

## 部件级表观模型的目标跟踪方法\*

王美华<sup>1</sup>, 梁云<sup>1,4</sup>, 刘福明<sup>1</sup>, 罗笑南<sup>2,3</sup>

<sup>1</sup>(华南农业大学 数学与信息学院, 广东 广州 510642)

<sup>2</sup>(国家数字家庭工程技术中心, 广东 广州 510006)

<sup>3</sup>(中山大学 信息科学与技术学院, 广东 广州 510006)

<sup>4</sup>(华南家禽疫病防控与产品安全协同创新中心, 广东 广州 510640)

通讯作者: 梁云, E-mail: sdliangyun@163.com

**摘要:** 因受遮挡、运动模糊、剧烈形变等因素的影响, 稳定且准确的目标跟踪是当前计算机视觉研究领域重要挑战之一。首先采用中层视觉线索的超像素描述目标/背景的部件, 以部件颜色直方图作为其特征, 并通过聚类部件库的特征集构建初始表观模型, 部件表达的局部性和灵活性使该模型能够准确描述目标/背景; 然后, 利用贝叶斯滤波模型计算目标框的初始状态, 并提出相似物体干扰的检测和处理算法以避免跟踪漂移, 得到更健壮的结果; 最后, 为了减弱形变、遮挡、模糊对表观模型的影响以更好地保持目标特征, 提出一种基于部件库的特征补集的在线表观模型更新算法, 根据部件变化实时反映目标/背景的变化情况。在多个具有跟踪挑战的视频序列上的实验结果表明(共 12 个视频序列): 与现有跟踪方法相比, 该算法跟踪结果的中心误差更小, 成功帧数更多, 能够更准确并稳定、有效地跟踪目标物体。

**关键词:** 部件库; 表观模型; 特征补集; 相似物体; 跟踪漂移

**中图法分类号:** TP391

中文引用格式: 王美华, 梁云, 刘福明, 罗笑南. 部件级表观模型的目标跟踪方法. 软件学报, 2015, 26(10): 2733-2747. <http://www.jos.org.cn/1000-9825/4737.htm>

英文引用格式: Wang MH, Liang Y, Liu FM, Luo XN. Object tracking based on component-level appearance model. Ruan Jian Xue Bao/Journal of Software, 2015, 26(10): 2733-2747 (in Chinese). <http://www.jos.org.cn/1000-9825/4737.htm>

## Object Tracking Based on Component-Level Appearance Model

WANG Mei-Hua<sup>1</sup>, LIANG Yun<sup>1,4</sup>, LIU Fu-Ming<sup>1</sup>, LUO Xiao-Nan<sup>2,3</sup>

<sup>1</sup>(College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China)

<sup>2</sup>(National Engineering Research Center of Digital Life, Guangzhou 510006, China)

<sup>3</sup>(School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510006, China)

<sup>4</sup>(South China Collaborative Innovation Center for Poultry Disease Control and Product Safety, Guangzhou 510640, China)

**Abstract:** Dealing with factors such as overlap, blurs from quickly moving and severe deformation, accurate and stable object tracking has become a critical challenge in compute vision field. First, in this paper, superpixels are used as middle level visual clue to describe the components of object/background with the color histograms of components as their features. The initial appearance model is proposed by clustering the features of a component library. The locality and flexibility of components representations allow the appearance model to describe object/background much more accurately. Then, the Bayesian filter model is used to compute the initial state of target region, and an algorithm is proposed to check and deal with the disturbance introduced by similar objects to avoid drift and obtain more robust tracking result. Finally, to reduce the influences of deformation, overlap and blurs to better preserve the features of object, an online appearance model update algorithm is developed based on the complementary set of the features of components library to enable the

\* 基金项目: 国家自然科学基金(61202293, U1301253); 广东省科技计划(2012B010100029, 2014A050503057)

收稿时间: 2013-10-17; 定稿时间: 2014-09-28

appearance model to reflect the real-time variation of object/background by the changes of components. Many experiments on video sequences with different tracking challenges (totally about 12 sequences) show that, compared with the existing object tracking methods, the proposed tracking algorithm results in less error of center position and more successful frame, and therefore can track an object more accurately, stably and effectively.

**Key words:** components library; appearance model; complementary set of features; similar object; tracking drift

目标跟踪是计算机视觉领域的研究热点之一,它在视频监控、普适计算、智能人机交互等领域具有广泛的应用.近年来,大量目标跟踪方法已被提出,但由于受光照、非刚性变化、同类物体干扰、遮挡等因素的影响,稳定且准确的目标跟踪方法仍是极具挑战性的研究课题<sup>[1,2]</sup>.大量研究证明,建立自适应表观模型,是实现目标跟踪的关键<sup>[3-11]</sup>.本文根据中层视觉线索超像素构建部件级表观模型,建立部件库特征补集以增强表观模型更新的自适应性和健壮性,提出相似物体干扰的检测和处理算法以避免漂移.与现有方法相比,本文方法可更准确、稳定地跟踪复杂环境下的目标.

## 1 相关工作

构建强健有效的表观模型,是目标跟踪算法的核心和关键.根据建模方法,当前目标跟踪方法可分为基于像素级的方法、基于整体模板级的方法和基于局部信息级的方法.

- 基于像素级的跟踪方法<sup>[3,4]</sup>根据目标区域的像素信息构建和更新表观模型.例如:Avidan 等人<sup>[3]</sup>在像素级上利用 AdaBoost 算法建立和更新强分类器以识别目标;施华等人<sup>[4]</sup>提出一种基于像素可信度和空间位置的跟踪方法,通过 Kalman 滤波在时域上的中值背景模型与空间域的交叉熵检测运动目标.尽管该类方法能够初步区分目标和背景,但它们在处理严重遮挡和复杂背景的跟踪时很不理想;
- 基于整体模板级的跟踪方法<sup>[6,7]</sup>根据目标区域建立描述物体的正/负模板,以模板为整体构建和更新物体的表观模型.例如:Babenko 等人<sup>[6]</sup>通过学习正样本和负样本模板构建判别式模型进行目标跟踪;Mei 等人<sup>[7]</sup>利用大量模板构造表观模型,把目标跟踪看作粒子滤波框架的稀疏逼近问题.该类方法极大地减少了复杂背景和遮挡时的跟踪漂移,但不能有效处理非刚性变化和同类物体干扰下的目标跟踪;
- 基于局部信息级的跟踪方法<sup>[8,10]</sup>通过记录和分析目标的局部变化建立和更新表观模型,其中,Kwon 等人<sup>[8]</sup>利用若干随机局部子块构建表观模型,通过增加和删除子块处理遮挡和非刚性;Wang 等人<sup>[10]</sup>基于超像素建立表观模型对目标进行跟踪,在遮挡、姿态变化和快速运动方面比较有效.该类方法能够灵活地表达视频序列中动态变化的目标和背景,近年来引起了人们的广泛关注.

本文提出一种新的基于局部信息级表观模型的跟踪方法,该方法通过学习、分类和更新物体的部件级信息构建和更新表观模型:首先对图像帧进行超像素分割,以超像素描述物体部件;然后,利用部件信息构造物体特征库以建立和更新目标的表观模型;最后,建立基于最大后验概率和同类物体检测策略以计算目标位置.本文方法的优势在于:

- (1) 与当前主流方法相比,本文方法能够更准确地跟踪复杂背景、光照变化、剧烈变形以及严重遮挡时的目标;
- (2) 与 Wang 等人<sup>[10]</sup>同样基于超像素建模的方法相比,本文通过建立目标部件的特征补集和进行同类物体分析,能够更实时、有效地反映动态变化的目标和背景,从而更有效地处理非刚性变化和同类物体遮挡下的跟踪.

## 2 系统框架

本文根据物体的局部信息建立和更新目标的表观模型,通过物体部件描述其局部信息,提出一种基于部件级的跟踪方法.为了有效计算物体部件,采用超像素集表示物体部件.超像素由于具有丰富的图像结构信息和明显的目标边界,比高层整体模型和低层视觉特征具有更出色的灵活性和自适应性,是一种有效的特征表达方式,已在物体的分割、分类、表示等领域表现出色<sup>[12-16]</sup>.本文算法的流程如图 1 所示,包含初始阶段、目标跟踪阶

段和更新阶段。

- 初始阶段:建立基于目标物体部件库的表观模型.首先,简单跟踪前  $m$  帧图像(如  $m=7$ ),以它们的目标区域为中心向四周扩展得到  $m$  个扩展区域,超像素分割这些扩展区域,用超像素表示物体部件,提取各部件 HSI 颜色空间的直方图作为部件特征,集合这些部件特征建立物体部件库;然后,利用均值聚类法聚类部件库内部件,将描述相同部件的超像素集合到一个聚类,用每个聚类的总体特征描述相应物体部件;最后,计算每个聚类属于目标的置信度描述对应部件属于目标的概率.如此,利用部件的特征和置信度来描述目标的初始表观模型;
- 目标跟踪阶段:根据部件匹配和相似干扰进行目标跟踪.对新一帧输入图像,首先根据其前一帧跟踪结果建立扩展区域;再对扩展区域进行超像素分割得到该帧的部件集,根据部件匹配算法计算每个部件的置信度;然后,在扩展区域采样多个候选样本,并在贝叶斯框架下根据各样本所含部件的置信度计算样本置信度;最后,根据相似物体干扰处理优化样本置信度,进而得到目标区域;
- 更新阶段:利用部件库的特征补集更新表观模型.首先,确定部件库被替换帧的特征集,分别计算该特征集与当前帧部件特征矢量的距离;然后,寻找被替换帧中不在当前帧内的部件特征,以构造当前帧的特征补集,合并当前帧的特征集及其补集,替换被替换帧的特征集,构造更新后的部件库;最后,对新部件库的特征集进行聚类并计算各聚类的置信度,完成对表观模型的更新.

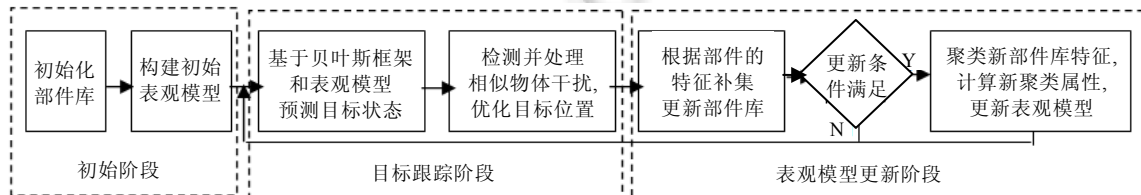


Fig.1 Flowchart of the tracking method

图1 本文目标跟踪算法的流程图

### 3 构建基于部件库的表观模型

本节首先描述建立表观模型的特征,进而给出如何根据目标/背景的特征构建目标物体的表观模型。

#### 3.1 特征提取和部件库的建立

本文采用超像素的颜色信息作为图像特征来构建表观模型.超像素是一种典型的中层视觉特征,一个超像素包含若干像素点,对应图像中一个具有相似结构的局部连通区域.由于中层视觉特征具有丰富的图像结构信息,与高层外观模型和低层视觉相比,它更灵活且适应性更强.因此,与目标跟踪中普遍采用的直方图、轮廓等特征相比,超像素对目标表达能力更强.本文利用超像素描述物体部件,首先采样 SLIC 方法<sup>[15]</sup>对输入图像进行超像素分割.SLIC 分割算法是一种简单的线性迭代聚类方法,它在一个 5 维空间内(3 维颜色空间+2 维位置空间)对像素点进行聚类,并提出新的距离函数约束超像素的划分.该方法得到的超像素尺寸均匀、内容紧凑、简单易用(只需超像素总数)且执行效率高(复杂度为  $O(N)$ ),被广泛应用于物体识别和分类.

将图像分割为若干超像素后,这些超像素具有相似的灰度值和纹理特性,表达了物体的局部部件.本文利用超像素描述物体部件,根据超像素构建表观模型.假设对第  $t$  帧图像进行超像素分割,得到  $N_t$  个超像素  $sp(t, j)$  ( $j=1, \dots, N_t$ ).有些超像素描述目标部件,如图 2(c)表示女孩黄色上衣的超像素和红色裤子的超像素;有些则描述背景,如图 2(c)描述灰色地板的超像素.由于 HSI 模型<sup>[17,18]</sup>能够准确反映人眼视觉系统对色彩的感知,本文提取超像素 HSI 颜色空间的直方图作为其特征.对于第  $t$  帧  $N_t$  个超像素的特征集,用  $F_t = \{f_t^r \mid r=1, \dots, N_t\}$  表示.

为了构建能够有效区分目标和背景的外观模型,需建立部件库作为先验知识.先验知识在当前跟踪场景中是不可得的,由已跟踪若干结果提供.首先,采用简单模板匹配方法<sup>[19]</sup>跟踪图像序列的前  $m$  帧,然后,把第  $i(1 \leq i$

$\leq m$ 帧目标区域的周围区域(以目标区域中心为中心,目标区域对角线长度的 $\lambda_s$ 倍作为边长)分割成 $N_i$ 个超像(本文实验中, $\lambda_s$ 取1.5, $m$ 取7, $N_i$ 取300),并计算超像素的特征;最后,集合 $m$ 帧超像素及其特征,建立初始部件库,并得到其特征集 $F = \{f_{i'}^r | t'=1, \dots, m, r=1, \dots, N_i\}$ . 由于采用对目标位置的周围区域进行分割而获取超像素集合,因此部件库中包含描述目标和背景的超像素(如图2(c)所示).

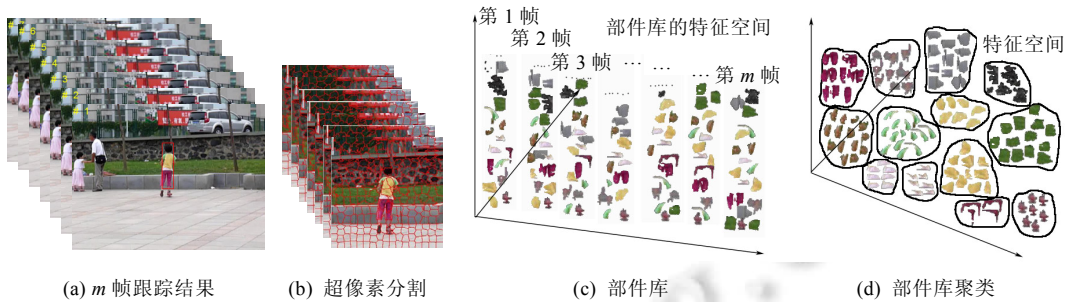


Fig.2 Procedure of clustering the component library

图2 部件库的聚类过程

### 3.2 初始表观模型的建立

本文根据描述先验知识的初始部件库建立初始表观模型,包括两步:聚类部件库的特征集和计算部件置信度.

步骤1. 聚类部件库的特征集.

部件库中的一个部件可能对应多个超像素,需要对部件库的特征集进行聚类,每个聚类描述一个部件,如图2(d)中的红色聚类描述目标物体的上衣,灰色聚类描述背景地板.利用均值聚类法<sup>[20-24]</sup>把部件库的特征集聚类成 $n$ 类 $clst(k)(k=1, \dots, n)$ ,描述 $n$ 个部件.第 $k$ 个部件包含信息 $f_c(k)$ (聚类 $k$ 中心的特征矢量,等于它所包含的所有特征矢量的均值,用于近似描述其特征)、 $r_c(k)$ (聚类 $k$ 的半径)和特征集 $\{f_{i'}^r | f_{i'}^r \in clst(k)\}$ (聚类 $k$ 的特征成员).

步骤2. 计算每个部件的置信度,描述部件属于目标的概率.

部件的置信度等于其所对应的聚类置信度,而聚类置信度取决于它所包含的超像素.当某个聚类中超像素全部位于目标区域时,该聚类完全属于目标;但当聚类的超像素全部位于背景时,该聚类则完全属于背景.但事实上,超像素常部分属于目标区域,部分属于背景,本文根据超像素属于目标和背景的面积来计算其所从属聚类的置信度.假设用 $S^+(k)$ 表示聚类 $k$ 所有超像素位于目标区域内的像素数,用 $S^-(k)$ 表示目标区域外的像素数,则聚类 $k$ 的置信度为

$$C_k = \frac{S^+(k) - S^-(k)}{S^+(k) + S^-(k)}, \forall k = 1, \dots, n \quad (1)$$

其中, $C_k \in [-1, 1]$ . $C_k$ 为1时,说明该部件所包含的超像素完全位于目标区域内;当 $C_k$ 为-1时,说明该部件所包含的超像素完全位于背景区域. $C_k$ 越大,说明部件属于目标区域的概率越大;反之,属于背景区域的概率越大.

## 4 基于表观模型与相似干扰检测的目标跟踪

本节中,根据已建立的初始表观模型提出目标跟踪方法.首先,利用贝叶斯滤波模型获取初始目标状态;然后,采用相似物体干扰检测与处理策略优化目标位置,得到更准确的跟踪结果.

### 4.1 贝叶斯框架

在目标跟踪领域,贝叶斯滤波原理是众多跟踪算法的理论基础,其中,粒子滤波在处理非线性、非高斯系统时表现得十分出色,具有较强的鲁棒性.本文基于贝叶斯滤波模型对目标进行跟踪,对于给定 $t$ 时刻的状态 $X_t$ 和到 $t$ 时刻的观察值 $Y_{1:t}$ ,可将目标跟踪问题转化为求解后验概率 $p(X_t | Y_{1:t})$ .

$$p(X_t | Y_{1:t}) \approx p(Y_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | Y_{1:t-1}) dX_{t-1} \quad (2)$$

其中,  $X_t = \{X_t^x, X_t^y, X_t^s\}$  为三维向量,  $X_t^x, X_t^y$  和  $X_t^s$  分别代表目标  $t$  时刻时所处的  $x, y$  坐标及其尺寸;  $p(Y_t | X_t)$  表示观察模型, 预测当前观察值的似然概率;  $p(X_t | X_{t-1})$  表示运动模型, 用于根据  $t-1$  时刻的目标状态预测当前状态. 本文采用高斯分布预测运动模型  $p(X_t | X_{t-1})$ , 去除跟踪中普通噪声的影响.

在贝叶斯框架下, 当前大量跟踪方法是针对观察模型进行研究<sup>[7-10, 25, 26]</sup>, 通过研究观察模型和运动模型, 利用贝叶斯滤波原理计算后验概率, 进而通过最大后验概率获得目标状态. 假设  $t$  时刻有  $N$  个候选样本, 根据贝叶斯滤波原理的最大后验概率,  $X_t^{(l)}$  代表状态  $t$  时刻的第  $l$  个候选样本, 则  $t$  时刻目标状态  $\hat{X}_t$  为

$$\hat{X}_t = \arg \max_{X_t^{(l)}} p(X_t^{(l)} | Y_{1:t}), \forall l = 1, \dots, N \quad (3)$$

根据公式(2)和公式(3), 构建观察模型  $p(Y_t | X_t)$  是基于贝叶斯框架目标跟踪的关键. 本文提出一种新的观察模型来预测候选样本属于目标或背景的可能性, 利用样本置信度描述这种可能性, 通过构建置信度图计算样本置信度, 具体方法见第 4.2 节.

## 4.2 计算样本的置信度

本文提出的观察模型用于估计样本置信度, 该置信度计算通过以下两步实现: 首先, 计算该样本所在图像的置信度图; 然后, 根据样本区域计算其置信度.

### (1) 构建图像置信度图

本节以如何获取第  $t$  帧的置信度图为例阐述置信度图的计算. 为了简化第  $t$  帧置信度图的计算, 通过保持第  $t-1$  帧目标区域的中心并增大目标区域的尺寸, 在第  $t$  帧上确定一个扩展区域. 因为视频序列中物体运动的连续性和一致性, 导致相邻帧目标区域的位置和尺寸的变化不大, 所以第  $t$  帧的所有候选样本均取自该扩展区域. 对第  $t$  帧的扩展区域进行超像素分割, 每个超像素近似对应一个物体部件, 得到  $N_t$  个部件(如图 3(b)所示). 第  $t$  帧的置信度图由这些部件的置信度确定. 每个部件根据以下两个因素得到其置信度: 所属聚类的置信度和该部件与对应聚类中心的距离. 第 1 因素的原理是如果部件在特征空间上属于表观模型中某个聚类, 那么部件的置信度应正比于该聚类的置信度; 第 2 因素考虑部件的特征矢量在特征空间上与相关聚类中心的距离, 距离越远, 说明该部件置信度与相关聚类置信度的相关系数越小. 假设第  $t$  帧的第  $r$  个部件的特征为  $f_t^r$ , 通过计算  $f_t^r$  与所有聚类中心特征的距离, 判断  $r$  所属的聚类. 当  $f_t^r$  与聚类  $clst(k)$  的中心特征  $f_c(k)$  距离最小时, 则认为部件  $r$  属于聚类  $clst(k)$ . 部件  $r$  的置信度  $C_t^r$  由聚类  $clst(k)$  的置信度  $C_k^c$  和  $r$  与聚类  $clst(k)$  的相关系数  $w(r, k)$  决定(见公式(5)), 其中,  $w(r, k)$  的计算如下:

$$w(r, k) = \exp\left(-\lambda_d \times \frac{\|f_t^r - f_c(k)\|^2}{r_c(k)}\right), \forall r = 1, \dots, N_t, k = 1, \dots, n \quad (4)$$

$$C_t^r = w(r, k) \times C_k^c, \forall r = 1, \dots, N_t \quad (5)$$

公式(4)的  $\lambda_d$  为常数(实验中取值为 2). 第  $t$  帧扩展区域内像素点的置信度由它所属的超像素的置信度决定(图 3(c)是图 3(a)的置信度图, 蓝色越深, 代表属于背景的可能性越大, 红色越深, 代表属于目标的可能性越大), 而扩展区域外的像素点则均视为背景, 它们的置信度赋值背景区域的置信度即为 -1.

### (2) 计算样本的置信度

假设  $\{A_t^{(l)}\}_{l=1}^{N_t}$  表示第  $t$  帧的第  $l$  个样本  $X_t^{(l)}$  在置信度图上的覆盖区域(如图 3(d)所示), 在该样本置信度图中, 第 1 个样本的置信度较高, 说明其属于目标区域的概率较大, 用  $V(i, j)$  表示位于  $(i, j)$  的像素点的置信度, 那么,  $X_t^{(l)}$  的置信度  $C_t(l)$  为

$$C_t(l) = \sum_{(i, j) \in A_t^{(l)}} V(i, j) \quad (6)$$

但直接使用样本置信度不能有效反映目标尺寸的变化, 需要对样本置信度作正则化处理, 将其归入到区间  $[0, 1]$ . 设  $S(X_t^{(l)})$  表示样本  $X_t^{(l)}$  的区域大小,  $S(X_{t-1})$  表示第  $t-1$  帧的目标区域  $X_{t-1}$  的区域大小, 则  $C_t(l)$  的正则化结果



为  $\hat{C}_t(l) = C_t^l(l) \times [S(X_t^{(l)}) / S(X_{t-1})]$ .  $\hat{C}_t(l)$  描述了样本  $X_t^{(l)}$  在观察模型  $p(Y_t | X_t^{(l)})$  中的取值, 满足  $p(Y_t | X_t^{(l)}) \propto \hat{C}_t(l)$ . 本文中取  $p(Y_t | X_t^{(l)}) = \hat{C}_t(l)$ , 并利用高斯模型计算  $p(X_t^{(l)} | X_{t-1})$ , 再根据公式(3)得到每个样本的后验概率, 并取后验概率最大的样本区域为第  $t$  帧的目标区域.

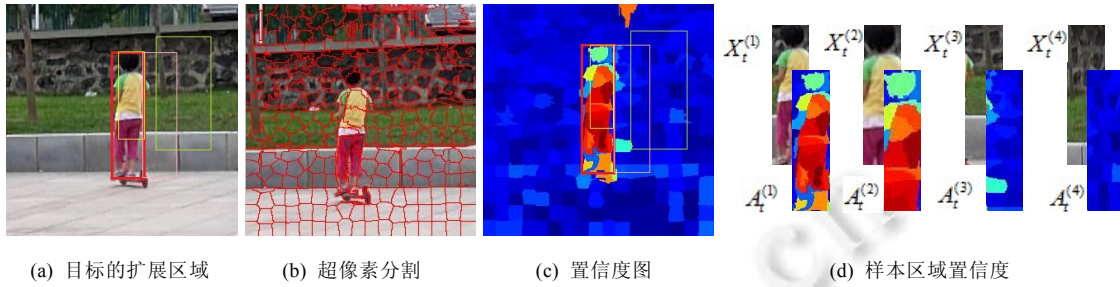


Fig.3 Procedure of object tracking based on Bayesian framework

图 3 贝叶斯框架的跟踪过程

### 4.3 相似物体干扰的检测与处理

在贝叶斯框架下, 通过第 4.2 节的观察模型和第 4.1 节的运动模型, 通常场景下能获取目标位置的准确状态. 然而, 跟踪中经常出现与目标非常接近的物体, 如图 4 中 653 帧处与红色矩形框内的目标区域相邻的绿色球衣运动员. 这些相似物体与目标的特征非常接近, 在观察模型中, 它们的置信度与目标区域的置信度接近相同(如下文图 5(a)第 1 列的置信度图所示). 当这类相似物体出现在目标周围时(如图 4 中 653 帧所示), 会对目标检测形成不利干扰, 处理不当时, 相似物体会被识别为目标物体, 从而产生跟踪漂移(如图 4(b)第 3 列所示).

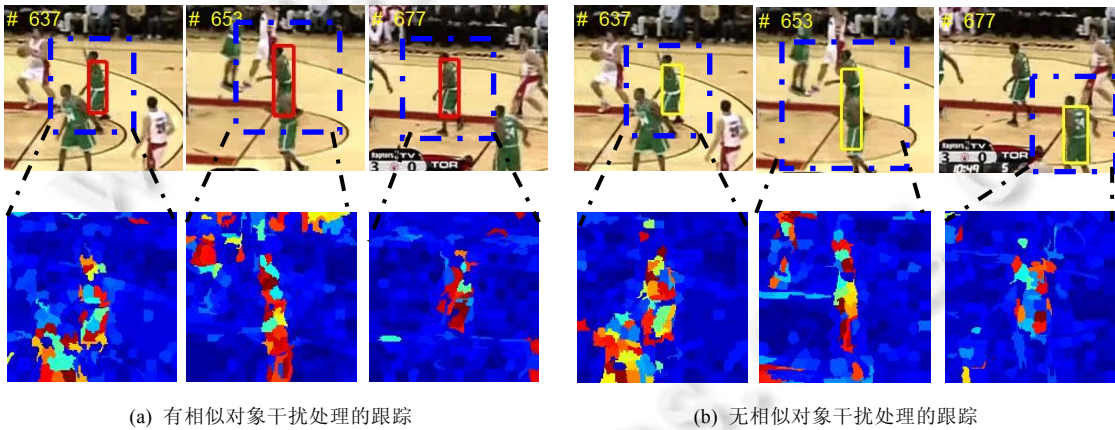


Fig.4 Tracking result with and without dealing with disturbance introduced by similar object

图 4 有无相似对象干扰处理结果

本文对相似物体的预测和处理, 通过以下 3 个步骤实现.

步骤 1. 相似物体干扰的预测.

对比第  $t$  帧与第  $t-1$  帧目标区域的 4 个方向的坐标, 判断第  $t$  帧目标区域是否发生最大限度的扩展. 设  $\{x_t^{(l)}, \dots, x_t^{(4)}\}$  分别代表第  $t$  帧目标框左上角的横、纵坐标和右下角的横、纵坐标(如图 5(a)所示),  $\{x_{t-1}^{(l)}, \dots, x_{t-1}^{(4)}\}$  表示第  $t-1$  帧目标框左上角的横、纵坐标和右下角的横、纵坐标,  $L$  为目标  $t$  时刻所允许的最大位移,  $\{X_t^{(h)}, X_t^{(w)}\}$  和  $\{X_{t-1}^{(h)}, X_{t-1}^{(w)}\}$  分别代表第  $t$  帧和第  $t-1$  帧目标框的高和宽. 当同时满足以下两个条件时, 说明可能出现相似物体干扰:

- (1) 目标区域在方向  $d$  外延伸了最大距离,即  $|x_t^{(d)} - x_{t-1}^{(d)}| \geq L, \forall d = 1, 2, 3, 4$ ;
- (2) 目标区域在高度或宽度上发生最大限度的扩展,即  $|X_t^{(s)} - X_{t-1}^{(s)}| \geq L, \forall s = h, w$ .

这种超过最大限度的扩展被定义为异常扩展.其中,  $L$  与目标框大小有关(本文实验中,当边框长度大于 40 时,  $L$  取 15;否则,  $L$  取 1/4 边框大小).

步骤 2. 相似性物体干扰的确定.

根据相似物体预测,当检测到目标框发生异常扩展时,需进一步确定该异常扩展是否为相似物体干扰.因为相似物体可能从目标的左/右/上/下和 4 个对角方向靠近,需从这 8 个方向判断.以相似物体从左边靠近为例解释异常扩展的判断.假设第  $t-1$  帧相似物体离目标较远,目标框能够恰好包围目标(如图 5(a)中蓝框所示);第  $t$  帧发生从左侧相似物体干扰(如图 4(b)所示),利用最大后验概率得到异常扩展的目标区域(如图 5(b)中绿框所示).但实际上图 5(b)蓝框所示才是正确的目标区域.是否发生异常扩展的判断,通过比较绿框本身的扩展区域(图 5(b)中黄框所示)和绿框左侧与扩展区域相同大小区域(如图 5(b)中粉色框所示)来实现.设  $N_{e1}$  表示扩展区域,即黄框中置信度大于 0 的像素数;  $N_{e2}$  表示粉色框中置信度大于 0 的像素数;  $\theta_1$  为常量,表示阈值(实验中取 0.6).当  $N_{e2}/N_{e1} \geq \theta_1$  时,说明粉色框内置信度大于 0 的像素比黄框多,即粉色框包含与黄框相似的物体,由此可以预测绿框左侧出现与目标相似的物体区域.据此断定黄框发生了异常扩展,该扩张由目标周围相似物体的干扰产生.

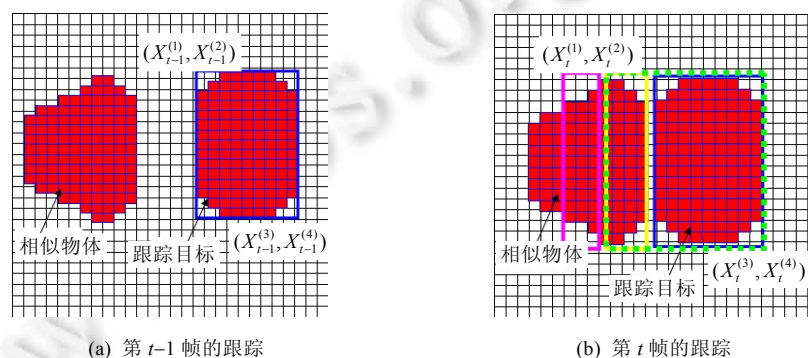


Fig.5 Detect the exception extension of target frame caused by disturbance introduced by similar object

图 5 检测相似对象干扰导致的异常扩展

步骤 3. 相似性物体干扰处理.

当检测到相似物体干扰后,需调整目标框以优化目标区域.本文把发生异常扩展的边向扩展反方向收缩  $L$  长度( $L$  为  $t$  时刻最大位移).如图 4 所示,第 1 行、第 2 行分别是跟踪结果和置信度图.图 4(a)有相似物体干扰结果,相似物体靠近目标(见第 1 列)、对目标遮挡(见第 2 列)和离开目标时(见第 3 列)都能有效跟踪目标;图 4(b)所示为无相似对象干扰处理策略结果,在相似物体靠近目标(见第 1 列)时可跟踪目标,在对目标部分遮挡(见第 2 列)时目标框偏向于干扰物体,遮挡消失时(见第 3 列)出现跟踪漂移.

## 5 基于部件特征补集的表现模型更新

随着时间的推移,光照变化、遮挡、运动模糊、目标物体变形等现象均有可能出现,需要根据背景和目标的变化不断调整目标表现模型.表现模型由其所包含的部件库的聚类、聚类属性(含特征矢量、半径和置信度)所确定.因此,对表现模型的更新也通过以下两步实现:更新构建表现模型部件库的聚类和更新这些聚类的属性值.

### 5.1 更新构建表现模型的部件库的特征聚类

传统的表现模型更新方法,每跟踪若干帧后就最新帧(时间最近帧)换掉最旧帧(时间最久帧),然后重新建模实现.为了处理严重遮挡和漂移问题,常通过寻找最佳被替换帧的方式以增强新表现模型的有效性.文献[10]

将发生严重遮挡帧作为被替换帧.但当更新表观模型只通过替换整帧来实现时,无法表示目标被遮挡或形变等区域,使表观模型不能完整地表达目标区域(如图 6(a)中第 3 列黄框区域所示).当恢复遮挡时,不完整表观模型无法识别曾经被遮挡或发生大形变的部件,从而产生不准确的跟踪结果(如图 6(a)中第 3 列所示).本文提出部件库特征补集方法以保留被替换帧中被丢弃的目标特征,使表观模型对目标的描述更加完整.本文对于表观模型的更新通过以下 3 步实现.

#### 步骤 1. 构建部件库的特征补集.

若不建立特征补集,这类特征将被舍弃,导致新表观模型无法表示相应的物体部件.为了处理严重遮挡或跟踪漂移问题,在没有发生严重遮挡或漂移时,选择最旧帧作为被替换帧;否则,选择距离当前最近的第  $k$  帧(实验中,  $k$  取 3)作为被替换帧.当被替换帧目标区域内置信度大于 0 的特征数量大于  $\beta$  时( $\beta$  为常数,用于限制特征补集所含特征的数量,实验中,  $\beta=15$ ),从中选出  $\beta$  个特征作为当前帧特征集的补集(要求它们与当前帧所属聚类中心的距离较大);当置信度大于 0 的特征小于  $\beta$  时,把这些特征直接作为补集.

本文根据公式(7)对严重遮挡或漂移进行判断.假设  $t$  时刻共有  $N_t$  个样本(实验中,  $200 \leq N_t \leq 600$ ),其置信度为  $C_t(l), l=1, \dots, N_t$ . 设  $\theta_0$  为检测严重遮挡或漂移的阈值,  $\mu$  为构建部件库的  $m$  帧图像置信度的平均值.  $C_t(l)$  为当前帧目标区域置信度,  $X_t^{(l)}$  为其对应的采样,  $S(X_t^{(l)})$  为该采样的面积值. 公式(7)左边分母使得左边值的范围为  $[-1, 1]$ . 当公式(7)左侧的比值大于  $\theta_0$  时,说明当前帧目标区域置信度远远小于用于构建部件库保留的各帧的置信度的平均值,意味着当前帧的目标区域大部分属于背景区域,即可断定当前帧发生了严重的遮挡或漂移.

$$\frac{\mu - C_t(l')}{2 \times S(X_t^{(l')})} > \theta_0, \text{ s.t. } C_t(l') = \max(\{C_t(l)\}_{l=1}^{N_t}) \quad (7)$$

#### 步骤 2. 更新部件库特征集.

得到部件库特征补集后,合并入当前帧特征集作为新特征集加入部件库,并删除被替换帧的特征集,完成一次部件库的特征集更新,得到描述新表观模型的部件库特征集.

#### 步骤 3. 更新表观模型的聚类.

根据第 3 节可知,表观模型由其特征聚类描述.因此,更新表观模型首先要更新其特征聚类.由步骤 2 可以得到新表观模型的特征集,对该特征集重新聚类,即得到新表观模型.

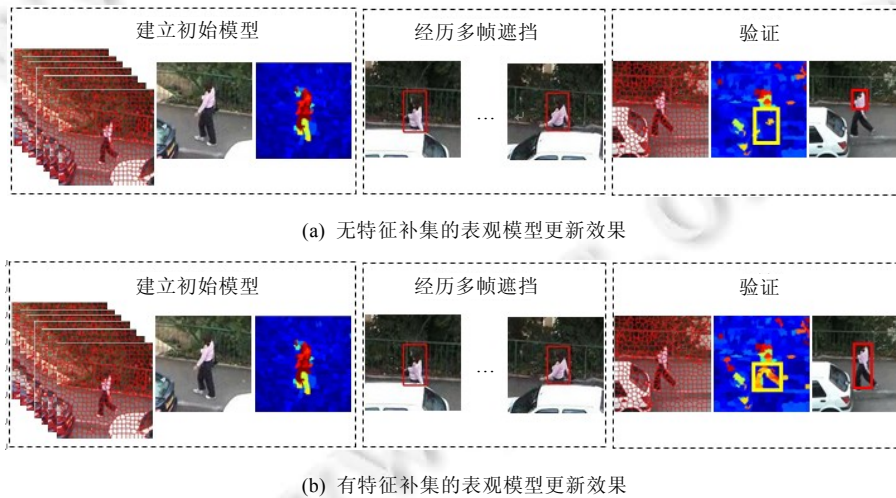


Fig.6 Influence of the complementary set to updating appearance model

图 6 特征补集在表观模型更新中的效果



## 5.2 表观模型的聚类属性的更新

聚类属性描述了目标物体的性质,主要包括聚类半径、聚类中心的特征和聚类置信度.更新表观模型时,必须更新相应的聚类属性.其中,聚类半径和聚类中心特征在重新聚类时得到,因此属性更新的关键是更新其置信度.

新表观模型通过重新聚类新部件库的特征集来实现,因此,新部件库每个聚类的特征集来自两个方面:部件库的特征补集和检测到的部件库的特征集.假设重建表观模型的特征中来自检测到的部件库的特征集为  $F$ ,来自部件库的特征补集为  $\tilde{F}$ .当任意一个新聚类  $clst(k)$  的特征全属于  $F$  时,计算其置信度的  $S^+(k)$  和  $S^-(k)$  的方法与第 3.2 节相同;但当新聚类  $clst(k)$  的特征既来自  $F$  又来自  $\tilde{F}$  时,根据公式(8)和公式(9)计算其  $S^+(k)$  和  $S^-(k)$ ,其中,  $S_F^+(k)$  和  $S_{\tilde{F}}^+(k)$  分别表示  $clst(k)$  中来自  $F$  和  $\tilde{F}$  的特征所对应部件属于目标的像素数,  $S_F^-(k)$  和  $S_{\tilde{F}}^-(k)$  则表示这些特征所对应部件属于背景的像素数.  $S_F^+(k)$  和  $S_{\tilde{F}}^-(k)$  的计算方法与第 3.2 节相同.

因为部件库的特征补集所对应的部件全部被识别为目标,这些部件所含像素数量应加入对应聚类位于目标的像素数量.因此,假设  $\sum_{\tilde{f} \in \tilde{F}} Area(\tilde{f})$  表示特征  $\tilde{f}(\tilde{f} \in \tilde{F})$  对应部件的面积,则  $S_{\tilde{F}}^+(k) = \sum_{\tilde{f} \in \tilde{F}} Area(\tilde{f})$ ,  $S_{\tilde{F}}^-(k) = 0$ .进而,包含特征补集的聚类的  $S^+(k)$  和  $S^-(k)$  为

$$S^+(k) = S_F^+(k) + S_{\tilde{F}}^+(k) = S_F^+(k) + \sum_{\tilde{f} \in \tilde{F}} Area(\tilde{f}) \quad (8)$$

$$S^-(k) = S_F^-(k) + S_{\tilde{F}}^-(k), \text{ s.t. } S_{\tilde{F}}^-(k) = 0 \quad (9)$$

图 6(a)所示为无特征补集更新表观模型,由图可知:经过多帧遮挡后,目标的腿部特征信息丢失,该部件在遮挡完成后被识别为背景(如图 6(a)中黄框所示).图 6(b)采用特征补集方法更新表观模型,腿部特征集作为部件库的特征补集而加以保留,所以当该部件重新出现时,能被正确识别(如图 6(b)中黄框所示).虽然图 6(a)与图 6(b)有初始表观模型和遮挡,但图 6(b)比图 6(a)的跟踪结果(对比两组第 3 列红色框)更准确.

## 6 实验与结果分析

### 6.1 本文实验设置

本文在 CPU 2.53GHz、内存 2G 的 PC 机, MATLAB(R2011b)下进行相关实验.利用 SLIC 算法<sup>[15]</sup>实现超像素分割(其权重参数和超像素个数分别设为 10 和 300),采用 HSI 颜色空间直方图描述部件的特征,采用 Mean Shift 聚类算法<sup>[20]</sup>对实现聚类(聚类的核宽参数设为[0.15~0.20]),并设定检测遮挡的阈值  $\theta_0$  的范围为[-0.25~0.3].

本文采用 12 组视频序列(见附录表 4)进行实验.这些序列涵盖了目标跟踪的主要挑战,包括复杂背景、目标快速移动、姿态和尺寸严重变形、部分或严重遮挡、严重的形状畸变、光照变化.为验证本文方法的优越性,我们将其与现有主流的 10 种跟踪方法(包括 IVT<sup>[5]</sup>, L1<sup>[7]</sup>, SCM<sup>[27]</sup>, LSAM<sup>[28]</sup>, Godec<sup>[9]</sup>, BHMC<sup>[26]</sup>, LOT<sup>[29]</sup>, VTD<sup>[8]</sup>, VTS<sup>[25]</sup>, SPT<sup>[10]</sup>)进行对比分析,具体见第 6.2 节和第 6.3 节.

### 6.2 目标跟踪结果对比的定性比较分析

分别从 7 个跟踪挑战的角度来分析本文方法的有效性(包括遮挡、快速移动、姿势变化、剧烈形变、相似对象干扰、复杂背景、光照变换),并将本文与现有方法进行对比分析.

#### (1) 遮挡实验

图 7 展示了 Woman\_sequence 视频中行人多次被汽车遮挡的跟踪效果,其中, Godec, IVT, L1, VTD, SPT 和 Ours 方法的跟踪结果分别用黑色(虚线)、紫色(虚线)、蓝绿色(虚线)、黑色(实线)、黄色(虚线)和红色(实线)表示.由于基于部件库特征补集的表观模型可同时学习目标 and 背景的特征,并通过特征补集保留被遮挡目标,即使经历长时间遮挡仍能识别出被遮挡部分(如第 167 帧的人腿),从而克服遮挡引起的漂移,在遮挡发生时仍能准确跟踪目标.而其他方法(如 IVT, VTD, Godec, L1 和 SPT)由于不能有效保留目标被遮挡部分的特征,在遮挡发生后无法跟踪到目标.如在 32 帧,行人的裤子与车窗色彩相似, Godec 方法没有限定目标导致目标区域过大,引起跟踪漂移. IVT 和 L1 方法采用整体模板为图像特征,遮挡发生时易将背景识别为目标,在 65 帧导致跟踪失败.

SPT 方法未用特征补集更新外观模型,长时间遮挡后被遮挡部件特征被丢弃,最终使目标区域不能准确覆盖物体(如 319 帧的黄色框).

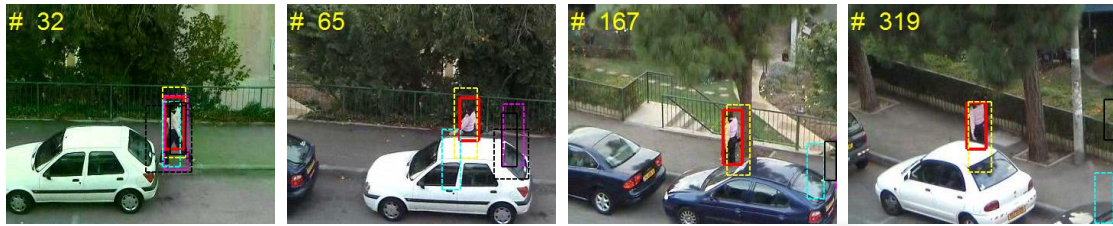


Fig.7 Tracked results of occlusion

图 7 遮挡跟踪结果

### (2) 快速运动实验

图 8 展示了 Liquor 序列快速运动引起运动模糊的跟踪结果,其中,LOT,VTS,VTD 和 Ours 方法的跟踪结果分别用白色(实线)、蓝色(虚线)、黑色(实线)和红色(实线)表示.本文方法基于部件级信息更新外观模型,当目标快速运动时,其外观模型的背景信息被迅速更新,从而在运动模糊情况下仍能准确跟踪到物体.LOT 方法有时能捕捉快速运动的物体(如第 872 帧的白色框),但当出现相似物体干扰时(如 1013 帧的黄色瓶子),则引起跟踪失败.VTD 方法和 VTS 方法因无法有效地响应背景快速变化,产生了跟踪漂移.



Fig.8 Tracked results of quick motions

图 8 快速运动跟踪结果

### (3) 姿势变化实验

图 9 展现了 Lemming 序列中目标姿势变化的跟踪效果,其中,VTD,LOT,IVT,VTS 和 Ours 方法的跟踪结果分别用黑色(实线)、白色(实线)、紫色(虚线)、蓝色(虚线)和红色(实线)表示.IVT 方法采用整体外观模型在没有姿势变化时能够很好地跟踪目标,但常在目标姿势和尺寸发生较大变化时跟踪失败(如第 1081 帧的紫色框).而 VTD 方法和 LOT 方法基于局部块的匹配,不能有效区分目标和背景,无法更新发生大变换的物体.本文方法在更新外观模型时采用部件集的替换方式,能够保留目标的特征,增强处理姿势和尺寸大变化时跟踪的健壮性.



Fig.9 Tracked results of variation of pose and scale

图 9 姿势和尺寸发生较大变化时的跟踪结果

### (4) 剧烈形变和复杂背景实验

图 10 展现了 Diving 序列目标剧烈形变的跟踪效果,其中,LOT,VTD,Godec,IVT,SPT 和 Ours 方法的跟踪结

果分别用白色(实线)、黑色(实线)、黑色(虚线)、白色(虚线)、黄色(虚线)和红色(实线)表示.IVT 方法因其表观模型的整体性,对刚性目标的跟踪效果良好,但在目标剧烈形变时会出现跟踪漂移(如第 76 帧紫色框).VTD 方法和 LOT 方法基于局部匹配跟踪模式,对非刚性变化不敏感,无法捕捉到剧烈形变下的目标区域(如第 76 帧的白色框和黑色框).Godec 方法对目标区域边界不确定,跟踪过程容易受复杂背景的影响,目标框常被扩大进而导致跟踪失败(如第 76 帧黑色虚线框).SPT 方法在超像素尺寸下区分目标和背景,对目标发生缓慢形变时有很好的跟踪效果,但常无法跟踪剧烈形变下的目标(如第 76 帧黄色虚线框).本文方法在部件级表观模型下,能够通过扩展和收缩采样框使目标框贴近物体,以获取准确的跟踪效果(如图 10 中红色框所示).



Fig.10 Tracked results of objects with non-rigid deformation

图 10 严重的形状畸变的跟踪结果

#### (5) 相似物体干扰实验

图 11 中展现了 Face 序列中目标在相似物体靠近或遮挡时的跟踪效果,其中,SPT,VTS,VTD,IVT,SCM,L1 和 Ours 方法的跟踪结果分别用黄色(实线)、蓝色(虚线)、黑色(实线)、紫色(虚线)、绿色(实线)、蓝绿(虚线)和红色(实线)表示.其中,相似物体(男孩)从目标左侧靠近并遮挡目标(女孩),并在最后返回左侧.现有主流的跟踪方法(如 SPT,SCM,VTD,IVT 和 VTS)因干扰物体的特征与目标相似,常把干扰对象识别为目标,引起跟踪漂移.L1 方法根据图像的灰度值进行跟踪,虽然可以避免相似物体靠近时的干扰,但无法处理遮挡情况(如第 38 帧蓝绿色框).本文方法通过检测和和处理相似物体,能够有效区分干扰对象和目标物体,得到更准确的跟踪结果.

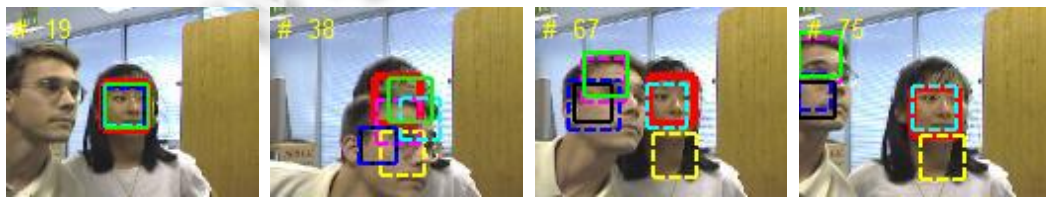


Fig.11 Tracked results of disturbance introduced by similar object

图 11 相似对象干扰的跟踪结果

#### (6) 光照变化实验

图 12 展现了 Singer1 序列中光照变化的跟踪效果,其中,LOT,L1,IVT,VTD 和 Ours 方法的跟踪结果分别用白色(实线)、蓝绿色(虚线)、紫色(虚线)、黑色(实线)和红色(实线)表示.其中,VTD 方法和 LOT 方法因无法适应剧烈的光照变化,导致目标框太大或太小(如第 135 帧黑色框和白色框).IVT 方法可适应光照变化下的目标运动,但当目标与光照的色彩接近时(白色目标和白色光照),导致跟踪漂移(如第 135 帧紫色框).本文方法基于部件级的表观模型建模机制,能够快速记录局部背景信息的变换,从而在光照变化时准确跟踪到物体.



Fig.12 Tracked results when there are large illumination variation

图 12 光照变化的跟踪结果



### 6.3 目标跟踪结果的定量比较分析

本文采用平均中心位置误差和平均跟踪成功率对目标跟踪进行定量分析.平均中心位置误差反映了跟踪结果与实际目标中心偏差的均值,该值越小,说明跟踪结果的中心与目标的中心越接近,跟踪精确度越高.平均跟踪成功率描述了视频序列中跟踪成功的帧数,该值越大,说明跟踪成功的帧数越多,跟踪准确率越高.

表 1 给出了多种方法(包括 IVT,L1,SCM,LSAM,Godec,BHMC,LOT,VTD,VTS,SPT)跟踪 12 组视频序列的跟踪结果中心位置误差(以像素为单位).平均中心位置误差越小,跟踪越准确.表中红色数值代表同一视频中的最小平均误差,蓝色数值表示次小值.由表 1 可知,本文方法(Ours 列)在所有视频序列中引起的平均中心误差都处于最小值和次小值.说明本文方法引起的平均误差均较小,能够产生稳定且准确的跟踪结果.

**Table 1** Average errors of center location in pixels

表 1 跟踪结果的平均中心点误差

序列	SPT	BHMC	Godec	IVT	L1	LOT	LSAM	SCM	VTD	VTS	Ours
basketball	5.6	8.8	8.1	221	113	69.6	83.1	12.9	7.9	7.9	5.0
bird_2	7.6	18.8	8.9	121.8	27.3	61.1	18.8	14.1	50.9	49.8	6.0
Bolt	6.5	118.3	177.7	42.4	249.5	8.3	393.7	10.9	13.4	15.1	6.9
Diving	79	14.3	48.7	75.2	80.1	33.1	77.9	82.5	100.9	99.2	12.3
Face	15.8	10	29.5	4.3	28.5	30	24.5	24.5	31.3	32.1	3.2
gril_mov	10.3	104.8	103.1	373.1	261.4	30.5	45.6	104.8	48.4	50.1	11.5
lemming	5.4	160.4	104.8	13.7	164.4	15.9	72.1	195.2	86.2	86.4	6.7
Liquor	8.8	37.3	76.7	236.3	137.6	10.3	177.1	248.8	94.6	49.5	7.9
Singer1	12.4	54.9	22.5	16.7	50.0	12.2	75.1	21.1	19.9	9.8	10.1
Skating2	33.1	66.3	28.8	206.5	196.6	35.8	51.8	184.1	18	19.1	16.6
transformer2_BMP	11.3	56.3	32.1	135.4	265.1	23.6	99.5	34.5	48.6	35.9	9.7
Woman_sequence	11.3	10.1	57.0	132.7	127.6	119.7	7.5	9	111.1	112.3	7.4

表 2 给出了表 1 中不同方法跟踪成功的帧数,其中,红色表示同一行内的最优值,蓝色表示次优值.本文采用 PASCAL VOC challenge 的检测准则<sup>[30]</sup>来判断是否跟踪成功.假设  $R_T$  为某一帧的跟踪结果, $R_G$  为该帧的标准目标位置,若  $score>0.5$ ,则说明该帧跟踪成功;否则,跟踪失败.其中, $score=area(R_T \cap R_G)/area(R_T \cup R_G)$ , $area()$  为面积函数.同一视频序列中,跟踪成功帧数越多,说明该方法越有效.表 2 中,红色数值代表成功帧数最多,对应最优跟踪方法,蓝色数值次之.由表 2 可知:本文方法(Ours 列)在 80% 以上的视频序列跟踪成功的帧数最多,其余序列跟踪成功的帧数也属于次多.这进一步说明,与现有主流跟踪方法相比,本文方法能够得到更好的跟踪结果.

**Table 2** Count of successful frame

表 2 跟踪结果的成功跟踪帧数

序列	SPT	BHMC	Godec	IVT	L1	LOT	LSAM	SCM	VTD	VTS	Ours
basketball	712	473	584	77	21	340	228	39	520	669	713
bird_2	94	28	84	5	5	8	5	5	10	3	94
Bolt	245	7	4	4	10	241	5	16	218	181	294
Diving	27	82	1	26	1	35	31	27	23	26	154
Face	35	-	33	34	72	39	31	51	34	30	88
gril_mov	1 453	6	1	88	90	1 072	1 158	95	858	759	1375
lemming	1 302	124	370	548	224	668	307	488	615	627	1298
Liquor	1 724	11	349	400	347	886	57	352	482	707	1725
Singer1	344	1	216	160	123	248	73	334	150	344	350
Skating2	93	35	20	40	16	324	324	78	504	473	598
transformer2_BMP	124	33	54	37	19	41	38	47	45	45	124
Woman_sequence	250	69	40	49	47	22	298	305	47	45	308

### 6.4 跟踪方法的计算效率比较

Wu 等人<sup>[31]</sup>提出 Benchmark 用以对比跟踪方法,采用每秒处理视频帧数度量计算效率,处理帧数越多,跟踪效率越高.本文采用相同方法进行效率对比,统计数据见表 3.由表 3 数据可知:本文方法执行效率明显优于 LOT 方法,在大部分情况下略优于 SPT 方法.主要因为本文优化了表观模型更新时的超像素,缩短了建模时间.虽然 VTD 方法和 VTS 方法的计算效率略高于本文方法,但它们是以牺牲质量来换取效率,常引起跟踪失败或漂移.



此外,在精度要求不高时,本文方法通过减少建模超像素数可大幅度提高计算效率.

**Table 3** Efficiency of algorithms, the value is frames per second

**表 3** 跟踪方法的计算效率,数值表示每秒跟踪帧数

序列	SPT	BHMC	LOT	LSAM	SCM	VTD	VTS	Ours
basketball	0.18	0.60	0.10	0.16	0.19	1.27	1.25	0.16
bird_2	0.20	1.15	0.06	0.14	0.15	1.12	1.21	0.23
Diving	0.15	0.92	0.14	0.32	0.18	1.31	1.20	0.13
face	0.18	-	0.42	0.15	0.22	1.28	1.21	0.14
gril_mov	0.12	0.73	0.07	0.30	0.16	1.27	1.17	0.10
lemming	0.15	0.69	0.06	0.21	0.12	1.31	1.30	0.17
liquor	0.13	0.79	0.07	0.24	0.15	1.22	1.25	0.11
Singer1	0.17	0.70	0.06	0.97	0.44	1.28	1.30	0.12
Skating2	0.13	0.99	0.09	0.26	0.18	1.10	1.20	0.15
Woman_sequence	0.21	0.99	0.13	1.74	0.20	1.30	1.25	0.23

## 7 结 论

本文针对目标跟踪中的若干挑战性问题展开研究,提出一种基于部件级表观模型的目标跟踪方法.该方法以超像素描述目标部件,通过分析部件库的特征构建表观模型,能够灵活、有效地表达不断变化的目标和背景.其中,基于部件库特征补集的表现模型更新方法,使本文方法能够在目标被长期遮挡、发生剧烈形变或遭受运动模糊时仍能准确记录目标特征,提供强健、有效的表现模型.此外,本文方法还检测和处理与目标相似的物体靠近和遮挡目标时所带来的干扰,有效地解决了相似物体引起的跟踪漂移问题.基于大量视频序列的实验结果表明:与现有主流跟踪方法相比,在多种跟踪挑战情况下(含遮挡、剧烈形变、光照变化、运动模糊、姿势变化、相似物体干扰、复杂背景),本文方法的跟踪结果具有更多的跟踪成功帧数和更小的跟踪结果中心误差,说明本文方法能够更有效地处理跟踪中的这些挑战,提供更准确且稳定的跟踪结果.

## References:

- [1] Yilmaz A, Javed O, Shah M. Object tracking: A survey. *ACM Computing Surveys*, 2006,38(4). [doi: 10.1145/1177352.1177355]
- [2] Yang HX, Shao L, Zheng F, Wang L, Song Z. Recent advances and trends in visual tracking: A review. *Neurocomputing*, 2011, 74(18):3823–3831. [doi: 10.1016/j.neucom.2011.07.024]
- [3] Avidan S. Ensemble tracking. In: *Proc. of the 18th IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR)*. San Diego, 2005. 494–501. [doi: 10.1109/CVPR.2005.144]
- [4] Hua S, Li CH, Wei FM, Wang HW. Moving object tracking based on location and confidence of pixels. *Journal of Computer Research and Development*, 2008,42(10):1726–1732 (in Chinese with English abstract).
- [5] Lim J, Ross D, Lin RS, Yang MH. Incremental learning for visual tracking. In: *Proc. of the Neural Information Processing Systems Foundation*. 2005. 793–800. <http://papers.nips.cc/book/advances-in-neural-information-processing-systems-18-2005>
- [6] Babenko B, Yang MH, Belongie S. Visual tracking with online multiple instance learning. In: *Proc. of the 22th IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2009. 983–990. [doi: 10.1109/CVPR.2009.5206737]
- [7] Mei X, Ling H. Robust visual tracking using  $l_1$  minimization. In: *Proc. of the 12th IEEE Int'l Conf. on Computer Vision (ICCV)*. Kyoto, 2009. 1436–1443. [doi: 10.1109/ICCV.2009.5459292]
- [8] Kwon J, Lee K M. Visual tracking decomposition. In: *Proc. of the 23th IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. San Francisco, 2010. 1269–1276. [doi: 10.1109/CVPR.2012.6247880]
- [9] Godec M, Roth PM, Bischof H. Hough-Based tracking of non-rigid objects. In: *Proc. of the 13th IEEE Int'l Conf. on Computer Vision (ICCV)*. Barcelona, 2011. 81–88. [doi: 10.1109/ICCV.2011.6126228]
- [10] Wang S, Lu H, Yang F, Yang MH. Superpixel tracking. In: *Proc. of the 13th IEEE Int'l Conf. on Computer Vision (ICCV)*. Barcelona, 2011. 1323–1330. [doi: 10.1109/ICCV.2011.6126385]
- [11] Wang JY, Chen XL, Gao W, Zhao DB. Adaptive appearance model robust to background variations. *Ruan Jian Xue Bao/Journal of Software*, 2006,17(5):1001–1008 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/17/1001.htm>

- [12] Ren X, Malik J. Learning a classification model for segmentation. In: Proc. of the 9th IEEE Int'l Conf. on Computer Vision (ICCV). Nice, 2003. 10–17. [doi: 10.1109/ICCV.2003.1238308]
- [13] Mori G, Ren X, Efros A, Malik J. Recovering human body configurations: Combining segmentation and recognition. In: Proc. of the 17th IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR). Washington, 2004,2:326–333. [doi: 10.1109/CVPR.2004.1315182]
- [14] Vedaldi A, Soatto S. Quick shift and kernel methods for mode seeking. In: Proc. of the 10th European Conf. on Computer Vision (ECCV). Marseille: Springer-Verlag, 2008. 705–718. [doi: 10.1007/978-3-540-88693-8\_52]
- [15] Levinshtein A, Stere A, Kutulakos K, Fleet D, Dickinson S, Siddiqi K. Turbopixels: Fast superpixels using geometric flows. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(12):2290–2297. [doi: 10.1109/TPAMI.2009.96]
- [16] Radhakrishna A, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S. Slicsuperpixels. Technical Report, 149300, EPFL, 2010.
- [17] Ai HZ, Liang LH, Xu GY, Zhang B. Face detection based on skin color and template. Ruan Jian Xue Bao/Journal of Software, 2001, 12(12):1784–1792 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/20011207.htm> [doi: 10.13328/j.cnki.jos.2001.12.007]
- [18] Li HQ, Liu ZK, Zhan S. A segmentation method of color texture image. Chinese Journal of Computers, 2001,24(9):965–971 (in Chinese with English abstract). [doi: 10.3321/j.issn:0254-4164.2001.09.011]
- [19] Li Y, Huang G, Nevatia R. Learning to associate: HybridBoosted multi-target tracker for crowded scene. In: Proc. of the 22th IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR). Miami, 2009. 2953–2960. [doi: 10.1109/CVPR.2009.5206735]
- [20] Comaniciu D, Ramesh V, Meer P. Kernel-Based object tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003, 25:564–577. [doi: 10.1109/TPAMI.2003.1195991]
- [21] Li XR, Wu FC, Hu ZY. Convergence of a mean shift algorithm. Ruan Jian Xue Bao/Journal of Software, 2005,16(3):365–374 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/365.htm>
- [22] Wen ZQ, Cai ZX. Convergence analysis of mean shift algorithm. Ruan Jian Xue Bao/Journal of Software, 2007,18(2):205–212 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/18/205.htm> [doi: 10.1360/jos180205]
- [23] Sun JG, Liu J, Zhao LY. Clustering algorithms research. Ruan Jian Xue Bao/Journal of Software, 2008,19(1):48–61 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/48.htm> [doi: 10.3724/SP.J.1001.2008.00048]
- [24] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002,24(5):603–619. [doi: 10.1109/34.1000236]
- [25] Kwon J, Lee KM. Tracking by sampling tracker. In: Proc. of the 13th IEEE Int'l Conf. on Computer Vision (ICCV). Barcelona, 2011. 1195–1202. [doi: 10.1109/ICCV.2011.6126369]
- [26] Kwon J, Lee KM. Tracking of a non-rigid object via patch-based dynamic appearance modeling and adaptive Basin Hopping Monte Carlo sampling. In: Proc. of the 22th IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR). Miami, 2009. 1208–1215. [doi: 10.1109/CVPR.2009.5206502]
- [27] Zhong W, Lu HC, Yang MH. Robust object tracking via sparsity-based collaborative model. In: Proc. of the 25th IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR). Providence, 2012. 1838–1845. [doi: 10.1109/CVPR.2012.6247882]
- [28] Jia X, Lu HC, Yang MH. Visual tracking via adaptive structural local sparse appearance model. In: Proc. of the 23th IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR). Providence, 2012.1822–1829. [doi: 10.1109/CVPR.2012.6247880]
- [29] Oron S, Bar-Hillel A, Levi D, Avidan S. Locally orderless tracking. In: Proc. of the 23th IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR). Providence, 2012. 1940–1947. [doi: 10.1109/CVPR.2012.6247895]
- [30] Everingham M, Gool LV, Williams C, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. Int'l Journal of Computer Vision (IJCV), 2010,88(2):303–338. [doi: 10.1007/s11263-009-0275-4]
- [31] Wu Y, Lim J, Yang MH. Online object tracking: A benchmark. In: Proc. of the 26th IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR). Portland, 2013. 2411–2418. [doi: 10.1109/CVPR.2013.312]

#### 附中文参考文献:

- [4] 施华,李翠华,韦凤梅,王华伟.基于像素可信度和空间位置的运动目标跟踪.计算机研究与发展,2005,42(10):1726–1732.

- [11] 王建宇,陈熙霖,高文,赵德斌.背景变化鲁棒的自适应视觉跟踪目标模型.软件学报,2006,17(5):1001-1008. <http://www.jos.org.cn/1000-9825/17/1001.htm>
- [17] 艾海舟,梁路宏,徐光祐,张钺.基于肤色和模板的人脸检测.软件学报,2001,12(12):1784-1792. <http://www.jos.org.cn/1000-9825/20011207.htm> [doi: 10.13328/j.cnki.jos.2001.12.007]
- [18] 李厚强,刘政凯,詹曙.一种彩色纹理图像的分割方法.计算机学报,2001,24(9):965-971. [doi: 10.3321/j.issn:0254-4164.2001.09.011]
- [21] 李乡儒,吴福朝,胡占义.均值漂移算法的收敛性.软件学报,2005,16(3):365-374. <http://www.jos.org.cn/1000-9825/16/365.htm>
- [22] 文志强,蔡自兴.Mean Shift 算法的收敛性分析.软件学报,2007,18(2):205-212. <http://www.jos.org.cn/1000-9825/18/205.htm> [doi: 10.1360/jos180205]
- [23] 孙吉贵,刘杰,赵连宇.聚类算法研究.软件学报,2008,19(1):48-61. <http://www.jos.org.cn/1000-9825/19/48.htm> [doi: 10.3724/SP.J.1001.2008.00048]

## 附录:图像序列来源汇总表

Table 4 Total image sequences

表 4 图像序列汇总表

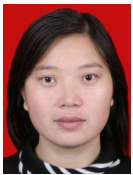
图像序列	来源	帧数	主要挑战
basketball	SPT <sup>[8]</sup>	725	遮挡、复杂背景
bird_2	SPT <sup>[8]</sup>	98	严重遮挡
bolt	SPT <sup>[8]</sup>	350	快速移动
Diving	BHMC <sup>[14]</sup>	230	剧烈形变
face	L1 <sup>[7]</sup>	100	相似对象干扰
gril_mov	SPT <sup>[8]</sup>	1 500	复杂背景
lemming	SPT <sup>[8]</sup>	1 336	姿势和尺寸严重变形
liquor	SPT <sup>[8]</sup>	1 741	姿势和尺寸严重变形
Singer1	VTS <sup>[13]</sup>	350	光照变化
Skating2	VTD <sup>[8]</sup>	706	姿势和尺寸严重变形
transformer2_BMP	SPT <sup>[8]</sup>	124	剧烈形变
Woman_sequence	SPT <sup>[8]</sup>	333	部分遮挡



王美华(1970—),女,江苏金坛人,副教授,主要研究领域为人工智能,优化设计,机器学习.



刘福明(1989—),男,硕士生,主要研究领域为人工智能,目标跟踪.



梁云(1981—),女,博士,副教授,CCF 会员,主要研究领域为图形图像处理,模式识别,计算机视觉.



罗笑南(1963—),男,博士,教授,博士生导师,主要研究领域为图形图像处理,数字家庭,数字通信.