

## 增强现实中的直线边缘对象位姿跟踪

潘绍松 左洪福

(南京航空航天大学民航学院, 南京, 210016)

**摘要:** 改进了增强现实中跟踪直线边缘对象位姿的方法。首先用扩展卡尔曼滤波器预测对象位姿并限制隐式边缘搜索的区域; 然后利用 M 估计子与权重直线拟合得到图像中直线边缘的方程系数; 最后利用平面的标准正交基构建关于对象位姿的目标函数, 并用奇异值分解法加以求解。实验表明本文位姿跟踪方法的旋转角最大的平均偏差为  $0.29^\circ$ , 绝对偏差为  $1.90^\circ$ , 跟踪速度大约为 10 帧/s, 综合性能明显优于基于 LM 的跟踪算法。

**关键词:** 计算机视觉; 互相关; 矩阵范数; 图像注册

**中图分类号:** TP751.2; TP391.44

**文献标识码:** A

**文章编号:** 1005-2615(2011)06-0744-05

## Pose Tracking for Object with Straight Lines in Augmented Reality

*Pan Shaosong, Zuo Hongfu*

(College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, Nanjing, 210016, China)

**Abstract:** The algorithm of pose tracking for object with straight lines in augmented reality is improved. Firstly the extended Kalman filter is used to predict object pose and the implicit searching region is restricted. Then, M-estimator and weight line fitting are used to obtain edge equation coefficients in image. Finally, the objective function about object pose is constructed using canonical basis of plane which is solved by singular value decomposition. Experiment shows that maximum mean deviation is  $0.29^\circ$  and absolute deviation  $1.90^\circ$  for rotation angles, and tracking rate is about 10 frame/s, so its comprehensive performance exceeds obviously the LM-based tracking algorithm.

**Key words:** computer vision; cross-correlation; matrix norm; image registration

在增强现实中, 采用机器视觉的方法基于对象的特征点或边缘轮廓跟踪对象位姿、实现图像注册是近年来的研究热点<sup>[1-2]</sup>, 但大部分工业对象因纹理贫乏而缺少稳定的特征点, 因而在工业环境中利用对象可见边缘的跟踪位姿方法更为常用。边缘跟踪方法大体如下: 首先将对象 3D 模型边缘用前一帧的位姿投影至当前帧图像上成为模型片段(segment), 再取片段上的点并沿其法向在当前图像帧内搜索边缘点, 最后用 LM 法最小化模型片段上点与边缘点间的最小距离实现对象当前位姿的估计, 该方法的缺点是计算量大且精度较低。为提高跟踪精度, Wuest 等人提出了对应的多重假设<sup>[3]</sup>, Kim 引入与模型边缘在同一平面的纹理<sup>[4]</sup>,

Keisuke 则将之推广到任意位置的纹理<sup>[5]</sup>, 文献[6, 7]则分别采用光流法和粒子滤波法以提高计算效率。这些方法最终仍使用 LM 法估计对象位姿。

本文改进了基于对象 3D 模型边缘的位姿跟踪方法: 利用扩展卡尔曼滤波(EKF)预测对象位姿, 提高边缘搜索的效率; 利用权重直线拟合与奇异值分解估算对象位姿, 而不再使用 LM 优化法。如果将 3D 模型理解为“地图”, 那么本文的对象位姿跟踪与 SLAM(同时定位与建图)<sup>[8]</sup>算法有共通之处; 不同之处是用于机器人路径规划<sup>[9-10]</sup>的 SLAM 法只要求确定相机与对象间的方位角与距离, 只有 4 个自由度, 而位姿跟踪则是求相机坐标系与对象坐标系间的 6 自由度的刚体变换。

**基金项目:** 国家自然科学基金联合资助基金(60939003/F01)资助项目。

**收稿日期:** 2010-09-01; **修订日期:** 2011-04-21

**通讯作者:** 左洪福, 男, 教授, 博士生导师, 1959 年生, E-mail: rms@nuaa.edu.cn。

本文首先介绍了利用 EKF 预测对象位姿并限制隐式直线边缘搜索的区域;接着利用 M 估计子和权重直线拟合获得直线边缘在图像中的方程,最后利用平面的标准正交基构造求解位姿的目标函数,并用 SVD 法求解。本文实验部分不仅与 LM 法做了对比,还给出了本文方法在增强现实中的应用实例。

## 1 构建直线边缘对应

本文的算法包括:隐式搜索直线边缘搜索、权重直线拟合、位姿估计 3 个部分,不采用显式方法提取直线边缘是因为该方法不仅耗时而且难以建立边缘的 2D-3D 对应。

### 1.1 边缘位置预测

EKF 广泛应用于视觉预测与目标跟踪<sup>[11]</sup>,在 AR 的连续视频应用中,相机相对于对象的位姿变化大都是缓慢而连续的,可认为在相邻时刻对象位姿的变化是线性小增量的。这里应用 EKF 预测的对象位姿不仅为位姿迭代计算提供初值,而且为直线边缘的搜索提供合理的限定范围。

设  $p^{k-1} \in R^{6 \times 1}$ ,  $\Delta \tilde{p}^{k-1} \in R^{6 \times 1}$  (旋转分量采用罗德里格斯表达)分别为对象前一时刻位姿以及当前的位姿增量,  $\Sigma^{k-1} \in R^{6 \times 6}$  为位姿增量的协方差矩阵,因此当前时刻的对象位姿可以预测为

$$\tilde{p}^k = p^{k-1} + \Delta \tilde{p}^{k-1} \quad (1)$$

3D 模型上的点  $X$  在当前图像中的位置预测为  $\tilde{x} = \text{proj}(\tilde{p}^k, X)$ ,  $\text{proj}(\cdot, \cdot)$  为相机针孔模型投影公式。由协方差前向传播原理<sup>[12]</sup>知道  $\tilde{x}$  的协方差矩阵为

$$\Sigma_{\tilde{x}}^k = J \Sigma^{k-1} J^T \quad (2)$$

式中  $J = \partial \text{proj}(\tilde{p}^k, X) / \partial \tilde{p}^k \in R^{2 \times 6}$ 。因此  $\tilde{x}$  可能的位置  $y$  将构成一个以  $\tilde{x}$  为中心的椭圆

$$(y - \tilde{x})^T \Sigma_{\tilde{x}}^k (y - \tilde{x}) \leq \kappa^2 \quad (3)$$

式中:  $\kappa^2$  表示自由度为 2、置信度为  $\alpha$  时  $\chi^2$  分布的逆,如  $\alpha = 0.95$  时  $\kappa^2 \approx 6$ 。

在利用式(11,12)得到对象位姿  $p^k$  后, EKF 的状态更新方程为<sup>[13]</sup>

$$G^k = \Sigma^{k-1} J^T (J \Sigma^{k-1} J^T + N) \quad (4)$$

$$\Delta \tilde{p}^k = \Delta \tilde{p}^{k-1} + G^k (\tilde{x} - \text{proj}(p^k, X)) \quad (5)$$

$$\Sigma^k = (I_6 - G^k J) \Sigma^{k-1} \quad (6)$$

式中:  $G \in R^{6 \times 2}$  是滤波器增益;  $N \in R^{2 \times 2}$  代表过程激励噪声。

### 1.2 直线边缘点隐式搜索

本文通过计算梯度归一化互相关系数的方式在图像中搜索与模型边相对应的直线边缘。具体

步骤如下:(1)用  $n$  个等间距分布的 3D 点  $X_i$  模拟 3D 模型片段  $L_j$ ; (2)称  $\text{proj}(p^{k-1}, X_i)$  为采样点,计算  $\text{proj}(p^{k-1}, L_j)$  法线方向  $\beta$ ; (3)沿  $\beta$  方向以  $[1 \ 0 \ -1]$  为掩码分别计算前一帧和当前帧图像的梯度; (4)沿  $\beta$  方向,在采样点周围取梯度的矩形区域作为模板; (5)沿  $\beta$  方向,在当前图像帧中从  $\text{proj}(\tilde{p}^k, X_i)$  开始在限定的区域内以 0.5 像素为步长,取与模板同方向、同尺寸的梯度矩形区域作为目标区域; (6)依次计算每个目标区域与模板区域的归一化互相关系数,最大值的位置作为与模型直线边缘相对应的图像边缘上点,记为集合  $S_j$ ; (7)对每条模型边的所有的采样点重复步骤(4~6)。图 1 为位置预测、搜索区域、搜索方向以及不同方向梯度模板的示意图。

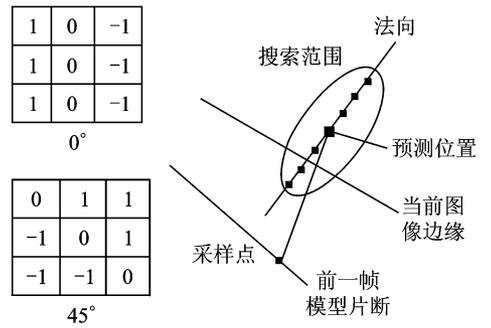


图 1 搜索方向与搜索位置示意图

边缘隐式搜索的方法源自文献[3,14],但原文献中搜索起点为采样点、搜索范围靠经验确定,由此导致无效搜索区域较大,为保证搜索结果正确,必须采用各向异性波高斯滤波器<sup>[15]</sup>平滑图像。该滤波器需要对每一个  $\beta$  方向的边缘进行一次耗时数百毫秒的滤波运算,而采用了 EKF 预测和限定搜索位置后则无需再采用该滤波器。

### 1.3 权重直线拟合

数据集  $S_j$  是当前图像帧中  $L_j$  上的点,设它们的归一化坐标为  $x_i = (u_i, v_i)$ , 其中  $i = 1, 2, \dots, n$ 。现需要利用  $S_j$  中的点拟合  $L_j$  的方程  $\alpha_j u + \beta_j v + \gamma_j = 0$ , 在拟合时应区别对待  $S_j$  中的野值和数据误差,为此用最小化权重误差(式(7))拟合  $L_j$ <sup>[16]</sup>

$$\varepsilon = \sum_{i=1}^n \omega_i (\alpha_j u_i + \beta_j v_i + \gamma_j)^2 - \lambda n (\alpha_j^2 + \beta_j^2) \quad (7)$$

式中:权重因子  $\omega_i$  用以评估拟合直线时点对误差(7)的“贡献”。 $\omega_i$  的值用 M 估计子<sup>[17]</sup>加以估计

$$\omega_i = \begin{cases} (1 - (\delta_i / \tau)^2)^2 & \delta_i \leq 2\sigma_j \\ 0 & \text{其他} \end{cases} \quad (8)$$

式中:  $\delta_i$  表示  $x_i$  到  $L_j$  的距离;  $\sigma_j$  表示距离的方差

它用绝对中值偏差<sup>[17]</sup>估计。由式(7)可得到 $\alpha_j, \beta_j$ 是矩阵 $\begin{bmatrix} \bar{u}^2 - (\bar{u})^2 & \bar{uv} - \bar{u}\bar{v} \\ \bar{uv} - \bar{u}\bar{v} & \bar{v}^2 - (\bar{v})^2 \end{bmatrix}$ 最小特征值所对应的特征向量中的元素<sup>[18]</sup>。矩阵中的元素上方的符号表示取权重均值,如 $\bar{u} = \frac{\sum_{i=1}^n \omega_i u_i}{\sum_{i=1}^n \omega_i}$ 。 $\omega_i$ 初值取1,实验表明3~5次迭代即可有效地剔除 $S_j$ 中的野值。

## 2 对象位姿估计

本文的目标是估计对象位姿,即对象坐标系与相机坐标系间的变换 $(R, t)$ 。

### 2.1 构造目标函数

如图2所示, $o_c$ 为相机光学中心, $\pi$ 为像平面, $I_c$ 为相机主轴与 $\pi$ 的交点, $l_j$ 为3D模型片段 $L_j$ 的投影, $l_j$ 和 $o_c$ 确定平面 $N_j$ 。在对象坐标系中 $X_j$ 是 $L_j$ 上任意一点, $D_j$ 是 $L_j$ 上单位向量。过 $I_c$ 向 $l_j$ 引垂线,垂足为 $d_j$ ,连接 $o_c d_j$ ,则有 $o_c d_j \perp l_j$ 。在相机坐标系内, $l_j$ 和 $o_c d_j$ 所构成的单位向量分别为 $a_j = [-\beta_j, \alpha_j, 0]^T$ 和 $b_j = [-\alpha_j \gamma_j, -\beta_j \gamma_j, 1]^T / \sqrt{1 + \gamma_j^2}$ ,因此 $a_j$ 和 $b_j$ 是 $N_j$ 的单位正交基,任给 $N_j$ 中向量 $\eta$ 都有 $\eta = (\eta, a_j)a_j + (\eta, b_j)b_j = (a_j a_j^T + b_j b_j^T)\eta$ 。定义 $H_j = a_j a_j^T + b_j b_j^T$ 则 $N_j$ 中向量 $RD_j$ 与 $RX_j + t$ 可分别表达成

$$RD_j = H_j RD_j \quad (9)$$

$$RX_j + t = H_j (RX_j + t) \quad (10)$$

若从式(9)能估算出 $R$ ,由式(10)可得出

$$t = \left( \sum_{j=1}^m (I - H_j) \right)^{-1} \left( \sum_{j=1}^m (H_j - I) RX_j \right) \quad (11)$$

式中: $m$ 为2D-3D直线边缘对应的数量。

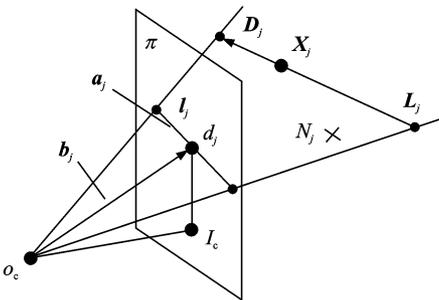


图2 构造目标函数示意图

### 2.2 对象位姿估算

构造 $A = [H_1 RD_1, \dots, H_m RD_m] \in R^{3 \times m}$ 和 $B = [D_1, \dots, D_m] \in R^{3 \times m}$ ,由式(9)自然地想到以 $\operatorname{argmin}_R \|A - RB\|_F^2$ 为目标函数求解 $R$ ,即

$\operatorname{argmin}_R (\operatorname{tr}(A^T A) + \operatorname{tr}(B^T B) - 2\operatorname{tr}(A^T RB))$ ,这表明若 $A$ 和 $B$ 固定, $\|A - RB\|_F^2$ 最小等价于 $\operatorname{tr}(A^T RB)$ 最大。设 $\operatorname{SVD}(BA^T) = UWV^T$ 且记 $Z = V^T R U$ ,则有 $\operatorname{tr}(A^T RB) = \operatorname{tr}(RBA^T) = \operatorname{tr}(RUWV^T) = \operatorname{tr}(V^T RUW) = \operatorname{tr}(ZW) = \sum_{i=1}^3 z_{ii} \omega_i \leq \sum_{i=1}^3 \omega_i$ 。这表明仅当 $Z$ 为单位阵,目标函数值最小,即

$$R = VU^T \quad (12)$$

从 $R$ 的初值出发构造 $A$ 和 $B$ ,然后对 $BA^T$ 进行奇异值分解得到最优的 $R$ ,重复该过程直至 $R$ 收敛到某个不动点。显然只有 $\operatorname{rank}(BA^T) = 3$ ,上述计算才是正确的,这要求模型直线边缘数量 $m \geq 3$ 且不是汇聚到同一点的直线束。

3D模型与图像的初始边缘对应由手工完成,采用平行透视投影法<sup>[19]</sup>的近似结果作为式(9)中 $R$ 的迭代初值;在后续的对象位姿连续视频跟踪中,边缘对应则由直线边缘隐式搜索的方式产生, $R$ 的迭代初值取EKF的预测值。

### 2.3 算法总结

基于直线边缘对象的位姿跟踪算法,流程综合如下:

**步骤1** 系统初始化,人工选择对象上的直线边缘;

**步骤2** 用式(1)预测当前相机位姿,用式(3)限制直线边缘的搜索区域;

**步骤3** 计算模板与目标区域的梯度归一化互相关系数,使系数达到最大值的位置记入 $S_j$ ;

**步骤4** 用基于M估计子的权重直线拟合得到 $l_j$ ;

**步骤5** 对直线边缘 $j=1, 2, \dots, m$ ,重复步骤2~4;

**步骤6** 迭代地构造 $A, B$ ,用式(11,12)获得 $(R, t)$ ;

**步骤7** 用式(4~6)对EKF作状态更新;

**步骤8** 转步骤(2),实现连续视频对象位姿跟踪。

## 3 实验与结果

实验先对比本文算法和LM<sup>[3-6]</sup>优化算法间的性能差异,这里选择的对象是1个电源控制盒,被跟踪的5条边如图3所示,选择该对象的原因是它的边缘明显、尺寸标准,本身位姿易于标定,方便实验对比。实验中的标准位姿由Fastrak位姿跟踪器获得,实验程序用C++基于OpenCV编写。

为体现2种算法的差异,将离线的连续视频图



图 3 用于算法比较的被跟踪对象

像按 3 : 1 的方式重新采样,以增加相邻图像间的变化量。图 4 显示了本文 Line 算法与 LM 算法所恢复的偏航角、俯仰角、横滚角的结果,其中纵轴为角度、横轴是帧序号,该图显示本文算法与标准位姿真值符合得较好,而 LM 法只是在变化趋势上与标准位姿相一致,尽管如此在 96 帧时 LM 法还是丢失了被跟踪对象。出现这样的结果是因为: LM 法本质是一种“抛光”技术,当相邻图像间的变化量超过其“抛光”限度后算法当然就会失效。

角度的最大绝对偏差、偏差的均值以及偏差的方差具体数值,这些数值表明本文的对象位姿跟踪精度满足 AR 图像注册需求。

表 1 两种方法跟踪对象位姿的偏差比较

位姿	最大绝对偏差		偏差均值		偏差的方差	
	Line 法	LM 法	Line 法	LM 法	Line 法	LM 法
偏航角	0.94	3.02	-0.22	-2.22	0.26	6.02
俯仰角	1.90	6.69	-0.11	4.09	0.62	11.30
横滚角	1.70	1.46	0.29	1.46	0.60	2.79

为了进一步验证本文算法的实用性,又选择了公园中的木制花坛作为被跟踪对象,花坛尺寸手工测绘。图 5 列出了在部分图像上利用位姿跟踪结果实现图像注册(白色细线条)以及对对象坐标的结果图像。从图中可以看到,利用跟踪位姿实现的图像注册结果是令人满意的。

实验中发现用 EKF 实现的直线边缘隐式搜索比用经验确定的位置和区域的搜索方法要节约 1 倍的时间,这说明 EKF 预测相机位姿的有效性。运行 1 次 2.3 节中步骤(2~8)耗时约 90~110 ms,即本文方法位姿跟踪的速度大约为 10 帧/s。

表 1 给出了图 4 中两种算法相对于标准位姿

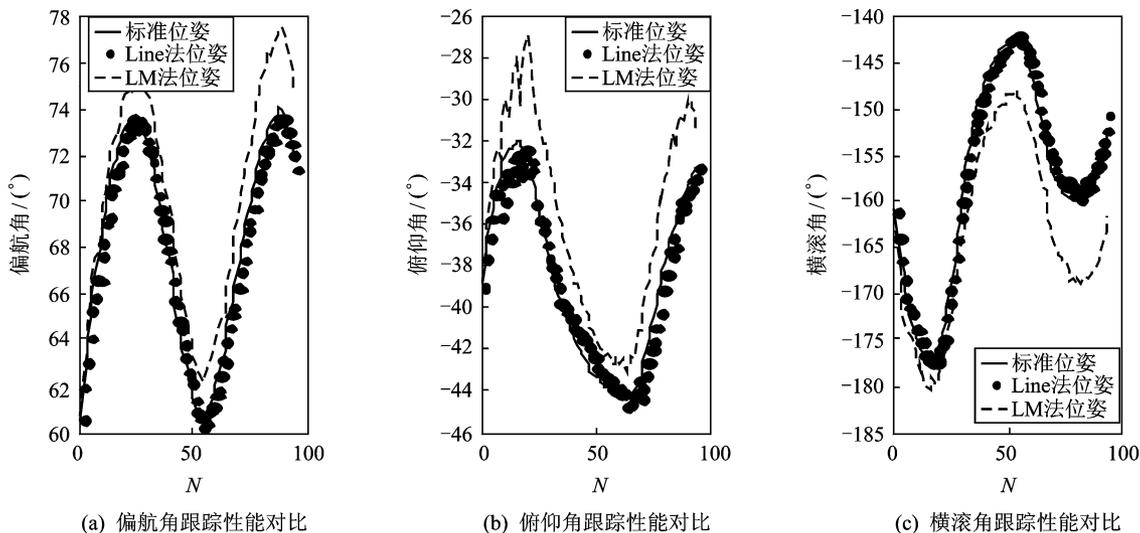


图 4 位姿跟踪算法的性能对比

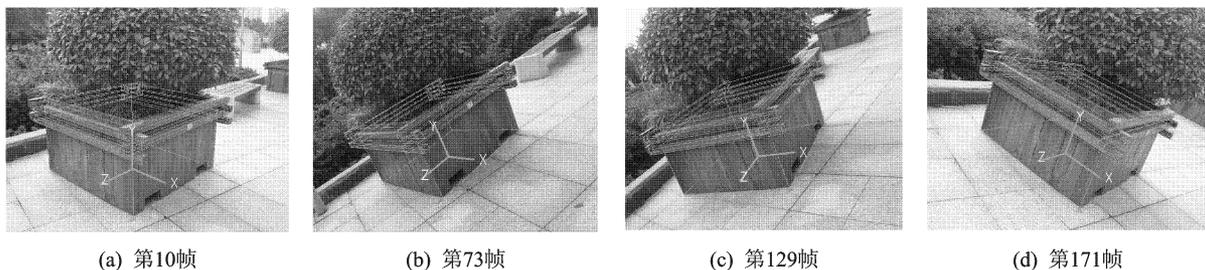


图 5 位姿跟踪与图像注册

## 4 结 论

综合文中的叙述,可以得到如下结论:

(1)利用 EKF 预测对象位姿并限制隐式直线边缘搜索的区域可以提高建立直线边缘对应的效率;

(2)利用 M 估计子的权重直线拟合可有效地消除野值对直线拟合精度的影响;

(3)利用平面正交基构造位姿目标函数并用 SVD 法求解的方式是可行的。

本文的对象位姿跟踪方法满足增强现实图像注册的要求。

### 参考文献:

- [1] 黄有群,李志鹏,胡超博,等.采用视觉注册的户外增强现实技术的研究[J].工程图学学报,2010(1):84-87.
- [2] 陈靖,王涌天,刘越,等.适用于户外增强现实系统的混合跟踪定位算法[J].计算机辅助设计与图形学学报,2010,22(2):204-209.
- [3] Harald W, Florent V. Adaptive line tracking with multiple hypotheses for augmented reality[C]//Proceedings Fourth IEEE and ACM International Symposium. Vienna, Austria: IEEE, 2005:62-69.
- [4] Kim S, Diverdi S. Implicit 3D modeling and tracking for anywhere augmentation[C]//The Symposium on Virtual Reality Software and Technology. Newport Beach, California: ACM, 2007:19-28.
- [5] Keisuke T, Daisuke K. Model-based 3D object tracking with online texture update[C]//MVA2009 IAPR Conference MVA Conf Comm on Machine Vision Applications. Yokohama, Japan: [s. n.], 2009:261-264.
- [6] Pressigout M, Marchand E. Hybrid tracking approach using optical flow and pose estimation[C]//IEEE Int Conf on Image Processing, ICIP' 08. San Diego, CA: IEEE, 2008:2720-2723.
- [7] Morwald T, Zillich M. Edge tracking of textured objects with a recursive particle filter[C]//19th International Conference on Computer Graphics and Vision. Moscow: Citeseerx, 2009:96-103.
- [8] 弋英民,刘丁.基于单目视觉的轮式机器人同步定位与地图构建[J].仪器仪表学报,2010,31(1):117-121.
- [9] 张文玲,朱明清.基于强跟踪 UKF 的自适应 SLAM 算法[J].机器人,2010,32(3):190-195.
- [10] 孙凤池,康叶伟,黄亚楼,等.基于综合互信息的视觉 SLAM 主动探索[J].模式识别与人工智能,2010,23(3):434-440.
- [11] 赵其杰,屠大伟.基于 Kalman 滤波的视觉预测目标跟踪及其应用[J].光学精密工程,2008,16(5):937-942.
- [12] Zisserman R H. 计算机视觉中的多视图几何[M]. 韦穗,杨尚骏,章权兵,等译.合肥:安徽大学出版社,2002.
- [13] Moreno-Noguer F, Lepetit V. Pose priors for simultaneously solving alignment and correspondence [C]//10th European Conference on Computer Vision (ECCV 2008). Marseille, FRANCE: IEEE, 2008: 405-418.
- [14] Comport A I, Pressigout M. Real-time markerless tracking for augmented reality the virtual visual servoing framework[J]. Vis and Comp Graph, 2006,12(4):615-628.
- [15] Geusebroek J M, Arnold W M. Fast anisotropic gauss filtering [J]. IEEE Trans Image Process, 2003,12(8):938-943.
- [16] Carsten S, Markus U. 机器视觉算法与应用[M]. 杨少荣,吴迪靖,段德山,等译.北京:清华大学出版社,2008.
- [17] Arya K V, Gupta P. Image registration using robust m-estimator[J]. Pattern Recognition Letters, 2007, 28(15):1957-1968.
- [18] Forsyth D A, Ponce J. Computer vision: a modern approach [M]. New Jersey: Pearson Education, 2002.
- [19] Horaud R, Dornaika F. Object pose: the link between weak perspective paraperspective, and full perspective[J]. International Journal of Computer Vision, 1997,22(2):173-189.