

分类号 _____
UDC _____

密级 _____
编号 _____

中国科学院研究生院

硕士学位论文

基于检测的多人体目标跟踪算法研究

陈孝罡

指导教师 焦建彬 教授、博导

杨 辉 研究员

申请学位级别 工程硕士 学科专业名称 软件工程

论文提交日期 2011.4 论文答辩日期 2011.5.15

培养单位 中国科学院研究生院工程教育学院

学位授予单位 中国科学院研究生院

答辩委员会主席 于华

中国科学院研究生院直属院系

研究生学位论文原创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

中国科学院研究生院直属院系

学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密的学位论文在解密后适用本声明。

作者签名：

日 期：

导师签名：

日 期：

摘 要

人体目标检测与跟踪作为计算机视觉中的重要研究方向，融合了检测与跟踪两种重要技术，在视频监控、智能辅助驾驶等领域拥有很大的应用前景。

但是，由于实际应用环境的多变与复杂性、人体跟踪过程中的遮挡、尺度变化等难点问题，人体目标检测跟踪仍然距离实际应用具有很大差距，具有较广阔的研究空间。如何解决上述难点问题，使得人体目标跟踪具有较高的鲁棒性，对于实际应用系统具有重要意义。本文针对人体跟踪的问题，提出了一种针对跟踪的局部块组合表示模型。该模型把人体目标分为整体和局部两个组成部分来考虑：整体上使用一个覆盖人体的椭圆表示，并采用颜色直方图作为特征；在局部表示上，将人体目标划分为多个局部块，每个块中提取梯度方向直方图（Histogram of Oriented Gradient, HOG）和颜色直方图(Histogram of color, HC)作为特征。在跟踪过程中，提出了将整体表示与局部表示之间相互约束的表达方法。

在局部块模型基础上，本文提出了一种依据局部块运动趋势计算目标尺度变化，以及利用局部块相似度计算整体目标遮挡度的计算模型，并且将该模型融入到基于检测的人体目标跟踪算法过程中。实验表明，该模型能够有效处理跟踪过程中的尺度变化及遮挡等情况，最终保证跟踪算法的鲁棒性。

关键词：人体检测、目标跟踪、数据关联、局部块模型

A Detection Based Multiple Pedestrians Tracking Algorithm

Chen Xiaogang (Software Engineering)

Directed by: **Jiao Jianbin** (Professor)

The research on pedestrian detection and tracking is one of the most important branches of computer vision research. The technology combines two significant methods: detection and tracking, which have a very broad application perspective in visual surveillance and autonomous drive assistance systems.

But, due to fast changing and complexity in environment, existing pedestrian tracking procedures usually suffer from the variation and complexity of object background, occlusion and scale change, which makes pedestrian detection and tracking is still far away from practical usage, and leave lots of question for research. An effective way to address such shortcomings, and improve the robustness of procedure, will definitely bring significant improvement to the application system. This thesis proposes a local part based model for pedestrian tracking. The model has two components: the whole one is an eclipse covering whole human body which uses color histogram as features; the local one is a multiple local-part presentation in which we use Histogram of Oriented Gradient (HOG) and Histogram of Color (HC) as features in a local block. In the tracking procedure, proposed a mutual constraint to combine the global and local representation.

Based on the local part model, we define a way to predict object's scale change though local parts' divergence tendency, and use the local feature similarity to calculate a proposed occlusion degree. We integrate the approach into detection based tracking framework, and valid it. Experimental results on real dataset shows that the proposed method effectively deals with scale change and occlusion problems, which improve the robustness of the tracking algorithm.

KEY WORDS: Pedestrian Detection, Object Tracking, Data Association, Part-Based Model



目 录

摘 要.....	I
目 录.....	V
图目录.....	VII
表目录.....	IX
第一章 绪论.....	1
1.1 课题背景及意义.....	1
1.1.1 课题来源.....	1
1.1.2 课题背景与研究意义.....	1
1.2 国内外研究现状.....	2
1.3 本文的主要研究内容.....	5
1.4 本文的组织结构.....	5
第二章 目标检测与跟踪算法基础.....	7
2.1 人体目标检测的主要算法.....	7
2.1.1 人体目标特征描述子.....	7
2.1.2 人体目标检测分类器.....	8
2.2 目标跟踪的主要算法.....	11
2.2.1 跟踪目标的表观模型.....	11
2.2.2 跟踪目标的运动模型.....	12
2.3 基于检测的多目标跟踪的主要算法.....	14
2.3.1 基于组合最优化的多目标跟踪算法.....	14
2.3.2 基于组合概率最大化的多目标跟踪算法.....	16
2.3.3 基于人体运动信息的人体目标跟踪算法.....	20
第三章 基于局部块组合的人体目标跟踪算法.....	23
3.1 传统整体模型的局限性.....	23
3.2 基于局部块组合的跟踪人体表示模型.....	23
3.3 局部块特征提取与相似性度量.....	25
3.4 基于局部块模型的跟踪过程.....	27
3.4.1 跟踪模型的建立.....	27
3.4.2 尺度更新与遮挡度的计算.....	30
第四章 基于检测的多人体目标跟踪算法.....	33
4.1 人体检测算法.....	33
4.2 跟踪的初始化与终止.....	34

4.3 检测与跟踪间的数据关联	34
4.4 基于检测的多人体目标跟踪算法	36
4.5 实验结果比较与算法性能分析	36
4.5.1 实验数据集合.....	36
4.5.2 实验结果分析与比较.....	37
4.6 算法实时性	40
总结与展望.....	43
参考文献.....	45
致 谢.....	51

图目录

图 1 HOG特征提取过程	7
图 2 支持向量机原理图	9
图 3 规划算法网络图	15
图 4 文献 ^[20] 中所使用的人体骨骼模型	20
图 5 跟踪目标遮挡导致整体偏移	23
图 6 人体检测结果、整体椭圆模型、局部块分解模型	24
图 7 HOG算子的计算示意图 ^[30]	25
图 8 局部块跟踪与目标尺度变化	31
图 9 遮挡度的计算结果	32
图 10 基于检测的多人体目标跟踪算法流程图	36
图 11 算法在视频 1 中的测试结果示例	38
图 12 算法在视频中的错跟踪及漏跟踪示例	38
图 13 算法在视频 2 中的测试结果	39



表目录

表 1 Metropolis-Hastings 抽样算法过程.....	28
表 2 基于局部块模型的跟踪算法过程.....	30
表 3 数据关联中的贪心选择算法.....	35
表 4 跟踪算法量化比较结果.....	40

第一章 绪论

1.1 课题背景及意义

1.1.1 课题来源

本研究受到以下课题资助：

1.“天目 360 度智能监控系统产业化开发”，哈尔滨产业技术创新与育成专项资金项目，2011.01-2013.12，在研；

2. “视频中人体目标可靠检性检测方法与稳定性跟踪算法研究”，国家自然科学基金面上项目，（课题编号：60872143），2009.01-2011.12，在研；

3. “基于多源数据的飞行器进近威胁目标检测跟踪及行为预测”，国家自然科学基金重点项目（课题编号：61039003），2011.01-2014.12，在研。

1.1.2 课题背景与研究意义

目标检测与跟踪作为计算机视觉中的重要基础性算法，是许多相关技术与应用的基础。该算法地进步推动着计算机视觉及模式识别领域的发展。

人体检测和跟踪作为计算机视觉中十分具有应用前景的方向，近年来受到了许多研究人员的关注，展开了广泛的研究，如：视频监控、保护行人安全的车载行人检测系统等等。该技术在以下领域将发挥重要作用，在国防领域，人体目标检测可以应用于自动识别危险目标，在战场监控中发挥作用；在民用领域，该技术可用于重点建筑（核电站、水电站、水库、银行、重要桥梁等）及重要地点（国家重点保护建筑、博物馆、体育场馆、地铁站、监狱等）的安全保卫工作；在智能交通领域，可以及时监控交通事故并可以在十字路口、新建高速路端等进行自动行人检测和安全预警；也可在车载的智能系统中融入行人检测与跟踪技术，对驾驶员进行驾驶状况安全预警，有效避免车祸的发生。在民众日常生活保障上，可用于检测可疑人物可疑行为，用于小区监控，减少犯罪率，为人民正常生活提供保障。

近年来目标检测与跟踪技术的长足进步，使其已开始逐步应用于社会生活的

方方面面。目标检测技术主要研究在图像中对特定种类的目标（人、车等）进行检测与定位，通过特征提取方法，提取目标所共有的特性，导入分类器（如支持向量机，SVM）或者通过聚类算法（如 K -均值聚类），以实现目标与非目标的判别分析。而目标跟踪技术主要着眼于在连续的视频帧之间，通过提取运动目标的特征表达值，在连续帧中通过特征值匹配进行搜索定位。通过确定性极值或者概率极值的方式，得到视频序列中运动目标之间的对应关系。

在构建智能化的计算机视觉系统时，检测与跟踪技术是相辅相成的两方面。跟踪技术通常受限于需要指定跟踪目标以及初始化过程，而目标检测技术正好可以将检测的结果输入，作为跟踪算法的初始条件；同时在跟踪算法运行过程中，目标检测的结果可以作为指导性信息，对跟踪结果进行修正；另一方面，目标跟踪算法计算过程所得到的空间分布概率也能对目标检测算法作为修正，加强正确结果的假设及排除可能错误的检测结果。

近年来，该方面的研究工作在许多模式识别及计算机视觉领域的顶级国际会议如，IEEE International Conference on Computer Vision (ICCV)，IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)，European Conference on Computer Vision (ECCV) 上都有发表。同时在该领域重要期刊 IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)，International Journal of Computer Vision (IJCV) 上也刊载了很多相关工作。

尽管近年来计算机视觉及模式识别领域中目标检测与跟踪技术已获得了很大的发展，但复杂环境下人体目标的检测与跟踪离实际应用仍有距离，其难点主要是首先由于人体目标姿态的多样性，外观的复杂性等，导致人体检测算法在实际应用中检测效率不高；另外由于目标运动的不规律性，目标自身与外界的遮挡等，使得跟踪算法也面临限制，而多目标的检测跟踪，提出了更高的要求，这也成为该领域当前亟待解决的问题。

1.2 国内外研究现状

在静止图像中的人体检测方面，研究人员经过多年的研究和发展，已形成了几种比较实用、有效的算法。Dalal^[1]在 05 年提出了梯度统计直方图 (Histogram of Gradient, HOG) 特征，HOG 特征基于对梯度幅值在梯度方向上的投影的分块统计结果，主要表达了目标的轮廓信息，对于人体目标着装颜色的变化具有一

定鲁棒性，其在人体目标检测中取得了较好的效果。在 HOG 的基础上，Zhu^[2]等人提出了变尺度 HOG (v-HOG) 特征，相比原始 HOG 特征的固定尺度特性，其对于局部轮廓的变化更加敏感。在文献^[3]中，作者使用 Covariance (协方差) 算子特征来表示人体，由于协方差算子是由不同位置、不同尺度的对称正定矩阵组成，作者认为这些对称正定矩阵处在连通的黎曼流形上，很好的表达了人体目标多姿态多外观的特性。在文献^[4]中，作者引入了在材质分析中广泛应用的共生矩阵 (Co-Occurrence Matrices) 特征，结合变尺度 HOG，取得了十分明显的改进。文献^[5]则引入了在指纹识别中的局部二值模式 (Local Binary Pattern) 特征，LBP 对单调灰度变化有不变性，用于表达人体目标的纹理特性，同样取得了很好的效果。

在目标跟踪方面，基于视觉特征的目标跟踪研究大致可分为表观模型特征提取及运动模型分析两个层次。表观模型特征提取主要研究对跟踪目标的表观进行特征提取，通过事先定义的相似性量函数，进行全局的搜索与模板匹配。Mean Shift^[6]是该类方法最为典型的代表，其对目标的颜色表观进行高斯加权统计，以巴氏系数 (Bhattacharyya coefficient) 作为相似性度量，在邻域内通过梯度下降方法搜索最匹配目标。运动模型分析一般引入自动控制系统中状态空间的概念，利用状态空间对当前跟踪系统的运动进行模型化，卡尔曼滤波 (Kalman Filter)^[7]和粒子滤波 (Particle Filter)^[8]是其典型代表。当目标系统的运动呈线性分布，且观察模型也为线性时，卡尔曼滤波可对后验概率分布进行时序上的传递与更新。当系统为非线性结构时，粒子滤波 (Particle Filter)，也被称为时序蒙特卡罗 (Sequential Monte Carlo) 方法，是目前最为广泛使用的方法。粒子滤波方法基于在空间中的随机采样，通过积分建立后验概率密度函数，获得系统运动模型。

为了解决跟踪算法需要手工初始化目标位置的问题，研究人员开始引入检测算法，将其结果作为跟踪的初始化，并在此基础上发展出了许多的检测跟踪融合与多目标跟踪算法。

在当前的研究中，主要有三个趋势。

第一类方法主要依赖于视频序列中的人体检测的结果，将跟踪问题转化为在整个视频序列中，检测结果的关联问题，该类方法主要利用数学规划来求解问题。Perera^[9]等为了解决跟踪中多目标长时间遮挡后轨迹融合的问题，引入了匈牙利算法 (Hungarian Algorithm)，其通过对轨迹的相似度建立关联函数，将融合问题转化为数学规划中的指派问题，采用匈牙利算法解决。Jiang^[10]基于同样的思路，

将运动轨迹视为图中的节点，并为目标间的遮挡情况建立遮挡节点，通过线性规划方法来解匹配问题。Huang^[11]等建立了一个多层级的检测响应数据关联方法，通过在不同时间间隔上的数据关联，获得较为鲁棒的多目标跟踪结果，且有效的去除了目标检测中漏检和错检。在该方法的基础框架上，许多后续的改进方法也被提出^{[12]-[14]}，Zhang^[12]在该框架基础上，将匈牙利算法改用最小费用网络流算法来进行匹配计算，Li^[13]采用混合 Boosting 方法（Hybrid Boosting），通过统计学习过程，建立多目标跟踪过程中的遮挡与关联模型。Kuo^[14]通过在线学习的表观区分模型（Online Learned Discriminative Appearance Models, OLDAMs）来进行多目标之间的区分。尽管该类方法取得了较好的跟踪效果，但其是延后性的，不适合实时系统的要求。

第二类方法区别于第一种，主要针对实时性系统进行算法设计，不做延后性的轨迹分析，一般通过将目标检测的结果与目标跟踪的概率分布合并到一起求联合概率分布的极大，来求解跟踪目标的对应。Breitenstein^[15]等将检测结果融合到粒子滤波的采样中，将有检测结果支持的粒子赋予高权重。Wu^[16]等首先进行人体分块检测，通过贝叶斯融合得到人体目标，在跟踪过程中，针对目标的表观、运动、数据关联等，设计了一种概率分布函数作为目标的表征，融入到 Mean Shift 的方法框架中。Ess^[17,40]等提出了一套用于多目标跟踪的实时系统结构，其中运用了双目立体视觉以及目标检测与跟踪技术，在文献^[18]中，他们融合了车辆与行人的检测算法，通过扩展卡尔曼滤波（Extended Kalman Filter）建立运动模型，运用颜色统计直方图建立数据关联关系，实现了智能辅助驾驶系统的原型，在文献^[19]中，通过图模型，利用置信传播（Belief Propagation）方法建立整个系统中检测与跟踪间的数学描述，通过二值二次型优化（Quadratic Binary Optimization）解数据关联问题。

第三类方法则是从人行走时的躯干运动特性出发，在跟踪过程中融入人的运动特性。Andriluka^[20-22]通过图模型，将人体躯干四肢分解成图模型中的节点，并针对节点间的关联设计了运动状态概率分布函数，在整个算法过程中，将检测、跟踪以及人行走时的连贯关系，融入到一个求概率极值问题中。

实际应用中，目标检测与跟踪算法的鲁棒融合将是该研究方向主要趋势。同时，目标检测和跟踪算法各自的发展都将为该系统带来提升。人体检测由于复杂的动态背景、人体视角的变化、人体形态的多变性、目标遮挡、光线亮度信息的变化、检测的速度等问题，目前尚有很大的提升空间。而目标检测的理论与方法

依然处于发展当中，其性能的提高与技术的改进无疑将对整个应用系统有大的提升。在未来，研究或许将着重于多目标之间的互动，相互间运动模型的影响，遮挡的建模，这样当在复杂的动态环境下进行多目标跟踪是能够获得更高的鲁棒性。

1.3 本文的主要研究内容

本文针对人体目标的跟踪过程，提出了一种新的针对跟踪的基于局部块组成和整体目标结合的目标表示模型，旨在通过该模型，有效处理跟踪过程中的目标遮挡、形变、尺度变化等问题，并将其融入到基于检测的多人体目标跟踪算法过程中，构建完整的行人检测跟踪算法。本文的主要研究内容总结如下：

- **基于局部块组成的目标跟踪模型：**针对跟踪目标本身，提出了一种新的跟踪目标局部块表示模型，将跟踪目标标识为多个局部块，每个局部块中分别提取梯度方向直方图和颜色直方图特征，作为单独的目标在后续视频帧检索最佳匹配目标。由各个局部块的跟踪结果对整体目标的跟踪进行修正，并通过计算局部块的匹配程度来计算整体目标的遮挡情况，由局部块运动外扩或内敛的趋势预测整体目标的尺度变化。实验结果表明了所提方法的有效性，能正确捕捉到跟踪目标的相应变化。
- **基于检测的多人体目标跟踪算法：**在提出局部块模型的基础上，将该跟踪模型融入到基于检测的多人体目标跟踪算法框架中，依据人体检测的结果进行跟踪目标初始化以及局部块初定义与修正。同时，实现了构建整体算法过程中所需的相关支撑算法，包括：跟踪目标的初始化与终止，检测结果与跟踪目标之间的数据关联等。在公开数据集上的实验验证了所提出的算法的有效性。

1.4 本文的组织结构

第一章，绪论。主要论述了人体检测与跟踪算法的研究现状及其应用背景和意义。从检测、跟踪、基于检测的跟踪三方面阐述了当前的研究现状，列举了本文的研究目的和方法及着重解决的问题。

第二章，目标检测与跟踪算法基础。详细介绍了检测、跟踪、检测与跟踪结

合三方面研究中的主要研究内容与进展，介绍了主流的方法。

第三章，基于局部块组合的人体目标跟踪算法。论述了本文所提出的用于人体跟踪的局部块表示方法。介绍了局部块的定位方式及特征提取过程，整体目标跟踪与局部块跟踪之间的相互修正关系。基于所提模型，给出了通过局部块的匹配度计算整体模型的遮挡度，及通过局部块的运动计算整体的尺度变化的方法。通过实例验证了所提模型的有效性。

第四章，基于检测的多人体目标跟踪算法。将第三章所提的跟踪模型，融入到一般基于检测的跟踪算法框架中，详细介绍了整个算法框架中数据关联、跟踪初始化与终止等构建完整系统所需求的各个部分及所使用的方法。最后，给出了该算法在公开测试集上的实验对比结果，对所提算法在处理遮挡和尺度变化方面的有效性进行了验证。

最后，对本文的工作进行了总结与展望，对该研究的进一步深入和实用化等进行了探讨。

第二章 目标检测与跟踪算法基础

2.1 人体目标检测的主要算法

人体检测作为计算机视觉中的重要研究领域，近年来受到许多研究人员的关注，并发表了一些代表性方法。当前，人体目标的特征表达提取以及判别的分类器构建，是该领域学者们主要关注的研究方向。下面将就这两方面介绍一些典型的方法。

2.1.1 人体目标特征描述子

当前人体检测算法中最有代表性的特征描述子当属梯度方向直方图（Histogram of Oriented Gradients, HOG）特征，最早由 Dalal^[1]于 2005 年提出。借鉴于 SIFT（Scale Invariant Feature Transformation）算子的思想，HOG 特征基于对梯度幅值在梯度方向上投影的分块统计结果，来表征区域内目标的边缘结构。由于 HOG 特征由区域内梯度方向直方图统计计算得到，其对于较小的目标形变以及噪声具有较强的鲁棒性。

HOG 特征的提取过程如下图所示，在某一块状区域内，首先计算每个像素在水平以及垂直方向上的梯度幅值和梯度方向，然后将梯度幅值依据梯度方向，投影到直方图上，归一化后形成梯度方向直方图描述子。

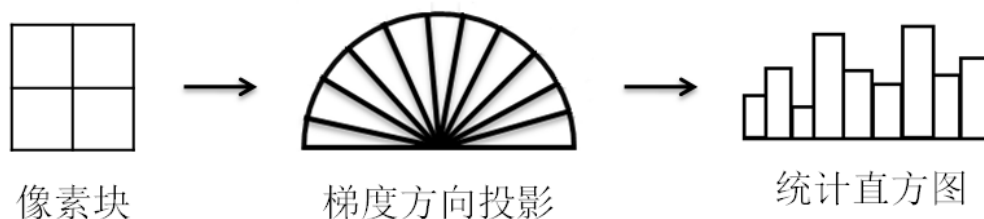


图 1 HOG 特征提取过程

在 Dalal 最早的工作中，将人体目标训练样本统一为 64×128 像素，并按照 8×8 像素为一个 Cell，田字型 4 个近邻的 Cell 组成一个 Block 的结构对整个训练样本进行分解，由此整个人体目标就形成了 105 个 Block。在每个 Cell 中有 9 个梯度投影方向，用 9 维的向量来表示，形成每个 Block 中就是 36 维的向量，最

最终以 $36 \times 105 = 3780$ 维的向量来表示人体。

在 2006 年, 为了改进 Dalal 算法中固定块模式下对于人体目标形变与姿态变化不敏感的缺点, Zhu^[2]提出了变尺度 HOG (V-HOG) 特征, 其对于 Cell 的处理不再拘泥于 8×8 像素的大小, 构造了如 8×16 、 16×8 等多种矩形局部块, 同时局部块的位置也可以在图像内任意选取, 不再定位于网格之上。V-HOG 特征是一种非常冗余的描述子, 一般要结合 Adaboost 分类器进行特征选取与分类器构建才能应用。

由于 HOG 特征是基于梯度统计得到的描述子, 其能很好的表达轮廓信息, 但对于颜色和纹理上的特性无法表达, 针对该缺点, 同时考虑到人体目标由于着装等因素, 其在颜色纹理上也表现出某种共性, Schwartz^[4]和 Wang^[5]先后在 HOG 特征的基础上引入了颜色共生矩阵和局部二值编码模式 (Local Binary Pattern, LBP) 作为纹理特征的补充, 而实验结果也表明, 在引入颜色纹理信息后, 检测的准确率有明显提高。

Tuzel^[3]提出了一种基于协方差矩阵 (Covariance Matrix, COV) 作为特征的 COV 算子。COV 算子对于目标, 选择像素位置, 像素灰度值、梯度以及角度等 8 个因素, 两两统计其协方差, 最终形成一个 8×8 维的矩阵作为目标的特征表示。其使用协方差的方式, 表示各个因素之间的统计关联, 形成了一个较完备的描述子。

近年来, 由局部块通过拓扑关系组合成整体表现目标的方法也越来越受到重视, 而新近的研究成果表明该方法在理论与应用上都存在研究价值^[34]。未来随着特征表达研究的深入, 人体检测也必将从中受益。

2.1.2 人体目标检测分类器

在人体检测领域常用的分类器主要是支持向量机 (Support Vector Machine, SVM) 和 AdaBoost 组合分类器。

SVM 由 Vapnik^[23]依据统计学习理论提出的模式分类和回归分析方法, 是一种基于结构风险最小化准则的机器学习算法。统计学习理论认为, 学习的风险有经验风险和结构风险两部分组成^[47]:

$$R_{emp}(w) + \sqrt{\frac{h(\ln(2n/h) + 1) - \ln(\eta/4)}{n}} \quad (2.1)$$

h 为学习机器的 VC 维。机器学习的过程不但要使经验风险最小化, 同时也要使

VC 维尽量减小以缩小置信范围，才能使实际风险减小，使得学习机器有好的推广性。SVM 基本原理如下图所示：

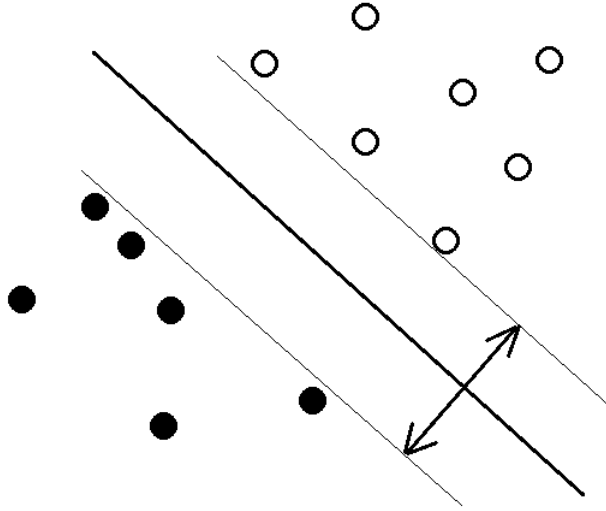


图 2 支持向量机原理图

图中以空心点和实心点分别代表两类样本，SVM 学习的过程为搜索一分类平面，使得两类样本间的分类间隔 $2/\|w\|$ 最大，等价于使 $\frac{1}{2}\|w\|^2$ 最小。从而整个分类问题就可归为一优化问题：

$$\text{minimize } \frac{1}{2}\|w\|^2 \quad (2.2)$$

$$\text{subject to } y_i(w^T x_i + b) \geq 1 \text{ and for all } i = 1, \dots, n \quad (2.3)$$

在线性可分的情况下，引入 Lagrange 优化过程，就可将上述问题转化为其对偶形式，即：

$$\max Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot y_j) \quad (2.4)$$

$$\text{subject to } \sum_{i=1}^n y_i \alpha_i = 0 \text{ and } \alpha_i \geq 0, i = 1, \dots, n \quad (2.5)$$

α_i 为与每个样本对应的 Lagrange 算子。求解该问题之后就可以得到分类函数为：

$$f(x) = \text{sgn}\{w \cdot x + b\} = \text{sgn}\left\{\sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x) + b^*\right\} \quad (2.6)$$

通常情况下，样本并不是线性可分的，因此，引入松弛项 ξ ，即考虑到在错分一定样本的情况下保持最大分类间隔。

$$\text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (2.7)$$

$$\text{subject to } y_i (w^T x_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \text{ for all } i = 1, \dots, n \quad (2.8)$$

其中 C 是惩罚系数，代表了对于错分样本的惩罚程度。对于非线性问题，另一种思路为将样本通过非线性变换到高维空间中，使得样本在高维空间中是线性可分的。但是非线性变换通常形式比较复杂，难以计算，而 Kernel 理论指出，通过符合 Mercer 条件的核函数实现非线性变换，该核函数就对应于高维空间中样本的内积。通过这种方式，就可以将原问题的样本映射到高维空间中求解。当前 SVM 研究中常用的核函数有线性核函数、径向基函数、Sigmoid 函数等。

AdaBoost^[25] 是一种弱分类器组合准则，其核心思想是由一组简单的弱分类器，以加权投票的方式，组合成一个强分类器。AdaBoost 分类器的训练过程就是一个迭代的训练分类器和样本权重更新的过程。

在训练分类器阶段，依据所有样本训练一弱分类器，利用该弱分类器对所有样本进行判别，对于分类错误的样本，加大其权重，也即样本权重更新过程。通过加权的方式，使分类器更加关注于难分类的样本。最终，各个弱分类器依其分类的准确率作为权重，组合成一强分类器。一般情况下，训练样本数量都比较大，造成弱分类器数量比较多，在应用 AdaBoost 时，通常要配合构建多级分类器的方式，以期通过更小的计算代价，达到与强分类器同样的判别效率。

Tuzel^[3] 在提出使用 COV 协方差算子作为人体特征的描述子之后，引入了流形中的判别方法。其所使用的 COV 算子为一正定矩阵，向量空间的分类器不适用于其分类问题。考虑到正定矩阵自然形成黎曼流形，同时黎曼流形在局部切空间同胚于一欧式空间，作者首先计算样本的均值，得到其在流形空间中的均值点，然后将样本映射到该均值点的切空间中，在该局部欧式空间中对样本进行分类。

近年来，随着压缩感知领域成果逐步被研究人员采纳，1 范数优化模型越来越

越多的应用到机器学习与模式识别领域中。Xu^[25]首先将 1-Norm Classifier 应用于人体检测领域，得到的实验结果表明，其与 2-Norm SVM 相比，具有更好的检测效率。在此基础上，其又提出了一种基于 1 范数最小化学习和线性规划组合的多级分类模型^[26]。实验结果表明，结合 V-HOG 的特征表达，该方法的检测结果比同样使用 V-HOG 特征加 Ada-Boost 组合分类器的方法有较大提高。

2.2 目标跟踪的主要算法

目标跟踪与目标检测一样，在计算机视觉领域也是研究热点。在几十年的发展历程中，科研人员引入了许多相关学科的内容如数学、自动控制、系统论等，而且跟踪技术具有的广阔应用前景，使其存在极大的研究价值。跟踪问题大体上也可归为两大研究方向，即目标的表观模型与目标的运动模型。

2.2.1 跟踪目标的表观模型

目标的表观模型也即如何表示目标，对其进行特征提取的过程。特征提取对于目标跟踪的过程有较大的影响，考虑跟踪目标与背景的差异，提取最具有差异性的特征以及目标所独具的特征，使目标的表观模型具有独特性，是该方向的研究重点。近年来，科研人员提出了几类表观模型，如下所述：

(1) 点模型

这类方法通过提取特征点来表示目标，其中，最著名的方法是 SIFT (Scale-Invariant Feature Transform, 尺度不变特征变换)算子，由 Lowe^[27]在 1999 年提出。SIFT 特征是图像的局部特征，其对于旋转、尺度变化、亮度变化保持不变性，这些特性使其在视觉领域得到广泛应用。其他应用较广的点模型还有 Harris 角点，KLT 等。

(2) 颜色模型

对目标颜色的 RGB 通道进行统计，并得到直方图描述，也是一种常用的特征。Comaniciu^[28]提出了一种带权式的颜色直方图统计方法，对离目标中心较远的像素赋予低权重，统计得到的颜色直方图配合 meanshift 方法，是视觉跟踪中最为典型的算法之一。

(3) 边缘模型

颜色模型的缺点在于其受光照的影响较大，而边缘模型对于光照变化适应性较好，因此也是一种优秀的特征提取方式。Canny 算子^[29]是边缘检测中应用较为

广泛的一种方法，基于对边缘信息进行统计的梯度方向直方图（HOG）在近年来也得到了广泛的应用，Han^[30]基于梯度直方图和颜色直方图，提出了一种综合直方图（HOGC）模型，并在跟踪的过程中，对于特征的权值调整，提出了一种依据判别力与卡尔曼滤波融合的特征权重更新模型，运用在目标跟踪上，取得了十分好的效果。

2.2.2 跟踪目标的运动模型

在提取目标的特征之后，就需要进行特征的比较与匹配用于进行跟踪目标的定位，而运动模型的好坏，将直接影响到目标定位的准确性。常用的运动模型有卡尔曼滤波（Kalman-Filter）^[7]和粒子滤波（Particle-Filter）^[8]。

卡尔曼滤波是由卡尔曼在 1960 年提出的一种用于解决离散数据线性滤波的递归方法。其计算量小，预测具有无偏、稳定和最优的特点。它的提出对现代控制理论的发展做出了巨大的贡献。在跟踪算法中，通常将目标的位置和速度设为系统的状态，由卡尔曼滤波进行预测与更新。卡尔曼滤波的基本原理如下：

令离散时间过程的状态变量 X 表征系统的状态，该变量由以下方程描述：

$$X_k = A \times X_{k-1} + B \times u_{k-1} + w_{k-1} \quad (2.9)$$

系统的观测方程为：

$$z_k = H \times X_k + v_k \quad (2.10)$$

w 和 v 分别为系统的状态噪声和观测噪声。假设它们相互独立，都为正态分布的白色噪声：

$$p(w) \sim N(0, Q) \quad (2.11)$$

$$p(v) \sim N(0, R) \quad (2.12)$$

状态的观测方程（先验估计）为：

$$\hat{X}(k|k-1) = A \times \hat{X}(k-1|k-1) + B \times u_{k-1} \quad (2.13)$$

经观测值修正过的后验估计为：

$$\hat{X}(k|k) = \hat{X}(k|k-1) + K_k(z_k - H\hat{X}(k|k-1)) \quad (2.14)$$

其中, $\hat{X}(k|k-1)$ 为 k 时刻的先验估计, 即系统由 $k-1$ 时刻的到 k 时刻的演化过程; $\hat{X}(k|k)$ 是 k 时刻的后验估计, 即由观测值 z_k 对状态向量进行修正;

$z_k - H\hat{X}(k|k-1)$ 是观测估计与实际观测之间的差值; K_k 成为滤波增益, 用来最小化 k 时刻的后验估计误差协方差。

卡尔曼滤波通过循环的预测-校正的过程, 来对系统状态进行追踪。卡尔曼滤波对于线性系统有良好的适用性, 但对于非线性系统, 其应用上存在不足。

粒子滤波 (Particle Filter), 即序贯蒙特卡罗 (Sequential Monte Carlo) 滤波, 是一种目前研究较多的非线性条件下的系统状态估计方法。其在计算机视觉中的应用^[8]与卡尔曼滤波一样基于贝叶斯时序滤波框架, 对于观测模型是非线性非高斯的系统时, 它提供了一种实现贝叶斯递归滤波的有效途径。其主要思想是通过带有相关权重的随机样本 (观测粒子), 基于这些样本来表示状态的后验概率分布 $p(x_k | z_{1:k})$ 。

令 $\{x_{0:k}^i, \omega_k^i\}_{i=1}^{N_s}$ 表示后验概率 $p(x_k | z_{1:k})$ 的随机观测值, 其中 N_s 是采样点数, ω_k^i 是 k 时刻第 i 个采样点的权重, 满足 $\sum_i \omega_k^i = 1$, 则 k 时刻后验概率分布 $p(x_k | z_{1:k})$ 可近似为:

$$p(x_{0:k} | z_{1:k}) \approx \sum_{i=1}^{N_s} \omega_k^i \delta(x_{0:k} - x_{0:k}^i) \quad (2.15)$$

粒子的权重 ω_k^i 可以通过重要采样 (Importance Sampling) 方法得到。若采样点集 $x_{0:k}^i$ 可由某重要度函数 $q(x_k | z_{1:k})$ 得到, 则第 i 个粒子的归一化权重为:

$$\omega_k^i \propto \frac{p(x_{0:k}^i | z_{1:k})}{q(x_{0:k}^i | z_{1:k})} \quad (2.16)$$

由 Markov 过程, $p(x_k | z_{1:k})$ 也可表示为:

$$\begin{aligned} p(x_{0:k} | z_{1:k}) &= \frac{p(z_k | x_k) p(x_k | x_{k-1})}{p(z_k | z_{1:k-1})} p(x_{0:k} | z_{1:k-1}) \\ &\propto p(z_k | x_k) p(x_k | x_{k-1}) p(x_{0:k} | z_{1:k-1}) \end{aligned} \quad (2.17)$$

2.3 基于检测的多目标跟踪的主要算法

随着计算机视觉、模式识别技术的研究与发展,在许多应用场景中,构建整个系统时要同时考虑目标检测与目标跟踪的问题,如车载的行人检测辅助驾驶系统、视频监控中人的行为分析等,研究人员开始关注于如何将这两种方法有机的结合到一个框架内,期望两者能够相互补充和彼此修正。下面就逐一介绍当前该研究中,主要的三种研究思路。

2.3.1 基于组合最优化的多目标跟踪算法

基于组合最优化的多目标跟踪算法,其着重于解决多目标跟踪中数据关联的问题,即在视频序列中,连续帧之间多个跟踪目标的匹配与互连问题。该类方法依据某种相似性度量,如表观模型和运动模型等,将该相似性度量作为目标间的关联系数,利用数学规划中的最优化算法来求解数据关联问题。该类方法的典型代表有匈牙利算法(Hungarian Algorithm)^[9,41]、线性规划(Linear Programming)^[10]和最小网络费用流(Min-Network Cost Flow)^[12],蒙特卡洛-马尔科夫链(Monte Carlo Markov Chain, MCMC)^[42]等。

匈牙利算法最早由 Perera^[9]应用到多目标跟踪上,用于解决多个跟踪目标在长时间遮挡丢失后,重新跟踪与轨迹连接的问题。所求解的问题可以描述为,假设有 n 个跟踪目标,而在当前帧同样有 n 个观测值,要将跟踪目标与新的观测值一一对应连接,问题就转化为了运筹学中的任务指派问题,该问题可用匈牙利算法求解。其计算过程首先是对各个目标与观测值之间,依据运动模型和外观的相似性度量赋予权值,建立基本的二分图,通过迭代寻找增广路径的方式求最大匹配。

线性规划与网络流方法都是基于网络图模型所实现的算法,其基本思想都是,对于视频序列,将每一帧视为网络中的每一层,而每一帧中由人体检测得到的结果就作为每一层中的节点,这样,整个视频序列就构建为一个多层的网络图。不同层的节点之间的连接边权值代表了两个检测结果之间的相似性,同样基于表观和运动模型,将相似性度量的函数值赋予各边。最后对整个网络图加上虚拟的源点与汇点,就能构建出该方法的基本模型。其示意图如下:

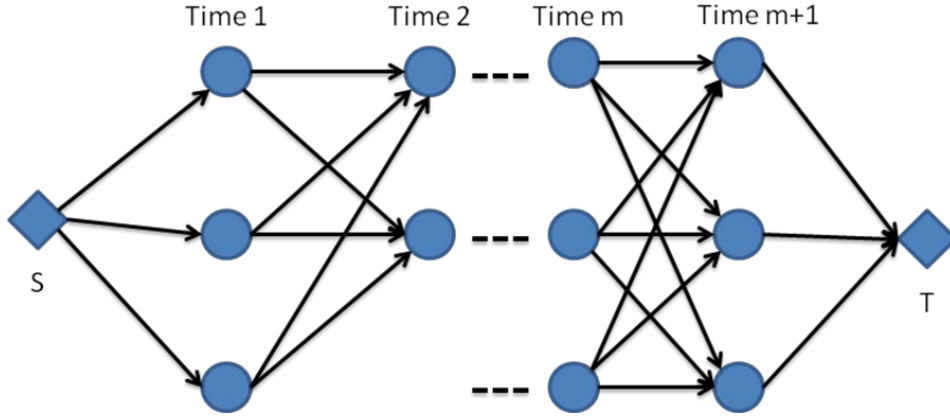


图 3 规划算法网络图

在 Jiang^[10]的论文中，引入了遮挡节点（Occluded node）用于处理跟踪过程中目标间出现遮挡和漏检测的情况，当一个目标在下一层中连接到一个遮挡节点时，可认为该跟踪目标在下一帧中处于遮挡状态或者是检测器发生了漏检，未来帧中该目标再次出现时，又可继续跟踪关联的步骤。这样的处理方式很好的解决了跟踪目标发生遮挡时的问题，提高了算法的鲁棒性。其建立的优化模型为：

$$\min_{\text{paths}} \sum_{n,m} \sum_{\substack{(v_{n,m,i}, v_{n,m+1,j}) \\ \text{on a path}}} \text{Cost}(v_{n,m,i}, v_{n,m+1,j}) + \mu \sum_m \sum_{\{n_1, n_2\} \in \mathbb{N}} \left\| \text{Size}_{n,m,i} - \text{Size}_{n,m+1,j} \right\| \quad (2.18)$$

Subject to at most one path goes through $O(v_{n,m,i}), \forall v_{n,m,i}$

其中 $\text{Cost}(v_{n,m,i}, v_{n,m+1,j})$ 即网络图中每条边的权值，代表了 $v_{n,m,i}$ 和 $v_{n,m+1,j}$ 节点的关联代价，通常包含两部分：选择该节点的代价和 i 与 j 节点的连接代价。目标方程中的第二项相当于是对跟踪目标跟待关联检测结果之间的大小提供了一致性约束。由于该优化问题为离散且不连续的，作者引入参数 $\xi_{(n,m,i),(n,m+1,j)}$ 表示某条边是否在目标 n 的路径上，是为 1 否则为 0，就可将上述问题松弛成为以下的线性规划问题：

$$\min_{\text{paths}} \sum_{\text{For all edges } (v_{n,m,i}, v_{n,m+1,j})} \xi_{(n,m,i),(n,m+1,j)} \cdot \text{Cost}(v_{n,m,i}, v_{n,m+1,j}) + \mu \sum_m \sum_{\{n_1, n_2\} \in \mathbb{N}} \left\| \text{Size}_{n,m,i} - \text{Size}_{n,m+1,j} \right\| \quad (2.19)$$

再通过一般的线性规划方法求解。

Zhang^[12]基于同样的网络图模型，但运用了不同的推理建模方式来构建数学模型，其过程为令 $\chi = \{x_i\}$ 表示检测算法得到的结果， $x_i = (x_i, s_i, a_i, t_i)$ ，其中 x_i 代

表位置, s_i 代表尺度, a_i 是表观模型, t_i 代表结果所在帧序。以 $T_k = \{x_{k_1}, x_{k_2}, \dots, x_{k_l}\}$ $x_{k_i} \in \chi$ 代表跟踪轨迹, 整个视频序列的跟踪轨迹集合为 $T = \{T_k\}$ 。整个数据关联问题就是求在给定 χ 的情况下求 T 的后验概率极大的问题, 即 $\arg \max_T P(T | \chi)$ 。

与 Jiang 的思路相同, 其首先对检测结果集扩充, 加入遮挡假设集, 与前面方法不同的是, 遮挡假设集的建立依赖于对现有检测结果之间的分析, 当两个目标之间的距离 $|x_i - x_j|$ 以及尺度比值 s_i / s_j 小于某个阈值时, 就在集合中加入遮挡假设, 由此形成最终的网络图模型并进行求解。

此类方法的实验结果验证了方法的有效性, 但其所共有的局限是方法的应用过程要使用视频序列中的多幅连续图像, 不是一个实时的算法, 不能应用在有实时性要求的场合, 而比较适用于行为分析等领域。

2.3.2 基于组合概率最大化的多目标跟踪算法

在目标检测的过程中, 通常都包含一个概率输出的过程, 如 SVM 的输出可认为是目标判别的置信度, 由于目标检测过程一般使用滑窗扫描的方式, 该过程就可以在图像上得到目标的概率分布。受此启发, 基于组合概率最大化的多目标跟踪算法利用检测过程得到的概率分布, 结合目标跟踪得到的概率分布, 将这两种概率分布融合到一个数学模型中去求解问题^[17-19,31,39]。

Leibe^[19]在构建车载的行人、车辆检测与跟踪系统时, 提出了一种目标检测与跟踪结合的新思路: 对目标跟踪而言, 在整个过程中若有足够多的检测结果关联支撑, 则可认为跟踪过程具有较高的置信度; 另一方面, 跟踪又给检测提供了某种场景先验, 给出了目标的分布概率, 可以对检测过程提供线索。考虑到这两者间的关系不是对称的, 即跟踪依赖于检测的结果来对目标进行更新, 但是对于新进入场景的目标, 现有的跟踪目标中没有能与其对应的跟踪轨迹, 因此需要设定一个虚拟轨迹与检测结果连接, 避免新的检测结果因为没有关联的跟踪目标而被算法去掉, 造成目标漏检。作者建立问题的模型为:

给定跟踪集 $\{m_i\}$, 检测结果集 $\{n_i\}$, 同时为每个检测结果构造一个虚拟轨迹, 集合为 $\{v_i\}$, 求:

$$\max_{m,v,n} \begin{bmatrix} m^T & v^T & n^T \end{bmatrix} \begin{bmatrix} \widetilde{Q} & U & V \\ U^T & R & W \\ V^T & W^T & S \end{bmatrix} \begin{bmatrix} m \\ v \\ n \end{bmatrix} \quad (2.20)$$

式 2.20 中矩阵的各个部分的含义如下：

\widetilde{Q} 构建了目标的跟踪过程，对其中每个元素，作者称为一种假设，表示跟踪目标自我更新及相互作用的过程，其中的元素为：

$$\widetilde{q}_{ii} = -\varepsilon_1 c(\mathcal{H}_{i,t_0:t}) + \sum_{H_{k,t_k} \in \mathcal{H}_{i,t_0:t-1}} ((1-\varepsilon_2) + \varepsilon_2 g_{k,l}) \quad (2.21)$$

$$\widetilde{q}_{ij} = -\frac{1}{2} \sum_{H_{k,t_k} \in (\mathcal{H}_i \cap \mathcal{H}_j)_{t_0:t-1}} ((1-\varepsilon_2) + \varepsilon_2 g_{k,l} + \varepsilon_3 O_{i,j}) \quad (2.22)$$

$$g_{k,i} = p^*(H_{k,t_k} | I_{t_k}) + \log p(H_{k,t_k} | \mathcal{H}_i) \quad (2.23)$$

$c(\mathcal{H}_{i,t_0:t})$ 是模型代价，对跟踪轨迹中的断裂部分进行惩罚。

H_{k,t_k} 代表 t_k 时刻的检测结果， I_{t_k} 则为图像信息， \mathcal{H}_i 为跟踪轨迹，表示为检测结果序列。

$g_{k,i}$ 代表了一种置信度量。

\widetilde{q}_{ii} 表示的是根据跟踪目标更新假设，表示目标模型自我更新，没有与别的目标发生遮挡。

\widetilde{q}_{ij} 表示的是目标之间的遮挡假设，其数值代表了该假设的置信度。

由于作者求解的模型是一个二次 0-1 规划问题，因此对于该矩形 \widetilde{Q} 的求解直观说明就是，若跟踪目标之间没有发生相互遮挡，则算法就会选择对角线上的元素；若存在目标遮挡情况，则会选择相应的非对角元素。通过这样巧妙的过程，对跟踪目标之间的遮挡作用进行了建模。

R 代表的是虚拟轨迹，其各元素为： $r_{ii} = \varepsilon$ ， $r_{ij} = 0$

如上所述，引入虚拟轨迹的意义在于使新的检测结果不会被算法剔除，在得到足够多的连续检测结果支撑之后，该检测目标也会形成新的轨迹，加入到跟踪

过程中。

U 代表了虚拟轨迹与实轨迹之间的作用, 类似于实轨迹之间作用的计算形式:

$$u_{ij} = -\frac{1}{2}((1-\varepsilon_2) + \varepsilon_2 g_{k,l} + \varepsilon_3 O_{i,j}) \quad (2.24)$$

\tilde{S} 物体检测矩阵

$$\tilde{s}_{ii} = -k_1 \varepsilon_1 - (1 - \varepsilon_2) \quad (2.25)$$

$$\tilde{s}_{ij} = -\frac{1}{2A_{s,v}} \sum_{p \in \text{Seg}(H_i \cap H_j)} ((1-k_2) + k_2 p(p = \text{fig} \cdot | H_k)) - \frac{1}{2} \log p(H_k) \quad (2.26)$$

对角线元素 \tilde{s}_{ii} 表示的是检测置信度, 非对角线元素 \tilde{s}_{ij} 含义为图像中检测结果之间存在的遮挡、对像素进行的竞争。 $A_{s,v}$ 表示的是表观模型。

V 和 W 表示的是跟踪轨迹与检测结果之间的相互作用矩阵:

$$v_{ij} = \frac{1}{2}((1-\varepsilon_2) + \varepsilon_2 p^*(H_j | I_t) + \varepsilon_2 \log p(H_j | \mathcal{H}_i)) \quad (2.27)$$

$$w_{ij} = \max_i [v_{ij}] \quad (2.28)$$

整个问题的求解是一个二次的 0-1 规划问题, 对于该优化问题, 有几点重要的性质可以帮助理解:

1. 如果某个跟踪轨迹被选为 1, 则说明其对应的 \tilde{q}_{ii} 大于某给定阈值。
2. 如果某个轨迹 \mathcal{H}_i 和相容的检测结果 H_j 都被选为 1, 则 H_j 对整个轨迹的置信度影响为 v_{ij} 。
3. 如果某个检测结果 H_j 没有与任何轨迹相关联, 但是其置信度 \tilde{s}_{ii} 大于某个阈值时, 其也会被选为 1, 且对其虚拟轨迹的置信度以 w_{ij} 影响。这样就保证了新的检测结果若有较高的置信度, 则可以生成新的跟踪目标。
4. 若是某个检测结果与跟踪轨迹存在对应性, 则其虚拟轨迹并不会被选为

1, 因为其对应的轨迹作用元素 u_{ij} 和轨迹置信度 r_{ii} 都会小于实轨迹的置信度。

5. 综上所述, 各个检测结果对于像素的竞争, 即图像中的区域属于哪个物体, 由 \tilde{s}_{ij} 来体现; 而各个跟踪目标之间的作用, 则由 \tilde{q}_{ij} 来进行计算。

作者通过这样一个完备的模型, 将整个系统中所可能遇到的问题, 如数据关联, 检测修正, 跟踪过程的更新, 新检测的加入等等, 统一到了一个数学模型中。

Breitenstein^[31]提出了一种将人体检测与粒子滤波框架相结合的在线多人跟踪算法, 其主要思想是将检测结果的置信度作为跟踪粒子的指导, 对粒子权重进行修正。因此该方法不再强调检测结果的作用, 检测只是对跟踪过程进行辅助, 提高了算法的鲁棒性。算法首先同样涉及检测结果与跟踪目标之间的数据关联问题, 作者在此使用了贪心搜索算法, 其过程如下: 首先, 对于每个跟踪-检测对, 首先通过

$$S(tr, d) = g(tr, d) \cdot (c_{tr}(d) + \alpha \cdot \sum_{p \in tr} p_N(d - p)) \quad (2.29)$$

求出两者间的匹配分数, 这其中: $g(tr, d) = p(size_d | tr)p(pos_d | tr)$ 依据距离和大小进行度量。

$c_{tr}(d)$ 是针对轨迹在线学习的分类器, 可以理解为一般的特征提取后求匹配的过程。

$p_N(d - p)$ 表示的是检测结果与跟踪目标之间距离的正态分布。

由所有的匹配分数形成关联矩阵, 后由贪心算法求解数据关联问题。作者的实验表明, 该贪心算法能取得与匈牙利算法相近的结果, 同时计算代价更小, 极大的提升了效率。在融入检测结果之后, 跟踪粒子的权重为:

$$w_{tr,p} = \beta \cdot I(tr) \cdot p_N(p - d^*) + \gamma \cdot d_c(p) \cdot p_0(tr) + \eta \cdot c_{tr}(p) \quad (2.30)$$

其中:

$I(tr) \cdot p_N(p - d^*)$ 代表关联的检测结果对粒子的修正, $I(tr)$ 是指示函数, 当有检测结果关联时为 1, 无检测关联时为 0, $p_N(p - d^*)$ 是一正态分布, 表示粒子跟检测结果之间距离远近与大小相似度。

$d_c(p) \cdot p_0(tr)$ 是检测置信度项，由检测算法直接输出。

$c_r(p)$ 是相似度度量函数，表示粒子在新位置得到的目标与原目标之间的相似度。

β 、 γ 、 η 是模型参数，由人为给定。

通过这样的建模过程，作者将人体检测和目标跟踪以组合概率的形式融入到了一个数学模型当中去求解。

该方法的创新之处在于，作者有效的利用了人体检测过程中得到的概率分布，将其融入到跟踪的过程中，相当于检测概率与粒子滤波得到的跟踪概率两者叠加得到最终所要的跟踪目标的概率分布，

基于组合概率最大化的多目标跟踪算法与前面所提到的基于组合最优化的算法不同的是，其针对的是实时的过程，对模型的更新、优化计算只在相邻帧之间进行，而不用考虑整个视频序列，因此可以应用在实时性的系统上。

2.3.3 基于人体运动信息的人体目标跟踪算法

近年来，随着图模型在目标检测与跟踪中的应用日渐广泛，研究人员对其也进行了深入的研究。**Andriluka**^[20-22]提出了一种基于图表示模型的人体目标检测与跟踪算法。其核心思想是将人体依据关节组合，分为多个局部块，各个块之间又连接成整体，其采用的骨骼模型如下所示：

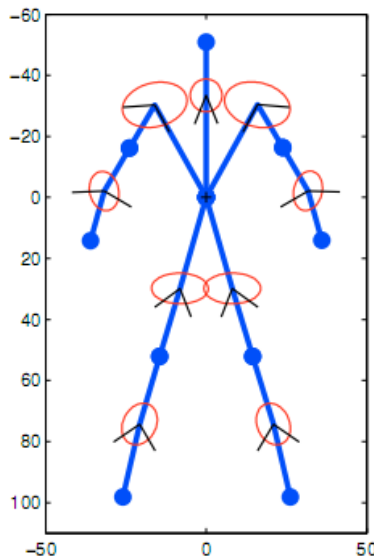


图 4 文献^[20]中所使用的人体骨骼模型

其算法的过程是，每个局部块都单独进行训练并在图像中进行检测，得到各个局部块的检测结果之后，通过定义的推理过程和关联准则，来求整个人体的检测和跟踪结果。其所构建的问题为：

在给定图像线索 $E = [E_1, E_2, \dots, E_m]^T$ 也即每个局部块的检测结果基础上，求各个行人的位置 $X^{O*} = [x_1^{O*}, x_2^{O*}, \dots, x_m^{O*}]^T$ 和各个局部块的位置和旋转角度组态 $Y^* = [y_1^*, y_2^*, \dots, y_m^*]$ ，即求 $p(Y^*, X^{O*} | E)$ 极大化。整个模型由贝叶斯准则推导为：

$$p(Y^*, X^{O*} | E) \propto p(Y^*)p(X^{O*})p(E | Y^*, X^{O*}) \propto p(Y^*)p(X^{O*}) \prod_{j=1}^m p(E_j | Y_j^*, X_j^{O*}) \quad (2.31)$$

$p(E_j | Y_j^*, X_j^{O*})$ 表示的是局部块检测的置信度，

$p(X^{O*})$ 表示的是整个人的运动模型，

$p(Y^*)$ 表示各个局部块，也即人体肢干的运动先验模型。

在论文中，作者使用了高斯过程隐参数模型（Gaussian Process Latent Variable Model, GPLVM）对局部块进行运动建模。

GPLVM 模型将 D 维的高维观测空间建模为 D 维的高斯过程，而其参数空间是 q 维的，其中 $q < D$ ，每个观测值 y_i 都对应于低维空间中的隐参数 z_i 。在给定隐参数序列 $Z = [z_1, z_2, \dots, z_m]^T$ 和模型参数 θ 下，观测序列 Y 的概率为：

$$p(Y | Z, \theta) = \prod_{i=1}^D \mathcal{N}(Y_{:,i} | 0, K_Z) \quad (2.32)$$

通过隐参数的方式，作者避免了在高维空间建模的困难，将行人行走过程中肢体的参数在低维空间利用隐参数进行建模。

因此，整个问题的解释就是，对于检测与跟踪过程来说，最终的结果，不仅涉及到各个局部块的检测结果能否组合成一个人体，同时在时间序列上整个人体的运动和各个肢干的运动也要符合一定的运动规律。

该方法对于检测与跟踪的结果的过程提出了一种包含物理含义的模型，存在

一定的合理性，但是该方法的局限性在于，局部块的检测结果一般较难取得良好效果，对于图片的分辨率要求较高。

第三章 基于局部块组合的人体目标跟踪算法

本文的研究中提出了一种将跟踪目标进行局部块分解，并通过对局部块的分析对目标进行状态更新和遮挡计算的算法，以下将就传统整体模型的限制性，以及本文新提出的局部块模型进行介绍。

3.1 传统整体模型的限制性

在传统的跟踪方法中，对于跟踪目标，通常采用某种几何包络的形式，如椭圆或者矩形来标识目标，相应的特征提取及匹配过程也在几何包络的范围内进行。整体模型方法虽然对于边缘噪声等具有较好的鲁棒性，但其对于遮挡等因素不够敏感。对于目标局部出现较大面积遮挡情况的时候，整体模型难以描述局部的遮挡与形变，且局部的遮挡像素也会对整体模型的特征表示造成影响，导致在搜索匹配的过程中，使跟踪目标产生偏移，如下图所示：



图 5 跟踪目标遮挡导致整体偏移

以 Mean Shift 方法为例，其所用的椭圆框定为目标的方式，在遮挡过程中，若遮挡过程发生较慢，则会出现由于模型逐步的更新，目标的特征颜色直方图会掺入大量的遮挡像素，使得目标表示出现偏差；若遮挡发生较为突然，则搜索匹配过程会使目标朝反方向搜索，以使模型排除遮挡像素，导致跟踪目标的位置出现偏离。因此，传统的整理几何包络表现目标的方式不适合复杂遮挡情况下的跟踪过程。

3.2 基于局部块组合的跟踪人体表示模型

基于图的局部块模型，将目标分解为多个关联的局部块，各个块之间存在着

拓扑约束的关系，以此来表示目标，该模型是当前目标识别领域中的研究重点^[32-34]。在本文的研究中，依据局部块分解的思想，同时借鉴文献^[45]的光流分析方法，提出了一种将人体划分为多个局部块，由整体跟踪和局部块跟踪相结合，对目标状态进行更新的方法。整体模型用于计算与检测结果的关联和约束局部块、保持整体运动的稳定性，而局部块跟踪则用于对整体运动的修正及遮挡分析。该模型并不按照人体关节的组成对局部块进行划分，避免了复杂的运动关系建模，而只依赖于表观特征的提取进行跟踪，以下将详细介绍该模型的构成。



图 6 人体检测结果、整体椭圆模型、局部块分解模型

在人体检测的结果基础上，使用一个短轴为 $0.3 * \text{Width}$ ，长轴为 $0.33 * \text{Height}$ 的椭圆作为人体的整体模型，并对其提取颜色直方图表示特征。

使用椭圆表示人体目标的依据为，在人体检测算法中，所有的训练样本都依据一定的准则进行标定，在本文所使用的人体检测算法中，每一个训练样本为 $64 * 128$ 像素大小，样本中的人体，与上下左右四个图像边界都保持 16 个像素的距离。因此检测结果的矩形框中，存在着许多非目标像素，本文中使用的矩形框大体上对于人体目标，从头覆盖到膝盖区域，没有涵盖小腿部分是因为，在行人行走过程中，小腿部分形变比较大，对于跟踪过程不是很有利。而人上半身由于着装的因素，一般具有较稳定的颜色分布，对于跟踪过程是十分有利的。

在最近的研究中，也有学者引入 Level Set 图割方法^[36]用于跟踪目标的精确分割，并应用到了人体目标跟踪上。本文使用的椭圆框在不引入过多干扰像素的情况下，极大的节省了计算开销。

在整体模型的基础上，将人体分解为 6 个正方形局部块，其初始化位置如图所示，其中一个涵盖头部的方块和四个位于躯干之上的方块，加上腿部一个方块，就构成了方法中的局部块组成。对每个局部块，算法中将其视为独立的跟踪对象，

提取 HOG 和颜色直方图特征，依照定义的相似性度量搜索匹配目标。

3.3 局部块特征提取与相似性度量

基于统计直方图的特征提取方式在近年来的研究与应用中，被证明了是一种十分有效的特征提取方式，无论是梯度方向直方图 HOG 算子，还是 RGB 颜色统计直方图，在计算机视觉的各种研究、应用中，都表现出了较好的性能。基于统计方式的特征提取，对于信息更强调整体性，而经过直方图归一化之后，能有效的避免噪声和特征扰动的影响。在本文的研究中，对于每个局部块，将提取两种直方图特征：梯度方向直方图特征和加权的颜色直方图特征。

梯度方向直方图（HOG）

梯度方向直方图通过统计图像中的边缘信息，对于光照的变化有较好的适应性，能够很好的表达目标局部的轮廓信息。对于每一个局部块，我们使用一个 105 维的 HOG 特征对其进行描述。

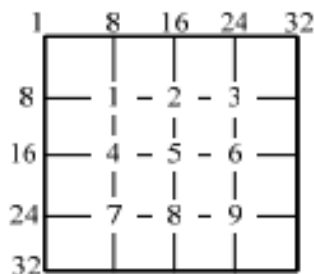


图 7 HOG 算子的计算示意图^[30]

下面介绍 HOG 算子的提取过程：首先将局部块缩放到 32*32 像素大小；以 8*8 像素为一个 cell 单元，将整个局部块划分为 16 个 cell；以邻近的 2*2 个 cell 组成一个 block 为基本统计块，将其中每个像素的梯度幅值按梯度方向投影到直方图中，本文中将 180° 按照 12° 为间隔，设定了 15 个投影方向。由上图可以看到，每个局部块将包含 9 个 block 单元，因此将得到一个 15*9=105 的描述向量，在统计梯度幅值完成后对整个向量进行归一化处理就得到该局部块的 HOG 特征。每个梯度的梯度幅值与梯度方向计算过程如下：

令 (w, h) 表示像素坐标，则

$$\begin{aligned}
 dx &= I(h, w+1) - I(h, w-1) \\
 dy &= I(h+1, w) - I(h-1, w) \\
 A(w, h) &= \sqrt{dx^2 + dy^2} \\
 ori(w, h) &= \arctan(dx / dy)
 \end{aligned} \tag{3.1}$$

其中 $I(w, h)$ 代表在 (w, h) 位置像素的灰度值， $A(w, h)$ 表示梯度幅值。

颜色直方图

参照 Comaniciu^[28] 提出的 kernel 化颜色直方图方法，对局部块的 RGB 颜色通道，各定义一个 16 维的颜色直方图，由此形成一个 48 维的颜色直方图特征。直方图的每一维计算方式为：

$$\tilde{q}_u = C \sum_{i=1}^n k(\|x_i^*\|^2) \delta[b(x_i^* - u)] \tag{3.2}$$

$$C = \frac{1}{\sum_{i=1}^n k(\|x_i^*\|^2)} \tag{3.3}$$

其中 $\delta[b(x_i^* - u)]$ 是 Kronecker delta 函数， $k(\|x_i^*\|^2)$ 是一个凸面函数，将离中心近的像素对直方图的贡献赋予高权重，离中心较远的粒子赋予低权重， C 是归一化系数。

考虑到离中心越远的像素越容易受到干扰及遮挡，依据距离对像素赋予不同权重的方式对于整个颜色特征的鲁棒性提供了保证，在实验中的应用效果也表明了该方法的有效性。

相似性度量

对于得到的特征描述子，需要有一种度量方式描述其相似性。

对于 HOG 特征，其描述向量位于欧式空间上，使用欧式距离（Euclidean Distance）对其相似性进行度量，对于两点 x_1 和 x_2 ，其距离定义为：

$$d(x_1, x_2) = \sqrt{\sum_i (x_{1,i} - x_{2,i})^2} \tag{3.4}$$

$x_{1,i}$ 和 $x_{2,i}$ 分别表示向量的每一维。

对于颜色直方图特征，文献^[28]使用的距离描述函数为：

$$d(y) = \sqrt{1 - \rho[\hat{p}(y), \hat{q}(y_0)]} \quad (3.5)$$

其中 $\rho[\hat{p}(y), \hat{q}(y_0)]$ 为巴氏系数 (Bhattacharyya Distance)，定义为：

$$\rho(y) \equiv \rho[\hat{p}(y), \hat{q}(y_0)] = \sum_{b=1}^m \sqrt{\hat{p}_b(y) \hat{q}_b(y_0)} \quad (3.6)$$

其几何意义为两个 m 维向量 $(\sqrt{\hat{p}_1}, \dots, \sqrt{\hat{p}_m})$ 和 $(\sqrt{\hat{q}_1}, \dots, \sqrt{\hat{q}_m})$ 夹角的余弦。在统计学中，巴氏系数用于度量两离散概率分布的相似性，原文作者认为 kernel 化后的颜色直方图代表颜色在目标上的分布概率，因此用巴氏系数对其相似性进行度量。

结合两个相似性度量，整个局部块的相似度为：

$$\begin{aligned} A &= \varepsilon_1 \cdot A_{HOG} + \varepsilon_2 \cdot A_{RGB} \\ \varepsilon_1 + \varepsilon_2 &= 1 \end{aligned} \quad (3.7)$$

ε_1 和 ε_2 表示两个特征相似度对整体相似度贡献的权重，在本文的方法中，对于目标分解成多个局部块，实验表明，对于不同的局部块，其所突出的特征不一样，设定不同的权值，使不同的局部块更侧重于其中更具区分度的特征。

本文的实验过程发现，对于头部的局部块，其相似度更依赖于 HOG 特征，直观上理解，由于头部与背景间存在明显的轮廓边缘，因此该部分更适合用边缘信息进行度量；对于躯干上的四个局部块而言，由于人体着装的关系，颜色的分布具有较稳定的状态，颜色直方图对其匹配发挥更大作用；对于腿部的块而言，边缘和颜色的作用相对而言较为平均。

在本研究中，对于头部块，经验的设定 HOG 和颜色直方图的权重为 0.7 和 0.3，躯干部分二者为 0.3 和 0.7，腿部则设为 0.5 和 0.5。

3.4 基于局部块模型的跟踪过程

3.4.1 跟踪模型的建立

在所提出的局部块模型基础上，在 t 时刻，对于跟踪目标，定义其状态为

$X_t = (X_t^c, X_t^1, \dots, X_t^6, D_t^1, \dots, D_t^6)$ 。其中 X_t^c 代表目标的中心位置， X_t^i 和 D_t^i 代表各个局部块的位置和与中心的相对位置。跟踪问题就可以表示为在 t 时刻，给定状态 X_t 和观察序列 $Y_{1:t}$ 下，求后验概率 $p(X_t | Y_{1:t})$ 极值的问题：

$$\hat{X}_t = \arg \max p(X_t | Y_{1:t}) \quad (3.8)$$

为简化模型，采用一阶马尔科夫过程建模，即：

$$\hat{X}_t = \arg \max p(X_t | Y_t) \quad (3.9)$$

由贝叶斯推理：

$$p(X_t | Y_t) = p(Y_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | Y_{t-1}) dX_{t-1} \quad (3.10)$$

其中：

$p(Y_t | X_t)$ 为观察模型，代表了观察与估计态之间的相似度。

$p(X_t | X_{t-1})$ 为状态转移模型，由前一时刻状态 X_{t-1} 来预测当前状态 X

为求解所提出的概率极值问题，通常采用马尔科夫链蒙特卡洛方法（Markov Chain Monte Carlo, MCMC）方法中的 Metropolis-Hastings 抽样算法来构建平稳马尔科夫链。MH 算法通过设定合适的提议分布（Proposal Distribution），然后迭代的运行提议（Proposal）和计算接受率（Acceptance Ratio）的过程，直到马尔科夫链达到平稳状态。其算法过程描述如下：

表 1 Metropolis-Hastings 抽样算法过程

1. 设定合适的提议分布 $q(\cdot | X^l)$
2. 设定初始迭代状态 X^0 ，并循环以下步骤，直到马尔科夫链达到平稳状态
 - (a) 由 $q(\cdot | X^l)$ 产生随机状态 X
 - (b) 计算接受率 $a = \min[1, \frac{f(X) \cdot q(X^l | X)}{f(X^l) \cdot q(X | X^l)}]$ ，以概率 a 令 $X^{l+1} = X$ ， $(1-a)$ 令 $X^{l+1} = X^l$
 - (c) l 增加，返回 (a) 步骤

由于状态空间存在多个变量，一般的 MCMC 方法并不适合于处理该类高维问题，文献^[49]中引入了 Basin Hopping Monte Carlo (BHMC) 的方法来进行求解。HPMC 与一般的 MCMC 方法一样，同样通过迭代的提议—计算接受率的循环方式来进行概率分布采样，其不同之处在于将高维的状态转换成低维状态来简化提议过程，并在计算接受率的时候通过状态变量的局部极值来进行。

针对提出的跟踪表示模型，考虑人体目标的整体性，每一个局部块的位置都可以由中心位置来进行确定，就可以将高维的状态空间问题转换成只针对中心位置的问题。在进行概率分布采样的过程中，只对中心位置 X_t^c 进行采样更新，各个局部块的位置由其相对于中心的距离进行确定。同 MH 算法，首先要设定提议分布 q ，此处采用高斯分布，即：

$$q(X_t^{c(l+1)} | X_t^{c(l)}) = \mathcal{N}(X_t^{c(l)}, \sigma^2) \quad (3.11)$$

l 代表的是当前的迭代次数，在得到中心位置之后，就可以由各个局部块的相对位置 D_t^i 计算出位置 X_t^i ：

$$\begin{aligned} X_t^{1(l+1)} &= X_t^{c(l+1)} + D_t^1 \\ &\vdots \\ X_t^{6(l+1)} &= X_t^{c(l+1)} + D_t^6 \end{aligned} \quad (3.12)$$

至此，得到了新的采样点 X_t^i ，计算接受率的方式如下：

$$a = \min\left[1, \frac{p(Y_t | f_o(X_t^{(l+1)}))p(f_o(X_t^{(l+1)}) | X_t^{(l+1)})}{p(Y_t | f_o(X_t^{(l)}))p(f_o(X_t^{(l)}) | X_t^{(l)})} \cdot \frac{q(X_t^{(l)} | X_t^{(l+1)})}{q(X_t^{(l+1)} | X_t^{(l)})}\right] \quad (3.13)$$

$p(Y_t | f_o(X_t^{(l)})) = \prod_{i=1}^6 p(Y_t | f_o(X_t^{i(l)}))$ 为目标的表观模型， f_o 代表各个局部块的

相似度极值点，由 3.3 节定义的相似性度量函数在各个局部块的临近区域内搜索匹配得到。

$p(f_o(X_t^{(l)}) | X_t^{(l)}) = \prod_{i=1}^6 p(f_o(X_t^{i(l)}) | X_t^{c(l)}, D_t^i)$ 代表了各个局部块与整体模型之

间空间位置上的关系约束。

在迭代完成之后，就可以得到原概率极值问题的解：

$$\hat{X}_t = \arg \max_{X_t^i} p(X_t^{(l)} | Y_t) \quad (3.14)$$

得到新的中心点之后需要对各个局部块的状态量进行更新，由于局部块的位

置 X_t^i 可由中心位置和相对位置 D_t^i 确定，因此只对相对位置进行更新：

$$D_{t+1}^i = D_t^i + \lambda \|f_o(X_t^i) - X_t^i\| \quad (3.15)$$

λ 类似于卡尔曼滤波中的增益因子，用于平衡观察与真实态。在实验中， λ 经验设定为 0.5。

至此，可以得到完整的基于局部块模型的跟踪算法：

表 2 基于局部块模型的跟踪算法过程

输入： $X_{t-1} = (X_{t-1}^c, X_{t-1}^1, \dots, X_{t-1}^6, D_{t-1}^1, \dots, D_{t-1}^6)$

输出： $\hat{X}_t = (X_t^c, X_t^1, \dots, X_t^6, D_t^1, \dots, D_t^6)$

采样阶段：

设定初始位置 $X_t^{c(1)}$

从 $l=1$ 到 $l=N-1$ 循环：

1. 利用 $q(X_t^{c(l+1)} | X_t^{c(l)})$ 提议新的采样点，并计算各个局部块的位置。
2. 在新的采样点由公式 (3.13) 计算接受率，由计算所得概率将其置为新的迭代起点。

循环结束

由 $\hat{X}_t = \arg \max_l p(X_t^{(l)} | Y_t)$ 求得目标位置

目标更新：

由 $\hat{X}_t - X_{t-1}$ 更新目标速度，及公式 (3.15) 更新各个局部块位置，并在新位置提取特征，更新表观模型。

迭代次数 N 参考文献^[49]的做法，设定为 20。过小的 N 使得算法难以得到极值点，而过大时又会影响算法的速度。

3.4.2 尺度更新与遮挡度的计算

在跟踪过程中，目标远离或靠近摄像机时，都会造成尺度的变化。对于目标的大小，其更新依赖于各个局部块是发散或收敛，可由上半部分的局部块和下半

部分的局部块的速度差，以及左右两部分速度差来进行更新修正：

$$Size_{x,t} = Size_{x,t-1} + \left| \sum_{left} u_i - \sum_{right} u_i \right| \quad (3.16)$$

$$Size_{y,t} = Size_{y,t-1} + \left| \sum_{up} v_i - \sum_{down} v_i \right| \quad (3.87)$$

如下图所示

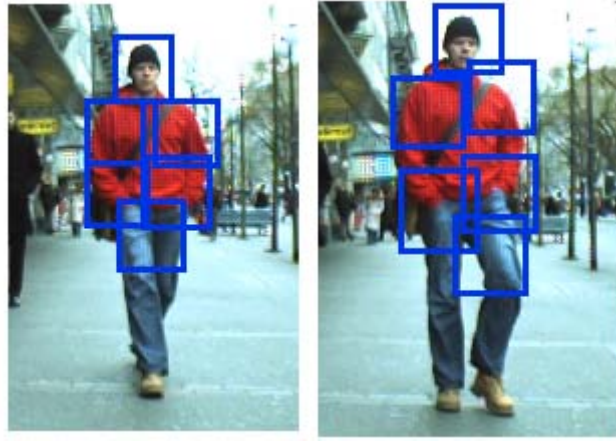


图 8 局部块跟踪与目标尺度变化

由图可以看到，在行人朝摄像机方向进行时，其尺度会逐渐变大，局部块会呈现出外扩的趋势，如最上方与最下方的块之间的垂直距离明显变大。通过判断局部块之间的趋向关系，就能很好的解决目标尺度变化的问题。

在目标跟踪的过程中，经常会发生目标被其他物体所遮挡的情况。对于整体模型来说，由于其涵盖的区域较大，其特征是在整个目标区域上提取的，对其进行遮挡分析很难表示出目标被部分遮挡的情况。

Bibby^[43, 44]在其研究中，提出了一种利用像素级统计模型计算目标间遮挡的模型，但是其依赖 Level Set 对目标进行前景与背景分割，且依赖每个像素的统计过程，计算量十分庞大，对跟踪过程不利。考虑到此点，在局部块模型上，本文提出了一种基于局部块相似度程度的目标遮挡状态计算方式。对于局部块，设定一个相似度阈值 θ ，当该局部块的相似度低于该阈值时，可认为该局部块存在遮挡。考虑整体跟踪的稳定性，遮挡度针对低于相似度阈值的局部块进行统计，整体目标的遮挡度计算为：

$$Occ = \sum_{A_i < \theta} (1 - A_i) \quad (3.18)$$

下图给出了一个跟踪过程中目标发生遮挡时遮挡度的计算结果：

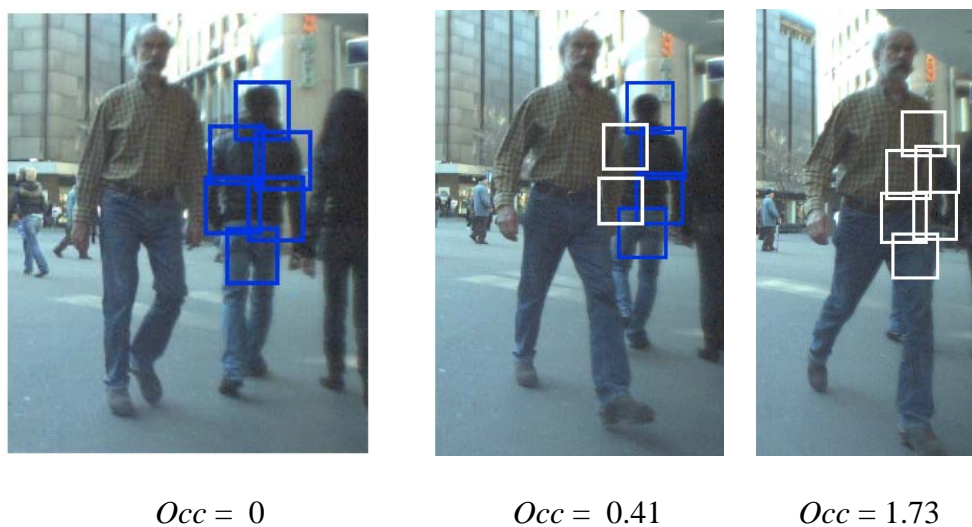


图 9 遮挡度的计算结果

在算法过程中，当某个局部块的出现较大的遮挡时，便停止对该局部块的表观模型进行更新，避免引入非目标像素，破坏整个目标的表观模型。同时在整体模型的计算中，不计入遮挡局部块的计算，只依赖于非遮挡状态下的局部块。

在跟踪算法过程中，每个时刻都对目标的表观模型进行更新，传统的简单几何包络表示目标的方式，在遮挡发生较缓慢时，模型更新的过程会逐步引入非目标像素，从而使模型逐渐出现偏离。使用局部块模型，通过局部块的判定，能有效避免遮挡像素对整体模型的影响，使跟踪过程更加鲁棒。

第四章 基于检测的多人体目标跟踪算法

本文旨在提出一种基于检测的多人体目标跟踪算法，应用到实时的行人检测跟踪系统当中。其算法依赖于人体检测算法作为输入，通过检测结果与跟踪目标的关联，实现检测与跟踪的融合与修正，依据局部块组合模型中的遮挡判别分析，能跟准确的感知目标的状态。在第三章所提出的基于局部块组合的跟踪模型基础上，整体算法的实现还有几方面的因素需要考虑：如人体检测算法及其对算法性能的影响，跟踪目标的初始化，检测结果与跟踪的数据关联过程，算法的实时性，实验性能等，以下将对这些方面进行论述。

4.1 人体检测算法

本文在实现过程中使用了 Dalal 提出的 HOG 特征加 SVM 分类器的人体检测算法，由于该算法在前面章节有了详细介绍，在此不再赘述。该方法已得到了广泛的应用推广，在最近计算机视觉工具包 OpenCV 中就附带有该算法的实现。在此主要阐述人体检测算法对整个检测与跟踪系统的影响。

人体检测算法作为整个算法的输入，其准确率越高，无疑对整个系统的性能提升越明显。检测算法中出现的误检可分为结构型误检与噪声型误检，结构型误检指场景中某些与人体极为相似的区域，如树木，建筑结构等，其在检测分类器中本身就难以与人体样本区分开来，其在整个算法过程中，会一直给出置信度较高的检测结果。噪声型误检指由于噪声的原因在检测过程中偶尔给出的人体假设，其偶尔出现在某帧中，不会给出连续性的结果。噪声型的误检可通过某些置信度累计的方法进行排除（下一小节将会论述该方法），所以整体算法对人体检测的准确率要求只在于尽量减少误报率。人体检测算法当前所面临的局限在于其对于检测到的人体定位不精确的问题，人体检测结果的矩形框并不能准确的表达人体目标，而传统的跟踪研究中，目标的初始化框选对于跟踪过程的稳定性也存在有影响。另外，在人群密集的地方，人体检测算法的输出往往存在指向不明确的问题，这容易造成后续处理的混乱。近年来，研究人员也开始关注这些问题^[48]，这两方面的局限若能够得到提高，无疑对整个系统性能的提升有重要作用。

4.2 跟踪的初始化与终止

考虑到检测过程中会存在噪声误检的问题，因此对于每一个新检测的目标就新建一个跟踪目标必然会产生许多误跟踪的情况。考虑到人体检测算法的误检率，文献^[16]中提出了一种依据连续帧间检测结果的相似度计算初始化置信度用于初始化目标的方法，其计算形式如下：

假定对同一人体假设，有连续 T 帧的检测结果对应，则初始化置信度为：

$$InitConf = (Track; Det_{1...T}) = \frac{1}{T-1} \sum_{t=1}^{T-1} A(\widehat{det}_{t+1}, det_{t+1}) \cdot (1 - e^{-\lambda_{ini} \sqrt{T}}) \quad (4.1)$$

式中的第一项代表 T 个检测结果的平均相似度，第二项则依赖于检测算法的准确度，若准确度越高，则参数 λ_{ini} 越大。由此就可以制定初始化策略：如果初始化置信度 $InitConf$ 大于某个阈值，则对该目标进行跟踪；否则，将该目标认为是可能的人体，待有后续的检测结果支持后再进行跟踪。也即对于目标来说，须有连续几帧相应的检测结果才能启动跟踪过程，这样的处理方式有效的去除了噪声型的误检测。在本文算法的实现过程中，设定目标需要连续 3 帧的检测支持才能启动跟踪过程。

跟踪的终止过程类似于初始化过程，对于跟踪的目标，同样定义一个置信度函数来决定是否终止跟踪：

$$EndConf = (Track; Det_{1...T}) = \left(1 - \frac{1}{T-1} \sum_{t=1}^{T-1} A(\widehat{det}_{t+1}, det_{t+1}) \right) (1 - e^{-\lambda_{ini} \sqrt{T}}) \quad (4.2)$$

即当某个跟踪目标连续 T 帧没有检测结果对应时，便会终止跟踪过程。在算法过程中，同样将该连续帧阈值设为 3 帧，考虑局部块模型能够计算出目标是否被遮挡，对于处于遮挡状态的目标来说，该连续影响阈值要设的更大一些，保证目标重新出现在图像中时，算法依然能够将其与之前串联起来。

4.3 检测与跟踪间的数据关联

检测结果与跟踪目标之间的数据关联，依赖于相似性度量的建立，本文使用

的相似度函数定义为:

$$A(tr, d^*) = p_N(Pos_{tr} - Pos_{d^*}) p_N(Size_{tr} - Size_{d^*}) \quad (4.3)$$

其中 p_N 代表的是正态分布。即两者的相似性依赖于相互间的距离以及大小的一致。在相似度建立的基础上, 求解数据关联的问题, 本文采用了与文献^[31]一样的贪心搜索算法来求解数据关联, 该算法流程如下:

表 3 数据关联中的贪心选择算法

假设: 当前帧有 n 个跟踪目标与 m 个待匹配检测结果
初始化: 依据相似度函数, 计算出两两间的相似度, 构成 $n*m$ 维的矩阵 A
迭代: 选择矩阵中相似度最高的元素 $A(i, j)$, 将跟踪目标 i 和检测结果 j 进行关联
删去矩阵中第 i 行元素和第 j 列元素, 形成新的矩阵
判定矩阵中是否所有元素都小于给定的相似度阈值, 若是, 结束迭代过程; 否则,
开始新一轮迭代

在本文的实验过程中, 该贪心算法已能保证结果的准确性。由于在检测—跟踪系统中, 问题的规模相对较小, 平均每帧需要关联的数据数量只有个位数, 而且建立的相似性度量函数对于不匹配的结果计算得到的函数值较小, 区分度很明显。相较其他规划算法如匈牙利算法等, 其他算法的结果最优虽然有理论支持, 但是计算量开销要更大些。从实用角度出发, 该贪心算法更符合实时性系统的要求。

在 3.4 节使用的跟踪算法过程中, 对于每个目标, 在采样阶段, 都要设定一代代的起始点。由此, 在基于检测的跟踪算法过程中, 就可以利用数据关联的结果, 将关联的检测结果, 作为该目标采样的起始点; 而对于没有检测结果关联的目标, 对其中心位置依据速度作更新之后得到新的位置作为起始点。

4.4 基于检测的多人体目标跟踪算法

在前文所论述的各个模型和算法的基础上，构建出一个完整的基于检测的多人体目标跟踪算法，流程图如下：

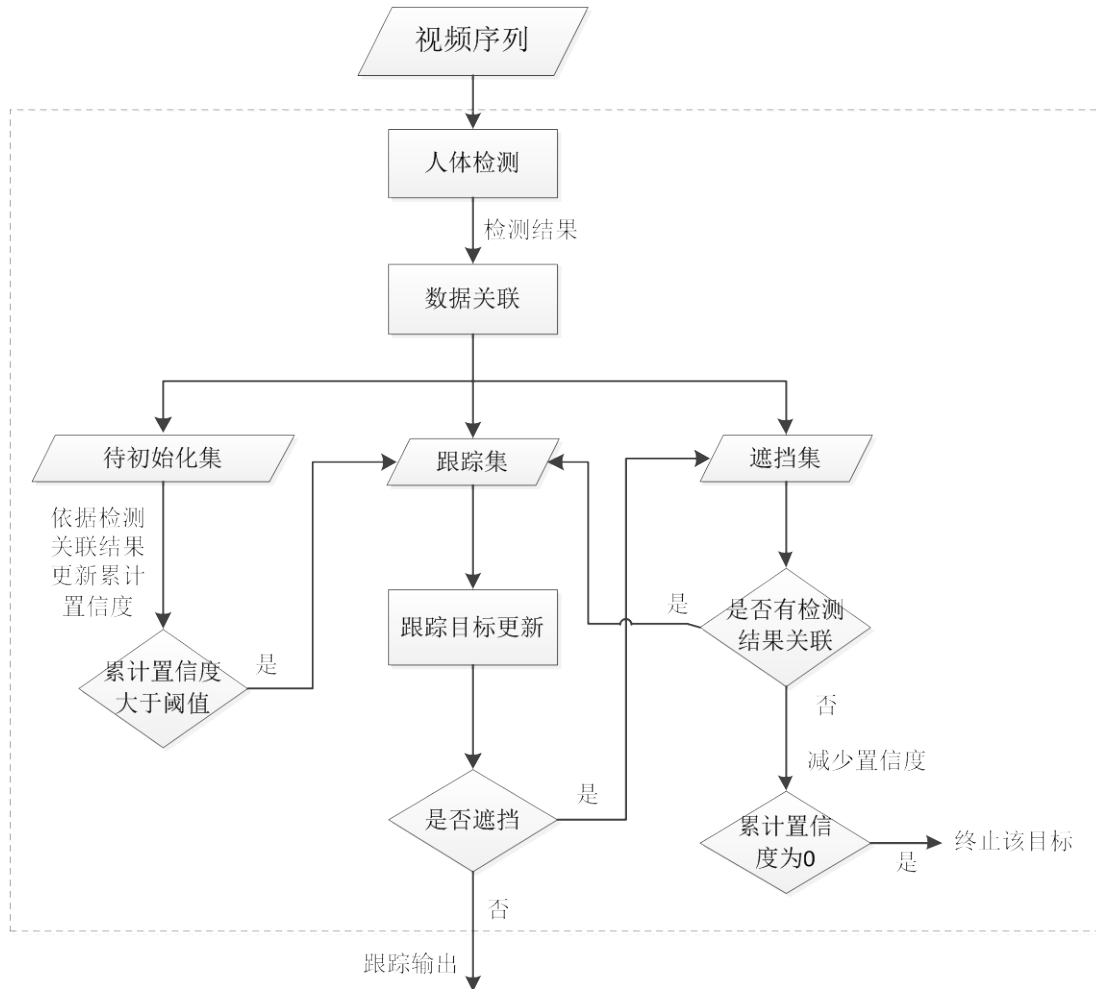


图 10 基于检测的多人体目标跟踪算法流程图

4.5 实验结果比较与算法性能分析

为验证所提方法的有效性，本文将所提方法在公开数据集上进行了实验验证。

4.5.1 实验数据集

本文所采用的实验数据集来自 ETH Zurich 实验室所发布的公开测试集^[35]。

该数据集中的视频系列利用其实验室开发的智能小车平台，沿着城市街道在行进中采集视频，其所用系统也给出了未来行人检测与跟踪系统的一种应用方式，即在智能辅助驾驶上，可以发挥其作用。该系列视频中含有大量的行人目标，同时行人运动之间经常发生遮挡，场景的光照变化也十分明显，十分适合于测试多人体目标跟踪系统的性能。在该数据集中，作者使用了双目视觉技术，因此采集的图像分左右两个摄像头所采集的，本文提出的方法并未考虑立体视觉的应用，因此测试过程只采用了单个摄像头的的数据。

在本文中，使用了“BANHHOF”（视频 1）和“Sunny Day”（视频 2）两个视频进行测试，视频 1 有 999 帧，总共有 5193 个标定的人体目标；视频 2 有 354 帧，总共有 1867 个标定的人体。视频 1 采集于阴天环境下，导致图像的颜色特征不够明显，同时场景中有许多树木等容易影响结果的结构；视频 2 采集于晴天，存在较多的阴影变化。

4.5.2 实验结果分析与比较

本文对于实验结果所使用的量化指标如下：

MT: (Mostly Tracked trajectories) 对于跟踪目标，成功捕捉到 80% 以上的轨迹。

PT: (Partially Tracked trajectories) 对于跟踪目标，捕捉到的轨迹在 20%-80% 之间。

ML: (Mostly Lost trajectories) 对于跟踪目标，捕捉到的轨迹不足 20%。

FA: (False Alarms) 误报

IDS: (ID Switch) 标识交换，代表多个目标之间在跟踪过程中标识交换的数量。

“BANHHOF”（视频 1）中的实验结果

下图给出了算法在视频 1 实验结果中一些代表性的视频帧：



(1)

(2)

(3)



图 11 算法在视频 1 中的测试结果示例

由结果可以看到，本文所提算法对于跟踪过程中的遮挡，尺度变化等有较好的适应性。在图 1-4 的过程中，右边穿红色衣服的人体在经过遮挡之后，又出现在视频中时，算法成功捕捉并将其与之前的跟踪过程关联起来，有效的处理了遮挡的问题；同时由图可以看到，当人体朝摄像机方向行进时，目标越来越大，而依据本文所提的局部块模型，能够很好的处理跟踪目标尺度变化的问题。

在测试过程中，所提方法也存在失效的时候，以下列出算法过程中存在的一些缺陷：



图 12 算法在视频中的错跟踪及漏跟踪示例

上图给出了算法过程中存在的一些错跟踪及漏跟踪的情况，在上图以标红注明。可以看到，算法的一些错跟踪是由于背景中的树木等造成的错检测造成的。由于这些结构在人体检测过程中给出了十分可信的检测结果，在连续多帧中，检测算法都将其判定为人体，因此不可避免的造成跟踪算法对其进行初始化并加入到跟踪的过程。在第一幅图中，跟踪框定位在两个人的中间，也是因为检测结果给出了中间位置，并没有将两人分别开来，在本章第一节中，已对人体检测结果对跟踪过程的影响进行了论述，如何通过跟踪过程和局部块模型对人体检测的结果进行修正，也是未来有待研究之处。

“SUNNYDAY”（视频 2）中的实验结果

视频 1 采集于阴天环境下，整个视频序列中基本没有出现阴影变化等，而视频 2 中光线较为充足，视频中的人体常常由于进入阴影区域，造成颜色表观上会发生明显变化，因此十分具有挑战性，本文在视频 2 中的实验结果如下所示：

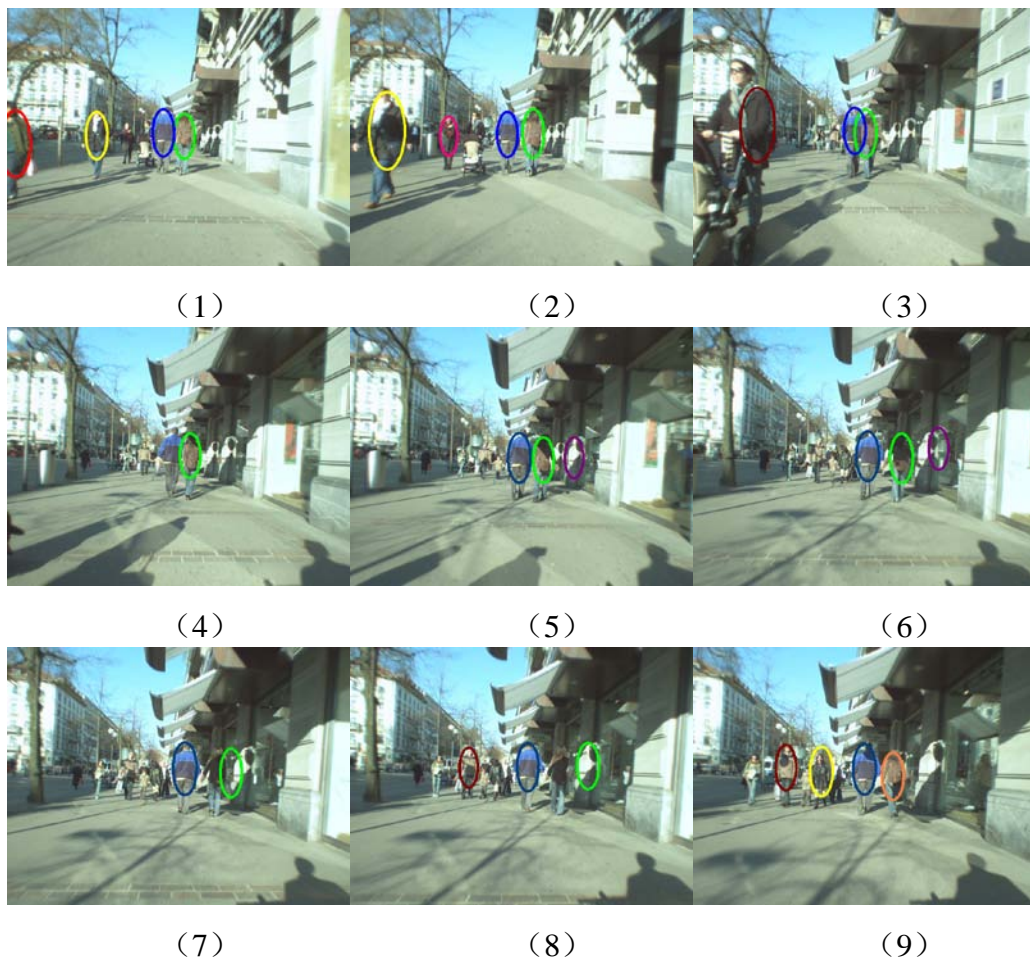


图 13 算法在视频 2 中的测试结果

由结果可以看到，算法大部分情况下运行良好，同时也存在一些漏跟踪的情况；在光照变化强烈的阶段中，算法跟踪出现了偏离目标的情况，由于跟踪目标进入阴影区域后，颜色上发生了较大的变化，同时，由于旁边的建筑结构在人体检测的阶段也给出了是人的判定，因此两种因素结合起来造成了跟踪结果的错误。

利用前文提到的量化比较方法，本文与文献^[36]的量化比较如下：

表 4 跟踪算法量化比较结果

	视频 1-文献 ^[36]	视频 1-本文	视频 2-文献 ^[36]	视频 2-本文
#persons	67	67	28	28
Mostly Tracked	54	52	23	21
Partially Tracked	4	10	1	2
Mostly Lost	9	5	4	5
False Alarms	9	11	6	4
ID Switch	6	3	3	4

4.6 算法实时性

作为跟踪算法，其运算速度须达到实时性，才能进行实际应用。基于检测的目标跟踪算法中，涉及到检测与跟踪两个过程。在本文的研究中，对于人体检测算法，使用了 OpecnCV 2.2 版本中所带的 HOG 人体检测算法，对于 640*480 像素大小的图像，其运算速度为 10s-15s 每帧，难以满足实时性的要求；在给定检测结果的情况下，在跟踪部分，本文所提的算法在 4 核心 2.93Ghz CPU 和 4GB 内存的计算机上，不采用并行化处理，运算速度为每秒 1-4 帧左右，运算速度依赖于图像中跟踪目标的多少，目标越多，其所需要的计算量越大。

近年来，随着 GPU 并行计算的广泛应用^[37,46]，研究人员也开始将其应用到人体检测中，并取得了很好的效果，文献^[37]给出了 GPU 加速人体检测过程后接近实时速度的效果。在跟踪过程方面，由于本文所提的局部块模型对每个局部块都要进行相应的匹配更新过程，也需要不小的计算量，文献^[38]对于一般局部块的模型，给出了一种级联式的搜索匹配结构用于加速，未来可参考该方法，并结

合 GPU 计算的使用，对整个模型的搜索匹配过程进行加速，使整个算法能够达到实时的效果。

总结与展望

人体目标的检测与跟踪是计算机视觉其中的一个研究热点方向，拥有着较为广泛的应用前景。本文提出了一种新的基于局部块组合的跟踪目标表示模型，有效的利用了人体检测的结果对跟踪目标进行模型初始化，同时利用各个局部块对跟踪过程进行修正与遮挡计算。实验结果表明，所提模型对于跟踪过程中的目标遮挡、尺度变化等，有良好的适应性。本文将该模型融入到基于检测的多人体目标跟踪算法框架中，在复杂场景下的实验结果表明，该算法相较其他方法，具有一定的优势。

本文所提的算法在某些方面依然存在不足之处，如当人体检测结果错误较多的时候，算法难以对这些结果进行排除，导致许多错跟踪的情况；同时，算法在实时性上还难以满足实时性系统的要求。未来，更准确、快速的人体检测算法将对系统性能有比较大的提升；使用更好的搜索策略以及并行化加速处理的方式，将对整体的算法的速度有很大的提高，使其达到实时性。

由于时间的关系，本文的算法在许多地方仍存在一些不足之处，如整体目标与局部块之间的关系建模，预测与修正过程的理论支撑等。现有的实验结果表明，该模型具有一定的研究价值，在未来，可对局部块之间的关系约束，以及局部块的特征选取、如何通过整个模型对于检测中的错误结果滤除等进行深入研究，从而将该算法向实用化目标推进。



参考文献

- [1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection”, In IEEE Computer Vision and Pattern Recognition (CVPR) 2005. PP. 886- 893.
- [2] Qiang Zhu, Shai Avidan, Mei-chen Yeh, Kwang-Ting Cheng, “Fast human detection using a cascade of Histograms of Oriented Gradients”, In IEEE Computer Vision and Pattern Recognition (CVPR) 2006.PP. 1491-1498.
- [3] O. Tuzel, F.Porikli, P.Meer. “Pedestrian detection via classification on Riemannian manifolds”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30, 1713-1727, 2008
- [4] W.R. Schwartz, A. Kembhavi, D. Harwood, L.S. Davis. “Human Detection Using Partial Least Squares Analysis”, In International Conference on Computer Vision (ICCV) 2009.PP.24-31.
- [5] Xiaoyu Wang, Tony X. Han and Shuicheng Yan, “An HOG-LBP Human Detector with Partial Occlusion Handling,” In IEEE International Conference on Computer Vision (ICCV) 2009.PP.32-39.
- [6] D. Comaniciu, V. Ramesh, and P. Meer. “Kernel-based object tracking”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003:564–577.
- [7] Brown, R. G. and P. Y. C. Hwang. 1992. Introduction to Random Signals and Applied Kalman Filtering, Second Edition, John Wiley & Sons, Inc.
- [8] M. Isard, and A. Blake.” CONDENSATION-conditional density propagation for visual tracking”. International Journal on Computer Vision, 1998.PP.5–28.
- [9] Perera, A.G.A., Srinivas, C., Hoogs, A., Brooksby, G., Hu, W. “Multi-Object Tracking Through Simultaneous Long Occlusions and Split-Merge Conditions”. In IEEE Computer Vision and Pattern Recognition (CVPR) 2006.PP.666-673.
- [10] Hao Jiang, Sidney Fels and James J. Little. “A Linear Programming Approach for Multiple Object Tracking”, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2007 .PP.1-8.
- [11] C. Huang, B. Wu, and R. Nevatia. “Robust object trackingby hierarchical association of detection responses”. In European Conference on Computer Vision

- (ECCV) 2008.PP.788-801.
- [12] L. Zhang, Y. Li, and R. Nevatia. "Global data association for multi-object tracking using network flows". In IEEE Computer Vision and Pattern Recognition (CVPR) 2008.PP.1-8.
- [13] Y. Li, C. Huang, and R. Nevatia. "Learning to associate: Hybridboosted multi-target tracker for crowded scene". In IEEE Computer Vision and Pattern Recognition (CVPR) 2009.PP.2953-2960.
- [14] Cheng-Hao Kuo, Chang Huang, Ram Nevatia. "Multi-Target Tracking by On-Line Learned Discriminative Appearance Models", In IEEE Computer Vision and Pattern Recognition(CVPR) 2010.PP.685-692.
- [15] Michael D. Breitenstein, Fabian Reichlin, Bastian Leibe, Esther Koller-Meier, Luc Van Gool, "Online Multi-Person Tracking-by-Detection from a Single, Uncalibrated Camera," IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 Dec. 2010.PP.267-282.
- [16] B. Wu and R. Nevatia. "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors". International Journal on Computer Vision, 75(2):247-266, 2007.
- [17] Ess, A.; Leibe, B.; Schindler, K.; Van Gool, L. "A mobile vision system for robust multi-person tracking". In IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2008.PP.1-8.
- [18] Andreas Ess, Bastian Leibe, Konrad Schindler, Luc van Gool, "Robust Multiperson Tracking from a Mobile Platform," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 10, pp. 1831-1846, May 2009.
- [19] B. Leibe, K. Schindler, N. Cornelis, and L. van Gool, "Coupled Detection and Tracking from Static Cameras and Moving Vehicles," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 10, pp. 1683-1698, Oct. 2008.
- [20] M. Andriluka, S. Roth, and B. Schiele, "People-Tracking-by-Detection and People-Detection-by-Tracking", In IEEE Conf. Computer Vision and Pattern Recognition(CVPR) 2008.PP.1-8.
- [21] M. Andriluka, S. Roth and B. Schiele, "Pictorial Structures Revisited: People Detection and Articulated Pose Estimation", In IEEE Conference on Computer

- Vision and Pattern Recognition (CVPR) 2009.PP.1014-1021.
- [22] M. Andriluka, S. Roth and B. Schiele, "Monocular 3D Pose Estimation and Tracking by Detection", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010.PP.623-630.
- [23] V. Vapnik. The nature of statistical learning theory. Springer-Verlag, 1995.
- [24] Yoav Freund and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting". In Computational Learning Theory: Eurocolt '95, pages 23–37. Springer-Verlag, 1995.
- [25] Ran Xu, Baochang Zhang, Qixiang Ye, Jianbin Jiao, "Human Detection in Image Via L1-Norm Minimization Learning" , In International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2010.PP.3566-3569.
- [26] Ran Xu, Baochang Zhang, Qixiang Ye, Jianbin Jiao, "Cascaded L1-norm Minimization Learning(CLML) Classifier for Human Detection", In IEEE Computer Vision and Pattern Recognition (CVPR) 2010.PP.89-96.
- [27] David G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.
- [28] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift," In IEEE Computer Vision and Pattern Recognition(CVPR) 2000.PP.142-149.
- [29] Canny, J., "A Computational Approach To Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986.
- [30] Zhenjun Han, Qixiang Ye, Jianbin Jiao, "Combined feature evaluation for adaptive visual object tracking," Computer Vision and Image Understanding, 2010.PP.69-80.
- [31] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. "Robust tracking-by-detection using a detector confidence particle filter". In IEEE International Conference on Computer Vision (ICCV) 2009.PP.1515-1522.
- [32] P. F. Felzenszwalb and D. P. Huttenlocher. "Pictorial structures for object recognition". International Journal of Computer Vision, 61(1):55–79, 2005.
- [33] P. F. Felzenszwalb, D. McAllester, and D. Ramanan. "A discriminatively trained, multiscale, deformable part model". In IEEE Computer Vision and Pattern

- Recognition(CVPR) 2008.PP.1-8.
- [34] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan. "Object Detection with Discriminatively Trained Part-Based Models", IEEE Pattern Analysis and Machine Intelligence, Sept 2009.PP.1627-1645.
- [35] A. Ess, B. Leibe, and L. van Gool, "Depth and Appearance for Mobile Scene Analysis," In International Conference on Computer Vision(ICCV) 2007.PP.1-8.
- [36] D. Mitzel, E. Horbert, A. Ess, B. Leibe. "Multi-Person Tracking with Sparse Detection and Continuous Segmentation", European Conference on Computer Vision (ECCV) 2010.PP.397-410.
- [37] Prisacariu, V., Reid, I. "fastHOG – a Real-Time GPU Implementation of HOG". Technical Report 2310/09, Dept. of Engineering Science, University of Oxford (2009).
- [38] P. Felzenszwalb, R. Girshick, D. McAllester, "Cascade Object Detection with Deformable Part Models" In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.PP.2241-2248.
- [39] Ting Yu; Ying Wu; Krahnstoever, N.O.; Tu, P.H., "Distributed data association and filtering for multiple target tracking". In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2008.PP.1-8.
- [40] A. Ess, K. Schindler, B. Leibe, L. van Gool. "Improved Multi-Person Tracking with Active Occlusion Handling", ICRA Workshop on People Detection and Tracking May 2009.
- [41] Junliang XING, Haizhou AI and Shihong LAO, "Multi-Object Tracking through Occlusions by Local Tracklets Filtering and Global Tracklets Association with Detection Responses", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2009.PP.1200-1207.
- [42] Qian Yu, Gérard Medioni, "Multiple-Target Tracking by Spatiotemporal Monte Carlo Markov Chain Data Association," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 12, pp. 2196-2210, Oct. 2009.
- [43] Charles Bibby and Ian Reid. "Robust Real-Time Visual Tracking using Pixel-Wise Posteriors," In European Conf on Computer Vision(ECCV) 2008.
- [44] C Bibby and I D Reid. "Real-time Tracking of Multiple Occluding Objects using

- Level Sets”, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010.PP.1307-1314.
- [45] T. Brox, J. Malik. “Large displacement optical flow: descriptor matching in variational motion estimation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, March 2011 .PP.500-513.
- [46] N. Sundaram, T. Brox, K. Keutzer. “Dense point trajectories by GPU-accelerated large displacement optical flow”, European Conference on Computer Vision (ECCV) 2010.
- [47]张学工, 关于统计学习理论与支持向量机,《自动化学报》,2000年01期32-42页。
- [48] Stefan Walk, Nikodem Majer, Konrad Schindler and Bernt Schiele . “New Features and Insights for Pedestrian Detection”, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010.PP.1030-1037.
- [49] Junseok Kwon and Kyoung Mu Lee, "Tracking of a Non-Rigid Object via Patch-based Dynamic Appearance Modeling and Adaptive Basin Hopping Monte Carlo Sampling," In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.PP1208-1215.

致 谢

在攻读硕士的三年期间，遇到了许多困难，付出了许多努力，值此毕业之际，对过去三年中帮助过我的人们表示衷心的感谢。

感谢我的导师焦建彬教授，在硕士学习期间，焦老师在学术上给予了我极大的指导与帮助，使我能有机会从事挑战性的课题，获益匪浅；同时焦老师在生活上也给予了我无微不至的关心，使我能够顺利的完成学业，知遇之恩，无以回报。焦老师严谨的治学态度、积极的进取精神以及平易近人的作风都是学生之榜样。

感谢杨辉研究员在论文写作过程中给予我的指导与帮助，使得论文能够顺利而圆满的完成。

感谢叶齐祥老师在硕士学习期间给予我的帮助，论文的工作离不开叶老师给予的指导，叶老师对于学术的严格要求，对于探索的孜孜不倦和精益求精的精神，常常给予我鼓励。

感谢所有模式识别与智能系统开发实验室的同学们，欢乐而融洽的实验室气氛离不开这些可爱的人们。

感谢那些在最困难时候给予我无私帮助的好朋友们，所有一切我都将铭记于心。

感谢我的父母与兄长，他们的爱与包容，是我坚实的后盾，感激之情无以言表，唯有谨记于心，拼搏。

陈孝罡

2011年4月