

基于PP-Matting抠图和增量式SfM的 三维重建方法

任梦欣^{1,2}, 杨剑锋^{1,3*}, 邓周灰^{2,4}, 邹琼⁵, 全天乐^{2,6}

¹贵州大学数学与统计学院, 贵州 贵阳

²贵大·贵安科创超级计算算力算法应用实验室, 贵州 贵阳

³贵州理工学院大数据学院, 贵州 贵阳

⁴贵安新区科创产业发展有限公司, 贵州 贵阳

⁵深圳瑞云科技股份有限公司, 广东 深圳

⁶贵州黔驴科技有限公司, 贵州 贵阳

收稿日期: 2023年6月16日; 录用日期: 2023年7月19日; 发布日期: 2023年7月26日

摘要

基于视觉的三维重建技术通过获取物体的真实图像来还原其三维模型。然而, 这些获取的图像通常包含大量无用的背景信息, 直接使用这样的图像进行三维重建将导致计算资源和存储空间的浪费。为了解决上述问题, 本文提出了一种融合PP-Matting抠图和增量式SfM的三维重建方法, 该方法在使用SfM和MVS算法完成三维重建之前, 对物体的原始图像进行抠图。本文利用Distinctions-646等多个图像集对PP-Matting抠图模型进行微调训练, 得到仅包含待重建物体的图像。实验结果表明, 本文提出的方法在重建效率方面取得显著提升, 并且能够降低存储空间需求。

关键词

三维重建, 抠图, PP-Matting, SfM + MVS

3D Reconstruction Method Based on PP-Matting and Incremental Structure-from-Motion

Mengxin Ren^{1,2}, Jianfeng Yang^{1,3*}, Zhouhui Deng^{2,4}, Qiong Zou⁵, Tianle Tong^{2,6}

¹School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

²Guizhou University and Gui'an Kechuang Supercomputing Power Algorithm Laboratory, Guiyang Guizhou

³School of Data Science, Guizhou Institute of Technology, Guiyang Guizhou

⁴Gui'an New District Science and Innovation Industry Development Co., Ltd., Guiyang Guizhou

⁵Shenzhen Rayvision™ Technology Co., Ltd., Shenzhen Guangdong

*通讯作者。

文章引用: 任梦欣, 杨剑锋, 邓周灰, 邹琼, 全天乐. 基于PP-Matting抠图和增量式SfM的三维重建方法[J]. 建模与仿真, 2023, 12(4): 4116-4126. DOI: 10.12677/mos.2023.124375

⁶Guizhou Qianlu Technology Co., Ltd., Guiyang Guizhou

Received: Jun. 16th, 2023; accepted: Jul. 19th, 2023; published: Jul. 26th, 2023

Abstract

Visual-based 3D reconstruction techniques aim to restore the three-dimensional models of objects by capturing their real images. However, these captured images often contain a significant amount of irrelevant background information, and directly using such images for 3D reconstruction results in wastage of computational resources and storage space. To address these issues, this paper proposes a three-dimensional reconstruction method that combines PP-Matting image matting and incremental Structure-from-Motion (SfM). The proposed method performs image matting on the original images of the objects before utilizing SfM and MVS algorithms for 3D reconstruction. The PP-Matting model is fine-tuned using multiple image datasets, including Distinctions-646, to obtain images that solely contain the objects to be reconstructed. Experimental results demonstrate that the proposed method significantly improves reconstruction efficiency and reduces storage space requirements.

Keywords

3D Reconstruction, Image Matting, PP-Matting, SfM + MVS

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

视觉三维重建即通过物体的二维图像重建出它的三维模型，能广泛应用于数字孪生城市、无人驾驶、元宇宙等新兴领域。相比于激光扫描等三维重建方法，视觉三维重建具有低成本、高效率的特点，适用场景更广泛。视觉三维重建的精度和速度受多种因素影响，包括输入数据的质量、算法的准确性和稳定性，以及应用场景的要求等。

基于视觉的三维重建以原始图像集作为基础，而往往图像中并非只有待重建的对象，还存在大量无用信息，如果基于原始图像直接进行计算，就会耗费大量计算资源去重建无用模型，如图 1 所示。

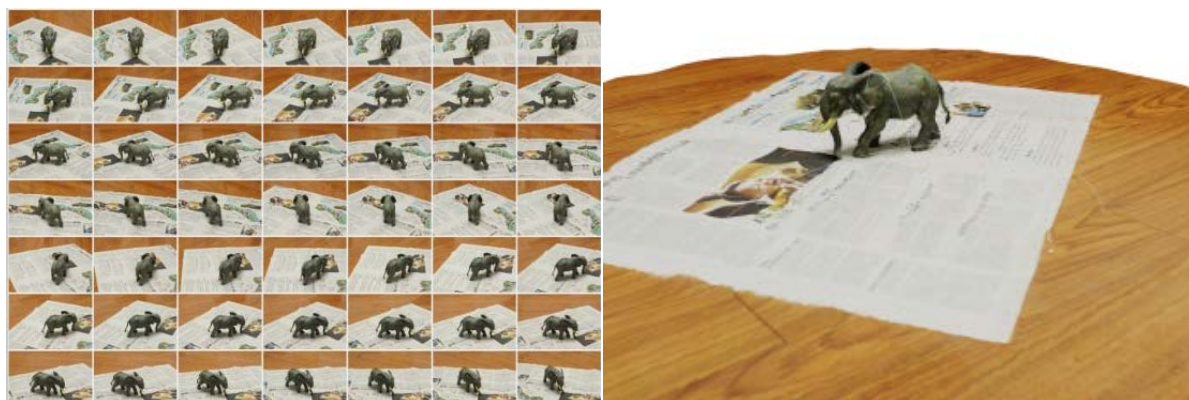


Figure 1. 3D reconstruction results with original image as input

图 1. 以原始图像为输入的三维重建结果

为了解决这一问题,有研究人员从图形获取层面进行无用信息削减(将环境背景设置为纯色),这样的方法对于针对小物品的三维重建可能会有效果,但不能适用于大场景(如建筑)。

本文的目标是从重建层面消除无用信息,提高计算效率。为此,本文以 SfM 进行稀疏重建,以 MVS 进行稠密重建,来完成三维重建任务,并在其中加入智能裁剪功能。具体来说,本文的智能裁剪功能有以下流程:

- 1) 获取原始图像集;
- 2) 对原始图像集进行智能裁剪(分割、抠图)以获得新的图像集;
- 3) 将原始图像的细节信息(相机信息、图像信息)加入到新图像集中;
- 4) 用新的图像集进行稀疏重建及后续工作。

本文采用 PP-Matting 抠图模型作为智能裁剪方法,以重建时间、重建精度等指标进行对比分析,探寻有效的方法,使得三维重建的效率和精度有所提高。

2. 国内外研究现状

2.1. 视觉三维重建

基于视觉的三维重建已经发展了数十年,取得了巨大的成功,无论是传统方法还是深度学习的方法都展示出了很好的效果。

传统的视觉三维重建算法发展历程较长,相关方法较为完整,目前为止依然是主流。M Farenzena 等[1]提出了一种基于分层聚类树的结构和运动恢复方法,用于从多张图像中估计场景中的点和相机的 3D 位置和姿态。N Snavely 等[2]介绍了一种基于计算机视觉和计算机图形学技术的方法,利用从互联网上获取的大量图片,生成高质量的三维重建模型,从而实现对不同场景的浏览和探索。

基于深度学习的视觉三维重建算法依据处理的数据形式主要分为体素、点云和网格的处理。Eigen 等[3]基于体素形式,直接用单张图像使用神经网络恢复深度图,将网络分为全局粗估计和局部精估计,并用一个尺度不变的损失函数进行回归。Choy 等[4]基于体素形式提出的 3D-R2N2 模型使用 Encoder-3DLSTM-Decoder 的网络结构建立 2D 图形到 3D 体素模型的映射,完成了基于体素的单视图/多视图三维重建。Yu 等[5]利用潜在空间中的特定类别的多模态先验分布训练变分自编码器,利用潜在空间的子集就可以找到先验分布的目标模态,获取类别的先验信息,随后将先验信息和图像特征共同送入解码器重建三维模型。Fan 等[6]提出了用点云做三维重建的开山之作,解决了训练点云网络时候的损失问题。Chen 等[7]通过对场景的点云进行处理,融合三维深度和二维纹理信息,提高了点云的重建精度。Wang 等[8]用三角网格来做单张 RGB 图像的三维重建。

2.2. 图像抠图

传统抠图方法可以分为三类。首先是基于色彩采样的方法,这类方法依赖于像素之间的强相关性,从已知的前景或背景颜色中采样并将它们应用于未知像素[9];其次是基于相似性的方法,这类方法通常计算相似度矩阵来表征相邻像素之间的相似性,并相应的将 α 值从已知区域扩散到未知区域[10][11][12];最后是基于色彩采样和相似性的优化方法,以获得更稳健的解决方案,达到更好的效果[13][14]。尽管这些方法通过综合设计在预测结果方面取得了显著的进步,但它们的表示能力受到低级颜色或结构特征的限制,难以将前景细节与复杂的自然背景区分开来。此外,由于这些方法中的大多数都需要手动标记辅助输入,因此抠图结果通常对未知区域的大小和模糊边界非常敏感。

基于深度学习的抠图算法可以分为两类,即基于辅助输入的抠图算法和自动抠图算法。基于辅助输入的抠图算法有三种方法论:一是使用单个 one-stage CNN 将输入图像和辅助输入的连接直接映射到

alpha 上[15] [16]; 二是将 one-stage CNN 与自定义的模块相结合使用, 以利用侧分支的辅助输入所提供的丰富特征来提高抠图的质量[17] [18]; 三是利用并行双模态或多模态结构将抠图任务分解为显式子任务[19] [20]。对于自动抠图方法, 也有三种主要的方法论: 一是 one-stage 结构, 可以选择性地包括一个全局模块作为指导, 直接从单个输入图像中预测出抠图结果[21]; 二是顺序两步结构, 首先生成中间分割掩码或 trimap, 然后与初始输入组合以产生最终的 alpha 遮罩[22]; 三是并行的双模态或多模态结构, 它将抠图任务分解为几个子任务, 例如前景和背景或全局语义掩码[23]和局部细节[24] [25]。

2.3. 结合图像分割的视觉三维重建

通过将图像分割技术应用于三维重建, 可以更好地理解场景的结构、提取物体的几何信息, 并实现更准确的三维重建结果。

J McCormac 等[26]提出了一种使用卷积神经网络进行密集的三维语义建图的方法, 将深度学习与稠密三维重建相结合, 实现了准确的三维语义建模。JJ Park 等[27]将深度学习与三维形状表示相结合, 通过学习连续的有符号距离函数来表示物体的几何形状。Shuairfeng Zhi 等[28]在基于 nerf 的三维重建训练中, 加入语义分割的训练, 并证明了联合 nerf 重建和语义分割, 可以对粗糙或者错误的语义 label 有矫正的作用, 并可以用于自动化标注。

抠图算法是图像分割中的一种, 其本质都是根据图像原始信息或标注信息来分离特定的对象, 但因其工作侧重不同, 两者之间也存在些许区别。具体来说, 图像分割是把图像分成若干个特定的、具有独特性质的区域并提出感兴趣目标的技术和过程; 图像抠图是将图像分为前景和背景, 提取出感兴趣的前景再进行后处理。对于三维重建而言, 往往只需要感兴趣的前景, 而不需要环境背景, 因此在抠图算法比分割算法更适合本文的研究内容。

3. 三维重建算法

3.1. 图像抠图

3.1.1. 抠图技术

图像抠图是指从图像中提取出我们所感兴趣的前景目标, 同时过滤掉背景部分。一张图像可以简单的看成是由两部分组成, 即前景和背景。简单来说, 抠图就是将一张给定图像的前景和背景区分开来。

假设原始图像用 S_i 来表示, α 表示对应的 Alpha 通道, F_i 和 B_i 分别表示对应的前景图像和背景图像, 那么一张具有 RGBA 通道的图像可以分解为如下几部分的组合:

$$S_i = \alpha \times F_i + (1 - \alpha) \times B_i \quad (1)$$

当 α 为 0 时, 图像为背景图像; 当 α 为 1 时, 图像为前景图像。对于图像中的每个像素点, 均可以表示为一个类似于上述的线性方程组。因此, 抠图的主要目标是根据原始输入图像, 来获得前景、背景和透明度。

3.1.2. PP-Matting 算法

PP-Matting 是 Guowei Chen 等[29]于 2022 年提出的一种能实现高精度自然图像抠图的无 trimap 网络架构, 由语义上下文分支(SCB)和高分辨率细节分支(HRDB)组成, 分别用于预测语义图和细节图。最后将语义图与细节图相融合以实现相互增强, 生成准确的 alpha 遮罩。

PP-Matting 网络由语义上下文分支(SCB)和高分辨率细节分支(HRDB)组成, 两个分支共享一个公共模块作为编码器, 即公共编码器。在 SCB 之前, 利用金字塔池化模块(PPM)来加强语义上下文。此外, 指导流(guidance flow)用于连接 SCB 和 HRDB, 有助于通过正确的语义指导进行细节预测。网络架构如

图 2 所示。

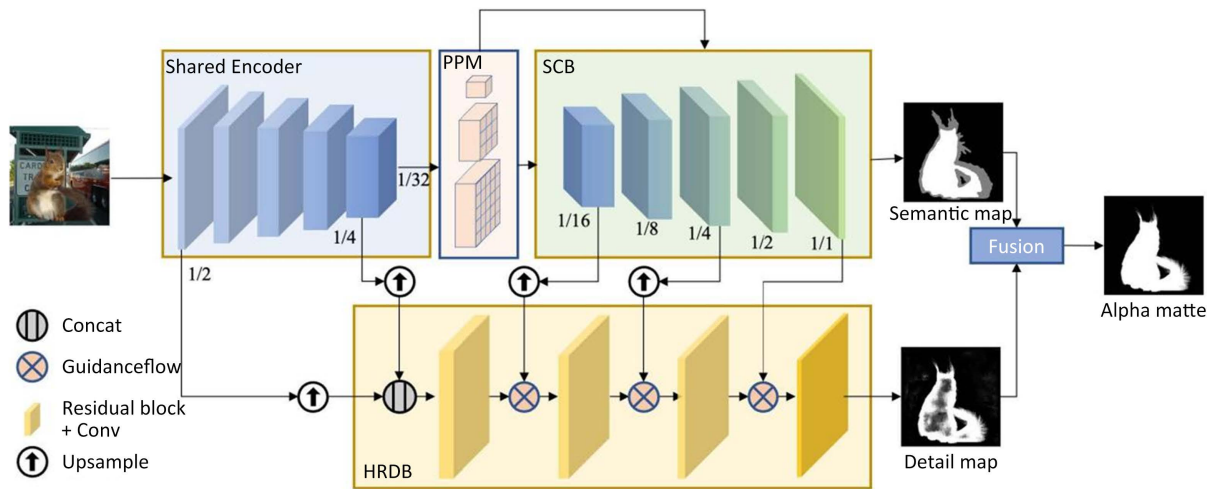


Figure 2. PP-Matting network architecture [29]

图 2. PP-Matting 网络架构[29]

模型运用了三种损失函数。第一个是 SCB 中的语义损失，表示为 L_s ，这是 3 类分割任务的交叉熵损失：

$$L_s = -\sum_{c=1}^3 \sum_{i=1}^{\Omega} g_c^i \log(p_c^i) \quad (2)$$

其中 $c \in \{1,2,3\}$ 表示语义图中的三个类， $p_c^i \in [0,1]$ 是第 i 个像素处第 c 类的预测概率， $g_c^i \in \{0,1\}$ 是相应的真实值， Ω 表示图像中所有的像素：

$$\Omega = \Omega^f \cup \Omega^b \cup \Omega^t \quad (3)$$

其中 $\Omega^f, \Omega^b, \Omega^t$ 分别表示前景像素、背景像素和过渡像素。

第二个损失是 HRDB 中的细节损失，表示为 L_d ：

$$L_d = \sum_{i=1}^{\Omega^t} (L_{\alpha}^i(d) + L_{grad}^i(d)) \quad (4)$$

其中 L_{α} 是 alpha 预测损失， L_{grad} 是梯度损失。

第三个损失是最终 alpha 遮罩中的融合损失，表示为 L_f ，由 alpha 预测损失、梯度损失和合成损失共同组成：

$$L_f = \sum_i^{\Omega} (L_{\alpha}^i(p) + L_{grad}^i(p) + L_{comp}^i) \quad (5)$$

其中 L_{comp}^i 是真实 RGB 颜色与真实前景、背景和预测的 alpha 遮罩合成的预测 RGB 颜色之间的绝对差异。

最终的加权损失计算如下：

$$L = \lambda_1 L_s + \lambda_2 L_d + \lambda_3 L_f \quad (6)$$

3.2. 增量式 SfM

增量式 SfM 是一边三角化(triangulation)和 pnp (perspective-n-points)，一边进行局部 BA (Bundle Adjustment, 捆绑调整)的一种稀疏重建方法，重建结果是稀疏点云。BA 算法是对初始结果进行非线性优化以均匀化误差和获得更精确的结果，从本质上来说，BA 是一个优化模型，其目的是最小化重投影误差。在已知结果的情况下对求解的参数进行误差最小化，使得求解结果更加精确的过程。一般使用重投影误差来对该问题进行优化：将特征匹配点记为观测点，三维空间点为求解点。根据求解的相机外参数(旋转

矩阵、平移向量)将三维空间点重投影到相机成像平面中。由于误差的存在,该点与观测点不会重合,优化的目标就是使得重投影误差最小。优化模型如下:

$$g(\theta) = \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^m \chi_{ij} \hat{u}_{ij} - u_{ij}^2 \quad (7)$$

$$u_{ij} = K \left[R_j | t_j \right] X_i = C_j X_i \quad (8)$$

上式中, X_i 代表三维空间中的第 i 个坐标点 $(x_i, y_i, z_i)^T$, C_j 为第 j 个相机的内参, u_{ij} 是 X_i 在第 i 个相机中的投影点, \hat{u}_{ij} 为观测点。由于并不是每一个视角的相机都能在另一个相机中找到对应的投影点,因此若 X_i 在第 j 个相机中可见,则 $\chi_{ij}=1$, 否则 $\chi_{ij}=0$ 。BA 算法的目的是最小化 $g(\theta)$, 找到使得 $g(\theta)$ 最小时的 u_{ij} , 即优化 C_j 。最终, 经过 BA 算法的优化可以得到最小化的 $g(\theta)$ 值, 在第 5 章中以 *Rmse* 体现, 作为重建质量评价指标之一。

4. 实验

4.1. 数据集

本文在 5 个数据集上进行实验: PPM-100 [30]、Composition-1k [31]、Distinctions-646 [32]、AIM-500 [33]、AM-2K [24]。在 PPM-100 中, 训练集包含 90 张原始图像和对应的 alpha 遮罩, 测试集包含 10 张原始图像和对应的 alpha 遮罩; 在 Composition-1k 中, 训练集包含 431 张原始图像和对应的 alpha 遮罩, 测试集包含 50 张原始图像和对应的 alpha 遮罩; 在 Distinctions-646 中, 训练集包含 596 张原始图像及其相应的 alpha 遮罩, 测试集包含 50 张原始图像及其对应的 alpha 遮罩; 在 AIM-500 中, 包含 500 张原始图像及其对应的 alpha 遮罩; 在 AM-2K 中, 训练集包含 1800 张原始图像和对应的 alpha 遮罩, 测试集包含 200 张原始图像和对应的 alpha 遮罩。本文选取上述 5 个数据集中 3417 张原始图像及其相应的 alpha 遮罩作为训练集, 310 张原始图像及其相应的 alpha 遮罩作为测试集。

4.2. 实验流程

4.2.1. 训练

本文使用上述的数据集对 PP-Matting 进行了微调。为了使得输入图像保持一致, 在训练阶段将输入图像尺寸调整为 512×512 (Mat(512)模型)和 1024×1024 (Mat(1024)); 随机梯度下降(SGD)优化器中的动量参数设置为 0.9, 权重衰减为 $4e-5$; 学习率初始化为 0.01, 并通过多项式策略进行调整, 指数为 0.9, 迭代次数为 100,000; 方程中的系数设为 $\lambda_1 = \lambda_2 = \lambda_3 = 1$ 。本文的所有实验都是在 RTX 3080 单个 GPU 上进行, batch_size 为 2。

训练结果如表 1 所示。由训练结果可知, Mat (512)模型最佳 SAD (绝对差总和)值为 108.7288, 最佳 MSE (均方误差)值为 0.004, 最佳 iter (迭代次数)为 37000; Mat (1024)模型最佳 SAD (绝对差总和)值为 63.5426, 最佳 MSE (均方误差)值为 0.003, 最佳 iter (迭代次数)为 58,000。Mat (1024)模型的 SAD 值与 MSE 值都小于 Mat (512)模型, 说明 Mat (1024)模型更为准确, 预计其预测效果优于 Mat (512)模型。

Table 1. Training results

表 1. 训练结果

	SAD	MSE	Iter
Mat (512)	108.7288	0.0040	37000
Mat (1024)	63.5426	0.0035	58000

4.2.2. 测试及预测

利用训练得到的 best_model 对测试集图像进行抠图测试，部分结果如图 3 所示。



Figure 3. Test example. From top to bottom are: original picture, predicted alpha picture, Foreground image after background removal

图 3. 测试示例。从上到下依次是：原始图片、预测的 alpha 图、去除背景后的前景图

从图 3 可以看到，模型在测试集上展示出了比较好的抠图效果，接下来将它用于三维重建中的原始图像抠图。如图 4 所示，使用抠图模型对待重建的原始图像进行抠图，得到抠图后的图像序列。可以发现，从该抠图模型获得的新图像序列很好的达到了本文实验要求，即仅包含待重建对象，不含其余无用背景信息。



Figure 4. Cuts out the original image. The top is the original image, and the bottom is the cutout image

图 4. 对原始图片进行抠图。上是原始图片，下是抠图之后的图片

4.2.3. 三维重建

经过以上抠图处理，获得抠图后的新图像集，将原始图像集的 exif 信息逐个写入该新图像集，得到用于三维重建的图像集。依次进行 SfM 以及 MVS，如图 5 所示，最终得到三维模型。



Figure 5. Reconstruction results of SfM and MVS

图 5. SfM 与 MVS 的重建结果

5. 结论分析

图 6 展示了不同流程的三维重建效果。首先仅从视觉上观察，未加入抠图流程(SfM + MVS)的重建结果相对比较完整，重建质量比较好，未发现明显缺失，但模型中存在大量无用的背景信息；加入 Mat (512) 抠图模型的重建结果不含有无用的背景信息，但重建质量相对于 SfM + MVS 而言稍差，存在部分模型缺失的情况；加入 Mat (1024) 抠图模型的重建结果同样不含无用信息，且重建质量相比于 Mat (512) 抠图模型的结果稍好，未出现明显的模型缺失。





Figure 6. 3D reconstruction results of different processes

图 6. 不同流程的三维重建结果

表 2 展示了三种不同三维重建流程的部分重建参数,由表可知,从重建时间和模型内存来看,Mat (512) + SfM + MVS 流程最优;从 SfM 场景 RMSE 来看, SfM + MVS 流程更优(“人像石雕”中的 RMSE 值存在意外,推测可能是因为原始图像中的背景被虚化,导致背景信息对重建过程造成了干扰)。

总体来说,加入抠图算法后的三维重建虽然其重建质量有所降低(降低的程度取决于原始图像的背景信息质量),但能获得更高的重建效率,同时能够减少大量的无用信息存储,只保留有用部分,大大节约了存储空间。这样的特性对于某些行业和研究方向来说具有很高的需求性,比如三维模型资产存档、文物数字化归档、三维重建业务平台、针对物品的三维重建等等。

Table 2. 3D reconstruction parameters of different processes

表 2. 不同流程的三维重建参数

重建对象	重建流程	重建时间	模型内存	SfM 场景 RMSE
大象摆件	① SfM + MVS	34min	34792.26 KB	0.397886
	② Mat(512) + SfM + MVS	8min	3107.58 KB	0.859279
	③ Mat(1024) + SfM + MVS	12min	5514.41 KB	0.801039
人像石雕	① SfM + MVS	45 min	37345.60 KB	0.242221
	② Mat(512) + SfM + MVS	30 min	23521.09 KB	0.219907
	③ Mat(1024) + SfM + MVS	31 min	23883.13 KB	0.213924
运动鞋	① SfM + MVS	90 min	74788.16 KB	0.712938
	② Mat(512) + SfM + MVS	21 min	6901.43 KB	0.731410
	③ Mat(1024) + SfM + MVS	24 min	6906.18 KB	0.718293

6. 结论与展望

本文提出了一种融合 PP-Matting 抠图和增量式 SfM 的三维重建方法,该方法在使用 SfM 和 MVS 算法完成三维重建之前,对物体的原始图像进行抠图。本文利用 Distinctions-646 等多个图像集对 PP-Matting 抠图模型进行微调训练,得到仅包含待重建物体的图像。实验结果表明,本文提出的方法在重建效率方面取得显著提升,并且能够降低存储空间需求。

但是经过抠图后的图像集存在三个问题:一是由于大量的信息被消除,因此在特征提取时所能提取的特征点减少,导致 SfM 计算三维坐标的精度下降,最终体现为 SfM 生成的稀疏点云不够精确, MVS

生成的三维模型质量会有所降低；二是抠图算法往往是沿着待处理对象的边缘进行提取，这导致了边缘像素模糊化，因此图像中的待重建对象的边缘信息会有所减少，在最终的三维模型中体现为表面纹理不平整，甚至有所缺失；三是以这样的方案来进行三维重建，其重建质量很大一部分取决于抠图质量的好坏，因此抠图模型必须要能准确无误地对前景进行提取，但就目前而言，想要在任何场景下达到这样的理想效果还比较困难。

基于此，本文在此提出对未来工作的几点展望：① 对抠图算法做特定修改，使其扩大边缘范围，不丢失边缘信息；② 在抠图算法中加入语义识别分割算法，能够智能选取特定对象；③ 不使用图像抠图，直接分割稀疏点云，不丢失辅助点信息。

参考文献

- [1] Farenzena, M., Fusiello, A. and Gherardi, R. (2009) Structure-and-Motion Pipeline on a Hierarchical Cluster Tree. 2009 *IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, Kyoto, 27 September-4 October 2009, 1489-1496. <https://doi.org/10.1109/ICCVW.2009.5457435>
- [2] Snavely, N., Seitz, S.M. and Szeliski, R. (2006) Photo Tourism: Exploring Photo Collections in 3D. *ACM Transactions on Graphics*, **25**, 835-846. <https://doi.org/10.1145/1141911.1141964>
- [3] Eigen, D., Puhrsch, C. and Fergus, R. (2014) Depth Map Prediction from a Single Image Using a Multi-Scale Deep Network. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Montreal, 8-13 December 2014, 2366-2374.
- [4] Choy, C.B., Xu, D.F., Gwak, J., Chen, K. and Savarese, S. (2016) 3D-R2N2: A Unified Approach for Single and Multi-View 3D Object Reconstruction. *Computer Vision—ECCV 2016: 14th European Conference*, Amsterdam, 11-14 October 2016, 1-17.
- [5] Yu, H. and Oh, J. (2022) Anytime 3D Object Reconstruction Using Multi-Modal Variational Autoencoder. *IEEE Robotics and Automation Letters*, **7**, 2162-2169. <https://doi.org/10.1109/LRA.2022.3142439>
- [6] Fan, H.Q., Su, H. and Guibas, L.J. (2017) A Point Set Generation Network for 3D Object Reconstruction from a Single Image. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 2463-2471. <https://doi.org/10.1109/CVPR.2017.264>
- [7] Chen, R., Han, S.F., Xu, J. and Su, H. (2019) Point-Based Multi-View Stereo Network. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 1538-1547. <https://doi.org/10.1109/ICCV.2019.00162>
- [8] Wang, N.Y., Zhang, Y.D., Li, Z.W., Fu, Y.W., Liu, W. and Jiang, Y.-G. (2018) Pixel2mesh: Generating 3D Mesh Models from Single Rgb Images. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 55-71. https://doi.org/10.1007/978-3-030-01252-6_4
- [9] Yang, C.J., Duraiswami, R., Gumerov, N.A. and Davis, L. (2003) Improved Fast Gauss Transform and Efficient Kernel Density Estimation. *Proceedings 9th IEEE International Conference on Computer Vision*, Nice, 13-16 October 2003, 664-671. <https://doi.org/10.1109/ICCV.2003.1238383>
- [10] Li, X., Li, J. and Lu, H. (2019) A Survey on Natural Image Matting with Closed-Form Solutions. *IEEE Access*, **7**, 136658-136675. <https://doi.org/10.1109/ACCESS.2019.2941832>
- [11] Levin, A., Rav-Acha, A. and Lischinski, D. (2008) Spectral Matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**, 1699-1712. <https://doi.org/10.1109/TPAMI.2008.168>
- [12] Rother, C., Kolmogorov, V. and Blake, A. (2004) “GrabCut” Interactive Foreground Extraction Using Iterated Graph Cuts. *ACM Transactions on Graphics (TOG)*, **23**, 309-314. <https://doi.org/10.1145/1015706.1015720>
- [13] Weiss, Y. and Freeman, W.T. (2001) On the Optimality of Solutions of the Max-Product Belief-Propagation Algorithm in Arbitrary Graphs. *IEEE Transactions on Information Theory*, **47**, 736-744. <https://doi.org/10.1109/18.910585>
- [14] Szeliski, R. (2006) Locally Adapted Hierarchical Basis Preconditioning. *ACM Transactions on Graphics*, **25**, 1135-1143. <https://doi.org/10.1145/1141911.1142005>
- [15] Li, Y. and Lu, H. (2020) Natural Image Matting via Guided Contextual Attention. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 11450-11457. <https://doi.org/10.1609/aaai.v34i07.6809>
- [16] Forte, M. and Pitié, F.F.B. (2020) Alpha Matting.
- [17] Ding, H.H., Zhang, H., Liu, C. and Jiang, X.D. (2022) Deep Interactive Image Matting with Feature Propagation. *IEEE Transactions on Image Processing*, **31**, 2421-2432. <https://doi.org/10.1109/TIP.2022.3155958>

-
- [18] Sun, Y., Tang, C.-K. and Tai, Y.-W. (2021) Semantic Image Matting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 11115-11124. <https://doi.org/10.1109/CVPR46437.2021.01097>
- [19] Sengupta, S., Jayaram, V., Curless, B., *et al.* (2020) Background Matting: The World Is Your Green Screen. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 2288-2297. <https://doi.org/10.1109/CVPR42600.2020.00236>
- [20] Yu, H.C., Xu, N., Huang, Z.L., Zhou, Y.Q. and Shi, H. (2021) High-Resolution Deep Image Matting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 3217-3224. <https://doi.org/10.1609/aaai.v35i4.16432>
- [21] Yu, Z.J., Li, X.H., Huang, H.J., Zheng, W. and Chen, L. (2021) Cascade Image Matting with Deformable Graph Refinement. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 7147-7156.
- [22] Liu, J.L., Yao, Y., Hou, W.D., Cui, M.M., Xie, X.S., Zhang, C.S. and Hua, X.-S. (2020) Boosting Semantic Human Matting with Coarse Annotations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 8560-8569.
- [23] Zhang, Y.K., Gong, L.X., Fan, L.B., Ren, P.R., Huang, Q.X., Bao, H.J. and Xu, W.W. (2019) A Late Fusion CNN for Digital Matting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 7461-7470. <https://doi.org/10.1109/CVPR.2019.00765>
- [24] Li, J.Z., Zhang, J., Maybank, S.J. and Tao, D.C. (2022) Bridging Composite and Real: Towards End-to-End Deep Image Matting. *International Journal of Computer Vision*, **130**, 246-266. <https://doi.org/10.1007/s11263-021-01541-0>
- [25] Li, J.Z., Ma, S.H., Zhang, J. and Tao, D.C. (2021) Privacy-Preserving Portrait Matting. *Proceedings of the 29th ACM International Conference on Multimedia*, 20-24 October 2021, 3501-3509.
- [26] McCormac, J., Handa, A., Davison, A. and Leutenegger, S. (2017) Semanticfusion: Dense 3D Semantic Mapping with Convolutional Neural Networks. *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, 29 May-3 June 2017, 4628-4635. <https://doi.org/10.1109/ICRA.2017.7989538>
- [27] Park, J.J., Florence, P., Straub, J., Newcombe, R. and Lovegrove, S. (2019) Deepsdf: Learning Continuous Signed Distance Functions for Shape Representation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 165-174.
- [28] Zhi, S.F., Laidlow, T., Leutenegger, S. and Davison, A.J. (2021) In-Place Scene Labelling and Understanding with Implicit Scene Representation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 15838-15847.
- [29] Chen, G.W., *et al.* (2022) PP-Matting: High-Accuracy Natural Image Matting.
- [30] Ke, Z.H., Sun, J.Y., Li, K.C., Yan, Q. and Lau, R.W.H. (2022) Modnet: Real-Time Trimap-Free Portrait Matting via Objective Decomposition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 1140-1147. <https://doi.org/10.1609/aaai.v36i1.19999>
- [31] Xu, N., Price, B., Cohen, S. and Huang, T. (2017) Deep Image Matting. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 2970-2979.
- [32] Qiao, Y., Liu, Y.H., Yang, X., Zhou, D.S., Xu, M.L., Zhang, Q. and Wei, X.P. (2020) Attention-Guided Hierarchical Structure Aggregation for Image Matting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 13676-13685. <https://doi.org/10.1109/CVPR42600.2020.01369>
- [33] Li, J.Z., Zhang, J. and Tao, D.C. (2021) Deep Automatic Natural Image Matting. *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, Montreal, 19-27 August 2021, 800-806. <https://doi.org/10.24963/ijcai.2021/111>