

实现基于卡尔曼滤波器的运动目标跟踪

一、运动目标跟踪算法的研究现状

基于检测的目标跟踪框架最早在 SVT 算法中被提出，此后，绝大部分的跟踪算法都采用这种框架进行设计。基于检测的跟踪框架把跟踪问题转化为检测问题，通过对所有的采样区域特征进行检测评分来得到最可靠的目标区域，从而更新目标所在的位置。在 2016 年以前，传统方法在跟踪领域占主导地位，而在 2016 年以后，深度学习方法逐渐崭露头角并最终占据主导地位。下面分别介绍传统和深度学习跟踪方法中的经典之作。

近十年来，国内外的研究者提出了许多经典的传统跟踪模型。1) 均值偏移 (MeanShift) 是早期的跟踪方法之一，均值偏移向量是指向局部密度中心的向量，这类方法通过迭代优化使跟踪位置收敛于密度极值点。均值偏移算法可以在简单场景下取得不错的跟踪效果，但如果背景中存在相似干扰物或目标移动距离较大时，这类方法容易收敛到局部最小值点，跟踪准确度大幅下降。2) 随机森林方法通过训练数据学习得到一颗决策树，决策树的子节点输出结果通过融合形成目标特征，最后该类方法一般会设计一个分类器对目标特征进行分类，得到最终的分类评分。3) 支持向量机 (SVM) 常被用于对目标和背景进行分类判别，基于 SVM 的跟踪方法一般会在线学习并更新一个 SVM 分类器，对所有检测区域的图像特征进行分类，取分类评分最大的区域作为目标所在的区域。由于 SVM 分类器对于小样本集的拟合能力较强，这类跟踪方法的准确度较高。4) 在基于稀疏表达的模型中，通常算法会预先建立一个目标字典，字典中的每一项都是目标的一个分块图像特征，在跟踪过程中，每一个搜索子区域都会用预先建立的字典进行稀疏重构，稀疏重构的误差越小，表明该区域的目标概率越大，由于该类算法采用 L1 范数作为正则化项，其算法运行速度一般较慢。5) 粒子滤波是一种离散的贝叶斯推断方法，这类跟踪方法的优化目标是后验概率最大化。由于跟踪目标的运动模式具有较大的不确定性，粒子滤波方法中的运动模型通常采用高斯模型进行近似。因此，粒子滤波方法的核心点在于外观模型的设计。6) 相关滤波模型把相关运算作为相似度计算方法，通过傅里叶变换把原问题转换到频域进行优化求解，极大提高了算法的运算速度，并通过循环移位增加了训练样本的数量。针对常规稀疏表达方法优化速度较慢的问题，CST 算法把循环移位数据结构引入到稀疏表达中并把优化问题转换到傅里叶域进行计算，有效地提高了稀疏表达方法的运算效率。传统目标跟踪方法的精度受限于特征图上的最小像素距离，C-COT 通过插值方法来学习空间上连续的相关滤波器，从而解决了跟踪精度的问题。

由于深度学习方法在目标检测、分类等领域的成功应用，一些研究者也把深度学习引入到跟踪问题中。1) 最早的深度学习跟踪方法采用 CNN 网络进行特征提取和分类。MDNet 是最优秀的算法之一，该网络框架包含共享的预训练特征提取层和多个域相关层，每一个域相关层对应于一个目标物体，尽管 MDNet 算法在通用数据库上取得了优异的效果，其算法运行速度为 1FPS，无法达到实时性的要求。2) SiamFC 算法把传统的相关滤波方法进行网络化实现并采用孪生网络进行设计，其运算速度可以满足实时性的要求，但跟踪准确度比 MDNet 差。3) 基于 RPN 的跟踪方法采用 RPN 网络进行候选区域提取并通过分类和回归方法来得到目标的外接矩形框。SiamRPN++ 根据孪生网络结构进行设计，把多层深度网络之间的逐通道相关进行加权融合，最后通过多层卷积网络得到最终的分类和回归结果。4) 目标分割和跟踪相结合是最新的发展趋势，这类方法一般通过全卷积网络 (FCN) 来得到目标的分割结果，然后根据分割结果来计算目标外接矩形框。FTMU 采用 MaskR-CNN 作为候选区域提取和 Mask 分割网络，并利用增强学习方法来选择当前跟踪的模式。

在简单场景下现有的算法具有很好的检测效果，但是在实际应用场景中难免会有很多复杂场景会干扰运动目标跟踪的结果，包括背景干扰、鬼影问题、目标尺度变化、目标被遮挡等，所以研究复杂环境下运动目标跟踪技术具有很重要的意义。

二、算法分析

2.1 线性卡尔曼滤波

经典的线性卡尔曼 (Kalman) 滤波是一种线性滤波与预测方法，卡尔曼滤波分为 2 个步骤，预测 (predict) 和校正 (correct)。预测是基于上一时刻状态估计当前时刻状态，而校正则是综合当前时刻的估计状态与观测状态，估计出最优的状态。预测与校正的过程如下：

预测：

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + B_k u_k \quad \text{公式 (1)}$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q_k \quad \text{公式 (2)}$$

校正：

$$K_k = P_{k|k-1}H_k^T(H_k P_{k|k-1}H_k^T + R_k)^{-1} \quad \text{公式 (3)}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H_k \hat{x}_{k|k-1}) \quad \text{公式 (4)}$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1}$$

公式（5）

公式（1）是状态预测，公式（2）是误差矩阵预测，公式（3）是卡尔曼增益计算，公式（4）是状态校正，其输出即是最终的卡尔曼滤波结果，公式（5）是误差矩阵更新。各变量说明如下表：

表 1 线性卡尔曼滤波参数说明	
变量	说明
x_k	k 时刻的状态
A	状态转移矩阵，和具体的线性系统相关
u_k	k 时刻外界对系统的作用
B	输入控制矩阵，外界的影响如何转化为对状态的影响
P	误差矩阵
Q	预测噪声协方差矩阵
R	观测噪声协方差矩阵
H	观测矩阵
K_k	k 时刻的 kalman 增益
z_k	k 时刻的观测值

其中，观测矩阵定义了系统状态和观测状态之间的转换关系。定义系统噪声 Q 和 R 是一个难点，即每一瞬间球状态估计的误差。一般可以简单认为系统状态各维之间相互独立，那么 Q 和 R 就可以设置为对角阵。而这两个对角线元素的大小将直接影响着滤波结果，若 Q 的元素远大于 R 的元素，则预测噪声大，从而更相信观测值，这样可能使得卡尔曼滤波结果与观测值基本一致；反之，则更相信预测，卡尔曼滤波结果会表现得比较规整和平滑。

2.2 小球追踪实验

我们选择一个黄色乒乓球作为实验对象，以跟踪它在球台上的运动，如图 1 所示。



图 1 实验场景

目标检测阶段采用经典的检测算法，涉及高斯平滑、HSV 转换、空间滤波、形态学滤波、二值化、轮廓检测，从而提取出包围球体的最小矩形框并作为后续处理的对象。

目标跟踪阶段，我们假设球的运动方式近似为线性的，对系统建立线性模型进行刻画，则可采用上述线性卡尔曼滤波方法实现跟踪。具体来说：

(1) 系统状态包括矩形框中心（球心）的坐标（ x,y ）、中心点的速度（像素/秒）（ v_x,v_y ）以及矩形框的大小（ w,h ），一共有 6 个状态量，即系统状态向量为 $[x,y,v_x,v_y,w,h]$ 。

(2) dt 用于度量从一个阶段到下一个阶段所经过的时间，则状态转移矩阵 A 为：

$$A = \begin{bmatrix} 1 & 0 & dt & 0 & 0 & 0 \\ 0 & 1 & 0 & dt & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(3) 观测状态包括球心的坐标（ x,y ）以及矩形框的大小（ w,h ），一共有 4 个状态量，即观测状态向量为 $[z_x,z_y,z_w,z_h]$ 。相应地，观测矩阵 H 为：

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(4) 系统没有外界输入，因此无需考虑 u_k 和 B ，均为 0 矩阵。

(5) 预测噪声协方差矩阵 Q 和观测噪声协方差矩阵 R 分别为：

$$Q = \begin{bmatrix} E_x & 0 & 0 & 0 & 0 & 0 \\ 0 & E_y & 0 & 0 & 0 & 0 \\ 0 & 0 & E_{vx} & 0 & 0 & 0 \\ 0 & 0 & 0 & E_{vy} & 0 & 0 \\ 0 & 0 & 0 & 0 & E_w & 0 \\ 0 & 0 & 0 & 0 & 0 & E_h \end{bmatrix} \quad R = \begin{bmatrix} E_{zx} & 0 & 0 & 0 \\ 0 & E_{zy} & 0 & 0 \\ 0 & 0 & E_{zw} & 0 \\ 0 & 0 & 0 & E_{zh} \end{bmatrix}$$

其中，设定 Q 矩阵中 $E_x = E_y = E_w = E_h = 0.01$ ， $E_{vx} = E_{vy} = 5$ ， R 矩阵的对角元素均为 0.1。

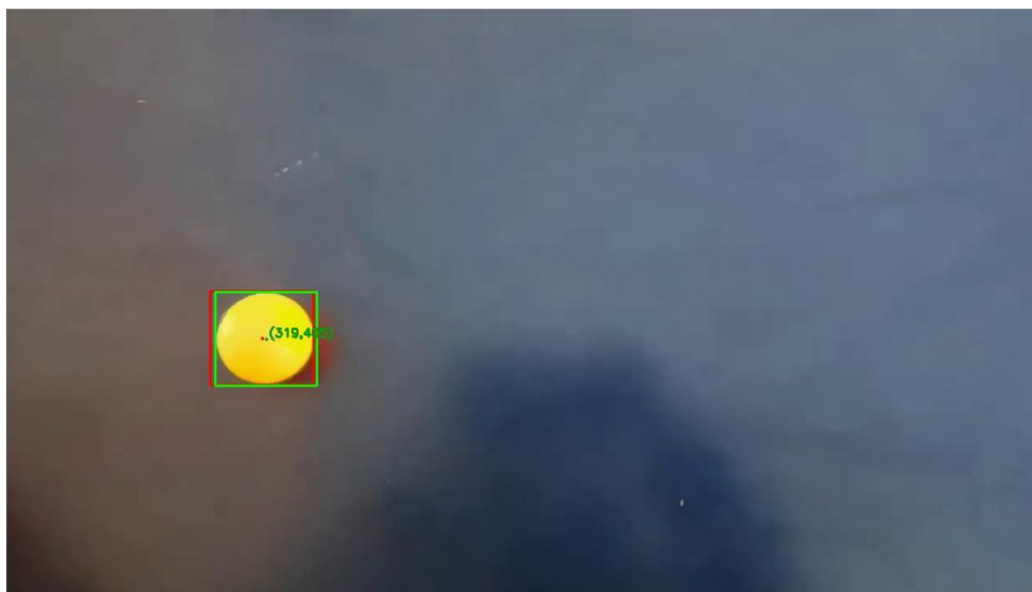
三、代码实现

实验采用 C++ 编程，并利用 OpenCV 库中的相关函数，对提前录制好的视频

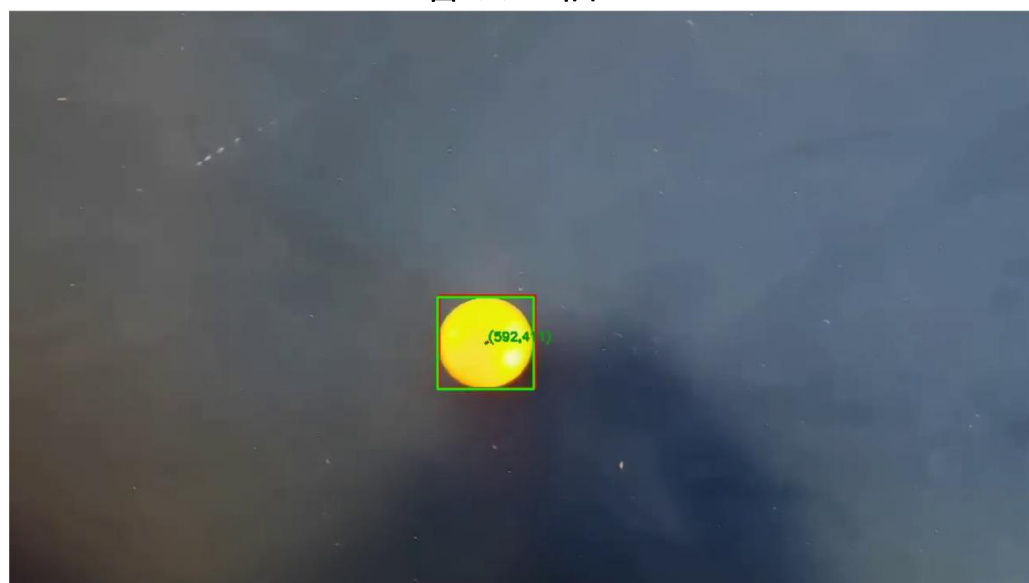
进行逐帧处理。

四、结果展示

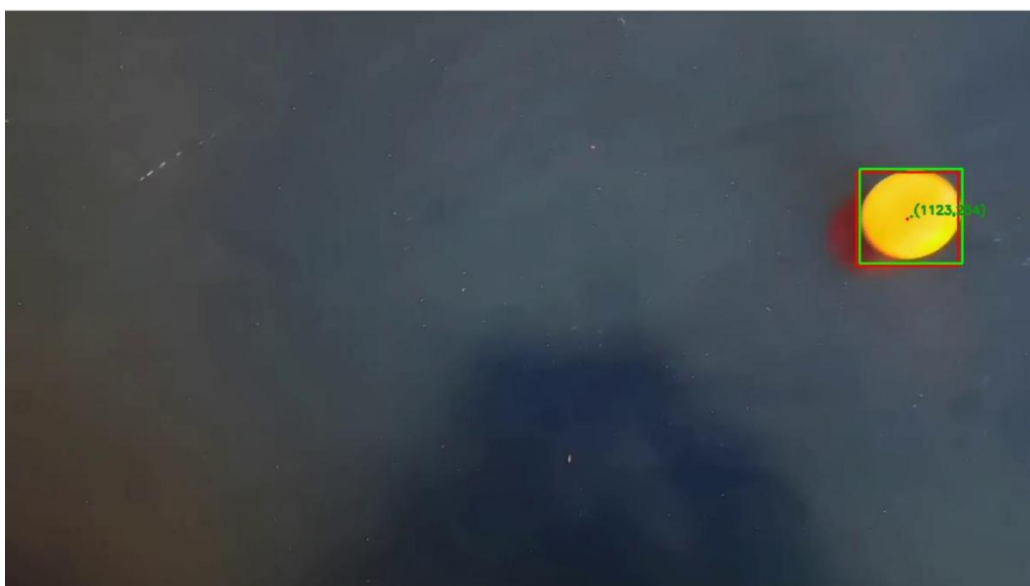
这里选取三帧跟踪结果画面进行展示，其中绿色部分是实际观测到的状态，包括球的最小包围框和中心点，红色部分显示的是卡尔曼滤波器预测的结果。可以看到，基本每帧画面中估计的状态与测量的状态都有较大重叠，跟踪效果较好，也说明球线性运动的假设是有效的。



图（a） 画面 1



图（b） 画面 2



图(c) 画面3

图 2 小球跟踪结果