

Vision based control of aerial robotic vehicles using the port Hamiltonian framework

Robert Mahony, Stefano Stramigioli, and Jochen Trumpf

Abstract—This paper investigates the formulation of sensor based control for aerial robotic vehicles based on the port Hamiltonian framework. The paper considers the particular case of vision based control and develops a model for an “infinite dimensional” *visual energy port* that uses optical flow in the image plane as a ‘velocity’ in the port Hamiltonian formalism.

I. INTRODUCTION

Small semi-autonomous aerial robots are becoming indispensable in a wide range of security, commercial and environmental monitoring tasks. Due to their low cost and weight, low power requirements, and rich information content, vision sensors are an attractive sensor modality for control of aerial robotic vehicles. The classical approach to directly control a robot using vision is termed image based visual servo (IBVS) control and uses features such as points, or lines, extracted from the image sequence [1] to generate a control signal that servos the robot such that the observed image matches a previously saved image. Image based control has the advantage that it is highly robust to errors in target geometry and noise models [2]. However, in common with general sensor based control [3], classical IBVS control algorithms were developed primarily in the context of robotic manipulators [2] and only the system kinematics are controlled in the image domain. Extending the classical approach to force controlled dynamic systems with no direct velocity measurement, such as aerial robotic vehicles, is difficult and there are very few results available in the literature [4], [5], [6], [7], [8], [9], [10], [11], [12]. In a separate research community, there is a considerable body of work published that uses optic or visual flow as a velocity cue for vehicle control. A strong motivation for this approach is drawn from the study of vision and flight in animals and insects [13], [14]. Visual flow in the image space is coupled to time-to-collision computations [15] and can be used for obstacle avoidance [16], [17], docking manoeuvres [18], and visual odometry [19]. Recent work has demonstrated terrain tracking [20], [21], obstacle avoidance [22], [23], and landing [24], [25] for aerial robots. The control design for these systems, however, often relies on dynamic reduction or high-gain cancellation of the second order system dynamics to regain an effectively first order

kinematic system. The formalism of port Hamiltonian systems offers a natural framework to deal with fully dynamic second order systems and interconnection of multiple control algorithms with competing objectives [26], however, to the authors’ knowledge there is no prior work on developing a port Hamiltonian framework for image based visual control.

In this paper, we present initial results on the development of a Hamiltonian energy port interpretation for control of dynamic vehicles based on optical flow. The port Hamiltonian approach has a number of clear potential advantages over vision based control algorithms that have been considered in the prior literature. The framework deals inherently with the vehicle dynamics and also with a dynamically changing environment. It provides a structure in which multiple control modalities can be *plugged* together to provide multi-modal control for a vehicle, ensuring robustness, fault tolerance and improving performance of the overall control. Finally, the port Hamiltonian framework has advantageous passivity characteristics that ensure highly robust and stable control, and simplifies the handling of transmission or communication delays [27]. The key idea in our approach is to treat optical or visual flow as the flow variable in an infinite dimensional energy port, the image plane of the camera. We use the theory of infinite dimensional port Hamiltonian systems that has received considerable attention in the systems and control community recently [28, Chap. 4]. Based on this formalism, we propose a dual *visual effort* to the visual flow, an infinite dimensional cotangent valued 2-form on the image plane, along with a bracket operation that enables the two variables to be contracted to compute power via an integral over the image surface. The mapping from velocity screw on the vehicle to visual flow on the image surface is modelled as a modulated transformer (MTF), with forward relationship given by the classical optical flow equations. The dual (adjoint) map takes visual efforts back to wrenches on the vehicle body and is uniquely defined by the choices made in the modelling of visual effort and the power bracket. Control of the vehicle through the visual energy port is then achieved by connecting the visual port to external driving forces and damping elements. We derive the form for a dissipative or damping element as a constitutive relation, or *visual impedance* [29], that maps visual flow to a visual effort such that the power dissipated in the port is always positive. We also show how an exogenous input (remote pilot) can be injected directly into the visual port input as a visual effort. This approach has significant advantages over allowing the pilot to control a remote vehicle directly, as it regulates vehicle motion relative to the observed environment

R. Mahony and J. Trumpf are with School of Engineering, Australian National University, ACT, 0200, Australia. {firstname}.lastname@anu.edu.au

S. Stramigioli is with Department of Electrical Engineering, University of Twente, 7500 AE Enschede, The Netherlands. S.Stramigioli@utwente.nl

rather than the absolute vehicle motion. The approach taken initially implicitly depends on knowledge of the scalar depth of the environment around the robot to implement the visual impedance and exogenous forcing control, information that is not normally available from visual flow measurements and, due to ambiguity of scale, cannot be fully reconstructed from observed flow. We show that with the initial approach we propose it is impossible to implement a simple static dissipative element that leaves the vehicle wrench independent of scene depth. To overcome this difficulty we propose an alternative model where the visual flow is split based on a coordinate decomposition of the vehicle screw into translational and rotational components. This allows us to implement dissipative impedances and forcing control that can be realized as a known wrench on the vehicle without knowledge of the scene depth. This paper should be viewed as an initial investigation of what we believe to be a highly fruitful direction of research and the results presented are of a preliminary nature.

After this introduction, Section II introduces the dynamic system models and proposes the concept of a visual energy port. Section III discusses how to define a proper modulated transformer which couples the finite dimensional mechanical port of 6D motion to the infinite dimensional one of the visual flow. Section IV discusses the structure of dissipative elements and exogenous forcing control and discusses issues with these elements. Section V considers an alternative model that allows us to realise wrenches on the vehicle even without knowledge of the scene depth.

II. VISION ENERGY PORT

In this work we consider a camera connected to a rigid body in motion. For ease of reference we consider the origin of the frame of reference for the rigid body $\{B\}$ to be at the focal point of the camera. The world reference frame $\{A\}$ is assumed to be inertial.

Although most cameras use a perspective projection to obtain an image from a flat CCD array, it is most convenient to use a spherical image representation for the study of dynamic systems [9]. In practice, wide angle lenses can be used together with an image warping algorithm that transforms the flat image array produced by a standard camera onto a spherical representation. For ease of analysis in the present paper, we will ignore all practical aspects of the physical camera and consider the imaging system as generating a time-varying positive scalar mapping (grey-scale image) defined on the sphere. That is the image obtained at time t is

$$I_t : S^2 \rightarrow \mathbb{R}_+$$

where $S^2 = \{\eta \in \mathbb{R}^3 \mid \|\eta\| = 1\}$ is the 2-sphere in \mathbb{R}^3 and \mathbb{R}_+ are the non-negative real numbers including zero and represent greyscale intensity. The optical focus of the imaging system is the origin of the sphere, located at the origin of $\{B\}$.

A key measurement that can be derived from a visual image sequence of a textured world but in the absence of a geometric model of the world is the visual flow or motion

of the image across the image plane. In fact, what we will use in the sequel is the projected motion field, that is the actual motion of the world scene projected onto the image plane. This can be obtained from a sequence of images using any of the common optic flow algorithms [30] that overcome the aperture problem. Although these algorithms are derived for planar images, it is straightforward to map them onto the spherical image surface using image warping techniques. Again we will ignore all practical aspects of the physical camera and assume that there is sufficient texture in the image at all points to compute a dense visual flow vector field. Under these assumptions the visual flow field is a time-varying mapping

$$\Phi_t : S^2 \rightarrow TS^2, \Phi_t(\eta) \in T_\eta S^2.$$

That is the visual flow at time t can be seen as a section of the tangent bundle of the sphere $\Phi_t \in \Gamma(TS^2)$. Where the context is clear we will drop the subscript denoting time and simply write Φ for Φ_t .

Let $T \in \mathfrak{se}(3)$ denote the velocity screw of $\{B\}$ with respect to $\{A\}$. That is

$$T = \begin{pmatrix} \omega \\ v \end{pmatrix}$$

where ω is the angular, and v is the linear, velocity of $\{B\}$ with respect to $\{A\}$ respectively, both expressed in the body-fixed-frame $\{B\}$. Assuming that the world surrounding the camera is stationary, the velocity T will result in an observed visual flow Φ on the image plane of the camera. The expression for visual flow is classical in the vision literature [30] and a spherical coordinates representation is given in, for example, Hamel *et al.* [9]. In this paper, we formalise visual flow as a modulated transformer that maps the vehicle velocity screw T to visual flow

$$n : \mathfrak{se}(3) \rightarrow \Gamma(TS^2), T \mapsto \Phi. \quad (1)$$

Visual flow is a 2-dimensional vector field and would normally be represented in two-dimensional local coordinates. It is convenient in the following derivation, however, to work with a representation of S^2 as the unit sphere embedded in \mathbb{R}^3 . The tangent space $T_\eta S^2 = \{X \in \mathbb{R}^3 \mid X^T \eta = 0\}$ is then a two dimensional subspace of \mathbb{R}^3 . Define

$$A(\eta) := \left(\eta_\times \quad \frac{\eta \eta^T - I}{\lambda(\eta)} \right) = \left(\eta_\times \quad -\frac{1}{\lambda(\eta)} \mathbb{P}_\eta \right) \quad (2)$$

where η_\times denotes the skew symmetric matrix associated with the vector cross product such that $\eta_\times w = \eta \times w$ for all $w \in \mathbb{R}^3$, $\lambda(\eta)$ denotes the *scene depth*, the distance from the focal point of the camera to the point in the world that projects onto η , and \mathbb{P}_η is the projector on the subspace of \mathbb{R}^3 orthogonal to η . One has

$$\Phi = n(T) = A(\cdot)T \in \Gamma(TS^2). \quad (3)$$

At an instant of time we would like to calculate a “visual power” as an integral on S^2 of a product of the visual flow and some dual “visual effort”. Geometrically, the quantity that must be integrated must have the structure of a 2-form on S^2 [28]. The duality product for scalar flow and effort

variables is classically calculated using the exterior (wedge) product [28]. As a consequence the two dual fields on the sphere representing the visual flow and its dual visual effort should be respectively a k and a $2 - k$ form, such that their dual product, expressed using the wedge product, can be integrated as a 2-volume over the sphere. This will return a scalar representing the visual power contained in the full image sphere.

In our case, the flow variable is a section of TS^2 . We can choose to interpret this as a TS^2 -valued differential 0-form, that is a smooth map from $S^2 \rightarrow TS^2$ that takes $\eta \mapsto \Phi(\eta) \in T_\eta S^2$. Following this interpretation, we choose to define the visual effort as a T^*S^2 -valued differential 2-form. The T^*S^2 valued element of the 2-form will contract in a natural manner with the visual flow field to yield a scalar valued 2-form that will be integrated over the sphere to compute power. Thus, we define a visual effort to be

$$\gamma \in \Gamma(T^*S^2 \otimes \wedge^2(T^*S^2)), \quad (4)$$

that is a section of T^*S^2 -valued differential 2-forms over the sphere.

We define a contraction operator by

$$c: \Gamma(TS^2) \times \Gamma(T^*S^2 \otimes \wedge^2(T^*S^2)) \rightarrow \Gamma(\wedge^2(T^*S^2)).$$

That is for $\gamma(\eta) = F(\eta) \otimes s^2(\eta)$ with $F(\eta) \in T_\eta^*S^2$ and $s^2 \in \wedge^2(T^*S^2)$ then

$$\begin{aligned} c(\Phi(\eta), \gamma(\eta)) &= F(\eta)(\Phi(\eta)) \otimes s^2(\eta) \\ &= F(\eta)(\Phi(\eta))s^2(\eta) \end{aligned}$$

since the tensor product for a scalar by a tensor is just scalar multiplication of the tensor. The contraction operator c can be used to define the dual product

$$\langle \gamma | \Phi \rangle := \int_{S^2} c(\Phi, \gamma) \in \mathbb{R} \quad (5)$$

This bracket operation provides a means to compute the power flowing through a visual port realised via a physical camera attached to the rigid-body.

III. COUPLING THE VISUAL PORT TO THE VEHICLE DYNAMICS

In order to make sense of a visual energy port it is necessary to understand how this port couples to the underlying mechanical port of the rigid-body to which it is attached.

The vehicle velocity screw T is associated with the velocity of the rigid-body $\{B\}$. Thus, the mapping $n: \mathfrak{se}(3) \rightarrow \Gamma(TS^2)$ (3), that takes the rigid-body velocity T to visual flow Φ , describes the contra-variant part of the modulated transformer (MTF) that links these two components (see Fig. 1). The dual (adjoint) map corresponding to the MTF will transform visual effort back to a virtual wrench W on the rigid-body such that the power $\langle W | T \rangle$ is equal to the power flowing through the energy port $\langle \gamma | \Phi \rangle$.

In order to compute this reverse mapping we must compute the adjoint, n^* , of the mapping n . The adjoint map is defined as

$$n^*: \Gamma(T^*S^2 \otimes \wedge^2(T^*S^2)) \rightarrow \mathfrak{se}^*(3), \quad \gamma \mapsto n^*(\gamma) \quad (6)$$

such that

$$\langle n^*(\gamma) | T \rangle = \langle \gamma | n(T) \rangle. \quad (7)$$

Consider an explicit expression of $\gamma(\eta) = F(\eta) \otimes s^2(\eta)$ for $F \in \Gamma(T^*S^2)$ while $s^2(\eta) \in \wedge^2(T^*S^2)$. In coordinates given by the embedding of S^2 in \mathbb{R}^3 one has

$$\gamma(\eta) = \sigma(\eta)F(\eta) * dr(\eta) \quad (8)$$

where $F(\eta)$ is a row vector in the subspace $\{Z \in \mathbb{R}^{1 \times 3} \mid Z\eta = 0\}$ corresponding to $T_\eta^*S^2$ embedded in \mathbb{R}^3 , σ is a positive scalar function on the sphere, and the notation $*dr$ is used to describe the canonical volume form¹ on the sphere S^2 embedded in \mathbb{R}^3 .

We now go ahead in computing the visual power for an arbitrary visual flow $\Phi = n(T)$ and visual effort γ in coordinates

$$\langle \gamma | \Phi \rangle = \int_{S^2} c(\Phi, \gamma) = \int_{S^2} \sigma F(\eta)A(\eta)T * dr$$

where $F(\eta)A(\eta)T$ is matrix multiplication and yields a scalar field. Recalling that T is constant with respect to the direction on the sphere one can factor the T outside the integral

$$\langle \gamma | \Phi \rangle = \left(\int_{S^2} \sigma(\eta)F(\eta)A(\eta) * dr(\eta) \right) T$$

Finally, using the inner product for power on $\mathfrak{se}(3)$ one obtains

$$\langle \gamma | \Phi \rangle = \left\langle \int_{S^2} \sigma(\eta)F(\eta)A(\eta) * dr(\eta) \middle| T \right\rangle$$

and by (7) the expression for the wrench $W \in \mathbb{R}^{1 \times 6}$ is given by

$$W = n^*(\gamma) = \int_{S^2} dW(\eta) \quad (9)$$

with

$$\begin{aligned} dW(\eta) &= \sigma(\eta)F(\eta)A(\eta) * dr(\eta) \\ &= -\sigma(\eta)F(\eta) \left(-\eta \times \frac{1}{\lambda(\eta)} \mathbb{P}_\eta \right) * dr(\eta) \end{aligned} \quad (10)$$

where we have substituted for the explicit expression for $A(\eta)$. Note that in this expression, the scene depth $\lambda(\eta)$ appears that represents the distance of an observed point in the direction $\eta \in S^2$. This distance is not directly measurable by means of vision and leads to difficulties in the implementation of arbitrary visual impedances.

IV. VISUAL PORT-BASED IMPEDANCE CONTROL

Passivity Based Control [31] or Energy Based Control commonly involves two steps [27]; firstly a shaping of the system energy function or addition of an exogenous forcing term to drive the system in the direction desired, and secondly an injection of damping to ensure stability. The paradigm we use here is based on what is called *control by interconnection* [28] that relies on the structural properties of

¹The notation comes from obtaining the S^2 volume element by applying the star-Hodge operator to the differential of the (constant) radial coordinate of the sphere r .

the port Hamiltonian framework. In this framework, damping is obtained by adding a dissipative energy port into the system and tuning its parameters. Energy shaping can be achieved by addition of more complex elements, while the forcing term is achieved by adding an external source of effort or flow that feeds energy into the system. In this section we consider firstly the structure of a dissipative energy port based on the vision sensor, and then go on to consider an exogenous control input or forcing term that is also realized through the vision port.

The proposed control scheme can be graphically represented as a bond graph as shown in Figure 1. The Hamiltonian vehicle dynamics are modelled by the inertia I and the associated velocity screw T and wrench W . The corresponding visual flow Φ and effort γ are seen on the right of the modulated transformer connecting the vehicle port with the visual port. According to equation (3) the modulation is given by $A(\cdot)$. The idea now is to measure the visual flow Φ and to determine γ based on this measurement, i.e. to design an impedance control scheme.

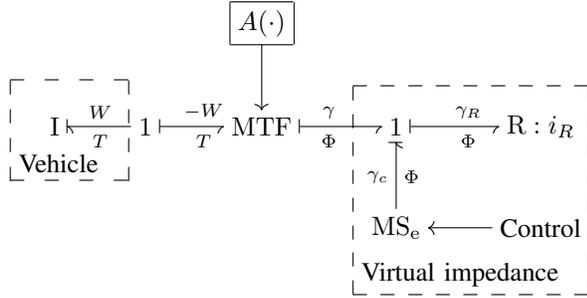


Fig. 1. Visual port-based impedance control

Damping is provided by connecting a dissipative element R to the visual 1-junction. This is particularly important in situations where the natural damping on a robotic vehicle's dynamics is minimal such as the case encountered in aerial robotic vehicle. Remote control is provided by supplying (virtual) energy generated by the modulated effort source MS_e to the vehicle where the modulation is the control signal. That is we propose to inject visual effort directly into the visual port in order that the control is explicitly undertaken with respect to the local environment with all the advantages that this brings in remote control of robotic vehicles [32], [33]. Note that superposition of an extra control action could also be done on the 1-junction on the left of the MTF, but we wish specifically to consider control action that is dependent on, and compatible with, data coming from vision sensors. It is important to understand that the final *physical implementation* of this control law happens by supplying energy via a wrench at the 1-junction in the vehicle. The energy that needs to be supplied at this port is only the difference between the virtual control energy and the virtual energy dissipated through the virtual damping resistor. The latter does hence not waste *real* energy, i.e. fuel. Given a desired control law, the implementing wrench can easily be computed using Eq. (10).

A. Visual Dissipation

We can use the construction of the visual port presented in Section II to create a physical behaviour in the visual space that resembles a dissipative energy port R . We will do this by introducing a suitable constitutive relationship that governs dependence of the visual effort on the observed visual flow.

A memoryless constitutive relationship between visual flow and visual effort is a function

$$i_R: \Gamma(TS^2) \rightarrow \Gamma(T^*S^2 \otimes \wedge^2(T^*S^2)), \Phi \mapsto \gamma_R. \quad (11)$$

That is, a function that maps a visual flow, Φ_t defined on the whole sphere at time t , to a visual effort $(\gamma_R)_t$ defined as cotangent field tensored with an S^2 volume element over the whole sphere at time t . In order for this relationship to be dissipative one requires that for all Φ_t then

$$\langle i_R(\Phi_t) | \Phi_t \rangle \geq 0.$$

That is that the total power flowing into the resistor R (cf. Figure 1) is always positive. Note that this is not a strict relation and it would also include situations in which a restricted part of the sphere is used.

We begin by defining a symmetric positive definite $(0, 2)$ -tensor, R , on the sphere S^2 . That is $R_\eta: T_\eta S^2 \times T_\eta S^2 \rightarrow \mathbb{R}$ is a symmetric bilinear map with $R_\eta(X, X) > 0$ for all non-zero $X \in T_\eta S^2$. The visual flow Φ_t can be lifted point wise to generate a cotangent field $F_t^R \in \Gamma(T^*S^2)$ by

$$F_t^R(\bullet) = R(\Phi_t, \bullet).$$

In practice, we will work with the embedded representation of the sphere. Thus, the coordinate representation of R_η should be a 3×3 positive definite matrix that we will also denote by $R_\eta \in \mathbb{R}^{3 \times 3}$ in an abuse of notation. The tensor acts on vectors $X, Y \in T_\eta S^2 \subset \mathbb{R}^3$ by $R_\eta(X, Y) = X^\top R_\eta Y$. Thus,

$$F_\eta^R = \Phi_t^\top R_\eta \mathbb{P}_\eta$$

is a row vector in $T_\eta^* S^2 = \{Z \in \mathbb{R}^{1 \times 3} \mid Z\eta = 0\}$. Note that since F_η^R is only ever applied to vectors $X \in T_\eta S^2$, the projection operator \mathbb{P}_η is algebraically unnecessary in the formulae, i.e. $\Phi_t^\top R_\eta \mathbb{P}_\eta X = \Phi_t^\top R_\eta X$ for $X \in T_\eta S^2$.

It is also necessary to choose a scaling function $\sigma_R: S^2 \rightarrow \mathbb{R}$ corresponding with the σ term in the coordinate representation of the visual effort (8). Since the tensor R is arbitrary it would be possible to roll the scaling function into the tensor and avoid the additional notation, however, it is convenient in the design process discussed later to have the two separate parameters available. Now set

$$i_R: \Gamma(TS^2) \rightarrow \Gamma(T^*S^2 \otimes \wedge^2(T^*S^2)), \\ i_R(\Phi_t(\eta)) := \sigma_R(\eta) F_t^R(\eta) * dr(\eta). \quad (12)$$

This representation of i_R is written in embedded coordinates. In the language of bond graphs this process produces the top right hand bond in Figure 1.

It is instructive to confirm that this construction yields a dissipative element. Compute the power in the bond defined

by i_R as

$$\begin{aligned} \langle i_R(\Phi_t) | \Phi_t \rangle &= \int_{S^2} c(\Phi_t, i_R(\Phi_t)) \\ &= \int_{S^2} \sigma_R(\eta) (F_t^R \Phi_t) * dr(\eta) \\ &= \int_{S^2} \sigma_R(\eta) \Phi_t^\top(\eta) R_\eta \Phi_t(\eta) * dr(\eta). \end{aligned}$$

Since $\Phi_t(\eta)^\top R_\eta \Phi_t(\eta) > 0$ for all $\Phi_t \neq 0$ and $\sigma_R > 0$ then the total power is always positive and the element designed is indeed dissipative.

B. Visual Forcing

The design of a visual effort that corresponds to a desired motion is now a matter of choosing a modulation function

$$i_c: \Gamma(TS^2) \rightarrow \Gamma(T^*S^2 \otimes \wedge^2(T^*S^2)), \quad \Phi \mapsto \gamma_c$$

for the virtual modulated effort source MS_e , cf. Figure 1. Note that the value $i_c(\Phi)(\eta) \in T_\eta^*S^2$ of γ_c at $\eta \in S^2$ can in principle depend on the whole visual flow field anywhere on the sphere and not just on the value $\Phi(\eta)$ of the visual flow field at η . This way localized effort can be injected based on the global scene geometry. An example for this would be a simple obstacle avoidance scheme where the visual flow is analyzed to identify a potential obstacle in direction $w \in S^2$. Control effort can be applied in the opposite direction, $-w$, to the identified obstacle. To implement this concept in the image domain we define a co-vector field

$$(Z_\eta^w) = \frac{1}{4\pi} w^\top \mathbb{P}_\eta \in T_\eta^*S^2$$

where the 4π term normalises for the surface area of S^2 . The co-vector field Z^w is expressed embedded in \mathbb{R}^3 but satisfies the constraints associated with T^*S^2 . Choose an arbitrary scaling function $\sigma_c: S^2 \rightarrow \mathbb{R}$ then the proposed forcing term for translation control expressed in embedded coordinates is

$$\gamma_c(\eta) := \sigma_c(\eta) Z_\eta^w * dr(\eta). \quad (13)$$

Note that since it can be easily shown that $\eta_\times(w^T \mathbb{P}_\eta)^T = 0$ for all w then

$$\int_{S^2} Z_\eta^w \eta_\times * dr(\eta) = 0.$$

Thus, if $\sigma_c(\eta) = 1$ is chosen equal to unity then the wrench on the vehicle due to the forcing term, with reference to (10), is

$$W_c = \left(\int_{S^2} \frac{1}{\lambda(\eta)} Z^w(\eta) \mathbb{P}_\eta * dr(\eta) \right) =: \begin{pmatrix} 0 \\ f_c \end{pmatrix}.$$

It is easy to prove that $\langle f_c, w \rangle$, that is the inner product between the applied linear force on the vehicle and the desired force direction w is always positive, hence the vehicle will move away from the obstacle. Note that the above construction is just one way to generate a driving signal in the image domain.

In the above calculations the scene depth λ is implicitly present in all the calculations. In practice, any control scheme

based on visual flow using this formalism will, *in principle*, require an estimate of scene depth. There are several classical computer vision methods available to obtain local world models [34] and estimates of depth can also be constructed from flow estimates [35]. However, it is undesirable to rely on scene depth estimates if possible. In practice, the only point where dependence matters in practice is when the actual wrench W on the vehicle is computed. It is not immediately clear in the above formulae whether this is the case, due to the integration over the sphere S^2 .

C. Dependence on scene depth

In this section we briefly demonstrate that the physical implementation of the dissipative element given by Eq. (12) *requires* knowledge of the scene depth λ . To this end we compute the wrench W corresponding to $\gamma = i_R(\Phi)$ via Eqs. (9), (10) and (3). We get

$$\begin{aligned} W(\eta) &= \\ &= -T^\top \int_{S^2} \sigma_R(\eta) \begin{pmatrix} \eta_\times R_\eta \eta_\times & -\frac{1}{\lambda(\eta)} \eta_\times R_\eta \mathbb{P}_\eta \\ \frac{1}{\lambda(\eta)} \mathbb{P}_\eta R_\eta \eta_\times & -\frac{1}{\lambda(\eta)^2} \mathbb{P}_\eta R_\eta \mathbb{P}_\eta \end{pmatrix} * dr(\eta) \end{aligned}$$

where T is the screw that generated Φ . Since $R_\eta > 0$, the upper left corner of the matrix in the integrand integrates to a matrix with one positive eigenvalue that is independent of λ while the lower right corner integrates to a matrix with two negative eigenvalues that depend on λ . Hence, not both of these terms can be made independent of λ at the same time by choosing the scalar function σ_R , the only remaining degree of freedom.

V. IMPEDANCE CONTROL IN SPLIT COORDINATES

Scene depth is difficult to compute a-priori and it is of interest to consider whether it is possible to design a control scheme that uses the port Hamiltonian framework but that doesn't require a scene depth estimate λ in order to be implemented. In the following we demonstrate an approach based on a splitting of visual flow in a translational and a rotational component.

It is possible to decompose the visual flow Φ_t into two components

$$\Phi_t = \Theta_t + \Psi_t.$$

corresponding to rotation and linear motion of the camera. Here

$$\Theta_t = \eta_\times \omega$$

is the component of the visual flow due to the angular velocity component of the screw T and

$$\Psi_t = \frac{-1}{\lambda} \mathbb{P}_\eta v$$

is the component of visual flow measurement due to the linear velocity component of the screw T . All aerial robotic vehicles are equipped with (calibrated) angular velocity gyrometers, providing a measurement of ω , necessary to stabilize the flight dynamics. Since the rotational component of optical flow Θ_t does not depend on scene depth, it can be estimated directly from the gyrometer angular velocity

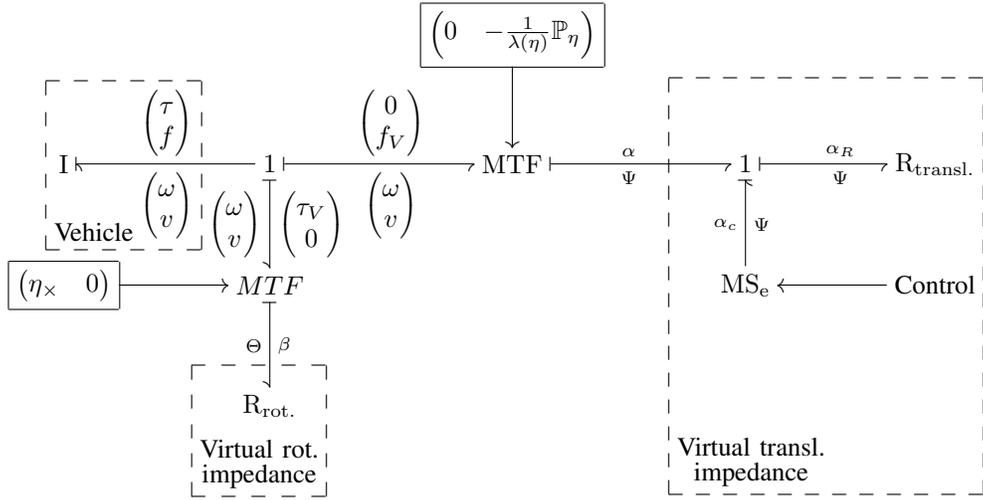


Fig. 2. Visual port-based impedance control using rotational and translational visual flow components

measurements, and the resulting translational component estimated as the difference $\Psi_t = \Phi_t - \Theta_t$.

In the following construction we assume that both the rotational (Θ) and translational (Ψ) visual flows are separately available and generate separate infinite dimensional ports to the Hamiltonian system. This leads to a different graphical representation of the system given by Figure 2 where two separate MTFs are used to express the separate visual ports. On the vehicle side, this splits the wrench W into a torque τ and a translational force f and the screw T into an angular velocity ω and a linear velocity v . The two bonds connected to the vehicle 1-junction symbolise the fact that we are now adding two separate controllers, the horizontal bond is associated with translational force and the vertical bond is associated with torque. Each controller consists of a virtual impedance, mapping visual flow to visual effort and thus generating the controlling force, respectively torque, through the respective MTF. The port Hamiltonian structure of the control architecture makes this splitting of control essentially trivial to implement, a significant advantage of the proposed approach.

In this configuration it is now straight forward to construct a control scheme that performs visual damping (through dissipation) and visual forcing in a way such that the resulting controlled translational force f_V and controlled torque τ_V do not depend on the unknown scene depth λ and can hence be implemented in a control system that is entirely based on optical flow measurements. As an example we provide the formulas for a setup as in Figure 2, where the rotational and translational movements are damped separately through visual resistors and translational control is supplied through a visual effort that is injected in the translational part of the visual port.

A rotational resistor $R_{\text{rot.}}$ is given by

$$\beta(\eta) = i_{R_{\text{rot.}}}(\Theta_t(\eta)) := \sigma_{R_{\text{rot.}}} \Theta_t(\eta)^\top R_{\text{rot.,}\eta} \mathbb{P}_\eta * dr(\eta)$$

in embedded coordinates, where $R_{\text{rot.,}\eta}$ is a 3×3 positive definite matrix and $\sigma_{R_{\text{rot.}}}$ is a positive scalar function, cf.

Eq. (12). It yields a torque

$$\tau_V = - \int_{S^2} \sigma_{R_{\text{rot.}}}(\eta) (\eta_\times R_{\text{rot.,}\eta} \eta_\times) * dr(\eta) \cdot \omega,$$

cf. Section IV-C, which is independent of scene depth λ as long as $\sigma_{R_{\text{rot.}}}$ is chosen independent of λ . Analogously, a translational resistor $R_{\text{transl.}}$ is given by

$$\begin{aligned} \alpha(\eta) &= i_{R_{\text{transl.}}}(\Psi_t(\eta)) \\ &:= \sigma_{R_{\text{transl.}}} \Psi_t(\eta)^\top R_{\text{transl.,}\eta} \mathbb{P}_\eta * dr(\eta) \end{aligned}$$

and yields a translational force

$$f_V = \int_{S^2} \sigma_{R_{\text{transl.}}} \frac{1}{\lambda(\eta)^2} \mathbb{P}_\eta R_{\text{transl.,}\eta} \mathbb{P}_\eta(\eta) * dr(\eta) \cdot v$$

which can be made independent of λ by choosing $\sigma_{R_{\text{transl.}}}(\eta) = \lambda(\eta)^2 > 0$. We leave it to the reader to work out the formulas for the translational forcing analogous to Eq. (13).

VI. CONCLUSIONS

This paper presents initial results on the development of a Hamiltonian energy port interpretation for vision based control of vehicles. The port Hamiltonian approach offers a control design paradigm that deals inherently with the vehicle dynamics and also with a dynamically changing environment. Multiple control modalities can be interconnected using the port structure of the Hamiltonian system, a significant practical advantage of the approach taken, and the natural passivity characteristics of the architecture ensure highly robust and stable control, as well as simplifying the handling of transmission or communication delays. We model the camera image plane as an infinite dimensional energy port with visual flow as the flow variable. We propose a structure for visual efforts and a bracket operation that computes power. From this we can derived the structure of the modulated transfer function that connects this port to the vehicle dynamics. A dissipative visual impedance is proposed and we show how to integrate an exogenous forcing function into the visual port. The initial approach taken leads

to a dependence on scene depth and we consider a second approach where the visual flow field is split into two separate energy ports. In this framework we can implement control that is independent of the scene depth.

VII. ACKNOWLEDGMENTS

This research was partly supported by the Australian Research Council through discovery grant DP0880509, “Image-based teleoperation of semi-autonomous robotic vehicles” and Future Fellowship FT0991771 “Foundations of Vision Based Control of Robotic Vehicles”.

REFERENCES

- [1] B. Espiau, F. Chaumette, and P. Rives, “A new approach to visual servoing in robotics,” *IEEE Transactions on Robotics and Automation*, vol. 8, no. 3, pp. 313–326, 1992.
- [2] S. Hutchinson, G. Hager, and P. Corke, “A tutorial on visual servo control,” *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 651–670, 1996.
- [3] C. Samson, M. Le Borgne, and B. Espiau, *Robot Control: The task function approach*, ser. The Oxford Engineering Science Series. Oxford, U.K.: Oxford University Press, 1991.
- [4] R. Kelly, “Robust asymptotically stable visual servoing of planar robots,” *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 759–766, 1996.
- [5] E. Zergeroglu, D. Dawson, M. de Queiroz, and S. Nagarkatti, “Robust visual-servo control of robot manipulators in the presence of uncertainty,” in *Proceedings of the 38th Conference on Decision and Control*, Phoenix, Arizona, USA., 1999.
- [6] A. Maruyama, H. Kawai, and M. Fujita, “Stability and tracking performance of dynamic visual feedback control for nonlinear mechanical systems,” in *Proc. 40th IEEE Conf. Decision and Control*, vol. 5, 2001, pp. 4415–4420.
- [7] A. Astolfi, L. Hsu, M. Netto, and R. Ortega, “Two solutions to the adaptive visual servoing problem,” *IEEE Transactions on Robotics and Automation*, vol. 18, no. 3, pp. 387–392, 2002.
- [8] T. Hamel and R. Mahony, “Robust visual servoing for under-actuated dynamic systems,” in *Proceedings of the Conference on Decision and Control, CDC’2000*, Sydney, N.S.W., Australia, 2000, pp. 3933–3938.
- [9] —, “Visual servoing of an under-actuated dynamic rigid-body system: An image based approach,” *IEEE Transactions on Robotics and Automation*, vol. 18, no. 2, pp. 187–198, April 2002.
- [10] H. Kawai, T. Murao, and M. Fujita, “Image-based dynamic visual feedback control via passivity approach,” in *Proc. IEEE Computer Aided Control System Design IEEE Int. Conf. Control Applications IEEE Int. Symp. Intelligent Control*, 2006, pp. 740–745.
- [11] M. Fujita, H. Kawai, and M. W. Spong, “Passivity-based dynamic visual feedback control for three-dimensional target tracking: Stability and l2-gain performance analysis,” *IEEE Control Systems Technology*, vol. 15, no. 1, pp. 40–52, 2007.
- [12] R. Mahony, P. Corke, and T. Hamel, “Dynamic image-based visual servo control using centroid and optic flow features,” *Journal of Dynamical Systems, Measurement and Control*, vol. 130, no. 1, p. 011005 (12 pages), January 2008.
- [13] F. Ruffier and N. Franceschini, “Optic flow regulation: The key to aircraft automatic guidance,” *Robotics and Autonomous Systems*, vol. 50, pp. 177–194, 2005.
- [14] M. Srinivasan and S. Zhang, “Visual motor computations in insects,” *Annual Review of Neuroscience*, vol. 27, pp. 679–696, July 2004.
- [15] D. Lee, “A theory of visual control of braking based on information about time to collision,” *Perception*, vol. 5, no. 4, pp. 437–459, 1976.
- [16] R. Nelson and J. Aloimonos, “Obstacle avoidance using flow field divergence,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 10, pp. 1102–1106, 1989.
- [17] D. Coombs, M. Herman, T. Hong, and M. Nashman, “Real-time obstacle avoidance using central flow divergence and peripheral flow,” *IEEE Transactions on Robotics and Automation*, vol. 14, no. 1, pp. 49–59, 1998.
- [18] C. McCarthy, N. Barnes, and R. Mahony, “A robust docking strategy for a mobile robot using flow field divergence,” *IEEE Transactions on Robotics*, vol. 24, no. 4, pp. 832–842, Aug. 2008.
- [19] M. Srinivasan, J. Chahl, K. Weber, S. Venkatesh, M. Nagle, and S. Zhang, “Robot navigation inspired by principles of insect vision,” *Robotics and Autonomous Systems*, vol. 26, pp. 203–216, 1999.
- [20] S. Humbert, A. Hyslop, and M. Chinn, “Experimental validation of wide-field integration methods for autonomous navigation,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2007)*, San Diego, CA, USA, 2007.
- [21] B. Herisse, T. Hamel, R. Mahony, and F. Russotto, “A terrain-following control approach for a vtol unmanned aerial vehicle using translational optical flow,” *Autonomous Robots*, vol. 29, no. 3-4, pp. 381–391, 2010, doi 10.1007/s10514-010-9208-x.
- [22] A. Beyeler, J.-C. Zufferey, and D. Floreano, “Vision-based control of near-obstacle flight,” *Autonomous Robots*, vol. 27, pp. 201–219, 2009, 10.1007/s10514-009-9139-6. [Online]. Available: <http://dx.doi.org/10.1007/s10514-009-9139-6>
- [23] R. Moore, S. Thurrowgood, D. Bland, D. Soccol, and M. Srinivasan, “A stereo vision system for uav guidance,” in *Proceedings, IEEE/RSJ International Conference on Intelligent Robots and Systems*, St. Louis, Missouri, USA., 11-15 October 2009.
- [24] A. Beyeler, J.-C. Zufferey, and D. Floreano, “optipilot: control of take-off and landing using optic flow,” in *Proceedings of the European Micro Air Vehicle conference and competition 2009 (EMAV 2009)*, Delft, The Netherlands, 2009, 8 pages (no pagination).
- [25] B. Herisse, T. Hamel, R. Mahony, and F.-X. Russotto, “Landing a vtol unmanned aerial vehicle on a moving platform using optical flow,” *IEEE Transactions on Robotics*, to appear, conditionally accepted as Regular paper 5 January 2011.
- [26] V. Duindam, A. Macchelli, S. Stramigioli, and H. Bryunickx, Eds., *Modeling and Control of Complex Physical Systems*. Springer, 2009.
- [27] C. Secchi, S. Stramigioli, and C. Fantuzzi, *Control of interactive robotic interfaces: A port-Hamiltonian approach*, ser. Springer Tracts in Advanced Robotics. Berlin: Springer, 2007, vol. 29.
- [28] V. Duindam, A. Macchelli, S. Stramigioli, and H. Bryunickx, Eds., *Modeling and Control of Complex Physical Systems*. Springer, 2009.
- [29] R. Mahony, F. Schill, P. Corke, and Y.-S. Oh, “A new framework for force feedback teleoperation of robotic vehicles based on optical flow,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2009.
- [30] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, “Performance of optical flow techniques,” *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.
- [31] R. Ortega, A. Loria, P. J. Necklasson, and H. Sira-Ramirez, *Passivity-based Control of Euler-Lagrange Systems: Mechanical, Electrical and Electro-Mechanical Applications*, ser. Communications and Control Engineering. London, U.K.: Springer-Verlag, 1998.
- [32] F. Schill, R. Mahony, P. Corke, and L. Cole, “Virtual force feedback teleoperation of the insectbot using optic flow,” in *Proceedings of the Australasian Conference on Robotics and Automation*, Canberra, Australia, December 2008.
- [33] F. Schill, R. Mahony, and P. Corke, “Estimating ego-motion in panoramic image sequences with inertial measurements,” in *Proceedings of the International Symposium on Robotics Research (ISRR)*, 2009.
- [34] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [35] J. Lim and N. Barnes, “Directions of egomotion from antipodal points,” in *Proceedings of CVPR*, 2008.