

人机交互基础教程

第4章 交互技术

中南大学计算机学院 2021年11月

内容摘要

- ▶ 人机交互输入模式
- ▶ 基本交互技术
- ▶ 二维图形交互技术
- ▶ 三维图形交互技术
- ▶ 自然交互技术

4.1 人机交互输入模式

- ▶ 由于输入设备是多种多样的，而且对一个应用程序而言，可以有多个输入设备，同一个设备又可能为多个任务服务，这就要求对输入过程的处理要有合理的模式。
 - 请求模式(Request Mode)
 - 采样模式 (Sample Mode)
 - 事件模式 (Event Mode)

▶ 请求模式

- 在请求模式下，输入设备的启动是在应用程序中设置的。应用程序执行过程中需要输入数据时，暂停程序的执行，直到从输入设备接受到请求的输入数据后，才继续执行程序。

● 范例：输入窗口

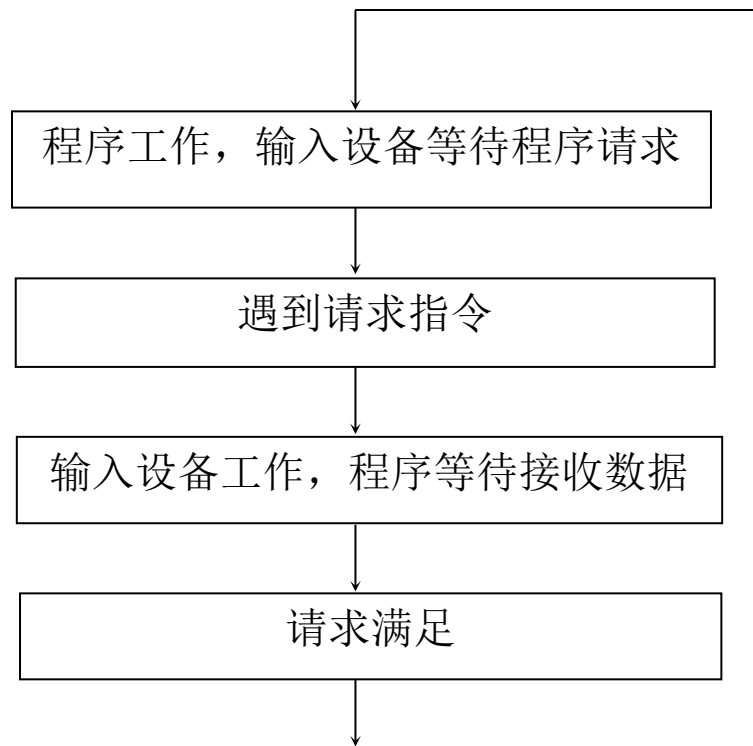


图4-1 请求模式的工作过程

▶ 采样模式

- 输入设备和应用程序独立地工作。输入设备连续不断地把信息输入进来，信息的输入和应用程序中的输入命令无关。应用程序在处理其它数据的同时，输入设备也在工作，新的输入数据替换以前的输入数据。当应用程序遇到取样命令时，读取当前保存的输入设备数据。
- 优点：这种模式对连续的信息流输入比较方便，也可同时处理多个输入设备的输入信息。
- 缺点：当应用程序的处理时间较长时，可能会失掉某些输入信息。

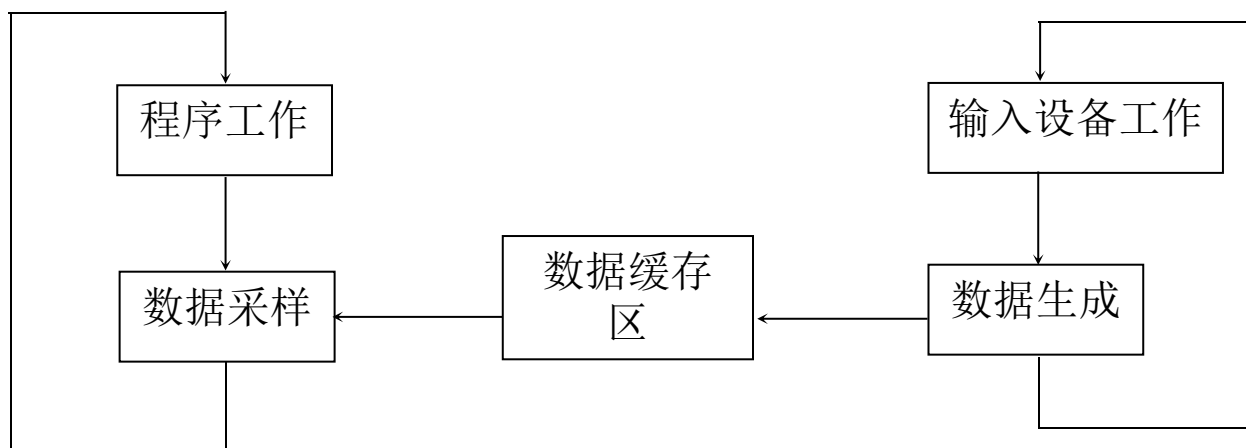


图4-2 采样模式的工作过程

● 范例：交通监控



▶ 事件模式

- 输入设备和程序并行工作。输入设备把数据保存到一个输入队列，也称为事件队列，所有的输入数据都保存起来，不会遗失。应用程序随时可以检查这个事件队列，处理队列中的事件，或删除队列中的事件。

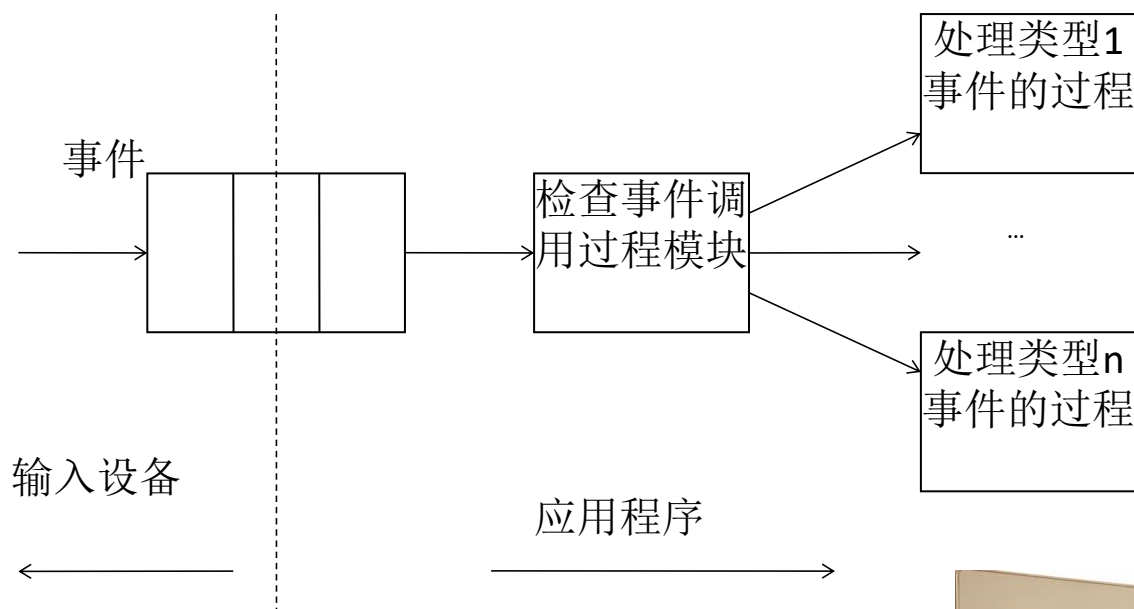


图4-3 事件模式

● 范例：门禁系统



4.2 基本交互技术

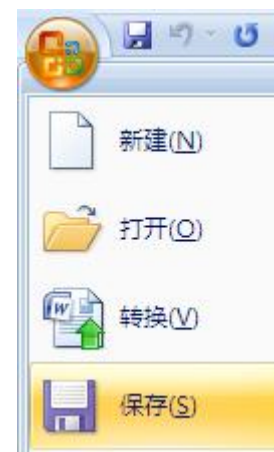
- ▶ 定位
- ▶ 笔划
- ▶ 定值
- ▶ 选择
- ▶ 字符串输入

▶ 定位

- 确定平面或空间的一个点的坐标，是交互中最基本的输入技术之一。
- 直接定位：用定位设备直接指定某个对象的位置，是一种精确定位方式。
- 间接定位：通过定位设备的运动控制屏幕上的映射光标进行定位，是一种非精确定位方式。其允许指定的点位于一个坐标范围内，一般用鼠标等指点设备配合光标来实现。

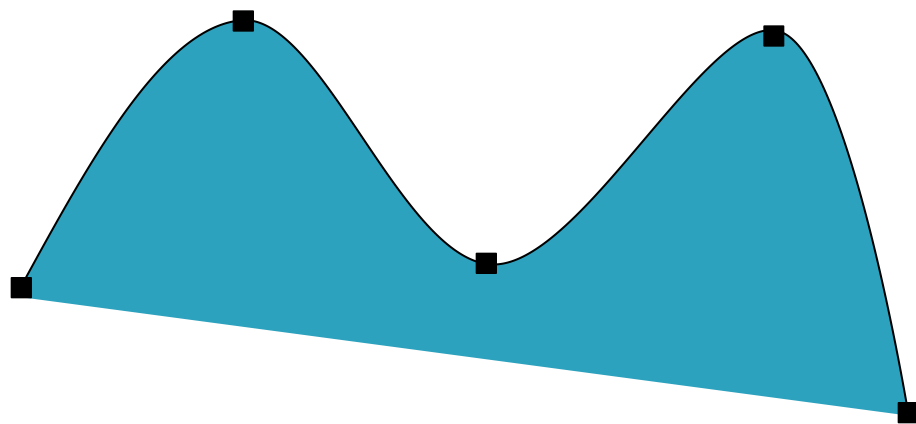


3DS Max中的精确定位



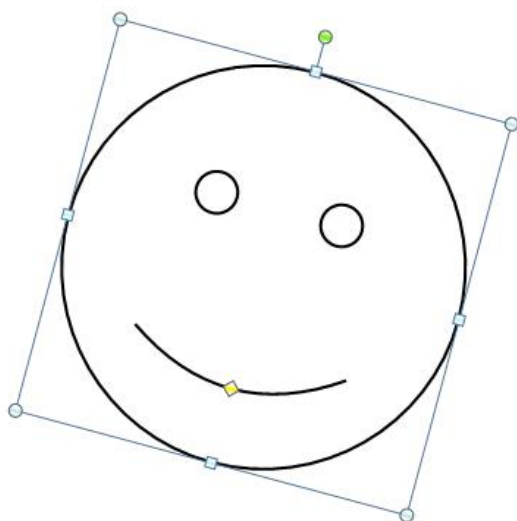
▶ 笔划

- 笔划输入用于输入一组顺序的坐标点。它相当于多次调用定位输入，输入的一组点常用于显示折线或作为曲线的控制点。



▶ 定值

- 定值（或数值）输入用于设置物体旋转角度、缩放比例因子等。



▶ 选择

- 单个元素选择
- 区域选择
 - 选框工具
 - 套索工具
 - 快速选择工具
 - 懒惰选择工具

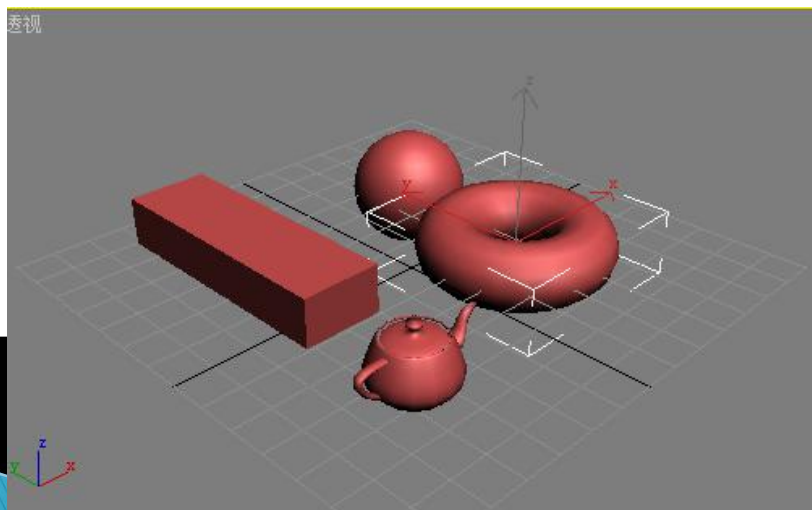
▶ 单个元素选择

- 单个元素选择是在某个选择集中选出一个元素，通过注视、指点或接触一个对象，使对象成为后续行为的焦点，是操作对象时不可缺少的一部分。

- 键盘 Ctrl+A
- 鼠标



选择

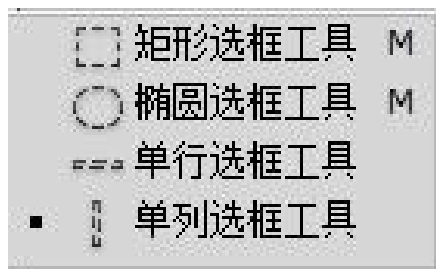


▶ 区域选择

- 区域选择是在选择集中选出一组元素或者选择一个区域，通过使用区域选择工具完成该交互操作。
 - 选框工具
 - 套索工具
 - 快速选择工具
 - 懒惰选择工具

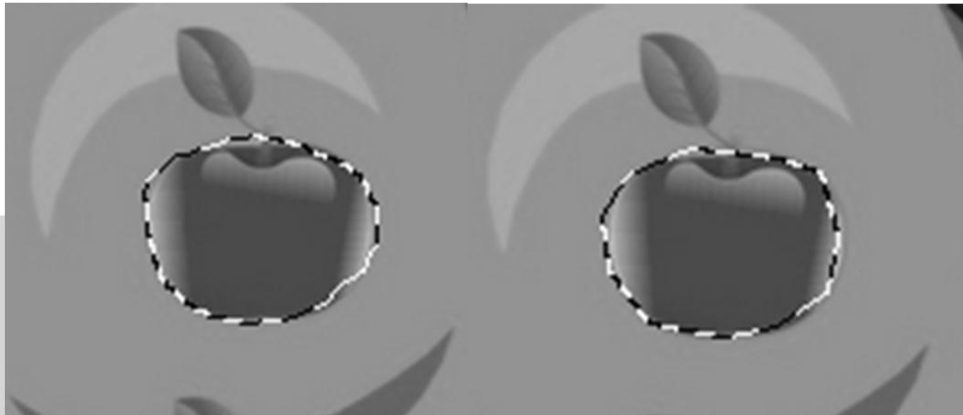
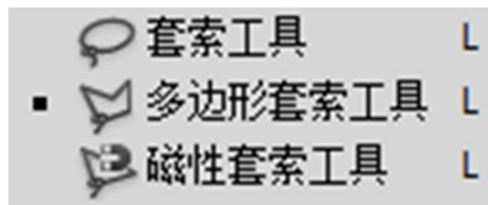
▶ 选框工具

- 目前常用的选框工具有矩形选框工具和椭圆选框工具，凡是和选框工具选择区域相交的元素均会被选择。该方法交互简单，但是选择对象往往不够精确



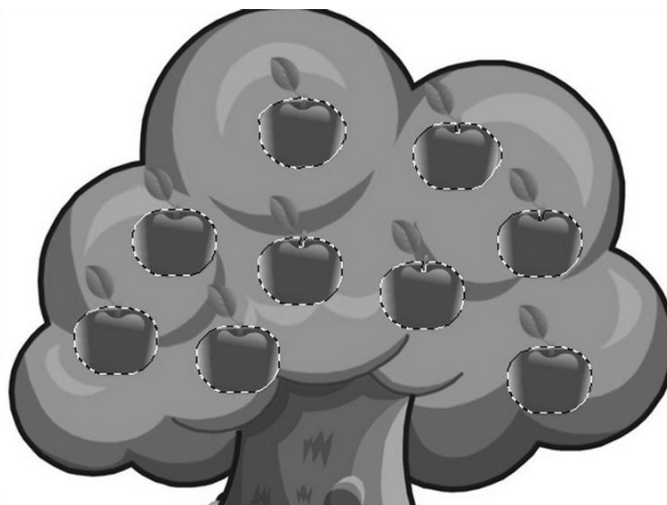
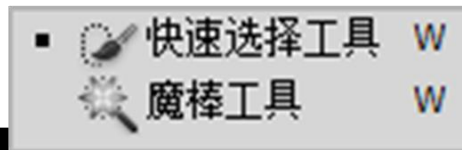
▶ 套索工具

- 为了使得选择区域更加精确，套索工具所勾画的封闭区域范围内的所有元素均被选中，该方法需要用户进行精细选择才能得到较为精确的选择结果，交互量比较大。



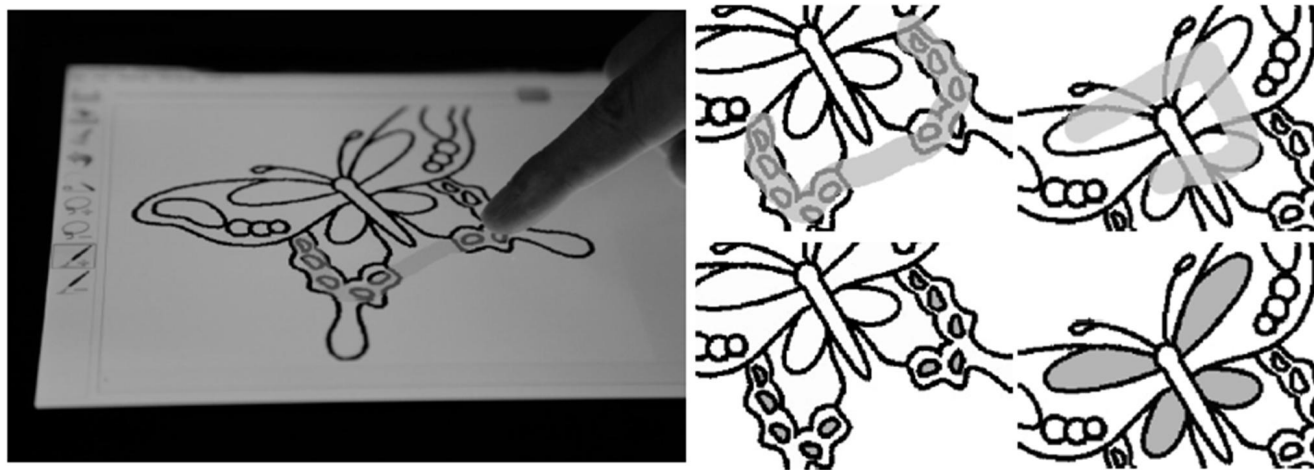
▶ 快速选择工具

- 为了提高选择精度且减少交互量，目前软件系统提供快速选择工具，这些工具采用快速选择算法，通过选择和选择点颜色相近的区域作为选择元素。



▶ 懒惰选择工具

- 懒惰选择方法，可以依据选择对象的形状、位置等关系信息，对用户手指划过的区域对象进行智能选择，但是这种交互技术尚不成熟，目前尚未广泛应用。



▶ 字符串

- 键盘是目前输入字符串最常用的方式，现在用写字板输入字符也已经很流行。
- 语音输入也是字符串输入以及功能选择的一种输入方法，语音输入需要使用语音识别技术。



4.3 二维图形交互技术

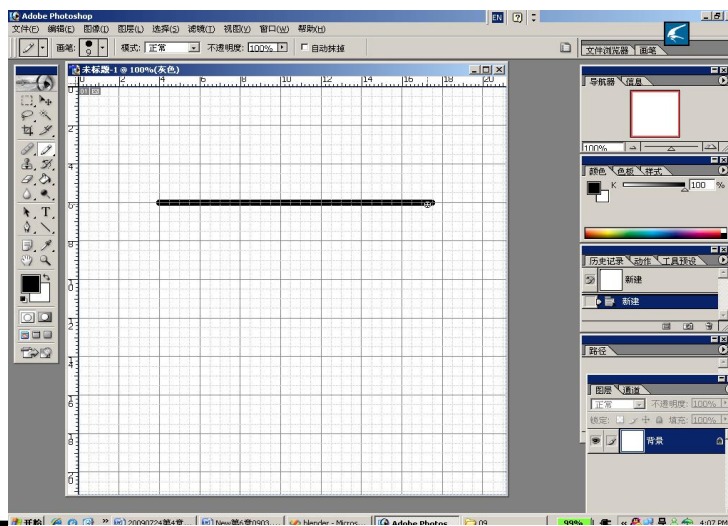
- ▶ 几何约束
- ▶ 引力场
- ▶ 拖动
- ▶ 橡皮筋技术
- ▶ 操作柄技术
- ▶ 三维交互技术

Sketchpad—人机交互的二维图形系统视频

<https://video.tudou.com/v/XMjQyNzE5MzgZNg==.html>

▶ 几何约束

- 几何约束可以用于对图形的方向、对齐方式等进行规定和校准。
- 对定位的约束（网格吸附）

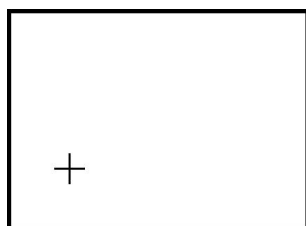


Adobe Photoshop 网格线

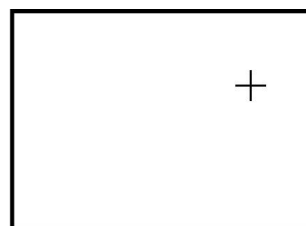
▶ 几何约束

◦ 方向约束

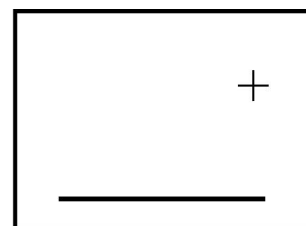
- 例如要绘的垂直或水平方向的线，当给定的起点和终点连线与水平线的交角小于 45° 时，便可绘出一条水平线，否则就绘垂直线。绘制印刷线路板、管网图或地籍图时非常有用。



移动
光标



按下按
钮开关



- 在Word绘图中，通过锁定纵横比，在拖动线段一个端点时，线段只是沿原来方向放缩

▶ 引力场

- 引力场也可以看作是一种定位约束，通过在特定图素（如直线段）周围假想有一个区域，当光标中心落在这个区域内时，就自动地被直线上最近的一个点所代替，就好像一个质点进入了直线周围的引力场，被吸引到这条直线上去一样。
- 引力场的大小要适中，太小了不易进入引力区，太大了线和线的引力区相交，光标在进入引力区相交部分时可能会被吸引到不希望选的线段上去，增大误接的概率。

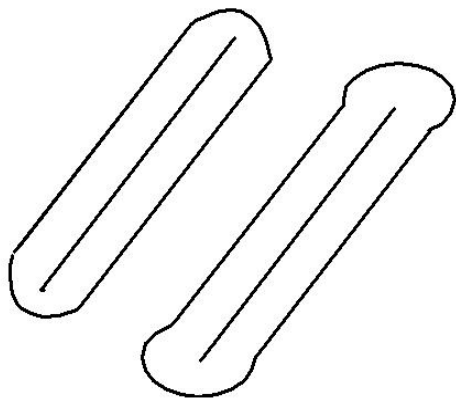


图 4-12 引力场

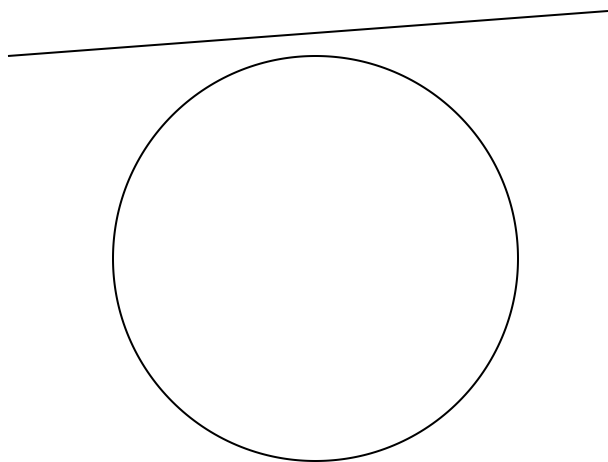
▶ 拖动

- 要把一个对象移动到一个新的位置时，如果我们不是简单地用光标指定新位置的一个点，而是当光标移动时拖动着被移动的对象，这样会使用户感到更直观，并可使对象放置的位置更恰当。
- 图形模式和图像模式



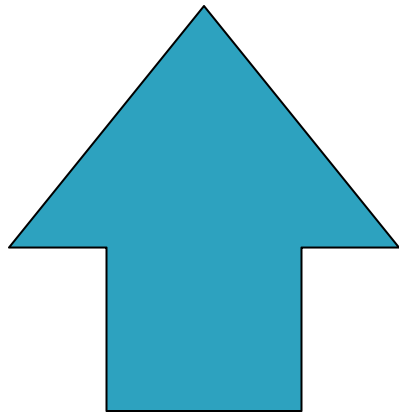
▶ 橡皮筋技术

- 被拖动对象的形状和位置随着光标位置的不同而变化。
- 不断地进行画图—擦除—画图的过程



▶ 操作柄技术

- 可以用来对图形对象进行缩放、旋转、错切等几何变换。
先选择要处理的图形对象，该图形对象的周围会出现操作柄，移动或旋转操作柄就可以实现相应的变换。



4.4 三维图形交互技术

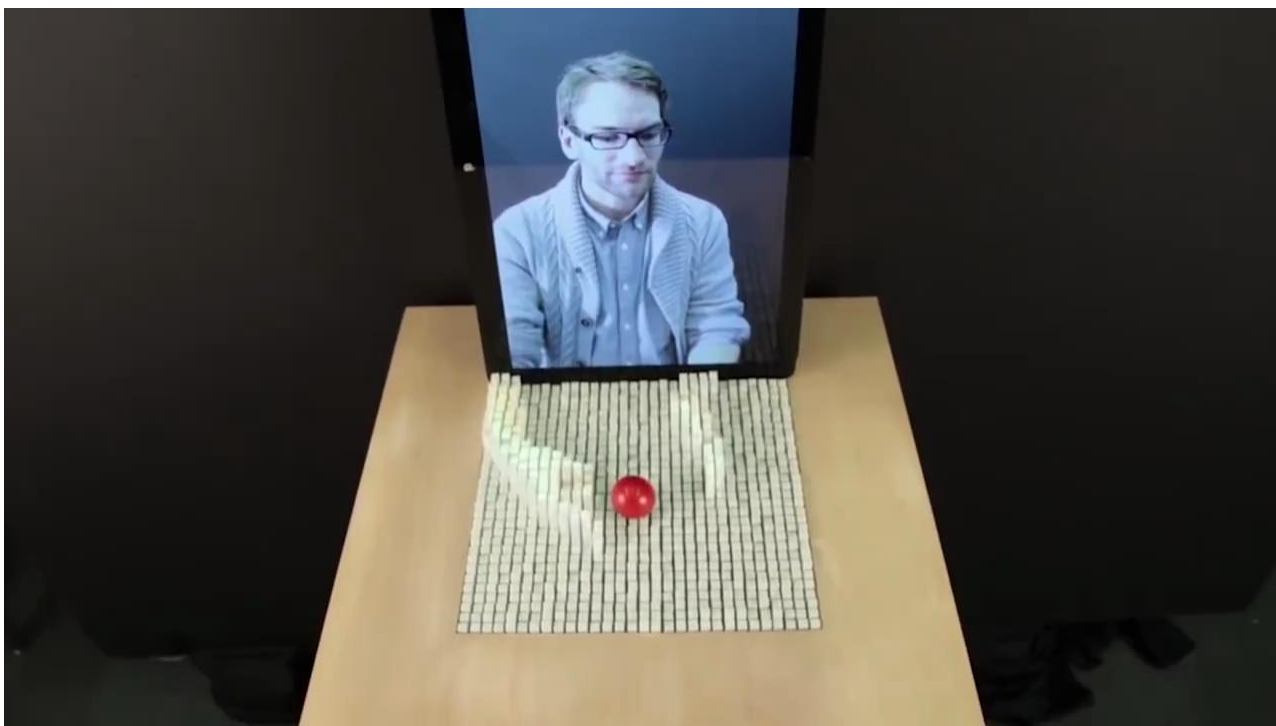
- ▶ 直接操作
- ▶ 三维Widgets
- ▶ 三视图输入



AutoCAD 2017三维图形交互功能的演示

一个三维图形查看功能的演示

<https://v.qq.com/x/page/a08192pmszc.html>



神奇的三维交互系统，将虚拟和现实直接连接

更多解释见

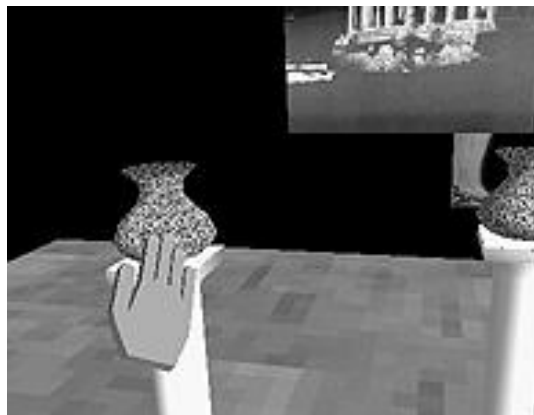
http://www.sohu.com/a/241235077_721246

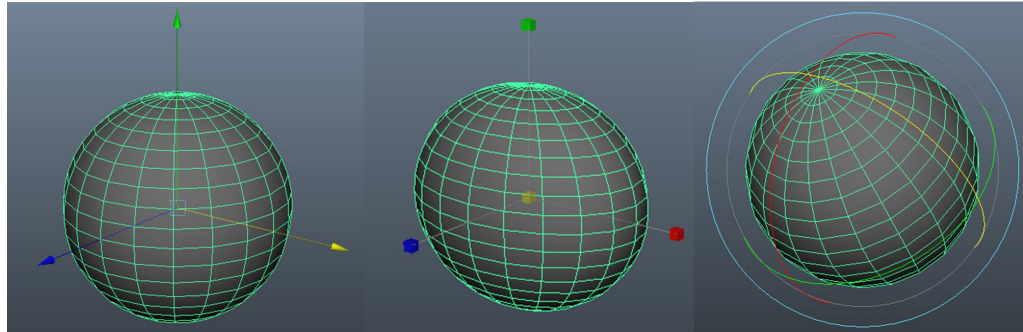
▶ 面临的问题

- 三维交互技术采用六自由度输入设备。所谓六自由度，指沿三维空间X、Y、Z轴平移和绕X、Y、Z轴旋转，而现在流行的用于桌面型图形界面的交互设备，如鼠标、轨迹球、触摸屏等只有两个自由度(沿平面X、Y轴平移)。
- 窗口、菜单、图符和传统的二维光标在三维交互环境中会破坏空间感，用户难以区分屏幕上光标选择到对象的深度值和其他显示对象的深度值，使交互过程非常不自然。

▶ 直接操作

- 三维光标必须有深度感，即必须考虑光标与观察者距离，离观察者近的时候较大，离观察者远的时候较小。
- 为保持三维用户界面的空间感，光标在遇到物体时不能进入到或穿过物体内部。
- 为了增加额外的深度线索，辅助三维对象的选择，可以采用半透明三维光标。
- 三维光标可以是人手的三维模型





▶ 三维Widgets

- 三维交互界面中的一些小工具
 - 三维空间中漂浮的菜单、用于拾取物体的手的三维图标、平移和旋转指示器等。
- 1992年美国Brown大学计算机系提出三维Widget设计原则
 - 三维Widget的几何形状应能表示其用途（eg:一个用来扭曲物体的Widget，最好本身就是一个扭曲的物体）
 - 适当选择Widget控制的自由度—由于三维空间有六个自由度，有时会使三维交互操作变得过于复杂，因此在用户使用某种Widget时，可以固定或者自动计算某些自由度的值
 - 根据三维用户界面的用途确定Widget的功能。例如，用于艺术和娱乐的Widget，只要能够完成使画面看起来像的操作就可以了；用于科学和制造的用户界面，则必须保证交互操作参数的精确性。

▶ 三视图输入

◦ 用二维输入设备在一定程度上实现三维的输入。

- 如果输入一个三维点，只要在两个视图上把点的对应位置指定后便唯一确定了三维空间中的一个点；
- 把直线段上两端点在三视图上输入后便可决定三维空间的一条直线；
- 把一个面上的各顶点在三视图上输入后，也唯一确定了三维空间中的一个面；
- 如果把一个多面体上的各面均用上述方法输入，也就在三维空间中输入了一个多面体。

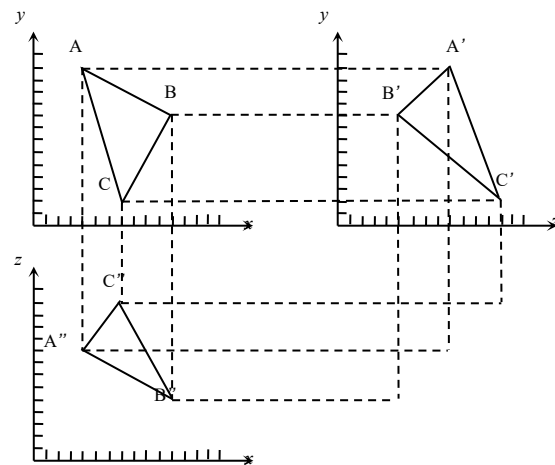
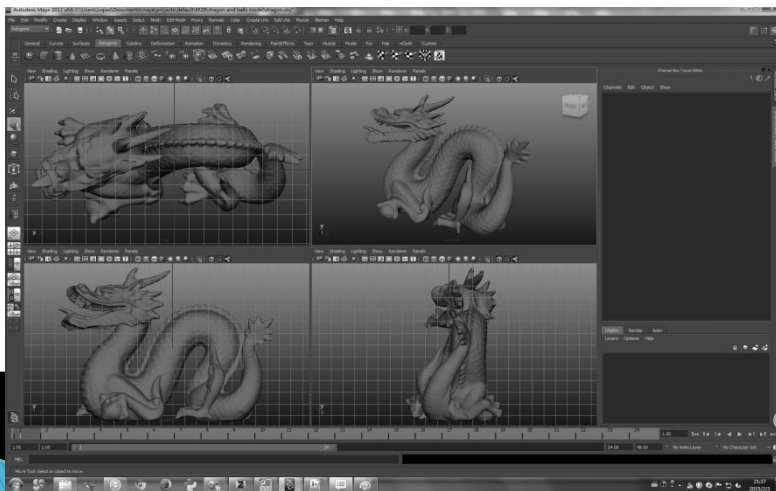



图4-30 用三视图输入三维图形

4.5 自然交互技术

- ▶ 多点触控技术
 - ▶ 手势识别技术
 - ▶ 表情识别技术
 - ▶ 语音交互技术
 - ▶ 眼动跟踪技术
 - ▶ 笔交互技术
- 

▶ 多点触控技术

- 多点触控技术是指借助光学和材料学技术，构建能同时检测多个触点的触控平台，使得用户能够运用多个手指同时操作实现基于手势的交互，甚至可以让多个用户同时操作实现基于协同手势的交互。
- 多点触控技术由硬件和软件两部分组成。
- 多点触控软件技术：
 - 触点检测和定位
 - 手指触点跟踪
 - 触摸手势识别

▶ 多点触控技术

◦ 触点检测和定位

- 图像预处理过程包括：灰度变换、平滑去噪、去除背景、图像分割等。
- 手指触点分割，如背景减除法。即将当前帧图像与背景图像相减，若差分图像中某个像素的灰度值大于某个阈值，则判断该像素点属于运动目标区域，即触点，否则属于背景区域。
- 触点定位：1) 对所有分割后的触点区域提取出其外轮廓，并对轮廓图进行筛选，把面积小于一定大小和外形不是凸包的触点轮廓去掉，保留真正的触点目标；2) 基于触点轮廓计算手指触摸点的信息，如重心坐标等，完成触点定位。

▶ 多点触控技术

◦ 手指触点跟踪

- 多点触控系统检测和定位出多个触点后，需要对每个触点进行跟踪，记录每个触点的轨迹信息，再做基于轨迹的动态手势识别，才能实现基于手势的自由交互。触点跟踪方法有Meanshift 算法、Kalman滤波、Kuhn-Munkres算法及CamShift算法等。

◦ 触摸手势识别

- 多点触控交互桌面上使用的多为单手多指手势，或者双手对称手势等
- 单手多指手势和双手对称手势识别，是在触点检测与稳定跟踪的基础上，通过标记、分析触点轨迹，识别手势含义。。

▶ 手势识别技术

- 手势具有生动、形象和直观的特点，具有很强的视觉效果，是一种自然的人机交流模式。根据不同的应用目的，手势可以分为控制手势、对话手势、通信手势和操作手势。
- 手势识别按照手势输入设备可以分为两类：
 - 以数据手套为输入设备的手势识别系统：
 - 数据手套反馈各关节的数据，并经一个位置跟踪器返回人手所在的三维坐标，从而来测量手势在三维空间中的位置信息和手指等关节的运动信息。
 - 以摄像机为输入设备的手势识别系统
 - 主要问题：手势分割、特征提取和选择以及手势识别。
- 手势识别,实现手势交互

CEVA及 eyeSight 自然用户界面手势识别演示

手势识别算法新突破, 新版Fingo SDK可实现双手

▶ 手势识别技术

◦ 手势分割

- 三类分割方法：一是基于直方图的分割，即阈值法；二是基于局部区域信息的分割；三是基于颜色等物理特征的分割方法。

◦ 特征提取和选择

- 常用的手势特征有：轮廓、边缘、图像矩、图像特征向量以及区域直方图特征等。

◦ 手势识别

- 模板匹配技术：它将待识别手势的特征参数与预先存储的模板特征参数进行匹配，通过测量两者之间的相似度来完成识别任务。
- 统计分析技术：通过统计样本特征向量来确定分类器的基于概率统计理论的分类方法。
- 神经网络技术：该技术具有自组织和自学习能力，具有分布性特征和处理不完整模式以及具有模式推广能力。

▶ 表情识别技术

- 从表情识别过程来看，表情识别可分为三部分：人脸图像的获取与预处理、表情特征提取和表情分类。
- 人脸面部表情识别特征主要方法：
 - 灰度特征：从表情图像的灰度值上提取的，利用不同表情用不同灰度值来得到识别的依据。
 - 运动特征：利用了不同表情情况下人脸的主要表情点的运动信息来进行识别。
 - 频率特征：利用了表情图像在不同的频率分解下的差别进行识别，速度快是其显著特点。

“MOTek” 工作室展示视频捕捉表情技术

▶ 表情识别技术

◦ 特征识别方法：

- 整体识别法：无论是从脸部的变形出发还是从脸部的运动出发，都是将表情人脸作为一个整体来分析，找出各种表情下的图像差别；局部识别法就是将人脸的各个部位在识别时分开，也就是说各个部位的重要性是不一样的。
- 形变提取法：根据人脸在表达各种表情时的各个部位的变形情况来识别的；运动法是根据人脸在表达各种特定的表情时一些特定的特征部位都会作相应的运动这一原理来识别的。
- 几何特征法：根据人的面部的各个部分的形状和位置(包括嘴、眼睛、眉毛、鼻子)来提取特征矢量，这个特征矢量来代表人脸的几何特征，根据这个特征矢量的不同就可以识别不同的表情；在容貌特征法中，主要是将整体人脸或者是局部人脸通过图像的滤波，以得到特征矢量。

▶ 语音交互技术

- 语音识别（Speech Recognition）是计算机通过识别和理解过程把语音信号转变为相应的文本文件或命令的技术，其所涉及的领域包括：信号处理、模式识别、概率论和信息论、发声机理和听觉机理、人工智能等。
- 一个完整的语音识别系统大致可分为语音特征提取，声学模型与模式匹配，以及语言模型与语义理解三部分。

▶ 语音交互技术



▶ 语音识别系统的组成

◦ 语音特征提取

- 从语音信号中提取语音的特征，既可以获得语音的本质特征，也起到数据压缩的作用。
- 输入的模拟语音信号首先要进行预处理，包括预滤波、采样和量化、加窗、端点检测、预加重等。

◦ 声学模型

- 声学模型对应于语音到音节概率的计算。在识别时将输入的语音特征同声学模型（模式）进行匹配与比较，得到最佳的识别结果。
- 目前采用的最广泛的建模技术是隐马尔科夫模型HMM建模和上下文相关建模。

▶ 眼动跟踪技术

◦ 眼动主要有三种形式：

- 注视（Fixations）：表现为视线在被观察目标上的停留，这些停留一般至少持续100ms~200ms。
- 跳动（Saccades）：注视点间的飞速跳跃，是一种联合眼动（即双眼同时移动），其视角为1到40度，持续时间为30到120ms，最高速度为400-600度/秒。
- 平滑尾随跟踪（Smooth Pursuit）：缓慢、联合追踪的眼动通常称为平滑尾随跟踪。平滑尾随追踪通常有一个缓慢移动的目标，在没有目标的情况下，一般不能执行。
- 视线跟踪技术中常用的主要参数有：注视次数、注视持续时间、注视点序列、第一次到达目标区的时间等。

▶ 眼动跟踪技术

◦ 视线跟踪基本原理：

- 由光源发出的光线经红外滤光镜过滤后只有红外线可以通过；红外线经过半反射镜后，部分到达反射镜，经反射镜发射到达眼球；眼球对红外线的反射光经同一反射镜到达能锁定眼睛的特殊的瞳孔摄像机，通过连续的记录从人的眼角膜和瞳孔反射的红外线，然后利用图像处理技术，得到眼球的完整图像；再经软件处理后获得视线变化的数据，达到视线跟踪的目的。
- 目前视线跟踪技术按其所借助的媒介分为以硬件为基础和以软件为基础两种。

▶ 眼动测量方法

- 瞳孔—角膜反射向量法。通过固定眼摄像机获取眼球图像，利用亮瞳孔和暗瞳孔的原理，提取出眼球图像内的瞳孔，利用角膜反射法校正眼摄像机与眼球的相对位置，把角膜反射点数据作为眼摄像机和眼球的相对位置的基点，瞳孔中心位置坐标就表示视线的位置。
- 眼电图法。眼球在正常情况下由于视网膜代谢水平较高，因此眼球后部的视网膜与前部的角膜之间存在着一个数十毫伏的静止电压，角膜区为正，视网膜区为负。当眼球转动时，眼球的周围的电势也随之发生变化；将两对氯化银皮肤表面电极分别置于眼睛左右、上下两侧，就能引起眼球变化方向上的微弱电信号，经放大后得到眼球运动的位置信息。

▶ 眼动测量方法

- 虹膜—巩膜边缘法。在眼部附近安装两只红外光敏管，用红外光照射眼部，使虹膜和巩膜边缘处左右两部分反射的光被两只红外光敏管接收。当眼球向左或向右运动时，两只红外光敏管接收的红外线会发生变化，利用这个差分信号就能测出眼动。
- 角膜反射法。角膜能反射落在它上面的光，当眼球运动时，光以变化的角度射到角膜，得到不同方向上的反光。角膜表面形成的虚像因眼球旋转而移动，实时检测出图像的位置，经信号处理可得到眼动信号。

▶ 眼动测量方法

- 双普金野象法。普金野图像是由眼睛的若干光学界面反射所形成的图像。角膜所反射出来的图像是第一普金野图像，从角膜后表面反射出来的图像微弱些，称为第二普金野图像，从晶状体前表面反射出来的图像称为第三普金野图像，由晶状体后表面反射出来的图像称为第四普金野图像。通过对两个普金野图像的测量可以确定眼注视位置。
- 接触镜法（ContactLens）。是将一块反射镜固定在角膜或巩膜上，眼球运动时将固定光束反射到不同方向，从而获得眼动信号。

▶ 笔交互技术

- 笔式输入具有连续性、使用笔的连续线条绘制可以产生字符、手势或者图形等特点。其优点是便于携带，输入带宽信息量大，输入延迟小；其缺点是翻译困难，再现精度低。
- 手写识别技术是笔交互中的一种基本技术，目前已经嵌入到各种设备中，得到广泛应用。

联机手写汉字识别：所处理的手写文字是书写者通过物理设备 (如数字笔、数字手写板或者触摸屏) 在线书写获取的文字信号, 书写的轨迹通过定时采样即时输入到计算机中.

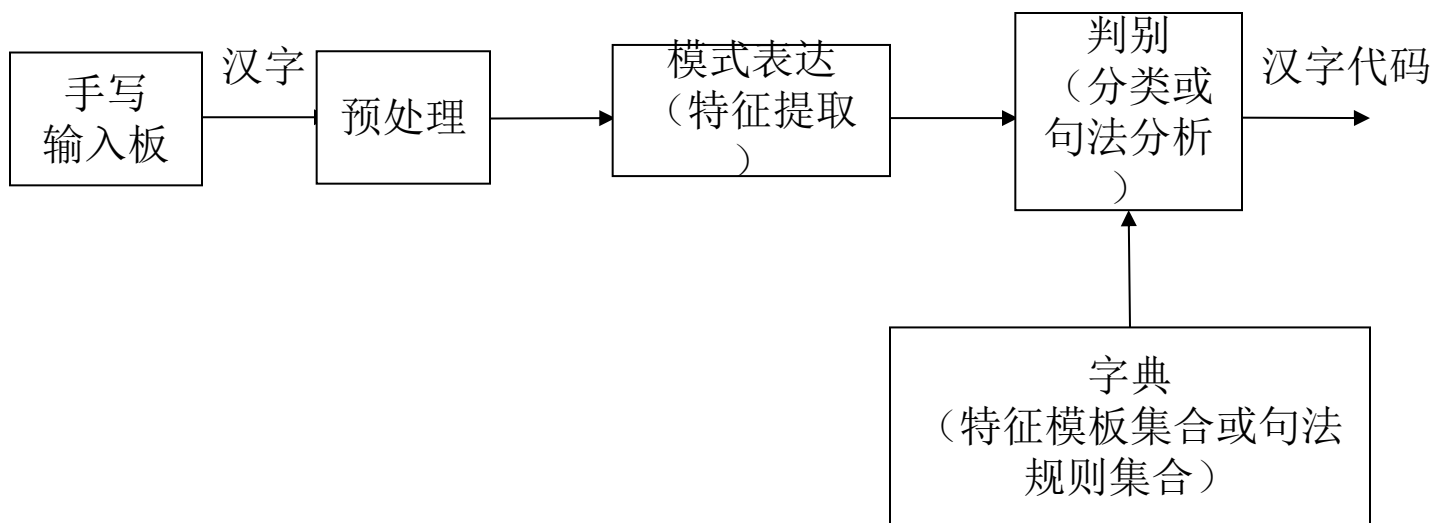
脱机手写文字识别：所处理的手写文字是通过扫描仪或摄像头等图像捕捉设备采集到的手写文字二维图片。

Online recognition: 联机识别, 基于笔画轨迹

Offline recognition: 脱机识别, 基于图像

▶ 联机手写识别

- 联机手写文字的识别过程通常分为四个阶段：预处理、特征抽取、特征匹配和判别分析

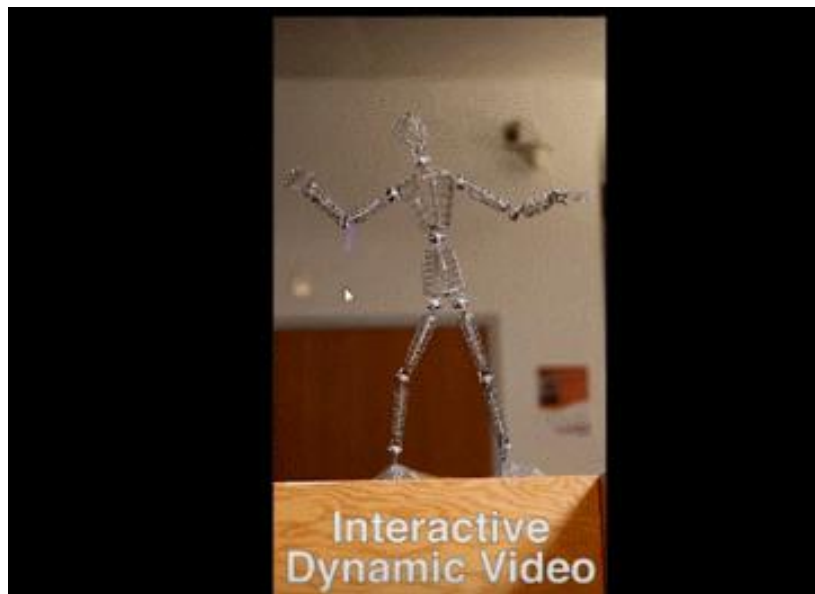


▶ 脱机手写识别

- 脱机手写识别比印刷体汉字识别、联机手写体识别都要困难。
- 脱机手写识别得到的描述则是点阵图像，要得到笔段的点阵通常需要细化运算。
- 细化会损失一些信息，并且不可能得到时间顺序信息。
- 脱机识别中，笔画与笔画之间经常粘连，很难拆分，而且笔段经过与另一笔段交叉分成两段后，也难以分清是否应该连起来。

动态交互视频技术

严格来说，视频并不是一个交互媒介，因为观众只是接收讯息，并不能跟视频里的画面实时互动。但是麻省理工学院（MIT）正在改变这一现实：他们研发了一项新技术，让观众可以伸出手“触碰”并操控影像里的物体，就像在真实世界里一样。如下图里一个小人，本来是静止的画面，但是如果你在上面点击并拖动鼠标，小人真的会因为你的动作而出现抖动或倾斜。这个技术可以让用户在设备上通过触摸对象来实现现实与虚拟之间的互动，也就是我们所说的MR（Mixed Reality，混合现实）



“动态交互视频”（Interactive Dynamic Video），简称IDV

► Interactive Dynamic Video演示视频:

<http://player.youku.com/embed/XMTY2OTY0NDc5Ng==>

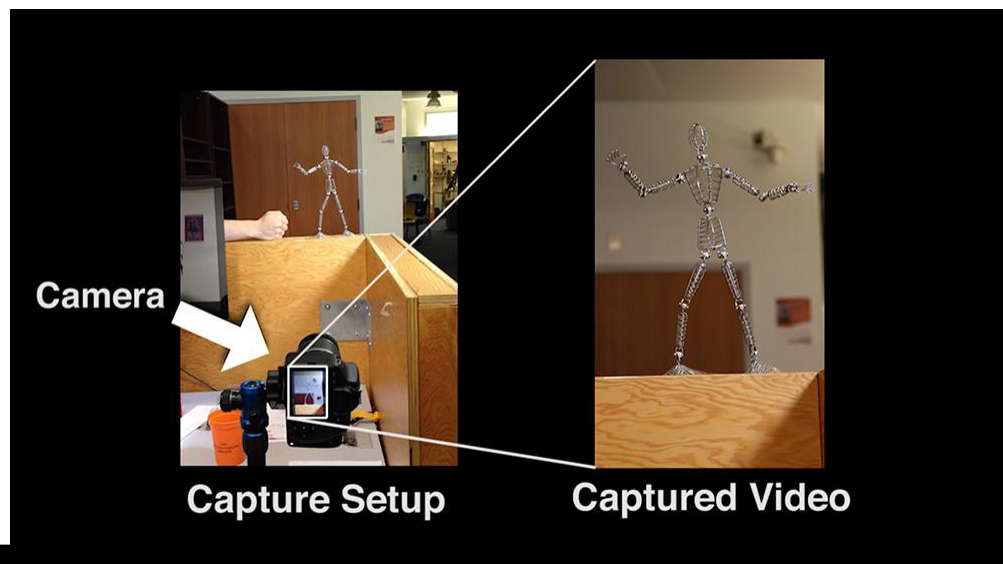
IDV系统可以捕捉到非常细小的物体振动频率。甚至一阵微风引起的振动也可以捕捉到，所以对于巨大的建筑物，这种模拟也可以实现。一般来说，模拟物体的运动，是需要3D建模的，但是这一过程非常昂贵，这套系统里根本不需要3D轮廓重建的步骤。而IDV只需要5秒钟的传统视频和一套算法，就可以实现运动模型的预测。

“这个技术可以让我们捕捉到物体的物理动态，让我们在虚拟空间中与拍摄的物体玩耍。”

IDV技术原理：5秒视频+振动模式+算法：

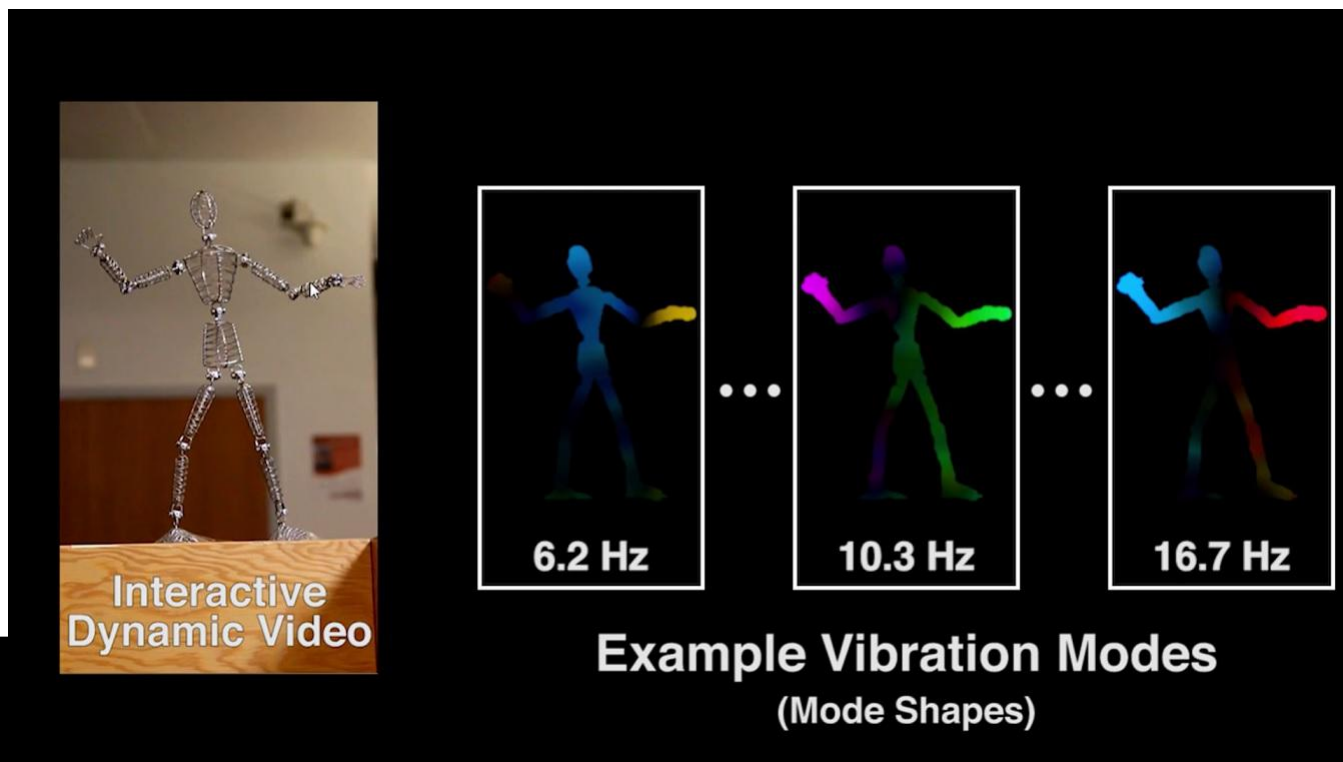
IDV系统的技术原理是这样的：每一个真实世界的物体都会受到来自外界的作用力，由此引发物体的振动。研究团队用摄像机拍摄该物体最短5秒的视频，然后用一个算法生成一个预测模型，预测出该物体在其他外力的作用下产生的振动模式。

所以上图中的室内的小人，在拍摄过程中，研究人员会在它站立的桌面上使劲拍打，使其产生振动。而这种振动，我们人类的肉眼几乎不可见。



IDV技术原理：5秒视频+振动模式+算法：

振动模式获取过程：首先他们会先让目标物体颤动起来，并用相机记录这种颤动。之后，他们特别开发的这套算法就可以通过分析物体颤动的频率，分析出这个物体在现实场景中可以进行何种运动，以及运动的幅度。

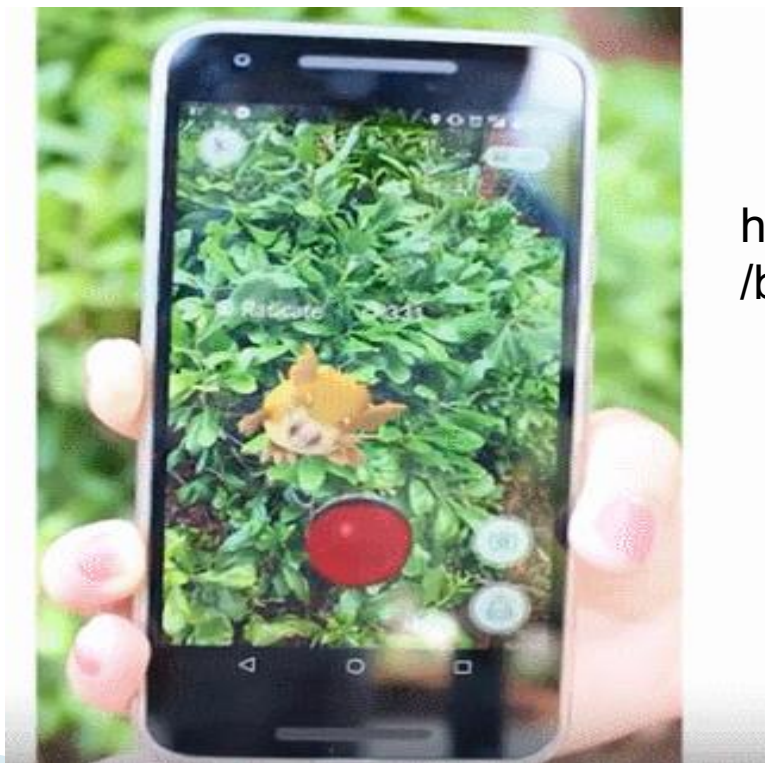


IDV应用领域

IDV的应用非常广泛，比如工程师可以利用IDV来模拟在遭遇大风或地震等灾害时的一座建筑物的状况，而对于CG电影动画制作者来说，制造一个虚拟的白宫或大桥倒塌的壮观景象，也无疑更加容易了。

但其中最令人激动的就是和AR的结合。以如今大火的Pokemon Go为例，在现在的游戏界面里，小精灵只是单纯“出现”在画面里，就像一个贴纸一样，如果你拨动这个小拉达后面的物体，它并不会受到任何影响。

<https://v.qq.com/x/page/y0312d8t724.html>



<https://v.qq.com/x/page/b3259l4l218.html>

而IDV可以让虚拟的动画角色跟真实的环境交互。一个虚拟的皮卡丘从草丛里一下子蹦出来，并带动了草丛里的草，这效果就太逼真了，形成了所谓的“动态AR”。



习题

- ▶ 请列出你所熟悉的软件系统（例如Microsoft Office）中涉及到的交互技术，若有本章中没有提及的交互技术，则对其进行进一步分析。
- ▶ 简述PhotoShop中的二维图形交互技术。
- ▶ 针对某一三维交互软件，简述其中的三维图形交互技术。
- ▶ 测试一下Windows XP自带的语音识别程序，看其识别率能达到多少。
- ▶ 在Windows XP操作系统下，使用Microsoft Word，可以通过移动鼠标使用手写体输入文字。测试一下识别率



