



文章编号:1001-9081(2018)04-1170-06

DOI:10.11772/j.issn.1001-9081.2017092312

## 基于静态-自适应外观模型纠正的目标跟踪算法

卫保国<sup>1\*</sup>, 葛 莹<sup>1</sup>, 武 宏<sup>1</sup>, 王高峰<sup>2</sup>, 韩文亮<sup>3</sup>

(1. 西北工业大学 电子信息学院, 西安 710129; 2. 贵州宇鹏科技有限责任公司, 贵阳 550000;

3. 中国人民解放军 93108 部队, 黑龙江 齐齐哈尔 161000)

(\*通信作者电子邮箱 wbg@nwpu.edu.cn)

**摘要:**针对单目标跟踪过程中难以长期稳定跟踪的问题,提出一种基于静态-自适应外观模型纠正的跟踪算法。首先将跟踪过程中可能遇到的干扰因素分为来自环境和目标本身两类,分别提出静态外观模型和自适应外观模型,静态外观模型用于全局匹配,自适应外观模型用于局部跟踪,静态模型纠正自适应模型的跟踪漂移问题;使用单链接层次聚类算法去除两种模型融合后引入的噪声;针对运动目标消失再出现时难以捕获的问题,使用静态模型进行全局搜索,捕获目标。对于实验中的视频序列,视频序列中目标的中心位置准确率为 0.9,计算机每秒能够处理 26 帧图像。实验结果表明,该跟踪算法框架可以实现长期稳定的跟踪,具有良好的鲁棒性和实时性。

**关键词:**目标跟踪;静态-自适应外观模型;层次聚类;跟踪算法框架

**中图分类号:**TP391.41    **文献标志码:**A

### Object tracking algorithm based on static-adaptive appearance model correction

WEI Baoguo<sup>1\*</sup>, GE Ping<sup>1</sup>, WU Hong<sup>1</sup>, WANG Gaofeng<sup>2</sup>, HAN Wenliang<sup>3</sup>

(1. College of Electronic Information, Northwestern Polytechnical University, Xi'an Shaanxi 710129, China;

2. Guizhou Yu Peng Technology Company Limited, Guiyang Guizhou 550000, China;

3. Unit 93108, the Chinese People's Liberation Army, Qiqihar Heilongjiang 161000, China)

**Abstract:** For long-term robust tracking to single target, a corrected tracking algorithm based on static-adaptive appearance model was proposed. Firstly, the interference factors that may be encountered in the tracking process were divided into two categories from environment and target itself, then a static appearance model and an adaptive appearance model were proposed respectively. The static appearance model was used for global matching while the adaptive appearance model was employed for local tracking, and the former corrected tracking drift of the latter. A single-link hierarchical clustering algorithm was used to remove the noise introduced by the fusion of the above two models. To capture the re-occurring target, static appearance model was applied for global search. Experimental results on standard video sequences show that the accuracy of tracking the target center is 0.9, and the computer can process 26 frames per second. The proposed tracking algorithm framework can achieve long-term stable tracking with good robustness and real-time performance.

**Key words:** object tracking; static-adaptive appearance model; hierarchical clustering; tracking algorithm framework

### 0 引言

目标跟踪是指在视频图像序列中对指定目标的运动轨迹进行估计,已经被广泛应用到视频监控、智能交通和人机交互等研究领域。尽管目前取得了很大的成就,但仍然存在着相似目标、光照、遮挡、目标形变等干扰因素的影响,使得跟踪算法在复杂场景下的效果较差<sup>[1]</sup>。

为此, Babenko 等<sup>[2]</sup>提出了基于多实例学习 (Multiple Instance Learning, MIL) 的目标跟踪算法,将训练样本扩大为多个样本的样本包,这样可以在一定程度上避免跟踪漂移的问题,但是需要大量实例进行学习。Kalal 等<sup>[3]</sup>提出了一种将跟踪、学习、检测 (Tracking-Learning-Detection, TLD) 相结合的目标跟踪算法,虽然该算法在一定程度上可以解决长期跟踪过程中目标漂移或丢失问题;但是实时性较差,并且在跟踪的初始阶段非常依赖关于目标的先验知识。Zhang 等<sup>[4]</sup>提出了基

于时空上下文 (Spatio-Temporal Context, STC) 的视觉跟踪算法,利用图像的低维信息来建立目标的外观模型,然后基于贝叶斯理论的时空上下文框架,选择当前帧中置信值最高的位置输出为目标位置;但是只使用简单的灰度特征,不能很好地对目标外观进行描述。Danelljan 等<sup>[5]</sup>提出了基于自适应选择颜色属性的实时视觉跟踪技术,将颜色的类别进行细分,在跟踪过程中通过一种类似于主成分分析的降维思想实时地选择比较显著的颜色作为跟踪的依据;但该方法仅对目标颜色特征进行跟踪,特征单一,不能很好地表达目标。

长期跟踪的关键就在于算法能否抵御实际复杂环境中可能出现的各种干扰因素,如:相似目标、光照、遮挡、目标形变。基于这一思路,本文从整体上对干扰因素进行归类,针对不同类别的干扰因素分别提出了静态外观模型和自适应外观模型,两个模型形成一个能够纠正的跟踪算法框架。实验结果表明,本文算法在各种干扰因素下,均能在保证实时性的基础

收稿日期:2017-09-25;修回日期:2017-12-04。

基金项目:贵州省科技厅工业攻关计划项目(2017GZ60903);西安市科技计划项目(2017086CG/RC049)。

作者简介:卫保国(1970—),男,陕西乾县人,副教授,博士,主要研究方向:图像处理、模式识别、计算机视觉; 葛莹(1992—),女,山西大同人,硕士研究生,主要研究方向:目标检测与跟踪。



上表现出良好的鲁棒性。

## 1 目标跟踪算法框架的设计

本文将复杂环境中影响目标跟踪的干扰因素分为两类:第一类来自环境,包括相似目标干扰、光照变化干扰、遮挡干扰、阴影干扰等,相应地提出静态外观模型,由初始模板保存;第二类来自目标本身,包括目标形变干扰、目标尺度变化干扰、目标旋转干扰、目标消失干扰等,相应地提出自适应外观模型。

本文使用静态模型和自适应模型相纠正的跟踪框架,如图1所示。

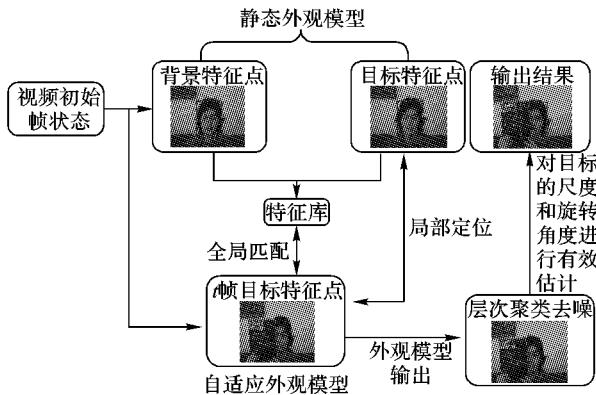


图1 本文的跟踪算法框架示意图

Fig. 1 Schematic diagram of the proposed tracking framework

在跟踪算法中,由于自适应外观模型在每一帧中都要进行更新,因此难免会逐渐将背景信息纳入外观模型中从而产生漂移,而静态外观模型不随跟踪过程进行更新,因此静态外观模型可以在每一帧对自适应外观模型的跟踪结果进行纠正。

## 2 基于静态-自适应外观模型的跟踪算法

### 2.1 静态外观模型的设计

考虑到目标受到局部遮挡时,全局特征提取比较困难,本文采用局部不变特征来作为外观模型的特征描述方式。文献[6]中针对常用的局部尺度不变特征变换(Scale Invariant Feature Transform, SIFT)、PCA-SIFT(Principal Component Analysis-SIFT)和SURF(Speed-up Robust Feature)进行了对比,综合时间复杂度和实际效果,本文选择SURF作为外观模型的特征描述方式,采用欧氏距离作为SURF特征点的距离计算方法。

静态外观模型是基于初始帧的目标外观建立的,由初始帧中感兴趣的目标区域的所有关键点 $x_i^0$ 所对应的描述子组成,其中: $x_i^0 \in P_0$ ,  $P_0 = \{x_1^0, x_2^0, \dots, x_m^0\}$ 是初始帧关键点集合。在跟踪过程中,本文针对初始帧中的每一个关键点 $x_i^0$ 都在当前帧中进行全局搜索以寻找与其匹配的候选关键点 $x_j^t$ ( $x_j^t \in P_t$ ,  $P_t = \{x_1^t, x_2^t, \dots, x_m^t\}$ ), $x_j^t$ 为 $t$ 帧图像关键点的位置,当目标消失再出现时,同样全局搜索当前帧特征,选择与初始帧目标特征匹配的候选特征点。由于初始帧与当前帧之间的帧数可以是任意的非负值,因此为了描述初始帧与之后任意帧中目标的关系,本文针对初始帧中目标的每一个关键点定义一个匹配关系,它体现了两帧中目标的变化情况,如式(1)所示:

$$\mathbf{m}_i = (x_i^0, x_j^t) \quad (1)$$

要求 $x_j^t$ 与 $x_i^0$ 初始帧中关键点满足以下关系:

$$d(x_i^0, x_j^t) < \theta \text{ 且 } \frac{d(x_i^0, x_j^t)}{d(x_i^0, x_k^t)} < \gamma; j \neq k \quad (2)$$

$x_j^t$ 与初始帧中关键点 $x_i^0$ 的绝对距离小于指定阈值 $\theta$ ,同时在第 $t$ 帧中, $x_j^t$ 与 $x_i^0$ 的距离相对其他关键点的距离最近( $\gamma \leq 1$ ),其中 $\theta$ 和 $\gamma$ 均为经验值。

由于本文中设定自适应外观模型需要在每一帧中都进行更新,为了令静态外观模型能够不断地纠正自适应外观模型在跟踪过程中纳入背景信息而产生的误差,需要在当前帧的静态外观模型中将那些与初始帧背景匹配的候选关键点去掉,最后得到静态对应关系集合:

$$L_i^a = \{\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_n\} \quad (3)$$

它表征的是初始帧与当前帧中由静态外观模型所提供的目标外观对应关系。

### 2.2 自适应外观模型的设计

面对来自目标本身的干扰因素以及全部遮挡时,仅依靠静态外观模型不能满足跟踪的鲁棒性,还需要自适应外观模型。

设计自适应外观模型的关键之一就是设置合适的遗忘因子,即自适应外观模型的更新速度,可以及时地感知目标的外观变化,所以本文在跟踪过程中针对每一帧都对自适应外观模型进行更新。由于静态外观模型不随着跟踪过程进行更新,因此静态外观模型可以在每一帧对自适应外观模型的跟踪结果进行纠正。

当前帧的自适应外观模型由上一帧中目标区域中所有关键点附近的区域块组成。自适应外观模型也存在一个自适应对应关系集合。假定相邻两帧之间时间很短,通过估计从第 $t-1$ 帧到第 $t$ 帧的稀疏光流,借助局部优化的方法,建立一个第 $t-1$ 帧与第 $t$ 帧的自适应对应关系集合。

此时在当前帧(第 $t$ 帧)中获取的自适应对应关系集合中的每一项匹配关系 $n_i$ 表征的是第 $t-1$ 帧与第 $t$ 帧的匹配关系:

$$n_i = \{x_i^{t-1}, x_i^t\} \quad (4)$$

为了和静态对应关系集合中的表示方式统一,在获取到每一帧的自适应对应关系结合后都要将其与初始帧对应,即由上一帧(第 $t-1$ 帧)的自适应对应关系 $L_{t-1}^a$ 和当前帧(第 $t$ 帧)的自适应对应关系 $L_t^{a*}$ 可得到:

$$\begin{cases} L_{t-1}^a = \{(x_1^0, x_1^{t-1}), (x_2^0, x_2^{t-1}), \dots, (x_n^0, x_n^{t-1})\} \\ L_t^{a*} = \{(x_1^{t-1}, x_1^t), (x_2^{t-1}, x_2^t), \dots, (x_n^{t-1}, x_n^t)\} \end{cases} \quad (5)$$

可得到 $L_t^a$ :

$$L_t^a = \{(x_1^0, x_1^t), (x_2^0, x_2^t), \dots, (x_n^0, x_n^t)\} \quad (6)$$

将过滤后的静态对应关系集合和自适应对应关系集合进行合并,将其称为对应关系集合,记作 $L_t^*$ :

$$L_t^* = L_t^a \cup L_t^s \quad (7)$$

### 2.3 针对外观模型的去噪

将静态外观模型和自适应外观模型输出的对应关系合并后,难免会将一些背景区域纳入目标区域中。实际上,解决目标跟踪问题的关键在于能否准确地划分出目标,因此本文选择聚类来作为去噪手段。由于背景噪声的类别未知,种类繁多,而 $k$ -均值聚类、混合高斯模型聚类、谱聚类都需要指定类别参数 $k$ ,都不适合作为此处的去噪手段,因此,本文利用单连接层次聚类算法作为去噪手段。

影响聚类效果的关键之一就是样本间距的计算方式,每



一帧中,对应关系集合中的任意两个元素,计算一个差异性度量 $D$ ,直观地反映了所跟踪目标的尺度变化和旋转角度,本文将 $D$ 定义为:

$$D(x_i, x_j) = \| (x_i^t - Hx_i^0) - (x_j^t - Hx_j^0) \| \quad (8)$$

其中: $H$ 是从 $L_t^*$ 估计得到的仿射变换矩阵,用于表征当前帧和初始帧相比目标形变情况。如果在 $L_t^*$ 中匹配关系的数目大于3,可以计算匹配点之间的仿射变换矩阵,经过仿射变换后,初始帧中关键点的相对位置关系被映射到当前帧的坐标系中。

假设 $d_{ij}$ 表示对应关系样本 $x_i$ 和 $x_j$ 间的差异性度量, $D_{ij}$ 表示样本 $x_i$ 和 $x_j$ 所在的簇 $C_i$ 和 $C_j$ 间的差异性度量,则 $D_{ij}$ 应为两簇间最近样本的距离,即 $D_{ij} = \min_{x_i \in C_i, x_j \in C_j} d_{ij}$ 。

针对外观模型的输出结果 $L_t^*$ 进行聚类:

1) 首先将每个样本作为一簇,簇 $C_i$ 和 $C_j$ 间的最短距离 $D_{ij} = d_{ij}$ 。

2) 令 $t = 0$ ,所有簇间的最短距离可以形成一个距离矩阵 $M(t)$ 。

3) 寻找 $M(t)$ 内非对角线上不大于距离阈值 $T$ 的最小元素,设为 $M_{pq}$ ,将 $C_p$ 和 $C_q$ 合并为一个新簇,记为 $C_r$ ,即 $C_r = \{C_p, C_q\}$ 。

4) 计算新簇 $C_r$ 与其他簇 $C_k$ 的最短距离 $D_{kr} = \min\{D_{kp}, D_{kq}\}$ ,对距离矩阵 $M(t)$ 进行更新,将其第 $p$ 、 $q$ 行及 $p$ 、 $q$ 列合并成一个新行新列,对应类 $C_r$ ,新行和新列上的新距离按照 $D_{kr} = \min\{D_{kp}, D_{kq}\}$ 进行计算,所得到的矩阵记为 $M(t+1)$ 。

5) 将 $t + 1$ 赋给 $t$ ,返回步骤3),直到 $M(t)$ 中所有元素均大于 $T$ 或所有样本聚为一簇时结束聚类。

对 $L_t^*$ 经过聚类后,对每一帧的跟踪结果进行估计时只用其中的最大簇 $L_t^{\text{large}}$ 即可。

## 2.4 目标信息的估计

在目标跟踪过程中,对每一帧的跟踪结果实际上由三个数据量组成。

### 2.4.1 目标的中心位置

本文使用目标位移参数 $\mu$ 来表示目标中心位置的变化,即假定目标在初始帧的中心位置为 $x_0$ ,则当前帧中目标的中心位置可以表示为:

$$x_t = x_0 + \mu \quad (9)$$

为了准确地衡量目标的中心位置,本文对聚类后输出的最大簇 $L_t^{\text{large}}$ 中所有匹配关系表征的位移参数之和取均值,即

$$\mu = \frac{1}{\text{Num}} \sum_{i=1}^{\text{Num}} \mu_i \quad (10)$$

其中: $\text{Num}$ 为最大簇 $L_t^{\text{large}}$ 中对应关系的个数,即 $\text{Num} = |L_t^{\text{large}}|$ ; $\mu_i$ 为每个匹配关系对目标位移的估计,即 $\mu_i = x_i^t - Hx_i^0$ 。

综上,每一帧的目标中心位置的位移参数估计如下:

$$\mu = \frac{1}{L_t^{\text{large}}} \sum_{m_i \in L_t^{\text{large}}} (x_i^t - Hx_i^0) \quad (11)$$

### 2.4.2 目标的尺度

本文使用尺度缩放系数来表征目标的尺度变化,使用 $s$ 来表示,即假定初始帧中目标尺度大小为 $S_0$ ,当前帧中目标的尺度大小为 $S_t$ ,则:

$$S_t = s \times S_0 \quad (12)$$

本文对尺度系数进行估计的方法如式(13)所示:

$$s = \text{med}(\{S_{ij}^t / S_{ij}^0, i \neq j\}) \quad (13)$$

其中: $S_{ij}^t$ 表示在 $L_t^{\text{large}}$ 中通过匹配关系估计得到的当前帧的目标尺度,即 $S_{ij}^t = \|x_i^t - x_j^t\|, i \neq j$ ; $S_{ij}^0$ 表示在 $L_0$ 中通过匹配关系估计得到的当前帧的目标尺度,即 $S_{ij}^0 = \|x_i^0 - x_j^0\|, i \neq j$ ; $\text{med}$ 表示取中位数。

综上,目标尺度变化系数的计算方法如式(14)所示:

$$s = \text{med}(\left\{ \frac{\|x_i^t - x_j^t\|}{\|x_i^0 - x_j^0\|}, i \neq j \right\}) \quad (14)$$

### 2.4.3 目标的旋转角度(弧度)

本文使用 $\alpha$ 来表示,即假定初始帧中目标的旋转角度为 $\alpha_0 = 0$ ,那么当前帧中目标的旋转角度为:

$$\alpha_t = \alpha_0 + \alpha = \alpha \quad (15)$$

本文对目标旋转角度的估计方法如式(16)所示:

$$\alpha = \text{med}(\{(\alpha_{ij}^0 - \alpha_{ij}^t), i \neq j\}) \quad (16)$$

其中: $\alpha_{ij}^0$ 表示初始帧中根据关键点的相对位置计算得到的目标相对于水平方向的角度,即 $\alpha_0 = \text{atan}2(x_i^0 - x_j^0), i \neq j$ ; $\alpha_{ij}^t$ 表示当前帧中根据关键点的相对位置计算得到的目标相对于水平方向的角度,即 $\alpha_t = \text{atan}2(x_i^t - x_j^t), i \neq j$ ; $\text{med}$ 表示取中位数。

综上,相对与初始帧,在当前帧中目标的旋转角度计算如式(17)所示:

$$\alpha = \text{med}(\{(\text{atan}2(x_i^0 - x_j^0) - \text{atan}2(x_i^t - x_j^t)), i \neq j\}) \quad (17)$$

## 2.5 算法流程

本文目标跟踪算法具体步骤如下:

初始化 目标的位置用 $(x, y, w, h)$ 表示,分别指目标在图像帧的横向、纵向坐标,以及目标矩形框的宽、高。

目标跟踪过程:

1) 静态外观模型。在初始帧确定矩形框,提取图像的SURF特征,根据目标框将图像分为目标特征点和背景特征点,构成特征库。

2) 自适应外观模型。通过金字塔LK(Lucas-Kanade)光流的前向光流(第 $t-1$ 到 $t$ 帧)和后向光流(第 $t$ 到 $t-1$ 帧)得到 $t$ 帧目标的特征点,在此过程中难免会纳入各种噪声信息。

3) 静态外观模型对自适应外观模型获取目标特征的纠正。通过自适应模型获得 $t$ 帧目标特征与初始帧目标特征点进行匹配,剔除一些背景的特征点,确定最终的目标特征点,作为外观模型的输出,用于估计外观信息和尺度信息。

4) 单连接层次聚类的去噪。对外观模型的输出进行去噪处理,得到的结果作为计算仿射变换矩阵的初始条件,对目标的尺度和旋转角度进行有效估计,实现跟踪算法的鲁棒性和实时性。

输出 下一帧目标位置。

## 3 实验分析

### 3.1 测试环境说明

本文使用台式计算机(Windows XP 32位操作系统,Intel Pentium E5400 CPU 2.7 GHz,内存2.00 GB),利用Python 2.7.9编程实现了基于静态-自适应外观模型的目标跟踪算法。所用图像序列源自文献[7]所提供的目标跟踪标准测试数据集。

文献[7]提供的测试数据集涵盖了目标跟踪过程中绝大多数的干扰因素,各干扰因素缩写含义如表1所示,由于不同视频序列的干扰因素不同,因此不同实验分别选取不同的视频序列进行实验,如表2所示。



表1 干扰因素含义  
Tab. 1 Meaning of interference factors

简称	全称	含义
IV	Illumination Variation	目标区域有明显的光照变化
SV	Scale Variation	目标尺度有明显的变化
OCC	Occlusion	目标被部分或全部遮挡
DEF	Deformation	目标存在明显的形变
MB	Motion Blur	目标区域有明显的模糊出现
FM	Fast Motion	目标存在快速运动情况
IPR	In-Plane Rotation	目标在图像平面上存在旋转
OPR	Out-of-Plane Rotation	目标在图像平面外有旋转
OV	Out-of-View	目标离开摄像头视野
BC	Background Clutters	存在与跟踪目标相似的物体

表2 测试所用视频序列及其包含的干扰因素  
Tab. 2 Experimental video sequences and their interference factors

视频	IV	SV	OCC	DEF	MB	FM	IPR	OPR	OV	BC
Suv		√				√	√		√	√
Football1		√					√	√		√
Dog1	√			√			√	√	√	
Singer1	√	√		√			√	√		
Coke	√	√			√	√	√			√
Skating1	√	√	√				√		√	

### 3.2 目标跟踪的评价方法

#### 3.2.1 目标中心位置估计的评价方法

给定目标的跟踪边界框  $r_T$  和真实边界框  $r_G$ , 则中心位置误差  $E$  被定义为  $r_T$  中心和  $r_G$  中心之间的欧氏距离。

对于一个给定的误差阈值  $T$ , 若  $E \leq T$  则认为该帧目标被准确跟踪, 否则即认为跟踪失败。为了能够客观地衡量跟踪算法的性能, 本文提出使用一个针对具体视频序列自适应的阈值  $T$  设定方法: 由于在针对目标中心位置估计的实验分析过程中不对目标的形变进行估计, 即对跟踪目标的估计结果中目标的尺度不发生变化, 假定初始人工指定的目标边界框大小为  $w \times h$  像素, 那么可以将阈值  $T$  定义为目标边界框对角线长度的  $\theta$  倍 ( $\theta$  根据经验取值):

$$T = \theta \times \sqrt{w^2 + h^2} \quad (18)$$

#### 3.2.2 目标形变估计的评价方法

由于在目标跟踪过程中, 脱离目标的旋转信息来单独考虑目标尺度的变化没有意义, 因此目前没有分别提供目标尺度变换和旋转角度真实值的标准测试数据集。

以图2和图3为例进行说明。图2中: 图(a)为视频序列“FaceOcc2”的初始帧, 图(b)为对应的目标区域放大图。图3中: 图(a)为针对该视频序列在第358帧时的跟踪结果, 图(b)是对尺度和旋转角度同时估计的跟踪结果, 图(c)仅对尺度估计的跟踪结果。比较图3中(b)和(c)可见, 图(c)由于没有考虑目标的旋转导致对尺度的估计不准确, 可以看到图(c)将背景信息纳入了目标区域中(跟踪对象的耳部), 同时原本应有的目标区域也被认为是背景(跟踪对象的部分下颚)。

考虑到以上因素, 采用文献[7]提供的基于边界框重叠度的评价方法, 给定任一帧目标的跟踪边界框  $r_T$  和真实边界框  $r_G$ , 则重叠度的定义如式(19)所示:

$$OP = \frac{|r_T \cap r_G|}{|r_T \cup r_G|} \quad (19)$$

其中:  $|r_T \cap r_G|$  表示  $r_T$  和  $r_G$  的交集区域;  $|r_T \cup r_G|$  表示  $r_T$  和  $r_G$  的并集区域;  $|\cdot|$  表示区域内的像素数。

对于一个给定的阈值  $T_{op} \in (0, 1)$ , 若  $OP \geq T_{op}$  则认为该帧目标被准确跟踪, 否则即认为跟踪失败。

使用某一固定的阈值  $T_{op}$  进行测试所得跟踪准确率来衡量跟踪算法的性能是不客观的, 也不具有代表性。对此, 本文提出通过积分值图来对跟踪算法进行衡量, 即绘制跟踪准确率与阈值  $T_{op}$  取值关系曲线图, 计算跟踪准确率关于  $T_{op}$  的积分, 积分值越大表示总体的跟踪精度越高, 以此就可以作为衡量跟踪算法在形变估计上的客观标准, 根据经验确定阈值  $T_{op}$ 。



(a) 第1帧 (b) 目标区域放大图

Fig. 2 Target deformation estimation on the 1st frame



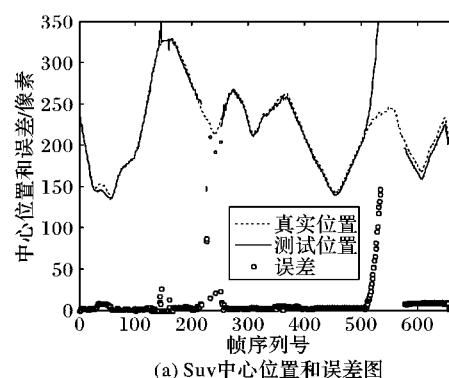
(a) 第358帧 (b) 进行尺度和角度估计的结果 (c) 进行尺度估计的结果

Fig. 3 Target deformation estimation on the 358th frame

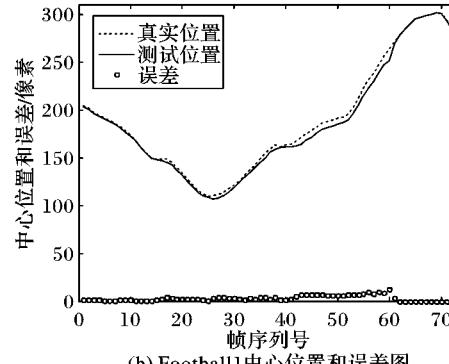
### 3.3 实验结果分析

#### 3.3.1 中心位置估计的实验分析

本节所用的视频序列及其所包含的干扰因素为表2的Suv和Football1。图4为针对Suv、Football1视频序列进行跟



(a) Suv中心位置和误差图



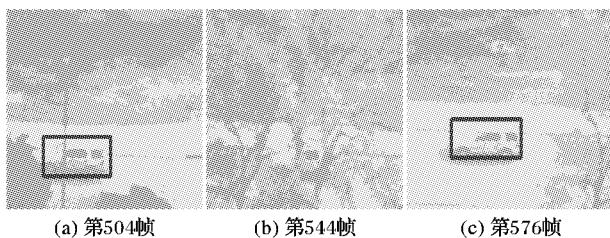
(b) Football1中心位置和误差图

Fig. 4 Position error curves of test video sequence



踪的位置误差曲线图。

对序列 Suv 的跟踪结果中第 504、544、576 帧图像如图 5 所示。在初始帧中将白色 SUV 作为跟踪目标, 在后续帧中该车与摄像头之间被树林所遮挡因此消失, 在穿过树林后, 目标即被迅速捕获。按照式(18)所述的阈值设定方法, 序列中目标位置估计准确率达到 0.93。因此针对目标消失再出现的干扰因素, 本文算法有很好的鲁棒性。

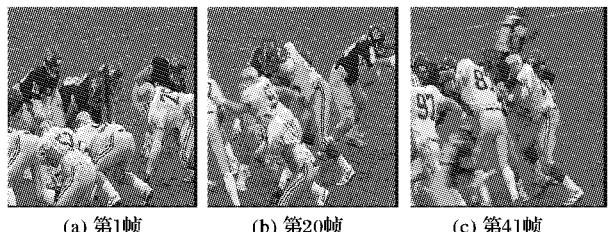


(a) 第504帧 (b) 第544帧 (c) 第576帧

图 5 对视频序列 Suv 的跟踪结果

Fig. 5 Tracking results of video sequence Suv

对序列 Football1 的跟踪结果中第 1、20、41 帧图像如图 6 所示, 视频序列中的主要干扰因素以相似目标干扰为主, 还存在严重的遮挡, 本文算法针对该序列的跟踪仍然比较准确, 目标中心位置估计的准确率为 1.00。



(a) 第1帧 (b) 第20帧 (c) 第41帧

图 6 对视频序列 Football1 的跟踪结果

Fig. 6 Tracking results of video sequence Football1

### 3.3.2 目标形变估计的实验分析

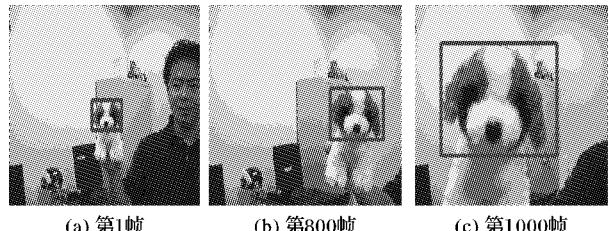
本节所用的视频序列及其所包含的干扰因素如表 2 中 Dog1 和 Singer1 序列所示。跟踪算法在各个视频序列上进行测试得到的积分值如表 3 所示。

表 3 在各个视频序列上进行测试得到的积分值 ( $T_{OP} \in [0, 0.5]$ )

Tab. 3 Integral values of test video sequences ( $T_{OP} \in [0, 0.5]$ )

视频序列	理想积分值	测试积分值
Dog1	1 000	978.711
Singer1	1 000	999.575

序列 Dog1 的跟踪结果的第 1、800、1000 帧图像如图 7 所示, 在初始帧中将男士手中的玩偶作为跟踪目标, 在跟踪过程中玩偶与摄像机的距离不断变动, 并不断地旋转。从跟踪结果可以看出, 本文算法可以实现非常准确的跟踪。



(a) 第1帧 (b) 第800帧 (c) 第1000帧

图 7 对视频序列 Dog1 的跟踪结果

Fig. 7 Tracking results of video sequence Dog1

序列 Singer1 的跟踪结果中第 1、175、350 帧图像如图 8 所示。在初始帧中将身着白色长裙的歌手作为目标进行跟踪, 由于镜头由近及远不断移动, 同时舞台灯光由暗转明, 又

由明转暗, 因此对于跟踪算法来说有两个非常明显的干扰因素: 目标尺度的变化和环境的光照变化。实验结果表明, 本文算法对目标的尺度估计都比较准确。



(a) 第1帧 (b) 第175帧 (c) 第350帧

图 8 对视频序列 Singer1 的跟踪结果

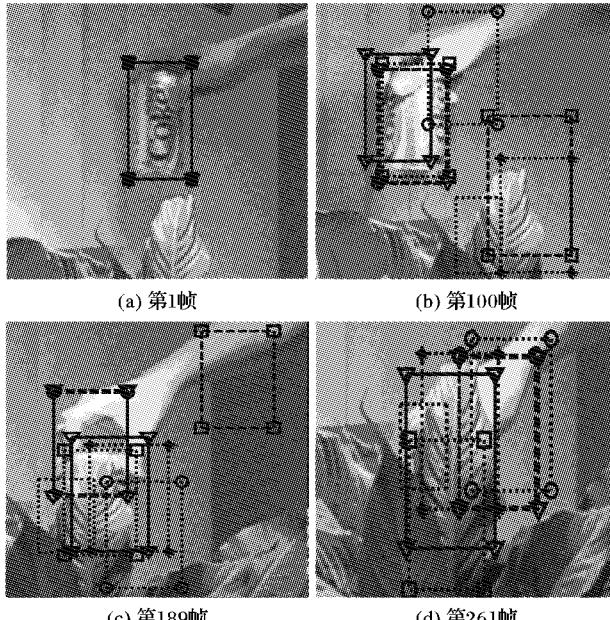
Fig. 8 Tracking results of video sequence Singer1

### 3.3.3 与其他算法对比实验

将本文的跟踪算法与目前跟踪效果最好的六种跟踪算法进行对比, 分别是: Struck ( Structured Output Tracking with Kernels)<sup>[8]</sup>、TLD、MOSSE ( Minimum Output Sum of Squared Error filter)<sup>[9]</sup>、CT (Compressive Tracking)<sup>[10]</sup>、STC 和 MIL。测试视频序列及其所包含的干扰因素如表 2 中的 Coke 和 Skating1 序列所示。在本节实验结果中, 各个跟踪算法在图像中对应的跟踪框如每幅图例所示, 为了呈现更清楚的实验结果, 所有跟踪结果采用局部裁剪之后图像显示。

针对 Coke 序列, 各个算法的跟踪结果中的第 1、100、189、261 帧图像如图 9 所示。在初始帧中将男士手中的易拉罐作为跟踪目标, 由于易拉罐在灯光下快速地运动, 在第 100 帧中目标被隐藏到叶子下方, 从跟踪开始时 TLD 就发生漂移, 到视频结束时只有本文算法依然能够锁定目标。

—▼ 本文算法    ⬤ CT    ⬤ MIL    ⬤ MOSSE  
—◆ STC    ⬤ Struck    ⬤ TLD    ⬤ Truth



(a) 第1帧 (b) 第100帧

(c) 第189帧 (d) 第261帧

图 9 对视频序列 Coke 的跟踪结果

Fig. 9 Tracking results of video sequence Coke

序列 Skating1 的跟踪结果中第 1、52、190、249 帧图像如图 10 所示。在初始帧中将身着蓝色上衣、白色短裙的女士作为跟踪目标, 由于女士滑冰的速度时快时慢, 顶棚的灯光色彩不断变化, 因此对跟踪的干扰很强; 从第 52 帧可以看到 CT、MIL 已经跟丢, 在视频序列的最后阶段只有本文算法能够持续准确地跟踪目标。

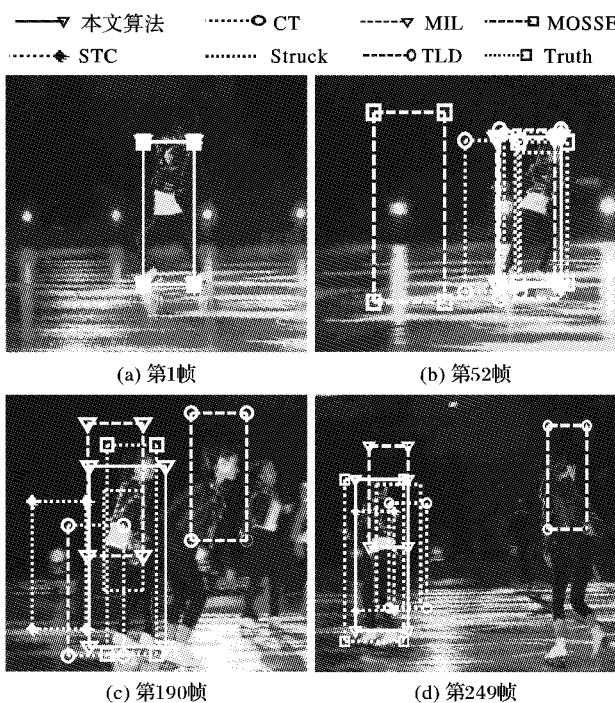


Fig. 10. Tracking results of video sequence Skating1

由实验可见,当跟踪序列中目标的帧间运动速度过快时,TLD 算法由于来不及获取足够的样本从而导致检测模块的分类器不够准确因而很快跟踪失效,由于其融合模块在将跟踪器和分类器的结果进行融合时,分类器的分类结果也占有投票权重,因此这时分类器不仅不能提高整体精度,反而拖累了跟踪算法的性能。但本文算法中,将静态外观模型与自适应模型相结合,能实现对目标的准确跟踪。

以上实验反映的是本文算法和其他目标跟踪算法在准确性、鲁棒性上的优势，其他跟踪算法采用原作者公开的源码；在实时性方面，本文算法也具有优势。表 4 是不同跟踪算法对 Coke 和 Skating1 视频序列（见表 2）进行跟踪时的处理速度对比。

表4 不同跟踪算法的处理速度比较

Tab. 4 Processing speeds comparison of different tracking algorithms

算法	平均每秒处理帧数	算法	平均每秒处理帧数
本文算法	26.47	CT 算法	45.33
Struck 算法	9.61	STC 算法	36.37
TLD 算法	18.16	MIL 算法	20.60
MOSSE 算法	39.81		

从表 4 中可以看到,在本文所用的计算机性能一般的情况下每秒平均能够处理 26 帧图像。由于人类眼睛的视觉暂留现象使得每秒 24 帧时的视频就已经满足实时性要求了,因此本文的实时性也是较好的。

4 结语

针对单目标长期跟踪,本文提出了一种基于静态-自适应外观模型纠正的目标跟踪算法,跟踪效果良好。取得的成果如下:针对不同类别的干扰因素分别提出了静态外观模型和自适应外观模型,能够抵御跟踪环境和目标本身的干扰因素;针对外观模型的噪声问题,将单连接层次聚类算法运用在主体跟踪算法的外观模型输出结果背景去噪过程中,有效地提

高了对目标跟踪的精度；最后，针对本文算法进行了多组实验分析，实验结果表明，在各种干扰因素存在的条件下，本文算法在长期跟踪过程中能够实现稳定鲁棒的跟踪。

## 参考文献 ( References )

- [1] SMEULDER A W, CHU D M, CUCCHIARA R, et al. Visual tracking: an experimental survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7): 1442 – 1468.
  - [2] BABENKO B, YANG M H, BELONGIE S. Robust object tracking with online multiple instance learning[ J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619 – 1632.
  - [3] KALAL Z, MIKOLAJCAYK K, MATAS J. Tracking-learning-detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409 – 22.
  - [4] ZHANG K, ZHANG L, YANG M H, et al. Fast tracking via spatio-temporal context learning[ EB/OL]. [2017-05-10]. <https://arxiv.org/abs/1311.1939>.
  - [5] DANELLIAN M, KHAN F, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking[C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 1090 – 1097.
  - [6] LUO J, GWUN O. A comparison of SIFT, PCA-SIFT and SURF [J]. International Journal of Image Processing, 2013, 3(4): 143 – 152.
  - [7] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C]// CVPR 2013: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2013: 2411 – 2418.
  - [8] HARE S, SAFFARI A, TORR P H S. Struck: structured output tracking with kernels[C]// Proceedings of the 2011 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2011: 263 – 270.
  - [9] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]// Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2010: 2544 – 2550.
  - [10] ZHANG K, ZHANG L, YANG M. Real-time compressive tracking [C]// ECCV 2012: Proceedings of the 12th European Conference on Computer Vision. Berlin: Springer, 2012: 864 – 877.
  - [11] KALAL Z, MIKOLAJCZYK K, MATAS J. Forward-backward error: automatic detection of tracking failures[C]// Proceedings of the 2010 20th International Conference on Pattern Recognition. Piscataway, NJ: IEEE, 2010: 2756 – 2759.
  - [12] SUPANCIC III, J S, RAMANAN D. Self-paced learning for long-term tracking[C]// CVPR 2013: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2013: 2379 – 2386.
  - [13] 朱明敏,胡茂海. 基于相关滤波器的长时视觉目标跟踪方法 [J]. 计算机应用, 2017, 37(5): 1466 – 1470. (ZHU M M, HU M H. Long-term visual object tracking algorithm based on correlation filter[J]. Journal of Computer Applications, 2017, 37 (5): 1466 – 1470.)

This work is partially supported by the Key Science and Technology Program of Guizhou Province (2017GZD60903), the Science and Technology Research Program of Xi'an City (2017086CG/RC049).

**WEI Baoguo**, born in 1970, Ph. D., associate professor. His research interests include image processing, pattern recognition, computer vision.

**GE Ping**, born in 1992, M. S. candidate. Her research interests include target detection and tracking.