

文章编号 1004-924X(2018)08-2122-10

## 基于场景辅助特征的 T-S 目标跟踪

宋 策\*, 张 葆, 宋玉龙, 钱 锋

(中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033)

**摘要:**针对遮挡同时目标附近出现相似目标干扰所导致的错跟问题,本文提出利用场景中辅助特征提升目标跟踪抗遮挡以及抗相似目标干扰性能。首先检测场景强特征及目标附近相似干扰,定义二者为场景辅助特征;其次,建立能够较好描述场景强特征及目标运动规律的动态模型以及相似干扰约束;最后,将场景辅助特征及目标的动态模型以粒子滤波的形式表达,提出 T-S 跟踪算法。采用 SPEVI 及 OTB100 数据库中若干典型测试视频,与近年来 6 种先进跟踪算法进行对比实验,并采用两种评价体系考量。实验结果表明,本文 T-S 算法对 SPEVI 多人脸、红外车辆的跟踪误差分别为 24 pixel 和 8 pixel;对 OTB100 数据库中 8 种视频跟踪测试时,在重叠率阈值为 0.5 时的跟踪成功率为 0.51,优于其它对比算法。本文 T-S 跟踪算法能够较好应对遮挡及相似目标干扰。

**关键词:**目标跟踪; 动态模型; 粒子滤波; 辅助特征

**中图分类号:**TP391.41 **文献标识码:**A **doi:**10.3788/OPE.20182608.2122

## T-S tracking algorithm based on context auxiliary feature

SONG Ce\*, ZHANG Bao, SONG Yu-long, QIAN Feng

(Changchun Institute of Optics, Fine Mechanics and Physics,  
Chinese Academy of Sciences, Changchun 130033, China)

\* Corresponding author, E-mail: songce528@163.com

**Abstract:** Occlusion and objects with the same appearance as the target (known as distractors) are extremely challenging in the tracking domain, and distractors appearing around the target during occlusion tend to cause tracking distractors. To resolve this problem, improving the tracking performance under conditions of serious occlusions and distractor by exploiting the context auxiliary feature was proposed. First, the context strength feature and distractors around the target were detected. Second, a dynamic model that can describe the movements of context strength feature and target well and the constraint of distractor were built. Finally, the dynamic model of context strength feature and target were described in particle filter, and the T-S tracking algorithm was proposed. Using a challenging test video from the SPEVI and OTB100 datasets, the proposed algorithm was compared with other six highly ranked tracking algorithms. Two types of evaluation were adopted during testing. The experimental results demonstrate that the error pixel is 24 pixel and 8 pixel when T-S algorithm tracking multifaces and infrared car from SPEVI datasheet, when tracking eight testing videos from OTB100 dataset, the success rate of T-S algorithm is 0.51 when overlap threshold is 0.5 and surpass other

收稿日期:2018-05-02;修订日期:2018-07-06.

基金项目:国家自然科学基金资助项目(No. 61705225)

compared algorithm. Our proposed T-S algorithm performs well when tracking targets under conditions of serious occlusions and distractors.

**Key words:** target tracking; dynamic model; particle filter; auxiliary feature

## 1 引 言

目标跟踪是计算机视觉领域一项重要技术,在智能监控、国防侦察等领域有着重要的应用。尽管国内外学者进行了数十年的研究,目标跟踪尚有较多难以解决的关键问题,如跟踪过程目标模板退化、目标表观特征剧烈变化(如因光照、视角变化)导致目标丢失、复杂背景干扰、相似目标干扰、遮挡等。

为解决上述难题,学者从目标表观特征出发,试图建立包含目标多种特征的表观模型<sup>[1-2]</sup>,并采用主成分分析、稀疏表示<sup>[3-5]</sup>等算法将多种混合特征压缩描述,此类模型被称为“生成”模型。该类模型通常计算量较大却效果有限。2009年,KA-LAL 等人首次将半监督学习引入到目标跟踪领域,提出 TLD 跟踪算法<sup>[6]</sup>,该算法的最大贡献在于将分类器引入跟踪领域,分类器并非仅对目标特征进行描述,而是对目标与其周围背景的特征差异进行建模,通常将该模型称为“差异”模型。这种模型以显著的跟踪效果引起众多学者研究,一系列基于差异模型的先进跟踪算法被快速提出<sup>[7-10]</sup>。差异模型的出现标志着学者从单纯研究目标特征,到开始研究目标与周围背景的特征差异,但这种研究也仅限于目标与其附近背景。

场景中通常包含较为丰富的特征,若能有效利用场景特征,则可显著提高目标跟踪性能。2011年,DINH T. B. 提出一种基于场景辅助的目标跟踪算法<sup>[11]</sup>,该算法在目标周围建立众多“支持”特征,当场景中出现与目标表观特征相似的干扰目标时,可依赖“支持”特征区分真实目标与干扰目标。但该算法有两处不足:首先,为简化“支持”特征与目标间的动态关系,仅选择目标附近特征,但目标周边若无有效特征,则该算法失效;其次,若目标发生遮挡,则目标周边的“支持”特征很大概率同样发生遮挡,此时无法预测目标位置。2016年,XIAO J. J. 提出一种基于场景辅助特征的目标跟踪算法<sup>[12]</sup>,该算法采用相似目标作为“支持”特征,建立“支持”特征与目标的动态

模型,但该算法的“支持”特征依靠分类器定位,定位精度较差,且其动态模型过于简单,“支持”特征较差的定位精度与简单的动态模型致使当目标周围出现相似目标时,难以准确排除干扰目标。

本文通过挖掘场景中的辅助特征来提升目标跟踪性能。首先,检测全局场景中稳定、显著的强特征以及与目标表观特征相似的干扰目标,将二者定义为场景辅助特征;其次,建立场景辅助特征与目标之间准确的动态模型;最后,将目标与场景辅助特征的动态模型以粒子滤波的形式表达,提出 T-S 跟踪算法。该算法能够较好应对遮挡及相似目标的干扰。

## 2 场景辅助特征

在目标跟踪过程中,场景中的信息包含被跟踪目标与目标周围的背景,而背景中很可能出现与目标表观特征较为相似的特征,比如在跟踪车辆过程中,被跟踪车辆周围的背景中很可能出现同型号的车辆。该类特征对目标的跟踪通常具有较大干扰,很可能导致错跟目标,本文将背景中与目标表观特征相似的区域定义为相似干扰特征。背景中除了相似干扰特征,通常还包含较为丰富的显著特征,本文将其定义为场景强特征。

### 2.1 相似干扰

相似干扰是指表观特征与被跟踪目标相似度极高的区域,通常与目标属同型物体。如图 1 所示(彩图见期刊电子版),其中红色椭圆框内为跟踪目标,而蓝色椭圆框内是与目标表观特征相似的相似干扰。相似干扰在接近目标时极易对跟踪产生干扰,尤其当目标被遮挡时极易错跟相似干扰,因此必须将相似干扰与目标区分开来。

若依靠二者的表观特征差异来区分,则需建立分类能力极强的分类器,该分类器需能够区分目标与相似干扰表观特征之间微弱的差别,这一方面需要分类器性能优异,另一方面需要获得目标、相似干扰二者大量的训练样本,而这种强大区分能力的缺点通常是计算量极大。然而,在目标跟踪领域除目标表观特征之外,还可以利用目标

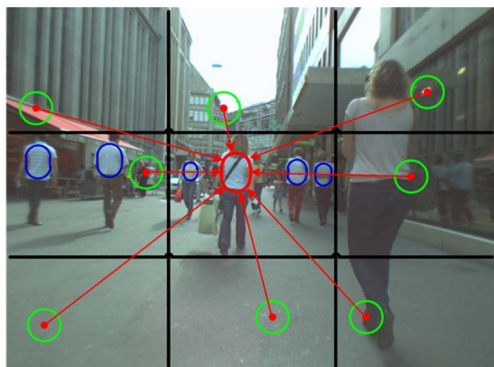


图 1 基于场景辅助特征的目标跟踪

Fig. 1 Target tracking by context auxiliary feature

的运动特征。目标与相似干扰之间虽表现特征相似,但运动特征必然不同,因此本文重点讨论如何依据二者的运动特征差异将其区分。

对相似干扰的检测采用分类器,该分类器不需区分出目标与相似干扰的微弱差异,只需将场景中相似干扰成功检测出即可,因此对该分类器并不需要过于复杂的设计,本文采用随机森林分类器<sup>[12]</sup>。

## 2.2 场景强特征

场景中除目标与相似干扰外,其它成分皆为背景,本文将背景中特征较为显著的区域定义为“场景强特征”,显著特征通常较为稳定,因此场景强特征是一种可靠的参考。若能建立场景强特征与目标之间的动态模型,则可利用稳定的场景强特征辅助提升对目标的跟踪性能,当目标被遮挡,可依据稳定的场景强特征对目标轨迹进行预测,此时即使场景中出现相似干扰亦不会发生错跟。

场景强特征采用快速海森检测器(Fast Hessian Detector)检测场景中 SURF 特征<sup>[13]</sup>,为保证检测出的场景特征代表全局而非某一局部,本文将全部场景分为 9 个区域,每个区域分别检测出特征最强位置作为该区域的场景强特征,如图 1 中绿色圆内区域所示。各区域的场景强特征依据特征强弱进行排序,最强的场景强特征通常最为稳定。

## 3 场景辅助特征与目标的动态模型

### 3.1 场景强特征与目标的动态模型

当检测到场景辅助特征,若想利用该特征提升对目标的跟踪性能,在目标被遮挡时能较好预

测目标位置并排除相似干扰,需建立场景辅助特征与目标间准确的运动模型。目标、场景强特征在图像中既存在绝对运动,又存在相对运动。如图 1 所示,绿色框内各区场景强特征的局部运动共同构成场景的全局运动,同时各区场景强特征间又存在相对运动。

场景强特征运动模型如下:

$$p_{s,k}^i = \bar{p}_{s,k} + \Delta p_{s,k}^i, \quad (1)$$

其中:  $p_{s,k}^i$  表示第  $k$  帧第  $i$  区场景强特征的绝对位置,  $\Delta p_{s,k}^i$  表示第  $k$  帧第  $i$  区场景强特征的相对位置,  $\bar{p}_{s,k}$  表示第  $k$  帧场景全局位置:

$$\bar{p}_{s,k} = \frac{1}{N_{s,k}} \sum_{i=0}^{N_{s,k}} p_{s,k}^i, \quad (2)$$

其中  $N_{s,k}$  表示场景辅助特征数。

定义目标第  $k$  帧绝对位置为  $p_{t,k}$ , 相对位置为  $\Delta p_{t,k}$ , 则满足:

$$p_{t,k} = \bar{p}_{s,k} + \Delta p_{t,k}. \quad (3)$$

在已知场景全局位置  $\bar{p}_{s,k}$  的情况下,若知道目标的相对位置  $\Delta p_{t,k}$ , 即可获得目标在图像中的绝对位置  $p_{t,k}$ 。

本文目标的相对运动模型采用线性回归模型<sup>[14]</sup>:

$$\Delta p_{t,k} = \beta_{c,0} + \beta_{c,1} k + \epsilon_{c,t}, \quad (4)$$

其中  $\Delta p_{t,k}$  表示目标的相对位置,  $\beta_{c,0}$ ,  $\beta_{c,1}$  为线性回归模型的系数,  $\epsilon_{c,t}$  为高斯白噪声。根据过去  $k-1$  帧图像序列中目标对于全局场景的相对位置,采用最小均方误差可估计出模型系数  $\hat{\beta}_{c,0}$ ,  $\hat{\beta}_{c,1}$ , 因此目标在第  $k$  帧的相对运动方程为:

$$\Delta \hat{p}_{t,k} = \hat{\beta}_{c,0} + \hat{\beta}_{c,1} k. \quad (5)$$

目标的绝对运动模型、场景各区强特征的相对运动模型、绝对运动模型均采用线性回归模型,即:

$$p_{t,k} = \beta_{a,0} + \beta_{a,1} k + \epsilon_{a,t}, \quad (6)$$

$$p_{s,i,k} = \beta_{a,s,0} + \beta_{a,s,1} k + \epsilon_{a,s,i}, \quad (7)$$

$$\Delta p_{s,i,k} = \beta_{c,s,0} + \beta_{c,s,1} k + \epsilon_{c,s,i}, \quad (8)$$

其中:  $\beta_{a,0}$ ,  $\beta_{a,1}$  表示目标绝对运动模型参数,  $\epsilon_{a,t}$  表示目标绝对运动模型噪声;  $p_{s,i,k}$  表示第  $k$  帧第  $i$  区场景强特征绝对运动位置,  $\beta_{a,s,0}$ ,  $\beta_{a,s,1}$  表示第  $i$  区绝对运动模型参数,  $\epsilon_{a,s,i}$  表示第  $i$  区场景强特征绝对运动模型噪声;  $\Delta p_{s,i,k}$  表示第  $k$  帧第  $i$  区场景强特征相对运动位置,  $\beta_{c,s,0}$ ,  $\beta_{c,s,1}$  表示第  $i$  区相对运动模型参数,  $\epsilon_{c,s,i}$  表示第  $i$  区场景强特征相对运动模型噪声。式(6)~式(8)相应参数估计方法与

式(4)参数估计方法相同。

需要指出的是绝对运动模型表示目标、场景强特征在图像中的运动轨迹,考虑成像过程中相机可能因无规则抖动造成目标、场景强特征在图像中的运动轨迹较为凌乱,而相机无规则抖动只能建模为噪声,因此目标、场景强特征的绝对运动模型的噪声  $\epsilon_{a,t}$ 、 $\epsilon_{a,s,i}$  必然较大;而相对运动模型仅考虑目标、各区域场景强特征对于全局场景的相对运动轨迹,所反应的是真实世界中的运动规律,采用线性回归模型必然较为准确,因此目标、场景强特征相对运动模型的噪声  $\epsilon_{c,t}$ 、 $\epsilon_{c,s,i}$  必然较小。

当目标被遮挡时,可首先根据场景强特征计算出场景全局运动位置  $\bar{p}_{s,k}$ ,根据式(3)和式(4)对目标进行轨迹预测。即使此时视场内出现相似干扰,本文算法亦不会出现错跟。

### 3.2 相似干扰的约束

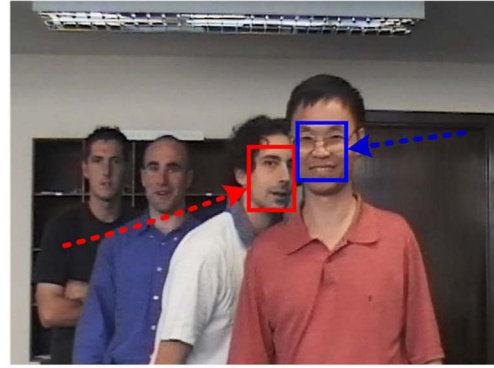
相似干扰的检测依靠随机蕨类分类器,一旦检测出相似干扰,即对它进行持续跟踪和运动建模,建立轨迹信息。对相似干扰的跟踪基于匹配算法,具体采用归一化互相关(Normalization Cross-Correlation, NCC)。跟踪过程采用 NCC 匹配而非分类器,是因为本文随机蕨类分类器被设计为一种宽松的检测算法,可快速检测出相似干扰却不能精确定位,而 NCC 匹配则可精确获取相似干扰位置,精确的位置配合准确的动态模型可对相似干扰轨迹进行准确预测。同理,对目标以及场景强特征的跟踪同样基于 NCC 匹配,匹配特征为 HOG 特征<sup>[15]</sup>。

为保证计算效率,仅在目标周围的有限范围内检测相似干扰,若与目标相距较远,则无法形成干扰。若相似干扰与目标绝对位置差异较大,则依靠相似干扰与目标动态模型的差异,即可区分二者;但在某些极端情况,某一时刻相似干扰与目标的绝对位置重合时,二者互相遮挡,如图 2(a)所示,此时跟踪极易出现错跟。

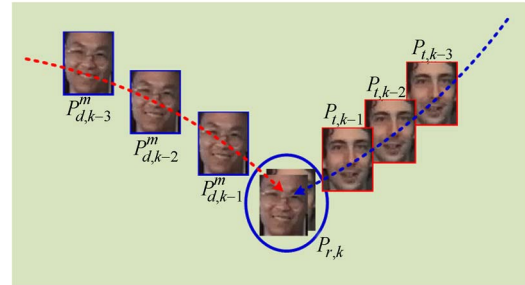
本文定义相似干扰约束能量函数  $E_{r,k}$  如下:

$$E_{r,k} = \begin{cases} E_{d,k} = L_{d,k}^m \cdot e^{-|p_{r,k} - \hat{p}_{d,k}^m|}, p_{r,k} = p_{d,k}^m \\ E_{t,k} = L_{t,k} \cdot e^{-|p_{r,k} - \hat{p}_{t,k}|}, p_{r,k} = p_{t,k} \end{cases}, \quad (9)$$

其中:  $p_{r,k}$  表示第  $k$  帧相似干扰与目标重叠区的绝对位置;  $p_{t,k}$  表示第  $k$  帧目标的绝对位置;  $p_{d,k}^m$  表示第  $k$  帧、第  $m$  个相似干扰的绝对位置,与目标绝对位置不构成重叠的相似干扰不必考虑;  $L_{d,k}^m$  表



(a) 目标与相似干扰  
(a) Target and distractor



(b) 目标与相似干扰的轨迹  
(b) Trajectory of target and distractor

图 2 目标与相似干扰相互遮挡

Fig. 2 Occlusion between target and distractor

示相似干扰模板与重叠区  $p_{r,k}$  表观特征进行 NCC 匹配获得的相似度;  $L_{t,k}$  表示目标相似度;  $\hat{p}_{t,k}$  为根据式(6)得到的目标第  $k$  帧绝对位置估计;  $\hat{p}_{d,k}^m$  表示第  $k$  帧相似干扰的绝对位置估计。

相似干扰约束能量函数  $E_{r,k}$  与目标及相似干扰的表观特征及运动特征相关,如图 2(b)所示,若相似干扰被遮挡,重叠区  $p_{r,k}$  表示目标绝对位置,则该区域表观特征与目标模板 NCC 匹配得到的相似度  $L_{t,k}$  大于  $L_{d,k}^m$ ,  $p_{r,k}$  与目标绝对位置估计  $\hat{p}_{t,k}$  的距离会更小,根据式(9)可知  $E_{t,k}$  大于  $E_{d,k}$ ;若目标被遮挡,重叠区  $p_{r,k}$  为相似干扰的绝对位置,则该区域表观特征与相似干扰模板 NCC 匹配所得的相似度  $L_{d,k}^m$  大于  $L_{t,k}$ ,  $p_{r,k}$  与相似干扰绝对位置估计  $\hat{p}_{d,k}^m$  会更小,根据式(9)可知  $E_{d,k} > E_{t,k}$ 。因此,根据最大化重叠能量函数  $E_{r,k}$  即可判定图 2(b)所示重叠区  $p_{r,k}$  是目标或者相似干扰,即:

$$p_{r,k} = \arg \max E_{r,k}. \quad (10)$$

## 4 T-S 粒子滤波跟踪算法

在众多跟踪算法中,粒子滤波能够同时描述目标的动态特征与观测特征,因此本文选择粒子滤波作为算法框架。基于粒子滤波的传统跟踪算法仅对单目标进行建模,而本文算法试图建立场景强特征、目标、相似干扰之间的动态关系,通过场景强特征、相似干扰约束提升相似干扰及遮挡时目标的跟踪性能。本文在传统粒子滤波基础上进行改进,提出了粒子滤波对目标及场景同时建模的 T-S(Target-Scene)粒子滤波。

### 4.1 粒子滤波

粒子滤波基于递归贝叶斯滤波理论<sup>[16]</sup>,主要有预测与更新两个过程,其方程表达如下:

$$p(x_k | y_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | y_{1:k-1}) dx_{k-1}, \quad (11)$$

$$p(x_k | y_{1:k}) \propto p(y_k | x_k) p(x_k | y_{1:k-1}), \quad (12)$$

其中:  $x_{k-1}$  表示  $k-1$  时刻目标的运动位置,  $y_{k-1}$  表示目标  $k-1$  时刻的观测值,  $y_{1:k-1}$  表示目标自初始时刻到  $k-1$  时刻的全部观测值,  $p(x_{k-1} | y_{1:k-1})$  为目标位置  $x_{k-1}$  的后验概率密度函数,  $p(x_k | x_{k-1})$  为运动模型的表达式,称为状态转移方程,  $p(y_k | x_k)$  表示观测模型。

所谓粒子滤波即是上述概率密度函数用带权重的有限粒子进行蒙特卡洛数值近似:

$$p(x_{k-1} | y_{1:k-1}) \approx \{x_{k-1}^{(i)}, \omega_{k-1}^{(i)}\}_{i=1}^N, \quad (13)$$

其中  $N$  表示粒子数,  $x_{k-1}^{(i)}, \omega_{k-1}^{(i)}$  分别表示  $k-1$  时刻第  $i$  个粒子的运动位置和粒子权重。跟踪过程中首先根据  $k-1$  时刻的后验概率密度函数  $\{x_{k-1}^{(i)}, \omega_{k-1}^{(i)}\}_{i=1}^N$  进行重采样,经过状态转移、权重更新过程获得  $k$  时刻的后验概率密度函数  $\{x_k^{(i)}, \omega_k^{(i)}\}_{i=1}^N$ , 进而确定目标的最优跟踪位置  $\hat{x}_k$ :

$$\hat{x}_k = \sum_{i=1}^N x_k^{(i)} \omega_k^{(i)}. \quad (14)$$

### 4.2 改进的 T-S 粒子滤波

本文粒子需要对目标、相似干扰、场景强特征的动态模型分别描述,上述三类粒子  $k$  时刻分别表示为  $\{x_{t,k}^{(j)}, \omega_{t,k}^{(j)}\}_{j=1}^{N_t}$ 、 $\{x_{d,i,k}^{(j)}, \omega_{d,i,k}^{(j)}\}_{j=1}^{N_{d,i}}$ 、 $\{x_{s,i,k}^{(j)}, \omega_{s,i,k}^{(j)}\}_{j=1}^{N_{s,i}}$ , 目标粒子的运动矢量表示如下:

$$\mathbf{X}_{t,k}^{(j)} = [p_{t,k}, \beta_{a,\theta}, \beta_{a,\phi}, \Delta p_{t,k}, \beta_{c,\theta}, \beta_{c,\phi}]^T.$$

其中  $p_{t,k}, \beta_{a,\theta}, \beta_{a,\phi}$  表示目标的绝对动态模型参

数,状态转移方程见式(6),  $\Delta p_{t,k}, \beta_{c,\theta}, \beta_{c,\phi}$  表示目标相对动态模型参数,状态转移方程为式(5)。相似干扰的粒子状态矢量表达式以及状态转移方程与目标相同;背景强特征粒子的状态矢量表达如下:

$$\mathbf{X}_{s,i,k}^{(j)} = [p_{s,i,k}^{(j)}, \beta_{a,s,\theta}^{(j)}, \beta_{a,s,\phi}^{(j)}, \Delta p_{s,i,k}^{(j)}, \beta_{c,s,\theta}^{(j)}, \beta_{c,s,\phi}^{(j)}]^T.$$

背景强特征粒子绝对状态转移方程见式(7),相对状态转移方程见式(8)。

根据式(3)可知,粒子状态转移过程采用的绝对动态模型与相对动态模型是等价的。由第 3 章分析可知绝对运动模型噪声  $\epsilon_{a,t}, \epsilon_{a,s,i}$  较大,相对运动模型噪声  $\epsilon_{c,t}, \epsilon_{c,s,i}$  较小,因此对目标、相似干扰、场景强特征状态转移时,采用绝对运动模型比相对运动模型需要更多粒子进行数值模拟。从算法的实时性角度考虑,本文趋向于采用粒子相对状态进行状态转移,但前提是当前场景需处于稳定状态。

在跟踪过程中,首先对最强场景强特征第  $i$  区采用绝对动态模型进行状态转移,后经观测模型进行粒子权重更新后,根据式(14)得到第  $i$  区域时刻强特征精确位置  $p_{s,i,k}$ 。此时定义场景处于稳定状态,场景全局运动位置估计为:

$$\bar{p}_{s,k} = p_{s,i,k}. \quad (15)$$

此后对其它区场景强特征状态转移时即可采用相对运动模型,后经观测模型更新粒子权重,根据式(14)得到各区场景强特征精确相对位置  $\Delta \hat{p}_{s,i,k}$ ,进而获得绝对位置  $p_{s,i,k}$ ,根据式(2)计算可得场景全局运动矢量  $\bar{p}_{s,k}$ 。针对目标采用相对动态模型进行状态转移,经过观测模型更新目标粒子权重后即可获得目标的精确位置  $p_{t,k}$ 。

本文所提算法与多目标跟踪较为类似却有本质差异,多目标跟踪需初始化多个目标,而本文跟踪仅初始化一个目标。场景强特征及相似干扰皆为图像中实时更新的特征,其目的是辅助提升目标跟踪的抗遮挡、抗相似干扰性能。

本文算法流程如下:

#### (1) 初始化

采用快速海森检测器检测场景强特征,并对特征强度进行排序,手动选择跟踪目标,初始化目标、场景强特征粒子状态矢量及粒子权重;

#### (2) 稳定场景

根据式(7)对最显著场景强特征进行状态转

移,并采用观测模型(NCC 匹配)计算状态转移后粒子权重,根据式(14)和式(15)得到场景全局动态矢量估计  $\bar{p}_{s,k}$ ,标记场景稳定;

(3)计算场景全局运动位置

根据式(8)对其它区域的场景强特征进行状态转移,并采用观测模型(NCC 匹配)计算状态转移后的粒子权重,最终根据式(14)得到全部区域的场景强特征绝对位置  $p_{s,k}^i$ ,根据式(2)更新场景全局运动位置  $\bar{p}_{s,k}$ ;

(4)目标定位

根据式(5)对目标粒子进行状态转移,根据观测模型(NCC 匹配)计算状态转移后的粒子权重,根据式(14)计算目标的绝对位置  $p_{t,k}$ ;

(5)状态更新

对目标、场景强特征区域进行粒子重采样;更新绝对动态模型、相对动态模型参数,即式(4)、式(6)~式(8)中的模型参数;

(6)模板更新

采用快速海森检测器检测全局特征,剔除各区域弱特征。运用随机蕨类分类器检测目标附近区域,若检测到相似干扰,则分配粒子对其进行运动建模;

(7)返回步骤(2)进行下一帧跟踪。

## 5 实验与分析

本文实验在 Inter Core i5-7200U 的 CPU、8G 内存配置电脑上使用 Visual Studio 2010 编程实现。本文相似干扰检测选用随机蕨类分类器,蕨类个数为 10,每个蕨类选择 4 个 6-bitBP 特征。粒子滤波的绝对动态模型数值模拟的粒子数为 300,相对动态模型数值模拟的粒子数为 200,粒子滤波的观测模型所选特征为 HOG 特征<sup>[15]</sup>。本文算法的计算量与场景强特征数量关系较大,实验表明,图 1 所示的 9 个强特征即可准确描述场景的运动规律。

对比算法选择近年来跟踪效果较为优秀的 6 种算法: VTD<sup>[1]</sup>, Struck<sup>[9]</sup>, HCF<sup>[17]</sup>, KCF<sup>[15]</sup>, DT<sup>[12]</sup>, PF<sup>[18]</sup>。其中,PF 是经典的基于颜色特征的粒子滤波跟踪算法,VTD 算法同样基于粒子滤波且选取复合特征、复合运动模型,Struck 为基于分类器的算法,KCF 是基于相关滤波的跟踪算

法,HCF 算法同样基于相关滤波且采用 CNN 特征,DT 算法与本文思想最为接近,同样通过挖掘背景辅助信息提高跟踪性能。上述 6 种算法基于 4 种模型。

本文所提 T-S 跟踪算法对遮挡及相似目标干扰有特殊的应对机制,因此本文实验分为两类:其一主要对目标被遮挡或目标周围具有相似目标干扰的视频进行跟踪;其二采用 8 种来自 OTB100<sup>[19]</sup>数据库的视频,该实验主要验证本文 T-S 算法应对跟踪领域其它常见干扰时的跟踪性能。在已知测试视频每帧图像目标真实坐标的情况下,本文采用目标跟踪平均误差<sup>[1,7]</sup>及目标真实位置重叠率曲线<sup>[12]</sup>两种评价体系对各类算法进行评价。

### 5.1 遮挡及相似干扰实验

此部分实验选择来自 SPEVI 数据库多人脸视频以及红外车辆视频,多人脸视频的跟踪难点在于视频中有 4 个外观特征极为相似的人脸,其余 3 个相似干扰人脸会对主目标形成较强干扰,且在跟踪过程相似干扰会对目标进行遮挡,常规算法极易在目标被遮挡时错跟相似干扰,进而跟踪失败。而本文 T-S 算法在跟踪过程中,可逐渐识别出相似干扰并分别进行动态建模,如图 3 所示(彩图见期刊电子版),蓝色标记为相似干扰,红色标记为目标。本文对相似干扰具有特殊约束,因此针对相似干扰对目标的遮挡,一方面可准确识别干扰而不会错跟,另一方面可对被遮挡目标进行轨迹预判,当目标重新出现在视场中时,即刻进行重新捕获,如图 3(c)所示。该视频难度较大,对比算法在视频中皆会丢失目标,本文算法可顺利完成对目标的跟踪,目标跟踪的初始尺寸为 50 pixel×80 pixel,本文 T-S 算法的跟踪平均误差为 24 pixel。

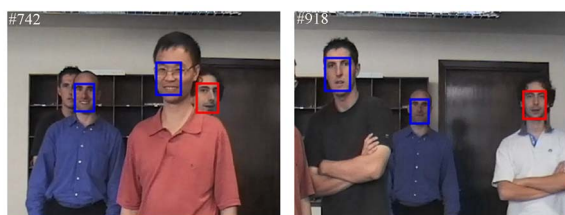


(a)第 25 帧

(a)25th

(b)第 72 帧

(b)72th



(c)第 742 帧 (d)第 918 帧  
(c)742th (d)918th

图 3 多人脸跟踪结果

Fig. 3 Tracking results of multifaces

该部分另一个视频为红外车辆跟踪视频,视频中目标有较长一段时间被树木遮挡,且目标后方即有相似目标跟随,如图 4(b)所示。对比算法在目标被树木遮挡时容易错跟相似干扰为真实目标。而本文算法在目标被遮挡时仍可依靠场景强特征的全局运动矢量预测目标的真实位置,如图 4(b)所示。在目标转弯、相似干扰逼近的过程中,本文算法皆可准确跟踪目标,如图 4(c)和 4(d)所示。对比算法中仅 DT 算法完成稳定跟踪而未丢失目标,目标初始跟踪尺寸为  $40 \text{ pixel} \times 40 \text{ pixel}$ ,本文 T-S 算法的跟踪平均误差为  $8 \text{ pixel}$ ,略优于 DT 算法。



(a)第 239 帧 (b)第 285 帧  
(a)239th (b)285th



(c)第 752 帧 (d)第 934 帧  
(c)752th (d)934th

图 4 红外车辆跟踪结果

Fig. 4 Infrared tracking results of cars

上述两种视频难度较大,目标被反复遮挡,且遮挡同时目标周围常有相似干扰,帧数较长,常规

算法通常在视频中间即丢失目标,表 1 所示为不同算法对两种视频的跟踪目标中心点平均误差,不难看出本文 T-S 算法效果较好。

表 1 跟踪平均误差

Tab. 1 Average center location error (pixel)		
跟踪算法	多人脸(1006 帧)	红外车辆(1594 帧)
PF	lost@62	lost@26
VTD	lost@65	lost@27
Struck standard	lost@65	lost@282
KCF conic	lost@735	lost@283
HCF conic	lost@736	lost@283
DT zernike	lost@97	11
本文 T-S standard	24	8

## 5.2 全面性能测试实验

本文 T-S 算法对遮挡及相似干扰具有针对性的动态建模及约束,实验证明 T-S 算法的模型及约束应对遮挡及相似干扰是有效的,但若该算法应对跟踪领域其它难点时效果不佳,显然难以应用。为进一步验证 T-S 算法的跟踪性能,本文选择 OTB100 数据库中若干典型视频进行算法对比测试。其中,shaking、singer 视频测试算法应对复杂背景对跟踪的干扰,sylvester、david 视频测试目标旋转及视角变化时的跟踪性能,motocross、soccer 视频测试目标突变运动、短时遮挡对跟踪的影响,animal、jumping 测试运动模糊对跟踪的影响。

图 5 为对比测试实验图,不难看出本文 T-S 算法效果较好,HCF 及 DT 算法仅次于本文 T-S 算法。HCF 算法采用特征极为丰富的 CNN 特征,本文算法采用的 HOG 特征描述能力虽比 CNN 特征稍弱,但 HCF 算法仅在目标周围进行特征采样,而本文算法实时检测全局场景显著特征,充分挖掘场景特征来辅助跟踪目标,即使在目标特征非常贫乏时,场景之中依然存在丰富的场景强特征。依靠场景强特征的支持,T-S 算法可有效应对复杂背景干扰、旋转、视角变化、突变运动以及运动模糊对跟踪的干扰。DT 算法虽与本文算法思想类似,试图利用相似目标提高真实目标的抗干扰能力,但其场景特征挖掘能力及场景特征与目标间的动态建模皆过于简单,因此跟踪效果稍差。

图 6 为不同算法针对上述 8 种测试视频的跟踪重叠率曲线<sup>[12]</sup>,曲线显示在不同重叠率阈值下

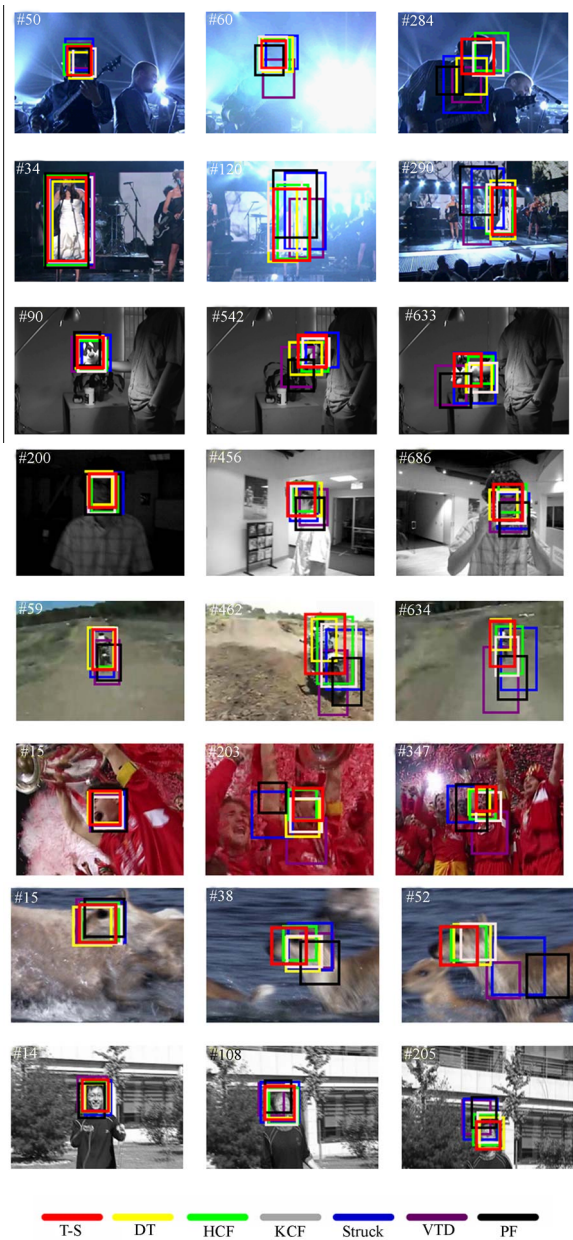


图 5 跟踪对比测试结果

Fig. 5 Comparison of tracking results between proposed and other algorithms

的跟踪成功率,由图 6 不难看出本文 T-S 算法的跟踪成功率优于其它算法。当重叠率阈值为 0.5 时,T-S 算法的跟踪成功率为 0.51,略优于 HCF (0.455)、DT(0.478)等算法。

参考文献:

[1] JUNSEOK K, LEE K M. Tracking by sampling and

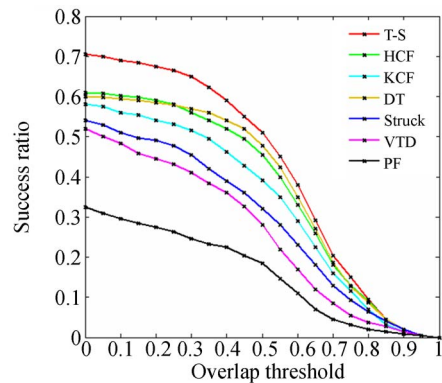


图 6 跟踪重叠率曲线

Fig. 6 Trade-off overlap rate curves

6 结 论

本文试图解决遮挡及相似目标干扰对目标跟踪的影响。传统跟踪算法仅关注目标本身特征,或者目标与周边背景的差异特征,而本文认为全局场景之中存在更为丰富的特征,且场景特征与目标存在动态关联,有效利用场景特征可提升目标跟踪性能。本文首先在场景中检测场景强特征,以及与目标表观特征相似的干扰特征,定义二者为场景辅助特征;其次建立场景辅助特征与目标的动态模型;最后,将场景辅助特征与目标的动态模型以粒子滤波形式表达,提出本文的 T-S 跟踪算法。本文通过两部分实验对 T-S 算法进行测试,首先选择目标被反复遮挡且相似目标干扰严重的两组视频验证 T-S 算法抗遮挡、抗相似目标干扰能力;其次,采用 OTB100 数据库视频测试 T-S 算法应对跟踪领域其它类型干扰的能力。实验结果表明,本文 T-S 算法对 SPEVI 多人脸、红外车辆的跟踪误差分别为 24 pixel 和 8 pixel;对 OTB100 数据库中 8 种视频跟踪测试时,在重叠率阈值为 0.5 时的跟踪成功率为 0.51,优于其它对比算法。T-S 算法的跟踪性能优异,应对遮挡及相似目标干扰时效果尤为显著,基于场景辅助特征提高目标跟踪性能是可行的。

integrating multiple trackers[J]. *IEEE Trans Pattern Anal Mach Intell*, 2014,36(7): 1428 - 1441.

[2] ADAM A, RIVLIN E, SHIMSHONI I. Robust

- fragments based tracking using the integral histogram[C]. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006 (1):798-805.
- [3] MEI X, LING H. Robust visual tracking using L1 minimization[C]. *IEEE 12th International Conference on Computer Vision*, 2009:1436-1443.
- [4] 毛建森, 屈玉福. 基于压缩感知的变尺度目标跟踪技术[J]. *液晶与显示*, 2016, 31(5): 497-505.
- MAO J S, QU Y F. Tracking of variable scale object based on compressive sensing [J]. *Chinese Journal of Liquid Crystals and Displays*, 2016,31 (5):497-505. (in Chinese)
- [5] 隋中山, 李俊山, 张姣, 等. 张量低秩表示和时空稀疏分解的视频前景检测[J]. *光学精密工程*, 2017, 25(2):529-536.
- SUI ZH SH, LI J SH, ZHANG J, *et al.*. Video foreground detection of tensor low-rank representation and spatial-temporal sparsity decomposition[J]. *Opt. Precision Eng.*, 2017, 25(2): 529-536. (in Chinese)
- [6] KALAL Z, MATAS J., MIKOLAJCZYK K. On-line learning of robust object detectors during unstable tracking[C]. *IEEE International Conference on Computer Vision Workshops*, IEEE, 2010: 1417-1424.
- [7] 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.
- [8] KUO C H, HUANG C, NEVATIA R. Multi-target tracking by on-line learned discriminative appearance models[C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010: 685-692.
- [9] HARE S, GOLODETZ S, SAFFARI A, *et al.*. Struck: Structured Output Tracking with Kernels [J]. *IEEE Trans Pattern Anal Mach Intell*, 2016, 38(10): 2096-2109.
- [10] 孙保基, 张葆, 宋策, 等. 基于角点增强改进的 TLD 目标跟踪算法[J]. *液晶与显示*, 2016, 31 (9):921-928.
- SUN B J, ZHANG B, SONG C, *et al.*. Improved TLD object tracking algorithm based on corner reinforced[J]. *Chinese Journal of Liquid Crystals and Displays*, 2016, 31(9): 921-928. (in Chinese)
- [11] DINH T B, VO N, MEDIONI G. Context tracker: Exploring supporters and distracters in unconstrained environments[C]. *CVPR 2011*, 2011: 1177-1184.
- [12] XIAO J, QIAO L, STOLKIN R, *et al.*. Distractor-Supported Single Target Tracking in Extremely Cluttered Scenes [J]. *Computer Vision-ECCV 2016*, 2016:121-136.
- [13] BAY H, ESS A, TUYTELAARS T, *et al.*. Speeded-Up Robust Features (SURF)[J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359.
- [14] DICLE C, CAMPS O I, SZNAIER M. The way they move: tracking multiple targets with similar appearance[C]. *2013 IEEE International Conference on Computer Vision*, 2013: 2304-2311.
- [15] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.*. High-speed tracking with kernelized correlation filters [J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, 37(3): 583-596.
- [16] BLAKE A, ISARD M. *Active Contours*[M]. Springer London, 1998.
- [17] MA C, HUANG J B, YANG X, *et al.*. Hierarchical convolutional features for visual tracking [C]. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015:3074-3082.

- [18] NUMMIARO K, KOLLER-MEIER E, VAN GOOL L. An adaptive color-based particle filter [J]. *Image and Vision Computing*, 2003, 21 (1): 99-110.
- [19] WU Y, LIM J, YANG M H. Online object tracking; a benchmark[C]. 2013 *IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 2411-2418.

作者简介:



宋 策(1987—),男,黑龙江齐齐哈尔人,博士,助理研究员,2009 年于哈尔滨工程大学获得学士学位,2014 年于长春光机所获得博士学位,主要从事目标跟踪与嵌入式系统方面的研究。E-mail: songce528@163.com



张 葆(1966—),男,吉林磐石人,博士,研究员,博士生导师,1989 年、1994 年、2004 年于长春光机所分别获得学士、硕士、博士学位,主要从事航空光电成像技术的研究。E-mail: cleresky@vip.sina.com

(本栏目编辑:秦 思)