

# 前景约束下的抗干扰匹配目标跟踪方法

刘大千<sup>1</sup> 刘万军<sup>2</sup> 费博雯<sup>3</sup> 曲海成<sup>2</sup>

**摘要** 传统模型匹配跟踪方法没有充分考虑目标与所处图像的关系,尤其在复杂背景下,发生遮挡时易丢失目标.针对上述问题,提出一种前景约束下的抗干扰匹配(Anti-interference matching under foreground constraint, AMFC)目标跟踪方法.该方法首先选取图像帧序列前  $m$  帧进行跟踪训练,将每帧图像基于颜色特征分割成若干超像素块,利用均值聚类组建簇集合,并通过该集合建立判别外观模型;然后,采用 EM (Expectation maximization) 模型建立约束性前景区域,通过基于 LK (Lucas-Kanade) 光流法框架下的模型匹配寻找最佳匹配块.为了避免前景区域中相似物体的干扰,提出一种抗干扰匹配的决策判定算法提高匹配的准确率;最后,为了对目标的描述更加准确,提出一种新的在线模型更新算法,当目标发生严重遮挡时,在特征集中加入适当特征补偿,使得更新的外观模型更为准确.实验结果表明,该算法克服了目标形变、目标旋转移动、光照变化、部分遮挡、复杂环境的影响,具有跟踪准确和适应性强的特点.

**关键词** 前景约束, 抗干扰匹配, 判别外观模型, 决策判定, 特征补偿

**引用格式** 刘大千, 刘万军, 费博雯, 曲海成. 前景约束下的抗干扰匹配目标跟踪方法. 自动化学报, 2018, 44(6): 1138–1152

**DOI** 10.16383/j.aas.2017.c160475

## A New Method of Anti-interference Matching Under Foreground Constraint for Target Tracking

LIU Da-Qian<sup>1</sup> LIU Wan-Jun<sup>2</sup> FEI Bo-Wen<sup>3</sup> QU Hai-Cheng<sup>2</sup>

**Abstract** The relation between a moving target and its image has not been fully considered in traditional model-matching tracking methods. The tracking drift problem may frequently occur when the target is occluded under a complex background. In this paper, a novel target tracking method, anti-interference matching under foreground constraint (AMFC), is proposed to solve this kind of problem. First, the method selects several initial frames from a video sequence for tracking training. Each of these frames is divided into several super-pixel blocks based on its color feature. These super-pixel blocks are combined into cluster sets by a mean shift algorithm to construct a discrimination appearance model. Then, a constrained foreground region is established using the expectation maximization (EM) model and a matching process is conducted based on the Lucas-Kanade (LK) optical flow method in order to select the optimum matching block. A decision-making algorithm is introduced to avoid the interference caused by similar targets in the foreground region, so as to increase the accuracy of the matching process. Moreover, in order to provide a more accurate target representation, an algorithm for appearance model online-updating is proposed. When a severe occlusion occurs, this algorithm can append appropriate feature compensations to the feature sets to improve the accuracy of the appearance model. Experimental results indicate that the proposed approach can provide superior tracking accuracy and adaptability, especially in the context of target deformation, target rotational movements, illumination changes, partial occlusion, and complex background.

**Key words** Foreground constraint, anti-interference matching, discrimination appearance model, decision-making, feature compensation

**Citation** Liu Da-Qian, Liu Wan-Jun, Fei Bo-Wen, Qu Hai-Cheng. A new method of anti-interference matching under foreground constraint for target tracking. *Acta Automatica Sinica*, 2018, 44(6): 1138–1152

收稿日期 2016-06-17 录用日期 2016-12-10  
Manuscript received June 17, 2016; accepted December 10, 2016

国家自然科学基金(61172144), 辽宁省科技攻关计划项目(2012216026)资助

Supported by National Natural Science Foundation of China (61172144), and Science and Technology Foundation of Liaoning Province (2012216026)

本文责任编辑 桑农

Recommended by Associate Editor SANG Nong

1. 辽宁工程技术大学电子与信息工程学院 葫芦岛 125105 2. 辽宁工程技术大学软件学院 葫芦岛 125105 3. 辽宁工程技术大学工商管理学院 葫芦岛 125105

1. School of Electronic and Information Engineering, Liaoning

基于视频序列的运动目标跟踪技术是计算机视觉领域研究的热点问题,在军事领域、人机交互、智能交通和医学图像处理<sup>[1-4]</sup>等领域发挥着重大作用.近年来,国内外的专家学者围绕基于目标自身特征匹配<sup>[5-6]</sup>、基于运动估计<sup>[7-8]</sup>和基于模型匹配<sup>[9-10]</sup>这三种主流跟踪方法进行深入研究.基于目标自身特性匹配的跟踪方法,在首帧中选取目标自身的特

Technical University, Huludao 125105 2. School of Software, Liaoning Technical University, Huludao 125105 3. School of Business and Management, Liaoning Technical University, Huludao 125105

征信息,并在其后帧中寻找该特征.寻找过程即为在视频序列中对目标的特征匹配和跟踪.常用的特征主要有几何形态、子空间信息、目标轮廓和特定特征点等,例如 SURF 特征点检测和 Harris 角点检测.由于目标在不断移动过程中,其自身特性(姿态、大小)也会发生改变,具有随机性,通常采用数字统计的方法来描述,例如像素灰度直方图、边缘方向梯度图等.

基于运动估计的跟踪方法主要是通过预测下一帧中目标可能出现的区域,并以区域中心向外扩充的方式对目标进行跟踪.在跟踪过程中如果对图像中所有场景信息逐一进行模型匹配,需要处理大量冗余信息,计算复杂度较高.基于运动估计的跟踪方法从一定程度上限制了目标的匹配范围,降低了计算复杂度,满足跟踪算法对时效性的要求.其中最具有代表性的是卡尔曼滤波方法.卡尔曼滤波是一种线性递归滤波,主要针对当前序列状态对下一时刻目标的状态做出最佳评估.施华等<sup>[11]</sup>提出一种利用像素可信度和空间位置的跟踪方法,通过卡尔曼滤波在时域中的背景模型与空间域的交叉熵协同实现对目标的检测和跟踪.

基于模型匹配的跟踪方法通过对目标的外观模型进行相似性分析来实现跟踪,其中外观模型用来评估候选目标在特定位置的似然性,在模型匹配中起着至关重要的作用.一种鲁棒的外观模型能够高效地解决场景中的光照变化、非目标物体遮挡、目标形态变化等问题.基于模型匹配的跟踪方法还分为基于整体模型的跟踪和基于局部模型的跟踪.基于整体模型的跟踪利用选定的目标区域建立整体模型进行模型匹配和跟踪. Babenko 等<sup>[12]</sup>通过对选定目标区域进行正/负样本的学习模型,从而构建判别式模型实现跟踪.基于局部模型的跟踪将目标区域进行分块处理,利用分割的局部子区域进行匹配跟踪,并实时更新局部模型,例如 Oron 等<sup>[13]</sup>提出的局部无序跟踪(Locally orderless tracking, LOT)算法,将目标分割为多个超像素块,通过超像素块的匹配对目标进行跟踪,同时利用粒子滤波对目标模型匹配加以约束,使得跟踪具有较高的鲁棒性.徐如意等<sup>[14]</sup>在 LK (Lucas-Kanade) 图像配准框架下,将稀疏表示目标跟踪算法利用求解的正则化来确定目标的特征模型,并创建动态字典用于模型的实时更新,从而准确跟踪目标. Wang 等<sup>[15]</sup>提出一种超像素跟踪(Super-pixel tracking, SPT)算法,从背景中提取目标构建一个具有区分度的判别外观模型,通过使用最大后验估计和超像素置信度计算下一时刻目标可能出现的位置,并实时更新外观模型.但基于模型匹配的跟踪方法仍存在问题,例如在模型匹配过程中,目标发生严重遮挡或图像背景较为复杂,使得构建外观模型的难度增加且目标容易淹

没在背景中,导致跟踪漂移,甚至丢失目标.

针对上述问题,提出一种前景约束下的抗干扰匹配(Anti-interference matching under foreground constraint, AMFC)目标跟踪方法.首先在前  $m$  帧图像序列中使用简单的模型匹配算法进行跟踪,确定  $m$  帧图像的目标区域,利用超像素分割所获得的目标区域,建立由若干超像素块组成的训练样本集,将具有相同颜色特征的超像素块集合进行均值聚类组建簇集合,计算每个簇的置信度概率,建立初始判别外观模型.然后利用基于 LK 光流法框架下的模型匹配,寻找最佳匹配块,引入高斯噪声模型防止强光照对目标跟踪的影响.同时,算法还引入决策判定,避免前景区域中相似物体的干扰,进而实现精确的目标跟踪.最后利用 EM (Expectation maximization) 模型估计下一帧图像的前景参数信息,并对目标是否发生严重遮挡进行判断,若发生遮挡,在特征集中加入适当补偿,即可获得新的判别外观模型.

本文的主要贡献为: 1) 将图像的前几帧作为训练样本集,通过聚类构建目标模型,滤除背景信息的干扰,得到较鲁棒的目标模型. 2) 对图像帧中的前景信息进行估计,对局部模型的匹配过程进行约束.只在前景区域中寻找最佳匹配的超像素块,达到局部最优匹配. 3) 在匹配过程中引入决策判定,只保留限定区间内的最佳匹配块,排除与目标相似物体的干扰,提高算法匹配抗干扰性. 4) 在模型更新中对目标是否发生严重遮挡进行判断,若发生遮挡,在特征集中加入适当补偿,使更新的外观模型对目标的描述更加准确.

## 1 算法概述

AMFC 算法分为建立判别外观模型、局部模型匹配和模型更新三个过程.算法流程如图 1 所示.

1) 建立判别外观模型.在前  $m$  帧图像序列中使用简单的模型匹配算法进行跟踪,确定  $m$  帧图像的目标区域,再利用超像素分割每帧获得的目标区域,建立由若干超像素块组成的训练样本集;然后利用均值聚类算法将具有相同颜色特征的超像素块集合到一个簇中,从而组建簇集合;最后,计算每个簇的置信度概率,判断其是否属于目标区域,从而利用超像素块的特征以及置信度概率建立初始判别外观模型.

2) 局部模型匹配.利用基于 LK 光流框架下的模型匹配,对于新的图像帧估计目标的前景区域,在此区域中进行最佳模型匹配,计算每个超像素块的匹配概率,寻找最佳匹配块,同时引入高斯噪声模型和决策判定,防止强光照以及前景区域中相似物体对目标跟踪过程的干扰.反复此过程,完成目标模型

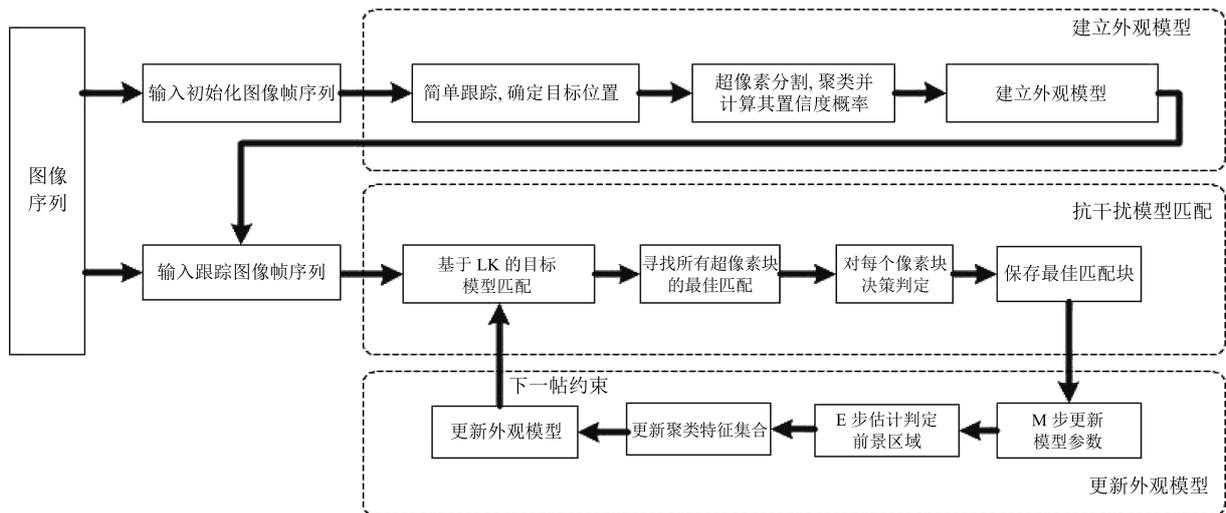


图1 AMFC 算法流程示例图

Fig.1 The flow diagram of AMFC algorithm

的匹配, 从而对目标进行准确跟踪。

3) 外观模型的更新. 首先利用 EM 模型估计下一帧图像的前景参数信息, EM 模型由两部分组成: E 步估计前景信息, M 步进行参数优化. 然后判断目标是否发生严重遮挡, 若没有发生严重遮挡, 则利用最新帧的跟踪结果替换外观模型中最旧帧, 若发生严重遮挡, 则选择与当前最近帧作为被替换帧, 并将被替换帧中的部分特征作为补偿集, 合并到当前帧的特征集合中, 对该聚类集合重新计算置信度获得新的判别外观模型。

## 2 相关工作

### 2.1 超像素分割

超像素分割是将视频帧图像分割成许多位置相近且特征大致相同(颜色、梯度、亮度等)的子区域, 即为像素点的集合. 超像素分割是一种过分割算法, 每个子区域具有某种近似相同的特征, 而相邻子区域在这种特性的度量下有所不同, 分割效果明显, 因此超像素块可以清楚地描述目标的局部特征. 超像素分割将许多像素点聚集成易于理解和计算的子区域, 通常情况下, 这些视频帧的子区域都是均匀分布的, 可将其作为一个整体来处理, 这些子区域也被称为超像素块。

基于像素级的目标跟踪算法往往不符合人类视觉注意机制的成像特点, 由于人类视觉注意机制的焦点往往不是某个单一像素点, 而是将这些像素点组合成一整块区域用于特征的识别. 以超像素块作为基本跟踪单元代替像素进行建模, 保留了丰富的图像结构特征信息, 与采用直方图、轮廓等特征的目标跟踪算法相比具有更强的特征描述能力. 本文利用 Achanta 等提出的基于颜色相似度和空间距离的

简单线性迭代聚类 (Simple linear iterative clustering, SLIC)<sup>[16]</sup> 进行超像素分割. 其特点为分割后的超像素块紧凑性好、边缘贴强度高、大小均匀且运算速度较快。

### 2.2 LK 算法

LK 算法<sup>[17]</sup> 是将目标跟踪问题视为在每一帧中某一已知邻域内寻找最佳匹配问题. 给定跟踪目标的模型, 通过求解最优化问题来定位跟踪目标。

设  $T$  为目标模型, 模型中的像素点集表示为  $PIXEL = \{pixel\}_{i=1}^n$ , 其中,  $pixel = (x, y)$ , 表示像素点的横纵坐标.  $I$  为第  $t$  帧图像, 图像像素  $I(W(pixel, w))$  由模型像素  $T(pixel)$  通过  $W(pixel, w)$  函数相似转化得到,  $w$  为转化位置参数. 已知在第  $t-1$  帧图像的估计位置, 即给出  $w^{t-1}$ . 在第  $t$  帧图像中寻找更新参数  $\Delta w$ , 使得  $w = w^{t-1} + \Delta w$  取最小值,  $\Delta w$  应满足

$$\Delta w = \arg \min \sum_{pixel \in PIXEL} \left[ I(W(pixel, w^{t-1} + \Delta w)) - T(pixel) \right]^2 \quad (1)$$

由于无法一次计算出最优参数  $w$ , 因此估计一组参数  $\Delta w$  用于每次迭代中更新前一次的参数  $w$ , 设定一阈值  $\varepsilon$  作为终止条件. 终止条件为: 当  $(\|\Delta w\| \leq \varepsilon)$  时

$$w = w^{t-1} + \Delta w \quad (2)$$

## 3 前景约束下的抗干扰匹配跟踪

AMFC 算法分为三个过程: 1) 建立基于超像素块的判别外观模型; 2) 引入决策判定的前景约束性

抗干扰局部模型匹配; 3) 引入补偿集的外观模型更新方法.

### 3.1 建立判别外观模型

为了建立可以区分背景的外观模型, 需要一个样本集进行训练. 首先利用简单的 EMD 模型匹配算法<sup>[18]</sup>跟踪前  $m$  帧 (在本文实验中  $m$  取 4) 视频图像序列, 确定每帧图像中目标的位置信息. 然后通过 SLIC 算法分割每个训练图像帧中扩展的目标区域 (以圈定目标区域的中心为中心点, 对角线长度为目标区域对角线的 1.5 倍进行扩展), 得到  $N_t$  个超像素块  $pixel_i$ . 在第  $t$  帧图像中, 通过集合  $F = \{f_t^x | t = 1, \dots, m; x = 1, \dots, N\}$  表示  $N_t$  个超像素块的特征集合.

由于分割算法是对目标的扩展区域进行分割, 获取的超像素块包含了对目标以及背景区域的描述, 所以应用均值漂移 (Mean shift, MS) 算法<sup>[19]</sup>对特征集合进行聚类, 获得  $n$  个簇  $clst(i)$  ( $i = 1, \dots, n$ ), 每个簇包括簇中心  $f_c(i)$  (其为特征矢量, 表示它包含的所有超像素块的特征均值)、簇半径  $r_c(i)$  和簇成员  $\{f_t^x | f_t^x \in clst(i)\}$ . 将  $n$  个簇对应到  $m$  个训练帧图像的局部区域.

通过聚类获得  $n$  个簇, 记每个簇区域包含  $S(i)$  个超像素块. 若这些超像素块都位于目标区域中, 则

该聚类簇属于目标区域, 若这些超像素块都位于背景区域中, 则该聚类簇属于背景区域. 但在大多数情况下, 超像素集合往往部分属于目标区域, 部分属于背景区域, 需要引入置信度的计算. 设与目标区域相对应的超像素块数记为  $Y(i)$ , 对应到背景区域的超像素块数记为  $E(i)$ , 即  $S(i) = Y(i) + E(i)$ . 为每个簇  $clst(i)$  ( $i = 1, \dots, n$ ) 分配一个权值, 记为聚类的置信度  $C_i^c$  ( $C_i^c \in [-1, 1]$ ).

$$C_i^c = \frac{Y(i) - E(i)}{S(i)}, \quad \forall i = 1, \dots, n \quad (3)$$

其中,  $C_i^c$  表示簇  $clst(i)$  ( $i = 1, \dots, n$ ) 是否属于目标区域. 当  $C_i^c \geq 0$  时,  $clst(i)$  属于目标区域, 值越大表明越偏向于目标区域. 即

$$pixel_i = \begin{cases} \text{目标区域,} & C_i^c \geq 0 \\ \text{背景区域,} & C_i^c < 0 \end{cases} \quad (4)$$

利用式 (3) 和式 (4) 判断聚类簇是否属于目标区域, 从而建立判别外观模型, 建立过程如图 2 所示. 需要指出的是, 在图 2(d) 中, 超像素块的置信度 ( $C_i^c$ ) 越高, 则表示该超像素块越偏向目标; 超像素块的置信度低, 则表示该超像素块越偏向背景.

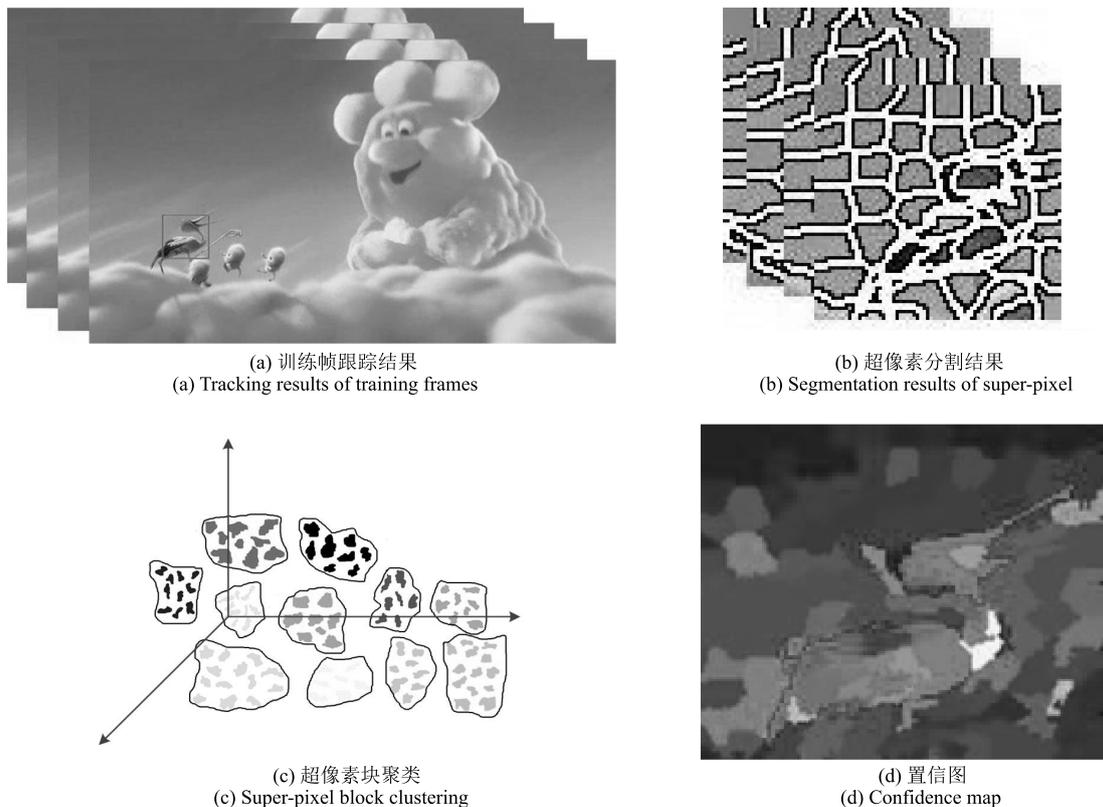


图 2 建立判别外观模型过程图

Fig. 2 Procedure of establishing discriminant appearance model

### 3.2 抗干扰模型匹配

#### 3.2.1 前景区域划分

**定义 1.** 在模型匹配时, 定义一个二进制变量  $h_{pixel}^I$  表示图像像素是否属于前景区域, 即

$$h_{pixel}^I = \begin{cases} 1, & pixel \in \text{目标区域} \\ 0, & pixel \in \text{背景区域} \end{cases} \quad (5)$$

利用带参数的伯努利分布表示图像像素是否属于前景, 即

$$p(h_{pixel}^I = 1) = v, \quad p(h_{pixel}^I = 0) = 1 - v \quad (6)$$

引入变量  $h_{pixel}^I$  有以下优点:

1) 利用目标模型进行匹配, 有时背景超像素块也可能与目标模型匹配, 导致目标圈定漂移, 甚至会丢失目标. 将匹配的图像帧分为前景和背景进行判定, 如果匹配的是背景像素, 则自动舍弃, 不做圈定, 从而达到局部特征自适应匹配, 避免了跟踪过程中的漂移现象.

2) 虽然 Oron 等也利用目标的前景/背景信息进行局部匹配, 但引入双重二进制判断前景/背景信息, 增加了算法的时间复杂度. 本文算法借助其思想, 引入前景区域判别, 并在目标模型建立时对前景/背景信息加以区别, 所以在匹配过程中只引入单一判断变量, 降低算法的时间复杂度.

#### 3.2.2 高斯噪声模型

利用超像素块的灰度特征进行模型匹配时, 易受到光照变化的影响. 当光照变化强烈时, 为了使算法能够保持较准确的跟踪, 本文引入高斯噪声模型抑制光照变化的影响.

**定义 2.** 加噪后的模板称为噪声模板, 定义为  $\tilde{T}$ . 加噪后的图像称为噪声图像, 定义为  $\tilde{I}$ , 构建的高斯模型为

$$\begin{aligned} p(\tilde{T}_{pixel}|T_{pixel}) &= G(\tilde{T}_{pixel}|T_{pixel}, \sigma) \\ p(\tilde{I}_{pixel}|I_{pixel}) &= G(\tilde{I}_{pixel}|I_{pixel}, \sigma) \end{aligned} \quad (7)$$

设定局部匹配子空间 (超像素块) 用  $\Omega$  表示, 则局部匹配概率为

$$p_{\Omega}(h_{pixel}^I, I_{pixel}, T_{pixel}, \tilde{I}_{pixel}, \tilde{T}_{pixel}) =$$

$$p(h_{pixel}^I)p(I_{pixel}|h_{pixel}^I)p(T_{pixel}) \times G(T_{pixel} - I_{pixel}|I_{pixel}, \sqrt{2}\sigma)^{h_{pixel}^I} \quad (8)$$

具体推导过程见附录 A. 由于 Singer1 图像序列中光照变化明显, 在本文实验中  $\sigma$  值的选取主要根据该图像序列来确定. Singer1 图像序列中  $\sigma$  值与平均中心误差之间的关系如图 3 所示.

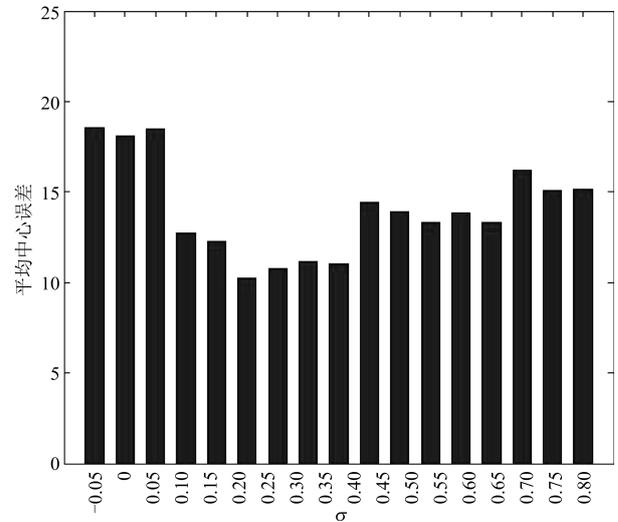


图 3  $\sigma$  值与平均中心误差之间的关系

Fig. 3 The relationship between the  $\sigma$  value and average center error

图 3 中, 横坐标表示选取  $\sigma$  值的大小, 纵坐标表示计算在 Singer1 视频图像序列中跟踪结果的平均中心误差. 由此可以看出, 当选定  $\sigma \in [0.25, 0.35]$  时, 算法的跟踪效果最佳. 加入噪声模型 ( $\sigma = 0.25$ ) 与不加入噪声模型的跟踪结果如图 4 所示. 从图中可以看出, 当光照变化剧烈时, 加入高斯噪声 (深色矩形框) 可以较为明显地提高跟踪准确性.

#### 3.2.3 最优化模型匹配

在局部模型匹配的过程中, 最重要的是寻找最佳匹配超像素块. 利用 LK 光流法进行模型匹配, 寻找 LK 最优化匹配问题等价于搜索最大似然像素集, 具体证明过程见附录 B.

采用式 (9) 计算最大似然匹配像素集.

$$\max_{\theta} \log p(T, I_w|\theta) = \max_{\theta} \log p(T, I_{w^{t-1}+\Delta w}|\theta) \quad (9)$$



图 4 加入噪声模型对比结果

Fig. 4 Comparison results of noise model

其中,  $\theta = \{v, \sigma, w\}$ . 由像素之间的独立性可以得到:

$$\log p(T, I_w) = \sum_{pixel \in PIXEL} \sum_{h_{pixel}^I \in \{0,1\}} \log p(h_{pixel}^I, I_{pixel}, T_{pixel}) \quad (10)$$

其中,  $p(h_{pixel}^I, I_{pixel}, T_{pixel})$  在附录 A 中给出.

$$p_{\Omega}(h_{pixel}^I, I_{pixel}, T_{pixel}) = p(I_{pixel})p(I_{pixel}|h_{pixel}^I)p(T_{pixel}) \times G(T_{pixel} - I_{pixel}|I_{pixel}, \sqrt{2}\sigma)^{h_{pixel}^I} \quad (11)$$

模型匹配即通过局部匹配概率计算最大似然匹配概率.

### 3.2.4 决策判定

在前景区域中有相似物体干扰时 (例如局部遮挡), 所找到的最佳匹配块往往不能准确地显示被跟踪目标, 因此算法分两步操作避免相似物体的干扰.

在匹配最佳块时, 前景区域中相似物体的干扰处理是必要的. 假设在目标连续运动的情况下, 计算相邻帧各个最佳匹配块中心距离, 即

$$d_i = \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \quad (12)$$

其中,  $d_i$  为相邻帧最佳匹配块间的中心距离,  $x_i$  和  $y_i$  分别为第  $t$  帧第  $i$  个像素块中心点的横纵坐标,  $x'_i$  和  $y'_i$  分别为第  $t-1$  帧第  $i$  个像素块中心点的横纵坐标. 对  $d_i$  取平均, 即  $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$ . 设定度量阈值区间为  $\bar{d} \pm \Delta d$  (在本文实验中,  $\Delta d$  取平均中心距离的  $1/5$ ), 对相邻帧间的最佳匹配块进行判定, 若距离  $d_i$  不在该区间范围内, 则自动舍弃, 只保留在该区间内的最佳匹配块. 在圈定目标时, 利用  $\bar{d}$  估计

被舍弃的匹配块位置, 将其与保留的匹配块相结合确定最优目标区域.

匹配块位置信息示意如图 5 所示. 图 5(a) 表示在  $t-1$  帧中圈定的最优前景区域. 假定由 4 个最佳匹配块组成, 第  $t$  帧中匹配块位置如图 5(b) 所示. 最佳匹配块 1 不在度量阈值区间 ( $\bar{d} \pm \Delta d$ ) 内, 则被舍弃, 改用  $\bar{d}$  重新估计匹配块 1 的位置从而得到第  $t$  帧的最优目标区域.

为了验证决策判定的有效性, 采用 Basketball 图像序列进行跟踪测试, 测试结果如图 6 所示. 图 6(a) 为加入决策判定的跟踪结果, 图 6(b) 为没有加入决策判定的跟踪结果. 从这两组图中可以明显看出引入决策判定较好地解决了前景区域相似物体干扰的情况, 使得算法的跟踪结果更为准确.

### 3.2.5 光流迭代寻优

光流迭代寻优即为给定跟踪目标的模型 ( $T$ ), 通过求解  $\max_{\Delta w} \log p(I(W(pixel, w^{t-1} + \Delta w))|T)$  最优化问题来定位跟踪目标. 由附录 B 的 LK 最优化匹配问题证明, 可以得到最优化解.

$$\begin{aligned} \max_{\Delta w} \log p(I(W(pixel, w^{t-1} + \Delta w))|T) = & \max_{\Delta w} \sum_{b \in (0,1)} \left[ \sum_{pixel \in PIXEL} p_{old}(h_{pixel}^I = b) \times \right. \\ & \left. \log p(I_{pixel}|h_{pixel}^I = b) - \frac{1}{2\sigma^2} \times \right. \\ & \left. \sum_{pixel \in PIXEL} p_{old}(h_{pixel}^I = 1)(T_{pixel} - I_{pixel})^2 \right] \end{aligned} \quad (13)$$

其中,  $p_{old} = p(h_{pixel}^I|I_{pixel}, T_{pixel}; \theta_{old})$ . 将式 (13) 进行一阶泰勒展开, 然后求导 (具体过程见附录 C), 可得

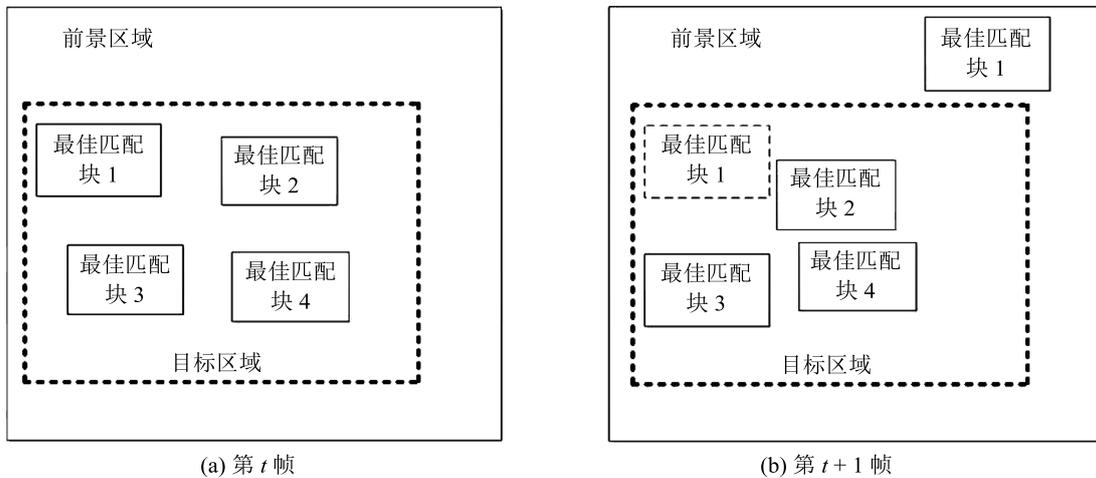


图 5 匹配块位置信息示意图

Fig. 5 Diagram of matching block location information

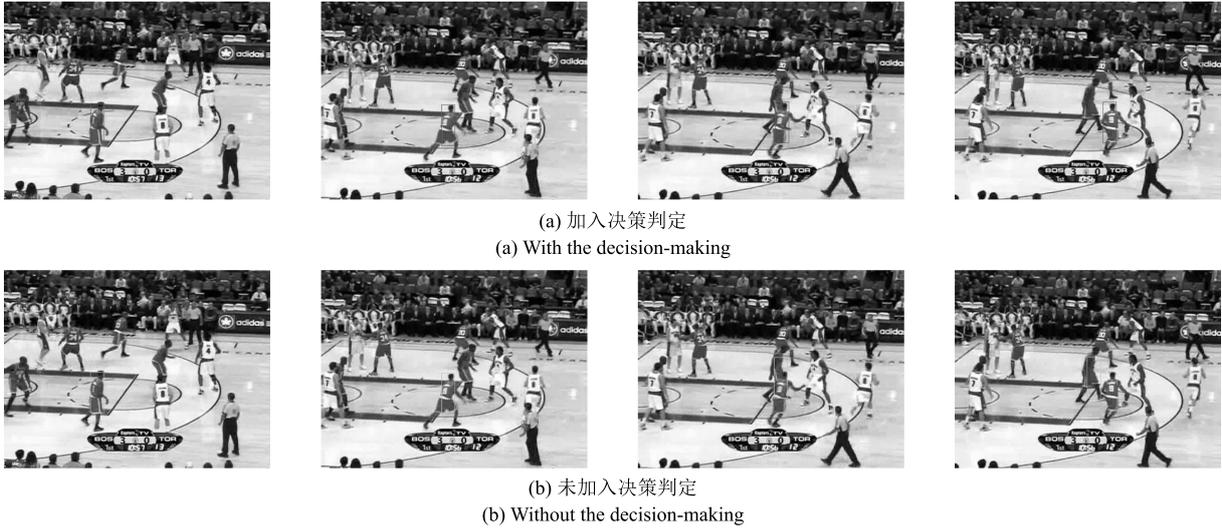


图 6 决策判定的有效性验证

Fig. 6 The effectiveness verification of decision-making

$$\Delta w = \left[ \sum_{pixel \in PIXEL} Q(pixel) [T(pixel) - I(W(pixel, w))] \times \nabla I \frac{dW(w)}{dw} - \frac{1}{2} \left( \sum_{pixel \in PIXEL} Q_0(pixel) \nabla I_0 \frac{dW(w)}{dw} + \sum_{pixel \in PIXEL} Q_1(pixel) \nabla I_1 \frac{dW(w)}{dw} \right) \right] \times \left[ \sum_{pixel \in PIXEL} Q(pixel) \left( \nabla I_0 \frac{dW(w)}{dw} \right)^T \times \left( \nabla I_0 \frac{dW(w)}{dw} \right) \right]^{-1} \quad (14)$$

其中,  $I_0 = \log p(h_{pixel}^I = 0)$  表示背景图像区域,  $I_1 = \log p(h_{pixel}^I = 1)$  表示前景图像区域,  $Q_0(pixel) = p_{old}(h_{pixel}^I(W(pixel, w) = 0))$  为当前帧背景图像  $I_0$  的似然估计的权重函数,  $Q_1(pixel) = p_{old}(h_{pixel}^I(W(pixel, w) = 1))$  为当前帧前景图像  $I_1$  的似然估计的权重函数,  $Q(pixel) = 1/(2\sigma^2) \times p_{old}(h_{pixel}^I(W(pixel, w) = 1, I(W(pixel, w) - T(pixel)) = 0))$  为对输入图像  $I$  通过传统 LK 算法目标函数的相似性度量。

### 3.3 模型更新

#### 3.3.1 前景模型更新

利用文献 [13] 解决存在隐含变量优化问题的 EM 算法估计隐含变量  $h_{pixel}^I$ . EM 算法利用优化参数  $\theta$  估计隐含变量  $h_{pixel}^I$ , 过程如下:

E 步: 已知前一帧参数  $\theta_{old}$ , 当前帧图像  $I$  和匹配模型  $T$ , 可求出隐藏变量  $h_{pixel}^I$  的概率分布  $\sum_{pixel \in PIXEL} p(h_{pixel}^I | I_{pixel}, T_{pixel}; \theta_{old})$ .  $h_{pixel}^I$  是二进制变量, 则所有像素的条件概率可以写成

$$p(h_{pixel}^I = b | I_{pixel}, T_{pixel}) = \frac{p(h_{pixel}^I = b, I_{pixel}, T_{pixel})}{\sum_{a \in \{0,1\}} p(h_{pixel}^I = a, I_{pixel}, T_{pixel})} \quad (15)$$

其中,  $b \in \{0, 1\}$ .

M 步: 对参数  $\theta$  进行优化.

$$\theta_{new} = \arg \max \sum_{pixel \in PIXEL} \sum_{h_{pixel}^I \in \{0,1\}} p_{old} \times \log p(h_{pixel}^I, I_{pixel}, T_{pixel}) \quad (16)$$

其中,  $p_{old} = p(h_{pixel}^I | I_{pixel}, T_{pixel}; \theta_{old})$ .

#### 3.3.2 外观模型的更新

传统的外观模型更新方法是在跟踪若干帧后 (通常取 5 帧或 10 帧), 用最新帧的跟踪结果代替聚类库中的最旧帧来重建外观模型. 这类方法当目标发生严重遮挡时易出现跟踪漂移甚至丢失目标的情况.

针对在跟踪过程中是否发生严重遮挡的问题, 本文定义一个遮挡阈值  $\theta_0$  ( $\theta_0$  取 0.7) 与当前帧计算的遮挡阈值  $\theta_t$  进行比较, 如果  $\theta_t < \theta_0$ , 则判定目标发生严重遮挡. 遮挡阈值为当前帧中目标区域置信度与前  $m$  帧跟踪的目标区域置信度均值之比, 即

$$\theta_t = \frac{C_t}{\text{mean}(C_t, C_{t-1}, \dots, C_{t-m+1})} \quad (17)$$

若未发生严重遮挡, 则利用最新帧的跟踪结果与外观模型中最旧帧进行替换, 重建外观模型(每10帧更新一次), 若发生严重遮挡, 则选择与模型中当前最近帧作为被替换帧, 并在此帧中选择  $n_{cluster}$  个聚类特征(其聚类特征的置信度大于0, 即属于目标区域的特征)作为跟踪当前帧的补偿集. 在实验中  $n_{cluster}$  取10, 若替换帧中属于目标区域的特征不足10个, 则将这些置信度大于0的特征直接作为补偿集. 将补偿集合并到当前帧的特征集中(严重遮挡下), 作为新的特征集合加入聚类特征集合, 完成聚类集合的更新. 然后对该聚类集合重新计算置信度, 即可获得新的判别外观模型.

加入补偿集的判别外观模型效果如图7所示. 从图7中可以看出, 只通过替换整帧图像来更新外观模型无法表示目标被严重遮挡的区域, 使得外观模型不能有效地描述目标, 当遮挡消失时, 建立的外观模型无法准确地识别目标被遮挡的部分, 这样导致跟踪结果不够准确. 而引入补偿集后, 外观模型能够识别目标被遮挡的部分, 使得算法的跟踪效果较为准确.

#### 4 实验结果与分析

本文算法在 MATLAB R2012b 环境下实现. 计算机环境: CPU 为 Intel Core i7, 内存为 16 GB. 算法对每一组图像序列都在第一帧人工标出目标区域. 为了验证本文算法的性能, 在吴毅提出的目

标跟踪评测标准<sup>[20]</sup>上选取12组具有代表性的图像序列以及8种跟踪算法进行对比实验, 以保证实验的可信性和全面性. 同时与近几年较为流行的 LOT (Locally orderless tracking) 和 SPT (Superpixel tracking) 跟踪算法进行对比, 验证算法的有效性和实用性. 12组图像序列分别是 Girl、Deer、Bird2、Football、Lemming、Woman、Bolt、CarDark、David1、David2、Singer1 和 Basketball. 这些图像序列基本涵盖光照变化、部分遮挡、目标形变、复杂背景等在跟踪过程中具有挑战性的因素. 8种跟踪算法分别为 ASLA (Adaptive structural local sparse appearance)<sup>[21]</sup>、FRAG (Fragments-based tracking)<sup>[22]</sup>、SCM (Sparsity-based collaborative model)<sup>[23]</sup>、VTD (Visual tracking decomposition)<sup>[24]</sup>、L1APG (L1 tracker using accelerated proximal gradient approach)<sup>[25]</sup>、CT (Compressive tracker)<sup>[26]</sup>、OAB (Online AdaBoost tracker)<sup>[27]</sup>、TLD (Tracking-learning-detection)<sup>[28]</sup>. 本文采用3种评估标准来评价算法的整体性能, 分别为中心误差<sup>[29]</sup>、跟踪重叠率<sup>[30]</sup>和运行效率. 中心误差是真实目标中心与算法跟踪的目标中心之间欧氏距离的误差, 以像素为测量单位来评价算法的精度; 跟踪重叠率是人工标注12组图像帧序列中目标的位置; 应用重叠指数  $score$  评估跟踪性能.

$$score = \frac{area(R_T \cap R_G)}{area(R_T \cup R_G)} \quad (18)$$



图7 判别外观模型的对比效果

Fig.7 The contrast effects of discrimination appearance model

式中,  $area$  表示区域面积,  $R_T$  表示标记实际面积,  $R_G$  表示跟踪输出面积.  $score$  值越大意味着算法具有良好的跟踪定位精度, 即跟踪的成功率越高. 实验中用到的图像帧序列信息见表 1.

#### 4.1 实验参数

在视频图像序列的前几帧中, 目标的运动并不会出现较大程度的偏移, 提取这些帧中的目标信息作为训练样本得到的目标模型更鲁棒. 由于跟踪实时性在整个算法的评价中是一个重要指标, 本文在保证算法跟踪成功率的前提下, 减少一些超像素分割、聚类、计算置信度的时间来满足算法实时性要求, 所以取  $m = 4$ .

超像素分割的块数  $N = 100$ . 利用 SLIC 算法分割每个训练图像帧中扩展的目标区域 (以圈定目标区域的中心为中心点, 对角线长度为目标区域对

角线的 1.5 倍进行扩展). 在小区域内进行分割, 若分割的块数过多会出现“过分割”现象, 而分割块数太少又无法完全分离无关背景信息. 本文借助前人的经验 (SPT 算法), 考虑到算法的实时性和准确性要求, 取  $N = 100$ .

在  $\sigma$  值的选择上, 不同的视频由于光照变化存在差异, 所以选择的  $\sigma$  值也会有所差异, 但为了保证实验的统一性和公平性, 考虑加入高斯噪声模型的意义在于抑制光照变化对跟踪的影响, 本文以实验中光照变化明显的 Singer1 视频序列中的  $\sigma$  值作为最优选择, 即  $\sigma = 0.25$ .

#### 4.2 实验结果

实验比较的 11 种跟踪算法在 12 个图像序列上平均中心误差见表 2, 11 种跟踪算法在 12 个图像序列上的跟踪重叠率见表 3.

表 1 实验图像序列信息

Table 1 The information of the test image sequences

图像序列	光照变化	遮挡	形变	复杂背景	旋转
Girl		✓			✓
Deer				✓	✓
Bird2		✓	✓		
Football		✓		✓	
Lemming	✓	✓		✓	✓
Woman		✓	✓		
Bolt		✓	✓	✓	
CarDark	✓				
David1	✓		✓		✓
David2					✓
Singer1	✓	✓		✓	
Basketball		✓	✓	✓	✓

表 2 不同跟踪算法的平均中心误差

Table 2 Average center errors of different tracking algorithms

图像序列	ASLA	FRAG	SCM	VTD	L1APG	CT	OAB	TLD	LOT	SPT	AMFC
Girl	36.76	24.27	<b>3.47</b>	8.64	25.51	19.45	4.68	7.66	20.28	4.73	4.76
Deer	<b>6.74</b>	87.64	37.62	7.93	38.76	52.13	16.77	25.34	29.44	28.76	11.92
Bird2	20.12	14.96	11.53	46.24	25.44	37.87	26.16	10.23	47.61	<b>6.32</b>	6.97
Football	16.62	15.38	9.64	13.57	11.31	17.44	9.28	13.82	7.15	11.27	<b>6.28</b>
Lemming	47.52	112.63	75.43	60.73	142.28	37.82	73.74	32.68	14.48	<b>9.41</b>	11.39
Woman	76.57	118.28	22.24	107.69	115.57	104.58	30.94	137.47	118.41	23.36	<b>21.71</b>
Bolt	62.42	73.62	9.37	49.17	132.48	38.74	129.56	141.24	17.62	15.23	<b>8.44</b>
CarDark	5.30	6.23	6.32	28.72	<b>3.44</b>	18.70	39.86	21.36	24.18	21.58	8.35
David1	<b>3.57</b>	84.41	4.38	48.95	5.76	9.69	31.26	8.92	37.84	23.29	9.36
David2	8.94	67.51	6.72	3.51	<b>3.23</b>	69.83	36.34	6.73	3.97	8.48	9.21
Singer1	45.62	57.83	17.85	12.53	49.84	31.07	36.27	22.17	16.64	<b>10.14</b>	10.25
Basketball	106.62	18.42	116.24	9.48	84.47	79.41	37.13	95.46	127.83	12.38	<b>7.73</b>
平均	36.40	56.77	26.73	33.10	53.17	43.06	39.33	43.59	38.79	14.58	9.70

注: 粗体为最优结果.

表 3 不同跟踪算法的跟踪重叠率  
Table 3 Tracking overlap ratio of different tracking algorithms

图像序列	ASLA	FRAG	SCM	VTD	L1APG	CT	OAB	TLD	LOT	SPT	AMFC
Girl	0.31	0.41	<b>0.74</b>	0.69	0.39	0.29	0.73	0.56	0.43	0.71	0.73
Deer	<b>0.69</b>	0.11	0.47	0.63	0.45	0.34	0.56	0.41	0.50	0.53	0.61
Bird2	0.66	0.70	0.74	0.41	0.63	0.44	0.58	0.80	0.43	<b>0.84</b>	0.83
Football	0.61	0.63	0.71	0.65	0.68	0.62	0.68	0.65	0.73	0.69	<b>0.75</b>
Lemming	0.71	0.43	0.53	0.57	0.39	0.74	0.56	0.77	0.83	<b>0.86</b>	0.86
Woman	0.27	0.14	0.59	0.15	0.15	0.17	0.48	0.12	0.14	0.59	<b>0.61</b>
Bolt	0.53	0.47	0.76	0.55	0.23	0.58	0.20	0.16	0.72	0.73	<b>0.77</b>
CarDark	<b>0.82</b>	0.82	0.81	0.43	<b>0.82</b>	0.72	0.38	0.46	0.44	0.43	0.79
David1	<b>0.83</b>	0.23	0.82	0.53	0.80	0.77	0.57	0.79	0.55	0.62	0.76
David2	0.68	0.21	0.69	0.73	<b>0.74</b>	0.02	0.33	0.69	0.73	0.68	0.64
Singer1	0.58	0.55	0.73	0.74	0.57	0.64	0.63	0.69	0.72	<b>0.76</b>	0.76
Basketball	0.19	0.63	0.17	0.67	0.25	0.27	0.59	0.21	0.14	0.68	<b>0.69</b>
平均	0.57	0.45	0.65	0.56	0.51	0.47	0.52	0.53	0.53	0.68	0.73

从表 2 和表 3 可以看出, 在大部分实验序列中, AMFC 算法优于同类算法 (LOT 算法和 SPT 算法). 不仅如此, 与其他 8 个当前主流的跟踪算法相比, AMFC 算法也具有较大优势, 而且 AMFC 算法在所有图像序列上的平均中心误差及跟踪重叠率均是最优秀的. 这表明本文提出的 AMFC 算法是合理有效的, 达到甚至超过了当前主流算法的跟踪效果. 图 8 是 11 种跟踪算法在 12 组图像序列中的跟踪结果.

#### 4.3 算法准确性分析

图像序列 David2, Girl 和 Basketball 主要用于测试当目标发生旋转时算法的跟踪效果. David2 序列的跟踪效果如图 8(j) 所示, Girl 序列的跟踪效果如图 8(a) 所示, Basketball 序列的跟踪效果如图 8(l) 所示. 在 David2 序列中, 当目标发生小范围旋转时, 大部分跟踪算法都能较准确地跟踪目标. 在 Girl 序列中, 当目标发生旋转且在部分图像帧中目标 (人脸) 背对摄像机时, SCM, VTD, OAB, TLD, SPT 和 AMFC 算法的跟踪效果较为理想. 在 Basketball 序列中, 当目标发生旋转且背景较为复杂时, 仅 VTD, SPT 和 AMFC 算法依然保持较好的跟踪效果. 综合这三组图像序列, AMFC 算法的跟踪效果最好, 平均重叠率为 0.69. 相比同类算法 (LOT 算法和 SPT 算法), AMFC 算法利用目标的外观信息构建局部先验模型, 通过前景信息进行目标匹配跟踪. 当发生旋转时, 目标的局部前景信息发生改变, 此时需要不断地进行模型更新. 在模型更新时, 选择与当前最近帧作为被替换帧, 并将被替换帧中的部分特征作为补偿集合并到当前帧的特征集中, 从而得到新的判别外观模型, 使得算法能够更准确地跟踪目标.

图像序列 Singer1, CarDark 和 David1 主要用于测试当发生光照变化时算法的跟踪效果. Singer1 序列的跟踪效果如图 8(k) 所示, CarDark 序列的跟踪效果如图 8(h) 所示, David1 序列的跟踪效果如图 8(i) 所示. 在 Singer1 序列中, 当图像中光照变化强烈时, 大部分跟踪算法都能较准确地跟踪目标. 在 CarDark 序列中, 当图像中存在较强烈光照变化且目标移动较快时, ASLA, FRAG, SCM, VTD, L1APG 和 AMFC 算法有较好的跟踪效果. 在 David1 序列中, 图像中存在较强烈光照变化且目标发生形变时, ASLA, SCM, L1APG, SPT 和 AMFC 算法能够保持较好的跟踪效果. 但相比较而言, AMFC 算法的平均跟踪重叠率 (0.77) 虽高于同类算法, 但略低于 SCM 算法 (0.79). 虽然 AMFC 算法加入了高斯噪声模型, 增强了模型的自适应性, 在进行局部模型匹配时具有较高的准确性, 但由于光照变化强烈, 在局部模型匹配时会发生轻度漂移, 并且算法利用目标的灰度特征, 这样的匹配成功率与 SCM 算法相比较低. 如果将本文算法的噪声模型系数  $\sigma$  上调, 跟踪效果优于其他跟踪算法, 但为了保持实验的公平性, 并没有将此结果作为参考.

图像序列 Bird2, Woman 和 Bolt 主要用于测试当目标发生形变时算法的跟踪效果. Bird2 序列的跟踪效果如图 8(c) 所示, Woman 序列的跟踪效果如图 8(f) 所示, Bolt 序列的跟踪效果如图 8(g) 所示. 在 Bird2 序列中, ASLA, FRAG, SCM, L1APG, TLD, SPT 和 AMFC 算法都获得了较好的跟踪重叠率. 在 Woman 和 Bolt 序列中目标存在严重的形变, 仅 SCM, SPT, AMFC 有较好的跟踪效果. 在这 3 组图像序列中, AMFC 算法的跟踪效果最佳, 平均重叠率为 0.74. 相比同类算法, 本文利用前景约束下的超像素块进行局部匹配, 可以较准确匹配目标的

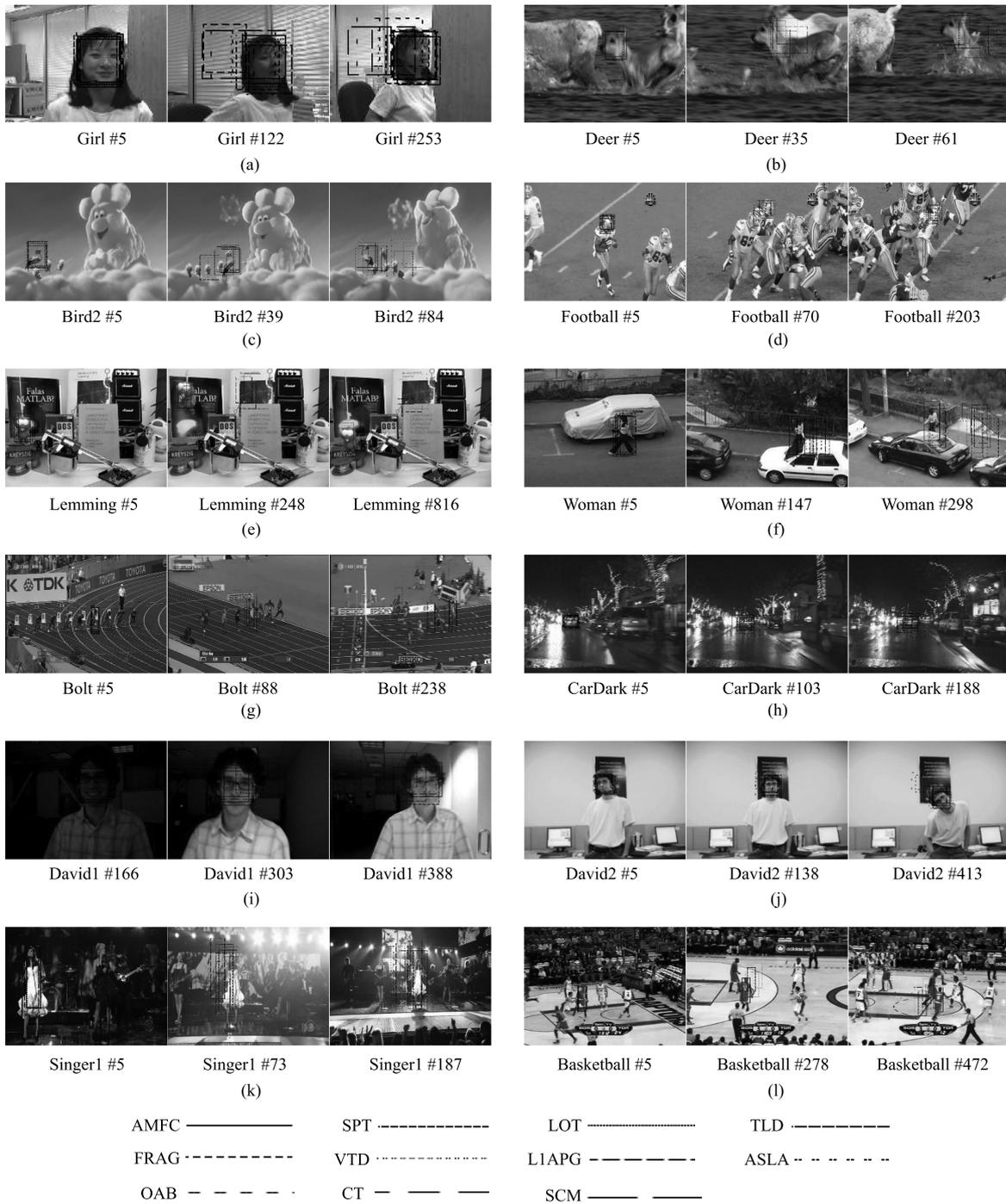


图 8 11 种跟踪算法在 12 组图像序列中的跟踪结果  
 Fig. 8 Tracking results of the 11 algorithms in the 12 image sequences

局部信息,并在更新外观模型时引入一种新的在线外观模型更新算法,对特征集加入适当特征补偿,使得在目标发生形变时外观模型能够更准确地描述目标.

图像序列 Football, Lemming 和 Woman 主要用于测试当目标发生遮挡时算法的跟踪效果. Football 序列的跟踪效果如图 8(d) 所示, Lemming 序

列的跟踪效果如图 8(e) 所示, Woman 序列的跟踪效果如图 8(f) 所示. 在 Football 和 Lemming 序列中,当图像中目标发生局部遮挡时,绝大多数跟踪算法都能较准确地跟踪目标. 在 Woman 序列中,目标发生严重遮挡时,大部分跟踪算法的跟踪重叠率都有所降低,但 AMFC 算法的跟踪效果较好. 与同类算法相比,AMFC 在前景区域判别时,总是寻找最

佳块匹配, 虽然发生严重遮挡 (甚至全局遮挡) 时会出现局部漂移, 但随着遮挡的减弱, 算法会快速重新跟踪目标, 避免了跟踪丢失现象. 此外, AMFC 算法引入决策判定, 只保留在该区间内的最佳匹配块, 避免前景区域中相似物体的干扰. 利用 EM 模型预测下一帧目标的前景信息, 即使目标与背景像素相近且运动较快时, 算法依然能够较好地估计目标位置, 并准确匹配目标的局部信息, 使得跟踪成功率较高.

图像序列 Bolt, Basketball 和 Deer 主要用于测试当目标处于复杂背景中算法的跟踪效果. Deer 序列的跟踪效果如图 8 (b) 所示, Bolt 序列的跟踪效果如图 8 (g) 所示, Basketball 序列的跟踪效果如图 8 (l) 所示. 在 Bolt 序列中, 当目标处于复杂背景且运动较快时, SCM, LOT, SPT 和 AMFC 算法的跟踪重叠率较高. 在 Basketball 序列中, 当目标处于复杂背景且发生局部遮挡时, 只有 VTD, SPT 和 AMFC 算法的跟踪重叠率较高. 在 Deer 序列中, 当目标处于复杂背景且背景颜色与目标相似时, 仅 ASLA, VTD 和 AMFC 算法的跟踪效果较好. 但综合这三组图像序列, AMFC 的跟踪效果最好, 平均重叠率为 0.69. 相比同类算法, AMFC 算法利用下一帧估计的前景信息作为约束条件, 防止复杂背景中与目标特征相似的物体干扰, 并在匹配过程中引入决策判定, 只保留在该区间内的最佳匹配块, 即使发生相似物体的遮挡, 算法依然能够较准确地匹配目标.

通过对算法准确性分析可以看出, 当场景变化较大时 (包括前/背景), AMFC 算法利用 EM 模型预测下一帧目标的前景信息, 即使目标与背景像素相近且较为复杂时, 算法估计目标在前景区域中的

位置, 并通过最优化模型匹配方法寻找在前景区域中的最佳块匹配, 此外, AMFC 算法引入决策判定, 只保留在该区间内的最佳匹配块, 避免前景区域中相似物体的干扰, 从而确定目标的最优位置. 在模型更新中对目标是否发生严重遮挡进行判断, 当发生严重遮挡时, 在特征集中加入适当补偿, 使更新的外观模型对目标的描述更加准确.

#### 4.4 跟踪效率分析

为了说明算法的实时性, 测试 11 种算法在处理 12 组图像帧序列时单位时间内算法跟踪的帧数, 即不同算法的平均运行速度, 结果见表 4.

通过本文算法与近几年目标跟踪算法的平均运行速度相比不难看出, 虽然 TLD、OAB、CT 算法的实时性较高, 但其跟踪成功率较低, 而 AMFC 算法的平均速度基本都在 3 帧左右, 跟踪重叠率较高. 相比较于同类算法, 在平均运行速度相当的前提下, 算法跟踪的平均中心误差较小、跟踪重叠率较高.

## 5 结论

本文提出了一种前景约束下的抗干扰匹配目标跟踪方法, 利用图像的前景判别, 突出了目标的重要区域, 为特定目标的识别跟踪提供了新的思路. 通过 EM 模型对前景信息的估计解决了模板匹配中的误识问题. 在匹配过程中, 引入决策判定, 排除与目标相似物体的干扰, 提高算法匹配的抗干扰性. 在模型更新中, 算法对目标是否发生严重遮挡进行判断, 当发生严重遮挡时, 在特征集中加入适当补偿, 使更新的外观模型对目标的描述更加准确, 提高了跟踪的可靠性.

表 4 不同跟踪算法的平均运行速度

Table 4 Average running speeds of different tracking algorithms

图像序列	ASLA	FRAG	SCM	VTD	L1APG	CT	OAB	TLD	LOT	SPT	AMFC
Girl	5.31	6.32	0.65	2.74	1.76	38.21	17.51	26.84	0.79	0.47	3.68
Deer	6.24	4.78	0.97	2.67	1.64	31.63	14.72	27.17	0.83	0.41	3.56
Bird2	5.74	5.43	0.67	2.58	1.49	27.06	9.94	26.58	0.65	0.56	3.09
Football	6.15	5.68	0.61	3.14	1.61	36.73	19.67	26.63	0.93	0.76	3.41
Lemming	6.78	6.27	0.69	2.77	1.75	28.15	10.46	26.70	0.71	0.37	3.39
Woman	8.45	6.41	0.57	2.16	1.53	32.42	11.31	26.32	0.66	0.43	4.37
Bolt	7.04	3.97	0.46	2.21	1.63	27.18	8.66	24.74	0.61	0.29	3.06
CarDark	7.23	4.23	0.48	2.49	1.74	26.79	10.35	25.13	0.59	0.36	2.95
David1	7.84	4.48	0.53	3.47	2.03	34.22	14.75	26.47	0.67	0.54	3.86
David2	5.87	5.25	0.48	2.68	1.44	36.36	16.68	25.89	0.73	0.66	2.97
Singer1	5.31	4.96	0.52	2.91	1.73	28.19	10.43	26.31	0.71	0.43	3.42
Basketball	7.91	6.23	0.89	2.34	2.04	25.81	9.09	24.53	0.62	0.34	2.89

通过对标准库中的 12 组视频图像序列进行实验对比分析, 结果表明, AMFC 算法在目标形变、目标旋转移位、光照变化、部分遮挡、复杂环境等干扰因素的影响下均可以得到较高的跟踪准确性。

AMFC 算法虽然引入了图像特征的前景信息, 但对于目标不规则运动和目标发生全局遮挡时跟踪的准确率不高. 在未来的工作中, 将针对目标的特征做进一步分析研究。

### 附录 A 对局部模型匹配过程的推导

$$p_{\Omega} \left( h_{pixel}^I, I_{pixel}, T_{pixel}, \tilde{I}_{pixel}, \tilde{T}_{pixel} \right) = p \left( h_{pixel}^I, I_{pixel}, T_{pixel}, \tilde{I}_{pixel}, \tilde{T}_{pixel} \right) 1_{\Omega}$$

其中,  $1_{\Omega} = \delta(\tilde{T}_{pixel} - \tilde{I}_{pixel})^{h_{pixel}^I}$  是狄克拉克函数. 由于像素的位置、颜色特征都是独立的, 上式可表示为

$$p_{\Omega} \left( h_{pixel}^I, I_{pixel}, T_{pixel}, \tilde{I}_{pixel}, \tilde{T}_{pixel} \right) = p \left( h_{pixel}^I \right) p \left( I_{pixel} | h_{pixel}^I \right) \times p \left( T_{pixel} \right) p \left( \tilde{I}_{pixel} | I_{pixel} \right) p \left( \tilde{T}_{pixel} | T_{pixel} \right) 1_{\Omega}$$

为了进一步简化, 对  $\tilde{I}_{pixel}$ ,  $\tilde{T}_{pixel}$  进行积分, 相应的推导过程如下:

$$p_{\Omega} \left( h_{pixel}^I, I_{pixel}, T_{pixel} \right) = \iint p_{\Omega} \left( h_{pixel}^I, I_{pixel}, T_{pixel}, \tilde{I}_{pixel}, \tilde{T}_{pixel} \right) d\tilde{I}_{pixel} d\tilde{T}_{pixel} = p \left( h_{pixel}^I \right) p \left( I_{pixel} | h_{pixel}^I \right) p \left( T_{pixel} \right) \times \iint G \left( \tilde{T}_{pixel} | T_{pixel}, \sigma \right) G \left( \tilde{I}_{pixel} | I_{pixel}, \sigma \right) \times \delta \left( \tilde{T}_{pixel} - \tilde{I}_{pixel} \right)^{h_{pixel}^I} d\tilde{I}_{pixel} d\tilde{T}_{pixel}$$

如果  $h_{pixel}^I = 0$ , 则上述公式的二重积分可分解为 2 个独立的高斯 CDF 积分. 如果  $h_{pixel}^I = 1$ , 二重积分可以合写为一个高斯分布, 即  $G(T_{pixel} - I_{pixel} | 0, \sqrt{2}\sigma)$ . 则

$$\iint G \left( \tilde{T}_{pixel} | T_{pixel}, \sigma \right) G \left( \tilde{I}_{pixel} | I_{pixel}, \sigma \right) \times \delta \left( \tilde{T}_{pixel} - \tilde{I}_{pixel} \right)^{h_{pixel}^I} d\tilde{I}_{pixel} d\tilde{T}_{pixel} = p \left( h_{pixel}^I \right) p \left( I_{pixel} | h_{pixel}^I \right) p \left( T_{pixel} \right) \times G \left( T_{pixel} - I_{pixel} | I_{pixel}, \sqrt{2}\sigma \right)^{h_{pixel}^I}$$

### 附录 B 寻找 LK 最优化匹配问题等价于搜索最大似然像素集问题

证明.

寻找与目标模型的最大似然匹配像素集  $I(W(PIXEL, w^{t-1} + \Delta w))$ . 即

$$\max_{\Delta w} \log p \left( I(W(PIXEL, w^{t-1} + \Delta w)) | T \right)$$

将上式代入到牛顿高斯模型中, 表示为

$$\max_{\Delta w} \sum_{pixel \in PIXEL} \log G \left( I(W(pixel, w^{t-1} + \Delta w)) - T | 0, \sigma \right)$$

对上述结果做进一步推导, 利用高斯表达式  $G(x | \mu, \sigma) = 1/\sqrt{2\pi}\sigma \exp(-1/2\sigma^2(x - \mu)^2)$  展开.

$$\max_{\Delta w} \sum_{pixel \in PIXEL} \log G \left( I(W(pixel, w^{t-1} + \Delta w)) - T | 0, \sigma \right) =$$

$$\max_{\Delta w} \sum_{pixel \in PIXEL} \log \left( \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{1}{2\sigma^2} \times \right. \right.$$

$$\left. \left. \left( I(W(pixel, w^{t-1} + \Delta w)) - T \right)^2 \right) \right) =$$

$$\max_{\Delta w} \sum_{pixel \in PIXEL} \log \left( \frac{1}{\sqrt{2\pi}\sigma} \right) +$$

$$\max_{\Delta w} \sum_{pixel \in PIXEL} -\frac{1}{2\sigma^2} \left( I(W(pixel, w^{t-1} + \Delta w)) - T \right)^2$$

其中,  $\max_{\Delta w} \sum_{pixel \in PIXEL} \log \left( \frac{1}{\sqrt{2\pi}\sigma} \right)$  为与  $w$  无关的常数, 由此可以看出, 寻找 LK 最优化匹配问题等同于搜索最大似然像素集问题.  $\square$

### 附录 C 光流迭代寻优过程

将  $\max_{\Delta w} \log p(I(W(pixel, w^{t-1} + \Delta w)) | T)$  进行一阶泰勒展开, 可得

$$L(\Delta w) = \sum_{pixel \in PIXEL} \left\{ Q_0(pixel) \left[ I_0(pixel, w) + \nabla I_0 \frac{dW(w)}{dw} \Delta w \right] + Q_1(pixel) \left[ I_1(pixel, w) + \nabla I_1 \frac{dW(w)}{dw} \Delta w \right] - Q(pixel) \left[ T(pixel) - I(W(pixel, w)) - \nabla I \frac{dW(w)}{dw} \Delta w \right]^2 \right\}$$

其中,

$$I_0 = \log p(h_{pixel}^I = 0)$$

$$I_1 = \log p(h_{pixel}^I = 1)$$

$$Q_0(pixel) = p_{old}(h_{pixel}^I(W(pixel, w) = 0))$$

$$Q_1(pixel) = p_{old}(h_{pixel}^I(W(pixel, w) = 1))$$

$$Q(pixel) = \frac{1}{2\sigma^2} p_{old}(h_{pixel}^I(W(pixel, w) = 1, I(W(pixel, w) - T(pixel)) = 0))$$

上述公式中,  $I_0$  表示背景图像区域,  $I_1$  表示前景图像区域,  $Q_0(pixel)$  为当前帧背景图像  $I_0$  的似然估计的权重函数,  $Q_1(pixel)$  为当前帧前景图像  $I_1$  的似然估计的权重函数. 第 1 行等式的含义为根据前一帧图像中匹配到的背景像素的权重 (由权重函数  $Q_0(pixel)$  给出) 获取背景图像  $I_0$  的似然估计. 第 2 行等式的含义根据前一帧图像中匹配到的前景像素的权重 (由权重函数  $Q_1(pixel)$  给出) 获取前景图像  $I_1$  的似然估计. 第 3 行等式为对输入图像  $I$  通过传统 LK 算法目标函数的相似性度量, 利用  $Q(pixel)$  加权获得模板与图像之间匹配关系。

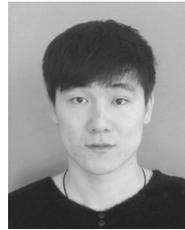
将  $L(\Delta w)$  求导, 设其导数为 0, 即可得到

$$\begin{aligned} \Delta w = & \left[ \sum_{pixel \in PIXEL} Q(pixel) [T(pixel) - \right. \\ & I(W(pixel, w))] \nabla I \frac{dW(w)}{dw} - \\ & \frac{1}{2} \left( \sum_{pixel \in PIXEL} Q_0(pixel) \nabla I_0 \frac{dW(w)}{dw} + \right. \\ & \left. \sum_{pixel \in PIXEL} Q_1(pixel) \nabla I_1 \frac{dW(w)}{dw} \right)] \times \\ & \left[ \sum_{pixel \in PIXEL} Q(pixel) \left( \nabla I_0 \frac{dW(w)}{dw} \right)^T \times \right. \\ & \left. \left( \nabla I_0 \frac{dW(w)}{dw} \right) \right]^{-1} \end{aligned}$$

## References

- Yin Hong-Peng, Chen Bo, Chai Yi, Liu Zhao-Dong. Vision-based object detection and tracking: a review. *Acta Automatica Sinica*, 2016, **42**(10): 1466–1489  
(尹宏鹏, 陈波, 柴毅, 刘兆栋. 基于视觉的目标检测与跟踪综述. 自动化学报, 2016, **42**(10): 1466–1489)
- Huang Dan-Dan, Sun Yi. Tracking via multitask discriminative local joint sparse appearance model. *Acta Automatica Sinica*, 2016, **42**(3): 402–415  
(黄丹丹, 孙怡. 基于判别性局部联合稀疏模型的多任务跟踪. 自动化学报, 2016, **42**(3): 402–415)
- Xu Jian-Qiang, Lu Yao. Robust visual tracking via weighted spatio-temporal context learning. *Acta Automatica Sinica*, 2015, **41**(11): 1901–1912  
(徐建强, 陆耀. 一种基于加权时空上下文的鲁棒视觉跟踪算法. 自动化学报, 2015, **41**(11): 1901–1912)
- Yuan X H, Kong L B, Feng D C, Wei Z C. Automatic feature point detection and tracking of human actions in time-of-flight videos. *IEEE/CAA Journal of Automatica Sinica*, 2017, **4**(4): 677–685
- Zoidi O, Tefas A, Pitas I. Visual object tracking based on local steering kernels and color histograms. *IEEE Transactions on Circuits and Systems for Video Technology*, 2013, **23**(5): 870–882
- Duffner S, Garcia C. Using discriminative motion context for online visual object tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016, **26**(12): 2215–2225
- Wang X C, Tuüretken E, Fleuret F, Fua P. Tracking interacting objects using intertwined flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(11): 2312–2326
- Khan Z H, Gu I Y H. Nonlinear dynamic model for visual object tracking on Grassmann manifolds with partial occlusion handling. *IEEE Transactions on Cybernetics*, 2013, **43**(6): 2005–2019
- Wang Mei-Hua, Liang Yun, Liu Fu-Ming, Luo Xiao-Nan. Object tracking based on component-level appearance model. *Journal of Software*, 2015, **26**(10): 2733–2747  
(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)
- Zhang Huan-Long, Hu Shi-Qiang, Yang Guo-Sheng. Video object tracking based on appearance models learning. *Journal of Computer Research and Development*, 2015, **52**(1): 177–190  
(张焕龙, 胡士强, 杨国胜. 基于外观模型学习的视频目标跟踪方法综述. 计算机研究与发展, 2015, **52**(1): 177–190)
- Shi Hua, Li Cui-Hua, Wei Feng-Mei, Wang Hua-Wei. Moving object tracking based on location and confidence of pixels. *Journal of Computer Research and Development*, 2005, **42**(10): 1726–1732  
(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基于外观模型学习的视频目标跟踪方法综述. 计算机研究与发展, 2005, **42**(10): 1726–1732)
- Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning. In: Proceedings of the 22nd IEEE International Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE, 2009. 983–990
- Oron S, Bar-Hillel A, Levi D, Avidan S. Locally orderless tracking. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012. 1940–1947
- Xu Ru-Yi, Chen Liang-Ying. Lucas-Kanade tracking based on sparse representation. *Journal of Image and Graphics*, 2013, **18**(3): 283–289  
(徐如意, 陈靓影. 稀疏表示的 Lucas-Kanade 目标跟踪. 中国图象图形学报, 2013, **18**(3): 283–289)
- Wang S, Lu H C, Yang F, Yang M H. Superpixel tracking. In: Proceedings of the 2011 IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 1323–1330
- Levinshtein A, Stere A, Kutulakos K N, Fleet D J, Dickinson S J, Siddiqi K, et al. Turbopixels: fast superpixels using geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(12): 2290–2297
- Baker Simon B, Matthews Iain M. Lucas-Kanade 20 years on: a unifying framework. *International Journal of Computer Vision*, 2004, **56**(3): 221–255
- Liu Wan-Jun, Liu Da-Qian, Fei Bo-Wen, Qu Hai-Cheng. Geometric active contour tracking based on locally model matching. *Journal of Image and Graphics*, 2015, **20**(5): 652–663  
(刘万军, 刘大千, 费博雯, 曲海成. 基于局部模型匹配的几何活动轮廓跟踪. 中国图象图形学报, 2015, **20**(5): 652–663)
- Comanicu D, Meer P. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(45): 603–619
- Wu Y, Lim J, Yang M H. Online object tracking: a benchmark. In: Proceedings of the 2013 IEEE International Conference on Computer Vision and Pattern Recognition. Oregon, Portland, USA: IEEE, 2013. 2411–2418

- 21 Jia X, Lu H C, Yang M H. Visual tracking via adaptive structural local sparse appearance model. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012. 1822–1829
- 22 Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 798–805
- 23 Zhong W, Lu H C, Yang M H. Robust object tracking via sparsity-based collaborative model. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012. 1838–1845
- 24 Kwon J, Lee K M. Visual tracking decomposition. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010. 1269–1276
- 25 Bao C L, Wu Y, Ling H B, Ji H. Real time robust  $L1$  tracker using accelerated proximal gradient approach. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012. 1830–1837
- 26 Zhang K H, Zhang L, Yang M H. Real-time compressive tracking. In: Proceedings of the 2012 European Conference on Computer Vision. Florence, Italy: Springer, 2012. 864–877
- 27 Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting. In: Proceedings of the 2006 British Machine Vision Conference. Edinburgh, UK: BMVA Press, 2006. 47–56
- 28 Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(7): 1409–1422
- 29 Gao Jun-Yu, Yang Xiao-Shan, Zhang Tian-Zhu, Xu Chang-Sheng. Robust visual tracking method via deep learning. *Chinese Journal of Computers*, 2016, **39**(7): 1419–1434  
(高君宇, 杨小汕, 张天柱, 徐常胜. 基于深度学习的鲁棒性视觉跟踪方法. 计算机学报, 2016, **39**(7): 1419–1434)
- 30 Li Qing-Wu, Zhu Guo-Qing, Zhou Yan, Huo Guan-Ying. Object compressive tracking via online feature selection. *Acta Automatica Sinica*, 2015, **41**(11): 1961–1970  
(李庆武, 朱国庆, 周妍, 霍冠英. 基于特征在线选择的目标压缩跟踪算法. 自动化学报, 2015, **41**(11): 1961–1970)



**刘大千** 辽宁工程技术大学电子与信息工程学院博士研究生. 2016 年获得辽宁工程技术大学硕士学位. 主要研究方向为图像与视觉信息计算, 目标的检测与跟踪. E-mail: liudaqianlntu@163.com  
(**LIU Da-Qian** Ph.D. candidate at the School of Electronic and Information Engineering, Liaoning Technical University. He received his master degree from Liaoning Technical University in 2016. His research interest covers image and visual information computing, target detection and tracking.)



**刘万军** 辽宁工程技术大学软件学院教授. 1991 年获得辽宁工程技术大学硕士学位. 主要研究方向为图像与视觉信息计算, 目标的检测与跟踪. 本文通信作者. E-mail: liuwanjun@lntu.edu.cn  
(**LIU Wan-Jun** Professor at the School of Software, Liaoning Technical University. He received his master degree from Liaoning Technical University in 1991. His research interest covers image and visual information computing, target detection and tracking. Corresponding author of this paper.)



**费博雯** 辽宁工程技术大学工商管理学院博士研究生. 2016 年获得辽宁工程技术大学硕士学位. 主要研究方向为数据挖掘, 模式识别.  
E-mail: feibowen2098@163.com  
(**FEI Bo-Wen** Ph.D. candidate at the School of Business and Management, Liaoning Technical University.

She received her master degree from Liaoning Technical University in 2016. Her research interest covers data mining and pattern recognition.)



**曲海成** 辽宁工程技术大学副教授. 2016 年获得哈尔滨工业大学博士学位. 主要研究方向为遥感高性能计算技术.  
E-mail: quhaicheng@lntu.edu.cn  
(**QU Hai-Cheng** Associate professor at the School of Software, Liaoning Technical University. He received his Ph.D. degree from Harbin Institute of Technology in 2016. His main research interest is remote sensing high performance computing technology.)