

# 基于视觉的三维重建技术综述\*

佟 帅, 徐晓刚, 易成涛, 邵承永

(海军大连舰艇学院 装备系统与自动化系, 辽宁 大连 116018)

**摘 要:** 基于视觉的三维重建仍然存在较大局限性。通过介绍基于视觉的三维重建技术的主要方法及其研究现状,对各种方法优缺点进行了比较分析,期望能够对该领域有较全面的把握,进一步明确未来的研究方向。

**关键词:** 基于视觉的三维重建; 单目视觉; 双目视觉; 三目视觉; 综述

**中图分类号:** TP391.9      **文献标志码:** A      **文章编号:** 1001-3695(2011)07-2411-07

**doi:** 10.3969/j.issn.1001-3695.2011.07.003

## Overview on vision-based 3D reconstruction

TONG Shuai, XU Xiao-gang, YI Cheng-tao, SHAO Cheng-yong

(Dept. of Equipment System & Automatization, Dalian Naval Academy, Dalian Liaoning 116018, China)

**Abstract:** As an developing technology, vision-based 3D reconstruction still got limitations in many aspects. Overviewed the main methods and relative study status of vision-based 3D reconstruction, and analyzed the advantages and disadvantages of these methods, hope to get a general understanding of this field and future indicate the study orientations in the future.

**Key words:** vision-based 3D reconstruction; monocular vision; binocular vision; trinocular vision; overview

## 0 引言

使用建模软件进行三维建模是常用的方法,但建模所需要耗费的大量人力、物力常常让人望而却步,重建效果也往往差强人意。基于视觉的三维重建技术为解决这一问题提供了一条新的思路。

基于视觉的三维重建技术,即采用计算机视觉方法进行物体的三维模型重建,是指利用数字摄像机作为图像传感器,综合运用图像处理、视觉计算等技术进行非接触三维测量,用计算机程序获取物体的三维信息。其优势在于不受物体形状限制,重建速度较快,可以实现全自动或半自动建模等,是三维重建的一个重要发展方向,能广泛应用于包括移动机器人自主导航系统、航空及遥感测量、工业自动化系统等在内的各个领域,由此项技术产生的经济效益极为可观。

作为计算机视觉技术的一个重要分支,基于视觉的三维重建技术以 Marr 的视觉理论框架为基础,形成了多种理论方法。例如,根据摄像机数目的不同,可分为单目视觉法、双目视觉法、三目视觉或多目视觉法;根据原理的不同,又可分为基于区域的视觉方法、基于特征的视觉方法、基于模型的方法和基于规则的视觉方法等;根据获取数据的方式,可分为主动视觉法和被动视觉法等。

本文根据最近几年来的国内外研究情况,选取了基于视觉的三维重建技术中相关研究和实际应用比较多的几种方法进行介绍和比较分析,指出了今后面临的主要挑战和未来发展方向。根据使用摄像机数目不同,本文将基于视觉的三维重建方法分为单目视觉方法、双目视觉方法和三目视觉方法三类进行

了介绍,重点介绍了单目视觉方法。

## 1 单目视觉方法

单目视觉方法(monocular vision)是指使用一台摄像机进行三维重建的方法。所使用的图像可以是单视点的单幅或多幅图像,也可以是多视点的多幅图像。前者主要通过图像的二维特征(用  $X$  表示)推导出深度信息,这些二维特征包括明暗度、纹理、焦点、轮廓等,因此也被统称为  $X$  恢复形状法(shape from  $X$ )。这一类方法设备结构简单,使用单幅或少数几张图像就可以重建出物体三维模型;不足的是通常要求的条件比较理想化,实际应用情况不是很理想,重建效果也一般。后者通过匹配不同图像中的相同特征点,利用这些匹配约束求取空间三维点坐标信息,从而实现三维重建。这种方法可以实现重建过程中的摄像机自标定,能够满足大规模场景三维重建的需求,且在图像资源丰富的情况下重建效果比较好;不足之处是运算量比较大,重建时间较长。下面将介绍几种主要的单目视觉方法。

### 1.1 明暗度法

明暗度法,即明暗度恢复形状法(shape from shading, SFS)。这种方法通过分析图像中的明暗度信息,运用反射光照模型,恢复出物体表面的法向信息进行三维重建。Horn<sup>[1]</sup>于1970年首次提出了SFS方法的概念,并给出了一个表示二维图像中各像素点明暗度与其对应的三维点的法向、反射率以及光照方向之间关系的非线性偏微分方程——明暗度方程。

但是这种SFS方法是一个欠约束问题<sup>[2]</sup>,需要其他约束才能进行求解。所以传统的SFS方法还要基于三个假设:a)

收稿日期: 2011-01-13; 修回日期: 2011-03-30      基金项目: 国家自然科学基金资助项目(60975016 61002052)

作者简介: 佟帅(1987-),男(满族),辽宁宽甸人,硕士研究生,主要研究方向为机器视觉、三维重建(tongs05@163.com);徐晓刚(1967-),男,浙江永康人,教授,博士,主要研究方向为信号分析、图像处理、虚拟仿真;易成涛(1974-),男,湖北随州人,副教授,博士,主要研究方向为图像处理、机器视觉;邵承永(1974-),男,河北石家庄人,博士,主要研究方向为图像图形处理、信息安全。

反射模型为朗伯特模型,即从各个角度观察,同一点的明暗度都相同的;b)光源为无限远处点光源;c)成像关系为正交投影。在这些假设条件下,物体表面明暗度只由光源入射角的余弦决定,因此可以由明暗度求解物体表面法向。但是基于这三个假设的模型存在两个问题:a)朗伯特模型是一种理想化的模型,真实物体通常不满足朗伯特模型;b)正交投影的病态性(解不唯一)问题——不同的表面也可以产生相同的图片,导致重建效果较差。对问题a),有研究者提出了使用非朗伯特模型的SFS方法<sup>[3]</sup>。对问题b),一种解决方案是添加辅助约束进行求解,但这类方法实用性较差,难以应用于实际图像。另一种方案是使用透视投影替代正交投影的SFS方法,即PS-FS(perspective shape from shading)方法<sup>[4]</sup>。这类方法在真实图像重建方面表现更好,也是目前普遍采用的方法。综合以上两个问题,Vogel等人<sup>[5]</sup>提出了使用基于非朗伯特模型的PSFS方法对真实图片进行重建,取得了不错的效果。此外,Ecker等人<sup>[6]</sup>还提出一种针对多面体三维重建的多项式SFS方法,避免了解偏微分方程比较困难的问题。

明暗度法的主要优势在于可以从单幅图像中恢复出比较精确的三维模型,可以应用于除镜面物体之外的几乎所有类型的物体。但是明暗度法重建单纯依赖数学运算,效果不佳,而且由于对光照条件要求比较苛刻,需要精确知道光源的位置及方向等信息,使得明暗度法很难应用在室外场景等光线情况复杂的三维重建上。

## 1.2 光度立体视觉法

虽然明暗度法支持从单幅图像中重建三维模型,但单幅图像中可获取的信息较少,实际重建效果一般。于是Woodham<sup>[7]</sup>对SFS方法进行了改进,提出了光度立体视觉法(photometric stereo)。这种方法的基础是亮度方程:

$$I(x, y) = k(x, y) \times N(x, y) \times S \quad (1)$$

其中: $k$ 是一个由物体表面反射系数、光源强度、摄像机光敏感度共同决定的系数; $N$ 为物体表面法向量; $S$ 为光源向量; $I$ 为图像亮度。

光度立体视觉法通过多个不共线的光源获得物体的多幅图像,再将不同图像的亮度方程联立,求解出物体表面法向量的方向,最终实现物体形状的恢复。从技术上讲,使用两个光源即可得到物体法向信息,但使用多个光源可以利用数据冗余解决由阴影以及镜面反射等因素造成无法求解的问题,鲁棒性较好,重建效果也能够得到提高,因此当前的方法基本上都采用多个(4~6个)光源进行三维重建。

Noakes等人<sup>[8]</sup>阐述了在光度立体视觉法中进行非线性与噪声减除的方法。Horovitz等人<sup>[9]</sup>在方法中引入了梯度场和控制点,提高了三维重建的精度。Tang等人<sup>[10]</sup>使用可信度传递与马尔可夫随机场<sup>[11]</sup>的方法进行优化,提高了重建效果。Basri等人<sup>[12]</sup>实现了光源条件未知情况下的三维重建。Sun等人<sup>[13]</sup>提出了采用非朗伯特模型的方法,避免了朗伯特模型比较理想化的问题。Hernandez等人<sup>[14]</sup>进一步提出了使用彩色光线进行重建的方法。Shi等人<sup>[15]</sup>提出了自标定的光度立体视觉法。此外,针对传统的光度立体视觉法均局限于对刚性物体重建的问题,还有人提出了针对动态表面<sup>[16]</sup>及非刚性不规则物体<sup>[17]</sup>进行三维重建的方法。

光度立体视觉法的优点与明暗度法相同,同时使用多幅图像避免了明暗度法存在的病态问题,而且采用多个光源也增加

了约束条件,提高了方法的精确度和鲁棒性;其缺点也是难以应用于镜面物体以及室外场景物体的三维重建。

## 1.3 纹理法

人类可以通过物体表面纹理在视网膜上投影感知物体的三维形状,因此视觉图像中纹理的梯度信息可以作为恢复物体形状及深度信息的线索。依据这一理论,可以通过分析图像中物体表面重复纹理单元的大小、形状,恢复出物体法向、深度等信息,得到物体的三维几何模型,即纹理恢复形状法(shape from texture, SFT),也就是纹理法。

纹理法的基本理论为:对于一个具有光滑表面并覆盖了重复纹理单元的物体,当被投影在二维图像上时,其上的纹理单元会发生变形,这种变形分为投影变形(projective distortion)和透视收缩变形(foreshortening distortion)。投影变形使离图像平面越远的纹理单元看上去越小,透视收缩变形使与图像平面角度越大的纹理单元看上去越短。由于这两种变形都可以从图像中测量得到,因此就可以分析变形后的纹理单元,反向求取物体表面法向和深度信息,进行三维重建。

Wiktin<sup>[18]</sup>最早提出了一种在正交投影条件下基于后验概率分布的SFT方法,不过这种方法只适用于具有等方向性纹理物体,而且受到正交投影固有二义性的影响,不适用于实际应用。这种通过分析图像纹理特征进行重建的方法被称为图像特征法。这方面最近的研究成果是Warren等人<sup>[19]</sup>给出的,其对Wiktin的方法进行了改进,使用了透视投影模型,消除了二义性,使重建效果有不小的提升。另一类SFT方法是谱分析的方法,即通过频域变换,分析纹理单元的谱信息,恢复物体表面法向,进行三维模型重建。例如,Brown等人<sup>[20]</sup>使用傅里叶变换方法对等方向纹理物体进行了重建,Clerc等人<sup>[21]</sup>使用小波变换进行纹理分析和重建等。由于传统方法只有在纹理单元结构不变的约束下才能发挥作用,Lobay等人<sup>[22]</sup>和Loh等人<sup>[23]</sup>分别给出了在纹理单元结构发生改变情况下基于正交投影和透视投影的重建方法。另一方面,由于SFT方法的本质是模拟人的感知方式,因此在SFT方法中引入生物学知识的方法也可以获得很好的效果。例如,Grossberg等人<sup>[24]</sup>提出了一种神经模型实现的SFT方法,Massotd等人<sup>[25]</sup>提出了一种生物拟真(biologically plausible)的频域分析法进行三维重建等。

纹理法的优势是重建精度和速度都比较高,可以从单幅图像中重建出物体的三维模型,同时对光照和噪声都不敏感,鲁棒性较好。但是纹理法只适用于具有规则纹理的物体,通用性较低,大大限制了实际应用。

## 1.4 轮廓法

轮廓法是轮廓恢复形状法(shape from silhouettes/contours)的简称。这种方法通过多个角度物体的轮廓图像得到物体的三维模型。轮廓法可以分为基于体素、基于视壳和基于锥素三种方法。

1) 基于体素的方法 这种方法将物体所在的三维空间离散化为体素(voxel),再将图像分割为前景(物体)和背景,通过投影测试判定一个体素点的投影是在前景还是背景中。如果一个体素点在不同视角都投影到了前景内,那么就认为它是物体的一部分,将所有这样的体素点集合起来就可以重构出物体的三维形状。这种方法简单且鲁棒性较好,但是由于量子化问题的影响,其在计算的时候需要占用大量(内存)空间。Martin

等人<sup>[26]</sup>首先提出了这种方法。Szeliski<sup>[27]</sup>使用了八叉树的体素存储结构,通过由粗到精的判断模式,极大地提高了建模速度。Tarini 等人<sup>[28]</sup>使用了 Marching Intersection 结构,提高了体素判断的效率。Snow 等人<sup>[29]</sup>使用背景剪除算子,结合全局优化,实现了直接从亮度图像建模的方法,避免了繁琐的图像前背景分割操作。Kuhn 等人<sup>[30]</sup>实现了一种方法,可以对已知环境中的未知物体进行重建。Landabaso 等人<sup>[31]</sup>提出了一种方法,可以解决由于标定不准确等误差导致的体素在不同视角图像中不一致的问题。Haro 等人<sup>[32]</sup>提出了一种可以在物体轮廓噪声较大情况下进行重建的方法。

2) 基于视壳的方法 这种方法使用的是一种被称为视壳 (visual hull)<sup>[33]</sup>的三维表示法。通过将所有视角上的轮廓图像反投影到三维空间中形成三维锥壳,所有三维锥壳相交就形成了视壳。物体的三维模型必然落在视壳内部,当轮廓图像足够多时,就可以近似认为视壳等同于物体三维模型。这种方法的不足是得到的视壳比较粗糙,会出现锯齿化效应,需要进一步处理。Matusik 等人<sup>[34]</sup>提出了一种可以进行实时重建的方法。Forbes 等人<sup>[35]</sup>实现了从未标定图像中生成视壳的方法。由于视壳是一种多面体的最大化估计,对视角较少的光滑物体重建效果较差。Franco 等人<sup>[36]</sup>也提出了一种引入假设(如物标表面光滑)的方法,提高了重建效果。

3) 基于锥素的方法 这种方法是基于视壳方法的引申。通过将三维空间中的视壳的概念引申到二维空间中<sup>[37]</sup>,即运用对极几何将三维空间中的运算变换为二维图像空间中的运算,避免了三维运算比较复杂的问题。Casas 等人<sup>[38]</sup>在此基础上提出了锥素 (conexel) 的概念,并应用四叉树进一步降低了运算量。

轮廓法虽然对输入信息的要求非常苛刻,但是由于其计算复杂度较低,重建效率还是非常高的。轮廓法需要输入摄像机内外参数和图像轮廓信息,因此除了需要进行精确的摄像机标定外,还需要从背景中精确提取物体轮廓。常用的方法有抠色法、背景剪除法等。轮廓法最主要的缺点是无法对物体表面的空洞和凹陷部分进行重建,因此通常应用于对模型细节精度要求不是很高的三维重建中。

## 1.5 调焦法

焦点恢复形状法 (shape from focus), 也称调焦法,是通过分析摄像机焦距、光圈与图像清晰度之间的关系恢复物体深度信息,进而得到三维模型的方法。

物体只有位于摄像机焦距处时才会投影出清晰的图像,因此,通过建立物体到投影中心的距离与图像清晰度之间的关系,就可以恢复出深度信息。由于这种关系是非线性的,且自由度非常大,因此需要多张图像才能进行重建。具体方法又可以分为聚焦法 (shape from focus) 和离焦法 (shape from defocus)。

1) 聚焦法 使摄像机相对于被测点处于聚焦位置,然后根据透镜成像公式可求得被测点相对于摄像机的距离。摄像机偏离聚焦位置会带来测量误差,寻求精确的聚焦位置比较困难。这种方法的关键是如何搜索每一点在各种焦距下的聚焦情况。已有的方法包括菲波拉齐搜索法<sup>[39]</sup>、金字塔结构搜索法<sup>[40]</sup>、拉普拉斯算子搜索法<sup>[41]</sup>以及动态规划搜索法<sup>[42]</sup>等。

2) 离焦法 这种方法不要求摄像机相对于被测点处于聚焦位置,而是根据标定出的离焦模型计算被测点相对于摄像机

的距离。这样就避免了由于寻求精确的聚焦位置而降低测量效率的问题,但离焦模型的准确标定是该方法的主要难点。早期的离焦法采用马尔可夫随机域对物体形状和外观进行重建<sup>[43]</sup>,效果不错,但是运算量非常大。研究情况方面,Lou 等人<sup>[44]</sup>给出了自标定和非标定的离焦重建方法。Pradeep 等人<sup>[45]</sup>提出了一种将聚焦法与离焦法结合的方法,综合利用了两种方法的优势。Mannan 等人<sup>[46]</sup>提出一种使用低通滤波器的重建方法。Hasinoff 等人<sup>[47]</sup>进一步提出共焦立体视觉法,能够对复杂物体进行重建。Liu 等人<sup>[48]</sup>提出了一种基于有向热流动概念的离焦扩散模型,可以有效避免假深度信息。Thelen 等人<sup>[49]</sup>提出了一种综合考虑焦距算子、邻域尺寸和高度差值的方法,获得了比较精细的建模效果。Sahay 等人<sup>[50]</sup>提出了一种去模糊的重建方法,可以实现高分辨率的重建。

调焦法可以使用少量图像计算物体表面的稠密深度信息,且对光源条件要求也比较宽松,重建效果比较精细;不足的是需要不断改变摄像机的焦距、光圈等设置,很难实现自动重建,同时调焦法对纹理复杂物体的重建效果也比较差。

## 1.6 运动法

运动法,即基于运动的建模 (structure from motion, SFM),是通过在多幅未标定图像中检测匹配特征点集,使用数值方法恢复摄像机参数与三维信息的一种方法。

运动法首先在图像中检测需匹配的特征点集,以恢复摄像机之间的位置关系。Harris 等人<sup>[51]</sup>首先提出了角点的定义;Shi 等人<sup>[52]</sup>在此基础上进行了改进,提出了一种效果更好的角度提取方法;而目前使用比较广泛的特征点提取和匹配方法是 SIFT (scale-invariant feature transform)<sup>[53]</sup>方法;在 SIFT 方法的基础上,又有研究者分别提出了 PCA-SIFT (principle component analysis SIFT)<sup>[54]</sup>、GLOH (gradient location-orientation histogram)<sup>[55]</sup>、SURF (speed up robust features)<sup>[56]</sup>等运算速度更快的方法。后来提出的算法虽然在速度上较之 SIFT 方法更快,但在稳定性和准确性方法都有所下降。在对运算速度没有太多要求的情况下,SIFT 方法仍然是最佳选择。

Tomasi 等人<sup>[57]</sup>首先提出了使用因式分解的方法,实现了射影层次的重建;但是这种方法的重建结果存在严重畸变,无法恢复平行、角度、距离等几何约束。Faugeras<sup>[58]</sup>通过引入几何约束实现了仿射、度量和欧式层次的重建;但这种方法只适用有几何约束的物体(如建筑物等),对一般物体重建则需要预先进行摄像机标定。Hartley<sup>[59]</sup>使用了 Kurppa 方程结合 SVD 分解,实现了一种在摄像机内参数不改变情况下的自标定方法。Pollefeys 等人<sup>[60]</sup>进一步给出了摄像机焦距改变情况下的自标定方法。

但是上述方法获得的数据精度不高,仍然需要进一步的优化以提高精度。目前最常用的优化方法是集束约束法 (bundle adjustment, BA)<sup>[61]</sup>。集束约束法采用的是 Levenberg-Marquardt 算法,这一算法也是目前解决非线性最小二乘问题最常用的算法。针对 BA 法运算量大,不适于大规模重建的问题,又有研究者分别提出了 SBA (sparse bundle adjustment)<sup>[62]</sup>等改进方法,可以应用于超大规模的三维重建。由于只计算出了特征点对应空间点的三维坐标,因此还需要通过插值及网格化等步骤才能最终得到三维模型。

运动法对图像的要求非常低,可以采用视频甚至是随意拍摄的图像序列进行三维重建。同时可以使用图像序列在重建

过程中实现摄像机的自标定,省去了预先对摄像机进行标定的步骤。而且由于各种特征点提取和匹配技术的进步,运动法的鲁棒性也极强。运动法的另一个巨大优势是可以对大规模场景进行重建,输入图像数量也可以达到百万级,非常适合自然地形及城市景观等三维重建。运动法的不足主要是运算量比较大,同时由于重建效果依赖特征点的密集程度,对特征点较少的弱纹理场景的重建效果比较一般。

## 2 双目视觉方法

立体视觉法(stereopsis)<sup>[63]</sup>,也称双目视觉方法(binocular vision)<sup>[64]</sup>,是一种将双目视差(binocular disparity)信息转换为深度信息的方法。这种方法使用两台摄像机从两个(通常是左右平行对齐的,也可以是上下竖直对齐的)视点观测同一物体,获取在物体不同视角下的感知图像,通过三角测量的方法将匹配点的视差信息转换为深度。这一重建过程与人类视觉的感知过程相似,比较直观和易于理解。

一般的双目视觉方法都是利用对极几何将问题变换到欧氏几何条件下,然后再使用三角测量的方法估计深度信息。这种方法可以大致分为图像获取、摄像机标定、特征提取与匹配、摄像机校正、立体匹配和三维建模六个步骤。

1) 图像获取 使用两台水平或垂直并列的摄像机获取图像,要求两台摄像机的位置关系尽可能满足前向平行对准,有利于降低摄像机校正时的运算量。同时可以对图像做必要的预处理,如去噪、平滑、增强等。

2) 摄像机标定 对摄像机进行标定,获得畸变向量,使用数学方法消除径向和切线方向上的镜头畸变,获得无畸变图像。同时获得摄像机内参数等信息,为后面计算本征矩阵做准备。

3) 特征提取与匹配 提取两幅图像中的特征点并对其进行匹配,为下一步获得摄像机位置关系并进行立体校正做准备。这一步匹配的特征点要尽可能保证匹配的准确性,数量只要能满足计算下一步计算基础矩阵的需求即可。

4) 摄像机校正 使用得到的匹配特征计算摄像机的基础矩阵,由摄像机内参数可以进一步求得本征矩阵,这样就获得了两台摄像机的位置关系信息,可以利用对极几何将图像校正为满足极线约束的行对准图像。

5) 立体匹配 获得了行对准图像后,匹配特征点的二维搜索变成了沿着极线的一维搜索,不仅节省大量计算,而且可以排除许多虚假匹配的点,以此为基础就可以获得两幅图像中点的稠密立体匹配,并计算出视差图。

6) 三维建模 使用三角测量的方法,由视差计算点的深度值,获得稠密的深度点云(point cloud),对点云进行插值和网格化就可以得到物体的三维模型。

由于这种在欧氏几何条件下的方法对摄像机的标定和校正的要求比较严格,也有一些双目视觉方法采用了在射影几何条件下,利用空间射线求交的方式来计算空间点的三维坐标。这种方法在摄像机标定和校正方面要求比较宽松,降低了运算量,但由于无法获得较准确的稠密立体匹配,其重建效果不如前面介绍的方法。

双目立体视觉法的优点是方法成熟,能够稳定地获得较好的重建效果,实际应用情况优于其他基于视觉的三维重建方法,也逐渐出现在一部分商业化产品上;不足的是运算量仍然偏大,而且在基线距离较大的情况下重建效果明显降低。双目

视觉方法整体上已经比较成熟,目前的研究主要集中在提高重建速度以及实际应用等方面。

## 3 三目视觉方法

双目视觉方法在重建过程中存在的主要问题是:a)图像中重复或者相似特征的存在易引起假目标的产生;b)使用外极线约束时,平行于外极线的边缘容易产生模糊;c)如果基线距离增大,使得遮挡严重,能重建的空间点减少,同时由于视差范围的增大,导致在较大搜索空间内产生错误匹配的可能性也增大。

针对这些问题,又提出了三目视觉方法(trinocular vision)<sup>[65-66]</sup>。其基本思想是通过增加一台摄像机提供额外约束,以此来避免上面提到的双目视觉方法的几个问题。根据摄像机的位置关系,三目视觉可分为直角三角形结构和共线结构两种。

1) 直角三角形结构 三台摄像机的投影中心形成一个等腰直角三角形排列。这种结构的特点是:a)左、右图像的外极线平行于 $X$ 轴,而上、下图像之间的外极线平行于 $Y$ 轴;b)对于每一点,左、右图像所获得的水平偏移等于上、下图像所获得的垂直偏移。这种方法相当于分别从相互垂直的两个方向上进行双目立体视觉深度估计,避免了边缘与外极线平行时的匹配模糊性。但如何综合两个图像对所得到的深度信息来求得更准确和完整的深度图,到目前为止,还未有一个很完善的方法。

2) 共线结构 三台摄像机的投影中心共线,设为沿着 $X$ 轴方向的左、中、右排序。设 $d_L$ 、 $d_C$ 、 $d_R$ 分别为同一空间点在左、中、右三幅图像中沿扫描线的偏移(视差),则搜索区域是一个三维立方体,但考虑到对于任何一点,由 $d_L$ 、 $d_C$ 或 $d_C$ 、 $d_R$ 所确定的深度应相等,即应满足如下约束: $2d_C - d_R - d_L = 0$ 。这就使搜索区域从三维立方体降低到由该约束方程所确定的一个二维空间中。然后可利用两层动态规划法,由全局搜索和局部搜索相结合得到最优匹配。

三目视觉方法可以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,在很多情况下重建效果优于双目视觉方法。但由于增加了一台摄像机,设备结构更加复杂,成本更高,控制上也难以实现,因此实际应用情况并不理想。

## 4 比较与分析

基于视觉的三维重建一直是计算机视觉领域的研究热点,随着光学、电子学以及计算机技术的发展,各种重建方法不断进步,成为生物医学、虚拟现实等领域的关键技术,同时也被广泛地应用于航天遥测、军事侦察等领域。基于视觉的三维重建技术优势在于效果逼真,自动化程度高,可以实现自动或半自动重建,极大地节约了人力并提高了效率。除了文中介绍的方法外,一些从其他角度出发的方法,如基于模型数据库的交互式方法<sup>[67-68]</sup>、多视点立体视觉方法<sup>[69]</sup>等,也都取得了不错的成果。

表1是各种基于视觉的三维重建方法的定性比较。从表中可以看出,除运动法以外,单目视觉方法的复杂度比较低,重建时间满足或接近实时要求;不足的是这些方法都依赖稀疏假设条件,方法通用性差,且易受到某些条件(如光照、纹理等)的影响,重建效果不稳定。运动法不依赖于特定假设和条件,

对输入图像要求不高,通用性比较好,其重建效果与图像数量相关,图像越多重建效果一般来说就越好,但相应的运算量也大大增加。双目视觉方法和三目视觉方法都有比较成熟的方法体系,重建效果稳定,适用范围也比较广,但是设备复杂,控制上难以实现,成本也比较高昂。

表 1 各种重建方法的定性比较

方法	复杂度	自动化程度	重建效果	应用情况	
明暗度法	使用单幅图像重建,空间和时间复杂度都较低,重建时间短,但微分求解的病态问题难以解决	可实现完全自动化	重建效果较差,易受光源影响,鲁棒性较差	对光源要求严格,难以应用于室外环境等光线条件下的重建	
光度立体视觉法	使用少量图像进行重建,空间和时间复杂度都较低,重建时间较短	可实现一定程度的自动化	重建效果较好,受光源影响,鲁棒性较差	对光源要求严格,难以应用于室外环境等光线条件下的重建	
单目视觉方法	纹理法	使用单幅图像重建,空间和时间复杂度都较低,重建时间短	可实现完全自动化	重建效果较差,但对光照、噪声等不敏感,鲁棒性较好	只适用于有规则纹理的物体,应用范围较小
	轮廓法	重建空间复杂度较高,时间复杂度低,重建时间可以满足实时要求	可实现完全自动化	取决于轮廓图像数量,轮廓图像越多重建越精确	无法应用于表面有凹陷、空洞物体的重建
	调焦法	使用少量图像进行重建,时间和空间复杂度较低	难以实现自动化重建	能计算每个点的深度,重建效果比较精细	对纹理信息少的物体重建效果较差
	运动法	空间和时间复杂度都很高,在图像数量多的情况下尤为明显	可实现完全自动化	取决于图像数量,图像数量越多重建效果越好	对图像获取要求比较低,能满足大多数情况的应用需求
	双目视觉方法	时间复杂度较高,重建时间较长,设备结构比较复杂	可实现完全自动化	在基线距离不大的情况下重建效果比较好	能满足大多数情况下的应用需求
三目视觉方法	时间复杂度较高,重建时间较长,设备结构复杂	可实现完全自动化	重建效果比较好	能满足大多数情况下的应用需求	

虽然从目前的实际情况及相关研究来看,双(三)目视觉方法在实际应用中的使用多于单目视觉方法,重建效果在多数情况下也优于单目视觉方法,但是在研究价值方面,单目视觉方法的研究价值更高。主要原因是设备复杂度和成本的问题:双(三)目视觉方法使用多台摄像机,摄像机之间的相对位置需要得到精确的固定,同时还需要保证多台摄像机在拍摄期间的同步性和稳定性,控制上难以实现,成本也非常高昂,基本上只能实现小范围的应用,无法进行推广。而单目视觉方法使用单摄像机拍摄图像,对设备的要求简单,成本低廉,易于实现,在推广应用方面有着巨大的优势。

单目视觉方法目前的主要问题和挑战是:

a) 鲁棒性问题。多数单目视觉方法易受到光线、噪声、模糊等问题的影响,鲁棒性较差。

b) 应用局限性问题。除运动法以外,其他单目视觉方法都依赖某些理想化假设,实际应用时受到很大限制。

c) 运算量问题。运动法目前存在的主要问题是运算量太大,算法运行开销对于个人计算机来说比较大,导致重建时间较长。

d) 相关问题。如摄像机标定、图像分割、图像特征提取与匹配等问题一直是计算机视觉领域中比较难以解决的问题,这些问题无法得到根本性解决,也是单目视觉算法易用性及重建效果等问题无法得到更大提高的一个重要原因。

针对以上这些问题,在未来一段时间内,基于单目视觉的三维重建方法的相关研究可以从以下几个方面展开:

a) 完善方法模型。从实际情况出发,对各种方法模型进行完善,使之更加符合真实情况,实现实际应用。

b) 混合方法。充分利用图像中包含的各种信息,使用不

同方法优势互补,构造混合式方法,提高建模效果及方法的通用性。如文献[70~72]提出的一些混合方法都获得了不错的重建效果。

c) 使用分布式计算。针对运算量过大的问题,采用计算机集群计算、网络云计算以及 GPU 计算等方式提高运行速度,缩短三维重建的时间。

d) 相关分支问题研究。如更加精确的摄像机标定方法、更加有效的图像分割方法、更加快速准确的图像特征提取与匹配算法等,这些问题对于提高三维重建的效果和速度都有着非常重大的意义。

除此之外,还有一些新的应用方向也将成为未来一段时间的研究热点:

a) 非刚性及动态物体的三维重建。传统三维重建方法针对的基本都是刚性物体,对非刚性物体(如液态物体、火焰)和动态物体等的重建问题一直无法较好地实现,因此这方面问题将会成为未来的一个研究热点。

b) 大规模场景的三维重建。对于自然地形(包括月表和火星地形)和城市景观等大规模场景的三维重建问题一直是一个比较困难的问题,如何使用卫星或航拍图片对这些场景进行快速准确的重建,也是一个非常有价值的研究方向。

c) 微小场景和微观物体的三维重建。如医学上使用内窥镜进行观察判定病情很大程度上依赖医生的经验,如果能够使用微型摄像机拍摄图像对之进行三维建模,将更清晰地呈现出人体内部情况,有助于医生诊断。此外,使用电子显微镜对生物大分子等的重建,加强对相关领域的研究也是非常有价值的。

d) 识别与理解。结合人工智能等技术,使用重建得到的三维模型进一步进行模式识别等方面的应用。

5 结束语

基于视觉的三维重建技术研究仍然处于探索阶段,各种方法距离实际应用还有一段距离,各种应用需求亟待被满足。因此,在未来很长的一段时间内,还需要在这一领域做更加深入细致的研究。

参考文献:

[1] HORN B. Shape from shading: a method for obtaining the shape of a smooth opaque object from one view [D]. Cambridge [s. n.], 1970.

[2] BELHUMEUR P, KRIEGMAN D, YUILLE A. The bas-relief ambiguity [J]. International Journal of Computer Vision, 1999, 35(1): 33-44.

[3] BAKSHI S, YANG Y. Shape from shading for non-lambertian surfaces [C]//Proc of International Conference on Image Processing. 1994: 130-134.

[4] PENNA M. A shape from shading analysis for a single perspective image of a polyhedron [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1989, 11(6): 545-554.

[5] VOGEL O, BREUB M, WEICKERT J. Perspective shape from shading with non-lambertian reflectance [C]//Proc of DAGM Symposium on Pattern Recognition. Berlin: Springer, 2008: 517-526.

[6] ECKER A, JEPSON A D. Polynomial shape from shading [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010.

[7] WOODHAM R J. Photometric method for determining surface orientation from multiple images [J]. Optical Engineering, 1980, 19(1):

- 139-144.
- [8] NOAKES L, KOZERA R. Nonlinearities and noise reduction in 3-source photometric stereo [J]. *Journal of Mathematical Imaging and Vision* 2003, 18(2): 119-127.
- [9] HOROVITZ I, KIRYATI N. Depth from gradient fields and control points: bias correction in photometric stereo [J]. *Image and Vision Computing* 2004, 22(9): 681-694.
- [10] TANG L K, TANG C K, WANG T T. Dense photometric stereo using tensorial belief propagation [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: [s. n. ] 2005: 132-139.
- [11] TANG L K, TANG C K, WANG T T. A Markov random field approach for dense photometric stereo [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005: 11-19.
- [12] BASRI R, JACOBS D W, KEMELMACHER I. Photometric stereo with general, unknown lighting [J]. *Journal of Mathematical Imaging and Vision* 2007, 72(3): 239-257.
- [13] SUN Jiu-ai, SMITH M L, SMITH L N *et al.* Object surface recovery using a multi-light photometric stereo technique for non-lambertian surfaces subject to shadows and specularities [J]. *Image and Vision Computing* 2007, 25(7): 1050-1057.
- [14] HERNANDEZ C, VOGIATZIS G, BROSTOW G J *et al.* Non-rigid photometric stereo with colored lights [C]//Proc of International Conference on Computer Vision. 2007: 1-8.
- [15] SHI Bo-xin, MATSUSHITA Y, WEI Yi-chen *et al.* Self-calibrating photometric stereo [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010.
- [16] KIM H, WILBURN B, BEN-EZRA M. Photometric stereo for dynamic surface orientations [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010.
- [17] HIGO T, MATSUSHITA Y, IKEUCHI K. Consensus photometric stereo [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Grete [s. n. ] 2010.
- [18] WIKTIN A. Recovering surface shape and orientation from texture [J]. *Artificial Intelligence*, 1981, 17(1/2/3): 17-45.
- [19] WARREN P A, MAMASSIAN P. Recovery of surface pose from texture orientation statistics under perspective projection [J]. *Biological Cybernetics* 2010, 103(3): 199-212.
- [20] BROWN R, SHVAYSTER H. Surface orientation from projective foreshortening of isotropic texture autocorrelation [J]. *IEEE Trans on PAMI*, 1990, 12(6): 584-588.
- [21] CLERC M, MALLAT S. Texture gradient equation for recovering shape from texture [J]. *IEEE Trans on PAMI* 2002, 24(4): 536-549.
- [22] LOBAY A, FORSYTH D. Shape from texture without boundaries [J]. *Journal of Mathematical Imaging and Vision* 2006, 67(1): 71-91.
- [23] LOH A M, HARTLEY R. Shape from non-homogeneous, nonstationary anisotropic perspective texture [C]//Proc of British Machine Vision Conference. Oxford, UK [s. n. ] 2005: 1-8.
- [24] GROSSBERG S, KUHLMANN L, MINGOLLA E. A neural model of 3D shape-from-texture: multiple-scaleclustering, boundary grouping, and surface filling-in [J]. *Vision Research* 2007, 47(5): 634-672.
- [25] MASSOTD C, HERAULT J. Model of frequency analysis in the visual cortex and the shape from texture problem [J]. *Journal of Mathematical Imaging and Vision* 2008, 76(2): 165-182.
- [26] MARTIN W N, AGGARWAL J K. Volumetric descriptions of objects from multiple views [J]. *IEEE Trans on PAMI*, 1983, 5(2): 150-158.
- [27] SZELISKI R. Rapid octree construction from image sequences [J]. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 1993, 58(1): 23-32.
- [28] TARINI M, CALLIERI M, MONTAN C *et al.* Marching intersections: an efficient approach to shape from silhouette [C]//Proc of Vision, Modeling and Visualization Conference. Erlangen [s. n. ] 2002: 10-15.
- [29] SNOW D, VIOLA P, ZABIH R. Exact voxel occupancy with graph cuts [C]//Proc of Eurographics Workshop on Rendering. London, UK: [s. n. ] 2001: 8-12.
- [30] KUHN S, HENRICH D. Multi-view reconstruction of unknown objects within a known environment [C]//Proc of International Symposium on Visual Computing. 2009.
- [31] LANDABASO J L, PARDAS M, CASAS J R. Shape from inconsistent silhouette [J]. *Computer Vision and Image Understanding* 2008, 112(2): 210-224.
- [32] HARO G, PARDAS M. Shape from incomplete silhouettes based on the reprojection error [J]. *Image and Vision Computing* 2010, 28(9): 1354-1368.
- [33] LAURENTINI A. The visual hull concept for silhouette-based image understanding [J]. *IEEE Trans on PAMI*, 1994, 16(2): 150-162.
- [34] MATUSIK W, BUEHLER C, MCMILLAN L. Polyhedral visual hulls for real-time rendering [C]//Proc of Eurographics Workshop on Rendering. London, UK [s. n. ] 2001: 115-125.
- [35] FORBES K, NICOLLS F, JAGER G *et al.* Shape-from-silhouette with two mirrors and an uncalibrated camera [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2006: 1-8.
- [36] FRANCO J S, LAPIERRE M, BOYER E. Visual shape of silhouette sets [C]//Proc of International Symposium on 3D Data Processing, Visualization and Transmission. 2006: 397-404.
- [37] MATUSIK W, BUEHLER C, MCMILLAN L. Polyhedral visual hulls for real-time rendering [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2000: 345-353.
- [38] CASAS J R, SALVADOR J. Image-based multi-view scene analysis using conexels [C]//Proc of HCSNet Workshop on Use of Vision in Human-computer Interaction. Graz [s. n. ] 2006.
- [39] XIONG Ya-lin, SHAFFER S A. Depth from focusing and defocusing [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1993: 68-73.
- [40] DARRELL T, WOHN K. Depth from focus using a pyramid architecture [J]. *Pattern Recognition Letters*, 1990, 11(12): 787-796.
- [41] NAYAR S, NAKAGAWA Y. Shape from focus [J]. *IEEE Trans on PAMI*, 1994, 16(8): 824-831.
- [42] YUN J, CHOI T S. Fast shape from focus using dynamic programming [C]//Proc of Three-Dimensional Image Capture and Applications. San José [s. n. ] 2000: 71-79.
- [43] RAJAGOPALAN A, CHAUDHURI S. Optimal selection of camera parameters for recovery of depth from defocused images [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1997: 219-224.
- [44] LOU Yi-fei, FAVARO F, BERTOZZI A L *et al.* Autocalibration and uncalibrated reconstruction of shape from defocus [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2007: 1-8.
- [45] PRADEEP K, RAJAGOPALAN A. Improving shape from focus using defocus cue [J]. *IEEE Trans on Image Processing* 2007, 16(7): 1920-1925.
- [46] MANNAN S M, MUTAHIRA H, CHOI T S. Shape recovery by focus based methods using low pass filter [J]. *Communication in Computer and Information Science* 2009, 56: 528-538.

- [47] HASINOFF S W ,KUTULAKOS K N. Control stereo[J]. *Journal of Mathematical Image and Vision* 2009 **81**( 1) : 82-104.
- [48] LIU Hong ,JIA Yi ,CHENG Hong *et al.* Depth estimation from defocus images based on oriented heat-flows[C]//Proc of the 2nd International Conference on Machine Vision. 2009: 212-215.
- [49] THELEN A ,FEERY S ,HIRSCH S *et al.* Improvements in shape-from-focus for holographic reconstructions with regard to focus operators , neighborhood-size , and height value interpolation [J]. *IEEE Trans on Image Processing* 2009 **18**( 1) : 151-157.
- [50] SAHAY R R ,RAJAGOPALAN A. Harnessing defocus blur to recover high-resolution information in shape-from-focus technique [J]. *Computer Vision* 2008 **2**( 2) : 50-59.
- [51] HARRIS C ,STEPHENS M. A combined corner and edge detector [C]//Proc of the 4th Alvey Vision Conference. Cambridge [s. n. ] , 1988: 147-151.
- [52] SHI Jian-bo ,TOMASI C. Good features to track [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Seattle [s. n. ] ,1994.
- [53] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. *Journal of Mathematical Image and Vision* ,2004 **60**( 2) : 91-110.
- [54] KE Yan ,SUKTHANKAR R. PCA-SIFT: a more distinctive representation for local image descriptors [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC [s. n. ] ,2004.
- [55] MIKOLAJCZYK K ,SCHMID C. A performance evaluation of local descriptors [J]. *IEEE Trans on PAMI* 2005 **27**( 10) : 31-47.
- [56] BAY H ,TUYTELAARS T ,Van Gool L *et al.* SURF: speeded up robust features [J]. *Computer Vision and Image Understanding* , 2008 **110**: 346-359.
- [57] TOMASI C ,KANADE T. Shape and motion from image streams under orthography: a factorization method [J]. *Journal of Mathematical Image and Vision* ,1992 **9**( 2) : 137-154.
- [58] FAUGERAS O. What can be seen in three dimensions with an uncalibrated stereo rig [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 1992: 563-578.
- [59] HARTLEY R. Kruppa's equations derived from the fundamental matrix [J]. *IEEE Trans on PAMI* ,1997 **19**( 2) : 133-135.
- [60] POLLEFEYES M ,KOCH R ,GOOL L V. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters [J]. *Journal of Mathematical Image and Vision* ,1999 **32**( 1) : 7-25.
- [61] TRIGGS B ,McLAUCHLAN P F ,HARTLEY R I *et al.* Bundle adjustment: a modern synthesis [C]//Proc of International Workshop on Vision Algorithms: Theory and Practice. 2000: 298-375.
- [62] LOURAKIS M ,ARGYROS A. SBA: a software package for generic sparse bundle adjustment [J]. *ACM Trans on Mathematical Software* 2009 **36**( 1) : 1-30.
- [63] AKIMOTO T ,SUENAGA Y ,WALLACE R. Automatic creation of 3D facial models [J]. *IEEE Computer Graphics & Applications* , 1993 **13**( 5) : 16-22 .
- [64] CHEN C L ,TAI C L ,LIO Y F. Visual binocular vision systems to solid model reconstruction [J]. *The International Journal of Advanced Manufacturing Technology* 2007 **35**( 3-4) : 379-384.
- [65] STEWART C V ,DYER C R. The trinocular general support algorithm: a three-camera stereo algorithm for overcoming binocular matching errors [C]//Proc of the 2nd International Conference on Computer Vision. 1988: 134-138.
- [66] 李智秀 ,张广军. 一种基于边缘线的三目立体视觉匹配方法 [J]. *光电工程* 2007 **34**( 2) : 22-26.
- [67] BLANZ V ,VETTER T. A morphable model for the synthesis of 3D faces [C]//Proc of SIGGRAPH. Los Angeles [s. n. ] ,1999: 187-194.
- [68] HASSNER T ,BASRI R. Example based 3D reconstruction from single 2D images [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York [s. n. ] 2006.
- [69] SEITZ S M ,CURLESS B ,DIEBEL J *et al.* A comparison and evaluation of multi-view stereo reconstruction algorithms [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006.
- [70] HERNANDEZ C ,VOGIATZIS G ,CIPOLLA R. Multi-view photometric stereo [J]. *IEEE Trans on PAMI* 2008 **30**( 3) : 548-554.
- [71] SINHA S N ,STEEDLY D ,SZELISKI R *et al.* Interactive 3D architectural modeling from unordered photo collections [J]. *ACM Trans on Graph* 2008 **27**( 5) : 1-10.
- [72] SALZMANN M ,FUA R. Reconstructing sharply folding surfaces: a convex formulation [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2009: 125-134.
- ( 上接第 2410 页)
- [15] SHIH K P ,WANG San-yuan ,CHEN H C *et al.* On target coverage in wireless heterogeneous sensor networks with multiple sensing units [C]//Proc of the 12th IEEE Symposium on Computers and Communications. 2007: 185-190.
- [16] CAMPBELL A ,EISENMAN S ,LANE N *et al.* People-centric urban sensing [C]//Proc of the 2nd Annual International Wireless Internet Conference. [S. l. ]: IEEE Computer Society 2006: 2-5.
- [17] CORNELIUS C ,KAPADIA A ,KOTZ D *et al.* AnonySense: privacy-aware people-centric sensing [C]//Proc of MobiSys'08. 2008.
- [18] ZOU Yi ,CHAKRABARTY K. Sensor deployment and target localization based on virtual forces [C]//Proc of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies. 2003.
- [19] TAN Guang ,JARVIS S A ,KERMARREC A M. Connectivity-guaranteed and Obstacle-adaptive deployment schemes for mobile sensor networks [J]. *IEEE Trans on Mobile Computing* 2009 **8**( 6) : 836-848.
- [20] JUANG P ,OKI H ,WANG Yong *et al.* Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with ZebraNet [J]. *ACM SIGOPS Operating Systems Review* ,2002 **36**( 5) : 96-107.
- [21] 苏金树 ,胡乔林 ,赵宝康 等. 容延容断网络路由技术 [J]. *软件学报* 2010 **21**( 1) : 119-132.
- [22] BURLEIGH S ,HOOKE A ,TORGERSON L *et al.* Delay-tolerant networking: an approach to interplanetary Internet [J]. *IEEE Communications Magazine* 2003 **41**( 6) : 128-136.
- [23] Delay Tolerant Networking Research Group [EB/OL]. ( 2009) . <http://www.dtnrg.org>.
- [24] Disruption tolerant networking ( DTN) [EB/OL]. ( 2008) . <http://www.darpa.mil/sto/solicitations/DTN/index.htm>.
- [25] BISWAS S ,MORRIS R. ExOR opportunistic routing in multi-hop wireless networks [C]//Proc of ACM SIGCOMM. 2006.
- [26] DUBOIS-FERRIERE H. Anypath routing [D]. [S. l. ]: EPFL 2006.
- [27] LUKOSIUS A. Opportunistic routing in multi-sink mobile Ad hoc wireless sensor networks [D]. [S. l. ]: University Bremen 2007.