Supplementary Material

Enhanced Scale-aware Depth Estimation for Monocular Endoscopic Scenes with Geometric Modeling

Paper ID: 1856

July 9, 2024

A Observation of Relative Depth Estimation Network

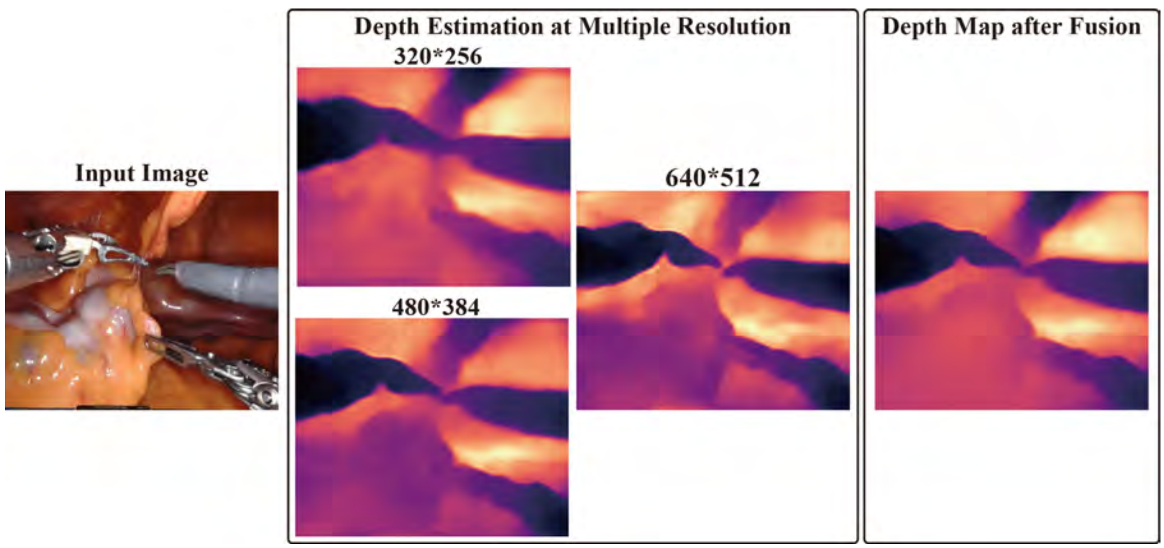

Figure 1: Observations on the relative depth estimation network behavior.

B Derivation of Eq. 3

Recall that:

$$ {}^{e}\widetilde{\mathbf{c}}_j^{\mathsf{T}} \cdot \begin{bmatrix} [{}^{e}\mathbf{s}_i]_\times [{}^{e}\mathbf{s}_i]_\times^{\mathsf{T}} & [{}^{e}\mathbf{s}_i]_\times {}^{e}\mathbf{m}_i \\ {}^{e}\mathbf{m}_i^{\mathsf{T}} [{}^{e}\mathbf{s}_i]_\times^{\mathsf{T}} & \|{}^{e}\mathbf{m}_i\|^2 - r_s^2 \end{bmatrix} \cdot {}^{e}\widetilde{\mathbf{c}}_j = 0. \tag{1} $$

Following the perspective projection theory:

$$ z({}^{e}\widetilde{\mathbf{c}}_j) \times [u_j \; v_j \; 1] = [\mathbf{K}|\mathbf{0}_{3\times1}] \cdot {}^{e}\widetilde{\mathbf{c}}_j. \tag{2} $$

where $z({}^{e}\widetilde{\mathbf{c}}_j)$ is the depth value of the 3D point, and ${}^{I}\widetilde{\mathbf{p}}_j = [u_j \; v_j \; 1]^{\mathsf{T}}$ represents the pixel coordinate in the image plane, so the point ${}^{e}\widetilde{\mathbf{c}}_j$ can be further converted to:

$$ {}^{e}\widetilde{\mathbf{c}}_j = z({}^{e}\widetilde{\mathbf{c}}_j) \cdot \begin{bmatrix} \mathbf{K}^{-1} \\ \mathbf{0}_{1\times3} \end{bmatrix} \cdot {}^{I}\widetilde{\mathbf{p}}_j + [0 \; 0 \; 0 \; 1]^{\mathsf{T}}. \tag{3} $$

Substituting (3) into (1) leads to a quadratic equation of the form $A \cdot z({}^{e}\widetilde{\mathbf{c}}_j) + 2B \cdot z({}^{e}\widetilde{\mathbf{c}}_j) + C = 0$, where $A = {}^{I}\widetilde{\mathbf{p}}_j^{\mathsf{T}}(\mathbf{K})^{-\mathsf{T}}[{}^{e}\mathbf{s}_i]_\times [{}^{e}\mathbf{s}_i]_\times^{\mathsf{T}} {}^{I}\widetilde{\mathbf{p}}_j$, $B = {}^{I}\widetilde{\mathbf{p}}_j^{\mathsf{T}}(\mathbf{K})^{-\mathsf{T}}[{}^{e}\mathbf{s}_i]_\times {}^{e}\mathbf{m}_i$, and $C = \|{}^{e}\mathbf{m}_i\|^2 - r_s^2$.

To solve the above equation for a unique solution, we have a constraint described as $B^2 - A \times C = 0$ which then leads to the Eq. 2 that includes the edge lines $\left(\mathbf{l}_i^-, \mathbf{l}_i^+\right)$ of the shaft.

C Calculation of Depth Value $z(\mathbf{c}_{f0})$

Recall that $\mathbf{c}_{s0}$ is the 3D point along the axis of the tool and ${}^e\mathbf{s}_i \in \mathbb{R}^{3\times1}$ is the direction of the tool's shaft. Therefore, we can obtain the depth of the point $\mathbf{c}_{f0}$ on the surface of the shaft as follows:

$$\begin{aligned} &\theta = \arccos \frac{{}^e\mathbf{s}_i^\mathsf{T} \cdot \mathbf{c}_{s0}}{\|{}^e\mathbf{s}_i\| \|\mathbf{c}_{s0}\|} \\ &\rightarrow \ z\left(\mathbf{c}_{f0}\right) = z\left(\mathbf{c}_{s0}\right) - r_s / \sin\theta \end{aligned} \tag{4}$$