

## 引入视觉注意机制的目标跟踪方法综述

黎万义<sup>1</sup> 王鹏<sup>1</sup> 乔红<sup>2</sup>

**摘 要** 视觉跟踪在无人飞行器、移动机器人、智能监控等领域有着广泛的应用,但由于目标外观和环境的变化,以及背景干扰等因素的存在,使得复杂场景下的鲁棒实时的目标跟踪成为一项极具挑战性的任务.视觉注意是人类视觉信息处理过程中的一项重要心理调节机制,在视觉注意的引导下,人类能够从众多的视觉信息中快速地选择那些最重要、最有用、与当前行为最相关的感兴趣的视觉信息,特别地,人类能够快速指向感兴趣的的目标,从而可以轻松地实现对目标的稳定跟踪.因此,将视觉注意机制引入到复杂场景下的目标跟踪中,有利于实现更为稳定和接近于人类认知机制的视觉跟踪算法.本文旨在对引入了视觉注意机制的目标跟踪方法进行综述.首先,介绍了视觉注意的基本概念及其代表性的计算模型;其次,对视觉注意与跟踪的内在关系进行了阐述;然后,对引入视觉注意机制的目标跟踪方法进行归纳、总结和分类,对代表性的方法进行介绍和分析;最后,对该类方法的特点和优势进行了讨论,并对未来的研究趋势进行了展望.

**关键词** 目标跟踪,视觉注意,显著性,选择性注意,视觉认知

**引用格式** 黎万义,王鹏,乔红.引入视觉注意机制的目标跟踪方法综述.自动化学报,2014,40(4):561–576

**DOI** 10.3724/SP.J.1004.2014.00561

## A Survey of Visual Attention Based Methods for Object Tracking

LI Wan-Yi<sup>1</sup> WANG Peng<sup>1</sup> QIAO Hong<sup>2</sup>

**Abstract** Visual tracking has been widely used in numerous applications, such as unmanned aerial vehicles, mobile robots and intelligent visual surveillance. Robust and real-time object tracking in complex scenes is a challenge task. Difficulties in tracking objects can arise due to changing appearance patterns of both the object and the environment, as well as factors such as background interference. Visual attention is one of the key mechanisms of visual perception which directs the processing resources to the visual data of the potentially most relevant, specially directs our gaze rapidly towards objects of interest in our visual environment and as a result humans can easily achieve stable object tracking. Therefore introducing the visual attention mechanism to the object tracking in complex scenes, will facilitate the realization of stable and humanoid tracking algorithms. This paper aims to review the state-of-the-art of visual attention based methods for tracking. Firstly, we introduce the basic concepts of visual attention and its representative computational models. Secondly, the relationship between visual attention and tracking is described. Thirdly, the attention-based visual tracking algorithms are classified into five categories and detailed descriptions of representative methods in each category are provided, and their pros and cons are examined. Finally, we highlight the advantages of attention-based tracking methods and provide insights for future.

**Key words** Object tracking, visual attention, saliency, selective attention, visual cognition

**Citation** Li Wan-Yi, Wang Peng, Qiao Hong. A survey of visual attention based methods for object tracking. *Acta Automatica Sinica*, 2014, 40(4): 561–576

视觉跟踪是计算机视觉中的一项重要任务,是指对视频序列中的目标状态进行持续推断的过程<sup>[1]</sup>,

其任务在于通过在视频的每一帧中定位目标,以生成目标的运动轨迹,并在每一时刻提供完整的目标区域<sup>[2]</sup>.视觉跟踪技术在军事和民用方面都有着十分广泛的应用<sup>[1–6]</sup>,军事方面包括无人飞行器、精确制导、空中预警、战场监视等,民用方面包括移动机器人、智能视频监控、智能交通系统、人机交互、虚拟现实等.

学术界和工业界在目标跟踪方面开展了大量工作,并取得了一定的进展<sup>[1–5,7]</sup>.然而由于目标及场景的动态变化、遮挡、相似物干扰或者光线剧烈变化等复杂外界环境的干扰、目标的突变运动等因素的存在,使得复杂环境下鲁棒实时的目标跟踪仍然是一个具有挑战性的问题.

收稿日期 2013-04-26 录用日期 2013-09-22  
Manuscript received April 26, 2013; accepted September 22, 2013

国家自然科学基金(61210009, 61100098, 61379097)资助  
Supported by National Natural Science Foundation of China (61210009, 61100098, 61379097)

本文责任编辑 贾云得  
Recommended by Associate Editor JIA Yun-De

1. 中国科学院自动化研究所精密感知与控制研究中心 北京 100190
2. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190
1. Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing 100190
2. State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

研究表明,人类具有异常突出的数据筛选能力,而视觉注意机制是保证人眼高效率工作的重要机制。在人类视觉信息处理中,总是迅速选择少数几个显著对象进行优先处理,而忽略或舍弃其他的非显著对象,这使我们能够有选择地分配计算资源,从而极大地提高视觉信息处理的工作效率。人类能够轻松地实现对目标的稳定跟踪,视觉注意机制扮演了重要的角色。因此,将视觉注意机制引入到目标跟踪算法中,以提高跟踪的鲁棒性和实时性,具有重要的意义。

近年来,视觉注意的计算建模取得了巨大进展<sup>[8-13]</sup>,并应用于目标检测<sup>[14-15]</sup>、目标识别<sup>[16]</sup>和视频质量评估<sup>[17]</sup>等领域,使得将视觉注意机制引入到目标跟踪算法的研究具备了可行性,并迅速成为机器视觉领域的研究热点之一,国内外研究者们对此开展了大量的研究。本文旨在对基于视觉注意的目标跟踪方法进行综述和分析。

本文余下的内容组织如下:第1节对视觉注意的基本概念和与本文综述到的跟踪方法有关的代表性的视觉注意计算模型进行介绍;第2节阐述了视觉注意与跟踪的内在关系;第3节提出了针对“引入视觉注意机制的目标跟踪方法”的有效分类标准,并将具有代表性的方法进行了分类讨论和分析,同时指出了各类方法的优缺点和可能改进的方向,并对本文列举分析到的跟踪方法的应用范围及跟踪性能进行了整理;第4节讨论了“引入视觉注意机制的目标跟踪方法”的特点和优势,并对未来的研究趋势进行了展望。第5节为结束语。

## 1 视觉注意计算模型

注意是人类信息加工过程中一项重要的心理调节机制,是指选择性地指向环境中的某一方面而忽视其他事物的认知过程<sup>[18]</sup>,它能够对有限的信息加工资源进行分配,使感知具备选择能力。注意存在于人类的包括视觉、听觉在内的所有知觉中,对视觉而言,即为视觉注意,其作用在于从众多的视觉信息中快速地选择那些最重要、最有用、与当前行为最相关的感兴趣的视觉信息<sup>[10]</sup>,或者说将人类注意快速指向感兴趣的目标<sup>[8]</sup>。显然,视觉注意机制对人类是极其有用的,可以说,进化选择了视觉注意机制<sup>[9]</sup>,因为人类往往需要在同一时间应对庞大数量的视觉数据,以至于难以完全地进行详细处理,这就使得人们必须高效地在众多的视觉数据中按重要性进行优先选择。机器感知同样面临人类类似的挑战,它们必须处理大量的数据,从而需要选择最有用的部分。因此,视觉注意机制对机器感知而言,具有十分重要的借鉴意义。

人类视觉系统指导注意力分配的因素<sup>[10]</sup>有两

个:自下而上(Bottom-up)的注意和自上而下(Top-down)的注意,图1列出了这两种注意。自下而上的注意,也称为数据驱动的注意选择,仅受感知数据的驱动,将人的视点指导到场景中的显著区域。通常与周围具有较强对比度或与周围有明显不同的区域吸引自下而上的注意;自上而下的注意,也称做任务驱动的注意选择,由人的“认知因素”决定,比如知识、预期和当前的目标。

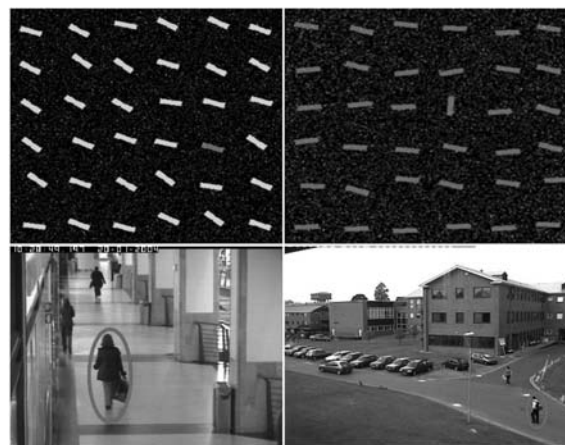


图1 指导视觉注意的两个因素(第1行为自下而上的注意,第1列浅灰色条和第2列的竖直接放的条形能立即引起人的注意(图片来自文献[19]);第2行为自上而下的注意,监控任务下,场景中的人体可引起注意(图片来自文献[20]中的视频序列“Walking1”和“Walking2”))

Fig. 1 Illustration of two factors that guide human attention (Items in the first row that attract bottom-up attention include the light gray bar among white bars and the vertical bar among horizontal bars (Images are from [19]). Pedestrians in the second row attract top-down attention (Images are from “Walking1” and “Walking2” sequences of [20]).)

在机器人和计算机视觉领域,研究者们对视觉注意机制的兴趣与日俱增,因为该机制可从大量的视觉数据中确定出最相关的部分。因此,近年来,研究者在特征综合理论<sup>[21]</sup>和 Guided search<sup>[22-23]</sup>等注意的心理学模型基础上,提出了大量的可计算的注意力选择模型,以用于模拟人类的视觉注意机制。这些模型包括:基于认知、贝叶斯、决策论、信息论、图模型、频域分析和基于模式分类的视觉注意模型,本文仅对较为典型,且与本文所综述的跟踪方法有关的注意计算模型进行介绍,包括认知注意模型、决策论注意模型、频域分析注意模型和图论注意模型,详细的综述可参阅文献[11]。

### 1.1 认知注意模型

实际上,几乎所有的注意模型都直接或间接地

受到认知概念的启发. 本节主要对那些和心理学或神经生理学研究结果有更多联系的模型进行介绍. 其中以 **Itti 的显著性模型** 最具代表性, 该模型已经成为了自下而上视觉注意模型的标准.

Itti 等<sup>[24]</sup> 于 1998 年, 根据早期原始视觉系统的行为和神经网络结构, 提出基于显著性的视觉注意模型, 并在 2001 年度 *Nature* 上对该模型理论作了进一步的完善<sup>[8]</sup>. 其基本结构如图 2 所示.

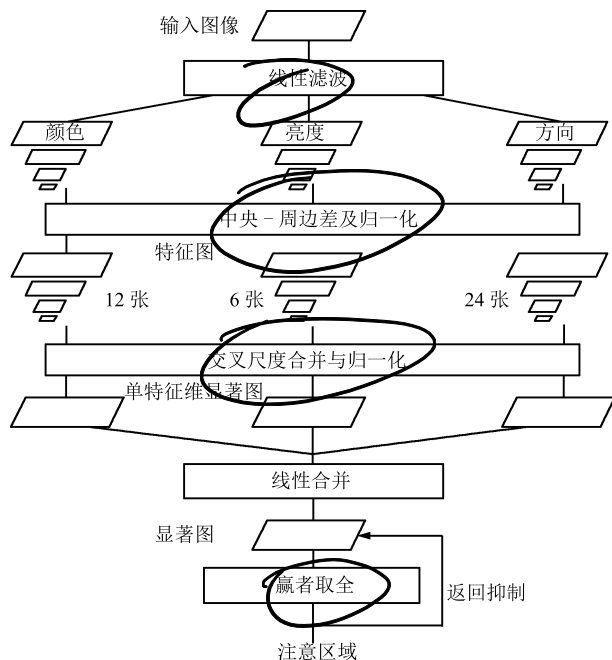


图 2 Itti 的显著性视觉注意模型结构图  
(英文版原图来自文献 [24])

Fig. 2 Architecture of Itti's saliency-based visual attention model (The figure is adapted from [24].)

对于一幅输入的图像, 该模型提取初级视觉特征: 颜色、亮度和方位, 在多种尺度下使用中央周边 (Center-surround)<sup>[24]</sup> 操作产生体现显著性度量的特征图, 将这些特征图合并得到最终的显著图 (Saliency map) 后, 利用生物学中赢家取全 (Winner-take-all) 的竞争机制得到图像中最显著的空间位置, 用来引导注意位置的选取, 最后采用返回抑制 (Inhibition of return) 的方法来完成注意焦点的转移.

目前为止, 该基本模型至少有四种实现: Itti 等<sup>[24]</sup> 的 iNVT、Walther<sup>[25]</sup> 的 Saliency Toolbox (STB)、Frintrop<sup>[26]</sup> 的用于视觉物体检测的计算注意系统 (Visual object detection with a computational attention system, VOCUS) 和 Harel<sup>[27]</sup> 实现的 Matlab 代码. 此外, 在文献 [28] 中, 通过加入运动和闪烁对比扩展该模型, 以应用到视频领域.

## 1.2 决策论注意模型

决策论的观点认为, 不断进化的感知系统能产生关于周围环境的、在决策论意义下的最优决策. 要点在于视觉注意应与当前任务有关的最优性所驱动.

Gao 等提出的判别显著性模型<sup>[29]</sup> 是最具代表性的决策论注意模型之一. 该模型把显著性问题看作是两类视觉刺激的最优决策, 这两类刺激为: 感兴趣的刺激和包含了不显著刺激的背景或零假设. 显著性的计算通过确定一个二分类问题得以实现. 视野内每个位置的显著性等同于一个视觉特征集在该位置上的判别力 (即期望分类误差). 具体而言, 给定特征集  $F = \{F_1, \dots, F_d\}$ , 位置  $l$ , 类别标签  $C$ ,  $C_l = 0$  对应于取自环绕区域的样本,  $C_l = 1$  对应于取自以  $l$  为中心的较小中心区域的样本, 然后显著性的评价对应于特征集与标签的互信息, 计算为  $I(F, C) = \sum_{i=1}^n I(F_i, C)$ , 互信息也可解释为两个类的特征响应分布之间的距离度量. 对于一个具体位置或点而言, 其期望分类误差越小 (即互信息越大), 显著值就越大.

决策论注意模型既能表达自下而上的注意, 也能表达自上而下的注意. 并已经在计算机视觉中得到了成功的应用, 如分类和注意定位预测, 均取得很高的准确率.

## 1.3 频域分析注意模型

与其他在空间域对图像进行处理的模型不同, 这类模型在频域上推导图像的显著性. Hou 等<sup>[30]</sup> 率先提出第一个通过频谱分析进行显著性检测的方法 — 谱残差法 (Spectral residual, SR), 随后 Guo 等<sup>[31]</sup> 提出相位谱傅里叶变换 (Phase spectrum of Fourier transform, PFT) 法; Achanta 等<sup>[32]</sup> 提出用于显著区域检测的频率调谐法 (Frequency-tuned); Li 等<sup>[33]</sup> 提出了超复数傅里叶变换法 (Hypercomplex Fourier transform, HFT). SR 和 PQFT 都只使用了相位信息, HFT 既利用了幅度信息, 也利用了相位信息.

Hou 等<sup>[30]</sup> 提出的谱残差法, 基于“相似隐含冗余”的思想. 他们提出, 频谱域上的统计奇异对应图像的异常区域, 在该区域, 原型物体变得显著. 因此从图像幅度谱中减去对幅度谱进行局部均值滤波的结果后, 结合原始相位, 利用傅里叶逆变换即得到空域显著图.

在 SR 方法提出后, Guo 等<sup>[31]</sup> 发现, 直接丢弃输入图像的幅度谱而只保留相位谱信息能得到更好的显著性预测, 该方法称为相位谱傅里叶变换. 随后, 他们将 PFT 进行扩展, 利用四元数傅里叶变换将亮度、颜色和运动信息组合在一起并行处理, 该方法



称为相位谱四元数傅里叶变换法 (Phase spectrum of quaternion Fourier transform, PQFT), 用于计算时空显著性, 并应用到视频序列。

Achanta 等<sup>[32]</sup> 提出频率调谐法 (Frequency-tuned) 用于显著区域检测, 该方法将输入的 RGB 图像转换到 CIELab 颜色空间并进行高斯平滑, 再减去图像特征向量的算术平均后, 按点求幅度即得到显著图。

Li 等<sup>[33]</sup> 提出了超复数傅里叶变换法。通过对信号进行频域分析发现, 幅度谱上的尖峰 (Spike) 对应原始信号中的重复模式, 进而可以通过对幅度谱进行低通滤波——谱滤波 (Spectral filtering, SF) 达到抑制重复模式, 突出显著信号的作用。该文表明, “图像的幅度谱与一个低通高斯核之间的卷积操作, 相当于图像显著性检测器”。使用在合适尺度进行低通滤波后的幅度谱和原始相位谱, 进行傅里叶逆变换重建的二维信号, 即为图像的显著图。与 Guo 等<sup>[31]</sup> 的做法类似, 他们将谱滤波进行扩展, 利用超复数表示图像的多维特征, 使用超复数傅里叶变换计算时空显著性。值得一提的是, 该文表明, SR 和 PFT 均是该方法的特例, 相当于谱滤波的尺度参数趋于无穷时的情况; 频率调谐法<sup>[32]</sup> 在一定程度上也是该方法的特例, 相当于谱滤波的尺度参数比较小时, 和均值对应的成分得到了明显抑制。谱滤波方法以自下向上的方式计算图像的显著图, 并取得了很好的效果。但在任务驱动的应用场景下, 自上向下的指导起着关键性的作用, 为此, Li 等<sup>[34]</sup> 提出了一种基于谱滤波的自上而下的时空显著性检测方法。

基于频谱分析的显著性模型, 形式简洁, 易于解释和实现, 并且在注意焦点预测和显著区域检测方面取得了很大的成功, 由于基于快速傅里叶变换实现, 能够满足实时要求, 和 iNVT 类似的模型相比, 运算速度可提高近 10 倍。美中不足的是, 其生物合理性不是非常清楚。

#### 1.4 图论注意模型

图模型是一个使用图对随机变量之间的条件依赖结构进行表示的概率框架。这种类型的注意模型, 把眼动看作一个时间序列。由于有大量的隐变量影响眼球运动的产生, 因此, 该类注意模型使用了隐马尔科夫模型、动态贝叶斯网和条件随机场等方法。

Liu 等<sup>[35]</sup> 使用一系列新颖的特征并采用条件随机场整合这些特征, 用于显著物体检测; Harel 等<sup>[27]</sup> 提出了基于图的视觉显著性 (Graph-based visual saliency, GBVS), 使用图结构表示图像之间的像素关联, 并用马尔科夫链计算显著性。

图模型可以对复杂的注意机制建模, 因此能取得较好的预测能力, 缺点在于模型的高复杂度, 尤其

涉及训练和可读性时。

## 2 视觉注意与跟踪的关系

生物视觉领域的研究表明: 注意在人类的目标跟踪中扮演了十分重要的角色, 主要体现在目标维持、错误恢复、目标增强和干扰抑制上; 另外, 计算机视觉领域的研究者对“显著性与跟踪的关系”提出目标跟踪的显著性假说<sup>[36]</sup>——“跟踪是通过判别性的中央-周边显著性机制进行自上而下调节完成的, 该调节基于目标的特征”, 并进行了一定程度的证实。

### 2.1 注意在跟踪中的角色

对于人类视觉系统而言, 目标跟踪极其依赖视觉注意<sup>[37-38]</sup>。Yantis<sup>[37]</sup> 声称, 注意在跟踪过程中, 涉及了视觉目标的维持。文献 [39] 指出, 注意过程通过参与跟踪过程, 或作为一种错误恢复机制, 起到促进跟踪的作用 (另见文献 [40])。

Makovski 等<sup>[41]</sup> 指出, 人类视觉的跟踪依赖于目标外观特征的辨别性, 外观特征对跟踪的性能具有增强作用, 这种增强作用是基于特征的, 从而体现了跟踪中的一定程度的特征绑定。Doran 等<sup>[42]</sup> 提供的证据表明, 视觉注意在跟踪中的作用主要体现在对目标的增强和对干扰的抑制上, 尤其在干扰物稠密的情况下。

### 2.2 跟踪的显著性假说

Mahadevan 等<sup>[36, 43]</sup> 假设跟踪是通过判别性的中央-周边显著性机制进行自上而下调节完成的, 该调节基于目标的特征。文献 [44] 中把这一假设称为“目标跟踪的显著性假说”, 并给出了三个断言, 如果“目标跟踪的显著性假说”成立, 则这三个断言必然成立: 1) 显著目标的跟踪可靠性比非显著目标的高; 2) 跟踪可靠性依赖于显著性的定义变量: 特征对比度和干扰物异质性, 并类似于显著性对这些变量的依赖; 3) 显著性和跟踪均可使用普通的低级神经机制实现。前两个断言通过一系列的关于“显著性和跟踪的关系”的人类行为研究结果进行证实。第三个断言的正确性通过构造一个普通的具有“神经心理学合理性”的架构进行说明。该架构可计算地求解显著性和跟踪, 并符合 1) 初级视觉皮层 (V1) 和后顶叶皮层区 (MT) 的标准生理学模型; 2) 外侧顶内沟区 (LIP) 的注意控制, 同时, 解释了人类行为实验的结果。

此外, 部分研究者将视觉跟踪问题归结为前景和背景的最大可分性问题<sup>[45-46]</sup>, 通过选取显著性跟踪特征, 将显著的目标区域加强, 将非显著的背景区域削弱, 从而将目标从背景区域中分离出来, 获得了稳定的跟踪效果。

由于注意机制在人类的跟踪过程中所起的重要作用, 以及近年来视觉注意计算模型取得的进展, 学术界对使用视觉注意的跟踪方法进行了大量研究. 下一节对这些方法进行分类介绍与评述.

### 3 引入视觉注意的目标跟踪方法

本节首先给出基于视觉注意的跟踪方法的几个分类标准. 分类标准与跟踪方法涉及的注意计算方式有关, 或者与所计算得到的显著图、提取的显著区域或显著点在跟踪中所起的作用有关.

#### 1) 自下而上 vs. 自上而下

注意力的分配依赖于自下而上的因素还是自上而下的影响或者是两者兼而有之. 定义见第 1 节.

#### 2) 空间 vs. 时空

估计显著性时仅使用空间信息还是时空信息都使用.

#### 3) 基于空间 vs. 基于物体 (或特征)

所使用的注意模型, 其注意的单元是空间位置还是物体 (或特征). 由于视觉特征与物体有着密切的关系, 因此把基于特征的和基于物体的归为一类.

#### 4) 估计显著性时所使用的特征

该标准指明计算显著性时使用哪些特征. 常用的特征有颜色、亮度、方位、轮廓等静态特征, 以及深度、物体尺寸、物体运动等动态特征.

#### 5) 注意计算模型的输出在跟踪中的作用

注意计算模型的输出通常为指明注意力分布情况的显著图、提取出的显著区域集或显著点集. 显著图、显著区域集、显著点集在跟踪算法中常见的作用包括: 目标检测或搜索; 目标外观表示; 作为一种加权方式; 作为一种特征增加到原来跟踪框架中, 如结合别的指标中用于训练样本的选择、衡量粒子的权重.

鉴于大多数基于视觉注意的跟踪方法需衡量图像中像素点级或区域级的显著性, 因此接下来按照获得显著性的机制对这些方法进行分类介绍, 包括: 1) 基于认知注意模型的跟踪; 2) 基于决策论注意模型的跟踪; 3) 基于频域分析注意模型的跟踪; 4) 基于图论注意模型的跟踪; 5) 基于其他注意模型的跟踪.

图 3 给出了本文所综述的跟踪方法的层次图. 表 1 给出了按本节提到的几个标准对它们进行分类的情况.

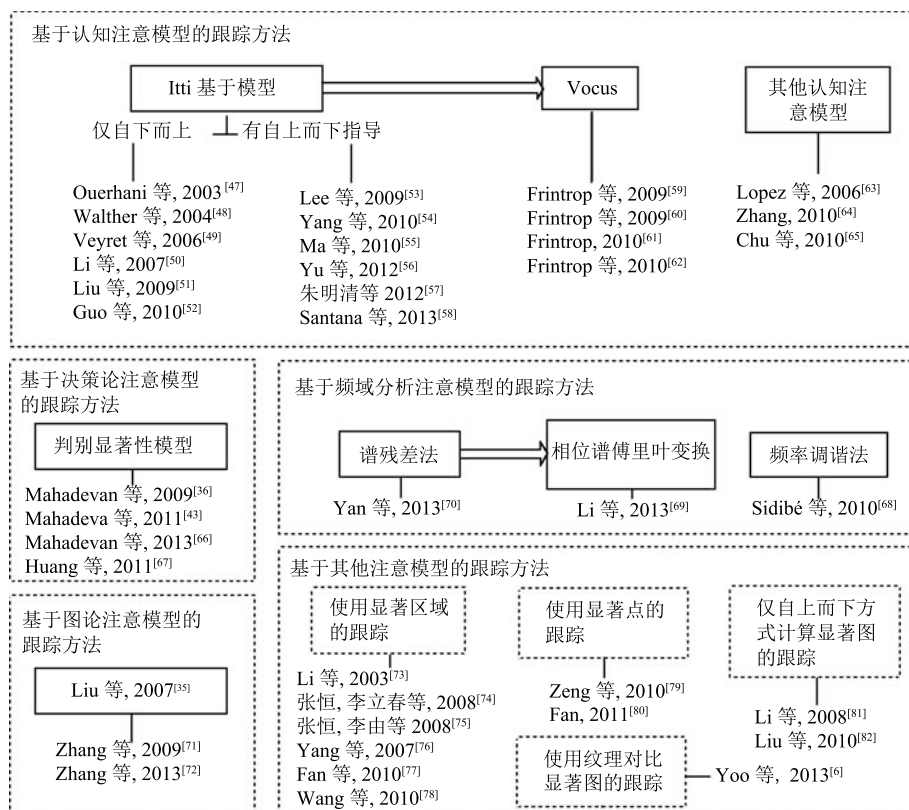


图 3 基于视觉注意机制的跟踪方法层次图 (实心矩形表示注意计算模型)

Fig. 3 A hierarchical illustration of visual attention based tracking methods (Solid rectangles show computational models of visual attention.)

表 1 引入视觉注意的跟踪方法一览表  
Table 1 Summary of visual attention based methods for object tracking

| 序号                 | 跟踪方法  | 自下而上 | 自上而下 | 空间 (-)/<br>时空 (+) | 基于空间 (+)/基于<br>物体或特征 (-) | 估计显著性时所用特征   | 注意计算模型的输出<br>在跟踪中的作用  |
|--------------------|---|------|------|-------------------|--------------------------|--|-----------------------|
| 1) 基于认知注意模型的跟踪方法   |   |      |      |                   |                          |  |                       |
| 1                  | Ouerhani 等, 2003 <sup>[47]</sup>  | +    | -    | +                 | +                        | 颜色, 亮度, 方向, 运动   | 显著区域检测                |
| 2                  | Walther 等, 2004 <sup>[48]</sup>   | +    | -    | -                 | +                        | 颜色, 亮度, 方向   | 目标检测                  |
| 3                  | Veyret 等, 2006 <sup>[49]</sup>  | +    | -    | +                 | +                        | 颜色, 亮度, 方向, 运动   | 目标检测                  |
| 4                  | Li 等, 2007 <sup>[50]</sup>  | +    | -    | +                 | -                        | 运动   | 运动区域检测                |
| 5                  | Liu 等, 2009 <sup>[51]</sup>   | +    | -    | +                 | -                        | 颜色, 亮度, 运动   | 目标检测                  |
| 6                  | Guo 等, 2010 <sup>[52]</sup>   | +    | -    | +                 | +                        | 颜色, 亮度, 运动 KL 变换分布 (KLT)   | 运动物体检测                |
| 7                  | Lee 等, 2009 <sup>[53]</sup>   | +    | +    | +                 | -                        | 两个静态特征 (亮度与色彩), 三个三维<br>态特征 (深度, 物体尺寸, 物体运动)                                 | 目标检测                  |
| 8                  | Yang 等, 2010 <sup>[54]</sup>  | +    | +    | -                 | +                        | 颜色, 亮度, 运动   | 目标检测                  |
| 9                  | Ma 等, 2010 <sup>[55]</sup>  | +    | +    | -                 | +                        | 颜色, 亮度, 方向   | 目标搜索                  |
| 10                 | Yu 等, 2012 <sup>[56]</sup>  | +    | +    | -                 | -                        | 颜色, 亮度, 方向, 轮廓   | 目标检测                  |
| 11                 | 朱明清等, 2012 <sup>[57]</sup>  | -    | +    | -                 | +                        | 目标的颜色直方图反向投影   | 目标全局搜索                |
| 12                 | Santana 等, 2013 <sup>[58]</sup>   | +    | +    | -                 | +                        | 颜色, 亮度   | 小路检测                  |
| 13                 | Frintrop 等, 2009 <sup>[59]</sup>  | +    | +    | -                 | +                        | 颜色, 亮度, 方向   | 目标搜索                  |
| 14                 | Frintrop 等, 2009 <sup>[60]</sup>  | +    | +    | -                 | +                        | 颜色, 亮度, 方向   | 作为粒子滤波的观测<br>模型       |
| 15                 | Frintrop 2010 <sup>[61]</sup> ; Frin-<br>trop 等, 2010 <sup>[62]</sup>   | +    | +    | -                 | +                        | 颜色, 亮度   | 目标外观表示                |
| 16                 | López 等, 2006 <sup>[63]</sup>   | +    | -    | +                 | -                        | 灰度, 形状, 运动   | 目标检测与分割               |
| 17                 | Zhang, 2010 <sup>[64]</sup>   | -    | +    | -                 | -                        | 稀疏基函数 (Sparse basis function)  | 特征选择                  |
| 18                 | Chu 等, 2010 <sup>[65]</sup>   | +    | +    | -                 | -                        | 颜色   | 目标检测                  |
| 2) 基于决策论注意模型的跟踪方法  |   |      |      |                   |                          |  |                       |
| 19                 | Mahadevan 等, 2009 <sup>[36]</sup> ; +<br>Mahadevan 等, 2011 <sup>[43]</sup> ;<br>Mahadevan 等, 2013 <sup>[66]</sup> | +    | +    | -                 | -                        | 离散余弦变换 (DCT)   | 特征选择, 目标检测            |
| 20                 | Huang 等, 2011 <sup>[67]</sup>   | +    | +    | -                 | -                        | 颜色, 边缘信息   | 关键点过滤和显著区<br>域检测      |
| 3) 基于频域分析注意模型的跟踪方法 |   |      |      |                   |                          |  |                       |
| 21                 | Sidibé等, 2010 <sup>[68]</sup>   | +    | -    | -                 | +                        | 高斯差分函数 (DOG)   | 参与衡量粒子的权重             |
| 22                 | Li 等, 2013 <sup>[69]</sup>  | +    | -    | -                 | -                        | 颜色, 亮度, 方向   | 原型物体检测                |
| 23                 | Yan 等, 2013 <sup>[70]</sup>   | +    | -    | -                 | +                        | 傅里叶变换 (FFT)  | 结合别的指标选择训<br>练样本      |
| 4) 基于图论注意模型的跟踪方法   |   |      |      |                   |                          |  |                       |
| 24                 | Zhang 等, 2009 <sup>[71]</sup> ;<br>Zhang 等, 2013 <sup>[72]</sup>  | +    | +    | +                 | -                        | 静态特征: 中央-周边直方图, 多尺度对<br>比, 颜色空间分布; 运动特征: 中央-周边<br>运动直方图, 多尺度运动对比, 运动空间<br>分布 | 显著物体检测; 粒子<br>滤波的观测模型 |
| 5) 基于其他注意模型的跟踪方法   |   |      |      |                   |                          |  |                       |
| 25                 | Li 等, 2003 <sup>[73]</sup>  | +    | -    | -                 | -                        | 颜色, 方向   | 显著区域检测                |
| 26                 | 张恒等, 2008 <sup>[74]</sup> ; 张恒等,<br>2008 <sup>[75]</sup>  | +    | -    | -                 | -                        | 颜色   | 作为一种加权方式              |
| 27                 | Yang 等, 2007 <sup>[76]</sup>  | +    | -    | -                 | +                        | 直方图  | 目标表示                  |
| 28                 | Fan 等, 2010 <sup>[77]</sup>   | +    | -    | -                 | +                        | 文脉流 (Contextual flow)  | 目标表示                  |
| 29                 | Wang 等, 2010 <sup>[78]</sup>  | +    | -    | -                 | +                        | 运动, 空间位置   | 目标检测                  |
| 30                 | Zeng 等, 2010 <sup>[79]</sup>  | +    | +    | +                 | -                        | 颜色, 亮度, 方向, 运动   | 显著区域检测                |
| 31                 | Fan 2011 <sup>[80]</sup>  | +    | -    | -                 | +                        | 像素点的文脉特征 (Context feature)   | 目标表示                  |
| 32                 | Li 等, 2008 <sup>[81]</sup>  | -    | +    | -                 | +                        | 颜色, 位置和亮度  | 人脸定位                  |
| 33                 | Liu 等, 2010 <sup>[82]</sup>   | +    | +    | +                 | -                        | 颜色反向投影, 运动   | 目标定位和搜索               |
| 34                 | Yoo 等, 2013 <sup>[6]</sup>  | +    | -    | -                 | +                        | 纹理   | 参与衡量粒子的权重             |

### 3.1 基于认知注意模型的跟踪方法

这类跟踪方法所涉及的注意模型受到认知概念的启发, 具有生物机制上的合理性. 本小节对与 Itti 的显著性模型相关、与 VOCUS 相关、或使用的注意计算模型与认知概念联系比较密切的跟踪方法进行介绍.

#### 3.1.1 基于 Itti 注意模型的跟踪

基于 Itti 基本注意模型的跟踪包括两种: 1) 仅按自下而上的方式计算显著图, 并用于显著区域<sup>[47]</sup>或运动区域检测<sup>[50, 52]</sup>、目标检测<sup>[48-49, 51]</sup>; 2) 对 Itti 基本模型进行扩展, 增加了自上而下的指导信息, 生成的显著图用于目标检测<sup>[53-54, 56, 58]</sup>或目标搜索<sup>[55, 57]</sup>.

首先对基于 Itti 的基本显著性模型以自下而上的方式计算显著图的跟踪方法进行介绍. Ouerhani 等<sup>[47]</sup>使用动态视觉注意, 对自然环境下的显著区域进行跟踪, 在 Itti 模型基础上, 增加了运动显著性, 通过注意模型提供的信息, 确定最具判别性的特征后, 用于后续的跟踪; Walther 等<sup>[48]</sup>使用 Itti 显著性模型进行目标检测, 卡尔曼滤波进行运动估计, 从而完成水下目标的跟踪; Veyret 等<sup>[49]</sup>对水族馆里的运动目标, 如鱼等, 进行检测和跟踪.

Li 等<sup>[50]</sup>针对对象和摄像机不规则运动, 提出了基于运动显著性检测机制的跟踪方法. 特定位置的运动显著性, 体现为中央区域的平均运动与环绕区域的平均运动的差别性, 通过图像局部邻域的灰度梯度张量秩亏进行衡量. 使用类似于 Itti 基本模型的金字塔式的采样获得中央区域与周边区域. 结合目标对象的多个空间特征来实现基于区域的跟踪, 跟踪结果由目标模板与在运动显著图中检测到的感兴趣区域之间的匹配误差决定.

Liu 等<sup>[51]</sup>把视觉目标跟踪问题视为人类视觉注意力的转移过程. 显著图通过 Itti 的基本模型计算得到. 早期的注意选择过程用来定位显著物体, 晚期的注意过程用于转移注意力, 以达到检测已跟踪丢失目标的目的.

Guo 等<sup>[52]</sup>通过合并运动显著特征、静态显著特征和 KL 变换 (Karhunen-Loeve transform, KLT) 特征计算显著图, 用于运动检测, 并结合核密度估计方法进行运动轨迹跟踪. 其中静态显著特征的计算和 Itti 的基本模型类似, 使用对比度和信息密度产生.

以上这些方法均基于自下而上的视觉注意, 因此它们只能跟踪场景里的显著区域, 即那些在颜色或灰度上具有很高对比度的物体或者运动显著的区域才能被跟踪到.

下面给出在 Itti 基本模型基础上增加了指导信

息的跟踪方法.

Lee 等<sup>[53]</sup>将视觉注意用于交互式虚拟环境下的物体跟踪. 该方法结合了自下而上的预注意特征和自上而下的上下文信息, 上下文信息通过对用户的时空行为进行推断得到. 根据计算得到的显著图, 从若干候选物体中确定最合理的注意物体. 本质上, 自下而上的显著图通过使用两个静态图像特征 (亮度、色度) 和三个三维动态特征 (深度、大小和物体运动) 扩展 Itti 的基本模型计算得到. 显著图用于物体检测, 该注意跟踪框架用于虚拟环境的细节层次管理 (Level of details, LOD), 效果得到了明显改善.

Yang 等<sup>[54]</sup>针对人体运动分析, 提出基于视觉注意和多线索融合的跟踪算法. 使用 Itti 的基本模型计算得到自下而上显著图, 然后使用目标相关的知识指导从视觉显著区域中筛选出有效的辅助性目标. 从而提高算法的协作跟踪的鲁棒性. 目标的最终位置通过辅助性物体与目标的竞争与结合得到.

Ma 等<sup>[55]</sup>提出了结合自上而下的注意和自下而上的注意的目标跟踪算法. 自下而上模块使用 Itti 的基本模型, 将图像分解成不同的特征图, 在学习阶段, 使用 Logistic 回归分析得到各特征图的权重系数, 以强化目标特征并抑制背景特征. 在跟踪阶段, 使用训练得到的权重系数进行自上而下的调节, 生成显著图. 根据生成的显著图搜索目标, 以满足一定条件的最显著区域作为最终的跟踪结果.

Yu 等<sup>[56]</sup>使用基于物体的注意机制用于单目标跟踪. 算法包括四个模块: 预注意分割、自上而下注意选择、后注意处理和目标模型在线学习. 该方法要点在于: 在目标模型的学习阶段, 使用 Itti 的基本模型估计目标在各个特征上的显著性, 依据显著性均值大小得出任务相关的显著特征, 以此作为显著性组件; 同时估计目标的外观组件; 在自上而下的注意阶段, 以显著性组件所体现的显著特征作为自上而下的指导, 根据图像中已分割的原型物体与目标外观组件在显著特征上的距离衡量最终的显著性, 选择最为显著的原型物体作为注意物体. 该方法对于杂乱和动态变化的环境下的单目标跟踪, 取得了较好的效果. 该方法的主要缺点在于, 依赖对众多参数的调整, 包括在线学习的参数和预注意分割模块的相似性阈值, 参数的调整会对算法的性能造成影响.

朱明清等<sup>[57]</sup>模拟人类视觉跟踪方式, 建立了与之对应的视觉目标跟踪框架, 提出一种自上而下的视觉注意机制的计算模型. 以目标的颜色直方图特征到图像的反向投影图为出发点, 使用和 Itti 基本模型类似的步骤, 建立了相应的视觉显著图, 对其进行多尺度下熵密度分析, 得到视觉注意力焦点区域; 该方法以检测视觉注意力焦点区域的方法来模拟人

类视觉系统的全局搜索过程,用粒子滤波跟踪来模拟人类视觉系统的局部跟踪过程,兼具局部性跟踪算法和全局性跟踪算法的优点,特别是在丢失目标之后能够自动恢复对目标的跟踪。

Santana 等<sup>[58]</sup>将基于群的视觉显著性用于小路检测与跟踪。自下而上的显著性,使用 Itti 的基本模型进行计算,得到颜色和亮度显著图(Conspicuity map),自上而下的注意,使用物体的整体布局作为主要线索,控制群体中个体的行为,计算信息素图,合并后得到最终的显著图,并用于小路位置的估计。

### 3.1.2 基于 VOCUS 的跟踪方法

Frintrop 等针对不需要长时间的训练阶段,可在线学习的,能在移动平台上实时工作的跟踪方法进行了研究。他们的跟踪方法基于 Frinrop 研究实现的注意系统—VOCUS<sup>[26]</sup>。该注意系统有两种执行模式:1)无指定任务的探测模式;2)有指定目标物体的搜索模式。在探测模式下,感兴趣区域通过特征的强烈对比(如颜色对比或亮度对比)和独特性定义。在搜索模式下,系统使用先前已学习到目标物体信息,进行对目标有偏的显著性计算,从而能够对目标物体进行主动搜索。

这类跟踪方法使用 VOCUS 计算显著性,注意计算模型的输出用于目标搜索<sup>[59]</sup>、作为粒子滤波的观测模型<sup>[60]</sup>或者目标外观表示<sup>[61-62]</sup>。下面按时间顺序对他们的系列工作进行介绍。

基于 VOCUS 视觉注意系统,Frinrop 和 Kessel 等<sup>[59]</sup>提出了最显著区域(Most salient region, MSR)跟踪器,目标外观通过对单帧图像进行快速在线学习获得,在后续帧中,物体的最显著部分通过自上而下信息指导的注意模块检测,从而完成目标的跟踪。同年,Frinrop 和 Königs 等<sup>[60]</sup>针对移动平台的人体跟踪,使用 VOCUS 作为观测模型并将其结合到粒子滤波算法中。

Frinrop<sup>[61]</sup>提出了一种具有认知观测模型、基于粒子滤波的跟踪方法。该方法的关键要素是一种能灵活地俘获目标的结构和外观的、基于部件的描述子。该描述子中的部件是基于颜色的,其计算方式源自 VOCUS,能从单个训练图像快速学习得到。他们还将该技术应用到移动机器人的人体跟踪中<sup>[62]</sup>,该工作可以看作是对文献[60]的扩展。

当背景和(或)目标外观发生剧烈变化时,目标描述子需要具备在线适应能力。Borji 等<sup>[83]</sup>给出了一种能让目标描述子具备适应能力的方法,该方法首先从训练序列中推导出一些类以及和这些类对应的目标表示。接下来对测试序列的每一帧图像,确定其最近的类,然后将对应的目标描述子用于该帧。此

工作的扩展版本见文献[84]。

Frinrop 等研究的系列跟踪方法,优点在于:1)能灵活地获得目标的外观和(或)结构;2)刻画目标外观和(或)结构的描述子能从单个训练图像快速学习得到,不需要长时间的训练阶段,从而能处理不同的目标物体和环境;3)能实时运行,可用于移动平台。但仍有进一步提升的空间:1)对于所使用的基于部件的描述子,如果部件的位置信息得以利用,目标的更为灵活的变形和旋转将能得到处理;2)当目标和背景的外观发生剧烈变化时,目标描述子应具有在线的适应能力。

### 3.1.3 基于其他认知注意模型的跟踪

除了以上基于 Itti 基本模型或使用 VOCUS 注意模型的跟踪方法之外,还有一些与认知心理学或神经生理学的研究结果关系比较紧密的跟踪方法。已有的这类方法把视觉注意机制用于目标检测<sup>[63, 65]</sup>或特征选择<sup>[64]</sup>。López 等<sup>[63]</sup>提出了一个用于在实际视觉监控场景下分割出车辆和行人的动态视觉注意模型。该系统受到人类视觉系统的启发,对局部特征(像素级)和全局特征(物体)结合永久响应的充电放电机理和侧向相互作用机制进行抽取和整合。注意模型的输出用于目标检测与分割。Zhang 等<sup>[64]</sup>使用基于特征的视觉注意机制进行目标跟踪过程中的特征选择,把从大量自然图像片断中学习得到的稀疏基函数作为特征,选择那些不增加系统熵的特征加入目标特征集,目标最终表示为相关特征集的概率分布。Chu 等<sup>[65]</sup>结合自下而上的注意和自上而下的注意用于机器人视觉跟踪。自下而上将场景分割成地面区域和显著区域,在自上而下注意的指导下,从显著区域中分离出物体区域和障碍物区域。

基于认知注意模型的跟踪,其优点在于:具有很好的生物合理性,也取得了不错的跟踪效果。不过,这类注意计算模型通常计算复杂度仍显较高,从而影响跟踪的实时性,需要在计算效率上作进一步的提高。此外,目前为止,视觉注意的生物机制并没有完全研究清楚,尤其是自上而下的任务驱动型注意的机制及计算原理,而目标跟踪是任务驱动的,因此,需要对自上而下的任务驱动型视觉注意计算模型展开进一步的研究,并将其应用到目标跟踪中,以进一步提升目标跟踪的性能。

### 3.2 基于决策论注意模型的跟踪方法

这类跟踪方法以基于判别显著性原理的为主,统称为判别显著性跟踪器。使用注意模型计算得到的显著图用于跟踪器的初始化、判别性特征选取和目标检测。

Mahadevan 等<sup>[36]</sup>基于判别性显著性原理<sup>[29]</sup>,提出了生物激发的跟踪框架。对每一帧图像,从背景



中区分出目标的问题被看成一个二分类问题. 从一组描述目标和背景的特征中, 使用最大化边际多样性原理 (Maximum marginal diversity, MMD), 选出最富含信息的一个子集, 用于分类. 在下一帧中, 通过使用这些特征计算的自上而下的显著图, 得出目标位置. 从而完成跟踪算法的一次迭代. 该文还对该框架进行了一个简单的扩展, 通过将运动特征包含到自下而上的显著图中, 对跟踪器进行自动初始化. 将提出的方法与当时最好的方法在一系列数据集上进行了比较, 该跟踪器取得了较好的性能.

Mahadevan 等<sup>[43]</sup> 对文献 [36] 中的基于显著性的判别跟踪算法作了进一步的阐述, 文中提出假设: 目标初始化和跟踪是指导视觉注意的显著性机制的结果, 该假设在文献 [44] 中被称为“跟踪的显著性假设”. 在这一假设下, Mahadevan 等<sup>[43]</sup> 提出了一个用于自动跟踪的基于判别显著性原理<sup>[29]</sup> 的跟踪框架. 在该框架中, 跟踪器的自动初始化通过使用运动特征的自下而上的显著性完成, 将跟踪问题表示为连续的目标-背景分类问题, 由使用显著性的两个阶段实现. 第一阶段为学习阶段, 结合注意焦点机制 (Focus of attention) 和自下而上的显著性, 确定用于目标检测的具有最大判别力的特征集; 第二阶段为检测阶段, 使用基于特征的注意机制和目标调节的自上而下的判别显著性检测器, 进行目标的检测. 总的来说, 跟踪器每次迭代执行操作: 1) 从当前图像帧的目标位置学习判别特征; 2) 在下一帧图像中检测目标的位置. 实验结果表明, 该框架取得了较好的性能.

Mahadevan 等<sup>[66]</sup> 继续对以上工作进行扩展, 说明了跟踪可以通过三个部件的组合而完成: 用于学习的自下而上的中央-周边判别显著性和空间注意、用于特征选择的基于特征的注意、用于目标检测的自上而下的显著性; 利用该原理, 构造了一个简单且计算高效的跟踪框架, 该框架与已经知道的生物视觉中的注意机制一致, 并提供了用于解决分类器设计、目标检测、跟踪器自动初始化和尺度自适应的整体方案. 该方法能自动进行跟踪器初始化, 并可以自适应尺度变化. 实验结果表明, 该框架和别的方法相比, 取得了较好的性能.

此外, Huang 等<sup>[67]</sup> 提出一种对主动视频中的感兴趣目标进行自动检测和跟踪的方法. 通过计算各颜色特征与类别标签之间的互信息得到颜色显著图后, 使用颜色显著性对边界概率图加权进行关键点过滤和显著区域检测. 该方法可以对感兴趣目标进行自动检测和可靠跟踪.

这类方法在取得较好性能的同时, 也存在一些不足, 包括: 1) 由于没有显式的保持在以前图像帧中出现的目标特征, 因此, 不能处理持续时间较长的

遮挡; 2) 由于其依赖于寻找能从背景中区分出目标的特征, 因此不适用于背景中存在有与目标相似的物体以及背景不当的大目标的跟踪; 3) 对于给定的位置, 其中心区域和环绕区域的不同尺度会导致不同的结果. 可能的改进思路包括: 1) 增加体现了稳定的目标特征的验证模块, 以判断注意区域是否确实为目标, 以应对遮挡的情况; 2) 充分利用运动信息, 通过运动信息实现相似物体的跟踪.

### 3.3 基于频域分析注意模型的跟踪方法

这类方法通过对频域进行分析得到显著图, 显著图用于参与衡量粒子权重<sup>[68]</sup>、物体检测<sup>[69]</sup> 或者训练样本的选择<sup>[70]</sup>.

Sidibe<sup>[68]</sup> 将视觉显著性信息结合到粒子滤波框架下. 显著图通过频率调谐的显著区域检测方法<sup>[32]</sup> 计算得到. 提出的方法结合颜色分布和显著性分布, 并自动衡量它们对似然函数的贡献. 实验表明, 结合视觉显著性后的方法在性能上超过了仅使用颜色作为特征的方法, 尤其对遮挡和光线变化的情况更为鲁棒.

Li 等<sup>[69]</sup> 提出了一个基于原型物体的生物启发的视觉跟踪框架. 对给定的图像序列, 先计算显著图: 使用颜色、亮度和方向三种初级特征 (提取方式同文献 [24]), 采用相位谱傅里叶变换 (PFT)<sup>[31]</sup>, 在不同尺度上计算特征显著图后合并成最终的显著图. 并对图像进行分割, 结合显著图、分割结果和话题模型 (Topic model), 检测得到原型物体, 然后基于原型物体的空间信息和显著信息, 对目标进行跟踪. 在提出的贝叶斯方法中, 统一地对目标和原型物体的状态进行估计. 在跟踪过程中, 使用了吉布斯采样优化估计. 该方法具有较好的鲁棒性, 在文中的实验中能处理遮挡, 干扰和光线变化. 不过, 由于显著图以自下而上的方式进行, 当所有原型物体特征丢失或较长时间不显著时, 该方法可能会失效.

Yan 等<sup>[70]</sup> 结合视觉显著性检测和 Online boosting 方法进行样本选择, 以用于鲁棒的在线目标跟踪. 其主要思想是: 结合当前跟踪位置和图像的显著性检测, 自适应地选择可靠的正样本和负样本, 以避免分类器受到不可靠样本的影响. 使用谱残差分析<sup>[30]</sup> 进行显著性检测. 该方法可以在目标位置不太准确的情况下, 有效避免漂移问题. 在颇具挑战性上的视频序列上与最新的方法进行了大量的对比实验, 实验结果表明, 提出的方法更为鲁棒和稳定.

使用频域分析注意模型的跟踪, 这类方法在计算显著图的部分, 运算速度快, 易满足实时性要求.

### 3.4 基于图论注意模型的跟踪方法

这类方法使用条件随机场等方式估计显著图,

显著图通常用于物体检测或作为粒子滤波的观测模型。

Zhang 等<sup>[71]</sup> 利用视觉注意机制来检测粒子的视觉显著性并作为粒子的权重, 以实现显著物体在线跟踪。目标的外观通过目标与背景的差异进行(即显著性)描述。显著性的计算方式: 使用一系列静态显著特征和运动显著特征用于描述显著物体, 这些显著特征按三种方式进行计算: 局部、区域和全局方式。系统初始化时, 通过学习一个条件随机场对这些特征进行组合, 以用于显著物体检测(文中实验参考了通过条件随机场已学习过的权值, 直接设置显著特征的权重)。跟踪时, 在线适应性地进行选择 and 组合, 并使用粒子滤波跟踪检测到的显著物体。该方法能跟踪任意的显著物体, 并对光照变化和局部遮挡鲁棒。此外, 注意可以自动重建, 而无需重检测。

Zhang 等<sup>[72]</sup> 提出了一种基于视觉显著性的在线关键物体发现与跟踪系统。将该问题表达为条件随机场上的二分类问题, 并用粒子滤波进行求解。还提出了上下文敏感度度量, 用于改善静态/动态显著图的准确性。改善后的显著图为关键物体所处位置提供了更为清晰的指示。基于良好的显著线索, 结合考虑时空上下文, 跟踪结果包围盒内的关键物体能得到进一步的分割。该方法的原始显著图的计算沿用了文献[35]和文献[85]的做法。实验表明, 在显著图质量和关键物体跟踪准确性上得到了明显的改善。

已有的该类显著物体跟踪方法<sup>[71-72]</sup> 在下面两个方面有改进的空间。1) 关于跟踪一致性。由于仅考虑前后两帧之间显著物体的相似性, 没有显式地维持第一帧检测到的显著物体, 因此不能较好地处理遮挡, 具体来说, 当前面跟踪的显著物体被部分遮挡时, 跟踪器会选择别的显著物体; 或者遮挡一段时间, 再次出现时, 跟踪器已经锁定别的显著物体, 而忽视该目标, 这种工作模式适用于视频摘要, 但不适用于机器人与环境的互动。针对此问题, 可以考虑作适当调整, 处理如下跟踪问题: 检测第一帧中的显著物体并持续跟踪该物体。2) 总是假设只能一个显著目标, 不符合许多实际场景。可以考虑对多个显著物体进行检测和持续跟踪, 研究目标进入、消失、合并、分离等情况的处理方法。

### 3.5 基于其他注意模型的跟踪方法

由于基于注意机制的跟踪属于比较新的研究领域, 因此有许多不能归入以上几类的方法, 本节对这些方法进行介绍。这些方法或使用显著区域进行跟踪<sup>[73-78]</sup>, 或使用显著点进行跟踪<sup>[79-80]</sup>, 或将完全以自上而下方式计算的显著图用于跟踪<sup>[81-82]</sup>, 或将使用纹理对比计算得到的显著图用于跟踪<sup>[6]</sup>。

对于使用显著区域的跟踪, 区域的显著性通常

使用该区域与其背景距离来衡量。Li 等<sup>[73]</sup> 提出用于在视频中自动检测和跟踪显著区域的方法。该方法基于使用方向 Gabor 滤波器和 K-means 方法得到的方向特征和颜色特征, 检测出显著区域, 然后对这些区域进行一小段时间的跟踪。张恒等针对复杂背景下运动目标跟踪问题, 使用显著性作为一种加权方式, 分别与最小二乘图像匹配<sup>[74]</sup> 及 Mean shift<sup>[75]</sup> 结合。显著性的计算方式为: 使用 K-means 将目标分割为不同区域, 然后计算各区域和背景的差别大小。Yang 和 Fan 等<sup>[76-77]</sup> 提出了使用显著区域进行目标表示的跟踪框架, 在他们的工作中, 目标通过一组注意区域进行表示, 基于低层特征使用自下而上方式检测出一系列具有较好局部判别力的区域作为注意区域后, 对这些区域进行跟踪, 目标的位置通过所有可用的显著区域的加权组合得到。Wang 等<sup>[78]</sup> 提出一种基于显著性选择的跟踪方法, 使用目标区域内的一组显著图像块和它们的空间上下文(相对于目标中心的位置信息)进行外观建模, 并在线更新, 图像块的显著性通过该图像块与其背景之间的概率分布距离来衡量。

使用显著区域进行跟踪, 其优点在于能在一定程度上处理外观变化、伪装环境及相似物干扰。不足之处在于, 通常依赖大量显著区域的共同出现<sup>[76]</sup>, 计算代价较为昂贵, 此外, 目标必须足够大<sup>[77]</sup>。

使用显著点的跟踪, 把目标表示为一系列显著点, 目标的跟踪通过显著点集的跟踪得到实现。Zeng 等<sup>[79]</sup> 通过提取易于跟踪的特征点, 并计算这些特征点的运动信息。综合考虑显著特征点的运动速度、方向和空间位置, 对特征点进行编组。从而得到基于物体的注意, 结合人体模型知识, 通过对目标组在当前帧的预测位置与实际存在的组进行匹配, 从而达到跟踪的目的。Fan 等<sup>[80]</sup> 定义了一种新的显著点, 该显著点包含了所给点的上下文信息, 从嵌在图像特征空间的流形结构上提取。实验表明, 该上下文特征点在许多具有挑战性的目标跟踪任务上取得了较好的性能。这种基于注意机制的思想所设计的显著点, 作为一种稳定新颖的图像特征, 具有很好的推广潜力。

完全以自上而下方式计算显著图的跟踪, 显著性通过图像特征分量信息与该特征上的先验分布之间的距离来体现, 或者计算为颜色直方图的反向投影。Li 等<sup>[81]</sup> 提出一种基于面部显著图的人脸分割算法。面部显著图的计算以自上而下的方式进行, 以测试图像在颜色、位置和亮度各个分量上的信息与先验分布的距离为依据计算显著图, 建立的人脸显著图用于对人脸区域进行定位, 最终实现基于跟踪的人脸分割。Liu 等<sup>[82]</sup> 将视觉注意用于处理机器人跟踪过程中摄像头快速运动和抖动问题。注意分布图

由颜色概率图和已分割运动区域作点对点相乘得到, 其中颜色概率图计算为判别性的颜色特征的反向投影. 计算得到的显著图用于正常情况下的目标定位和发生突变运动时的全局目标搜索.

Koo 等提出一种结合显著性与基于颜色的粒子滤波的飞行器跟踪方法<sup>[6]</sup>. 基本思想为: 采用判别方向一致性 (Discriminative directional consistency, DDC) 作为显著性, 以局部方式去补偿跟踪过程中的推断误差, 显著性参与衡量粒子的权重. 给定位置的显著性通过纹理对比进行计算, 具体而言, 显著性体现为判别方向一致性, 计算为能体现主方向及其能量的结构张量的两个特征值之差的平方.

为了让读者对引入视觉注意机制的跟踪方法有一个直观的印象, 我们对本文列举分析到的跟踪方法的应用范围及跟踪性能进行了整理, 如表 2 所示.

本文综述到的“引入视觉注意的目标跟踪方法”, 其应用范围主要涉及:

1) 军事应用, 如飞行器跟踪<sup>[6, 74-75]</sup>;

2) 机器人, 这类应用中需要跟踪的对象通常为如人体<sup>[59-60, 65, 82]</sup>、小路<sup>[58]</sup>和常规物体<sup>[61]</sup>等作为机器人视觉反馈的目标.

以上两种场景下的跟踪难点主要有快速突变运动、动态背景、目标外观变化以及要求较高的跟踪精度和实时性.

3) 视频监控<sup>[50, 52, 56, 63]</sup>, 这类应用下跟踪对象通常为人体、车辆等. 这类视频的特点包括: 遮挡, 光线及尺度变化引起的目标外观变化等.

4) 增强现实<sup>[49]</sup>及虚拟环境<sup>[53]</sup>下的物体跟踪.

除以上应用外, 本文所综述的方法中, 相当一部分没有具体指定应用场景, 仅针对某些特点的视频进行处理. 对这些方法, 我们根据原文中使用的实验视频, 对视频特点进行了分析并列入了表 2 中. 跟踪性能主要包括跟踪精度和运行速度 (实时性), 量化指标计算为相应跟踪方法在对应文献中报道的所有视频上的平均值, 运行速度体现为处理帧率, 并列出了所使用的硬件资源. 下面给出表 2 中使用到的跟踪精度指标.

1) F 测度

$$F_{\alpha} = \frac{(1 + \alpha) \cdot \text{precision} \cdot \text{recall}}{\alpha \cdot \text{precision} + \text{recall}}, \quad \alpha = 0.5 \quad (1)$$

式 (1) 中, *precision* 为精度, *recall* 为召回率.

2) 平均跟踪误差 1

$$\epsilon = \frac{1}{T} \sum_t \left( \frac{\sum_{i,j} G_{i,j}^t (1 - B_{i,j}^t)}{\sum_{i,j} G_{i,j}^t} \right) \quad (2)$$

3) 平均跟踪误差 2

$$\epsilon = \frac{1}{T} \sum_t \left( 1 - \frac{|G_t \cap B_t|}{|G_t \cup B_t|} \right) \quad (3)$$

式 (2) 和式 (3) 中, 跟踪误差定义为跟踪结果区域与真实目标区域的未重叠率, 值为 0 时表示完全重叠, 值为 1 时表示无重叠.  $G^t$  为第  $t$  帧真实目标区域的外接矩形,  $B^t$  为跟踪结果的外接矩形. 式 (2) 和式 (3) 定义跟踪误差时有所不同, 两式的大括号中的分子均与“跟踪结果区域  $B^t$  与真实目标区域  $G^t$  的交集区域面积”有关, 不同之处在于, 式 (2) 的分母仅考虑  $G^t$ , 而式 (3) 的分母为  $G^t$  和  $B^t$  的并集区域的面积.

4) 跟踪错误率

$$E = \frac{1}{T} \left( \sum_{t=1}^T \frac{|rx_t - gx_t|}{w_t} + \sum_{t=1}^T \frac{|ry_t - gy_t|}{h_t} \right) \quad (4)$$

式 (4) 中,  $(rx_t, ry_t)$  为跟踪结果的目标中心位置坐标,  $(gx_t, gy_t)$  为真实目标区域的中心位置坐标,  $w_t, h_t$  为真实目标区域的宽度和高度; 位置误差为估计的目标位置与真实目标位置的空间距离, 以像素作计量单位.

5) 重叠率

$$S_t = \frac{|r_t \cap r_t^{GT}|}{|r_t \cup r_t^{GT}|} \quad (5)$$

式 (5) 中,  $r_t^{GT}$  为第  $t$  帧真实目标区域,  $r_t$  为跟踪结果区域,  $\cap$  和  $\cup$  表示两个区域的交和并,  $|\cdot|$  表示区域中的像素个数.

从表 2 中可以看出, “引入视觉注意机制的目标跟踪方法”整体上取得了不错的跟踪效果. 一些方法取得了很高的精度<sup>[36, 43, 56, 58, 66, 69-70, 76-77]</sup>, 原文报道和当时最好的方法相比, 跟踪精度有明显提高; 部分方法<sup>[6, 53, 58-62, 76-77, 82]</sup>在保证较高精度的同时可满足或基本满足实时性要求.

## 4 总结与展望

人类视觉系统能够轻易完成复杂场景下的可靠持续的目标跟踪, 视觉注意机制起到了很大的作用. 将注意机制引入到计算机视觉跟踪技术中, 可以使其得到很大程度地改进和完善. 从综述的文献中可以看出, 引入注意机制的目标跟踪具有很多优点, 同时, 有许多值得进一步深入研究的问题, 本节对该类跟踪方法的优势和未来方向进行总结与展望.

表 2 本文综述的跟踪方法的应用范围及跟踪性能

Table 2 Application and tracking performance of visual attention based tracking methods

| 序号                 | 跟踪方法  | 应用范围   | 精度  | 实时性 (fps)                   | 软件平台    | 硬件资源  |
|--------------------|---|--|---|-----------------------------|---------|---|
| 1) 基于认知注意模型的跟踪方法   |   |  |   |                             |         |   |
| 1                  | Ouerhani 等, 2003 <sup>[47]</sup>  | 自然场景下显著区域跟踪  | 定性  | —                           | —       | —   |
| 2                  | Walther 等, 2004 <sup>[48]</sup>   | 水下动物检测与跟踪  | 跟踪成功率: 80 %   | 3 fps, 720 × 480 pixels     | C++     | 8 dual Xeon 2.4 GHz servers   |
| 3                  | Veyret 等, 2006 <sup>[49]</sup>  | 增强现实, 水族馆鱼类跟踪  | 定性  | 5 fps, 320 × 240 pixels     | C++     | —   |
| 4                  | Li 等, 2007 <sup>[50]</sup>  | 室外监控, 运动人体跟踪, 视频特点: 目标和相机的不规则运动                          | 定性  | —                           | —       | —   |
| 5                  | Liu 等, 2009 <sup>[51]</sup>   | 带遮挡的人体跟踪   | 跟踪成功率: 93.1 %   | —                           | —       | —   |
| 6                  | Guo 等, 2010 <sup>[52]</sup>   | 视频监控, 人体、车辆跟踪  | 定性  | —                           | —       | —   |
| 7                  | Lee 等, 2009 <sup>[53]</sup>   | 虚拟环境下的物体跟踪, 细节层次管理                                       | 定性  | 30 fps                      | OpenGL  | 3.2 GHz Pentium 4 PC, GPU 加速  |
| 8                  | Yang 等, 2010 <sup>[54]</sup>  | 人体跟踪   | 跟踪成功率: 60.95 %  | —                           | —       | —   |
| 9                  | Ma 等, 2010 <sup>[55]</sup>  | 人体跟踪   | 定性  | —                           | —       | —   |
| 10                 | Yu 等, 2012 <sup>[56]</sup>  | 视频监控, 人体跟踪   | 跟踪成功率: 94.49 %  | —                           | —       | —   |
| 11                 | 朱明清等, 2012 <sup>[57]</sup>  | 带遮挡的跟踪   | 跟踪成功率: 75.56 %  | —                           | —       | —   |
| 12                 | Santana 等, 2013 <sup>[58]</sup>   | 移动机器人, 小路检测与跟踪   | 跟踪成功率: 97 %   | 20 fps, 640 × 480 pixels    | C++     | Core2 Duo 2.53 GHz P8700  |
| 13                 | Frintrop 等, 2009 <sup>[59]</sup>  | 移动机器人物体跟踪  | 跟踪成功率: 88 %   | 30 fps, 320 × 240 pixels    | —       | 1.66 GHz PC   |
| 14                 | Frintrop 等, 2009 <sup>[60]</sup>  | 移动机器人人体跟踪  | 跟踪成功率: 73 %   | 15 fps, 320 × 240 pixels    | —       | 2 GHz dual core PC onboard the robot  |
| 15                 | Frintrop 2010 <sup>[61]</sup> ; Frin-trop 等, 2010 <sup>[62]</sup>   | 移动机器人常规物体跟踪 <sup>[61]</sup><br>移动机器人人体跟踪 <sup>[62]</sup> | 跟踪成功率: 81 % <sup>[61]</sup> , 跟踪成功率: 94 % <sup>[62]</sup> | 12.5 fps; 15 fps            | —       | 2.5 GHz dual core PC <sup>[61]</sup> 2 GHz dual core PC <sup>[62]</sup> onboard the robot |
| 16                 | López 等, 2006 <sup>[63]</sup>   | 视觉监控, 车辆和行人分割  | 分割精度: 83.19 %, 分割精度定义为: 注意焦点与实际目标区域交集像素数/实际目标区域像素数        | —                           | —       | —   |
| 17                 | Zhang 2010 <sup>[64]</sup>  | 常规物体跟踪, 部分遮挡, 伪装和姿态变化                                    | 定性  | 27 fps, 320 × 240 pixels    | —       | Standard PC with a 3.0 GHz processor and 1 GB memory                                      |
| 18                 | Chu 等, 2010 <sup>[65]</sup>   | 移动机器人目标跟踪  | 定性  | 5 ~ 8 fps, 752 × 480 pixels | —       | Intel(R) Core(TM) Duo CPU T2300 1.66 GHz, Memory: 256 M                                   |
| 2) 基于决策论注意模型的跟踪方法  |   |  |   |                             |         |   |
| 19                 | Mahadevan 等, 2009 <sup>[36]</sup> ; Mahadevan 等, 2011 <sup>[43]</sup> ; Mahadevan 等, 2013 <sup>[66]</sup> | 视频特点: 外观剧烈变化   | 平均跟踪误差 1: 0.073; 平均跟踪误差 2: 0.15; 平均跟踪误差 3: 0.35           | 1.5fps; 1.5fps; —           | C; C; — | —; A standard PC; A standard PC   |
| 20                 | Huang 等, 2011 <sup>[67]</sup>   | 主动视频中的人, 车辆跟踪  | 定性  | —                           | —       | —   |
| 3) 基于频域分析注意模型的跟踪方法 |   |  |   |                             |         |   |
| 21                 | Sidibé等, 2010 <sup>[68]</sup>   | 视频特点: 背景混淆和光线变化  | 跟踪成功率: 88.61 %, 式(5) 定义的重叠率大于 0.25 时, 认为跟踪成功.             | —                           | —       | —   |
| 22                 | Li 等, 2013 <sup>[69]</sup>  | 人体跟踪, 视频特点: 遮挡, 背景杂乱和光线变化                                | 跟踪错误率: 0.19   | 1fps, 384 × 288 pixels      | Matlab  | Pentium4 computer with 1GB of RAM   |
| 23                 | Yan 等, 2013 <sup>[70]</sup>   | 视频特点: 遮挡, 背景杂乱, 目标外观变化                                   | 平均位置误差: 7.83 像素   | —                           | —       | —   |
| 4) 基于图论注意模型的跟踪方法   |   |  |   |                             |         |   |
| 24                 | Zhang 等, 2009 <sup>[71]</sup> ; Zhang 等, 2013 <sup>[72]</sup>   | 显著物体跟踪, 人体, 车辆, 视频特征: 外观变化                               | 定性 <sup>[71]</sup> ; F 测度 <sup>[72]</sup> : 0.7           | —                           | —       | —   |
| 5) 基于其他注意模型的跟踪方法   |   |  |   |                             |         |   |
| 25                 | Li 等, 2003 <sup>[73]</sup>  | 常规视频中的显著区域检测与跟踪  | 定性  | —                           | —       | —   |
| 26                 | 张恒等, 2008 <sup>[74]</sup> ; 张恒等, 2008 <sup>[75]</sup>   | 飞机跟踪; 飞机跟踪   | 定性; 定性  | 32.5fps; 53.8fps            | —       | —   |
| 27                 | Yang 等, 2007 <sup>[76]</sup>  | 常规物体长时跟踪, 视频特点: 快速运动, 遮挡, 伪装, 尺度光照变化                     | 平均位置误差: 1 像素  | 10 ~ 15fps                  | C++     | Pentium-IV 3 GHz PC   |
| 28                 | Fan 等, 2010 <sup>[77]</sup>   | 视频特点: 背景干扰, 遮挡, 旋转, 尺度变化                                 | 平均位置误差: < 10 像素   | 15 fps, 320 × 240 pixels    | C++     | Pentium 3 G PC  |
| 29                 | Wang 等, 2010 <sup>[78]</sup>  | 人体和一般物体跟踪, 视频特点: 遮挡, 姿态, 尺度变化和相机运动                       | 同上  | —                           | —       | —   |
| 30                 | Zeng 等, 2010 <sup>[79]</sup>  | 行人跟踪   | 定性  | 原文称可实时                      | —       | —   |
| 31                 | Fan 2011 <sup>[80]</sup>  | 人体, 动物跟踪, 实验视频特点: 视角变化, 背景干扰, 快速运动                       | 定性  | —                           | —       | —   |
| 32                 | Li 等, 2008 <sup>[81]</sup>  | 人脸分割与跟踪  | 定性  | —                           | —       | —   |
| 33                 | Liu 等, 2010 <sup>[82]</sup>   | 移动机器人目标跟踪, 视频特点: 突变运动                                    | 跟踪成功率: 89 %   | 14fps                       | —       | 1.6 GHz AMD Turion Dual CPU and 1 GB memory   |
| 34                 | Yoo 等, 2013 <sup>[6]</sup>  | 飞行器跟踪, 视频特点: 突变运动: 动态背景, 目标变形, 姿态变化, 不可预知的自运动            | 平均欧氏距离: 6.6 像素  | 7fps                        | Matlab  | Low-end PC  |



#### 4.1 引入注意机制的跟踪方法的优势

##### 1) 快速的信息选择

注意机制可以高效地从众多的视觉信息中, 按照与跟踪任务的相关性及其重要性, 进行优先选择, 将跟踪器快速指向感兴趣的目标或区域. 从而极大地减少需进一步处理的视觉信息, 提高数据处理效率, 增强跟踪的实时性.

##### 2) 自动初始化

当没有事先指定需要跟踪的目标时, 注意机制能够以自下而上的方式, 自动的检测显著区域或显著物体, 使得跟踪系统的初始化可以自动进行.

##### 3) 自动恢复机制

由于存在诸如背景杂乱、运动突变、完全遮挡等各种原因可能导致跟踪失效, 事先设计好的跟踪系统通常不能应对引发跟踪失效的所有可能原因, 因此, 对失效的自动恢复显得尤为重要. 通过可靠的验证模块自动地确定跟踪是否失效, 一旦失效, 在注意机制的辅助下, 跟踪系统可以快速地进行全局搜索, 达到目标重检测的目的.

##### 4) 判别特征选择

从注意计算模型得到的显著性, 可以近似衡量目标的局部或全局判别性, 从而可以用来自动地选择判别特征. 判别性可以改善自上而下的注意选择和跟踪恢复过程的目标检测性能.

##### 5) 完整的目标区域

基于物体的注意机制可以获得完整目标区域, 为在线学习提供完整的纯净的训练样本.

##### 6) 稳定的特征提取

基于注意机制的思想, 可以设计如显著点、区域描述子等稳定新颖的图像特征, 对目标进行描述.

#### 4.2 需要进一步研究的问题

尽管引入注意机制的视觉跟踪技术在近几年已经取得了很大的研究进展, 但是仍然有许多问题值得作进一步的深入研究.

1) 目前提出的基于注意机制的方法, 大多直接把注意到的区域作为跟踪结果, 这与人的跟踪过程不符, 人的视觉系统通过注意机制选择与当前任务相关度高的信息后, 送入工作区进行处理; 未来的研究应体现出注意对目标的保持能力, 选择后的信息送入保持有稳定的目标特征的验证模块作精细处理, 并对验证后的样本在线学习.

2) 由目标确定的判别性信息指导生成的显著图, 可以有不同的利用方式, 如: 可直接作为目标置信图, 在上面直接施加, 如 Mean shift<sup>[86]</sup> 之类的搜索策略; 或者作为先验融合到别的方法中, 从而提高跟踪的鲁棒性; 或者作为一种错误恢复机制. 因此, 可以深入发掘显著图的利用方式, 以助于提高跟踪

的性能.

3) 注意机制具备很好的特征选择能力, 而且已有的基于注意机制思想所设计的显著点, 也取得了不错的效果, 因此, 可以基于注意机制, 深入挖掘其特征提取的潜力, 设计新颖的具有很好表示能力的特征.

4) 视觉注意计算模型的时间复杂度, 对其能否应用到跟踪中有很大的影响, 因为, 可用的目标跟踪方法, 通常有实时性的要求. 由于频域分析在运算效率在有很大的优势, 因此研究使用频域分析的基于物体的注意模型并将其应用到跟踪中, 是一个值得研究的课题.

5) 目前的大多数视觉注意计算模型主要集中在自下而上方面, 尚缺乏任务驱动注意 (即自上而下注意) 的计算原理, 而跟踪是任务驱动的, 因此, 研究适用于包括跟踪在内的, 满足时变任务需求的任务驱动注意模型, 以进一步提升目标跟踪的性能, 是一个很有前景的研究方向.

6) 目前为止, 仅有少量文献将“基于视觉注意机制的目标跟踪方法”用到军事应用中, 未见将这类方法应用到如导弹图像末制导等需要对高速运动目标进行跟踪的场合的文献报道. 该类应用具有目标和相机高速运动、成像质量较差、精度和实时要求高等特点. 因此, 针对该类应用, 研究快速注意计算模型并设计相应的基于注意机制的跟踪算法, 是一个值得尝试的研究课题.

#### 5 结束语

视觉注意机制在人类视觉认知过程中扮演了重要角色. 通过视觉注意力的引导, 人类可以在短时间内搜寻到视野中的含有潜在目标的重要区域, 有效压缩了完成复杂视觉认知任务所需的计算量. 使用人类的视觉注意机制改进和完善计算机视觉跟踪技术, 具有重要的意义.

将人的视觉注意机制引入到复杂场景下的目标跟踪中, 有利于实现更为稳定和接近于人类认知机制的视觉跟踪算法. 同时, 模仿人的视觉认知机制, 把视觉认知与信息科学实现融合和交叉是目前的重要趋势之一. 随着认知心理学、计算机视觉、机器学习等学科的交叉研究的不断深入, 在不远的将来, 和人类视觉处理方式更加接近的鲁棒实时的目标跟踪系统将得以实现.

#### References

- 1 Zhang S P, Yao H X, Sun X, Lu X S. Sparse coding based visual tracking: review and experimental comparison. *Pattern Recognition*, 2013, **46**(7): 1772–1788
- 2 Yilmaz A, Javed O, Shah M. Object tracking: a survey. *ACM Computing Surveys*, 2006, **38**(4): 13

- 3 Hou Zhi-Qiang, Han Chong-Zhao. A survey of visual tracking. *Acta Automatica Sinica*, 2006, **32**(4): 603–617  
(侯志强, 韩崇昭. 视觉跟踪技术综述. 自动化学报, 2006, **32**(4): 603–617)
- 4 Maggio E, Cavallaro A. *Video Tracking: Theory and Practice*. West Sussex: Wiley, 2011
- 5 Zou Hai-Rong, Gong Zhen-Bang, Luo Jun. The status quo and prospect of tracking system of ground moving object by UAV. *Journal of Astronautics*, 2006, **27**(z1): 233–236  
(邹海荣, 龚振邦, 罗均. 无人飞行器地面移动目标跟踪系统研究现状与展望. 宇航学报, 2006, **27**(z1): 233–236)
- 6 Yoo S, Kim W, Kim C. Saliency combined particle filtering for aircraft tracking. *Journal of Signal Processing Systems*, 2013, doi: 10.1007/s11265-013-0803-x
- 7 Liu Wei-Feng, Chai Zhong, Wen Cheng-Lin. Multi-measurement target tracking by using random sampling approach. *Acta Automatica Sinica*, 2013, **39**(2): 164–174  
(刘伟峰, 柴中, 文成林. 基于随机采样的多量测目标跟踪算法. 自动化学报, 2013, **39**(2): 164–174)
- 8 Itti L, Koch C. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2001, **2**(3): 194–203
- 9 Frintrop S, Rome E, Christensen H I. Computational visual attention systems and their cognitive foundations: a survey. *ACM Transactions on Applied Perception*, 2010, **7**(1): 1–39
- 10 Frintrop S. Computational visual attention. *Computer Analysis of Human Behavior*. London: Springer, 2011. 69–101
- 11 Borji A, Itti L. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(1): 185–207
- 12 Guo Ying-Chun, Yuan Hao-Jie, Wu Peng. Image saliency detection based on local and regional features. *Acta Automatica Sinica*, 2013, **39**(8): 1214–1224  
(郭迎春, 袁浩杰, 吴鹏. 基于 Local 特征和 Regional 特征的图像显著性检测. 自动化学报, 2013, **39**(8): 1214–1224)
- 13 Ma Ru-Ning, Tu Xiao-Po, Ding Jun-Di, Yang Jing-Yu. To evaluate salience map towards popping out visual objects. *Acta Automatica Sinica*, 2012, **38**(5): 870–875  
(马儒宁, 涂小坡, 丁军娣, 杨静宇. 视觉显著性凸显目标的评价. 自动化学报, 2012, **38**(5): 870–875)
- 14 Navalpakkam V, Itti L. An integrated model of top-down and bottom-up attention for optimizing detection speed. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2006. 2049–2056
- 15 Hu Zheng-Ping, Meng Peng-Quan. Graph presentation random walk salient object detection algorithm based on global isolation and local homogeneity. *Acta Automatica Sinica*, 2011, **37**(10): 1279–1284  
(胡正平, 孟鹏权. 全局孤立性和局部同质性图表示的随机游走显著目标检测算法. 自动化学报, 2011, **37**(10): 1279–1284)
- 16 Gao D S, Han S Y, Vasconcelos N. Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(6): 989–1005
- 17 Feng Xin, Yang Dan, Zhang Ling. Saliency variation based quality assessment for packet-loss-impaired videos. *Acta Automatica Sinica*, 2011, **37**(11): 1322–1331  
(冯欣, 杨丹, 张凌. 基于视觉注意力变化的网络丢包视频质量评估. 自动化学报, 2011, **37**(11): 1322–1331)
- 18 Anderson J R. *Cognitive Psychology and Its Implications*. New York: Worth Publishers, 2004
- 19 Itti L. Visual salience [Online], available: [http://www.scholarpedia.org/article/Visual\\_salience](http://www.scholarpedia.org/article/Visual_salience), July 24, 2013
- 20 Wu Y, Lim J, Yang M H. Online object tracking: A benchmark. In: Proceedings of the 2003 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, OR, USA: IEEE, 2013. 2411–2418
- 21 Treisman A M, Gelade G. A feature-integration theory of attention. *Cognitive Psychology*, 1980, **12**(1): 97–136
- 22 Wolfe J M. Guided Search 2.0. A revised model of visual search. *Psychonomic Bulletin and Review*, 1994, **1**(2): 202–238
- 23 Wolfe J M. Guided search 4.0. *Integrated Models of Cognitive Systems*, 2006, **44**(3): 99–120
- 24 Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, **20**(11): 1254–1259
- 25 Walther D, Koch C. Modeling attention to salient proto-objects. *Neural Networks*, 2006, **19**(9): 1395–1407
- 26 Frintrop S. *VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search*. Berlin: Springer, 2006
- 27 Harel J, Koch C, Perona P. Graph-based visual saliency. In: Proceedings of the 20th Annual Conference on Neural Information Processing Systems, NIPS 2006. New York: Neural Information Processing System Foundation, 2007. 545–552
- 28 Itti L, Dhavale N, Pighin F. Realistic avatar eye and head animation using a neurobiological model of visual attention. In: Proceedings of the SPIE's 48th Annual Meeting Optical Science and Technology. New York: International Society for Optics and Photonics, 2004. 64–78
- 29 Gao D, Vasconcelos N. Decision-theoretic saliency: computational principles, biological plausibility, and implications for neurophysiology and psychophysics. *Neural Computation*, 2009, **21**(1): 239–271
- 30 Hou X, Zhang L. Saliency detection: a spectral residual approach. In: Proceedings of the 2007 Computer Vision and Pattern Recognition. Minneapolis, MN: IEEE, 2007. 1–8
- 31 Guo C L, Ma Q, Zhang L M. Spatio-temporal Saliency detection using phase spectrum of quaternion Fourier transform. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK: IEEE, 2008. 1–8
- 32 Achanta R, Hemami S, Estrada F, Susstrunk S. Frequency-tuned salient region detection. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, 2009. 1597–1604
- 33 Li J, Levine M D, An X, Xu X, He H. Visual saliency based on scale-space analysis in the frequency domain. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(4): 996–1010
- 34 Li W Y, Wang P, Qiao H. Top-down spatiotemporal saliency detection using spectral Filtering. In: Proceedings of the 5th International Conference on Digital Image Processing (ICDIP 2013). Beijing, China: SPIE, 2013. 88782G–88782G-5
- 35 Liu T, Sun J, Zheng N N, Tang X O, Shum H Y. Learning to detect a salient object. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN: IEEE, 2007. 1–8

- 36 Mahadevan V, Vasconcelos N. Saliency-based discriminant tracking. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, 2009. 1007–1013
- 37 Yantis S. Multielement visual tracking: attention and perceptual organization. *Cognitive Psychology*, 1992, **24**(3): 295–340
- 38 Allen R, McGeorge P, Pearson D, Milne A B. Attention and expertise in multiple target tracking. *Applied Cognitive Psychology*, 2004, **18**(3): 337–347
- 39 McKeever P, Pylyshyn Z. Nontarget Numerosity and Identity Maintenance with FINSTs: A Two Component Account of Multiple Target Tracking, Technical Report, Centre for Cognitive Science, University of Western Ontario, Canada, 1993
- 40 Sears C R, Pylyshyn Z W. Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Experimentale*, 2000, **54**(1): 1–14
- 41 Makovski T, Jiang Y V. Feature binding in attentive tracking of distinct objects. *Visual Cognition*, 2009, **17**(1–2): 180–194
- 42 Doran M M, Hoffman J E. The role of visual attention in multiple object tracking: evidence from ERPs. *Attention, Perception, and Psychophysics*, 2010, **72**(1): 33–52
- 43 Mahadevan V, Vasconcelos N. Automatic initialization and tracking using attentional mechanisms. In: Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2011). Colorado Springs, CO: IEEE, 2011. 15–20
- 44 Mahadevan V, Vasconcelos N. On the connections between saliency and tracking. In: Proceedings of the 26th Annual Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, USA: Curran Associates, Inc., 2012. 1673–1681
- 45 Collins R T, Liu Y, Leordeanu M. Online selection of discriminative tracking features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10): 1631–1643
- 46 Wang P, Qiao H. Online appearance model learning and generation for adaptive visual tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, **21**(2): 156–169
- 47 Ouerhani N, Hügli H. A model of dynamic visual attention for object tracking in natural image sequences. *Computational Methods in Neural Modeling*, 2003, **2686**: 702–709.
- 48 Walther D, Edgington D R, Koch C. Detection and tracking of objects in underwater video. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE, 2004. I-544–I-549
- 49 Veyret M, Maisel E. Attention-based target tracking for an augmented reality application. In: Proceedings of the 14th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2006. Václav Skala, Union Agency, 2006. 101–108
- 50 Li S, Lee M C. Fast visual tracking using motion saliency in video. In: Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing. Honolulu, HI: IEEE, 2007. I1073–I1076
- 51 Liu H, Shi Y. A robust visual tracking based on selective attention shift. In: Proceedings of the 2009 IEEE Control Applications, (CCA) and Intelligent Control. New York, USA: IEEE, 2009. 1176–1179
- 52 Guo W, Xu C S, Ma S D, Xu M. Visual attention based motion object detection and trajectory tracking. In: Proceedings of the 11th Pacific Rim Conference on Multimedia. Berlin: Springer, 2010. 462–470
- 53 Lee S, Kim G J, Choi S. Real-time tracking of visually attended objects in virtual environments and its application to LOD. *IEEE Transactions Visualization and Computer Graphics*, 2009, **15**(1): 6–19
- 54 Yang G, Liu H. Visual attention and multi-cue fusion based human motion tracking method. In: Proceedings of the 6th International Conference on Natural Computation (ICNC). Yantai, China: IEEE, 2010. 2044–2054
- 55 Ma L L, Cheng J, Liu J, Wang J Q, Lu H Q. Visual attention model based object tracking. In: Proceedings of the 2010 Advances in Multimedia Information Processing, and the 11th Pacific Rim Conference on Multimedia. Berlin: Springer, 2010. 483–493
- 56 Yu Y L, Mann G K I, Gosine R G. A single-object tracking method for robots using object-based visual attention. *International Journal of Humanoid Robotics*, 2012, **9**(4): 50030
- 57 Zhu Ming-Qing, Wang Zhi-Ling, Chen Zong-Hai. Human visual intelligence and particle filter based robust object tracking algorithm. *Control and Decision*, 2012, **27**(11): 1720–1724  
(朱明清, 王智灵, 陈宗海. 基于人类视觉智能和粒子滤波的鲁棒目标跟踪算法. 控制与决策, 2012, **27**(11): 1720–1724)
- 58 Santana P, Correia L, Mendonça R, Alves N, Barata J. Tracking natural trails with swarm-based visual saliency. *Journal of Field Robotics*, 2013, **30**(1): 64–86
- 59 Frintrop S, Kessel M. Most salient region tracking. In: Proceedings of the 2009 IEEE International Conference on Robotics and Automation. Kobe: IEEE, 2009. 1869–1874
- 60 Frintrop S, Königs A, Hoeller F, Schulz D. Visual person tracking using a cognitive observation model. In: Proceedings of the 2009 ICRA Workshop on People Detection and Tracking. Kobe, Japan: IEEE, 2009
- 61 Frintrop S. General object tracking with a component-based target descriptor. In: Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA). Anchorage, AK: IEEE, 2010. 4531–4536
- 62 Frintrop S, Königs A, Hoeller F, Schulz D. A component-based approach to visual person tracking from a mobile platform. *International Journal of Social Robotics*, 2010, **2**(1): 53–62
- 63 López M T, Fernández-Caballero A, Fernández M A, Mira J, Delgado A E. Visual surveillance by dynamic visual attention method. *Pattern Recognition*, 2006, **39**(11): 2194–2211
- 64 Zhang S P, Yao H X, Liu S H. Robust visual tracking using feature-based visual attention. In: Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing. New York: IEEE, 2010. 1150–1153
- 65 Chu J K, Li R H, Li Q Y, Wang H Q. A visual attention model for robot object tracking. *International Journal of Automation and Computing*, 2010, **7**(1): 39–46
- 66 Mahadevan V, Vasconcelos N. Biologically inspired object tracking using center-surround saliency mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(3): 541–554

- 67 Huang J W, Li Z N. Automatic detection of object of interest and tracking in active video. *Journal of Signal Processing Systems*, 2011, **65**(1): 49–62
- 68 Sidibé D, Fofi D, Mériaudeau F. Using visual saliency for object tracking with particle filters. In: Proceedings of the 18th European Signal Processing Conference (EUSIPCO 2010), 2010. 1776–1780
- 69 Li Z D, Wang W H, Wang Y, Chen F, Wang Y. Visual tracking by proto-objects. *Pattern Recognition*, 2013, **46**(8): 2187–2201
- 70 Yan J, Chen X, Zhu Q. Robust online tracking via adaptive samples selection with saliency detection. *EURASIP Journal of Advances in Signal Processing*, 2013, **2013**(1): 1–11
- 71 Zhang G, Yuan Z J, Zheng N N, Sheng X D, Liu T. Visual saliency based object tracking. In: Proceedings of the 9th Asian Conference on Computer Vision (ACCV 2009). Berlin: Springer, 2009. 193–203
- 72 Zhang G, Yuan Z, Zheng N. Key object discovery and tracking based on context-aware saliency. *International Journal of Advanced Robotic Systems*, 2013, **10**(5): 1–12
- 73 Li Y, Ma Y F, Zhang H J. Salient region detection and tracking in video. In: Proceedings of the 2003 International Conference on Multimedia and Expo. Baltimore, MD: IEEE, 2003. 269–272
- 74 Zhang Heng, Li Li-Chun, Li You, Yu Qi-Feng. Tracking method based on the significance weighted least square image matching. *Opto-Electronic Engineering*, 2008, **35**(4): 23–27  
(张恒, 李立春, 李由, 于起峰. 显著性加权最小二乘图像匹配跟踪方法. 光电工程, 2008, **35**(4): 23–27)
- 75 Zhang Heng, Li You, Li Li-Chun, Yu Qi-Feng. A new significance weighted mean shift tracking method. *Optical Technique*, 2008, **191**(3): 404–407  
(张恒, 李由, 李立春, 于起峰. 基于显著性加权的 Mean Shift 跟踪方法. 光学技术, 2008, **191**(3): 404–407)
- 76 Yang M, Yuan J S, Wu Y. Spatial selection for attentional visual tracking. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN: IEEE, 2007. 1–8
- 77 Fan J L, Wu Y, Dai S Y. Discriminative spatial attention for robust tracking. *Computer Vision-ECCV 2010, Pt I*. Berlin: Springer-Verlag, 2010. 480–493
- 78 Wang Q, Chen F, Xu W L. Saliency selection for robust visual tracking. In: Proceedings of the 17th IEEE International Conference on Image Processing (ICIP). Hong Kong, China: IEEE, 2010. 2785–2788
- 79 Zeng J H, Sun Y R. Real-time pedestrian tracking by visual attention and human knowledge learning. In: Proceedings of the 2010 IEEE International Conference on Progress in Informatics and Computing (PIC). Shanghai, China: IEEE, 2010. 345–348
- 80 Fan J L. Contextual saliency with an application to visual tracking. In: Proceedings of the 4th International Congress on Image and Signal Processing (CISP 2011). Shanghai, China: IEEE, 2011. 1416–1419
- 81 Li H L, Ngan K N. Saliency model-based face segmentation and tracking in head-and-shoulder video sequences. *Journal of Visual Communication and Image Representation*, 2008, **19**(5): 320–333
- 82 Liu H, He H J. A salient feature and scene semantics based attention model for human tracking on mobile robots. In: Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA 2010). Anchorage, AK: IEEE, 2010. 4545–4552
- 83 Borji A, Frintrop S. Learning context-based feature descriptors for object tracking. In: Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction. Osaka: IEEE, 2010. 79–80
- 84 Borji A, Frintrop S, Sihite D N, Itti L. Adaptive object tracking by learning background context. In: Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Providence, RI: IEEE, 2012. 23–30
- 85 Liu T, Yuan Z J, Sun J, Wang J D, Zheng N, Tang X. Learning to detect a salient object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(2): 353–367
- 86 Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift. In: Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head Island: IEEE, 2000. 142–149



**黎万义** 中国科学院自动化研究所助理研究员. 2014 年获得中国科学院自动化研究所博士学位. 主要研究方向为视觉检测与跟踪. 本文通信作者.

E-mail: wanyi.li@ia.ac.cn

**(LI Wan-Yi)** Assistant professor at the Institute of Automation, Chinese Academy of Sciences. He received his

Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences in 2014. His research interest covers visual detection and tracking. Corresponding author of this paper.)



**王 鹏** 中国科学院自动化研究所副研究员. 2010 年获得中国科学院自动化研究所博士学位. 主要研究方向为视觉检测与跟踪, 视觉注意力模型.

E-mail: peng.wang@ia.ac.cn

**(WANG Peng)** Associate professor at the Institute of Automation, Chinese Academy of Sciences. He received his

Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences in 2010. His research interest covers visual detection, object tracking, and visual attention modeling.)



**乔 红** 中国科学院自动化研究所研究员. 1995 年获得英国 De Montfort 大学机器人与人工智能博士学位. 主要研究方向为智能机器人, 机器学习和模式识别. E-mail: hong.qiao@ia.ac.cn

**(QIAO Hong)** Professor at the Institute of Automation, Chinese Academy of Sciences. She received her Ph.D. degree from De Montfort University, UK, in 1995. Her research interest covers intelligent robotics, machine learning, and pattern recognition.)