

# 基于决策的人脸检测与自动跟踪方法

史双飞,张震

(上海大学 机电工程与自动化学院,上海 200072)

**摘要:** 对人脸检测与跟踪的研究现状进行了简要介绍,发现传统的MDP(Markov decision processes)跟踪算法需要手动初始化,这不利于实际场景中的灵活运用。因此,提出一种基于Viola-Jones人脸检测算法和改进的MDP自动跟踪算法。从视频序列中提取人脸的类Haar特征,采用AdaBoost算法构建强分类器,并利用级联方式将强分类器进行联合,从而提高人脸检测率。MDP跟踪算法将在线的多目标跟踪问题规划成MDP中的决策,为每一个人脸目标建立一个MDP模型,并用VJ检测器的输出来初始化该模型,将人脸的出现到消失看作是MDP模型中的状态转移,在跟踪过程中采用光流法结合Kalman运动估计提高人脸跟踪的准确性和鲁棒性,减少目标ID的分配错误。在此过程中VJ检测器作为监督指标,与跟踪器的输出进行关联。实验结果表明,该算法可以稳定地检测并跟踪场景内的人脸目标,其速率也能满足应用要求。

**关键词:** VJ算法;AdaBoost算法;MDP模型;光流法;Kalman运动估计

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2019)09-0035-05

doi:10.3969/j.issn.1673-629X.2019.09.007

## A Method of Face Detection and Automatic Tracking Based on Decision Making

SHI Shuang-fei, ZHANG Zhen

(School of Mechanical and Electrical Engineering and Automation, Shanghai University, Shanghai 200072, China)

**Abstract:** The present status of face detection and tracking is briefly introduced. It is found that the traditional MDP (Markov decision processes) tracking algorithm needs to be initialized manually, which is not conducive to the flexible use of the actual scene. Therefore, we propose an algorithm based on Viola-Jones algorithm and improved MDP automatic tracking algorithm. First, the Haar-like features of human faces are extracted from video sequences, and a strong classifier is constructed by using AdaBoost algorithm. The strong classifiers are combined by cascade method to improve the rate of face detection. Second, MDP tracking algorithm divides the online multi-target tracking problem into decision. An MDP model is established for each face target and initialized with the output of the VJ detector. The appearance and disappearance of human faces is regarded as state transition in MDP model. In the process of tracking, optical flow method combined with Kalman motion estimation is used to improve the accuracy and robustness of face tracking and reduce the ID switches. In this process, the VJ detector is also used as a monitoring index to correlate with the output of the tracker. Finally, experiment shows that the proposed algorithm can detect and track human faces stably in the scene, and its speed can meet the application requirements.

**Key words:** VJ algorithm; AdaBoost algorithm; MDP model; optical flow; Kalman motion estimation

## 0 引言

人脸检测与跟踪一直都是计算机视觉领域的研究热点,在视频监控、人机交互、运动分析等视频分析场景中应用广泛。人脸检测作为第一步骤非常关键,它的目标是找出图像中所有人脸对应的位置,并输出人脸外接矩形在图像中的位置。而人脸跟踪则需要在视

频序列中实时确定已被检测的人脸在后续帧中的运动轨迹。目前为止,在简单稳定的环境中,人脸检测与跟踪已取得了较好的成果。而在实际人流量较大的复杂场景中,遮挡、尺度变化、光照变化、杂乱背景、人脸姿态及表情等仍会在很大程度上影响人脸检测与跟踪的效果,如何增强检测跟踪算法的鲁棒性仍是研究的

收稿日期:2018-10-26

修回日期:2019-03-04

网络出版时间:2019-04-24

基金项目:国家自然科学基金(51005143)

作者简介:史双飞(1994-),女,硕士,研究方向为计算机视觉、机器学习等;张震,博士,副研究员,研究方向为运动仿生机理、移动机器人视觉、康复机器人、救援机器人等。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190424.1005.016.html>

重点。

目前已有的人脸检测算法分为基于知识和基于统计的方法<sup>[1]</sup>。基于知识的人脸检测算法主要是利用人脸特征的先验知识及规律来检测人脸,包括眼睛、眉毛、嘴巴、鼻子等器官特征,各器官相互之间的几何位置关系,人脸形状、边缘、纹理、肤色等特征。如建立 YCbCr 彩色空间的肤色模型<sup>[2]</sup>,利用人脸灰度特征和几何特征来定位人脸的方法<sup>[3]</sup>等,这些方法虽然特征明显,但易受光照、遮挡等外界因素干扰。基于统计的人脸检测算法是将人脸看作一个整体,根据统计学原理通过大量人脸图像样本构造人脸模式空间,根据相似度量来判断人脸是否存在。该方法主要通过机器学习算法训练大量人脸和非人脸样本,从而得到表征人脸的有效分类器。常用方法有神经网络法<sup>[4-5]</sup>、支持向量机法<sup>[6]</sup>、AdaBoost 算法<sup>[7-8]</sup>等,这类方法检测精度更高,适应性和鲁棒性均有了很大提升,只是需要事先训练大量样本,相较于实际应用过程中的大量在线计算,文中倾向于后者,因此选择 AdaBoost 算法进行人脸检测。

复杂场景中的多人脸跟踪方法种类繁多,其中应用较多的有:光流法<sup>[9]</sup>、Meanshift 算法<sup>[10]</sup>、Camshift 算法<sup>[11]</sup>、Kalman 滤波<sup>[12]</sup>和粒子滤波算法<sup>[13]</sup>等。然而仅单独使用这些传统的跟踪算法,不仅计算量大,鲁棒性不够,且不能在线学习,从而不能适应跟踪目标的变化,因此需要将这些方法进行组合或在此基础上形成鲁棒性较好且能在线学习新目标的算法。文中采用一种 MDP 算法<sup>[14]</sup>,将在线多目标跟踪规划成一个马尔可夫决策过程(Markov decision processes)中的决策,将每个目标的生命周期(从出现到消失)建模成一个 MDP,即将目标的出现和消失看作是 MDP 中的状态转移,然后使用现有的单目标跟踪方法来处理问题,可以泛化使用到各个场景的跟踪中。

文中在原始 MDP 算法的基础上,通过基于 AdaBoost 的 Viola-Jones (VJ)<sup>[8]</sup>算法快速检测人脸,将该结果作为 MDP 跟踪算法的目标检测输入,实现人脸自动检测。同时,为了增强新算法的鲁棒性,还在跟踪过程中加入了 Kalman 滤波进行运动估计,从而更好地处理目标 ID 分配错误的情况。

## 1 基于 AdaBoost 的 VJ 人脸检测算法

人脸检测作为人脸跟踪的第一道工序,其准确性将对后续人脸跟踪系统的性能和稳定性产生重大影响。人脸检测的目的是为了获取任务目标人脸区域的位置、大小等信息,为后续的跟踪奠定良好的基础。文中主要利用基于 AdaBoost 的 VJ 算法<sup>[8]</sup>进行人脸检测。

VJ 检测算法的整体框架分成三个部分。

### 1.1 引入积分图像计算目标的类 Haar 特征

与使用像素来描述人脸相比,使用类 Haar 特征不仅包含一定领域的信息,且计算速度快,因而在人脸检测算法中被大量使用。此外还引入积分图计算类 Haar 特征,只需遍历一次图像便可求出图像中所有区域的像素和。

### 1.2 采用 Adaboost 算法构建分类器

VJ 算法采用基于 Adaboost 的分类器学习方法,其核心思想是针对同一个训练集训练不同的弱分类器,将弱分类器相结合,从而构成一个强分类器。为了提升弱分类器,在一次学习结束后,对错误分类的样本通过增加其权重并重新学习以得到比上一次更好的分类器。具体步骤如下:

(1) 给定人脸图片  $(x_1, y_1), \dots, (x_n, y_n)$ , 其中  $x_i$  表示第  $i$  个样本;  $y_i = 1$  表示正样本(人脸);  $y_i = 0$  表示负样本(非人脸)。

(2) 初始化样本权值。

$$w_{1,i} = \begin{cases} \frac{1}{2m}, y_i = 1 \\ \frac{1}{2n}, y_i = 0 \end{cases} \quad (1)$$

其中,  $m$  为正样本数;  $n$  为负样本数。

(3) 首先,归一化权值:

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}} \quad (2)$$

其中,  $t$  是迭代次数。

其次,对特征  $j$ , 训练一个弱分类器  $h_j$ , 然后对每个特征生成的弱分类器计算权重误差:

$$\varepsilon_j = \sum_i w_i |h_j(x_i) - y_i| \quad (3)$$

最后,选择具有最小误差  $\varepsilon_i$  的分类器  $h_i$  叠加到强分类器中,并更新权值:

$$w_{t+1,i} = w_{t,i} \beta_i^{1-e_i}$$

其中,  $e_i = 0, \beta_i = \frac{\varepsilon_i}{1 - \varepsilon_i}$ , 表示样本分类正确;  $e_i = 1$  表示样本分类错误。

(4) 输出强分类器。

$$h(x) = \begin{cases} 1, \sum_{i=1}^T \alpha_i h_i(x) \geq \frac{1}{2} \sum_{i=1}^T \alpha_i \\ 0, \text{otherwise} \end{cases} \quad (4)$$

其中,  $\alpha_i = \log \frac{1}{\beta_i}$ 。

### 1.3 构建级联分类器

人脸检测中,只靠一个强分类器还不足以保证检测的正确率,需要一连串的强分类器联合在一起来提高检测正确率。文中利用筛选式级联分类器,先将多

个强分类器按复杂度排序,简单的在前面,类似于一个决策树,如图1所示。第一个简单而高效的分类器输出 True 结果就会触发同样具有较高检测率的第二个分类器对窗口图像做出评价。以此类推,第二个分类器输出 True 结果将触发第三个分类器对窗口图像做出评价。只要有一个分类器节点输出 False 结果,直接认为该窗口图像不包含目标物,从而拒绝该窗口。

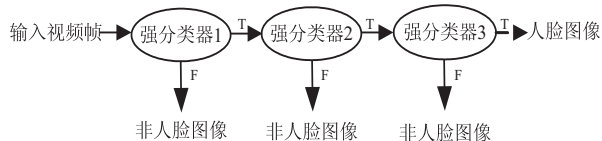


图1 级联分类器模型

## 2 MDP 跟踪算法

### 2.1 MDP 算法原理

基于 MDP 的跟踪算法将在线的多目标跟踪问题规划成 MDP 中的决策,为每一个目标建立一个 MDP 模型,将目标的出现到消失看作是 MDP 中的状态转移,并采用强化学习的方法学习相似数据的关联性,然后使用现有的单目标跟踪方法来处理问题。

### 2.2 MDP 的组成

用 MDP 来建模一个目标的生命周期,包括四个部分:目标的状态集  $S$ 、目标可采取的动作  $A$ 、状态转移函数  $T$  (描述各状态进行各动作后的影响)、实值奖励函数  $r$  (描述在状态  $s$  下执行动作  $a$  的实时奖励)。

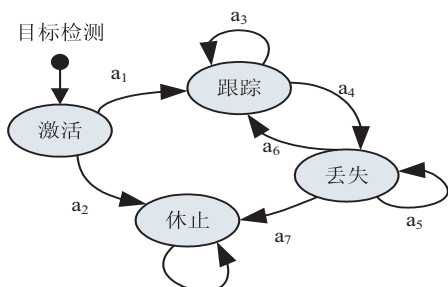


图2 目标的MDP模型

MDP 模型将目标分为四个状态:激活,跟踪,丢失,休止。如图2所示,当 VJ 检测器检测出人脸,随即进入激活状态,然后根据事先训练好的 SVM 分类器<sup>[15]</sup>判断该目标转移到跟踪或休止状态。处于跟踪状态的人脸目标通过光流法及 Kalman 运动估计继续跟踪,若人脸有遮挡或者离开相机的视野范围导致目标跟踪失败,则进入丢失状态。目标处于丢失状态时,先利用强化学习训练得到的二类 SVM 分类器,计算丢失目标与 VJ 检测器传来的目标之间的相似度分数,然后通过匈牙利算法<sup>[16]</sup>获得丢失状态下的目标集合与 VJ 检测目标集合间的最优分配。根据分配,可以将丢失状态的目标和某些检测目标关联,然后将其转

移到跟踪状态,否则就继续保持丢失状态。当目标处于丢失状态超过一定的阈值帧数后,就将其转移到休止状态,这是一个终结状态,不能再进行转移。文中着重介绍跟踪状态下所采用的策略及其改进方案。

### 2.3 MDP 跟踪状态下的策略

在跟踪状态下,MDP 主要决定目标是继续保持跟踪状态还是转移到丢失状态。若目标无遮挡且还在相机视野内,则保持跟踪,否则转移到丢失状态。

#### 2.3.1 光流计算

决策过程采用类似 TLD 跟踪算法<sup>[17]</sup>,首先用检测边界框在线初始化目标的外观模型,构建目标模板,利用光流信息进行跟踪,以正反向误差 (forward-backward) 作为评判标准。光流法用于计算两帧图像间对应点的运动矢量,即给定目标模板  $I$  上的一点  $u = (u_x, u_y)$ ,通过 Lucas-Kanade 方法<sup>[18]</sup>计算该点的光流找到新视频帧  $J$  上的对应点  $v = u + d$ ,其中  $d = (d_x, d_y)$  是  $u$  处的光流。在计算了所有采样点的正向光流之后,根据正反向误差来判断光流是否稳定,即根据  $J$  帧的目标检测值  $v$  反向计算目标模板  $I$  对应点  $u$  的光流,获得一个新的预测  $u'$ 。若光流稳定,则  $u$  和  $u'$  应该彼此接近,因此一个点的正反向误差可以用正反向预测之间的欧氏距离来表示:  $e(u) = \|u - u'\|^2$ 。若所有采样点的 FB 的中值  $e_{\text{medFB}}$  小于设置的阈值,则认为跟踪稳定,反之则不稳定。借此可以过滤掉不稳定的点,并用剩下的匹配点形成一个新的边界框,作为新的目标位置,并且在跟踪过程中,MDP 会收集目标的历史模板,便于后续操作。

#### 2.3.2 重叠区域计算

考虑到 VJ 检测器可能会存在错检,仅仅依靠光流法进行跟踪是不稳定的,因此引入一个新机制:当跟踪目标和检测目标有一段时间的不匹配,就认为其是错误目标。检查目标的历史模板,计算前  $k$  帧目标边界框  $t_k$  与相应的检测框  $D_k$  的重叠区域面积,并计算平均值  $o_{\text{mean}} = \text{mean}(\{o(t_k, D_k)\}_{k=1}^K)$ ,将其作为决策的另一个标准。

根据上述两个指标,可以得到特征表达式  $\varphi_{\text{Tracked}}(s) = (e_{\text{medFB}}, o_{\text{mean}})$ ,奖励函数如下:

$$R_{\text{Tracked}}(s, a) = \begin{cases} y(a), e_{\text{medFB}} < e_0; o_{\text{mean}} < o_0 \\ -y(a), \text{otherwise} \end{cases} \quad (5)$$

其中,  $e_0$  和  $o_0$  是指定的阈值。当  $a = a_3$  时,  $y(a) = +1$ ;  $a = a_4$  时,  $y(a) = -1$ 。也就是说当  $e_{\text{medFB}}$  小于设定的阈值,  $o_{\text{mean}}$  大于设定阈值的时候,目标会继续保持跟踪状态,反之转移到丢失状态。

#### 2.3.3 运动估计

在实际应用中,目标处于运动状态且每个目标的运动方向是随机的,因此若是目标之间产生相对运动



导致交叉重叠,那将会产生目标 ID 分配错误的情况,这是光流法的误差遗留问题。于是文中在原来的基础上引入了 Kalman 滤波,根据目标的历史信息来估算下一帧该目标可能出现的位置,然后将光流估计得出的边界框与运动估计得出的边界框进行比较,判断是否会出现反向,如果出现反向较大的情况,就认为跟踪失败,该目标应处于丢失状态。

在相邻帧内,运动目标的加速度可视为近似相等,因此可研究匀变速运动模型,并且假设目标没有运动突变。根据 Kalman 模型,假定目标运动方向的状态向量为  $\mathbf{X}_t = \begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix}$ ,其中  $x_t$  为  $\Delta t$  时刻的位置,  $\dot{x}_t$  为  $\Delta t$  时刻的速度,以此建立目标的运动方程:

$$\begin{cases} x_t = x_{t-1} + (\dot{x}_t \times \Delta t) + \frac{1}{2} a (\Delta t)^2 \\ \dot{x}_t = \dot{x}_{t-1} + a \times \Delta t \end{cases} \quad (6)$$

其中,  $a$  为运动目标在该方向的加速度;  $\Delta t$  为采样时间。其矩阵形式为:

$$\begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \dot{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} (\Delta t)^2 \\ \Delta t \end{bmatrix} a \quad (7)$$

则 Kalman 系统状态方程如下:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}u_{k-1} + w_{k-1} \quad (8)$$

其中,  $\mathbf{x}_{k-1}$  为状态向量;  $\mathbf{A}$  为转换矩阵;  $\mathbf{B}$  为系统参数;  $u_{k-1}$  为系统输入;  $w_{k-1}$  为系统噪声。

由以上两式可得出:  $\mathbf{A} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$ ,  $\mathbf{B} = \begin{bmatrix} \frac{1}{2} (\Delta t)^2 \\ \Delta t \end{bmatrix}$ , 系统噪声  $w_{k-1}$  可忽略。则  $k$  时刻的目标估计位置由  $k-1$  时刻的位置  $\mathbf{x}_{k-1}$  得到,并将此位置与光流计算得到的位置进行比较,若方向相反,则将目标转移至丢失状态。

### 3 实验测试与分析

为了验证算法的有效性,文中进行了 8 个不同场景的多人脸跟踪实验,并选择 3 个指标对该算法进行评价,分别是多人脸跟踪准确度 (MFTA)、多人脸跟踪精度 (MFTP) 以及跟踪速率 (Hz)。实验在 Inter Core i5 CPU 1.6GHz 和内存 8 G 的 PC 机上进行。

图 3 分别展示了 8 个不同场景的视频序列多人脸跟踪的实验结果,按顺序分别是 Crossing, Dance, Door, Elevator, Exercise, Lab, Road, Subway。可以看出, Dance, Elevator, Lab 等场景较为简单,且目标人物



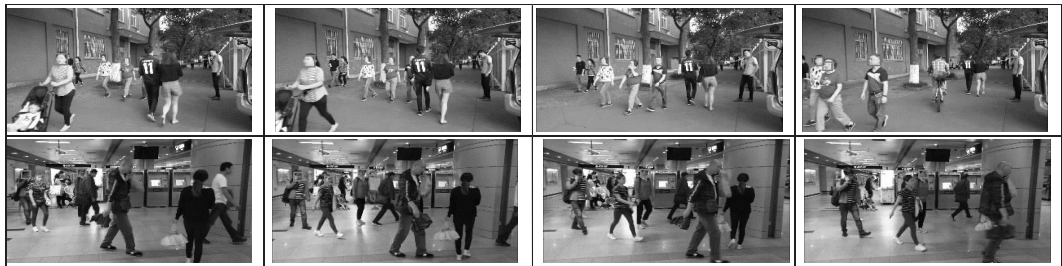


图3 视频序列的正脸检测与跟踪

基本都为正脸的情况下,人脸检测与跟踪的效果明显比其他场景复杂、人脸姿态各异的跟踪效果要好一些,且在目标被遮挡后仍能继续跟踪。可见该算法对目标人脸的跟踪具有一定的鲁棒性。

表1 改进算法的跟踪性能

视频序列	帧数	MFTA	MFTP	Hz
Crossing	222	33.6	71.8	8.0
Dance	276	79.4	82.1	8.8
Door	192	69.3	77.3	8.3
Elevator	197	78.9	76.5	8.5
Exercising	232	66.7	74.6	8.6
Lab	208	88.2	98.5	8.9
Road	152	77.1	71.1	7.8
Subway	132	44.3	70.2	7.6

表1显示了该算法的性能,除了 Crossing 和 Subway 两个场景的准确率稍低(推测可能是人脸与背景相似度较高,检测器出现了误检),总体来看,该算法基本可以稳定地检测并跟踪场景内的正脸目标,其速率也能满足应用要求。

4 结束语

MDP 算法是一种结合在线与离线学习的视觉跟踪算法,在其基础上,文中提出一种基于 VJ 人脸检测优化的 MDP 多目标跟踪算法,并且在跟踪过程中使用光流法结合 Kalman 运动估计来提高人脸跟踪的准确性和鲁棒性。此外,对于没有被跟踪的 VJ 检测目标,会初始化一个 MDP 继续跟踪。该算法能够有效适用于复杂场景中人脸目标的稳定跟踪。

参考文献:

[1] 龙伶敏. 基于 Adaboost 的人脸检测方法 & 眼睛定位算法研究[D]. 成都: 电子科技大学, 2008.

[2] 徐 翠, 李 然. 基于 YCbCr 空间的高斯肤色模型的人脸检测[J]. 软件导刊, 2009, 8(2): 179-181.

[3] 陈 莉, 刘智明, 周激流, 等. 一种基于颜色和几何特征的人脸检测方法[J]. 计算机工程与应用, 2003, 39(27): 80-82.

[4] SUNG K K, POGGIO T. Example-based learning for view-based human face detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(1): 39-51.

[5] BALUJA S, ROWLEY H A, KANADE T. Neural network

based face detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(1): 23-38.

[6] OSUNA E, FREUND R, GIROSI A F. Training support vector machines: an application to face detection[C]//IEEE computer society conference on computer vision & pattern recognition. Puerto Rico: IEEE, 1997: 130-136.

[7] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//IEEE computer society conference on computer vision & pattern recognition. Kauai, HI, USA: IEEE, 2001: 511-518.

[8] VIOLA P, JONES M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.

[9] 吴 垠, 李良福, 肖樟树, 等. 基于尺度不变特征的光流法目标跟踪技术研究[J]. 计算机工程与应用, 2013, 49(15): 157-161.

[10] 闫德莹. 基于 MeanShift 跟踪算法的交通流检测系统研究[D]. 合肥: 中国科学技术大学, 2011.

[11] 杨 杰, 穆平安, 戴曙光. 一种改进 Camshift 算法的研究[J]. 计算机应用与软件, 2014, 31(2): 167-170.

[12] 孙旺旺. 基于 Kalman 滤波跟踪的车道偏离预警方法研究[D]. 西安: 长安大学, 2015.

[13] 相入喜, 李见为. 多特征自适应融合的粒子滤波跟踪算法[J]. 计算机辅助设计与图形学学报, 2012, 24(1): 97-103.

[14] XIANG Yu, ALAHI A, SAVARESE S. Learning to track: on-line multi-object tracking by decision making[C]//IEEE international conference on computer vision. Santiago, Chile: IEEE, 2015: 4705-4713.

[15] BOSER B E, GUYON I M, VAPNIK V N. A training algorithm for optimal margin classifiers[C]//Proceedings of the 5th annual workshop on computational learning theory. New York: ACM, 1992: 144-152.

[16] MUNKRES J. Algorithms for the assignment and transportation problems[J]. Journal of the Society for Industrial and Applied Mathematics, 1957, 5(1): 32-38.

[17] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(7): 1409-1422.

[18] BOUGUET J Y. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm[J]. Intel Corporation, 2001, 5: 1-10.