Noname manuscript No.
(will be inserted by the editor)

3D reconstruction of colon structures and textures from stereo colonoscopic images

Supplementary material for RAL with IROS option submission

Shuai Zhang · Liang Zhao · Shoudong Huang · Qi Hao · Menglong Ye

the date of receipt and acceptance should be inserted later

1 Abstract

This supplementary material gives details of the proposed optimization algorithm for the geometric and photometric scan to colon model registration and barycentric based texture rendering algorithm for texturing and shading of the reconstructed colonic surface in the paper.

2 Details of the supplementary material

2.1 Optimization

In Section 3.3 of the paper, we formulate a joint optimization objective by combining the geometric constraint and photometric constraint together:

$$E(T) = (1 - \sigma)E_G(T) + \sigma E_F(T), \tag{1}$$

where $E_G(T)$ is the geometric term, and the $E_F(T)$ is the photometric feature term provided by the pair-wise 3D sparse anchor points generated from 2D SIFT features described in Section 3.2 of the paper, $\sigma \in [0, 1]$ is a weight that balances the two terms. Our goal is to find the optimal transformation T that best aligns the reconstructed scan to the colon model.

The geometric term $E_G(T)$ sums all the squared distances between each source point $\mathbf{s}_i = [s_{ix}, s_{iy}, s_{iz}, 1]^T$ in a scan and the tangent plane at its closet point $\mathbf{d}_i = [d_{ix}, d_{iy}, d_{iz}, 1]^T$ in the colon model, and $\mathbf{n}_i = [n_{ix}, n_{iy}, n_{iz}, 0]^T$ is the unit normal vector at $\mathbf{d}_i$.

Similarly, the photometric term $E_F(T)$ sums all the point-to-point distances between the 3D anchor point $\mathbf{s}_i^f = [s_{ix}^f, s_{iy}^f, s_{iz}^f, 1]^T$ in a scan and its corresponding 3D anchor point $\mathbf{d}_j^f = [d_{jx}^f, d_{jy}^f, d_{jz}^f, 1]^T$ in the colon mesh, described in Section 3.2 of the paper:

Address(es) of author(s) should be given

$$E_G(T)=\sum^{i}((T\cdot\mathbf{s}_i-\mathbf{d}_i)\bullet\mathbf{n}_i)^2,\ E_F(T)=\sum^{j}(T\cdot\mathbf{s}_j^f-\mathbf{d}_j^f)\bullet(T\cdot\mathbf{s}_j^f-\mathbf{d}_j^f)\quad(2)$$

We minimize the objective function $E(T)$ of the non-linear least-squares problem by linear approximation to the rotation matrix [1]. At the k^{th} iteration, T can be expressed as following:

$$T=\Delta T\cdot T^k\tag{3}$$

where T^k is the global transformation estimated in the last iteration and ΔT is the incremental 3D rigid-body transformation which is composed of a rotation matrix $R(\alpha,\beta,\gamma)$ and a translation matrix $\mathbf{t}(t_x,t_y,t_z)$:

$$\Delta T=\mathbf{t}(t_x,t_y,t_z)\cdot R(\alpha,\beta,\gamma)\tag{4}$$

where

$$\mathbf{t}(t_x,t_y,t_z)=\begin{bmatrix}1&0&0&t_x\\0&1&0&t_y\\0&0&1&t_z\\0&0&0&1\end{bmatrix}\tag{5}$$

and

$$\begin{aligned}&R(\alpha,\beta,\gamma)=R_z(\gamma)\cdot R_y(\beta)\cdot R_x(\alpha)\\&=\begin{bmatrix}\cos\gamma\cos\beta & -\sin\gamma\cos\alpha+\cos\gamma\sin\beta\sin\alpha & \sin\gamma\sin\alpha+\cos\gamma\sin\beta\cos\alpha & 0\\ \sin\gamma\cos\beta & \cos\gamma\cos\alpha+\sin\gamma\sin\beta\sin\alpha & -\cos\gamma\sin\alpha+\sin\gamma\sin\beta\cos\alpha & 0\\ -\sin\beta & \cos\beta\sin\alpha & \cos\beta\cos\alpha & 0\\ 0&0&0&1\end{bmatrix}\end{aligned}\tag{6}$$

$R_x(\alpha)$, $R_y(\beta)$, $R_z(\gamma)$ are rotations of the angles α, β and γ around the x-axis, y-axis and z-axis, respectively. When the incremental rotations of each iteration are small, it can be approximated as following:

$$\boldsymbol{R}(\alpha,\beta,\gamma)\approx\begin{bmatrix}1&\alpha\beta-\gamma&\alpha\gamma+\beta&0\\ \gamma&\alpha\beta\gamma+1&\beta\gamma-\alpha&0\\ -\beta&\alpha&1&0\\0&0&0&1\end{bmatrix}\approx\begin{bmatrix}1&-\gamma&\beta&0\\ \gamma&1&-\alpha&0\\ -\beta&\alpha&1&0\\0&0&0&1\end{bmatrix}\tag{7}$$

Then, T is approximated by:

$$T\approx\begin{bmatrix}1&-\gamma&\beta&t_x\\ \gamma&1&-\alpha&t_y\\ -\beta&\alpha&1&t_z\\0&0&0&1\end{bmatrix}\cdot T^k\tag{8}$$

Each $(T\cdot\mathbf{s}_i-\mathbf{d}_i)\bullet\mathbf{n}_i$ in (2) can be written as a linear expression of the six parameters α, β, γ, t_x, t_y and t_z:

$$(T\cdot\mathbf{s}_i-\mathbf{d}_i)\bullet\mathbf{n}_i=[\bar{\mathbf{s}}_i\times\mathbf{n}_i,\mathbf{n}_i^T]\cdot[\alpha,\beta,\gamma,t_x,t_y,t_z]^T-[\mathbf{d}_i-\bar{\mathbf{s}}_i]\bullet\mathbf{n}_i\tag{9}$$

where $\bar{\mathbf{s}}_i=T^k\cdot\mathbf{s}_i$. Given N_1 pairs of point correspondences in term $E_G(T)$, we can arrange all $(T\cdot\mathbf{s}_i-\mathbf{d}_i)\bullet\mathbf{n}_i$, $1\le i\le N_1$, into a matrix expression:

$$A_1\mathbf{x}-\mathbf{b}_1\tag{10}$$

where A_1 is a N_1 by 6 matrix, $\mathbf{b}_1$ is a N_1 by 1 vector and $\mathbf{x} = [\alpha, \beta, \gamma, t_x, t_y, t_z]^T$ is a 6 by 1 vector:

$$A_1 = \begin{bmatrix} \overline{\mathbf{s}}_1 \times \mathbf{n}_1,\, \mathbf{n}_1^T \\ ... \\ \overline{\mathbf{s}}_i \times \mathbf{n}_i,\, \mathbf{n}_i^T \\ ... \\ \overline{\mathbf{s}}_{N_1} \times \mathbf{n}_{N_1},\, \mathbf{n}_{N_1}^T \end{bmatrix},\ \mathbf{b}_1 = \begin{bmatrix} [\mathbf{d}_1 - \overline{\mathbf{s}}_1] \bullet \mathbf{n}_1 \\ ... \\ [\mathbf{d}_i - \overline{\mathbf{s}}_i] \bullet \mathbf{n}_i \\ ... \\ [\mathbf{d}_{N_1} - \overline{\mathbf{s}}_{N_1}] \bullet \mathbf{n}_{N_1} \end{bmatrix} \tag{11}$$

Similarly, each $(T \cdot \mathbf{s}_j^f - \mathbf{d}_j^f)$ in (2) can also be written as a linear expression group of $\mathbf{x}$:

$$T \cdot \mathbf{s}_j^f - \mathbf{d}_j^f = \begin{bmatrix} [\overline{s}_{jz}^f \cdot \beta - \overline{s}_{jy}^f \cdot \gamma + t_x] - [d_{jx}^f - \overline{s}_{jx}^f] \\ [\overline{s}_{jx}^f \cdot \gamma - \overline{s}_{jz}^f \cdot \alpha + t_y] - [d_{jy}^f - \overline{s}_{jy}^f] \\ [\overline{s}_{jy}^f \cdot \alpha - \overline{s}_{jx}^f \cdot \beta + t_z] - [d_{jz}^f - \overline{s}_{jz}^f] \end{bmatrix} \tag{12}$$

where $\overline{\mathbf{s}}_j^f = T^k \cdot \mathbf{s}_j^f$. Given N_2 pairs of anchor point correspondences in term $E_F(T)$, we can arrange all $T \cdot \mathbf{s}_j^f - \mathbf{d}_j^f$, $1 \le j \le N_2$, into a matrix expression:

$$A_2\mathbf{x} - \mathbf{b}_2 \tag{13}$$

where A_2 is a $N_2 \times 3$ by 6 matrix and $\mathbf{b}_2$ is $N_2 \times 3$ by 1 vector:

$$A_2 = \left[A_{21}^T\ ...\ A_{2j}^T\ ...\ A_{2N_2}^T\right]^T,\ \mathbf{b}_2 = \left[\mathbf{b}_{21}^T\ ...\ \mathbf{b}_{2j}^T\ ...\ \mathbf{b}_{2N_2}^T\right]^T \tag{14}$$

with $A_{2j} = \begin{bmatrix} 0 & \overline{s}_{jz}^f & -\overline{s}_{jy}^f & 1\ 0\ 0 \\ -\overline{s}_{jz}^f & 0 & \overline{s}_{jx}^f & 0\ 1\ 0 \\ \overline{s}_{jy}^f & -\overline{s}_{jx}^f & 0 & 0\ 0\ 1 \end{bmatrix}$ and $\mathbf{b}_{2j} = \begin{bmatrix} d_{jx}^f - \overline{s}_{jx}^f \\ d_{jy}^f - \overline{s}_{jy}^f \\ d_{jz}^f - \overline{s}_{jz}^f \end{bmatrix}$.

Therefore, we can obtain the optimal $\mathbf{x}$ by solving for:

$$\min_{\mathbf{x}} (1-\sigma)|A_1\mathbf{x} - \mathbf{b}_1|^2 + \sigma|A_2\mathbf{x} - \mathbf{b}_2|^2, \tag{15}$$

which is a linear least-squares problem, and can be solved by setting the derivative of the objective function with respect to the $\mathbf{x}$ to zero. Then, the solution is:

$$\mathbf{x}_{opt} = ((1-\sigma) \cdot A_1^T \cdot A_1 + \sigma \cdot A_2^T \cdot A_2)^{-1} \cdot ((1-\sigma) \cdot A_1^T \cdot \mathbf{b}_1 + \sigma \cdot A_2^T \cdot \mathbf{b}_2) \tag{16}$$

Since the obtained solution is an approximation, we will apply it to (4) to map the estimated transformation into $SE(3)$. In each iteration, we solve the linear system in (15), and update T by applying the incremental transformation ΔT to T^k using (8). In the next iteration, we rep-linearize T around T^{k+1} and repeat.

In Equation (16), We use the parameter σ to balance the geometric and the photometric term. If the value of σ is too large, the optimization objective will be more close to the point-to-point objective in the proposed registration algorithm and the optimal solution will mainly depend on the relatively small number of the 3D anchor correspondences provided by the 2D SIFT approach which represent texture features, but the optimal solution may not be reliable when some anchor correspondences are not correct or less accurate. However, if the value of σ is too small, the optimization objective will become close to the geometric term and the feature based regulation term becomes less effect on the optimal solution, and this causes inconsistency of texture matching in the overlapping region of two scans. In this paper, we set σ to 0.5.

2.2 Texture mapping using barycentric coordinates

In Section 3.5 of the paper, we use a barycentric based texture rendering technique to map textures from colonoscopic images to the reconstructed colonic surface. As we can see from Fig. 1, for three vertices **A**, **B**, **C** of one triangular $\triangle\mathbf{ABC}$ face in the colon mesh, we can extract their matched points in the 3D reconstructed scan by referring to the established point correspondences between the scan and the vertices of colon mesh. Furthermore, as each 3D point in a scan corresponds to a 2D pixel in a 2D image when reconstructs the scan, we can extract a triangular texture region $\triangle\mathbf{abc}$ (where **a**, **b**, and **c** are the 2D location of three vertices of the triangle) in 2D images corresponding to each triangle $\triangle\mathbf{ABC}$ face in the colon mesh.

After that, we use barycentric mapping technique [2] to map pixel color from the 2D texture region $\triangle\mathbf{abc}$ to the 3D triangle $\triangle\mathbf{ABC}$ face. For an arbitrary point $\mathbf{P}(x_\mathrm{P}, y_\mathrm{P}, z_\mathrm{P})$ inside the triangle $\triangle\mathbf{ABC}$, there is a unique sequence of three numbers, $\lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0$ to represent it:

$$\begin{cases} x_\mathrm{P} = \lambda_1 x_\mathrm{A} + \lambda_2 x_\mathrm{B} + \lambda_3 x_\mathrm{C} \\ y_\mathrm{P} = \lambda_1 y_\mathrm{A} + \lambda_2 y_\mathrm{B} + \lambda_3 y_\mathrm{C} \\ z_\mathrm{P} = \lambda_1 z_\mathrm{A} + \lambda_2 z_\mathrm{B} + \lambda_3 z_\mathrm{C} \\ 1 = \lambda_1 + \lambda_2 + \lambda_3 \end{cases} \tag{17}$$

where λ_1, λ_2, λ_3 indicate the barycentric coordinates of the point **p** with respect to the triangle. Once we have the barycentric coordinates, the texture coordinates of **P** can be determined by interpolating the texture values at the vertices using the barycentric coordinates as weights:

$$\begin{cases} u_\mathrm{p} = \lambda_1 u_\mathrm{a} + \lambda_2 u_\mathrm{b} + \lambda_3 u_\mathrm{c} \\ v_\mathrm{p} = \lambda_1 v_\mathrm{a} + \lambda_2 v_\mathrm{b} + \lambda_3 v_\mathrm{c} \end{cases} \tag{18}$$

Overall, it takes the following steps to texturize the reconstructed colonic surface from multiple colonoscopic 2D images:

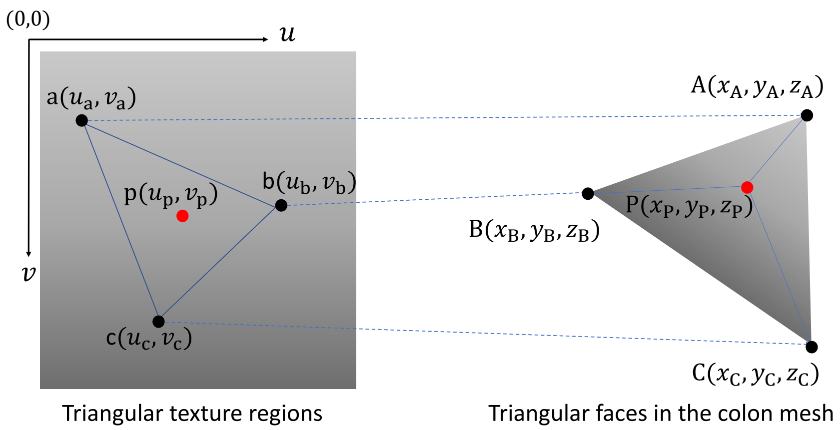
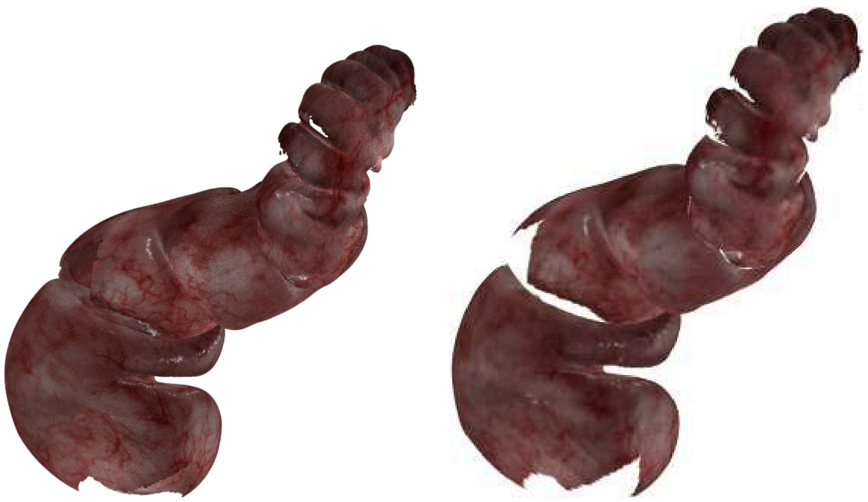


Fig. 1 Barycentric coordinates based texture mapping

1. Establishing triangular texture region in 2D texture images for each triangle face in the colon model;
2. Using a set of barycentric coordinates to interpolate arbitrary points inside each triangle face in the colon model;
3. Calculating each interpolated point's texture coordinates in its corresponding triangular texture region based on its barycentric weights;
4. Mapping textures from the triangular texture region to the triangle in the colon model.

Fig. 2 shows the texture quality comparison between the proposed approach and patch coloring approach. We can find that the texture quality from the proposed texture rendering method is more clear and accurate.



(a) (b)

Fig. 2 Texturized rectum colon using two different texture rendering approaches: (a) Barycentric coordinates based texture rendering; (b) Patch rendering.

References

1. Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *3dim*, volume 1, pages 145–152, 2001.
2. Ofir Weber, Mirela Ben-Chen, Craig Gotsman, and Kai Hormann. A complex view of barycentric mappings. In *Computer Graphics Forum*, volume 30, pages 1533–1542. Wiley Online Library, 2011.