

基于双目视觉的运动目标跟踪与测量

张娟,潘建寿,吴亚鹏,刘继艳

ZHANG Juan, PAN Jian-shou, WU Ya-peng, LIU Ji-yan

西北大学 信息科学与技术学院, 西安 710127

Department of Information Science and Technology, Northwest University, Xi'an 710127, China

E-mail: zhangjuan_sj@sina.com

ZHANG Juan, PAN Jian-shou, WU Ya-peng, et al. Tracking and measurement of moving object in binocular stereo vision. *Computer Engineering and Applications*, 2009, 45(25): 191-194.

Abstract: This paper designs an experimental system for tracking and measurement of moving object in binocular stereo vision based on the researching of the tracking combination algorithm of the position-forecasting of moving object in the Mean Shift and the localization algorithm of spatial points with binocular vision. Then, 3D coordinate of the centroid to the moving object is extracted, and the depth information and velocity are measured by using the system.

Key words: binocular vision; localization algorithm; tracking of moving object

摘要: 在研究融合运动目标位置预测的 Mean Shift 跟踪算法和双目立体视觉中的空间点定位算法的基础上, 基于双目视觉设计了双目立体视觉运动目标跟踪和测量系统, 并在所进行的跟踪与测量实验中, 提取了运动目标质心的三维坐标序列, 实现了对目标深度和速度的测量。

关键词: 双目视觉; 定位算法; 运动目标跟踪

DOI: 10.3778/j.issn.1002-8331.2009.25.059 文章编号: 1002-8331(2009)25-0191-04 文献标识码: A 中图分类号: TP391

1 引言

立体视觉技术是计算机视觉研究的核心内容。双目立体视觉系统由两台摄像机组成, 直接模拟人类双眼处理景物的方式和立体感知过程, 简便可靠, 具有极大的应用前景。

运动目标跟踪技术是计算机视觉研究的另一个重要内容, 在机器人视觉导航、公共场景监控、军事视觉制导、智能交通等许多方面都有广泛的应用^[1-3]。其实质是通过对摄像机拍摄的视频序列进行分析, 提取场景中运动目标的特征, 通过匹配在每一帧图像中对其进行定位。而目前对该技术的实际研究中, 大多数工作都是基于单目视觉而展开的, 单目视觉的信息量小, 每次只需处理一幅图像, 运算速度相对较快, 但由于在图像投影过程中, 丢失了实际场景的三维信息, 因此得到的目标运动信息只是相对的; 双目立体视觉通过提取立体图像对之间的视差信息, 能够在一定程度上恢复实际场景的三维信息, 因此当需要对目标进行实际的三维位置和深度(即目标与观察者之间的距离)测量时, 双目立体视觉有其不可替代的优点。

2 融合运动目标位置预测的 Mean Shift 跟踪算法

Mean Shift 算法^[4-6]是目前比较流行的运动目标跟踪算法, 它利用梯度优化方法来减少特征搜索匹配的时间, 实现对目标的快速定位, 同时利用 Bhattacharyya 系数作为对目标模板和候

选目标的相似性测度, 完成特征的匹配。

但在运动目标跟踪过程中, Mean Shift 跟踪算法要求当前帧目标搜索区域与上一帧目标模板区域之间具有一定的重叠部分。而当被跟踪目标快速、变速运动或目标发生遮挡时, 此要求将不一定得到满足(重叠部分较少或不存在重叠)。在这种情况下, 由于当前帧搜索区域范围内缺少足够的运动目标信息或者根本没有了运动目标的信息, 目标模板也就不知道要移动到何处去进行匹配了。另外, Mean Shift 算法在跟踪过程中只关心候选目标和目标模板在颜色分布上的相似性, 而没有利用其运动信息。基于上述分析, 为改善目标快速运动以及目标被遮挡情况下的跟踪效果, 将卡尔曼滤波器^[6-7]引入其中, 从而修改了文献[8]中的算法步骤, 并完成了实验验证。

2.1 Mean Shift 算法过程修改

步骤(1)修改为: 确定被跟踪目标, 选取目标模板, 计算目标模板在上一帧 y_0 处的概率密度 $\{\hat{q}_u(\hat{y}_0)\}_{u=1 \dots m}$, 同时初始化卡尔曼滤波器参数: 确定目标模板中心位置, 目标初始速度可以假设为 0。

步骤(2)修改为: 使用卡尔曼滤波器, 获得目标在当前帧的预测位置 y_0' , 并计算在 y_0' 处的概率密度 $\{\hat{p}_u(\hat{y}_0')\}_{u=1 \dots m}$ 和 Bhattacharyya 系数 $\rho[\hat{p}(y_0'), \hat{q}]$;

基金项目: 陕西省自然科学基金(the Natural Science Foundation of Shaanxi Province of China under Grant No.F0306)。

作者简介: 张娟(1984-), 女, 硕士研究生, 主要研究领域为信号处理与模式识别; 潘建寿(1953-), 男, 教授, 主要研究领域为信号处理与模式识别、计算机视觉及应用; 吴亚鹏(1982-), 男, 硕士研究生, 主要研究领域为计算机视觉与模式识别。

收稿日期: 2008-05-14 修回日期: 2008-07-28

步骤(3)~(7)同Mean Shift 算法;确定目标最终位置后,提取当前目标模板灰度信息,更新卡尔曼滤波器参数进行下一次跟踪。

2.2 遮挡情况判断

Bhattacharyya 系数表征的是当前候选目标区域灰度信息与目标模板灰度信息的相似度,系数值越大说明候选目标与目标模板就越匹配。在运动过程中,当目标被其他物体遮挡时,候选目标与目标模板匹配程度会减小,Bhattacharyya 系数自然会减小。利用 Bhattacharyya 系数值的变化作为判断跟踪过程中是否发生遮挡情况的准则,具体算法过程为:给定阈值 T ,计算第 k 帧卡尔曼滤波器预测位置 $Y_k=[xw_k, yw_k]$ 处的 Mean Shift 向量 $\hat{\rho}_u(Y_k), \hat{q}_u$ 与 Bhattacharyya 系数 $\rho(\hat{p}_u(Y_k), \hat{q}_u)$;比较 $\rho(\hat{p}_u(Y_k), \hat{q}_u)$ 与给定阈值 T 的大小,若 $\rho(\hat{p}_u(Y_k), \hat{q}_u) \geq T$ 则认为发生遮挡,目标直接定位在预测位置 $Y_k=[xw_k, yw_k]$ 处,反之,则认为正常跟踪,继续执行 Mean Shift 算法步骤(3)~步骤(7)。

3 双目立体视觉中的空间点定位

采用“TSAI”两步标定算法^[9-10]分别对左右摄像机进行标定,获得参数,具体如下:

左摄像机内参数:焦距 f_l ,实际图像中心坐标 (ul_0, vl_0)

左摄像机外参数:旋转矩阵 R_l ,平移向量 T_l

右摄像机内参数:焦距 f_r ,实际图像中心坐标 (ur_0, vr_0)

右摄像机外参数:旋转矩阵 R_r ,平移向量 T_r

需要注意的是,此时获得的左右摄像机外参数都是相对于各自标定过程中设定的世界坐标系而言的。为了确定双目视场内空间点的三维坐标,根据人类视觉习惯,将左右摄像机统一到同一世界坐标系下。现假定左摄像机光心为世界坐标原点,即观察者处于左摄像机光心位置,则右摄像机坐标系 $o_r-x_yz_r$ 与此时的世界坐标系 $o-xyz$ 可通过空间转换矩阵 $[R T]$ 表示为:

$$\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = [R T] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 & t_x \\ r_4 & r_5 & r_6 & t_y \\ r_7 & r_8 & r_9 & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (1)$$

其中 R, T 为左右摄像机相对于同一世界坐标系的旋转矩阵和平移向量。其中, $R=R_r R_l^{-1}; T=T_r-R_r R_l^{-1} T_l$ 。由计算机图像坐标 (u, v) 和世界坐标 (X_w, Y_w, Z_w) 之间的关系^[11],结合式(1)获得双目视觉下的空间点三维坐标的提取公式,具体推导过程如下:

(1)将左右摄像机内参数分别代入矩阵

$$A_l = \begin{bmatrix} f_l/dx & 0 & ul_0 \\ 0 & f_l/dy & vl_0 \\ 0 & 0 & 1 \end{bmatrix}; A_r = \begin{bmatrix} f_r/dx & 0 & ur_0 \\ 0 & f_r/dy & vr_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

其中 dx, dy 分别为图像在 x, y 方向上单位像素的尺寸大小。例如,CCD 芯片成像面积大小为 $4.8 \text{ mm} \times 3.6 \text{ mm}$,图像分辨率大小为 768×576 像素,则 $dx=4.8 \text{ mm}/768=0.00625; dy=3.6 \text{ mm}/576=0.00625$ 。

(2)设空间点 P 的世界坐标为 (X_w, Y_w, Z_w) (世界坐标系原点为左摄像机光心);在左右摄像机图像上的投影点分别为 $(ul, vl), (ur, vr)$,得到如下两组矩阵方程:

$$z_l \begin{bmatrix} ul \\ vl \\ 1 \end{bmatrix} = H_l \times \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix}; z_r \begin{bmatrix} ur \\ vr \\ 1 \end{bmatrix} = H_r \times \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \quad (3)$$

其中,

$$H_l = A_l \times E = \begin{bmatrix} hl_{11} & hl_{12} & hl_{13} & hl_{14} \\ hl_{21} & hl_{22} & hl_{23} & hl_{24} \\ hl_{31} & hl_{32} & hl_{33} & hl_{34} \end{bmatrix} \quad (4)$$

$$H_r = A_r \times [R T] = \begin{bmatrix} hr_{11} & hr_{12} & hr_{13} & hr_{14} \\ hr_{21} & hr_{22} & hr_{23} & hr_{24} \\ hr_{31} & hr_{32} & hr_{33} & hr_{34} \end{bmatrix} \quad (5)$$

E 为 4 阶单位矩阵, H_l 和 H_r 的所有分量均为已知,将式(5)展开并分别消去 z_l, z_r 可得:

$$\begin{aligned} &(u_l hl_{31} - hl_{11})X_w + (u_l hl_{32} - hl_{12})Y_w + (u_l hl_{33} - hl_{13})Z_w = hl_{14} - u_l hl_{34} \\ &(v_l hl_{31} - hl_{21})X_w + (v_l hl_{32} - hl_{22})Y_w + (v_l hl_{33} - hl_{23})Z_w = hl_{24} - v_l hl_{34} \\ &(u_r hr_{31} - hr_{11})X_w + (u_r hr_{32} - hr_{12})Y_w + (u_r hr_{33} - hr_{13})Z_w = hr_{14} - u_r hr_{34} \\ &(v_r hr_{31} - hr_{21})X_w + (v_r hr_{32} - hr_{22})Y_w + (v_r hr_{33} - hr_{23})Z_w = hr_{24} - v_r hr_{34} \end{aligned} \quad (6)$$

取

$$M = \begin{bmatrix} u_l hl_{31} - hl_{11} & u_l hl_{32} - hl_{12} & u_l hl_{33} - hl_{13} \\ v_l hl_{31} - hl_{21} & v_l hl_{32} - hl_{22} & v_l hl_{33} - hl_{23} \\ u_r hr_{31} - hr_{11} & u_r hr_{32} - hr_{12} & u_r hr_{33} - hr_{13} \\ v_r hr_{31} - hr_{21} & v_r hr_{32} - hr_{22} & v_r hr_{33} - hr_{23} \end{bmatrix}; U = \begin{bmatrix} hl_{14} - u_l hl_{34} \\ hl_{24} - v_l hl_{34} \\ hr_{14} - u_r hr_{34} \\ hr_{24} - v_r hr_{34} \end{bmatrix}$$

得 $M[X_w Y_w Z_w]^T = U$,利用最小二乘法计算得空间目标点 P 在以左摄像机光心为原点的世界坐标系下的三维坐标:

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = (M^T \times M)^{-1} \times M^T \times U \quad (7)$$

4 双目系统的设计与实现

4.1 运动目标的三维跟踪

由第 2 章和第 3 章的讨论,可以提取运动目标质心在一段时间内的二维坐标序列和某一时刻的三维坐标。但为了实现目标的三维跟踪,必须将第 2, 3 章系统地结合,提取一段时间内运动目标质心的三维坐标序列。跟踪系统主要功能模块关系如图 1,系统对第 k 帧图像的处理程序流程如图 2,具体描述如下:

(1)初始化模块(initialize_data()):读入左右视频序列所有帧图像和立体视觉标定参数 (R, T) ;提取第 k 帧左右图像对(get_current_frame_data())。

(2)运动目标检测与跟踪模块:对左右视频图像序列分别进行运动目标检测(get_objects_in_frame()),记录目标(ObjTrack(001, 002, …))与目标个数(Count OBJ)并提取目标特征并与前一帧(第 $k-1$ 帧)图像做匹配,实现左右视频内的运动目标跟踪,记录目标质心在二维图像中的运动轨迹。

(3)立体匹配与三维坐标提取模块:以左摄像机为主摄像机,以左视场(主视场)下运动目标质心作为匹配点;右摄像机为副摄像机,右视场(副视场)内运动目标质心作为候选匹配点。对左、右(主、副)视场内的运动目标质心进行匹配(Match(ObjTrackL, ObjTrackR)),对于匹配成功的目标,系统保存其质心在左右视频序列中对应的二维坐标,并提取该目标质心的三维坐标;对于未匹配目标,系统也暂时保存其在左右视频中的二维跟踪坐标。

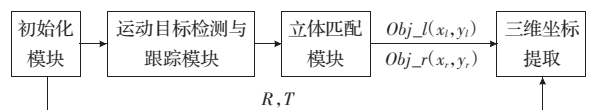


图 1 系统功能模块关系

4.2 运动目标的深度与速度测量

根据第 3 章的讨论,在双目视场下可以获得目标在真实场

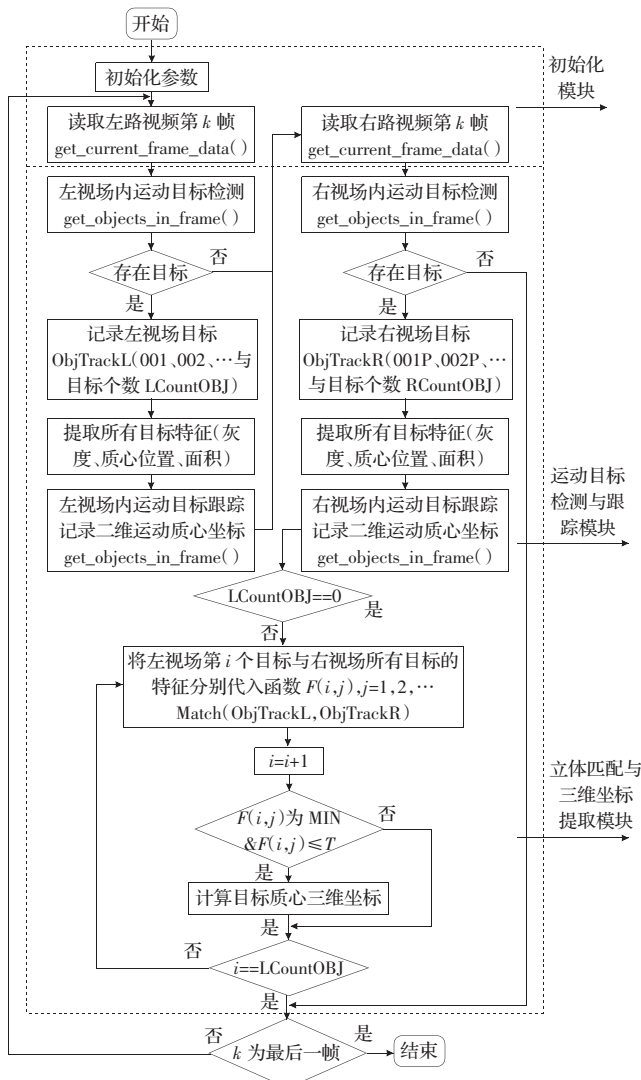
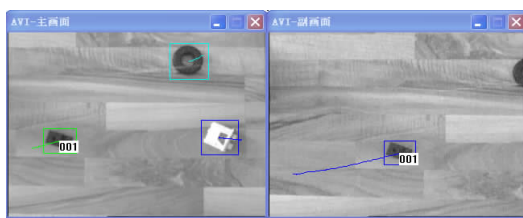
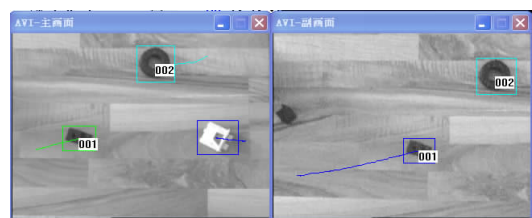


图2 跟踪系统程序流程图



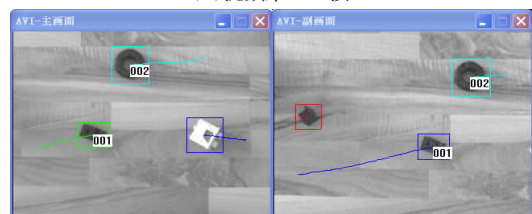
(a) 视频第 210 帧



(b) 视频第 250 帧



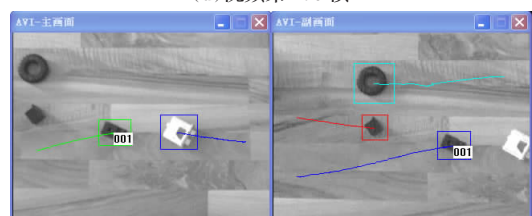
(c) 视频第 260 帧



(d) 视频第 270 帧



(e) 视频第 280 帧



(f) 视频第 380 帧

图3 主、副摄像机视频序列跟踪结果

景中的空间坐标位置。结合空间欧氏几何知识,运动目标的深度测量可由公式(8)计算:

$$d_k = \sqrt{x_k^2 + y_k^2 + z_k^2} \quad (8)$$

其中, d_k 表示第 k 帧目标质心的深度(mm); (x_k, y_k, z_k) 表示第 k 帧目标质心的三维坐标。

根据物理学中平均速度的定义,要实现对运动目标实际速度的测量需要获得两个信息量,一个是运动目标运动的时间间隔 T ,另一个是运动目标运动的路程 S 。对于第 k 帧到第 l 帧的时间间隔 T_{l-k} , 可以通过计算目标运动期间所经历视频序列的帧数得到,即 $T_{l-k} = |l-k| \times \Delta t$

对于运动目标从第 k 帧到第 l 帧的运动路程 S_{k-l} , 同样根据空间欧氏几何知识来计算,即

$$S_{k-l} = \sqrt{(x_k - x_l)^2 + (y_k - y_l)^2 + (z_k - z_l)^2} \quad (10)$$

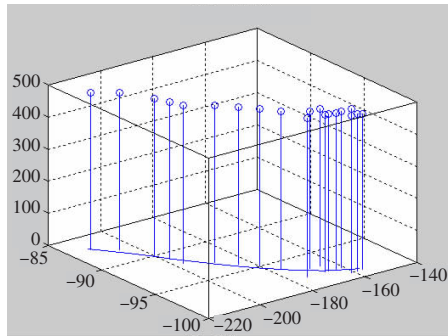
综上,可以通过公式(11)获得目标在第 k 帧至第 l 帧之间的平均速度(mm/s)

$$v_{k-l} = \frac{S}{T} = \frac{\sqrt{(x_k - x_l)^2 + (y_k - y_l)^2 + (z_k - z_l)^2}}{|l-k| \times \Delta t} \quad (11)$$

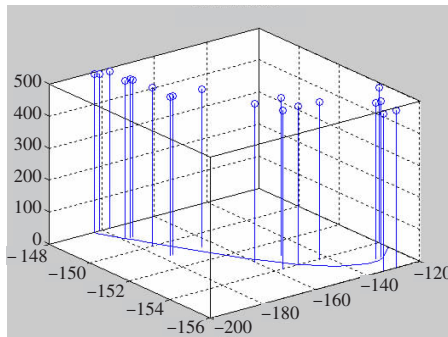
5 实验结果与分析

在室内环境下以拉动的圆形、方形工件作为运动目标,主、副摄像机视频序列跟踪结果如图3。其中,图3(a)显示了第210帧时刻的跟踪结果,此时目标001(长方体黑色工件)已匹配成功,系统保存其质心、三维坐标序列并在主、副视场中描绘其二维运动轨迹;由于存在视差,目标002(空心圆黑色工件)并未完全进入公共视场,仅在主视场中已被检测并跟踪,此时系统预先为其开辟了内存空间以保留其质心、三维坐标序列。图3(b)显示了第250帧时刻的跟踪结果,此时目标002在副视场中被检测到并与主视场内的对应目标相匹配,由于002已进入公共视场,此时系统开始记录其三维运动序列。图3(c)、(d)、(e)显示了第260、270、280帧的跟踪结果,这段时间内系统对目标001和002连续跟踪。图3(f)显示了第380帧的

跟踪结果,此时目标 002 退出公共视场,而新目标已经出现。图 4 显示了目标 001 和目标 002 在空间的三维运动轨迹。表 1 给出了实验对运动目标的深度和一定时间段内目标的运动速度的测量结果。



(a)目标 001 空间运动轨迹



(b)目标 002 空间运动轨迹

图 4 目标空间运动轨迹

根据实际观察,并使用刻度尺和坐标纸定点对目标深度和速度进行了测量和估计,结果与该文获得的测量结果基本一致。此外,从获得的目标 001 和 002 的三维坐标序列变化规律上看,由于实验中目标 001 和 002 都是在地板上做近似直线运动,因此在它们的三维坐标序列中, z 坐标应该基本不变(目标高度没有变化), y 坐标变化不大(目标做近似直线运动), x 坐标变化显著,而图 4 符合上述分析。

6 结论

将视频序列中运动目标跟踪与双目立体视觉相结合,在双目环境下实现了空间运动目标的跟踪,测量了目标的深度和运动速度,获得了预期的实验结果。通过对该问题的研究表明:应用双目立体视觉技术能够将传统的运动目标跟踪从二维图像拓展至三维空间,具有重要的实际应用意义。完善系统算法,进一步扩大系统测量范围,研究复杂场景下运动目标的跟踪与三维测量是今后工作的重点。

(上接 153 页)

参考文献:

- [1] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995.
- [2] Cristiniani N, Shawe-Taylor J. An introduction to support vector machines[M]. Cambridge: Cambridge University Press, 2000.
- [3] 邓乃扬,田英杰.数据挖掘中的新方法—支持向量机[M].北京:科学出版社,2004.
- [4] Vapnik V N. Statistical learning theory[M]. New York: Wiley, 1998.
- [5] Zhang Xue-gong. Using class-center vectors to build support vector

表 1 目标 001、002 质心测距、测速结果

目标 001		目标 002	
视频序列帧号	深度	视频序列帧号	深度
210	540.937 1	252	522.820 2
226	536.842 6	260	531.091 3
242	538.815 4	268	533.250 4
258	531.781 9	276	535.714 5
274	529.426 1	284	540.741 2
290	527.323 4	292	540.453 5
306	525.624 4	300	546.950 6
322	521.889 8	308	541.550 6
338	524.254 7	316	544.590 7
354	523.644 0	324	545.244 0
370	520.236 5	332	546.298 5
平均速度=8.82 mm		平均速度=19.934 mm	

参考文献:

- [1] Collins R T, Lipton A J, Kanade T. Introduction to the special section on video surveillance[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(8): 745-746.
- [2] Wei G Y, Li C F. Design and implementation of visual servoing system for realistic air target tracking[C]//Proceedings of IEEE International Conference on Robotics and Automation, 2001, 1: 229-234.
- [3] 邵文坤,黄爱民,韦庆.目标跟踪方法综述[J].影像技术,2006,18(1):17-20.
- [4] Comaniciu D, Meer P. Mean shift analysis and application[C]//Proceedings of the 7th IEEE International Conference: Computer Vision, 1999.
- [5] Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using Mean Shift[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2000.
- [6] Bar-Shalom Y, Fortmann T. Tracking and data association[M]. Boston: Academic Press, 1988: 106.
- [7] 王江涛.遮挡情况下基于 Kalman 均值偏移的目标跟踪[J].系统仿真学报,2007,19(18).
- [8] 王宾,潘建寿,王琳,等.基于多特征融合的均值偏移视频目标跟踪算法[J].小型微型计算机系统,2006(9).
- [9] Tsai R Y. Accuracy analysis and prediction for 3d robotics vision metrology, Technical Report RC 11348[R]. IBM Research.
- [10] Tsai R Y. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses[J]. IEEE Trans Rob Autom, 1987, 3: 323-344.
- [11] 马颂德,张正友.计算机视觉——计算理论与算法基础[M].北京:科学出版社,1998.
- [12] Lin Chun-fu, Wang Sheng-de. Fuzzy support vector machines[C]//Wisconsin: IEEE Conference on Neural Networks for Signal Processing, 1999: 3-11.
- [13] Lin Chun-fu, Wang Sheng-de. Fuzzy support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2): 464-471.
- [14] 张桂香,费岚,等.非均衡数据去噪模糊支持向量机新方法[J].计算机工程与应用,2008,44(16): 142-144.
- [15] 裴继红,范九伦,谢维信.一种新的高效软聚类方法:截集模糊 C-均值(S2FCM)聚类算法[J].电子学报,1998,26(2): 83-86.
- [16] Bezdek J C. Pattern recognition with fuzzy objective function algorithms[M]. New York: Plenum Press, 1981.
- [17] 边肇祺,张学工.模式识别[M].北京:清华大学出版社,2007.