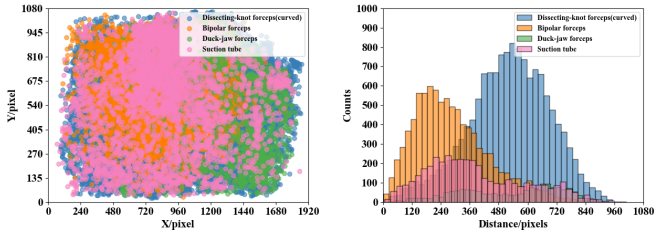
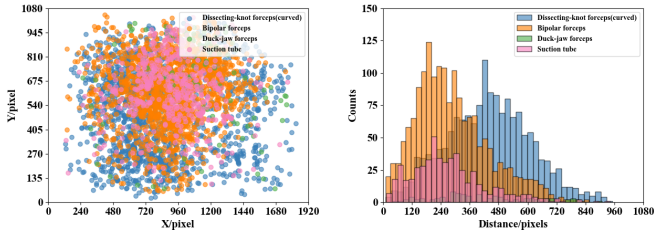


(a) (b)

Fig. 1. Tips of the surgical instrument statistical results. (a) Position distribution in the image. (b) Distance from the image center to the tip center.



(a) (b)

Fig. 2. Tips of the surgical instrument statistical results when there is only one surgical instrument. (a) Position distribution in the image. (b) Distance from the image center to the tip center.

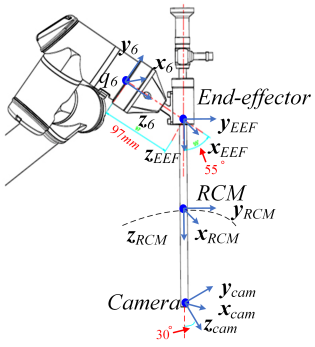


Fig. 3. Schematic diagram of laparoscope installation.

Appendix A
Clinical Data Analysis

The laparoscopic video data obtained by the first-class hospitals at grade 3 in China, including the First Affiliated Hospital of Anhui Medical University and the First Affiliated Hospital of the University of Science and Technology of China, were analyzed. Four types of minimally invasive surgeries, including hysteromyomectomy, cholecystectomy, sleeve gastrectomy, and radical cystectomy, were selected for detailed analysis. It was found that one to four types of surgical instruments are usually used simultaneously during the operations. Dissecting-knot forceps(curved), bipolar forceps, duck-jaw forceps, and suction tube are the most commonly used surgical instruments in these four surgeries. The position distribution diagram of the surgical instrument tip in the image was obtained, and it is shown in Fig. 1(a). The distance from the image center to the tip center is shown in Fig. 1(b). When there is only one surgical instrument in the laparoscopic image, its pixel distribution in the image and the distance from the image center to the tip center are as shown in Figs. 2(a) and 2(b), respectively.

In addition, the statistical values of distances between the image center and the tip center, including the 2nd percentile, the 98th percentile, quartile, and mean, was calculated. The results are shown in Table I. The distance of the tip center from the image center was analyzed for five cases: cumulative case (i.e., all images were considered, regardless of the number of surgical instruments), one-instrument case (images where only one instrument was used were considered), two-instruments case (images where two instruments were used were considered), three-instruments case (images where three instruments were used were considered), and four-instruments case (images where four instruments were used were considered). For each of these cases, four different types of instruments were tracked (i.e., dissecting-knot forceps(curved), bipolar forceps, duck-jaw forceps, and suction tube), and a non-instrumentally-discriminative sub-total was calculated. For each instrument in Table I, the first line describes the six statistics of the distance between the image center and tip center, while the second line describes the radio of the distance to the maximum distance that the instrument in the image can achieve.

Appendix B
Kinematics

During laparoscopic surgery, laparoscope held by UR3 is inserted into the patient's body through a small incision point, which is presented in Fig. 3. We study the kinematic model with fixed RCM constraint to relate the velocities of endpoint of laparoscope with each joint of UR3. The joint variable of UR3 is denoted as q_i(for $i = 1, 2, \ldots, 6$),and $\boldsymbol{q} = \begin{bmatrix} q_1 & q_2 & q_3 & q_4 & q_5 & q_6 \end{bmatrix}$. Please note that all the six joints of UR3 revolute.

Suppose the pedestal of UR3 as the base frame, i.e., frame 0. The position and orientation of the frame i with respect to the frame $i-1$ is $\boldsymbol{A}_i^{i-1}(\theta_i)$. $\boldsymbol{T}_{EEF}^6$ is the coordinate transformation that describes the position and orientation of end-effector frame with respect to frame 6. Therefore, the position and orientation of end-effector relative to the base frame is a function of $\boldsymbol{q}$, and is calculated as:

$$\begin{aligned} \boldsymbol{T}_{EEF}^0 = f(\boldsymbol{q}) &= \prod_{i=1}^{6} \boldsymbol{A}_i^{i-1}(q_i)\boldsymbol{T}_{EEF}^6 \\ &= \begin{bmatrix} \boldsymbol{x}_{EEF} & \boldsymbol{y}_{EEF} & \boldsymbol{z}_{EEF} & \boldsymbol{p}_{EEF} \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned}, \tag{B.1}$$

where $\boldsymbol{T}_{EEF}^0$ and $\boldsymbol{T}_{EEF}^6$ represent the coordinate transformations describing the position and orientation of end-effector frame with respected to the frame 0 and frame 6, respectively. $\boldsymbol{x}_{EEF}$, $\boldsymbol{y}_{EEF}$, and $\boldsymbol{z}_{EEF}$ represent the unit vectors of a frame attached to the end-effector. $\boldsymbol{p}_{EEF}$ represents the position vector of end-effector at frame 0.

The coordinate transformations describing the position and orientation of the camera frame of laparoscope with respect

TABLE I
STATISTICAL VALUES OF DISTANCES BETWEEN THE IMAGE CENTER AND THE TIP CENTER

Case	Instrument	2nd percentile	25th percentile	50th percentile	75th percentile	98th percentile	mean
Cumulative case	Dissecting-knot forceps(curved)	134.42	415.69	529.47	641.08	835.67	521.24
		0.12	0.38	0.48	0.58	0.76	0.47
	Bipolar forceps	47.67	169.41	261.11	376.97	676.16	286.71
		0.04	0.15	0.24	0.34	0.61	0.26
	Duck-jaw forceps	116.5	349.93	555.04	693.34	851.9	522.8
		0.11	0.32	0.5	0.63	0.77	0.47
	Suction tube	59.3	227.66	347.32	531.93	838.26	386.85
		0.05	0.21	0.32	0.48	0.76	0.35
	Total	69.21	261.12	418.05	578.72	813.93	423.84
		0.06	**0.24**	**0.38**	**0.53**	**0.74**	**0.38**
One-instrument case	Dissecting-knot forceps(curved)	98.33	315.35	435.69	555.96	807.25	438.39
		0.09	0.29	0.4	0.5	0.73	0.4
	Bipolar forceps	39.71	175	260.46	373.51	659.61	287.19
		0.04	0.16	0.24	0.34	0.6	0.26
	Duck-jaw forceps	41	282.24	363.85	555.43	801.39	414.44
		0.04	0.26	0.33	0.5	0.73	0.38
	Suction tube	38.91	166.87	246.57	345.49	674	272.81
		0.04	0.15	0.22	0.31	0.61	0.25
	Sub-total	52.74	211.26	327.41	469.55	750.56	350.24
		0.05	**0.19**	**0.3**	**0.43**	**0.68**	**0.32**
Two-instruments case	Dissecting-knot forceps(curved)	115.16	388.17	493.6	601.01	813.71	490.4
		0.1	0.35	0.45	0.55	0.74	0.45
	Bipolar forceps	50.3	168.69	260.87	376.43	657.06	283.72
		0.05	0.15	0.24	0.34	0.6	0.26
	Duck-jaw forceps	112.07	331.11	504.51	698.38	853.33	510.93
		0.1	0.3	0.46	0.63	0.77	0.46
	Suction tube	56.06	204.8	294.05	392.51	768.74	323.1
		0.05	0.19	0.27	0.36	0.7	0.29
	Sub-total	65.91	239.53	378.34	528.36	786.98	390.26
		0.06	**0.22**	**0.34**	**0.48**	**0.71**	**0.35**
Three-instruments case	Dissecting-knot forceps(curved)	188.75	478.22	580.98	682.55	851.63	570.87
		0.17	0.43	0.53	0.62	0.77	0.52
	Bipolar forceps	48.62	166.37	264.97	384.88	702.32	293.35
		0.04	0.15	0.24	0.35	0.64	0.27
	Duck-jaw forceps	107.17	360.12	575.09	696.19	851.95	531.38
		0.1	0.33	0.52	0.63	0.77	0.48
	Suction tube	127.65	390.08	541.91	687.05	883.3	534.02
		0.12	0.35	0.49	0.62	0.8	0.48
	Sub-total	79.44	320.79	502.96	643.89	843.91	480.6
		0.07	**0.29**	**0.46**	**0.58**	**0.77**	**0.44**
Four-instruments case	Dissecting-knot forceps(curved)	244.75	490.15	574.5	677.28	851.11	578.93
		0.22	0.45	0.52	0.61	0.77	0.53
	Bipolar forceps	48.59	167.11	241.75	362.1	710.22	283.83
		0.04	0.15	0.22	0.33	0.64	0.26
	Duck-jaw forceps	138.86	375.03	590.82	693.34	848.8	539.1
		0.13	0.34	0.54	0.63	0.77	0.49
	Suction tube	126.89	368.3	519.68	712.5	901.86	526.68
		0.12	0.33	0.47	0.65	0.82	0.48
	Sub-total	105.35	352.85	522.93	659.77	847.22	499.17
		0.1	**0.32**	**0.47**	**0.6**	**0.77**	**0.45**

to the end-effector frame is $\boldsymbol{T}_{cam}^{EEF}$, so the position and orientation of camera relative to the base frame is:

$$\begin{aligned}\boldsymbol{T}_{cam}^{0} &= \boldsymbol{T}_{EEF}^{0}\boldsymbol{T}_{cam}^{EEF} \\ &= \begin{bmatrix} \boldsymbol{x}_{cam} & \boldsymbol{y}_{cam} & \boldsymbol{z}_{cam} & \boldsymbol{p}_{cam} \\ 0 & 0 & 0 & 1 \end{bmatrix},\end{aligned} \tag{B.2}$$

where $\boldsymbol{x}_{cam}$, $\boldsymbol{y}_{cam}$, and $\boldsymbol{z}_{cam}$ are the unit vectors of a frame attached to the camera, $\boldsymbol{p}_{cam}$ is the position vector of camera at frame 0.

Considering fixed RCM, we have:

$$\begin{cases} \boldsymbol{p}_{EEF} &= \boldsymbol{p}_{cam} - L\frac{\boldsymbol{p}_{cam}-\boldsymbol{p}_{RCM}}{\|\boldsymbol{p}_{cam}-\boldsymbol{p}_{RCM}\|} \\ \boldsymbol{z}_{EEF} &= \frac{\boldsymbol{p}_{cam}-\boldsymbol{p}_{RCM}}{\|\boldsymbol{p}_{cam}-\boldsymbol{p}_{RCM}\|} \end{cases}, \tag{B.3}$$

where $\boldsymbol{p}_{RCM}$ represents the position vector of RCM at frame 0. L represents the shaft length of laparoscope.

Hand-eye coordination is critical to surgeons in laparoscopic surgery. During laparoscopic surgery, the orientation of the endoscope should remain relatively fixed. To ensure the stability of the picture captured by the laparoscope, the $\boldsymbol{x}_{EEF}$ is fixed as:

$$\boldsymbol{x}_{EEF} = \begin{cases} \frac{\boldsymbol{y}_0\times\boldsymbol{z}_{EEF}}{\|\boldsymbol{y}_0\times\boldsymbol{z}_{EEF}\|} & \|\boldsymbol{y}_0\times\boldsymbol{z}_{EEF}\| \neq 0 \\ \boldsymbol{x}_0 & \|\boldsymbol{y}_0\times\boldsymbol{z}_{EEF}\| = 0 \end{cases}. \tag{B.4}$$

where $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ are the X axis unit vector and Y axis unit vector at the base coordinates, respectively.

$$\boldsymbol{y}_{EEF} = \frac{\boldsymbol{z}_{EEF}\times\boldsymbol{x}_{EEF}}{\|\boldsymbol{z}_{EEF}\times\boldsymbol{x}_{EEF}\|}. \tag{B.5}$$

Substituting Eq. (B.3), (B.4) and (B.5), into (B.1), we note that $\boldsymbol{T}^0_{EEF}$ is a function of $\boldsymbol{p}_{cam}$, and is expressed as $\boldsymbol{T}^0_{EEF} = g(\boldsymbol{p}_{cam})$. By solving the inverse function, the inverse kinematics model of UR3 under RCM constraint is obtained:

$$\boldsymbol{q} = f^{-1}(g(\boldsymbol{p}_{cam})). \tag{B.6}$$

Differentiating the Eq.(B.6) with respect to time, we get:

$$\dot{\boldsymbol{q}} = \frac{\partial f^{-1}(g(\boldsymbol{p}_{cam}))}{\partial(g(\boldsymbol{p}_{cam})}\frac{\partial(g(\boldsymbol{p}_{cam})}{\partial\boldsymbol{p}_{cam}}\dot{\boldsymbol{p}}_{cam}, \tag{B.7}$$

where $\dot{\boldsymbol{q}}$ represents the velocity of each joint, and $\dot{\boldsymbol{p}}_{cam}$ represents the velocity of camera. According to Eq. (B.7), by setting the velocity of each joint, the desired camera velocity can be obtained.

APPENDIX C
ARCHITECTURE OF MODIFIED SURGICAL INSTRUMENT DETECTION ALGORITHM

The architecture of surgical instrument detection algorithm is shown in Table II, and consists of convolutional layers, residual blocks, convolutional block, multiscale blocks and yolo layers. The convolutional block is composed of a plurality of convolutional layers. The multiscale block is composed of a upsample layer, a route layer and a plurality of convolutional layers, and is used to multi-scale feature information extraction. Three yolo layers detect the tip of surgical instrument from three scales (i.e. 20×15,40×30,80×60).

APPENDIX D
PROOF OF THEOREM 1

A Lyapunov function candidate is defined as $V = d_i^2 = (u_i-u_0)^2+(v_i-v_0)^2$, and the first derivative of this exponent is:

$$\dot{V} = (u_i-u_0)\dot{u_i} + (v_i-v_0)\dot{v_i}. \tag{D.1}$$

TABLE II
ARCHITECTURE OF SURGICAL INSTRUMENT DETECTION

Type	Filter shape	Input shape
Convolutional layer	$640\times480\times32$	$640\times480\times3$
Convolutional layer	$320\times240\times64$	$640\times480\times32$
Residual block	$320\times240\times64$	$320\times240\times64$
Convolutional layer	$160\times120\times128$	$320\times240\times64$
2×Residual block	$160\times120\times128$	$160\times120\times128$
Convolutional layer	$80\times60\times256$	$160\times120\times128$
8×Residual block	$80\times60\times256$	$80\times60\times256$
Convolutional layer	$40\times30\times512$	$80\times60\times256$
8×Residual block	$40\times30\times512$	$40\times30\times512$
Convolutional layer	$20\times15\times1024$	$40\times30\times512$
4×Residual block	$20\times15\times1024$	$20\times15\times1024$
Convolutional block	$20\times15\times21$	$20\times15\times1024$
Yolo	-	$20\times15\times21$
Multiscale block	$40\times30\times21$	$20\times15\times512$
Yolo	-	$40\times30\times21$
Multiscale block	$80\times60\times21$	$40\times30\times256$
Yolo	-	$40\times30\times21$

According to Eq. (8) and (9), the camera velocity is:

$$\boldsymbol{v}_i = \begin{bmatrix} v_x^i \\ v_y^i \\ v_z^i \end{bmatrix} = V_s^i(d_i)\begin{bmatrix} \frac{x_iz_i}{s} \\ \frac{y_iz_i}{s} \\ \frac{-x_i^2-y_i^2}{s} \end{bmatrix}. \tag{D.2}$$

where $s = \sqrt{(x_i^2+y_i^2)^2+x_i^2z_i^2+y_i^2z_i^2}$. The relative velocity of surgical instrument to camera $\boldsymbol{v}_{ins}^{'}$ is:

$$\boldsymbol{v}_{ins}^{'} = \begin{bmatrix} \dot{x}_i - v_x^i + \frac{v_y^iz_i}{l} \\ \dot{y}_i - v_y^i + \frac{v_x^iz_i}{l} \\ \dot{z}_i - v_z^i + \frac{v_y^ix_i}{l} + \frac{v_x^iy}{l} \end{bmatrix}. \tag{D.3}$$

where $l = \|\boldsymbol{p}_{cam}-\boldsymbol{p}_{RCM}\|$. Substituting Eq. (D.3) into (11) in the body part and (12) in the body part yields the following equations:

$$\dot{u_i} = \frac{f_x(\dot{x}_iz_i-dotz_1x_i-V_s^i(d_i)(x_iz_1^2+x_1^3+x_iy_1^2))/s}{z_i^2} + \frac{f_xV_s^i(d_i)(y_iz_i^3-2x_i^2y_iz_i)/ls)}{z_i^2} \tag{D.4}$$

$$\dot{v_i} = \frac{f_y(\dot{y}_iz_i-dotz_1y_i-V_s^i(d_i)(y_iz_i^2+y_i^3+y_ix_i^2))/s}{z_i^2} + \frac{f_yV_s^i(d_i)(x_iz_i^3-2y_i^2x_iz_i)/ls)}{z_i^2} \tag{D.5}$$

Substituting Eq. (D.4) and (D.5) to (D.1), we get:

$$\dot{V} = \frac{f_x^2x_i}{z_i^3}(\dot{x}_iz_i - \dot{z}_ix_i - \frac{V_s^i(d_i)x_iw}{s} + \frac{V_s^i(d_i)y_iz_i(z_i^2-x_i^2)}{ls}) + \frac{f_y^2y_i}{z_1^3}(\dot{y}_iz_i - \dot{z}_iy_i - \frac{V_s^i(d_i)y_iw}{s} + \frac{V_s^i(d_i)x_iz_i(z_1^2-y_1^2)}{ls}), \tag{D.6}$$

where $w = x_i^2 + y_i^2 + z_i^2$. When $l \to +\infty$, the maximum speed of surgical instrument that the proposed vector has the ability to track is:

$$\lim_{l\to+\infty} \dot{V} = \frac{f_x^2}{z_i^3} C + \frac{f_y^2}{z_i^3} E, \tag{D.7}$$

$$C = \dot{x}_i x_i z_i - \dot{z} x_i^2 - \frac{V_s^i(d_i) x_1^2 w}{s}, \tag{D.8}$$

$$E = \dot{y}_i y_i z_i - \dot{z} y_i^2 - \frac{V_s^i(d_i) y_1^2 w}{s}. \tag{D.9}$$

Because,

$$\begin{aligned} C &= \dot{x}_i x_i z_i - \dot{z} x_i^2 - \tfrac{V_s^i(d_i) x_1^2 w}{s} \\ &= x_i^2 (\dot{x} \tfrac{z_i}{x_i} - \dot{z} - V_s^i(d_i) \sqrt{\tfrac{w}{x^2+y^2}}) \\ &\leq x_1^2 \left[\max\{\|\boldsymbol{v}_{ins}\| \tfrac{z_i}{\|x_i\|}, \|\boldsymbol{v}_{ins}\|\} - V_s^i(d_i) \sqrt{\tfrac{w}{x^2+y^2}} \right] \end{aligned} \tag{D.10}$$

Case 1: $\frac{z_i}{\|x_i\|} \leq 1$, then:

$$\begin{aligned} C &\leq x_1^2 (\|\boldsymbol{v}_{ins}\| - V_s^i(d_i) \sqrt{\tfrac{w}{x_i^2+y_i^2}}) \\ &< x^2 (\|\boldsymbol{v}_{ins}\| - V_s^i(d_i)) \end{aligned} \tag{D.11}$$

Case 2: $\frac{z_i}{\|x_i\|} > 1$, then:

$$\begin{aligned} C &\leq x_1^2 (\|\boldsymbol{v}_{ins}\| \tfrac{z_i}{\|x_i\|} - V_s^i(d_i) \sqrt{\tfrac{w}{x_i^2+y_i^2}}) \\ &< x_i^2 (\|\boldsymbol{v}_{ins}\| \tfrac{z_i}{\|x_i\|} - V_s^i(d_i) \sqrt{\tfrac{x_i^2+z_i^2}{x_i^2}}) \\ &< x_i^2 (\|\boldsymbol{v}_{ins}\| \tfrac{z_i}{\|x_i\|} - V_s^i(d_i) \tfrac{z}{\|x_i\|}) \\ &= z_i \|x_i\| (\|\boldsymbol{v}_{ins}\| - V_s^i(d_i)) \end{aligned} \tag{D.12}$$

When $d_i = R$, $V_s^i(d_i) = V_{max}$. Considering that $\|\boldsymbol{v}_{ins}\| \leq V_{max}$, there is $C < 0$. Besides, when $d_i \in (r, R)$ $\frac{dV_s^i(d_i)}{dd_i} < 0$, so $\frac{dC}{dd_i} > 0$. $\exists \tau_1 \in (r, R)$, such that $C = 0$. Similarly, $\exists \tau_2 \in (r, R)$, such that $E = 0$. When $\tau = \max\{\tau_1, \tau_2\}$, there is $\dot{V} < \lim_{l\to+\infty} \dot{V} \leq 0$. Hence, $\exists \tau \in (r, R)$, such that $\dot{d}_i < 0$ when $d_i \in (\tau, R)$.

Appendix E
Proof of Theorem 2

According to definition 2, $\exists F_1 \in (0, \lambda r(\theta, \varphi))$, when $\forall l < F_1$, $V_c = V_{max}$.

$$\begin{aligned} \dot{l} &= \textstyle\sum_{i=1}^n \gamma_i (\boldsymbol{v}_i \cdot \boldsymbol{z}_{EEF} + \frac{\boldsymbol{v}_c}{\boldsymbol{z}_{EEF}}) \\ &= \textstyle\sum_{i=1}^n \gamma_i (\boldsymbol{v}_i \cdot \boldsymbol{z}_{EEF} + V_c) \\ &\geq \textstyle\sum_{i=1}^n \gamma_i (-V_{max} + V_{max}) = 0 \end{aligned} \tag{E.1}$$

Empathy, $\exists F_2 \in (L_1, r)$, when $\forall l > F_2$, $V_c = -V_{max}$.

$$\begin{aligned} \dot{l} &= \textstyle\sum_{i=1}^n \gamma_i (\boldsymbol{v}_i \cdot \boldsymbol{z}_{EEF} + \frac{\boldsymbol{v}_c}{\boldsymbol{z}_{EEF}}) \\ &= \textstyle\sum_{i=1}^n \gamma_i (\boldsymbol{v}_i \cdot \boldsymbol{z}_{EEF} + V_c) \\ &\leq \textstyle\sum_{i=1}^n \gamma_i (V_{max} - V_{max}) = 0 \end{aligned} \tag{E.2}$$

Appendix F
Detailed Description of Simulation

In the simulation, the robot base coordinate system is taken as the global coordinate system, and $\boldsymbol{P}_{RCM} = \begin{bmatrix} 350 & -350 & 60 \end{bmatrix}^T$, while the initial pose of end-effector is

$$\boldsymbol{T}_{EEF}^0 = \begin{bmatrix} 0 & -1 & 0 & 350 \\ -1 & 0 & 0 & -350 \\ 0 & 0 & -1 & 150 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{F.1}$$

The trajectory of the moving surgical instrument is expressed as:

$$\begin{bmatrix} 435 + 30\sin(\pi t) \\ -285 + 40\cos(0.5\pi t) \\ -184 + 15\cos(0.2\pi t) \end{bmatrix}. \tag{F.2}$$

In Sections IV.A and IV.B, the initial pose of the end-effector is the same as in Eq. (F.1).

In Section IV.A, RCM is located at $\boldsymbol{P}_{RCM} = \begin{bmatrix} 350 & -350 & 60 \end{bmatrix}^T$, and the trajectory of the moving surgical instrument is expressed as:

$$\begin{bmatrix} 430 + 25\sin(\pi t) \\ -380 + 25\cos(0.5\pi t) \\ -262 + 15\cos(0.5\pi t) \end{bmatrix}. \tag{F.3}$$

In Section IV.B, RCM is located at $\boldsymbol{P}_{RCM} = \begin{bmatrix} 350 & -350 & 40 \end{bmatrix}^T$, and the trajectory of the moving surgical instrument is expressed as:

$$\begin{bmatrix} 425 + 65\sin(0.35\pi t) \\ -249 + 65\sin(0.35\pi t) \\ -262 + 20\sin(0.35\pi t) \end{bmatrix}. \tag{F.4}$$

In Section IV.C-IV.E, RCM is located at $\boldsymbol{P}_{RCM} = \begin{bmatrix} 250 & -250 & -61 \end{bmatrix}^T$, and the initial pose of end-effector is expressed as:

$$\boldsymbol{T}_{EEF}^0 = \begin{bmatrix} 0 & -1 & 0 & 250 \\ -1 & 0 & 0 & -250 \\ 0 & 0 & -1 & 50 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{F.5}$$

The trajectories of the three moving surgical instruments are expressed as:

$$\boldsymbol{P}_1 = \begin{bmatrix} 258 + 5\sin(\pi t) \\ -239 + 5\cos(1.5\pi t) \\ -216 + 0.5\sin(5\pi t) \end{bmatrix}, \tag{F.6}$$

$$\boldsymbol{P}_2 = \begin{bmatrix} 259 + 3\sin(1.5\pi t) \\ -248 + 3\sin(\pi t) \\ -221 + 0.5\sin(5\pi t) \end{bmatrix}, \tag{F.7}$$

and

$$\boldsymbol{P}_3 = \begin{bmatrix} 269 + 3\cos(1.5\pi t) \\ -244 + 2\cos(\pi t) \\ -226 + 0.5\sin(5\pi t) \end{bmatrix}. \tag{F.8}$$

In Section IV.C, the two moving surgical instruments are denoted as $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, while in Section IV.E, the two moving surgical instruments are denoted as $\boldsymbol{P}_1$ and $\boldsymbol{P}_3$.

APPENDIX G
RECOMMENDED RANGE OF PARAMETERS

When setting the parameters of the proposed visual tracking space vector method, in order to ensure the stability of the laparoscopic image, R and r are recommended to set according to the image resolution. A laparoscopic image with resolution of $2u_0 \times 2v_0$, reasonable R and r need to satisfy:

$$ratio_{0.02} \times \sqrt{u_0^2 + v_0^2} \leq r \leq ratio_{0.5} \times \sqrt{u_0^2 + v_0^2}, \quad \text{(G.1)}$$

and

$$r < R \leq ratio_{0.98} \times \sqrt{u_0^2 + v_0^2}. \quad \text{(G.2)}$$

where $ratio_{0.02}$ is the radio of distance at the 2nd quantile to the maximum distance that the instrument in the image can achieve, $ratio_{0.5}$ is the radio of distance at the 50th quantile or mean to the maximum distance that the instrument in the image can achieve, $ratio_{0.98}$ is the radio of distance at the 98th quantile to the maximum distance that the instrument in the image can achieve. The simulation and experimental results have shown that the proposed algorithm has good robustness. The setting of K has little effect on the experimental results. Thus, it can be set according to personal preference. A reasonable range is greater than zero and less than 10. The speed of a surgical instrument is related to the surgery type and surgeon, and limited by the narrow operating space, it does not exceed 300 mm/s. Due to the design principle of the proposed algorithm, only when the value of V_{max} is greater than the speed of the surgical instrument, the surgical instrument can be tracked. At the same time, a larger value of V_{max} should be set in the range of the maximum speed of the surgical instruments' tip of 300 mm/s.

The parameter settings can be different and depend on a surgeon, surgery, and surgical scene, etc. Users can use this method to adjust the parameters' values. In addition, the proposed method can achieve autonomous FOV adjustment without special settings of parameters. In other words, as long as these parameters are within reasonable ranges, the performance of the proposed method can be guaranteed.