

## Article

# Visual Tracking Control of Cable-Driven Hyper-Redundant Snake-Like Manipulator

Qisong Zhou <sup>1</sup>, Jianzhong Tang <sup>1</sup>, Yong Nie <sup>1,\*</sup>, Zheng Chen <sup>2</sup>  and Long Qin <sup>2</sup> 

<sup>1</sup> School of the State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China; qszhou@zju.edu.cn (Q.Z.); jztang@zju.edu.cn (J.T.)

<sup>2</sup> School of Ocean College, Zhejiang University, Zhoushan 316021, China; zheng\_chen@zju.edu.cn (Z.C.); long\_qin@zju.edu.cn (L.Q.)

\* Correspondence: ynie@zju.edu.cn

**Abstract:** The cable-driven hyper-redundant snake-like manipulator (CHSM) inspired by the biomimetic structure of vertebrate muscles and tendons, which consists of numerous joint units connected adjacently driven by elastic materials with hyper-redundant DOF, performs flexible kinematic skills and competitive compound capability under complicated working circumstances. Nevertheless, the drawback of lacking the ability to perceive the environment to perform intelligently in complex scenarios leaves a lot to be improved, which is the original intention to introduce visual tracking feedback acting as an instructor. In this paper, a cable-driven snake-like robotic arm combined with a visual tracking technique is introduced. A visual tracking approach based on dual correlation filter is designed to guide the CHSM in detecting the target and tracing after its trajectory. Specifically, it contains an adaptive optimization for the scale variation of the tracking target via pyramid sampling. For the CHSM, an explicit kinematics model is derived from its specific geometry relationships and followed by a simplification for the inverse kinematics based on some assumption or limitation. A control scheme is brought up to combine the kinematics with visual tracking via the processing tracking errors. The experimental results with a practical prototype validate the availability of the proposed compound control method with the derived kinematics model.

**Keywords:** cable-driven; visual tracking; kinematics; correlation filter



**Citation:** Zhou, Q.; Tang, J.; Nie, Y.; Chen, Z.; Qin, L. Visual Tracking Control of Cable-Driven Hyper-Redundant Snake-Like Manipulator. *Appl. Sci.* **2021**, *11*, 6224. <https://doi.org/10.3390/app11136224>

Received: 2 June 2021

Accepted: 28 June 2021

Published: 5 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Robotic arms have been widely developed in the industrial manufactured application and gradually reduced human participation since the first sophisticated robotic arm was designed by da Vinci in 1495. There is a definite trend in the manufacture of robotic arms toward more dexterous devices, more degrees of freedom (DOF), and capabilities beyond the human arm [1]. The redundant DOF tends to perform more competently under some complicated scenarios such as narrow deep cavities or underwater pipe networks, especially with the discrete sets of joints imitating the structure of muscles and tendons with similar functions [2]. High flexibility of avoiding obstacles, good loading capacity, and easy maintenance should be taken into consideration for the structural design of robotic arms to satisfy the requirement of tasks under complicated working circumstances. A snake-like continuum manipulator with redundant DOF inspired by the biomimetic structure of vertebrate limbs has attracted increasing attention of researchers but limited to the drawbacks such as slight load capacity, imprecisely controlled and limited measurements, alleviated after the cable-driven technique introduced [3].

Cable-driven hyper-redundant snake-like manipulator (CHSM) which has numerous joint units connected adjacently driven by elastic cables with hyper-redundant DOF can achieve the desired positions with different postures [4], which means the flexible motion and good obstacle avoidance capability under complicated working circumstances [5]. The biggest difference from traditional rigid manipulators composed of actuated joints

is that the CHSM has no driving actuator located in joints, which leads to the rather intractable issue of inverse kinematics [6,7]. Zhao Zhang et al. [8] proposed an approach based on the Product-Of-Exponential (POE) formula to model instantaneous kinematics for a cable-driven snake-like manipulator, computing the numerical solution of inverse kinematics via the Newton-Raphson method. Andres Martin et al. [9] involved the cyclic coordinate descent method named natural-CCD to simply achieve the best result among the innumerable solutions for the inverse kinematics, mapping the hyper-redundant DOF into three spatial dimensions. Since the snake-like manipulator has different kinematic and dynamic behaviors from traditional rigid robots, it is worthy to develop a visual servo system that can improve stability and precision for controlling [10]. Furthermore, the computer vision techniques endow the traditional robotic arms with the ability to perceive the environment to perform intelligently in complex scenarios. Khatib et al. [11] achieved great success in underwater tasks, in which the computer vision techniques play a critical role as the bridge between robotic arms and human cognitive guidance. Considering the advantage of cable-driven snake-like manipulator, visual tracking is recommended as an assistant target-oriented method for guiding the CHSM to lock on mission objectives under tasks such as salvage or rescue, especially in the scenarios such as narrow and tortuous caverns or pipelines.

Most research achievements in the visual tracking field have blossomed over the past few decades, divided into two branches called generative model [12–14] and discriminative model [15,16]. The generative approaches establish models or templates with respect to the area of target in the current frame, aim at describing the targets' appearance and find the most similar areas in the next frame as the estimated new position (e.g., Kalman filtering [17], Particle filtering [18], Mean-shift [19], etc.). The discriminative approaches depend on the extracted features combined with learning evolution, discriminate between the target and background in the current frame as a positive sample and negative sample respectively, and aim at using a classifier training by machine learning algorithm such as SVM to estimate the optimum areas in the subsequent frames. In recent years approaches based on correlation filter stand out among a series of discriminative approaches in many competitions such as VOT, rank top and surpass the other branches with a great superiority not only in accuracy but also on FPS. Correlation is used to measure the similarity between two signals in signal processing subject, involved in visual tracking by Bolme et al. [20] and achieved great success. After a filter with extracted features constructed at the beginning of tracking, the commonly used classifier is replaced by a cross-correlation score which will be computed between the filter and a subsequent frame. The target's position can be predicted locating on the maximum of the response score as the similar characteristic in signal processing. Henriques [21] introduced the circulant matrices which implement a great simplification for the matrices operation in complex field due to its diagonalizable characteristics by Fourier matrices. The kernel trick commonly used in SVM was also introduced in [21], enriched the extracted features with great diversification, and improved performance further. Danelljan Martin et al. [22] and Li et al. [23] focused on the variation of target's scale in tracking, greatly ameliorated the drifting issue caused by the variation of scale. Danelljan Martin et al. [24] and Kiani Galoogahi et al. [25] expanded the ratio of detecting area to filter area and punished the filter coefficient around the border of the tracking box which makes effort to alleviate the boundary effects. Chao Ma et al. [26] and Wang [27] deeply studied on long-term tracking direction through a confidence level evaluated on correlation. They constructed a third filter besides translation and scale used to assess the confidence degree which will be activated to correct the filter reloading when the confidence degree descends below a threshold.

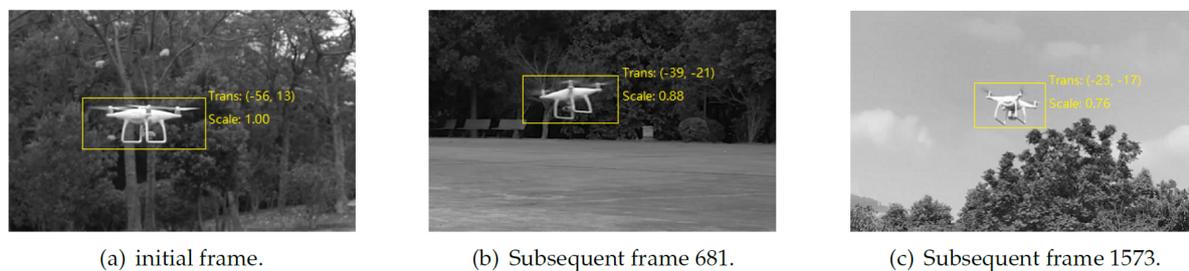
In this work, the main contribution is described as follows. We develop a prototype of CHSM for complicated tasks such as salvage or rescue, which is controlled based on a visual servo framework and consists of a separating structure between the power subsystem and motion subsystem. A simplified forward and inverse kinematics model under the visual servo control is derived under some assumptions to avoid the time-consuming

computation of matrix operation. A visual tracking algorithm based on dual correlation filter (DCF) is presented and customized for visual tracking control of the CHSM.

The remainder of this paper is organized as follows. The implementation of the tracking component is proposed in Section 2. In Section 3, the overall structure of the CHSM and kinematics model is introduced, meanwhile, the control method is proposed and analyzed. In Section 4, an experiment is carried out and the experimental results validate the availability of the proposed compound control method with the derived kinematics model. Finally, a conclusion is presented in Section 5.

## 2. Implementation of Tracking Component

In general, visual tracking means distinguishing the target between foreground objects and background environment in the camera's field of view (FOV) while the target is moving unrestrictedly with latent variance both in appearance and scale, additionally orienting the camera towards the target. As shown in Figure 1, the moving target is accompanied by an attached bounding box, which highlights the target without many gaps between border and target. The implementation in this paper is derived from DCF which is a classical discriminative method in correlation filter field. At the beginning of tracking, the tracking target is highlighted by a bounding box in an artificial marking manner to train and generate an optimum template called filter which can describe explicit or implicit features of the target as much as possible, as shown in Figure 1a. A translation filter detects the areas of the target in the subsequent frames, marks with a bounding box in a suitable scale, and records the center of bounding box as the estimated result. In addition, a scale filter matches the target at the estimated position and adjusts the size of bounding box with gaps as few as possible, records as a size ratio to the original box in the initial frame. After the estimates both in translation and scale, the filters will learn some features according to the current appearance of the target and update itself to enhance the robustness. To ensure that the target always locates in the camera's FOV, the control of orientation is combined with the kinematics of CHSM which will be stated in Section 3.4.



**Figure 1.** A visual tracking application for the flight of an unmanned aerial vehicle.

### 2.1. Translation Filter

This filter is used to estimate the instantaneous target's position in the subsequent frames. In the training processing, some fragments are sampled around the original sample (the segment in the bounding box) to construct the training sample dataset. The linear bridge regression is used to find the optimum filter  $w$  which minimizes the squared error over samples  $x_i$  and their regression targets  $y_i$  stated as Equation (1). Two-dimensional Gaussian Distribution is considered to be the ideal hypothesis for  $y_i$  and the peak of Gaussian Distribution is placed in the center of the bounding box which can be presented as the position of the target. In other words, the original sample should be filtered by the optimum filter to formulate a two-dimensional Gaussian Distribution.

$$\min_w \sum_i (w^T x_i - y_i)^2 + \lambda \|w\|^2 \quad (1)$$

$\lambda$  is a regularization parameter which controls overfitting, the same characteristic as in the SVM. The minimizer has a closed-form solution. Given that Discrete Fourier Transform (DFT) is involved in the calculation of correlation for accelerating, the linear bridge regression should be solved in complex field. The generic solution is shown in Equation (2).

$$W = (X^H X + \lambda I)^{-1} X^H Y \quad (2)$$

$X$  and  $Y$  represent the training dataset and regression dataset, respectively.  $X^H$  is the Hermitian transpose, where  $X^H = (X^*)^T$  and  $X^*$  is the complex-conjugate of  $X$ . Circulant matrices are considered to be an excellent implementation in avoiding the extremely high expense for matrices' time-consuming calculation, due to its diagonalizable characteristics by Fourier matrices. In that way, the matrices' Matmul product can be converted into Hadamard product, under the precondition that the training dataset is constructed with cyclically shifted samples around the original sample. The derivation procedure is showed in Appendix A. The derivation is shown in Equation (3):

$$\hat{W} = \frac{\hat{x} \odot \hat{y}}{\hat{x}^* \odot \hat{x} + \lambda} \quad (3)$$

$\hat{x}$  represents the DFT of the original sample. In this way, the time complexity is reduced from  $O(n^3)$  to  $O(n \log(n))$ . KCF (kernel correlation filter) propelled Equation (3) forward by introduced the kernel trick which can map the sample space into a high-dimension and non-linear feature space. Kernel trick plays a critical role in SVM and achieves an effective promotion in KCF as well. However, this trick is not adopted in our implementation by contrasting the insufficient improvement of performance with the accompanied expensively time-consuming increase, considering that the FPS of detection should be guaranteed primarily.

## 2.2. Multi-Channel Features

Pixel data are the original form of  $x$  as shown in Equation (3). However, the DCF approach has recently been extended to multidimensional feature representations based on various feature operators for several applications [22]. It means that  $x$  consists of some  $d$ -dimensional feature vector  $f(n) \in R^d$  shaped as a rectangular, and the filter  $w$  also has a third dimension  $d$ . Therefore, Equation (3) is modified by summing over all the channels in the Fourier domain:

$$\hat{w}_l = \frac{\hat{x}_l \odot \hat{y}}{\sum_{k=1}^d \hat{x}_k^* \odot \hat{x}_k + \lambda} \quad l = 1, 2, 3 \dots d \quad (4)$$

we can calculate the Hadamard product separately for every feature channel and concatenate them by channels to obtain a 3-dimensional filter. As the literature [28] did, we consider two widely used features in visual tasks besides the grayscale pixel of the original sample and select one suited for the working circumstances and the images' quality in the experimental validation stage. HOG (Histogram of Gradient) extracts the gradient information from a region of pixels and counts the discrete orientation to form the histogram. HOG has been confirmed that it is sensitive to the variation of appearance between target and background [29]. CN (Color-Naming) is a perspective space which abstracts the color attributes of objects but is more similar to human sense than RGB space [28]. HOG takes more attention to the edge of objects and CN emphasis on the color information.

## 2.3. Adaptive Scale

A robust tracking algorithm should have the ability to respond to the change of target's size in the pixel space. A standard approach usually applies a tracker at multiple resolutions. We reference the principles as stated in the literature [22] to figure out the changing scale of the target. When the translation filter (obtained as above described) locks the instantaneous position of the target in a new frame then several patches will be sampled

by different resolutions centered around the new position. The patches are cropped for each size in  $S \in [s_1, s_2, \dots, s_n]$  in the manner of a scale pyramid. All patches are filtered by the scale filter to sift out the most compatible resolution for the target's size at that moment. The scale filter has the same form of Equation (4) and just flatten the multi-channel features into a 1-dimensional vector. The scale filter finds the optimal solution with the highest correlation response score (introduced in Section 2.4) among all patches.

#### 2.4. Tracking by Detection

At first frame when a rectangular region of an object is selected as the target that will be tracked in the next frame sequence, we use Equation (4) to obtain the optimal correlation filter of the target both for translation and scale. In a subsequent frame, the translation filter is applied on  $z_t$  (called test sample) centered around the target's position inherited from the latest previous frame.  $z_t$  is processed similarly to the training sample  $x$ . (feature representation, pre-processing, cyclic sampling). Then we use:

$$\hat{y}_t = \sum_{l=1}^d \hat{w}_l^T \odot \hat{z}_t \quad (5)$$

to compute the DFT of correlation scores  $y_t$  in the Fourier domain. The location of the maximum value in the correlation score can be regarded as the target's new position, due to the similar characteristic of correlation in signal processing. Around the new target's position, a scale pyramid sampling is applied for the scale detection (as stated in Section 2.3). Like the translation filter, we calculate the scale correlation response scores for all the patches:

$$\hat{y}_{st} = \sum_{l=1}^{d_s} \hat{w}_s^T \odot \hat{z}_{st} \quad (6)$$

The most compatible scale is derived by finding the maximum  $y_{st}$  among all the  $z_{st}$ .

#### 2.5. Learning and Update

Although the optimal filter is obtained at the original frame, it is necessary to consider the robustness in the following time instances due to the latent variation of the target's appearance and size. Usually, this can be achieved by weighted averaging the filter of all training samples but in our case, it is more reasonable to update the correlation filter at the new target position and scale after detection. An adjustment is made for the classical machine learning algorithm  $w_t = (1 - \eta)w_{t-1} + \eta w_t$  where  $\eta$  is the learning rate and  $t$  represents the index of tracking view frame sequence:

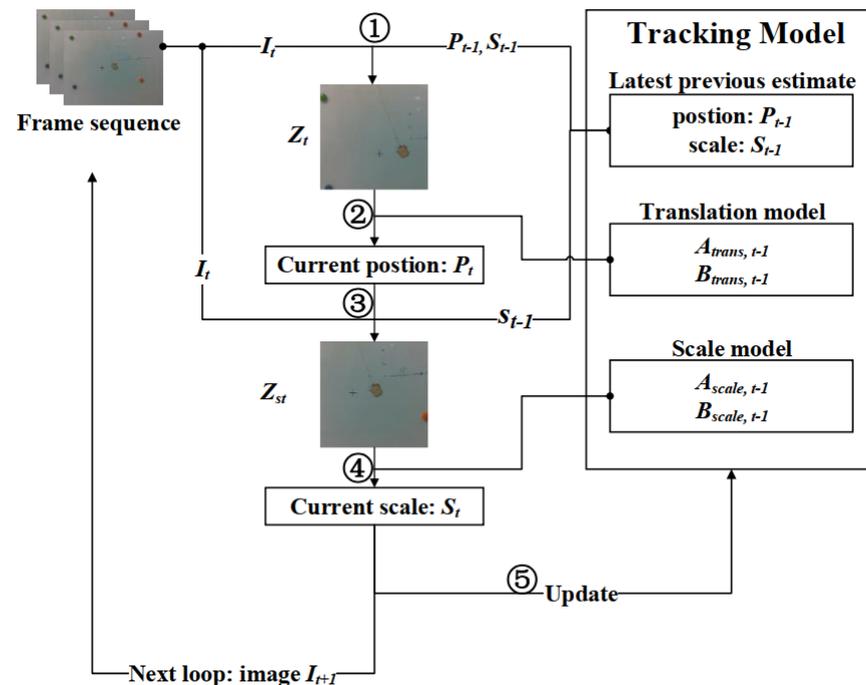
$$\hat{w}_t^l = \frac{\hat{x}_t^l \odot \hat{y}}{\sum_{k=1}^d (\hat{x}_t^k)^* \odot \hat{x}_t^k + \lambda} = \frac{A_t^l}{B_t + \lambda} \quad l = 1, 2, 3 \dots d \quad (7)$$

$$A_t^l = (1 - \eta)A_{t-1}^l + \eta \hat{x}_t^l \odot \hat{y} \quad l = 1, 2, 3 \dots d \quad (8)$$

$$B_t = (1 - \eta)B_{t-1} + \eta \sum_{k=1}^d (\hat{x}_t^k)^* \odot \hat{x}_t^k \quad l = 1, 2, 3 \dots d \quad (9)$$

Equations (8) and (9) are used to update the correlation filter both for the translation and scale.

Figure 2 has shown the tracking processing in our implementation.

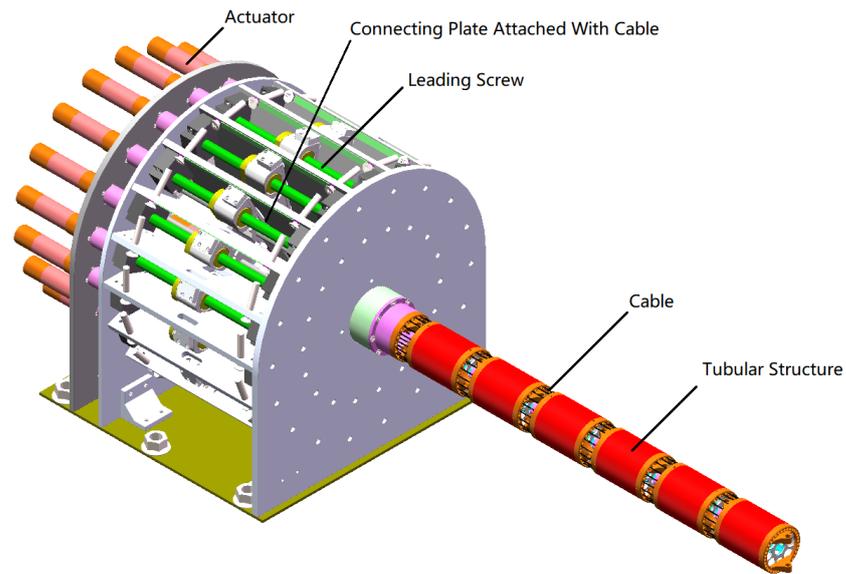


**Figure 2.** Illustration for the flow graph: ① Extract sample  $Z_t$  from  $I_t$  at previous position  $P_{t-1}$  and scale  $S_{t-1}$  with multi-channel features; ② Compute translation correlation scores  $y_{trans,t}$  according to Equation (5) and find the current estimated target position  $P_t$  where the maximum located in  $y_{trans,t}$ ; ③ Extract sample  $Z_{st}$  from  $I_t$  at current position  $p_t$  and scale  $S_{t-1}$ ; ④ Compute scale correlation scores  $y_{scale,t}$  according to Equation (6) and find the current estimated target scale  $S_t$  where the maximum located in  $y_{scale,t}$ ; ⑤ Update translation model and scale model according to Equations (7)–(9), then save the current position and scale as the latest previous estimate.

### 3. Kinematics Modeling and Control Methodology

#### 3.1. Overall Structure Design of CHSM

High flexibility of avoiding obstacles, good loading capacity, and easy maintenance should be taken into consideration for the structural design of snake-like manipulators to satisfy the requirement of tasks under complicated working circumstances. The proposed overall structure design of CHSM is presented in Figure 3. A separating structure is adopted in dividing the power and motion subsystem on purpose protecting the power subsystem away from harsh working circumstances but remaining the motion subsystem deploying its task normally. The motion subsystem is composed of repeated tubular structures connected to a universal joint by two fixed endplates as shown in Figure 4. Driving cables from the power subsystem get through a series of piercing holes which are uniformly distributed along the circumferential periphery in the shell of the tubular structure. Three cables are attached on the rear endplate, equally spaced at  $120^\circ$  and used to drive this tubular unit meanwhile the others are maintained piercing to serving for followed tubular units respectively as shown in Figure 5. In the power subsystem, the cables are attached to the sliders which are mounted on the screw nut seat of the leading screw and drive to control the pose of each tubular by rotating upon the universal joint with 2-DOF. The motion subsystem is assembled flexibly to conveniently extend the number of joint units according to diverse task demands.

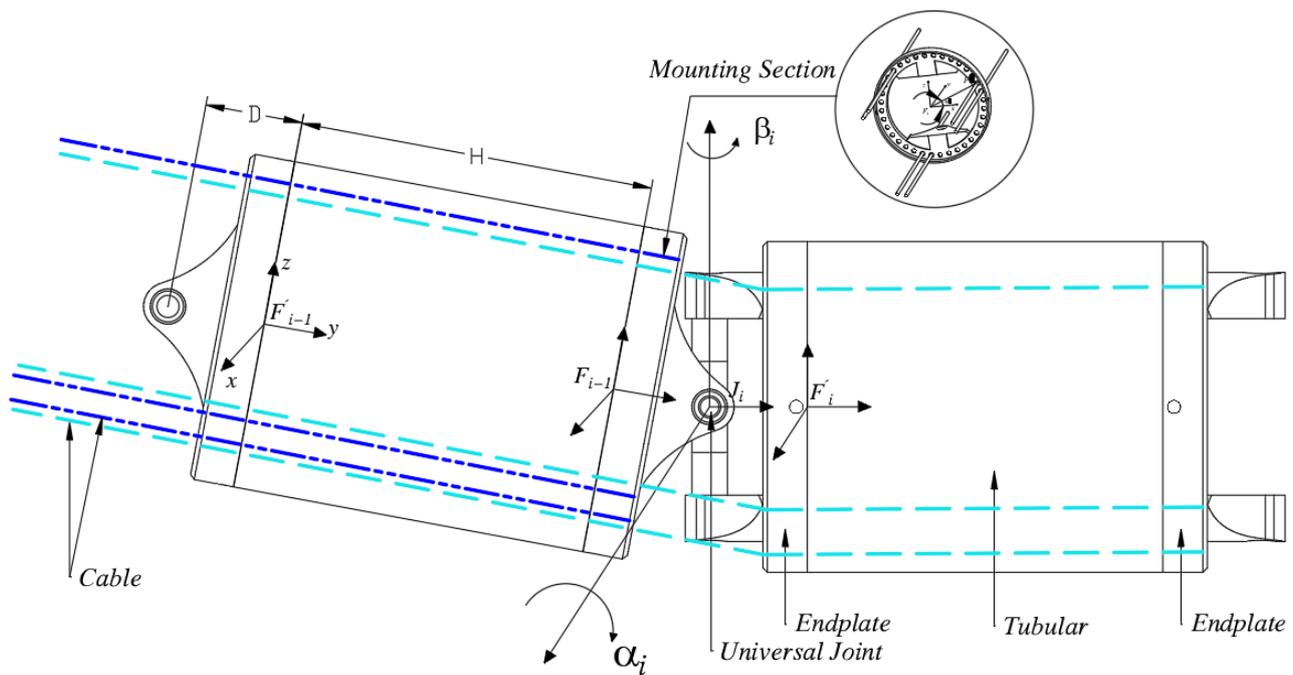


**Figure 3.** Overall structure design of cable-driven hyper-redundant snake-like manipulator (CHSM).

### 3.2. Forward Kinematics Analysis

the cable-joint kinematics can be derived from the established geometric model presented in Figure 4. The Frame  $F_i$  and  $F'_i$  are fixed on the center of where rear and proximal endplate concatenates the  $i$ -th tubular structure whose axial direction is parallel to the Y-axis. The universal joints in the geometric model can be mounted in an odd or even manner but do not influence the formulation. The cross-section for cable mounting is illustrated in Figure 5.

The main purpose of forward kinematics is to infer the end-effector's position based on the given joint angles  $[\alpha_i, \beta_i](i = 1, 2, \dots, N)$  but for CHSM the mapping transformation from unique and defined cable length  $L_{i,j}(j = 1, 2, 3)$  to the joint angles is critical and principal.



**Figure 4.** A geometric model between two tubular structures.

### 3.2.1. Coordinate Transformation Matrix between Adjacent Tubular

the pose transformation matrix of Frame  $F_{i-1}$  with reference to Frame  $F'_i$  can be formulated as:

$${}^i_{i-1}T = Transl(0, D, 0)RotZ(\beta_i)RotX(\alpha_i)Transl(0, D, 0) \tag{10}$$

where  $Transl(0, D, 0)$  represents the translation function with respect to Y-axis by shifting displacement  $D$ .  $RotZ(\beta_i)$  and  $RotX(\alpha_i)$  represent the rotation function regarding Z-axis and X-axis,  $\alpha_i$  and  $\beta_i$  denote respectively the corresponding counterclockwise rotation angle around axis. Equation (10) can be rewritten as homogeneous matrix expression:

$${}^i_{i-1}T = \begin{bmatrix} \cos \beta_i & -\cos \alpha_i \sin \beta_i & \sin \alpha_i \sin \beta_i & -D \cos \alpha_i \sin \beta_i \\ \sin \beta_i & \cos \alpha_i \cos \beta_i & -\cos \beta_i \sin \alpha_i & D + D \cos \alpha_i \cos \beta_i \\ 0 & \sin \alpha_i & \cos \alpha_i & D \sin \alpha_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{11}$$

### 3.2.2. Mapping Relation between Cable and Joint

As shown in Figure 5,  $P_{i,j}$  represents the mounted point of  $j$ -th ( $j = 1, 2, 3$ ) cable on the rear endplate in the Frame  $F_i$  and the counterclockwise angle  $\varphi_{i,j}$  can be derived as Equations (12) and (13) where  $M$  represents the total number of the holes in an endplate.

$$p_{i,j} = [R \times \cos \varphi_{i,j}, 0, R \times \sin \varphi_{i,j}]^T \tag{12}$$

$$\varphi_{i,j} = i \times \frac{360^\circ}{M} + (j - 1) \times 120^\circ \tag{13}$$

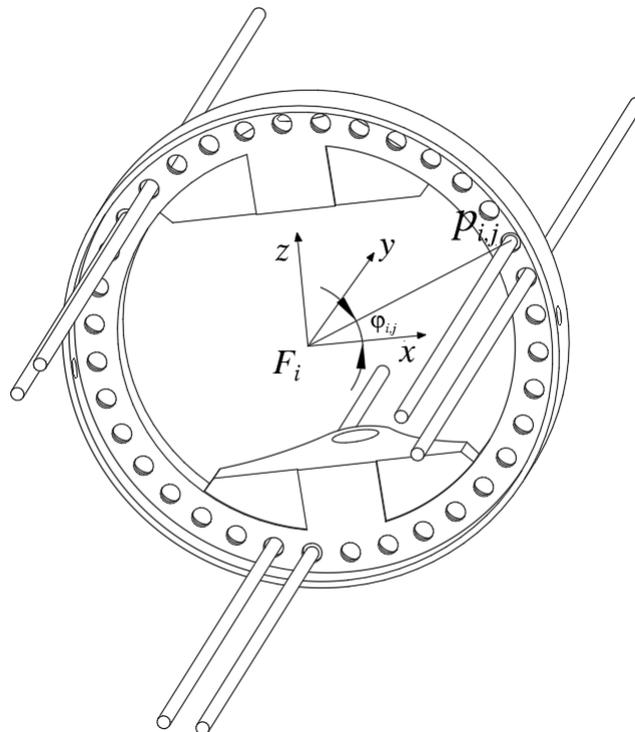


Figure 5. The cable’s mounted point on the rear endplate.

The through hole (instead of the mounted point if this cable is used for driving subsequent tubular unit) is denoted as  $p'_{i,j}$  in the Frame  $F'_i$  and the Euclidean distance from  $p_{i-1,j}$  to  $p'_{i,j}$  denoted as  $l_{i,j}$  which represents the cable’s length between the  $i$ -th tubular unit and the  $(i + 1)$ -th one in the condition of considering the cable is always straight through

the tube and ignoring its deformation. The total lengths of cables in k-th tubular units can be derived as Equation (15).

$$l_{i,j} = \left\| {}^i_{i-1}T \times p_{i,j} - p'_{i,j} \right\| \tag{14}$$

$$L_{m,j} = \sum_i^m (l_{i,j} + H) \quad m = 1, 2 \dots N \tag{15}$$

$L_{m,j}$  is formulated as a function  $f(\varphi_{i,j}, \alpha_i, \beta_i)$  once all the cables' lengths are given, each joint angle  $[\alpha_i, \beta_i]$  will be derived uniquely and definitely with respect to Equations (10)–(15).

### 3.2.3. End-Effector Pose Expression

The end-effector is fixed on the  $N$ -th tubular unit and the coordinate transformation matrix is denoted as  ${}^E_N T$  similarly to  ${}^i_{i-1} T$ . The coordinate transformation matrix from base to the end-effector can be derived as Equation (16) where the  $F_0$  represents the global coordinate Frame.

$${}^E_0 T = \prod_{i=1}^N ({}^i_{i-1} T \times Trans(0, H, 0)) \times {}^E_N T \tag{16}$$

The pose of end-effector can be achieved from Equations (10)–(16) due to the forward kinematics model is stated explicitly from cable space to the working space based on the equations described above. The camera for visual tracking undertakes the obligation of the end-effector in the scenario of this article.

### 3.3. Inverse Kinematics

In general, inverse kinematics is much more complicated than forward kinematics especially for CHSM due to its high level of redundant DOF which means numerous analytical solutions. Otherwise, the numerical method suffers severe penalty from solving the pseudo-inverse of the Jacobian matrix which costs increasing rapidly computation time. Under these circumstances, an efficient, simplified, and accessible solution could be taken advantage instead of the numerical method.

Two assumptions are introduced to simplify the joints' motion. First, the last tubular unit keeps horizontal direction paralleling to the  $Y$ -axis of the world coordinate system. This restriction improves stability for the visual sampling and keeps away from image distortion. Furthermore, the  $N$ -th joint angle  $[\alpha_N, \beta_N]$  could be easily obtained from  $[\alpha_N, \beta_N] + \sum_{i=0}^{N-1} [\alpha_i, \beta_i] = 0$  if the other ones are definite. Second, adjust each joint as slightly as possible and primarily use the last couple of joints during moving the end-effector to the desired position. That means if the desired position is contained in the working space of the last three tubular units, the last three joints should be adjusted, and the others should keep remaining unchanged. When the desired position is beyond working space, a requirement of adding some new left-neighbor tubular unit arises. Under this condition, we simplify the model into a Three-Connecting-Rod Mechanism by forcing the middle tubular units to form a straight line which means their joint angles are set to zero.

In the simple model, the tubular units and universal joints are substituted by lines (with a length of  $L = H + 2D$ ) and dots respectively where  $A_{Nd}$  represents the  $N$ -th joint's position. The subscripts  $d$  and  $o$  denote the desired position and the initial position. After the tracking target's moving trajectory is captured by the tracking component,  $A_{Nd}$  is definite and the number  $m$  of demanding joints to be adjusted can be inferred according to Equation (17). Equation (17) means the desired position is beyond the working space of the last  $N - m - 1$  joints but is contained in the working space of the last  $N - m$  ones.

$$\begin{cases} \|A_{Nd} - A_{(m+1)o}\| \geq (N - m - 1) \times L \\ \|A_{Nd} - A_{mo}\| \leq (N - m) \times L \end{cases} \tag{17}$$

Based on the simplified Three-Connecting-Rod Mechanism the  $A_{md}$ ,  $A_{(m+1)d}$  and  $A_{Nd}$  satisfies Equation (18).

$$\begin{cases} \|A_{Nd} - A_{(m+1)d}\| = (N - m - 1) \times L \\ \|A_{(m+1)d} - A_{md}\| = L \end{cases} \quad (18)$$

$A_{(m+1)d}$  is located on an intersecting circle between a sphere (center at  $A_{md}$ , radius of  $L$ ) and another (center at  $A_{Nd}$ , radius of  $(N - m - 1)L$ ). Following the second assumption the  $A_{(m+1)d}$  is accessible and the other  $A_{id}, i = m, m + 2, \dots, N - 1$  could be inferred via the coordinate transformation introduced earlier.

### 3.4. Control Method

Based on the kinematics discussed previously, a control scheme of CHSM is shown in Figure 6. For the object tracking task, the end-effector is trying to follow the object’s moving trajectory and keep a constant distance from the target. The tracking component is applied to infer the target’s position in the working space according to that in the pixel space by multiplying the inverse instinct matrix of the camera and then calculate the desired end-effector’s pose  $A_{Nd}(x_{Nd}, y_{Nd}, z_{Nd})$ . The desired cable’s length  $l_{di}(i = 1, 2, \dots, 3N)$  are calculated from Equations (11)–(15) with desired joint angle  $[\alpha_{di}, \beta_{di}](i = 1, 2, \dots, N)$  obtained from the inverse kinematics with  $A_{Nd}$ . The linear magnetic encoder mounted on the actuator measures each practical cable length  $l_{ri}(i = 1, 2, \dots, 3N)$  as inputs for the closed-loop PID controller which is designed as follows:

$$u = K_P(l_d - l_r) + K_I \int_0^t (l_d - l_r) d\tau + K_D(\dot{l}_d - \dot{l}_r) \quad (19)$$

where  $K_P, K_I, K_D$  are PID parameters and the output has a linear proportional relation with the cable’s stretching speed.

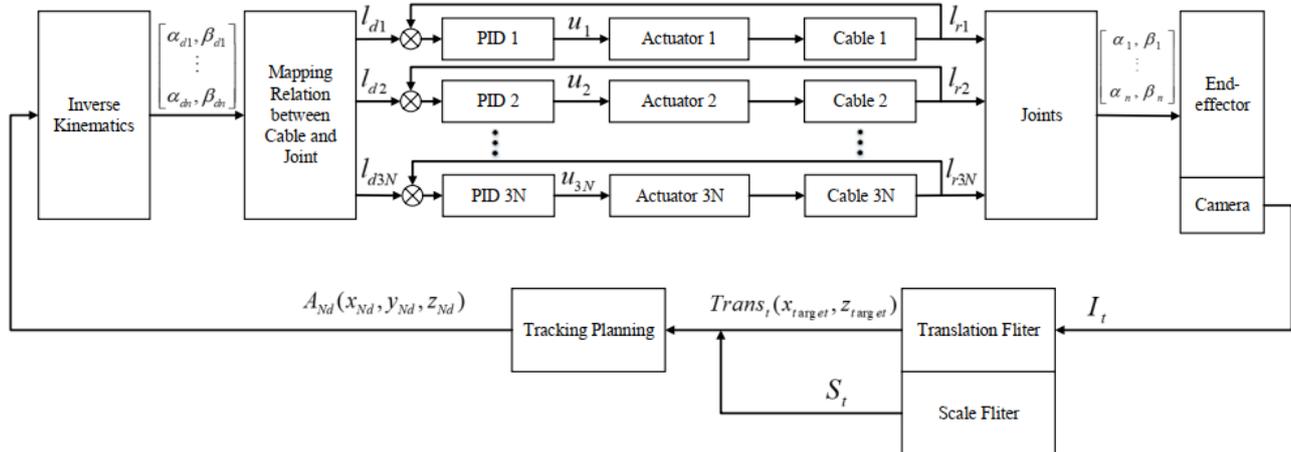


Figure 6. The control scheme of CHSM.

## 4. Experiment and Validation

In this section, an experiment is implemented to validate the moving and controlling performances of CHSM based on the introduced visual tracking.

The prototype platform used for the experiment is shown in Figure 7. The primary parameters are illustrated in Table 1. Moreover, the PID parameters  $K_P, K_I, K_D$  are selected as  $K_P = 5, K_I = 2, K_D = 0.05$  in a comprehensive consideration both for rapid response and stability of each actuator. The displacement signal of cable collected by Quanser QUARC software from the linear magnetic encoder to PID controller and the reverse controlling signal are all aggregated and passed via the real-time platform Simulink/Matlab with a

sampling frequency at 100 HZ. The visual servo feedback referring signal is captured by the camera (attached on the end tubular structural unit) which is working in BGR color space and 30 HZ refresh frequency. HOG features perform better and are adopted in the next experiments.

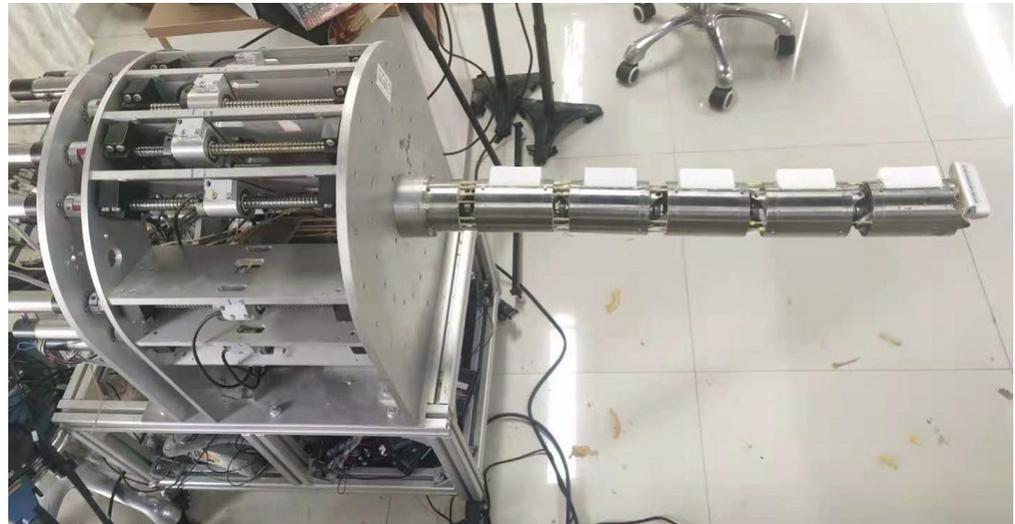


Figure 7. The prototype entity.

Table 1. Some parameters selected in the experiment.

Parameters	Description	Values
$\lambda$	Regularization parameter of ridge regression	0.01
$\eta$	Learning rate for updating correlation filter	0.025
$N$	Number of tubular units	5
$D$	Distance from endplate to universal joint	12 mm
$R$	Radius of tubular structure	33.8 mm
$H$	Length of tubular structure	106 mm

The experimental results of tracking target are shown in Figures 8–11, where the target is moving towards a random direction with a slowly increasing velocity earlier and then reduced to a normal level later. The tracking component can accurately capture the target in the visual field and driving the CHSM to adjust its pose (according to the Kinematics Equations (11)–(18)) to focus on the target, which means the target always appears at the center of the camera’s perceived screen in every frame. In that case, the moving trajectory of CHSM’s endpoint should be identical to the target as closely as possible. The CHSM’s moving in the Y-axis is limited to a narrow range due to the scale adaptation of the tracking component. Figures 9 and 10 show the experimental result of the X-axis and Z-axis, respectively.

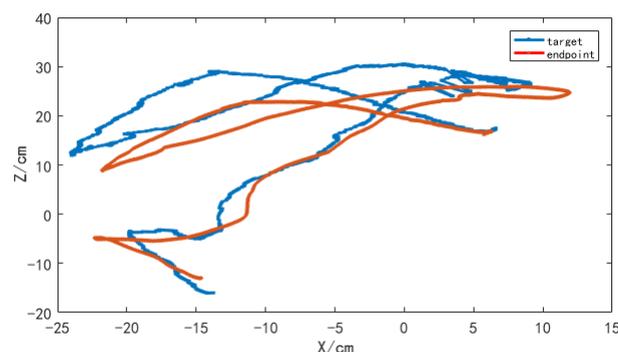


Figure 8. The moving trajectory of target and endpoint of CHSM.

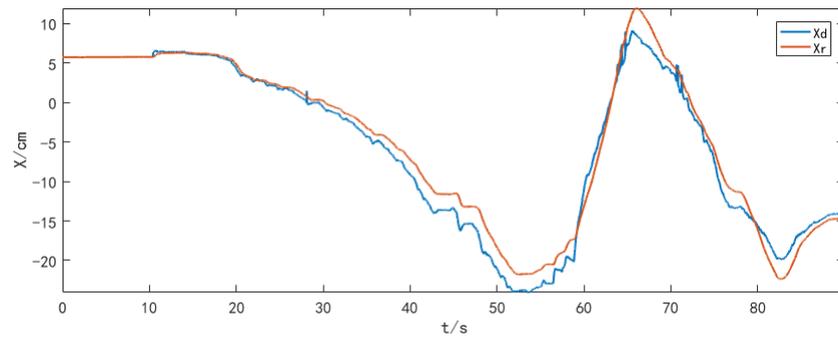


Figure 9. The result of X-axis.

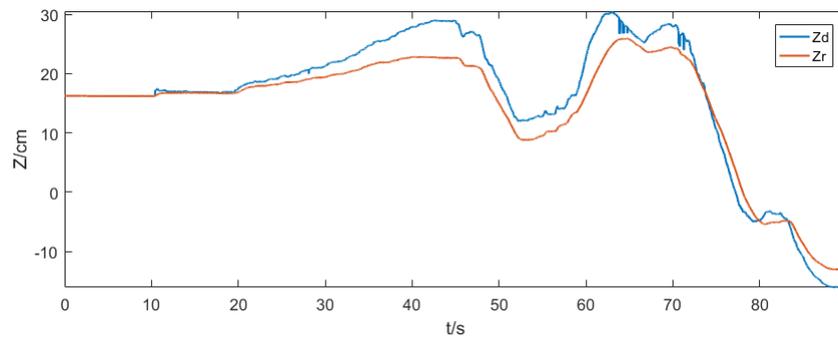
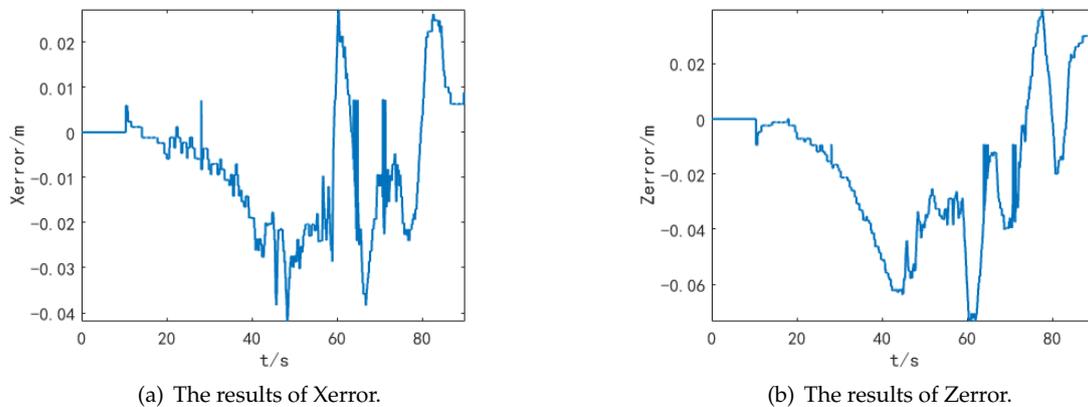


Figure 10. The result of Z-axis.

The subscript labels  $d, r$  represent the target and the CHSM’s endpoint respectively and have the same meaning in the Figure 11.



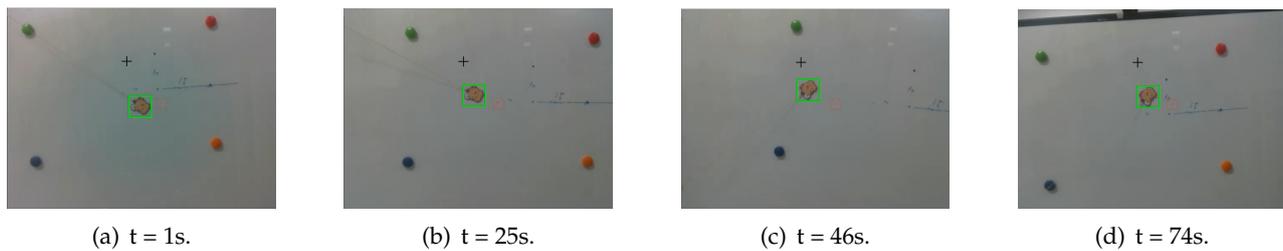
(a) The results of Xerror.

(b) The results of Zerror.

Figure 11. The result of error.

CHSM has a good tracking performance with the target except for a little delay. The absolute value of error has the same variation tendency as the target’s moving speed which may be caused by the limited sampling scope of the tracking component.

Figure 12 shows the recorded perspective from the camera when the target is moving dragged by a rope and four frames are selected as representatives. The target is detected and highlighted by a green bounding box which is nearly located in the center of the frames. When the target moves, the tracking component detects the target’s new position and guides the CHSM to focus on the target so that the bounding box looks such as rooting on the center. The bounding box’s deviation from the center has the same tendency as the error (shown in Figure 11).



**Figure 12.** Observation window of the camera.

## 5. Conclusion and Future Work

In this paper, we introduce a cable-driven hyper-redundant snake-like manipulator (CHSM), then build up a rigorous forward kinematics model base on coordinate transformation and a simplified inverse kinematics model via geometry relationships which has a low time-consuming for real-time control. The correlation filter technique is introduced to get on well with the tracking task, optimized for the adaptive scale variation, and improved on the robustness under the visual servo. The experimental result shows that CHSM has a good trajectory tracking performance and is endowed with some intelligence by the visual tracking technique. This work brings an advanced computer vision technique into the cable-driven snake-like manipulator to form an intelligent scheme. This scheme can be conveniently expanded for more advanced computer vision techniques involved on comprehensive tasks.

The future work will be carried out as follows. The weight of one tubular unit is a little beyond our expectation which restricts the extension for several units and impacts loading capacity. We have considered two aspects for optimizing it, reducing the radius of the tubular structure or finding an advanced composite material (e.g., carbon fiber reinforced polymer, etc.) instead of steel. The study of this paper is launched on the assumption that the end-effector is restricted to keeping horizontal orientation so that the camera's field of view is limited to some degree. In future work, this restriction will be relieved, and the end-effector can orientate toward any direction to assure that the manipulator can respond with minimal translation or rotation which is critical under the high-speed tracking circumstance. Meanwhile, the inverse kinematics and control loop should be redesigned.

**Author Contributions:** Conceptualization, J.T. and Z.C.; methodology, J.T. and Z.C.; software, Q.Z. and L.Q.; validation, Q.Z. and L.Q.; formal analysis, Y.N.; investigation, Q.Z. and L.Q.; resources, Z.C. and Y.N.; data curation, Q.Z.; writing—original draft preparation, Q.Z.; writing—review and editing, J.T. and Z.C.; visualization, Y.N.; supervision, J.T. and Y.N.; project administration, Y.N., J.T. and Z.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** This paper reports no data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Theorem A1.** *The eigenvalues of a circulant matrix  $C_x$  are given by the DFT  $\hat{x} = F(x)$ , and the eigenvectors by the unitary DFT matrix  $U$ , proved in [30]. Equivalently,*

$$C(x) = U \text{diag}(\hat{x}) U^* \quad (\text{A1})$$

where  $C(x)$  represents the circulant matrix constructed from the signal  $x$ .  $U^*$  is the complex-conjugate of the unitary DFT matrix  $U$ .

**Theorem A2.** Convolution theorem: The convolution operation can be viewed as the circulant matrix constructed from one signal multiply with the other, proved in [31]. Equivalently,

$$x * y = C(\bar{x})y \quad (\text{A2})$$

where  $\bar{x}$  represents the reversed sequence of signal  $x$ .

By the Theorem A1 and Equation (2),

$$\begin{aligned} w &= (Fdiag(\hat{x}^*)F^H Fdiag(\hat{x})F^H + \lambda FF^H)^{-1} Fdiag(\hat{x}^*)F^H y \\ &= (Fdiag(\hat{x}^* \odot \hat{x} + \lambda)F^H)^{-1} Fdiag(\hat{x}^*)F^H y \\ &= Fdiag\left(\frac{\hat{x}^*}{\hat{x}^* \odot \hat{x} + \lambda}\right)F^H y \end{aligned} \quad (\text{A3})$$

Then according to Theorem A2,

$$F(C(x)y) = F(\bar{x} * y) = F(x)^* \odot F(y) \quad (\text{A4})$$

applied this transformation in Equation (A3),

$$\hat{W} = \frac{\hat{x} \odot \hat{y}}{\hat{x}^* \odot \hat{x} + \lambda} \quad (\text{A5})$$

## References

- Moran, M.E. Evolution of robotic arms. *J. Robot. Surg.* **2007**, *1*, 103–111. [[CrossRef](#)] [[PubMed](#)]
- Wooten, M.; Frazelle, C.; Walker, I.D.; Kapadia, A.; Lee, J.H. Exploration and inspection with vine-inspired continuum robots. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 5526–5533.
- Tang, J.; Zhang, Y.; Huang, F.; Li, J.; Chen, Z.; Song, W.; Zhu, S.; Gu, J. Design and kinematic control of the cable-driven hyper-redundant manipulator for potential underwater applications. *Appl. Sci.* **2019**, *9*, 1142. [[CrossRef](#)]
- Buckingham, R.; Chitrakaran, V.; Conkie, R.; Ferguson, G.; Graham, A.; Lazell, A.; Lichon, M.; Parry, N.; Pollard, F.; Kayani, A.; et al. *Snake-Arm Robots: A New Approach to Aircraft Assembly*; SAE International: Warrendale, PA, USA, 2007.
- Qin, L.; Huang, F.; Chen, Z.; Song, W.; Zhu, S. Teleoperation Control Design with Virtual Force Feedback for the Cable-Driven Hyper-Redundant Continuum Manipulator. *Appl. Sci.* **2020**, *10*, 8031. [[CrossRef](#)]
- Liljebäck, P.; Pettersen, K.Y.; Stavdahl, Ø.; Gravdahl, J.T. A review on modelling, implementation, and control of snake robots. *Robot. Auton. Syst.* **2012**, *60*, 29–40. [[CrossRef](#)]
- Xu, W.; Liu, T.; Li, Y. Kinematics, dynamics, and control of a cable-driven hyper-redundant manipulator. *IEEE/ASME Trans. Mechatron.* **2018**, *23*, 1693–1704. [[CrossRef](#)]
- Zhang, Z.; Yang, G.; Yeo, S.H. Inverse kinematics of modular Cable-driven Snake-like Robots with flexible backbones. In Proceedings of the 2011 IEEE 5th International Conference on Robotics, Automation and Mechatronics (RAM), Qingdao, China, 17–19 September 2011; pp. 41–46.
- Martin, A.; Barrientos, A.; Del Cerro, J. The natural-CCD algorithm, a novel method to solve the inverse kinematics of hyper-redundant and soft robots. *Soft Robot.* **2018**, *5*, 242–257. [[CrossRef](#)] [[PubMed](#)]
- Wang, H.; Chen, W.; Yu, X.; Deng, T.; Wang, X.; Pfeifer, R. Visual servo control of cable-driven soft robotic manipulator. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 57–62.
- Khatib, O.; Yeh, X.; Brantner, G.; Soe, B.; Kim, B.; Ganguly, S.; Stuart, H.; Wang, S.; Cutkosky, M.; Edsinger, A.; et al. Ocean one: A robotic avatar for oceanic discovery. *IEEE Robot. Autom. Mag.* **2016**, *23*, 20–29. [[CrossRef](#)]
- Vojir, T.; Noskova, J.; Matas, J. Robust scale-adaptive mean-shift for tracking. *Pattern Recognit. Lett.* **2014**, *49*, 250–258. [[CrossRef](#)]
- Karavasiliis, V.; Nikou, C.; Likas, A. Visual tracking using the Earth Mover’s Distance between Gaussian mixtures and Kalman filtering. *Image Vis. Comput.* **2011**, *29*, 295–305. [[CrossRef](#)]
- Rao, G.M.; Nandyala, S.P.; Satyanarayana, C. Fast visual object tracking using modified Kalman and particle filtering algorithms in the presence of occlusions. *Int. J. Image Graph. Signal Process.* **2014**, *6*, 43–54. [[CrossRef](#)]
- Hare, S.; Golodetz, S.; Safari, A.; Vineet, V.; Cheng, M.M.; Hicks, S.L.; Torr, P.H. Struck: Structured output tracking with kernels. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 2096–2109. [[CrossRef](#)] [[PubMed](#)]
- Kalal, Z.; Mikolajczyk, K.; Matas, J. Tracking-learning-detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1409–1422. [[CrossRef](#)] [[PubMed](#)]

17. Vidal, F.B.; Alcalde, V.H.C. Window-matching techniques with Kalman filtering for an improved object visual tracking. In Proceedings of the 2007 IEEE International Conference on Automation Science and Engineering, Scottsdale, AZ, USA, 22–25 September 2007; pp. 829–834.
18. Kwon, J.; Park, F.C. Visual tracking via particle filtering on the affine group. *Int. J. Robot. Res.* **2010**, *29*, 198–217. [[CrossRef](#)]
19. Li, S.X.; Chang, H.X.; Zhu, C.F. Adaptive pyramid mean shift for global real-time visual tracking. *Image Vis. Comput.* **2010**, *28*, 424–437. [[CrossRef](#)]
20. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550.
21. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 583–596. [[CrossRef](#)] [[PubMed](#)]
22. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Discriminative scale space tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1561–1575. [[CrossRef](#)] [[PubMed](#)]
23. Li, Y.; Zhu, J. A scale adaptive kernel correlation filter tracker with feature integration. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 254–265.
24. Danelljan, M.; Hager, G.; Shahbaz Khan, F.; Felsberg, M. Learning spatially regularized correlation filters for visual tracking. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4310–4318.
25. Kiani Galoogahi, H.; Sim, T.; Lucey, S. Correlation filters with limited boundaries. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4630–4638.
26. Ma, C.; Yang, X.; Zhang, C.; Yang, M.H. Long-term correlation tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5388–5396.
27. Wang, M.; Liu, Y.; Huang, Z. Large margin object tracking with circulant feature maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4021–4029.
28. Danelljan, M.; Shahbaz Khan, F.; Felsberg, M.; Van de Weijer, J. Adaptive color attributes for real-time visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1090–1097.
29. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
30. Henriques, J.F. Circulant Structures in Computer Vision. Ph.D. Thesis, Universidade de Coimbra, Coimbra, Portugal, 2015.
31. Lyons, R.G. *Understanding Digital Signal Processing, 3/E*; Pearson Education: Chennai, India, 2004.