 0$ . This is reasonable since  $\alpha$  is a functional on the attitude error such that  $\alpha \in [0,1]$ . Thus, the assumption in (57) is simply a bound on the attitude error. The determinant can be expressed as a quadratic function of  $K_d$  such that

$$\det(W_v) = \beta_0 + \beta_1 K_d + \beta_2 K_d^2 \tag{83}$$

and  $\beta_i$  is a function of  $c_1$ ,  $K_p$ ,  $\gamma_1$ ,  $\gamma_2$ ,  $\gamma_3$ , and m. The critical point of the quadratic occurs when

$$K_d = \frac{K_p m}{c_1} + \frac{K_p m + \alpha c_1 \gamma_1 \gamma_3}{c_1 \left(1 - \alpha \gamma_1 \gamma_2 m\right)}$$
(84)

24

Toward Autonomous Avian-Inspired Grasping for Micro Aerial Vehicles

and has a value of

$$\det(W_v) = \frac{K_p \left(1 - \alpha \gamma_1 \gamma_2 m\right) \left(K_p m - c_1^2\right)}{m}.$$
(85)

In both equations,  $(1 - \alpha \gamma_1 \gamma_2 m) > 0$  as a result of the assumption in (57). Thus (84) is positive, and by (55), (85) is positive and  $W'_v > 0$ . Now, we consider the combined Lyapunov candidate for the translational and rotational error dynamics,  $\mathcal{V} = \mathcal{V}_v + \mathcal{V}_R$ . From (46) and (71), we have,

$$\mathbf{z}_{v}^{T} M_{v} \mathbf{z}_{v} + \mathbf{z}_{\theta}^{T} M_{\theta} \mathbf{z}_{\theta} \leq \mathcal{V} \leq \mathbf{z}_{\theta}^{T} M_{\Theta} \mathbf{z}_{\theta} + \mathbf{z}_{v}^{T} M_{V} \mathbf{z}_{v}.$$
(86)

Further, we see that

$$\dot{\mathcal{V}} \leq -\mathbf{z}_{v}^{T}W_{v}\mathbf{z}_{v} + \mathbf{z}_{v}^{T}W_{v\theta}\mathbf{z}_{\theta} - \mathbf{z}_{\theta}^{T}W_{\theta}\mathbf{z}_{\theta},$$
(87)

$$\leq -\lambda_{min}(W_v) \|\mathbf{z}_v\|^2 + \|W_{v\theta}\| \|\mathbf{z}_v\| \|\mathbf{z}_{\theta}\| -\lambda_{min}(W_{\theta}) \|\mathbf{z}_{\theta}\|^2,$$
(88)

and from (56), we have  $\dot{\mathcal{V}}$  to be negative definite, and the zero equilibrium of the closed-loop system is locally exponentially stable.

# References

- Justin Thomas, Joe Polin, Koushil Sreenath, and Vijay Kumar. Avian-Inspired Grasping for Quadrotor Micro UAVs. In *IDETC/CIE 2013*, Portland, 2013.
- [2] Justin Thomas, Giuseppe Loianno, Koushil Sreenath, and Vijay Kumar. Toward Image Based Visual Servoing for Aerial Grasping and Perching. In *submitted to International Conference on Robotics and Automation*, 2014.
- [3] Daniel Mellinger, Michael Shomin, and Vijay Kumar. Control of Quadrotors for Robust Perching and Landing. In *International Powered Lift Conference*, Philadelphia, PA, 2010.
- [4] Quentin Lindsey, Daniel Mellinger, and Vijay Kumar. Construction of Cubic Structures with Quadrotor Teams. In *Robotics: Science and Systems*, Los Angeles, CA, USA, 2011.
- [5] Koushil Sreenath, Taeyoung Lee, and Vijay Kumar. Geometric control and differential flatness of a quadrotor UAV with a cable-suspended load. In *IEEE Conference on Decision and Control* (CDC), pages 2269–2274, Florence, Italy, December 2013.
- [6] Daniel Mellinger, Michael Shomin, Nathan Michael, and Vijay Kumar. Cooperative grasping and transport using multiple quadrotors. In International Symposium on Distributed Autonomous Robotic Systems, 2010.
- [7] Koushil Sreenath and Vijay Kumar. Dynamics, Control and Planning for Cooperative Manipulation of Payloads Suspended by Cables from Multiple Quadrotor Robots. In Proceedings of Robotics: Science and Systems, Berlin, Germany, 2013.
- [8] Caitlin Powers, Daniel Mellinger, Alex Kushleyev, Bruce Kothmann, and Vijay Kumar. Influence of Aerodynamics and Proximity Effects in Quadrotor Flight. In Proceedings of the International Symposium on Experimental Robotics, 2012.
- [9] A. M. Dollar and R. D. Howe. The Highly Adaptive SDM Hand: Design and Performance Evaluation. The International Journal of Robotics Research, 29(5):585–597, February 2010.
- [10] Courtney E Doyle, Justin J Bird, Taylor A Isom, C Jerald Johnson, Jason C Kallman, Jason A Simpson, Raymond J King, Jake J Abbott, and Mark A Minor. Avian-inspired passive perching mechanism for robotic rotorcraft. In 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 4975–4980. IEEE, September 2011.

- [11] Daniel Mellinger, Quentin Lindsey, Michael Shomin, and Vijay Kumar. Design, modeling, estimation and control for aerial grasping and manipulation. In 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 2668–2673. IEEE, September 2011.
- [12] Alexis Lussier Desbiens and Mark R. Cutkosky. Landing and Perching on Vertical Surfaces with Microspines for Small Unmanned Air Vehicles. Journal of Intelligent and Robotic Systems, 57(1-4):313–327, October 2009.
- [13] A. Lussier Desbiens, A. T. Asbeck, and M. R. Cutkosky. Landing, perching and taking off from vertical surfaces. *The International Journal of Robotics Research*, 30(3):355–370, January 2011.
- [14] Joseph Moore and Russ Tedrake. Control synthesis and verification for a perching UAV using LQR-Trees. In 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), pages 3707– 3714. IEEE, December 2012.
- [15] The Slow Mo Guys. "Red Kites in Slow Motion", http://youtu.be/AYOx-iCMZhk.
- [16] Karen Bass, Brian Leith, Justin Anderson, Peter Bassett, Joe Stevens, Hugh Pearson, and Jeff Turner. Nature's Most Amazing Events. [DVD]. BBC Worldwide Ltd. Programs, 2009.
- [17] Anastasios I Mourikis, Stergios I Roumeliotis, and Joel W Burdick. SC-KF Mobile Robot Localization: A Stochastic Cloning Kalman Filter for Processing Relative-State Measurements. *IEEE Transactions on Robotics*, 23(4):717–730, August 2007.
- [18] Stephan Weiss, Davide Scaramuzza, and Roland Siegwart. Monocular-SLAM-based navigation for autonomous micro helicopters in GPS-denied environments. *Journal of Field Robotics*, 28(6):854–874, November 2011.
- [19] S. Hutchinson, G.D. Hager, and P.I. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–670, 1996.
- [20] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, June 1992.
- [21] Francois Chaumette and Seth Hutchinson. Visual servo control. I. Basic approaches. IEEE Robotics & Automation Magazine, 13(4):82–90, December 2006.
- [22] F. Chaumette and S. Hutchinson. Visual servo control. II. Advanced approaches [Tutorial]. IEEE Robotics & Automation Magazine, 14(1):109–118, March 2007.
- [23] C.J. Taylor and J.P. Ostrowski. Robust vision-based pose control. In Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065), volume 3, pages 2734–2740. IEEE, 2000.
- [24] N.J. Cowan, J.D. Weingarten, and D.E. Koditschek. Visual servoing via navigation functions. IEEE Transactions on Robotics and Automation, 18(4):521–533, August 2002.
- [25] Tarek Hamel and Robert Mahony. Visual Servoing of an Under-Actuated Dynamic Rigid-Body System: An Image-Based Approach. *IEEE Transactions on Robotics & Automation*, 18(2):187– 198, 2002.
- [26] Tarek Hamel and Robert Mahony. Image based visual servo control for a class of aerial robotic systems. Automatica, 43(11):1975–1983, November 2007.
- [27] Daewon Lee, Tyler Ryan, and H. Jin. Kim. Autonomous landing of a VTOL UAV on a moving platform using image-based visual servoing. In 2012 IEEE International Conference on Robotics and Automation, pages 971–976. IEEE, May 2012.
- [28] "Ascending Technologies, GmbH," http://www.asctec.de.
- [29] Taeyoung Lee, Melvin Leok, and N.H. McClamroch. Control of Complex Maneuvers for a Quadrotor UAV using Geometric Methods on SE(3). Asian Journal of Control, 2011.
- [30] Koushil Sreenath, Nathan Michael, and Vijay Kumar. Trajectory Generation and Control of a Quadrotor with a Cable-Suspended Load A Differentially-Flat Hybrid System. In International Conference on Robotics and Automation, 2013.
- [31] Yi Ma, Stefano Soatto, Jana Kosecka, and S. Shankar Sastry. An Invitation to 3-D Vision: From Images to Geometric Models. Interdisciplinary Applied Mathematics. Springer, New York, 2004.
- [32] Hamed Jabbari, Giuseppe Oriolo, and Hossein Bolandi. Dynamic IBVS control of an underactuated UAV. 2012 IEEE International Conference on Robotics and Biomimetics

(*ROBIO*), pages 1158–1163, December 2012.

- [33] Hong Zhang and J.P. Ostrowski. Visual Servoing with Dynamics: Control of an Unmanned Blimp. In International Conference on Robotics and Automation, volume 1, pages 618–623. IEEE, 1999.
- [34] R.M. Murray, M. Rathinam, and Willem Sluis. Differential flatness of mechanical control systems: A catalog of prototype systems. In ASME International Congress and Exposition, 1995.
- [35] Daniel Mellinger and Vijay Kumar. Minimum snap trajectory generation and control for quadrotors. In 2011 IEEE International Conference on Robotics and Automation, pages 2520– 2525. IEEE, May 2011.
- [36] Daniel Mellinger, Nathan Michael, and Vijay Kumar. Trajectory Generation and Control for Precise Aggressive Maneuvers with Quadrotors. In International Symposium on Experimental Robotics, 2010.
- [37] "Vicon Motion Systems, Inc," http://www.vicon.com.
- [38] A. M. Dollar and R. D. Howe. Joint coupling design of underactuated hands for unstructured environments. The International Journal of Robotics Research, 30(9):1157–1169, June 2011.
- [39] "Gumstix, Inc.," http://www.gumstix.com.
- [40] Andrew I. Comport, Eric Marchand, Muriel Pressigout, and François Chaumette. Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *IEEE transactions on visualization and computer graphics*, 12(4):615–28.
- [41] Nathan Michael, Daniel Mellinger, Quentin Lindsey, and Vijay Kumar. The GRASP multiple micro-UAV testbed. Robotics & Automation Magazine, IEEE, 17(3):56-65, 2010.
- [42] Justin Thomas, Joe Polin, Koushil Sreenath, and Vijay Kumar. February 7, 2013, Avian-Inspired Grasping for Quadrotor Micro UAVs, from https://www.youtube.com/watch?v=ol8c9bdp7YI, 2013.
- [43] D. Shiokata, A. Namiki, and M. Ishikawa. Robot dribbling using a high-speed multifingered hand and a high-speed vision system. In 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 2097–2102. IEEE, 2005.