

# 基于两级隐式形状模型的抗遮挡目标跟踪\*

张路平, 李 飏, 王鲁平, 韩建涛

(国防科技大学 电子科学与工程学院, 湖南 长沙 410073)

**摘要:**为解决遮挡条件下的目标跟踪问题,提高定位精度,提出一种基于两级隐式形状模型的目标跟踪算法。利用 Fast Hessian 检测子提取待跟踪目标及周围区域的局部关键点区域构建码本字典,用 SURF 描述子获取码本的特征描述矢量建立码本支持模型,利用广义 Hough 变换建立码本字典集与目标之间的共生关系,通过隐式形状模型进行在线学习更新,通过寻找投票空间中的极大值来对目标进行定位,根据跟踪过程中目标遮挡程度的不同,分别赋予目标自身码本投票及周围码本投票不同权重,提高不同遮挡状态下目标定位精度。实验结果表明,在目标被遮挡甚至不可见或者丢失后重新回到视场时,该算法均能鲁棒地定位出目标。

**关键词:** 目标抗遮挡;隐式形状模型;码本字典;SURF 描述子;广义 Hough 变换

中图分类号:TP311 文献标志码:A 文章编号:1001-2486(2013)06-0096-07

## Anti-occlusion object tracking based on two-level ISM model

ZHANG Luping, LI Biao, WANG Luping, HAN Jiantao

(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

**Abstract:** A target tracking algorithm based on two-level implicit shape model is proposed to solve the tracking problem under occlusion and improve the location accuracy. Firstly, partial key points about the target and surrounding areas were extracted to build the codebook dictionary by Fast Hessian detector, and the feature description vectors of the codebook were constructed by SURF descriptor to establish the codebook support model. Secondly, the symbiotic relationship between the codebook dictionary set and target was established through the generalized Hough transform, and the online updating was accomplished by the implicit shape model. Finally, by finding the maximum value in the voting space, the target was located. According to the occlusion states in the tracking process, different voting weights were assigned to the codebook of target itself and surrounding area respectively, in order to improve the location accuracy under different occlusions. Experiments show the algorithm can locate the target robustly even though the target is occlusive, or even not visible, or returns to the field of view after missing.

**Key words:** object anti-occlusion; implicit shape model; codebook dictionary; SURF descriptor; generalized Hough transform

视频跟踪在人机交互、视频监控、虚拟世界、运动识别、视频索引、车辆定位等方面有着广泛的应用。根据使用信息的不同,可将跟踪分为基于目标表示的跟踪与基于目标定位的跟踪,主要包括模型匹配法、特征匹配法、分割法、基于检测的跟踪算法以及 Meanshift、粒子滤波及其改进算法。由于目标自身状态及周围环境的改变,对目标进行长时间精确跟踪面临诸多挑战,如目标姿态变化、光照变化、帧丢失、目标遮挡等。

视频序列中的运动目标总是与一定的上下文信息相关联,当出现遮挡或目标外观模型改变时,利用上下文信息不仅能预测目标位置,还能有效抑制由相似表现引起的跟踪漂移问题。近年来,出现了许多利用上下文信息进行目标跟踪的方

法。文献[1]将目标外观表示为多个局部子区域的集合,通过建立在线更新的混合高斯投票模型得到多区域联合投票结果,但它仅能处理部分遮挡。文献[2]使用数据挖掘技术,在线寻找有利的辅助目标,将其运用到跟踪中解决遮挡及杂乱背景等问题,这种融入辅助目标的合作式跟踪方法具有较强的鲁棒性,但由于每帧都需要寻找辅助目标,计算复杂耗时,且找到的目标不一定都能有效地预测目标位置。文献[3]将目标时空关系融合到跟踪中,通过一个动态的 Markov 随机域对上下文关系建模。文献[4]提出一种级联信任度滤波,它依次将对目标大小约束、背景优势约束、轨迹平滑性约束结合起来,以增强算法的鲁棒性。文献[5]使用强运动将各种约束与目标联系在一

\* 收稿日期:2013-04-28

基金项目:国家部委资助项目;国家 863 计划资助项目

作者简介:张路平(1985—),男,云南曲靖人,博士研究生,E-mail:zhangluping-002@163.com;

李飏(通信作者),男,研究员,博士,博士生导师,E-mail:libiao\_cn@163.com

起,即使目标不可见时,也能用上下文辅助信息定位出目标,缺点是对所有的局部特征进行检测和匹配比较费时,且不同观测角度下的目标运动不容易预测。文献[6]用局部图像块的直方图来表示目标,利用目标自身的结构信息来处理部分遮挡问题,不足之处是模板不随时间更新,没有构建目标与周围环境的相关性。文献[7]提出一种利用生成辨别混合的合作训练方法对目标进行在线跟踪,但遮挡时目标外观模型未更新,辨别分类器性能下降,导致定位不准确。文献[8]提出一种基于时空约束的鲁棒的 TLD (Tracking Learning Detecting) 跟踪算法,利用 PN 学习对目标运动轨迹进行约束,能很好地捕获视场中消失的目标,但由于缺少空间上下文约束,容易跟踪到相似的运动目标上。文献[9]提出一种时空结构上下文学习方法,通过构建增量空间模型表示目标低维特征矢量,同时引入贡献者来预测目标位置,由于增量空间涉及特征值分解等复杂运算,实时性不强。Yang 等构建了一种上下文可知的跟踪器<sup>[10]</sup>来跟踪目标周围的任意区域而非目标本身,辅助目标与目标之间一致的运动相关性能在很大程度上抑制漂移。文献[11]在考虑相似目标之间空间关系的同时也对这些相似目标进行跟踪,但是该方法忽略了目标的时间维信息,对目标外观改变异常敏感。文献[12]提出一种多级 LPBoost 算法,将跟踪视为多类别分类问题,该方法在约束环境中性能良好,但在遮挡、杂乱背景及光照变化等复杂环境中处理失败。Dinh<sup>[13]</sup>等提出了一种基于转移者与支持者的上下文框架,转移者是与目标有相似外观的区域,支持者是与目标有短时间运动相关的局部关键点,不足之处是很难界定目标与上下文之间的运动相关性。

本文提出一种基于两级隐式形状模型的目标定位方法解决遮挡目标的跟踪问题,其中一级模型源自目标自身,另一级模型来自目标周围。根据遮挡程度不同,分别赋予来自目标自身的码本投票得分及目标周围码本投票得分不同的权重,以提高不同遮挡状态下目标定位精度,然后利用与目标有运动一致性的码本在概率框架下对目标位置投票,只要这些码本可见且与目标间的相对运动耦合关系保持不变,即使目标被遮挡,也能利用这些码本有效预测目标位置。

## 1 隐式形状模型基本原理

隐式形状模型 (Implicit Shape Model, ISM) 主要用于行人检测<sup>[14-15]</sup>,首先通过感兴趣点检测算

子提取表征行人的表观码本,然后通过训练获得码本模型的空间共生分布模型,并将行人分割融入一个概率框架中以产生一个基于像素的置信测度,指明其所在的区域能在多大程度上支持一个行人候选区域,然后使用空间共生分布信息进行广义 Hough 变换 (Generalized Hough Transform, GHT)。利用每一个码本向可能的中心位置投票,第  $i$  个码本字典  $C_i$  对目标位置投票如下所示:

$$\begin{aligned} x_c &= x_i + r_i \cos(\phi_i) \\ y_c &= y_i + r_i \sin(\phi_i) \end{aligned} \quad (1)$$

其中  $(x_c, y_c)$ 、 $(x_i, y_i)$  分别为目标  $T$  及  $C_i$  的中心位置,  $r_i$ 、 $\phi_i$  分别为  $C_i$  相对于  $T$  的距离及角度,所有  $C_i$  都对目标位置投票会得到一个有关投票位置的分布,通过寻找投票峰值得到目标真实位置。

隐式形状模型是在概率框架下将码本字典集作为输入对目标位置进行 Hough 投票,投票得分  $S$  表明目标位于某一位置  $\mathbf{x} = (x, y)$  的概率,即

$$P(\mathbf{x} | I) \propto S = \sum_{f^{(i)} \in F} P(\mathbf{x} | f^{(i)}) P(f^{(i)} | I) \quad (2)$$

其中  $F$  是码本字典特征集,  $f^{(i)}$  是其中的某个特征,出现在图像  $I$  中的概率为  $P(f^{(i)} | I)$ ,  $f^{(i)}$  可以向图像的不同位置投票,对位置  $\mathbf{x}$  投票的概率为  $P(\mathbf{x} | f^{(i)})$ 。所有  $f^{(i)}$  对  $\mathbf{x}$  投票,通过分析投票空间找出概率密度分布的峰值,就可以得到当前目标位置,即

$$\hat{\mathbf{X}} = \max_{f^{(i)} \in F} \sum P(\mathbf{x} | f^{(i)}) P(f^{(i)} | I) \quad (3)$$

本文将隐式形状模型引入跟踪领域,在训练阶段首先利用 Fast Hessian 检测子<sup>[16]</sup>提取与目标有潜在运动相关性的特征点,用 SURF (Speed Up Robust Feature) 描述子对这些特征点建立码本字典特征集,得到码本与目标间的相对关系,并为所有码本建立索引值。当目标被遮挡时,先提取与目标相关的可见兴趣点,获取码本特征,再从码本字典特征集中寻找对应的匹配码本,通过训练阶段记录的该码本相对于目标中心的信息,在概率框架下对中心位置进行投票,投票分数最高的点即为目标中心点。

## 2 建立码本特征集与共生关系

### 2.1 构建码本特征集

利用目标自身及周围显著点的局部表观码本建立码本字典特征集。基于尺度空间理论,用 Fast Hessian 检测子提取稳定的特征点及对应的尺度,用 SURF 算子对特征点进行描述。为了保证旋转

不变性,以特征点为中心计算半径为  $6s$  ( $s$  为特征点对应的尺度) 的邻域的 Haar 小波响应,通过对响应值进行高斯加权,计算每一个特征点的主方向  $\phi_0$ 。以特征点为中心将坐标轴旋转到主方向,根据主方向在每一个特征点周围提取  $20 \times 20$  大小的图像块,将其分成  $4 \times 4$  的子区域,计算每一个子区域的 Haar 小波响应,将权值系数赋予响应值,以增强对几何变换的鲁棒性。将每个子区域对应的绝对值相加,形成 64 维特征描述矢量,再进行归一化处理,增强对光照变化的鲁棒性。

得到特征点及其特征描述矢量后,利用 GHT 建立特征点与目标的共生关系。

## 2.2 利用 GHT 建立共生关系

GHT 是一种强有力的描述方法,主要用于解决不易用解析式表达的曲线或目标轮廓的检测问题。利用 GHT 建立特征点与目标之间的共生关系。将平面坐标系中目标与特征点之间的几何关系映射成极坐标下的距离  $r$  及角度  $\phi$  (相对于主方向角度  $\phi_0$ ), 将所有的共生关系以  $(\mathbf{d}, r, \phi, \Sigma)$  形式存入数据库  $DB$ , 其中四个参数分别表示特征点描述子、距离、角度及方差矩阵。利用特征点对目标进行投票,通过寻找投票峰值,把强弱耦合运动分开,从而实现对目标的定位。对于有限个特征点,将特征  $f^{(i)}$  对目标位置的投票近似为高斯分布,即

$$P(\mathbf{x} | f^{(i)}) \propto \frac{1}{\sqrt{2\pi} |\Sigma|} \cdot \exp(-0.5(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)) \quad (4)$$

其中  $\mu$  为均值,  $\Sigma$  为方差矩阵。

为了估计  $P(f^{(i)} | I)$ , 将特征  $f^{(i)}$  与数据库  $DB$  中存储的特征采用点乘方式进行匹配,如果最佳匹配值小于某一固定门限  $Th$ , 则认为该特征为匹配上的特征点,即:

$$\max_{\mathbf{d} \in DB} (f^T \cdot \mathbf{d}) > Th \quad (5)$$

如果没有特征可以匹配上,则将其作为一个新的码本特征添加到数据库中。

检测到每一帧图像中的特征点后,利用中值流跟踪<sup>[17]</sup> (Median Flow Tracker) 算法对这些特征点进行跟踪筛选,对小于中值 50% 的特征点的投票矢量进行更新,更新准则如下所示:

$$\begin{aligned} r_t^{(i)} &= \alpha r_{t-1}^{(i)} + (1 - \alpha) r^{(i)} \\ \varphi_t^{(i)} &= \alpha \varphi_{t-1}^{(i)} + (1 - \alpha) (\varphi^{(i)} - \varphi_0^{(i)}) \\ \Sigma_t^{(i)} &= \alpha \Sigma_{t-1}^{(i)} + (1 - \alpha) \Sigma^{(i)} \end{aligned} \quad (6)$$

其中:

$$r^{(i)} = \|\mathbf{x}_t^{(i)} - \mathbf{x}_t^*\|, \varphi^{(i)} = \angle(\mathbf{x}_t^{(i)}, \mathbf{x}_t^*) \quad (7)$$

$$\Sigma^{(i)} = (\mathbf{x}_t^* - \mu^{(i)})(\mathbf{x}_t^* - \mu^{(i)})^T$$

对于强相关特征,方差随之减小,而投票峰变得更尖锐。

## 3 两级隐式形状投票模型

假定目标及特征集中某个特征  $f^{(i)}$  的位置分别为  $\mathbf{x}^* = (x^*, y^*)$ 、 $\mathbf{x}^{(i)} = (x^{(i)}, y^{(i)})$ , 通过一定的运动模型将特征与目标联系在一起,并假定短时间内特征与目标的位置相对固定。利用 ISM 中的局部特征来对目标位置投票的同时,也对码本字典特征库及其与目标位置之间的耦合关系进行更新。在估计  $f^{(i)}$  对 ISM 模型的投票得分时,也估计与目标的相对位置  $\bar{x} = x^* - x^{(i)}$ 。采用指数遗忘准则<sup>[5]</sup>, 通过在线方式对 ISM 模型进行估计,即:

$$P_t(f^{(i)} | I) \propto \alpha P_{t-1}(f^{(i)} | I) + (1 - \alpha) P(f^{(i)} | I_t) \quad (8)$$

$$\begin{aligned} P_t(\bar{x} | f^{(i)}) &\propto \alpha P_{t-1} P(\bar{x} | f^{(i)}) \\ &+ (1 - \alpha) P(\mathbf{x}^* - \mathbf{x}^{(i)} | f^{(i)}) \end{aligned} \quad (9)$$

其中  $\alpha \in [0, 1]$ , 决定了前一时刻的估计权重。当前帧图像  $I_t$  及目标位置  $\mathbf{x}^*$  都包含在  $P(f^{(i)} | I_t)$  中,  $P(\mathbf{x}^* - \mathbf{x}^{(i)} | f^{(i)})$  是当前帧中的相对目标位置。

本文仅在目标位置  $\mathbf{x}^*$  或对应的局部图像特征位置  $\mathbf{x}^{(i)}$  发生改变时,才对指示函数进行更新,即:

$$\|\mathbf{x}_t^i - \mathbf{x}_{t-1}^i\|_2 > \theta \quad \text{or} \quad \|\mathbf{x}_t^* - \mathbf{x}_{t-1}^*\|_2 > \theta \quad (10)$$

其中  $\theta$  为运动判决门限。

当目标被遮挡时,应用 ISM 模型对目标位置进行投票,将式(8)、式(9) 带入式(2), 得到最终的投票概率为:

$$P(\mathbf{x} | I_t) \propto S_t = \sum_{f^{(i)} \in F} P_t(\mathbf{x} | f^{(i)}) P_t(f^{(i)} | I_t) \quad (11)$$

当目标未被遮挡时,可只利用目标自身信息进行定位,用 TLD<sup>[8]</sup>、CT<sup>[18]</sup> (Compressive Tracking) 等方法鲁棒地跟踪目标;当目标完全被遮挡时,用目标周围码本信息对其进行预测。根据遮挡程度的不同,分别赋予目标自身及周围码本不同的投票权值,即遮挡不严重时,目标自身码本投票权重较大,遮挡严重时,目标周围码本投票权重较大。本文利用重叠度来衡量目标遮挡程度,通过自适应尺度采样,提取与目标模板最相似的区域,重叠度  $\lambda_{AB}$  计算如下所示:

$$\lambda_{AB} = \frac{B_A \cap B_B}{B_A \cup B_B - B_A \cap B_B} \quad (12)$$

其中  $B_A \cap B_B$ 、 $B_A \cup B_B$  分别表示目标图像块  $B_A$  与  $B_B$  的交集与并集。目标未遮挡时,  $\lambda_{AB} = 1$ , 完全遮挡时,  $\lambda_{AB} = 0$ , 部分遮挡时,  $0 < \lambda_{AB} < 1$ 。分别用  $\lambda_{AB}$ 、 $(1 - \lambda_{AB})$  对码本信息投票加权, 得到二级隐式形状投票概率为

$$P(\mathbf{x} | I_t) \propto S_t = \lambda_{AB} \sum_{f^{(i)} \in F^i} P_t(\mathbf{x} | f^{(i)}) P_t(f^{(i)} | I_t) + (1 - \lambda_{AB}) \sum_{f^{(j)} \in F^o} P_t(\mathbf{x} | f^{(j)}) P_t(f^{(j)} | I_t) \quad (13)$$

其中,  $F^i$ 、 $F^o$  分别表示来自于目标自身及周围区域的码本字典特征集。

### 4 算法实现

基于短时间内目标与特征点间的相对运动耦合关系保持不变且特征点可见的前提假设, 实现本文算法。

(1) 交互式选取第一帧图像中待跟踪目标, 用 SURF 算子提取目标自身及周围区域特征点, 建立码本特征集, 利用 GHT 建立目标与特征点间的共生关系。

(2) 检测当前帧中的特征点, 通过中值流跟踪方法对特征点进行筛选, 对满足条件的特征点, 寻找码本集中与之相匹配的特征及其相应的共生关系, 更新投票矢量。

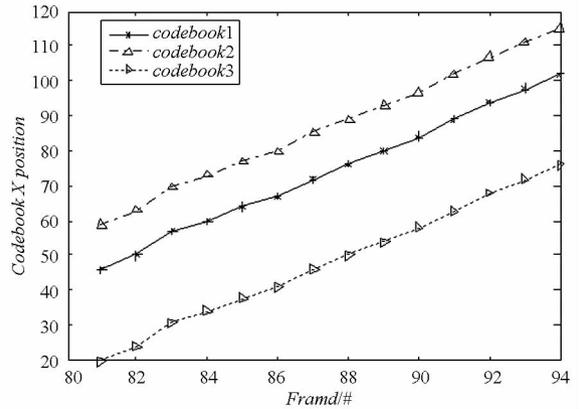
(3) 对未匹配上的特征点, 将其存入 DB, 构建新的共生关系, 并初始化投票矢量。

(4) 当前帧中所有满足条件特征点都更新或初始化完毕后, 计算目标重叠度, 获取两级模型的不同权重, 根据式 (13) 计算最终的投票概率分布, 寻找投票峰值估计目标位置, 同时利用目标位置更新投票矢量。

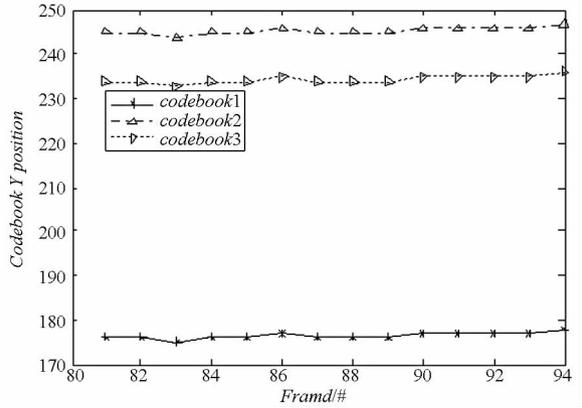
### 5 实验结果分析

在 2GB 内存、2.8GHz 主频的电脑上利用 MATLAB 与 C 混合编程方式实现本文算法。为了验证本文算法对目标定位的可行性及有效性, 选取一组具有典型运动特征的瓶子序列进行测试, 序列长度为 361, 图像大小为  $150 \times 170$ 。瓶子在手的作用下逐渐运动至一杯子后面, 遮挡后被抬起, 被放下, 再从杯子后面移出, 经历未遮挡 (1 ~ 120 帧)、部分遮挡 (121 ~ 140 帧)、完全遮挡 (141 ~ 231 帧)、部分遮挡 (232 ~ 241 帧)、未遮挡 (242 ~ 361 帧) 等 5 个状态。图 1 统计了 14 帧未遮挡图像中目标与周围三个码本 (手表上的特征点区域) 间  $r$ 、 $\phi$  随时间的变化曲线, 图 1(a)、图 1

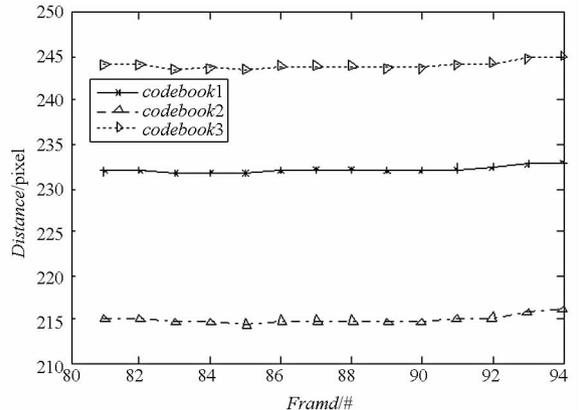
(b) 分别给出了三个运动码本的  $x$  及  $y$  坐标, 由图可得, 码本朝水平方向运动。图 1(c)、图 1(d) 为 3 个码本与目标之间的距离  $r$  及角度  $\phi$  在前后帧的变化曲线, 它们基本保持为一固定值, 即码本与目标之间的共生关系基本保持不变。图 1(e)、图 1(f) 为利用运动码本及其与目标间的耦合关系, 通过隐式形状模型估计得到的目标中心坐标, 不同码本估计得到的目标位置差异较小, 会在投票空间某一位置形成很高的峰值, 该峰值就对应当前帧目标位置。只要目标与码本间的运动耦合关系保持不变且码本可见, 就能利用码本来预测遮挡条件下的目标位置。



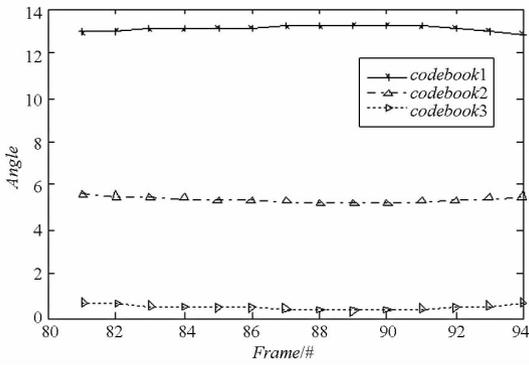
(a) 码本  $x$  坐标



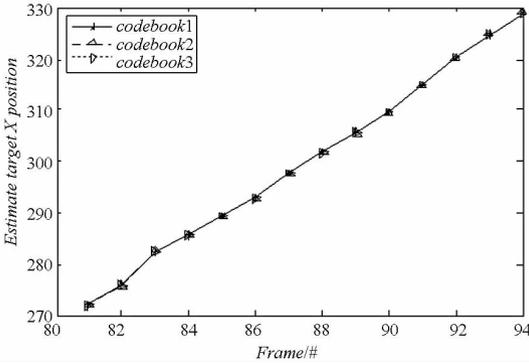
(b) 码本  $y$  坐标



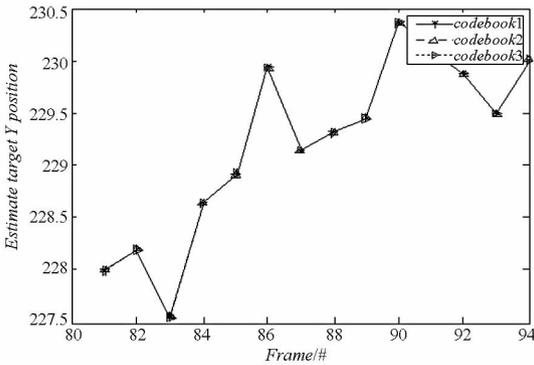
(c) 距离



(d) 角度



(e) 码本估计出的目标 x 坐标



(f) 码本估计出的目标 y 坐标

图 1 码本与目标间的共生关系

Fig. 1 The Co-occurrence relation between codebooks and object



图 3 基于隐式形状模型的位置估计

Fig. 3 Position estimation based on ISM model

选用三种经典跟踪算法与本文方法进行比较: Frag<sup>[6]</sup>、CoGD<sup>[7]</sup>、TLD<sup>[8]</sup>, 其中前两种方法利用了上下文信息。本文主要研究遮挡情况下的目标定位问题, 选用 TLD 作为目标未被遮挡情形下的基本跟踪方法。图 4 给出了采用上述三种方法

图 2 给出了图像序列重叠度变化曲线, 随着目标逐渐被遮挡, 重叠度逐渐减小, 完全被遮挡时重叠度变为 0, 重叠度可较好地反映出目标遮挡程度。

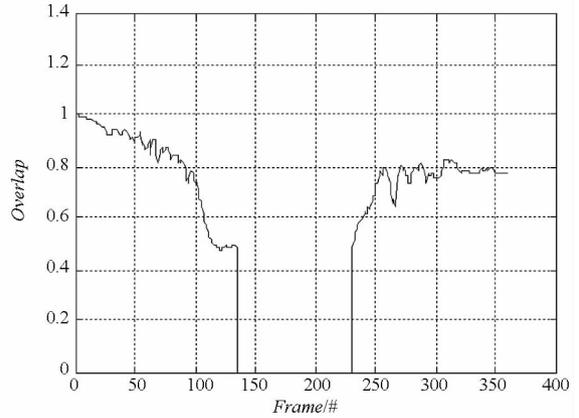


图 2 序列图像重叠度

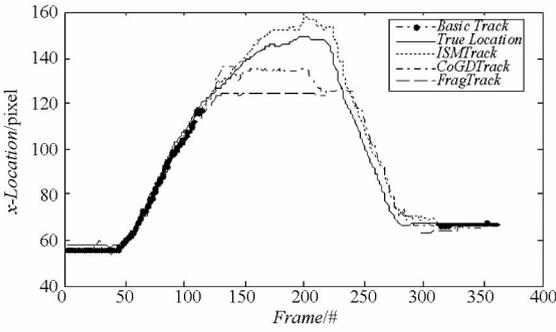
Fig. 2 The overlap of image sequences

利用第一帧获取的信息构建码本特征集及与目标间的共生关系, 之后通过在线学习不断地对码本进行更新。部分参数取值如下: 调节因子  $\alpha = 0.8$ , 以保持前后两帧间的运动耦合相关性。初始方差矩阵  $\Sigma = \sigma^2 I$ ,  $I$  为  $2 \times 2$  单位矩阵,  $\sigma^2 = 10$ , 判决门限  $\theta = 3$ 。

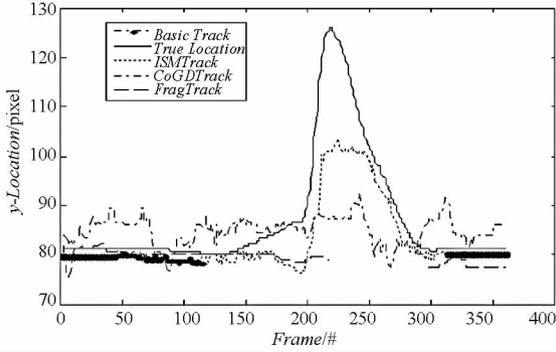
图 3 给出瓶子未被遮挡、部分遮挡、完全遮挡状态下利用码本对目标位置进行的估计, 图中直线表示码本对目标位置的投票, 圆圈表示估计得到的中心位置。完全遮挡时, 利用周围 (袖口及手表区域) 码本对目标位置进行投票估计, 由于投票码本与目标有较强的运动耦合性, 且这种特性在短时间内保持不变, 投票空间成单峰分布, 通过寻找投票峰值就可得到目标估计位置。

及本文方法估计得到的目标位置与真实位置之间的对比, 从图上直观地看出, 目标未被遮挡 (1 ~ 120 帧) 时四种方法得到的估计位置与真实位置较为接近, 完全遮挡时 TLD 不返回目标任何信息。在这四种方法中, 本文方法估计得到的位置

与真实位置最为接近。



(a) 目标  $x$  方向位置



(b) 目标  $y$  方向位置

图 4 目标跟踪结果对比

Fig. 4 The comparison of tracking results

记  $\hat{X}_k$  为第  $k$  帧估计得到的目标位置,  $X_k, \sigma_k$  分别为采用金本位法<sup>[4]</sup>得到的真实位置均值与标准差, 中心定位误差 (Center Location Error, CLE) 定义为:  $\hat{X}_k$  与  $X_k$  间的欧式距离。若  $|\hat{X}_k - X_k| < 2\sigma_k$ , 则认为当前帧估计位置有效。记  $TN$  为序列长度,  $RP$  为返回信息帧数,  $TP$  为有效帧数, 用检测率  $R$  (recall) 及准确率  $P$  (precision) 来衡量算法性能, 其定义如下:

$$\begin{aligned} R &= RP/TN \\ P &= TP/RP \end{aligned} \quad (14)$$

令  $\sigma_k = 12$ , 表 1 对四种算法的跟踪性能进行了比较。

表 1 跟踪性能比较

Tab. 1 The comparison of tracking performance

	TLD	Frag	CoGD	ISM
$TN$	361	361	361	361
$RP$	163	282	361	361
$TP$	163	251	300	352
$R$	45%	78%	100%	100%
$P$	100%	89%	83%	97%

TLD 通过在线学习机制获取目标最新的外观特征, 只要目标外观模型不发生突变, 即使目标短

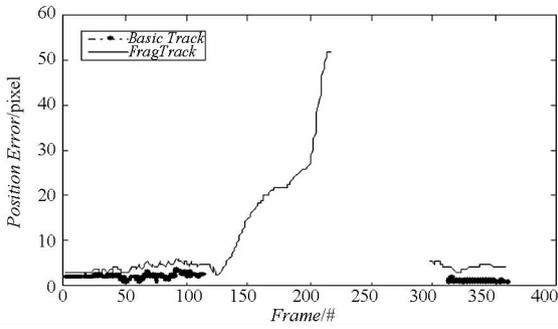
暂消失, 当其再次出现时也能很快重新捕捉到目标。完全遮挡时 TLD 失效, 因此其检测率较低,  $R = 45\%$ , 但返回信息的图像帧中目标估计位置与真实位置之间的偏差较小, 返回信息的图像帧均为有效帧, 准确率  $P = 100\%$ 。

FragTrack 用多个任意组合的图像块来表示目标, 通过比较每一个图像块与其对应图像块的直方图, 来对当前帧中目标可能的位置及尺度进行投票, 综合多个图像块的投票来获取最终的目标位置, 即使目标被部分遮挡, 仍可以用未遮挡部分来估计目标中心, 相比于 TLD 其检测率明显提高,  $R = 78\%$ , 不足之处是仅利用部分目标自身信息得到的估计, 其精度较低,  $P = 89\%$ 。

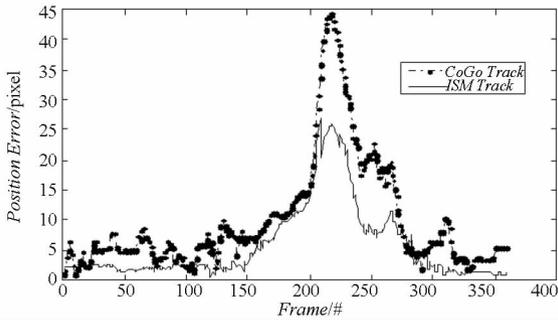
CoGD 利用生成、辨别混合方法来更新目标外观模型, 利用已有数据通过合作训练方式对新数据进行标记。因此即使目标被遮挡, 也能利用周围合作训练数据估计目标位置, 检测率  $R = 100\%$ , 由于遮挡时目标外观模型未更新, 辨分分类器性能下降, 准确率降低,  $P = 83\%$ 。

本文通过构建两级码本与目标间的共生关系, 在概率框下对目标位置进行 Hough 投票, 即使目标被完全遮挡, 也能利用周围码本有效预测目标位置, 检测率  $R = 100\%$ , 准确率  $P = 97\%$ , 估计的目标位置与真实位置之间的偏差为  $6.2\text{pixel}$ 。完全遮挡条件下的目标由周围码本来预测其位置, 当目标突然运动 (瓶子突然被抬起) 时, 前后帧中周围码本与目标间的运动共生关系未及时更新, 估计出的位置与真实位置有一定偏差, 导致准确率比 TLD 低。只要运动平缓, 运动耦合性得以保持, 即使目标不可见, 也能利用周围码本估计出其位置。本文方法实际是以估计精度的降低换取检测率的提高, 确保遮挡情况下也能估计目标位置。

图 5 比较了四种方法得到的 CLE, 目标未被遮挡时, 本文方法与 TLD 得到的 CLE 最小, 约为  $4\text{pixel}$ 。遮挡时, Frag 及 TLD 不返回目标位置信息, 未给出 CLE 具体数值。无论目标是否遮挡, 本文方法及 CoGD 都能估计出目标位置, 但本文方法的 CLE 比 CoGD 的 CLE 小得多, 特别是在完全遮挡条件下, 前者在  $25\text{pixel}$  以内, 后者远大于  $25\text{pixel}$ , 本文方法不仅能准确跟踪未遮挡条件下的目标, 而且也能估计出遮挡条件下的目标位置, 且估计位置与目标真实位置之间的偏差较小, 在四种方法中性能最优。



(a) BasicTrack and FragTrack



(b) CoGDTrack and ISMTrack

图 5 中心定位误差

Fig. 5 The center location error

## 6 结束语

码本特征实质上是一种上下文信息,利用这些信息对目标进行时空约束,从而对目标进行定位。从空间维来看,码本与目标组成一定的共生关系,从时间维上分析,这种共生关系在短时间保持不变。本文正是基于这种暂时不变性来构建隐式形状模型,假定短时间内码本特征集与目标之间形成的运动耦合关系保持不变,通过提取目标周围特征点,构建目标与自身、目标与周围码本特征集间的两级隐式形状模型及共生关系,并依据遮挡程度的不同分别赋予不同的权值,以提高不同遮挡状态下目标定位精度。通过码本投票及更新寻找峰值估计目标位置,即使目标不可见时也能准确估计出目标位置,但一旦目标与码本间的运动耦合关系发生剧变或消失时,例如当目标周围可用码本也被完全遮挡时,将会得到错误的估计位置,这种情形将是本文未来的研究方向。

## 参考文献 (References)

- [1] 孙钟前. 基于上下文信息的运动目标跟踪算法的研究[D]. 哈尔滨:哈尔滨工业大学,2008.  
SUN Zhongqian. Research on contextual information based tracking approach[D]. Harbin:Harbin Institute of Technology, 2008. (in Chinese)
- [2] Yang M, Wu H, Lao S. Intelligent collaborative tracking by mining auxiliary objects [C]//Proceedings of International

- Coference on Computer Vision and Pattern Recognition, 2006: 697 - 704.
- [3] Li Y, Nevatia R. Key object driven multi-category object recognition, location and tracking using spatio-temporal context [C]//Proceedings of European Coference on Computer Vision, 2008;31 - 39.
- [4] Stalder S, Grabner H, Gool L. Cascaded confidence filtering for improved tracking by detection [C]//Proceedings of European Coference on Computer Vision, 2010;369 - 382.
- [5] Grabner H, Matas J, Gool L, et al. Tracking the invisible: learning where the object might be [C]//Proceedings of International Coference on Computer Vision and Pattern Recognition, 2010;1285 - 1292.
- [6] Adam A, Rivlin E, Shimshoni I. Robust fragment-based tracking using the intergral histogram [C]//Proceedings of International Coference on Computer Vision and Pattern Recognition, 2006;798 - 805.
- [7] Yu Q, Dinh T B, Vo N, et al. Online tracking and reacquisition using co-trained generative and discriminative trackers [C]// Proceedings of European Coference on Computer Vision, 2008: 678 - 691.
- [8] Kalal Z, Mikolajczyk K, Matas J. Tracking- Learning- Detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012,34(7):1409 - 1421.
- [9] Wen L, Cai, Z, Zhen L, et al. Online spatio-temporal structural context learning for visual tracking[C]//Proceedings of European Coference on Computer Vision, 2012;716 - 729.
- [10] Yang M, Wu Y, Hua G. Context-aware visual tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 0(0):1195 - 1209.
- [11] Gu S, Tomasi C. Branch and track [C]//Proceedings of International Coference on Computer Vision and Pattern Recognition, 2011;1169 - 1174.
- [12] Saffari A, Godec M, Pock M, et al. Online multi-class LPBoost [C]//Proceedings of International Coference on Computer Vision and Pattern Recognition, 2010: 3570 - 3577.
- [13] Dinh T B, Vo N, Medioni G G. Context tracker: Exploring supporters and distracters in unconstrained environments [C]//Proceedings of International Coference on Computer Vision and Pattern Recognition, 2011;1177 - 1184.
- [14] Leibi B, Leonardi A, Schiele B. Robust object detection with interleaved categorization and segmentation [J]. IJCV Special Issue on Learning for Vision and Vision for Learning,2007.
- [15] Vecera S P, Reilly R C. Figure-ground organization and object recognition progresses: An interactive [J]. J. Exp. Psych:Human perception and Performance,1998,24(2):441 - 462.
- [16] Bay H, Tuytelars T, Van G. Speed-up robust features (SURF) [J]. Computer Vision anf Image Understanding, 2008(110):346 - 359.
- [17] Kalal Z, Mikolajczyk K. Forward-backward error: automatic detection of tracking failures [C]//Proceedings of International Coference on Pattern Recognition, 2010;1 - 4.
- [18] Zhang K, Zhang L. Real-time compressive tracking [C]// Proceedings of European Coference on Computer Vision, 2012;866 - 879.