

一种基于 Kalman 滤波的视频对象跟踪方法

张江山 朱光喜

(华中科技大学电子与信息工程系图象信息处理与智能控制国家教委开放研究实验室, 武汉 430074)

摘要 为了更加准确地预测对象的位置和运动,利用刚体运动模型导出最佳 Kalman 系数,通过 Kalman 反馈滤波器对 Moscheni 等人提出的视频对象分割与跟踪算法^[1]进行改进,提出了一种将离散 Kalman 滤波技术用于视频序列的对象跟踪方法.这种方法可用于有关场景描述的各种应用领域中,如在机器视觉的研究中,对动态场景进行分析与理解;在基于对象的视频编码中(如 MPEG-4),对视频对象进行分割后,分别进行编码,从而改善编码的可分级性及编码效率.实验结果表明,采用这种方法可以有效地改善时间-空间分割和目标跟踪,有助于更好地理解动态场景,并表现出良好的鲁棒性.

关键词 Kalman 滤波 视频对象跟踪 视频编码 运动预测

中图分类号: TN919.81 文献标识码: A 文章编号: 1006-8961(2002)06-0606-04

Kalman Filter for Video Object Segmentation and Tracking

ZHANG Jiang-shan, ZHU Guang-xi

(Dept. of Electronics & Information Engineering., Huazhong Univ. of Science & Tech., Wuhan 430074)

Abstract In this paper, a technique based on a discrete Kalman filter algorithm is proposed to follow the trajectory of the objects. The aim is to obtain a precise prediction of their position and motion. The accurate prediction improves both the recursive spatio-temporal segmentation and object tracking performances, enabling a high level understanding of the scene dynamics. The derived scene representation obtained finds applications in various domains. For instance, it is very well suited for dynamic scene analysis where a deep scene understanding is required. Typical examples are scene understanding and robot vision. It is also very appealing in the context-based video coding(MPEG-4). Experimental results have shown that this method is able to integrate over time the temporal information for each object and to interpolate or extrapolate its trajectory, correctly predict the position and the motion of temporal coherent objects. However, if the object has performed maneuvers, the Kalman filter fails in its prediction. In order to decide when is convenient to use the last estimated motion of the object instead of the Kalman prediction a test based on motion compensation error is used. Finally the proposed algorithm has shown its robustness in the presence of object occlusions.

Keywords Kalman filter, Video object tracking, Video code, Motion prediction

0 引言

随着 MPEG-4 视频压缩标准的提出,基于视频内容的编码方法逐渐成为目前的研究热点,而其关键问题在于如何准确地实现运动视频对象的分割及跟踪. Moscheni 等人提出了利用空间、时间特性来

描述运动对象,并实现对运动对象的分割与跟踪的方法^[1~3].该方法主要由两部分组成,即时-空分割和目标跟踪.如图 1 所示,首先,通过 k 帧图象,对 $k+1$ 帧图象进行初步分割;然后,比较 $k+1$ 帧与 k 帧图象的分割区域;最后利用空间、时间特性来描述运动对象,并确定其运动轨迹.

由于是通过运动对象的空间和时间特性的

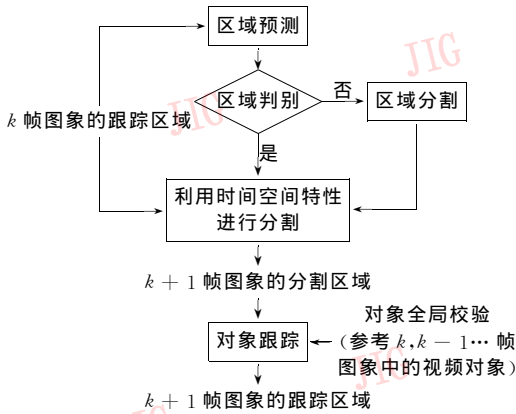


图 1 Moscheni 的时-空分割与对象跟踪方法

预测,来实现运动对象的分割与跟踪,因此,空间、时间信息的准确预测对整个分割与跟踪算法至关重要,一个错误的预测将会导致整个算法的失败.而上述方法只是简单地采用基于像素的区域搜索匹配方法进行对象的运动预测和判别,所以只有对象在 2 个连续帧的运动变化很小时,算法才适用,但当对象运动发生跳跃变化或受到其他因素干扰时,如被其他对象遮挡或在视频传输时发生丢帧现象,算法便会失败,这种现象称之为“死锁”.为了获得对象运动的可靠预测,引入离散 Kalman 滤波对上述分割与跟踪方法进行了改进.

1 基于 KALMAN 滤波的对象运动预测

为了充分利用每个对象的内在时间信息,采用刚体运动学模型描述每个运动矢量的时间变化特征,即对象的运动可用以下离散线性方程来描述

$$S_{k+1} = \Phi_k S_k + W_k \quad (1)$$

其中, S_k 表示系统状态,它描述了在 k 时刻每个运动矢量的值; Φ_k 矩阵为运动模型,用于描述刚体平移及旋转运动; W_k 表示运动模型的预测误差,其统计特性与白噪声或高斯噪声相似.

根据式(1),可以定义度量公式为

$$X_k = H_k S_k + V_k \quad (2)$$

式(2)描述了度量状态 X_k 与系统状态 S_k 之间的关系.其中, H_k 为关系矩阵, V_k 为运动估值过程中所产生的误差.

V_k 和 W_k 的协方差矩阵为

$$E[W_k W_k^T] = \begin{cases} Q_k & i = k \\ 0 & i \neq k \end{cases} \quad (3)$$

$$E[V_k V_k^T] = \begin{cases} R_k & i = k \\ 0 & i \neq k \end{cases} \quad (4)$$

$$E[W_k V_k^T] = 0 \quad (5)$$

通过式(1)用运动模型对 k 时刻的运动状态进行估计,得到一个先验估值 \hat{S}_k ,其估值误差为

$$\hat{e}_k = S_k - \hat{S}_k \quad (6)$$

其协方差矩阵为

$$\hat{C}_k = E[\hat{e}_k \hat{e}_k^T] = E[(S_k - \hat{S}_k)(S_k - \hat{S}_k)^T] \quad (7)$$

然后,便可用度量数据 X_k 来改善先验估值 \hat{S}_k ,可以采用 S_k 的线性递归估计方法,利用最小均方误差,确定最佳估值,其表达式为

$$S_k^* = \hat{S}_k + K_k (X_k - H_k \hat{S}_k) \quad (8)$$

其中, S_k^* 为先验估值的修正值, K_k 为调和因子.

现在的问题是如何得到 K_k ,使得修正值(称为后验估值)与实际值的误差最小,即,使得协方差矩阵

$$C_k = E[e_k e_k^T] = E[(S_k - S_k^*)(S_k - S_k^*)^T] \quad (9)$$

最小.

将式(2)代入式(8),然后将结果代入式(9)得

$$C_k = E\{[(S_k - \hat{S}_k) - K_k (H_k S_k + V_k - H_k \hat{S}_k)] \times [(S_k - \hat{S}_k) - K_k (H_k S_k + V_k - H_k \hat{S}_k)]^T\} \quad (10)$$

式(10)中, $(S_k - \hat{S}_k)$ 为先验估值误差,它与度量误差 V_k 无关,因此有

$$C_k = (I - K_k H_k) \hat{C}_k (I - K_k H_k)^T + K_k R_k K_k^T \quad (11)$$

从中可以得到使估值误差最小的 K_k 为

$$K_k = \hat{C}_k H_k^T (H_k \hat{C}_k H_k^T + R_k)^{-1} \quad (12)$$

即 Kalman(最佳)系数,详细论述参阅文献[4].

将式(12)代入式(11)得

$$C_k = (I - K_k H_k) \hat{C}_k \quad (13)$$

要使用式(8)对先验估值进行修正,就必须得到 \hat{S}_k 和 \hat{C}_k ,因此,首先在 k 时刻用式(9)给出的初始误差协方差矩阵对估值进行初始化,并采用类似方法对下一个度量数据 X_{k+1} 进行预测.

由于 W_k 描述的是 k 到 $k+1$ 时刻处理前的累计误差,因此,可以通过下式进行预测处理

$$\hat{S}_{k+1} = \Phi_k S_k^* \quad (14)$$

这样便可通过先验误差的计算得到 \hat{S}_{k+1} 以及误差协方差矩阵

$$\hat{e}_{k+1} = S_{k+1} - S_{k+1}^* = (\Phi_k S_k + W_k) - \Phi_k S_k^* = \Phi_k e_k + W_k \quad (15)$$

由于 W_k 是 k 时刻以前的误差,所以它与 e_k 没有相关性.这样,与 \hat{S}_{k+1} 相关误差的协方差矩阵为

$$\begin{aligned}\hat{C}_{k+1} &= E[\hat{e}_{k+1}\hat{e}_{k+1}^T] \\ &= E[(\Phi_k e_k + W_k) + (\Phi_k e_k + W_k)^T] \\ &= \Phi_k C_k \Phi_k^T + Q_k\end{aligned}\quad (16)$$

综上所述,式(8)、式(12)~式(15)等构成了

Kalman 反馈滤波公式.其执行过程如图 2 所示.

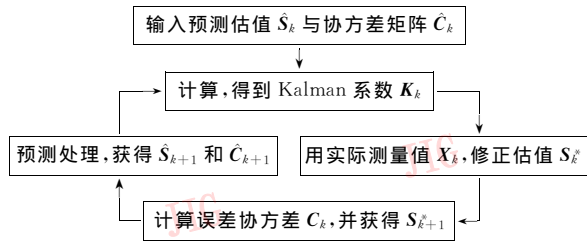


图 2 Kalman 循环滤波过程

利用上述离散 Kalman 滤波算法对 Moscheni 的方法进行改进,便实现了一种反馈式运动跟踪,每执行一次运动估值操作,便获得一个新的对象运动的度量数据 X_k ,根据度量数据和运动模型,修正 Kalman 滤波器的状态矢量 S_k .这使得滤波器可以充分利用全局时间信息进行运动估值的计算.使用式(1),通过 S_k 预测下一个状态 S_{k+1} ,并用 $S_{k+1/k}$ 表示这种预测方式.

然而,当运动对象的运动方式发生较大的改变时(运动模型参数发生变化),就需要重新设置 Kalman 滤波的初始条件,利用对象实际运动轨迹,对运动模

型参数进行调整.这种预测方式,用 S_k 表示.

从以上所述的 Kalman 滤波算法可以看出,它具有良好的数学特性.首先,在进行时间分析处理时,不需要定义一个精确的运动模型.它能够通过一个简单的运动学模型有效地跟踪真实对象的运动轨迹. Kalman 滤波能够根据输入参数进行自适应调整处理.其次,由于本文采用的 Kalman 滤波算法是一种基于运动学模型的反馈结构算法,因此当发生类似于“死锁”现象时,可以根据对象以前的运动轨迹和运动模型继续对其运动进行预测.当这些对象再次在场景中出现时,可以帮助跟踪系统识别这个对象,因此非常适合于视频序列编码与视频监视等应用.

2 实验结果分析

以“Table-Tennis”视频序列中球的运动为例,使用 Moscheni 算法和 Kalman 滤波改进算法对“乒乓球”分别进行分割和跟踪.“乒乓球”的实际运动轨迹(圆)和预测运动轨迹(叉)如图 3 所示.图 4 为两种方法的预测误差比较.在采用 Kalman 滤波改进算法时,需通过对补偿后误差大小的比较,来确定是使用 S_k 方式,还是使用 $S_{k+1/k}$ 方式进行预测.例如,当小球到达其轨迹顶部或与球拍碰撞时,使用 S_k 方式进行预测,可以获得更理想的估值结果.

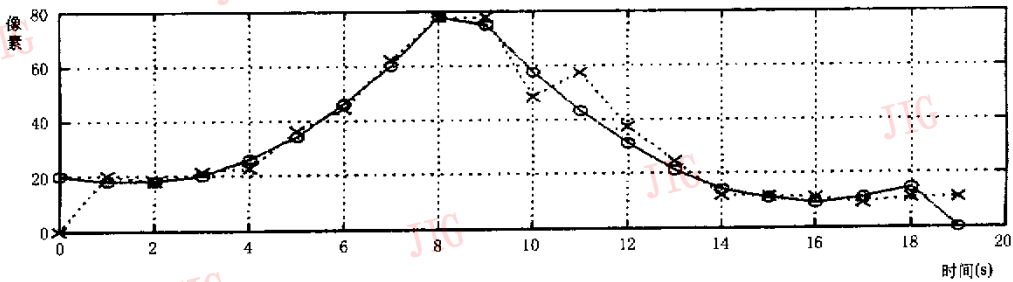
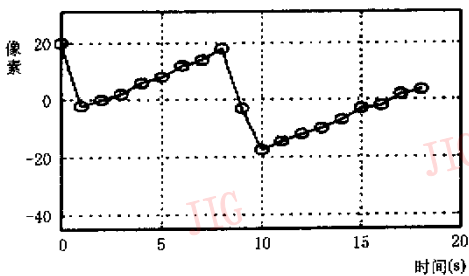
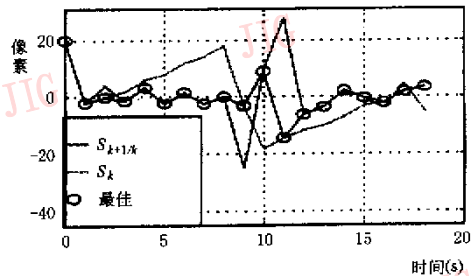


图 3 乒乓球的实际运动轨迹和预测运动轨迹



(a) 没有采用 Kalman 滤波的预测误差



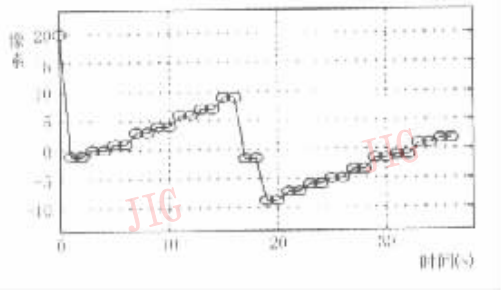
(b) Kalman 预测误差

图 4 两种方法对乒乓球的运动预测误差比较

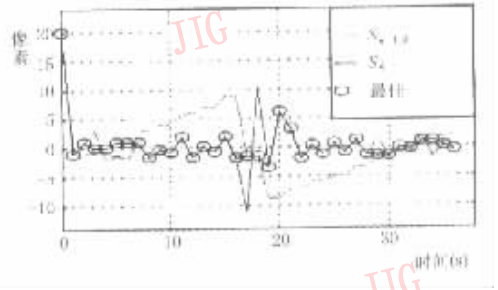
为了更好地使用 Kalman 滤波,可以在对象运动轨迹上线性地增加一些观测点,并将这些点用于 Kalman 滤波,其结果如图 5 所示.

最后,为了考察算法的鲁棒性,预先将“Table-Tennis”图象序列的 23 帧到 27 帧中的乒乓球用挡

板进行遮挡处理后,采用两种方法进行测试比较,其预测结果如图 6 所示.可以看到,使用原算法对“乒乓球”进行跟踪时,在 23 帧~27 帧中发生“死锁”现象(图 6(a)),且误差很大.而使用本文改进后的算法,克服了这种现象(图 6(b)).

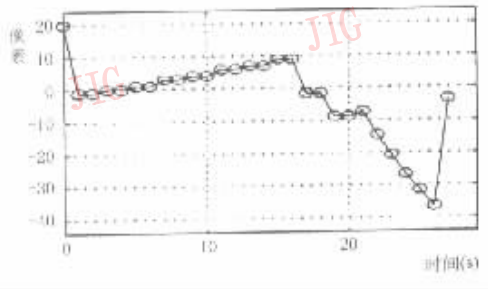


(a) 没有采用 Kalman 滤波的预测误差

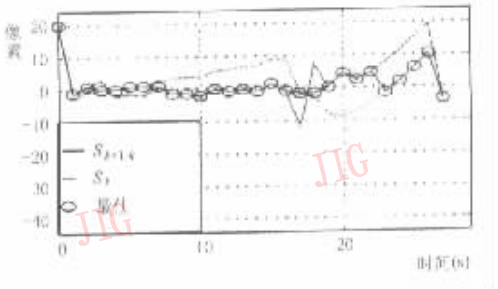


(b) Kalman 预测误差

图 5 添加测试点后两种方法的预测误差比较



(a) 没有采用 Kalman 滤波的预测误差



(b) Kalman 预测误差

图 6 乒乓球被遮挡后(23 帧到 27 帧)两种方法对其进行运动预测的误差比较

3 结 论

本文通过引入 Kalman 滤波,对 Moscheni 提出的视频对象分割与跟踪算法进行了改进.实验结果表明,这种方法能够更准确地预测对象的运动.而且提供给预测的参考点越多,预测准确性越高,但当对象的运动方式发生较大的改变时(运动模型参数发生变化),需要重新设置 Kalman 滤波的初始条件,利用对象的实际运动轨迹,对运动模型参数进行调整.同时,这种方法也很好解决了 Moscheni 算法中的“死锁”现象,表现出良好的鲁棒性.

参 考 文 献

- 1 Moscheni F. Spatio-temporal segmentation and object tracking: An application to second generation video coding[D]. Lausanne: Swiss Federal Institute of Technology, 1997.
- 2 Denzler J, Niemann H. Combination of simple vision modules for

robust real-time motion tracking[J]. European Transaction on Telecommunications, 1995,6(3):121~137.

- 3 Taubin G, Cooper D B. Object recognition based on moment (or algebraic) invariants[M]. MIT Press, 1992:375~397.
- 4 Shalom Y B, Fortmann T E. Tracking and data association[M]. Academic Press, Inc., 1988.



张江山 1970 年生,1992 年获华中科技大学(华中理工大学)电子与信息工程学士学位,1997 年获硕士学位,于 1999 年攻读博士学位,现为华中科技大学电子与信息工程系讲师.主要研究领域为视频处理、多媒体通信、计算机视觉等.



朱光喜 1945 年生,1969 年获华中科技大学(华中工学院)学士学位,现为华中科技大学电子与信息工程系教授,博士生导师.主要研究领域为计算机图形图像处理、多媒体通信、计算机协同工作等.