第 14 卷 第 1 期
2025 年 1 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 14 No. 1
Jan. 2025

引文格式：
许涛, 王顺成, 钟建文, 等. 基于内窥镜图像的儿童腺样体场景三维重建 [J]. 集成技术, 2025, 14(1): 65-77.
Xu T, Wang SC, Zhong JW, et al. Three dimensional reconstruction for pediatric adenoid based on endoscopic images [J]. Journal of Integration Technology, 2025, 14(1): 65-77.

基于内窥镜图像的儿童腺样体场景三维重建

许 涛[1#] 王顺成[2#] 钟建文[2] 刘大波[2] 周宜龙[2*] 刘 畅[3*]

[1]（南方科技大学 深圳 518055）
[2]（南方医科大学深圳医院 深圳 518000）
[3]（中国科学院深圳先进技术研究院 深圳 518055）

摘 要 腺样体肥大是导致儿童阻塞性睡眠呼吸暂停综合征的主要因素。医生通过内窥镜影像评估腺样体肥大程度及其对气道的阻塞程度。然而，内窥镜影像为二维图像，医生只能主观推测患者腺样体区域的三维结构。这种方法严重依赖医生的相关经验和对腺样体的观察角度。腺样体表面为人体黏膜组织，覆盖鼻腔分泌物，其内窥镜影像具有强反光、特征稀少、场景光滑、模糊等特点。根据腺样体表面特点，该文提出一种基于腺样体鼻咽腔内镜图像序列的多视图三维重建算法。算法首先采用多视图立体匹配技术获取图像对应深度图的粗糙估计；其次，使用网格曲面在深度空间中对粗糙的深度信息进行拟合，从而得到平滑、精细的深度图；最后，通过点云融合算法获得腺样体区域稠密、精确的三维重建。仿真与真实实验表明，该文算法基于腺样体内窥镜图像序列，可实现精确、稠密、平滑的腺样体区域三维重建，且重建结果显著优于现有三维重建算法。

关键词 腺样体；三维重建；曲面拟合；多视图立体重建
中图分类号 R765.04+1；TP391.4 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20240307001

Three Dimensional Reconstruction for Pediatric Adenoid Based on Endoscopic Images

XU Tao[1#] WANG Shuncheng[2#] ZHONG Jianwen[2] LIU Dabo[2] ZHOU Yilong[2*] LIU Chang[3*]

[1](*Southern University of Science and Technology*, *Shenzhen* 518055, *China*)
[2](*Shenzhen Hospital of Southern Medical University*, *Shenzhen* 518000, *China*)
[3](*Shenzhen Institute of Advanced Technology*, *Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[*]*Corresponding Authors*: *yilongzhou2992@126.com*; *chang.liu2@siat.ac.cn*
[#]*Equal Contribution*

Abstract Adenoid hypertrophy (AH) is a key contributor to pediatric obstructive sleep apnea syndrome

收稿日期：2024-03-07 **修回日期：**2024-04-23
基金项目：国家重点研发计划项目（2020YFC2004400）；广东省自然科学基金项目（2021A1515011678）；国家自然科学基金项目（62273324）；南方医科大学深圳医院重点学科建设基金科研提升计划项目（ZDXKKYTS0021）；深圳市卫生健康菁英人才培养计划项目（2021）
作者简介：许涛（共同第一作者），硕士研究生，研究方向为三维重建；王顺成（共同第一作者），主治医师，研究方向为儿童耳鼻喉疾病；钟建文，主任医师，研究方向为儿童睡眠呼吸障碍疾病；刘大波，主任医师，研究方向为儿童耳鼻喉疾病；周宜龙（通讯作者），副主任医师，研究方向为儿童睡眠呼吸障碍疾病，E-mail：yilongzhou2992@126.com；刘畅（通讯作者），副研究员，研究方向为模式识别与智能系统，E-mail：chang.liu2@siat.ac.cn。

(OSAS). Physicians rely on nasopharyngeal endoscopy to identify AH and the obstruction of adenoid to the airway. However, due to the limitations of 2D endoscope images, physicians have to subjectively infer the 3D structure of the adenoid region, which heavily relies on their expertise and the angle at which the adenoids are observed. The adenoid surface is composed of mucosal tissue covered by nasal secretions, and thus strongly reflective, smooth, and lack features. Furthermore, the endoscope image of adenoid is relatively blurred. Based on these unique characteristics of the adenoids, this paper introduces a multi-view stereo algorithm based on endoscopic image sequences of the adenoid nasopharyngeal cavity. The algorithm employs multi-view stereo matching to first estimate the depth maps corresponding to the images. Subsequently, it utilizes mesh surfaces to fit the rough depth information in the depth space, and thereby generates the smooth and refined depth maps. Eventually, fusing the obtained depth maps leads to a dense and precise reconstruction of the adenoid region. Both synthetic and real experimental results demonstrate that the algorithm can achieve accurate, dense, and smooth reconstruction of the adenoid area, surpassing the existing reconstruction algorithms significantly.

Keywords adenoid; 3D reconstruction; surface fitting; multi-view stereo

Funding This work is supported by National Key Research and Development Program of China (2020YFC2004400), Natural Science Foundation of Guangdong Province (2021A1515011678), National Natural Science Foundation of China (62273324), Scientific Research Promotion Project of Key Discipline Construction Fund of Shenzhen Hospital of Southern Medical University (ZDXKKYTS0021), Elite Talent Training Program of Shenzhen Municipal Health Commission (2021)

Ethical Statement The study was approved by the Institutional Review Board of the Medical Ethics Committee of Shenzhen Hospital of Southern Medical University (ID number: NYSZYYEC20200037), and written informed consent was obtained from each participant

1 引 言

儿童阻塞性睡眠呼吸暂停综合征是一种导致儿童频繁出现呼吸暂停和低通气的睡眠障碍，不仅扰乱儿童的正常睡眠模式，还可能影响儿童面容，如出现牙列不齐等腺样体面容，甚至引发生长发育障碍、行为和学习能力损害，以及心血管和代谢问题[1-3]。美国儿科学会认为腺样体肥大是导致儿童阻塞性睡眠呼吸暂停综合征的主要病因[1]。腺样体是儿童鼻咽部的淋巴组织，当过度增生肥大时，会阻塞上呼吸道，导致儿童在睡眠时出现呼吸障碍。因此，腺样体的诊断对治疗儿童阻塞性睡眠呼吸暂停综合征意义重大。腺样体的检查方法主要包括影像学检查和电子鼻咽内镜检查。影像学检查因其辐射性而不易被患儿家长接受。当前，专家共识与学科指南更倾向于推荐电子鼻咽内镜作为评估腺样体状况的首选手段[1]。电子鼻咽内镜可全面观察腺样体区域的组织情况及腺样体形态，但仅提供二维光学图像，具有一定局限性。二维光学图像无法精确反映腺样体区域的空间位置和几何结构，导致诊断结果严重依赖医生的临床经验和对腺样体的观察角度，甚至可能对同一腺样体产生不同的评估结果[4-5]。在使用电子鼻咽镜评估腺样体大小时，

医生主要依据二维图像进行分度判断，从而给出腺样体大小的概数估计，无法提供精确数值。为保证结果的客观性，有时需要多名医生共同评估，增加了医疗资源的消耗[6]。因此，利用二维图像信息重构腺样体区域的三维结构，为医生的评估与诊断提供更客观、精确的指导，具有重要的现实意义。

近 20 年来，计算机视觉领域蓬勃发展，基于内窥镜影像的三维重建被广泛应用于医学领域，然而，基于计算机视觉的腺样体区域三维重建研究尚属空白。目前，基于内窥镜的三维重建技术主要使用阴影恢复形状算法[7-9]、基于多视角的三维重建(multi-view stereo，MVS)算法[10-11]和基于特殊相机的三维重建算法[12-13]。阴影恢复形状算法基于光照模型和照相机成像模型，构建图像强度、光照方向与强度和目标深度三者之间的偏微分方程，然后通过偏微分方程的数值解获得目标的深度信息[8-9]。该方法的精度严重依赖边界条件、光照方向和相关光度学参数，因此无法在现实中解决基于内窥镜图像的三维重建问题。现有 MVS 算法通过匹配不同视角图像中的特征点，重建物体或场景的三维结构[10-11]。但是 MVS 算法很难实现弱纹理区域的特征匹配，这将导致相关方法无法有效恢复三维结构。腺样体是人体的淋巴组织，具有黏膜表面，其内窥镜影像具有强反光、特征稀少、场景光滑、模糊等特点，因此现有的 MVS 算法无法准确重建腺样体区域的三维结构。另外，有学者通过使用基于 ToF(Time of Flight)相机或双目立体相机的内窥镜，直接获得内腔的三维结构信息[12-13]。然而，这类内窥镜的体积较大，难以插入儿童的鼻咽腔。

事实上，腺样体表面为光滑曲面，且曲面上任一点的局部邻域可被视为一个平面。针对腺样体这一特点，本文提出一种基于空间网格曲面拟合的多视图三维重建算法，原理见第 3 节。本研究内容主要包括：(1)设计了一种高效、鲁棒的局部平面参数优化算法，可优化每幅图像中每个像素的深度值；(2)提出了一种深度空间网格曲面拟合算法，通过构建深度空间的三维曲面，对深度场进行拟合，以获得平滑、细致的三维重建结果；(3)设计了一个完整的腺样体三维重建架构，有效重建了腺样体区域。仿真与真实实验表明，与现有方法相比，本文所提算法的重建结果更稠密、平滑、精确。

2　相关研究现状

随着三维重建技术的发展，越来越多的研究者将其应用于内窥镜影像的三维重建，从而辅助诊疗[7-10]。MVS 算法通过多个视角采集的图片恢复场景的三维结构。这类方法无须目标场景的先验信息，因此通常是唯一可用的基于内窥镜影像的三维重建方法。现有的多视图重建方法主要基于体素[14-15]、轮廓[16-17]、点云[18-19]和深度图[20-21]。基于体素的方法将三维空间划分成若干小体素，以雕刻三维场景。该方法受体素粒度大小的影响，占用大量内存资源。基于轮廓的方法仅能对单一目标进行三维重建，无法对场景进行三维重建。基于点云的方法直接对点云进行操作，但点云的不规则性限制了该方法的性能和表现。基于深度图的方法灵活高效，广受青睐。Shen[22]首先提出了基于块匹配的多视图重建框架。Zheng 等[23]提出了基于图模型的块匹配多视图重建算法，不仅加快了重建速度，还提高了重建精度。在此基础上，Schönberger 等[21]综合考虑了深度、法向量和图像像素之间的几何约束，提高了三维重建精确性。Schönberger 等[21]的算法虽然取得了令人满意的精度，但其序列式深度信息传播的计算效率严重受图像大小的影响。为克服上述问题，Galliani 等[20]提出了红黑棋盘深度信息传播策略，可同时处理图像中一半的

像素点，极大地提高了算法运算效率。然而，该算法的代价计算方法较简单，导致算法重建效果不佳。为解决这一问题，Xu 等[24]提出了新的自适应深度信息传播方案，扩大了深度传播范围，并提出了基于多个深度假设的代价融合策略，充分利用了多视图信息，提高了重建效果。近年来，多视图立体匹配方法中的研究热点和技术难点是弱纹理区域的重建。为精确重建弱纹理区域，学者们已经提出了多种方法。其中，有研究者提出构建多尺度的图像金字塔，在较粗糙的尺度上对弱纹理区域进行匹配，以期获得更精确的结果[25]。另外，还有学者利用超像素分割的方法，通过对超像素进行匹配提高重建的精确性[26]。此外，也有研究者尝试扩大弱纹理区域的匹配块，使其包含更多的纹理信息，从而提高匹配的准确性[27]。现有基于 MVS 算法的方法可以很好地应用于普通场景，但是无法实现对腺样体区域的精确重建，主要原因如下：(1)这些方法通常严重依赖多视图之间的图像特征匹配，而腺样体表面为黏膜组织，非常光滑且缺少纹理，因此无法为算法提供充足的特征匹配；(2)现有算法虽然可在一定程度上改善弱纹理平坦区域的重建困难问题，但重建方法仅考虑场景的局部信息，因此依旧无法有效处理腺样体上大面积的平坦、光滑区域。

3 基于内窥镜图像的腺样体重建算法原理

给定一组经过标定的鼻咽腔内窥镜拍摄的腺样体图像序列，每一幅图像对应的照相机位姿参数可通过运动恢复结构方法获得[28]。本文认为腺样体表面为光滑曲面，且曲面上的任一点 $\boldsymbol{x}$ 的局部邻域可被视为一个平面，该平面可由点 $\boldsymbol{x}$ 在照相机坐标系下的深度值与法向量表示，本文称该深度值与法向量为点 $\boldsymbol{x}$ 的局部平面参数。基于这一假设，本文提出了基于网格曲面拟合的腺样体三维重建算法，算法流程图如图 1 所示：(1)对一组输入图像，本文依次将每一幅图像作为参考图像，其余图像作为参考图像的源图像；(2)对参考图像中每个像素点对应的局部平面参数进行随机初始化(图 1 中红框)；(3)基于参考图像中每一点对应的局部平面参数，建立该点邻域与所有源图像之间的单应变换，由此构建该点的光度学代价；(4)根据每个像素点的光度学代价，通过参数传播和随机扰动策略，更新参考图像中每个像素点的局部平面参数(图 1 中绿框)；(5)基于深度一致性过滤掉明显错误的局部平面参数(图 1 中黄框)；(6)使用网格曲面在深度空间对参考图像的深度图进行拟合，从而获得平滑的局部平面参数(图 1 中蓝框)；(7)使用局部平面参数优化算法对拟合的局部平面参数进行更新，如

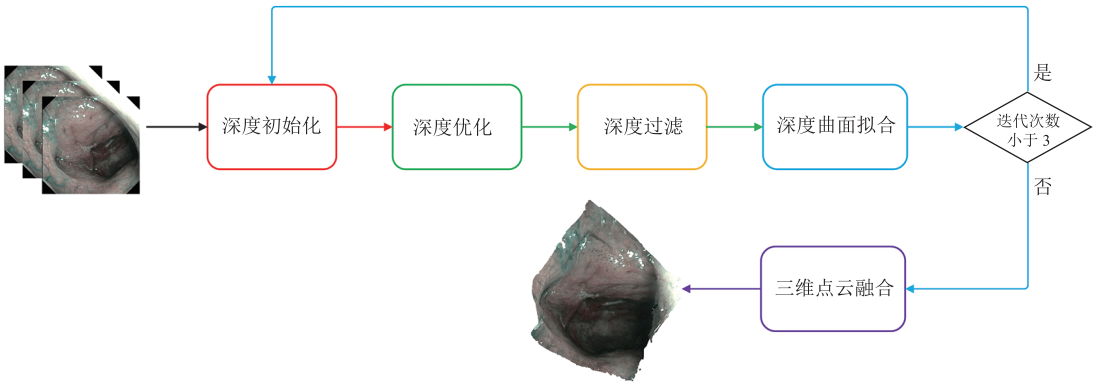


图 1 本文算法框架图

Fig. 1 The pipeline of our method

此迭代 3 次；(8)基于深度一致性和重投影一致性，通过融合所有输入图像的深度信息，获得腺样体区域的三维点云。

3.1　局部平面参数初始化

给定图像序列$\{I_s,s=1,2,\cdots,N\}$，I_s 所对应的照相机坐标系记为 C_s。对于任意参考图像 I_i，本文为 I_i 的每个像素点随机赋予深度值 d 和法向量 $\boldsymbol{n}$，从而构建该像素点的局部平面参数(图 1 中红框)。深度 d 是像素点对应的空间点在相机坐标系 C_i 下的 Z 分量，法向量 $\boldsymbol{n}$ 为像素点对应空间点所在空间平面的法向量。本文假设参考图像 I_i 的相机参数为 $P_i=\{\boldsymbol{K}_i,\boldsymbol{R}_i,\boldsymbol{t}_i\}$，源图像 I_j 的照相机参数为 $P_j=\{\boldsymbol{K}_j,\boldsymbol{R}_j,\boldsymbol{t}_j\}$，其中，$\boldsymbol{K}_i$、$\boldsymbol{K}_j$ 为相机的内参数；$\boldsymbol{R}_i$、$\boldsymbol{R}_j$ 分别为相机坐标系 C_1 到相机坐标系 C_i、C_j 的旋转矩阵；$\boldsymbol{t}_i$、$\boldsymbol{t}_j$ 分别为 C_1 到 C_i、C_j 的平移向量。那么，对于 I_i 中的任一像素点 $\boldsymbol{p}$，假设其对应的齐次坐标为 $\boldsymbol{p}=[u,v,1]^{\mathrm{T}}$，对应的深度值和法向量分别为 d_i 和 $\boldsymbol{n}_i$，通过平面单应映射，可得到其在源图像 I_j 中的对应点 $\boldsymbol{q}=\boldsymbol{H}_{ji}\boldsymbol{p}$，其中，$\boldsymbol{H}_{ji}$ 可表示如下[29]：

$$\boldsymbol{H}_{ji}=\boldsymbol{K}_j\left(\boldsymbol{R}_j\boldsymbol{R}_i^{-1}-\frac{\left(\boldsymbol{t}_j-\boldsymbol{R}_j\boldsymbol{R}_i^{-1}\boldsymbol{t}_i\right)\boldsymbol{n}_i^{\mathrm{T}}}{d_i\boldsymbol{n}_i^{\mathrm{T}}\boldsymbol{K}_i^{-1}\boldsymbol{p}}\right)\boldsymbol{K}_i^{-1} \tag{1}$$

基于式(1)，本文对局部平面参数进行初始化。本文在深度范围$[d_{\min},d_{\max}]$中随机采样得到 d，其中 $d_{\min}$ 和 $d_{\max}$ 分别为预先设定的深度最小值和最大值。对于法向量，本文令 $\boldsymbol{n}=[\cos\theta\sin\phi,\sin\theta\sin\phi,\cos\phi]^{\mathrm{T}}$，其中，$\theta\in[0,2\pi]$；$\phi$ 为像素点的反投影线与 Z 轴的夹角，本文认为，如果腺样体上的某一点是可见的，那么这个点对应的 ϕ 值应落在区间$[0,\pi/3]$内，因此，本文在该区间内随机采样获得 ϕ 的初始值。本文利用式(1)计算参考图像中每个像素点相对于每幅源图像的零均值归一化互相关系数(zero normalized cross-correlation，ZNCC)，然后以 3 个最小的 ZNCC 的平均值作为该点的初始光度学代价。

3.2　局部平面参数优化

根据每个像素点的光度学代价，本文将使用参数传播和随机扰动策略，以 3.1 节中获得的平面参数为初始值，更新参考图像中每个像素点的局部平面参数(图 1 中绿框)，算法原理如下。为提高算法的鲁棒性和运行效率，本文采用如图 2 所示的棋盘格模式进行参数传播和更新。对于参考图像 I_i 中的任意像素点 $\boldsymbol{p}$(图 2 中红点)，本文按照图 2 中绿框涵盖的范围定义其 8 个邻域。在每个邻域上搜索光度学代价最小的像素点，每个邻域参与比较的像素由图 2 蓝点表示。定义集合$\left\{\theta_{i_p}=\left\{d_{i_p},\boldsymbol{n}_{i_p}\right\},i_p=0,1,\cdots,8\right\}$，其中，$\theta_0$ 为像素点 $\boldsymbol{p}$ 对应的局部平面参数，而 $\theta_1\sim\theta_8$ 分别为像素点 $\boldsymbol{p}$ 的第 1～8 个邻域中光度学代价最小的像素点对应的局部平面参数。对于每一幅源图像 I_j，通过 θ_{i_p}，可计算像素点 $\boldsymbol{p}$ 关于 I_j 的 ZNCC 值 $\rho_{i_p j}$。显然，对于每一幅 I_j，像素点 $\boldsymbol{p}$ 都有 9 个对应的 ZNCC 值$\left\{\rho_{i_p j},i_p=0,1,\cdots,8\right\}$，因此像素点 $\boldsymbol{p}$ 总共有 $9\times(N-1)$ 个 ZNCC 值。接下来，本文将根据$\left\{\rho_{i_p j},i_p=0,1,\cdots,8\right\}$计算源图像 I_j 相对于像素点 $\boldsymbol{p}$ 的权重。对于每一幅源图像 I_j，本文定义集合 $S_p{}^j=\left\{\rho_{i_p j}{}^*\middle|\rho_{i_p j}{}^*\in\left\{\rho_{i_p j},i_p=0,1,\cdots,8\right\},\rho_{i_p j}{}^*\geqslant\tau\right\}$，其中 τ 为预设的阈值。如果 $S_p{}^j$ 中元素个数大于 n(预设阈值)，本文就认为源图像 I_j 对参考图像是有贡献的，因此将索引 j 放入集合 I_{sel}^p 中。然后，通过式(2)计算 $S_p{}^j$ 中每个 $\rho_{i_p j}{}^*$ 的置信度，该置信度将用于计算源图像 I_j 相对于像素点 $\boldsymbol{p}$ 的权重。

$$C\left(\rho_{i_p j}{}^*\right)=\mathrm{e}^{-\frac{\rho_{i_p j}{}^{*2}}{2\beta^2}},\ j\in I_{\mathrm{sel}}^p \tag{2}$$

其中 β 为一个常数。源图像 I_j 相对于像素点 $\boldsymbol{p}$ 的权重可表示如下：

$$w_j=\underset{\rho_{i_p j}{}^*\in S_p{}^j}{\mathrm{median}}\left(C\left(\rho_{i_p j}{}^*\right)\right),\ j\in I_{\mathrm{sel}}^p \tag{3}$$

其中 median 表示集合的中值。最终，像素点 $\boldsymbol{p}$

关于 θ_{i_p} 的光度学代价为

$$h_{i_p}=\frac{\sum_{j\in I_{\mathrm{sel}}^p} w_j \rho_{i_p,j}}{\sum_{j\in I_{\mathrm{sel}}^p} w_j},\ i_p=0,1,\cdots,8 \tag{4}$$

接下来，本文使用最小的光度学代价对应的局部平面参数更新当前像素点 $\boldsymbol{p}$ 的局部平面参数。

上述更新过程未引入新的局部平面参数。因此在上述更新过程完成后，设像素点 $\boldsymbol{p}$ 对应的局部平面参数为 $\theta_{\mathrm{c}}=\{d_{\mathrm{c}},\boldsymbol{n}_{\mathrm{c}}\}$，本文采用文献[21]中的方法，通过扰动 θ_c 和随机生成新的局部平面参数，构建以下 7 个备选平面参数：

$$\{\{d_{\mathrm{c}},\boldsymbol{n}_{\mathrm{c}}\},\{d_{\mathrm{c}},\boldsymbol{n}_{\mathrm{p}}\},\{d_{\mathrm{c}},\boldsymbol{n}_{\mathrm{r}}\},\{d_{\mathrm{p}},\boldsymbol{n}_{\mathrm{c}}\},\{d_{\mathrm{r}},\boldsymbol{n}_{\mathrm{c}}\},\{d_{\mathrm{r}},\boldsymbol{n}_{\mathrm{r}}\},\{d_{\mathrm{p}},\boldsymbol{n}_{\mathrm{p}}\}\} \tag{5}$$

其中，d_{p} 和 $\boldsymbol{n}_{\mathrm{p}}$ 是在 d_{c} 和 $\boldsymbol{n}_{\mathrm{c}}$ 上稍加扰动形成的；而 d_{r} 和 $\boldsymbol{n}_{\mathrm{r}}$ 则是完全随机生成的。然后，本文利用式(4)计算这 7 个局部平面参数对应的光度学代价，并将最小代价对应的局部平面参数作为当前像素点 $\boldsymbol{p}$ 的局部平面参数。这样既可以搜寻最优局部平面参数，又可以避免算法估计陷入局部最小值。

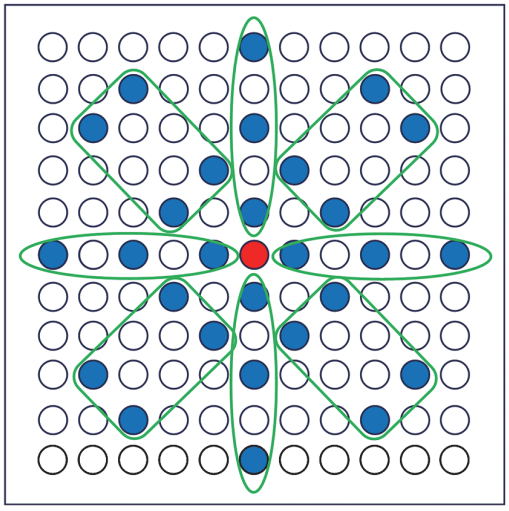


图 2 棋盘采样图案

Fig. 2 Pattern of checkboard sample

3.3 深度过滤

上述优化过程仅考虑不同图像之间的光度一致性，并未考虑不同深度图之间的几何一致性。由于腺样体图像的特征稀少、色彩单一，由此获得的深度图中存在许多明显的噪声点和错误点。因此在深度优化后，本文基于深度相对差过滤掉所有视图对应的深度图中明显的噪声点和错误点(图 1 中黄框)。对于参考图像 I_i 中的任一点 $\boldsymbol{p}$，本文通过式(1)找到点 $\boldsymbol{p}$ 在源图像 I_j 中的对应点 $\boldsymbol{q}$，其深度值为 d_q。点 $\boldsymbol{p}$ 对应的空间点在源图像 I_j 的相机坐标系下的深度值为 d_p，深度相对差为 $d_{\mathrm{r}}^j=\left|d_p-d_q\right|/d_p$。若至少有 n 幅源图像满足 $d_{\mathrm{r}}^j\leqslant 0.01$，则本文认为点 $\boldsymbol{p}$ 是可靠的，否则就删除点 $\boldsymbol{p}$ 的局部平面参数。该过滤算法移除了大部分明显的噪声点和错误点，为后续的网格曲面拟合提供了可靠的数据(具体算法见 3.4 节)。

3.4 空间网格曲面拟合

第 3.3 节的过滤算法去除了深度图中不可靠的深度信息，导致深度图中某些区域的深度缺失。为恢复这些深度信息，本文提出了空间网格曲面拟合方法(图 1 中蓝框)。具体来讲，本文认为鼻咽腔表面的每一点的邻域可以近似为一个小平面，且相近的两点的邻域的法向量是相近的。因此，本文在深度空间中使用一个平滑的三角网格曲面拟合过滤后的可靠点像素的深度值，从而获得图像深度值的平滑估计。首先，本文定义深度空间的 xy 平面与图像平面重合。在深度空间中，对于任意一个三角面片 T_i，其内部的点满足如下平面方程：

$$d=ax+by+c \tag{6}$$

其中，d 为 T_i 上某点的深度值；$[x,y]$为 T_i 上某点对应的像素坐标；$[a,b,c]$为平面参数。设三角面片 T_i 上的顶点由 $\boldsymbol{p}_{1,T_i}$，$\boldsymbol{p}_{2,T_i}$ 和 $\boldsymbol{p}_{3,T_i}$ 这 3 个点构成，其中，$\boldsymbol{p}_{j,T_i}=\left[x_{j,T_i},y_{j,T_i},d_{j,T_i}^*\right]^{\mathrm{T}}$，$j=1,2,3$；$x_{j,T_i}$ 和 y_{j,T_i} 为 $\boldsymbol{p}_{j,T_i}$ 的横纵坐标；d_{j,T_i}^* 为点 $\boldsymbol{p}_{j,T_i}$ 的深度值。因此，将 $\boldsymbol{p}_{1,T_i}$，$\boldsymbol{p}_{2,T_i}$ 和 $\boldsymbol{p}_{3,T_i}$ 代入式(6)得

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}=\begin{bmatrix} x_{1,T_i} & y_{1,T_i} & 1 \\ x_{2,T_i} & y_{2,T_i} & 1 \\ x_{3,T_i} & y_{3,T_i} & 1 \end{bmatrix}^{-1}\begin{bmatrix} d_{1,T_i}^* \\ d_{2,T_i}^* \\ d_{3,T_i}^* \end{bmatrix} \tag{7}$$

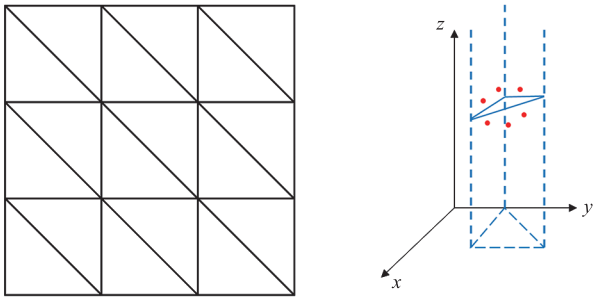
将式(7)代入式(6)可得三角面片 T_i 内的任意一个点的深度值，表示如下：

$$d=\begin{bmatrix}x & y & 1\end{bmatrix}\begin{bmatrix}x_{1,T_i} & y_{1,T_i} & 1\\ x_{2,T_i} & y_{2,T_i} & 1\\ x_{3,T_i} & y_{3,T_i} & 1\end{bmatrix}^{-1}\begin{bmatrix}d^*_{1,T_i}\\ d^*_{2,T_i}\\ d^*_{3,T_i}\end{bmatrix} \tag{8}$$

其中，$\left[x_{j,T_i},y_{j,T_i}\right],j=1,2,3$ 可通过对深度空间的 xy 平面(图像平面)进行三角剖分获得，如图 3(a)所示，因此为定值。d^*_{j,T_i} 为待求变量，不同的 d^*_{j,T_i} 决定了三角面片在深度空间中的不同位置。本文用三角面片逼近场景三维信息，尽可能使三角面片曲面拟合可靠像素点的深度，如图 3(b)所示，x 轴和 y 轴为图像平面的横纵坐标，z 轴表示深度，红色点表示图像中可靠像素点在深度空间中对应的位置，虚线三角形表示对像素平面剖分得到的三角面片(剖分原理见图 3(a))，实线三角形表示拟合可靠点后三角面片的空间位置。同时，为平滑三角面片曲面，本文要求相邻两个三角面片的参数尽量一致，最终本文构建了如下方程

$$\min\underbrace{\sum_{T_i\in T}\sum_{\boldsymbol{q}_{\varepsilon_i}\in pro(T_i)}\left(d_{\varepsilon_i,T_i}-\tilde{\boldsymbol{q}}_{\varepsilon_i}{}^{\mathrm{T}}\begin{bmatrix}x_{1,T_i} & y_{1,T_i} & 1\\ x_{2,T_i} & y_{2,T_i} & 1\\ x_{3,T_i} & y_{3,T_i} & 1\end{bmatrix}^{-1}\begin{bmatrix}d^*_{1,T_i}\\ d^*_{2,T_i}\\ d^*_{3,T_i}\end{bmatrix}\right)^2}_{\text{数据项}}+$$
$$\underbrace{\lambda\sum_{(T_u,T_v)\in adj}\left\|\begin{bmatrix}x_{1,T_u} & y_{1,T_u} & 1\\ x_{2,T_u} & y_{2,T_u} & 1\\ x_{3,T_u} & y_{3,T_u} & 1\end{bmatrix}^{-1}\begin{bmatrix}d^*_{1,T_u}\\ d^*_{2,T_u}\\ d^*_{3,T_u}\end{bmatrix}-\begin{bmatrix}x_{1,T_v} & y_{1,T_v} & 1\\ x_{2,T_v} & y_{2,T_v} & 1\\ x_{3,T_v} & y_{3,T_v} & 1\end{bmatrix}^{-1}\begin{bmatrix}d^*_{1,T_v}\\ d^*_{2,T_v}\\ d^*_{3,T_v}\end{bmatrix}\right\|_2^2}_{\text{平滑项}} \tag{9}$$

其中，T 为所有三角面片的集合；$pro(T_i)$为三角面片 T_i 在 xy 平面上的正交投影；$\boldsymbol{q}_{\varepsilon_i}$ 为 $pro(T_i)$内的可靠像素点坐标；$\tilde{\boldsymbol{q}}_{\varepsilon_i}$ 为 $\boldsymbol{q}_{\varepsilon_i}$ 的齐次坐标；adj 为相邻平面对构成的集合；d_{ϵ_i,T_i} 为 $\boldsymbol{q}_{\varepsilon_i}$ 对应的深度值；λ 为平滑参数。式(9)中的数据项表示在可靠像素点处，拟合的深度值与可靠像素点深度值的差异，平滑项表示相邻两个三角面的参数差异。式(9)为二次规划问题，本文通过使用稀疏矩阵理论，可以很快得到所有三角面片的顶点对应的深度值$\left\{d^*_{j,T_i},j=1,2,3,T_i\in T\right\}$，从而得到每个三角面片的平面方程。然后，本文用所有三角面片的平面方程估计 xy 平面(图像平面)内所有像素点的深度值，在填补深度图空洞的同时，又能优化原本的可靠深度值。完成深度拟合后，本文再使用局部平面参数优化算法对拟合结果进行更新，如此迭代 3 次(如图 1 所示)。



(a) 像素平面分割策略　　(b) 深度空间拟合

图 3　空间网格曲面拟合

Fig. 3　Spatial mesh surface fitting

3.5　三维点云融合

本文将拟合后的深度图融合成点云(图 1 中紫框)。融合的条件包括深度相对差和重投影误差。深度相对差的计算方法如 3.3 节所述。参考图像 I_i 中一点 $\boldsymbol{x}_i$ 的邻域到源图像 I_j 的单应矩阵 $\boldsymbol{H}_{ji}$ 可由式(1)确定。假设点 $\boldsymbol{x}_i$ 在源图像 I_j 中的对应点为 $\boldsymbol{x}_j=\boldsymbol{H}_{ji}\boldsymbol{x}_i$，由 $\boldsymbol{x}_j$ 处的局部平面参数易得 $\boldsymbol{x}_j$ 的邻域到图像 I_i 的单应矩阵 $\boldsymbol{H}_{ij}$。那么 $\boldsymbol{x}_j$ 在 I_i 中的对应点为 $\boldsymbol{x}^*_i=\boldsymbol{H}_{ij}\boldsymbol{x}_j$，则 $\boldsymbol{x}_i$ 关于 I_j 的重投影误差表示如下：

$$e_{\text{repro}}{}^j=\left\|\hat{\boldsymbol{x}}_i-\hat{\boldsymbol{x}}^*_i\right\|_2 \tag{10}$$

其中 $\hat{\boldsymbol{x}}$ 为点 $\boldsymbol{x}$ 的非齐次坐标。若至少有 m 幅源图像满足：$e_{\text{repro}}{}^j\leqslant 2$，$d^j_r\leqslant 0.01$，则保留 $\boldsymbol{x}_i$ 对应的三维点，并将 $\boldsymbol{x}_i$ 在所有满足上述条件的源图像中的对应点删除，以避免重复计算。最终获得的三维点云即为腺样体区域的三维重建结果。

4　实验与结果分析

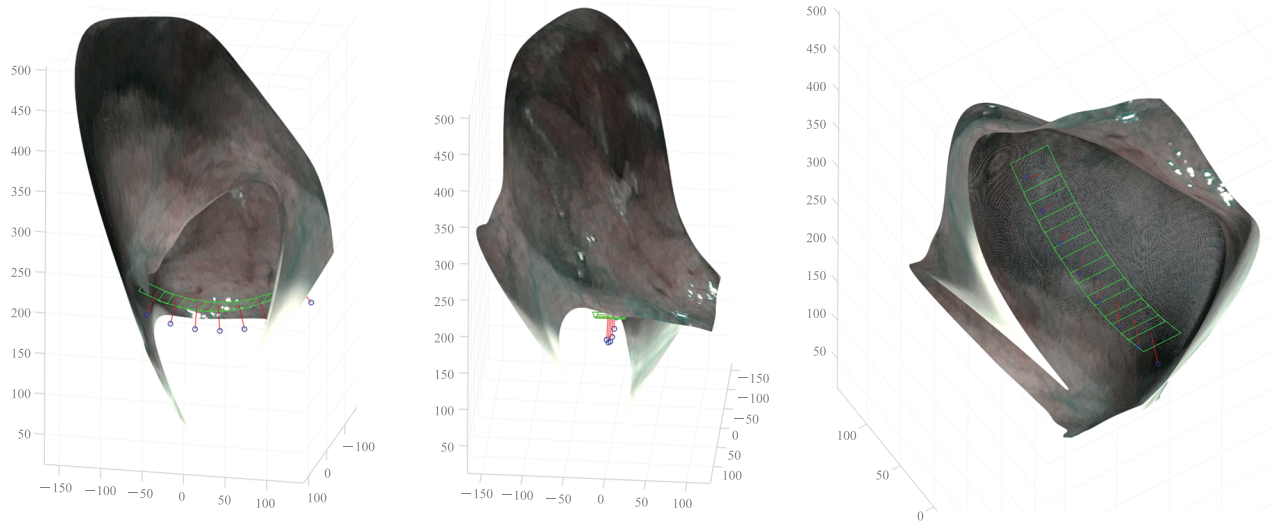
本文通过仿真实验与真实实验定性与定量地

评估本文算法的有效性。本文的实验在一台配备 Intel(R) Xeon(R) Silver 4114 处理器、32 GB RAM 和两张 NVIDIA Quadro RTX 4000 显卡的联想工作站上进行，实验代码基于 MATLAB 和 C++ 语言编写并优化。本实验与 ACMH[24]、ACMM[24]、ACMMP[25]和 APD[30]方法进行了对比，其中 ACMMP 和 APD 是近两年的先进方法，它们在公开数据集上表现优异。

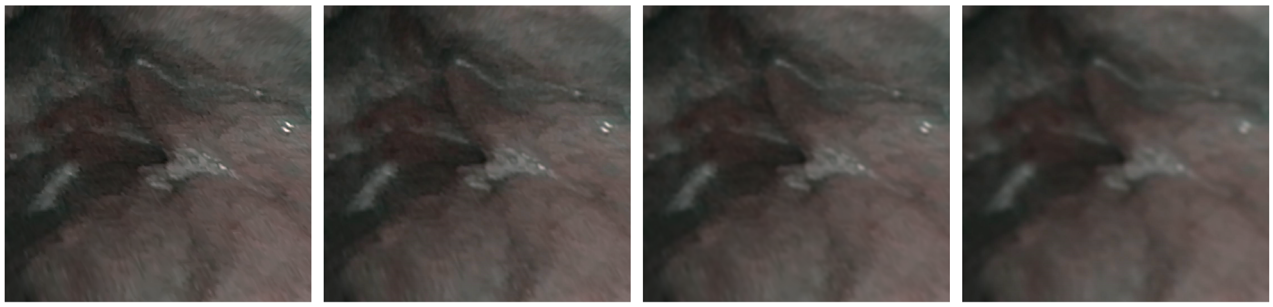
4.1 仿真实验

在仿真实验中，本文使用基于经典光线追踪法的计算机图形学方法获取腺样体的仿真模型和图像。具体来说，本文首先根据相关的临床经验，人为构造出与若干张内窥镜图像中腺样体区域对应的三维面片模型(三维模型如图 4(a)所示)；然后设置照相机的拍摄位置与拍摄角度(照相机的拍摄位置如图 4(a)蓝点所示)，并通过三维面片模型与内窥镜图像的纹理映射关系，使用光线追踪法渲染出腺样体区域的仿真图像。光线追踪法的技术细节见文献[31]。鼻腔内分泌物会使鼻咽腔内镜图像变模糊。为模拟这种情况，本文对仿真的腺样体图像加了不同程度的高斯模糊，模糊尺度从 1 增至 8，不同模糊尺度的图像如图 4(b)所示。

本文使用精度、完整度和 F 分数这 3 个指标评估本文算法与对比算法的三维重建结果。假设真实点云为 $\mathcal{G}$，重建点云为 $\mathcal{R}$，对于重建点云中的任意一点 $\boldsymbol{r}\in\mathcal{R}$，它到真实点云 $\mathcal{G}$ 的距离定义为 $e_{\boldsymbol{r}\to\mathcal{G}}=\min\limits_{\boldsymbol{g}\in\mathcal{G}}\|\boldsymbol{r}-\boldsymbol{g}\|_2$。同理，真实点云中的点 $\boldsymbol{g}$



(a) 腺样体的仿真三维模型



(b) 不同模糊尺度的仿真腺样体采集图像

图 4 腺样体的仿真模型和图像

Fig. 4 Simulation models and images of the adenoids

到重建点云 $\mathcal{R}$ 的距离定义为 $e_{g\to\mathcal{R}}=\min\limits_{r\in\mathcal{R}}\|\boldsymbol{g}-\boldsymbol{r}\|_2$。那么，当给定误差阈值 d 时，精度可反映重建点云的相对误差情况，其计算方法如下

$$P(d)=\frac{100}{|\mathcal{R}|}\sum_{r\in\mathcal{R}}\left|\left\{e_{r\to\mathcal{G}}\,|\,e_{r\to\mathcal{G}}<d\right\}\right| \tag{11}$$

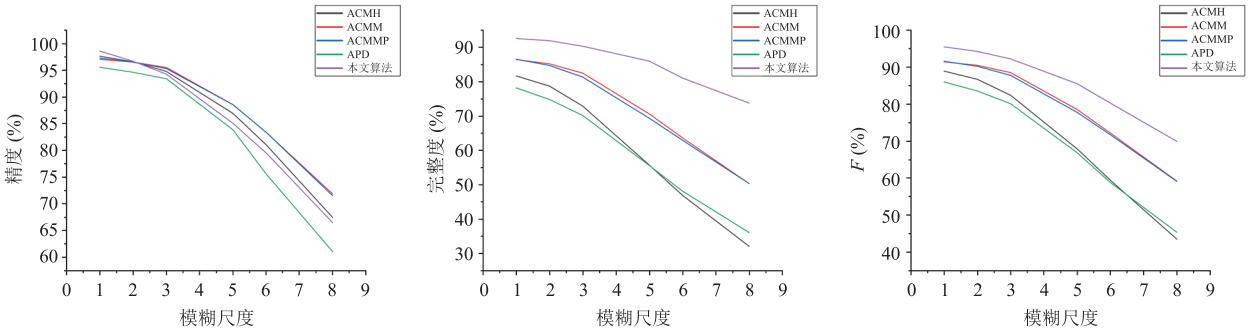
其中 $|*|$ 为集合中元素的个数。完整度可反映重建点云的稠密情况，表示如下：

$$R(d)=\frac{100}{|\mathcal{G}|}\sum_{g\in\mathcal{G}}\left|\left\{e_{g\to\mathcal{R}}\,|\,e_{g\to\mathcal{R}}<d\right\}\right| \tag{12}$$

F 分数是精度和完整度的综合指标，表示如下：

$$F(d)=\frac{2P(d)R(d)}{P(d)+R(d)} \tag{13}$$

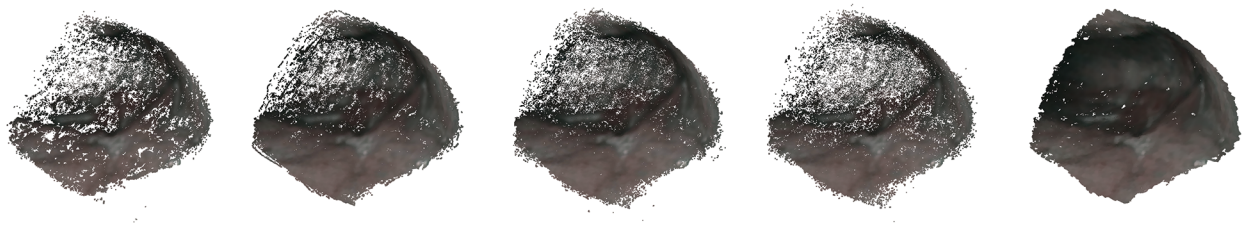
仿真实验结果如图 5 所示。由图 5 可知，所有方法的精度、完整度与 F 分数都随着模糊尺度的增加而显著下降。这是因为模糊尺度的增大加大了特征匹配的难度，在匹配过程中容易出现误匹配点，这些误匹配点严重影响三维重建点云的准确性。在精度方面，各种方法差别不大，当模糊尺度为 8 时，所有方法的精度都在 60% 以上。在完整度方面，本文方法明显优于对比方法。在各个模糊尺度下，本文算法的完整度都大幅高于所有对比方法。随着模糊尺度从 1 增至 8，本文方法的完整度仅下降了 19%。相比之下，所有对比方法的完整度下降了超过 35%。当模糊尺度为 1 时，本文算法的完整度接近 95%。当模糊尺度为 8 时，本文方法的完整度仍然将近 75%。此时，作为次优算法的 ACMM 的完整度仅约 50%。本文方法的 F 分数亦显著优于对比方法，当模糊尺度为 1 时，本文算法的 F 分数大于 95%。当模糊尺度为 8 时，本文算法的 F 分数约为 70%，比次优算法高约 10%。由图 5 可知，本文算法的精确性和鲁棒性均优于现有最新方法，原因是本文提出的空间网格曲面拟合方法充分满足了腺样体表面的平滑特性。另外，本文重建算法的整体框架具有对外点出色的鲁棒性。当模糊尺度为 8 时，各方法的重建结果如图 6 所示。可以看出，本文算法的重建结果更稠



(a) 精度比较　(b) 完整度比较　(c) F 分数比较

图 5　不同模糊尺度下的仿真结果比较

Fig. 5　Comparison of simulation results under different fuzziness scales



(a) ACMH　(b) ACMM　(c) ACMMP　(d) APD　(e) 本文算法

图 6　模糊尺度为 8 时仿真实验的重建结果比较

Fig. 6　Comparison of reconstruction results of simulation experiments at a fuzziness scales of 8

密、平滑。不同模糊尺度下，各算法的重建点云数量如表 1 所示，该表定量地反映了本文算法的重建结果更稠密。

4.2 真实实验

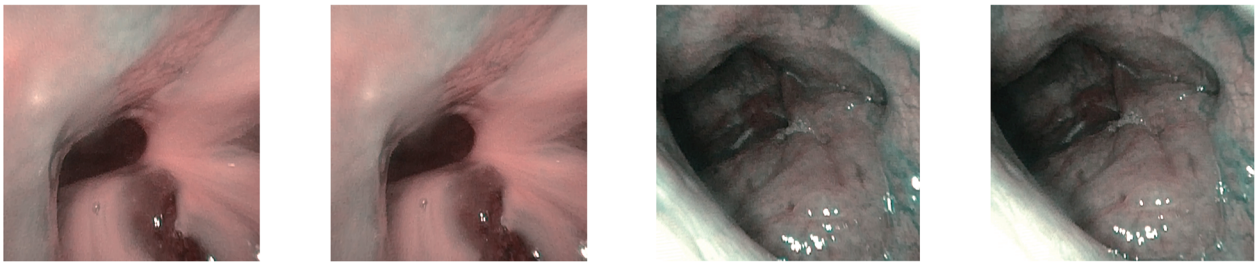
真实实验使用的鼻咽内窥镜图像数据集由南方医科大学深圳医院儿童耳鼻喉科提供。本文使用文献[32]中的算法标定鼻咽内窥镜内参数，并对内窥镜图像进行畸变校正。真实实验使用的图像如图 7 所示，图像的尺寸为 1 080×1 080。本实验分别使用本文方法和对比方法对两位患者的腺样体区域进行重建，并进行了比较。本文算法和对比算法的深度图如图 8 所示，其中蓝色越深的点，深度值越小，反之，深度值越大。由图 8 可知，对比方法的结果中存在较多噪声点，相比之下，本文算法生成的深度图更平滑，保证了深度信息的全局连续性，表明本文算法在深度估计准确性和一致性方面较优。

使用本文算法与对比算法对不同患者的腺样体区域进行三维重建的多视角结果如图 9 所示。由图可知，对比方法生成的点云较稀、噪声大，仅能粗略地重建腺样体区域。本文算法的重建结果较稠密、平滑，几何结构一致性较好。综上所述，真实实验定性地证明了本文算法在弱纹理区域的深度估计问题上的优越性，非常适用于腺样体区域的三维

表 1 不同模糊尺度下的重建点云的数量

Table 1 The number of reconstructed point cloud under different fuzziness scales

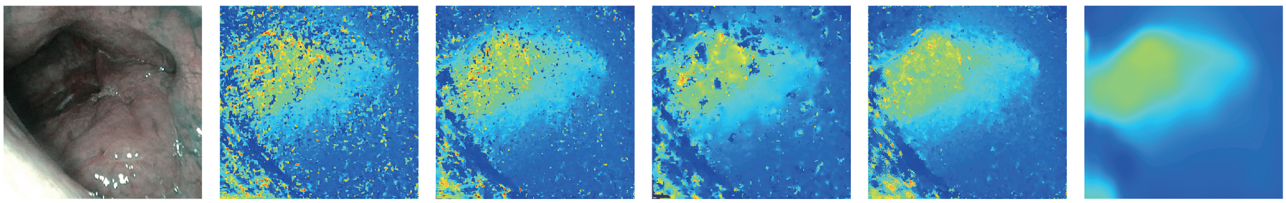
模糊尺度	1	2	3	5	6	8
ACMH	1 182 445	1 046 608	830 121	452 966	337 502	211 423
ACMM	1 397 162	1 336 374	1 218 969	892 294	752 559	528 702
ACMMP	827 314	691 682	574 841	413 635	360 587	279 131
APD	498 413	391 218	311 850	210 213	180 776	142 797
本文算法	1 608 002	1 566 867	1 552 566	1 412 461	1 405 668	1 374 924



(a) 患者 1 的腺样体图像 (b) 患者 2 的腺样体图像

图 7 腺样体图片预处理结果

Fig. 7 Results for adenoid image preprocessing



(a) 参考图像 (b) ACMH (c) ACMM (d) ACMMP (e) APD (f) 本文算法

图 8 不同算法估计的深度图

Fig. 8 Depth maps estimated by different algorithms

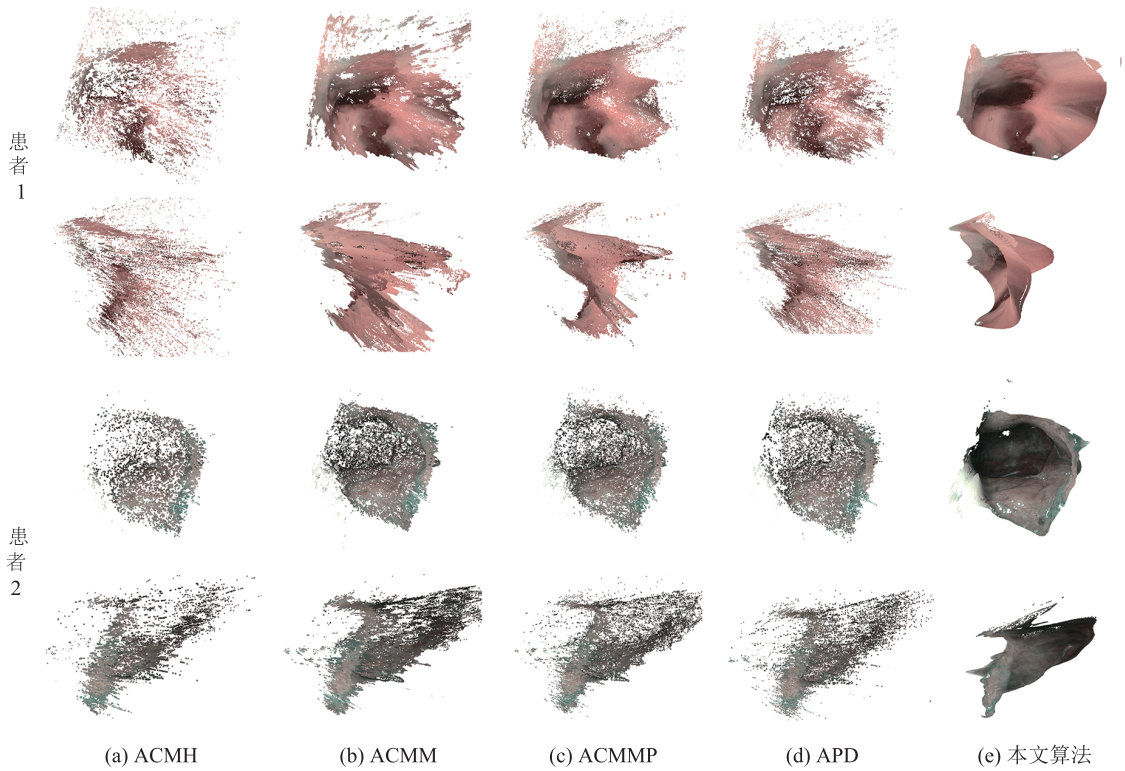


图 9　不同算法重建的腺样体点云比较

Fig. 9　Comparison of adenoid point cloud reconstructed by different algorithms

重建。本文算法虽然在腺样体重建任务上取得了一定进展，但仍存在局限性。与现有技术相比，本文算法虽然在应对图像模糊方面的鲁棒性较好，但重建结果仍然不可避免地受模糊的影响。由于腺样体表面常附有鼻腔分泌物，因此模糊成为腺样体图像的一个固有特征，严重影响重建图像的质量。为克服这一问题，未来，本团队考虑结合深度学习技术对图像模糊进行针对性处理，以期改善重建结果。此外，通过仔细观察真实实验的重建结果可知，本文算法的重建结果在某些区域过于平滑，微小的细节未能有效重建。为进一步提升重建图像的质量，本团队将在未来的工作中优化曲面拟合算法，力求在构建平滑场景的同时保留场景的细节部分。

5　结　论

本文提出一种基于网格曲面拟合的多视图立体匹配算法。该算法首先对图像对应的深度图进行粗糙估计；其次，针对腺样体的平滑特征，在深度空间中应用网格曲面拟合技术对可靠的深度信息进行精细化拟合，实现了深度图的平滑和优化；最后，融合所有深度信息，实现了对腺样体区域的密集且精确的三维重建。实验结果表明，与现有算法相比，本文提出的算法在腺样体三维重建上具有优越性。然而，在物体边缘，即深度突变区域和大范围的弱纹理区域，本文算法仍存在局限，这是本研究团队未来的研究工作。

参 考 文 献

[1]　Marcus CL, Brooks LJ, Ward SD, et al. Diagnosis and management of childhood obstructive sleep apnea syndrome [J]. Pediatrics, 2012, 130(3): e714-e755.

[2]　Davies CR, Harrington JJ. Impact of obstructive

sleep apnea on neurocognitive function and impact of continuous positive air pressure [J]. Sleep Medicine Clinics, 2016, 11(3): 287-298.

[3] Smith DF, Amin RS. OSA and cardiovascular risk in pediatrics [J]. Chest, 2019, 156(2): 402-413.

[4] Zhao QY, Price T, Pizer S, et al. The endoscopogram: a 3D model reconstructed from endoscopic video frames [C] // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016: 439-447.

[5] 张雅丽, 崔晓波. 儿童阻塞型睡眠呼吸暂停低通气综合征诊断及治疗的研究进展 [J]. 临床医学进展, 2021, 11(6): 2596-2602.
Zhang YL, Cui XB. Overview on diagnosis and treatment of obstructive sleep apnea-hypopnea syndrome in children [J]. Advances in Clinical Medicine, 2021, 11(6): 2596-2602.

[6] 张明君, 徐宏鸣, 浦诗磊, 等. 儿童腺样体肥大的评估与分级研究进展 [J] . 中华耳鼻咽喉头颈外科杂志, 2023, 58(2): 188-192.
Zhang MJ, Xu HM, Pu SL, et al. Evaluation and grading progression of adenoid hypertrophy in children [J]. Chinese Journal of Otorhinolaryngology Head and Neck Surgery, 2023, 58(2): 188-192.

[7] Richa R, Poignet P, Liu C. Three-dimensional motion tracking for beating heart surgery using a thin-plate spline deformable model [J]. The International Journal of Robotics Research, 2010, 29(2-3): 218-230.

[8] Tankus A, Sochen N, Yeshurun Y. Reconstruction of medical images by perspective shape-from-shading [C] // Proceedings of the 17th International Conference on Pattern Recognition, 2004: 778-781.

[9] 王瑞胡. 基于 Shape from Shading 的医学图像三维重建 [J]. 计算机工程与应用, 2008, 44(6): 222-224.
Wang RH. 3-D reconstruction of median image Shape from Shading based [J]. Computer Engineering and Applications, 2008, 44(6): 222-224.

[10] Grasa OG, Bernal E, Casado S, et al. Visual SLAM for handheld monocular endoscope [J]. IEEE Transactions on Medical Imaging, 2013, 33(1): 135-146.

[11] Widya AR, Monno Y, Okutomi M, et al. Whole stomach 3D reconstruction and frame localization from monocular endoscope video [J]. IEEE Journal of Translational Engineering in Health and Medicine, 2019, 7: 3300310.

[12] Penne J, Höller K, Stürmer M, et al. Time-of-flight 3-D endoscopy [C] // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2009: 467-474.

[13] Zhou HY, Jagadeesan J. Real-time dense reconstruction of tissue surface from stereo optical video [J]. IEEE Transactions on Medical Imaging, 2019, 39(2): 400-412.

[14] Faugeras O, Keriven R. Variational principles, surface evolution, PDE's, level set methods and the stereo problem [J]. IEEE Transactions on Image Processing, 1998, 7(3): 336-344.

[15] Sinha SN, Mordohai P, Pollefeys M. Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh [C] // Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, 2007: 1-8.

[16] Cremers D, Kolev K. Multiview stereo and silhouette consistency via convex functionals over convex domains [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(6): 1161-1174.

[17] Esteban CH, Schmitt F. Silhouette and stereo fusion for 3D object modeling [J]. Computer Vision and Image Understanding, 2004, 96(3): 367-392.

[18] Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereopsis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362-1376.

[19] Goesele M, Snavely N, Curless B, et al. Multi-view stereo for community photo collections [C] // Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, 2007: 1-8.

[20] Galliani S, Lasinger K, Schindler K. Massively parallel multiview stereopsis by surface normal

diffusion [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 873-881.

[21] Schönberger JL, Zheng EL, Frahm JM, et al. Pixelwise view selection for unstructured multi-view stereo [C] // Proceedings of the European Conference on Computer Vision, 2016: 501-518.

[22] Shen SH. Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes [J]. IEEE Transactions on Image Processing, 2013, 22(5): 1901-1914.

[23] Zheng EL, Dunn E, Jojic V, et al. Patchmatch based joint view selection and depthmap estimation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1510-1517.

[24] Xu QS, Tao WB. Multi-scale geometric consistency guided multi-view stereo [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5483-5492.

[25] Xu QS, Kong WH, Tao WB, et al. Multi-scale geometric consistency guided and planar prior assisted multi-view stereo [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(4): 4945-4963.

[26] Romanoni A, Matteucci M. TAPA-MVS: textureless-aware PatchMatch multi-view stereo [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 10413-10422.

[27] Xu ZY, Liu YG, Shi XL, et al. MARMVS: matching ambiguity reduced multiple view stereo for efficient large scale scene reconstruction [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5981-5990.

[28] Schonberger JL, Frahm JM. Structure-from-motion revisited [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4104-4113.

[29] Hartley R, Zisserman A. Multiple view geometry in computer vision [M]. Cambridge: Cambridge University Press, 2003.

[30] Wang YS, Zeng ZJ, Guan T, et al. Adaptive patch deformation for textureless-resilient multi-view stereo [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 1621-1630.

[31] Hearn D. Computer graphics with OpenGL [M]. Beijing:Publishing House of Electronics Industry, 2005.

[32] Zhang ZY. Flexible camera calibration by viewing a plane from unknown orientations [C] // Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999: 666-673.