

基于TLD改进框架的视频目标跟踪算法

史殊凡, 孙光民

(北京工业大学信息学部, 北京 100124)

摘要: 研究了传统跟踪—学习—检测(Tracking-Learning-Detecting)目标跟踪算法的结构和特点, 提出改进思路; 虽然TLD算法采用P-N学习机制, 在应对长时间跟踪方面有很好的鲁棒性, 但是当目标发生严重遮挡、形变, 或者场景发生较大的光照、旋转变化时, 也会导致跟踪的失败。基于对以上问题的研究, 提出TLD改进跟踪算法。改进算法在跟踪模块运用SIFT特征匹配算法来代替原算法中LK光流法, 减少了计算的复杂度, 提高了算法的环境适应能力。

关键词: 目标跟踪; TLD算法; LK光流法; P-N学习; SIFT

中图分类号: TP391 **文献标识码:** A

Video Target Tracking Algorithm Based on Improved TLD Framework

SHI Shufan, SUN Guangmin

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: In this paper, the structure and characteristics of the traditional tracking-learning-detecting (TLD) target tracking algorithm are studied and the improved idea is proposed. Although the TLD algorithm adopts the PN learning mechanism, it has good robustness in dealing with long-term tracking. However, when the target is severely occluded, deformed, or when there is a large illumination or rotation change for the scene, it may also lead to failure of tracking. Based on the research on the above-mentioned problems, an improved TLD tracking algorithm is proposed. The improved algorithm uses the SIFT feature matching algorithm in the tracking module to replace the LK optical flow method in the original algorithm, which reduces the computational complexity and improves the environment adaptability of the algorithm.

Keywords: target tracking; TLD algorithm; LK optical flow method; P-N learning; SIFT

1 引言(Introduction)

视觉作为人类认知世界的重要方式之一, 人类所获得的外部世界信息大多来自人类视觉系统, 人类视觉系统承担了人类生活中的大量繁杂的信息与处理工作, 它可以快速并且正确地完图形描述、识别与处理等任务。因此, 在洞悉人类视觉本质的基础上, 计算机视觉技术应运而生, 该技术用各种成像系统代替人类视觉器官, 用计算机取代大脑, 来完成对人类接收的图像信息的处理和理解。计算机视觉技术的最终研究目标就是能够使计算机拟人化, 使计算机能够像人类那样思考, 通过视觉观摩和了解世界, 且具有自主的环境适应能力^[1]。

对于视频跟踪来说, 常用的方法主要有两种, 方法之一

是使用跟踪器^[2], 运用物体在上一帧的位置信息预测它在下一帧的位置情况, 但这样做有一个很大的缺点, 就是会积累误差, 并且一旦物体在视频帧中消失, 跟踪器就会失效, 即使物体再出现也无法完成追踪; 方法之二是使用检测器^[3], 对视频中的每帧单独进行处理, 来达到检测物体的位置的目的, 但是同样的, 这种方法的不足之处在于需要提前对检测器进行离线训练, 因此只能用于跟踪事先已知的目标。因此, TLD算法便被提出^[4]。该算法结合了传统的目标跟踪技术与目标检测技术, 当跟踪目标发生遮挡或者目标在镜头中消失从而导致跟踪失败时, 然后运用检测器检测出再次出现的目标, 再利用跟踪器去重新跟踪, 如此循环下去, 便可以满足长时间跟踪的需求。并且加入一种在线学习机制, 使整个系

统更加稳定、可靠。

但是,当目标出现严重遮挡、形变、光照变化时,仍会出现跟踪失败的问题,基于以上问题,提出了本文的改进方法。

2 TLD目标跟踪算法(TLD target tracking algorithms)

TLD算法是英国萨里大学的博士生Zdenek Kalal提出^[5]。主体结构由跟踪模块、检测模块和学习模块构成。结构如图1所示。

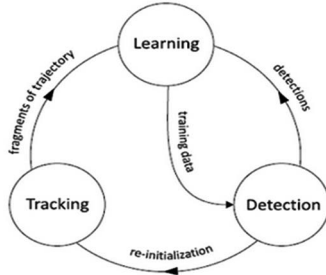


图1 TLD算法框架结构图

Fig.1 TLD algorithm framework diagram

2.1 跟踪模块

在TLD算法中,跟踪模块采用的是一种基于双向轨迹的LK光流跟踪法^[6]。光流,是背景中可见点的三维速度矢量在二维平面上的投影,它表示了背景表面的点在图像中位置的瞬间变化趋势。LK方法的三个假设:亮度恒定、时间连续和空间一致。

(1)假设原图为 $I(x, y, z, t)$ (这里是扩展到三维空间的,所以还有 z 值),移动后的图像是 $I(x + \delta x, y + \delta y, z + \delta z, t + \delta t)$,两者满足:

$$I(x + \delta x, y + \delta y, z + \delta z, t + \delta t) = I(x, y, z, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial z} \delta z + \frac{\partial I}{\partial t} \delta t + H.O.T \quad (1)$$

(H.O.T.指更高阶,可以忽略)

(2)其中图像移动可以认为:

$$I(x, y, z, t) = I(x + \delta x, y + \delta y, z + \delta z, t + \delta t) \quad (2)$$

即是满足:

$$\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial z} \delta z + \frac{\partial I}{\partial t} \delta t = 0 \quad (3)$$

(3)从这个方程中我们可以得到:

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y + \frac{\partial I}{\partial z} V_z + \frac{\partial I}{\partial t} = 0 \quad (4)$$

其中, $V_x = u$, $V_y = v$, 也就是光流的值(二维图像没有 z), $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, $\frac{\partial I}{\partial z}$ 则是图像在 (x, y, z, t) 这一点的梯度($\frac{\partial I}{\partial t}$ 是两帧图像块之间差值)。

(4)假设流(V_x, V_y, V_z)在一个大小为 $m*m*m(m>1)$ 的小窗中是一个常数,那么从像素1, $\dots$, n , $n=m*m*m$ 中可以得到下列一组方程:

$$I_{x1}V_x + I_{y1}V_y + I_{z1}V_z = -I_{t1} \quad (5)$$

$$I_{x2}V_x + I_{y2}V_y + I_{z2}V_z = -I_{t2} \quad (6)$$

$$I_{xn}V_x + I_{yn}V_y + I_{zn}V_z = -I_{tn} \quad (7)$$

通过观察,方程的个数比未知数的个数还要多,因此判定该方程组为超定方程组,换句话说就是方程组内有冗余,所以,方程组可以表示为:

$$\begin{bmatrix} I_{x1} & I_{y1} & I_{z1} \\ I_{x2} & I_{y2} & I_{z2} \\ \dots & \dots & \dots \\ I_{xn} & I_{yn} & I_{zn} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \\ V_z \end{bmatrix} = \begin{bmatrix} -I_{t1} \\ -I_{t2} \\ \dots \\ -I_{tn} \end{bmatrix} \quad (8)$$

也就是 $A\vec{v} = -b$,为了解决这个问题,采用最小二乘法^[7]:

$$\vec{v} = (A^T A)^{-1} A^T (-b) \quad (9)$$

得到:

$$\begin{bmatrix} V_x \\ V_y \\ V_z \end{bmatrix} = \begin{bmatrix} \sum I_{xi}^2 & \sum I_{xi} I_{yi} & \sum I_{xi} I_{zi} \\ \sum I_{xi} I_{yi} & \sum I_{yi}^2 & \sum I_{yi} I_{zi} \\ \sum I_{xi} I_{zi} & \sum I_{yi} I_{zi} & \sum I_{zi}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_{xi} I_{ti} \\ -\sum I_{yi} I_{ti} \\ -\sum I_{zi} I_{ti} \end{bmatrix} \quad (10)$$

其中,求和是从1到 n 。

在TLD算法里,把视频帧均匀画成网格状,并且将每个网格中的左上角位置的顶点作为特征点,接着采用双向轨迹光流法^[8],将这些特征点在下一帧中的位置通过跟踪预测出来。假设当前帧为 I_t 下一帧为 I_{t+1} ,普通光流法仅由当前帧 I_t 预测得到 I_{t+1} ,缺乏反向校验。而双向轨迹光流法则在由 I_t 预测出 I_{t+1} 之后,利用 I_{t+1} 中预测得到的点,再做反向预测,即由 I_{t+1} 预测到 I_t ,由这种双向轨迹预测法,得到一个位移差值。假如由反向预测得到的 I_t 中的特征点和之前得到的 I_t 中的特征点位移差值超过了设定阈值,则把 I_{t+1} 中预测出的偏差较大的特征点排除^[9]。如图2所示,即为双向轨迹光流法的基本原理图。

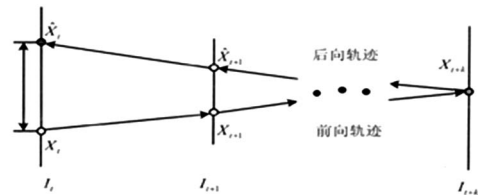


图2 双向轨迹光流预测原理图

Fig.2 Principle of optical flow prediction for bidirectional trajectory

2.2 检测模块

检测模块的主要功能是定位视频中每一帧目标的具体位置。TLD中的检测模块可以看作是一个级联分类器,它由方差滤波器、集成分类器和最近邻分类器这三部分构成。

方差滤波器作为级联分类器的第一部分,采用方差比较法,比较候选区域的方差与目标区域的方差,将候选区域中方差小于目标区域方差一半区域纳入被拒区域,通过这样的操作,大量的非目标区域便被排除在外。

集成分类器是第二部分。该部分采用随机森林^[10]进行分类。随机森林是一个包含多个随机森林的分类器。每一个

随机藏 K 都初始目标的像素点和在候选区域采集的像素点对灰度做差,生成二进制编码 X 。 X 指向决策树的后验概率。对于所有的决策树的这些后验概率取半朴素贝叶斯输出值 M ,进入下个分类器条件为 M 大于50%。当该种情况发生时那么就认为当前图像块含有目标,否则就进入拒绝区域。

最近邻分类器^[11]为第三部分,将通过前两个分类器剩下的矩形框,计算其相关相似度。相关相似度大于0.6的区域,即认定为最终前景目标。检测模块原理如图3所示。

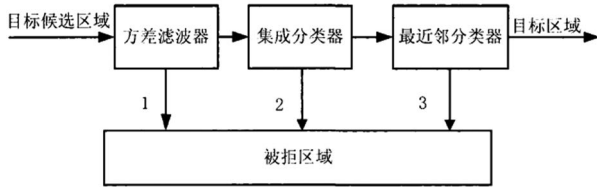


图3 TLD检测模块原理图

Fig.3 Principle diagram of TLD detection module

2.3 学习模块

TLD学习模块采用的机器学习方法是P-N学习^[12]。P-N学习是一种半监督模式^[13]的机器学习算法,它的独到之处在于,检测模块对样本分类时大多会有错误发生,错误主要分为两种,P-N学习采用P和N这两种专家对错误进行纠正:

P专家:测出漏检(正样本误分类为负)的正样本;

N专家:修正误检(负样本误分类为正)的正样本。

其原理图如图4所示。

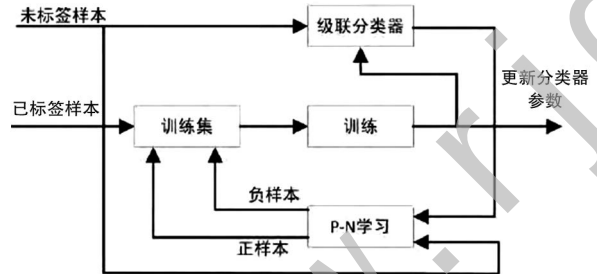


图4 P-N学习原理图

Fig.4 P-N learning principle diagram

3 TLD目标跟踪算法的改进(Improvement of TLD target tracking algorithms)

经典TLD算法的跟踪模块采用的是基于双向轨迹的LK光流法,其中假设亮度守恒是光流法使用的前提,但在实际应用中,由于遮挡、光度变化和噪声等干扰源的存在,亮度守恒定律往往不能满足。较大的光照变化可能导致目标的丢失,并且目标平面旋转、形变等情况也会导致光流法无法形成运动矢量场,大大增加目标位置的确定难度,甚至导致跟踪失败。

为了解决上面的问题,采用SIFT特征匹配算法^[14]来进行目标的跟踪。SIFT全称尺度不变特征变换。它的原理是将在图像中检测到的特征点,采用一个维数为128的特征向量来进行描述,所以根据上面的说法,一幅图像经由SIFT算法描述后,即成为一个维数是128的特征向量集,转化后的特征向量

集具有下列性质:(1)对图像缩放、平移、旋转不变等特征,具有一定不变性;(2)对于应对光照变化,仿射和投影变换,稳定性同样上一个台阶。

该算法的实现主要可分为六个步骤:

(1)建立高斯差分金字塔模型,进行尺度空间的构建。对原始图像进行尺度变换,得到图像多尺度下的尺度空间表示序列,并从这些序列中,提取尺度空间主轮廓部分。

(2)空间极值点检测^[15]。

(3)关键点的精确定位。由于DOG值对噪声和边缘敏感^[16]这一特点,在上步检测到的局部极值点还要进一步的筛选,以便去掉不稳定和误检出的极值点。

(4)关键点方向信息分配。通过求极值点的梯度来完成对方向的分配。对于上述关键点,其梯度幅值表述为:

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \quad (11)$$

梯度方向为:

$$\theta(x,y)=\tan^{-1}\left[\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\right] \quad (12)$$

(5)关键点表征。David G.Lowed^[17]的实验结果表明:对每个关键点,运用128维向量的描述子进行关键点表征时,综合效果最好。

(6)特征点匹配。量比较两组特征点的128维关键点的欧式距离^[18]来实现。得到的结果越小,说明相似度越高,当结果小于设定的阈值时,便可以判定匹配成功。

4 实验结果及分析(Experimental results and analysis)

实验以Visual Studio 2010、OpenCV 2.4.10作为开发平台,在Windows 7系统上进行测试。测试采用了五个基本测试集对TLD和TLD改进算法进行测试,并以carchase测试集为例进行了测试结果具体分析。表1对TLD算法和TLD改进算法成功跟踪的目标帧数进行了较为详细的对比。

表1 TLD算法和TLD改进算法成功跟踪目标帧数

Tab.1 TLD algorithm and TLD improved algorithm track target frame number successfully

视频库	视频总帧数	TLD	TLD改进
David	770	630	699
Football	362	168	329
FaceOcc2	812	727	811
Jumping	313	303	313
Carchase	3659	2655	3349

SIFT特征作为图像的局部特征,应对较差的跟踪环境时稳定性很好。首先对于David视频测试集,在跟踪过程中目标发生了较大的光照变化、平面旋转和目标短时间被遮挡的问题,TLD算法跟踪到了628帧,改进算法跟踪到699帧。对于Football测试集,在跟踪过程中,视频目标发生了自身形变和遮挡等问题,且目标消失一段时间后,又出现在视野范

围中,针对该种情况TLD算法跟踪效果较差,只成功跟踪了168帧,而改进算法跟踪效果较好,成功跟踪了329帧。对于FaceOcc2测试集,目标几乎发生了在前两个测试集中所遇到的全部变化,测试结果为:TLD算法跟踪到727帧,而改进算法几乎成功跟踪到每帧的目标。对于Jumping测试集,出现摄像头抖动问题,TLD算法成功跟踪了303帧,改进算法成功跟踪到了每一帧目标。最后对于Carchase测试集,由于帧数很多,TLD跟踪到了2655帧,改进后的TLD则成功跟踪了3349帧的目标。以下为具体分析。

图5为TLD算法在Carchase测试集下的跟踪结果。



图5 TLD算法跟踪结果

Fig.5 Tracking results of TLD algorithm

以上分别为在第60、98、216、673和1259帧的跟踪结果。如图5所示,TLD算法在216和673帧中并没有成功的跟踪到目标,在216帧因为车辆过桥的原因发生了较长时间的遮挡,在673帧则是摄像头发生旋转后才恢复正常拍摄。由此可见,TLD在应对长时间跟踪方面效果不错,但是当目标发生严重遮挡、形变、旋转等变化时,仍然会有跟踪失败的可能性。图6为TLD改进算法同样在Carchase测试集,上述帧的跟踪结果。



图6 TLD改进算法跟踪结果

Fig.6 Tracking results of improved TLD algorithm

如图6所示,改进算法在216和673帧同样跟踪到了目标。综上所述,说明改进算法确实提高了TLD算法在应对不利环境时的鲁棒性。

5 结论(Conclusion)

TLD跟踪算法是近些年来提出的高性能跟踪方法之一,受到了人们的极大关注。采用在线学习的方法,即使先验信息很少,也能通过学习追踪到目标的位置,为长时间跟踪开拓了道路。本文在原始TLD算法框架的基础上,引入SIFT特征匹配算法,使其与TLD进行结合从而得到了一种TLD改进跟踪方法。本文通过相应的视频数据集分别对原始TLD算法和TLD改进算法进行了测试,实验结果显示改进后的TLD算法在应对严重遮挡、形变,或者场景发生较大的光照、旋转变化时,同样有很好的跟踪效果,并且在长时间跟踪方面也不受影响。下一步对TLD的研究方向可以转向目标选择自动化和多目标跟踪,大大节约跟踪时间和成本,构建更加完整的跟踪体系。

参考文献(References)

[1] Dalal N, Triggs B. Histograms of oriented gradients for human

detection[C]. Computer Vision and Pattern Recognition, 2005. IEEE Computer Society Conference on. IEEE, 2005, 1: 886-893.

[2] Park J M, Choi H C, Oh S Y. Real-time vehicle detection in urban traffic using Ada Boost[C]. International Conference on Intelligent Robots and Systems, 2010: 3598-3603.

[3] X. Li, W. Hu, C. Shen, et al. A Survey of Appearance Models in Visual Object Tracking[C]. TIST, 2013.

[4] 冈萨雷斯. 数字图像处理[M]. 北京: 电子工业出版社, 2004.

[5] Yilmaz A, Javed O, Shah M. Object tracking: A survey[J]. ACM computing surveys (CSUR), 2006, 38(4): 13.

[6] Moravec H P. Visual mapping by a robot rover[C]. Proceedings of the 6th international joint conference on Artificial intelligence—Volume 1. Morgan Kaufmann Publishers Inc., 1979: 598-600.

[7] Shi C, Stephens M. A combined corner and edge detector[C]. Alvey vision conference, 1988, 15: 50.

[8] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.

[9] Kuehnl A. Symmetry-based recognition of vehicle rears[J]. Pattern recognition letters, 1991, 12(4): 249-258.

[10] Srinivasa N. Vision-based vehicle detection and tracking method for forward collision warning in automobiles[C]. Intelligent Vehicle Symposium, 2002. IEEE, 2002(2): 626-631.

[11] 伍健荣, 杜向龙, 刘海涛. 一种基于Kalman滤波器的自适应背景建模改进算法[J]. 传感器与微系统, 2012, 31(1): 52-54.

[12] 张蕾. 基于视频的交通运动目标检测与跟踪系统的设计与实现[D]. 哈尔滨工程大学, 2007.

[13] 李敏敏. 基于TLD模型的目标跟踪方法[D]. 西安电子科技大学, 2013.

[14] 张佳威, 支瑞峰. 光流算法比较分析研究[J]. 现代电子技术, 2013, 36(13): 39-42.

[15] Funk N. A study of the Kalman filter applied to visual tracking[R]. University of Alberta, Project for CMPUT, 2003(1): 26.

[16] 薛陈, 朱明, 刘春香. 遮挡情况下目标跟踪算法综述[J]. 中国光学与应用光学, 2009, 2(5): 388-394.

[17] 张帅领. 基于TLD的视频目标跟踪算法研究[D]. 西安电子科技大学, 2014.

[18] 薛亚阳. 基于图像处理的运行目标跟踪算法研究[D]. 西安工业大学, 2012.

作者简介:

史殊凡(1994-), 男, 硕士生. 研究领域: 信号与图像处理, 神经网络与模式识别.

孙光民(1960-), 男, 博士, 教授. 研究领域: 信号与图像处理, 神经网络与应用, 人工智能与模式识别.