Hybrid Eye-in-Hand/Eye-to-Hand Image Based Visual Servoing for Soft Continuum Arms

Ali AlBeladi, Evan Ripperger, Seth Hutchinson, *Fellow, IEEE*, and Girish Krishnan

***Abstract*—Soft continuum arms (SCAs) that are controlled by visual servoing (VS) present trade-offs between the camera range and tracking accuracy. Cameras placed at a distance (eye-to-hand) can observe a larger workspace area and the SCA tip, while a camera at the end effector (eye-in-hand) can more accurately survey the target. In this letter, we present a hybrid eye-to-hand and eye-in-hand VS scheme to track a desired object in the SCA's worksapce. When the target is not in the field-of-view of the tip camera, hand-to-eye VS is implemented using a wide field-of-view camera on the soft robot's base, to servo the soft robot's tip to a feasible region where the target is expected to be seen by the tip camera. This region is estimated by solving an optimization problem that finds the best region to place the SCA assuming a constant curvature model for the SCA. When the target is seen by the tip camera, the system switches to a hand-in-eye controller that keeps the target in the desired image position of the tip camera. Experimental results on the popular BR^2 SCA demonstrates the effectiveness of the hybrid VS scheme under practical settings that include external disturbances.**

***Index Terms*—Soft robotics, visual servoing.**

I. Introduction

SOFT continuum arms (SCAs) are hard to accurately control due to their inherent nonlinearity, hysterisis and deformable structure that can be affected by the environment. The use of vision systems in controlling soft robotic motion, i.e. visual servoing (VS), has been shown to be a viable choice due to their compactness and ease of implementation [1]–[17]. Typical visual servo systems are composed of a camera that matches features extracted from the current image with that of a desired image, and a control system that moves the robot such that the feature positions move closer to their desired positions in the image [18]. Most of the work in visual servoing for soft robots is concentrated on one of two setups. The first, traditionally called *eye-in-hand* VS, has the camera fixed on the tip of the soft robot and VS is used to control the soft robot to achieve the desired feature positions in the image. This camera can move close to the desired scene and capture *local* details. However, it has limited sight and cannot interact with the entire workspace of the robot. Eye-in-hand VS has been successfully applied to cable-driven soft robots [4], [19], concentric tube robot (CTR) [5], [8], catheters [13], series pneumatic artificial muscle (sPAM) [6], [11] and parallel pneumatic actuators [12], [15], [17].

The second setup, called *eye-to-hand* VS, has the camera fixed in the workspace with the soft robot in its field of view and VS is used to control the robot's tip to a desired position in the image. While the camera can capture a *global* view of the workspace it cannot be used to explore more details in the soft robot's environment. Eye-to-hand VS has been applied to tendon-driven SCAs [1], [3], [10], [14], [16], CTRs [2], [20], bending pneumatic actuators [9], and elephant trunk robots [21].

In this letter, a hybrid eye-in-hand and eye-to-hand VS scheme that controls an SCA to track a target in its workspace is proposed. A camera that is fixed to the SCA's base is used for eye-to-hand VS, which moves the tip of the soft robot closer to a region where the target can be seen by a camera that is attached to the SCA's tip. This region is obtained by solving an optimization problem that finds where the tip of a constant curvature SCA needs to be (in the base image) to see the target in the center of the tip camera. We emphasize here that the constant curvature assumption is not used to control the SCA, but rather to obtain a fast approximation of the region where the target could be seen by the tip camera. This will only ensure that the tip is close to the region. Whenever the target is observed in the tip camera's field-of-view, the system switches to the eye-in-hand VS controller which keeps the target in the desired image position. For brevity, these two systems will be called the global visual servoing (GVS) and the local visual servoing (LVS) systems, respectively. An illustration of this setup is shown in Fig. 1.

Various approaches for hybrid eye-in-hand/eye-to-hand VS have been investigated for rigid robots. In [22], cooperation between eye-to-hand and eye-in-hand has been proposed where the former controls the position of the end effector and the latter controls the orientation of the end effector. A similar setup was proposed in [23], however the fixed camera was only used to estimate the robot's end-effector position. In [24], multiple eye-to-hand cameras and eye-in-hand cameras (on different robotic arms) were used to estimate the pose of the desired object using an extended Kalman filter for sensor fusion. This pose estimate was used for position based VS. Sensor fusion

Manuscript received 24 February 2022; accepted 6 July 2022. Date of publication 28 July 2022; date of current version 29 August 2022. This letter was recommended for publication by Associate Editor C. Duriez and Editor C. Laschi upon evaluation of the reviewers' comments. This work was supported in part by NSF-USDA COALESCE under Grant USDA 2021-67021-34418, and in part by the Jump Applied Research for Community Health through Engineering and Simulation (ARCHES) program, an endowment partnership between OSF HealthCare and the University of Illinois Urbana-Champaign. *(Corresponding author: Ali AlBeladi.)*

Ali AlBeladi is with the Electrical Engineering, King Fahd University of Petroleum & Minerals, Dhahran 61822, Saudi Arabia (e-mail: albelad2@illinois.edu).

Evan Ripperger is with the Mechanical Science and Engineering, University of Illinois Urbana-Champaign, Champaign, IL USA (e-mail: evanjr2@illinois.edu).

Seth Hutchinson is with the School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: seth@gatech.edu).

Girish Krishnan is with the Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign, Champaign, IL USA (e-mail: gkrishna@illinois.edu).

Digital Object Identifier 10.1109/LRA.2022.3194690

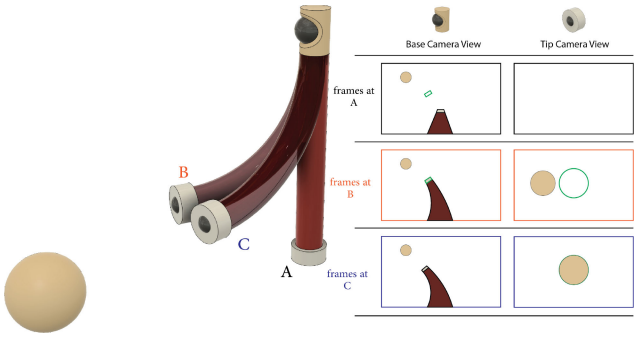

Fig. 1. An example of the Global and Local Visual Servoing Systems. The goal is to observe the target (spherical ball in this illustration) and keep it at the center of the tip camera's view (green circle in frame B). The SCA starts at the straight configuration (A). If the desired target is not seen in the tip camera, the system estimates where the SCA's tip should move to have the target in its field of view (green rectangle at frame A). The GVS controls the SCA to achieve the desired end tip position in the image and the SCA reaches (B). The view of the tip camera at (B) is slightly off target; the LVS servos the SCA to achieve the desired view and it reaches (C).

for image based VS using both schemes has also been proposed in [25]. Most of these methods require the object to be viewed in the eye-in-hand camera, therefore [26] proposed an initialization step for these methods using an epiplolar geometry based VS scheme that centers the epipolar line in the image and searches for the target along the line. In [27], multiple RGBD eye-to-hand cameras and a stereo eye-in-hand camera are used for hybrid position-based VS control of a robotic arm. The system initially is in eye-to-hand control and switches to eye-in-hand control when the distance between the robot and target is less than some threshold.

Although hybrid VS has been developed for rigid robots, the methods in [22–27] cannot be applied directly to SCAs. These methods utilize forward kinematics to get an accurate estimate of the end-effector pose, which is not possible with SCA due to the lack of accurate sensors and the fact that SCA's deformation behaviour is dependent on external forces and disturbances. In the context of SCA's, utilizing both eye-in-hand and eye-to-hand sensors has been achieved in [28] to guide an OCT probe along a desired path in the eye-to-hand camera, while keeping the probe in tangential contact with a target surface.

In this letter, we propose a hybrid VS scheme where image projections of the target and the SCA tip alone are used to formulate control laws. Unlike the work in [28], the main goal of this letter is to track a target and keep it in the center of the tip (eye-in-hand) camera, even if initially it is only in the field-of-view of the base (eye-to-hand) camera.

The main contributions of this work are:

1) developing a novel method to get a fast approximation of where the SCA's tip should be in the base camera in order to see the target in the tip camera.
2) developing a coarse to fine switch control system with the goal of tracking a target with the tip camera even if the target is not in its field of view. The system starts with an eye-to-hand VS controller that moves the SCA's tip to the approximated region (mentioned in the first contribution). When the target is seen in the local camera, the system switches to an eye-in-hand controller that keeps the target in the center of the image.

In the next sections, background information is presented. In Section III, the proposed hybrid visual servoing system for soft robots is introduced. Experimental evaluation of this method is then presented in Section IV. and and further work to be done.

II. Background

A. Soft Robot Model

The shape of the SCA can be described by a position vector $\mathbf{x}(s) \in \mathbb{R}^3$ and a rotation frame $\mathbf{R}(s) \in SO(3)$ at each cross-section $s \in [0, L]$ along its length L. For convenience, the position and orientation are expressed in a compact way by joining them into a single matrix in the special Euclidean group $SE(3)$ [29]

$$\mathbf{X}(s) = \begin{bmatrix} \mathbf{R}(s) & \mathbf{x}(s) \\ \mathbf{0} & 1 \end{bmatrix} \in SE(3).$$

The position and orientation of the center-points on the SCA's cross-sections evolve, with respect to its length parameter s, according to

$$\mathbf{X}'(s) = \mathbf{X}(s) \left[\mathbf{K}(s)\right]_\times, \tag{1}$$

$$\mathbf{K} = \begin{bmatrix} \boldsymbol{\kappa} \\ \mathbf{q} \end{bmatrix} \in \mathbb{R}^6, \quad \left[\mathbf{K}\right]_\times = \begin{bmatrix} \hat{\boldsymbol{\kappa}} & \mathbf{q} \\ 0 & 0 \end{bmatrix} \in \mathfrak{se}(3),$$

where $\boldsymbol{q} \in \mathbb{R}^3$ is a vector that contains the stretching/shearing strains, $\boldsymbol{\kappa} \in \mathbb{R}^3$ contains the bending/twisting strains, and $\hat{(\cdot)}$ is the usual mapping of a vector in $\mathbb{R}^3$ to a skew-symmetric matrix in $\mathfrak{so}(3)$, and $\left[\cdot\right]_\times$ is the map from $\mathbb{R}^6$ to the lie algebra $\mathfrak{se}(3)$. For the SCA used in this research, it is safe to assume that there is no stretching nor shearing and only bending and twisting is happening. In this case, we have $\mathbf{q} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$.

B. Traditional Visual Servoing

The goal of an image-based visual servoing (IBVS) system is to decrease the norm of the error, $\mathbf{e}(t)$, between the current, $\mathbf{p}(t)$, and desired, ${}^*\mathbf{p}$, positions of the observed visual features in the image,

$$\mathbf{e}(t) = \mathbf{p}(t) - {}^*\mathbf{p}. \tag{2}$$

Traditionally, $\mathbf{p}$ is the coordinates of a set of feature points (they can also be other features like lines, segments, etc.) in the image, $\mathbf{p} = \begin{bmatrix} \mathbf{p}_1^T & \mathbf{p}_2^T \dots \mathbf{p}_N^T \end{bmatrix}^T$ where $\mathbf{p}_n = \begin{bmatrix} x_n & y_n & z_n \end{bmatrix}^T$ is the image sphere coordinates of the n^{th} point, and N is the number of image feature points. Note that we use an omnidirectional camera on the base of the robot and thus, as in [30], [31], we are considering the spherical projection model for the camera. More specifically, the point features are projections of 3-D points of interest, with coordinates $\mathbf{x}_n = \begin{bmatrix} X_n & Y_n & Z_n \end{bmatrix}^T$ in the camera frame, onto the unit image sphere,

$$\mathbf{p}_n = \mathbb{P}(\mathbf{x}_n) = \frac{1}{r_n}\mathbf{x}_n$$
$$\text{where } r_n = \|\mathbf{x}_n\|. \tag{3}$$

The image sphere velocities of the features are related to the angular and linear velocities of the camera, $\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\omega}^T & \mathbf{v}^T \end{bmatrix}^T$,

through the relationship

$$\dot{\mathbf{p}}_n(t) = \mathbf{L}_n \boldsymbol{\Omega}(t), \tag{4}$$

where $\mathbf{L}_n \in \mathbb{R}^{3\times 6}$ is the interaction matrix for $\mathbf{p}_n$. The interaction matrix depends on the setup used for visual servoing. In an eye-to-hand VS setting, the camera is observing the robotic arm and the arm moves to achieve a desired image position for its end-effector. In eye-in-hand VS, the camera is on the tip of the robot's arm and the arm moves to achieve a desired view of an object in the image. Equations for the interaction matrices of the proposed setup will be presented in the next section.

To achieve a decrease in the feature position error, the most common approach is to apply the following control law

$$\boldsymbol{\Omega}(t) = -\lambda \mathbf{L}^{+} \mathbf{e}(t), \tag{5}$$

where λ is a positive constant and $\mathbf{L}^{+}$ is the pseudoinverse of the stacked interaction matrices

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 \\ \vdots \\ \mathbf{L}_N \end{bmatrix} \in \mathbb{R}^{3N\times 6}. \tag{6}$$

For underactuated systems, it is not possible to control all 6 degrees-of-freedom and therefore applying equation (5) would not be possible. Hence, the following relationship is used

$$\boldsymbol{\Omega}(t) = \mathbf{J}(\mathbf{u})\dot{\mathbf{u}}(t), \tag{7}$$

where $\mathbf{u} \in \mathbb{R}^{M}$ is the actuation control inputs and $\mathbf{J} \in \mathbb{R}^{6\times M}$ is the Jacobian matrix that maps actuation-space velocities to task-space velocities. From equations (4) and (7) we obtain the relationship between actuation velocities and image space velocities

$$\dot{\mathbf{p}} = \mathbf{L}\mathbf{J}\dot{\mathbf{u}}. \tag{8}$$

The following control law can be applied to servo the soft robot

$$\dot{\mathbf{u}}(t) = -\lambda \left(\mathbf{L}\mathbf{J}\right)^{+} \mathbf{e}(t). \tag{9}$$

Since the interaction matrix of a single feature is of rank two, the number of visual features required to obtain a velocity input are given by $N \geq \frac{M}{2}$ In the next section, the proposed hybrid VS system is introduced.

III. Methodology

In the proposed hybrid system, both eye-to-hand and eye-in-hand VS work together to track a target. Eye-to-hand VS utilizes a camera near the base of the SCA (*base camera*) to control its tip position, while eye-in-hand VS utilizes a camera on the SCA's tip (*tip camera*). The main goal is to move the SCA's tip in a position where the tip camera can keep a specific target in the center of the image (or another desired position in the image). Since eye-to-hand VS enables us to set a global position for the SCA's tip, we call it global VS (GVS). On the other hand, eye-in-hand VS enables more accurate positioning within a local region of the workspace, thus we call it local VS (LVS). Fig. 2 shows the reference frames for this setup and Fig. 3 shows a flow diagram of the developed system.

More specifically, the user sees the images from both cameras and can click on any target he/she wants the SCA to track. If the chosen target is in the tip camera's field of view, the LVS system will control the SCA to keep the target in the desired image position. If the target is only detected in the base camera, the GVS system moves the tip to an estimated image position that ensures the target will be in the tip camera's field of view. This desired tip position is obtained through solving an optimization problem and will be presented in Section III-C. If the target is detected in the tip camera, the controller will directly switch to the LVS system.

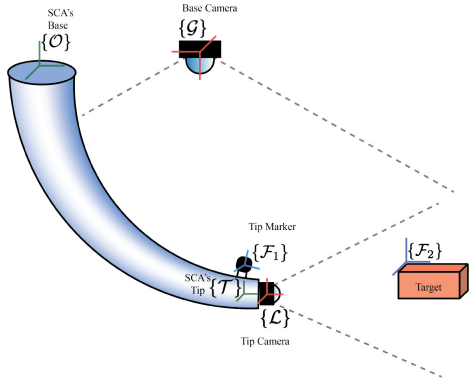

Fig. 2. An illustration of the coordinate frames for the hybrid visual servoing setup. These frames are for the SCA's base $\{\mathcal{O}\}$, SCA's tip $\{\mathcal{T}\}$, base camera $\{\mathcal{G}\}$, tip camera $\{\mathcal{L}\}$ tip marker $\{\mathcal{F}_1\}$, and target $\{\mathcal{F}_2\}$.

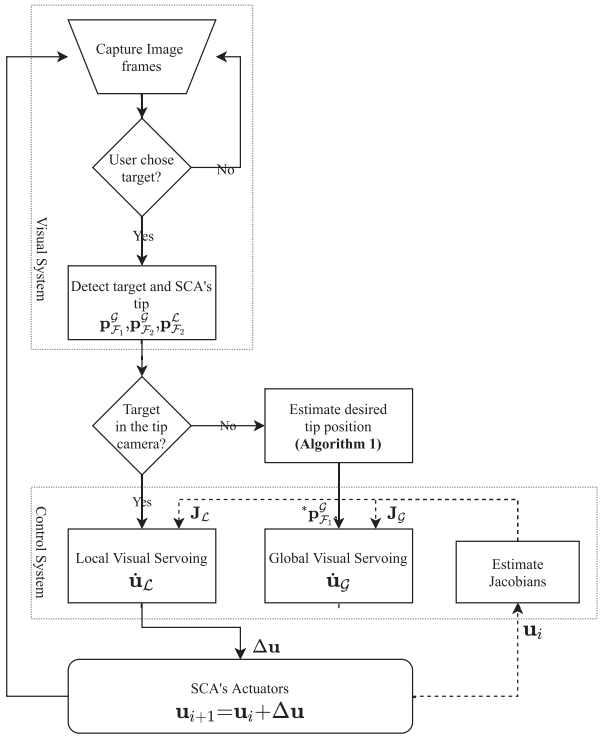

Fig. 3. Flow diagram of the Hybrid Visual Servoing scheme.

A. Hybrid Visual Servoing

Note that the GVS system tracks features on the tip of the SCA, frame $\mathcal{F}_1$, while the LVS system tracks features on the target, frame $\mathcal{F}_2$. The image sphere velocities of the features in both cameras are related to the SCA's control inputs, $\mathbf{u}$, through

the relationship

$$\dot{\mathbf{p}}^{\mathcal{L}}_{\mathcal{F}_2} = \mathbf{L}_{\mathcal{L}} Ad_{\mathbf{X}^{\mathcal{L}}_{\mathcal{T}}} \mathbf{J} \dot{\mathbf{u}}, \tag{10}$$

$$\dot{\mathbf{p}}^{\mathcal{G}}_{\mathcal{F}_1} = \mathbf{L}_{\mathcal{G}} Ad_{\mathbf{X}^{\mathcal{G}}_{\mathcal{T}}} \mathbf{J} \dot{\mathbf{u}}, \tag{11}$$

where $\mathbf{p}^{\mathcal{L}}_{\mathcal{F}_2}$ and $\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1}$ are the target and tip feature point projections on the image sphere of the tip and base cameras, respectively. The interaction matrices, $\mathbf{L}_{\mathcal{L}}$ and $\mathbf{L}_{\mathcal{G}} \in \mathbb{R}^{3\times 6}$, map the camera velocities to feature velocities in the spherical image space and are given by [30], [31]

$$\mathbf{L}_{\mathcal{L}} = \begin{bmatrix} \hat{\mathbf{p}}^{\mathcal{L}}_{\mathcal{F}_2} & -\frac{1}{r^{\mathcal{L}}_{\mathcal{F}_2}} \left(\mathbf{I} - \mathbf{p}^{\mathcal{L}}_{\mathcal{F}_2} {\mathbf{p}^{\mathcal{L}}_{\mathcal{F}_2}}^T \right) \end{bmatrix} \tag{12}$$

$$\mathbf{L}_{\mathcal{G}} = \begin{bmatrix} -\hat{\mathbf{p}}^{\mathcal{G}}_{\mathcal{F}_1} & \frac{1}{r^{\mathcal{G}}_{\mathcal{F}_1}} \left(\mathbf{I} - \mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1} {\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1}}^T \right) \end{bmatrix}. \tag{13}$$

where $r^{\mathcal{A}}_{\mathcal{B}} = \|\mathbf{x}^{\mathcal{A}}_{\mathcal{B}}\|$, and the Jacobian, $\mathbf{J}$, maps between the actuation space of the soft robot to velocities in the SCA's tip position and orientation. The matrices $Ad_{\mathbf{X}^{\mathcal{L}}_{\mathcal{T}}}$ and $Ad_{\mathbf{X}^{\mathcal{G}}_{\mathcal{T}}}$ are Adjoint maps in SE(3), defined by

$$Ad_{\mathbf{X}} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \hat{\mathbf{x}}\mathbf{R} & \mathbf{R} \end{bmatrix} \in \mathbb{R}^{6\times 6},$$

which map velocities from the tip frame $\{\mathcal{T}\}$ to the camera frames $\{\mathcal{L}\}$ and $\{\mathcal{G}\}$, respectively.

For notation simplicity, and without loss of generality, we take the case where the SCA has two degrees-of-freedom (i.e. $M = 2$). Since $N \geq \frac{M}{2}$, we see that one image feature ($N = 1$) is enough to obtain a velocity input. This can be easily generalized for higher degrees-of-freedom by stacking multiple feature points as done in (6). The LVS and GVS control inputs are, respectively, given by

$$\dot{\mathbf{u}}_{\mathcal{L}}(t) = -\lambda_{\mathcal{L}} \left(\mathbf{L}_{\mathcal{L}} Ad_{\mathbf{X}^{\mathcal{L}}_{\mathcal{T}}} \mathbf{J} \right)^{+} \left(\mathbf{p}^{\mathcal{L}}_{\mathcal{F}_2} - {}^{*}\mathbf{p}^{\mathcal{L}}_{\mathcal{F}_2} \right) \tag{14}$$

$$\dot{\mathbf{u}}_{\mathcal{G}}(t) = -\lambda_{\mathcal{G}} \left(\mathbf{L}_{\mathcal{G}} Ad_{\mathbf{X}^{\mathcal{G}}_{\mathcal{T}}} \mathbf{J} \right)^{+} \left(\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1} - {}^{*}\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1} \right). \tag{15}$$

where $\lambda_{\mathcal{L}}$ and $\lambda_{\mathcal{G}}$ are positive gain constants, ${}^{*}\mathbf{p}^{\mathcal{L}}_{\mathcal{F}_2}$ is the desired target position in the tip camera, and ${}^{*}\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1}$ is the desired tip position in the base camera. The hybrid control law is given by

$$\dot{\mathbf{u}}(t) = \begin{cases} \dot{\mathbf{u}}_{\mathcal{L}}(t) & \text{if target in f.o.v. of tip camera} \\ \dot{\mathbf{u}}_{\mathcal{G}}(t) & \text{otherwise} \end{cases} \tag{16}$$

B. Offline Jacobian Learning

To implement the individual VS control laws in (14) and (15), an estimate of the Jacobian is needed. In this letter, the Jacobians were estimated through a simple offline model-free approach that builds on the work of [3]. The actuation space of the SCA is divided into segments that form a grid. At each node in this grid, the actuation inputs are slightly perturbed and the resulting changes in image feature positions are measured (for both the base camera and the tip camera). Several of these measurements are obtained and the Jacobians map to the camera frames (i.e. $\mathbf{J}_{\mathcal{L}} := Ad_{\mathbf{X}^{\mathcal{L}}_{\mathcal{T}}} \mathbf{J}$ and $\mathbf{J}_{\mathcal{G}} := Ad_{\mathbf{X}^{\mathcal{G}}_{\mathcal{T}}} \mathbf{J}$) are estimated using least squares, with the assumption that the Jacobians are relatively equivalent for subsequent time-steps. All the Jacobians are stored in memory with their corresponding actuation space grid points. In the servoing process, the Jacobians of the nearest grid point to the current actuation input are used in equations (14) and (15).

Algorithm 1: Estimate Desired Tip Position.

Inputs:

$\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_2}$ (Image position of the target in the base camera)

$\mathcal{A}, i_c$ (Set of α's, and current index of α)

Output:

${}^{*}\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1}$ (Desired image position of tip in base camera)

i_n (next index of α)

1: Obtain a constant curvature vector $\hat{\boldsymbol{\kappa}}(\alpha_{i_c})$ from (23)
2: Integrate curvatures to obtain the tip pose $\mathbf{X}^{\mathcal{O}}_{\mathcal{T}}$ from (20)
3: Find the estimated tip position with respect to the base camera's frame ${}^{*}\mathbf{x}^{\mathcal{G}}_{\mathcal{F}_1}$ from (24)
4: Find the desired image projection ${}^{*}\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1}$ from (25)
5: $i_n \leftarrow i_c + 1$

C. Estimating Desired Tip Position

In this section, the method used to estimate the desired tip position in the base camera, ${}^{*}\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1}$, is detailed. Algorithm 1 summarizes this method. When the target is not in the field view of the tip camera, the GVS system moves the tip of the SCA to a position where the target is expected to be in the tip camera's field of view. This is achieved with only the image space projections of the target, $\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_2}$, and the SCA's tip feature, $\mathbf{p}^{\mathcal{G}}_{\mathcal{F}_1}$, on the base camera (no 3D information is required). The main observation that is utilized is that when a feature is in the center of the tip camera's image, its position with respect to the camera's frame should be along the z-axis (i.e. $\mathbf{x}^{\mathcal{L}}_{\mathcal{F}_2} = \begin{bmatrix} 0 & 0 & \beta \end{bmatrix}^T$ for some $\beta > 0$). In other words, the following is true for a feature in the center of the tip camera's image

$$\begin{bmatrix} \mathbf{e}_x & \mathbf{e}_y \end{bmatrix}^T \mathbf{x}^{\mathcal{L}}_{\mathcal{F}_2} = \mathbf{0}$$
$$\mathbf{e}_z^T \mathbf{x}^{\mathcal{L}}_{\mathcal{F}_2} > 0, \tag{17}$$

where $\begin{bmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \end{bmatrix} = \mathbf{I}_3$, and the inequality constraint ensures that the target is in-front of the tip camera. The feature's position in the tip camera's frame, $\mathbf{x}^{\mathcal{L}}_{\mathcal{F}_2}$, is related to its position in the base camera's frame, $\mathbf{x}^{\mathcal{G}}_{\mathcal{F}_2}$, through the following relationship

$$\bar{\mathbf{x}}^{\mathcal{L}}_{\mathcal{F}_2} = \mathbf{X}^{\mathcal{L}}_{\mathcal{T}} \mathbf{X}^{\mathcal{O}^{-1}}_{\mathcal{T}} \mathbf{X}^{\mathcal{O}}_{\mathcal{G}} \bar{\mathbf{x}}^{\mathcal{G}}_{\mathcal{F}_2}, \tag{18}$$

where $\bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x}^T & 1 \end{bmatrix}^T$; the relative pose between the tip camera and the SCA's tip, $\mathbf{X}^{\mathcal{L}}_{\mathcal{T}}$, and between the SCA's base and the base camera, $\mathbf{X}^{\mathcal{O}}_{\mathcal{G}}$, are known; the relative pose between the SCA's base and its tip, $\mathbf{X}^{\mathcal{O}}_{\mathcal{T}}$, is unknown; and $\mathbf{x}^{\mathcal{G}}_{\mathcal{F}_2}$ is known up to a scale factor $\alpha = \|\mathbf{x}^{\mathcal{G}}_{\mathcal{F}_2}\|$,

$$\mathbf{x}^{\mathcal{G}}_{\mathcal{F}_2} = \alpha \mathbf{p}^{\mathcal{G}}_{\mathcal{F}_2}. \tag{19}$$

Since α is unknown (i.e. we do not have information about the distance of the target's feature from the base frame), we consider a set of possible α's, $\mathcal{A} := \{\alpha_i | \alpha_i \in [\alpha_{\min} \quad \alpha_{\max}], i =$

$1, \ldots, K\}$, where $\alpha_{\min}$ and $\alpha_{\max}$ are the minimum and maximum expected distances for the target and K is the number of samples. For each $\alpha \in \mathcal{A}$, we would like to find $\mathbf{X}_{\mathcal{T}}^{\mathcal{O}}$ that satisfies (17). However, since we do not have a model for the actual SCA, we consider a virtual SCA that has a constant curvature vector, $\boldsymbol{\kappa}$, throughout its length. By integrating (1) along its length, the virtual SCA's tip pose is given by

$$\tilde{\mathbf{X}}_{\mathcal{T}}^{\mathcal{O}}(\boldsymbol{\kappa}) = e^{L[\mathbf{K}]_\times}, \tag{20}$$

where $[\mathbf{K}]_\times$ is defined in (1), L is the length of the soft robot, $e^{L[\mathbf{K}]_\times}$ is the exponential map in $SE(3)$ defined by

$$e^{L[\mathbf{K}]_\times} := \mathbf{I}_4 + L[\mathbf{K}]_\times + \frac{1}{\theta^2}\left(1 - \sin(L\theta)\right)[\mathbf{K}]_\times^2 + \frac{1}{\theta^3}\left(L\theta - \sin(L\theta)\right)[\mathbf{K}]_\times^3, \tag{21}$$

and $\theta = ||\boldsymbol{\kappa}||$.

The feature's position with respect to the tip of this virtual SCA is obtained by substituting (19) and (20) into (18)

$$\bar{\tilde{\mathbf{x}}}_{\mathcal{F}_2}^{\mathcal{L}}(\boldsymbol{\kappa}; \alpha) = \mathbf{X}_{\mathcal{T}}^{\mathcal{L}} \tilde{\mathbf{X}}_{\mathcal{O}}^{\mathcal{T}}(\boldsymbol{\kappa}) \mathbf{X}_{\mathcal{G}}^{\mathcal{O}} \overline{\alpha \mathbf{p}}_{\mathcal{F}_2}^{\mathcal{G}}(\alpha). \tag{22}$$

To find the curvature of the virtual SCA that satisfies (17) for a given α, the following optimization routine is applied

$$\tilde{\boldsymbol{\kappa}}(\alpha) = \arg\min_{\boldsymbol{\kappa}} \left\| \begin{bmatrix} \mathbf{e}_x & \mathbf{e}_y \end{bmatrix}^T \tilde{\mathbf{x}}_{\mathcal{F}_2}^{\mathcal{L}}(\boldsymbol{\kappa}; \alpha) \right\|^2 \quad \text{subject to} \quad \mathbf{e}_z^T \tilde{\mathbf{x}}_{\mathcal{F}_2}^{\mathcal{L}}(\boldsymbol{\kappa}; \alpha) > 0, \tag{23}$$

where (20)–(22) are used to calculate $\tilde{\mathbf{x}}_{\mathcal{F}_2}^{\mathcal{L}}(\boldsymbol{\kappa}; \alpha)$.

Once the curvature of the virtual SCA is estimated for a specific α, the desired tip position in the base camera image is obtained by first finding the position of the virtual SCA's tip with respect to the base camera,

$$\bar{\tilde{\mathbf{x}}}_{\mathcal{F}_1}^{\mathcal{G}}(\alpha) = \mathbf{X}_{\mathcal{G}}^{\mathcal{O}^{-1}} \tilde{\mathbf{X}}_{\mathcal{T}}^{\mathcal{O}}(\tilde{\boldsymbol{\kappa}}(\alpha)) \bar{\mathbf{x}}_{\mathcal{F}_1}^{\mathcal{T}}, \tag{24}$$

where $\mathbf{x}_{\mathcal{F}_1}^{\mathcal{T}}$ is the position of the tip marker with respect to the SCA's tip, and then projecting it onto the base camera,

$${}^*\mathbf{p}_{\mathcal{F}_1}^{\mathcal{G}}(\alpha) = \mathbb{P}\left(\bar{\tilde{\mathbf{x}}}_{\mathcal{F}_1}^{\mathcal{G}}(\alpha)\right). \tag{25}$$

This is repeated for all $\alpha \in \mathcal{A}$ and a corresponding set of image projections is obtained,

$$\mathcal{P} := \left\{ {}^*\mathbf{p}_{\mathcal{F}_1}^{\mathcal{G}}(\alpha) | \alpha \in \mathcal{A} \right\}. \tag{26}$$

The goal of the GVS system is to move the tip of the SCA along this set to scan for the target in the tip camera. This is achieved by iterating between the points in $\mathcal{P}$ and choosing a single point as the desired tip position for the controller. To insure the SCA has minimum vibrations, a low value is chosen for the period to go from the first point to the last point in $\mathcal{P}$.

Although the actual SCA will not have a constant curvature, the proposed approach still works since the goal is to get the target somewhere within the tip camera's field view. Our experiments, as detailed in the next section, show that this model is sufficient even with the affect of external forces such as gravity and tip loads.

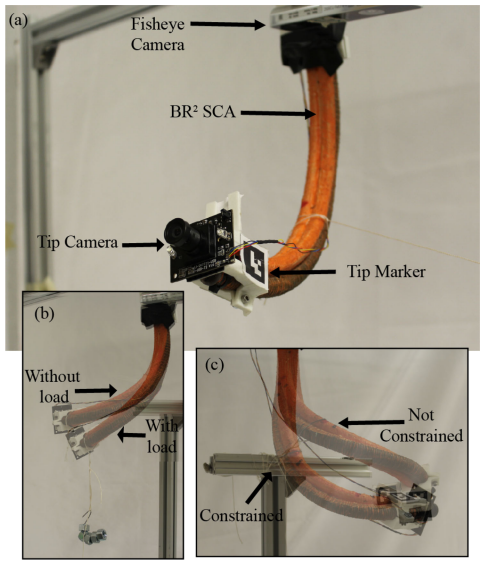

Fig. 4. Experimental setup used to validate the proposed method. The (a) main components of the system, the shape of the SCA when (b) load is applied to its tip, and (c) when it is constrained from its center, under the same pressure input.

IV. Results

A. Experimental Setup

To test the proposed method, an experimental rig shown in Fig. 4 is used. The SCA used is a BR^2 [32] that weighs 35 grams and is 30 cm long. This soft arm is composed of a parallel combination of three pneumatically actuated fiber reinforced actuators, where one is responsible for bending the SCA (B) and two for rotating (R^2) it about its length (only one is actuated at a given time to rotate it either clockwise or counterclockwise). Note that since this SCA has bending in only one direction and twisting along its length, the second component of the constant curvature $\boldsymbol{\kappa}$ is constrained to zero when solving (23) (i.e. $\kappa_y = 0$). The Jacobians where obtained by dividing the actuation space into a 4×5 grid with a step size of 8 psi (i.e. maximum values are 24 and 32 psi for the bending and rotating actuators, respectively). Since only one rotating actuator is active at a single time, two of these grids were obtained for each rotating actuator independently with the bending actuator.

This SCA is chosen because of its high dexterity and workspace (ability to spirally deform) despite having a compact parallel architecture. To actuate each section we have created a control board that drives three digital pressure regulators which correspond to the segments of our SCA. With this setup we are able to access a workspace suitable for conducting experiments between the LVS and GVS regimes. The SCA is suspended upside down with its base attached to a fixed structure. For our base camera, we mounted an Intel RealSense Tracking Camera T265 with the primary wide angle lens aligned with the base of the SCA. The tip camera consists of a USB Mini Camera Module, with a 1280×720 resolution, chosen for it's compactness and light weight. To obtain the spherical image projections, both cameras where calibrated using the method in [33]. On the tip of the soft robot, a marker composed of several AprilTags [34] was fixed to simplify the process of detecting the SCA's tip in the base camera. To accurately evaluate the proposed method throughout the soft robot's workspace, a $55\,\mathrm{cm} \times 55\,\mathrm{cm}$

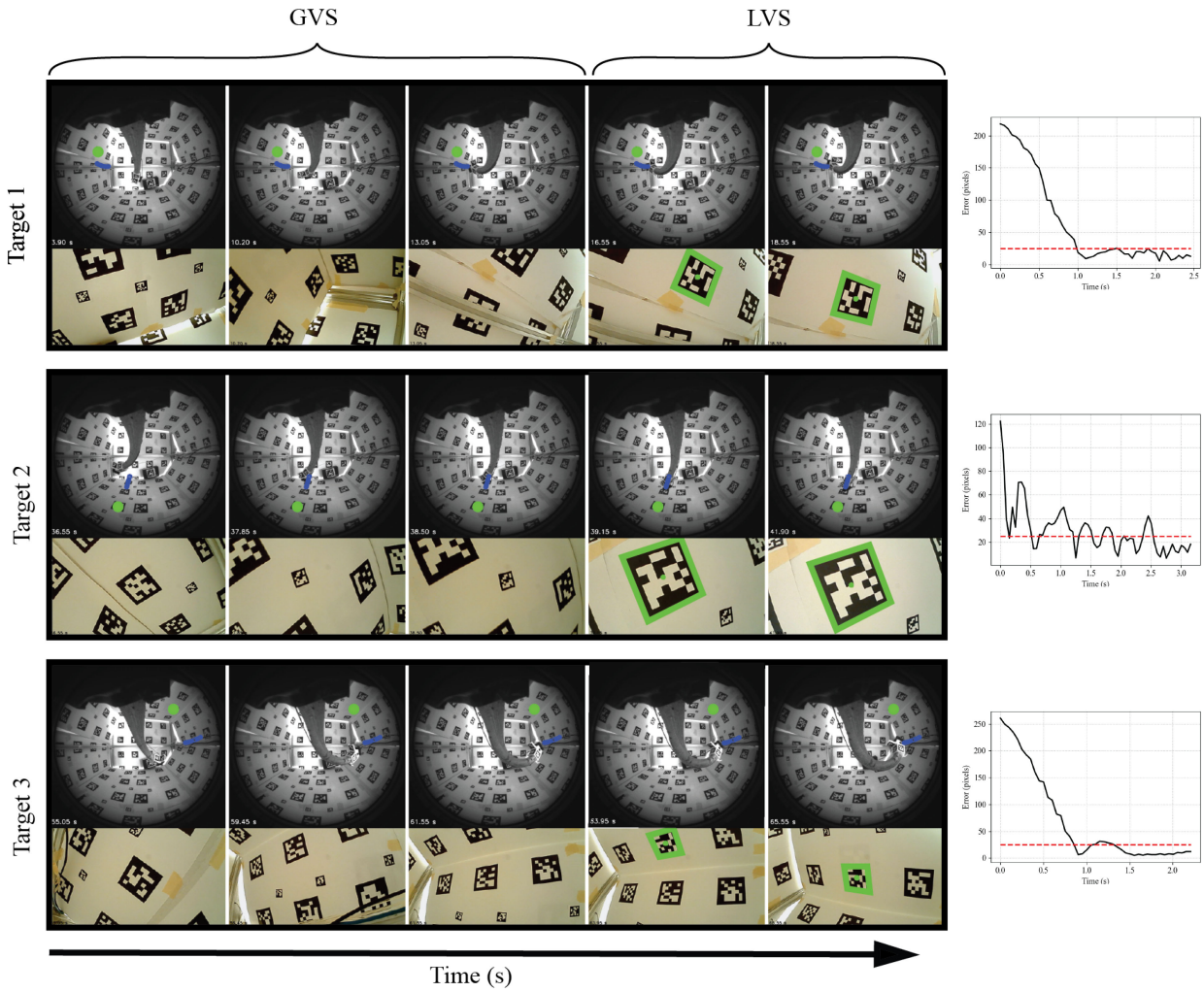

Fig. 5. Demonstration of the hybrid Visual servoing method for three distinct targets. The chosen target is highlighted with a green circle in the global images and a green square in the local images, and the blue curve represents the set of desired image projections in Equation (26). The plots on the right represent the norm pixel errors between the position of the target in the tip camera's frame and the center of the frame. The 25 pixel error tolerance is plotted in dotted red.

box with markers in its inside (that are considered targets) was placed around the SCA. Note that the proposed method does not require markers on the SCA's tip or on the targets, however the markers are used here to allow for a consistent evaluation across the SCA's workspace without requiring object detection and tracking. Equation (23) was solved using the sequential least square quadratic programming (SLSQP) optimizer in [35]. The detection frequency of the imaging system was 30 hz and the control frequency of the proposed controller was 10 hz.

B. Results[1]

1) Hybrid vs Without Disturbances: The first set of experiments, shown in Fig. 5, demonstrates the ability for the proposed scheme to utilize both the GVS and LVS control regimes to track a desired target without any disturbances. The target is selected in the global frame and tracked with GVS. Once it is viewed in the local frame, the target is tracked using LVS. At the beginning of the experiment a target marker, identified with a green circle in the global image of Fig. 5, is chosen in the view of the global camera. The blue curve in the image shows the expected position of the SCA tip so that the selected target is visible in the local camera's frame. The GVS control then sends the tip to the positions on the blue curve until the target is located. Once the tip camera has identified the target (fourth column in Fig. 5), the control scheme is switched to LVS which centers the target in the tip camera frame (as shown in the column 5 of Fig. 5 and the corresponding error plots). We set a 25 pixel error tolerance for the target to reach the center of the image. From the error plots, we observe that the second target has more vibrations than the other two case. This is due to the close proximity of the target to the camera compared to the other two targets (as seen in the size of the target in the last frame of the local camera). This experiment was repeated for a total of 30 targets throughout the workspace to evaluate the effectiveness of the estimated tip positions in getting the local camera to see the target. A success rate of 100% was achieved towards this goal. Fig. 7(a) shows the positions of these targets in the base camera.

2) Hybrid vs With Disturbances: In these set of experiments, we evaluate the robustness of the proposed hybrid VS method towards disturbances. Two types of disturbances are considered:

[1]A video presenting some of the results: https://youtu.be/a10x9BHHKxE.

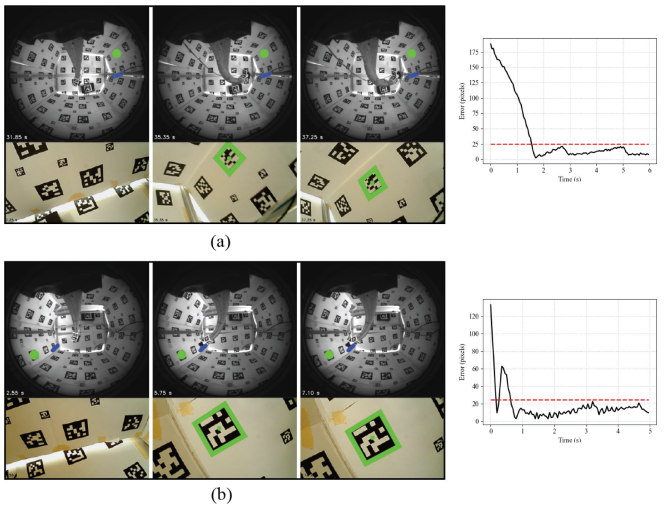

Fig. 6. Demonstration of the hybrid VS method when (a) a load of 35 grams is on the tip, and (b) the center of the SCA is constrained. The plots on the right represent the norm pixel errors between the position of the target in the tip camera's frame and the center of the frame. The 25 pixel error tolerance is plotted in dotted red.

a load of 35 grams (approximately equal to the SCA's weight) on the SCA's tip, as shown in Fig. 4(b), and a string that is tied to the center of the SCA that constrains the movement of its upper half, as shown in Fig. 4(c). Such disturbances might be encountered in applications that require manipulation or maneuvering in complex environments ridden with obstacles. Fig. 6 shows two successful scenarios for the hybrid system under these disturbances. For further evaluation, a set of 30 experiments were conducted for targets distributed throughout the SCA's workspace. The number of successful experiments were 25 and 24 for tip load and constrained cases, respectively. Fig. 7(b) and (c) show the positions of the targets in all the trials. The parts of the workspace that performed poorly were the regions that required extreme bending and twisting. In these cases, a relatively large load acting on the SCA's tip will cause significant changes in the SCA's deformations. Therefore, the offline Jacobians that are used in the control law become inaccurate, thus making it difficult for the controller to take the SCA's tip to the desired region. This was the case for the 5 unsuccessful cases in Fig. 7(b). When the center of the SCA is constrained with a string, the SCA workspace reduces and it cannot reach the part of the workspace where it can see the targets behind it. Therefore, the failure is due to the SCA's limited workspace. In the contrary, it was surprising that the method works in most cases even with the first half of the SCA not moving. The parts of the workspace that performed poorly were the regions behind the SCA for both cases. For the tip camera to see these regions, the SCA has to undergo a significant amount of twisting. When under tip load, significant deformation of the SCA's shape is observed. Thus when the tip reaches the desired image position, the target is not observed in the tip camera's view. On the other hand, when the upper half of the SCA is constrained, not enough twisting can be achieved to observe the regions it. In general, when major deformations in the SCA's shape are caused by the disturbances, the desired tip positions that are based on the virtual constant curvature SCA do not lead to desirable results.

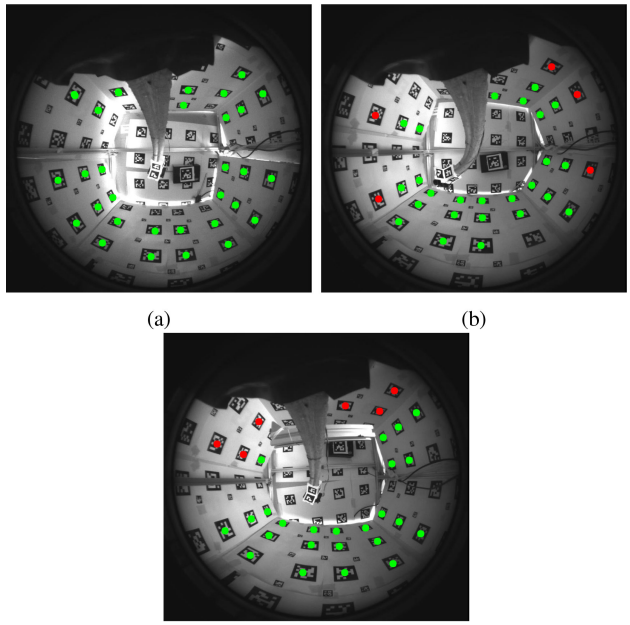

Fig. 7. The targets that correspond to the successful (green circle) and unsuccessful (red circle) cases for (a) the disturbance free case, (b) tip load disturbance case, and (c) constrained movement disturbance case.

V. Conclusion

Although accurate control for soft continuum arms (SCAs) can be achieved with eye-to-hand or eye-in-hand visual servoing, each has its own limitations. The former has a global view and can place the SCA's tip anywhere within its workspace, while the latter can achieve better positioning for exploration and manipulation tasks. In this letter, a method for controlling SCAs through a hybrid eye-to-hand/eye-in-hand visual servoing scheme is presented. The advantage of this hybrid system is the ability to track a target with the soft robot's tip camera even if the target is not in the camera's field-of-view. The crux of the method involves an optimization-based estimation of feasible positions for the SCA tip that has maximum probability of observing the target from the eye-in-hand camera. Results show the effectiveness of the proposed method, which was able to track targets throughout the soft robot's workspace even with the presence of disturbances. In cases where the SCA is under major disturbances, the target might not be seen by the tip camera. This could be resolved by investigating other models for the virtual SCA used to estimate the best tip positions in the image.

The SCA used in this work has only two degrees-of-freedom, however it can be easily extended for applications which require soft robot's with more degrees-of-freedom. This work can also be extended to robotic arms that have both rigid and soft links, such as the arm presented in [36]. Furthermore, a single camera on the soft robot's base may not be sufficient to observe the entire workspace for some applications. Therefore, extending this work to account for multiple global cameras would be desirable. Since it is possible to structure the soft robot as needed, having a marker on its tip should be practical for most applications. However, for applications where this is not possible, the pose of the tip camera with respect to the base camera can be obtained

by finding corresponding image features in both cameras and solving a bundle adjustment problem.

The work presented in this letter will be used to build autonomous capabilities for robotic systems with SCA's that could be applied to agriculture [36] and health applications. Therefore, we will investigate the performance of the hybrid visual servoing system under unstructured environments that require the flexibility of SCAs.

REFERENCES

[1] A. Vandini, A. Salerno, C. J. Payne, and G.-Z. Yang, "Vision-based motion control of a flexible robot for surgical applications," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2014, pp. 6205–6211.

[2] K. Wu, L. Wu, and H. Ren, "An image based targeting method to guide a tentacle-like curvilinear concentric tube robot," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2014, pp. 386–391.

[3] M. C. Yip and D. B. Camarillo, "Model-less feedback control of continuum manipulators in constrained environments," *IEEE Trans. Robot.*, vol. 30, no. 4, pp. 880–889, Aug. 2014.

[4] H. Wang, B. Yang, Y. Liu, W. Chen, X. Liang, and R. Pfeifer, "Visual servoing of soft robot manipulator in constrained environments with an adaptive controller," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 1, pp. 41–50, Feb. 2017.

[5] Y. Lu, C. Zhang, S. Song, and M. Q.-H. Meng, "Precise motion control of concentric-tube robot based on visual servoing," in *Proc. IEEE Int. Conf. Inf. Automat.*, 2017, pp. 299–304.

[6] J. D. Greer, T. K. Morimoto, A. M. Okamura, and E. W. Hawkes, "Series pneumatic artificial muscles (sPAMs) and application to a soft continuum robot," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2017, pp. 5503–5510.

[7] Z. Zhang, T. M. Bieze, J. Dequidt, A. Kruszewski, and C. Duriez, "Visual servoing control of soft robots based on finite element model," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 2895–2901.

[8] A. V. Kudryavtsev et al., "Eye-in-hand visual servoing of concentric tube robots," *IEEE Robot. Automat. Lett.*, vol. 3, no. 3, pp. 2315–2321, Jul. 2018.

[9] H. Ru, J. Huang, W. Chen, C. Xiong, and J. Wang, "Design and control of a soft bending pneumatic actuator based on visual feedback," in *Proc. IEEE 13th World Congr. Intell. Control Automat.*, 2018, pp. 26–31.

[10] F. Xu, H. Wang, W. Chen, and J. Wang, "Adaptive visual servoing control for an underwater soft robot," *Assem. Automat.*, vol. 38, no. 5, pp. 669–677, 2018.

[11] J. D. Greer, T. K. Morimoto, A. M. Okamura, and E. W. Hawkes, "A soft, steerable continuum robot that grows via tip extension," *Soft Robot.*, vol. 6, no. 1, pp. 95–108, 2019.

[12] G. Fang et al., "Vision-based online learning kinematic control for soft robots using local gaussian process regression," *IEEE Robot. Automat. Lett.*, vol. 4, no. 2, pp. 1194–1201, Apr. 2019.

[13] M. Verghese, F. Richter, A. Gunn, P. Weissbrod, and M. Yip, "Model-free visual control for continuum robot manipulators via orientation adaptation," in *Proc. Int. Symp. Robot. Res.*, pp. 959–970, 2019. [Online]. Available: http://arxiv.org/abs/1909.00450

[14] F. Xu, H. Wang, J. Wang, K. W. S. Au, and W. Chen, "Underwater dynamic visual servoing for a soft robot arm with online distortion correction," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 3, pp. 979–989, Jun. 2019.

[15] X. Wang et al., "Eye-in-hand visual servoing enhanced with sparse strain measurement for soft continuum robots," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 2161–2168, Apr. 2020.

[16] J. Lai, K. Huang, B. Lu, and H. K. Chu, "Toward vision-based adaptive configuring of a bidirectional two-segment soft continuum manipulator," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics*, 2020, pp. 934–939.

[17] S. Kamtikar, S. Marri, B. Walt, N. K. Uppalapati, G. Krishnan, and G. Chowdhary, "Visual servoing for pose control of soft continuum arm in a structured environment," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 5504–5511, Apr. 2022. [Online]. Available: https://arxiv.org/abs/2202.05200

[18] F. Chaumette and S. Hutchinson, "Visual servo control. I. basic approaches," *IEEE Robot. Automat. Mag.*, vol. 13, no. 4, pp. 82–90, Dec. 2006.

[19] H. Wang, W. Chen, X. Yu, T. Deng, X. Wang, and R. Pfeifer, "Visual servo control of cable-driven soft robotic manipulator," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 57–62.

[20] R. J. Webster, J. P. Swensen, J. M. Romano, and N. J. Cowan, "Closed-form differential kinematics for concentric-tube continuum robots with application to visual servoing," *Springer Tracts Adv. Robot.*, vol. 54, pp. 485–494, 2009.

[21] V. K. Chitrakaran, A. Behal, D. M. Dawson, and I. D. Walker, "Setpoint regulation of continuum robots using a fixed camera," *Robotica*, vol. 25, no. 5, pp. 581–586, 2007.

[22] G. Flandin, F. Chaumette, and E. Marchand, "Eye-in-hand/eye-to-hand cooperation for visual servoing," in *Proc. IEEE Int. Conf. Robot. Automat. Symposia*, 2000, vol. 3, pp. 2741–2746. [Online]. Available: https://hal.inria.fr/inria-00352159

[23] M. Elena, M. Cristiano, F. Damiano, and M. Bonfe, "Variable structure PID controller for cooperative eye-in-hand/eye-to-hand visual servoing," in *Proc. IEEE Conf. Control Appl.*, 2003, pp. 989–994.

[24] V. Lippiello, B. Siciliano, and L. Villani, "Position-based visual servoing in industrial multirobot cells using a hybrid camera configuration," *IEEE Trans. Robot.*, vol. 23, no. 1, pp. 73–86, Feb. 2007.

[25] O. Kermorgant and F. Chaumette, "Multi-sensor data fusion in sensor-based control: Application to multi-camera visual servoing," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2011, pp. 4518–4523.

[26] C. Dune, E. Marchand, and C. Leroux, "One click focus with eye-in-hand/eye-to-hand cooperation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2007, pp. 2471–2476.

[27] H. Cuevas-Velasquez, N. Li, R. Tylecek, M. Saval-Calvo, and R. B. Fisher, "Hybrid multi-camera visual servoing to moving target," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 1132–1137.

[28] Z. Zhang, B. Rosa, O. Caravaca-Mora, P. Zanne, M. J. Gora, and F. Nageotte, "Image-guided control of an endoscopic robot for OCT path scanning," *IEEE Robot. Automat. Lett.*, vol. 6, no. 3, pp. 5881–5888, Jul. 2021.

[29] K. Lynch and F. Park, *Modern Robotics: Mechanics, Planning, and Control*. Cambridge, MA, USA: Cambridge Univ. Press, 2017.

[30] T. Hamel and R. Mahony, "Visual servoing of an under-actuated dynamic rigid-body system: An image-based approach," *IEEE Trans. Robot. Automat.*, vol. 18, no. 2, pp. 187–198, Apr. 2002.

[31] R. T. Fomena and F. Chaumette, "Visual servoing from spheres using a spherical projection model," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2007, pp. 2080–2085.

[32] N. K. Uppalapati and G. Krishnan, "Design and modeling of soft continuum manipulators using parallel asymmetric combination of fiber-reinforced elastomers," *J. Mechanisms Robot.*, vol. 13, no. 1, Feb. 2021, doi: 10.1115/1.4048223.

[33] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2006, pp. 5695–5701.

[34] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 4193–4198.

[35] J. Nocedal and S. Wright, "Sequential Quadratic Programming," in *Proc. Springer Ser. Operations Res. Financial Eng.*, 2006, pp. 529–562.

[36] N. K. Uppalapati and G. Krishnan, "VaLeNS: Design of a novel variable length nested soft arm," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 1135–1142, Apr. 2020.