



第3章 环境感知



第09讲 景物理解

一般人的视觉系统都能够看到三维景物的世界。对于人类的生存和活动而言，这种把握空间信息的能力是至关重要的，否则我们将在这个世界上寸步难行。

所以，机器如果要具备类似于人类视觉能力的话，同样地，也必须首先获得这种根据二维图像获得三维景物的能力。应该说，从纯理论意义上看，景物理解能力的实现是整个机器视觉研究最为核心的研究主题。



1、获取景物空间线索

景物理解的第一步就是要从输入的图像中来获取三维空间信息。

空间信息的三维线索，从观察者的角度看，主要是指图像景物的表面朝向和深度距离。

视觉生理和心理研究表明，人类的空間信息主要是通过双目视差和环境光流的分布差异产生的。所有这些信息来源在视觉系统中的分散与聚合的相互影响加工过程中，形成了三维形体的形象景物。



1、获取景物空间线索

一般，双眼视物时，通过两只眼睛同时辐辏和双眼视差的协调运作，从而可以对刺激物获得深度知觉的线索。

所谓辐辏作用，是指由两眼球转动以聚合视线从而获得深度知觉；

所谓双眼视差，则指注视同一物体时，因两眼视线的角度不完全相同，故而在两眼网膜上所构成的两个影像彼此稍有差别的现象。



1、获取景物空间线索

对于下图，当你将两幅图像分别呈现给左、右眼时，比如用一张薄纸将左图与右图隔开来观看，就能够获得这种三维的深度信息，从而感知到空间景物的立体感来。

。



体视匹配图



1、获取景物空间线索

除了考虑双目视差因素外，我们还应利用其它空间线索。这时运动视差就成为另一个深度线索的重要信息源。

因为当运动时，注视的物体方向就会改变：

如果物体在近处，其方向变化大；

反之，如果物体在远处，其方向变化就小，

而运动视觉反映的正是不同距离处物体方向的变化速率之差。

而环境光流的因素则包括：

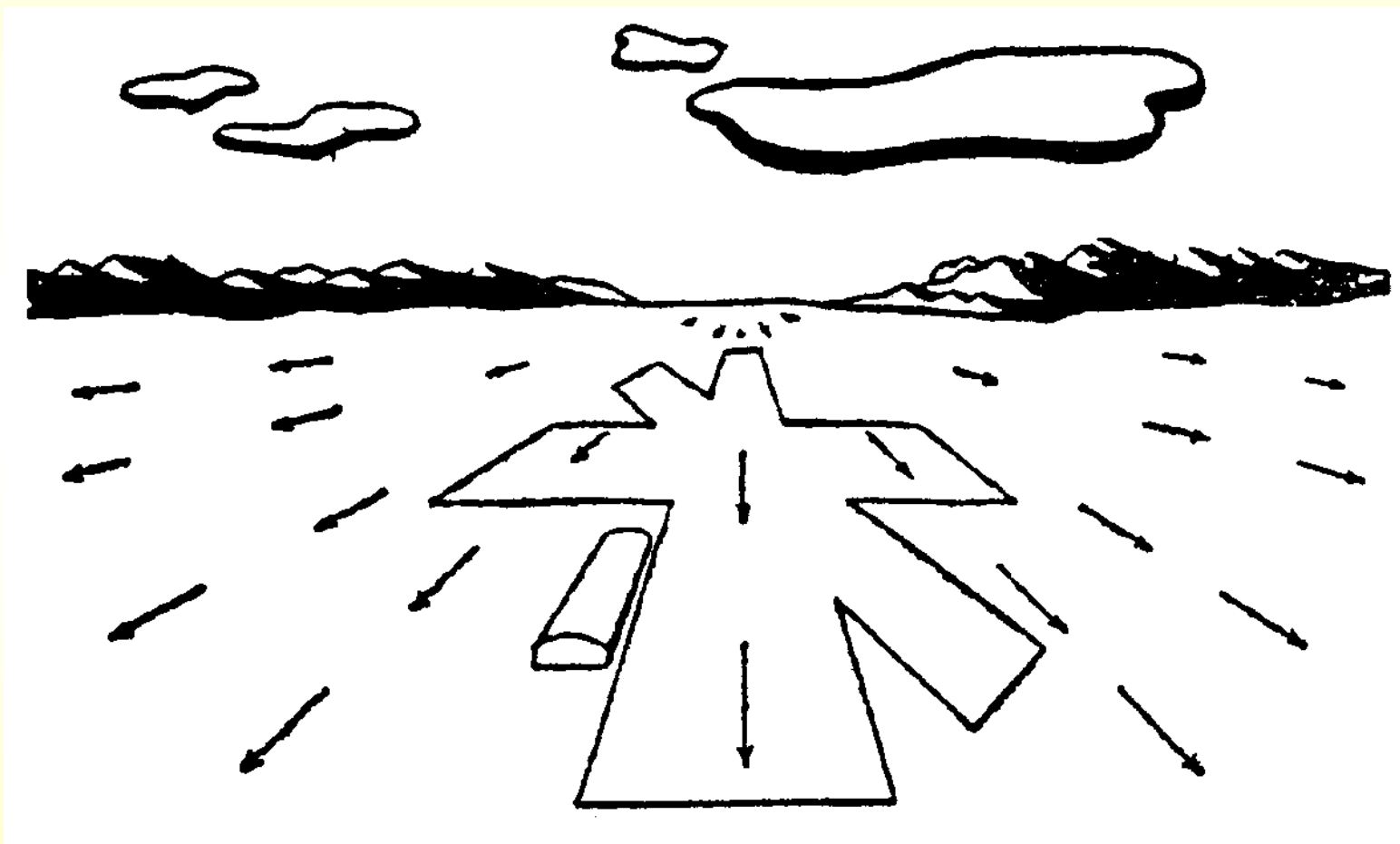
来自运动的结构信息、

来自质地的形状信息，

以及来自明暗和轮廓的形状信息等。



1、获取景物空间线索





1、获取景物空间线索

在立体视觉的**深度信息表征**方面，**阴影**，尤其是**附着阴影**，也是决定物体深度知觉的一个重要因素。

实际上，阴影作为深度线索，其效力十分强烈，即使在没有物体存在的情况下，阴影也能引起对这些不存在事物的知觉。

除了阴影线索外，**遮挡重叠线索**、**明暗纹理线索**，以及**大小透视线索**等等，都可以作为立体视觉恢复的依据。

不过，在人类的立体视觉加工中，往往是综合利用各种线索来根据二维形状恢复三维形体的。



1、获取景物空间线索

我们视觉深度知觉也会受到恒常性经验的影响。至于阴影无中生有而产生立体知觉的例子，更加说明任何体视阴影提供的线索都是不牢靠的。

很明显，人类的体视机制，远要比我们计算假设所要想象的还要复杂，这其中许多因素都是与主观经验所密不可分的。

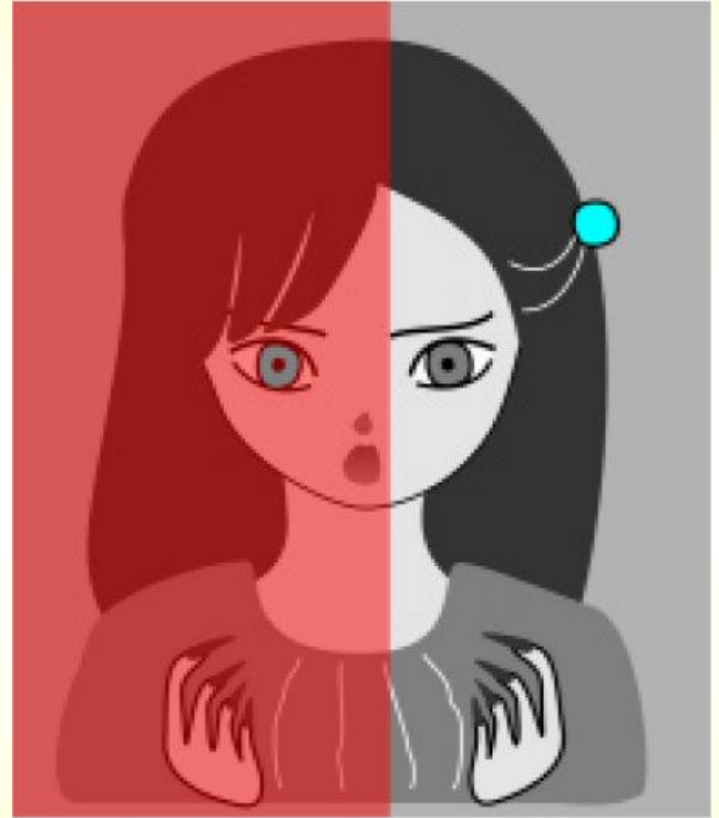


1、获取景物空间线索

例如，物体的颜色具有明显的主观深度效果。在日光的照耀下，暖色距离向前靠，冷色距离向后靠。

而在同等距离内，鲜红色的物体比鲜蓝色物体似乎要靠前些，而饱和度（如红色）要比靠色（如粉红色）显得也要近些。

当然，这些依赖于主观经验的视觉信息，目前很难为机器所能利用。



颜色主观效应



2、马尔视觉计算理论

对于机器而言，要实现三维景物的理解计算，首先必须要给出可以进行**形式化表征**的计算策略和方法。

根据美国**视觉计算理论**提出者马尔的观点，**视觉**感知首先是一个**信息处理过程**，是要从**图像**中发现外部世界中**有什么**以及**处在什么位置**。

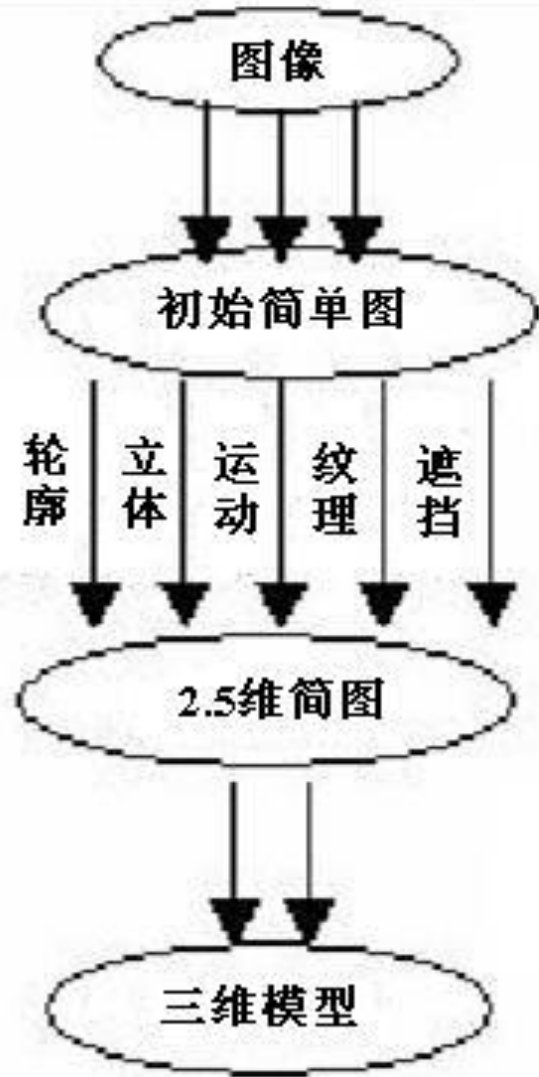
因此，视觉对象的**内部表征**就成为视觉计算的主要载体。于是视觉计算任务就成为如何**根据给定的图像**，来获取各个层次的**内部表征**，直至恢复图像的**三维景物**。



2、马尔视觉计算理论

根据马尔的观点，视觉感知首先是一个信息处理过程，是要从图像中发现外部世界中有什么以及处在什么位置。因此，视觉对象的内部表征就成为视觉计算的主要载体。于是视觉计算任务就成为如何根据给定的图像，来获取各个层次的内部表征，直至恢复图像的三维景物。

马尔的视觉计算理论



视觉系统的三个表象层次

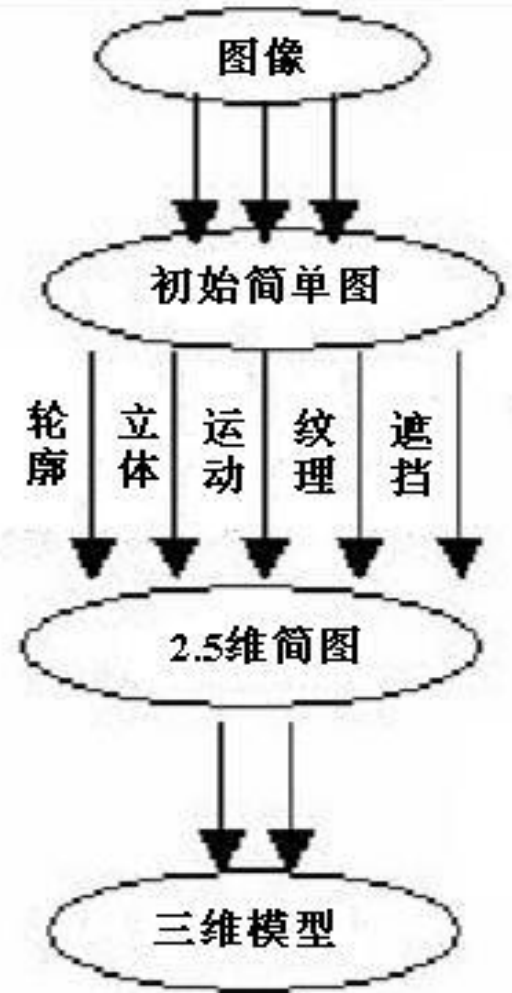


2、马尔视觉计算理论

马尔便是从三个层次来建立视觉计算理论：

- (1) **计算理论**：确定视觉计算的目的；
- (2) **表征与算法**：如何实现视觉计算任务，确定输入输出的表征，给出不同表征转换之间的算法；
- (3) **硬件实现**：在物理上如何实现视觉表征及其转换算法。

特别是第二个层次，就是视觉计算理论的核心内容，为此，马尔提出了**具体四级表征**，分别是：

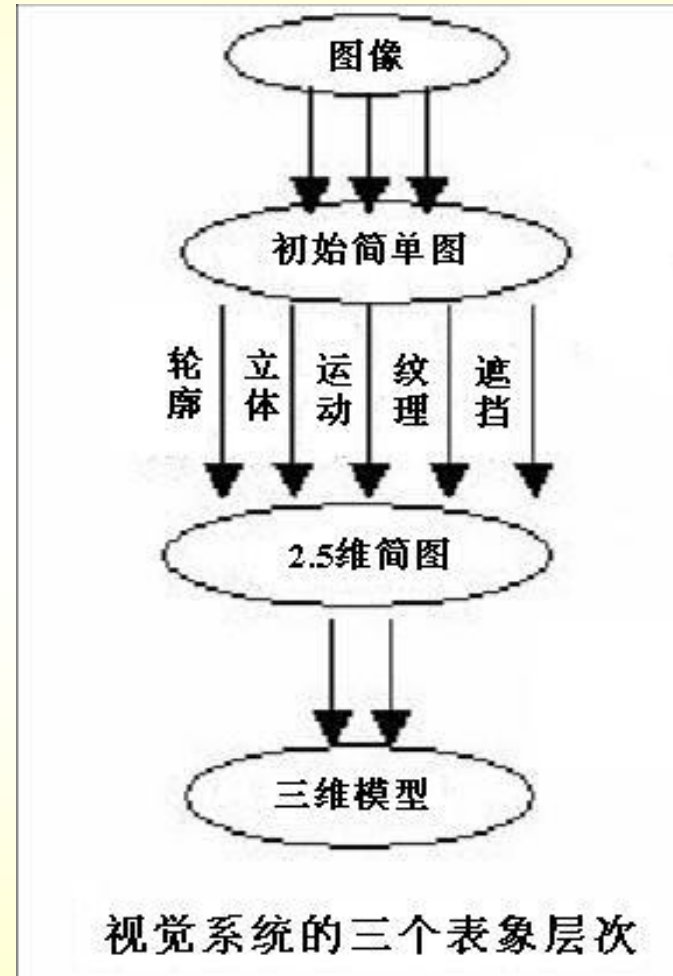


视觉系统的三个表象层次
马尔的视觉计算理论



2、马尔视觉计算理论

- (1) **图像**：表达光强，用像素表征；
- (2) **要素图**：表达二维图像中的重要**变化信息**及其**分布**，比如零交叉、斑点、端点、不连续点、边缘片段、有效线段、组合群、曲线组织、边界等；
- (3) **2.5维图**：在以**观察者为中心**的坐标系中，将可见**朝向**、大致**深度**及其不连续**轮廓**表达清楚，比如表面要素的朝向、距离观察者的深度、深度上不连续点、表面朝向不连续点等。
- (4) **3维模型**：在以**物体为中心**的坐标系中，景物用**体积**基元和**面积**基元给出的**层次模块化表征**。





2、马尔视觉计算理论

对上述各个层次表征的获取，正好对应视觉计算不同步骤的计算分析处理阶段，涉及到：

图形检测、运动检测、空间检测、分形检测等差异性
低层处理技术，

边线合成、区域生成、纹理识别、表面恢复等相似性
中层处理技术，

以及景物匹配、含义推断、知识习得、目标规划等理解性
高层处理技术。

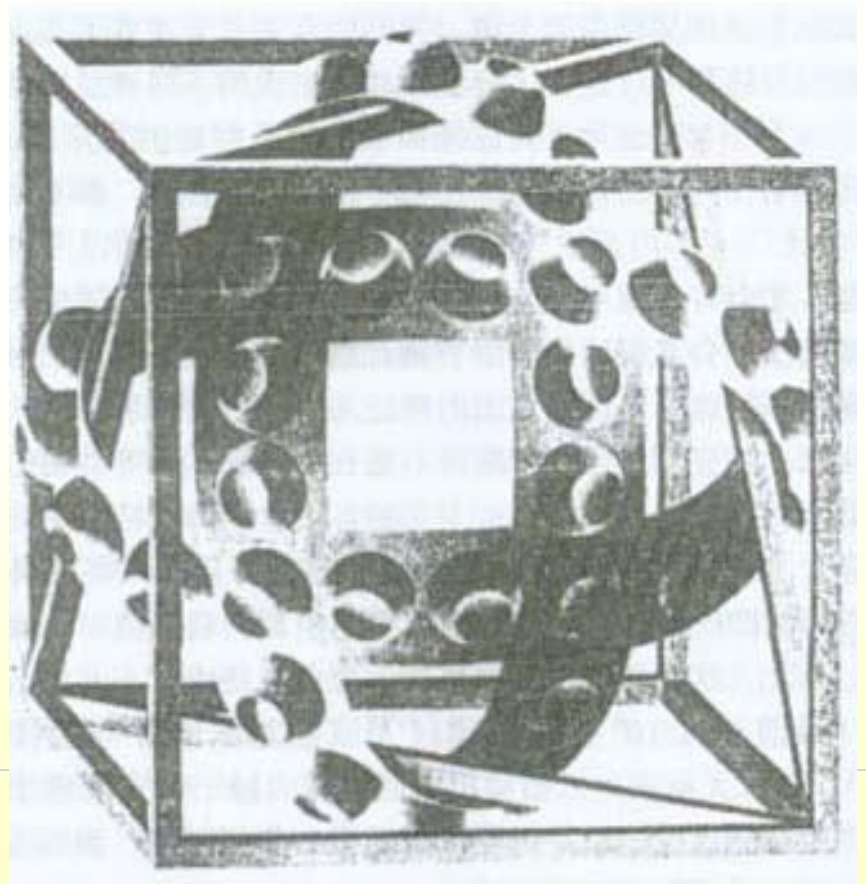
这样，按照马尔计算理论的四级步骤，在一定程度上是可以完成视觉景物的深度计算任务的，从而恢复三位景物的立体形状。



3、视觉主动计算问题

应该说，大多数**体视线索**，只要是客观存在的，或多或少机器都可以加以**利用**，来为深度计算提供**依据**。

但是如果遇到像艾舍尔创作的下面这幅作品情况，你就根本**无法客观**地计算出其**深度距离**。这幅作品深度信息的理解，一切都依然于你**主观意念**是如何准备接收它的，是起念为“凹陷物”呢还是起念为“凸起物”？强调客观信息表征的机器，对此将做何为之呢？



是「凹陷物」还是「凸起物」



3、视觉主动计算问题

的确，相对于机器而言，人类有着复杂的视觉活动能力，特别是人类视觉中对视觉含意的把握理解以及主观经验在视觉活动中所起作用的微妙机理，远远超过目前机器所能拥有的水平。

这也意味着，要让机器真正拥有人类视觉能力，就必定要涉及到完整的人类视觉思维能力的计算实现问题。



3、视觉主动计算问题

人类在用视觉器官收受信息时，并非对环境中所接触到的一切刺激特征全盘接收的。

人类的眼睛之所以不同于照相机，以及人类的视觉系统之所以不同于目前的各类机器视觉系统，其根本原因就是人类所获取的视觉信息、所获得的视觉经验，都是主动选择性的。

从这个意义上讲，人类的视觉系统不是简单地记录外部世界的精确映像，而是创造性地给出渗透有自己主观加工的“画卷”。



3、视觉主动计算问题

美国科学家欧文·洛克在《知觉之谜》一书中指出：“大脑并不是简单地记录下世界的准确的映象，而是创造出自己的“照片”来。”

无独有偶，英国的格里高里在《视觉心理学》中也指出：“知觉不是简单地被刺激模式决定的，而是对有效的资料能动地寻找最好的解释。”

最近，著名科学家克里克在《惊人的假设》中更是强调：“看是一个建构过程。在此过程中，大脑以并行的方式对景物的很多不同“特征”进行响应，并以以往的经验为指导，把这些特征组合成一个有意义的整体。看涉及大脑中的某些主动过程；它导致景物明晰的、多层次的符号化解释。”



3、视觉主动计算问题

很明显，对于一个物体的**知觉**，必须通过对这一物体的**各组成要素**进行**感知把握**之后才能完成；

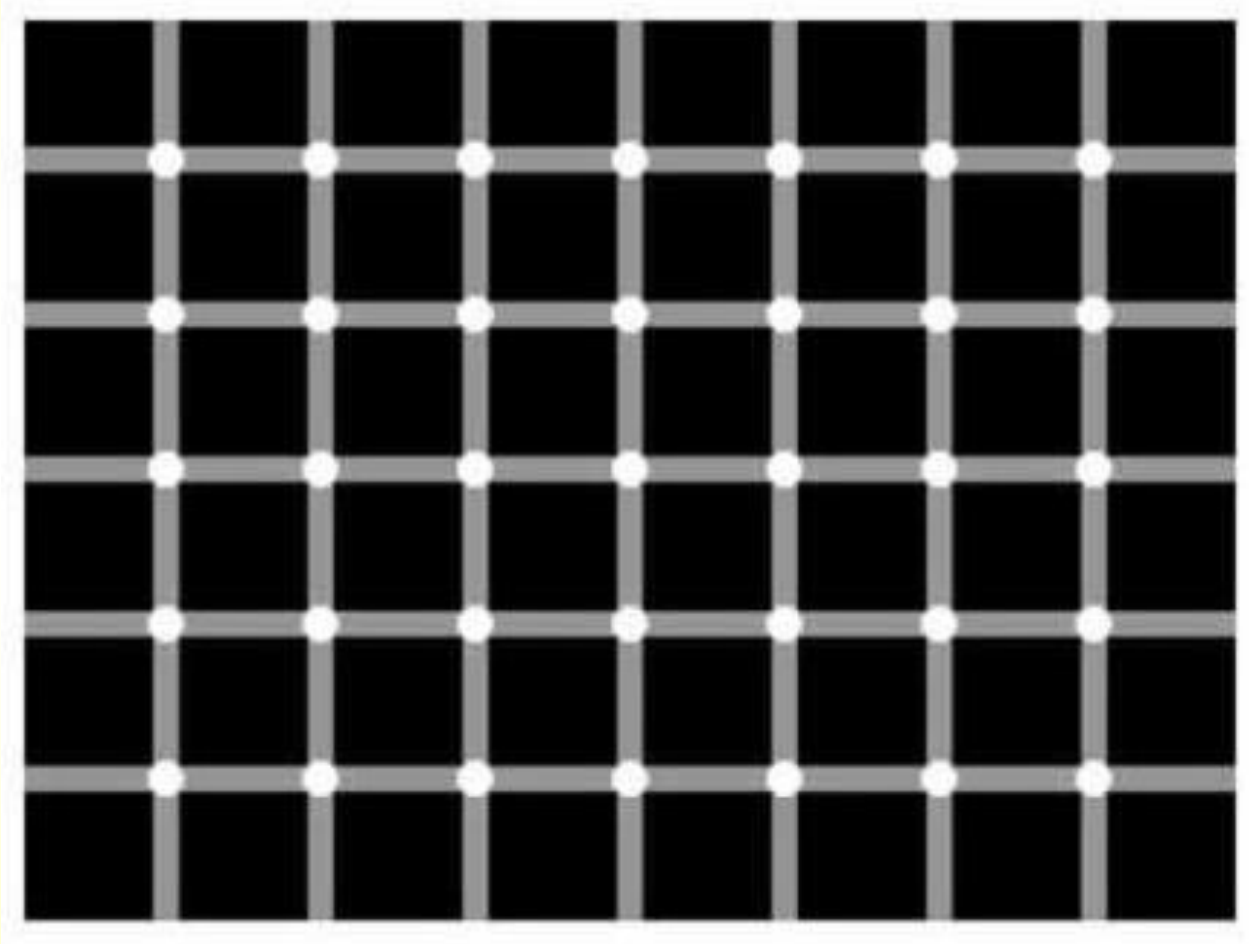
然而如果在感知把握时，没有一个**整体**的概念作指导，那么对这个物体的知觉就连一步也不能深入下去。

观看并不是对视觉对象的**机械复制**，而是对其总体结构特征**积极主动、有选择的把握**。

通俗一点讲就是说，观察者能看见什么，不仅取决于**外界**呈现的**视觉刺激**，还取决于他**主观的注意**和**意向指导**。外界刺激只有在主观意识活动的参与下，才能形成视觉形象的显现。



3、视觉主动计算问题



似动视觉效应



3、视觉主动计算问题

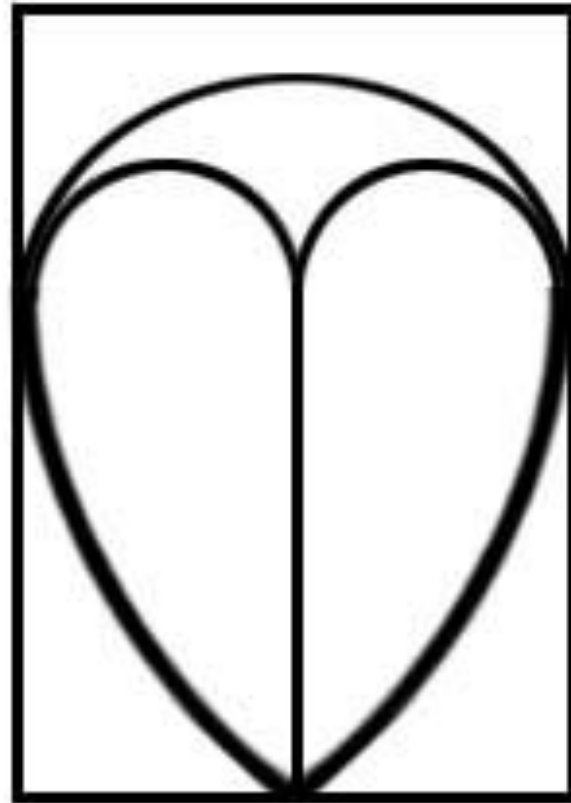
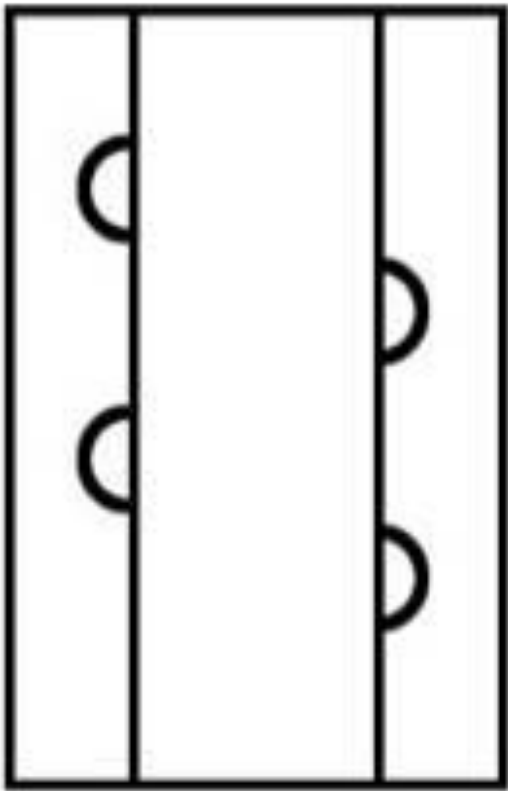
其实，从某种意义上讲，一切理解都必然是主观性的、是个性化的。

古代有一则寓言，说的是：“人有亡鈇者，意其邻人之子，视其行步，窃鈇也；颜色，窃鈇也；言语，窃鈇也；动作态度，无为而不窃鈇也。俄而掘其谷而得其鈇，他日复见其邻人之子，动作态度，无似窃鈇者。”（《吕氏春秋·去尤篇》）

在主动视觉中，主观意念一旦产生有时会如此强烈，令你挥之不去，顽固盘旋在你的脑海，左右着你的感知活动。



3、视觉主动计算问题



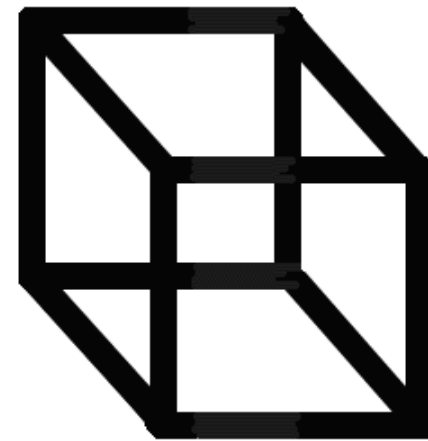
画谜：主观意念测试图



3、视觉主动计算问题

当然，在这种主观意念起作用的情况下，除了视觉会主动有选择地去“发现”线索以构成有意义的整体感知外，还有一个重要的特点就是主观意念往往并非总是唯一的，

其中，往往象著名的Necker立方体的感知一样，存在着多个意念都代表了一种合理的感知理解，而最终赢得主导作用的那个结果知觉形象就是在竞争中获胜的那个意念最终起到了主导地位。



Necker立方体



3、视觉主动计算问题

美国画家 Jasper Johns 利用这种**竞争主导机制**曾创作过一幅抽象画，见右图。

在这幅图画中你可以任意选择你自己喜欢的幸运数字，然后你总能在这其中清晰确定地找到它。但如果你不先入为主地设定任何数字在脑海里，那么你会看到一团凌乱的涂鸦而不知所云。

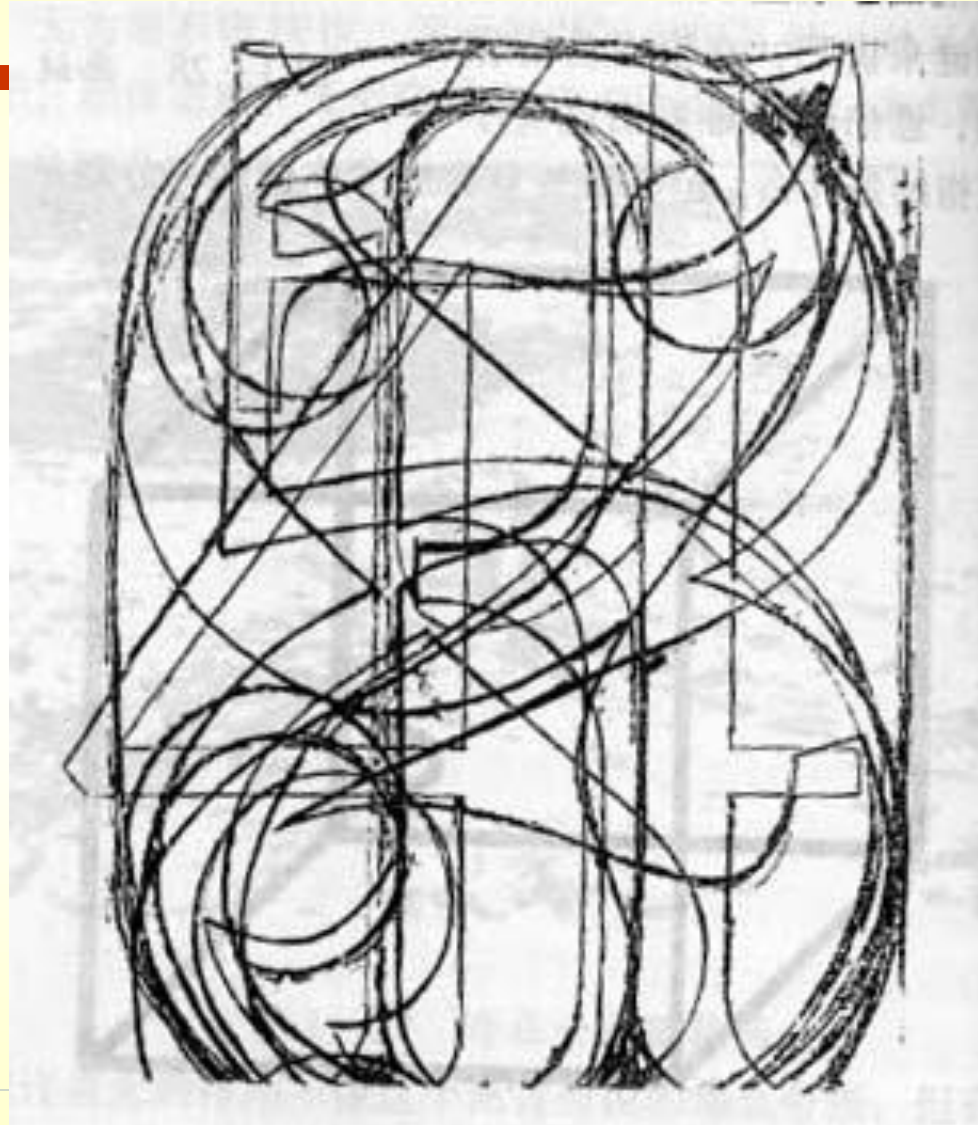


图 1-30 《0-9》作品 Jasper Johns 作，1960



3、视觉主动计算问题

从视觉神经机制上讲，由于要有意地主动跟踪和搜寻有效线索，因此视觉第二通路是必不可少的。

从事机器视觉研究的科学家们已经认识到这一点，也已经开始了主动视觉的机器实现研究。

遗憾的是，通过运动序列图像的分析 and 跟踪，机器确实可以去选择有效的线索，但由于机器缺乏主观意向性，因此客观视觉刺激不存在的线索，机器是不可能无中生有的。这样，主观意念就将成为机器视觉的真正困境，诸位有什么方法能够使机器拥有主观性？



3、视觉主动计算问题

那么，面对如此复杂的主动视觉理解问题，我们如何构建有效的视觉计算系统来部分实现有效的视觉景物理解任务呢？借助于对人类视觉原理的了解和运用，这里我们试探性地给出了一种动态场景主动理解的视觉计算模型，抛砖引玉，希望能够有助于读者了解机器视觉研究未来的发展方向。

应该指出，目前机器视觉系统的构建，主要都是建立在视觉信息处理的bottom-up策略之上的，如马尔的计算策略那样，很少运用人类视觉经验的top-down策略。为了弥补这样的不足，构建未来的机器视觉系统，主要应该依据人类视觉认知机理研究的成果，通过引入联想觉知机制，来形成一种具有觉知能力的视觉感知动态计算模型

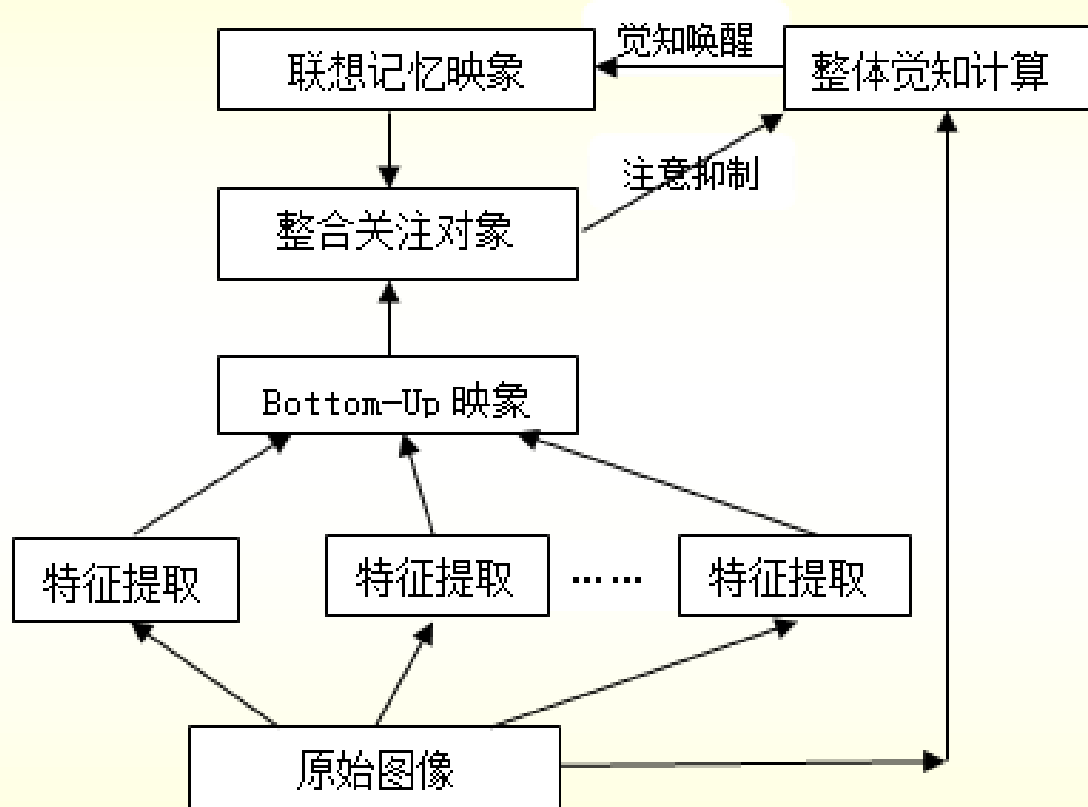


3、视觉主动计算问题

为此，我们给出了一种按照这样构想的一种视觉动态觉知的计算模型。在该模型中，除了必须的Bottom-Up视觉加工处理外，主要加入了体现top-down计算策略的“联想记忆”和“整体觉知”模块。然后将Bottom-Up和top-down两者加工策略相汇合，整合形成实时动态场景中关注对象觉知，从而解决动态场景的视觉计算问题。我们希望这样的研究思路，能够为主动视觉的机器实现，提供一种新的计算途径。



3、视觉主动计算问题



一种视觉动态觉知的计算模型



3、视觉主动计算问题

总之，景物理解的**主动视觉**，特别是**主观意念参与**的知觉过程是与人类整个心智能力，包括**意识、情感、经验**等等在内的机能密不可分，而其中的**视觉选择性注意**是人类视觉系统能够开展主动感知活动的基础。因此，希望机器视觉也能够部分地模仿人类景物理解能力，首先必须解决**主动视觉机制**的计算实现问题。我们期待有朝一日，在机器主动视觉的计算模型及其系统应用方面，有长足的进步。