

# 基于层次结构的 3D 室内场景结构重组

张嘉培 梁 成

(中国科学院深圳先进技术研究院 深圳 518055)

**摘 要** 室内场景, 连同建筑和植物, 是构成人类生存环境的三大重要因素。但是, 目前可获取的三维室内场景并不像人们期望的那样是经过合理性语义层面上的分割, 而是存在着大量的拓扑结构上的错误。这些拓扑错误是由模型重用产生的碎片以及模型转换中导致的信息缺失等因素造成的。本文通过考虑拓扑结构关系(支撑关系和主次关系)以及几何关系(对称、对齐、相交、连接等), 循环迭代地利用图割方法, 构建了一个具有层次结构的场景树。该层次结构树对场景进行了层级的重整, 用户可以根据需要得到不同细节层次上的有意义的语义分割, 从而纠正了原有模型拓扑结构上的错误, 为进一步的场景编辑提供了便利。

**关键词** 3D 室内场景; 图割算法; 层次结构; 基于支撑关系的图割; 基于主次关系的图割

## Hierarchy-Based Reorganization for 3D Indoor Scene

ZHANG Jia-pei LIANG Cheng

(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

**Abstract** Indoor scene plays an important role in human living environment which constitutes the urban landscape together with building and flora reconstruction. However, the accessible 3D indoor scene data are usually not as semantic and reasonable as expected. They may contain a multitude of topological errors and omissions. This easily results in fragments production during model reusing and incompatible transform during format conversion. We creatively build a scene hierarchy using “Graph Cuts” method iteratively by considering the topological relation (“support & non-support” and “primary & non-primary”) and geometry relation (symmetry, alignment, intersection and connectivity). The hierarchy reorganizes the scene in different levels and users could select the required level of detail which contains semantic segmentation. It corrects the topological errors and offer convenience for further scene edit.

**Keywords** 3D indoor scene; graph-cut; hierarchy; Support Cuts; Primary Cuts

## 1 引 言

在众多的三维数据模型中, 室内场景因其复杂性(室内场景物体种类繁多)和规律性(基本由墙、地板、桌椅等组成)越来越受到研究人员的瞩目。很多图形学相关领域, 如 3D 游戏、虚拟现实、数字城市等, 也在很大的程度上依赖于这些室内场景方面的研究。很多现成的大型三维模型场景数据库(如 Google 3D Warehouse)为我们获取室内场景的数据提供了方便, 但是这些室内场景数据实际上存在着很多拓扑结

构上的混乱和错误, 比如不该属于同一物体的部件被成组为同一物体, 原本完整的物体被肢解为不同部件等。这些成为我们对室内场景的高层次的理解或者进一步的编辑的一大障碍。本文旨在通过引入人对于室内场景的理解, 借用计算机视觉中普遍用到的“图割算法”对场景进行层次化结构重组, 从而为室内场景后续研究工作铺平道路。

从互联网上三维模型库下载的室内场景数据并非如我们期望的那样经过了语义层面上的分割, 有的网格可能分割得并不彻底, 有的网格则属于过分割。这主要由两方面的因素造成的: (1) 建模人员经常重用

基金项目: 国家自然青年科学基金资助项目(61103166)。

作者简介: 张嘉培, 硕士研究生, 研究方向为计算机图形学, E-mail: jiapeonline.zhang@gmail.com; 梁成, 助理研究员, 研究方向为计算机图形学。

些已有的网格模型,在此基础上进行复制重组等操作来构建新的模型,因此如果没有经过人工成组的话,那么新模型则是由一些零碎的网格构成的;(2)不同三维模型格式(3ds, obj, stl, dxf, wrl 等)之间的转换可能会导致某些原始几何信息的错误和缺失,一个原本完整的物体因为缺失了小部分几何信息从而被分裂成若干个网格碎片。

室内场景模型最好是在建模期间,建模人员就能够将各个物体的网格进行有意义的成组,但是如果是非常复杂的大规模室内场景,对于建模人员来说将是一项非常繁杂的工作,所以我们期望有一种方法能够按照人们对室内场景的理解将具有不规整分割布局的原始模型进行自动重组,从而减少人工操作,同时使得对室内场景的高层次理解成为可能。为了解决这个问题,我们必须充分利用原始数据已有的信息,其中拓扑信息和几何信息是最直接也是最具有普适性的。因此我们通过分析原始数据中的拓扑信息和几何信息,生成一个具有层次结构的场景树,利用这个场景树对场景进行层级的组织管理。下面我们从三个方面来阐述本研究工作的创新点:

(1)几何信息和拓扑信息是我们考虑利用原始数据的最主要的两个内容。原始的三维网格模型还包含了其他的有用信息,如标识(Tag)、纹理(Texture)、材质(Material)等,但是因为包含过多的人为因素所以并不具有普适性。而几何信息和拓扑信息相对而言是三维模型中更加稳定更加易于用数学的手段提取信息的属性,因此利用它们可以适用于大部分的室内场景。

(2)我们将整个场景表达为一种层次结构。借用 LOD(Level of Detail)的思想,通过迭代的分割,我们得到不同细节程度的场景,从代表整个场景的根节点往下,层层细分,逐层递进,并且每一个中间层都引入了正确的拓扑关系对最细节层次上的拓扑错误进行隐藏和纠正。

(3)我们利用的分割手段是图割方法。在计算机视觉领域,图割方法被广泛应用于图像分割,而在三维模型中利用图割方法进行分割的工作却很少。我们不仅把图割方法应用到了三维室内场景模型的分割中,并且还迭代的生成层次结构。在图割算法的设计中,我们融入人对于室内场景的理解,将几何关系和拓扑关系通过对各种数据项的巧妙定义,从而使得我们的算法不仅具有普适性而且具有灵活性。

近年来有一些出色的室内场景方面的研究工作,如家具配置<sup>[1-4]</sup>和 3D 模型搜索<sup>[5-7]</sup>,但是这些工作都

依赖于一个重要前提,即室内场景是经过语义分割的,每一个物体都是具有意义的独立个体,而不是一些零碎的网格碎片。我们将本文提出的方法试验了大量的三维室内场景,对于大部分室内场景,我们都能够对其进行正确的语义分割,对于极少数分割错误的情况也允许通过少量的用户交互进行修正。当然,我们的工作只是室内场景研究工作的一个前期自动处理过程,但是它是一项不可或缺的有重大意义的工作,因为它不仅解决了现实中存在的三维模型不规整分割的现象,而且也为后续的室内场景高层次理解提供了前提条件。

## 2 相关研究

### 2.1 家具配置

由于 3D 室内场景物体中家具居多,所以家具的配置和布局问题是室内场景研究的热点问题。Yu 等人开发了一套能够自动合成室内场景中合理的家具配置的系统<sup>[1]</sup>,该方法首先提取了一个层次化空间场景结构,然后用模拟退火算法对能量代价函数进行优化。另外还有一个基于室内场景设计的交互式家具布局系统<sup>[2]</sup>,它是将室内场景设计中的一些准则融入其中,这启发我们可以在层次结构中引入空间布局的某种关系来帮助进行分割。还有的工作是从一组建筑平面图自动生成建筑的布局<sup>[3]</sup>,这其实也是一种和家具布局相似的空间分配问题,它是用一种基于概率图模型的数据驱动的方法来提取节点间的语义关系。虽然在本课题中我们没有用到数据驱动的方法,但是它对于在三维场景中提取有用信息是一种很有效的方法。虽然以上对室内场景的研究形式多样并且具有一定的实用性,但是它们都是以室内场景都是由有意义的物体组成的这一前提为基础的。

### 2.2 3D 模型搜索

目前网络上的大规模 3D 模型数据库使得用户能够越来越容易获得想要的 3D 模型。为了适应大规模 3D 模型数据库发展的需求,3D 模型搜索技术显得越来越重要。现有的 3D 模型搜索方法有关键字搜索(Keyword-Based Search)、基于形状搜索(Shape-Based Search)、基于素描搜索(Sketch-Based Search)、组合式搜索(Assembly-Based Search)以及上下文环境搜索(Context-Based Search)等。由于关键字搜索、基于形状搜索和基于素描搜索要么只是用到文字方面的描述信息,要么只是关注于搜索模型自身的

特征信息,所以它们不是我们的重点考虑对象。组合式搜索<sup>[4]</sup>是对于一个给定的3D模型(缺失其中某些组成部件),运用“贝叶斯网络”概率图模型的方法来代表部件间的关系,进而搜索缺失的部件。上下文环境搜索参考文献[5]与本课题的研究更相关,它是指用户在场景中放置一个搜索盒子,算法通过比较物体的形状特征、标签信息和空间关系等,用一个权值来描述待匹配模型和场景中的其他模型间的匹配程度,然后以匹配程度从大到小的顺序向用户展示搜索结果。本课题借鉴了文献[5]中组织语义信息的方法,例如如何衡量两个物体间的相似性等。以上的3D模型搜索方法仍然以室内场景都是由有意义的物体组成这一前提为基础的,所以对于目前并不理想的3D室内原始模型来说仍具有很大的局限性。

### 2.3 图割方法(Graph Cuts)

Graph Cuts是基于图论中的最大流最小割理论的一种解决能量函数最小化问题的最优化算法,它广泛应用于计算机视觉中的立体视觉匹配(Stereo Matching)、图像重建(Image Restoration)、纹理合成(Texture Synthesis)等问题中。

Graph Cuts的开山之作<sup>[6]</sup>系统介绍了如何构造图和能量项来解决立体视差、视觉运动等问题,此后Graph Cuts还应用于立体视觉的匹配<sup>[7]</sup>,改善了传统Graph Cuts计算耗时的缺点。KOLMOGOROV等人提供了最小化能量函数的通用构造方法<sup>[8]</sup>,并给出了最小化二进制能量函数的必要条件。Graph Cuts也被广泛应用于图像分割领域<sup>[9]</sup>,这也是Graph Cuts最主要的应用领域。

Graph Cuts算法大致是对一个由三个能量项构成的能量代价函数:

$$E_l = \sum_p D_p(l_p) + \sum_{(p,q)} V_{(p,q)}(l_p, l_q) + \sum_{L'} h_{L'}(l)$$

进行最优化的过程,  $\sum_p D_p(l_p)$  是数据项(Data Cost), 其中  $D_p$  是将结点  $p$  赋予  $l_p$  的代价值;

$\sum_{(p,q)} V_{(p,q)}(l_p, l_q)$  是平滑项(Smooth Cost), 其中  $V_{(p,q)}(l_p, l_q)$  是指将结点  $p$  和结点  $q$  分别赋予  $l_p$  和  $l_q$  的代价值;

$\sum_{L'} h_{L'}(l)$  是标签项(Label Cost), 其中  $L'$  是所有的标签集合  $L = \{l_0, l_1, \dots, l_n\}$  的非空子集,  $h_{L'}(l)$  是指如果标签  $l$  出现在子集  $L'$  中的代价值, 这一项主要用在多标签的图割算法中, 我们的算法里没有用到这一项。

在图像分割和立体视觉领域, Graph Cuts 是区分前

景和背景的有力工具<sup>[10,11]</sup>。但是在三维图形领域却很少用到 Graph Cuts, 其中有一项工作<sup>[12]</sup>是利用 Graph Cuts 来分割三维网格模型, 用户先粗略地在模型上画线来决定包含部分和剔除部分作为分割的参考, 然后 Graph Cuts 再迭代地对构建的有权图进行分割, 该文还利用了层次化结构加速来适应大规模的网格模型, 因为传统的 Graph Cuts 方法只能适用于小的模型。我们在本课题中也是利用 Graph Cuts 这把“剪刀”对场景进行多种具有语义的分割(Support Cuts 和 Primary Cuts)。

Graph Cuts 方法允许用户自定义能量代价函数来适应不同问题的需求, 并且能够产生很好的分割效果。正因为它的有效性、灵活性和广泛性, 所以我们在本课题中选用 Graph Cuts 作为构建层次化结构场景图的核心算法。

## 3 技术细节

### 3.1 预处理

预处理阶段主要是构建基本单元和进行粗分割。这一步为后续的工作提供了基本的数据结构处理单元, 另外粗分割也为层次结构树提供了一个雏形。

#### 3.1.1 基本单元(Patch)

我们把原始的三维模型网格按初态封装在一个称之为“Patch”的结构单元里, 这个结构单元里包含了一些我们后续工作需要的属性值: (1) 包围盒, 用来衡量相邻的两个“Patch”之间的对齐程度等; (2) PCA 值, 用来衡量两个“Patch”是否在几何上是相同的; (3) label 值, 用来标记每次 Graph Cuts 之后产生的拓扑关系(支撑关系或主次关系)等等。

#### 3.1.2 粗分割(Rough Group)

对于一般的室内场景, 墙、地板、天花板作为最主要的支撑物直接或者间接地支撑着场景中其他的物体。所以, 我们首先根据高度和法线分布来检测出墙、地板和天花板。受到“分治法”思想的启发, 我们假设移除了墙、地板、天花板, 再将场景中其他物体按照连接关系分成几组, 称之为“簇”(Cluster)。在此基础上, 我们构建层次结构场景树的雏形, 见图1。在接下来的步骤中, 我们以“簇”为单位利用图

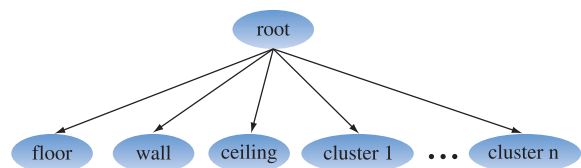


图1 “粗分割”后的结构场景树示意图



割算法进行层级的分割。

### 3.2 层次结构场景

#### 3.2.1 Support Cuts

观察发现, 室内场景普遍存在的一个拓扑关系是支撑与被支撑的关系。例如茶几和茶几上的花瓶就是支撑和被支撑的关系。据此我们根据传统的图割算法, 设计了一个引入了支撑关系的图割方法, 我们称之为“Support Cuts”。对于粗分割产生的每一个“簇”, 我们需要最小化能量函数:

$$E = \sum_p D_p(l_p) + \lambda \sum_{p,q} V_{p,q}(l_p, l_q)$$

其中, 前一项是数据项(Data Cost),  $D_p(l_p)$  表示把  $l_p$  赋给结点  $p$  的代价值(cost); 后一项是平滑项(Smooth Cost),  $V_{p,q}(l_p, l_q)$  表示把  $l_p$  赋给结点  $p$  并且把  $l_q$  赋给结点  $q$  的代价值。  $l_i \in \{SP, NSP\}$ ,  $SP$  代表被支撑物体,  $NSP$  代表支撑物体。

数据项是根据一个结点属于支撑物体的概率来定义的。我们观察到, 基于上一层次分割产生的支撑物体, 如果一个结点离这个支撑物体越近, 那么它越有可能作为这一次分割的支撑物体。比如粗分割以后, 待分割的结点有茶几和茶几上的花瓶, 地板是上一层次的支撑物体。茶几离地板比花瓶离地板更近, 那么茶几就有很大的概率在这一次分割中属于支撑物体。对于一群未经标记的结点  $P = \{p_0, p_1, \dots, p_n\}$ , 我们分别计算出它们离上一层次分割产生的支撑物体的最短步长  $L = \{l_0, l_1, \dots, l_n\}$ , 那么数据项就定义为:

$$D_p(SP) = \frac{l_p}{\max(L)}, D_p(NSP) = 1 - D_p(SP)$$

平滑项是根据两个结点归属于同一物体的概率来定义的。这里我们考虑了三个因素: 相似度(Similarity Term)、对齐度(Alignment Term)、连接度(Connectivity Term)。对于任意两个结点  $p$  和  $q$ , 我们定义:

$$V_{p,q}(NSP, NSP) = V_{p,q}(SP, SP) = 0$$

$$V_{p,q}(NSP, SP) = V_{p,q}(SP, NSP) = \alpha S_{p,q} + \beta A_{p,q} + \gamma C_{p,q}$$

其中,  $S_{p,q}$  是相似度,  $A_{p,q}$  是对齐度,  $C_{p,q}$  是连接度,  $\alpha, \beta, \gamma$  分别是它们的比例因子。下面对平滑项的三个因素如何设置进行详细叙述:

(1) 相似度通过比较两个结点的 PCA 的长度以及结点下的“Patch”的材质信息是否一致, 来衡量两者间几何形状的相似程度。

(2) 对齐度通过比较两个结点包围盒在某个平面上的重叠比例, 来衡量两个结点的对齐程度。

(3) 连接度通过计算两个结点之间的最短步长来衡量两个结点连接是否紧密。如结点  $i_0$  和结点  $i_n$  之间的最短路径是  $path = \{i_0, i_1, \dots, i_n\}$ , 那么它们之间的连接度定义为:

$$C_{p,q} = \frac{1}{n}, p, q \in Path$$

“Support Cuts”是一个二分操作, 每一次切割对应产生一个新的场景层次, 该层由一个支撑物体和若干个被支撑物体组成。图 2 显示了“Support Cuts”作用在一群待分节点以后产生的场景结构图。

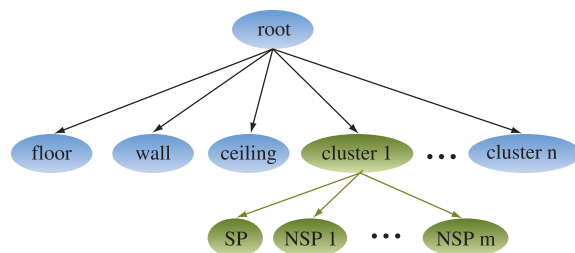


图 2 “Support Cuts”分割后的场景结构图示意图 (绿色部分为“Support Cuts”作用在“Cluster 1”结点后分割出了一个支撑结点“SP”和若干非支撑结点“NSP 1”至“NSP m”)

#### 3.2.2 Primary Cuts

虽然“Support Cuts”能够对大多数场景进行有效的切割, 但是对于某些场景切割的结果仍然不理想。在研究中我们发现, 有些物体并排横置且相互连接, 这样就不存在支撑关系, 那么用“Support Cuts”就无法将它们分割开来。虽然这种情况下物体之间没有支撑关系, 但是我们观察到, 其中有的物体会比另一些物体起主导作用。比如, 卧室中的床和床两边的床头柜连成了一片, 它们之间没有支撑关系, 但是我们明显觉察到床相对于床头柜来说在这群物体中起主导作用。因此, 我们对图割方法又引入了另外一种关系——主次关系, 并且称这个图割方法为“Primary Cuts”。同样, 对于用“Support Cuts”无法分割出来的一群待分割结点, 我们需要最小化能量函数:

$$E = \sum_p D_p(l_p) + \lambda \sum_{p,q} V_{p,q}(l_p, l_q)$$

数据项和平滑项的意义同“Support Cuts”, 而  $l_i \in \{PR, NPR\}$ ,  $PR$  代表主要物体,  $NPR$  代表次要物体。

这种主次关系是怎样来衡量的呢? 我们观察到, 如果一个物体在某一参考面上的表面积越大, 并且连接它的物体数量越多, 那么它在这群物体中起主导作用的概率也就越大。据此, 我们可以如下设计

“Primary Cuts”的数据项。

对于每一个待分割物体，我们首先算出它们在某个参考面上的最大表面积和与它们相连物体的数目，并且分别将其标准化到[0,1]范围内，得到一组标准化的表面积  $S=\{s_0,s_1,\dots,s_n\}$  和连接数  $C=\{c_0,c_1,\dots,c_n\}$  以后，我们依据下列公式来跟别计算出它们的“主要程度分数”：

$$p_p = \varphi \cdot c_p + \omega \cdot s_p$$

$$\varphi + \omega = 1, 0 < \varphi < 1, 0 < \omega < 1$$

这样得到的一组标准化的“主要程度分数”  $P=\{p_0,p_1,\dots,p_n\}$  后，我们定义“Primary Cuts”的

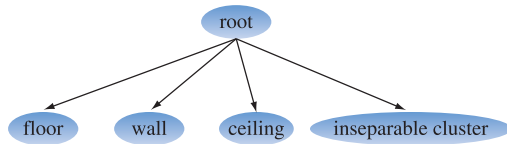
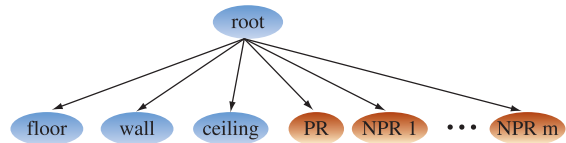


图3 同一层次上产生的具有主次关系的结点后场景结构树的示意图

(左图的场景结构树中包含了一个用“Support Cuts”不能分割的“簇”；右图是将“Primary Cuts”作用在“Support Cuts”不能分割的“簇”上)



### 3.2.3 Support Cuts&Primary Cuts

我们算法的核心是两种类型的图割算法，它们分别引入了室内场景中两种关系，即“支撑关系”和“主次关系”，并且在场景的切割中循环迭代。此部分主要阐述两种类型的图割算法在整个算法中的地位和作用。

“支撑关系”是具有普适性和传递性的。

首先，“支撑关系”在室内场景是普遍存在的拓扑关系。从物理意义上说，任何物体都需要被其他物体直接或间接支撑，所以以前研究工作提出的很多具体的拓扑关系如“包含”、“插入”、“悬挂”等都可以抽象为“支撑关系”。因此，我们就将室内场景中各种复杂的拓扑关系简化成了一种主要关系。

其次，“支撑关系”是可以传递的，迭代这种关系很容易能够构建层次结构。例如地面支撑着餐

桌，餐桌支撑着水果盘，水果盘又支撑着若干水果，那么这种递进关系可以描述为  $floor \xrightarrow{support} table \xrightarrow{support} tray \xrightarrow{support} fruits$ ，那么我们可以很容易根据这种三次递进关系形成三层的层次结构，如图4所示。

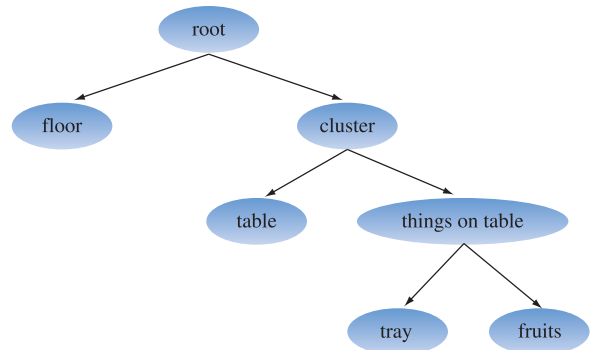


图4 根据“支撑关系”的递进生成的三层场景结构树

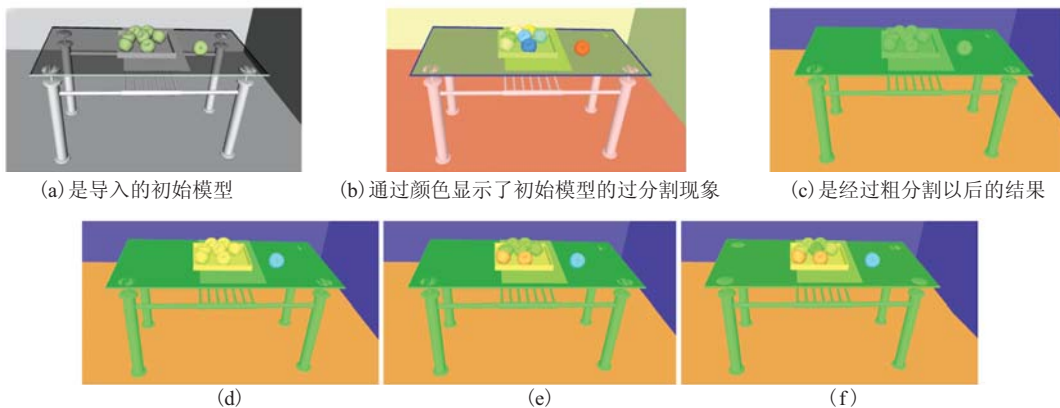


图5 一个简单的室内场景

((d) (e) (f) 是不断迭代分割以后，场景层次不断深入的结果)

而“主次关系”在室内场景中并不具有这种普适性和传递性, 它只是基于观察而得出的经验性结论, 用来补充“支撑关系”, 辅助解决“支撑关系”不能解决的问题。因此, 在迭代图割算法进行场景切割的时候, 我们优先使用“Support Cuts”, 并且生成新的场景层次, 只有当“Support Cuts”切不出来的时候, 我们才使用“Primary Cuts”进行辅助切割, 而且并不生成新的场景层次, 只是将原来的 NSP 结点在同层次上分裂成若干结点。

### 4 结果

我们先以一个简单的室内场景为例, 展示本文的算法的每一步执行的结果(见图 5)。图 6 是针对图

5 的场景建立的场景层次结构树, 它的每一层的深入对应着图 5 中的每一次分割。图 7 为用本文的算法对 Google Warehouse 中的一个较复杂的室内场景进行切割的实验结果。

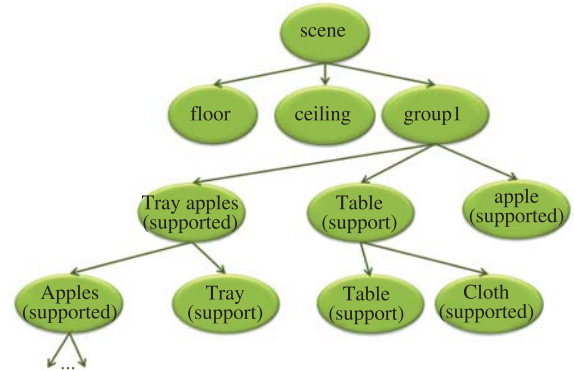


图 6 图 5 中的场景建立的场景层次结构树

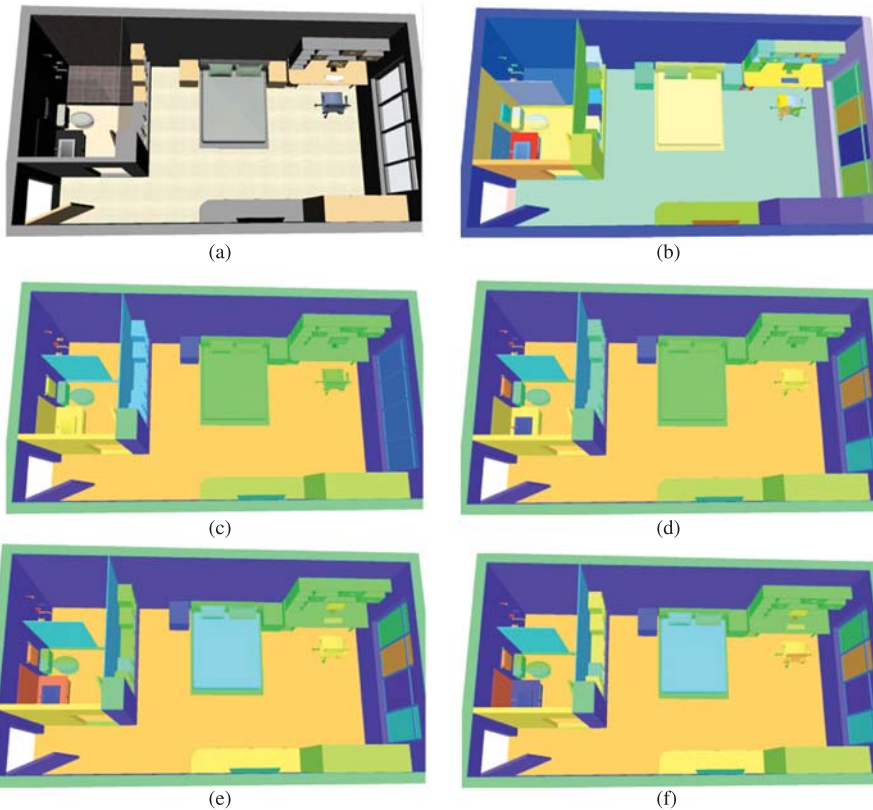


图 7 一个复杂的室内场景

(各子图的意义对应于图 5 各子图的意义)

我们看出, 本文提出的算法可以有效的对室内场景的三维模型进行有效的语义切割, 并且建立合理的层次结构场景图。

### 5 结 语

本文针对目前 3D 室内场景模型所存在的拓扑结

构混乱和过分割的现象, 将视觉领域常用的图割算法应用到三维场景中, 通过引入人对室内场景的理解(“支撑关系”和“主次关系”), 设计了符合三维室内场景模型的两类图割算法——“Support Cuts”和“Primary Cuts”进行迭代切割, 实验证明该方法能够有效地对大部分室内场景进行语义切割, 少数场景需要加入用户的手动交互。

在未来的工作中, 我们期望能够利用本文建立的层次结构场景树对室内场景进行更高层次的理解。

### 参考文献

- [1] Yu L F, Yeung S K, Tang C K, et al. Make it home: automatic optimization of furniture arrangement [C] // Proceedings of ACM SIGGRAPH, 2011: 86.
- [2] Merrell P, Schkufza E, Li Z, et al. Interactive furniture layout using interior design guidelines [C] // Proceedings of ACM SIGGRAPH, 2011: 87.
- [3] Merrell P, Schkufza E, Koltun V. Computer-generated residential building layouts [C] // Proceedings of ACM SIGGRAPH Asia, 2010: 181.
- [4] Chaudhuri S, Kalogerakis E, Guibas L, et al. Probabilistic reasoning for assembly-based 3d modeling [C] // Proceedings of ACM SIGGRAPH, 2011: 35.
- [5] Fisher M, Hanrahan P. Context-based search for 3d models [C] // Proceedings of ACM SIGGRAPH Asia, 2010: 182.
- [6] Boykov Y, Veksler O., Zabih R. Fast approximate energy minimization via graph cuts [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23 (11): 1222-1239.
- [7] Kolmogorov V, Zabih R. Graph based algorithms for scene reconstruction from two or more views [D]. NY, USA: Cornell University Ithaca, 2004.
- [8] Kolmogorov V, Zabih R. What energy functions can be minimized via graph cuts [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26 (2): 147-159.
- [9] Boykov Y, Jolly M-P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D Images [J]. International Conference on Computer Vision. 2001, I: 105-112.
- [10] Rother C, Kolmogorov V, Blake A. Graph-cut: interactive foreground extraction using iterated graph cuts [C] // Proceedings of ACM SIGGRAPH, 2004: 309-314.
- [11] Park A, Hong K, Jung K. Better foreground segmentation for 3d face reconstruction using graph cuts [C] // Proceedings of the 3rd Pacific Rim Conference on Advances in Image and Video Technology, 2007: 715-726.
- [12] Brown S, Morse B, Barrett W. Interactive part selection for mesh and point models using hierarchical graph-cut partitioning [C] // Proceedings of Graphics Interface, 2009: 23-30.