

# 基于单幅图像的三维浏览图生成算法

张巧芳,李光耀,丁美林,丁志强

(同济大学 CAD 研究中心,上海 201804)

**摘 要:**传统的建模方式非常复杂,而日常生活中拍摄的多数照片本身包含着足够几何信息,于是文中提出了从单幅图像中直接生成具有三维动态效果的系列图片的设想。首先需要用户简化勾画出图像的遮挡关系,然后将图像按遮挡关系进行分层,针对每一个图层,根据提取出的特征边缘来识别出有意义的空间信息,最后从新的视角出发,重定景物间的关系和相对位置,进而生成具有三维浏览效果的图像。实验结果表明,该方法在用户简单参与下能得到良好效果。

**关键词:**三维浏览;图像分层;空间信息;单幅图像

**中图分类号:**TP301.6

**文献标识码:**A

**文章编号:**1673-629X(2010)01-0022-03

## Three - Dimensional Scanning Map Generating Algorithm from Single Image

ZHANG Qiao-fang, LI Guang-yao, DING Mei-lin, DING Zhi-qiang

(CAD Research Center, Tongji University, Shanghai 201804, China)

**Abstract:** Traditional modeling methods are extremely complex, and daily lives of the majority of photos contain sufficient geometric information. Therefore in this paper, directly generate a series of three - dimensional map from a single image. Firstly, do image stratify according to the block relationships. Secondly, identify meaningful spatial information in each layer. Finally, generate a new 3D scanning image according to the new perspective. Experimental results show that this method can get good effect with some simple participation of users.

**Key words:** 3D scanning; image stratification; spatial information; single image

### 0 引 言

随着计算机图形学、数字图像处理、多媒体通信和计算机网络等技术的不断发展,以及现实生活中的各种需求,虚拟现实技术越来越受到各界关注并且得到较大突破。然而,由于网络流量的限制、对真实再现的高要求、对模型细腻程度的较低要求等,如果用现在已经很成熟的三维软件来建模是不符合需求的。现实生活中拍摄的照片,其本身获取的就是非常真实的实体纹理,如果能通过一些方法,从拍摄的照片中直接生成虚拟浏览,效果将是非常好的。

目前基于照片构建虚拟现实思想的主流包括有基于包含足够几何信息的单幅照片和基于同一实物不同角度的照片集这两种方式。

通过相机沿某一特定路线匀速移动所拍摄的系列照片重构三维场景<sup>[1,2]</sup>,该方法的确能达到较好的效

果,但是对拍摄技术要求太高。刘钢等<sup>[3]</sup>提出了以普通相机自由运动拍摄的照片,按照几何空间层次分级重建的方法,充分利用场景中存在的几何约束信息对重建结果进行优化,但是该方法有一定的适用性。

基于单幅照片的建模,明显将避免上述提到的如图像融合等问题,并且具有制作简单,需要采集的源数据也简单,简易设备即可采集数据等优势。目前理论上已经证明,如果单幅照片中包含足够的三维信息,则自动推断照片中场景的三维信息是可行的<sup>[4-7]</sup>。L. Lin等<sup>[8]</sup>提出了一种在二维层面上,应用不同等级的贝叶斯模型从单幅照片中提取并恢复缺省的几何信息重构三维模型的有效方法,但是该方法并没有把照片中实体的空间几何关系考虑在内。A. Saxena等<sup>[5]</sup>提出了一种新的方法,通过推导图像中的三维结构信息,有效地搭建了与图像相对应的三维模型,该方法的适用范围很广,即使图像中无明显的几何结构也能达到三维浏览的效果。而 D. Hoiem等<sup>[9]</sup>则通过统计学的方法,对图像进行分隔、变换、重整等步骤,得到了较好的三维效果图,但对导入的图像等作了特定假设。

许多单照片建模方法都试图通过在照片中寻找精

收稿日期:2009-04-13;修回日期:2009-08-05

作者简介:张巧芳(1985-),女,硕士研究生,研究方向为计算机仿真、数字图像处理;李光耀,研究员,博士生导师,研究方向为计算机仿真、数字图像处理。

确的几何信息重构三维模型来实现虚拟浏览。但二维图像中本身就缺乏三维信息,这也是造成所有照片建模困难重重的本质原因。因文中最终所要达到的是三维浏览的效果,所以将直接跃过传统的建模阶段。

## 1 图像分层

图像分层的主要目的是区分图像中明显的空间关系,并且视同一个图层中的像素具有相近的空间  $Z$  值。文中所使用的空间坐标系采用符合右手螺旋定律的坐标,指向用户的方向规定为  $z$  轴的正方向。

文中选取同类方法中效果较精确的 Levin 等提出的方法<sup>[10,11]</sup>,该方法对树木等边缘模糊的区域有较好的提取效果。目前已知的单照片三维浏览生成方法中,有直接推导图像中的三维结构信息<sup>[5]</sup>;基于贝叶斯方法<sup>[7]</sup>;有将图像分割成特征区域,然后寻找空间信息等<sup>[9]</sup>。而将抠像技术应用到单照片建模中还属新颖方式。实践证明利用遮挡关系能精确地识别出  $Z$  方向上的层次关系,在精确生成三维浏览图时有重要贡献。

### 1.1 抠像原理

假设图像满足模型

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i \quad (1)$$

其中,  $I_i$  代表实际图像;  $\alpha_i$  代表前景层的透明度;  $F_i$  代表前景层;  $B_i$  代表背景层。抠像技术的目的是求出函数(1)的最优解  $B$  和  $\alpha, F$ 。针对灰度图像,首先假设任意一个像素周围的小窗口为  $w$ , 且  $F$  和  $B$  近似不变。将式(1)重写为

$$\alpha_i \approx a I_i + b, \forall i \in w \quad (2)$$

其中  $a = (F - B)^{-1}, b = -B * (F - B)^{-1}$ 。经过变换可得到优化函数

$$J(\alpha, a, b) = \sum_{j \in I} \left( \sum_{i \in w_j} (\alpha_i - a_j I_i - b_j)^2 + \epsilon a_j^2 \right) \quad (3)$$

其中,  $w_j$  是像素  $j$  为中心的一个  $N * N$  小窗口,通常取  $N = 3$ 。对于彩色图像,可将上述理论应用到 RGB 模型中,简单地可取  $H_i = (R + G + B) * 3^{-1}$  代替  $I_i$ 。也可以构建新的模型

$$\alpha_i \approx \sum_c a^c I_i^c + b, \forall i \in w \quad (4)$$

其中,假设  $B$  是两种颜色  $B_a, B_b$  的线性组合,  $F$  同理,于是可得到类似式(3)的优化函数

$$J(\alpha, a, b) = \sum_{j \in I} \left( \sum_{i \in w_j} (\alpha_i - \sum_c a_j^c I_i^c - b_j)^2 + \epsilon \sum_c a_j^c^2 \right) \quad (5)$$

最终得到仅含未知数  $\alpha$  的函数  $J(\alpha) = \alpha^T L \alpha$ ,  $L$  为  $N * N$  矩阵,它的每一项  $(i, j)$  为

$$\sum_{k: (i, j) \in w_k} (\delta_{ij} - |w_k|^{-1} * (1 + (I_i - \mu_k) * (\sum_k +$$

$$\epsilon * |w_k|^{-1} * I_3)^{-1} * (I_j - \mu_k))) \quad (6)$$

其中,  $\sum_k$  是  $3 * 3$  的协方差矩阵;  $\mu_k$  是  $3 * 1$  的颜色均值向量;  $I_3$  是  $3 * 3$  的单位矩阵;  $\epsilon$  与  $|w_k|$  的作用请详见文献[9]和文献[10]。

### 1.2 应用示例

为添加用户限定信息,将  $J(\alpha) = \alpha^T L \alpha$  扩展为

$$\alpha = \alpha^T L \alpha + \lambda (\alpha^T - b_s^T) D_s (\alpha - b_s) \quad (7)$$

其中,常数  $\lambda$  表示用户约束起作用的大小;向量  $b_s$  中,用户标示过的像素对应值为1,否则为0;对角阵  $D_s$ ,用户标示了的像素的相应对角线上元素取值为1,其余为0。式(7)可改写成线性方程组  $(L + \lambda D_s) \alpha = \lambda b_s$ ,最终提取出前景图  $\alpha$  并完成图像的分层。若该图像存在多个层次,可分多次提取,提取过程中将标定各个图层间的层次关系,文中默认的层次关系为距离摄像机由近及远。以下将给出实例图,图1在原图基础上含有用户给定的粗略轮廓线;图2根据抠像原理获取的前景图,相应的部分由其它颜色填充。

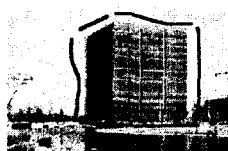


图1 原始图

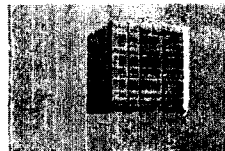


图2 前景图

## 2 空间信息识别

输入的图像经过分层后,将每一层的信息单独保存。假设同一图层中的景物在坐标系中的  $Z$  值近似相等,除设置了各图层间默认的  $Z$  方向差距外,为了增强交互性,用户也可自行设置图层间的距离。空间信息识别就是在每一单独的图层中识别出具有共面性的各个区域。

### 2.1 特征边缘提取

为了突出主要的特征边缘,文中对传统的边缘检测算子进行改进,更接近真实的梯度信息。改进的算子分别为如下四个方向:

$\begin{bmatrix} 1 & 2 & 2 & 2 & 1 \\ & 1 & 2 & 1 & \\ & & * & & \\ & -1 & -2 & -1 & \\ -1 & -2 & -2 & -2 & -1 \end{bmatrix}$ <p>水平</p>	$\begin{bmatrix} 1 & & & & -1 \\ 2 & 1 & & & -2 \\ 2 & 2 & * & & -2 \\ 2 & 1 & & -1 & -2 \\ 1 & & & & -1 \end{bmatrix}$ <p>垂直</p>
$\begin{bmatrix} 2 & 2 & 2 & 1 & \\ 2 & 2 & 1 & & -1 \\ 2 & 1 & * & -1 & -2 \\ 1 & & -1 & -2 & -2 \\ -1 & -2 & -2 & -2 & \end{bmatrix}$ <p>主对角</p>	$\begin{bmatrix} & 1 & 2 & 2 & 2 \\ -1 & & 1 & 2 & 2 \\ -2 & -1 & * & 1 & 2 \\ -2 & -2 & -1 & & 1 \\ -2 & -2 & -2 & -1 & \end{bmatrix}$ <p>副对角</p>

在对提取边缘后的图像进行连接和细化时,提出了边界点确定的优化搜索算法。该方法提高了噪声消除和主要边缘识别的效率。方法的主要实现步骤:

(1) 设  $M_0$  为图层中有效区域内任意一个像素点。记  $w(M_0)$  是以  $M_0$  为中心的  $5 \times 5$  的小窗口,  $T_1 \sim T_4$  依次为上述所提的水平、垂直、主对角和副对角算子, 则点  $M_0$  的梯度值为:

$$\text{dif}(M_0) = \text{abs}(\max\{T_i \cdot w(M_0)\}), i \in [1, 4]$$

由  $\max\{T_i \cdot w(M_0)\}$  还可得出  $M_0$  的梯度方向  $\text{ori}(M_0)$ 。

(2)  $\text{Len}(M_0)$  表示以点  $M_0$  为起始点的边界长度, 初值为 0。边界轨迹点集初值为  $\text{Bou}(M_0) = \{M_0\}$ 。点  $M_1$  为  $\text{ori}(M_0)$  所指向的像素点。

(3) 新的边界点添加到  $\text{Bou}(M_0)$  中, 累计边界长度  $\text{Len}(M_0)$ 。由  $M_{i-2}$  和  $M_{i-1}$  可确定边界的下一搜索方向为  $\text{ori}(M_{i-2}M_{i-1}) = \overrightarrow{M_{i-2} \rightarrow M_{i-1}}$ 。扩充下一边界点  $M_i$  可能方向为  $\text{ori}(M_{i-2}M_{i-1})$  和  $\text{ori}(M_{i-2}M_{i-1}) \pm 45^\circ$ ,  $M_i$  取此三个方向中  $\text{dif}$  值最大的点。若  $\text{dif}(M_i) < \text{Min}$ , 搜索终止, 转至步骤(4); 否则, 记  $M_i$  为新的边界点, 转至步骤(3)。Min 为边界阈值。

(4) 若  $\text{Len}(M_0) > \text{LEN}_{\text{Min}}$ , 边界  $\text{Bou}(M_0)$  有效。 $\text{LEN}_{\text{Min}}$  为边界长度阈值。 $\forall M_i \in \text{Bou}(M_0)$ , 则  $M_i$  被确认为边界点的次数为  $\text{count}(M_i) + 1$ ; 反之, 边界无效。

(5) 选取新起始点, 转至步骤(2)。遍历图层有效区域至完毕。

(6) 合并步骤(5) 所得的各条边界片段。选择多次重复的(真正的边缘是才会被多次检测), 且边界宽度介于  $\alpha \sim \beta$  像素的为最终边界  $\text{End}(e)$ 。

采用如上所述方法, 综合各次训练所得到的最佳效果, 能得到较理想的图片。图像分层后得图 2, 图 3 是在图 2 为原图的基础上训练后得到的图。

## 2.2 空间信息约束

结合文中所提到的边缘提取的方法, 针对不同的输入图像, 结合一些基本常识, 可以提高空间信息识别的效率。如若图像的主题是建筑物等规则图案, 则可以在边缘提取时加入形态学信息, 根据识别出的边缘所构成的封闭或近似封闭区域的亮度信息判定共面性。



图 3 边缘提取

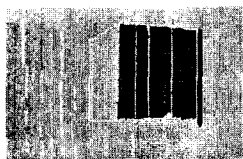


图 4 空间定位

如图 4 所示, 是图 1 的空间信息识别效果图, 根据亮度信息提取出共面区域, 相异面用不同颜色区分。

## 3 视角转换

依据相机平行于  $XOY$ 、 $YOZ$ 、 $ZOX$  面旋转的不同, 分为三类。假设相机平行于  $XOZ$  面旋转, 之前的步骤已经将图像划分成几个特征面后, 提取出各面之间的垂直或近似垂直分界线  $l_i$ , 按从上到下、从左到右顺序排列。 $p_L$  和  $p_R$  是由  $l_i$  分隔而成的两个异面区域。假设拍摄角度平行水平面逆时针旋转  $\theta$  角, 则  $p_L$  和  $p_R$  将绕  $l_i$  顺时针旋转  $\theta$  角度, 如有必要,  $p_R$  还须做平移等操作。如图 5 所示, 实线是为经过视角转换的图像信息, 虚线部分则对应于视线转变后的图像。

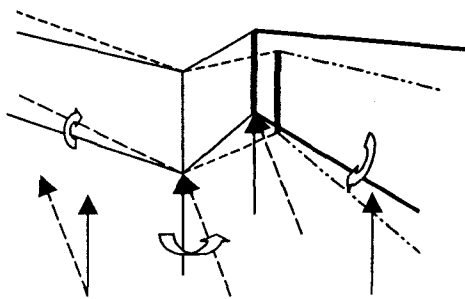


图 5 视角变换

## 4 实验结果及总结

文中所提方法已用 OpenCV 实现。实验结果表明, 文中所提方法已能收到理想的效果。

当然, 文中也存在很多不足之处。首先, 无法实现自动生成, 需要用户参与某些操作, 严格意义上该程序是一个半自动的过程。如预先对图像进行分类, 因为有的操作是可以省略的, 例如一张只含有树、天空、地面的图, 在进行三维浏览图生成时, 可通过抠像原理将整棵树识别为一个单独的面, 则不必再进行空间信息识别这一步骤; 在空间信息识别时, 文中提出了针对规则图案的图像加入几何信息以提高识别效率, 当然, 针对非规则图案, 结合分形技术将会是今后进一步研究的方向; 空间信息识别对于边缘提取效果的依赖性很高; 如何进一步提高共面判定的效率等不足之处。

以下给出实例图, 图 6 左边部分是原图, 右边部分



图 6 原图和  $30^\circ$  视角图

和有效的。在对数据集 1,2,3 的计算中,NSDT 算法的压缩比 CR 比 SDT 算法平均增加了约 40%,压缩误差 CE 平均减小了约 15%。这就表明,NSDT 算法确实可以有效地提高压缩效果。另外,注意到,当记录限  $E$  为 1.0 时 NSDT 算法的压缩比 CR 比 SDT 算法平均增加了约 30%,而当  $E$  为 0.5 时平均增加了约 50%,这说明,NSDT 算法比 SDT 算法更适合压缩要求比较高的情况。在对数据集 4,5,6 的计算中,NSDT 算法的压缩比 CR 比 SDT 算法平均增加了约 30%,压缩误差 CE 几乎相等。这就说明,在有噪声的情况下,NSDT 算法与 SDT 算法相比仍然能显著地提高压缩效果。

表 3 显示了对实际过程数据进行压缩时,在相同的记录限  $E$  的条件下,NSDT 算法的压缩比 CR 比 SDT 算法平均增加了约 20%。表明在处理实际过程数据时,使用 NSDT 算法可以明显地减少对存储空间的需求,减小现场总线网络发生堵塞的可能性,提高控制系统性能。另外,对不同组的实际过程数据,在 NSDT 算法压缩比 CR 相同的条件下,SDT 算法的压缩比 CR 也很接近。这就是说,NSDT 算法对压缩效果的提高是比较稳定的。可以相信,在大多数工业生产条件下,NSDT 算法与 SDT 算法相比压缩效果会有较为显著的提高。

(上接第 24 页)

是具有动态效果的图,假设人的视线与 XOY 面成  $30^\circ$  视角所看到的场景效果。

#### 参考文献:

- [1] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision [M]. Cambridge: Cambridge University Press, 2004.
- [2] Seitz S M, Kim J. Multiperspective Imaging[J]. Computer Graphics and Applications. IEEE, 2003, 23(6): 16-19.
- [3] 刘 钢, 彭群生, 鲍虎军. 基于多幅图像的场景交互建模系统[J]. 计算机辅助设计与图形学学报, 2004, 16(10): 1419-1424.
- [4] Saxena A, Chung S H, Ng A Y. Learning depth from single monocular images[M]//Neural Information Processing Systems. [s.l.]: [s.n.], 2005.
- [5] Saxena A, Sun M, Ng A Y. Learning 3d scene structure from a single still image[C]//International Credit Card Validation Workshop on 3D Representation for Recognition (3dRR-07). [s.l.]: [s.n.], 2007.
- [6] Hoiem D, Efros A, Herbert M. Geometric context from a

## 4 结束语

该研究提出了一种基于 SDT 算法的新的过程数据压缩算法,编写了仿真软件并对合成数据和实际过程数据进行了仿真计算。与 SDT 算法相比,新的算法能在不增加压缩误差的条件下有效地提高压缩比。

#### 参考文献:

- [1] 徐 慧. 实时数据库中数据压缩算法的研究[D]. 杭州: 浙江大学, 2006.
- [2] 赵光煌, 赵 平. 实时数据库的现状与发展趋势[J]. 天津农学院学报, 2002(16): 58-63.
- [3] Bristol E H. Swing Door Trending: Adaptive Trending Recording[R]. Institute Society of American: Research Triangle Park. NC, 1990: 749-754.
- [4] Mah R S H. Process Trending with Piecewise Linear Smoothing[J]. Computer Chemical Engineering, 1995, 19(2): 129-137.
- [5] Hales J C, Sellars H L. Historical data recording for process computers[J]. Chem. Eng. Prog., 1981(11): 38-43.
- [6] 王正洪. 改进的 SDT 算法及其在过程数据压缩中的应用[J]. 煤矿自动化, 2000(6): 9-11.
- [7] 冯晓东, 邵惠鹤. 一种改进的过程数据压缩算法及其性能分析[J]. 测控技术, 2002, 21(1): 136-139.
- [8] 嵇月强. 工业历史数据库的研究[D]. 杭州: 浙江大学, 2007.
- [9] single image[C]//Tenth IEEE International Conference on Computer Vision. [s.l.]: [s.n.], 2005: 654-661.
- [7] Delage E, Lee H, Ng A Y. A dynamic bayesian network model for autonomous 3d reconstruction from a single indoor image[C]//Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. [s.l.]: [s.n.], 2006: 2418-2428.
- [8] Lin L, Zeng K, Wang Y, et al. 3D structure inference by integrating segmentation and reconstruction from a single image[J]. Computer Vision, IET, 2008, 2(1): 15-22.
- [9] Hoiem D, Efros A, Hebert M. Automatic photo pop-up[C]//Proceedings of ACM SIGGRAPH. [s.l.]: [s.n.], 2005: 577-584.
- [10] Levin A, Lischinski D, Weiss Y. A closed form solution to natural image matting[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: [s.n.], 2006: 61-68.
- [11] Levin A, Lischinski D, Weiss Y. A closed-form solution to natural image matting[J]. IEEE transactions on pattern analysis and machine intelligence, 2008, 30(2): 228-242.