

# 交互式动态体三维显示关键技术

潘文平, 沈春林, 蔡亮, 邢建芳, 李莉

南京航空航天大学自动化学院, 南京 210016

**摘要** 研究了交互式动态体三维显示的关键技术, 包括三维数据体素化、体素激活、体空间生成和人机交互等, 实现了体三维显示系统样机。其中, 采用体立方栅格(BCC)采样策略以减少体素化生成的体显示数据, 以数字微镜器件(DMD)作为空间光调制器(SLM), 并针对DMD设计了二值切片的快速激活算法; 分析了基于旋转螺旋面的体三维空间生成方法, 并将基于隐马尔科夫模型(HMM)的动态手势控制应用于交互式体三维显示系统。针对原理样机设计的各组实验表明, BCC体素化策略减少了40%以上的体显示数据, 基于DMD的切片快速激活算法使得在500mm×250mm半圆柱形成像空间内显示的三维图像刷新率达16Hz, 并且可以选择任意视点裸眼观看, 准确率超过93%、响应时间低于40ms的动态手势识别算法能够较好地实现实时交互式体三维显示。

**关键词** 体三维显示; 数字微镜器件; 手势识别

**中图分类号** TP391, TN141

**文献标识码** A

**doi** 10.3981/j.issn.1000-7857.2011.01.05

## Key Technologies of Interactive Dynamic Volumetric 3D Display

PAN Wenping, SHEN Chunlin, CAI Liang, XING Jianfang, LI Li

College of Automation Engineering, Nanjing University of Astronautics and Aeronautics, Nanjing 210016, China

**Abstract** Key technologies of dynamic Volumetric Three-Dimensional (V3D) display is implemented in our prototype, which include 3D data voxelization, voxel activation, volumetric space creation, and human-computer interface, where 3D models are voxelized through Body-Cubic Center (BCC) sampling strategy for data reduction. Digital Micro-mirror Device (DMD) is used as the Spatial Light Modulator (SLM) and an algorithm is designed for binary slice projection based on DMD. An approach for volumetric space creation based on a rotating helical screen is also adopted. Furthermore, a hand gesture control-based HCI subsystem for the prototype is designed and implemented. Experimental results demonstrate that, a data reduction of more than 40% is achieved through utilizing BCC sampling strategy, and objects can be displayed with true three dimensions at a refresh rate of 16Hz in a half-cylinder display space with a diameter of 500mm and a height of 250mm, and can be viewed from almost any angle without any special viewing aids. With the hand gesture control subsystem, the average recognition rate is as high as 93%, the average response time is less than 40ms, and the real-time interactions can be well performed from an arbitrary viewpoint.

**Keywords** volumetric three-dimensional display; digital micro-mirror device; hand gesture recognition

### 0 引言

自然界中任何物体都具有三维物理尺寸, 然而传统的二维平面显示器无法呈现出物体真实的深度信息, 只能达到伪三维的显示效果。与传统的二维平面显示器如CRT、LCD等不同, 体三维显示系统可以呈现具有真实物理深度的三维图

像<sup>[1]</sup>。而且有别于立体及自动立体显示系统, 体三维显示系统可以使多个观察者在无须佩戴任何辅助装置的情况下, 从360°范围内任意视角同时观察被显示的物体, 并且每个观察者看到被显示物体的不同侧面。体三维显示作为三维数据可视化的重要手段, 在计算机辅助设计、辅助放射治疗规划、空

收稿日期: 2010-08-17; 修回日期: 2010-11-17

基金项目: 国家高技术研究发展计划(863计划)项目(2007AA01Z338)

作者简介: 潘文平, 博士后, 研究方向为体三维显示和人机交互, 电子信箱: paninxian@gmail.com; 沈春林(通信作者, 中国科协所属全国学会个人会员登记号: E280411838M), 教授, 研究方向为体三维显示、低空飞行与突防, 电子信箱: clshenac@nuaa.edu.cn

中交通管制、体育和影视娱乐等众多领域具有广阔的应用前景<sup>[1]</sup>。近年来,国外先后制成的体三维显示系统样机包括:德国的 Felix 系统<sup>[2]</sup>、美国的 Perspecta 系统<sup>[3]</sup>和 DepthCube 系统<sup>[4]</sup>。在国内,浙江大学开发过基于 LED 的主动发光式体三维显示系统<sup>[5]</sup>,南京航空航天大学先后制成基于振镜的体三维显示系统、基于旋转平板的体三维显示系统<sup>[6]</sup>。体三维显示按实现方式的不同,可分为静态体显示和动态体显示。

最早的静态体显示技术是 Swainson 提出的基于固体介质能量跃迁的体显示技术<sup>[7]</sup>。此技术利用两只红外激光器同步扫描充满固体介质的三维成像空间,在两道光线的交汇处固体介质发生能级跃迁而发出可见荧光。另一种基于能级跃迁的静态体显示技术利用气体介质构成成像空间<sup>[8]</sup>,当两束激光同时照射在同一个气体原子上时,原子发生能级跃迁而发出可见荧光。此外,文献[9]选用了液晶材料作为成像介质。静态体显示技术使用的激光束容易对人眼造成伤害,并且图像的清晰度受到激光器扫描速度的限制。

最早的动态体显示技术是 Ketchpel 提出的基于阴极射线球的体显示技术<sup>[1]</sup>。此技术将荧光物质镀在旋转屏幕上,利用电子射线束扫描处于真空容器中的旋转屏幕。当电子射线束的扫描时序与屏幕旋转位置同步时,旋转屏幕构成的成像空间即可显示出三维图像。电子射线束的扫描速度、真空容器壁对光线的折射、荧光物质的发光启动时间和余晖等都是影响图像质量的重要因素。另一种动态体显示技术利用旋转发光二极管(LED)阵列构建成像空间<sup>[5]</sup>。当每个 LED 的发光时序与旋转面板的旋转位置同步时,在成像空间内即可显示出具有真实物理深度的三维图像。LED 阵列的密度限制、LED 的开关时间以及成像空间内体素分布的不均匀性等构成了提高图像质量的瓶颈。Actuality 公司开发的 Perspecta<sup>[3]</sup>则是另外一种动态体三维显示系统。Perspecta 利用空间光调制器(Spatial Light Modulator, SLM)技术,能以 24Hz 刷新率向高速旋转的平面成像屏投射包含多达 198 个二维切片的图像序列,在直径 10" 的柱形体空间内显示出性能优异的彩色三维图像。为克服体素重叠引起的视觉死区, Perspecta 引入的转向光学部件增加了其制作难度和成本。Perspecta 已在放射治疗等少数医学领域获得应用<sup>[10]</sup>,是目前体三维显示器中的高端产品。

另外,因体三维显示系统具有 360° 观察视角,其人机交互问题也引起了人们的关注。Perspecta 系统的作者最早关注从多方位与体三维显示系统进行人机交互的问题<sup>[11]</sup>,通过在手指上佩戴特殊颜色标记,可在 Perspecta 系统半球形外壳上实现基于视觉的空间多点触控式人机交互。Felix 系统采用 3D 鼠标,适合从固定方位与显示系统进行人机交互。DepthCube 系统则以二维平面显示器为辅助,采用普通鼠标和键盘从固定方位实现人机交互。

综合以上分析可知,动态体三维显示的关键技术包括:三维数据体素化、体素激活、体空间生成、人机交互等。本文

比较并借鉴国内外已有的技术成果,构建了基于数字微镜器件(Digital Micro-mirror Device, DMD)和旋转螺旋屏的动态体三维显示系统,并针对体三维显示的关键技术,将此系统分为体素化子系统、体素激活子系统、成像空间子系统和人机交互子系统。其中的体素化子系统采用的体立方栅格(Body-Cubic Center, BCC)采样策略则用于降低体显示数据。体素激活子系统采用 DMD 作为 SLM,设计的 DMD 控制算法适于向成像屏高速投射复杂三维物体的二值切片图像序列。成像空间子系统采用旋转螺旋屏构建成像空间,克服了旋转 LED 阵列等旋转平板式体显示方式引起的体素分布不均匀、视觉死区大等缺点。此外,针对体三维显示系统具有的 360° 的观察视角,建立了基于手势控制的人机交互子系统,使操作者能够从围绕显示系统任意角度进行人机交互。

## 1 交互式动态体三维显示系统

### 1.1 体素化子系统

体三维图像上的每个像素点都具有三维坐标,这些像素点称为体素<sup>[1](voxel)</sup>。数据体素化子系统的任务是根据成像空间构成的特点,通过体素化方法获得体素数据并输出到体素激活子系统。体素化方法需要把三维物体的几何模型通过采样,离散为最接近的体素模型。根据多维信号采样<sup>[12]</sup>的研究结果,六边形栅格是二维平面的最优采样。将六边形扩展到三维空间,则体心立方栅格 BCC 是对应三维空间的最优采样。体素化与三维重构互为逆过程,文献[13]在研究三维重构时指出,为重构得到同样精度的三维物体,按 BCC 采样策略进行体素化得到的体素数量,仅为由笛卡儿采样得到的体素数量的 70.7%。本文将 BCC 采样策略运用于体三维显示系统中,以减少体显示的数据量。

记 BCC 的采样矩阵<sup>[14]</sup>为

$$M_{\text{BCC}} = \begin{bmatrix} T & 0 & T/2 \\ 0 & 0 & T/2 \\ 0 & T & T/2 \end{bmatrix} \quad T = \frac{2\pi}{u} \quad (1)$$

其中,  $T, u$  分别为采样周期和频率。设体素空间的分辨率为  $R_x \times R_y \times R_z$ , 记采样点在采样空间的物理坐标为  $(x, y, z)$ , 逻辑坐标为  $(i, j, k)$ , 即  $i, j, k$  分别为采样点在  $x, y, z$  方向上的列、层和行的序号。其中  $0 \leq i < R_x, 0 \leq j < R_y, 0 \leq k < R_z$ 。假定行、列长度不限,层序号  $j$  为偶数和奇数时分别有

$$\begin{cases} x = iT - 0.5 + \frac{T}{2} \\ y = j\frac{T}{2} - 0.5 + \frac{T}{2} \\ z = kT - 0.5 + \frac{T}{2} \end{cases} \quad (2)$$

$$\begin{cases} x = iT - 0.5 + \frac{T}{2} \\ y = j\frac{T}{2} - 0.5 + \frac{T}{2} \\ z = kT - 0.5 + T \end{cases} \quad (3)$$

以三维物体的三角面片模型为例,可按以下方法确定其每个面片对应的可见体素。给定三角面片  $P$  和体素  $v$ ,记  $P$  所在平面方程为  $Ax+By+Cz+D=0$ ,  $v$  的中心点  $p$  的坐标为  $(x', y', z')$ ,过  $p$  点引  $P$  的垂线,垂足为  $q$ ,则  $p$  到  $P$  的欧氏距离为

$$d = \frac{|Ax'+By'+Cz'+D|}{\sqrt{A^2+B^2+C^2}} \quad (4)$$

当  $d$  小于给定的阈值且  $q$  落在三角面片内部时,将  $v$  标记为可见。

## 1.2 体素激活子系统

### 1.2.1 DMD

美国 TI 公司研制的 DMD<sup>[15]</sup> 是一种基于大规模集成电路技术和微机电系统技术的空间光调制器。DMD 由高达百万个可独立寻址和转动的微型反射镜组成,每个微镜安装在极小的铰链上,单个微镜的尺寸为  $16\mu\text{m} \times 16\mu\text{m}$ ,如图 1 所示。DMD 工作时,图像信息转化为控制信号序列载入到 RAM 中,通过 CMOS 驱动电路改变每个微镜的偏转状态。在基于 DMD 的投影装置中,当某个微镜的控制信号为 1 时,微镜绕固定轴偏转  $+12^\circ$ ,处于“开”状态,使得反射光线刚好沿光轴方向通过投影物镜到达成像屏。控制信号为 0 时,则微型反射镜偏转  $-12^\circ$ ,处于“关”状态,反射光线不能到达成像屏。1 和 0 信号的占空比,决定了图像像素的灰阶。

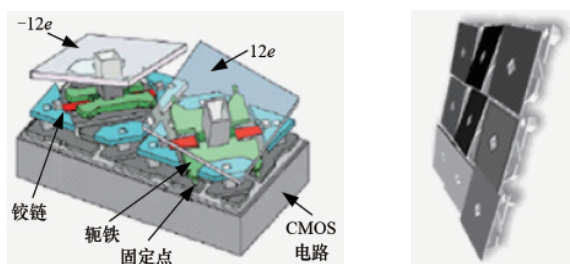


图 1 DMD 的微镜结构  
Fig. 1 Microstructure of DMD

### 1.2.2 体素激活

本文采取基于点绘制的方法对 DMD 进行控制,即通过寻址切片上每个像素点对应的微镜并设置其缓存单元的值,实现对微镜的控制。DMD 主要通过行寻址寄存器和列移位寄存器实现寻址,并通过数据分路器实现位寻址,具体步骤如下。首先行寻址寄存器指向要访问的目标行,数据分路器为 15,先寻址第 15,31, ..., 1023 列;然后数据分路器置为 14,继续寻址第 14,30, ..., 1022 列;当数据分路器为 0 时,寻址 0,16, ..., 1008 列,此时目标行对应的 1024 列寻址结束。行寄存器再指向下一目标行进行寻址,如此反复直到访问完所有微镜。每行寻址需 16 个时钟周期,寻址完整个 DMD 需  $768 \times 16$  个时钟周期,当时钟为 120MHz 时约为 0.1ms。由以上分析可知,物理上相邻的 2 个微镜在存储单元上并不相邻。当每个微镜占 1 位,即显示二值图像时,对于分辨率为  $1024 \times 768$

的 DMD,其寻址空间为 96KB,每行微镜对应 128B。物理空间中第  $i$  行  $j$  列的微镜,对应缓存单元的地址是第  $m$  字节  $n$  位,其中,

$$\begin{cases} m=128i+8(15-j\%16)+j/128 \\ n=(j\%128)/16 \end{cases} \quad (5)$$

对于二维切片图像,直线采用 Bresenham 算法进行绘制,曲线则通过多条直线拟合进行绘制,离散的点则逐点绘制。以图像数据作为成像掩膜,与对应字节进行或操作并写入相应缓存单元,实现对微镜的“开”/“关”状态设置,从而激活体素单元。

## 1.3 成像空间子系统

### 1.3.1 切片式投影

图 2 揭示了切片式投影的基本原理。成像屏可沿  $z$  轴在  $(z_{\min}, z_{\max})$  区域内做高速往复运动,扫描出长方体形的成像空间。在成像屏沿  $z$  轴正向或逆向运动的过程中,高速投影装置依次向成像屏投射三维物体的  $N$  个二维切片,成像屏在沿  $z$  轴不同的位置截获这些切片图像,在真实的三维空间形成一个包括多个切片的“切片栈”。如果成像屏往复运动的速度足够快,而且高速投影装置投射的切片数目足够大,由于视觉暂留效应,人眼便能感知到在成像空间内具有真实物理深度的立体图像。

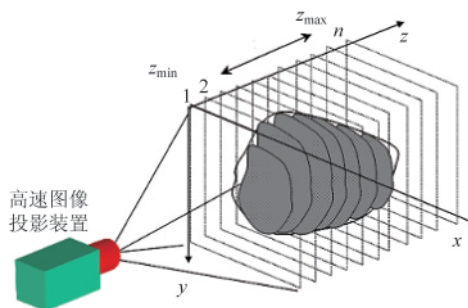


图 2 切片式投影  
Fig. 2 Slice-styled projection

### 1.3.2 基于旋转螺旋屏的成像空间子系统

切片式投影通过成像屏的高速往复运动构成成像空间,原理简单但缺点明显:在与成像屏平行的方向易形成视觉死区,高速往复运动易引起振动和噪声。为利用切片式投影的基本原理同时改进其不足,本文采用旋转螺旋屏构建成像空间子系统,如图 3 所示。螺旋屏的曲面方程为

$$\begin{cases} y=x\tan\theta \\ z=\frac{H\theta}{180} \end{cases} \quad (6)$$

其中,  $x \in [0, L]$ ,  $\theta \in [0, 180^\circ]$ ,  $H$  和  $L$  分别为螺旋屏的高和半径,且有  $H < 2L$ 。螺旋屏绕中心轴旋转时,投影装置依次向其投射三维物体的  $N$  个螺旋切片,螺旋屏在不同旋转角度截获这些切片,显示出三维物体的轮廓。如果螺旋屏转速足够快

且切片数目足够大,人眼同样能在圆柱成像空间内感知到具有物理深度的立体图像。螺旋屏的各个位置是相互平行的,因此螺旋屏利用了切片式投影的原理,但将成像屏的往复运动转化成螺旋屏的旋转运动。螺旋屏材质采用高强度光敏树脂,光敏树脂对光线的透射率与反射率可以通过表面打磨进行调控,达到半透射半反射的效果,实现被动式发光,构建成像空间。由于螺旋屏通过金属阶梯轴由电机带动高速旋转,为提高运动稳定性,最终制定的旋转屏,采用了双螺旋面的结构,如图3所示,其中 $H=250\text{mm}$ , $L=250\text{mm}$ ,屏厚 $d=3\text{mm}$ 。为避免电机转轴等对成像空间入射光线的阻挡,采用对单侧螺旋面底部投影的方式显示立体图像。

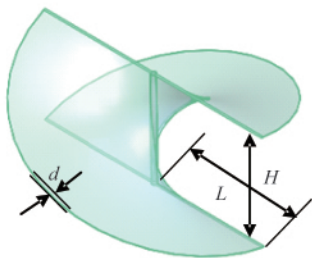


图3 螺旋屏  
Fig. 3 Helical screen

按第1.1节,将三维物体的体素模型,用可见体素的集合表示,记为 $O$ 。同样可得到上述螺旋屏的体素模型。给定螺旋屏的旋转角度 $\theta$ ,记其体素模型为 $H(\theta)$ ,则三维切片的体素模型为 $S(\theta)=O \cap H(\theta)$ 。其中, $\theta=2.727k$ , $k \in [1, 132]$ ,即用132个三维切片描述三维物体。

### 1.3.3 成像子空间性能分析

#### 1) 图像刷新率

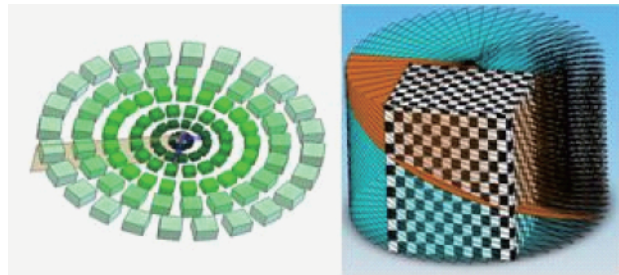
螺旋屏每转过一周,两个螺旋面分别扫描过半柱形的立体空间一次。如果电机转速为 $N \text{ r/s}$ ,则立体图像的刷新频率为 $2N \text{ Hz}$ 。若螺旋屏采用单螺旋面结构,则图像刷新频率与电机转速一致。可见,在电机转速一定时,螺旋屏的双螺旋面结构使得图像刷新率提高一倍。

#### 2) 体素量与体素分布

根据第1.3.2节,每个螺旋切片的分辨率为 $1024 \times 768$ ,每个三维图像由132个切片构成。因此,总体素量为 $1024 \times 768 \times 132$ 。

通过旋转平面屏构建成像空间,是早期动态体三维显示系统普遍采用的方法。以旋转LED阵列系统为例<sup>[5]</sup>,当LED阵列绕中心轴旋转时,LED发光单元的运用轨迹是多个半径不等的同心圆,并且越靠近转轴半径越小。因此在其成像空间内,体素分布不均匀,体素大小不等,如图4所示。

假设螺旋屏的中心轴沿竖直方向,当螺旋屏绕中心轴转过 $\alpha$ ,则螺旋面在竖直方向的位移为 $H\alpha/180$ ,同样螺旋面上每个像素在竖直方向的位移也为 $H\alpha/180$ 。另外,当 $\alpha$ 变化时,螺旋面上的像素数目保持不变。因此在成像空间内,体素分布均匀、大小均等,如图4所示。



(a) 旋转LED阵列  
(a) Voxel distribution of spinning LED array  
(b) 旋转螺旋屏  
(b) Voxel distribution of spinning helical screen

图4 体素分布

Fig. 4 Voxel distribution

### 3) 成像死区

对于动态体三维显示系统,体素重叠、部件遮挡等都可能形成成像空间的死区<sup>[1]</sup>。对于旋转平面屏构建的成像空间,由于体素分布不均匀,在靠近转轴部分会因体素重叠造成死区。在螺旋屏扫描形成的成像空间内,体素分布均匀且大小均等,因此部件遮挡是死区的主要成因。部件遮挡包括两个螺旋面的相互遮挡、金属阶梯轴对螺旋面的遮挡。实验结果显示,由于本文制成的螺旋屏对光线具有半透射半反射的能力,因此螺旋面相互遮挡引起的死区对显示效果影响不大。如果提高成像空间入射光线的强度,则可进一步降低螺旋面相互遮挡产生的死区。金属阶梯轴引起的死区则不可避免。

### 1.4 人机交互子系统

体三维显示系统的人机交互子系统将操作者的动态手势转化为控制指令,实现对显示对象的切换、平移、缩放、旋转等基本操作。人机交互子系统首先从相邻帧差图像中检测特定的静态手势获得初始跟踪区域,并针对此区域建立手部肤色的分布模型。然后通过光流法更新跟踪区域内KLT (Kanade-Lucas-Tomasi)<sup>[6]</sup>特征点的位置,并按特征点群算法结合手部肤色的分布模型对特征点的位置进行修正<sup>[7]</sup>。对于跟踪获得的动态手势轨迹,经过坐标变换并按8方向Freeman链码进行量化编码,获得手势观测序列。最后利用经过训练建立手势隐马尔科夫模型库,实现动态手势的识别。

#### 1.4.1 手势检测与跟踪

##### (1) 手势检测

按以下步骤可以检测完全张开的手型。

① 对于当前帧与前一帧帧差的二值图像,提取其外部轮廓,图5显示了轮廓内实际的手部图像。

② 计算外部轮廓的凸包,凸包的顶点即为指尖点(已去除位于图像边界的凸包顶点)。

③ 计算外部轮廓的凸起缺陷及其深度,即各个缺陷点到凸包的距离。记最大深度值为 $D_{\max}$ ,取深度超过 $0.6D_{\max}$ 的缺陷点为指间缝隙点,如图5中的红色点。

④ 若指尖数为5且指间缝隙数为4,静态手势检测成功。以凸起缺陷点集(如图5,包括白色点)所确定的矩形区域

为初始跟踪区域  $R_{ini}$ , 以所有指尖点和凸起缺陷点确定的矩形区域为手部边界区域  $R_{bound}$  (如图 5 中方框所示), 退出手势检测; 否则读取新的输入帧并进入步骤①。

在初始跟踪区域  $R_{ini}$  内, 按直方图反向投影法建立手部肤色概率分布图, 具体步骤如下:

① 将当前帧图像从 RGB 空间转换到 HSV 空间。

② 对其中的  $H$  分量作直方图, 并将  $H$  分量的值量化到  $[0, 255]$ 。

③ 将图像中每个像素的值用其颜色出现的概率做替换, 即可得到肤色概率分布图。

### (2) 手势跟踪

在初始化阶段, 在手势检测获得的初始跟踪区域  $R_{ini}$  中, 找到  $n \times k$  个 KLT 特征点<sup>[16]</sup>。根据第 1.1 节建立的肤色分布模型, 提取前  $n$  个特征点, 构成特征点群  $S$ , 并将  $S$  的几何中心设置为手势轨迹序列起始点。

在手势跟踪阶段, 首先运用金字塔光流算法更新特征点的位置, 然后利用特征点群算法结合肤色分布模型调整部分特征点的位置。特征点群算法适于跟踪快速运动且形状多变的铰链式物体, 它可保证物体表面的各个特征点既不离其中心, 又不发生相互重叠<sup>[17]</sup>。肤色模型则可保证所有特征点都在肤色区域内。手势跟踪的具体步骤如下:

① 利用金字塔光流算法更新  $S$  中各个特征点的位置。

② 计算  $S$  的几何中心, 并将手部边界区域  $R_{bound}$  的中心更新到  $S$  的几何中心。

③ 对其中的特征点, 如果不能同时满足下列约束条件则转入④, 否则转入⑤。

(i) 与其他特征点之间的最小距离大于阈值  $D_{min}$ ;

(ii) 位于  $R_{bound}$  之内;

(iii) 具有较高的肤色概率。

④ 在  $R_{bound}$  内重随机重定位特征点, 使其满足上述约束条件并且位于具有高肤色概率的位置。

⑤ 重新计算  $S$  的几何中心, 并将其添加到手势轨迹序列末尾。

### (3) 动态手势特征提取

对于具有时空特征的动态手势序列, 可以提取手部形心坐标、面积、所在椭圆区域的偏心率、速度、加速度等为每一帧输入图像建立高维特征向量, 为提高手势识别的实时性能, 在此仅对手部形心的轨迹曲线进行特征提取。对于手势轨迹线上相邻的两个采样点(形心)  $P_a, P_b$ , 计算其连线  $\overline{P_a P_b}$  与水平线的夹角  $\theta$ , 根据  $\theta$ , 可按 8 方向 Freeman 链码对  $\overline{P_a P_b}$  进行编码, 码值为  $V = \theta / 45 (\theta \in [0, 360^\circ], 0 \leq V \leq 7)$ , 如图 6 所示。此外, 为避免采样点过于密集、编码过长, 只有在当前采样点与前一有效采样点的间距大于阈值  $D_T$  时, 当前采样点才为被设置为有效采样点。

手势检测、跟踪及特征提取均以开源计算机视觉库 OpenCV 为基础实现, 其中凸包计算采用 Sklansky 算法, 点群  $S$  包含的特征点数  $n$  为 25,  $k$  取 2,  $D_{min}$  为 3 像素,  $D_T$  为 5 像

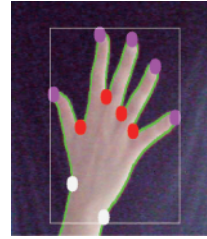


图 5 静态手势特征

Fig. 5 Features of the static hand gesture

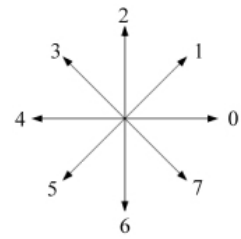


图 6 8 方向 Freeman 链码

Fig. 6 Eight-directional Freeman chain code

素。此外,  $R_{ini}$  的具体计算方式是: 获得所有凸起缺陷点构成的多边形的内接圆, 此圆的内接正方形所在区域即为  $R_{ini}$ 。

### 1.4.2 基于 HMM 的手势识别

HMM 作为典型的统计学模式识别方法, 在自动语音识别等系统中获得了成功的应用, 基于 HMM 的人脸识别、手势识别等也引起了人们的广泛关注<sup>[18-19]</sup>。一个恰当构建的 HMM 对类内偏移具有较强的鲁棒性, 而对类间模式的差异具有较强的敏感性。HMM 是一个双重随机过程, 其中之一是马尔科夫链, 这是基本的随机过程, 描述状态的转移; 另一个是观测值(通常称为观测符号)在各状态中的分布概率, 它描述状态和观测值之间的统计对应关系。用两个随机过程更能有效地描述客观世界中事件发生的复杂关系。按照观测值概率取值与状态转移描述方式的不同, 可以分为连续 HMM (Continuous HMM, CHMM)、半连续 HMM (Semi Continuous HMM, SCHMM) 和离散 HMM (Discrete HMM, DHMM)。对于 CHMM, 其观测概率矩阵的每一项通常为多高斯模型, 而 DHMM 的观测概率矩阵为实数阵。相对来说, 半连续、连续 HMM 都能更细致地刻画事件发生的关系, 但也需要更多的训练样本数据。本文采用 DHMM, 记为  $\lambda = (\pi, A, B)$ 。其中, 初始概率向量  $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ ,  $N$  为状态数;  $A = [a_{ij}]_{N \times N}$  为状态概率转移矩阵;  $B = [b_j(v_k)]_{N \times M}$  为观测值概率矩阵,  $M$  为观测符号数。

#### (1) 手势 HMM 的基本问题及其相关算法

HMM 的基本问题包括估计、评估、解码 3 个方面。对于基于 HMM 的手势识别, 需解决如何建立手势 HMM, 即估计问题; 以及如何对手势观测序列进行识别, 即评估问题。

##### ① 估计问题

对于模型  $\lambda = (\pi, A, B)$ , 将观测样本作为训练输入, 不断调整模型参数, 使得观测样本的似然函数取值最大, 这是参数估计 (estimation) 问题, 也就是模型训练问题。

通常, 一个动态手势数据完成特征提取后, 手势被表示为  $E$ , 如果  $E$  由高维列向量构成, 对  $E$  中各列向量进行向量量化 (vector quantization) 可得到该手势的观测序列  $O = \{b_1, b_2, \dots, b_T\}$  ( $T$  为序列长度)。如第 1.4.1 节, 本文按 8 方向 Freeman 链码对二维手势轨迹进行编码, 码值为 0~7, 故  $b_i \in [0, 7]$ 。对于一个特定的手势, 多个不同的观测序列构成其训练样本集  $\{O_1, O_2, \dots, O_s\}$ , 其中  $s$  为样本个数。本文利用 Baum-Welch 算法对手势 HMM 进行训练。对  $l$  个不同动态手势的训练样本

集分别进行训练,即可建立手势的 HMM 库 $\{\lambda_1, \lambda_2, \dots, \lambda_l\}$ 。

② 评估问题

给定观测序列  $O=\{O_1, O_2, \dots, O_T\}$  和模型  $\lambda=(\pi, A, B)$ , 计算  $P(O|\lambda)$ 。这是评估 (evaluation) 问题, 也就是给定模型和观测序列, 如何计算从模型生成观测序列的概率的问题。也可以把它看作是评估一个模型和给定观测输出序列的匹配程度。对于基于 HMM 的手势识别, 其任务就是对于给定的观测值序列  $O$  和所有动态手势 HMM, 计算由各模型产生  $O$  的概率, 并取其中最大概率对应的模型为识别结果。本文采用前向-后向算法计算  $P(O|\lambda)$ , 而且为避免数值太过密集、输出概率过小, 对计算出的概率值按下式取自然对数后进行比较:

$$P' = \ln(P(O|\lambda)) \quad (7)$$

(2) 手势识别流程

手势识别流程如下:

- ① 对输入的动态手势  $I$  包含的轨迹按 8 方向 Freeman 链码进行量化处理, 得到其观测值序列  $O=\{b_1, b_2, \dots, b_m\}$ ;
- ② 计算动态手势 HMM 库中各模型  $(\lambda_i, i \in [1, l])$  产生  $O$  的概率  $P'_i, i \in [1, l]$ ;
- ③ 取  $\{P'_1, P'_2, \dots, P'_l\}$  中最大值  $P'_j = \max\{P'_1, P'_2, \dots, P'_l\}$ , 如果  $P'_j > T$  ( $T$  为一确定阈值),  $j$  即为识别出的手势编号。

2 实验

图 7 为体三维显示系统原理图, 图 8 为最终制成的样机。样机采用明基 PB6245 投影仪的系统光路, 并去除了色轮。DMD 芯片的尺寸、分辨率和开关偏角分别为 0.7"、XGA (1024x768)、12°。DMD 控制算法在 DMD 引擎上实现, DMD 引擎具有 1GB 的 RAM, 用于缓存切片数据。选用额定转速为 3000r/s、额定功率为 0.75kW 的富士 GYS 电机, 其配套的伺服放大器型号为 RYC751D3-VVT2。另外采用 PIO-DA16 作为数据采集卡, 用于采集伺服电机转角为 0 和 180° 的信号。当投影控制板通过采集卡接收到伺服电机的信号时, 按特定间隔依次向旋转螺旋屏投射切片序列数据, 这样, 伺服电机每转过一周, 切片序列投影两次。

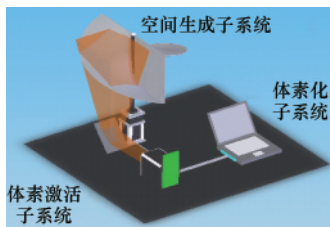


图 7 体三维显示系统原理  
Fig. 7 Schematic diagram of the volumetric 3D display



图 8 体三维显示系统样机  
Fig. 8 Prototype of the volumetric 3D display

表 1 比较了本文样机与 Perspecta 系统的技术参数。可知本文样机在成像空间尺寸、体素数目、刷新频率和最大切片数等方面低于 Perspecta 系统, 在切片分辨率方面超过了

表 1 本文样机与 Perspecta 系统技术参数比较

Table 1 Comparison of technical parameters comparison between our prototype and Perspecta

	本文样机	Perspecta
屏幕形状	双螺旋面	圆平面
成像空间	<10" (半圆柱形)	10" (圆柱形)
体素量	103809024	116785152
颜色位数/b	1	24
刷新率/Hz	16	24
最大切片数	132	198
切片分辨率	1024x768	768x768

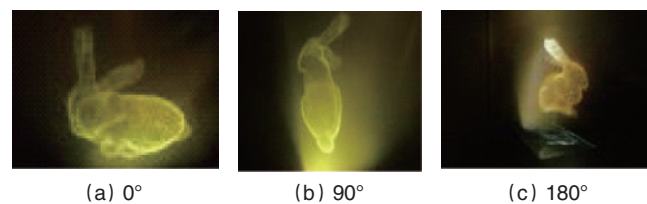
Perspecta 系统。另外 Perspecta 系统采用 3 块 DMD 芯片分别调制彩色图像的 R、G、B 分量, 实现了彩色图像的显示。本文系统目前仅支持二值图像的显示。

第一组实验比较 BCC 采样策略与传统笛卡儿采样方法。设定相等的采样点间隔, 当笛卡儿栅格的分辨率为  $R^3$  时, BCC 栅格的分辨率为  $2R^3$ 。在此, 将 BCC 栅格的采样点数目设定为笛卡儿采样点数量的 70%, 分别按两种采样方法进行模型体素化, 表 2 表明, 在体素量、时间耗费、内存耗费等方面, BCC 采样策略较笛卡儿采样方法均有大幅降低, 其中体素数量减少 40%~50%。其中的 Rabbit 模型采用 Stanford Bunny, 不同角度观看的效果如图 9 所示。

表 2 体素化性能比较 (BCC 策略与笛卡儿方式)

Table 2 Comparison of voxelization performance comparison between BCC strategy and Cartesian scheme

模型		Rabbit	Cow	Elephant
三角面数		65533	5804	39290
体素量	BCC	9579	5091	6282
	笛卡儿	19326	9135	11722
时间/ms	BCC	219	36	94
	笛卡儿	359	47	203
内存/MB	BCC	15.76	11.652	13.84
	笛卡儿	19.946	15.52	17.684



(a) 0° (b) 90° (c) 180°

图 9 体三维兔模型显示效果

Fig. 9 Volumetric display of a rabbit model

第二组实验验证基于手势控制的人机交互子系统。以水平面内围绕柱形显示空间的平台为背景, 当观众或操作者对摄像头 (位于螺旋屏外壳顶部正上方 3.6m 处) 做出动态手势时, 根据第 1.4.1 节获得手势观测序列, 再利用训练好的 HMM 计算手势观测序列的输出概率, 判断出具有最大可能的手

势。如果识别结果正确,向投影控制板发送控制指令,调整投影切片。如果识别结果错误,投影切片不作调整,摄像头重新进行捕获。为提高系统的实时性,本文通过判断手势中心的速度和加速度确定手势起止点,提取手势输入序列,因此仅需处理相邻3帧图像。由于交互式体三维显示系统可在360°范围内观看与交互,因此手势轨迹在进行编码前应根据观众或操作者的位置进行坐标变换,具体过程如下:

(1) 计算手势轨迹最小包围扇形区域  $MON$ , 如图 10 所示。

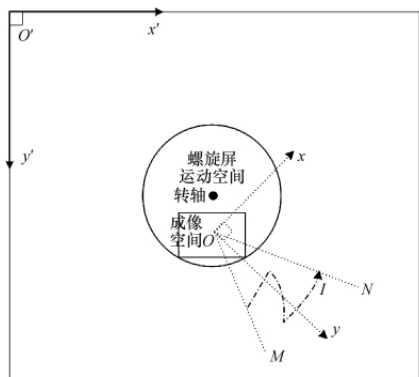


图 10 手势轨迹的坐标变换

Fig. 10 Coordinate transformation of a gesture trajectory

(2) 以  $O$  为中心建立直角坐标系  $xOy$ , 其中  $Oy$  平分  $\angle MON$ ,  $Oy$  与观众或操作者的视角方向一致。将手势轨迹数据  $I$  从图像坐标系  $x'O'y'$  转换到  $xOy$  坐标系。

(3) 在  $xOy$  坐标系下对手势轨迹数据进行 8 方向 Freeman 链码编码,得到动态手势的观测序列。

共设计了 8 种动态手势, 它们表示的意义如表 3 所示。对于第 3、4 号手势,显示模型每次移动幅度为显示空间高度的  $1/20$ , 约 1.25cm。对于第 5、6 号手势,  $z$  轴为投影切片图像的中心轴,沿垂直方向。对于第 7、8 号手势,放大、缩小指将显示模型调整到下一尺寸,每次调整幅度为 0.05 倍。进行手势识别的主机系统具有 Intel Pentium 4 2.8GHz CPU、Kingston 1GB DDR RAM。表 4 给出了动态手势识别实验结果,其中  $M$  为状态数。从表中可以看出,对不同手势的平均识别正确率超过 93%,对不同手势的平均动响应时间小于 40ms,具有较

表 3 动态手势定义

Table 3 Definition of dynamic hand gestures

编号	轨迹样式	意义	编号	轨迹样式	意义
1	$\leftarrow$	前一模型	2	$\rightarrow$	后一模型
3	$\uparrow$	沿 $z$ 轴上移	4	$\downarrow$	沿 $z$ 轴下移
5	$\curvearrowright$	绕 $z$ 轴正向旋转	6	$\curvearrowleft$	绕 $z$ 轴逆向旋转
7	$\approx$	放大	8	$\approx$	缩小

表 4 手势识别结果

Table 4 Results of dynamic hand gesture recognition

编号	训练样本数	测试样本数	准确率/%		响应时间/ms	
			$M=4$	$M=5$	$M=4$	$M=5$
1	60	50	92	92	31.1	31.2
2	60	50	94	96	30.2	30.9
3	60	50	92	96	31.7	33.7
4	60	50	98	94	29.6	31.0
5	60	50	94	98	30.7	31.8
6	60	50	90	90	29.1	32.0
7	60	50	92	94	38.7	39.4
8	60	50	94	100	36.8	37.3
合计	480	400	93.3	95.0	32.2	34.1

好的实时性,可以满足交互式体三维显示的要求,实现围绕体三维显示系统 360°范围的人机交互。图 11 是在手势控制下,某型飞机模型按不同旋转角度和尺寸显示的效果,所有图片均在固定视角拍摄得到。其中,第 1 行从左至右为模型在标准尺寸下,根据第 5、6 号手势逐次旋转 20°的显示效果;第 2 行从左至右为根据第 7、8 号手势,模型按 0.3 倍、0.6 倍和标准尺寸显示的效果。

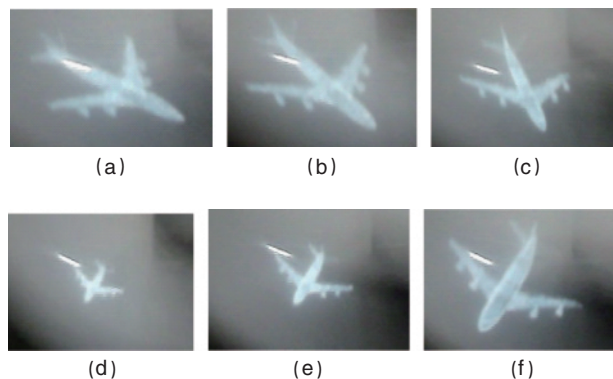


图 11 动态手势控制下的飞机模型

Fig. 11 A plane model controlled by dynamic hand gestures

### 3 结论

本文在分析体三维显示方面前人工作的基础上,针对动态体三维显示的多项关键技术,包括三维数据体素化、体素激活、体空间生成、人机交互等,设计并实现多个子系统。通过采用双螺旋旋转屏结构、底部投影、半柱形空间被动式发光等方法,构建了体素分布均匀的成像空间,降低了视觉死区,提高了系统刷新频率;通过设计 DMD 控制算法,实现了二值体素切片的快速激活;通过 BCC 采样策略,降低了体三维显示数据量;通过引入手势控制,实现了围绕体三维显示系统样机全方位的人机交互。针对样机设计的各组实验,进一步验证了本文设计的有效性。下一步的工作重点是彩色图像的体三维显示和基于双手连续动态手势控制的交互式体三维显示。

## 参考文献 (References)

- [1] Bundell B G, Schwarz J A. Volumetric three-dimensional display systems [M]. New York: John Wiley&Sons, 2000.
- [2] Knut Langhans. New portable Felix 3D display [J]. *Proceedings of SPIE*, 1998, 3296: 204-216.
- [3] Favalora G E, Dorval R K, Hall D M, et al. Volumetric three dimensional display system with rasterization hardware[J]. *Proceedings of SPIE*, 2001, 4297: 227-235.
- [4] Sullivan A. A solid-state multi-planar volumetric display [R]/SID 2003 Digest, Baltimore: SID, 2003: 354-356.
- [5] 林远芳, 刘向东, 刘旭, 等. 基于二维旋转屏的体三维显示系统像素属性分析[J]. *光子学报*, 2004, 33(4): 476-487.  
Lin Yuanfang, Liu Xiangdong, Liu Xu, et al. *Acta Photonica Sinica*, 2004, 33(4): 476-487.
- [6] 李莉, 李玉峰, 沈春林, 等. 基于数字微镜的旋转体三维显示装置研究 [J]. *仪器仪表学报*, 2008, 29(1): 67-72.  
Li Li, Li Yufeng, Shen Chunlin, et al. *Chinese Journal of Scientific Instrument*, 2008, 29(1): 67-72.
- [7] Downing E, Hesselink L, Ralston J, et al. Three-color, solid-state, three-dimensional display[J]. *Science*, 1996, 273(5279): 1185-1189.
- [8] Kim I I, Korevaar E, Hakakha H. Three-dimensional volumetric display in rubidium vapor[J]. *Proceedings of SPIE*, 1996, 2650: 274-284.
- [9] Refai H H. Static volumetric three-dimensional display [J]. *Journal of Display Technology*, 2009, 5(10): 391-397.
- [10] Zhang C Y, Yurkewicz K, Khan A, et al. 3D display of treatment planning and anatomy data: Initial observation using a promising technical advance [C]/IFMBE Proceedings, World Congress on Medical Physics and Biomedical Engineering 2006, Seoul, Korea, August 27-September 1, 2006: 1729-1732.
- [11] Grossman T, Wigdor D, Balakrishnan R. Multi-finger gestural interaction with 3D volumetric displays[C]/UIST'04: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology. Santa Fe, NM: UIST, 2004: 61-70.
- [12] Dudgeon D E, Mersereau R M. Multidimensional digital signal processing[M]. Englewood Cliffs, New Jersey: Prentice-Hall, 1984.
- [13] Theußl T, Möller T, Gröller E. Optimal regular volume sampling[C]/IEEE Visualization 2001: Proceedings of IEEE Visualization. San Diego, California: IEEE, 2001: 91-98.
- [14] Fujimoto A, Takayu T, Iwata K. ARTS: Accelerated ray-tracing system [J]. *IEEE Computer Graphics and Applications*, 1986, 6(4): 16-26.
- [15] Meier R E. DMD pixel mechanics simulation [J]. *TI Technical Journals*, 1998, 15(3): 64-74.
- [16] Shi J, Tomasi C. Good features to track[C]/CVPR 1994: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 1994: 593-600.
- [17] Kolsch M, Turk M. Fast 2D hand tracking with flocks of features and multi-cue integration [C]/CVPR'04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop. Washington: IEEE, 2004.
- [18] Just A, Marcel S. A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition[J]. *Computer, Vision and Image Understanding*, 2009, 113(4): 532-543.
- [19] Martinez A. Face image retrieval using HMMs [C]/CBAIVL 1999: Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries. Fort Collins, Colorado: IEEE, 1999: 35-39.

(责任编辑 朱宇)

## ·学术动态·



## “第十三届国际体视学大会”征文

由国际体视学会主办,中国体视学学会承办的“第十三届国际体视学大会”将于2011年10月19-23日在北京市召开。会议主题:交流体视学理论方法和体视学应用领域的最新研究成果。

征文内容:体视学方法:Stereology, geometrical and topological tools, Main mathematical and physical transforms, Mathematical morphology, stochastic geometry, Spatial statistics, space-time dynamics pattern analysis, Texture evaluation; Fractals, image enhancement, 3D reconstruction, Color images; 体视学应用: Micro- and nano-structures, Biology and medicine, Biotechnology, Agriculture, Materials tomography, X-ray scanner, PET-scan, High resolution imaging in remote sensing, Vision, industrial control, real time aspects, electronic architectures, Micro-electronics, Secure information systems。

全文截止时间:2011年7月15日。

联系方式:北京市海淀区清华大学工物系(100084) 刘卿楼;电话:010-62776336; 电子信箱:tscss@mail.tsinghua.edu.cn; 会议网站:www.tscss.org。

## 《科技导报》“走向职场”栏目征稿

“走向职场”栏目期望通过业内专家撰文,或有关人士讲述自己求职、就业的经历和成功事例,帮助博士生和博士后学会如何应聘、如何创业和发展,给即将走上职场的读者朋友启迪和帮助。欢迎投稿,择优录用,每篇文章约2200字。栏目责任编辑:陈广仁,投稿信箱:chenguangren@cast.org.cn。

## 《科技导报》“科学共同体要闻”栏目征稿

“科学共同体要闻”栏目报道中国科协、地方科协和中国科协所属全国学会,以及其他学术团体的重大科技事件、学术活动成果等新闻。每条科技新闻字数为300字左右,要求标注新闻来源、日期。欢迎各级科协 and 全国学会推荐或自荐新闻。栏目责任编辑:陈广仁,投稿信箱:chenguangren@cast.org.cn。