

# 基于视觉的三维重建关键技术研究综述

郑太雄<sup>1</sup> 黄帅<sup>1</sup> 李永福<sup>2</sup> 冯明驰<sup>1</sup>

**摘要** 三维重建在视觉方面具有很高的研究价值, 在机器人视觉导航、智能车环境感知系统以及虚拟现实中被广泛应用. 本文对近年来国内外基于视觉的三维重建方法的研究工作进行了总结和分析, 主要介绍了基于主动视觉下的激光扫描法、结构光法、阴影法以及 TOF (Time of flight) 技术、雷达技术、Kinect 技术和被动视觉下的单目视觉、双目视觉、多目视觉以及其他被动视觉法的三维重建技术, 并比较和分析这些方法的优点和不足. 最后对三维重建的未来发展作了几点展望.

**关键词** 三维重建, 主动视觉, 被动视觉, 关键技术

**引用格式** 郑太雄, 黄帅, 李永福, 冯明驰. 基于视觉的三维重建关键技术研究综述. 自动化学报, 2020, 46(4): 631–652

**DOI** 10.16383/j.aas.2017.c170502

## Key Techniques for Vision Based 3D Reconstruction: a Review

ZHENG Tai-Xiong<sup>1</sup> HUANG Shuai<sup>1</sup> LI Yong-Fu<sup>2</sup> FENG Ming-Chi<sup>1</sup>

**Abstract** 3D reconstruction is important in vision, which can be widely used in robot vision navigation, intelligent vehicle environment perception and virtual reality. This study systematically reviews and summarizes the progress related to 3D reconstruction technology based on active vision and passive vision, i. e. laser scanning, structured light, shadow method, time of flight (TOF), radar, Kinect technology and monocular vision, binocular vision, multi-camera vision, and other passive visual methods. In addition, extensive comparisons among these methods are analyzed in detail. Finally, some perspectives on 3D reconstruction are also discussed.

**Key words** 3D reconstruction, active vision, passive vision, key techniques

**Citation** Zheng Tai-Xiong, Huang Shuai, Li Yong-Fu, Feng Ming-Chi. Key techniques for vision based 3D reconstruction: a review. *Acta Automatica Sinica*, 2020, 46(4): 631–652

三维重建经过数十年的发展, 已经取得巨大的成功. 基于视觉的三维重建在计算机领域是一个重要的研究内容, 主要通过使用相关仪器来获取物体的二维图像数据信息, 然后, 再对获取的数据信息进行分析处理, 最后, 利用三维重建的相关理论重建出真实环境中物体表面的轮廓信息. 基于视觉的三维重建具有速度快、实时性好等优点, 能够广泛应用于人工智能、机器人、无人驾驶、SLAM (Simultaneous localization and mapping)、虚拟现

实和 3D 打印等领域, 具有重要的研究价值<sup>[1–3]</sup>, 也是未来发展的重要研究方向.

1963 年, Roberts<sup>[4]</sup> 首先提出了使用计算机视觉的方法从二维图像获取物体三维信息的可能性, 也就是从这时开始, 基于视觉的三维重建快速发展, 涌现出了许多新方法. 从发表在 ICCV (International Conference on Computer Vision)、ECCV (European Conference on Computer Vision) 和 CVPR (International Conference on Computer Vision and Pattern Recognition) 等知名国际会议上的相关论文数量增长情况便可看出其研究发展程度. 发达国家对于三维重建技术的研究工作起步比较早, 研究相对比较深入. 1995 年, 日本东京大学的 Kiyasu 等<sup>[5]</sup> 利用物体反射的 M-array coded 光源影像对物体表面进行三维重建. 随着研究更进一步的深入, 2006 年, Snavely 等<sup>[6]</sup> 开发出了 Photosynth 和 Photo Tourism 两个三维重建系统. 这两个系统的优点是能够自动计算每一帧图像的视点, 从而可以重建出物体的稀疏三维模型. 遗憾的是, 稀疏三维模型重建的效果并不是很清晰, 可视化程度较低, 需要进行稠密三维模型重建. 2008 年,

收稿日期 2017-10-24 录用日期 2018-07-05  
Manuscript received October 24, 2017; accepted July 5, 2018  
国家自然科学基金 (61773082, 51505054), 重庆市基础与前沿技术项目 (cstc2018jcyjAX0684), 重庆邮电大学交叉项目 (A2018-02), 重庆市重点产业共性关键技术创新专项项目 (cstc2015zdcy-ztzz60002) 资助

Supported by National Natural Science Foundation of China (61773082, 51505054), Basic Science and Emerging Technology of Chongqing (cstc2018jcyjAX0684), Project of Crossing and Emerging Area of CQUPT (A2018-02), and Chongqing Science and Technology Commission (cstc2015zdcy-ztzz60002)

本文责任编辑 桑农

Recommended by Associate Editor SANG Nong

1. 重庆邮电大学先进制造工程学院 重庆 400065 2. 重庆邮电大学自动化学院 重庆 400065

1. College of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065  
2. College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065

Pollefeys 等<sup>[7]</sup> 在相机焦距不变的对重建物体的周围拍摄多幅图像, 通过特征提取、匹配和多视几何关系等步骤对相机进行标定并重建三维模型. 2009 年, Furukawa 等<sup>[8]</sup> 提出了一种基于面片的多视图立体重建方法, 这种方法的优点是重建出的物体轮廓完整性较好、适应性较强, 而且不需要初始化数据. 此外, 2013 年, 微软研究院推出的 Kinect Fusion 项目<sup>[9]</sup> 在三维重建领域取得了重大突破, 与三维点云拼接不同, 它主要采用一台 Kinect 围绕物体进行连续扫描, 并且实时地进行物体的三维模型重建, 这样做有效地提高了重建精度. 微软研究院 (Microsoft Research) 在 ISMAR 2015 会议上公布了 Mobile Fusion 项目<sup>[10]</sup>, 这个项目使用手机作为一台 3D 扫描仪, 可以拍摄出各种 3D 场景图像. 国内对于三维重建的研究虽然相对落后, 但也取得了不错的成果. 1996 年, 中科院的李利等<sup>[11]</sup> 提出了恢复室内场景的三维重建技术. 2002 年, 中科院的 Zhong 等<sup>[12]</sup> 提出了一种新的匹配方法—半稠密匹配法, 这种方法解决了稀疏匹配重建出的物体信息较少和稠密匹配重建出的点云物体信息较多等问题. 2003 年, 中科院的 Lei 等<sup>[13]</sup> 利用 Kruppa 方程进行相机的自标定, 成功研发出了 CVSuite 软件<sup>[14]</sup>, 该软件实现了利用不同视角的影像进行三维建模. 2014 年, 西安电子科技大学的张涛<sup>[15]</sup> 提出了一种基于单目视觉的三维重建方法, 这种方法利用获取的空间稀疏三维点云, 再使用欧氏重建和射影重建方法, 从而重建出获取图像的真实场景.

近年来, 三维重建技术的研究和应用得到了快速的发展, 但仍然面临着许多问题. 为此, 本文将对近些年来基于视觉的三维重建技术方法的主要进展和部分具有代表性的研究成果进行介绍, 为科研人员提供参考, 并以此为基础, 通过对比和分析, 探究三维重建技术研究中的难点和热点, 以及可能的发展趋势.

在接下来章节中, 本文将从现有技术分析总结和未来发展方向两个方面讨论三维重建关键技术问题, 具体安排如下: 第 1 节总结了三维重建的方法; 第 2 节对各种方法进行了分析, 并比较了各种方法的优缺点、自动化程度、重建效果、实时性以及应用场景; 第 3 节总结了三维重建关键技术以及未来的发展方向, 并总结概括了本文内容.

## 1 三维重建方法

从整体上来看, 三维重建技术主要通过视觉传感器来获取外界的真实信息, 然后, 再通过信息处理技术或者投影模型得到物体的三维信息, 也就是说, 三维重建是一种利用二维投影恢复三维信息的计算机技术<sup>[16-17]</sup>. 1997 年, Várady 等<sup>[18]</sup> 将数据获取

方式分为接触式和非接触式两种. 2005 年, Isgro 等<sup>[19]</sup> 又将非接触式方法分为主动式和被动式两类. 主动式需要向场景中发射结构光源, 然后再通过计算和提取光源在场景中的投影信息来检测目标位置并进行测量. 被动式不使用任何其他能量, 而是通过获取外界光源的反射来进行三维测量.

接触式方法其实就是利用某些仪器能够快速直接测量场景的三维信息<sup>[20]</sup>, 主要包括触发式测量、连续式测量、CMMs (Coordinate measuring machines) 和 RA (Robotics arms) 等. 虽然, 接触式方法有其独特的优点, 但是该方法只能应用于仪器能够接触到测量场景的场合. 而且, 在测量某些加工精密物体表面时, 很可能会划伤被测物体的表面, 造成被测物体某种程度的损坏, 影响其性能. 非接触式方法是在不接触被测量物体的前提下, 利用影像分析模型原理来获取被测物体的数据信息. 虽然, 这种方法的精度并没有接触式高, 但是, 这种方法的应用范围比接触式方法更广泛. 由于接触式测量不属于视觉测量, 因此本文只对非接触式方法进行详细介绍. 非接触式主要包括主动视觉法和被动视觉法; 主动视觉又包括激光扫描法、结构光法、阴影法、TOF 技术、雷达技术、Kinect 技术等; 被动视觉法根据摄像机数目的不同分为单目视觉法、双目视觉法和多目视觉法; 根据原理 (匹配方法) 不同又可以分为区域视觉法、特征视觉法等; 根据应用方法也可以分为运动恢复结构法和机器学习法等. 三维重建技术的分类如图 1 所示.

### 1.1 基于主动视觉的三维重建技术

基于主动视觉的三维重建技术主要包括激光扫描法<sup>[21-22]</sup>、结构光法<sup>[23]</sup>、阴影法<sup>[24]</sup> 和 TOF 技术<sup>[25]</sup>、雷达技术<sup>[26]</sup>、Kinect 技术<sup>[27]</sup> 等. 这些方法主要利用光学仪器对物体表面进行扫描, 然后, 通过分析扫描数据, 重建物体表面的三维结构. 此外, 这些方法还可以获取目标表面的其他一些细节信息, 从而能够精确地重建出目标物的三维结构.

#### 1.1.1 激光扫描法

激光扫描法其实就是利用激光测距仪来进行真实场景的测量. 首先, 激光测距仪发射光束到物体的表面, 然后, 根据接收信号和发送信号的时间差确定物体离激光测距仪的距离, 从而获得测量物体的大小和形状. 该方法的优点是不仅可以建立简单形状物体的三维模型, 还能生成不规则物体的三维模型, 而且生成的模型精度比较高. 激光扫描数据处理流程如图 2 所示, 首先, 通过激光扫描法获取点云数据, 然后与原始获得的数据进行配准获得配准后的点云数据, 最后对获取的点云数据进行一系列的处理, 从而获取目标物的三维模型.

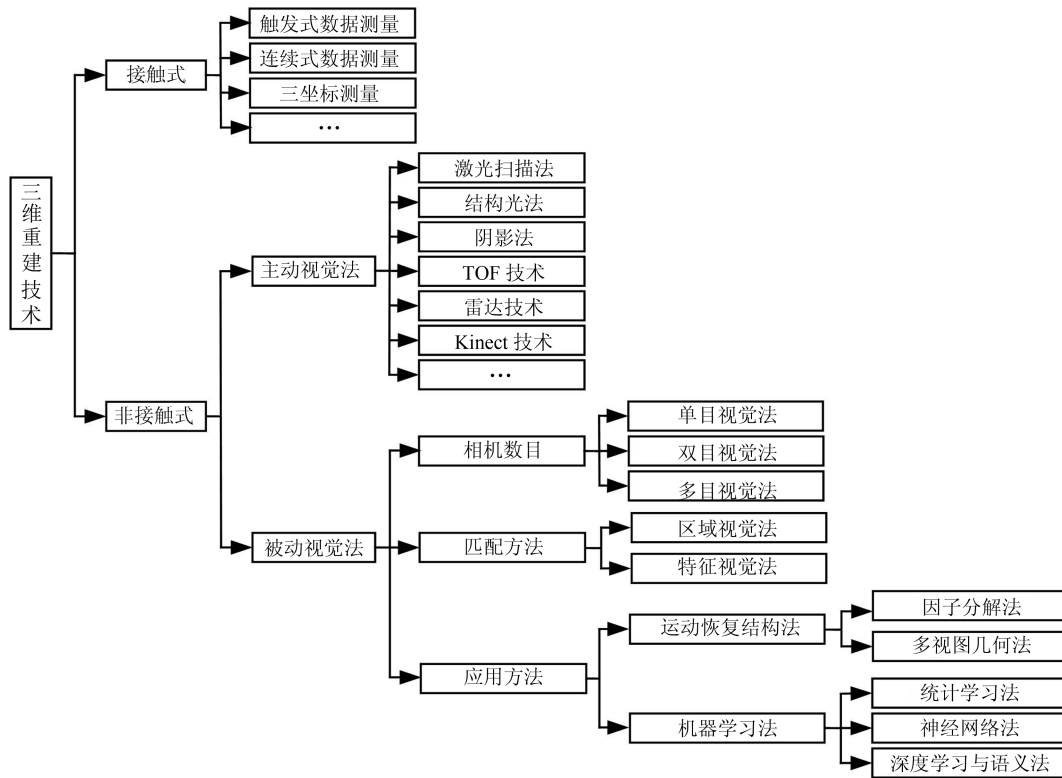


图 1 三维重建技术分类

Fig. 1 Classification of 3D reconstruction technology

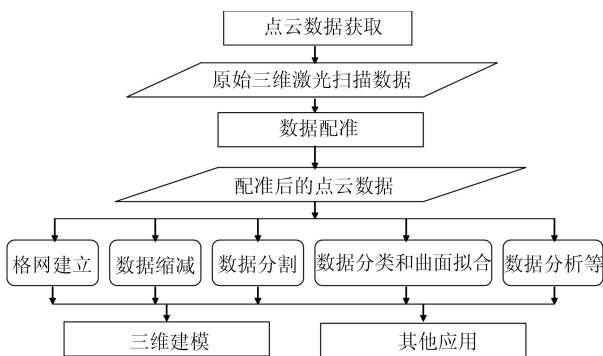


图 2 激光扫描数据处理流程

Fig. 2 The process of laser scanning data processing

20 世纪 60 年代, 欧美一些国家就已经对三维激光扫描技术进行了研究. 在很早以前, 斯坦福大学就已经开展了大规模的地面固定激光扫描系统的研究, 获得了较精确的实验结果. 1999 年, Yang 等<sup>[28]</sup>介绍了三角法激光扫描, 详细地论述了在大型曲面测量原理的基础上影响激光扫描测量精度的几个因素. 2003 年, Boehler 等<sup>[29]</sup>分析并验证了使用不同种类的三维激光扫描仪对实验结果的影响. 更进一步, 2006 年, Reshetyuk<sup>[30]</sup>详细地分析了脉冲式地面激光扫描仪的误差来源以及影响程度, 并对该误

差模型进行了评价. 2007 年, Voisin 等<sup>[31]</sup>研究环境光线对三维激光扫描的影响. 至此, 三维激光扫描仪步入了一个新的里程碑.

### 1.1.2 结构光法

随着科技的不断进步, 三维重建技术涌现出了许多研究方向, 其中结构光法就是三维重建技术的主要研究方向之一<sup>[32]</sup>. 结构光法的原理是首先按照标定准则将投影设备、图像采集设备和待测物体组成一个三维重建系统; 其次, 在测量物体表面和参考平面分别投影具有某种规律的结构光图; 然后再使用视觉传感器进行图像采集, 从而获得待测物体表面以及物体的参考平面的结构光图像投影信息; 最后, 利用三角测量原理、图像处理等技术对获取到的图像数据进行处理, 计算出物体表面的深度信息, 从而实现二维图像到三维图像的转换<sup>[33-36]</sup>. 按照投影图像的不同, 结构光法可分为: 点结构光法、线结构光法、面结构光法、网络结构光和彩色结构光.

基于结构光法的三维重建主要利用光学三角测量原理来计算物体的深度信息. 它主要通过扫描仪中的光源、光感应器和反射点构成的三角关系来计算目标物体的深度信息, 从而实现目标物体的三维重建. 三角测量又可以分为: 单光点测量、单光条测

量和多光条测量. 如图 3 为结构光三角测量原理示意图.

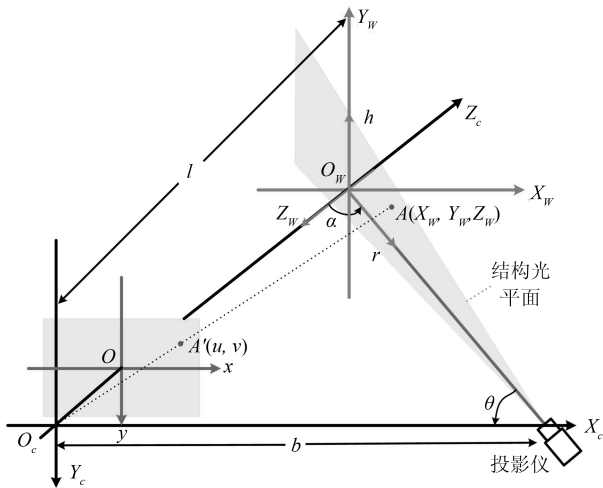


图 3 结构光三角测量原理示意图

Fig. 3 Schematic diagram of the principle of structured light triangulation

如图 3 所示, 假设物体坐标  $(X_w, Y_w, Z_w)$  为世界坐标与被测量的图像坐标  $(u, v)$  以及投影角  $\theta$  之间的关系如下:

$$[X_w, Y_w, Z_w] = \frac{b}{f \cos \theta - u} [u, v, f] \quad (1)$$

自 20 世纪 80 年代以来, 基于结构光法的三维重建越来越受到国外研究人员的关注. 2000 年, Kowarschik 等<sup>[37]</sup> 采用了一种光栅结构法的三维测量系统, 解决了结构光在测量中存在的遮挡问题. 2002 年, Shakhnarovich 等<sup>[38]</sup> 提出了利用多种点结构光投影的光点法进行三维重建. 2004 年, Salvi 等<sup>[39]</sup> 采用结构光条法, 将激光发射的光束直接通过圆柱体透镜, 然后, 再使用步进电机匀速转动圆柱体透镜, 使光束能够完全扫过测量物体的表面, 进而可以获得物体的图像信息并进行信息的提取和三维测量. 国内也在这方面做了大量的研究, 2002 年, 张广军等<sup>[40]</sup> 建立了结构光三维双视觉 RBF (Radial basis function) 神经网络模型, 这种模型的优点是不需要考虑外在因素的影响, 从而使该模型具有较高的精度. 同年, 天津大学首先研制了可以应用于生物医学、工业测量等领域的线结构光轮廓传感器<sup>[41]</sup>. 2004 年, 清华大学研究出了线结构光的多用途传感器, 这种传感器的优点是可以对运动的物体以及腐蚀性的物体进行三维测量和重建, 特别适合于对移动物体和腐蚀性表面的快速、在线、非接触的测量与重建<sup>[42]</sup>.

### 1.1.3 阴影法

阴影法是一种简单、可靠、低功耗的重建物体

三维模型的方法<sup>[43-44]</sup>. 这是一种基于弱结构光的方法, 与传统的结构光法相比, 这种方法要求非常低, 只需要将一台相机面向被灯光照射的物体, 通过移动光源前面的物体来捕获移动的阴影, 再观察阴影的空间位置, 从而重建出物体的三维结构模型. 这种方法的优点是检测速度快、精度高. 阴影法主要分为这几种类型: 平行光的直接阴影法、点光源发光的直接阴影法、微观阴影法、聚焦阴影法、立体和全息阴影法和大型阴影法. 最经典的平行光阴影法如图 4 所示, 该方法使用点光源通过聚焦透镜和针孔, 再利用凹透镜原理使其转换成平行光投影到毛玻璃片上, 其中  $\varepsilon$  表示平行光投影到毛玻璃片上产生的误差.

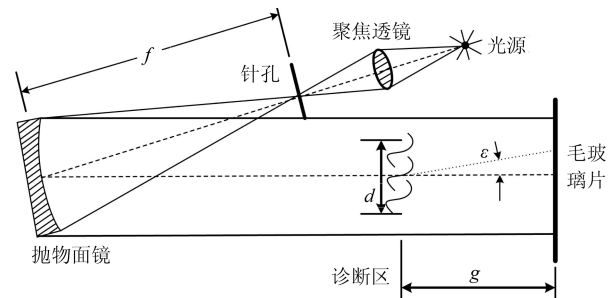


图 4 平行光阴影法

Fig. 4 Parallel photocathode

从国内外的研究来看, 阴影被分为硬阴影和软阴影. 与硬阴影相比, 软阴影要考虑物体之间的几何特征, 更加难以实现, 但是, 显示效果更加真实. 在真实的世界中, 由于光源比较复杂以及物体之间有光照的影响, 形成的阴影明暗程度并不是唯一的, 所以, 使用阴影法实现三维空间的物体重建是非常复杂的过程<sup>[45-48]</sup>, 该方法不适合于实时性较高的三维场景.

### 1.1.4 TOF 技术

TOF (Time of flight) 法是主动测距技术的一种, 可从发射极向物体发射脉冲光, 遇到物体反射后, 接收器收到反射光时停止计时, 由于光和声在空气中的传播速度是不变的, 从而通过发射到接收的时间差来确定物体的距离, 进而确定产生的深度信息, 其原理如式 (2) 所示:

$$d = \frac{n + \frac{\varphi}{2\pi}}{2} \lambda \quad (2)$$

其中,  $\lambda$  表示脉冲的波长;  $n$  表示波长的个数;  $\varphi$  表示脉冲返回时的相位;  $d$  表示物体离发射之间的距离.

TOF 相机的研究相对比较早, 与二维测距仪相比具有较大的优势, 它可以从三维点云中直接获取场景的几何信息. 2014 年, 微软推出了 Kinect 2.0

传感器, 采用 TOF 技术来计算深度, 从而获得三维点云信息. 文献 [49–50] 使用 TOF 相机获取的深度信息提取出场景中的几何信息. 2008 年, May 等<sup>[49]</sup> 使用两帧之间匹配数据中对应的方向向量来提高定位精度. 2009 年, Hedge 等<sup>[50]</sup> 运用提取的方向向量来探测不容易识别的路平面. 同年, Pathak 等<sup>[51]</sup> 利用方向向量建立三维地图, 为移动机器人提供导航信息. 然而, 由于 TOF 相机获取的三维点云信息存在比较多的误差点, 只依靠几何信息来构建地图和定位会产生较大的误差. Stipes 等<sup>[52]</sup> 采用 ICP (Iterative closest point) 算法拼接 TOF 两帧之间的数据, 通过获取的三维点云来实现 ICP 的迭代过程. May 等<sup>[53]</sup> 通过 SLAM 算法解决两帧之间的数据匹配问题.

### 1.1.5 雷达技术

雷达作为一种很常见的主动视觉传感器, 可以通过发射和接收的光束之间的时间差来计算物体的距离、深度等信息. 如式 (3) 所示:

$$d = \frac{c\Delta t}{2} \quad (3)$$

式中,  $c$  为光速;  $\Delta t$  为发射与接受的时间间隔;  $d$  表示雷达到物体之间的距离.

在 20 世纪 60 年代激光雷达传感器迅速发展, 这种传感器通过激光束的扫描, 可以得到周围环境的深度信息. 本部分仅介绍激光雷达的相关应用, 其他雷达不再赘述. 激光雷达的数学模型可以表示为:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \lambda \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} X_S \\ Y_S \\ Z_S \end{pmatrix} \quad (4)$$

其中,  $X, Y, Z$  是空间点的三维坐标;  $a_i, b_i, c_i$  为 3 个空间姿态角组成的方向余弦;  $x, y, z$  为空间扫描点坐标;  $X_S, Y_S, Z_S$  为激光雷达扫描器的直线外方位元素; 通过式 (4) 可以获得物体的空间三维坐标.

2004 年, Streller 等<sup>[54]</sup> 对激光雷达获取的扫描点进行聚类, 从而实现智能车前方目标的检测. 2005 年, Schwalbe 等<sup>[55]</sup> 利用激光雷达获取点云数据, 然后采用线追踪近邻面将点云数据进行分段投影, 最后重建出建筑物的三维模型. 2007 年, Weiss 等<sup>[56]</sup> 使用激光雷达聚类的方法来提取智能车前方车辆的轮廓信息, 然后对目标车辆进行三维重建, 从而获取形状信息, 最后采用模式识别算法, 结合得到的轮廓和形状信息对目标车辆进行检测. 2010 年, 胡明<sup>[57]</sup> 提出了边界保持重建算法, 利用激光雷达获取的点云数据选取二次曲面进行局部拟合, 再使用单元分

解的方法对拟合曲面进行点云拼接, 从而实现了点云的三维重建. 2012 年, 魏征<sup>[58]</sup> 使用车载激光雷达获取建筑物的点云数据进行了几何重建.

### 1.1.6 Kinect 技术

Kinect 传感器是最近几年发展比较迅速的一种消费级的 3D 摄像机, 它是直接利用镭射光散斑测距的方法获取场景的深度信息<sup>[59]</sup>, Kinect 在进行深度信息获取时采用的是第 1.1.2 节所介绍的结构光法, 下面主要是对 Kinect 技术研究现状进行简要概述. 由于 Kinect 价格便宜, 自 2010 年发售以来, 受到了国内外的广泛关注, 并开始使用 Kinect 进行三维重建的研究. Kinect 传感器如图 5 所示.



图 5 Kinect 传感器

Fig. 5 Kinect sensor

Kinect 传感器中间的镜头为摄像机, 左右两端的镜头被称为 3D 深度感应器, 具有追焦的功能, 可以同时获取深度信息、彩色信息、以及其他信息等. Kinect 在使用前需要进行提前标定, 大多数标定都采用张正友标定法<sup>[60]</sup>. 2011 年, Smisek 等<sup>[61]</sup> 为了解决 Kinect 传感器无法找到棋盘格角点问题, 对 Kinect 深度相机自身的红外发射器进行遮挡, 并使用卤素灯生成红外图像, 从而标定 Kinect 传感器两个相机之间的位置. 2014 年, Zollhöfer 等<sup>[62]</sup> 为了解决 Kinect 获取的深度信息含有噪声的问题, 使用高斯滤波器进行滤波处理, 从而减小了噪声影响.

目前, 使用 Kinect 进行三维重建的研究比较流行. 2014 年, Henry 等<sup>[63]</sup> 最早使用 Kinect 相机对室内环境进行三维重建, 得到的效果不是很好, 重建的模型有很多黑色斑点, 实时性也较差, 需要进一步提高其性能. 为了解决这些问题, 2012 年, Henry 等<sup>[64]</sup> 使用了重投影误差的帧间配准、FAST 特征等优化方法对其进行了改进, 实时性得到了显著提高. 2011 年, Newcombe 和 Izadi 等<sup>[65–66]</sup> 开发了 Kinect Fusion 系统, 该系统利用获取的深度信息生成三维点云及法向量, 从而可以对场景进行三维重建, 其结果更加精确. 2013 年, 吴侗<sup>[67]</sup> 采用体密度变化率直方图的方法对点云数据进行分割和检测, 然后, 对于 Kinect 采集到的纹理信息使用卷包裹算法, 从而完成了对点云数据的三维重建. 表 1 所示为主动视觉常用方法优缺点的对比.

表 1 主动视觉方法对比  
Table 1 Active visual method comparison

方 法	激光扫描法 <sup>[28-31]</sup>	结构光法 <sup>[32-42]</sup>	阴影法 <sup>[43-48]</sup>	TOF 技术 <sup>[49-53]</sup>	雷达技术 <sup>[54-58]</sup>	Kinect 技术 <sup>[59-67]</sup>
优点	1. 重建结果很精确; 2. 能建立形状不规则物体的三维模型.	1. 简单方便、无破坏性; 2. 重建结果速率快、精度高、能耗低、抗干扰能力强.	1. 设备简单, 图像直观; 2. 密度均匀, 简单低耗, 对图像的要求非常低.	1. 数据采集频率高; 2. 垂直视场角大; 3. 可以直接提取几何信息.	1. 视场大、扫描距离远、灵敏度高、功耗低; 2. 直接获取深度信息, 不用对内部参数进行标定.	1. 价格便宜、轻便; 2. 受光照条件的影响较小; 3. 同时获取深度图像和彩色图像.
缺点	1. 需要采用算法来修补漏洞; 2. 得到的三维点云数据量非常庞大, 而且还需要对其进行配准, 耗时较长; 3. 价格昂贵.	1. 测量速度慢; 2. 不适用室外场景.	1. 对光照的要求较高, 需要复杂的记录装置; 2. 涉及到大口径的光学部件的消像差设计、加工和调整.	1. 深度测量系统误差大; 2. 灰度图像对比度差、分辨率低; 3. 搜索空间大、效率低; 4. 算法扩展性差, 空间利用率低.	1. 受环境的影响较大; 2. 计算量较大, 实时性较差;	1. 深度图中含有大量的噪声; 2. 对单张图像的重建效果较差.

## 1.2 基于被动视觉的三维重建技术

### 1.2.1 根据相机数目分类

基于被动视觉的三维重建技术是通过视觉传感器(一台或多台相机)获取图像序列,进而进行三维重建的一种技术.这种技术首先通过视觉传感器(一台或多台相机)获取图像序列,然后提取其中有用的信息,最后,对这些信息进行逆向工程的建模,从而重建出物体的三维结构模型.该方法的优点是能够应用于各种复杂的环境中,对主动视觉法具有很好的补足.另外,它具有价格较低,操作简单,实时性较高,对光照要求较低以及对场景没有要求的优点,容易实现;不足的是重建精度不是很高.由于主动视觉方法受环境及设备等因素的限制,近几年,人们投入大量精力用于被动视觉方法的研究上.根据相机数量的不同,被动视觉的三维重建技术可以分为单目视觉、双目视觉和多目视觉,这一部分将重点从相机数目的角度对被动视觉的三维重建技术进行总结和分类.

#### 1.2.1.1 单目视觉法

单目视觉是仅使用一台相机进行三维重建的方法,该方法简单方便、灵活可靠、处理时间相对较短,而且价格便宜,使用范围比较广,能够使用在三维测量和检测等领域.为了进一步表示空间中任意一个三维点  $P$  在世界坐标系转换到二维图像坐标系之间

的关系,关系坐标可以表示为:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (5)$$

其中,  $(X_W, Y_W, Z_W)$  为空间中的三维点;  $(R \ t)$  称为旋转矩阵和平移向量;  $f_x$  和  $f_y$  是摄像机在两个方向上的焦距;  $(u_0, v_0)$  是摄像头主点在图像坐标系下的坐标;  $(u, v)$  是图像坐标系下的坐标; 从而通过式(5)可以求解出任意空间一点的三维坐标.基于单目视觉的三维重建流程如图 6 所示.

单目视觉主要提取图像中的亮度、深度、纹理、轮廓、几何形状、特征点等特征信息.由于这些特征信息已经在文献 [68] 中详细阐述过,为了使相关研究人员以及读者能够更好地了解发展趋势以及能够清楚它们之间在三维重建中的优缺点,这一部分简要的概述图像中所包含的特征信息.

#### 1) 明暗度恢复形状法

明暗度恢复形状法,简称 SFS (Shape from shading),即通过分析图像中的明暗度信息,利用表面的反射模型,获取物体表面的法向信息,从而恢复出物体的三维轮廓,图像在  $(u, v)$  处的像素强度

$I_{uv}$  可以表示为:

$$I_{uv} = R_I(\rho, n, s, v) \quad (6)$$

其中,  $R_I$  表示反射图;  $\rho$  为表面反射率;  $n$  是表面法向量;  $s$  表示入射光方向;  $v$  表示反射光方向.

明暗度恢复形状法的概念最早由 Horn<sup>[69]</sup> 于 1970 年提出. 1989 年, Penna<sup>[70]</sup> 提出了 PSFS (Perspective shape from shading) 方法, 这种方法其实就是用透视投影替代正交投影的明暗度恢复法. 1994 年, Bakshi 等<sup>[71]</sup> 提出了使用非朗伯特模型的明暗度法. 2008 年, Vogel 等<sup>[72]</sup> 综合以上两种方法又提出了基于非朗伯特模型的 PSFS 方法.

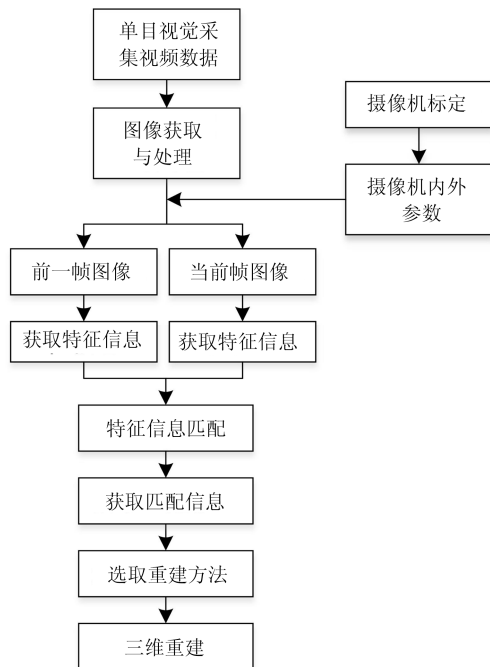


图 6 基于单目视觉的三维重建流程

Fig. 6 3D reconstruction process based on monocular vision

## 2) 光度立体视觉法

虽然 SFS 可以从单幅图像中获取物体的三维信息, 但是其信息量比较少, 而且重建出来的三维模型的效果也不是很好. 于是, Woodham<sup>[73]</sup> 于 1980 年对明暗度恢复形状法的不足进行改进, 提出了光度立体视觉法, 简称 PS (Photometric stereo). 光度立体视觉法首先将单个摄像机固定在目标物体的正上方, 然后通过光源发出的光线从不同的角度射到目标物体的表面, 最后通过摄像机获取多幅图像, 从而得到图像的灰度值与目标物体的关系以此来恢复三维物体的形状. 随后, 许多研究人员在光度立体视觉法的基础上又有了进一步的研究. 2003 年, Noakes 等<sup>[74]</sup> 在光度立体视觉法中提出非线性与噪声减除的方法. 2004 年, Horovitz 等<sup>[75]</sup> 在光度立体视觉

法中引入了控制点和梯度场的概念. 2005 年, Tang 等<sup>[76]</sup> 使用可信度传递与马尔科夫随机场<sup>[77]</sup> 的方法对光度立体视觉法进行了优化. 2007 年, Sun 等<sup>[78]</sup> 采用非朗伯特模型的光度立体视觉法. 2009 年, Vlastic 等<sup>[79]</sup> 提出了使用多视角进行三维重建的方法. 2010 年, Shi 等<sup>[80]</sup> 提出了自标定的光度立体视觉法. Morris 等<sup>[81]</sup> 使用了动态折射立体法对物体表面进行三维重建. Higo<sup>[82]</sup> 提出了对非刚性不规则物体进行三维重建的方法. 这些方法在一定程度上提高了三维重建的精度. 这种方法可以用亮度方程进行表示:

$$I(x, y) = k(x, y) \times N(x, y) \times S \quad (7)$$

其中,  $I$  为图像亮度;  $S$  为光源向量;  $N$  为物体表面的法向量;  $k$  是由物体表面反射系数、光源强度、摄像机对光敏感度共同决定的系数.

光度立体视觉法在不同光照的条件下通过摄像机拍摄多幅图像, 再根据不同图像的亮度方程进行联立, 从而求解出物体表面的法向量, 进而恢复物体的几何形状.

## 3) 纹理法

纹理法简称 SFT (Shape from texture). 这种方法通过分析图像中物体表面的纹理大小和形状, 来获取物体的三维信息, 进而重建出物体的三维模型.

纹理法分为两种, 一种是基于频谱分析的方法, 这种方法主要通过频域变换分析纹理单元的谱信息来恢复物体表面的法向, 利用这些法向重建出物体的三维模型. 1988 年, Brown 等<sup>[83]</sup> 采用傅里叶变换对物体的纹理进行了三维重建. 2002 年, Clerc 等<sup>[84]</sup> 使用小波变换对物体表面进行了纹理分析和三维重建. 另外一种则是在正交投影条件下基于后验概率分布的方法, 这个方法是由 Wiktin<sup>[85]</sup> 于 1981 年最早提出的. 2010 年, Warren 等<sup>[86]</sup> 为了使重建效果有进一步的提高, 采用了透视投影模型对 Wiktin 的方法进行了改进, 通过实验验证了这种方法的可行性.

## 4) 轮廓法

轮廓法简称 SFS/SFC (Shape from silhouettes/contours). 该方法主要是通过一个相机从多个角度拍摄图像来获取物体的轮廓信息, 通过这些轮廓信息恢复物体的三维结构模型. 轮廓法又可以分为体素法<sup>[87]</sup>、视壳法<sup>[88]</sup> 和锥素法<sup>[89-91]</sup> 三种.

采用轮廓进行三维重建是由 Martin 等<sup>[87]</sup> 于 1983 年首次提出的方法, 这种方法首先将物体所在的三维几何空间离散化为体素, 然后再使用正向试探法, 消除投影在轮廓区域以外的体素, 进而可以获得物体的三维信息. 为了进一步研究轮廓法的相关

理论, 1994 年, Laurentini<sup>[88]</sup> 提出了可视壳 (Visual hull) 的概念. 2006 年, Forbes 等<sup>[92]</sup> 实现了从未标定图像中生成三维可视壳的方法. 此外, 文献 [93] 中还提出了一些基于图形硬件的可视壳生成方法.

### 5) 调焦法

调焦法简称 SFF (Shape from focus), 这种方法通过分析相机的光圈、焦距和拍摄图像的清晰度之间的关系获取物体表面的深度信息, 从而重建出物体的三维模型.

由于相机镜头具有光学聚焦的原理, 因此, 当物体经过相机镜头时, 其产生的表面深度信息与相机焦距之间的关系可以确定该物体在图像上的清晰程度. 所以, 该方法又可以分为聚焦法<sup>[94-95]</sup> 和离焦法<sup>[96]</sup> 两种.

### 6) 亮度法

亮度法简称 SFI (Shape from illumination). 这种方法主要是用来分析物体在多个视角下通过相机拍摄来获取物体的多张图像, 然后计算图像中的亮度特征, 通过这些亮度特征恢复出物体表面的深度信息, 利用深度信息从而可以重建出物体的三维模型. 该方法具体又可以分为正向法<sup>[97]</sup> 和逆向法<sup>[98]</sup>.

正向法是将物体三维空间中的采样点投影到二维图像中, 通过判断是否满足亮度一致性来判断该点是否属于物体的表面. 目前已有的正向法有体素颜色法<sup>[99]</sup>、空间雕刻法<sup>[100]</sup> 等.

逆向法是搜索图像中能够满足亮度一致性的匹配点, 然后再使用立体视觉中的三角测量原理, 用这些获取的匹配点来反算其对应的三维点位置.

#### 1.2.1.2 双目视觉法

双目视觉的工作原理来源于人类的双目视觉系统<sup>[101-102]</sup>, 也就是说从不同的视角通过两个相同的相机捕获同一个位置下的左右两侧图像, 然后再利用三角测量原理获取物体的深度信息, 通过这些深度信息重建出物体的三维模型. 目前, 基于双目视觉的三维重建方法是三维重建技术中的热点和难点<sup>[103-104]</sup>.

20 世纪 60 年代, 麻省理工学院的 Roberts 首次将二维图像扩展到三维图像. 此外, MIT 人工智能实验室的 Marr 等<sup>[105]</sup> 提出了视觉相关理论, 这种理论为双目视觉的研究奠定了坚实的基础. 根据两个相机安装位置的不同可分为两种双目视觉系统, 一种称为平行式光轴双目视觉系统<sup>[106]</sup>, 另一种被称为汇聚式光轴双目视觉系统<sup>[107]</sup>. 图 7 显示了这两种系统.

#### 1) 平行式光轴双目视觉系统

平行式光轴双目视觉系统是比较理想的一种系统. 在平行式光轴视觉系统中, 左右相机互相对齐,

它们的光轴也要互相平行, 形成一个共面的成像平面. 由于左右相机只在轴上的位置不同, 而焦距等其他参数是相同的, 因此, 左右相机拍摄的同一物点所成的像分别在左右两图像上对应的对集线上, 可以较好的实现立体匹配. 图 7(a) 所示, 这一成像特点极大地解决了立体匹配问题.

假设空间点  $P$  的三维坐标为  $(X_W, Y_W, Z_W)$ , 对应左右摄像机中的图像坐标分别为  $p_1(x_1, y_1), p_2(x_2, y_2)$ , 由此可以得到式 (8):

$$\begin{cases} X_W = \frac{\Delta x(x_1 - u_0)}{x_1 - x_2} \\ Y_W = \frac{\Delta x a_x (y_1 - v_0)}{a_y (x_1 - x_2)} \\ Z_W = \frac{\Delta x a_x}{x_1 - x_2} \end{cases} \quad (8)$$

其中,  $a_x, a_y, u_0, v_0$  为摄像机的内部参数;  $x_1 - x_2$  称为视差;  $\Delta x$  为两台摄像机光心之间的距离. 因此, 在得到摄像机的内部参数以后, 只要再知道空间中任意一点投影到左右摄像机平面中的图像坐标, 然后利用视差图就可以恢复空间中任意一点的三维坐标.

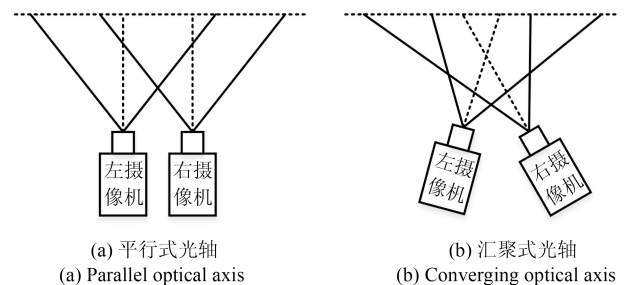


图 7 双目视觉系统

Fig. 7 Binocular vision system

#### 2) 汇聚式光轴双目视觉系统

汇聚式光轴双目视觉系统是将平行式光轴双目视觉系统中的左右相机分别绕光心顺时针和逆时针旋转一定角度, 从而形成汇聚式双目视觉系统. 此系统的优点是能够获得更大的视场, 图 7(b) 所示. 大视场的好处就是能够提高计算视差的精度, 从而可以提高三维重建的精度. 但是, 在匹配过程中视差太大会产生误匹配点增多、匹配范围扩大等问题, 需要更加复杂的算法来处理这些问题, 并且实时性会降低. 最重要的是, 当相机摆放的角度太大会造成立体匹配相当困难, 进而不能精确地重建出物体的三维模型.

同样假设空间点  $P$  的三维坐标为  $(X_W, Y_W, Z_W)$ , 左摄像机的坐标系为  $O_1 - X_1 Y_1 Z_1$ , 图像坐标系为  $o_1 - x_1 y_1$ ; 右摄像机的坐标系为  $O_2 - X_2 Y_2 Z_2$ , 图像坐标系为  $o_2 - x_2 y_2$ ; 原点  $O_1, O_2$  分别为左右摄像机的光心; 左右摄像机的焦距分别



设为  $f_1, f_2$ . 如图 8 所示为汇聚式双目视觉理论模型.

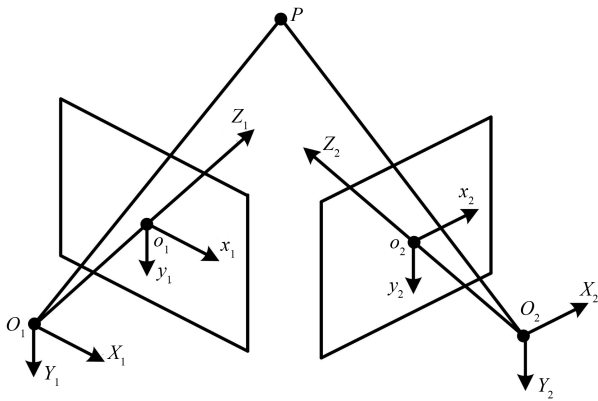


图 8 汇聚式双目视觉理论模型

Fig. 8 Convergent binocular vision theory model

根据成像模型可以得出汇聚式双目视觉模型为:

$$\begin{cases} X_W = \frac{Z_W x_1}{f_1} \\ Y_W = \frac{Z_W y_1}{f_1} \\ Z_W = \frac{f_1(f_2 t_x - x_2 t_z)}{x_2(r_7 x_1 + r_8 y_1 + r_9 f_1) - f_2(r_4 x_1 + r_2 y_1 + r_3 f_1)} \end{cases} \quad (9)$$

或者:

$$\begin{cases} X_W = \frac{Z_W x_1}{f_1} \\ Y_W = \frac{Z_W y_1}{f_1} \\ Z_W = \frac{f_1(f_2 t_x - y_2 t_z)}{y_2(r_7 x_1 + r_8 y_1 + r_9 f_1) - f_2(r_4 x_1 + r_5 y_1 + r_6 f_1)} \end{cases} \quad (10)$$

其中,  $r_1, \dots, r_9$  为旋转分量;  $t_x, t_y, t_z$  为平移分量. 式 (9) 和式 (10) 为汇聚式双目视觉理论模型中三维坐标的解. 因此, 可以通过汇聚式双目视觉理论模型求出空间中任意点的三维坐标.

双目视觉方法大都是利用对极几何将问题转换到欧氏几何条件下, 然后, 再采用三角测量原理估计其深度信息, 该方法大致可以分为 5 个步骤, 分别是图像获取、摄像机标定、图像校正、立体匹配和三维重建<sup>[108-109]</sup>, 图 9 为双目视觉三维重建系统组成.

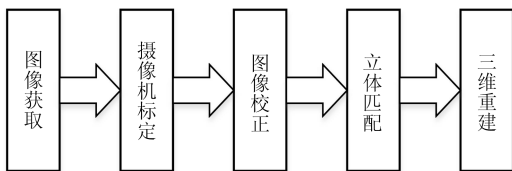


图 9 双目视觉三维重建系统组成

Fig. 9 The composition of the binocular vision 3D reconstruction system

1) 图像获取: 双目视觉方法中的图像获取是采用两台相同的相机同时获取同一个场景中的目标物

体图像, 对于相机的安装位置并没有什么要求. 为了使获取的结果容易计算, 一般情况下使左右相机能够平行放置, 并且要保证基线不能太长, 这样有利于降低图像配准时间以及运算量. 随后, 对获取的图像需要进行预处理, 有利于后面的立体匹配运算.

2) 摄像机标定<sup>[60]</sup>: 双目视觉方法中的摄像机标定主要是为了获取摄像机的畸变向量以及相机的内外参数等信息. 对于获取的畸变向量, 需要消除它的径向和切线方向上的镜头畸变, 以获得无畸变图像, 同时建立摄像机的成像模型, 进而可以确定目标点与像素点之间的对应关系, 为后面计算本征矩阵做准备.

3) 图像校正<sup>[110]</sup>: 双目视觉方法中的图像校正主要通过线性变换使处理后的无畸变图像中的共轭极线位于同一水平线上, 从而把二维空间的匹配问题化简为一维空间的求解问题, 以提高后续立体匹配的鲁棒性和效率<sup>[111]</sup>.

4) 立体匹配<sup>[112]</sup>: 双目视觉方法中的立体匹配是采用立体匹配算法得到校准后的图像与原图像的视差值, 然后利用这个视差值得到每两幅图像之间的稀疏匹配, 再通过优化算法, 获得稠密匹配. 立体匹配是三维重建中最关键的一步, 匹配问题的好坏决定着三维重建的效果和精度.

5) 三维重建: 双目视觉方法中的三维重建是采用三角测量原理计算获取的立体匹配图像的深度值, 从而可以得到稠密的三维空间点云, 随后, 再对获取的三维空间点云进行网格化和差值计算, 进而可以得到物体的三维结构模型. 基于双目视觉的三维重建获取深度信息流程如图 10 所示.

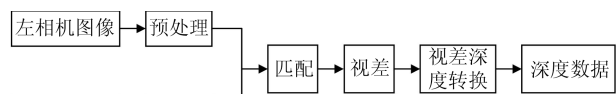


图 10 双目视觉获取深度信息流程

Fig. 10 Process of access to depth information by binocular vision

### 1.2.1.3 多目视觉法

20 世纪 60 年代中期, MIT 的 Roberts 对二维图像的研究开始转到对三维目标物的研究上, 标志着多目视觉技术的出现. 在随后的几十年中, 多目视觉技术的快速发展已经形成从图像获取到可视表面重建的一门新的研究领域.

多目视觉是双目视觉的一种延伸, 它是在双目视觉的基础上, 增加一台或者多台摄像机作为辅助进行测量<sup>[111]</sup>, 从而获得不同角度下同一物体的多对图像. 多目视觉法大多数的理论与双目视觉法是相同的<sup>[113]</sup>, 唯一不同的是, 多目视觉采用了三个或三

个以上的摄像头进行环境中目标物体的获取。多目视觉的优点是当测量物体的表面倾斜的角度太大导致其中的一个或两个 CCD 摄像机不能接收到漫反射光时, 其他的摄像机可继续工作<sup>[114]</sup>。

多目视觉研究最为关键的问题是依据人眼视差的图像匹配问题。目前, 许多学者在多目视觉中对相机标定以及立体图像对的匹配算法进行了大量研究, 形成了一系列相对来说比较完善的理论。2000 年, 中科院的雷成等<sup>[14]</sup> 研发出了 CVSuite 软件。该软件可以实现相机的自标定、特征点提取和匹配以及模型的三维可视化等功能。2002 年, 天津大学的陈明舟<sup>[115]</sup> 采用主动光栅投影的立体视觉方法来获得被测自由曲面的外形轮廓, 这种方法获得了较高的精度。之后, 清华大学的王磊<sup>[116]</sup> 也提出了一种以平面投影变换和遗传算法为基础的匹配算法。

多目视觉法不需要人为地对相关辐射源进行设置, 能够在不接触的情况下进行自动在线检测。这种方法的优点是可以减少测量中的盲区, 获得更大的视野范围, 更高的识别精度; 此外, 该方法还能解决双目视觉中的误匹配现象, 能够适应各种场景。但由于在双目的基础上增加了一台或多台相机, 在进行三维重建时需要大量的数据进行处理, 导致处理时间的增加, 无法满足实时性的要求; 此外, 在重建过程中需要进行大量的匹配运算, 运算量偏大, 而且易受环境光照条件的影响; 另外, 在基线距离比较大时, 重建效果会降低。目前, 多目视觉法在车辆自动驾驶、机器人视觉、多自由度机械装置控制等很多领域获得了广泛的应用<sup>[117-119]</sup>。表 2 对单目、双目和多目视觉方法的优缺点进行了对比。

### 1.2.2 根据匹配方法分类

三维环境重建技术一直是机器视觉和数字图像处理领域的重点研究对象, 众多学者针对三维环境重建提出了很多不同的算法。而图像特征信息匹配的质量在三维环境重建过程中起着十分关键的作用。图像特征信息的匹配首先提取待重建图像的匹配信息, 并使用相应的算法在提取出的匹配信息集中寻找最佳匹配集, 根据最佳匹配集求解变换模型。根据匹配的方法不同可以分为区域视觉法和特征视觉法。

#### 1.2.2.1 区域视觉法

区域视觉法就是基于区域立体匹配算法的三维重建技术。该算法利用对极几何约束和连续性, 提高了稠密匹配的效率和三维重建的质量。1986 年, Goshtasby 等<sup>[120]</sup> 提出了一个由粗到细的分割方法, 该方法减小了图像分割的质量对区域检测的影响。1994 年, Flusser 等<sup>[121]</sup> 首先对获取的图像进行分割并提取封闭区域, 然后再用质心表示其封闭区域, 最后, 获取具有旋转不变特性的区域。进一步的, 2003

年, Alhichri 等<sup>[122]</sup> 提出一种称为虚圆的图像特征。为了提取这种特征, 需要对图像的边缘进行检测, 从而获取其二值化的边缘图像, 然后再使用欧氏距离变换提取图像特征。为了使检测算法具有仿射不变性, 1997 年, Schmid 等<sup>[123]</sup> 提出局部灰度值不变量。2004 年, Matas 等<sup>[124]</sup> 又提出一种仿射不变区域检测算法。同年, Tuytelaars 等<sup>[125]</sup> 提出了两种具有仿射变换不变性的特征区域的方法。与此同时, Kadir 等<sup>[126]</sup> 也提出了一种显著区域的方法, 遗憾的是, 这种方法计算量比较大, 并且检测区域的重复率也不如其他检测算法。

#### 1.2.2.2 特征视觉法

基于特征视觉的三维重建技术其实就是通过相机获取二维图像, 然后提取图像中的角点作为特征点, 以双目立体视觉理论为基础, 利用匹配算法, 获得特征点匹配对, 再通过三角测量原理获取深度值, 从而获得物体表面的三维模型。这种方法的优点是很好地解决了立体匹配的精度和速度, 从而能够较精确地对物体实现三维重建。

1988 年, Harris 等<sup>[127]</sup> 首先提出具有旋转不变特性的角点检测算法, 这种算法是对 Moravec<sup>[128]</sup> 提出的检测算法的改进。由于该算法在结果检测中存在许多的误检测点, 为此 Schmid 等<sup>[129]</sup> 和 Weijer 等<sup>[130]</sup> 又在该算法的基础上进行了改进, 减少了误检测点的数量。与此同时, 为了适应图像分辨率的变化, 2004 年, Mikolajczyk 等<sup>[131]</sup> 又在算法的基础上加入了尺度参数, 将 Harris 角点算法扩展为 Harris-Laplace 算法, 这种算法具有尺度不变性。在此基础上又提出了具有仿射不变的 Harris-Affine 算法。这种算法的优点是在较大的仿射变形情况下仍能保持不变性。

为了使角点的检测效率更高, 速度更快, 1997 年, Smith 等<sup>[132]</sup> 提出 SUSAN (Smallest univalue segment assimilating nucleus) 角点检测算法。随后, Lindeberg 等<sup>[133-134]</sup> 提出了一种基于尺度归一化微分算子的检测方法, 这种方法的优点是对于提取的特征点的尺度和形状使其能够满足仿射不变性。Baumberg<sup>[135]</sup> 提出了一种新的迭代算法用来调整特征点的形状, 使其能够满足图像的局部结构。Lowe<sup>[136]</sup> 最先提出了 SIFT 算子, 该算子具有旋转不变性、尺度不变性以及光照不变性等优点, 是目前最稳定的特征算子, 但是, 它的缺点也很明显, 运算复杂, 处理速度慢。2004 年, Ke 等<sup>[137]</sup> 在传统的 SIFT 特征算子的基础上, 提出了基于主成分分析 (Principal component analysis, PCA) 的 SIFT 特征描述算子方法, 该方法降低了特征描述子的维度, 提高了匹配的精度。2006 年, Bay 等<sup>[138]</sup> 提出了

表 2 单目、双目和多目视觉方法对比

Table 2 Comparison of monocular, binocular and multiocular vision methods

	单目视觉 <sup>[68]</sup>	双目视觉 <sup>[101-110, 112]</sup>	多目视觉 <sup>[111, 113-119]</sup>
优点	1. 简单方便、灵活可靠、使用范围广; 2. 可以实现重建过程中的摄像机自标定, 处理时间短; 3. 价格便宜.	1. 方法成熟; 2. 能够稳定地获得较好的重建效果; 3. 应用广泛.	1. 避免双目视觉方法中难以解决的假目标、边缘模糊及误匹配等问题; 2. 在多种条件下进行非接触、自动、在线的测量和检测; 3. 简单方便、重建效果更好, 能够适应各种场景;
缺点	1. 不能够得到深度信息, 重建效果较差; 2. 重建速度较慢.	1. 运算量大; 2. 基线距离较大时重建效果降低; 3. 价格较贵.	1. 设备结构复杂, 成本更高, 控制上难以实现; 2. 实时性较低, 易受光照的影响.

SURF 算法, 这个算法是 SIFT 算子的一种改进, 解决了 SIFT 算子计算速度慢的缺点. 2014 年, 葛盼盼等<sup>[139]</sup> 在研究 Harris 算法和 SURF 算法各自的特点后, 提出了 Harris-SURF 算法, 这种方法结合了 Harris 算法定位精度高以及 SURF 算法的旋转不变性、尺度不变性以及光照不变性等两者的优势, 提高了三维重建的精度, 使重建的效果更好.

### 1.2.3 根据应用方法分类

基于被动视觉的三维重建技术根据所选取方法不同, 所重建的效果有明显差别, 但每种方法都有不同的优点和缺点. 因此, 根据应用方法可以分为运动恢复结构法和机器学习法.

#### 1.2.3.1 运动恢复结构法

如果获取的图像是从多个视点捕获的多张图像, 可以通过匹配算法获得图像中相同像素点的对应关系, 再利用匹配约束关系, 结合三角测量原理, 获得空间点的三维坐标信息, 进而重建出物体的三维模型. 这个过程被称为运动恢复结构法, 即 SfM (Structure from motion).

运动恢复结构法是通过三角测量原理来恢复场景的三维结构. 这种方法不仅是三维重建的一种重要手段, 而且也是一种结构测量的方法, 能够广泛地应用在测绘、军事侦查等领域. 目前 SfM 主要分为两类: 增量式 SfM (Incremental structure from motion)<sup>[140]</sup> 和全局式 SfM (Global structure from motion)<sup>[141]</sup>.

增量式 SfM 是最早提出的 SfM 方法, 首先通过相机获取多幅图像, 选取其中一对作为“种子”, 其次求解出相机参数并恢复部分三维点, 再次加入相关图像进行重加, 最后使用 BA (Bundle adjustment) 进行优化, 从而实现对图像重建的过程.

全局式 SfM 首先使用旋转一致性求解相机全局

旋转, 然后计算相机的位移, 最后通过 BA 进行优化提升重建质量. 全局式是由 Sturm 等<sup>[142]</sup> 在 1996 年提出的. 随后, 在 2011 年, Crandall 等<sup>[143]</sup> 提出了一种基于马尔科夫随机场的全局式方法. 同年, Irschara 等<sup>[144]</sup> 引入 GPS 坐标作为相机坐标, 并使用全局式 SfM 进行三维重建. 2015 年, Cui 等<sup>[141]</sup> 在 Crandall 等人提出的方法基础上提出了一种全新的全局式 SfM 理论, 主要通过辅助信息对求解过程进行优化, 使其具备处理多种数据的能力.

目前, 运动恢复结构法主要有因子分解法和多视图几何法两种.

#### 1) 因子分解法

因子分解法是 Tomasi 等<sup>[145]</sup> 在 1992 年最早提出的. 该方法主要是把摄像机近似为正投影模型, 再通过奇异值分解 SVD (Singular value decomposition) 对获取的观测矩阵进行处理, 从而获得结构矩阵和运动矩阵. 但是, 这种方法的不足是忽略了大部分相机的真实投影模型. Poelman 等<sup>[146]</sup> 将因子分解法扩展到弱透视投影和平行透视投影模型中, 但得到的效果仍不是很好. Triggs<sup>[147]</sup> 又进一步将因子分解法推广到一般的透视投影模型中, 这种模型需要首先恢复射影的深度, 然后再通过因子分解法恢复其结构, 该模型得到的效果要比以前研究者所使用的方法进步了很多, 但其精度仍不是很高. 然而, Han 等<sup>[148]</sup> 于 2001 年提出了在透视投影模型下考虑相机投影矩阵约束的因子分解法, 该方法取得了不错的效果.

因子分解法的优点是简便灵活, 对周围的环境并没有什么特别的要求, 也不依赖于任何一种具体的模型, 抗噪能力相对来说也比较强, 唯一的缺点就是精度不高.

#### 2) 多视图几何法

基于多视图几何的三维重建在计算机视觉领域

中一直都是最热门的研究问题之一<sup>[149]</sup>. 多视图几何法是描述同一场景从不同视角下获取的多幅图像与物体之间投影关系的几何模型, 该方法主要应用于单相机多视角的三维重建中. 近十几年来, 多视图几何理论把在数学中的推导引入到计算机视觉中, 主要包括射影变换、小孔成像以及捆绑调整等的推导, 大大拓展了计算机视觉理论的研究和应用, 为图像的三维重建提供了理论依据. 通常, 多视图几何法研究内容主要包括不同层次的射影变换(两层摄影变换、三层射影变换等)、摄像机模型中的内外参数的求取、三维坐标计算等, 该方法还涉及一些三维重建中基本的数学公式表达, 例如, 基础矩阵、本质矩阵和  $N$  视点关系等<sup>[17]</sup>. 多视图几何法主要解决运动恢复结构法中的一些问题. 文献 [150–151] 对多视图几何理论做出了很多贡献, 解决了一些匹配以及视点对应关系等问题, 为三维重建提供了更多的理论基础.

通常多视图几何法主要包括以下 4 个步骤:

#### 1) 特征提取与匹配

特征提取与匹配首先利用局部不变性对图像特征信息进行检测, 然后再用描述算子来提取特征点, 最后对同一场景中表示同一空间位置的特征点进行匹配, 从而计算出正确的匹配特征点对. 特征提取与匹配是计算机视觉的研究基础, 而图像特征信息匹配的质量在三维环境重建过程中也起着十分重要的作用, 被广泛应用于图像检索、图像拼接、目标识别与跟踪及三维重建等领域<sup>[152–154]</sup>.

#### 2) 多视图几何约束关系计算

多视图几何约束关系计算就是通过对极几何将几何约束关系转换为本质矩阵的模型参数估计的过程. 1981 年, Longuet-Higgins<sup>[155]</sup> 最早提出多视图间的几何约束关系, 这种关系在欧氏几何中可以用本质矩阵表示. 1990 年, Faugeras 等<sup>[156]</sup> 详细研究了本质矩阵的性质. 1993 年, Luong 等<sup>[157]</sup> 提出了解决两幅图像之间几何关系的基础矩阵, 并阐述了对极几何在射影空间下的基本性质. 1996 年, Luong 等<sup>[158]</sup> 将基础矩阵的概念进行推广, 开启了多视图几何学的新篇章.

#### 3) 优化估计结果

当得到初始重建结果后, 为了均匀化误差和获得更精确的结果, 通常需要对初始结果进行非线性优化. 在 SfM 中对误差应用最精确的非线性优化方法就是捆绑调整法, 该方法是大规模三维重建中最为关键的一步, 也是 SfM 的核心问题<sup>[159]</sup>, 它主要利用代价函数对多台相机模型的内外参数以及获取的三维点云数据同时进行优化调整.

2009 年, Lourakis 等<sup>[160]</sup> 提出了稀疏捆绑调整法 (Sparse bundle adjustment, SBA), 它利用增量

标准方程的稀疏结构来提高解方程的效率. 然而, 随着获取的三维点云数量的增加, 捆绑调整方法需要大量的计算时间, 为了解决这种问题, Choudhary 等<sup>[161]</sup> 结合 CPU 和 GPU 加速运算的方式使用在 BA 算法上, 这种方法有利于加速捆绑调整. 2012 年, Hu 等<sup>[162]</sup> 对大规模三维点云的优化进行了捆绑调整改进, 可以在 SfM 中使用高分辨率图像, 以及进行稠密三维重建<sup>[163]</sup>. 目前, 捆绑调整法已经成为基于特征的多视图三维重建、结构和运动估计算法最后必须采用的步骤.

#### 4) 稠密场景描述

多视图几何法经过射影重建和度量重建后会生成一个稀疏的三维结构模型, 这种稀疏的三维结构模型已经可以应用于要求精度不高的场景, 但其不具有可视化效果. 因此, 为了获得应用范围更广的三维模型, 要进行表面稠密估计, 恢复稠密的三维点云结构模型.

### 1.2.3.2 机器学习法

机器学习其实就是使机器具有学习的能力, 从而不断获得新知识以及新技能得到有效提升. 机器学习在三维环境重建中一直是重点研究对象, 因此, 根据机器学习可以分为常用的三种方法, 分别是统计学习法、神经网络法和深度学习与语义法.

#### 1) 统计学习法

统计学习法就是需要通过不断地学习再学习的过程. 该方法是以大型数据库为基础, 例如, 人脸数据库、场景数据库等. 首先, 这种方法需要对数据库中的每一个目标进行特征统计, 这些特征主要包括亮度、纹理、几何形状、深度等, 然后, 再对重建目标的各种特征建立概率函数, 最后, 计算重建目标与数据库中相似目标的概率大小, 取概率最大的目标深度为重建目标的深度, 再使用差值计算和纹理映射进行目标的三维重建. 目前, 常用的概率模型有马尔科夫模型 (Markov model, MM)<sup>[164]</sup>、隐马尔科夫模型 (Hidden Markov model, HMMs)<sup>[165]</sup> 和 PHMMs 模型 (Pseudo 2D hidden Markov models)<sup>[166]</sup> 等. 统计学习法可以对大型场景<sup>[167–168]</sup>、人脸<sup>[169–171]</sup> 及人体<sup>[172–173]</sup> 进行重建, 并可以应用到视频检索和识别系统等其他领域.

#### 2) 神经网络法

基于神经网络法的三维重建是利用神经网络具有较好的泛函逼近能力, 能够以任意的精度逼近任何非线性关系的优点来进行三维重建<sup>[174]</sup>. 如, Do 等<sup>[175]</sup> 提出了两种利用神经网络进行三维重建的方法. 第一种方法是首先通过线性关系进行三维重建, 然后再使用神经网络的优点来拟合重建的误差, 从而提高三维重建的精度. 第二种方法是先利用神经

网络拟合图像的像差, 从而获得理想的像点, 然后再利用线性关系进行三维重建. 不足的是这两种方法进行三维重建的过程比较复杂, 需要进行相机的标定, 来获取相机的参数. 为了使三维重建的过程简单, 而且更能符合人的双眼认知机理, 袁野等<sup>[176]</sup>提出了一种基于 BP 网络的三维重建算法, 如图 11 所示.

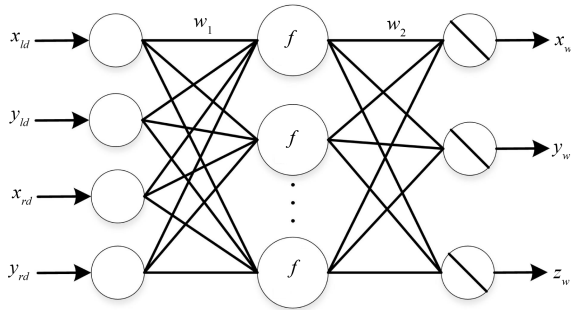


图 11 基于 BP 网络结构的三维重建

Fig. 11 3D reconstruction based on BP network structure

图 11 中  $(x_{ld}, y_{ld}), (x_{rd}, y_{rd})$  分别为左图像和右图像点坐标, 把它们作为输入样本,  $(x_w, y_w, z_w)$  为目标物体的空间三维坐标作为输出样本, 然后对网络进行训练, 把训练后的立体视觉成像几何模型中的非线性映射关系存储在神经网络所建立的连接权值和阈值中. 这种算法不需要复杂的标定过程, 用神经网络建立隐式视觉模型获得像平面到视场之间的映射关系, 从而提高了三维重建过程的复杂度. 该方法在文献 [177] 得到了进一步的应用.

### 3) 深度学习与语义法

基于深度学习的三维重建最近几年取得了非常大进展, 是当前计算机视觉领域比较流行的方法之一. 学习系统是由多层卷积层和多层全连接层组成的卷积神经网络 (CNN) 进行学习图像的层次化特征表示, 这种方法能够精确地重建出物体的几何形状. 与传统的三维重建相比较, 基于深度学习的三维重建技术不需要进行特征提取与匹配以及复杂的几何运算, 使用深度学习的方法更加方便, 使得实时性进一步提高.

基于语义的三维重建可以运用在移动的行人或车辆等大的场景, 这种方法能够精确地对环境中的目标物体进行识别, 而深度学习技术也是最近几年刚刚兴起的比较有优势的识别方法, 因此, 深度学习和语义相结合的三维重建是未来几年的研究趋势, 也会受到该领域的研究者们广泛关注.

2015 年, Savinov 等<sup>[178]</sup>提出了一个稠密的语义三维重建方法. 2016 年, Bláha 等<sup>[179]</sup>提出了一个自适应多分辨率的语义三维重建方法. 该方法的目的是在给定一组场景图像中进行 3D 场景稠密重

建, 并分割成语义目标类. 随后, Sünderhauf 等<sup>[180]</sup>提出了面向对象的语义建图方法, 该方法的主要目的是构建环境的 3D 点云地图, 然后再通过卷积神经网络对关键帧图像中的物体进行识别检测, 最后对 3D 目标物体进行点云分割, 从而在地图中更新或添加目标物体的 3D 点云等信息<sup>[181]</sup>.

## 2 对比与分析

随着科技的发展, 以上所述的研究方向取得了不同程度的进展, 国内外的研究人员也开始关注三维重建的效果、稳定性以及实时性等问题, 相信以后三维重建的相关研究会越来越好. 表 3 给出了各种基于视觉的三维重建技术的对比与分析.

从表 3 中可以看出, 主动视觉法重建的三维模型精度要比被动视觉法高, 但是, 只能应用在某些场景中, 不能被广泛应用. 基于被动视觉的三维重建方法种类较多, 其中基于双目和多目视觉方法在实际应用中的使用多于单目视觉方法, 然而, 在研究价值方面, 似乎单目视觉的研究价值更高一些. 而双目和多目视觉方法需要使用多台相机才能完成相关实验, 成本大大地增加, 并且相机在安装时也要确定好它们之间的位置, 而且相机在拍摄图像时能够同时保证稳定性和同步性, 控制上难以实现. 此外, 双目和多目视觉方法应用场景相对来说较小, 不易于推广使用. 被动视觉中的单目视觉方法简单、可靠、灵活、使用范围广, 克服了立体视觉中视场小、立体匹配难的不足. 单目视觉分类方法体系庞大, 分类众多, 但是必须要满足一些假设条件, 而且容易受到光照、纹理等其他一些因素的影响, 通用性较差, 重建效果不太好. 被动视觉中的双目视觉方法能够较好地获取重建效果, 但是, 必须要满足基线的范围才能重建出更好的效果. 被动视觉中的多目视觉方法视野范围大, 识别精度较高, 重建效果稳定, 适应性较强, 能够应用在各种场景中, 不足的是设备复杂, 价格昂贵, 计算量较大, 控制上难以实现. 运动恢复结构法对图像的要求较低, 通用性非常好, 鲁棒性较强, 一般来说其重建效果依赖于特征点的密集程度, 特征点越密集重建效果就越好, 但相应的运算量也大大增加. 统计学习法重建质量和效率都非常高, 但其适用范围较窄, 重建效果取决于数据库的完备程度, 数据库越完备重建越精确. 其他视觉中的区域视觉法计算简单, 匹配精度较高, 但是, 对图像的要求较高, 受光线的干扰较大. 其他视觉中的特征视觉法抗干扰能力强, 鲁棒性较高, 应用范围广, 但是, 不能够对图像信息进行全面描述. 其他视觉中的神经网络法重建效果较好, 具有很强的鲁棒性, 应用范围较广, 不足的是, 在重构之前需要进行训练, 收敛速度

表 3 基于视觉的三维重建技术对比与分析  
Table 3 Comparison and analysis of 3D reconstruction based on vision

方法	优点	缺点	自动化程度	重建效果	实时性	应用场景
接触式方法 <sup>[18]</sup>	快速直接测量物体的三维信息; 重建结果精度比较高	必须接触测量物体, 测量时物体表面容易被划伤	难以实现自动化重建	重建质量效果较好	实时	不能被广泛的应用, 只能应用到测量仪器能接触到的场景
激光扫描法 <sup>[28-31]</sup>	重建的模型很精确; 重建形状不规则物体的三维模型	形成的三维点云数据量非常庞大, 不容易处理; 重建的三维模型会产生漏洞; 设备比较复杂, 价格非常昂贵	一定程度的自动化重建	重建的三维模型很好	实时	目前主要应用在工厂的生产和检测中, 无法被广泛使用
结构光法 <sup>[32-42]</sup>	仅需要一幅图像就能获得物体形状; 简单方便; 无破坏性	重建速度较慢	一定程度的自动化重建	重建效果的精度比较高	实时	适用于室内场景
阴影法 <sup>[43-48]</sup>	设备简单低耗; 对图像的要求非常低	对光源有一定的要求	自动化重建较低	重建效果较差, 重建过程比较复杂	实时	无法被广泛使用
TOF技术 <sup>[49-53]</sup>	数据采集频率高; 垂直视场角大; 可以直接提取几何信息	深度测量系统误差大; 灰度图像对比度差、分辨率低; 搜索空间大、效率低; 算法扩展性差, 空间利用率低	一定程度的自动化重建	重建效果的精度较低	实时	能够广泛应用在人脸检测、车辆安全等方面
雷达技术 <sup>[54-58]</sup>	视场大、扫描距离远、灵敏度高、功耗低; 直接获取深度信息, 不用对内部参数进行标定	受环境的影响较大; 计算量较大, 实时性较差; 价格较贵	一定程度的自动化重建	重建效果一般	实时	能够广泛应用于各行各业
Kinect技术 <sup>[59-67]</sup>	价格便宜、轻便; 受光照条件的影响较小; 同时获取深度图像和彩色图像	深度图中含有大量的噪声; 对单张图像的重建效果较差	一定程度的自动化重建	重建效果较好	实时	能够被广泛应用于室内场景
明暗度法 <sup>[69-72]</sup>	重建结果比较精确应用范围广泛	易受光源影响; 依赖数学运算; 鲁棒性较差	完全自动化重建	在光源比较差的情况下重建效果较差	非实时	难以应用于镜面物体以及室外场景物体的三维重建
光度立体视觉法 <sup>[73-82]</sup>	避免了明暗度法存在的一些问题; 重建精度较高	易受光源影响; 鲁棒性较差	一定程度的自动化重建	重建效果较好	非实时	难以应用于镜面物体以及室外场景物体的三维重建
纹理法 <sup>[83-86]</sup>	对光照和噪声都不敏感; 重建精度较高	通用性较低; 速度较快; 鲁棒性较好	完全自动化重建	重建效果的精度较高	非实时	只适用于具有规则纹理的物体
轮廓法 <sup>[87-93]</sup>	重建效率非常高; 复杂度较低	对输入信息的要求很苛刻; 无法对物体表面的空洞和凹陷部分进行重建	完全自动化重建	重建效果取决于轮廓图像数量, 轮廓图像越多重建越精确	非实时	通常应用于对模型细节精度要求不是很高的三维重建中
调焦法 <sup>[94-96]</sup>	对光源条件要求比较宽松; 可使用少量图像测量物体表面信息	很难实现自动重建; 需要多张图片才能进行重建	不能实现自动化重建	重建效果比较好	非实时	对纹理复杂物体的重建效果较差, 不能广泛应用
亮度法 <sup>[97-100]</sup>	可全自动、无手工交互地进行高精度建模; 对光照条件要求宽松	鲁棒性较低; 灵活性较低; 复杂度较高	自动化重建	重建效果比较精细	非实时	可应用于文物数字化和人脸自动建模等领域

表 3 基于视觉的三维重建技术对比与分析(续)

Table 3 Comparison and analysis of 3D reconstruction based on vision (Continued)

单目 视觉 法 <sup>[68]</sup>	简单方便、价格便宜、 灵活可靠、使用范围 广;可以实现重建过程 中的摄像机自标定, 处理时间短	不能够得到深度信息, 重建速度较慢	自动化重建	重建效果 较差	实时	可应用于各种场景
双目 视觉 法 <sup>[101-110,112]</sup>	方法成熟;能够获得 较好的重建效果	运算量大;价格较贵; 在基线距离较大时重 建效果降低	完全自动化 重建	基线在一定 条件下重建 效果较好	实时	适用于室外场景,应 用范围广泛
多目 视觉 法 <sup>[111,113-119]</sup>	识别精度高,适应性 较强,视野范围大	运算量较大;价格昂 贵,重建时间长	完全自动化 重建	基线距离较 大的情况下 重建效果明 显降低,而且 测量精度下 降,速度受限	实时	能够适应各种场景, 在很多范围内都可以 使用
区域 视觉 法 <sup>[120-126]</sup>	计算简单;匹配速度 有所提高;匹配精度 较高;提高了稠密匹 配效率	受光线干扰较大;对 图像要求较高;实验 对象偏少	一定程度的 自动化重建	重建结果 较好	非实时	适用于各种领域,例 如,视觉导航、遥感 测绘
特征 视觉 法 <sup>[154-166]</sup>	提取简单;抗干扰能 力强;鲁棒性好;时 间和空间复杂度低	不能够对图像信息进 行全面的描述	完全自动化 重建	能够较精确 地对物体实 现三维重建	实时	应用范围较广
运动恢 复结构 法 <sup>[127-139]</sup>	实用价值较高;鲁棒 性较强;对图像的 要求较低	计算量较大,重建时 间较长	完全自动化 重建	重建效果取 决于获取图 像数量,图 像越多重建 效果越好	实时	一般适用于大规模场 景中
因子 分解 法 <sup>[145-148]</sup>	简便灵活,抗噪能力 强,不依赖于其他模 型	精度较低,运算时间 较长	完全自动化 重建	重建效果精 度较低	实时	一般适用于大场景中
多视图 几何 法 <sup>[149-163]</sup>	实用性较高;通用性 较强;能够解决运动 恢复结构法中的一些 问题	计算量较大,重建时 间较长	一定程度完 全自动化重 建	重建效果比 较好	实时	一般应用于静止的场 景
统计 学习 法 <sup>[164-173]</sup>	重建质量和效率都很 高;基本不需要人工 交互	获取的信息和数据库 目标不一致时,重建 结果与目标相差甚远	一定程度的 自动化重建	重建效果取 决于数据库的完 整程度,数据库 越完备重建 效果越好	非实时	适用于大场景、识别 和视频检索系统
神经 网络 法 <sup>[174-177]</sup>	精度较高,具有很强 的鲁棒性	收敛速度慢,运算量 较大	一定程度完 全自动化重 建	重建效果较好	实时	能够应用于各种领 域,例如计算机视觉、 军事及航天等
深度学 习与语 义法 <sup>[178-181]</sup>	计算简单,精度较高, 不需要进行复杂的几 何运算,实时性较好	训练时间较长,对 CPU 的要求较高	一定程度完 全自动化重 建	重建结果取 决于训练的 好坏	实时	适用于各种大规模场 景

较慢,运算量较大.另一方面,三维重建技术方法要求的功能也更加具体化,不仅要能精确地重建出物体的形状,还要能够对物体的形状进行辨识,尤其是能够对智能车重建出的前方环境中的动态目标进行判别.同时,在人工智能、机器人技术、无人驾驶等领域中,三维重建已成为不可或缺内容.

### 3 总结与展望

本文从几个方面对三维重建关键技术进行了分析,分别从主动视觉法和被动视觉进行详细的介绍,并总结出了其他被动视觉法在三维重建中的研究进展,以及三维重建效果.在分析前文三维重建方法的基础上,可以大致归纳出不同方法其相应的应用价值:

1) 基于主动视觉的三维重建技术如激光扫描法、结构光法、阴影法和 TOF 技术、雷达技术、Kinect 技术是目前比较成熟而有效的主动方法之一.应用此类方法,可用于不同环境下的三维重建.该类方法不足的是成本高昂,需要购买扫描仪等专用设备,如果操作稍有差错就会导致重构的结果不精确.另外,由于环境的限制主动视觉法不大可能对大规模复杂场景进行扫描,导致其只能应用在小规模领域,并且其后期处理过程也较为复杂.

2) 基于被动视觉的三维重建技术方法,根据相机数目分为单目视觉法、双目视觉法、多目视觉法等,这些方法首先对获取的图像序列中的各种信息进行分析,然后再对物体的建模进行逆向处理,最后,获得场景中物体表面的三维模型.相对于主动视觉法,这种方法对光照的要求并不是很高,而且成本较低、容易操作、易于实现,能够广泛应用于各种复杂的场景.同时也是对激光扫描仪等主动视觉方法的有益补充.进一步地,通过这种方法所估计的参数可直接用于在线三维重建.例如,通过预先估计相机的内外参数,应用到三维重建中,使重建的效果更精确.然而,不足的是,这类方法对物体的细节特征重建还不够精确.

3) 基于其他被动视觉的三维重建关键技术方法,根据应用方法分为运动恢复结构法、因子分解法、多视图几何法、捆绑调整法、统计学习法、区域视觉法、特征视觉法、神经网络法、深度学习与语义法等,此类方法在三维重建中的时间比较长,实时性不高.应用此类方法,需要相机精确的内外参数,因此在相机内外参数估计的过程上花费了较长的时间,使得整体三维重建时间加长.但此过程进一步提高了三维重建的效果,有效地解决了复杂环境条件下的三维重建问题.

三维重建一直是视觉领域研究的热点问题,出现了许多三维重建方法,但是相比于静态的三维重

建,可能有时候更需要实时的三维重建.从目前基于视觉的三维重建关键技术的发展状况来看,以下几个方面的工作值得关注:

1) 目前基于单目视觉的三维重建的应用环境非常广泛,然而大多数是以静态室内环境为主.应用于室外单目视觉的三维重建研究相对比较少,因此,对于动态的室外大规模三维场景重建,包括城市建设等是一个重要的研究方向.

2) 更高效地将视觉传感器获取的三维信息与其他多传感器信息进行融合,可以应用到智能车的环境感知系统,提高智能车周围环境的识别能力,这也是一个非常有价值的研究方向.

3) 基于视觉的三维重建主要是由视觉特征的检测与匹配完成的,然而,目前的视觉特征匹配还存在着许多缺点,例如匹配精度比较低、速度慢以及无法适应重复的纹理等问题.因此,还需要更深一步的研究来发掘新的视觉特征检测与匹配方法,从而满足基于视觉的三维重建在复杂环境中的应用.

4) 传统三维重建方法针对的基本都是刚性物体,对于非刚性物体(如液态物体、火焰)和动态物体等的重建问题一直无法较好的实现,因此这也是未来的一个研究热点.

5) 三维重建在医学上的应用也具有巨大的研究价值,是未来相关领域的一个研究方向.

6) 基于视觉的三维重建技术研究和应用已从传统的工业领域快速扩展到其他领域,如医疗康复、外星探索等,这些领域也是未来发展的新方向.

由于三维重建中的方法各有缺点,因此在重建方法上需要改进的地方有如下几点: 1) 计算量过大,比较消耗资源; 2) 抗干扰性不强,易受到各种外在条件的影响; 3) 对光线的要求比较高,太过于理想化,不符合实际的要求.又由于各种方法离实际应用还有一段距离,各种应用需求亟待被满足.也正因为如此,越来越多的研究人员进入该领域并积极推动其发展.总而言之,在未来的工作中,将在现有工作基础上,重点需要在信息融合、虚拟现实、多机协作系统、室外动态环境的三维实时重建算法等方面开展进一步的研究,相信三维重建的发展一定能够促进其该领域的发展.

### References

- 1 Shen S H. Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes. *IEEE Transactions on Image Processing*, 2013, **22**(5): 1901–1914
- 2 Qu Y F, Huang J Y, Zhang X. Rapid 3D reconstruction for image sequence acquired from UAV camera. *Sensors*, 2018, **18**(1): 225–244
- 3 Lee D Y, Park S A, Lee S J, Kim T H, Heang S, Lee J H, et al. Segmental tracheal reconstruction by 3D-printed scaffold: Pivotal role of asymmetrically porous membrane. *The Laryngoscope*, 2016, **126**(9): E304–E309



- 4 Roberts L G. Machine Perception of Three-Dimensional Solids [Ph. D. dissertation], Massachusetts Institute of Technology, USA, 1963
- 5 Kiyasu S, Hoshino H, Yano K, Fujimura S. Measurement of the 3-D shape of specular polyhedrons using an m-array coded light source. *IEEE Transactions on Instrumentation and Measurement*, 1995, **44**(3): 775–778
- 6 Snavely N, Seitz S M, Szeliski R. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics*, 2006, **25**(3): 835–846
- 7 Pollefeys M, Nistér D, Frahm J M, Akbarzadeh A, Mordohai P, Clipp B, et al. Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision*, 2008, **78**(2–3): 143–167
- 8 Furukawa Y, Ponce J. Carved visual hulls for image-based modeling. *International Journal of Computer Vision*, 2009, **81**(1): 53–67
- 9 Han J G, Shao L, Xu D, Shotton J. Enhanced computer vision with Microsoft Kinect sensor: a review. *IEEE Transactions on Cybernetics*, 2013, **43**(5): 1318–1334
- 10 Ondruška P, Kohli P, Izadi S. Mobilefusion: real-time volumetric surface reconstruction and dense tracking on mobile phones. *IEEE Transactions on Visualization and Computer Graphics*, 2015, **21**(11): 1251–1258
- 11 Li Li, Ma Song-De. On the global quadric shape from contour. *Chinese Journal of Computers*, 1996, **19**(6): 401–408 (李利, 马颂德. 从二维轮廓线重构三维二次曲面形状. 计算机学报, 1996, **19**(6): 401–408)
- 12 Zhong Y D, Zhang H F. Control points based semi-dense matching. In: Proceedings of the 5th Asian Conference on Computer Vision. Melbourne, Australia: ACCV, 2002. 23–25
- 13 Lei Cheng, Hu Zhan-Yi, Wu Fu-Chao, Tsui H T. A novel camera self-calibration technique based on the Kruppa equations. *Chinese Journal of Computers*, 2003, **26**(5): 587–597 (雷成, 胡占义, 吴福朝, Tsui H T. 一种新的基于 Kruppa 方程的摄像机自标定方法. 计算机学报, 2003, **26**(5): 587–597)
- 14 Lei Cheng, Wu Fu-Chao, Hu Zhan-Yi. A new camera self-calibration method based on active vision system. *Chinese Journal of Computer*, 2000, **23**(11): 1130–1139 (雷成, 吴福朝, 胡占义. 一种新的基于主动视觉系统的摄像机自标定方法. 计算机学报, 2000, **23**(11): 1130–1139)
- 15 Zhang Tao. 3D Reconstruction Based Monocular Vision [Master thesis], Xidian University, China, 2014 (张涛. 基于单目视觉的三维重建 [硕士学位论文], 西安电子科技大学, 中国, 2014)
- 16 Ebrahimnezhad H, Ghassemian H. Robust motion from space curves and 3D reconstruction from multiviews using perpendicular double stereo rigs. *Image and Vision Computing*, 2008, **26**(10): 1397–1420
- 17 Hartley R, Zisserman A. *Multiple View Geometry in Computer Vision*. New York: Cambridge University Press, 2003
- 18 Várady T, Martin R R, Cox J. Reverse engineering of geometric models – an introduction. *Computer-Aided Design*, 1997, **29**(4): 255–268
- 19 Isgro F, Odone F, Verri A. An open system for 3D data acquisition from multiple sensor. In: Proceedings of the 7th International Workshop on Computer Architecture for Machine Perception. Palermo, Italy: IEEE, 2005. 52–57
- 20 Williams C G, Edwards M A, Colley A L, Macpherson J V, Unwin P R. Scanning micropipet contact method for high-resolution imaging of electrode surface redox activity. *Analytical Chemistry*, 2009, **81**(7): 2486–2495
- 21 Kraus K, Pfeifer N. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 1998, **53**(4): 193–203
- 22 Göbel W, Kampa B M, Helmchen F. Imaging cellular network dynamics in three dimensions using fast 3D laser scanning. *Nature Methods*, 2007, **4**(1): 73–79
- 23 Rocchini C, Cignoni P, Montani C, Pingi P, Scopigno R. A low cost 3D scanner based on structured light. *Computer Graphics Forum*, 2001, **20**(3): 299–308
- 24 Al-Najdawi N, Bez H E, Singhai J, Edirisinghe E A. A survey of cast shadow detection algorithms. *Pattern Recognition Letters*, 2012, **33**(6): 752–764
- 25 Park J, Kim H, Tai Y W, Brown M S, Kweon I. High quality depth map upsampling for 3d-tof cameras. In: Proceedings of the 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 1623–1630
- 26 Schwarz B. LIDAR: mapping the world in 3D. *Nature Photonics*, 2010, **4**(7): 429–430
- 27 Khoshelham K, Elberink S O. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 2012, **12**(2): 1437–1454
- 28 Yang Yao-Quan, Shi Ren, Yu Xi-Ning, Gao Tang-Nian. Laser scanning triangulation for large profile measurement. *Journal of Xi'an Jiaotong University*, 1999, **33**(7): 15–18 (杨耀权, 施仁, 于希宁, 高饒年. 激光扫描三角法大型曲面测量中影响参数分析. 西安交通大学学报, 1999, **33**(7): 15–18)
- 29 Boehler W, Vicent M B, Marbs A. Investigating laser scanner accuracy. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2003, **34**(5): 696–701
- 30 Reshetyuk Y. Investigation and Calibration of Pulsed Time-of-Flight Terrestrial Laser Scanners [Master dissertation], Royal Institute of Technology, Switzerland, 2006. 14–17
- 31 Voisin S, Foufou S, Truchetet F, Page D L, Abidi M A. Study of ambient light influence for three-dimensional scanners based on structured light. *Optical Engineering*, 2007, **46**(3): Article No. 030502
- 32 Scharstein D, Szeliski R. High-accuracy stereo depth maps using structured light. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Madison, WI, USA: IEEE, 2003. I-195–I-202
- 33 Chen F, Brown G M, Song M M. Overview of 3-D shape measurement using optical methods. *Optical Engineering*, 2000, **39**(1): 10–22
- 34 Pollefeys M, Van Gool L. Stratified self-calibration with the modulus constraint. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, **21**(8): 707–724
- 35 O’Toole M, Mather J, Kutulakos K N. 3D shape and indirect appearance by structured light transport. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(7): 1298–1312
- 36 Song Z, Chung R. Determining both surface position and orientation in structured-light-based sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(10): 1770–1780
- 37 Kowarschik R, Kuehmstedt P, Gerber J, Schreiber W, Notni G. Adaptive optical 3-D-measurement with structured light. *Optical Engineering*, 2000, **39**(1): 150–158
- 38 Shakhnarovich G, Viola P A, Moghaddam B. A unified learning framework for real time face detection and classification. In: Proceedings of the 5th IEEE International Conference on Automatic Face Gesture Recognition. Washington, USA: IEEE, 2002. 14–21

- 39 Salvi J, Pagès J, Batlle J. Pattern codification strategies in structured light systems. *Pattern Recognition*, 2004, **37**(4): 827–849
- 40 Zhang Guang-Jun, Li Xin, Wei Zhen-Zhong. A method of 3D double-vision inspection based on structured light. *Chinese Journal of Scientific Instrument*, 2002, **23**(6): 604–607, 624  
(张广军, 李鑫, 魏振忠. 结构光三维双视觉检测方法研究. 仪器仪表学报, 2002, **23**(6): 604–607, 624)
- 41 Wang Bao-Guang, He Zhong-Hai, Chen Lin-Cai, Ni Yong. Model and performance analysis of structured light sensor. *Acta Optica Sinica*, 2002, **22**(4): 481–484  
(王宝光, 贺忠海, 陈林才, 倪勇. 结构光传感器模型及特性分析. 光学学报, 2002, **22**(4): 481–484)
- 42 Luo Xian-Bo, Zhong Yue-Xian, Li Ren-Ju. Data registration in 3-D scanning systems. *Journal of Tsinghua University (Science and Technology)*, 2004, **44**(8): 1104–1106  
(罗先波, 钟约先, 李仁举. 三维扫描系统中的数据配准技术. 清华大学学报(自然科学版), 2004, **44**(8): 1104–1106)
- 43 Savarese S, Andreetto M, Rushmeier H, Bernardini F, Perona P. 3D reconstruction by shadow carving: theory and practical evaluation. *International Journal of Computer Vision*, 2007, **71**(3): 305–336
- 44 Wang Y X, Cheng H D, Shan J. Detecting shadows of moving vehicles based on HMM. In: Proceedings of the 19th International Conference on Pattern Recognition. Tampa, FL, USA: IEEE, 2008. 1–4
- 45 Rüfenacht D, Fredembach C, Süsstrunk S. Automatic and accurate shadow detection using near-infrared information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(8): 1672–1678
- 46 Daum M, Dudek G. On 3-D surface reconstruction using shape from shadows. In: Proceedings of the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Santa Barbara, CA, USA: IEEE, 1998. 461–468
- 47 Woo A, Poulin P, Fournier A. A survey of shadow algorithms. *IEEE Computer Graphics and Applications*, 1990, **10**(6): 13–32
- 48 Hasenfratz J M, Lapierre M, Holzschuch N, Sillion F, Gravier/Imag-Inria A. A survey of real-time soft shadows algorithms. *Computer Graphics Forum*, 2003, **22**(4): 753–774
- 49 May S, Droschel D, Holz D, Wiesen C. 3D pose estimation and mapping with time-of-flight cameras. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Nice, France: IEEE, 2008. 120–125
- 50 Hegde G P M, Ye C. Extraction of planar features from swissranger sr-3000 range images by a clustering method using normalized cuts. In: Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, MO, USA: IEEE, 2009. 4034–4039
- 51 Pathak K, Vaskevicius N, Poppinga J, Pfingsthorn M, Schwertfeger S, Birk A. Fast 3D mapping by matching planes extracted from range sensor point-clouds. In: Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, MO, USA: IEEE, 2009. 1150–1155
- 52 Stipes J A, Cole J G P, Humphreys J. 4D scan registration with the SR-3000 LIDAR. In: Proceedings of the 2008 IEEE International Conference on Robotics and Automation. Pasadena, CA, USA: IEEE, 2008. 2988–2993
- 53 May S, Droschel D, Fuchs S, Holz D, Nüchter A. Robust 3D-mapping with time-of-flight cameras. In: Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, MO, USA: IEEE, 2009. 1673–1678
- 54 Streller D, Dietmayer K. Object tracking and classification using a multiple hypothesis approach. In: Proceedings of the 2004 IEEE Intelligent Vehicles Symposium. Parma, Italy: IEEE, 2004. 808–812
- 55 Schwalbe E, Maas H G, Seidel F. 3D building model generation from airborne laser scanner data using 2D GIS data and orthogonal point cloud projections. In: Proceedings of ISPRS WG III/3, III/4, V/3 Workshop “Laser Scanning 2005”. Enschede, the Netherlands: IEEE, 2005. 12–14
- 56 Weiss T, Schiele B, Dietmayer K. Robust driving path detection in urban and highway scenarios using a laser scanner and online occupancy grids. In: Proceedings of the 2007 IEEE Intelligent Vehicles Symposium. Istanbul, Turkey: IEEE, 2007. 184–189
- 57 Hu Ming. Algorithm Reconstruction Based on Cloud Data [Master thesis], South China University of Technology, China, 2010  
(胡明. 基于点云数据的重建算法研究 [硕士学位论文], 华南理工大学, 中国, 2010)
- 58 Wei Zheng. Automated Extraction of Buildings and Facades Reconstruct from Mobile LiDAR Point Clouds [Ph. D. dissertation], Wuhan University, China, 2012  
(魏征. 车载 LiDAR 点云中建筑物的自动识别与立面几何重建 [博士学位论文], 武汉大学, 中国, 2012)
- 59 Zhang Z Y. Microsoft kinect sensor and its effect. *IEEE Multimedia*, 2012, **19**(2): 4–10
- 60 Zhang Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(11): 1330–1334
- 61 Smisek J, Jancosek M, Pajdla T. 3D with Kinect. In: Proceedings of the 2011 IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 1154–1160
- 62 Zollhöfer M, Nießner M, Izadi S, Rehmann C, Zach C, Fisher M, et al. Real-time non-rigid reconstruction using an RGB-D camera. *ACM Transactions on Graphics*, 2014, **33**(4): Article No. 156
- 63 Henry P, Krainin M, Herbst E, Ren X F, Fox D. RGB-D mapping: using depth cameras for dense 3D modeling of indoor environments. *Experimental Robotics: the 12th International Symposium on Experimental Robotics*. Heidelberg, Germany: Springer, 2014. 477–491
- 64 Henry P, Krainin M, Herbst E, Ren X F, Fox D. RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments. *The International Journal of Robotics Research*, 2012, **31**(5): 647–663
- 65 Newcombe R A, Izadi S, Hilliges O, Molyneaux D, Kim D, Davison A J, et al. KinectFusion: real-time dense surface mapping and tracking. In: Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality. Basel, Switzerland: IEEE, 2011. 127–136
- 66 Izadi S, Kim D, Hilliges O, Molyneaux D, Newcombe R, Kohli P, et al. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. Santa Barbara, California, USA: ACM, 2011. 559–568

- 67 Wu Tong. Research on Key Technologies of 3D Reconstruction Based on Multi-plane Detection in Point Clouds [Master thesis], Nanchang Hangkong University, China, 2013 (吴侗. 基于点云多平面检测的三维重建关键技术研究 [硕士学位论文], 南昌航空大学, 中国, 2013)
- 68 Tong Shuai, Xu Xiao-Gang, Yi Cheng-Tao, Shao Cheng-Yong. Overview on vision-based 3D reconstruction. *Application Research of Computers*, 2011, **28**(7): 2411–2417 (佟帅, 徐晓刚, 易成涛, 邵承永. 基于视觉的三维重建技术综述. 计算机应用研究, 2011, **28**(7): 2411–2417)
- 69 Horn B K P. Shape from Shading: A Method for Obtaining the Shape of A Smooth Opaque Object from One View [Ph. D. dissertation], Massachusetts Institute of Technology, USA, 1970
- 70 Penna M A. A shape from shading analysis for a single perspective image of a polyhedron. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, **11**(6): 545–554
- 71 Bakshi S, Yang Y H. Shape from shading for non-Lambertian surfaces. In: Proceedings of the 1st International Conference on Image Processing. Austin, USA: IEEE, 1994. 130–134
- 72 Vogel O, Breuß M, Weickert J. Perspective shape from shading with non-Lambertian reflectance. *Pattern Recognition*. Heidelberg, Germany: Springer, 2008. 517–526
- 73 Woodham R J. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 1980, **19**(1): Article No. 191139
- 74 Noakes L, Kozera R. Nonlinearities and noise reduction in 3-source photometric stereo. *Journal of Mathematical Imaging and Vision*, 2003, **18**(2): 119–127
- 75 Horovitz I, Kiryati N. Depth from gradient fields and control points: bias correction in photometric stereo. *Image and Vision Computing*, 2004, **22**(9): 681–694
- 76 Tang K L, Tang C K, Wong T T. Dense photometric stereo using tensorial belief propagation. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 132–139
- 77 Xie H, Pierce L E, Ulaby F T. SAR speckle reduction using wavelet denoising and Markov random field modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 2002, **40**(10): 2196–2212
- 78 Sun J A, Smith M, Smith L, Midha S, Bamber J. Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities. *Image and Vision Computing*, 2007, **25**(7): 1050–1057
- 79 Vlasic D, Peers P, Baran I, Debevec P, Popović J, Rusinkiewicz S, et al. Dynamic shape capture using multi-view photometric stereo. *ACM Transactions on Graphics*, 2009, **28**(5): Article No. 174
- 80 Shi B X, Matsushita Y, Wei Y C, Xu C, Tan P. Self-calibrating photometric stereo. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 1118–1125
- 81 Morris N J W, Kutulakos K N. Dynamic refraction stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(8): 1518–1531
- 82 Higo T, Matsushita Y, Ikeuchi K. Consensus photometric stereo. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 1157–1164
- 83 Brown L G, Shvaytser H. Surface orientation from projective foreshortening of isotropic texture autocorrelation. In: Proceedings of the 1988 Computer Society Conference on Computer Vision and Pattern Recognition. Ann Arbor, USA: IEEE, 1988. 510–514
- 84 Clerc M, Mallat S. The texture gradient equation for recovering shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(4): 536–549
- 85 Witkin A P. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 1981, **17**(1–3): 17–45
- 86 Warren P A, Mamassian P. Recovery of surface pose from texture orientation statistics under perspective projection. *Biological Cybernetics*, 2010, **103**(3): 199–212
- 87 Martin W N, Aggarwal J K. Volumetric descriptions of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1983, **PAMI-5**(2): 150–158
- 88 Laurentini A. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994, **16**(2): 150–162
- 89 Bongiovanni G, Guerra C, Levioldi S. Computing the Hough transform on a pyramid architecture. *Machine Vision and Applications*, 1990, **3**(2): 117–123
- 90 Darell T, Wohn K. Depth from focus using a pyramid architecture. *Pattern Recognition Letters*, 1990, **11**(12): 787–796
- 91 Kavianpour A, Bagherzadeh N. Finding circular shapes in an image on a pyramid architecture. *Pattern Recognition Letters*, 1992, **13**(12): 843–848
- 92 Forbes K, Nicolls F, De Jager G, Voigt A. Shape-from-silhouette with two mirrors and an uncalibrated camera. In: Proceedings of the 9th European Conference on Computer Vision. Graz, Austria: Springer, 2006. 165–178
- 93 Lehtinen J, Aila T, Chen J W, Laine S, Durand F. Temporal light field reconstruction for rendering distribution effects. *ACM Transactions on Graphics*, 2011, **30**(4): Article No. 55
- 94 Nayar S K, Nakagawa Y. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994, **16**(8): 824–831
- 95 Hasinoff S W, Kutulakos K N. Confocal stereo. *International Journal of Computer Vision*, 2009, **81**(1): 82–104
- 96 Pradeep K S, Rajagopalan A N. Improving shape from focus using defocus cue. *IEEE Transactions on Image Processing*, 2007, **16**(7): 1920–1925
- 97 Slabaugh G G, Culbertson W B, Malzbender T, Stevens M R, Schafer R W. Methods for volumetric reconstruction of visual scenes. *International Journal of Computer Vision*, 2004, **57**(3): 179–199
- 98 de Vries S C, Kappers A M L, Koenderink J J. Shape from stereo: a systematic approach using quadratic surfaces. *Perception & Psychophysics*, 1993, **53**(1): 71–80
- 99 Seitz S M, Dyer C R. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 1999, **35**(2): 151–173
- 100 Kutulakos K N, Seitz S M. A theory of shape by space carving. *International Journal of Computer Vision*, 2000, **38**(3): 199–218
- 101 Li D W, Xu L H, Tang X S, Sun S Y, Cai X, Zhang P. 3D imaging of greenhouse plants with an inexpensive binocular stereo vision system. *Remote Sensing*, 2017, **9**(5): Article No. 508

- 102 Helveston E M, Boudreaault G. Binocular vision and ocular motility: theory and management of strabismus. *American Journal of Ophthalmology*, 1986, **101**(1): 135
- 103 Qi F, Zhao D B, Gao W. Reduced reference stereoscopic image quality assessment based on binocular perceptual information. *IEEE Transactions on Multimedia*, 2015, **17**(12): 2338–2344
- 104 Sizintsev M, Wildes R P. Spacetime stereo and 3D flow via binocular spatiotemporal orientation analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(11): 2241–2254
- 105 Marr D, Poggio T. A computational theory of human stereo vision. *Proceedings of the Royal Society B: Biological Sciences*, 1979, **204**(1156): 301–328
- 106 Zou X J, Zou H X, Lu J. Virtual manipulator-based binocular stereo vision positioning system and errors modelling. *Machine Vision and Applications*, 2012, **23**(1): 43–63
- 107 Li Zhan-Xian, Xu Zhe. Analysis of imaging model of binocular vision. *Mechanical Engineering & Automation*, 2014, (4): 191–192  
(李占贤, 许哲. 双目视觉的成像模型分析. *机械工程与自动化*, 2014, (4): 191–192)
- 108 Zhang Wen-Ming, Liu Bin, Li Hai-Bin. Characteristic point extracts and the match algorithm based on the binocular vision in three dimensional reconstruction. *Optical Technique*, 2008, **34**(2): 181–185  
(张文明, 刘彬, 李海滨. 基于双目视觉的三维重建中特征点提取及匹配算法的研究. *光学技术*, 2008, **34**(2): 181–185)
- 109 Bruno F, Bianco G, Muzzupappa M, Barone S, Rationale A V. Experimentation of structured light and stereo vision for underwater 3D reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011, **66**(4): 508–518
- 110 Fusiello A, Trucco E, Verri A. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 2000, **12**(1): 16–22
- 111 Baillard C, Zisserman A. A plane-sweep strategy for the 3D reconstruction of buildings from multiple images. In: Proceedings of the 2000 International Archives of Photogrammetry and Remote Sensing. Amsterdam, Netherlands: ISPRS, 2000. 56–62
- 112 Hirschmuller H, Scharstein D. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(9): 1582–1599
- 113 Zhang T, Liu J H, Liu S L, Tang C T, Jin P. A 3D reconstruction method for pipeline inspection based on multi-vision. *Measurement*, 2017, **98**: 35–48
- 114 Saito K, Miyoshi T, Yoshikawa H. Noncontact 3-D digitizing and machining system for free-form surfaces. *CIRP Annals*, 1991, **40**(1): 483–486
- 115 Chen Ming-Zhou. Research on Active Grating Projection Stereo Vision Sensor [Master thesis], Tianjin University, China, 2002  
(陈明舟. 主动光栅投影双目视觉传感器的研究 [硕士学位论文], 天津大学, 中国, 2002)
- 116 Wang Lei. The Research and Application of 3D Reconstruction Technology [Master thesis], Tsinghua University, China, 2002  
(王磊. 三维重构技术的研究与应用 [硕士学位论文], 清华大学, 中国, 2002)
- 117 Hernández C, Vogiatzis G, Cipolla R. Overcoming shadows in 3-source photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(2): 419–426
- 118 Park H, Lee H, Sull S. Efficient viewer-centric depth adjustment based on virtual fronto-parallel planar projection in stereo 3D images. *IEEE Transactions on Multimedia*, 2014, **16**(2): 326–336
- 119 Bai X, Rao C, Wang X G. Shape vocabulary: a robust and efficient shape representation for shape matching. *IEEE Transactions on Image Processing*, 2014, **23**(9): 3935–3949
- 120 Goshtasby A, Stockman G C, Page C V. A region-based approach to digital image registration with subpixel accuracy. *IEEE Transactions on Geoscience and Remote Sensing*, 1986, **GE-24**(3): 390–399
- 121 Flusser J, Suk T. A moment-based approach to registration of images with affine geometric distortion. *IEEE Transactions on Geoscience and Remote Sensing*, 1994, **32**(2): 382–387
- 122 Alhichri H S, Kamel M. Virtual circles: a new set of features for fast image registration. *Pattern Recognition Letters*, 2003, **24**(9–10): 1181–1190
- 123 Schmid C, Mohr R. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(5): 530–535
- 124 Matas J, Chum O, Urban M, Pajdla T. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 2004, **22**(10): 761–767
- 125 Tuytelaars T, Van Gool L. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 2004, **59**(1): 61–85
- 126 Kadir T, Zisserman A, Brady M. An affine invariant salient region detector. In: Proceedings of the 8th European Conference on Computer Vision. Prague, Czech Republic: Springer, 2004. 228–241
- 127 Harris C, Stephens M. A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference. 1988. 1475–151
- 128 Morevec H P. Towards automatic visual obstacle avoidance. In: Proceedings of the 5th International Joint Conference on Artificial Intelligence. Cambridge, USA: ACM, 1977. 584
- 129 Schmid C, Mohr R, Bauckhage C. Evaluation of interest point detectors. *International Journal of Computer Vision*, 2000, **37**(2): 151–172
- 130 Van de Weijer J, Gevers T, Bagdanov A D. Boosting color saliency in image feature detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(1): 150–156
- 131 Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 2004, **60**(1): 63–86
- 132 Smith S M, Brady J M. SUSAN — a new approach to low level image processing. *International Journal of Computer Vision*, 1997, **23**(1): 45–78
- 133 Lindeberg T, Gårding J. Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure. *Image and Vision Computing*, 1997, **15**(6): 415–434
- 134 Lindeberg T. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 1998, **30**(2): 79–116
- 135 Baumberg A. Reliable feature matching across widely separated views. In: Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head Island, USA: IEEE, 2000. 774–781

- 136 Lowe D G. Object recognition from local scale-invariant features. In: Proceedings of the 7th IEEE International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1999. 1150–1157
- 137 Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2004. II-506–II-513
- 138 Bay H, Tuytelaars T, Van Gool L. SURF: speeded up robust features. In: Proceedings of the 9th European Conference on Computer Vision. Graz, Austria: Springer, 2006. 404–417
- 139 Ge Pan-Pan, Chen Qiang, Gu Yi-He. Algorithm of remote sensing image matching based on Harris corner and SURF feature. *Application Research of Computers*, 2014, **31**(7): 2205–2208  
(葛盼盼, 陈强, 顾一禾. 基于 Harris 角点和 SURF 特征的遥感图像匹配算法. *计算机应用研究*, 2014, **31**(7): 2205–2208)
- 140 Wu C C. Towards linear-time incremental structure from motion. In: Proceedings of the 2013 International Conference on 3D Vision. Seattle, WA, USA: IEEE, 2013. 127–134
- 141 Cui H N, Shen S H, Gao W, Hu Z Y. Efficient large-scale structure from motion by fusing auxiliary imaging information. *IEEE Transactions on Image Processing*, 2015, **24**(11): 3561–3573
- 142 Sturm P, Triggs B. A factorization based algorithm for multi-image projective structure and motion. In: Proceedings of the 4th European conference on computer vision. Cambridge, UK: Springer, 1996. 709–720
- 143 Crandall D, Owens A, Snavely N, Huttenlocher D. Discrete-continuous optimization for large-scale structure from motion. In: Proceedings of CVPR 2011. Colorado Springs, CO, USA: IEEE, 2011. 3001–3008
- 144 Irschara A, Hoppe C, Bischof H, Kluckner S. Efficient structure from motion with weak position and orientation priors. In: Proceedings of CVPR 2011 WORKSHOPS. Colorado Springs, CO, USA: IEEE, 2011. 21–28
- 145 Tomasi C, Kanade T. Shape and motion from image streams: a factorization method. Carnegie Mellon University, USA, 1992. 9795–9802
- 146 Poelman C J, Kanade T. A paraperspective factorization method for shape and motion recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(3): 206–218
- 147 Triggs B. Factorization methods for projective structure and motion. In: Proceedings of the 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 1996. 845–851
- 148 Han M, Kanade T. Multiple motion scene reconstruction from uncalibrated views. In: Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, Canada: IEEE, 2001. 163–170
- 149 Yu Hai-Yan. 3D Reconstruction Based on Multiple View Geometry [Ph.D. dissertation], Harbin Institute of Technology, China, 2007  
(于海雁. 基于多视图几何的三维重建研究 [博士学位论文], 哈尔滨工业大学, 中国, 2007)
- 150 Sivic J, Zisserman A. Video Google: a text retrieval approach to object matching in videos. In: Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003. 1470–1477
- 151 Faugeras O. *Three-dimensional Computer Vision: A Geometric Viewpoint*. Cambridge: MIT Press, 1993
- 152 Xie R P, Yao J, Liu K, Lu X H, Liu X H, Xia M H, et al. Automatic multi-image stitching for concrete bridge inspection by combining point and line features. *Automation in Construction*, 2018, **90**: 265–280
- 153 Yan W Q, Hou C P, Lei J J, Fang Y M, Gu Z Y, Ling N. Stereoscopic image stitching based on a hybrid warping model. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, **27**(9): 1934–1946
- 154 Pei J F, Huang Y L, Huo W B, Zhang Y, Yang J Y, Yeo T S. SAR automatic target recognition based on multiview deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, **56**(4): 2196–2210
- 155 Longuet-Higgins H C. A computer algorithm for reconstructing a scene from two projections. *Nature*, 1981, **293**(5828): 133–135
- 156 Faugeras O D, Maybank S. Motion from point matches: multiplicity of solutions. *International Journal of Computer Vision*, 1990, **4**(3): 225–246
- 157 Luong Q T, Deriche R, Faugeras O D, Papadopoulou T. On determining the fundamental matrix: analysis of different methods and experimental results. RR-1894, INRIA, 1993. 24–48
- 158 Luong Q T, Faugeras O D. The fundamental matrix: theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 1996, **17**(1): 43–75
- 159 Wu C C, Agarwal S, Curless B, Seitz S M. Multicore bundle adjustment. In: Proceedings of CVPR 2011. Providence, RI, USA: IEEE, 2011. 3057–3064
- 160 Lourakis M I A, Argyros A A. SBA: a software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software*, 2009, **36**(1): Article No. 2
- 161 Choudhary S, Gupta S, Narayanan P J. Practical time bundle adjustment for 3D reconstruction on the GPU. In: Proceedings of the 11th European Conference on Trends and Topics in Computer Vision. Heraklion, Crete, Greece: Springer, 2010. 423–435
- 162 Hu Z Y, Gao W, Liu X, Guo F S. 3D reconstruction for heritage preservation [Online], available: <http://vision.ia.ac.cn>, March 29, 2012.
- 163 Fang T, Quan L. Resampling structure from motion. In: Proceedings of the 11th European Conference on Computer Vision. Crete, Greece: Springer, 2010. 1–14
- 164 Tanimoto J, Hagishima A. State transition probability for the Markov model dealing with on/off cooling schedule in dwellings. *Energy and Buildings*, 2005, **37**(3): 181–187
- 165 Eddy S R. Profile hidden Markov models. *Bioinformatics*, 1998, **14**(9): 755–763
- 166 Chang M T, Chen S Y. Deformed trademark retrieval based on 2D pseudo-hidden Markov model. *Pattern Recognition*, 2001, **34**(5): 953–967
- 167 Saxena A, Chung S H, Ng A Y. 3-D depth reconstruction from a single still image. *International Journal of Computer Vision*, 2008, **76**(1): 53–69
- 168 Handa A, Whelan T, McDonald J, Davison A J. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation. Hong Kong, China: IEEE, 2014. 1524–1531
- 169 Kemelmacher-Shlizerman I, Basri R. 3D face reconstruction from a single image using a single reference face shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(2): 394–405

- 170 Lee S J, Park K R, Kim J. A SFM-based 3D face reconstruction method robust to self-occlusion by using a shape conversion matrix. *Pattern Recognition*, 2011, **44**(7): 1470–1486
- 171 Song M L, Tao D C, Huang X Q, Chen C, Bu J J. Three dimensional face reconstruction from a single image by a coupled RBF network. *IEEE Transactions on Image Processing*, 2012, **21**(5): 2887–2897
- 172 Seo H, Yeo Y I, Wohn K. 3D body reconstruction from photos based on range scan. In: Proceedings of the 1st International Conference on Technologies for E-Learning and Digital Entertainment. Hangzhou, China: Springer, 2006. 849–860
- 173 Allen B, Curless B, Popovi Z. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics*, 2003, **22**(3): 587–594
- 174 Funahashi K I. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 1989, **2**(3): 183–192
- 175 Do Y. Application of neural networks for stereo-camera calibration. In: Proceedings of the 1999 International Joint Conference on Neural Networks. Washington, USA: IEEE, 1999. 2719–2722
- 176 Yuan Ye, Ou Zong-Ying, Tian Zhong-Xu. 3D reconstruction of stereo vision using neural networks implicit vision model. *Journal of Computer-Aided Design & Computer Graphics*, 2003, **15**(3): 293–296  
(袁野, 欧宗瑛, 田中旭. 应用神经网络隐式视觉模型进行立体视觉的三维重建. 计算机辅助设计与图形学学报, 2003, **15**(3): 293–296)
- 177 Li X P, Chen L Z. Research on the application of BP neural networks in 3D reconstruction noise filter. *Advanced Materials Research*, 2014, **998–999**: 911–914
- 178 Savinov N, Ladický L, Häne C, Pollefeys M. Discrete optimization of ray potentials for semantic 3D reconstruction. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 5511–5518
- 179 Bláha M, Vogel C, Richard A, Wegner J D, Pock T, Schindler K. Large-scale semantic 3D reconstruction: an adaptive multi-resolution model for multi-class volumetric labeling. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 3176–3184
- 180 Sünderhauf N, Pham T T, Latif Y, Milford M, Reid I. Meaningful maps with object-oriented semantic mapping. In: Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Vancouver, BC, Canada: IEEE, 2017. 5079–5085
- 181 Zhao Yang, Liu Guo-Liang, Tian Guo-Hui, Luo Yong, Wang Zi-Ren, Zhang Wei, et al. A survey of visual SLAM based on deep learning. *Robot*, 2017, **39**(6): 889–896  
(赵洋, 刘国良, 田国会, 罗勇, 王梓任, 张威, 等. 基于深度学习的视觉 SLAM 综述. 机器人, 2017, **39**(6): 889–896)



郑太雄 重庆邮电大学教授, 工学博士。主要研究方向为汽车电子相关研究。本文通信作者。

E-mail: zhengtx@cqupt.edu.cn

(ZHENG Tai-Xiong Ph.D., professor at Chongqing University of Posts and Telecommunications. His research interest covers automotive electronics related research. Corresponding author of this paper.)

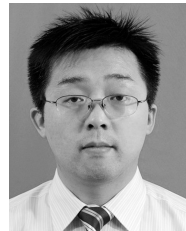


黄帅 重庆大学自动化学院博士研究生。2018 年获得重庆邮电大学硕士学位。主要研究方向为智能车环境感知, 信息物理系统, 智能交通, 数据挖掘。

E-mail: huangs316@163.com

(HUANG Shuai Ph.D. candidate in the School of Automation, Chongqing University. He received his

master degree from Chongqing University of Posts and Telecommunications in 2018. His research interest covers environment perception for intelligent vehicle, cyber-physical systems, intelligent transportation systems, and data mining.)



李永福 重庆邮电大学副教授, 工学博士, 普渡大学博士后。主要研究方向为车联网与智能交通, 汽车电子, 控制理论与应用。E-mail: laf1212@163.com

(LI Yong-Fu Ph.D., associate professor at Chongqing University of Posts and Telecommunications. Since 2014 to 2016, Dr. Li has been worked as the

Post-doc Research Associate at Purdue University, West Lafayette, IN 47906, USA. His research interest covers connected and autonomous vehicles, intelligent transportation systems, automotive electronics, control theory and application.)



冯明驰 重庆邮电大学副教授, 工学博士。主要研究方向为多相机视觉测量。

E-mail: fengmc@cqupt.edu.cn

(FENG Ming-Chi Ph.D., associate professor at Chongqing University of Posts and Telecommunications. His research interest covers vision measurement of multi-cameras.)