

基于 3 维模型的单视图不规则物体定位

李成龙^{1,2)}, 钟 凡^{1,2)}, 秦学英^{1,2)}

¹⁾(山东大学计算机学院 济南 250101)

²⁾(山东省软件工程重点实验室 济南 250101)
(lcl19880926@163.com)

摘 要: 单幅图像中物体的定位需要估计物体在 3 维空间中的位置和姿态, 在静态场景中等价于摄像机定标, 具有广泛的应用价值. 针对具有 3 维模型, 但缺失纹理信息的不规则物体的定位问题, 提出一种基于轮廓匹配的定位方法. 首先使用基于图像分割的方法提取输入图像中物体轮廓线, 然后将图像轮廓线与给定位置和姿态参数下渲染 3 维模型的轮廓线进行匹配, 匹配误差可以表示为位置与姿态参数的函数. 由于该函数不能解析表达与求解, 需要通过离散采样计算导数及目标函数值. 位置与姿态参数的最优值可以通过 LM(Levenberg-Marquardt)方法进行求解. 实验结果表明, 该方法可以快速收敛, 并具有很高的精确性和鲁棒性.

关键词: 姿态估计; 轮廓线; 匹配函数; Levenberg-Marquardt 方法
中图分类号: TP391

Model-Based Irregular Object Localization from a Single View Image

Li Chenglong^{1,2)}, Zhong Fan^{1,2)}, Qin Xueying^{1,2)}

¹⁾(School of Computer Science and Technology, Shandong University, Jinan 250101)

²⁾(Shandong Provincial Key Laboratory of Software Engineering, Jinan 250101)

Abstract: Object localization aims to estimate the position and orientation of the object in 3D space. In static scenes, it is equivalent to camera tracking, which has a wide range of applications. In this paper, we propose a contour-based approach to localize texture-less irregular objects with known 3D models. The target object is first extracted from the input image via image segmentation, the contour is then matched with rendered 3D model with given position and orientation parameters. The matching error can be expressed as a function of the position and orientation parameters. Since the function cannot be analytically expressed and solved, we calculate the matching score and derivative by discrete sampling, the optimal score and derivative parameters can be solved efficiently via LM (Levenberg-Marquardt) solver. Experimental results show that this method can converge quickly, and can achieve very high accuracy and robustness.

Key words: pose estimation; contour; energy function; Levenberg-Marquardt

3 维物体的空间定位一直是计算机视觉中的一个重要问题. 物体的定位是要确定其与相机的相对空间关系, 包括位置与姿态的参数估计. 当物体静止时, 物体的定位等价于相机外部参数的标

定, 从而也可以用于摄像机定标. 因此, 其在增强现实、机器人导航、运动跟踪以及医学上的外科手术等众多领域有重要的应用.

多年来, 研究人员对物体的定位问题进行了

收稿日期: 2013-11-17; 修回日期: 2014-05-28. 基金项目: 国家自然科学基金广东省联合基金(U1035004); 国家自然科学基金(61173070, 61202149). 李成龙(1988—), 男, 博士研究生, 主要研究方向为运动捕捉; 钟 凡(1982—), 男, 博士, 讲师, 主要研究方向为图像视频处理、计算机视觉; 秦学英(1966—), 女, 教授, 博士生导师, 主要研究方向为增强现实、计算机视觉、计算机图形学.

深入的研究. 在已知物体3维模型的前提下, 其主要困难在于目标模型3维点与其成像点的对应问题, 特别地, 当目标的3维模型缺失纹理信息时, 点的匹配几乎不可能. 目前物体定位问题的方法主要分为3类:第1类是已知若干模型与其成像的对应点对^[1-7];第2类是已知若干模型与其成像的对应线段^[8-12];第3类是对应性未知算法^[13-18].

目前, 已有的研究方法主要集中在第1类方法上. Huang 和 Netravali^[1]对现有的方法进行了综述. 若已知3~5对不共面的对应点, 可以通过求解多项式方程组得到位置和姿态参数的封闭解. 针对3点透视投影问题(p3p)和4点透视投影问题(p4p), Fishier 等^[2]提出了使用几何透视法求解多项式方程组. Huttenlocker 等^[3]在弱透视投影的情况下利用3个不共线模型点求解多项式方程组. Horaud 等^[4]提出了一种利用4个不共面的对应点的方法, 该方法通过求解四次多项式方程得到姿态参数. 若已知6对或更多对应点对, 可以采用线性或非线性逼近的方法求解参数. Quan 等^[5]和 Faugeras^[6]都采用线性方法来求解姿态问题. 在确定物体姿态时, 线性方法可以避免迭代. 然而, Quan 等^[5]的方法没有考虑点对之间的特殊关系约束, 算法的准确性和稳定性需要提高. Faugeras^[6]的工作只能应用于采用6个或更多对应点的情况, 且模型点不能共面. 郝颖明等^[7]对点匹配方法的鲁棒性进行了分析.

第1类方法需要已知对应点的匹配关系, 但是, 当3维模型的纹理缺失时, 很难获得匹配点, 导致算法的失效.

第2类方法主要利用模型线与其对应投影线产生的约束关系, 比第1类方法理论要复杂很多. 最少需要3条线段来求解姿态估计问题, 3条线段可以建立3个非线性方程, 之后姿态估计问题可以转化为求解非线性方程组的解. 目前, 第2类方法主要分为2类:分析方法和数值方法. 为了简化分析方法, Dhorne 等^[8]和 Chen^[9]建立了一个特定的模型坐标系和一个特定的视点坐标系, 利用3条线段建立一个8维的多项式方程来确定姿态参数的封闭解. 其中 Chen^[9]所利用的线为3条特殊的线或者特意指定的线(包括平行线, 3条中有2条平行线、共面线、垂直线等). 刘昶等^[10]利用了3条相互垂直的直线. 分析方法的缺点在于参数求解时存在多组解, 并且对噪声比较敏感. 对于数值方法, Yuan^[11]对于单摄像机的情况采用牛顿迭代法估计位置和

姿态参数. Christy 等^[12]同样使用迭代方法求解参数, 他们所采用的方法需要初始化一个与真实值较接近的初始值. 与分析方法相比, 数值方法通常具有较高的精度, 然而数值方法需要较好的初始化并且效率较低. 第2类方法要求3维模型是比较规则的物体, 从而容易提取特征. 对于不规则物体, 很难找到匹配的模型线与投影线, 此类方法容易失效.

在第3类方法中, 问题的求解难度更大, 目前主要的研究方法有:基于训练集的方法^[13]、基于假设检验的方法^[14]、能量函数优化求解的方法^[15-18]. 基于训练集的方法^[13], 首先利用计算机学习大量不同姿态参数的图像特征并存储于数据库中, 之后通过测试图像中提取的目标特征到数据库中搜索相似的对应特征, 进而得到对应的姿态参数. 基于假设检验的方法^[14]先利用 RANSAC(random sample consensus)方法^[2], 随机假设若干对应点, 然后根据这些对应点计算姿态参数, 根据求解参数得到模型投影图像, 若投影图像与原始图像匹配, 则求解结束, 否则调整假设的对应点, 重复之前的步骤直到求得正确姿态参数. 能量函数优化求解的方法, 首先定义一个包含姿态参数与对应性的全局能量函数, 之后再通过优化算法求解姿态参数. 如结合 POSIT(pose from orthography and scaling with iterations)算法^[15]和 SoftAssign 算法^[16]来进行目标函数的优化求解. 这类方法在训练和检验时也需要将虚拟物体的渲染结果与原图像进行比对, 从而在光照和纹理产生差异的情况下, 无法准确地判断是否匹配, 影响算法的精度. Iwashita 等^[17]和 Dambreville 等^[18]提出在目标轮廓上定义能量优化函数, 然后通过数值优化的方法最小化能量函数进行位置和姿态参数的估计. 但是他们的方法都存在收敛速度较慢的问题. 主要原因在于:此类方法的能量函数高度非线性导致数值优化过程易陷入错误的局部极小值;因为没有显著的对应性关系, 参数的估计完全依赖于能量函数的数值优化, 导致算法复杂度过高、效率低.

综上所述, 现有的大部分方法都需要解决外观特征点的对应性问题, 且求解计算量较大. 针对不规则物体的定位问题, 本文提出了一种新的基于3维模型的单视图姿态估计方法, 该方法利用了物体在图像上的轮廓线匹配来估计姿态. 该匹配过程自动进行, 并且具有很高的求解精度与算法效率.

1 算法描述

算法的目标是:已知不规则物体的 3 维模型及相机内部参数,从该物体的单幅图像中估计其位置与姿态参数。

本文算法根据 3 维模型渲染图像与输入图像中物体轮廓的匹配程度,建立一个包含位置与姿态参数的匹配度量,并以此作为目标函数,将物体的定位问题转化为目标函数的最优化求解问题。算法步骤如下所示:

Step 1. 任意输入实物照片,使用图像分割方法提取实物照片的物体轮廓;

Step 2. 任意给定一组初始位置和姿态参数;

Step 3. 针对 1 组位置和姿态参数,对模型进行坐标变换;

Step 4. 将变换后的模型投影到 2 维平面,得到对应 CG 图像;

Step 5. 提取 CG 图像中物体的轮廓;

Step 6. 计算由 Step 1 和 Step 5 所得轮廓重合区域积;

Step 7. 计算轮廓重合区域占 2 轮廓的比例,将其作为相似性度量函数;

Step 8. 判断相似性度量函数值是否等于 1,若等于 1,则算法结束,输出参数,否则,继续执行 Step 9;

Step 9. 针对当前位置和姿态参数,选取充分小增量,在参数空间进行采样;

Step 10. 对每个采样所得参数,执行 Step 3 到 Step 7,得到不同采样的相似性度量函数值;

Step 11. 比较相似性度量函数值,选取其中最大值,并判断其是否等于 1,若等于 1,则算法结束,输出对应参数值,否则,继续执行 Step 12;

Step 12. 将 Step 11 中最大值所对应的参数作为下次计算的初始参数值,转到 Step 3。

2 相机投影模型

采用针孔相机模型,空间中任意一个 3 维点 $X = [x, y, z, 1]^T$,在成像平面上的投影点为 $u = [u, v, 1]^T$,则有

$$u = K[R|t]X,$$

其中, K 为相机内参数矩阵, $[R|t]$ 为相机坐标系与物体坐标系的坐标变换矩阵,即旋转矩阵 R 与平移向量 t 。内部参数矩阵

$$K = \begin{bmatrix} f & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix},$$

其中 (c_x, c_y) 是基准点, f 为焦距。本文算法中, (c_x, c_y) 选为图像的中心点。 $[R|t]$ 和 f 为需要确定的参数。本文算法中,相机已经校准过,即已知 f 。

对于平移向量 t ,可以用 (t_1, t_2, t_3) 来表示。旋转矩阵 R 常用的表示方法有欧拉角表示法和四元数表示法。本文所采用的方法为欧拉角表示法,则所要求解的参数 Θ 可以用 6 个参数来表示,即

$$\Theta = (\alpha, \theta, \omega, t_1, t_2, t_3),$$

其中 α, θ, ω 分别为绕 x, y, z 3 个轴所旋转的角度。

3 轮廓的匹配

算法的基本思想,是采用搜索方式,通过轮廓匹配的方法,确定 3 维物体的位置与姿态。

3.1 轮廓提取

图像的视觉特征通常包含有颜色特征、纹理特征和形状特征等。颜色特征容易受背景、光照等影响,当图像的纹理信息相差不大时,纹理特征很难准确反映,纹理特征还存在计算量大的缺点。相比颜色特征与纹理特征,形状特征刻画了物体的几何特征,具有良好的稳定性,不易受光照、背景颜色和纹理信息的影响。本文的算法采用目标轮廓线作为形状特征,通过轮廓线的匹配来判断 2 幅图像的匹配。

对于实物照片,由于受光照和背景的影响,很难用传统的边缘检测方法进行轮廓提取,本文采用 Zhong 等^[19]提出的基于图像分割的方法提取轮廓,获得图像上目标物体的区域并提取轮廓线。

对于 3 维模型,通过设定不同相机的参数,利用 OpenGL 渲染其 CG 图像,将 CG 图像阈值化,之后通过链码跟踪存储轮廓信息,最终得到具有单像素边缘的轮廓图像。

3.2 轮廓匹配

本文的目标是要寻找最佳的相机参数,使得 3 维模型在投影平面上的轮廓与图像上目标物体的轮廓重合。于是需要定义轮廓的相似性函数。

通过轮廓提取分别得到输入图像与 CG 图像所对应的轮廓图像之后,本文根据轮廓线包围区域的重合程度来判断 3 维模型的匹配程度。如图 1 所示。

定义 $M(\Theta)$ 表示在某个姿态参数下的 3 维模型,则 3 维模型在 2 维图像平面上的投影 CG 图像为 $A(M(\Theta))$, CG 图像轮廓提取所得的轮廓图像为

$C_{A(M(\theta))}$. 对于输入图像 D , 轮廓提取所得的轮廓图像为 C_D . 本文通过 $C_{A(M(\theta))}$ 中轮廓包围区域与 C_D 中轮廓包围区域的重合程度定义相似度函数

$$S(\theta) = \frac{A_0^2(\theta)}{A(C_{A(M(\theta))})A(C_D)},$$

其中 $A_0(\theta)$ 为 2 条轮廓线 $C_{A(M(\theta))}$ 与 C_D 包围区域重合部分面积, $A(C_{A(M(\theta))})$ 与 $A(C_D)$ 分别为对应轮廓所包围区域面积. $0 \leq S \leq 1$, S 越大, 2 个区域越相似, 当且仅当 2 区域重合时, $S = 1$.

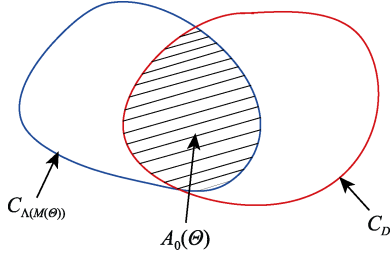


图 1 轮廓的匹配

4 位置与姿态估计

在求解位置与姿态参数 θ 时, 由于待定参数是 6 维变量, 如果采用全局搜索法求解, 其时间复杂度太高, 且搜索步长很难给出合适的设定.

针对全局搜索法效率低的问题, 可以采用改进的迭代搜索法, 设 μ_i 为第 i 次迭代的步长, 给定 $\mu_{i+1} = \mu_i / n$, 每次搜索得到的局部最优解作为下一次迭代的初始值, 并将搜索空间缩小. 迭代搜索法相对全局搜索法在效率上得到提高, 但是由于参数空间是 6 维的, 搜索的时间复杂度还是太高. 并且迭代搜索法还容易陷入局部最优解, 不能确保得到全局最优解.

基于轮廓相似度的定义, 本文将位置和姿态估计问题归结为函数的最优化问题. 即寻找最优的参数 $\hat{\theta}$, 使得相似度函数最大, 即

$$\hat{\theta} = \arg \max_{\theta} S(\theta).$$

由于相似度函数依赖于参数 θ 的选取, 及其对应的 3 维模型投影轮廓线, 因此函数 $S(\theta)$ 不是解析函数. 但对任意给定的参数 θ , 都可以渲染出其对应的 CG 图像, 从而计算相似度函数.

为了在保证算法精度的基础上又保证算法效率, 本文采用 LM(Levenberg-Marquardt)算法思想来求解^[20]. LM 算法是高斯-牛顿法的改进形式, 既

有高斯-牛顿法的局部特性, 又具有梯度法的全局特性. 由于利用了近似的二阶导数信息, 算法比梯度法快得多, 适合解决非线性问题.

给定参数 θ 和能量函数 $S(\theta)$, 需要找到 $\Delta\theta$, 使得 $S(\theta + \Delta\theta)$ 逐步增大, 直到 $S(\theta')$ 最大, 则 θ' 为所要求解的最优参数. LM 算法中 $\Delta\theta$ 为

$$\Delta\theta = -(\mathbf{J}^T \mathbf{J} + \lambda \mathbf{J})^{-1} \mathbf{J}^T (S(\theta + \Delta\theta) - S(\theta)),$$

其中 \mathbf{J} 为雅可比矩阵, 即 $\frac{\partial S(\theta + \Delta\theta)}{\partial \theta}$. 由于 S 为单值函数, 雅可比矩阵退化为向量, 即在各参数方向的方向导数为

$$\mathbf{J} = \left[\frac{\partial S}{\partial \alpha}, \frac{\partial S}{\partial \theta}, \frac{\partial S}{\partial \omega}, \dots \right]^T.$$

由于相似度函数不是解析函数, 不可微, 其雅可比矩阵 \mathbf{J} 无法直接通过求偏导得到. 因此在求解过程中, 本文采用采样的方法逼近雅可比矩阵 \mathbf{J} 的值. 选取充分小增量

$$\begin{cases} \nabla_{\alpha} = [d_{\alpha}, 0, 0, 0, 0, 0]^T, \\ \nabla_{\theta} = [0, d_{\theta}, 0, 0, 0, 0]^T, \\ \nabla_{\omega} = [0, 0, d_{\omega}, 0, 0, 0]^T, \\ \nabla_{t_1} = [0, 0, 0, d_{t_1}, 0, 0]^T, \\ \nabla_{t_2} = [0, 0, 0, 0, d_{t_2}, 0]^T, \\ \nabla_{t_3} = [0, 0, 0, 0, 0, d_{t_3}]^T, \end{cases}$$

采样计算 $S(\theta)$ 与 $S(\theta + \nabla_i)$ 的值, $i = 1, 2, \dots, 6$, 其方向导数可以用下式逼近:

$$\begin{cases} \frac{\partial S(\theta)}{\partial \alpha} \approx \frac{S(\theta) - S(\theta + \nabla_{\alpha})}{d_{\alpha}}, \\ \frac{\partial S(\theta)}{\partial \theta} \approx \frac{S(\theta) - S(\theta + \nabla_{\theta})}{d_{\theta}}, \\ \frac{\partial S(\theta)}{\partial \omega} \approx \frac{S(\theta) - S(\theta + \nabla_{\omega})}{d_{\omega}}, \\ \frac{\partial S(\theta)}{\partial t_1} \approx \frac{S(\theta) - S(\theta + \nabla_{t_1})}{d_{t_1}}, \\ \frac{\partial S(\theta)}{\partial t_2} \approx \frac{S(\theta) - S(\theta + \nabla_{t_2})}{d_{t_2}}, \\ \frac{\partial S(\theta)}{\partial t_3} \approx \frac{S(\theta) - S(\theta + \nabla_{t_3})}{d_{t_3}}. \end{cases}$$

对于参数 λ 的选择, 如果某次迭代的 $\Delta\theta$ 使得

$S(\theta)$ 增大, 则本次迭代的 $\Delta\theta$ 可接受, λ 的值要减小, 如果 $S(\theta)$ 减小, 则要增大 λ 来重新计算 $\Delta\theta$, 其过程与标准 LM 算法相似.

5 实验结果

为了验证本文算法, 采用激光扫描及多视角 3 维重建方法获得的不规则 3 维模型, 进行了 3 组实验. 实验中位置参数单位为 cm, 姿态参数单位为“°”. 实验最后给出了本文方法与文献[17]所用方法精度的比较. 运行环境为 Inter(R) Core(TM)2 Duo 2.33 GHz 处理器和 2 GB RAM, 以及 32 位 Windows 7 操作系统. 开发环境为 Visual Studio 2010, 基于 OpenGL 和 OpenCV 开发库.

首先, 本文采用激光扫描对得到的肝脏 3 维模型和 1 组输入图像进行实验. 该模型包含 4 000 个三角面片, 直径大约为 20 cm. 图 2a 为相机拍摄的肝脏图像, 其中每个图像的分辨率均为 2 500×3 000, 图 2b 为激光扫描的 3 维肝脏模型. 为了验证算法的精度, 本文按 1 组设定的姿态参数生成相应的 CG 图像作为输入. 图 3 给出了 6 幅任意的 CG 图像, 其中每个图像的分辨率均为 400×400. 表 1 给出了求解参数值和相应的误差值, 平移误差基本上控制在 0.1 cm 内, 角度误差基本上控制在 0.5°以内, 通过分析实验结果, 本文算法可以保证较好的精度.

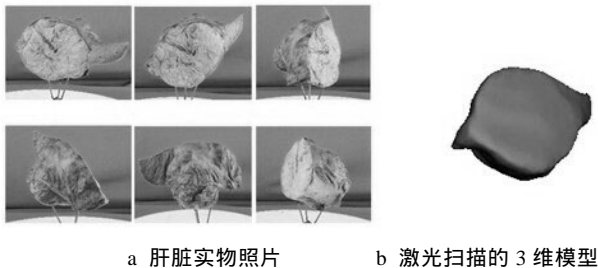


图 2 肝脏数据

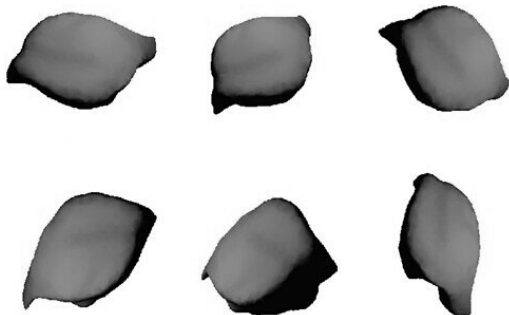


图 3 CG 图像

表 1 肝脏数据参数求解与误差

图像	真实参数值					
	t_1/cm	t_2/cm	t_3/cm	$\alpha/(\text{°})$	$\theta/(\text{°})$	$\omega/(\text{°})$
1	-20.000	50.000	100.000	-20.000	20.000	0.000
2	-20.050	52.000	100.010	-20.050	22.000	1.010
3	-20.100	54.000	100.050	-20.100	24.000	2.050
4	-20.150	56.000	100.030	-20.150	26.000	3.030
5	-20.200	58.000	100.120	-20.200	28.000	4.120
6	-20.250	60.000	100.080	-20.250	30.000	5.080
图像	求解参数值					
	t_1/cm	t_2/cm	t_3/cm	$\alpha/(\text{°})$	$\theta/(\text{°})$	$\omega/(\text{°})$
1	-19.986	50.037	99.883	-19.786	19.737	0.583
2	-20.034	51.738	100.103	-20.134	21.838	1.113
3	-20.287	54.054	99.978	-20.387	24.154	1.778
4	-20.098	55.978	99.978	-20.198	25.908	3.362
5	-20.234	57.786	100.232	-20.034	27.686	4.132
6	-20.357	59.877	100.006	-19.957	30.277	4.786
图像	误差值					
	$\Delta t_1/\text{cm}$	$\Delta t_2/\text{cm}$	$\Delta t_3/\text{cm}$	$\Delta \alpha/(\text{°})$	$\Delta \theta/(\text{°})$	$\Delta \omega/(\text{°})$
1	0.014	0.037	-0.117	0.214	-0.263	0.583
2	0.016	-0.262	0.093	-0.084	-0.162	0.103
3	-0.187	0.054	-0.072	-0.287	0.154	-0.272
4	0.052	-0.022	-0.468	-0.048	-0.092	0.362
5	-0.034	-0.214	0.112	0.166	-0.314	0.012
6	-0.107	-0.123	-0.074	0.293	0.277	-0.294

通过第 1 组实验验证了本文算法的精度, 下面要确定实物照片中物体的姿态参数. 首先将图 2a 中的图像进行分割处理, 结果如图 4 所示. 对于姿态参数的求解, 搜索法在理论上可以得到最优解. 图 5a~c 给出了搜索法的参数变化情况, 每个图中其他 4 个参数均已取得最优解. 通过观察可以发现这些数据有明显的峰值, 即有最优解存在. 表 2 给出了求解参数值和采用不同方法的时间复杂度, 搜索法效率最低, 平均时间为 10^5 数量级, 迭代搜

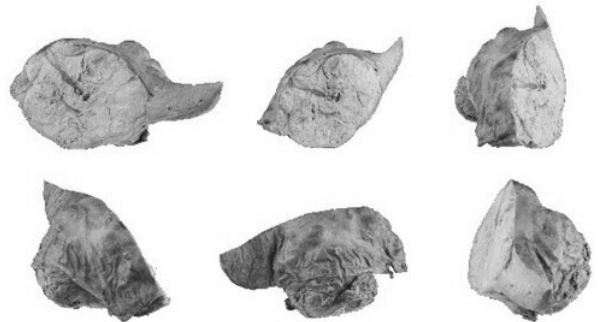


图 4 输入图像对应的分割图像

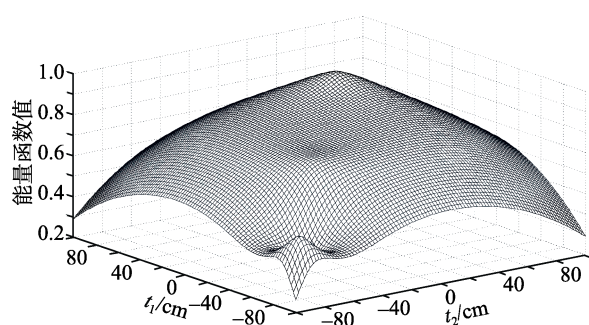
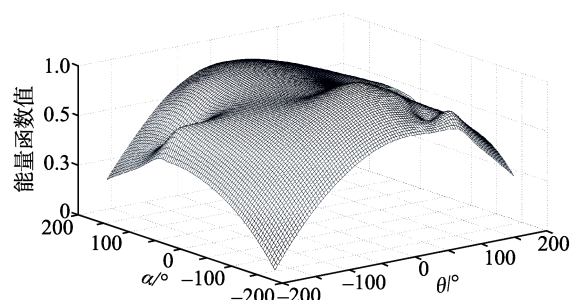
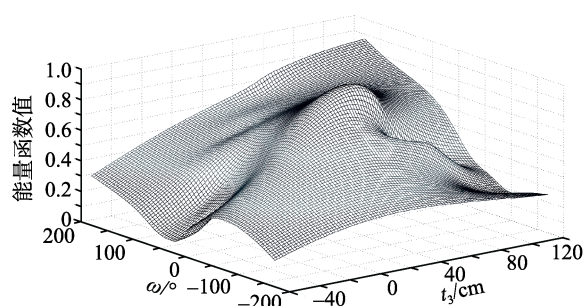
a 参数 $t_3, \alpha, \theta, \omega$ 取得最优时参数 t_1 与 t_2 对应能量函数值b 参数 t_1, t_2, t_3, ω 取得最优时参数 α 与 θ 对应能量函数值c 参数 t_1, t_2, α, θ 取得最优时参数 t_3 与 ω 对应能量函数值

图 5 参数变化曲面

表 2 求解参数值与迭代次数

图像	求解参数值					
	t_1/cm	t_2/cm	t_3/cm	$\alpha/(\text{°})$	$\theta/(\text{°})$	$\omega/(\text{°})$
1	-24.370	53.103	10.220	-33.653	40.337	-15.318
2	-28.753	54.441	15.235	-25.186	12.144	37.614
3	-43.107	46.235	22.351	22.034	10.220	7.082
4	-33.653	-30.234	-40.317	-70.298	-30.413	35.951
5	-50.223	-44.938	42.144	25.657	37.186	40.822
6	-57.852	65.313	30.148	-20.879	-20.152	19.186

图像	迭代次数		
	全局搜索	迭代搜索	本文算法
1	10^5	1 567	35
2	10^5	3 541	40
3	10^5	4 876	20
4	10^5	6 331	29
5	10^5	2 178	44
6	10^5	2 607	36

索法也需要几千次迭代, 而本文算法在效率上有极大提高, 只需要几十次迭代, 平均时间为 45 ms. 对于任一输入图像, 表 3 给出了使用本文算法求解过程中不同迭代次数的参数差值与匹配残差, 随着迭代次数的增多, 匹配残差逐步减小, 直到收敛. 图 6 给出了某次求解 LM 算法的迭代过程, 图 6 中其他 4 个参数已取得最优解, 其中左下角点为任意选择的初始点, 右上角点为最优解, 其他点为每次采样计算的中间点. 图 7 给出了某次求解过程中不同采样点的虚拟轮廓与图像轮廓的匹配情况, 随着迭代次数的增加轮廓匹配程度变高, 直到 2 轮廓完全匹配, 其中形状不变曲线为图像轮廓, 形状变化的曲线为虚拟轮廓.

前面 2 组实验所用 3 维模型通过 3 维扫描仪得到, 模型精度比较高. 下面本文使用通过 3 维重建所得模型来进行实验. 图 8 给出了多视点几何重

表 3 不同迭代次数的参数差值与匹配残差

迭代次数	Δt_1	Δt_2	Δt_3	$\Delta \alpha$	$\Delta \theta$	$\Delta \omega$	匹配残差
1	71.35	102.07	52.98	59.86	94.97	84.72	0.637
3	64.55	85.97	36.87	42.31	63.52	70.62	0.538
8	44.15	46.17	25.36	33.86	40.07	57.30	0.334
12	30.15	32.37	13.62	25.14	21.36	31.22	0.214
16	9.55	9.17	7.54	10.03	9.85	11.35	0.089
20	0.03	0.05	0.02	0.07	0.12	0.23	0.001

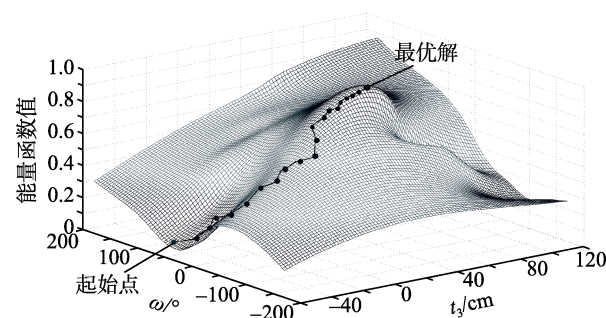


图 6 LM 算法的迭代过程

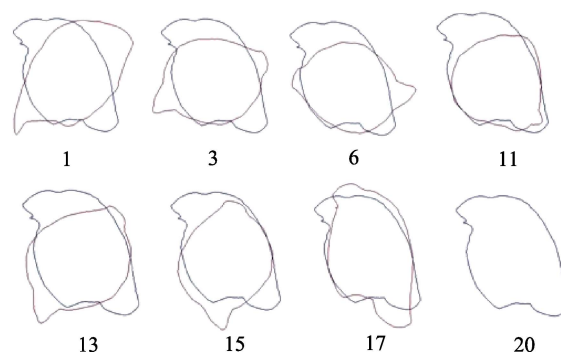
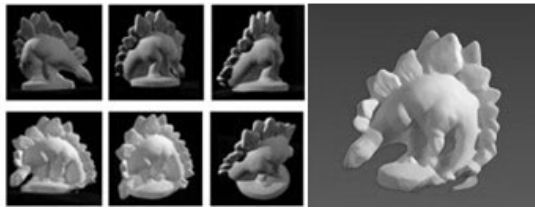


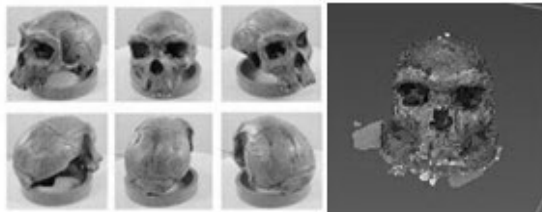
图 7 算法迭代过程中的虚拟轮廓与图像轮廓



a 恐龙雕塑实物照片 b 重建恐龙雕塑模型

图 8 恐龙雕塑数据

建所用的恐龙雕塑图像和对应的重建模型, 图像的分辨率为 640×480 , 恐龙雕塑模型包含 3 536 个三角面片, 直径大约为 15 cm. 图 9 给出了多视点几何重建所用的人体头骨图像和对应的重建模型, 图像的分辨率为 $1\,900 \times 1\,700$, 人体头骨模型包含 5 329 个三角面片, 直径大约为 23 cm. 通过图片可以发现 3 维重建的模型有较大误差, 尤其是图 9 中人体头骨模型. 重建误差会对算法执行过程中的轮廓匹配造成影响, 导致实验结果误差变大. 表 4 和表 5 给出了使用 3 维重建模型求解结果与误差, 其误差相对于之前实验的求解误差要大. 图 9 中的人体头骨模型, 在进行正面的姿态估计时, 其左侧部分区域会导致无法匹配到最优姿态, 求解参数误差较大. 对于 3 维重建的模型, 其准确程度对算法有直接的影响.



a 人体头骨实物照片 b 重建人体头骨模型

图 9 人体头骨数据

文献[21]也使用轮廓线信息来估计单幅图像中物体的位置和姿态参数. 与本文方法不同之处是文献[21]没有直接使用物体 3 维模型, 而是通过训练得到不同视点的图像模板, 使用图像模板检测目标并估计其位置和姿态参数. 进行参数求解时, 文献[21]使用了一种中间级特征——BOB(边界包), 通过输入图像与已训练的图像模板之间 BOB 的匹配求解参数. 为了验证其精度, 文献[21]在训练时, 姿态参数空间每 10° 进行采样, 对每一个物体生成 1 944 幅训练图像. 为了验证算法精度, 使用了合成数据, 通过求解参数值与真实参数值的比较, 给出了平均误差. 表 6 给出了文献[21]与本文算法的

表 4 恐龙数据参数求解与误差

图像	真实参数值					
	t_1/cm	t_2/cm	t_3/cm	$\alpha/(\circ)$	$\theta/(\circ)$	$\omega/(\circ)$
1	-22.000	15.000	50.000	-5.000	-20.000	0.000
2	-22.050	17.000	50.010	-5.050	-2.000	1.210
3	-22.100	18.000	50.050	-5.100	24.000	2.350
4	-22.150	20.000	50.030	-5.150	65.000	3.060
5	-22.200	21.000	50.120	-5.200	87.000	4.150
6	-22.250	23.000	50.080	-5.250	110.000	5.100

图像	求解参数值					
	t_1/cm	t_2/cm	t_3/cm	$\alpha/(\circ)$	$\theta/(\circ)$	$\omega/(\circ)$
1	-21.886	15.137	49.873	-5.186	-19.737	0.473
2	-22.134	16.738	50.203	-5.134	-1.838	1.313
3	-22.297	18.054	49.878	-5.307	24.154	2.478
4	-22.048	20.378	50.238	-5.298	64.908	3.192
5	-22.334	21.286	50.232	-5.024	86.686	4.172
6	-22.357	22.867	50.006	-4.958	110.877	4.986

图像	误差值					
	$\Delta t_1/\text{cm}$	$\Delta t_2/\text{cm}$	$\Delta t_3/\text{cm}$	$\Delta \alpha/(\circ)$	$\Delta \theta/(\circ)$	$\Delta \omega/(\circ)$
1	0.114	0.137	-0.127	-0.186	0.263	0.473
2	-0.084	-0.262	0.193	-0.084	0.162	0.103
3	-0.197	0.054	-0.172	-0.207	0.154	0.128
4	0.102	0.378	0.208	-0.148	-0.092	0.132
5	-0.134	0.286	0.112	0.176	-0.314	0.022
6	-0.107	-0.133	-0.074	0.292	0.877	-0.114

表 5 人体头骨数据参数求解与误差

图像	真实参数值					
	t_1/cm	t_2/cm	t_3/cm	$\alpha/(\circ)$	$\theta/(\circ)$	$\omega/(\circ)$
1	-12.000	10.000	35.000	-3.000	-15.000	-1.000
2	-12.150	12.000	35.100	-3.150	-2.000	0.100
3	-12.100	13.000	35.150	-3.100	25.000	2.050
4	-12.150	14.000	35.030	-3.250	56.000	3.050
5	-12.200	15.000	35.120	-3.200	88.000	4.020
6	-12.250	16.000	35.400	-3.250	100.000	5.000

图像	求解参数值					
	t_1/cm	t_2/cm	t_3/cm	$\alpha/(\circ)$	$\theta/(\circ)$	$\omega/(\circ)$
1	-11.886	10.037	34.883	-2.786	-14.737	-0.583
2	-12.034	11.738	35.203	-3.334	-2.138	0.213
3	-12.287	13.154	35.378	-3.387	24.755	1.678
4	-12.098	13.878	34.678	-3.198	55.908	3.362
5	-12.334	15.386	35.432	-3.034	87.676	4.132
6	-12.367	15.857	35.206	-3.457	100.372	4.786

图像	误差值					
	$\Delta t_1/\text{cm}$	$\Delta t_2/\text{cm}$	$\Delta t_3/\text{cm}$	$\Delta \alpha/(\circ)$	$\Delta \theta/(\circ)$	$\Delta \omega/(\circ)$
1	0.114	0.037	-0.117	0.214	0.263	0.417
2	0.116	-0.262	0.103	-0.184	-0.138	0.113
3	-0.187	0.154	0.228	-0.287	-0.245	-0.372
4	0.052	-0.122	-0.352	0.052	-0.092	0.312
5	-0.134	0.386	0.312	0.166	-0.324	0.112
6	-0.117	-0.143	-0.194	-0.207	0.372	-0.214

表 6 精度比较(平均误差)

方法	平均误差
本文	0.24
文献[21]	3.10

平均精度比较, 本文算法精度要比文献[21]方法高很多. 效率上与文献[21]相比, 本文算法不需要训练且可以快速收敛.

6 结 论

本文提出了一种新的基于轮廓的单幅图像姿态估计方法. 该方法避免了特征点的对应性问题, 精确地估计单幅图像物体位置和姿态. 算法采用原始图像与 3 维模型渲染轮廓对齐的思想, 寻找最佳的位置与姿态估计参数, 并用扩展的 LM 方法来加速求解. 算法对 3 维模型非规则、无纹理的情况下, 比较其他算法有很大优势, 可以获得准确的结果. 算法精度与 3 维模型的误差有关, 当 3 维模型是多视角视图重建结果时, 误差会变大, 但仍然可以容忍. 实验结果显示了良好的精度和速度.

参考文献(References):

- [1] Huang T S, Netravali A N. Motion and structure from feature correspondence: a review [J]. *Proceedings of the IEEE*, 1994, 82(2): 252-268
- [2] Fishier M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. *Communications of the ACM*, 1981, 24(6): 381-395
- [3] Huttenlocker D P, Ullman S. Recognizing solid objects by alignment with an image [J]. *International Journal of Computer Vision*, 1990, 5(2): 195-212
- [4] Horaud R, Conio B, Le Boulleux O, *et al.* An analytic solution for the perspective 4-point problem [C] // *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 1989: 500-507
- [5] Quan L, Lan Z D. Linear N-point camera pose determination [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(8): 774-780
- [6] Faugeras O. Three-dimensional computer vision: a geometric viewpoint [M]. Cambridge: MIT Press, 1993
- [7] Hao Yingming, Zhu Feng, Ou Jinjun, *et al.* Robustness analysis of pose estimation based on points [J]. *Computer Applications*, 2008, 28(7): 1713-1716+1720 (in Chinese)
(郝颖明, 朱 枫, 欧锦军, 等. 基于点特征的位姿测量系统鲁棒性分析[J]. *计算机应用*, 2008, 28(7): 1713-1716+1720)
- [8] Dhome M, Richetin M, Lapresté JT, *et al.* Determination the attitude of 3D objects from a single perspective view [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(12): 1265-1278
- [9] Chen H H. Pose determination from line-to-plane correspondences: existence condition and closed-form solutions [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 13(6): 530-541
- [10] Liu Chang, Zhu Feng, Ou Jinjun. Monocular pose determination from three perpendicular lines [J]. *Pattern Recognition and Artificial Intelligence*, 2012, 25(5): 737-744 (in Chinese)
(刘 昶, 朱 枫, 欧锦军. 基于三条相互垂直直线的单目位姿估计[J]. *模式识别与人工智能*, 2012, 25(5): 737-744)
- [11] Yuan J S C. A general photogrammetric method for determining object position and orientation [J]. *IEEE Transactions on Robotics and Automation*, 1989, 5(2): 129-142
- [12] Christy S, Horaud R. Iterative pose computation from line correspondences [J]. *Computer Vision and Image Understanding*, 1999, 73(1): 137-144
- [13] Beis J S, Lowe D G. Indexing without invariants in 3D object recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(10): 1000-1015
- [14] Grimson W E L. Object recognition by computer: the role of geometric constraints [M]. Cambridge: MIT Press, 1990
- [15] David P, Dementhon D, Duraiswami R, *et al.* Evaluation of the SoftPOSIT model-to-image registration algorithm [R]. Maryland: University of Maryland, 2002
- [16] David P, Dementhon D, Duraiswami R, *et al.* SoftPOSIT: simultaneous pose and correspondence determination [J]. *International Journal of Computer Vision*, 2004, 59(3): 259-284
- [17] Iwashita Y, Kurazume R, Konishi K, *et al.* Fast alignment of 3D geometrical models and 2D grayscale images using 2D distance maps [J]. *Systems and Computers in Japan*, 2007, 38(14): 52-62
- [18] Dambreville S, Sandhu R, Yezzi A, *et al.* A geometric approach to joint 2D region-based segmentation and 3D pose estimation using a 3D shape prior [J]. *SIAM Journal on Imaging Sciences*, 2010, 3(1): 110-132
- [19] Zhong F, Qin X Y, Peng Q S. Robust image segmentation against complex color distribution [J]. *The Visual Computer*, 2011, 27(6/7/8): 707-716
- [20] Moré J J. The Levenberg-Marquardt algorithm: implementation and theory [M] // *Lecture Notes in Mathematics*. Berlin: Springer, 1978, 630: 105-116
- [21] Payet N, Todorovic S. From contours to 3D object detection and pose estimation [C] // *Proceedings of IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2011: 983-990