Automatic Tracking of Surgical Instruments with a Continuum Laparoscope Using Data-Driven Control in Robotic Surgery

Kong, Xiaowen; Mo, Hangjie; Dong, Erbao; Liu, Yunhui; Sun, Dong

Published in:
Advanced Intelligent Systems

Published: 01/02/2023

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

License:
CC BY

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1002/aisy.202200188

Publication details:

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

Download date: 27/01/2024
Automatic Tracking of Surgical Instruments with a Continuum Laparoscope Using Data-Driven Control in Robotic Surgery

Xiaowen Kong, Hangjie Mo, Erbao Dong, Yunhui Liu, and Dong Sun*

In existing surgery process, surgeons need to manually adjust the laparoscopes to provide a better field of view (FOV) during operation, which may distract surgeons and slow down the surgery process. Herein, a data-driven control method that uses a continuum laparoscope to adjust the FOV by tracking the surgical instruments is presented. A Koopman-based system identification method is first applied to linearize the nonlinear system. Shifted Chebyshev polynomials are used to construct observation functions that transfer low-dimensional observations to high-dimensional ones. The Koopman operator is approximated using a finite-dimensional estimation method. An optimal controller is further developed according to the trained linear dynamic model. Furthermore, a learning-based pose estimation framework is designed to detect keypoints on surgical instruments and provides visual feedback for the control system. Compared with other detection methods, the proposed scheme achieves a higher detection precision and provides more optional keypoints for tracking. Simulation and experiments validate the feasibility of the proposed control method. Experimental results show that the proposed method can automatically adjust the field of continuum laparoscope by tracking surgical instruments in real time and satisfy the clinical requirements.

1. Introduction

Robot-assisted minimally invasive surgery (RMIS) has received increasing attention because of its unique advantages compared with traditional open surgery.[7–9] In RMIS, laparoscopes used to display the surgical scenario on a screen are held by a robotic arm instead of an assistant. Surgeons need to adjust the laparoscope frequently to provide a proper field of view (FOV) in surgery. This procedure distracts the attention of surgeons during the surgery, thereby affecting the progress of the laparoscopic operations.

A control method for adjusting the FOV of laparoscopy by tracking surgical instruments automatically during surgery needs to be developed. Currently, rigid laparoscopy is widely applied to RMIS with automatic FOV adjusting algorithms. Yang et al. proposed a region-based visual servoing method to automatically manipulate laparoscopy with colored markers, which can improve the control efficiency and safety of FOV.[10] An autonomous surgical instrument tracking method without any markers was further proposed based on the visual tracking space vector.[11] An eye-tracking camera control method is used in Senhance surgical robotic system (Asensus Surgical, NC, USA) for adjusting the FOV of the laparoscope. However, these methods encounter the workspace problem. When operating a rigid laparoscope using a robotic arm, avoiding collision between the robot arm and the other surgical instruments is necessary. It is difficult to operate a rigid laparoscope using a robotic arm in a narrow workspace, leading to the limited FOV of the rigid laparoscope.

Continuum manipulators have been widely applied in robotic surgical applications due to their higher dexterity and less workspace required.[8–10] Recently, continuum manipulators have been used for automatic FOV adjustment with visual servoing in RMIS.[11] However, the dynamics of continuum manipulator are always highly nonlinear and high dimensional due to the mechanical compliance of its structure.[12] These characteristics bring challenges to the precise control of continuum manipulators. Existing methods usually simplify the continuum manipulator based on physical assumptions in establishing dynamic models, such as piecewise constant curvature model, pseudo-rigid-body, quasistatic, and simplified geometry.[13–20]
Assumptions in these simplified models may lead to deviation under actual conditions and inaccurate results, which are not feasible for use in practice, especially for scenarios with high precision requirements.

Recently, data-driven control methods, such as neural networks and reinforcement learning, have shown great potential for controlling continuum manipulators. The advantage of these methods lies in the input–output mapping of the system derived from sensing data without analytical modeling and complex computation. Given enough input–output data, data-driven models can describe the behavior of the system over its entire operating range. However, these methods usually require many tuning parameters and repeated trials to establish accurate models. Other concerns include low real-time performance and computational complexity. Koopman operator provides an alternative solution for establishing the dynamic model of a continuum manipulator based on its unique linear structure. Koopman operator lifts the nonlinear dynamic model of the system into an infinite-dimensional space and evolves the state functions, also called observation functions, in the new space. In this way, the dynamic model of the nonlinear system can be easily propagated in a linear manner, relying on input–output data only. As a result, linear control methods can be applied to control the continuum manipulator with high precision.

Apart from accurate system identification, close-loop control with high precision visual feedback is also essential. Visual feedback in laparoscopic instruments tracking can be divided into two types: marked methods and unmarked methods. Marked methods manually add a characteristic marker on the instrument for easy detection. Although this method can localize the target quickly, uncertainty exists due to the presence of blood and gas during surgery. Furthermore, this method provides surgeons with a poor experience and has a low tracking precision because markers are usually located at a nonclient rod of the instrument. Unmarked methods usually choose the whole metal part of the instrument as the detecting area and then detect the area as an object detection method. This method can increase the precision of keypoints detection and benefit subsequent control tasks. The second unit is the visual feedback and optimal control unit, which provides control feedback for the surgical instruments tracking task. In this unit, a deep keypoints detection network is developed to predict the pixel positions of keypoints on surgical instruments. Unlike the existing object detection methods, a pose estimation method is developed to detect the keypoints on surgical instruments. The pose estimation method can directly regress the pixel coordinates of the keypoints on surgical instruments instead of detecting the whole area as the object detection method. This method can increase the precision of keypoints detection and benefit subsequent control tasks. The keypoints needed as the tracking point can be selected flexibly according to different surgery stages in the following control system.

In this present work, we focus on autonomous control of a continuum laparoscope to adjust the FOV and keep the surgical instruments at the view center in RMIS. To address this critical issue, an automatic surgical instrument tracking framework is proposed based on a Koopman-based control scheme and learning-based vision feedback. As shown in Figure 1, this framework can be divided into two units. The first is the data-driven system identification unit, which applies the Koopman operator to transfer a nonlinear dynamics system into a linear closed-loop control. Unlike the Taylor-based method, we introduce the Chebyshev polynomials to choose observation functions. Chebyshev polynomials is a global approximation method dependent on high-order derivatives of the system state as existing methods. The approximation error of the proposed method is also analyzed. A linear quadratic regulator controller is further used for real-time control based on the linear representation of the continuum laparoscope.

The second unit is the visual feedback and optimal control unit, which provides control feedback for the surgical instruments tracking task. In this unit, a deep keypoints detection network is developed to predict the pixel positions of keypoints on surgical instruments. Unlike the existing object detection methods, a pose estimation method is developed to detect the keypoints on surgical instruments. The pose estimation method can directly regress the pixel coordinates of the keypoints on surgical instruments instead of detecting the whole area as the object detection method. This method can increase the precision of keypoints detection and benefit subsequent control tasks. The keypoints needed as the tracking point can be selected flexibly according to different surgery stages in the following control system. In addition, weights of different surgical instruments can be

Figure 1. Framework of the automatic tracking system of the surgical instruments with a continuum laparoscope.
set by surgeons when multiple instruments are used according to their requirements.

The rest of this article is organized as follows. Section 2 reviews the Koopman operator and describes the linear quadratic regulator (LQR) controller design. The method to select the observation functions through shifted Chebyshev polynomials is introduced, and the approximation error is analyzed. In Section 3, the pose estimation task of surgical instruments and the architecture of the learning-based keypoint detection network are described in detail. Section 4 evaluates the proposed method in a simulation environment. Section 5 introduces the surgical instruments tracking method on a real platform, and the experimental results are also discussed. Finally, conclusions of the work and a discussion of future work are given in Section 6.

2. Data-Driven System Identification and LQR Control Design

2.1. Preliminary

Consider the following general discrete nonlinear dynamic system

\[ s_{k+1} = F(s_k, u_k) \]  

(1)

where \( k \) is the time step, \( s_k \in \mathbb{R}^d \) represents the state of the system at step \( k \) and \( u_k \in \mathbb{R}^q \) is the input of the system at step \( k \), and \( F \) denotes the transition function, which advances the state \( s_k \) to \( s_{k+1} \) with the corresponding input \( u_k \). Furthermore, we defined an observation function \( \Psi \), which is used to transfer observed \( q \)-dimension states \( s_k \) into \( \Psi(s_k) \), in an infinite-dimensional space \( \mathcal{F} \).

The Koopman operator \( \mathcal{K} \), known as an infinite-dimensional linear operator, is applied to approximate the nonlinear system of the continuum laparoscope.\(^{[12-15]}\) The linear system in Equation (1) can be expressed by a set of Koopman operators \( \mathcal{K}_d \) and the defined observation function \( \Psi \) in the infinite-dimensional function space \( \mathcal{F} \)

\[ \mathcal{K}_d \Psi = \Psi \circ F \]  

(2)

where \( d = 1, 2, \ldots \) is the index of the Koopman operator. Therefore, Koopman operator \( \mathcal{K} \) advances measurements of the states linearly. Equation (2) can be written as follows

\[ \mathcal{K}_d \Psi(s_k, u_k) = \Psi(s_{k+1}, u_{k+1}) \]  

(3)

As a result, the nonlinear dynamical system given by (1) can be controlled in a linear manner by a set of Koopman operators. However, the Koopman operator \( \mathcal{K} \) is infinite-dimensional, which may be difficult to be applied in practice.

Recent studies use the least-squares method to approximate the infinite-dimensional operator \( \mathcal{K} \) with a finite-dimensional representation with tolerable error. In order to get an approximation to the Koopman operator, the observation function \( \Psi(s, u) \) in space \( \mathcal{F} \) can be expressed as

\[ \Psi(s, u) = [\psi_1(s, u), \psi_2(s, u), \ldots, \psi_p(s, u)] \]  

(4)

where \( p \) is the number of the observable functions.

The observed state and its corresponding input are assumed to have been collected in advance and expressed in vector form, as follows

\[ V = [(s_1, u_1), (s_2, u_2), \ldots, (s_i, u_i)] \]  

(5)

where \( i \) is the number of collected data with the same time step. Koopman operator \( \mathcal{K} \) can be calculated using the least-squares method,\(^{[29]}\) expressed as follows

\[ \mathcal{K}_d = \arg \min \sum_{i=1}^{i-1} \frac{1}{2} \left\| \Psi(s_{i+1}, u_{i+1}) - \mathcal{K}_d \Psi(s_i, u_i) \right\|^2 \]  

(6)

Solving the least-squares problem yields

\[ \mathcal{K}_d = A G^\dagger \]  

(7)

where \( \dagger \) is the Moore–Penrose pseudoinverse and

\[ A = \begin{pmatrix} 1 & \sum_{i=0}^{i-1} \Psi(s_{i+1}, u_{i+1}) \Psi(s_i, u_i)^T \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} 1 & \sum_{i=0}^{i-1} \Psi(s_i, u_i)^T \end{pmatrix} \]  

(8)

This approach generally yields a better approximation with the increase of output dimension of \( \Psi \). Furthermore, the number of collected data and their distribution across the state space will have a significant effect on the computed \( \mathcal{K}_d \). Thus, it is crucial to choose the observation function \( \Psi \) and collect enough data of the unknown system.

For simplicity, the observation function can be re-written as follows

\[ \Psi(s, u) = [\Psi(s), \Psi(u)] \]  

(9)

where \( \Psi(s) \in \mathbb{R}^{q_s} \) and \( \Psi(u) \in \mathbb{R}^{q_u} \) are the functions depending on the states \( s \) and inputs \( u \) of the system, respectively. \( \mathbb{R}^{q_s} \) and \( \mathbb{R}^{q_u} \) are the subsets of space \( \mathbb{R}^q \), and \( q = q_s + q_u \). \( q \) is the dimension of invariant subspace of \( \mathcal{K}_d \). Then we can rewrite (3) as follows

\[ \begin{bmatrix} \Psi(s_{k+1}) \\ \Psi(u_{k+1}) \end{bmatrix} = \begin{bmatrix} \mathcal{K}_d^A \Psi(s_k) \\ \mathcal{K}_d^B \Psi(u_k) \end{bmatrix} \]  

(10)

where \( \mathcal{K}_d^A \in \mathbb{R}^{q_s \times q_s} \) and \( \mathcal{K}_d^B \in \mathbb{R}^{q_u \times q_u} \) are submatrices of \( \mathcal{K}_d \). \( \Psi(s_{k+1}) \) can be expressed as follows

\[ \Psi(s_{k+1}) = \mathcal{K}_d^A \Psi(s_k) + \mathcal{K}_d^B \Psi(u_k) \]  

(11)

To ensure the control linearly, we use \( \Psi(u_k) = u_k \) and Equation (11) is re-written as follows

\[ \Psi(s_{k+1}) = \mathcal{K}_d^A \Psi(s_k) + \mathcal{K}_d^B u_k \]  

(12)

Algorithm 1 summarizes the system identification process with data-driven methods.
Algorithm 1. Nonlinear system identification with data-driven methods.

Input: Collected system states and inputs \((s_k, u_k)\), penalty coefficient \(\lambda\)

Step 1: Lift the system states \(s_{k+1}\) to \(\Psi(s_{k+1})\) using Chebyshev polynomials with (16)

Step 2: Build the observation function \(\Psi(s, u)\) with (4)

Step 3: Calculate the Koopman operator \(\tilde{K}_d\)

Step 4: Extract submatrices \(\tilde{K}_d^1\) and \(\tilde{K}_d^2\) belong to the states \(s\) and inputs \(u\) with (10)

Step 5: Build the linear model of the system with (11).

Output: Koopman operator \(\tilde{K}_d\)

Now, the nonlinear system has been expressed in a linear manner with the Koopman representation. A LQR controller is applied to choose optimal control inputs over a finite time horizon. The minimized objective is defined as follows

\[
J = \sum_{k=0}^{\infty} (\Psi(s_k) - \Psi(s_{k+1}))^T Q (\Psi(s_k) - \Psi(s_{k+1})) + u_k^T R u_k
\]

where \(Q \geq 0 \in \mathbb{R}^{n_s \times n_s}\) and \(R \geq 0 \in \mathbb{R}^{n_u \times n_u}\) are positive definite weight matrices on states and inputs, which penalize the deviation from the desired observable functions \(\Psi(s_{k+1})\) at step \(k\).

The LQR feedback law for (13) becomes the following

\[
u_k = -K(\Psi(s_k) - \Psi(s_{k+1}))
\]

where \(K \in \mathbb{R}^{n_s \times n_u}\) represents the state-feedback gain metric of the LQR control. Metric \(K\) can be calculated by solving the discrete-time algebraic Riccati equation.[10]

Algorithm 2 summarizes the optimal control process of the nonlinear system.

2.2. Observation Function Selection

As described earlier, a set of observation functions is chosen to approximate the Koopman operator. The error between two adjacent can be expressed as follows

\[
\varepsilon = \Psi(s_{k+1}, u_{k+1}) - \tilde{K}_d \Psi(s_k, u_k)
\]

Algorithm 2. Optimal control of the nonlinear system.

Input: Initial state \(s_1\), target state \(s_{k+1}\), weight matrices \(Q\) and \(R\), linear model of the system with (11)

Step 1: Build linear quadratic regulator (LQR) controller and cost function of the system with (13) for \(t = 1, 2, \ldots, k + 1\)

Step 2: Calculate the needed input \(u_i\)

Step 3: Apply \(u_i\) to the system

Step 4: Set the time from \(t\) to \(t + 1\)

Step 5: Get and update the new state \(s_{k+1}\)

Output: System input sequences \(u_1, u_2, \ldots, u_k\) and state sequences \(s_1, s_2, \ldots, s_k\)

To minimize the error \(\varepsilon\), selecting an appropriate observation function is important. Taylor series, which is continuously differentiable up to \(n\)th order, has been applied to approximate the observation function. Considering that the values of derivatives in Taylor series can be evaluated using numerical estimation from state measurements, this method is easy to implement. However, the Taylor series maintains high precision only at the expansion point \(s_k\). Moreover, it has strict conditions, such as continuously differentiable.

To solve this problem, we introduce shifted Chebyshev polynomials into the selection of observation functions.[11] Chebyshev polynomial approximation is a global method that gives a better approximation for the truncated series than other techniques. For \(t \in [-1, 1]\), the well-known Chebyshev polynomials \(T_m(t)\) are defined through the following identity

\[
T_m(t) = \cos(m\theta)
\]

where \(\theta = \arccos(t)\) and \(m = 0, 1, 2, \ldots\). The polynomials form an orthogonal basis with the corresponding weighting function \(\rho(t) = 1/\sqrt{1 - t^2}\) and can be expressed as follows

\[
\int_{-1}^{1} T_m(t) T_n(t) \rho(t) dt = \begin{cases} 0, & m \neq n \\ \frac{\pi}{2}, & m = n \neq 0 \\ \pi, & m = n = 0 \end{cases}
\]

The polynomials can be generated by the following

\[
T_0(t) = 1, \\
T_1(t) = t, \\
T_{m+1}(t) = 2tT_m(t) - T_{m-1}(t), m \geq 1
\]

A function \(f(t) \in \mathcal{D}^2(-1, 1)\) can be expanded in terms of the Chebyshev polynomials, as follows

\[
f(t) = \sum_{m=0}^{\infty} a_m T_m(t)
\]

where the coefficients \(a_m\) can be calculated by the following

\[
a_m = \frac{2}{\pi} \int_{-1}^{1} f(t) T_m(t) \rho(t) dt, \ m \geq 1
\]

Now, we can shift the domain of Chebyshev polynomials from \([-1, 1]\) to \([x_k, x_{k+1}]\) via the transformation \(x = \frac{2x - (x_{k+1} + x_k)}{x_{k+1} - x_k}\), where \(k = 0, 1, 2, \ldots\). The recursion formula of Chebyshev polynomials can be written as

\[
T_0(x) = 1, \\
T_1(x) = \frac{2x - (x_{k+1} + x_k)}{x_{k+1} - x_k}, \\
T_{m+1}(x) = 2\left(\frac{2x - (x_{k+1} + x_k)}{x_{k+1} - x_k}\right) T_m(x) - T_{m-1}(x), m \geq 1
\]

and the corresponding weighting function is changed into

\[
\rho(x) = 1/\sqrt{1 - \left(\frac{2x - (x_{k+1} + x_k)}{x_{k+1} - x_k}\right)^2}
\]

on the interval \([x_k, x_{k+1}]\). Hence, for any arbitrary function, \(f(x) \in \mathcal{D}^2[x_k, x_{k+1}]\), it can be approximated as

\[
f(x) \approx \sum_{m=0}^{\infty} a_m T_m(x)
\]
be written as follows

$$f(x) = \sum_{m=0}^{\infty} a_m T_m(x)$$

where $m \geq 1$, and the coefficients $a_m$ can be computed as follows

$$a_m = \frac{2}{\pi} \int_{x_k}^{x_{k+1}} f(x)T_m(x) \rho(x) \, dx$$

Truncating the series in Equation (22), we can rewrite it as follows

$$f(x) \approx \sum_{m=0}^{M} a_m T_m(x) = O^T \Phi(x)$$

where $T$ denotes matrix transposition. $O$ and $\Phi(x)$ represent the column vectors as follows

$$O = [a_0, a_1, \ldots, a_M]^T$$

$$\Phi(x) = [T_0(x), T_1(x), \ldots, T_M(x)]^T$$

Equation (25) can be analytically written as follows

$$T_m(x) = \sum_{i=0}^{m} b_m x^i$$

$$b_m = (-1)^{m-i} \frac{(m+i-1)!(2i)!}{(m-i)!(x_{k+1} - x_k)^i(2i)!}$$

From Equation (25)–(27), $\Phi(x)$ can be denoted as follows

$$\Phi(x) = P Q(x)$$

where $Q(x) = (1, x, x^2, \ldots, x^m)$, and $P$ is the coefficient matrix given as follows

$$P = \begin{bmatrix} p_{0,0} & 0 & \cdots & 0 \\ p_{1,0} & p_{1,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p_{m,0} & p_{m,1} & \cdots & p_{m,m} \end{bmatrix}$$

and

$$p_{0,0} = 1$$

$$p_{ij} = 0, \quad \text{for } i < j$$

$$p_{ij} = 2 \left( \frac{2}{x_{k+1} - x_k} p_{i-1,j-1} \frac{x_{k+1} + x_k}{x_{k+1} - x_k} p_{i-1,j} \right) - p_{i-2,j}$$

For any arbitrary $x \in [x_k, x_{k+1}]$, there exists coefficient $q$ ($0 < q < 1$) that satisfies $x_{k+1} = \frac{x_k}{q}$. Then we have the following

$$T_m(x_{k+1}) = T_m \left( \frac{x_k}{q} \right) = R T_m(x_k)$$

The parameter $R$ can be written as following

$$R = \begin{bmatrix} r_{0,0} & 0 & \cdots & 0 \\ r_{1,0} & r_{1,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ r_{m-1,0} & r_{m-1,1} & \cdots & r_{m-1,m-1} \\ r_{m,0} & r_{m,1} & \cdots & r_{m,m} \end{bmatrix}$$

where

$$r_{ij} = \begin{cases} 1 & \text{for } i = j \\ f_j^{-1}(1) \sum_{l=j+1}^{M} f_l^{-1}(a_{l,j+1}), & \text{for } i < j \\ 0, & \text{for } i > j \end{cases}$$

and

$$a_{i,j+1} = -p_{i,j+1} r_{i,j+1}, \quad i = 1, 2, \ldots, m$$

$$a_{i,j} = -p_{i,j-1}^{-1} \left( p_{i,j-1}^{-1} + \sum_{l=j+1}^{M} a_{l,j} p_{l-1,j-1} \right),$$

$$i \in [1, m-1], \quad j \in [1, m-1]$$

So there exists a matrix $S$ that satisfies the following

$$\sum_{m=0}^{M} a_m T_m(x_{k+1}) = S \sum_{m=0}^{M} a_m T_m(x_k)$$

where $S$ can be calculated by, respectively, equating the coefficients of $x^i$ and solving the linear equations.

It is found that Equation (3) is similar to the Equation (35). Here we choose the Chebyshev polynomials and corresponding coefficients as the observables, as follows

$$\Psi(x_{k+1}) = [a_0 T_0(x_{k+1}), a_1 T_1(x_{k+1}), \ldots, a_M T_M(x_{k+1})]^T$$

The method can also be used for a system with multiple states.

### 2.3. Error Estimation and Overfitting

Considering the shifted Chebyshev polynomials of $f(x) \in \mathbb{R}[x_k, x_{k+1}]$, if $f(x)$ has $M + 1$ derivatives, the max error in Equation (15) induced by Chebyshev polynomials to calculate Koopman operator $\hat{K}_{ad}$ across one time step can be calculated as follows

$$\max_{x_k \leq x \leq x_{k+1}} |e| \leq \frac{(x_{k+1} - x_k)^{M+1}}{2^{M+1}(M+1)!} \max \left| \sum_{m=0}^{M} a_m T_M(x) \right|$$

One of the main factors affecting the data-driven modeling approach is overfitting. Although the least square method can minimize the $L^2$-norm error in the training process, it is easily affected by noise and singularity. To solve this problem, we apply least absolute shrinkage and selection operator (LASSO)\cite{222}, which is an $L^1$-regularization method.
Using the LASSO, the Equation (6) can be written as follows
\[
\hat{K}_d = \arg\min_{K_d} ||\Psi(s_{i+1}, u_{i+1}) - K_d \Psi(s_i, u_i)||_2^2 + \lambda ||K_d||_1
\]  
(38)

where \( \lambda \) is the weight of the \( L^1 \) penalty coefficient.

### 3. Learning-Based Keypoint Detection

In this section, a learning-based framework is developed to detect and localize the keypoints on the surgical instrument without any artificial markers. Then the pixel position of the keypoint in the image plane is used as the feedback in a closed-loop control. As shown in Figure 2, five keypoints, namely: RightClasper, LeftClasper, Head, Shaft and End, are chosen with consideration of the articulated structure of surgical instruments.[32]

#### 3.1. Network Architecture

Existing approaches for pose estimation tasks mainly use heatmap regression to detect keypoints.[33,34] Then the grouping method is used to group the keypoints that belong to the same object. Alternatively, each object can be segmented in advance, and keypoints detection of each object is further carried out.[35] Heatmap regression has several known drawbacks, such as computational inefficiencies, inherent quantization error, and spatial resolution sensitivities.[36] Therefore, current methods cannot achieve real-time performance with high accuracy.

In order to solve this problem, a real-time pose estimation framework is proposed to detect keypoints of surgical instruments. Small bounding boxes are used to represent the keypoint regions and the position of keypoints located at the center of the region. Details of the architecture are described as follows.

In Figure 3, GhostNet is used as the backbone to extract features, which can generate more features by using fewer parameters.[37] A spatial attention bottleneck module is then applied to provide abundant spatial information on the keypoints.[38] Feature pyramid network (FPN) and path aggregation network (PAN) are used to fusion features with different scales, which can fusion spatial information from low-level feature maps and semantical information from high-level feature maps.[39] Meta-ACON is used as the activation function.[40] Nonmaximum suppression (NMS) is applied to obtain the candidate keypoint regions. Associative embedding is used for keypoints grouping after detecting module (group keypoint that belongs to the same instrument). The grouping process clusters the identity-free keypoints into individuals by grouping keypoints whose labels have small L2 distance.

#### 3.2. Loss Function

To enhance the pose estimation performance, the loss function \( \mathcal{L}_{total} \) for a single image is computed as follows
\[
\mathcal{L}_{total} = \mathcal{L}_{keypoint} + \mathcal{L}_{grouping}
\]  
(39)

where \( \mathcal{L}_{keypoint} \) represents the complete intersection over union (CIoU) of keypoint between ground truth regions and predicted regions. \( \mathcal{L}_{grouping} \) represents the loss function of associative embedding.

### 4. Simulation

To verify the proposed control method in controlling a continuum laparoscope to adjust the FOV, a set of simulations are conducted in the instrument tracking task using Mujoco, which provides a simulation environment and physiomechanics. As shown in Figure 4a, a continuum laparoscope model is established first. This model consists of 4 joints, and the size of the simulated continuum laparoscope is same as the real ones. The two cylinders in Figure 4b represent the surgical instruments. The red markers represent the target keypoints on the surgical instruments. There is a fixed global camera in the simulation environment to monitor the position of tracking point and the FOV center of the continuum laparoscope.

Simulation performances are compared to determine the appropriate parameters in the LQR controller. Then two types of simulation tasks were performed to evaluate the performance by using the proposed method through tracking instruments in static and moving, respectively. For each type of simulation task,

![Figure 2. Keypoints and skeletons of surgical instruments. a) Large needle driver. b) Grasping retractor.](image_url)
The proposed method is evaluated in two surgical scenarios with single and double instruments. Finally, the tracking error caused by constructing the observation function of Koopman operator using Taylor series and the proposed shifted Chebyshev polynomials were compared.

The state of the automatic tracking system is

$$s = [h, w, d]^T,$$

where $h$ and $w$ are the pixel positions of tracking point in the image plane, and $d$ is the distance between the tracking point and the center of FOV. $u = [u_1, u_2]^T$ represents the rotation angle of motors that applied to the continuum laparoscope. The parameter $M$ we choosing in Equation (35) is 5 in simulation. The penalty coefficient $\lambda$ is 10. The weight metrics $Q$ and $R$ in the LQR controller mainly affects the convergence speed of the system state and the error with the target state. In order to choose the appropriate weight metrics in the LQR controller, repeated simulations are performed with different $Q$ and $R$ from the same initial position of a single static surgical instrument, respectively. As shown in Figure 5, the cumulative error to reach the equilibrium state of the system is shorter, and likewise, the error of the system at the equilibrium state is larger. Considering the convergence speed and error of the system state in the equilibrium state, the weight metrics $Q$ and $R$ in the LQR controller are chosen as $\text{diag}(10, 10, 1)$ and $\text{diag}(1, 1)$, respectively.

### 4.1. Tracking the Static Instruments

First, we evaluated the proposed method by tracking a single stationary surgical instrument from different initial positions. The continuum laparoscope moved automatically under the proposed control methods until the tracking point reached the center of FOV. The pixel coordinate of the laparoscopic FOV center was $(200, 200)$ in the image plane. As shown in Figure 6a, several repeat trials were performed, while the instrument was located at different initial positions in four different areas of the image plane. The initial positions were marked as black points. All initial positions are chosen randomly, so the selected initial positions are basically representative of the entire workspace. The distance between the tracking point and the center of FOV is also called the tracking error. The time interval between each sample data is 0.05 s. The continuum laparoscope can be controlled to adjust the view until the tracking point is located at the center of FOV in the image plane. As shown in Figure 6b, the distance-step curve oscillates in the first few steps, and then the states of the system will become stable with the increase of steps. After approximately 25 steps, the tracking point locates at the center of FOV with an error of 10 pixels. This status also indicates that tracking is successful, and the continuum laparoscope maintains the current state unless the instrument moves. The errors of the three repeat trials are almost the same, which shows the stability of the proposed optimal control methods.

Then we simulate the double static surgical instruments tracking task while the instruments have random positions. We set the center of the two detected keypoints on each instrument as the new tracking point. As shown in Figure 6c,d, the tracking point moves relatively from the initial state (black point) to the view center. The tracking error is approximately 4 pixels.
4.2. Tracking the Moving Instruments

In this section, moving instruments tracking will be conducted. Figure 7a shows the trajectory of the tracking point on a single instrument. When the surgical instrument moves in a circular trajectory, the FOV center of the continuum laparoscope in the image plane of the globe camera also moves along with the surgical instrument. It is seen from the global camera that the continuum laparoscopy can track the keypoints with the movement of the surgical instrument. Figure 7b shows the distance between the tracking point and center of FOV in the image plane when tracking the moving instrument.

Another simulation of tracking two moving surgical instruments is further performed. Figure 7c shows trajectory while tracking double instruments. The center of the two detected keypoints on the instruments is set as the tracking point. Figure 7d shows the distance between the tracking point and center view while tracking two instruments simultaneously. In addition, Figure 7d also shows the distance from the detected tracking points on the two surgical instruments to the view center in the simulation environment.
The simulation results show the feasibility of our approach while controlling a continuum laparoscope by tracking moving surgical instruments. In addition, the laparoscopic FOV can be well adjusted based on the position of the tracking point. The simulation results also show that the continuum laparoscope can provide a stable FOV. The distances between the tracking point and the center of FOV while tracking a single moving instrument and two moving instruments are approximately 6.02 and 5.84 pixels, respectively.

4.3. Comparison with Taylor-Based Method

In this section, we will compare the tracking error while using the Taylor series and Chebyshev polynomials in constructing observation functions of the Koopman operator.

Figure 8a shows the comparison between the two methods while tracking a static instrument. Chebyshev polynomials used in our method are more complex. Thus, reaching the balanced state of the system (the surgical instrument appears in the center of FOV) is slower than using the Taylor series method. On the other hand, the error of our method is smaller than that of the Taylor-based method. We then compare the distances between the tracking point and center of FOV while tracking a moving surgical instrument using the two methods. As shown in Figure 8b, the tracking error of the Taylor-based method is 14.76 pixels, whereas our Chebyshev-based method has an error of 11.23 pixels. The small error variation shows that our Chebyshev-based method can provide a more stable FOV by controlling the laparoscope automatically.

5. Experimental Section

Experimental Setup: An experiment platform that is consistent in the simulation environment is built using a continuum laparoscope based on the proposed automatic tracking system to validate the data-driven control method. As shown in Figure 9, a 2 mm diameter pinhole camera with a resolution of 400 × 400 pixels is fixed at the end effector of the cable-driven continuum manipulator (Intuitive Surgical, California, USA). Sensing image can be obtained with a frequency of 30 Hz through a USB port. The continuum manipulator has four connected joints, which can be divided into two groups. Joint1 and Joint4 control the movement in the X-axis direction, and both are actuated by a brushless motor (Maxon Group, Sachseln, Switzerland). Joint2 and Joint3 control the movement in the Y-axis direction and are actuated by another brushless motor. The continuum manipulator is fixed in the Z-axis direction, which could ensure safety once the initial position is determined. Elmo drivers (Elmo Motion Control Ltd., Israel) are used to actuate motors precisely by receiving command from TwinCAT3 (Beckhoff Automation GmbH & Co. KG, Germany) through the EtherCAT bus. The large needle driver and the grasping retractor (Intuitive Surgical, California, USA) are used in this work. The large needle driver holds the needle while forcing the tip through...
the tissue to complete the suturing task in RMIS. Grasping retractor is used to retract the tissue to reveal the surgical scope so that the surgeon can explore the surgical area and perform surgery.

Data Collection and Implementation: First, in order to train the learning-based pose estimation model, we collected an in-house dataset containing 3000 images with the two surgical instruments using the experimental platform. The in-house dataset is labeled following the label rules for the pose estimation task by using the LabelImg software. We performed data augmentations to enlarge the dataset and prevent the overfitting problem. Mosaic, horizontal flips, random crop, color transform, and translation are used as augmentation skills in this work. The hyperparameters during training are as follows: the training step is 300, the batch size is 32, the polynomial decay learning rate scheduling strategy is adopted with an initial learning rate of 0.1, the momentum is 0.9, and the weighted decay is 0.005. We implemented our method using Pytorch on an Ubuntu 18.04 LTS workstation with an Intel Xeon(R) 2.30 GHz CPU and an NVIDIA Tesla P100 16 GB graphics card. The implementation takes around 6.5 h to train the entire framework. The processing speed achieves an average of 34.3 fps for videos.

Second, in order to calculate the infinite-dimensional approximation of the Koopman Operator, a set of system states and
corresponding inputs were collected. We collected 100 trials with random initial inputs of the motors. The state of the automatic tracking system and the input of two motors are the same as that selected in the previous simulation environment. Then the input of the two motors varies one cycle according to the trigonometric function to generate the states. Following this rule, the continuum laparoscope runs 500 steps a cycle in the workspace. Finally, we collected 50,000 pairs of data about system states and inputs. The Koopman operator of the continuum laparoscope system is estimated according to the collected system states and inputs in Equation (7). Then the dynamic model of the system in Equation (12) is determined. Parameters used in the LQR are the same as that selected in the simulation environment. Parameter $M$ in Equation (35) is 5, and the penalty coefficient $\lambda$ is 10. The weight metrics $Q$ and $R$ in the LQR controller are $\text{diag}(10, 10, 1)$ and $\text{diag}(1, 1)$, respectively. According to the control strategy of LQR control, states of the continuum laparoscope system will change with inputs of the motors. Once the updated system state is determined, the new input of system can be calculated using Equation (14). Then the motors are controlled through the Elmo drivers and turn to the specified angle, thus achieving the goal of adjusting the FOV of the continuum laparoscope.

Furthermore, we also evaluated the dynamic uncertainty of the continuum laparoscope system based on the data collection. Dynamical uncertainty means that the same inputs may lead to slightly different outputs. This uncertainty is caused by motor encoder error and system assembly error. Specifically, the system gets an input that changes periodically. The output of the system over 100 cycles is then collected, and the standard error over 100 cycles is calculated. The standard error of the system is 19 pixels. The internal error of this system will affect our tracking experiment, which cannot be eliminated by the designed controller.

Experiments on Keypoint Detection: Motion blur, specular reflections of surgical instruments, blood, surgical smoke, and tissue occlusions are the main challenges in medical image analysis, which are also applicable to the keypoint detection task of surgical instruments. In order to evaluate the capability of the proposed pose estimation model in facing these challenges, experiments are first performed on a public EndoVis Challenge dataset. This dataset is collected from several laparoscopic colorectal surgeries. A large needle driver is also used in this dataset, which is the same as the in-house collected dataset. The EndoVis Challenge dataset contains 1850 images: 940 for training and 910 for testing. The frame resolution is $720 \times 576$ pixels. To evaluate our proposed framework, we use a standard metric, namely, object keypoint similarity (OKS) for the pose estimation task. We report the standard average precision (AP) and average recall (AR) scores, including AP$^{0.5}$ (AP at OKS = 0.5), AP$^{0.75}$ (AP at OKS = 0.75), AP (mean of AP scores from OKS = 0.50 to OKS = 0.95 with the increment as 0.05), AP$^M$ (AP scores for the instrument of medium sizes), AP$^L$ (AP scores for instruments of large sizes), and AR (the mean of AR scores from OKS = 0.50 to OKS = 0.95 with the increment as 0.05). Results are shown in Table 1. Figure 10 more intuitively shows the pose estimation results of the surgical instruments on the two datasets. More intuitive results of instruments are available in the Video SI, Supporting Information.

Table 1. Results of the proposed keypoint detection model on the public dataset and in-house dataset with metrics of object keypoint similarity.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>AP</th>
<th>AP$^{0.5}$</th>
<th>AP$^{0.75}$</th>
<th>AP$^M$</th>
<th>AP$^L$</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-house dataset</td>
<td>79.6</td>
<td>93.1</td>
<td>87.6</td>
<td>62.8</td>
<td>82.4</td>
<td>75.3</td>
</tr>
<tr>
<td>Public dataset</td>
<td>83.0</td>
<td>95.0</td>
<td>88.7</td>
<td>64.5</td>
<td>85.7</td>
<td>78.2</td>
</tr>
</tbody>
</table>

Figure 10. Examples of keypoint detection results using the proposed learning-based pose estimation method on the two datasets. From top to bottom, we present the public dataset and the in-house dataset.
Besides, in order to compare the proposed method with other methods, we also used the same evaluation criteria, as follows: mean average precision (mAP), mean localization error (mLE), and detection time. As shown in Table 2, our method achieves 96.27% of the mAP, 1.1 pixels of the mLE, and 34.3 fps. Experimental results show that our keypoint detection method based on pose estimation task can obtain relatively high precision of keypoint position and avoid the shortcomings of other methods mentioned in Section 3.1.

Through the learning-based pose estimation method, the pixel positions of the different parts of the surgical instrument could be obtained. For different surgical phases, different surgical instruments are used. We can adjust the tracking point automatically for a better view due to the difference between the shapes of the instruments. Tips of the surgical instrument are usually close to the target tissue to be operated. The target tissue located in the center of FOV can bring a better surgical experience to the surgeon. Considering the effect of the large needle driver and the grasping retractor and tissue occlusions in surgery, we choose the head as the reference point for a better view. In addition, we set the weight of the two surgical instruments in the FOV to 1:1 while tracking two instruments.

**Experiments on Tracking Static Instruments:** Experiments on static surgical instruments are performed to evaluate the autonomous laparoscopic control method to adjust the FOV of a continuum laparoscope. Repeat trials are conducted for different surgical instruments. Only the initial position of the surgical instrument is different. As shown in Figure 11a, the black point represents the initial position of the tracking keypoint on the surgical instrument. The scatters mean the position of the tracking point relative to the center of the FOV at each step. Figure 11b shows the distance between the tracking point and the FOV center of the laparoscope. When the continuum laparoscopy was controlled to adjust the FOV based on visual feedback automatically, it needs approximately 25 steps to approach the center of the FOV. This is consistent with our verification in the simulation environment. After 25 steps, the continuum laparoscopy remains largely stationary, indicating that our approach provides a stable FOV after reaching the tracking purpose. The tracking error is approximately 39.1 pixels when the system is stable.

Figure 11c shows the keypoints on the two surgical instruments and the pixel position of our tracking points relative to the center of the FOV at each step. Notably, the weights of both surgical instruments in the FOV are the same. Figure 11d shows changes of the tracking error and relative position of the two surgical instruments, which is consistent with the performance when a single surgical instrument tracking task is performed.

---

**Table 2.** Results of the proposed pose estimation model compared with other methods.

<table>
<thead>
<tr>
<th>Methods</th>
<th>mAP</th>
<th>mLE (pixel)</th>
<th>fps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker</td>
<td>86.8%</td>
<td>12.7</td>
<td>45.46</td>
</tr>
<tr>
<td>YOLOv3</td>
<td>92.64%</td>
<td>1.6</td>
<td>31.25</td>
</tr>
<tr>
<td>Pose estimation</td>
<td>96.27%</td>
<td>1.1</td>
<td>34.3</td>
</tr>
</tbody>
</table>

---

**Figure 11.** Experimental results while the continuum laparoscope tracking the static instruments (large needle driver and grasping retractor). Black points mean the initial position of tracking point in FOV. a) Tracking the two surgical instruments, respectively, with different initial position. b) Distances between the tracking point and the center of FOV while tracking the two surgical instruments, respectively. c) Tracking the two surgical instruments simultaneously. d) Distances between the tracking point (center of the two instruments) and the center of FOV.
The performance of tracking static surgical instruments proves the feasibility of our proposed data-driven control approach.

Experiments on Tracking Moving Instruments: First, we evaluate the proposed method with one moving surgical instrument. As shown in Figure 12a, the scatters represent the relative position of the tracking points on the surgical instrument in the FOV of the continuum laparoscope. The number of color bar means the density of the tracking points in the image plane. The higher the value, the more times the tracking point locates in the area with the movement of the surgical instrument. It is seen that most of the tracking points locate near the center of the FOV. The distance between the tracking point and the FOV center of the laparoscope is shown in Figure 12b. The average distance while tracking a moving surgical instrument is approximately 45.77 pixels.

Second, we extend the tracking task to two surgical instruments. Similar to the setting in Section 5.4, we set the same weights for the two surgical instruments in the tracking process. The same weights mean that the tracking point is located at the midpoint of the key-point on the two surgical instruments. Figure 12c shows the relative pixel position of the tracking points in the image plane. Figure 12d shows tracking errors while tracking double surgical instruments. The average distance is about 28.47 pixels.

Discussion with Clinical Requirements: Researchers reported the clinical requirements for adjusting the FOV of the laparoscope with the movement of surgical instruments.[7] The researchers analyzed collected videos of different clinical surgeries. They found that the average distance from the tip of the surgical instruments to the FOV center of the laparoscope is about 423.84 pixels, even though the surgical assistant was constantly adjusting the laparoscopic field of vision during the procedure. These errors are mainly caused by the lack of understanding between assistant and surgeons. Considering the resolution of collected surgical videos, the error is approximately 22.08% of the horizontal resolution and 39.24% of the vertical resolution. In addition, the average moving speed of surgical instruments during the surgery in FOV of the laparoscope is about 131.95 pixel s⁻¹.

We have demonstrated the feasibility of our proposed autonomous laparoscopic control approach to adjust FOV regardless of whether the surgical instruments are moving. However, the fast movement of surgical instruments leads to the accuracy of keypoint detection and then affects the accuracy of surgical instrument tracking. In the experiment, the average moving speed of surgical instruments is about 173.54 pixels s⁻¹, which is sufficient in clinical surgery. From the experimental results on the automatic tracking task of surgical instruments, the distances between the tracking point and the center of FOV while tracking single moving instruments and double moving instruments are about 45.77 and 28.47 pixels, respectively. The results are approximately 11.44% and 7.12% of our continuum laparoscopic FOV, which is much smaller than the FOV error of clinical surgery. Considering the dynamic uncertainty of the continuum laparoscope system, the internal error of this system is about 19 pixels. This indicates the proposed autonomous FOV adjusting method with a continuum laparoscope can satisfy the clinical requirements. In addition to laparoscopic FOV adjustment as part of the surgical procedure, our approach can promote the automation of the robotic surgery process.

Figure 12. Pixel position of the moving surgical instruments (large needle driver and grasping retractor) compared with initial position. Color bar represents the position density. a) Pixel coordinates density of the tracking point while the single surgical instrument moving. b) Distances between the target position and center of the view while tracking the single surgical instrument. c) Pixel coordinates density of the tracking point while the double surgical instrument moving simultaneously. d) Distances between the target position (center of the two instruments) and center of the view.
6. Conclusion

This article presents a data-driven control method for a continuum laparoscope system with learning-based visual feedback to adjust the FOV automatically in RMIS. We first developed a nonlinear system identification method using the Koopman operator and Chebyshev polynomials. Then we build an LQR controller based on the trained Koopman Operator with the visual feedback. A learning-based keypoint detection method is designed without manual markers to provide precise visual feedback for the control system. This method provides more options for selecting keypoints on surgical instruments while ensuring detecting accuracy during the different surgical processes. Simulation and experiments are performed to evaluate the proposed methods. The pose estimation method provides higher accuracy than other keypoint detection methods. Tracking experiments show the feasibility of the proposed data-driven control method of a continuum laparoscope for adjusting the FOV automatically. Experimental results show that the proposed method can also satisfy the clinical requirements. In the future, the freedom of continuum laparoscope in Z-axis direction to enlarge the view will be considered to provide surgeons a better experience. Constrained workspaces and inputs will be studied to ensure safety during robotic surgery. In addition, we will further explore scene segmentation in surgical images or videos. According to the results of scene segmentation, the target tissue was chosen as visual feedback instead of surgical instruments to adjust the FOV of the laparoscope automatically.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the corresponding author upon reasonable request.

Acknowledgements

X.K. and H.M. contributed equally to this work. This work was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Ref. no. T21-409/18R and no.12111421).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

continuum laparoscope, data-driven control, keypoint detection, robotic surgery, visual feedback

Received: July 2, 2022
Revised: October 2, 2022
Published online: December 23, 2022


