

文章编号: 1002-0446(2000)03-0217-07

基于图像矩的运动目标 3D 平动视觉跟踪

林 靖 陈辉堂 王月娟 徐强华 蒋 平

(同济大学电气工程系 上海 200092)

摘要: 区别于图像的简单几何特征, 本文利用图像的全局特征描述子-图像矩特征作为图像特征信息, 实现了基于图像的运动目标 3D 平动的视觉跟踪. 针对任务要求, 本文选取了一组矩特征用以完成任务. 基于所选的矩, 本文给出了矩特征变化量与相对位姿变化量之间的关系矩阵, 即图像雅可比矩阵, 然后利用所推导的图像雅可比矩阵, 设计了由图像反馈与目标运动自适应补偿组成的视觉伺服控制器, 实现了在未知目标成-深度及摄像机焦距的情况下对运动目标的 3D 平动跟踪.

关键词: 视觉伺服; 机器人; 图像矩; 图像雅可比矩阵

中图分类号: TP24

文献标识码: B

1 引言

如同人类视觉系统的作用, 机器人视觉系统赋予了机器人一种高级的感知功能. 视觉系统提供的经量化后的数字图像, 给出了机器人末端执行器与目标之间的相对位姿信息. 利用这个相对位姿信息作为反馈量而构造的机器人视觉伺服系统, 通常分为两类, 即基于位置(position-based)和基于图像(image-based)的控制方式. 这两类控制方式都需要从图像中抽取特征信息. 所不同的是, 基于位置的控制方式其期望给定值以工作空间中的值表示, 所以需要将特征信息投影逆变换到工作空间; 而基于图像的控制方式其期望给定值直接以图像特征信息表示, 所以不需要有一投影逆变换过程. 因此相对基于位置的控制方式而言, 基于图像的控制方式对系统的标定误差不敏感, 具有更高的定位精度, 为多数的视觉伺服系统所采用.

基于图像的控制方式需要根据所选用的视觉特征求取相应的图像雅可比矩阵. 由于一幅数字图像含有大量的数据, 为减少运算量及降低系统的复杂性, 目前多数的视觉伺服系统利用了物体的简单几何特性, 如点、线、圆、面积等, 以简化从图像抽取特征信息的过程^[1-3]. 然而, 这些特征的可靠抽取很大程度上依赖于物体的形状以及周围环境, 并且通常需要在场景中插入特殊定义的视觉信号^[1,2]. 另外, 在图像序列中需要进行特征的匹配, 这需要复杂的算法及大量的运算^[3], 特别是物体的外形及纹理、遮挡、噪声、照明条件等极大程度上影响了匹配的精度. 文[4]指出应用全局的图像特征相对于简单的几何特征, 由于考虑了所有的图像数据, 而不是简单的几个点、几条线, 因此具有更好的鲁棒性. 该文应用了图像傅立叶描述子及图像矩. 然而它需要利用大量的样本图像来训练神经网络, 并利用训练好的神经网络来依据实际的图像特征给出完成定位任务所需的控制量.

虽然计算全局图像特征需要处理整幅图像数据, 然而近年来在图像矩方法的研究方面有了很大的进展, 计算图像矩的运算量有了显著的降低^[5]. 针对利用视觉跟踪运动目标的 3D 平

基金项目: 博士点基金资助项目.

收稿日期: 1999-05-31

动这一任务, 本文选取了一组矩特征用以完成任务. 基于所选的矩, 本文给出了矩特征变化量与相对位姿变化量之间的关系矩阵, 即图像雅可比矩阵, 然后利用所推导的图像雅可比矩阵, 设计了由图像反馈与目标运动自适应补偿组成的视觉伺服控制器, 实现了在未知目标成像深度及摄像机焦距的情况下对运动目标的 3D 平动跟踪. 文中最后给出了仿真与实验例子, 表明了矩特征的有效性.

2 基于矩特征的图像雅可比矩阵

基于图像的视觉伺服控制系统, 其结构如图 1 所示, 控制量 u 为机器人手部摄像机的运动控制量. 为实现视觉伺服控制任务, 首先利用示教盒引导机器人运动到与目标成期望构形, 获取在此期望构形下的期望图像. 定义一完成任务所需的视觉特征集, 从期望图像中抽取期望特征集, 作为视觉伺服控制系统的期望输入. 在实时的控制过程中, 由视觉传感器获取实时采样图像, 从中获取实时视觉特征集, 从而构成视觉反馈, 引导机器人完成任务. 区别于图像的简单几何特征, 本文选用的视觉特征集为全局图像描述——图像矩. 本节首先根据任务要求选取了一组矩特征, 然后基于这一组矩, 给出了矩特征变化量与相对位姿变化量之间的关系矩阵, 即图像雅可比矩阵. 在下一节, 本文将利用所推导的图像雅可比矩阵, 设计由图像反馈与目标运动自适应补偿组成的视觉控制器, 以实现运动目标的 3D 平动跟踪.

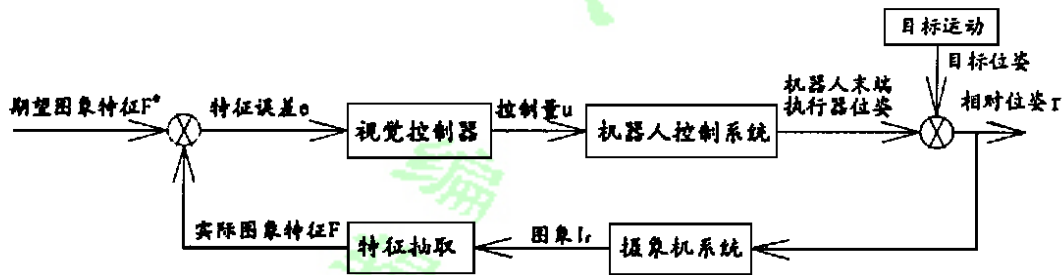


图 1 视觉伺服系统结构

一幅 $N \times M$ 的图像其 $p+q$ 阶几何矩定义为

$$m_{pq} = \sum_{x=1}^N \sum_{y=1}^m x^p y^q g(x, y)$$

x, y 为图像像素的坐标, $g(x, y)$ 为图像像素灰度值. 当 $g(x, y)$ 在 $x-y$ 平面的有限区域非零且分段连续时, 几何矩集合 $\{m_{pq} | p, q = 0, 1, 2, \dots\}$ 与 $g(x, y)$ 是互相唯一确定的^[6]. 对于某一具体的应用, 需要从几何矩集合中找出能提供足够的信息以识别图像的一组矩.

两个一阶矩 $\{m_{10}, m_{01}\}$ 表征了图像的质心位置, 当摄像机与目标之间发生平行于图像平面的相对平动时, 质心位置随之出现变化, 即该运动在一阶矩 $\{m_{10}, m_{01}\}$ 得到反映. 当摄像机与目标之间存在垂直于光轴的相对平动时, 目标在图像上投影面积的大小, 即 m_{00} 出现相应的变化. 因此, 针对运动目标 3D 平动的视觉伺服控制问题, 选取矩集合 $\{m_{00}, m_{10}, m_{01}\}$ 可反映摄像机与目标之间的相对平动.

对于如图 2 所示的摄像机系统 R_s , 具有焦距 f . 假定目标为刚体而且仅做平动运动, 则整个目标的运动完全可以由目标内任一点的运动来确定. 设目标上任一点 P 在 R_s 坐标系中为 r

$= (X_s, Y_s, Z_s)$, 且对应于图像平面中的投影点 p 的坐标为 (x, y) , 满足如下透视关系

$$x = f(X_s/Z_s), y = f(Y_s/Z_s) \quad (1)$$

设背景灰度一致且不变, 机械手运动引起手部摄像机运动, 其平动速度为 $T_c = (T_{xc}, T_{yc}, T_{zc})^T$, 目标的平动速度为 $T_0 = (T_{x0}, T_{y0}, T_{z0})^T$. 令 $T = (T_x, T_y, T_z)^T = T_c - T_0$, 则图像平面光流 $u = \dot{x}, v = \dot{y}$ 可计算如下

$$u = \frac{d}{dt} \left(f \frac{X_s}{Z_s} \right) = f \left[\frac{\dot{X}_s}{Z_s} - \frac{X_s \dot{Z}_s}{Z_s^2} \right] = x \frac{T_z}{Z_s} - f \frac{T_x}{Z'_s} \quad (2)$$

$$v = \frac{d}{dt} \left(f \frac{Y_s}{Z_s} \right) = f \left[\frac{\dot{Y}_s}{Z_s} - \frac{Y_s \dot{Z}_s}{Z_s^2} \right] = y \frac{T_z}{Z_s} - f \frac{T_y}{Z'_s}$$

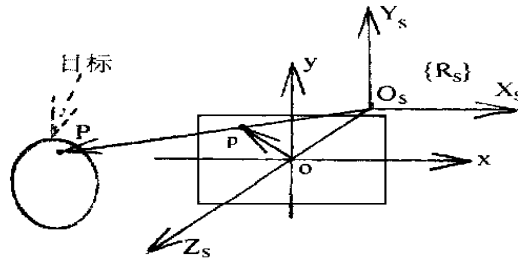


图 2 摄像机坐标系

设 t 及 $t + \Delta t$ 时刻图像平面点 (x, y) 处的灰度值分别为 $g(x, y, t)$ 与 $g(x, y, t + \Delta t)$. 对于 $p + q$ 阶矩 m_{pq} , 假定 Δt 足够小, 有

$$m_{pq}(t + \Delta t) - m_{pq}(t) = \sum_{x=1}^N \sum_{y=1}^M x^p y^q g(x, y, t + \Delta t) - \sum_{x=1}^N \sum_{y=1}^M x^p y^q g(x, y, t)$$

$$= \sum_{x=1}^N \sum_{y=1}^M [x^p y^q g(x, y, t) + x^p y^q g_t(x, y, t) \Delta t - x^p y^q g(x, y, t)]$$

方程两边同除以 Δt , 令 $\Delta t \rightarrow 0$, 有

$$\frac{dm_{pq}(t)}{dt} = \sum_{x=1}^N \sum_{y=1}^M x^p y^q g_t(x, y, t) \quad (3)$$

由光流约束方程可得图像上任一点灰度变化率^[7]

$$g_t(x, y) = -g_x(x, y)u - g_y(x, y)v \quad (4)$$

其中 $g_x(x, y), g_y(x, y), g_t(x, y)$ 分别为像素点的灰度沿 x, y, t 三个方向的梯度值.

将式(2)(4)代入(3)可得

$$\frac{dm_{pq}(t)}{dt} = \sum_{x=1}^N \sum_{y=1}^M [-x^p y^q g_x(x, y, t)u - x^p y^q g_y(x, y, t)v] = J_{(pq)}(t)(\tau_c - \tau_0) \quad (5)$$

其中: $\tau_c = T_c, \tau_0 = T_0$, 分别为摄像机运动控制量和目标未知运动参数;

$J_{(pq)}(t) = [J_{(pq)1}, J_{(pq)2}, J_{(pq)3}]$ 为图像特征雅可比矩阵,

$$J_{(pq)1} = \sum_{x=1}^N \sum_{y=1}^M \frac{f}{Z_s} [x^p y^q g_x(x, y, t)], \quad J_{(pq)2} = \sum_{x=1}^N \sum_{y=1}^M \frac{f}{Z_s} [x^p y^q g_y(x, y, t)],$$

$$J_{(pq)3} = - \sum_{x=1}^N \sum_{y=1}^M \frac{1}{Z_s} [x^{p+1} y^q g_x(x, y, t) + x^p y^{q+1} g_y(x, y, t)]$$

当摄像机光轴保持与工作平面垂直时, 有

$$J_{(pq)}(t) = \frac{1}{Z_s} | J'_{(pq)1} \ J'_{(pq)2} \ J'_{(pq)3} | \text{diag}(f, f, 1) = \frac{1}{Z_s} J'_{(pq)}(t) \text{diag}(f, f, 1)$$

定理 假定图像的背景灰度一致, 设为 d , 当目标落在视场中且不在边界上时, 下面等式成立

$$\sum_{x=1}^N \sum_{y=1}^M x^p y^q g_x(x, y) = \begin{cases} 0 & (p = 0) \\ p \times (m_{(p-1)q}^0 - m_{(p-1)q}) & (p > 0) \end{cases} \quad (6)$$

$$\sum_{x=1}^N \sum_{y=1}^M x^p y^q g_y(x, y) = \begin{cases} 0 & (q = 0) \\ q \times (m_{p(q-1)}^0 - m_{p(q-1)}) & (q > 0) \end{cases} \quad (7)$$

其中, m_{pq}^0 为图像平面上无目标时的 $p+q$ 阶几何矩, $m_{pq}^0 = \sum_{x=1}^N \sum_{y=1}^M (y^q x^p d)$.

证明 考虑连续图像的情况, 设图像平面为 $I = \{(x, y), x \in [0, a], y \in [0, b]\}$. 由于图像的背景灰度一致, 设为 d , 则当目标落在视场中且不在边界上时, 有

$$g(x, 0) = g(x, b) = g(0, y) = g(a, y) = d \quad (x \in [0, a], y \in [0, b])$$

当 $p = 0$ 时,

$$\iint_I x^p y^q g_x(x, y) dx dy = \int_0^b y^q \left(\int_0^a g_x(x, y) dx \right) dy = \int_0^b y^q (g(a, y) - g(0, y)) dy = 0 \quad (8)$$

当 $p > 0$ 时,

$$\begin{aligned} \iint_I x^p y^q g_x(x, y) dx dy &= \int_0^b y^q \int_0^a x^p g_x(x, y) dx dy \\ &= \int_0^b y^q a^p g(a, y) dy - \int_0^b p y^q x^{p-1} g(x, y) dx dy \\ &= \frac{1}{q+1} b^{q+1} a^p d - \int_0^b p y^q x^{p-1} g(x, y) dx dy \\ &= p \left(\int_0^b \int_0^a (y^q x^{p-1} d) dx dy - \int_0^b \int_0^a y^q x^{p-1} g(x, y) dx dy \right) \end{aligned} \quad (9)$$

将连续图像 I 量化成大小为 $N \times M$ 的图像, 且合并式(8)(9)为

$$\sum_{x=1}^N \sum_{y=1}^M x^p y^q g_x(x, y) = \begin{cases} 0 & (p = 0) \\ p \times (m_{(p-1)q}^0 - m_{(p-1)q}) & (p > 0) \end{cases}$$

同理, 可证式(7). 证毕.

选取视觉特征为 $F = (m_{10}, m_{01}, m_{00})^T$, 由上面定理有

$$\frac{dF}{dt} = J(\tau_c - \tau_0) \quad (10)$$

其中, $J = [J_{(10)}^T \ J_{(01)}^T \ J_{(00)}^T]^T = \frac{1}{Z_s} J' \text{diag}(f, f, 1)$

$$J' = \begin{bmatrix} m_{00}^0 - m_{00} & 0 & 3(m_{10} - m_{10}^0) \\ 0 & m_{00}^0 - m_{00} & 3(m_{01} - m_{01}^0) \\ 0 & 0 & 2(m_{00} - m_{00}^0) \end{bmatrix} \quad (11)$$

显然 J' 为上三角形矩阵. 当目标落在图像上时, $m_{00}^0 - m_{00} \neq 0$, J' 可逆, 而且其逆矩阵也为上三角形矩阵.

3 视觉伺服控制器

令控制量 $\tau_c = \tau_f + \hat{\tau}_0$, 由视觉反馈控制 τ_f 和运动目标的自适应补偿 $\hat{\tau}_0$ 构成.

建立 Lyapunov 函数

$$S = \frac{1}{2} e^T Q e + \frac{1}{2} (\hat{\tau}_0 - \tau_0)^T P^{-1} (\hat{\tau}_0 - \tau_0)$$

其中, $e = F - F^*$, 为图像特征误差; F 与 F^* 分别为实时采样图像与期望图像的视觉特征, $Q = \text{diag}(q_1, q_2, q_3)$, $q_1, q_2, q_3 > 0$; P 为对称正定矩阵.

假定目标匀速运动, 有

$$\dot{S} = e^T Q \frac{de}{dt} + \hat{\tau}_0^T P^{-1} (\hat{\tau}_0 - \tau_0) = e^T Q J \tau_f + (\hat{\tau}_0^T + e^T Q J P) P^{-1} (\hat{\tau}_0 - \tau_0)$$

选取运动目标自适应跟踪控制律

$$\tau_f = - K J^{-1} e, \quad \hat{\tau}_0 = - P J^T Q e \tag{12}$$

其中 $K = \text{diag}(k_1, k_2, k_3)$, $k_1, k_2, k_3 > 0$. 由于 J' 为上三角形矩阵, 当目标在控制过程中不超出视场时, 易证矩阵 $Q J K J^{-1}$ 的特征值大于零, 即 $Q J K J^{-1}$ 为正定矩阵. 从而可得 $\dot{S} = - e^T Q J K J^{-1} e \leq 0$, 而且 $\dot{S} = 0$ 当且仅当 $e = 0$, 系统渐近稳定.

为避免在线计算逆矩阵即目标成像深度, 下面对控制律(12)进行简化.

假定对应于期望位姿构形的图像雅可比矩阵为 J_d , 成像深度为 Z_d , 零阶矩为 $(m_{00})_d$. 假定在某时刻, 成像深度为 Z_s , 则由式(1)有

$$x = f \left(\frac{X_s}{Z_s} \right) = \frac{Z_d}{Z_s} f \left(\frac{X_s}{Z_d} \right), \quad y = f \left(\frac{Y_s}{Z_s} \right) = \frac{Z_d}{Z_s} f \left(\frac{Y_s}{Z_d} \right)$$

显然, 在成像深度为 Z_s 时所采集图像相对于期望图像, 其图像平面 X, Y 方向各放大了 $\frac{Z_d}{Z_s}$ 倍. 从而有文[7]: $m_{00} = \left(\frac{Z_d}{Z_s}\right)^2 (m_{00})_d$, 即 $Z_s = \sqrt{\frac{(m_{00})_d}{m_{00}}} \times Z_d$. 由于不难验证 $Q J K J_d^{-1}$ 为正定的, 从而可选用如下控制律

$$\begin{aligned} \tau_f &= - K J_d^{-1} e = - \text{diag}(k'_1, k'_2, k'_3) J_d'^{-1} e = - K' J_d'^{-1} e \\ \hat{\tau}_0 &= - \frac{1}{Z_s} P \text{diag}(f, f, 1) J'^T Q e = - \frac{1}{m_{00} P'} J'^T Q e \end{aligned} \tag{13}$$

其中

$$k'_1 = \frac{k_1 Z_d}{f}, \quad k'_2 = \frac{k_2 Z_d}{f}, \quad k'_3 = k_3 Z_d, \quad P' = \frac{1}{Z_d (m_{00})_d} P \text{diag}(f, f, 1),$$

可得摄像机运动控制量

$$\tau_c = - K' J_d'^{-1} e - P \int_0^t m_{00} J'^T Q e \tag{14}$$

由式(11)可知, J'_d, J'^T 可以利用期望图像与实时采样图像的视觉特征 F^* 与 F 来计算. 因此, 控制律(14)只与实时采样的图像及期望图像矩特征向量有关, 在控制过程中不需要知道物体在期望位置及实际位置时的成像深度, 而且也不需要知道摄像机焦距.

4 仿真结果

利用控制律(14), 由固连于机器人末端执行器上的摄像机引导机器人跟踪如图 3 所示的无明显特征的不规则目标. 设摄像机焦距为 16mm, 物体在摄像机标架中的期望高度为 160mm. 视觉伺服控制周期为 100ms, 目标在摄像机标架中以 $T_{x0} = 7\text{cm/s}$, $T_{y0} = 8\text{cm/s}$, $T_{z0} = 9\text{cm/s}$ 速度在 3 维空间做 3D 平动. 期望图像二值化后的矩特征值约为 $(m_{10}, m_{01}, m_{00}) = (-$

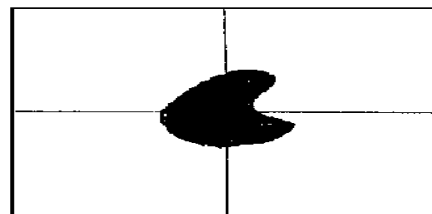


图 3 运动目标期望图像

311, - 2569, 3816), 取控制参数为

$$q_1 = q_2 = 1.0, q_3 = 3000.0, k'_1 = k'_2 = 4.0, k'_3 = 600.0$$

$$P' = \text{diag}(p'_1, p'_2, p'_3), p'_1 = p'_2 = 5.0 \times 10^{-10}, p'_3 = 5.0 \times 10^{-12}$$

控制结果如图4所示, 各轴的最终跟踪误差不超过2mm, 而在期望高度时图像上一个像素所能反映的距离为10mm, 可见利用矩特征作为视觉特征信息可获得亚像素级的控制精度. 基于简单几何特征的控制通常在z轴方向上有较大的误差^[3], 而本文所提的基于矩特征的控制z轴方向也具有较高的精度.

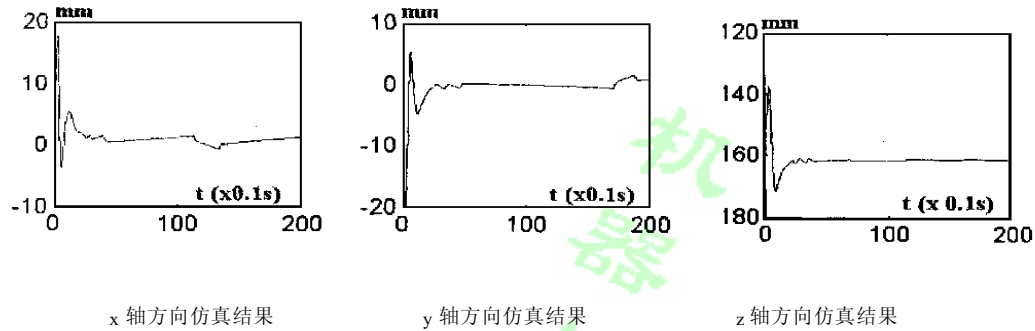


图4 摄像机引导机器人跟踪结果

我们对各种复杂形状的物体进行过仿真实验, 结果显示所提控制方案具有很好的广泛性和一般性, 而不象利用物体简单几何特征的方法那样, 对于不同形状的物体需要人为选择不同的特征, 以便能充分反映物体外形.

5 结束语

针对3D平动视觉伺服控制问题, 本文利用图像的全局特征描述子-图像矩特征作为图像特征信息, 从而避免了复杂的特征匹配, 而且不需要借助特定的方法来标识目标. 根据任务要求, 本文首先选取了一组矩特征用以完成任务. 基于所选的一组矩, 本文给出了矩特征变化量与相对位姿变化量之间的关系矩阵, 即图像雅可比矩阵, 然后利用所推导的图像雅可比矩阵, 设计了由图像反馈与目标运动自适应补偿组成的视觉伺服控制器, 实现了在未知目标成像深度及摄像机焦距的情况下对运动目标的3D平动跟踪. 从仿真结果可看出, 在视觉伺服控制中以矩做为视觉特征可获得较高的控制精度.

参 考 文 献

- 1 Feddema J T, Lee C S G, Mitchell O R. Weighted Selection of Image Features for Resolved Rate Visual Feedback Control, IEEE Trans, on Robotics and Automation, 1991, 7(1): 31- 47
- 2 Hashimoto K, Kimoto T, Ebine T, Kimura H. Manipulator Control with Image-Based Visual Servo. in Proc. of the 1991 IEEE Int'l Conf, on Robotics and Automation, Sacramento, California, April 1991, 2267- 2272
- 3 Papanikolopoulos N P, Khosla K P. Feature Based Robotic Visual Tracking of 3-D Translational Motion. in Proc of the 30th on Decision and Control, 1991. 1877- 1882
- 4 Wells G, Venaille C, Torras C. Promising Research Vision-based Robot Positioning Using Neural Networks. Image and Vision Computing, 1996, 14(10): 715- 732
- 5 Martinez J, Thomas F. A Reformation of Gray-Level Image Geometric Moment Computation for Real-time Application.

- Proc of IEEE Int. Conf on Robotics and Automation, 1996. 2315- 2320
- 6 Prokop R J, Reeves A P. A Survey of Moment-based Object Representation and Recognition, *CVGIP: Graphical Models and Image Processing*, 1992, **54**(5): 438- 460
- 7 李介谷. 计算机视觉的理论和应用. 上海交通大学出版社, 1991

IMAGE MOMENTS BASED VISUAL TRACKING OF 3D TRANSLATIONAL MOTION

LIN Jing CHEN Hui-tang WANG Yue-juan XU Qiang-hua JIANG Ping

(Dept. of E. E., Tongji University, Shanghai, 20092)

Abstract This paper describes a visual tracking method which uses image moment as image feature for tracking a target that moves in 3D with translational motion. According to the specific tasks, a set of moment is selected as image feature. Then, the moment-based Jacobian is deduced. With the moment-based Jacobian, the visual servoing controller composed of image-based feedback and adaptive motion compensation is designed to track the moving target. In the tracking process, the knowledge of camera focus and depth of object is unnecessary. The simulation results show that using image moment as image feature can avoid the complex feature matching process, and acquire a satisfied tracking accuracy.

Keywords: Visual servoing, robot, image moments, image Jacobian

作者简介:

林 靖 (1973-), 男, 博士生. 研究领域: 机器人视觉.

陈辉堂 (1934-), 男, 博士生导师. 研究领域: 机器人控制与智能机器人.

王月娟 (1936-), 女, 博士生导师. 研究领域: 机器人控制与智能机器人.