PROOF OF THEOREM

On the one hand, the proposed method decouples and controls the position and attitude of the endoscopic camera at the same time. On the other hand, it adjusts the endoscopic depth to optimize the image definition under the premise of meeting the position-level safety constraint.

Theorem 1:

According to the principle of pinhole imaging, it can be obtained:

$$\begin{bmatrix} u^i_{\text{tips}} \\ v^i_{\text{tips}} \\ 1 \end{bmatrix} = \frac{1}{{}^{\text{lap}}z_i} \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} {}^{\text{lap}}x_i \\ {}^{\text{lap}}y_i \\ {}^{\text{lap}}z_i \end{bmatrix} \tag{1}$$

As shown in Fig.5, the direction vector of the endoscopic camera velocity is determined by:

$$\boldsymbol{v}^i_r = \frac{\boldsymbol{z}_{\text{lap}} \times \hat{\boldsymbol{p}}^i_{\text{tips}} \times \hat{\boldsymbol{p}}^i_{\text{tips}}}{\left\| \boldsymbol{z}_{\text{lap}} \times \hat{\boldsymbol{p}}^i_{\text{tips},i} \times \hat{\boldsymbol{p}}^i_{\text{tips}} \right\|} \tag{2}$$

where, $\boldsymbol{z}_{\text{lap}} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\text{T}}$, $\hat{\boldsymbol{p}}^i_{\text{tips}} = [\, x^i_{\text{tips}} / z^i_{\text{tips}},\ y^i_{\text{tips},i} / z^i_{\text{tips}}, 1\,]^{\text{T}}$, and $\boldsymbol{p}^i_{\text{tips}} = [\, x^i_{\text{tips}}, y^i_{\text{tips}}, z^i_{\text{tips}}\,]^{\text{T}}$ is the 3D coordinate of the ith device tracking point.

Therefore, the linear velocity of the endoscopic camera can be expressed as:

$$\boldsymbol{v}^i_c = \begin{bmatrix} v^i_{cx} \\ v^i_{cy} \\ v^i_{cz} \end{bmatrix} = \upsilon^i_s \boldsymbol{v}^i_r = \upsilon^i_s \begin{bmatrix} x^i_{\text{tips}} z^i_{\text{tips}} / l \\ y^i_{\text{tips},i} z^i_{\text{tips}} / l \\ \left(x^{i,2}_{\text{tips}} + y^{i,2}_{\text{tips}} \right) / l \end{bmatrix} \tag{3}$$

where, $l = \sqrt{\left(x^{i\,2}_{\text{tips}} + y^{i\,2}_{\text{tips}} \right)} \left\| \boldsymbol{p}^i_{\text{tips}} \right\|$, $\upsilon^i \geq 0$ is the magnitude of the camera velocity, $\boldsymbol{v}^i_r = \left[v^i_{rx}, v^i_{ry}, v^i_{rz} \right]^{\text{T}}$.

When the endoscopic camera moves and the surgical instrument tip is stationary, the velocity of the surgical instrument tip relative to the endoscopic camera is:

$$\boldsymbol{v}^i_f = \begin{bmatrix} -v^i_{rx} - z^i_{\text{tips}} \omega^i_{ry} \\ -v^i_{ry} + z^i_{\text{tips}} \omega^i_{rx} \\ -v^i_{rz} - y^i_{\text{tips}} \omega^i_{rx} + x^i_{\text{tips}} \omega^i_{ry} \end{bmatrix} \tag{4}$$

where, $\omega^i_{rx} = -v^i_{ry} / d_{in}$, $\omega^i_{ry} = v^i_{rx} / d_{in}$.

When both the endoscopic camera and the surgical instrument tips are moving, the linear velocity of the surgical instrument tips relative to the endoscopic camera is:

$$\hat{\boldsymbol{v}}^i_f = \begin{bmatrix} \dot{x}^i_{\text{tips}} - v^i_{rx} - \dfrac{v^i_{rx} z^i_{\text{tips}}}{d_{in}} \\ \dot{y}^i_{\text{tips}} - v^i_{ry} - \dfrac{v^i_{ry} z^i_{\text{tips}}}{d_{in}} \\ \dot{z}^i_{\text{tips}} - v^i_{rz} + \dfrac{v^i_{ry} y^i_{\text{tips}}}{d_{in}} + \dfrac{v^i_{rx} x^i_{\text{tips}}}{d_{in}} \end{bmatrix}$$
$$= \begin{bmatrix} \dot{x}^i_{\text{tips}} - \dfrac{\upsilon^i}{l} x^i_{\text{tips},i} z^i_{\text{tips},i} - \dfrac{\upsilon^i}{d_{in} l} x^i_{\text{tips},i} z^{i2}_{\text{tips}} \\ \dot{y}^i_{\text{tips}} - \dfrac{\upsilon^i}{l} y^i_{\text{tips}} z^i_{\text{tips}} - \dfrac{\upsilon^i}{d_{in} l} y^i_{\text{tips}} z^{i,2}_{\text{tips}} \\ \dot{z}^i_{\text{tips}} + \dfrac{\upsilon^i}{l} \left(x^{i,2}_{\text{tips}} + y^{i,2}_{\text{tips}} \right) + \dfrac{\upsilon^i}{d_{in} l} \left(x^{i,2}_{\text{tips}} + y^{i,2}_{\text{tips}} \right) z^i_{\text{tips}} \end{bmatrix} \tag{5}$$

Differentiate (1) and substitute (5) into the result of the derivation to get:

$$\begin{aligned} \dot{u}^i_{\text{tips}} &= \frac{f_u}{z^i_{\text{tips}}} v^i_{fx} - \frac{f_u\, x^i_{\text{tips}}}{z^{i2}_{\text{tips}}} v^i_{fz} \\ &= \frac{f_u}{z^{i2}_i} \Big[\dot{x}^i_{\text{tips}} z^i_{\text{tips}} - x^i_{\text{tips}} \dot{z}^i_{\text{tips}} \\ &\quad - \frac{\upsilon^i}{l} x^i_{\text{tips}} \left(x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}} + z^{i2}_{\text{tips}} \right) \\ &\quad - \frac{\upsilon^i}{d_{in} l} x^i_{\text{tips}} z^i_{\text{tips}} \left(x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}} + z^{i2}_{\text{tips}} \right) \Big] \end{aligned} \tag{6}$$

$$\begin{aligned} \dot{v}^i_{\text{tips}} &= \frac{f_v}{z^i_{\text{tips}}} v^i_{fy} - \frac{f_v\, y^i_{\text{tips}}}{z^2_i} v^i_{fz} \\ &= \frac{f_v}{z^{i2}_{\text{tips}}} \Big[\dot{y}^i_{\text{tips}} z^i_{\text{tips}} - y^i_{\text{tips}} \dot{z}^i_{\text{tips}} \\ &\quad - \frac{\upsilon^i}{l} y^i_{\text{tips}} \left(x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}} + z^{i2}_{\text{tips}} \right) \\ &\quad - \frac{\upsilon^i}{d_{in} l} y^i_{\text{tips}} z^i_{\text{tips}} \left(x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}} + z^{i2}_{\text{tips}} \right) \Big] \end{aligned} \tag{7}$$

The established Lyapunov function is the distance between the pixel coordinates of the actual instrument tips and the center of the camera imaging image, that is,

$$L = \left(u^i_{\text{tips}} - u_0 \right)^2 + \left(v^i_{\text{tips}} - v_0 \right)^2 \tag{8}$$

Taking the derivative of (8), we can get:

$$\dot{L} = 2\left(u^i_{\text{tips}} - u_0 \right) \dot{u}^i_{\text{tips}} + 2\left(u^i_{\text{tips}} - v_0 \right) \dot{v}^i_{\text{tips}} \tag{9}$$

Substituting (1), (6), and (7) into (9), yield:

$$\begin{aligned} \dot{L} = &\frac{f^2_u x^{i2}_{\text{tips}}}{z^{i3}_{\text{tips}}} \left[\frac{z^i_{\text{tips}}}{x^i_{\text{tips}}} \dot{x}^i_{\text{tips}} - \dot{z}^i_{\text{tips}} - \left(1 + \frac{z^i_{\text{tips}}}{d_{in}} \right) \upsilon^i \frac{\left\| \boldsymbol{p}^i_{\text{tips}} \right\|}{\sqrt{x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}}}} \right] \\ + &\frac{f^2_v y^{i2}_{\text{tips}}}{z^{i3}_{\text{tips}}} \left[\frac{z^i_{\text{tips}}}{y^i_{\text{tips}}} \dot{y}^i_{\text{tips}} - \dot{z}^i_{\text{tips}} - \left(1 + \frac{z^i_{\text{tips}}}{d_{in}} \right) \upsilon^i \frac{\left\| \boldsymbol{p}^i_{\text{tips}} \right\|}{\sqrt{x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}}}} \right] \end{aligned} \tag{10}$$

When $d_{in} \to +\infty$, (10) can be further simplified as:

$$\begin{aligned} \dot{L} = &\frac{f^2_u x^{i2}_{\text{tips}}}{z^{i3}_{\text{tips}}} \left[\frac{z^i_{\text{tips}}}{x^i_{\text{tips}}} \dot{x}^i_{\text{tips}} - \dot{z}^i_{\text{tips}} - \frac{\left\| \boldsymbol{p}^i_{\text{tips}} \right\|}{\sqrt{x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}}}} \upsilon^i \right] + \\ &\frac{f^2_v y^{i2}_{\text{tips}}}{z^{i3}_{\text{tips}}} \left[\frac{z^i_{\text{tips}}}{y^i_{\text{tips}}} \dot{y}^i_{\text{tips}} - \dot{z}^i_{\text{tips}} - \frac{\left\| \boldsymbol{p}^i_{\text{tips}} \right\|}{\sqrt{x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}}}} \upsilon^i \right] \\ \leq &\frac{f^2_u x^{i2}_{\text{tips}}}{z^{i3}_{\text{tips}}} A + \frac{f^2_v y^{i2}_{\text{tips}}}{z^{i3}_{\text{tips}}} B \end{aligned} \tag{11}$$

with

$$\begin{aligned} A &= \frac{z^i_{\text{tips}}}{x^i_{\text{tips}}} \dot{x}^i_{\text{tips}} - \dot{z}^i_{\text{tips}} - \frac{\left\| \boldsymbol{p}^i_{\text{tips}} \right\|}{\sqrt{x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}}}} \upsilon^i \\ B &= \frac{z^i_{\text{tips}}}{y^i_{\text{tips}}} \dot{y}^i_{\text{tips}} - \dot{z}^i_{\text{tips}} - \frac{\left\| \boldsymbol{p}^i_{\text{tips}} \right\|}{\sqrt{x^{i2}_{\text{tips}} + y^{i2}_{\text{tips}}}} \upsilon^i \end{aligned} \tag{12}$$

Since $z^i_{\text{tips}} > 0$, (12) has the following characteristics:

$$A \le \max\left\{ \frac{z_{\text{tips}}^i}{\left\| x_{\text{tips}}^i \right\|} \upsilon_s^i, \upsilon^i \right\} - \frac{\left\| \boldsymbol{p}_{\text{tips}}^i \right\|}{\sqrt{x_{\text{tips}}^{i2} + y_{\text{tips}}^{i2}}} \upsilon^i$$
$$B \le \max\left\{ \frac{z_{\text{tips}}^i}{\left\| y_{\text{tips}}^i \right\|} \upsilon^i, \upsilon^i \right\} - \frac{\left\| \boldsymbol{p}_{\text{tips}}^i \right\|}{\sqrt{x_{\text{tips}}^{i2} + y_{\text{tips}}^{i2}}} \upsilon^i \tag{13}$$

Further,

$$A: \begin{cases} A \le \upsilon^i - \dfrac{\left\| \boldsymbol{p}_{\text{tips}}^i \right\|}{\sqrt{x_{\text{tips}}^{i2} + y_{\text{tips}}^{i2}}} \upsilon^i \le 0, \text{ if } \dfrac{z_{\text{tips}}^i}{\left\| x_{\text{tips}}^i \right\|} \le 1 \\ A \le \dfrac{z_{\text{tips}}^i}{\left\| x_{\text{tips}}^i \right\|} \upsilon^i - \sqrt{\dfrac{x_{\text{tips}}^{i2} + z_{\text{tips}}^{i2}}{x_{\text{tips}}^{i2}}} \upsilon^i \le 0, \text{ else} \end{cases}$$
$$B: \begin{cases} B \le \upsilon^i - \dfrac{\left\| \boldsymbol{p}_{\text{tips}}^i \right\|}{\sqrt{x_{\text{tips}}^{i2} + y_{\text{tips}}^{i2}}} \upsilon^i \le 0, \text{ if } \dfrac{z_{\text{tips}}^i}{\left\| y_{\text{tips}}^i \right\|} \le 1 \\ B \le \dfrac{z_{\text{tips}}^i}{\left\| y_{\text{tips}}^i \right\|} \upsilon^i - \sqrt{\dfrac{y_{\text{tips}}^{i2} + z_{\text{tips}}^{i2}}{y_{\text{tips}}^{i2}}} \upsilon^i \le 0, \text{ else} \end{cases} \tag{14}$$

Based on (14) and (11), $\dot{L} \le 0$ can be obtained, so the distance error between the pixel coordinates of the actual instrument end and the center of the camera image is always convergent.

Theorem 2:

According to the hand-eye coordination model, $\exists \beta_1 \ge 0$, when $\forall \beta \ge \beta_1$, $0 \le {}^{\text{con}}\omega_{\text{coor}} \le \omega_{\max}$.

$$\dot{\beta} = -\eta_{\text{coor}} {}^{\text{con}}\omega_{\text{coor}} \le 0 \tag{15}$$

Empathy, $\exists \beta_2 \le 0$, when $\forall \beta \le \beta_2$, $-\omega_{\max} \le {}^{\text{con}}\omega_{\text{coor}}^t \le 0$.

$$\dot{\beta} = -\eta_{\text{coor}} {}^{\text{con}}\omega_{\text{coor}} \ge 0 \tag{16}$$

Theorem 3:

Let $\eta_{\text{def}} \ge \eta_{\text{track}}$. According to the optimization model considering image definition, $\exists \left| \eta_q \right| \ge \eta_0$, when $\forall d_{obj} < d_{foc}$, $\lambda_{\text{def}} = v_{\max}$.

$$\begin{aligned} \dot{d}_{\text{obj}} &= \eta_{\text{track}} {}^{lap}\boldsymbol{v}_{\text{track}} \cdot \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\text{T}} + \eta_{\text{def}} \lambda_{\text{def}} \\ &\ge \left(-\eta_{\text{track}} + \eta_{\text{def}} \right) v_{\max} \\ &\ge 0 \end{aligned} \tag{17}$$

Empathy, $\exists \left| \eta_q \right| \ge \eta_0$, when $\forall d_{\text{obj}} > d_{\text{foc}}$, $\lambda_{\text{def}} = -v_{\max}$.

$$\begin{aligned} \dot{d}_{\text{obj}} &= \eta_{\text{track}} {}^{\text{lap}}\boldsymbol{v}_{\text{track}} \cdot \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\text{T}} + \eta_{\text{def}} \lambda_{\text{def}} \\ &\le \left(\eta_{\text{track}} - \eta_{\text{def}} \right) v_{\max} \\ &\le 0 \end{aligned} \tag{18}$$

Theorem 4:

Let $\eta_{\text{safe}} \ge \eta_{\text{track}}$. According to the position-level safety constraint, $\exists d_{in1} \in \left[d_{\text{alart,min}}, d_{\text{safe,min}} \right]$, when $\forall d_{in} \in \left[d_{\text{alart,min}}, d_{in1} \right]$, $\lambda_{safe} = v_{\max}$.

$$\begin{aligned} \dot{d}_{in} &= \eta_{\text{track}} {}^{\text{con}}\boldsymbol{v}_{\text{track}} \cdot \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\text{T}} + \eta_{\text{safe}} \lambda_{\text{safe}} \\ &\ge \left(-\eta_{\text{track}} + \eta_{\text{safe}} \right) v_{\max} \\ &\ge 0 \end{aligned} \tag{19}$$

Empathy, $\exists d_{in2} \in \left[d_{\text{safe,max}}, d_{\text{alart,max}} \right]$, when $\forall d_{in} \in \left[d_{in2}, d_{\text{alart,max}} \right]$, $\lambda_{safe} = -v_{\max}$.

$$\begin{aligned} \dot{d}_{in} &= \eta_{\text{track}} {}^{\text{con}}\boldsymbol{v}_{\text{track}} \cdot \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\text{T}} + \eta_{\text{safe}} \lambda_{\text{safe}} \\ &\le \left(\eta_{\text{track}} - \eta_{\text{safe}} \right) v_{\max} \le 0 \end{aligned} \tag{20}$$