Modular Design of Image Based Visual Servo Control for Dynamic Mechanical Systems

Robert Mahony

Abstract This paper presents a modular framework for design of image based visual servo control for fully actuated dynamic mechanical systems. The approach taken uses the formalism of port Hamiltonian systems to track energy exchanged between the mechanical system and virtual potentials or Hamiltonians associated with each image feature. Asymptotic stability of the system is guaranteed by injecting damping to the otherwise conservative system. A simple approach based on full state measurement is presented and then extended to deal with unmeasured relative depth of image features.

1 Introduction

Visual servoing algorithms have been extensively developed in the robotics field over the last twenty years. *Image-Based Visual Servo* (IBVS) control regulates the dynamics of features in the image plane directly [1], resolving the cartesian motion planning task implicitly [2, 3]. IBVS control methods offer advantages in robustness to camera and target calibration errors, reduced algorithmic complexity and are easily extended multiple camera scenarios [4]. Classical visual servo control was principally developed for industrial serial-link robotic manipulators [4] where the dynamics of the system can be easily compensated using a computed torque (or high gain) control design. As a consequence, classical IBVS control is framed in the kinematic setting [1]. The last decade has seen a number of developments of visual servo control algorithms that consider the full dynamics of a mechanical system motivated by non-stiff robotic manipulators, such as those used in human safe environments, and other applications such as visual servo control of aerial vehicles. Kelly *et al.* considered the full dynamics of a robot and used a simple image based error feedback along with damping to prove asymptotic stability of the full

Robert Mahony
Australian National University, ACT, 0200, Australia, e-mail: Robert.Mahony@anu.edu.au

system, firstly for planar robots [5], then later for full 6-DOF manipulators with known image depth [6]. Zergeroglu *et al.* [7] used robust backstepping methods and adaptive estimation to deal with unknown calibration and system dynamics. Bishop *et al.* [8] considered a similar problem using non-linear PD control techniques. Hamel *et al.* [9] used an image centroid feature to ensure passivity properties of system and then applied robust backstepping to control a dynamic under-actuated model of an aerial robotic vehicle. Maruyama *et al.* [10] took an approach similar to that of Kelly [5, 6] and worked with a general Euler-Lagrange model of a robotic manipulator. A key issue in all visual servo control schemes is overcoming the loss of relative depth information associated with using an imaging system to observe a target point. Maruyama and Fujita further developed their work to include an observer that estimates the camera pose [11, 12], effectively estimating the unkown depths of the image points. An advantage of this approach is that a positive-definite control Lypunov function for the closed-loop system is available and provides an estimate of L_2 gain of the visual servo control loop [12]. Since this work incorporates an explicit estimate of the system pose it is natural to use a pose based error for the control criteria, leading to position based visual servo control [4]. Kawai *et al.* [13], however, showed that an image based regulation error can be used if desired; although the control algorithm still contains the complexity of a full pose estimator. Several authors have also considered using navigating functions [14, 15, 16] for visual servo control of dynamic systems.

In this paper, present a novel modular framework for design of image based visual servo control for dynamic systems. The proposed approach uses the structure of port Hamiltonian systems represented graphically using the bondgraph formalism. Each image feature is associated with a separate branch of the bondgraph, ensuring a modularity and structural simplicity to the design framework. The natural pairing of generalized forces (efforts) and generalized velocities (flows) associated with the mechanical system are transformed into image flow and image effort in a power preserving modulated transformer. Image Hamiltonians are introduced in the image space that represent stored energy associated with the image variables. The control objective is assigned by shaping of the total potential energy of the complete bondgraph using the flexibility available in choosing the image Hamiltonian potentials. Asymptotic stability of the system is obtained by injecting damping into the otherwise conservative system. Since each image branch of the graph is independent of the others the approach is inherently modular and image points can be added or removed arbitrarily as long as care is taken that the total system energy is preserved and that the system potential is always shaped with a minimum at the desired pose. A further advantage of the approach is that it is straightforward to interface the proposed design framework with other control algorithms that have a port-Hamiltonian representation. For example, bilateral force-feedback teleoperation of the system can be achieved by simply connecting a external source port representing the master system to the bondgraph.

The initial results of the paper are presented under the assumption that all system variables are measured. In practice, the relative depth of each image feature is never directly measured due to the physical nature of imaging systems. The second part

of the paper presents a modification of the design framework that incorporates an on-line estimate of the relative depth.

2 Classical Image Based Visual Servo Control

This section provides a brief review of classical IBVS methods for a serial manipulators.

Classical Image Based Visual Servo (IBVS) control was developed in the kinematic setting for serial manipulator devices [4]. Let $\{A\}$ denote the base frame or inertial frame of reference. Let $\{B\}$ denote the body-fixed frame, or end-effector frame of reference, and let ${}^A\xi_B$ (resp. ${}^AR_B$) denote the position (resp. orientation) of $\{B\}$ with respect to $\{A\}$. Note that ${}^A\xi_B \in \mathbb{R}^3$ while ${}^AR_B$ is a rotation matrix. For an N link manipulator the generalized coordinates or joint variables of the manipulator are denoted $q \in \mathbb{R}^N$. Let ${}^BV = {}^B_AV_B$ and ${}^B\Omega = {}^B_A\Omega_B$ denote the linear and angular velocity of $\{B\}$ with respect $\{A\}$ expressed in $\{B\}$. Let ${}^BU = ({}^BV, {}^B\Omega)$ denote the combined spatial velocity of $\{B\}$ with respect to $\{A\}$ expressed in $\{B\}$. Then the velocity Jacobian of the manipulator, $J(q) \in \mathbb{R}^{6\times N}$ gives the relationship between spatial and generalized velocities

$$ {}^BU = J(q)\dot{q}. \tag{1}$$

Any physical point p can be given coordinates either in the world frame ${}^Ap \in \{A\}$ or in the end-effector frame ${}^Bp \in \{B\}$. The mapping between these coordinate representations of p is given by the transformation mapping ${}^AH_B : {}^Bp \mapsto {}^Ap$,

$$ {}^Ap = {}^AH_B({}^Bp) = {}^AR_B\,{}^Bp + {}^A\xi_B.$$

with inverse ${}^Bp = {}^AH_B^{-1}({}^Ap) = {}^AR_B^\top\left({}^Ap - {}^A\xi_B\right)$. The time variation of ${}^Bp$ is given by

$$ {}^B\dot{p} = {}^Bp \times {}^B\Omega - {}^BV + {}^AR_B^\top\,{}^A\dot{p}. \tag{2}$$

In the sequel, we will only be concerned with the case where the point p is stationary in the world frame, that is ${}^A\dot{p}$. In this case, it is convenient to write a matrix form for (2)

$$ {}^B\dot{p} = A({}^Bp)\ {}^BU := \begin{pmatrix} -I_3 & {}^Bp_\times \end{pmatrix}\begin{pmatrix} {}^BV \\ {}^B\Omega \end{pmatrix}$$

where ${}^Bp_\times$ is the 3×3 matrix such that ${}^Bp_\times v = {}^Bp \times v$ for any vector $v \in \mathbb{R}^3$.

Classical eye-in-hand IBVS control uses a set of features s_i observed by a camera attached to the end-effector frame of reference as its measurements. Many different visual features have been considered in the past, however, for this paper I only consider point features[1], that is the image coordinates $s_i = (u_i, v_i)$ of the image of an

[1] The principles presented in this paper should generalize to any image based feature.

observed point p_i for $i = 1, \ldots, n$. Let $(x_i, y_i, z_i)^\top = {}^B p \in \mathbb{R}^3$. Then for point features observed with a calibrated pinhole camera model one has

$$s_i := \begin{pmatrix} u_i \\ v_i \end{pmatrix} = \sigma \begin{pmatrix} \frac{x_i}{z_i} \\ \frac{y_i}{z_i} \end{pmatrix} \tag{3}$$

where $\sigma > 0$ is a scalar that represents the focal length of the camera. That is s_i is a non-linear function $f(x_i, y_i, z_i) = \sigma(\frac{x_i}{z_i}, \frac{y_i}{z_i})$. We will term this function f the *image map*.

The image interaction matrix or image Jacobian, L_i, is the linear relationship between the instantaneous variation of the image feature and the generalized coordinates, (Remark 1 below justifies why I write $L_i := L_i(s_i, z_i, q)$)

$$\begin{aligned} L_i(s_i, z_i, q) &= Df({}^B p_i)\, A({}^B p_i)\, J(q) \\ &= \frac{1}{z_i} \begin{pmatrix} \sigma & 0 & -u_i \\ 0 & \sigma & -v_i \end{pmatrix} \begin{pmatrix} -I_3 & ({}^B p_i)_\times \end{pmatrix} J(q) \in \mathbb{R}^{2\times N} \end{aligned} \tag{4}$$

recalling that $s_i = (u_i, v_i)$ and rewriting the equations in the standard form. The image interaction matrix is the linear mapping between generalized velocities of the manipulator and the velocity of the observed image point

$$\dot{s}_i = L_i(s_i, z_i, q)\dot{q}. \tag{5}$$

Let $s = \left(s_1^\top, \ldots, s_n^\top\right) \in \mathbb{R}^{2n}$ be the concatenated vector of stacked image features and define

$$L(s, z_1, \ldots, z_n, q) := \begin{pmatrix} L_1(s_1, z_1, q) \\ \vdots \\ L_n(s_n, z_n, q) \end{pmatrix}. \tag{6}$$

Then the combined image kinematics can be written

$$\dot{s} = L(s, z_1, \ldots, z_n, q)\dot{q} \tag{7}$$

Remark 1. Recalling Eq. 4 it appears to be the case that $L_i := L_i(s_i, z_i, {}^B p_i, q)$ depends on the full target position ${}^B p_i$ as well as s_i, z_i and q. In fact the term $({}^B p_i)_\times$ in the rightmost block of $A({}^B p_i)$ can be combined with the depth dependence of $Df({}^B p_i)$ to yield

$$\frac{1}{z_i} \left({}^B p_i\right)_\times = \begin{pmatrix} s_i \\ 1 \end{pmatrix}_\times .$$

Although this trick cancels both the z_i and ${}^B p$ dependence in the rightmost 3×3 block of $Df({}^B p_i)A({}^B p_i)$ the z_i dependence of the leftmost block remains. Thus, to compute L_i one requires the variables s_i, z_i and q.

Let $s_i^* \in \mathbb{R}^2$ be a set of desired image coordinates. The task considered in IBVS control is to move the robot until the camera is in a position such that $s_i = s_i^*$ for $i = 1, \ldots, n$. The desired image points $\{s_i^*\}$ should be chosen to be feasible, that is that

there should exist a physical pose of the camera in which the observed points match the desired $\{s_i^*\}$. Typically image based visual servo control is used to return to a pose that has been visited before and the desired image coordinates are computed directly from a reference image obtained at the desired pose, ensuring that this pose is feasible. The image error is defined to be

$$\tilde{s} = s - s^* = \left((s_1 - s_1^*)^\top, \ldots, (s_n - s_n^*)^\top \right) \in \mathbb{R}^{2n}. \tag{8}$$

The control is chosen to stabilize the image error

$$\dot{q} = -kL^\dagger(s, z_1, \ldots, z_n, q)\tilde{s}$$

where $k > 0$ is a position scalar gain and $L^\dagger = (L^\top L)^{-1} L^\top$ is the matrix pseudo-inverse. The image features s and manipulator joint coordinates q are measured and are available and can be used in computing L. The z_i coordinates of ${}^B p_i$, however, are usually not available due to the nature of imaging devices. The simplest work around to this issue used in practice is to use approximations $(\hat{z}_1, \ldots, \hat{z}_n)$ to the depth and compute an estimate $\hat{L} := L(s, \hat{z}_1, \ldots, \hat{z}_n, q)$ of the true interaction matrix. There are plethora of schemes proposed in the literature to computer estimates of $\{\hat{z}_i\}$ or directly compute estimates of $\hat{L}$ or even its pseudo-inverse [4, 17, 18, 19].

3 Modular Image Based Visual Servo Control for a Dynamic Systems

This section introduces a port Hamiltonian frame work for design image based visual servo control algorithms. A key advantage of this framework is that each image feature is treated as a separate energy port attached to the system dynamics. Thus, the separate image features can be treated as detachable modules in the control design leading to a number of advantages that will be discussed in future sections. The theory is developed starting from a general Euler-Lagrange model of a mechanical system expressed with respect to generalized coordinates. Although this model is directly applicable to the standard dynamic models of robotic manipulators, it is not necessarily this class of systems that is of most interest as potential applications for the results. For example, the port Hamiltonian approach has recently been applied aerial robotic vehicles [20] and there are a large number of additional systems that can easily and naturally be modeled using this framework.

Let $q \in \mathbb{R}^N$ denote generalised coordinates for a fully actuated mechanical system with generalised inertia matrix $M(q)$ and potential function $U(q)$. Lagrange's equations yield dynamics

$$M(q)\ddot{q} = -C(q, \dot{q})\dot{q} - \frac{\partial U}{\partial q}(q) + \tau \tag{9}$$

where $\tau \in \mathbb{R}^N$ are the generalised forces. The Hamiltonian associated with the mechanical system is

$$H_0(q,p) := \frac{1}{2} p^\top M^{-1}(q) p + U(q), \qquad p = M(q)\dot{q} \tag{10}$$

and the Euler-Lagrange dynamics (9) is equivalent to the standard port Hamiltonian equations with flow and effort variables

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p}(q,p) \\ \dot{p} &= -\frac{\partial H}{\partial q}(q,p) + \tau \end{aligned} \tag{11}$$

where $(\tau, \dot{q})$ is the effort-flow pairing of the energy multi-port of the Hamiltonian. That is

$$\frac{d}{dt} H_0(q,p) = \langle \tau \,|\, \dot{q} \rangle = \tau^\top \dot{q} \tag{12}$$

where the $\langle \cdot \,|\, \cdot \rangle$ is the power in the port. In the graphical representation of bondgraphs[2] the port Hamiltonian sytem is represented as a single storage bond attached to a 1-junction as shown in Figure 1. The causality stroke of the bond (horizontal line at the tip of the arrow) indicates that the effort τ is the input to the Hamiltonian system while the flow $\dot{q}$ is the output, as is natural in the differential equation (9).

Fig. 1 Single bond storage element indicating the Hamiltonian dynamics associated with a mechanical system with Hamiltonian $H_0(q,p)$ and Dirac structure as specified in (11).

The approach taken is to treat each image feature as a separate bond interconnection in a bondgraph with the Hamiltonian dynamics of the system present as initial element. Thus, each image feature s_i will result in a bond attached to the 1-junction in Figure 1 with an associated effort denoted τ_i as shown in Figure 2. Following the rules of 1-junctions in bondgraphs the flow $\dot{q}$ in each of the bonds is equal while the forces add $\tau + \sum_{i=1}^{n} \tau_i = 0$.

In order to interface the ith image bond to the image feature a non-linear modulated transformer is used to transform the flow $\dot{q}$ to the image flow $\dot{s}_i$. The transformer relationship is defined by the relationship between image flow and the generalized velocity given by (5). That is the image interaction matrix $L_i(s_i, z_i, q)$ is the

[2] In Figure 1 the arrow on the bond should be a half arrow with the hook on the side of the flow variable. However, I was unable to typeset this effectively, and I have chosen to make all bonds full arrows in this paper to make the notation consistent.

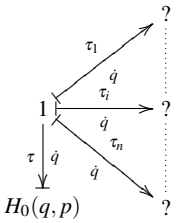

Fig. 2 Attachment of a multiple image feature bonds. Each image feature contributes an effort τ_i to the 1-junction that couples with the flow $\dot{q}$ of the mechanical system. These efforts sum to produce the IBVS control action. The question marks indicate that this part of the bondgraph remains to be defined.

defining matrix for the transformer relationship

$$\dot{s}_i = L_i(s_i, z_i, q)\dot{q}, \qquad \tau_i = L_i(s_i, z_i, q)^\top e_i \tag{13}$$

where e_i is the effort variable that will be associated with the flow $\dot{s}_i$. The relationship between image effort e_i and τ_i is implied by the principle of power conservation in the transformer. The resulting bondgraph has the form shown in Figure 3 where only the ith image bond is shown and where the line without an arrowhead indicates exchange of information without flow of power. The transformer is modulated by

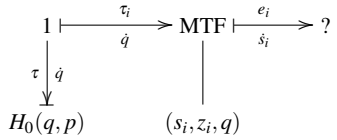

Fig. 3 Modulated transformer to relate image flow $\dot{s}_i$ with the flow $\dot{q}$ at the 1-junction.

the signals (s_i, z_i, q) via the dependence of $L_i(s_i, z_i, q)$ on these variables.

In Figure 3 the causality of the bonds indicates that the effort variable τ_i depends on e_i (13), and that the effort variable e_i depends on the element placed at the question mark. The specification of the flow variable $\dot{s}_i$ indicates that the image coordinates s_i should act as the energy variables associated with any storage associated with the image effort e_i. The approach taken in this paper is to use an *image Hamiltionian*, that is a spring like storage element defined in the image space, to define the relationship between the image flow $\dot{s}_i$ and the image effort e_i. The natural image Hamiltonian to use is the squared norm of image error. Assume that the target image coordinates s_i^* are known and that they are constant. Define an image storage element (or Hamiltonian) by

$$H_i(s_i) := \frac{1}{2} k_i ||s_i - s_i^*||^2. \tag{14}$$

where k_i is a positive scalar. Treating this element in the classical manner then the natural interconnection (Dirac structure) for the bondgraph is to assign

$$e_i := \frac{\partial H_i}{\partial s_i} = k_i(s_i - s_i^*) \tag{15}$$

It follows that

$$\frac{d}{dt}H_i = \left(\frac{\partial H_i}{\partial s_i}\right)^\top \dot{s}_i = \langle e_i \,|\, \dot{s}_i \rangle = k_i(s_i - s_i^*)^\top \dot{s}_i$$

as expected for a storage element. The resulting bond graph is shown in Figure 4.

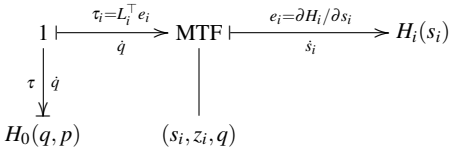

Fig. 4 Complete structure of the modular image branch of the proposed bondgraph design framework.

The proposed image Hamiltonian $H_i(s_i)$ (14) is chosen to such that the minimum energy in of the image Hamiltonians occurs when $s_i = s_i^*$, normally corresponding to correct positioning of the camera. In more generality the image Hamiltonians should be chosen to shape the energy of total system, including the potential energy in the mechanical system and any other storage elements that may be integrated into the system, in order that its minimum corresponds to the desired regulation point of the system. Shaping the energy of the total system may require additional insight in a given system configuration or control task, and indeed may utilize more than just the image variables. In particular, in the present case where the mechanical system itself has potential energy then this must be balanced either by shaping the image Hamiltonians or by adding additional storage that cancels the potential $U(q)$ of the mechanical system. In the present development I will compensate the potential $U(q)$ directly by adding a separate Hamiltonian storage element $-U(q)$ attached the 1-junction that depends on the known generalized coordinates q. The final bondgraph for the closed-loop system (Figure 5) is obtained by adding an additional dissipation term R, such that $\delta = R\dot{q}$ with $R > 0$ is a positive definite matrix.

For the bondgraph given in Figure 5 the generalized forces τ to the mechanical system are specified by the relationship implied by the 1-junction

$$\begin{aligned}\tau &= -\sum_{i=1}^{n} \tau_i - \delta + \frac{\partial U}{\partial q}(q) \\ &= -\sum_{i=1}^{n} k_i L_i(s_i, z_i, q)^\top (s_i - s_i^*) - R\dot{q} + \frac{\partial U}{\partial q}(q)\end{aligned} \tag{16}$$

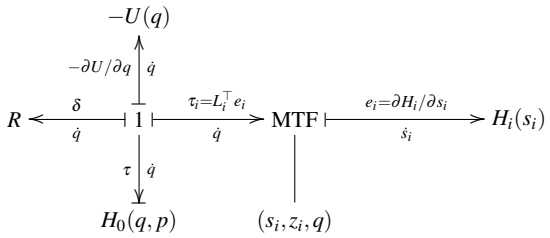

Fig. 5 Proposed modular framework for dynamic image based visual servo control.

The total energy of the closed-loop system is given by sum of all the storage elements

$$H(q,p,s_i) = H_0(q,p) + \sum_{i=1}^{n} H_i(s_i) - U(q) = \frac{1}{2} p^\top M^{-1}(q) p + \frac{1}{2} \sum_{i=1}^{n} k_i ||s_i - s_i^*||^2 \quad (17)$$

The theory of port-Hamiltonian system now guarantees that

$$\frac{d}{dt} H(q,p,s_i) = -\langle \delta \,|\, \dot{q} \rangle = -\dot{q}^\top R \dot{q}. \quad (18)$$

Lemma 1. *Consider the generalized mechanical system* (9) *with input* τ *given by* (16)*. Assume that the total energy* $H(q,p,s_i)$ (17) *has a unique minimum at* (q^*, p^*, s_i^*) *and is radially unbounded in* $(q,\dot{q})$*. Assume furthermore that* $L(s,z_1,\ldots,z_n,q)$ (6) *is full rank. Then, for any initial condition* $(q(0),\dot{q}(0))$ *the closed-loop trajectory* $(q(t),\dot{q}(t)) \to (q^*,0)$.

Proof. Since $H(q,p,s_i)$ is radially unbounded in $(q,\dot{q})$ then (12) implies that trajectories of $(q,\dot{q})$ are bounded and exist for all time. From (18) along with Lyapunov's theorem and LaSalle's principle then $(q,\dot{q})$ converges to the largest forward invariant set contained in $\dot{q}^\top R \dot{q} = 0$ proving that $\dot{q}(t) \to 0$.

From (9) it follows that on LaSalle's invariant set $\tau = \partial U/\partial q$. Substituting from (16) one has

$$\sum_{i=1}^{n} k_i L_i(s_i,z_i,q)^\top (s_i - s_i^*) = L(s,z_1,\ldots,z_n,q)^\top \operatorname{diag}(k_1 I_2,\ldots,k_n I_2)(s - s^*) = 0$$

where $\operatorname{diag}(k_1 I_2,\ldots,k_n I_2)$ is the diagonal $2n \times 2n$ matrix with block diagonal entries $k_i I_2$ and $(s - s^*)$ is given by Eq. 8. Since $L(s,z_1,\ldots,z_n,q)$ is full rank it follows that $\tilde{s} = 0$ and the uniqueness of the minimum of the Hamiltonian energy implies that $q(t) \to q^*$. *QED.*

The proposed control scheme provides a simple and intuitive design methodology for image based visual servo control of a mechanical system. A key advantage of the proposed control is that, unlike the classical IBVS approach, it is not a linearizing control design. In classical IBVS the pseudo inverse of the image Jacobian

$L(s, z_1, \ldots, z_n, q)$ and may lead to significant stability issues in the closed-loop system if the estimate of L is poor. The proposed scheme benefits from the natural robustness and passivity of the port Hamiltonian framework. Good performance of the closed-loop response depends on tuning the damping and spring coefficients.

The proposed approach is highly modular with each image feature dealt with as a separate branch of the bondgraph. This structure has considerable advantages in the practical implementation of image based visual servo control. In particular, the proposed framework allows one to develop heuristic schemes that can switch between features, drop old features, or add new features as long as the total energy of the system is conserved. For example, if at time t_1 it is wished to switch from image feature s_i to a new image feature $\bar{s}_i$ then by choosing $\bar{k}_i$ such that

$$\bar{k}_i = k_i \frac{||s_i - s_i^*||^2}{||\bar{s}_i - \bar{s}_i^*||^2}$$

the value of the total Hamiltonian $H(t)$ is continuous at time t_1. As a consequence the basic Lyapunov stability of the system is certainly preserved and $\dot{q} \to 0$. The time variation that is introduced by a heuristic switching scheme invalidates the LaSalle argument used in the proof of Lemma 1, however, for reasonable choices of switching strategies and feature choices it is to be expected that the full stability result will hold.

A modification of this idea can be used if an image feature is leaving the field of view. As an image feature i exits the field of view at time t_1, the energy in the ith image Hamiltonian

$$H_i(s_i(t_1)) = \frac{1}{2}||s_i(t_1) - s_i^*||^2$$

can be partitioned into $(n-1)$ portions H_i^j for $j = 1, \ldots, n$ with $j \neq i$. Each $k_j^{\text{after}} := k_j^{\text{before}} + \Delta_j$ can then be augmented by

$$\Delta_j = 2||s_i - s_i^*||^2 H_i^j, \quad j = 1, \ldots, n, \quad j \neq i$$

that increases the energy in the image Hamiltonians $H_j(s_j)$ to compensate for the energy lost when $H_i(s_i)$ is removed from the bondgraph. A new image feature s_{n+1} can be added in a similar manner by taking energy from the existing image Hamiltonians, with the caveat that it may be necessary to choose the scaling k_{n+1} of the new feature sufficiently small to ensure there is sufficient energy available to cover its initial potential.

Another classical problem in image based visual servo control, that of ensuring that no image feature leaves the field of view during the evolution of the closed-loop system is also easily addressed. In this case the image Hamiltonian can be augmented by a barrier function

$$H_i(s_i) = \frac{k_i}{2}||s_i - s_i^*||^2 + \Phi(s_i)$$

where Φ is a positive definite function that is radially unbounded on the boundary of the image. It is still necessary to ensure that the total Hamiltonian of the system is shaped such that the minimum energy corresponds to the desired pose of the system.

A final advantage of the proposed approach is that it can be easily interfaced with other port-Hamiltonian control modalities. For example, exogenous user input such as a haptic interface can be added by including an additional bond to an exogenous source associated with the mode of input of the user. In this way it is straightforward to integrate bilateral force-feedback teleoperation into the image based visual servo control framework.

The proposed approach will still suffer from certain of the failings of classical IBVS control. In particular, the assumption that the total Hamiltonian $H(q,p,s_i)$ is radially unbounded in $(q,\dot{q})$ made in Lemma 1 is not satisfied in many classical configurations of IBVS, and without this assumption it is possible that the trajectory in q may become unbounded. For example, the now classical 'Chaumette conundrum' [21] has four image points located in a square configuration with the camera initial condition such that the observed points are exactly 180° rotated around the centre axis. In this case the proposed control will act to reduce the distance of all four points to the target vector causing them to contract into the centre of the image, corresponding to the camera moving infinitely far away along image axis, an unbounded motion in q. In this case the Hamiltonian is indeed decreasing along the closed-loop trajectories of the system, however, the potential is not radially unbounded in q.

4 On-line estimation of image depth

This section presents a extension of the development in Section 3 that incorporates an estimator for the unknown relative depth signals required in the implementation of (16).

The generalized forces τ_i (Fig. 5) associated with each individual image feature are given by the relationship

$$\tau_i = -k_i L_i(s_i, z_i, q)^\top (s_i - s_i^*).$$

In order to implement the proposed control exactly it is necessary to have measurements of the variables s_i, s_i^*, z_i and q. Of these signals it is only the relative depth z_i that is classically considered unknown. The image coordinates s_i and s_i^* are the primary measurements of the IBVS servo control problem, while the classical IBVS control problem is formulated in the context of serial robotic manipulator for which the generalized coordinates q are typically available to the control algorithm. In some of the more modern applications of IBVS control, such as those involving aerial robotic vehicles, the generalized coordinates q may be more difficult to measure, however, in the present paper I will focus on the classical problem formulation where the only unmeasured variable is the relative range z_i of an image feature.

The most common approach to dealing with unknown depth z_i in IBVS control is to use an estimate $\hat{z}_i$ in the algorithm implementation. This estimate is classically chosen *a priori* as a constant estimate based on a best guess available given the expected environmental configuration [4]. Typically the expected depth at the regulation point is used as a constant estimate of depth in the control algorithm. An alternative approach is to build an estimator that uses information obtained during the trajectory motion to estimate the relative depth z_i online [17]. Other approaches include partial pose estimation [18] and optimization based methods [19]. The approach taken in this paper is most closely coupled to an observer, or adaptive control [17] based approach. The novelty in the proposed approach is the integration of this idea into the port-Hamiltonian framework.

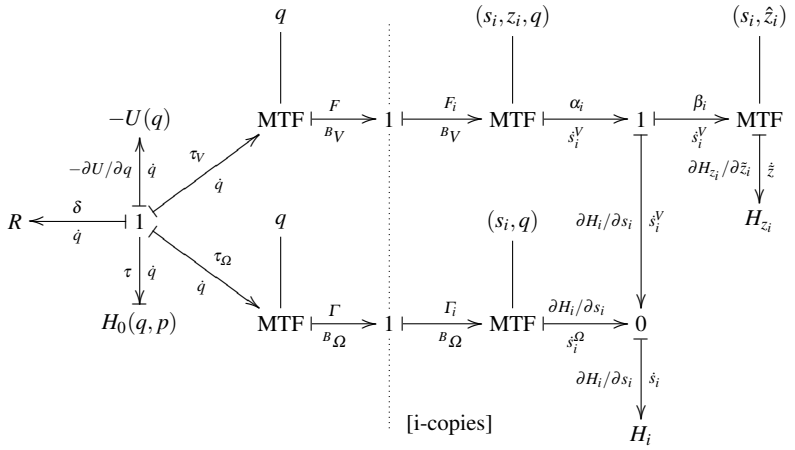

Fig. 6 The proposed architecture for modular image based visual servo control with depth estimation.

Consider the bondgraph shown in Figure 6. The leftmost 1-junction, the Hamiltonian H_0 for the mechanical system, the $-U(q)$ potential, and the damping R are identical to the bondgraph in Figure 5 and motivated in the same way. The difference in this section lies on the righthand side of the graph where the contributions from the translational and rotational motion of camera frame to the generalized mechanical system are separated into two modulated transformers. Recalling (1) it is clear that the relationship between rigid-body and generalized coordinate velocity can be decomposed as

$$\begin{pmatrix} {}^B V \\ {}^B \Omega \end{pmatrix} = \begin{pmatrix} J_V(q) \\ J_\Omega(q) \end{pmatrix} \dot{q},$$

and the matrices $J_V(q)$ and $J_\Omega(q)$ define the two modulated transformers in Fig. 6. The principle of conservation of power implies that

$$\tau_V = J_V^\top(q) F, \qquad \tau_\Omega = J_\Omega^\top(q) \Gamma.$$

providing a decomposition of the generalized force τ into a component due to translational motion and one due to rotational motion of the camera frame. This decomposition is undertaken before the separate image branches of the bondgraph are created. The additional 1-junctions in the translational and rotational branch of the bondgraphs are added to provide a junction for the n image feature branches of the bondgraph, however, now each image feature generates a pair of image bonds, one associated with translational motion and one associated with rotational motion. Only the ith image bonds are shown in Figure 6 and the dotted line is used to indicate that there are n modular pairs of image bonds attaching to the two 1-junctions. The transformation into image velocity is achieved using a pair of modulated transformers based on a decomposition of Eq. 4. Firstly, write (4) as

$$L_i(s_i,z_i,q) = \left(K_i^V(s_i,z_i,q)\ K_i^\Omega(s_i,q)\right) J(q)$$

where

$$K_i^V(s_i,z_i,q) = -\frac{1}{z_i}\begin{pmatrix}\sigma & 0 & -u_i\\ 0 & \sigma & -v_i\end{pmatrix}, \tag{19}$$

$$K_i^\Omega(s_i,q) = \begin{pmatrix}\sigma & 0 & -u_i\\ 0 & \sigma & -v_i\end{pmatrix}\begin{pmatrix}s_i\\ 1\end{pmatrix}_\times. \tag{20}$$

Here (20) is obtained by factoring the $1/z_i$ dependence into the ${}^B p_\times$ to obtain the term $(s_i,1)_\times$ as was discussed in Remark 1. Define two new image flow variables, $\dot{s}_i^V$ and $\dot{s}_i^\Omega$ for translational and rotational image flow respectively, by

$$\dot{s}_i^V = K_i^V(s_i,z_i,q)\,{}^B V, \qquad \dot{s}_i^\Omega = K_i^\Omega(s_i,q)\,{}^B\Omega. \tag{21}$$

The two relationships (21) define the second set of modulated transformers (to the right of the dotted line) in the bondgraph Figure 6. The reason for taking this approach is to separate the part of L_i that depends on the relative depth z_i into a single modulated transformer (the translational motion) branch that has a *scalar* dependence on the unknown variable. The modulated transformer in the rotational motion branch of the bondgraph depends only on known variables and can be implemented explicitly.

The translational and rotational image flows add to generate the full image flow

$$\dot{s}_i = \dot{s}_i^V + \dot{s}_i^\Omega. \tag{22}$$

This summation is implemented in the 0-junction in the bottom right of the bondgraph Figure 6. Here the image effort $\partial H_i/\partial s_i$ is equal in each of the branches of the 0-junction while the image flows add in accordance to the directions of the bond arrows. This completes the separation of a modular image feature branch of the bondgraph into two branches that rejoin to implement the image Hamiltonian H_i in the bottom right of the bondgraph. The justification for this decomposition of the bondgraph is to enable the inclusion of an estimator for the unknown depth into each translational motion image feature sub-branch of the bondgraph. In mod-

ifying the bondgraph, the image flow $\dot{s}_i^V$ must be preserved in-order to implement the image Hamiltonian. The image effort associated with the translational motion, however, can be modified. Access to the relevant signal is achieved by introducing a 1-junction into the upper branch of the bondgraph just before the translational and rotational image velocities are recombined in the 0-junction. The 1-junction introduced preserves flow in each branch ensuring that $\dot{s}_i^V$ is preserved. The 1-junction does introduce a relationship between the three efforts associated with bonds on the 1-junction

$$\alpha_i = \beta_i + \frac{\partial H_i}{\partial s_i}. \tag{23}$$

Let $\hat{z}_i$ denote a new variable that is an estimate of the relative depth z_i. Define

$$\alpha_i := \frac{z_i}{\hat{z}_i} \frac{\partial H_i}{\partial s_i}.$$

This assignment fixes the effort β_i according to (23)

$$\beta_i = \frac{(z_i - \hat{z}_i)}{\hat{z}_i} \frac{\partial H_i}{\partial s_i} \tag{24}$$

and fully defines the three bonds attached to the 1-junction in the top right of Figure 6. The reason for the choice of α_i can be seen by computing the relationship between F_i and α_i given by the modulated transformer

$$F_i = \frac{1}{z_i} \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \\ -u_i & -v_i \end{pmatrix} \alpha_i = \frac{1}{\hat{z}_i} \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \\ -u_i & -v_i \end{pmatrix} \left(\frac{\hat{z}_i}{z_i} \alpha_i \right) = \frac{1}{\hat{z}_i} \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \\ -u_i & -v_i \end{pmatrix} \frac{\partial H_i}{\partial s_i}$$

Thus, the effort F_i can now be computed based on a 'virtual' modulated transformer (obtained by replacing the unknown depth z_i by its estimate $\hat{z}_i$) acting on the image effort $\partial H_i / \partial s_i$. All variables here are available and the control can be implemented.

It remains to terminate the bond $(\dot{s}_i^V, \beta_i)$ in the top right of the bondgraph. To save notational complexity I will assume the image Hamiltonian is given by $H_i(s_i) = k/2 \, ||s_i - s_i^*||^2$ as was discussed in Section 3. This implies that $\partial H_i / \partial s_i = k(s_i - s_i^*)$.

Define a general form of dynamics for the observer $\hat{z}_i$

$$\dot{\hat{z}}_i = w_i \tag{25}$$

where w_i is an arbitrary driving term. The state $\hat{z}_i$ becomes an internal dynamic state in the implementation of the control algorithm. Let

$$\tilde{z}_i = \hat{z}_i - z_i. \tag{26}$$

denote the observer error.

Define a Hamiltonian for the relative depth error to be

$$H_{z_i}(\tilde{z}_i) := \frac{\gamma}{2} ||\tilde{z}_i||^2.$$

The natural flow variable associated with H_{z_i} is $\dot{\tilde{z}}_i$ while the corresponding effort is

$$\frac{\partial H_{z_i}}{\partial \tilde{z}}(\tilde{z}) = \gamma \tilde{z}_i. \tag{27}$$

The idea is to hypothesize a modulated transformer to relate the effort β_i with the effort $\partial H_{z_i}/\partial \tilde{z}$ and then implement this transformer by choosing the observer dynamics w_i to ensure that the flows $\dot{s}_i^V$ and $\dot{\tilde{z}}_i$ conform to the transformer relationship. Let $P(\cdot)$ denote the transformer relationship then

$$\beta_i = P(\cdot)\frac{\partial H_{z_i}}{\partial \tilde{z}}$$

and hence substituting from (24) and (27) and using $\partial H_i/\partial s_i = k_i(s_i - s_i^*)$ (14) one obtains

$$P(\hat{z}_i, s_i, s_i^*) := \frac{k}{\gamma \hat{z}_i}(s_i - s_i^*) \in \mathbb{R}^2 \tag{28}$$

Applying the principle of conservation of power then

$$\dot{\tilde{z}}_i = \frac{k}{\gamma \hat{z}_i}(s_i - s_i^*)^\top \dot{s}_i^V$$

Recalling (25) and differentiating (26) one has

$$\dot{\hat{z}}_i = w_i = \dot{z}_i + \dot{\tilde{z}}_i = {}^B V_z + \frac{k}{\gamma \hat{z}_i}(s_i - s_i^*)^\top \dot{s}_i^V \tag{29}$$

since the velocity $\dot{z}_i = {}^B V_z$ is just the z-axis velocity of the body fixed frame velocity, remembering that the target points are stationary. This process assigns observer dynamics to $\hat{z}_i$.

Space limitations prevent a formal statement of stability of the closed-loop system in this configuration. It is clear, however, that an analogous argument to that in Lemma 1 will guarantee firstly that $\dot{q} \to 0$ and secondly (using LaSalle's principle) that $q \to q^*$. There is no guarantee that the observer error $\tilde{z}_i$ actually converges to zero as the modulated transformer $P(\hat{z}_i, s_i, s_i^*)$ (28) decreases to zero as $s \to s^*$. Thus, the observer Hamiltonian H_{z_i} becomes decoupled from the full Hamiltonian system in the limit as $q \to q^*$ even if $\hat{z}_i \neq z_i$. Hence, even though the energy minimum of the total Hamiltonian occurs for $\tilde{z}_i = 0$, the modulated transformer can act to partition off energy in the observer Hamiltonian H_{z_i} that remains locked in storage as the system comes to equilibrium.

5 Conclusions

This paper used the port-Hamiltonian formalism to provide a design methodology for dynamic image based visual servo control. The approach is conceptually simple

once the underlying bondgraph formalism is accepted. The graphical representation provides a powerful visualization of design framework and I believe it will provide considerable insight leading to the development of practical control algorithms for a range of relevant problems.

Acknowledgements This research was supported by the Australian Research Council through Future Fellowship FT0991771 "Foundations of Vision Based Control of Robotic Vehicles"

References

1. B. Espiau, F. Chaumette, P. Rives, IEEE Transactions on Robotics and Automation **8**(3), 313 (1992)
2. C. Samson, M. Le Borgne, B. Espiau, *Robot Control: The task function approach*. The Oxford Engineering Science Series (Oxford University Press, Oxford, U.K., 1991)
3. R. Pissard-Gibollet, P. Rives, in *Proceedings of the IEEE International Conference on Robotics and Automation, ICRA'95* (Nagasaki, JAPAN, 1995), pp. 166–171
4. S. Hutchinson, G. Hager, P. Corke, IEEE Transactions on Robotics and Automation **12**(5), 651 (1996)
5. R. Kelly, IEEE Transactions on Robotics and Automation **12**(5), 759 (1996)
6. R. Kelly, R. Carelli, O. Nasisi, B. Kuchen, F. Reyes, IEEE/ASME TRANS. ON MECHATRONICS **5**(1), 39 (2000). DOI 10.1109/3516.828588
7. E. Zergeroglu, D. Dawson, M. de Queiroz, A. Behal, IEEE/ASME TRANSACTIONS ON MECHATRONICS **6**, 322 (2001)
8. B. Bishop, M. Spong, Control Engineering Practice **7**, 423 (1999)
9. T. Hamel, R. Mahony, IEEE Transactions on Robotics and Automation **18**(2), 187 (2002)
10. A. Maruyama, M. Fujita, Advanced Robotics **12**(14), 67 (1998)
11. A. Maruyama, H. Kawai, M. Fujita, in *Proc. 40th IEEE Conf. Decision and Control*, vol. 5 (2001), vol. 5, pp. 4415–4420. DOI 10.1109/.2001.980897
12. M. Fujita, H. Kawai, M.W. Spong, IEEE Control Systems Technology **15**(1), 40 (2007). DOI 10.1109/TCST.2006.883236
13. H. Kawai, T. Murao, M. Fujita, in *Proc. IEEE Computer Aided Control System Design IEEE Int. Conf. Control Applications IEEE Int. Symp. Intelligent Control* (2006), pp. 740–745. DOI 10.1109/CACSD-CCA-ISIC.2006.4776738
14. N. Cowan, J. Weingarten, D. Koditschek, Robotics and Automation, IEEE Transactions on **18**(4), 521 (2002)
15. R. Kelly, J. Moreno, R. Campa, in *43rd IEEE Conference on Decision and Control* (Atlantis, Paradise Island, Bahama, 2004), pp. 4028–4033
16. T. Murao, H. Kawai, M. Fujita, in *IEEE International Conference on Control Applications* (Yokohama, Japan, 2010)
17. N. Papanikolopoulos, P.K. Khosla, T. Kanade, in *Proceedings of the American Control Conference* (1991), pp. 962–967
18. E. Malis, F. Chaumette, S. Boudet, IEEE Transactions on Robotics and Automation **15**(2), 238 (1999)
19. J.A. Piepmeier, A dynamic quasi-newton method for model independent visual servoing. Phd, Georgia Institute of Technology, Atlanta, USA (1999)
20. S. Stramigioli, R. Mahony, P. Corke, in *Procedings of the IEEE International Conference on Robotics and Automation (ICRA)* (2010), pp. 5302–5308
21. F. Chaumette, in *Conflux of vision and control LNCIS* (1998), 237, pp. 66–78