

基于多尺度特征提取的均值漂移目标跟踪算法

孔 军^{1,2,3,4}, 汤心溢², 蒋 敏¹, 葛运建³

(1. 江南大学物联网工程学院, 江苏 无锡 214122; 2. 中国科学院上海技术物理研究所, 上海 200083;
3. 中国科学院合肥智能机械研究所, 合肥 230031; 4. 中国科学院研究生院, 北京 100049)

摘要: 为在图像对比度较低、相似目标过多等情况下较好地实现目标跟踪, 提出一种基于多尺度特征提取的均值漂移跟踪算法。前一帧目标区域的特征点经匹配得到后续帧目标区域的特征点, 利用所得特征点集的中心坐标修正均值漂移搜索窗位置, 以此为约束条件, 减小均值漂移迭代产生的偏差。实验结果表明, 该算法可以提高跟踪精度、鲁棒性及实时性。

关键词: 多尺度特征; 特征提取; 特征点匹配; 均值漂移; 目标跟踪

Mean Shift Target Tracking Algorithm Based on Multi-scale Feature Extraction

KONG Jun^{1,2,3,4}, TANG Xin-yi², JIANG Min¹, GE Yun-jian³

(1. School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China; 2. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China; 3. Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China; 4. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

【Abstract】 This paper proposes Mean Shift algorithm based on multi-scale feature extraction for fulfilling the target tracking in complex environment such as images with low contrast and to many similar targets. After the feature points being matched, next frame feature points are gotten. The center of next frame feature points is took as the center of searching window by which Mean Shift searching windows are continually modified and iteration deviation is reduced. Experimental results show that the robustness, precision and real-time performance of the algorithm are improved, and its iteration frequency is reduced.

【Key words】 multi-scale feature; feature extraction; feature point matching; Mean Shift; target tracking

DOI: 10.3969/j.issn.1000-3428.2011.22.054

1 概述

均值漂移(Mean Shift)由 Fukunaga^[1]等人首次提出, 后由 Cheng^[2]和 Comaniciu^[3]等人发展与推广, 应用于图像滤波、分割、目标跟踪等领域。其中, 目标跟踪意义上的均值漂移算法被描述为: 提出经核函数加权的色彩(或灰度)直方图概率分布, 把对目标跟踪的问题描述成对色彩直方图概率分布函数不断逼近的最优化问题, 以色彩直方图为特征, 对目标进行跟踪。色彩直方图具有缩放和旋转不变性等优点, 但当目标与背景对比度较低、色彩相近或相似目标较多时, Mean Shift 跟踪算法稳定性不高, 甚至失败。针对文献[4-5]所提及的 SIFT 算法及其不足, 本文设计了一种优化的多尺度特征提取(Multi-scale Feature Extraction, MSFE)算法^[6]。该算法通过基于尺度因子变化高斯卷积模板自适应调整, 以及连续帧双向配准的办法, 来快速有效地完成连续帧图像的配准。并以此为约束条件引入传统的 Mean Shift 跟踪算法中, 通过减小迭代偏差和迭代次数来提升均值漂移跟踪算法的效率。

2 多尺度特征点提取与匹配

2.1 多尺度空间建立

多尺度(三维)空间^[7] (x, y, σ) 采用图像金字塔来表达, 以具有不同尺度因子 σ_n 的高斯函数二阶梯度形式对原始图像进行卷积得到的一系列图像, 见式(1):

$$D(x, y, \sigma_n) = [\sigma_n^2 \nabla^2 G(x, y, \sigma_n)] * I(x, y) \quad (1)$$

其中, $G(x, y, \sigma_n)$ 是以 σ_n 为尺度因子的高斯滤波函数; $I(x, y)$

为原始图像。实际中, 尺度因子的递增过程如式(2)所示, 其中, s 取经验值 3。所以, 每个图像所对应的金字塔有 6 层。

$$\sigma_n = \sigma \times 2^{n/s}, n = 0, 1, \dots, s+2 \quad (2)$$

高斯卷积模板尺寸为 $N \times N$, N 与尺度因子符合如下经验公式:

$$N = \frac{9 \times \sigma_n}{1.2} \quad (3)$$

本文取 σ 的初始化值为 0.98, 表 1 给出 σ_n 和 N 的关系。

表 1 尺度因子与模板尺寸关系

σ_n	N_n	高斯滤波器模板尺寸
0.98	7.35	7×7
1.23	9.26	9×9
1.56	11.66	11×11
1.96	14.70	15×15
2.47	18.52	19×19
3.11	23.34	23×23

理论上, 尺度因子 σ_n 的值越大, 对应的模板尺寸越大, 图像的模糊化效果越好, 但卷积时间越长。传统的 SIFT 算法

基金项目: 国家自然科学基金资助项目(60910005); 中央高校基本科研业务费基金资助项目(JUSRP211A36, JUSRP111A41)

作者简介: 孔 军(1974—), 男, 副教授、博士, 主研方向: 目标跟踪, 图像识别; 汤心溢, 研究员、博士; 蒋 敏, 副教授、博士; 葛运建, 研究员、博士

收稿日期: 2011-06-13 **E-mail:** j-kong@163.com

是根据模板尺寸随着 σ_n 的增大不断加大的原则来构建多尺度空间的。另外, 实验统计发现, 随着 σ_n 的增加, 特征点数目并非一味增多, 其达到临界值后就会减少, 该临界值对应的尺度因子一般出现在 (1.20, 1.70) 范围内。可见, 模板尺寸超过一定范围时, 高斯滤波不仅无法增加稳定特征点的数量, 还消耗了多余的时间, 不利于算法的实时性要求。根据表 1, σ_n 出现临界值所对应的模板尺寸正好落在 9×9 、 11×11 上, 所以, 本文算法在高斯图像金字塔生成过程中, σ 由初始值 0.98 递增时, 模板尺寸进行自适应调整: $\sigma_n < 1.20$ 时, 直接选用 9×9 模板; $\sigma_n > 1.70$ 时, 直接选用 11×11 模板, 具体算法如下:

步骤 1 当前帧图像作为第 1 层, 以 9×9 卷积模板对其进行高斯平滑处理, 得到第 2 层, 依然用 9×9 卷积模板处理第 2 层图像, 得到第 3 层。然后改用 11×11 模板处理第 3 层, 得到第 4 层, 并在第 4 层基础上得到第 5 层、第 6 层。该帧图像生成的 6 层平滑图像即为第 1 组图像高斯金字塔。

步骤 2 第 1 组图像高斯金字塔中首幅图像进行下采样, 得到第 2 组图像高斯金字塔的首幅图像, 重复步骤 1, 构建第 2 组高斯图像金字塔, 以此类推, 得到后续图像高斯金字塔, 共 6 组。

步骤 3 对以上每组图像高斯金字塔相邻尺度图像两两相减, 得到该组的图像高斯差分金字塔, 共 6 组。

步骤 4 对后续帧重复步骤 1~步骤 3。

2.2 兴趣点检测与特征点确定

多尺度空间中的每一个采样点与立体邻域中所有点进行比较, 若灰度值之差满足设定阈值, 则视为兴趣点。通过对兴趣点的泰勒级数展开以及对偏移量的极值化处理, 将低对比度和不稳定的边缘点剔除, 最终确定出稳定的特征点集。

2.3 特征点的矢量化及其向量表示

选取以特征点为中心的 16×16 邻域窗口, 并将该窗口分割成 16 个 4×4 子窗口(每个子窗口视为一个种子)。在每个子窗口中计算出所有点的大小和方向, 然后对大小进行加权化处理, 并通过直方图统计出所有点的方向。这样, 一个特征点由 16 个子窗口(16 个种子)的点维持着, 而每个种子又可以用 8 个既有方向又有大小的数据来表示, 最终, 一个特征点将由一个 128(即 16×8) 维的向量来表示。

2.4 特征点的双向匹配

连续帧图像的特征向量生成之后, 可用多维空间向量欧氏距离作为匹配准则, 见式(4):

$$d = \sqrt{\sum_{i=1}^{128} (x_{i1} - x_{i2})^2} \quad (4)$$

其中, x_{i1} 表示其中一幅图像上某个特征点的特征向量的第 i 维元素值; x_{i2} 表示另一幅图像上某个特征点的特征向量的第 i 维元素值。具体算法如下:

步骤 1 建立极线约束条件, 减小图像配准的搜索范围。

步骤 2 求出前一帧图 A 中某一个特征点 F_1 与后续帧图 B 中在对应极线上所有特征点的欧氏距离, 搜索出最近点 F_2 , 如果该欧氏距离小于匹配阈值, 保留 F_2 , 进入步骤 3; 否则, 删除 F_1 , 重复步骤 2。

步骤 3 求出后续帧图 B 中 F_2 与前一帧图 A 中对应极线上所有特征点的欧氏距离, 搜索出最近点 F_3 。

步骤 4 判断 F_1 和 F_3 是否相等, 若是, 匹配成功, 否则, 重复步骤 2。

3 基于多尺度特征提取的 Mean Shift 算法

均值漂移算法是一种统计迭代算法, 在起始帧中通过手动方式确定搜索区域, 并以该区域的色彩直方图概率分布作为目标模型, 迭代计算出后续帧中候选搜索窗的直方图概率分布, 以 2 个直方图概率分布距离最小原则计算出后续帧的 Mean Shift 向量, 从而估计出目标的最佳位置。算法具体步骤如下:

(1) 以起始帧确定目标模型

将像素色彩直方图分成若干特征区间, 计算起始帧手动提取的搜索窗中所有像素 n 个样本点 $x_i = (x_i^s, x_i^r)$, $i = 1, 2, \dots, n$, 其中, x_i^s 表示像素点坐标; x_i^r 表示像素点有关色彩的多维向量; 起始搜索窗中心为 x_0 , 则起始搜索窗模型所有像素属于第 $u = 1, 2, \dots, m$ 个特征值的概率为:

$$\hat{q}_u = C \sum_{i=1}^n K\left(\left\|\frac{x_i^s - x_0}{h}\right\|\right) \delta[b(x_i^s) - u] \quad (5)$$

其中, 函数 $b(x_i^s)$ 是样本点 x_i 的坐标 x_i^s 处的颜色, 如果颜色值为 u , 则 $\delta[b(x_i^s) - u] = 1$, 否则为 0; $K(\cdot)$ 为核函数。

(2) 计算后续帧目标模型

后续帧搜索窗中心为 y , 则其模型所有像素属于第 u 个特征值的概率为:

$$\hat{p}_u(y) = C_h \sum_{i=1}^n K\left(\left\|\frac{x_i^s - y}{h}\right\|\right) \delta[b(x_i^s) - u] \quad (6)$$

其中, C_h 是不同于 C 的归一化常数。

(3) 计算相似性函数(Bhattacharrya 系数)

相似性函数描述的是起始帧目标模型和后续帧模型之间的距离, 因此, Mean Shift 算法的本质是寻找最优的 y , 使 $\hat{p}_u(y)$ 与 \hat{q}_u 最相似。 $\hat{p}_u(y)$ 和 \hat{q}_u 的相似性用 Bhattacharrya 系数 $\hat{\rho}(y)$ 来度量分布, 即:

$$\hat{\rho}(y) = \rho[\hat{p}_u(y), \hat{q}_u] = \frac{m}{\sum_{u=1}^m \sqrt{\hat{p}_u(y) \hat{q}_u}} \quad (7)$$

(4) 确定后续帧目标位置

色彩直方图概率密度函数的核函数估计为:

$$\hat{f}(x) = \frac{\sum_{i=1}^n G\left(\frac{x_i^s - x}{h}\right) \omega(x_i^s)}{h^d \sum_{i=1}^n \omega(x_i^s)} \quad (8)$$

其中, $\omega(x_i) \geq 0$ 是赋予采样点 x_i 的权重。取单位均匀核函数 $G(x)$ 为核函数, 对 $\hat{f}(x)$ 的梯度函数 $\nabla \hat{f}(x)$ 进行估计, 并提取其中的 Mean Shift 向量, 得到式(9):

$$\hat{y}_1 = \frac{\sum_{i=1}^n x_i^s G\left(\left\|\frac{x_i^s - \hat{y}_0}{h}\right\|\right) \omega(x_i^s)}{\sum_{i=1}^n G\left(\left\|\frac{x_i^s - \hat{y}_0}{h}\right\|\right) \omega(x_i^s)} \quad (9)$$

其中, \hat{y}_0 是当前帧搜索窗中心坐标; \hat{y}_1 表示通过 Mean Shift 算法一次迭代计算出的后续帧的搜索窗中心坐标。

4 基于 MSFE 的 Mean Shift 目标跟踪

Mean Shift 算法基于连续帧目标色彩直方图概率密度函数梯度的无参数估计, 计算量较小, 但算法对目标与背景色彩差和对比度要求较高。本文提出基于 MSFE 的 Mean Shift 跟踪算法, 其实质是在跟踪过程中, MSFE 算法完成前一帧图像对后续帧图像的配准, 并标记出后续帧图像中目标区域已经匹配成功的单个特征点, 或若干个最密集特征点集外接邻域中心, 并以此为约束条件, 修正 Mean Shift 算法的搜索

窗。具体算法如下:

步骤1 设定 Mean Shift 迭代容许误差 ϵ , 设定 MSFE 修正阈值 δ 。

步骤2 读入视频序列, 选择跟踪目标, 在第 k 帧初始化 Mean Shift 搜索窗, 记录下窗口中心坐标 \hat{y}_0 。

步骤3 完成第 k 帧目标与第 $k+1$ 帧目标的 MSFE 特征点匹配, 记录第 $k+1$ 帧目标的特征点集中心坐标 \hat{x}_1 。

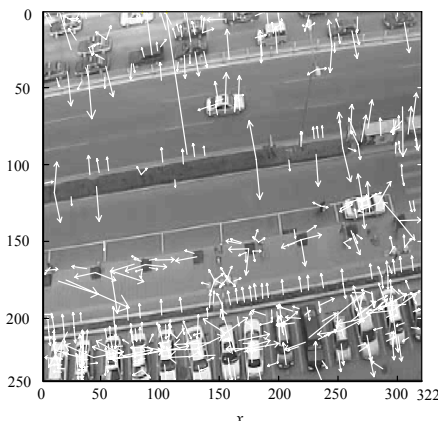
步骤4 利用均值漂移算法计算 \hat{y}_1 , 如果 $\|\hat{y}_1 - \hat{y}_0\| > \epsilon$, 且 $\|\hat{y}_1 - \hat{x}_1\| < \delta$, 则 $\hat{y}_0 = \hat{y}_1$, 重复步骤 4; 如果 $\|\hat{y}_1 - \hat{y}_0\| > \epsilon$, 但 $\|\hat{y}_1 - \hat{x}_1\| > \delta$, 则 $\hat{y}_0 = \frac{\hat{x}_1 + \hat{y}_1}{2}$, 重复步骤 4; 如果 $\|\hat{y}_1 - \hat{y}_0\| < \epsilon$, 记录 \hat{y}_1 的值, 进入步骤 5。

步骤5 以 \hat{y}_1 为中心, 在第 $k+1$ 帧确定搜索窗, $k = k+1$, 重复步骤 3、步骤 4。

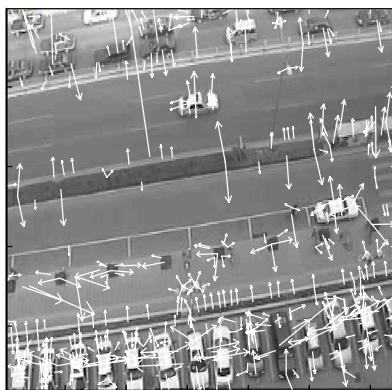
5 实验结果与分析

为了验证本文基于 MSFE 的 Mean Shift 跟踪算法的性能, 利用某单目摄像机采集视频图像序列, 以完成对“2 辆在高速公路上行驶的小汽车”的跟踪, 帧频为 28 f/s, 每帧图像分辨率为 320×240 像素。运行机器为 Intel Core 2 Duo E4700 2.6 GHz 的 PC, 内存为 2 GB, 操作系统为 WinXP, 开发平台为 VC++6.0, 结合 OpenCV 开源库函数。

对视频流的每一帧图像, 利用 2.1 节的算法构建高斯图像金字塔, 获得目标区域上的稳定特征点集, 由于篇幅所限, 图 1 仅给出视频第 1 帧、第 2 帧的特征描述矢量。



(a) 第 1 帧



(b) 第 2 帧

图 1 特征点描述矢量示意图

对视频流相邻帧以 MSFE 算法(阈值选 0.53)完成两两匹配, 即帧 1 匹配帧 2, 帧 2 匹配帧 3。由于篇幅限制, 图 2

只给出第 1 辆车出现在视场时帧 1 与帧 2 的目标配准结果, 其余帧类似。由于引入了 2.4 节的极线约束条件, 因此减小了图像配准的搜索范围。



图 2 基于 MSFE 算法的目标配准结果

图 3 给出传统 Mean Shift 算法对小车的跟踪情况。在第 1 帧, 第 1 辆小车进入视场, 手动在目标模型上初始化搜索窗, 在第 11 帧, 搜索窗后移目标中心, 在第 21 帧, 搜索窗脱靶, 在第 74 帧, 第 2 辆小车进入视场, 同样手动初始化搜索窗, 此时目标与背景对比度较低, 所以, 跟踪效果较浅色小车差, 在第 84 帧、第 94 帧上, 搜索窗均有脱靶现象。

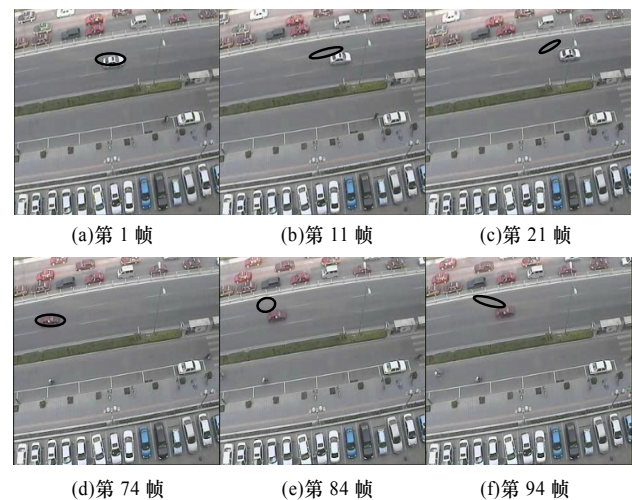


图 3 Mean Shift 算法跟踪效果

图 4 给出基于 MSFE 的 Mean Shift 算法对小车的跟踪情况。可以看出, 搜索窗能精确地跟踪到先后驶入的 2 辆小车。

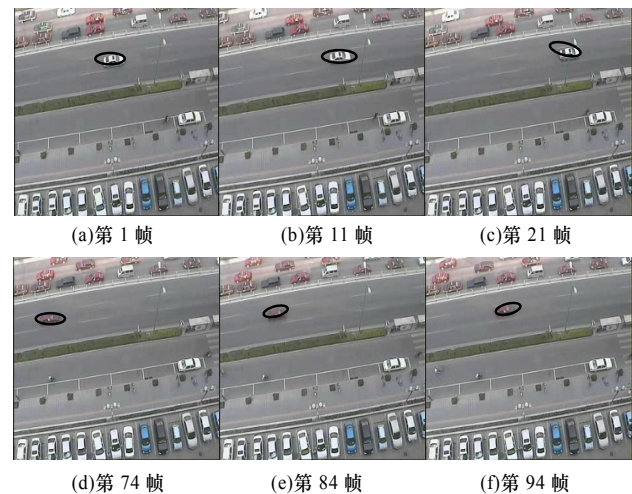


图 4 基于 MSFE 的 Mean Shift 算法跟踪效果

为了进一步对比 2 种算法的跟踪效率, 图 5 给出传统算

法和改进算法在确立搜索窗时每帧所完成的迭代次数。由于目标模型特征点坐标通过设定的 MSFE 修正阈值 δ 不断“修复”迭代时搜索窗的中心坐标, 因此后者确定最终搜索窗的迭代次数少于前者, 扣除相邻帧配准所花费的时间, 基于 MSFE 的 Mean Shift 算法跟踪速度可提高近 45%~50%, 从而提高了均值漂移算法的实时性。

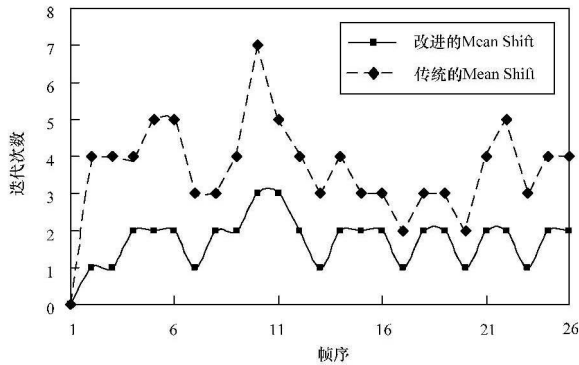


图5 2种算法的迭代次数比较

6 结束语

传统的 Mean Shift 算法在目标与背景对比度较低、色彩接近或相似目标较多的情况下不能很好地跟踪目标。为此, 本文将多尺度特征点配准的思想引入 Mean Shift 框架中, 提出一种改进的算法。以 MSFE 特征点配准为约束条件, 前一帧目标区域特征点匹配出后续帧目标区域特征点, 以后者的特征点集中心坐标不断修正均值漂移搜索窗的中心位置, 最

终获得连续帧中最优的均值漂移搜索窗位置。实验结果表明, 基于多尺度特征提取的均值漂移跟踪算法不仅提高了跟踪精度, 改善了传统的均值漂移跟踪算法的鲁棒性, 同时降低了迭代次数和迭代偏差, 实时性也得到一定的提升。

参考文献

- [1] Fukunaga K, Hostetler L D. The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition[J]. IEEE Trans. on Information Theory, 1975, 21(1): 32-40.
- [2] Cheng Yizong. Mean Shift, Mode Seeking, and Clustering[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790-799.
- [3] Comaniciu D, Ramesh V, Meer P. Kernel-based Object Tracking[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.
- [4] Lowe D G. Object Recognition from Local Scale-invariant Features[C]//Proc. of International Conference on Computer Vision. Corfu, Greece: [s. n.], 1999.
- [5] Lowe D G. Distinctive Image Features from Scale-invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [6] 孔 军, 汤心溢, 蒋 敏. 基于多尺度特征提取的运动目标定位研究[J]. 红外与毫米波学报, 2011, 30(1): 21-26.
- [7] 于丽莉, 戴 青. 一种改进的 SIFT 特征匹配算法[J]. 计算机工程, 2011, 37(2): 210-212.

编辑 张 帆

(上接第 163 页)

参数收敛性实验则选择建模相对误差来反映参数收敛性与迭代步数的关系。建模相对误差的定义见文献[3-4]。

表 1 是本文辨识方法和文献[4]的方法在 4 种情况下的建模相对误差, 为节省篇幅, 仅列出了前 6 次迭代的建模相对误差。2 种方法的迭代初值均由子空间辨识方法确定(采用 Matlab 的标准函数 n4sid()得到)。从表 1 可知, 由于迭代初值相同, 2 种方法的初始建模相对误差(第 1 步迭代的误差)相同, 均为 28.6。对于同一迭代步数, 2 种方法在 4 种数据缺失程度下所对应的建模相对误差不相同(情况(1)的建模相对误差最小, 情况(4)最大), 表明 2 种方法均受到数据缺失程度的影响。但对于每一种情况, 本文方法的建模相对误差小于文献[5]的方法。

表 1 2 种方法的建模相对误差

迭代次数	本文方法				文献[5]方法			
	情况(1)	情况(2)	情况(3)	情况(4)	情况(1)	情况(2)	情况(3)	情况(4)
1	28.60	28.60	28.60	28.60	28.60	28.60	28.60	28.60
2	25.80	26.10	26.30	27.60	26.20	26.50	26.80	27.60
3	16.90	19.30	20.10	17.40	18.30	20.60	21.10	18.80
4	10.40	□5.50	11.20	11.30	□4.24	□5.26	10.40	16.70
5	□5.20	□5.63	11.60	14.10	□6.20	□5.35	10.80	16.80
6	□1.18	□5.71	11.80	14.50	□2.18	□5.48	11.30	15.60

5 结束语

针对输出存在状态缺失的状态空间模型, 本文提出了极

大似然辨识方法。首先构建以观测数据为条件概率的似然函数, 接着给出极大化似然函数的系统矩阵计算方法, 最后进行了数值仿真实验。实验结果证明该辨识方法是有效的。本文假定状态缺失模式已知, 对随机状态缺失模式的系统辨识将是后续的研究工作。

参考文献

- [1] Sinopoli B, Schenato L, Franceschetti M, et al. Kalman Filtering with Intermittent Observations[J]. IEEE Trans. on Automatic Control, 2004, 49(9): 1453-1463.
- [2] 王 超, 高 珍, 王 睿. 基于增量维纳滤波的图像插值算法辨识[J]. 计算机工程, 2010, 36(7): 224-226.
- [3] 衷路生, 宋执环. 基于正交梯度搜索的动态系统递阶优化辨识[J]. 自动化学报, 2008, 34(6): 711-715.
- [4] Gibson S, Ninness B. Robust Maximum-likelihood Estimation of Multivariable Dynamic Systems[J]. Automatica, 2005, 41(10): 1667-1682.
- [5] Wang Zidong, Daniel W D, Liu Xiaohui. Variance-constrained Filtering for Uncertain Stochastic Systems with Missing Measurements[J]. IEEE Trans. on Automatic Control, 2003, 48(7): 1254-1258.

编辑 张 帆