

·信号与信息处理·

基于计算机视觉的深度估计方法

江 静^{1,2}, 张雪松³

(1.中国矿业大学(北京)煤炭资源与安全开采国家重点实验室,北京 100083;
2.华北科技学院机电工程系,北京 101601;3.光电信息控制和安全技术重点实验室,河北 三河 065201)

摘 要:深度估计方法可分为主动视觉和被动视觉两类。介绍了被动视觉测距中的五种深度估计方法,包括立体视觉、运动测距、描影法、聚焦法和散焦法,分析了各自的成像原理、数学模型和测量方法。最后,对基于计算机视觉的深度估计方法进行总结,分析了现有测量方法各自的局限性,提出算法进一步改进的方向是应用计算机视觉系统中需要的不同模板,联合依赖于实际环境的权重因子与被动深度估计算法相结合。这是当前计算机视觉研究的一个热点领域。

关键词:计算机视觉;深度估计;被动测距;双目视觉;单目视觉

中图分类号:TP391.4

文献标识码:A

文章编号:1673-1255(2011)01-0051-05

Depth Estimation Methods Based on Computer Vision

JIANG Jing^{1,2}, ZHANG Xue-song³

(1. State Key Laboratory of Coal Resources and Safe Mining, China University of Mining and Technology(Beijing), Beijing 100083, China; 2. Department of Mechanics and Electricity Engineering, North China Institute of Science and Technology, Beijing 101601, China; 3. Science and Technology on Electro-optical Information Security Control Laboratory, Sanhe 065201, China)

Abstract: The depth estimation methods can be divided into two categories, active vision and passive vision. The five different passive vision range finding methods are introduced, including stereo vision, structure from motion, shape from shading, depth from focus and depth from defocus, and their imaging principles, mathematical models and measurement methods are also analyzed. Finally, the depth estimation methods based on computer vision are summarized, and the limitation of the current measurement methods is analyzed. The further direction of improvement algorithm is using the different modules in the computer vision system. The association depends on combining the different weighting factors in the actual environment with passive depth estimation methods. This is a hot point in the present computer vision.

Key words: computer vision; depth estimation; passive ranging; binoculars stereo; monocular vision

计算机视觉是一门研究如何使机器实现人类视觉功能的科学,通过摄像机和计算机代替人眼对目标进行识别、跟踪和测量,也称为机器视觉。计算机视觉中的一个基本问题是研究如何从景物的二维图像来获得三维世界的结构和属性等信息^[1-4]。因此,从景物的一幅或多幅图像中自动、有效地提取深度信息,

即深度估计,是计算机视觉中的一项重要研究内容。目前,基于计算机视觉的深度估计方法,根据成像光源不同可分为两大类:主动视觉(active vision, AV)法与被动视觉(passive vision, PV)法。主动视觉是指向被测物体发射可控制的光束,然后拍摄光束在物体表面上所形成的图像,通过几何关系计算出被测物体距

收稿日期:2010-11-26

基金项目:国家自然科学基金(51074169)

作者简介:江静(1979-),女,江苏连云港人,博士,研究方向为机器视觉、图像处理与模式识别;张雪松(1977-),男,河北香河人,博士,CCF会员,高级工程师,研究方向为数字图像复原与超分辨率、模式识别和图像实时处理。

离的方法^[5-8]。被动式方法一般不采用特殊光源进行照明,仅从一个或多个摄像系统获取的二维图像信息中确定空间信息,形成三维轮廓数据。文中将介绍被动视觉深度估计的成像原理、数学模型和测量方法。

被动视觉属于被动传感的一种,所需景物的照明是靠环境提供的,包括双目立体视觉、运动视觉、描影法、聚焦法和散焦法。在距离的测量与3D景物的恢复中,双目立体视觉是最为常用的方法^[9]。它是利用两台摄像机得到同一目标的两幅图像的视差来计算深度。为了计算视差,就必须进行特征点提取和对应点匹配。立体视觉的距离估计精度较高,但是两幅图像中特征点的鲁棒匹配是个难题,而且计算量大。运动视觉是研究如何从变化着的场景的一系列不同时刻的图像中,提取出有关场景中的物体的形状、位置和运动的信息,这在计算机视觉的理论上与实践上都有重要的作用^[10]。对于具有同一反射特性的光滑表面物体,图像的灰度分布主要受照射光线、表面形状、表面材料的反射率和成像投影方向的影响。描影法就是要从图像灰度中抽取物体的形状信息,主要是利用反射图与物体表面每点灰度的方向有关^[11],可以利用双向反射分布函数和光源进行计算。焦距一直被认为是获得深度信息的主要方法之一。当图像被正确聚焦,摄像机参数可用来估计目标点的深度。通过不断改变像检测器到镜头的距离可得到场景的一系列图像。对于不同的焦距设置,场景中的点将随每帧图像焦距设置的不同,而呈现不同程度的模糊,对于图像序列中的每一帧,利用邻域窗计算每个像素的聚焦尺度。使用有最大清晰度像素点的那帧图像的摄像机参数值来估计该点的深度值^[12]。该方法不存在立体视觉中的匹配问题,主要缺点是要花费大量的时间获得图像。基于图像散焦的深度测量是利用景深有限的区域图像的散焦进行深度估计的。散焦程度是镜头的参数与景深的函数。因此,假定摄像机参数已知,如果图像的散焦量得到,则深度值可得到,这就是基于散焦图像深度估计的基本原理。该方法首先被Pentland提出^[13]。这种方法的优点是绝对深度可通过两幅图像得到,克服了图像间的匹配问题与遮挡问题,而这些问题是立体视觉与运动视觉的主要问题。

1 立体视觉

双目立体视觉(depth from stereo, DFS)是人类

获取距离信息的主要方式,属于被动式的视觉方法,该方法是研究如何利用二维投影图像来恢复三维景物世界,即由不同位置的两台或者一台摄像机(CCD)经过移动或旋转拍摄同一幅场景,通过计算空间点在两幅图像中的视差,根据立体视差获得该点的三维坐标值。所谓立体视差就是目标点在左、右摄像机像平面上成像点位置的差异。

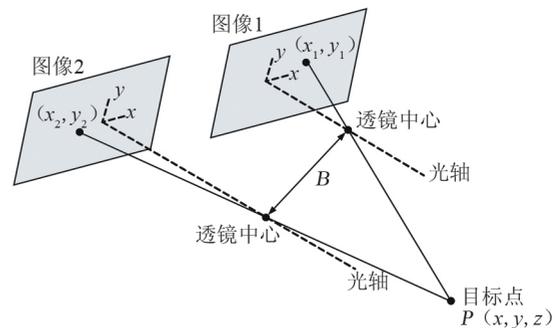


图1 双目立体视觉成像原理图

双目立体视觉成像系统如图1所示。左右摄像机成像平面中心的连线确定为 x 轴, y 轴垂直于 x 轴, z 轴是光轴。原点设置在图像1和图像2的平面中心,两相机透镜中心点的距离为 B 。空间被测目标点 P 的坐标为 (x, y, z) ,点 P 在图像1和图像2的图像坐标分别为 $(x_1, y_1), (x_2, y_2)$,同一目标点的两幅图像取自不同的位置,假设相机的焦距是 f ,则该目标点在两幅图像中的立体视差为 d ,进而获得距离目标点的深度^[14]。由

$$\begin{cases} \frac{x_1}{f} = \frac{x + B/2}{z} \\ \frac{x_2}{f} = \frac{x - B/2}{z} \\ \frac{y_1}{f} = \frac{y_2}{f} = \frac{y}{z} \end{cases} \quad (1)$$

$$d = x_1 - x_2 \quad (2)$$

化简公式可得

$$z = \frac{f(B-d)}{d} \quad (3)$$

由此可见,图像间立体视差反比于相机测量的深度,说明这种方法只有在近距离时,测量才较准确,而远距离时测量精度就难以保证。双目立体视觉成像系统通常包括相机建模、特征提取、图像匹配、深度计算以及图像插值运算,具体测量过程分

为如下几步:

- (1)从图像1(或图像2)中选出某些特征点;
- (2)找出该特征点在图像2(或图像1)中的匹配点,一般称为立体匹配;
- (3)根据以上点对的匹配关系,计算匹配点在摄像机坐标系中的空间三维坐标;
- (4)由于在(3)步中只能计算出匹配点的三维坐标,因此其他点则需要用插值的方法来获得,实现三维重建。

立体匹配问题始终是双目视觉测量的一个主要难点,国内外众多学者对此进行深入而持久的研究,提出了大量的匹配算法并进行了实验验证。例如利用外极线约束、兼容性约束、唯一性约束、连续性约束、形状连续性约束、偏差梯度约束等约束条件减小匹配搜索范围和确定正确对应关系。

2 运动视觉

运用被测对象与相机之间的相对运动恢复三维表面信息的技术被称为运动视觉技术。运动视觉是基于因子分解的运动估计结构(structure from motion, SFM),主要研究一系列不同时刻变化着的场景图像,提取有关场景中物体的形状、位置和运动的信息。主要包括两个步骤:一是从不同时刻相邻的两幅或多幅图像中抽取特征点,并建立对应关系;二是根据这些特征点之间的函数关系,计算物体的结构和运动。运动视觉成像系统如图2所示。

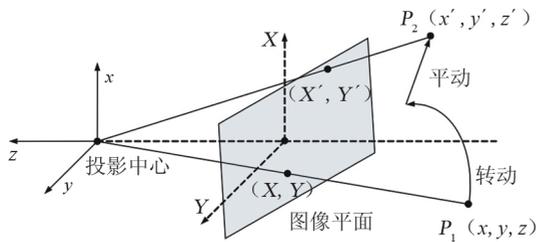


图2 运动视觉成像原理图

空间中的任意一被测目标点 $P_1(x, y, z)$,在三维空间经过水平运动、垂直运动以及旋转运动到点 $P_2(x', y', z')$,点 (X, Y) 和 (X', Y') 为 $P_1(x, y, z)$ 和 $P_2(x', y', z')$ 在图像平面的成像点,得到函数关系

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \mathbf{R} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \mathbf{T} \quad (4)$$

式中, $\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$ 为旋转矩阵; $\mathbf{T} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix}$ 为平移向量。

任意选择时刻 t_1, t_2 ,则

$$\begin{cases} x' = x + \Delta x \\ y' = y + \Delta y \end{cases} \quad (5)$$

$$\begin{cases} x = f \frac{X}{Z} \\ y = f \frac{Y}{Z} \end{cases} \quad (6)$$

式中, f 为相机成像的焦距。由式(4)~式(6),令 $\Delta X, \Delta Y, \Delta Z$ 为位移矢量 \mathbf{T} 的增量,则

$$\Delta Z' = \frac{\Delta Z}{f}; \Delta X' = \frac{\Delta X}{\Delta Z}; \Delta Y' = \frac{\Delta Y}{\Delta Z}; \bar{Z} = \frac{Z}{\Delta Z} \quad (7)$$

代入旋转矩阵 \mathbf{R} ,可得

$$X' = \frac{(-r_{11}X - r_{12}Y + r_{13})z + \Delta x}{(-r_{31}X - r_{32}Y + r_{33})z + \Delta z} \quad (8)$$

$$Y' = \frac{(-r_{21}X - r_{22}Y + r_{23})z + \Delta y}{(-r_{31}X - r_{32}Y + r_{33})z + \Delta z} \quad (9)$$

消去变量 z ,可得

$$\frac{\Delta x + X' \Delta z}{\Delta y + Y' \Delta z} = \frac{X'(r_{31}X + r_{32}Y - r_{33}) + (r_{11}X + r_{12}Y - r_{13})}{Y'(r_{31}X + r_{32}Y - r_{33})z + (r_{21}X + r_{22}Y - r_{23})} \quad (10)$$

其中,方程(10)含有5个未知量,任意选中5个特征点对,就可确定深度 z

$$z = \frac{\Delta x + X' \Delta z}{X'(r_{31}X + r_{32}Y - r_{33}) + (r_{11}X + r_{12}Y - r_{13})} \quad (11)$$

3 描影法

描影法是基于单一图像灰度信息恢复表面三维轮廓(shape from shading, SFS)的方法,即从明暗恢复形状。主要原理是利用单幅灰度图像中明暗变化,并根据对物体的局部或整体形状的约束条件,来获取其表面各点的相对高度或表面法向量等参数值。

描影法涉及从单一图像中提取物体表面辉度方向的问题,三维目标成像取决于物体的外形、表面反射特性以及光源的空间分布。物体表面的反射特性可以通过光源的双向反射分布函数来确定,

综合上面3个多项式,得到模糊圆的半径和目标点在镜头前的位置无关。

$$r_{b1} = r_{b2} = \frac{r_0(v_0 - v)}{v} \quad (15)$$

根据成像法则: $\frac{1}{f} = \frac{1}{v} + \frac{1}{u}$, 代入式(15), 得

$$r_b = r_0 \left(\frac{v_0}{v} - 1 \right) = r_0 v_0 \left(\frac{1}{f} - \frac{1}{u} - \frac{1}{v_0} \right) \quad (16)$$

因此,被测目标点距离镜头的深度为

$$u = \frac{fv_0}{v_0 - f - \frac{r_b}{r_0}f} \quad (17)$$

散焦法不要求摄像机相对于被测点处于聚焦位置,而是根据标定出的散焦模型计算被测点相对于摄像机距离。这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,但是散焦模型的精确标定是该方法的主要难点。实际进行深度估计中,由于聚焦测距需要拍摄大量图像,且计算量大;散焦测距只要拍摄一个或两个图像,速度快,但是测量精度没有聚焦测距好,通常把两者结合起来使用。首先通过散焦测距对场景深度作一个粗略计算,再通过聚焦测距得到精度较高的3D场景深度值。

6 结论

经过几十年的研究,国内外学者们已提出许多深度估计的算法模型,但是仍然具有很强的局限性:(1)双目立体视觉能够测出较准的绝对深度,但要需要耗费大量的时间来解决匹配问题。另外这种方法只能对那些具有可匹配的特征的地方才能计算深度;(2)运动视觉大都是在假定图像序列间的配准问题已经解决的前提下进行的。此外,运动视觉要求解系统的非线性方程,并且解的收敛性又依赖于系统方程的初始值,故鲁棒性较差;(3)景物深度的测量和目标跟踪主要采用立体视觉,该方法需要特征点的选取与匹配,计算复杂,计算量大;(4)聚焦法的深度估计计算量大而且需要大量的图像。并且实际的摄像机系统并不是针孔摄像机,而是有一定孔径大小的摄像机镜头。目标物体如果偏离聚焦距离就会在摄像机的像检测器上形成散焦图像,且偏离距离不同,图像的散焦程度也不同;(5)散焦法需图像的分辨率较高,要求图像中的频谱内容丰富。散焦图像中隐含着景物的深度

信息。但是对于复杂的自然景物由于图像点和点的形成参数有很大的任意性,故需要算法的改进。

各种各样的测距方法中,没有一种方法明显的优于其他方法,它们都有明显的缺点与适用条件,需要研究人员不断地提高和改进。应用计算机视觉系统中需要的不同模板,联合依赖于实际环境的权重因子与被动深度估计算法相结合。这是当前计算机视觉研究的一个热点领域。随着对新的测距原理、求解方法的不断探索和创新,计算机视觉的深度估计将进入一个更加繁荣的时代。

参考文献

- [1] 贾云得. 机器视觉[M]. 北京: 科学出版社, 2000.
- [2] B K P Hom, B G Schunck. Determining optical flow, Artificial Intelligence[J]. 1981,17:185-203.
- [3] B D Lucas, T Kanade. An iterative image registration technique with an application to stereovision[C] //Proc. Of the International Joint Conf. On Artificial Intelligence, 1981.
- [4] E P Simoneelli, E H Adelson, D J Heeger. Probability distribution of optical flow[C] //IEEE Conference of Computer Vision and Pattern Recognition, 1991:310-315.
- [5] Asada M, Ichikawa H, Tsuji S. Determining Surface Orientation by Projecting a Stripe Pattern [J]. IEEE Trans. PAMI, 1988, 10(5):749-754.
- [6] Wang Y F, Mitiche A, Aggarwal J K. Computation of Surface Orientation and Structure of Objects Using Grid Coding [J]. IEEE Trans. PAMI, 1987, 9(1):129-137.
- [7] 郑南宁. 计算机视觉与模式识别[M]. 北京: 国防工业出版社, 1998.
- [8] 金声, 李允明. 三维计算机视觉中距离的获取[J]. 机器人, 1989, 3(1):57-64.
- [9] 吴立德. 计算机视觉[M]. 上海: 复旦大学出版社, 1993.
- [10] A 罗森菲尔德. 数字图像分析[M]. 北京: 科学出版社, 1987.
- [11] 李介谷. 计算机视觉的理论与实践[M]. 2版. 上海: 上海交通大学出版社, 1999.
- [12] S K Nayar. Shape from focus system [C] //Proc. IEEE CVPR, 1992: 302-308.
- [13] A P Pentland. A new sense for depth of field [J]. IEEE Trans. PA MI, 1987, 9:523-531.
- [14] J L C Sanz. Advances in machine vision[M]. Springer-Verlag, New York, 1989.