

基于 CAD 模型的目标跟踪和定位算法研究^①张鹏程^② 徐 德^③

(中国科学院自动化研究所精密感知与控制中心 北京 100190)

摘 要 基于 Eye-to-Hand 单目视觉配置,研究了一种三维目标跟踪和定位算法,实现了在笛卡尔空间下三维目标的精确跟踪和定位。在目标跟踪过程中,采用基于模型的三维跟踪算法,利用虚拟视觉伺服算法使 CAD 模型在图像空间中的投影收敛于目标边缘,进而实现在笛卡尔空间中与三维目标精确对准,获得实时、鲁棒的三维跟踪。在目标定位过程中,在笛卡尔空间设计增量 PI 控制算法,使当前目标位姿收敛到期望位姿。实验结果表明,该算法能精确地跟踪和定位三维目标,位置精度控制在 0.2mm 以内,姿态精度控制在 0.1° 以内。

关键词 CAD 模型, 位姿估计, 三维跟踪, 视觉反馈控制, 视觉定位

0 引 言

三维跟踪技术在计算机视觉领域有重要的应用价值,可用于目标定位、视觉伺服、增强现实、人机交互等方面。基于模型的跟踪是通过一定先验知识对所跟踪的目标离线建立模型,如三维点云、计算机辅助设计(CAD)模型和体模型等,然后通过实时更新模型的三维位姿,实现模型与目标的对准。由于模型含有精确的三维信息,对目标的表面、边缘和连接关系都有定性和定量和描述,因此,基于模型的三维跟踪更加鲁棒,应用广泛。现有文献提出的多种基于模型的三维跟踪方法。大都基于相邻帧或匹配关键帧获得当前目标三维位姿。Harris 最早提出了实时姿态和位置确定(real-time attitude and position determination, RAPD)方法,它以模型投影与图像边缘特征误差最小为目标函数,实时估计目标位姿,但相邻帧之间位姿变化需限制在一定范围内^[1]。Mooser 等利用 SFM (Structure from Motion) 建立点云模型,并附加尺度不变特征转换(SIFT)局部描述量;在跟踪阶段,实时匹配输入图像特征与点云特征,并通过 Lenzenbeng- Marquardt 优化位姿^[2],但若输入图像与模型位姿偏差较大,很难获得鲁棒的匹配。Drummond 将位姿表达成李群代数空间矩阵基的线

性组合,优化权值,通过迭代更新边缘距离向量^[3],但受上一帧目标位姿影响较大。Vacchetti 等采用 CAD 模型离线建立一系列的目标关键帧,结合关键帧和在线信息恢复摄像机参数,然后结合输入图像 Harris 特征与最接近的关键帧的特征进行在线匹配,恢复目标的位姿^[4, 5],但初始化时,目标必须位于关键帧附近。Marchand 等利用边缘信息提出了虚拟视觉伺服(virtual visual servoing, VVS)技术,利用鲁棒的伺服控制率,迭代收敛使模型与目标实现三维对准^[6-8]。Vacchetti、Pressigout 等融合目标纹理和边缘信息,跟踪性能更加稳定^[9-11]。

本文研究目标是以图像空间 CAD 模型的投影特征与目标的边缘特征偏差为目标函数,采用虚拟视觉伺服算法迭代最小化此特征偏差,获得笛卡尔空间中模型当前位姿与上一帧的位姿变化,实时跟踪三维目标。最后,采用鲁棒的视觉反馈控制,实现精确的三维定位。

1 基于模型的三维跟踪算法

1.1 目标边缘跟踪

三维跟踪需要从目标图像检测相应特征,例如,边缘特征、关键点和局部描述子等。边缘检测方法

① 国家自然科学基金(61227804, 61105036)资助项目。

② 男,1985 年生,博士;研究方向:视觉跟踪和定位,视觉伺服,机器视觉;E-mail: pengcheng. zhang@ia. ac. cn

③ 通讯作者, E-mail: de. xu@ia. ac. cn

(收稿日期:2013-12-15)

有梯度算子、Canny、LoG 等,关键点有 Harris 角点,局部特征描述子有 Sift、Surf 等。但是机器视觉的大多数目标缺乏纹理信息,很难获得鲁棒的关键点和描述子描述。因此,考虑到目标特性和实时性要求,

采用基于边缘的特征检测方法,即在图像平面采用 Moving-Edge 方法跟踪目标边缘轮廓,沿轮廓法方向搜索边缘像素点,实时获得目标边缘^[6],如图 1 所示。

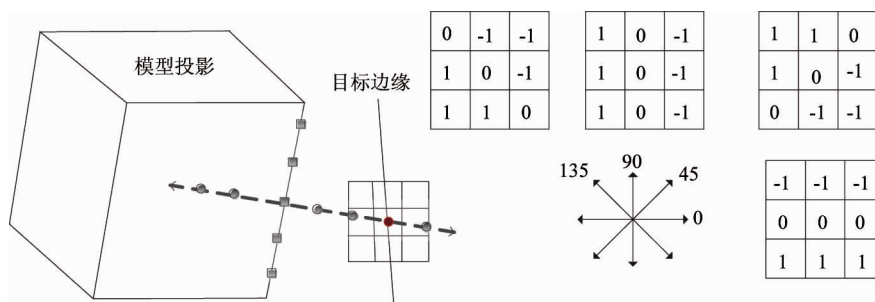


图 1 Moving-edge 边缘跟踪原理及其检测算子

基于边缘的特征检测方法首先利用上一帧的目标位姿矩阵,求得当前模型边缘在图像空间可见边缘的投影直线,即通过模型节点的投影计算直线极坐标方程参数。然后,沿直线方向采样,设定计算采样步长,并计算各采样点的像素坐标。利用每个采样点像素坐标和直线的角度参数,选择边缘检测单元。其中,根据直线角度选择相应角度的边缘检测算子,本文采用边缘检测算子 $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$,如图 1 所示。对每个采样单元,沿直线法方向搜索,并设定搜索像素范围,求解每个检测单元法线方向像素的卷积最大值,并计算相应的像素坐标。在下一帧边缘检测过程中,利用当前帧的目标位姿求模型可见边缘在图像平面的投影,并沿直线法线方向搜索最大卷积值,获得下一帧的边缘像素点,循环直到跟踪结束。

在跟踪过程中,需要判断模型各表面相对于摄像机的可见性。因此,针对于每个表面,在摄像机坐标系中构建从坐标原点到各平面中心的三维向量 OA_i ,如图 2 所示,并根据此三维向量与相应的平面

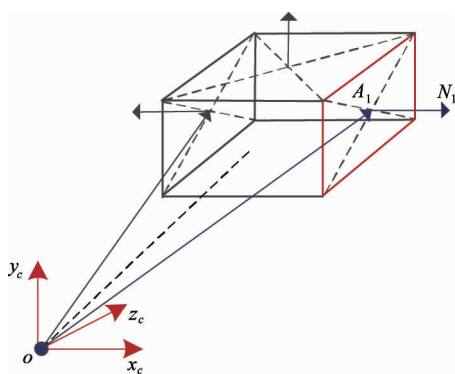


图 2 判断模型的可见表面

法向量 $A_i N_i$ 的夹角来判断此平面相对于摄像机是否可见或邻近可见。在边缘跟踪过程中,根据模型可见边缘的图像空间投影,沿其法方向搜索相应的目标边缘点。

1.2 点到直线交互矩阵

下面将讨论如何将图像空间的特征变化转化到笛卡尔空间的位姿变化。本文参考文献[6],采用点到直线距离的交互矩阵表达图像空间特征变化与笛卡尔空间位姿变化的关系。假设摄像机坐标系中点 $P(x_c, y_c, z_c)$,其相应的焦距归一化平面上的投影点为 $p(x, y)$,则对投影关系求导如下:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1/z_c & 0 & -x_c/z_c^2 \\ 0 & 1/z_c & -y_c/z_c^2 \end{pmatrix} \begin{pmatrix} \dot{x}_c \\ \dot{y}_c \\ \dot{z}_c \end{pmatrix} \quad (1)$$

可得目标在三维空间平移和旋转方向的速度与在图像空间坐标速度的关系,即点的交互矩阵^[12]。其中,对其估计需要利用相机内参数以及在摄像机坐标系中的深度信息。

对于图像空间的直线方程可表达为参数方程形式,即 $x \cos \theta + y \sin \theta - \rho = 0$,其中, θ 为直线与图像坐标系 u 轴的夹角, ρ 为图像坐标原点到直线的距离。在摄像机坐标系中,对于空间直线所在的平面方程可表示为 $A_1 x_c + B_1 y_c + C_1 z_c + D_1 = 0$,则焦距归一化图像平面的投影直线方程为

$$Ax + By + C = 1/z_c \quad (2)$$

其中, $A = -A_1/D_1$, $B = -B_1/D_1$, $C = -C_1/D_1$ 。对直线极坐标参数方程取导数操作,即

$$\dot{\rho} + (x \sin \theta - y \cos \theta) \dot{\theta} = \dot{x} \cos \theta + \dot{y} \sin \theta \quad (3)$$

由式(1),(2),(3),并参考文献[6]得

$$\begin{pmatrix} \dot{\rho} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \lambda_\rho \cos\theta & \lambda_\rho \sin\theta & -\lambda_\rho \rho & (1+\rho^2)\sin\theta \\ \lambda_\theta \cos\theta & \lambda_\theta \sin\theta & -\lambda_\theta \rho & -\rho \cos\theta \\ -(1+\rho^2)\cos\theta & 0 & -\rho \sin\theta & -1 \end{pmatrix} \quad (4)$$

其中, $\lambda_\rho = -A\rho \cos\theta - B\rho \sin\theta - C$, $\lambda_\theta = -A\sin\theta + B\cos\theta$, 则直线交互矩阵为

$$L_l = \begin{pmatrix} \lambda_\rho \cos\theta & \lambda_\rho \sin\theta & -\lambda_\rho \rho & (1+\rho^2)\sin\theta \\ \lambda_\theta \cos\theta & \lambda_\theta \sin\theta & -\lambda_\theta \rho & -\rho \cos\theta \\ -(1+\rho^2)\cos\theta & 0 & -\rho \sin\theta & -1 \end{pmatrix} \quad (5)$$

对于点到直线距离的交互矩阵, 表达为图像空间中点到直线的距离与笛卡尔空间直线位姿变化的关系, 其中, 点到直线的距离可由图像坐标原点到两条斜率相等直线的距离差得出, 即点 (x_d, y_d) 到直线 (ρ, θ) 的距离可表达为

$$d = \rho - \rho_d = \rho - x_d \cos\theta + y_d \sin\theta \quad (6)$$

上式求导后化为

$$\dot{d} = \dot{\rho} - (x_d \cos\theta + y_d \sin\theta) \dot{\theta} = \dot{\rho} - \alpha \dot{\theta} \quad (7)$$

其中, $\alpha = x_d \cos\theta + y_d \sin\theta$, 则由直线交互矩阵推导得到点到直线距离的交互矩阵 $L_d = L_\rho - \alpha L_\theta$, 即

$$L_d = \begin{bmatrix} \lambda_d \cos\theta \\ \lambda_d \sin\theta \\ -\lambda_d \rho \\ (1+\rho^2)\sin\theta - \alpha \rho \cos\theta \\ -(1+\rho^2)\cos\theta - \alpha \rho \sin\theta \\ -\alpha \end{bmatrix} \quad (8)$$

其中, $\lambda_d = \lambda_\rho + \alpha \lambda_\theta$ 。对于每一个跟踪单元中的图像特征表达为边缘点与相应的模型投影直线之间的距离。每个跟踪单元都有相应的交互矩阵, 其表达该距离变化和笛卡尔空间位姿变化的关系。然而, 在目标中存在多个跟踪单元, 需要融合目标所有图像特征, 构造系统的交互矩阵, 实现所有可见模型投影边缘特征收敛于所有图像边缘点。因此, 通过堆垛每个跟踪单元的子交互矩阵得到系统的交互矩阵, 即

$$L_s = [L_1 \ L_2 \ \cdots \ L_m] \quad (9)$$

每个跟踪单元特征的交互矩阵表示为

$$L_i = \begin{bmatrix} \lambda_{di} \cos\theta_i \\ \lambda_{di} \sin\theta_i \\ -\lambda_{di} \rho_i \\ (1+\rho_i^2)\sin\theta_i - \alpha_i \rho_i \cos\theta_i \\ -(1+\rho_i^2)\cos\theta_i - \alpha_i \rho_i \sin\theta_i \\ -\alpha_i \end{bmatrix} \quad (10)$$

其中, $\lambda_{di} = \lambda_{\rho_i} + \alpha_i \lambda_{\theta_i}$ ($i = 1, \dots, m$), 行数为在可见模型投影边缘采样的特征单元个数, 列数为模型笛卡尔空间上自由度的个数。然后, 通过相机内参数和直线极坐标参数实时估计当前的系统交互矩阵, 并利用其伪逆将图像空间特征偏差转化为笛卡尔坐标中模型速度变化, 如下所示:

$$[v_x \ v_y \ v_z \ w_x \ w_y \ w_z]^T = L_{n \times 6}^+ [d_1 \ d_2 \ \cdots \ d_m]^T \quad (11)$$

其中, $L_{n \times 6}^+$ 为目标交互矩阵的伪逆, $v_x, v_y, v_z, w_x, w_y, w_z$ 表示当前模型在旋转和平移方向的速度, $d_1, d_2, \dots, d_m$ 表示当前目标边缘点到相应的模型可见投影边缘直线的距离。

1.3 鲁棒的虚拟视觉伺服

为实现笛卡尔空间的三维对准, 需要最小化模型投影特征和目标边缘特征之间的偏差, 即在图像空间使目标边缘点到直线的距离收敛到零, 可将其看为迭代最小化偏差问题。考虑到实时性等因素, 本文采用鲁棒的 M 估计, 即带权重的最小二乘法, 在图像空间建立目标函数, 表达为目标边缘特征与模型投影特征之间的偏差

$$\Delta = \sum_{i=1}^N \rho(s_i(r) - s_i^*)^2 = \sum_{i=1}^N \rho(p(K, M_i) - s_i^*)^2 \quad (12)$$

其中, ρ 为鲁棒估计参数, M_i 为当前模型的投影矩阵, N 为边缘跟踪特征个数, K 为相机内参数。

基于模型的三维目标跟踪过程如图 3 所示, 通过目标的 CAD 模型获得节点和表面信息, 判断其表面可见性, 并跟踪可见边缘。实时的目标位姿估计问题, 即三维模型与目标对准过程, 可看作为虚拟视觉伺服 (VVS) 控制过程^[6]。假设以一定的控制律控制目标模型在摄像机坐标系中运动, 多次迭代后, 使模型位姿收敛到当前目标的位姿, 获得模型的位姿变化量, 通过上一帧模型的位姿计算当前帧模型的位姿, 即为当前目标的位姿。其控制律可表达为

$$v = -\lambda L_s^+ D(s(r) - s^*) \quad (13)$$

其中, v 是模型的速度控制率, λ 为比例因子, L_s^+ 是图像特征交互矩阵, 表示目标在笛卡尔空间的速度变化和图像空间特征的变化关系, $D = \text{diag}(w_1, w_2, \dots, w_m)$ 为对角阵, w_i ($i = 1, 2, \dots, m$) 表达各图像特征的权重。获得模型瞬时速度向量 v 后, 通过 Rodrigues 变换将其转化为其瞬时位姿矩阵增量, 如下所示:

$${}^c M_i^{n+1} = {}^c M_i^n e^v \quad (14)$$

其中, n 为当前的迭代收敛次数。在迭代过程中, 若

当前跟踪单元数所占总采样点数的百分比小于设定阈值时,迭代中止。经过多次迭代后,模型位姿收敛于当前目标的位姿,从而实现精确三维跟踪。

在迭代最小化过程中,首先,初始化每个特征偏差权重,比较当前点到直线的距离特征与上一帧中的距离特征。若距离变化在设定阈值范围内,则增加当前特征的权重,否则,减小当前特征的权重。但在实际情况下,特征边缘点误差不可能都服从高斯分布,存在一些边缘点偏离真实边缘。因此,采用 M 估计再次优化位姿矩阵。M 估计的输入为各跟踪单元的特征偏差,输出为相应的权重值。对于 M 估计中的标准偏差 σ 在每次迭代中给出其估计,通常取平均绝对残差的常数倍,即, $\sigma = 1.4826 Med(\delta_i - Med(\delta))$ 。其中, $\delta_i (i = 1, \dots, m)$ 为各特征的偏差, $Med(\delta)$ 为特征平均绝对偏差。本文采用基于 Tukey 函数的 M 估计量,即

$$\varphi(x_i) = \begin{cases} [1 - (x_i/c)^2]^2, & \text{若 } x_i < 4.6851 \\ 0, & \text{其它} \end{cases} \quad (15)$$

其中, $x_i = \delta_i/\sigma$, 各特征的权重为 $w_i = \varphi(x_i)/x_i$, ($i = 1, \dots, m$)。最后,将各特征偏差分别乘以其相应权重值。若权重值近似为零,则意味着当前图像边缘点特征为外点。经多次迭代后,模型投影直线收敛于目标边缘特征,则获得当前帧目标位姿,其将用于下一帧位姿估计,直至跟踪结束。

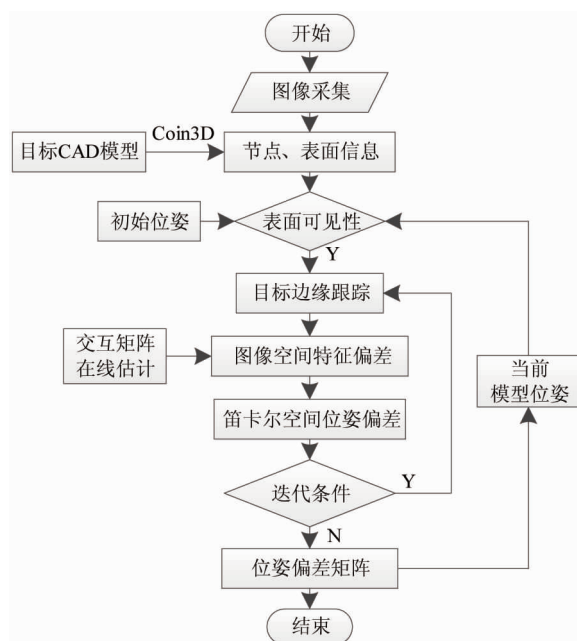


图3 基于模型跟踪的程序流程图

2 三维定位算法

2.1 视觉系统的手眼标定

机器人手眼标定的基本思路是控制机器人末端运动到不同位姿,从而推导机器人基坐标系和摄像机坐标系的平移和旋转关系。常用的手眼标定方法有 Tsai 两步法、Ma 自标定法和 Lei 自标定法等。其中, Ma 和 Lei 的自标定法需要运动多次纯平移和旋转,不需要摄像机内参数。Tsai 法需要标定板,仅运动两次,通过摄像机内参数获得摄像机与机器人基坐标系的变换关系。本系统为 Eye-to-Hand 手眼配置,为了获得较高的标定精度,采用 Tsai 两步法^[13,14]。首先,利用标定板标定机器人基坐标系和摄像机坐标系的变换关系,然后,利用目标标定机器人末端与目标的位姿变换关系。其中,标定坐标系如图4所示, $O_c-X_cY_cZ_c$ 表示是摄像机坐标系, $O_w-X_wY_wZ_w$ 表示机器人基坐标系, O_f 表示标定坐标系原点, O_e 表示机器人末端坐标系原点。

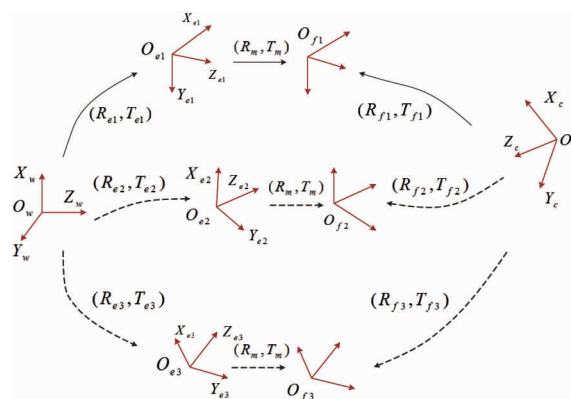


图4 手眼标定坐标系系统

摄像机坐标系 $\{C\}$ 中的点坐标和图像空间点坐标变换关系为 $X_c = \lambda K^{-1}u$ 。机器人末端坐标系 $\{R\}$ 与坐标系 $\{C\}$ 的变换关系为 $X_r = {}^R R_c X_c + {}^R t_c$ 。世界坐标系 $\{W\}$ 下和坐标系 $\{R\}$ 下的变换关系为 $X_w = {}^W R_r X_r + {}^W t_r$ 。图像空间坐标与世界坐标系 $\{W\}$ 的变换关系为 $X_w = \lambda {}^W R_r R_c K^{-1}u + {}^W R_r t_c + {}^W t_r$ 。从而, 机器人基坐标系到摄像机坐标系的变换关系表示为 $(R_c, T_c) = (R_e, T_e)(R_m, T_m)(R_f, T_f)^{-1}$ 。控制机器人末端经过两次运动,通过三个姿态得以下方程:

$$\begin{aligned} & (R_{e(i-1)}, T_{e(i-1)})^{-1}(R_{ei}, T_{ei}) \\ &= (R_m, T_m)(R_{f(i-1)}, T_{f(i-1)})^{-1}(R_{fi}, T_{fi})(R_m, T_m)^{-1} \end{aligned} \quad (16)$$

即, $P_{li} = P_m P_{ri} P_m^{-1}$, 其中, $i = 1, 2, 3, P_{li} = (R_{li}, T_{li}) = (R_{e(i-1)}, T_{e(i-1)})^{-1} (R_{ei}, T_{ei}), P_{ri} = (R_{ri}, T_{ri}) = (R_{f(i-1)}, T_{f(i-1)})^{-1} (R_{fi}, T_{fi}), P_m = (R_m, T_m)$ 。由于 P_{li}, P_{ri}, P_m 均为正交矩阵, 则 P_{li}, P_{ri} 为相似矩阵, 具有相同特征值。根据通用旋转变换, 任意姿态可由绕空间单位向量的旋转表示, $P_{li} = Rot(k_{li}, \theta_{li}), P_{ri} = Rot(k_{ri}, \theta_{ri})$, 则从机器人末端坐标系到标定板坐标系的旋转变换为

$$R_m = [k_{l1} \ k_{l2} \ k_{l1} \times k_{l2}] [k_{r1} \ k_{r2} \ k_{r1} \times k_{r2}]^{-1} \quad (17)$$

同理, 平移变换为

$$T_m = \begin{bmatrix} R_{r1} - I \\ R_{r2} - I \end{bmatrix}^{-1} \begin{bmatrix} R_m \cdot T_{l1} - T_{r1} \\ R_m \cdot T_{l1} - T_{r2} \end{bmatrix} \quad (18)$$

从而得到机器人基坐标系与摄像机坐标系的位姿变换关系为 $(R_c, T_c) = (R_e, T_e)(R_m, T_m)(R_f, T_f)^{-1}$ 。至此, 摄像机坐标系与机器人基坐标系的位姿变换关系标定结束。

下面标定机器人末端坐标系与目标坐标系的位姿变换关系。计算当前目标在摄像机坐标系中的位姿 (R_β, T_β) , 通过已标定的摄像机坐标系和机器人基坐标系位姿变换矩阵, 获得目标坐标到机器人末端坐标系的变换关系为 $(R, T) = (R_{e3}, T_{e3})^{-1} (R_c, T_c) (R_\beta, T_\beta)$ 。

2.2 视觉反馈控制算法

通过三维跟踪实时获得的目标位姿, 在笛卡尔空间设计增量 PI 控制器, 控制目标是使当前目标位姿收敛到期望的目标位姿, 实现精确的三维定位。

由于已知目标笛卡尔空间中位姿, 因此, 采用基于位置的视觉控制方案, 其中, 增量 PI 控制器如下所示:

$$\begin{bmatrix} \Delta u_{tx}(k) \\ \Delta u_{ty}(k) \\ \Delta u_{tz}(k) \\ \Delta u_{\theta_x}(k) \\ \Delta u_{\theta_y}(k) \\ \Delta u_{\theta_z}(k) \end{bmatrix} = K_p \begin{bmatrix} \Delta t_x(k) \\ \Delta t_y(k) \\ \Delta t_z(k) \\ \Delta \theta_x(k) \\ \Delta \theta_y(k) \\ \Delta \theta_z(k) \end{bmatrix} - \begin{bmatrix} \Delta t_x(k-1) \\ \Delta t_y(k-1) \\ \Delta t_z(k-1) \\ \Delta \theta_x(k-1) \\ \Delta \theta_y(k-1) \\ \Delta \theta_z(k-1) \end{bmatrix} + K_i \begin{bmatrix} \Delta t_x(k) \\ \Delta t_y(k) \\ \Delta t_z(k) \\ \Delta \theta_x(k) \\ \Delta \theta_y(k) \\ \Delta \theta_z(k) \end{bmatrix} \quad (19)$$

其中, $\Delta u_{tx}(k), \Delta u_{ty}(k), \Delta u_{tz}(k), \Delta u_{\theta_x}(k), \Delta u_{\theta_y}(k), \Delta u_{\theta_z}(k)$ 为增量 PI 控制器的输出, $\Delta t_x(k), \Delta t_y(k), \Delta t_z(k), \Delta \theta_x(k), \Delta \theta_y(k), \Delta \theta_z(k)$ 为当前目标位姿与期望位姿之间的偏差, K_p, K_i 分别为 PI 控制器比例和微分参数, 其均为对角阵。控制系统框图如图 5 所示, 通过图像采集和特征提取, 经三维跟踪实时获得目标位姿后, 通过坐标系变换得到机器人末端的位置增量 $\Delta t_x, \Delta t_y, \Delta t_z$ 和姿态增量 $\Delta \theta_x, \Delta \theta_y, \Delta \theta_z$, 然后, 在笛卡尔坐标系中设计增量 PI 控制器, 其控制输出 $\Delta u_{tx}, \Delta u_{ty}, \Delta u_{tz}, \Delta u_{\theta_x}, \Delta u_{\theta_y}, \Delta u_{\theta_z}$ 作为机器人控制器的输入, 同时, 需设定控制量的最大最小值。

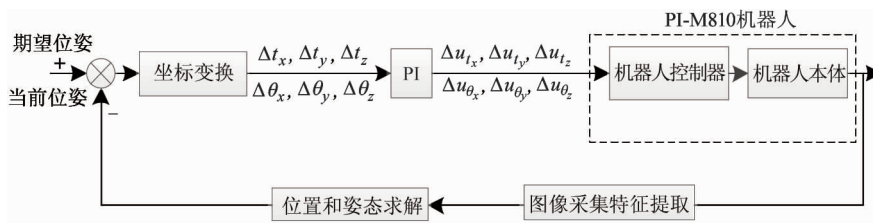


图 5 视觉反馈控制框图

3 实验系统及结果分析

3.1 视觉系统硬件介绍

本系统采用 Eye-to-hand 配置, 三维目标被固定在六自由度机器人的末端, 单目视觉系统被固定在气浮平台上, 如图 6 所示。图 7 为目标的 CAD 实体模型和线框模型。实验系统所涉及的硬件设备介绍如下: 视觉系统包括 Navistar 镜头和 Point Gray CCD (GRAS-50S5M-C), 其分辨率为 $2448(H) \times 2048$

(W); 机器人系统为 PI-M850 六自由度机器人, 包三个平移自由度和三个旋转自由度。由于机器人系统的高精度定位特性, 其重复定位精度在 x, y, z 平移方向分别为 $\pm 2\mu\text{m}, \pm 2\mu\text{m}, \pm 0.5\mu\text{m}$; 在 u, v, w 旋转方向分别为 $\pm 3\mu\text{rad}, \pm 3\mu\text{rad}, \pm 15\mu\text{rad}$ 。因此, 可通过经机器人控制器获得的末端位姿来评价本文跟踪和定位算法的精度。实验过程中采用点到点的运动形式, 通过三维跟踪实时获得目标位姿, 并将此作为反馈量, 引导目标到达期望位姿。

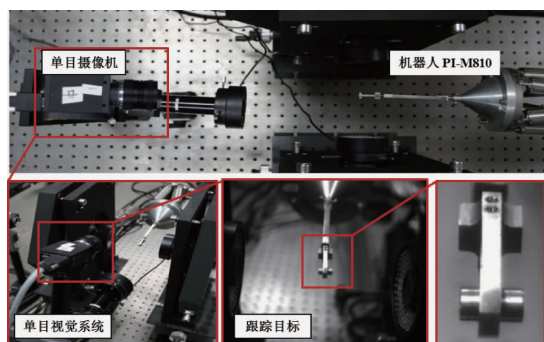


图 6 实验系统配置

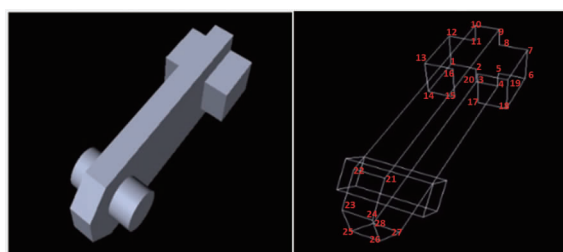


图 7 目标的 CAD 实体模型和线框模型

3.2 手眼标定实验结果

手眼标定实验系统如图 8 所示,棋盘格角点间距为 8mm,角点个数为 81。摄像机内参数 K 采用棋盘标定法^[12],结果如下所示:

$$K = \begin{bmatrix} 3624.57 & 0 & 1259.22 \\ 0 & 3616.82 & 947.81 \\ 0 & 0 & 1 \end{bmatrix} \quad (20)$$

利用标定板标定摄像机坐标和机器人基坐标之间的旋转平移关系 (R_c, T_c) ,结果如下:

$$(R_c, T_c) = \begin{bmatrix} 0.0067 & -0.9702 & -0.2307 & 69.7062 \\ -1.0005 & -0.0005 & -0.0072 & 34.5059 \\ -0.006 & -0.2312 & -0.9707 & 371.1767 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (21)$$

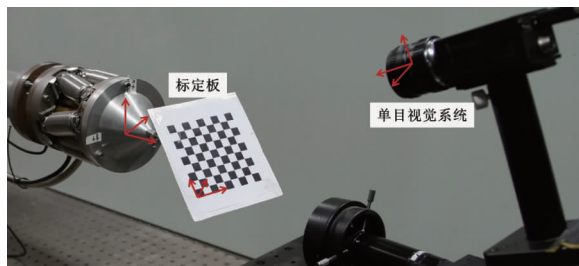


图 8 手眼标定实验系统

驱动目标运动一次,标定目标坐标系和机器人末端坐标系之间的旋转平移关系 (R, T) ,结果如下:

$$(R, T) = \begin{bmatrix} -0.0211 & -0.8091 & -0.5826 & 14.3399 \\ -0.9987 & 0.0496 & -0.0327 & 43.0148 \\ 0.0557 & 0.5814 & -0.8090 & 148.8364 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (22)$$

3.3 目标跟踪和定位实验结果及分析

本文采用 Solidworks 三维软件对目标进行 CAD 建模,通过适当数目的模型控制点,获得虚拟现实建模 (VRML) 格式的模型信息描述文件。Coin3D 是基于 OpenGL 的面向对象的三维图形软件开发包,可用来读取目标模型的表面控制点和面信息,获得所有控制点在模型坐标系中的三维坐标。本文采用目标 CAD 线框模型,如图 7 所示,目标所含直线数为 28。目标初始位姿可采用 perspective-n-point (PNP) 算法求得,并采用正交迭代 (OI) 优化初始位姿^[14]。对于模型可见直线,通过初始帧目标位姿对应的投影矩阵,求得其图像空间的投影,最小化特征偏差获得当前帧目标位姿,并将其应用到下一帧跟踪过程中。目标跟踪过程中, Moving-edge 跟踪算法的投影直线采样间隔为 4 个像素,沿投影轮廓法线方向的边缘点搜索范围为 20 个像素,特征采样单元的数目,即交互矩阵的行数变化如图 9 所示。

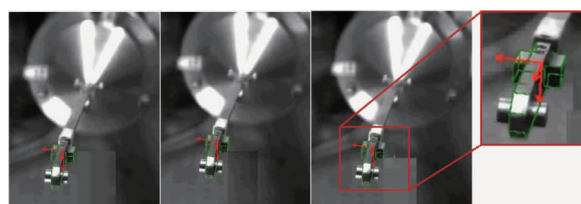


图 9 靶的三维跟踪实验结果

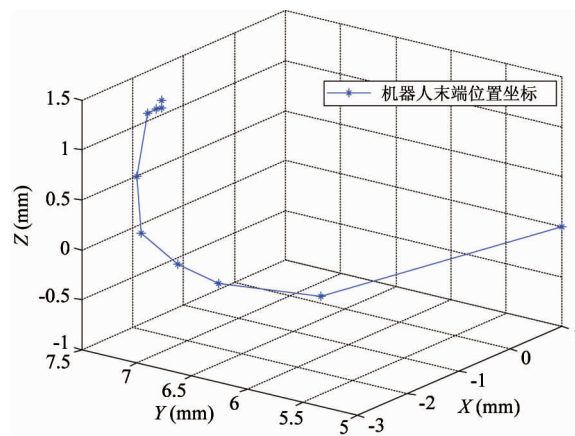
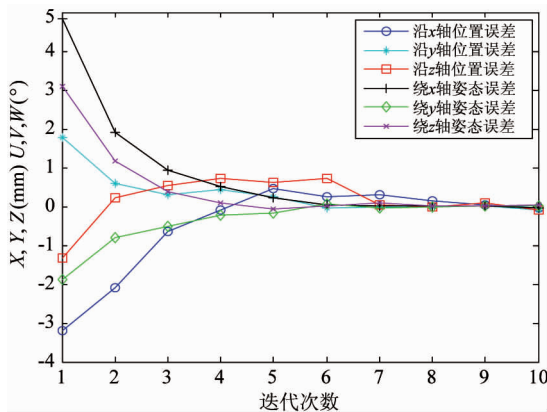


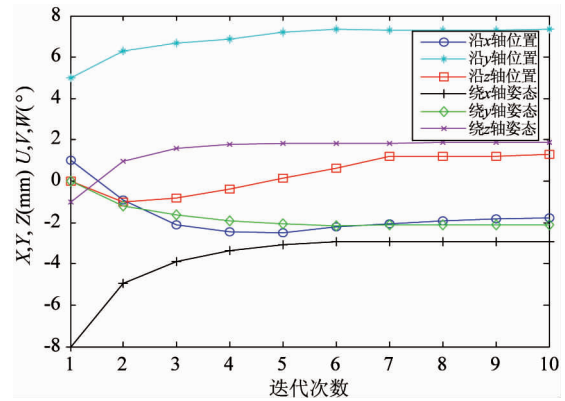
图 10 基于位置的视觉控制中机器人末端在基坐标系中的轨迹曲线

在目标定位过程中,通过目标在摄像机坐标系中的位姿求解机器人末端位姿,然后,设计 PI 控制器,通过视觉反馈定位目标。其中,六个自由度方向增量 PI 比例和积分参数为 $\{0.7, 0.04\}$, $\{0.6, 0.03\}$, $\{0.7, 0.04\}$, $\{0.6, 0.03\}$, $\{0.6, 0.03\}$, $\{0.6, 0.03\}$ 。图 10 为机器人的笛卡尔空间运动轨迹。图 11(a)所示为通过单目视觉位姿估计获得的机器人末端在机器人坐标系中的位姿偏差变化曲线。同时,机器人末端的位姿可由机器人控制器读出,其变化曲线如图 11(b)所示。在三维目标运动过程中,相对于摄像机的可见平面发生变化,可见边缘上的采样点个数也相应地变化,图 12 所示为连续帧中目标可见边缘上采样点个数变化情况。为验证

目标实际控制精度,机器人初始位姿设定为 $\{1\text{mm}, 5\text{mm}, 0, -8^\circ, 0^\circ, -1^\circ\}$, 期望位姿设定为 $\{-2\text{mm}, 7\text{mm}, 1\text{mm}, -3^\circ, -2^\circ, 2^\circ\}$, 由视觉反馈控制收敛位姿为 $\{-1.802\text{mm}, 7.284\text{mm}, 1.280\text{mm}, -2.919^\circ, -2.132^\circ, 1.900^\circ\}$ 。由此可知,基于模型跟踪的定位精度在平移方向控制在 0.2mm 以内,在旋转方向控制在 0.1° 以内。在本实验的初始和期望位姿条件下,目标定位所用时间在 30s 以内,收敛步数控制在 10 步以内,跟踪和定位图像序列如图 13 所示。对于分辨率为 2448×2048 图像序列,本算法平均跟踪时间为 36ms ,符合系统实时性要求。其中,连续帧序列中跟踪时间变化曲线如图 14 所示。



(a) 经单目视觉系统目标跟踪获得



(b) 从机器人控制器获得

图 11 机器人末端位姿变化曲线

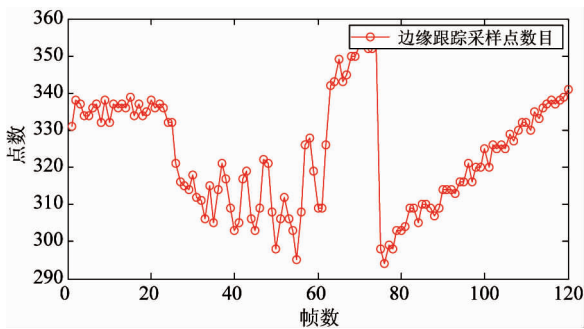


图 12 连续帧目标可见边缘上采样点个数变化曲线

通过跟踪实时获得目标位姿,经坐标变换求得机器人末端在笛卡尔空间中位姿,由于控制器设计在笛卡尔空间,在平移方向和旋转方向独立控制,性能更加稳定、鲁棒。定位误差主要来源于系统的标定误差,包括摄像机内参数误差和坐标系位姿变换标定误差。由于目标坐标系到机器人末端坐标系的变换关系的标定是借助目标本身位姿经一次标定而获得,精度较低,因此,由视觉跟踪系统获得的末端

位姿的起始位姿偏差 $\{-3.193\text{mm}, 1.771\text{mm}, -1.320\text{mm}, 4.833^\circ, -1.869^\circ, 3.083^\circ\}$ 与由机器人自身控制器所得末端起始位姿偏差 $\{-3\text{mm}, 2\text{mm}, 1\text{mm}, 5^\circ, -2^\circ, 3^\circ\}$ 有所不同,但受 PI 控制器的作用,二者均收敛于零值附近。

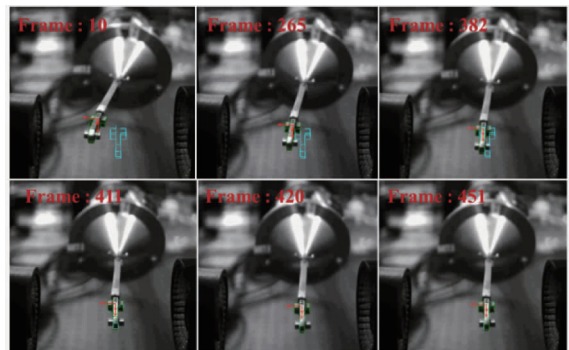


图 13 第 10,265,382,411,420,451 帧目标跟踪和定位的图像序列。目标从初始位姿运动到期望位姿,图像空间的绿色投影对应模型的当前位姿,蓝色投影对应模型的期望位姿

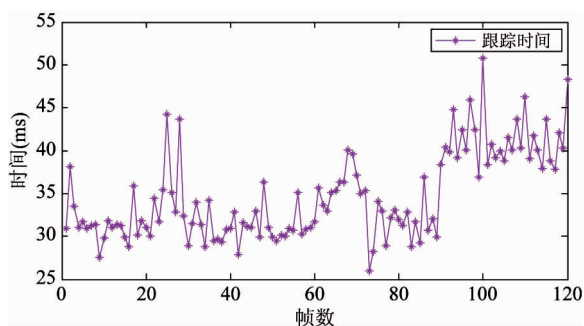


图 14 连续帧中三维跟踪所需时间变化曲线

实验结果表明,沿 Z 轴方向的位置偏差曲线不够平滑,容易超调,相应的定位精度较低。一方面,由于受镜头景深的影响,在 Z 轴方向所容许的运动范围较小,则摄像机坐标系和机器人基坐标系关系在 Z 轴方向的标定结果不精确,或者目标坐标系和机器人基坐标系关系在 Z 轴方向标定误差较大;另一方面,目标边缘特征检测的精确性也影响深度的估计。综上所述,单目视觉系统虽然对目标深度的估计存在一定局限性,但其避免了多目视觉系统的数据融合问题,易于工程实践,有着重要的应用价值。

4 结 论

本文研究了一种单目视觉下基于模型的目标跟踪和定位算法。在跟踪过程中,采用虚拟视觉伺服技术最小化图像空间中的特征偏差,使目标边缘特征收敛于模型投影边缘,并通过在线求解交互矩阵,将图像空间中特征变化转化为笛卡尔空间位姿的变化,进而实现模型与目标的三维对准,实时跟踪三维目标。在定位过程中,通过目标当前位姿,求解机器人末端位姿,然后设计增量 PI 控制器,使机器人当前末端位姿收敛于期望位姿,完成精确的三维定位。实验表明,该算法具有较高的精度和一定的鲁棒性。但是,此种方法适用于少量的、特定目标的跟踪和定位。由于当前目标的三维信息依赖上一帧的位姿信息,需给定初始位姿,在下一步的研究中需要对目标初始位姿进行估计,以及跟踪失败后的位姿重新估计等问题。

参考文献

[1] Lepetit V, Fua P. Monocular model-based 3d tracking of rigid objects: a survey. *Foundations and Trends in Computer Graphics and Vision*, 2005, 1(1): 1-89

[2] Mooser J, You S, Neumann U. Applying robust structure from motion to markerless augmented reality. In: *Proceeding of the IEEE Workshop on Applications of Computer Vision*, Snowbird, USA, 2009. 1-8

[3] Drummond T, Cipolla R. Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 932-946

[4] Vacchetti L, Lepetit V, Fua P. Stable real-time 3d tracking using online and offline information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(10): 1385-1391

[5] Park Y, Lepetit V, Woo W. Multiple 3d object tracking for augmented reality. In: *Proceeding of the IEEE/ACM International Symposium on Mixed and Augmented Reality*, Cambridge, UK, 2008. 117-120

[6] Comport A I, Marchand E, Pressigout M. Real-time markerless tracking for augmented reality the virtual visual servoing framework. *IEEE Transactions on Visualization and Computer Graphics*, 2006, 12(4): 615-628

[7] Antoine Petit E M, Kanani K. A robust model-based tracker combining geometrical and color edge information. In: *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Tokyo, Japan, 2013

[8] Coutard L, Chaumette F, Pflimlin J M. Automatic landing on aircraft carrier by visual servoing. In: *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, USA, 2011. 2843-2848

[9] Vacchetti L, Lepetit V, Fua P. Combining edge and texture information for real-time accurate 3d camera tracking. In: *Proceeding of the IEEE/ACM International Symposium on Mixed and Augmented Reality*, Arlington, USA, 2004. 48-56

[10] Pressigout M, Marchand E. Real-time 3d model-based tracking: Combining edge and texture information. In: *Proceeding of the IEEE International Conference on Robotics and Automation*, Orlando, USA, 2006. 2726-2731

[11] Coutard L, Chaumette F. Visual detection and 3d model-based tracking for landing on an aircraft carrier. In: *Proceeding of the IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011. 1746-1751

[12] Zhang Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334

[13] 马颂德, 张正友. 计算机视觉. 北京: 科学出版社, 1998. 173-192

[14] 徐德, 谭民, 李原. 机器人视觉测量与控制, 北京: 国防工业出版社, 2011. 51-56

Research on CAD model-based tracking and positioning algorithm

Zhang Pengcheng, Xu De

(Research Center of Precision Sensing and Control, Institute of Automation,
Chinese Academy of Sciences, Beijing 100190)

Abstract

Based on the eye-to-hand monocular vision configuration, a 3D target tracking and positioning algorithm was developed, and the high precise 3D tracking and positioning in a Cartesian space was realized. In the tracking procedure, the model-based 3D tracking algorithm is introduced, and the virtual visual servoing (VVS) technique is employed to make the CAD model's projection lines converge to the target edges in the image space. Correspondingly, the precise registration can be achieved between the model and the target in the Cartesian space and the real-time, robust 3D tracking can be realized. In the positioning procedure, the increment PI controller is employed, which can make the current pose converge to the desired one. The experimental results show the positioning accuracy of the algorithm can be controlled within 0.2mm along the translational directions and 0.1° along the rotational directions.

Key words: CAD model, pose estimation, 3D tracking, visual feedback control, visual positioning