

计算机视觉

Introduction

1. 引言

- **计算机视觉**是人工智能一个重要分支，视觉感知是智能机器的核心功能
- **智能机器**：能模拟人类的功能，能感知外部世界并有效地解决人所能解决问题
- **感知系统**：人类感知外部世界主要是通过视觉、触觉、听觉和嗅觉等感觉器官，其中约80%的信息是由视觉获取的。因此,对于智能机器来说，赋予机器以人类视觉功能对发展智能机器是及其重要的，也由此形成了一门新的学科—计算机视觉(也称计算机视觉或图像分析与理解等)。计算机视觉的发展不仅将大大推动智能系统的发展，也将拓宽计算机与各种智能机器的研究范围和应用领域
- **计算机视觉**：研究用计算机来模拟生物外显或宏观视觉功能的科学和技术。计算机视觉系统的首要目标是用图像创建或恢复现实世界模型，然后认知现实世界

2. 计算机视觉发展

- **20世纪50年代**：主要集中在二维图像分析和识别上，如，光学字符识别，工件表面、显微图片和航空图片的分析和解释等。是模式识别的重要内容
- **60年代**MIT 的Roberts通过计算机程序从数字图像中提取出诸如立方体、楔形体、棱柱体等多面体的三维结构，并对物体形状及物体的空间关系进行描述。Roberts的研究工作开创了以理解三维场景为目的的三维计算机视觉的研究。Roberts对积木世界的创造性研究给人们以极大的启发，许多人相信，一旦由白色积木玩具组成的三维世界可以被理解，则可以推广到理解更复杂的三维场景
- **70年代**，已经出现了一些视觉应用系统。70年代中期，麻省理工学院(MIT)人工智能(AI)实验室正式开设“计算机视觉”(Machine Vision)课程，由B. K. P. Horn教授讲授

MIT AI Lab（现在和计算机科学实验室合并为CSAI Lab）吸引了国际上许多知名学者参与计算机视觉的理论、算法、系统设计的研究，David Marr教授就是其中的一位。他于1973年应邀在MIT AI 实验室领导一个以博士生为主体的研究小组，1977年提出了不同于“积木世界”分析方法的计算视觉理论 (computational vision)，该理论在80年代成为计算机视觉研究领域中的一个十分重要的理论框架

研究热潮是从 20世纪80年代开始的，到了80年代中期，计算机视觉获得了蓬勃发展，新概念、新方法、新理论不断涌现；基于感知特征群的物体识别理论框架，主动视觉理论框架，视觉集成理论框架等，Marr的计算理论是目前比较认可的视觉框架

2012年以来深度学习在ImageNet比赛的结果大幅超过了传统方法，推动了人工智能快速发展

3. Marr的视觉计算理论

Marr 的视觉计算理论[Marr1982]立足于计算机科学，系统地概括了心理生理学、神经生理学等方面取得的所有重要成果，是视觉研究中迄今为止最为完善的视觉理论。Marr建立的视觉计算理论，使计算机视觉研究有了一个比较明确的体系，并大大推动了计算机视觉研究的发展。人们普遍认为，计算机视觉这门学科的形成与Marr的视觉理论有着密切的关系。

Marr的视觉理论中的哲学意义和方法论。

Marr 奖（**ICCV best paper**）

Fu奖（**ICPR best paper**）

3.1 视觉表示框架

第一阶段(也称为早期阶段)是将输入的原始图像进行处理, 抽取图像中诸如角点、边缘、纹理、线条、边界等**基本特征**, 这些特征的集合称为**基元图(primitive sketch)**;

第二阶段(中期阶段)是指在以观测者为中心的坐标系中, 由输入图像和基元图恢复场景可见部分的深度、法线方向、轮廓等, 这些信息的包含了深度信息, 但不是真正的物体三维表示, 因此, 称为**二维半图(2.5 dimensional sketch)**;

第三阶段(后期阶段)是在以物体为中心的坐标系中, 由输入图像、基元图、二维半图来恢复、表示和识别**三维物体**。

3.2 Marr 视觉理论的不足

Marr理论是计算机视觉研究领域的划时代成就，但该理论不是十分完善的，许多方面还有争议。比如：

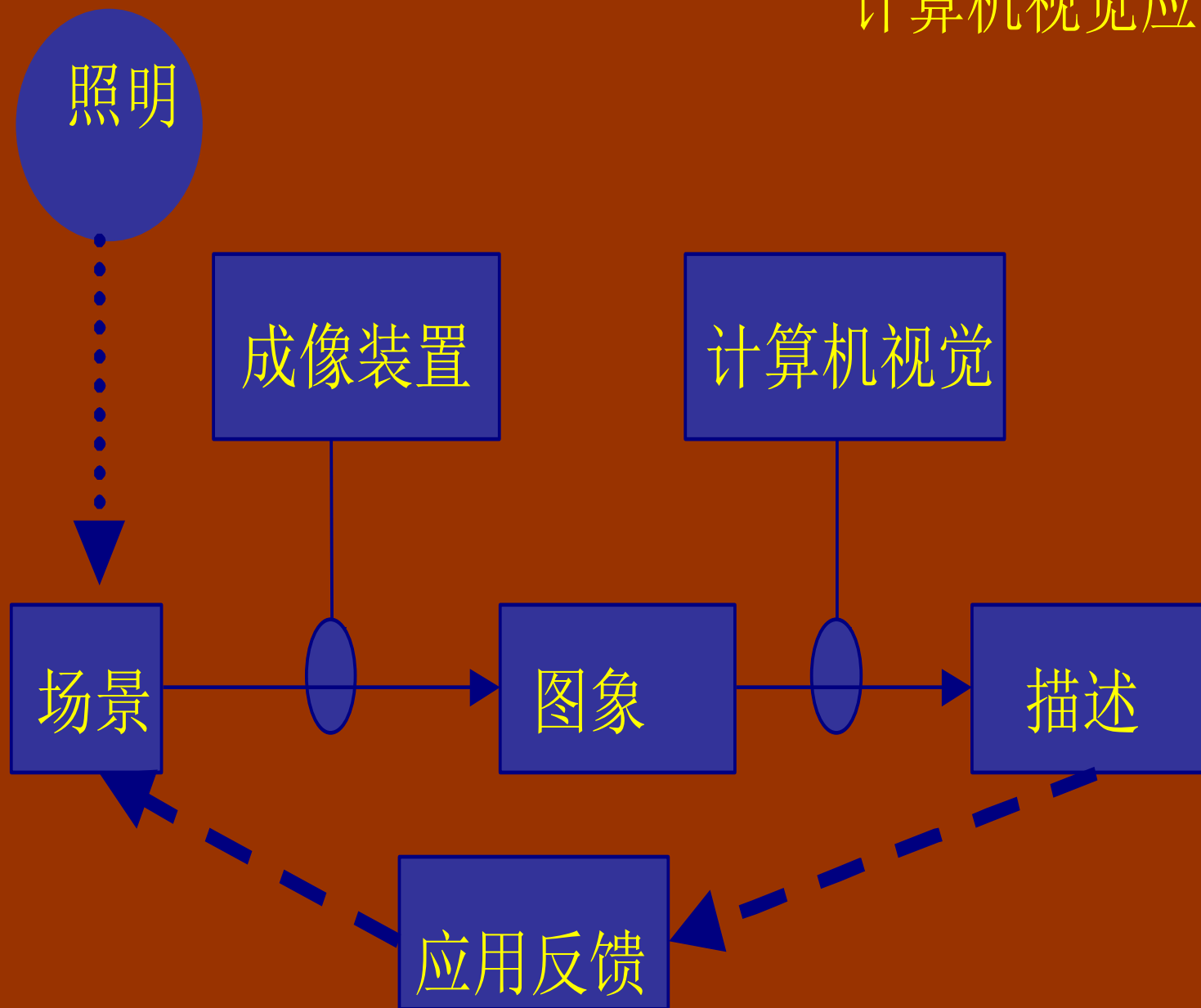
视觉处理框架基本上是自下而上，没有反馈；

没有足够地重视**知识的应用**。

Marr理论给了我们研究计算机视觉许多珍贵的哲学思想和研究方法，同时也给计算机视觉研究领域创造了许多研究起点。

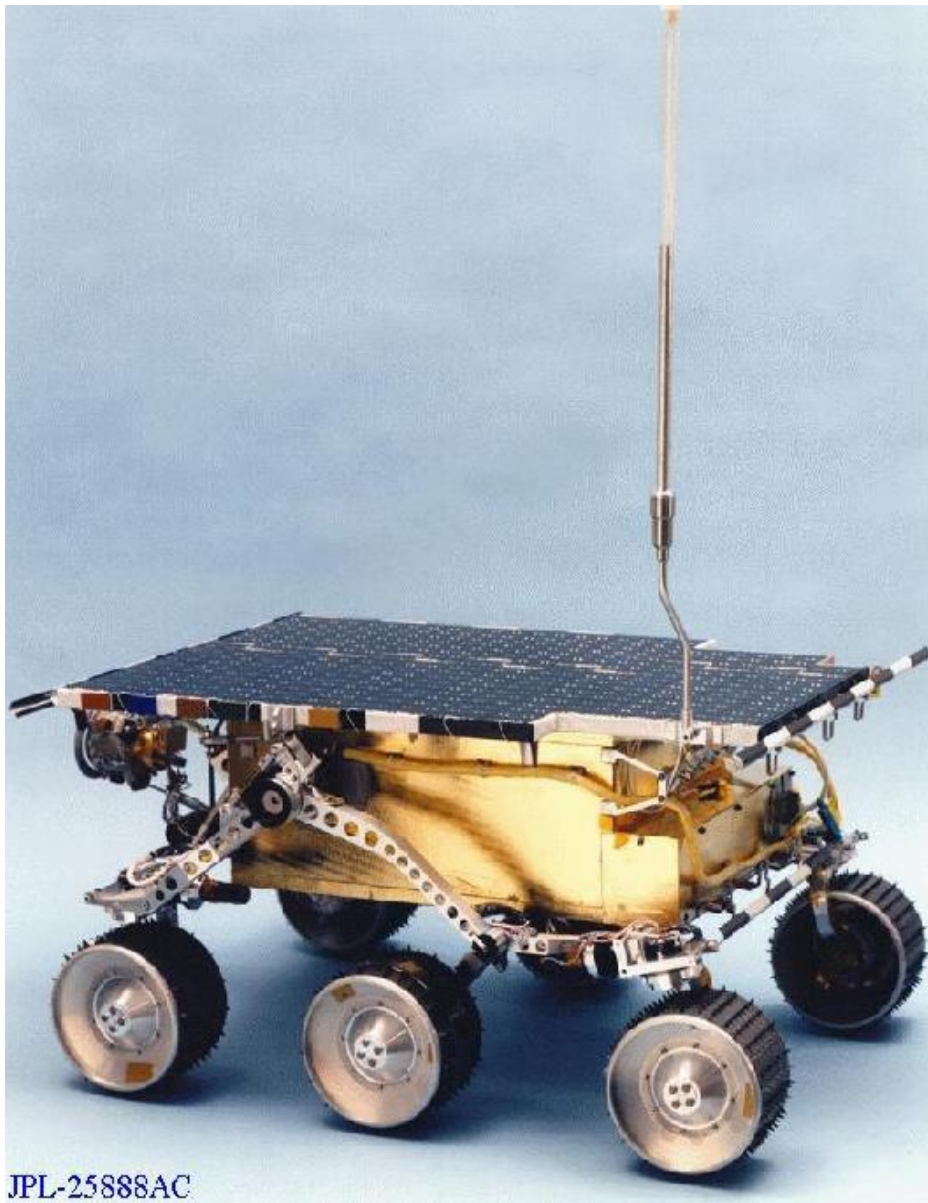
视觉系统

计算机视觉应用系统



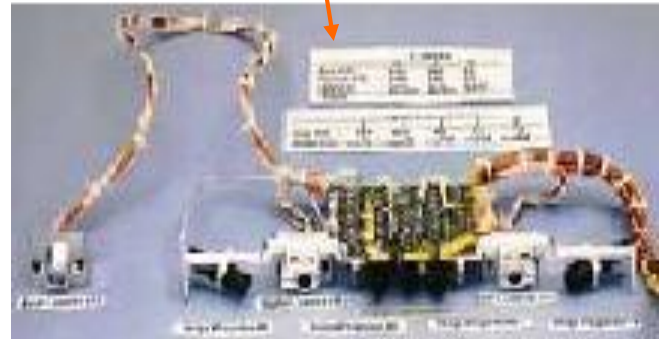
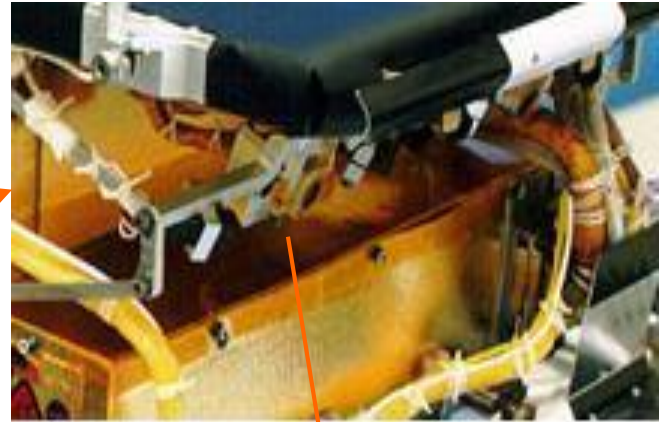
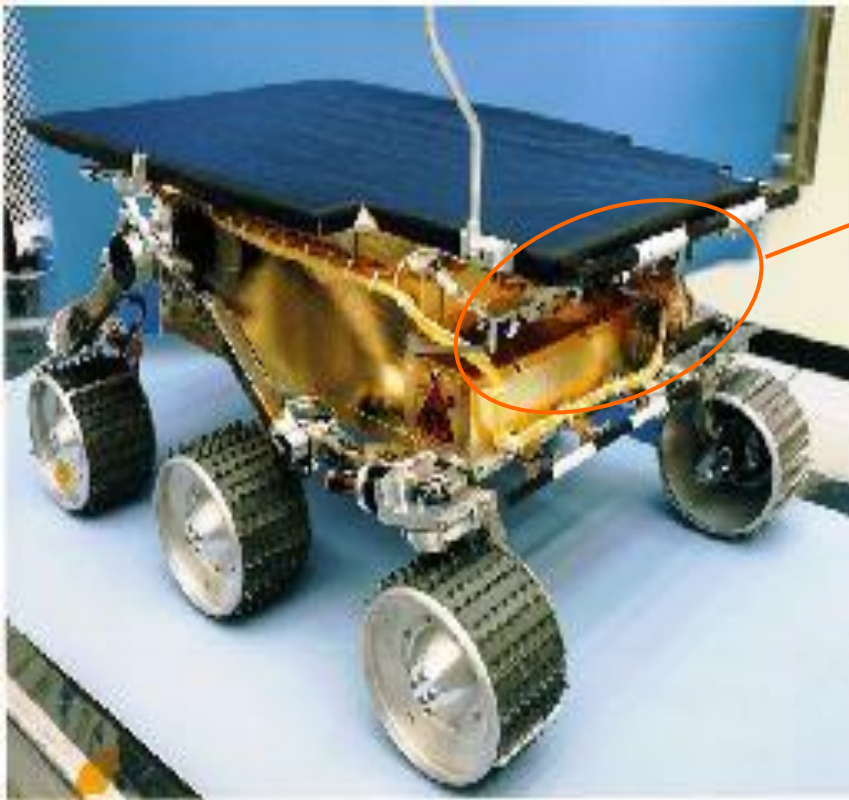
4. 计算机视觉的应用

- 零件识别与定位（工业生产线）
- 产品检验（纺织工业棉花质量检验）
- 移动机器人导航（星球机器人）
- 遥感图像分析（植被分析）
- 医学图像分析（骨骼定位）
- 安全鉴别、监视与跟踪（门禁系统）
- 国防系统（目标自动识别ATR与目标跟踪）
- 其它（动画、体育、考古）



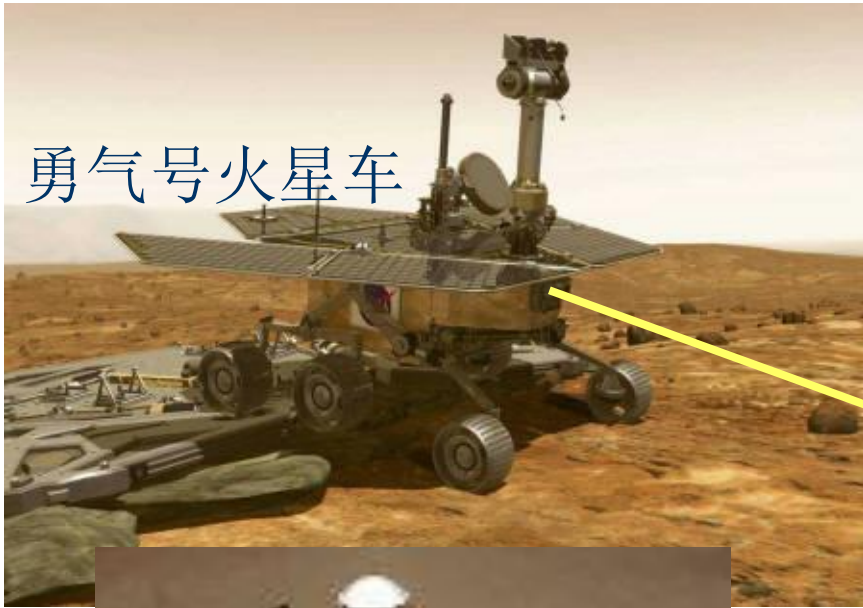
JPL-25888AC

Sojourner
火星车前部
图，中部的
两个小突出
是两个黑白
CCD摄像机

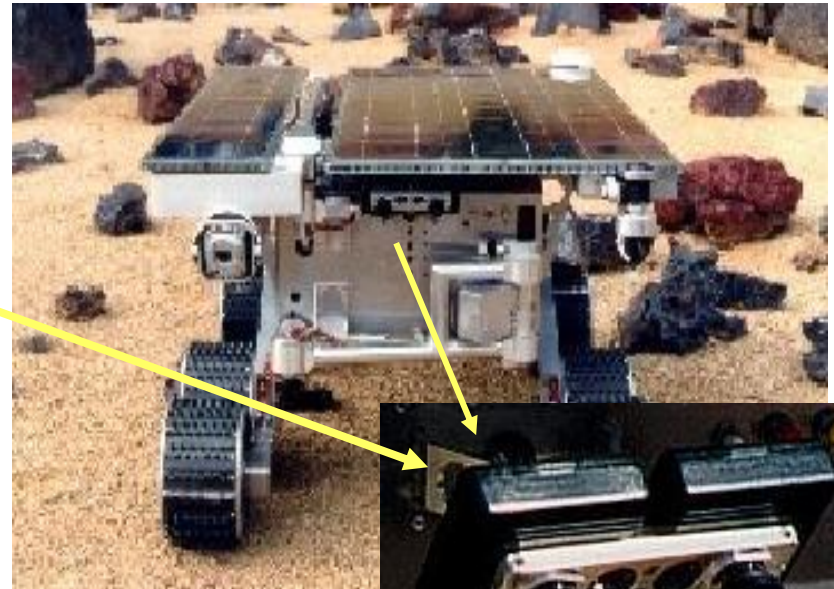


Computer Vision

勇气号火星车



Rocky 7 火星车

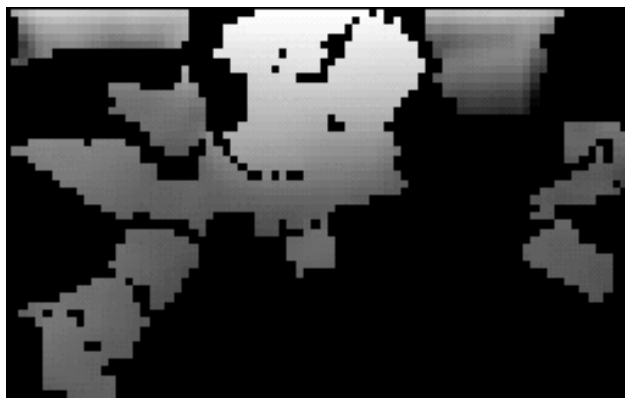


Mobile Robot Testbed

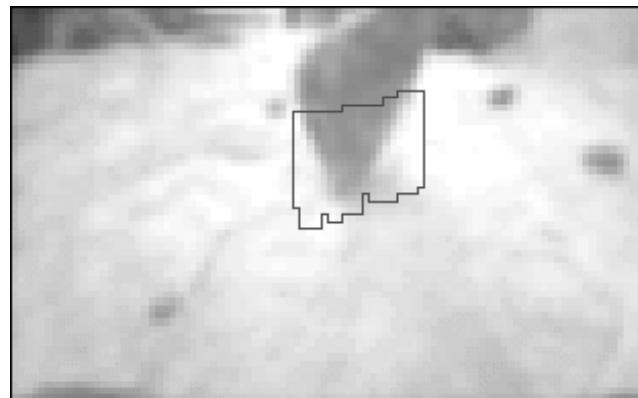
Computer Vision



Rocky7视觉系统获取的立体图象对



(a) 深度图象



(b)障碍物探测示意图

Rocky7 视觉系统对场景的深度恢复



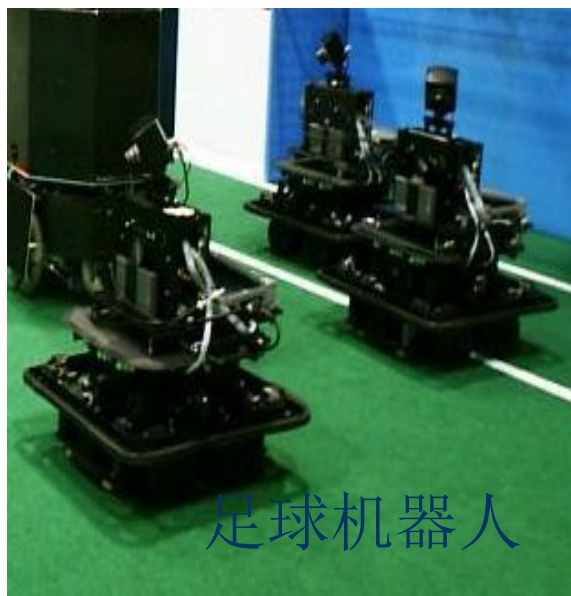
CMU月球探测实验车Nomad漫游者

Computer Vision



月球探测实验车Nomad漫游者

Computer Vision



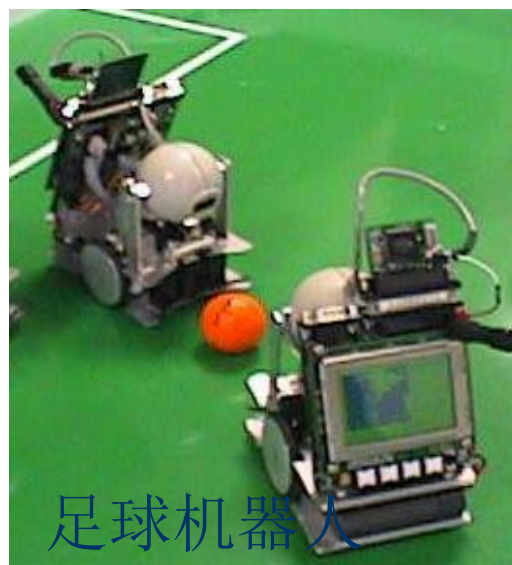
足球机器人



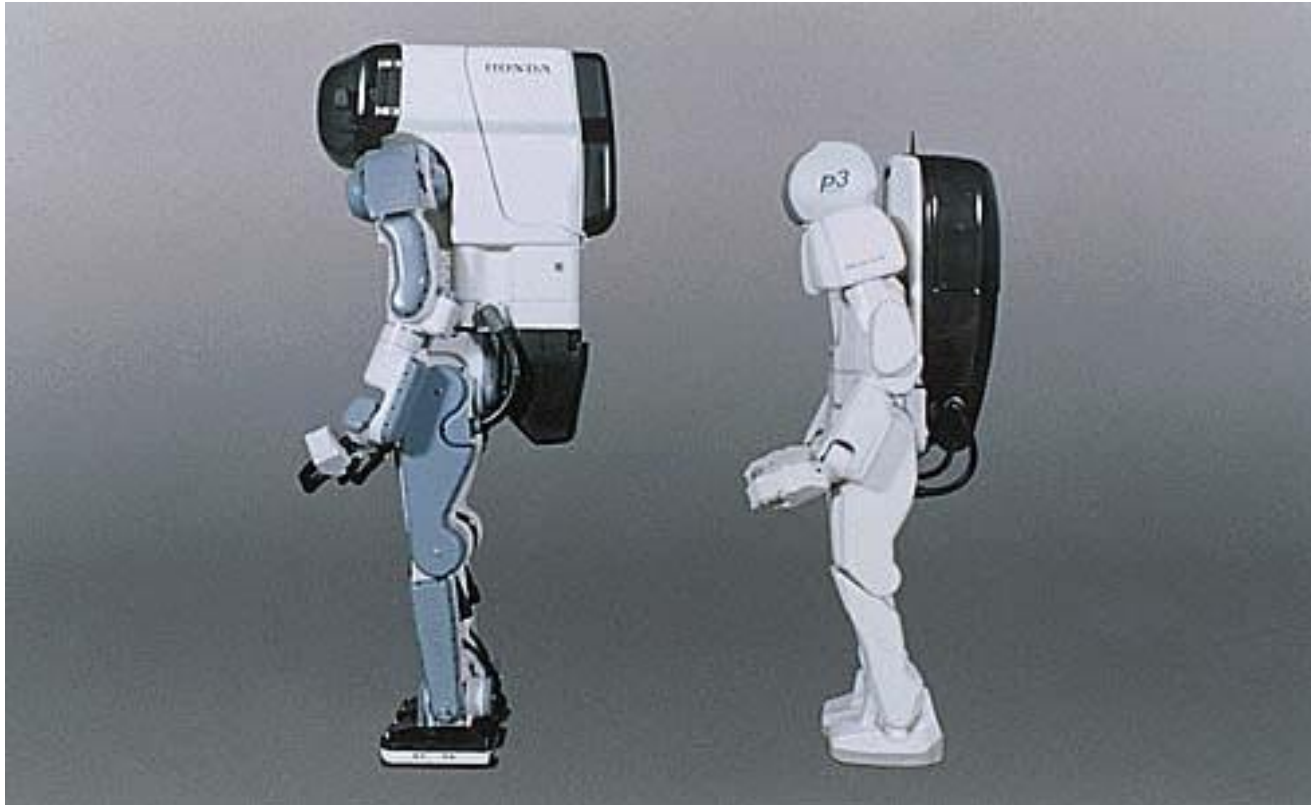
昆虫机器人



Computer Vision



足球机器人



日本 Honda 仿人机器人

Computer Vision

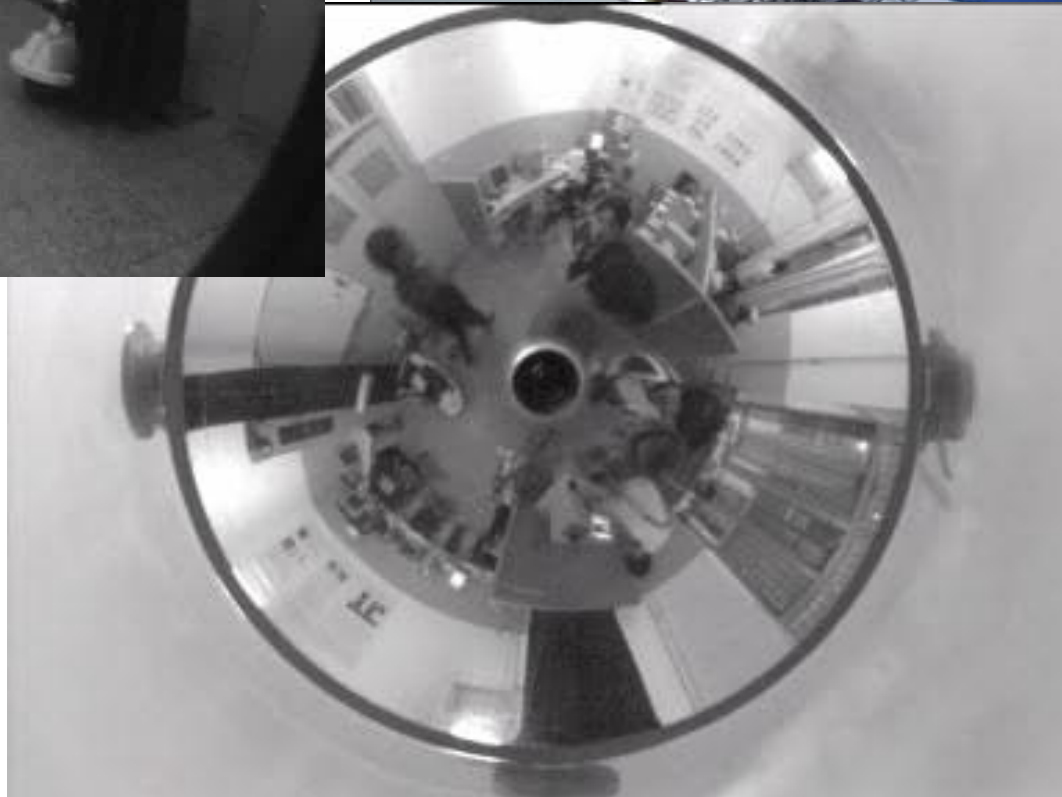


具有立体视觉的机器人

Computer Vision



人头部跟踪演示



全像图像

Computer Vision





MIT Media Lab ， 与虚拟生物交互演示

Computer Vision



基于图象序列的三维人脸恢复



深度学习



5. 计算机视觉研究的五大研究内容

一、输入设备(input device) 包括成像设备和数字化设备. 成象设备是指通过光学摄像机或红外、激光、超声、X射线对周围场景或物体进行探测成象, 得到关于场景或物体的二维或三维数字化图像.

二、低层视觉(low level) 主要是对输入的原始图像进行处理. 这一过程借用了大量的图像处理技术和算法, 如图像滤波、图像增强、边缘检测等, 以便从图像中抽取诸如角点、边缘、线条、边界以及色彩等关于场景的基本特征; 这一过程还包含了各种图像变换(如校正)、图像纹理检测、图像运动检测等.

6. 计算机视觉研究的五大研究内容

三、中层视觉（middle level） 主要任务是恢复场景的深度、表面法线方向、轮廓等有关场景的2.5维信息，实现的途径有立体视觉（stereo vision）、测距成像（rangefinder）运动估计（motion estimation）、明暗特征、纹理特征等所谓的从X恢复形状的估计方法。系统标定、系统成像模型等研究内容一般也是在这个层次上进行的。

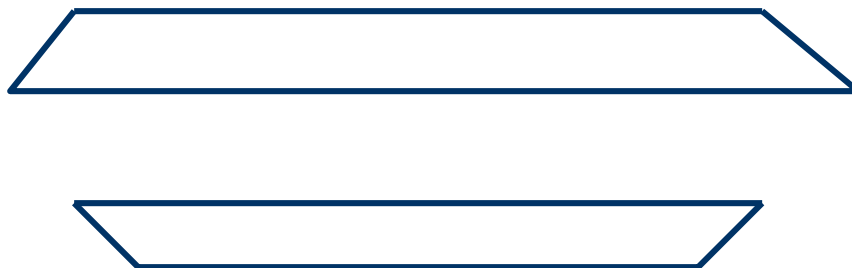
四、高层视觉（high level） 主要任务是在以物体为中心的坐标系中，在原始输入图像、图像基本特征、2.5维图的基础上，恢复物体的完整三维图，建立物体三维描述，识别三维物体并确定物体的位置和方向。

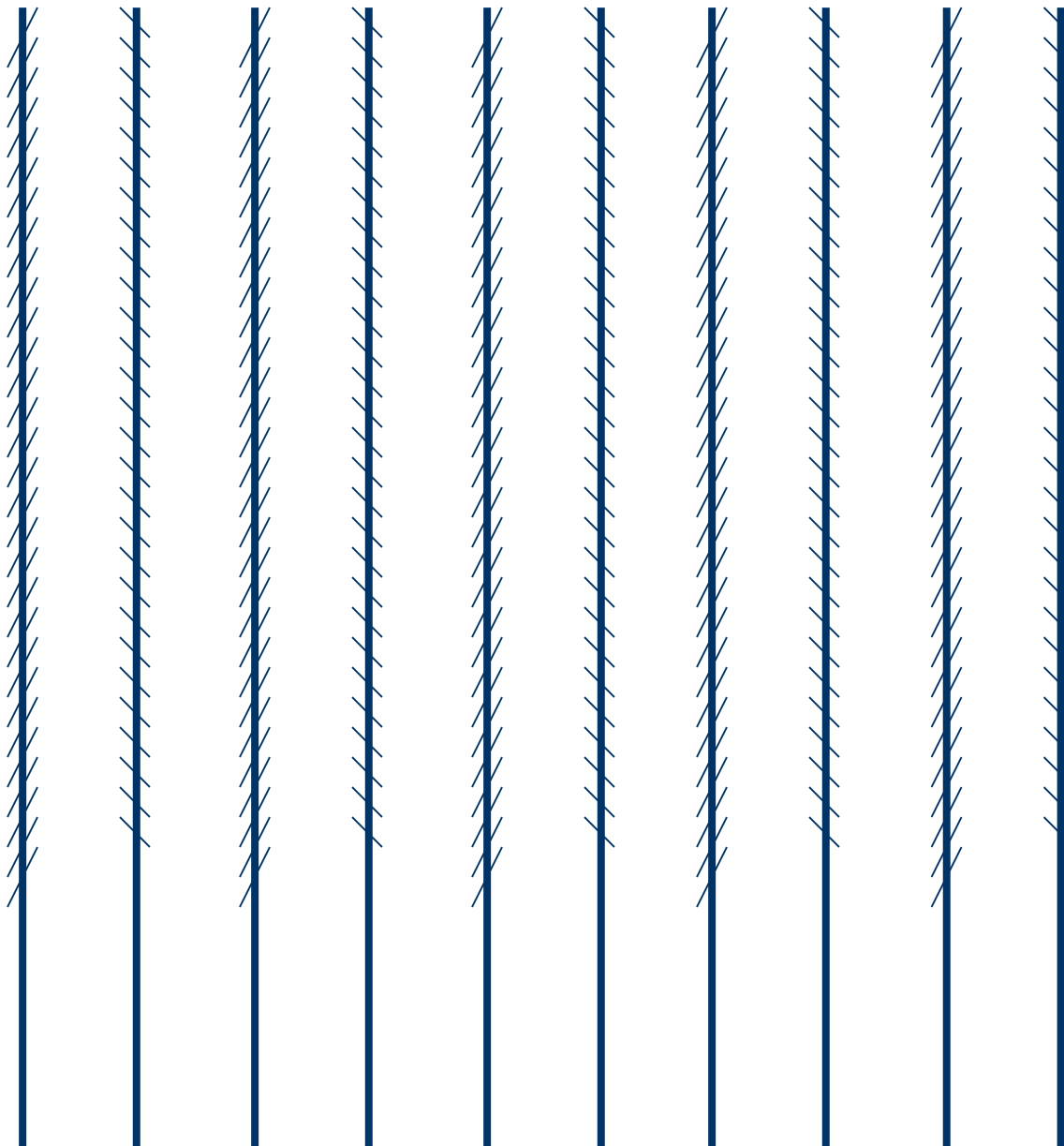
7. 计算机视觉研究的五大研究内容

五、体系结构（system architecture） 这一术语最通常的含义系指在高度抽象的层次上，根据系统模型而不是根据实现设计的具体例子来研究系统的结构。体系结构研究涉及一系列相关的课题：并行结构、分层结构、信息流结构、拓扑结构以及从设计到实现的途径。

8. 计算机视觉研究面临的困难

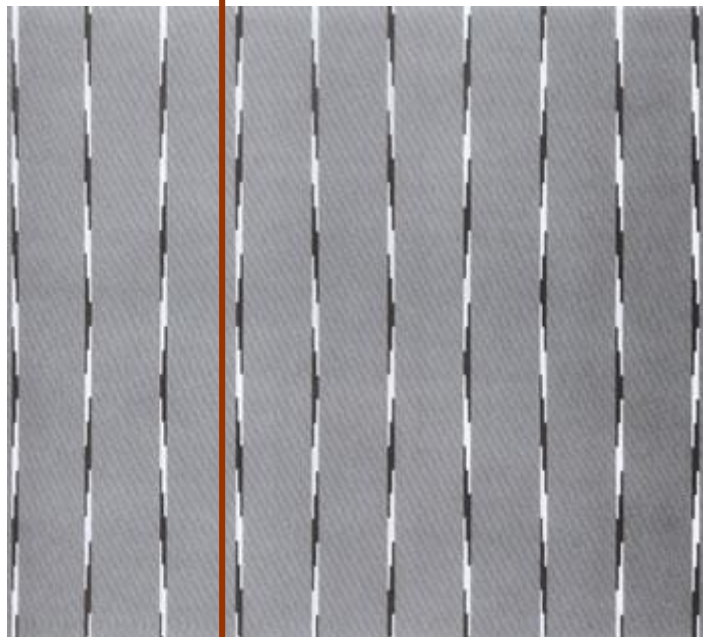
- (1) 图像多义性： 三维场景被投影为二维图像，深度和不可见部分的信息被丢失，因而会出现不同形状的三维物体投影在图像平面上产生相同图像的问题。另外，在不同角度获取同一物体的图像会有很大的差异。
- (2) 环境因素影响： 场景中的诸多因素，包括照明、物体形状、表面颜色、摄像机以及空间关系变化都会对成像有影响，
- (3) 知识导引： 同样的图像在不同的知识导引下，将会产生不同的识别结果。
- (4) 大量数据： 灰度图像，彩色图像，深度图像的信息量十分巨大，巨大的数据量需要很大的存贮空间，同时不易实现快速处理。





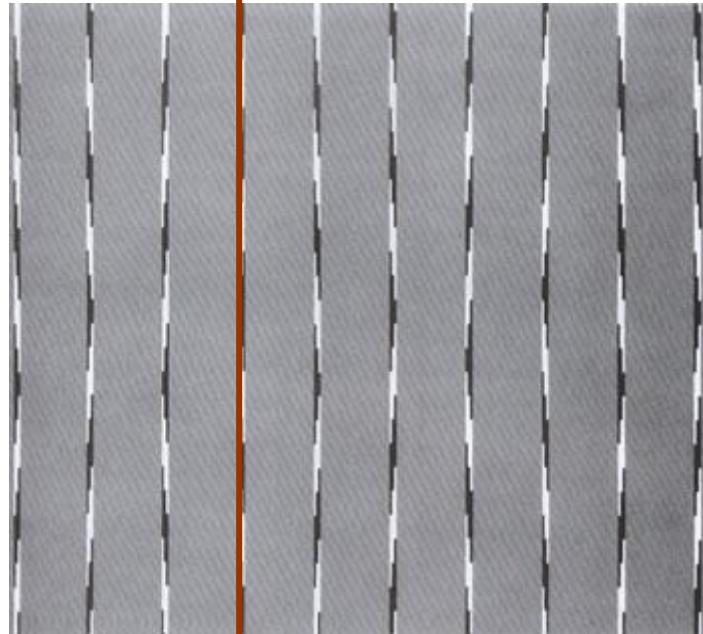
Computer Vision

平行

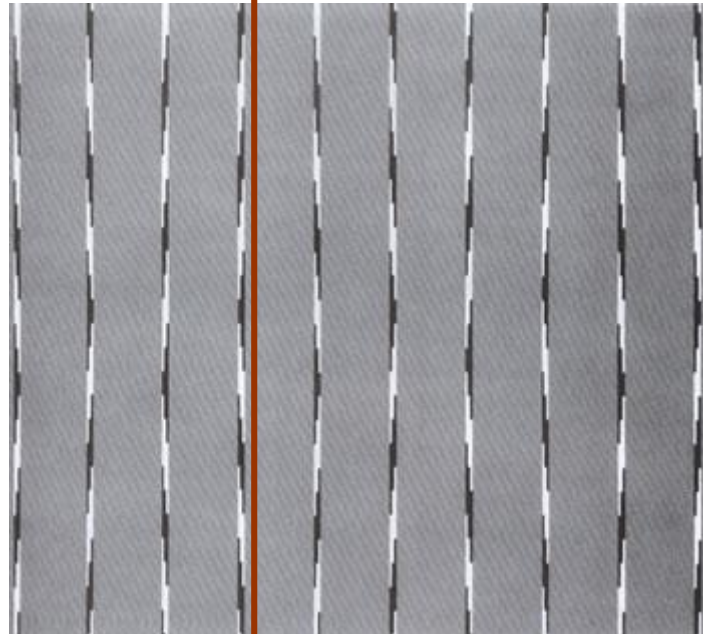


Computer Vision

平行



Computer Vision



Computer Vision

9. 计算机视觉与其它学科领域的关系

(1) **图像处理**：图像处理通常是把一幅图像变换成另外一幅图像，也就是说，图像处理系统的输入是图像，输出仍然是图像，信息恢复任务则留给人来完成

(2) **计算机图形学**：通过几何基元，如线、圆和自由曲面，来生成图像，它在可视化（Visualization）和虚拟现实(Virtual Reality)中起着很重要的作用。计算机视觉正好是解决相反的问题，即从图像中估计几何基元和其它特征。因此，计算机图形学属于图像综合，计算机视觉属于图像分析。

(3) **模式识别**：用于识别各种符号、图画等平面图形。模式一般指一类事物区别于其它事物所具有的共同特征。

10. 计算机视觉与其它学科领域的关系

(3) **人工智能 (AI)**：涉及到智能系统的设计和智能计算的研究。在经过图像处理和图像特征提取过程后，接下来要用人工智能方法对场景特征进行表示，并分析和理解场景。人工智能有三个过程：感知、认知和行动。

(4) **媒体计算 (MC)**：文字\图形\图像\动画\视频\音频等各类感觉媒体的共同基础计算理论、计算方法、算法，媒体系统实现技术。基础性和综合性，是以实现下一代计算机能听、能看、会说、会学习为目标。

(5) **神经物理学与认知科学**：将人类视觉作为主要的研究对象。计算机视觉中已有的许多方法与人类视觉极为相似。许多计算机视觉研究者对研究人类视觉计算模型比研究计算机视觉系统更感兴趣，希望计算机视觉更加自然化，更加接近生物视觉。

11. 计算机视觉研究对策

研究人员不断寻求新的途径和手段，比如，
主动视觉(active vision)，
面向任务的视觉(task-oriented vision)，
基于知识、基于模型的视觉，
以及多传感融合和集成视觉等方法

11. 计算机视觉研究对策

我们会看到，计算机视觉系统的最大特征是，在视觉的各个阶段，系统尽可能地进行自动运算。

为此，系统需要使用各种知识，包括特征模型、成像过程、物体模型和物体间的关系。如果计算机视觉系统不用这些知识，则其应用的范围及其功能将十分有限。

因此，视觉系统应该使用那些可以被明确表示的知识，以使系统具有更高的适应性和鲁棒性。

合理地使用知识不仅可以有效地提高系统的适应性和鲁棒性，而且可以求解计算机视觉中较难的问题。