高技术通讯 2013 年 第23 卷 第8 期:781 ~788

doi:10.3772/j.issn.1002-0470.2013.08.002

基于形状检索的场景图像三维建模①

樊亚春② 谭小慧 周明全 陆兆老

(北京师范大学信息科学与技术学院 北京 100875)

摘 要 提出一种利用二维场景图像快速构建三维虚拟场景的三维建模方法,该方法有以下特点:利用改进的图割算法对场景图像进行分割,提取对象轮廓形状;基于尺度不变特征点进行轮廓形状局部特征描述,用量化后的视觉词汇向量描述三维模型轮廓及对象轮廓;计算不同形状轮廓之间的距离,从三维数据库中检索出与对象相似的三维模型;计算对象模型和背景场景模型的坐标变换,合成虚拟场景。实验证明,该方法检索效果好,用户操作简单,能实现复杂虚拟场景的快速建模。

关键词 三维建模,图割,形状检索,视觉词汇,多尺度特征描述

0 引 言

目前虚拟现实场景的三维建模通常是利用商业的建模系统完成,这些系统操作复杂,操作要求很高,构建场景需要专业人员花费大量时间。为了节省三维建模的开发成本,可利用网络上已有的大量、三维模型数据,但是如何从大量三维数据中寻找需要的模型数据用于虚拟现实场景的快速构建,已经成为当前迫切需要解决的问题。随着数码相机的普及以及二维图像处理技术的进步,目前的场景表达多来自于二维图像媒体。用户在要求构建三维场景时,常常会利用已有的二维图像场景作为参考,实现与具体场景相匹配的三维虚拟场景的构建。基于以上分析,本文提出了一种利用二维场景图像快速构建三维虚拟场景的三维建模方法,此方法根据已有大型三维模型数据库中的数据对象,快速检索出与图像场景中对象匹配的三维模型,通过场景合成的方式建立新的虚拟场景。提出此方法并建立系统平台应用的最初目的是为了让更多普通用户能够实现虚拟场景的构建,降低构建虚拟场景所耗费的人力及物力资源。不同于以往的建模工具,此方法仅仅需要用户进行选择、搜索、合成等操作就能够实现一个复杂虚拟场景的生成,再结合用户的进一步个性化编辑就能够完成一个满足用户需求的虚拟场景的产品。

1 相关工作

本文提出的新的三维建模方法,能够确保没有任何专业基础的人在短时间内快速建立需要的三维场景模型。不同于3DSMax[1],SketcuUp[2]等三维建模工具,该方法利用三维模型数据库以及二维图像快速建立三维场景模型。Funkhouser[3]在文献[3]中提出了基于样例的三维建模方法,用此方法,用户可通过检索三维模型数据库,得到模型的不同部分,利用模型不同部分的不同组合方式建立新的不同的三维几何模型,它的主要优点是建模过程便于用户操作,用户不需要处理细节性的建模操作。快速三维建模的另外一类方法是在已有二维图像基础上增加三维信息,建立与之相对应的三维模型。基于图像的三维建模方法多是通过多幅图像的精确摄像机标定来实现[4]。Xu[5]提出,可利用用户绘制的模型图像建立三维几何模型,该方法对图像中物体的不同部分进行分割后,利用一组与该物体相似的三维模型的不同区域的变形建立新的三维模型。

利用已有的检索技术合成新的可视场景是由图像技术的发展而来,并在近几年取得了很好的成果。Eitz[6]利用简单的手绘草图快速建立复杂场景图像,此方法利用交互方法分割对象,重新组合检索到

① 国家863 计划(2008AA01Z301)和国家自然科学基金(61001168,61170203)资助项目。

② 女,1978 年生;博士,讲师;研究方向:虚拟现实,计算机图形图像技术;联系人,E-mail:fanyachun@hotmail.com。

(收稿日期:2012-12-17)

的多个对象,最终生成用户需要的二维场景图像。Chen[7]在利用用户草图的同时,给每个草图赋予了一个文本标注,并从 Internet 上直接检索图片,增加了图像检索的信息,提高了检索的匹配效果。利用草图轮廓检索,草图绘制结果直接影响了检索的效果,这种检索方式不稳定。Chen[8]在之前工作的基础上,提出对已有图像分割后提取对象轮廓并标注后进行检索,从而生成新的组合后图像。同样是快速场景图像获取,Lan[9]在利用支持向量机对图像集学习训练中,不仅考虑图像中物体对象,还考虑了对象之间的结构信息,用户只需要提供简单的结构对象信息即可检索到需要的二维场景图像。Engel[10]利用草图区域的不同颜色表示及语义标识描述场景图像并进行检索,从而获取相似的二维场景图像。这两种方法没有考虑场景图像的重构,仅检索出已有二维场景信息。本文提出的三维建模方法,在二维图像分割的基础上,搜索与对象轮廓相似模型,并最终组合为与输入图像相似的三维场景模型。

图像分割及轮廓提取作为图像处理的基本问题,已经研究了很多年。其主要方法有均值偏移方法[11]、马尔科夫场方法[12]、区域增长方法[13]、分水岭方法[14]、图割方法[15]。图割方法是利用 Gibbs 能量最小化解决图像分割问题,这种表示方法的分割效果已经在交互式分割领域得到了充分证明[16,17]。已有图割方法对于对象和背景颜色接近的彩色图像分割效果不佳,本文提出利用 n 连接分量和 t 连接分量分别表示不同边的能量信息,进而计算最小分割代价,实现更加适合轮廓表示的对象分割结果。

基于形状特征的三维模型检索分为全局特征描述和局部特征描述[18]。全局特征描述算法通常利用边缘形状的统计信息或频率变换后系数作为特征描述子。Zhang[19]将模型的一些几何参数如面积、体积等作为特征向量。Osada[20]根据不同几何形体表面顶点间的相互关系呈现出不同的分布特征,将一个任意的三维模型中复杂的特征提取转换成相对简单的形状概率分布问题。全局特征提取简单易操作,但是对于表面复杂模型,其检索的准确率并不理想。局部特征提取方法充分考虑了复杂曲面上点及其周边形状信息,检索更加准确。Chua[21]计算三维点周边曲面累积信息作为区域形状特征。Johnson[22]提出了利用旋转图像统计三维点周边局部曲面的二维直方图信息。由于它们提到的描述方法复杂,因此长期以来局部特征表示方法并没有很大的发展。基于视觉词汇的方法是一种二维图像检索的局部特征描述方法[23],这种方法被证明是一种非常有效的形状描述方法。为了构造一种通用特征描述二维轮廓形状及三维轮廓形状,本文提出利用视觉词汇生成三维模型特征描述符,从而建立二维图像与三维模型相似度比较的桥梁。本文在已有相关工作的基础上,针对场景图像的三维建模提出了新方法:首先提出改进的图割方法对场景图像的物体进行分割,得到物体轮廓信息,其次提出基于尺度不变特征变换(scale-invariant feature transform, SIFT)的视觉词汇特征描述方法,构造二维图像和三维模型的特征描述符,最后利用检索出的相似模型,合成与图像场景相似的三维模型场景。

2 系统框架

按本文方法,建模系统输入为一幅场景图像,该图像中包含有对象及背景场景。系统输出是一个包含有三维物体对象的三维场景模型。二维图像和三维模型的主要区别是深度信息的缺失。为了解决这个问题,提出利用大规模三维模型数据库及三维模型检索技术实现二维到三维的转换。在转换过程中,有对象分割和三维模型检索两个主要问题需要解决。

对象分割。采用了改进的图割方法来解决此问题,用户只需要进行简单操作便可以分割出物体轮廓。系统提供画矩形及曲线功能帮助用户确定要分割的对象及背景信息,使用鼠标左键画出红色线表示要分割的对象信息,使用鼠标右键画出蓝色线表示背景信息,用户还可以选择使用矩形指定对象大概范围帮助分割结果更加精确。系统提供场景中多个物体的分割及检索功能。

三维模型检索。采用一种视觉词汇特征方法统一描述二维轮廓和三维模型特征,并最终检索出与二维对象相似的三维模型。在视觉词汇生成阶段,选择了 500 个三维模型作为样本数据,提取 200 个不同的局部特征描述作为视觉词汇,利用不同形状的不同视觉词汇的距离计算,得到与图像轮廓相似的三维模型投影轮廓,继而得到相似的三维模型排序序列。

3 对象分割

图像中的对象分割可以看做是一个二值化标定

问题[24]。对象 O 和背景 B 分别赋予 1 和 0 值。图割 $GC = \{S,T\}$ 将无向图 G 分为 S 和 T 两个集合。源端点 s 和目标端点 t 分别位于两个集合中。图中的边分为两种,一种是相邻节点之间的边叫做 n 连接,一种连接节点与端点之间的边叫 t 连接。每条边都拥有一个权值 w,分割的代价定义为所有边的权值之和:

$$C(S,T) = \sum_{p\in S,q\in T} w(p,q) \tag{1}$$

其中 (p,q) 代表节点 p 和 q 的边,$w(p,q)$ 为边 (p,q) 的代价,$C(S,T)$ 表示分割的代价。

分割的代价越小表示能够获得更好的分割效果,最好的分割就是代价最小的分割。Gibbs 能量函数能够保证分割的代价最小:

$$E(X) = \sum_{p\in\delta} E_1(x_p,s\mid t) + \lambda \sum_{(p,q)\in\varepsilon} E_2(x_p,x_q) \tag{2}$$

此能量函数包括两部分内容,即 t 连接分量的能量总值和 n 连接分量的能量总值。其中 $E_1(x_p,s\mid t)$ 表示 t 连接分量能量值,$E_2(x_p,x_q)$ 表示 n 连接分量能量值,$\lambda \geqslant 0$ 是用来平衡这两个分量的参数。对于节点 p 或 q,x_p 和 x_q 为该节点的像素值。$E(X)$ 为图像中的能量值。X 表示图像中的所有像素集合。δ 是图中节点集合,ε 为图中的边集合。

t 连接分量。分割中用户的交互确定最初的 S 和 T 集合,为了计算能量值 E_1,将用户画线的像素进行聚类,分别对对象和背景交互内容使用 k-means 方法进行聚类,对象的聚类中心用 $o_n \in S$ 表示,背景的聚类中心用 $b_m \in T$ 表示。节点到两个端点的距离分别定义为 $d_p^o = \min_n \| x_p - o_n \|$ 和 $d_p^b = \min_m \| x_p - b_m \|$。距离越短意味着节点和端点越相似。因此节点 $p \notin \{O\} \cup \{B\}$ 到端点 s 的距离定义为 $\upsilon = \dfrac{d_p^o}{d_p^o + d_p^b}$。最后,$E_1(x_p,s\mid t)$ 的描述如下:

$$\begin{cases} & p\in\{O\} & p\in\{B\} & p\notin\{O\}\cup\{B\} \\ E_1(x_p=1) & 0 & \infty & \upsilon \\ E_1(x_p=0) & \infty & 0 & 1-\upsilon \end{cases} \tag{3}$$

n 连接分量。$E_2(x_p,x_q)$ 是利用像素及其周边区域关系来确定其属于对象或者背景。对于相邻节点 p,q,通常当 p 和 q 越相似时 $E_2(x_p,x_q)$ 值越大,相反,当这两个节点完全不同时,$E_2(x_p,x_q)$ 的值趋向于 0。图像中像素分布可以用高斯混合模型表示,考虑到两个像素之间的颜色距离,p 和 q 之间的 n 连接分量定义为

$$E_2(x_p,x_q) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\| x_p - x_q \|^2}{2}\right) \tag{4}$$

最后利用最优化方法中的网络最大流法求解出能量函数,得到分割后的结果。

4 对象的三维模型检索

4.1 三维模型的轮廓形状表示

利用图割方法进行场景图像的对象分割后,物体的轮廓信息随即得到。为了获取与对象相匹配的三维模型数据,本文首先利用轮廓线条描述三维模型形状特征[25],随后提取 SIFT 特征描述建立视觉词汇库用于表示三维模型特征。

对于三维模型的轮廓线描绘,除了外围轮廓线外,还要获得更多的内部特征,包括谷线、脊线等。径向曲率是判断三维模型轮廓线的重要参数,在径向曲率的基础上,实现了与视点相关的特征线提取方法,这种特征线能够有效地表示三维模型的形状特征。图 1 为利用该方法提取的三维模型轮廓线表示效果。

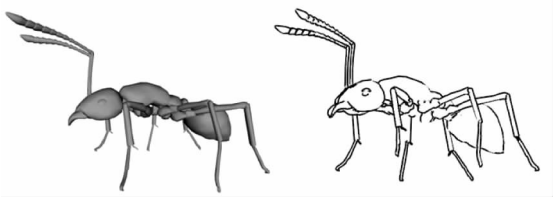


图 1 三维模型的轮廓线表示

提取 SIFT 特征描述之前的重要一步是完成三维轮廓模型的二维表示,三维模型的二维投影方法最常用的是三视图投影[26],这种方法本质上是一种基于正八面体顶点的一种投影方式,实现简单,但是仅仅依赖于三幅投影无法获取三维模型内部更详细的视觉信息。三维空间中正多面体仅有 5 种,即正四面体、正八面体、正六面体、正十二面体和正二十面体。正十二面体相比于其它 4 种正多面体在视觉观察领域具有独特的优势。首先正十二面体是仅次于正二十面体的一种接近于球体的正多面体,非常适合视觉观察,能够更加全面地覆盖物体的各角度平面细节信息;其次,正十二面体的每个面由一个五边形组成,是 5 种正多面体中平面最接近于圆形的一种,进一步增强了物体的平面视觉覆盖效果。

本文利用正十二面体的 20 个顶点信息作为摄

像机的位置,对三维模型进行二维图像投影。图 2 为对三维模型的投影示例。

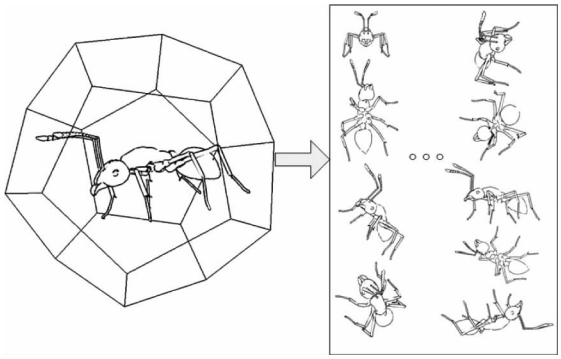


图 2 三维模型的二维轮廓投影

4.2 局部特征描述

对于轮廓图像的特征计算,SIFT 算法[27]具有非常大的优势。首先,它提取的特征是局部特征,对旋转、尺度缩放、亮度变化等都具有很好的不变性,对于视角变化、仿射变换、噪声也保持很好的稳定性;其次,少数的轮廓信息就可以在特征点基础上获得大量的特征向量信息;再次,特征描述向量信息丰富,适用于海量特征数据库的快速准确匹配。本文以 SIFT 算法为基础构造适合轮廓形状相似度比较的局部特征描述符。

局部特征的计算依赖于尺度不变的特征点定位。利用高斯模糊模拟不同的尺度空间,并在多维尺度空间中寻找关键点。一个图像的尺度空间可以看作是一个变化尺度的高斯函数与图像本身的卷积。由于高斯差分(Difference of Gaussian, DoG)函数与尺度归一化的高斯拉普拉斯函数的近似性[28],利用 DoG 函数代替高斯拉普拉斯函数检测不同尺度空间的极值点。

特征点的检测分为三个过程:首先,比较尺度空间中点与其相邻 26 个点来计算极值点,这样产生的极值点并不全都是稳定的特征点,因为某些极值点响应较弱,而且 DoG 算子会产生较强的边缘响应;其次,对尺度空间 DoG 函数进行曲线拟合,计算相对插值中心的偏移量,当任一维度 x, y, σ 上的偏移量大于 0.5 时,改变当前关键点的位置,其中 x, y 为像素位置,σ 为高斯标准差,另外删除像素值小于 $0.04/S$ 的极值点,其中 S 为组内层数;最后,剔除不稳定的边缘响应点,计算特征点处的 Hessian 矩阵 H,并且满足公式 $Tr(H)^2/Det(H) < (r+1)^2/r$ 则保留关键点,否则删除,其中 $Tr(H)$ 表示矩阵 H 对角线元素之和,$Det(H)$ 表示矩阵 H 的行列式,r 为常数取值为 10。

特征点拥有三个信息:位置、尺度以及梯度方向。为了描述特征点及周边区域特征,为每个特征点建立一个描述符。首先需要确定特征点周边局部区域范围,将特征点附近的领域划分为 $d \times d$ 个子区域,其中 $d = 4$,每个子区域大小为 $3\sigma \times 3\sigma$,并有 8 个方向。为了确保旋转不变性,将坐标轴旋转为关键点的方向。旋转后的采样点坐标在半径为 r 的圆内被分配到 $d \times d$ 的子区域,计算每个子区域 8 个方向的梯度。统计的 $4 \times 4 \times 8 = 128$ 个梯度信息进行归一化后所得到的向量值即为该特征点的描述符。

4.3 视觉词汇计算

不同的局部特征向量代表了不同的视觉形状,图 3 为利用 SIFT 描述子提取的局部区域以及区域中关键点特征值及方向。由于图中是轮廓图像在不同尺度上局部区域特征在原图中的映射结果,区域大小和尺度变换因子 σ 相关,在大尺度图像上有较大区域,在小尺度图像上有较小的区域,因此图中区域大小不同,且会发现在原图的空白区域中有关键点出现。不同区域特征描述了不同的轮廓形状,而不同区域特征描述向量代表了不同的视觉词汇。

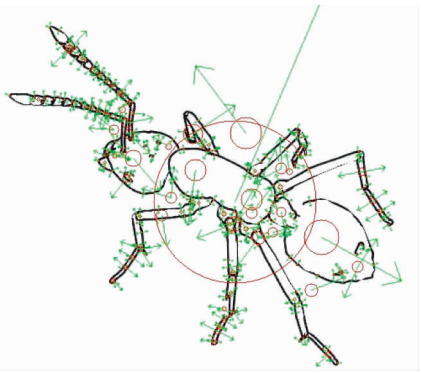


图 3 图像轮廓区域特征表示

相近的视觉形状具有相似的特征向量描述,为了将这些相近的视觉形状用相同的数据向量表示出来,那么这个数据向量看做是这些视觉形状的词汇描述。为了获取足够的视觉词汇描述,本文选取 500 个三维模型数据作为样本数据。每个三维模型数据有 20 个方向的轮廓投影,对每个投影轮廓提取关键点及其局部区域特征描述,采用 K-means 方法对 10000 幅图像中的所有局部特征进行聚类,每个聚类中心定义为一个视觉词汇。聚类中特征描述子的距离计算公式为

$$d(F_1, F_2) = \sqrt{\frac{(F_1 - F_2)^{\mathrm{T}}(F_1 - F_2)}{Cov(F_1, F_2)}} \qquad (5)$$

其中 F_1 为特征向量, Cov 为两向量的协方差值。

局部特征聚类后产生 200 个聚类簇,每个聚类簇中心描述为一个视觉词汇值,并将此视觉词汇值加入视觉词汇库中。

4.4 三维模型形状检索

将视觉词汇库内容作为标准值对轮廓中的所有局部特征值进行衡量。对同一个轮廓图像,统计与不同视觉词汇相似的局部特征数量,并将统计后的量化结果作为轮廓图像的特征值。对于图像特征提取,对象经过图割方法分割后形成二维轮廓图像,提取此二维轮廓图像的关键点及关键点周边局部区域特征,对图像中所有的关键点局部区域特征与 200 个视觉词汇进行比较,选择距离最小的视觉词汇用于描述此关键区域特征,并进行直方图统计,形成一个 200 维的量化特征向量用来描述图像特征。对于三维模型特征提取,则首先对三维模型提取轮廓线表示,其次利用十二面体投影方法生成轮廓线模型的 20 幅轮廓图像,对每个轮廓产生一个 200 维的量化特征向量,即生成 20 × 200 维向量作为三维模型的特征向量。图 4 为图 3 中蚂蚁轮廓的量化特征值表示结果,可以看出此特征值是一个稀疏向量。

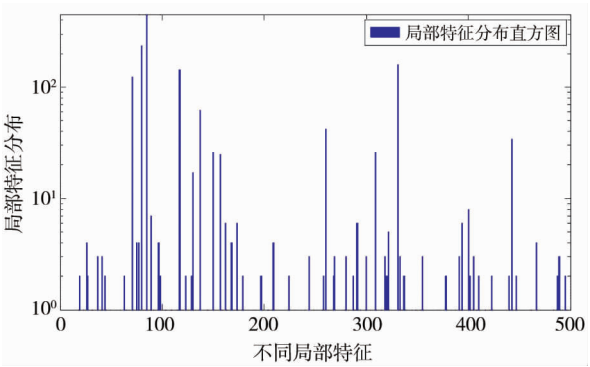


图 4 特征值表示结果

对象轮廓与三维模型相似度比较被抽象为特征值的比较,图像特征值向量与三维模型的 20 个不同方向特征向量分别进行距离计算,距离最小的那个结果作为图像与模型之间的相似距离:

$$D(I,M) = \min_{j=1}^{20}(d(F_I,F_{Mj})) \tag{6}$$

其中 I 和 M 分别表示图像和三维模型, j 为三维模型的不同投影图像序号。相似距离越小,表示对象与三维模型越相似,三维模型越能代表对象内容。

5 三维场景合成

利用轮廓形状相似度,从模型数据库中检索出了与物体最相似的前 20 个三维模型。一般情况,系统默认相似度最高的模型为被选模型,对被选模型进行坐标尺度变换。同时,提供用户反馈选择模式,用户可以根据主观判断,选择其它相似模型代替备选模型。为了和已选择的背景模型相适应,备选模型需要进行坐标及尺度变换 $\boldsymbol{P}' = \boldsymbol{P} \times \begin{bmatrix} \boldsymbol{S} & 0 \\ \boldsymbol{T} & 1 \end{bmatrix}$,其中 $\boldsymbol{P}$ 为备选模型原始空间齐次坐标 $(x,y,z,1)$, $\boldsymbol{P}'$ 为变换后的新坐标。变换矩阵中 $\boldsymbol{S}$ 为尺度缩放矩阵, $\boldsymbol{T}$ 为平移矩阵。

尺度缩放矩阵是利用图像中对象与背景的相对尺度,调整备选模型包围矩阵的缩放比例,从而使得备选模型与背景模型的尺度比例符合正常视觉效果。缩放矩阵 $\boldsymbol{S}$ 为一个 3 × 3 的对角矩阵,对角元素为缩放比例 s 。其中 s 为图像场景中物体与背景长宽比例的较大值: $s = \max(\frac{H_o}{H_c},\frac{W_o}{W_c})$,其中 H_o、W_o 代表物体在图像场景中的长宽值, H_c、W_c 为图像场景的长宽值。

平移矩阵 $\boldsymbol{T}$ 是一个 1 × 3 的矩阵, $\boldsymbol{T} = (x_c \quad y_c \quad z_c)$,其中平移矩阵的元素来源于背景模型的中心位置。

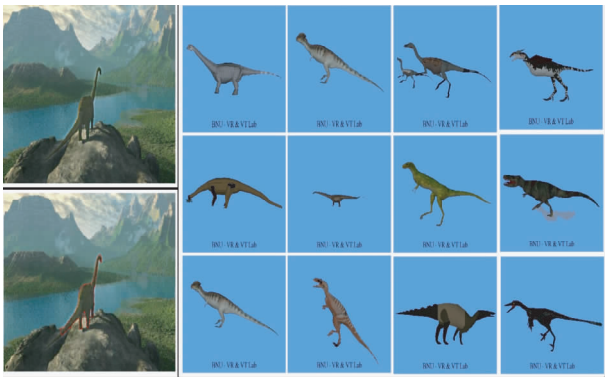
经过坐标变换后的备选模型加入到背景场景中,其空间位置仍需要用户进行二次微调方可被至于合适的位置。为了达到用户操作的真实感觉,系统实现过程中加入了物理碰撞检测模拟。

6 实验结果

6.1 检索结果

检索的三维模型数据来源于拥有 3 万多条三维模型数据的数据库,每个三维模型均拥有缩略图。数据分为 25 个大类,409 个小类。

本文提出了基于图像分割结果检索物体三维模型,完成三维场景建立的建模新方法。图 5 为利用图像对象分割后检索的三维模型结果。(a)图检索的是恐龙对象,(b)图检索的是花瓶图像。由于恐龙图像比花瓶图像拥有更多的形状轮廓信息,因此检索的结果更符合用户需求,花瓶轮廓单一且较为平滑,因此利用 SIFT 特征描述效果不如恐龙轮廓形



(a)恐龙对象轮廓分割后检索结果,左下为分割后轮廓显示效果

(b)花瓶对象检索结果

图 5　图像中物体分割后三维模型检索结果

状特征描述丰富,检索结果稍差。

为了验证基于视觉词汇三维模型检索算法的有效性,将本算法检索效果与基于深度图像三维模型检索算法[29]以及形状分布 D2 算法[20]进行了比较。比较了三种算法在汽车分类中的查准率及查全率,此分类中共有数据 1643 个。图 6 中横轴为查全率,纵轴为查准率,在查全率越来越高的情况下,曲线仍然保持更高的查准率,则表示该曲线所代表的检查方法越好。

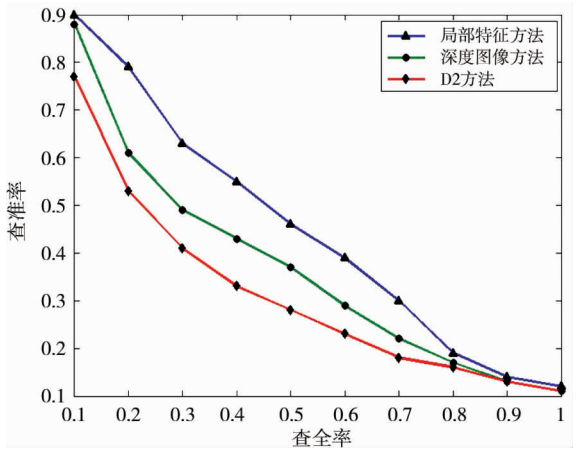


图 6　特征检索算法的查全率查准率比较

基于深度图像三维模型检索算法和 D2 算法均是面向于模型的全局特征提取的。其中深度图像特征中按照三视图原理提取模型的三个方向深度图像,利用图像傅里叶变换后低频系数作为模型的特征进行比较。此算法只考虑到三个方向投影,对模型的空间信息丢失过多,另外算法在进行特征计算过程中没有考虑到局部细节信息对于模型投影的影响,且对于汽车模型本身细节信息丰富,因此按照深度图像算法的查询结果略差于本文方法。D2 算法作为三维模型检索中最经典的方法,该方法简单和有效,实验中取 D2 特征的 200 维数据作为三维模型形状描述子,这种算法在三维模型顶点数较多时为随机取点进行运算,在一定程度上影响了形状相似度比较的效果。

6.2　建模结果

三维场景建立最终效果依赖于两个方面的结果,第一是物体三维模型检索及选择的结果,第二是用户选择的背景模型的结果,只有这两个模型都恰当,且在模型合成位置合理,才会得到一个理想的三维场景。目前数据库中有三维背景模型 40 个,分别是对常见的室内室外场景的建模结果。图 7 为对不同场景图像进行三维重建的结果,左侧图为原始二维场景图像,红线勾勒的是物体分割结果,右侧图为合成后的三维场景。合成后的三维场景可根据用户的具体需求进行二次编辑。生成的三维场景可根据用户需要修改光照变化及渲染效果,由于算法实施过程中并未考虑二维图像场景与三维模型场景的完全一致效果,因此未曾将二维图像作为三维模型纹理图像处理并计算 UV 坐标信息,本文中的三维模型背景及检索出的模型均以用户最终选择的为主。

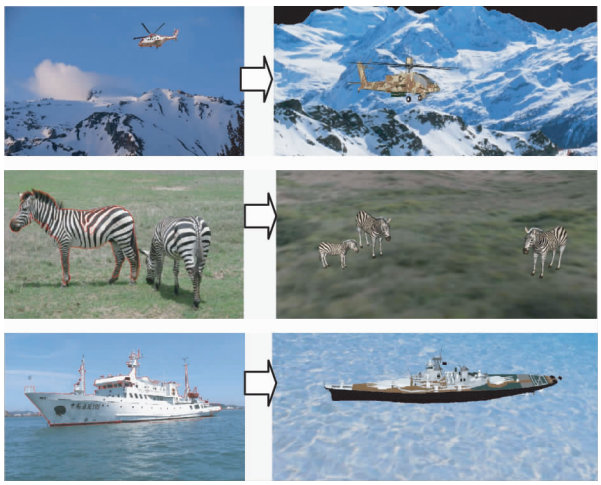


图 7　三维场景建模结果

图 8 为系统实现界面图,其中界面左侧为轮廓形状检索出的三维模型缩略图列表,右侧有四个视图,分别为图像场景视图,三维虚拟场景视图,加载物体模型视图以及加载背景视图,为了节省系统内存开销,如图 8 中所示背景模型被加载后系统将不再显示。

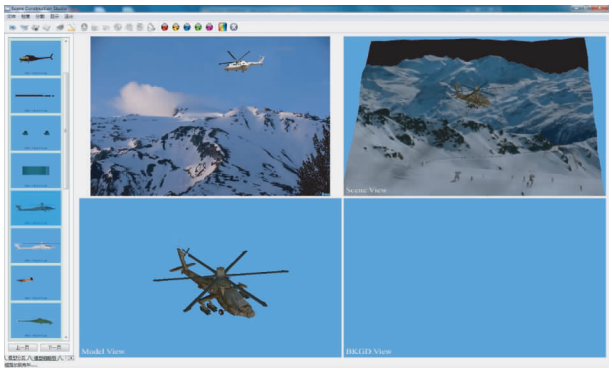


图 8　系统实现示例

7　结　论

本文提出了一种新的三维场景建模方法:首先利用图像场景作为参考,分割图像场景的物体轮廓信息,然后依据物体轮廓信息检索与它相似的三维模型数据,最后在检索出的待选模型中选择合适的三维模型,并从背景场景中选择合适的背景场景与此三维模型数据进行合成,形成与二维图像场景相匹配的三维虚拟场景。实验结果证明,检索算法有效,三维建模效果理想。该方法的实现过程仍需进行改进,以适应更加广泛的应用。第一,由于聚类算法的局限性,依据 SIFT 特征的局部特征方法并不是一种稳定方法,不能够保证对同一形状能够获得相同的量化结果,因此下一步仍需要提高特征提取算法的稳定性。第二,为了系统应用更加广泛,背景模型库还需要大量进行扩充。三维物体数据量以及数据种类虽然已经比较丰富,但仍需要不断扩充。第三,在三维场景绘制方面为了取得更好的真实感效果,需要结合图像的明暗效果调整三维场景的光照及阴影显示效果。相信经过进一步的改进,本文方法会具有更好的应用效果。

参考文献

[1] Henrichs S. 3ds max environment modeling #1:Procedural stone. http://saschahenrichs. blogspot. co m/2010/03/3dsMax-environment-modeling-1. html,2010

[2] Gaspar J. Google SketchUp Pro 8 Step by Step. USA:VectorPro Publisher,2011. 1-3

[3] Funkhouser T, Kazhdan M, Shilane P, et al. Modeling by Example. In:Proceedings of the 2004 Acm Siggraph, New-York, USA,2004. 652-663

[4] Remondino F, Sabry E. Image-based 3D Modelling:A Review. *The Photogrammetric Record*, 2006, 21(115):269-291

[5] Xu K, Zheng H, Zhang H, et al. Photo-inspired model-driven 3D object modeling. *ACM Transactions on Graphics*, 2011, 30(4):80

[6] Eitz M, Hildebrand K, Boubekeur T, et al. PhotoSketch:A Sketch Based Image Query and Compositing System. In SIGGRAPH 2009 Talk Program. Artical No 60

[7] Chen T, Cheng M, Tan P, et al. Sketch2Photo:Internet Image Montage. *ACM Transactions on Graphics*, 2009, 28(5):Artical No 124

[8] Chen G, Chen T, Zhang F, et al. Data-Driven Object Manipulation in Images. *Computer Graphics Forum*, 2012, 31(2):265-274

[9] Lan T, Yang W, Wang Y, et al. Image retrieval with structured object queries using latent ranking SVM. In ECCV 2012 Lecture Notes in Computer Science. 2012, 7577:129-142

[10] Engel D, Herdtweck C, Browatzki B, et al. Image retrieval with semantic sketches. Human computer interaction 2011, *Lecture Notes in Computer Science*. 2011, 6946:412-425

[11] Tao W, Jin H, Zhang Y. Color Image Segmentation Based on Mean Shift and Normalized Cuts. *IEEE Transactions on Systems, Man, and Cybernetics Part B*, 2007. 37(5):1382-1389

[12] Provost J, Collet C, Rostaing P, et al. Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps. *Computer Vision and Image Understanding*, 2004, 93(2):155-174

[13] Kang C, Wang W. A novel edge detection method based on the maximizing objective function. *Pattern Recognition*, 2007, 40(2):609-618

[14] Li N, Liu M, Li Y. Image segmentation algorithm using watershed transform and level set method. In:Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2007, 613-616

[15] Kolmogorov V, Zabih R, Gortler S. Generalized multi-camera scene reconstruction using graph cuts. In:Proceedings of the European Conference on Computer Vision, 2002, 501-516

[16] Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization. *Vi-*

sion IEEE Transactions on Pattern Analysis and Machine, 2004,26(9):1124-1137

[17] Rother C, Kolmogorov V, Blake A. Grabcut interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*,2004,23(3):309-314

[18] Tangelder J, Veltamp R. A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications*,2008,39(3):441-471

[19] Zhang C, Chen T. Efficient feature extraction for 2D/3D objects in mesh representation. In: Proceedings of the International Conference on Image Processing,2001,3:935-938

[20] Robert O, Thomas F, Bernard C, et al. Shape distributions. *ACM Transactions Graphics*,2002,21(4):807-832

[21] Chua S, Jarvis R. Point signatures: a new representation for 3D object recognition. *International Journal of Computer Vision*,1997,25(1):63-65

[22] Johnson A, Hebert M. Using spin images for efficient object recognition in cluttered 3D scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*,1999,21(5):635-651

[23] Squire D, Mueller W, Mueller H, et al. Content-based query of image databases: inspirations from text retrieval. *Pattern Recognition Letters*,2003,21(13):1193-1198

[24] Greig D, Porteous B, Seheult A. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society*,1989,51(2):271-279

[25] Judd T, Durand F, Adelson E. Apparent ridges for line drawing. *ACM Transactions on Graphics*,2007,26(3): Artical No 19

[26] Wang Y, Liu R, Baba T, et al. An images-based 3d model retrieval approach. In: Proceedings of the 14th International Conference on Advances in Multimedia Modeling. 2008,90-100

[27] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*,2004,60(2):91-110

[28] Lindeberg, T. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*,1994,21(1-2):225-270

[29] Liu Y, Zhou M, Fan Y. Using depth image in 3D model retrieval system. *Advanced Materials Research*, 2011, 268-270:981-987

3D Modeling of scene images based on shape retrieval

Fan Yachun, Tan Xiaohui, Zhou Mingquan, Lu Zhaolao

(College of Information Science and Technology, Beijing Normal University, Beijing 100875)

Abstract

A new method for quick building of a 3D virtual scene by using 2D scene images is proposed. It has the charcteristis below: an improved graph cutting algorithm is used to segment objects and extract contour information in the scene image; a new local shape description is used to describe the contour of 3D model projections and 2D image by using quantized visual vocabulary vectors which are based on scale-invariant feature transform; 3D models are retrieved from 3D databases by computing the distances between shape contours of different 3D models; coordinate position is transferred to combine object models to build the visual scene. The experimental results show that the performance of the proposed approach is quite well and it can effectively build complicated virtual scenes.

Key words: 3D modeling, graph cut, shape retrieval, visual vocabulary, scale-invariant feature transform