

基于相关滤波器的视觉目标跟踪综述

魏全禄 老松杨 白亮

(国防科学技术大学信息系统工程重点实验室 长沙 410073)

摘要 视觉跟踪是一个重要的计算机视觉任务,有着广泛的应用,由于现实场景中存在着众多困难,视觉跟踪仍是一个活跃的研究领域。判别式分类器是现代跟踪方法中的一个核心组成部分,其在线学习一个二值分类器以在每一帧中区分目标与背景,充分利用机器学习丰富的学习算法,取得了许多突破。相关滤波器已成功应用到目标检测和识别中,其由于计算效率高,近年来作为一种判别式跟踪方法被应用到视觉跟踪领域,取得了很好的效果。首先简要介绍了判别式跟踪算法;然后对相关滤波器基本理论及几种典型的相关滤波器构造方法进行了描述;最后重点介绍了近年来相关滤波器在视觉跟踪中的应用及研究进展,并总结了可能的研究方向和发展趋势。

关键词 视觉跟踪,判别式学习方法,相关滤波器

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.11.001

Visual Object Tracking Based on Correlation Filters: A Survey

WEI Quan-lu LAO Song-yang BAI Liang

(Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China)

Abstract Visual object tracking is a fundamental task in many computer vision applications which is still an active research field due to the challenges in real scenes. The core component of most modern trackers is a discriminative classifier which can use the abundant algorithms in machine learning to learn a binary classifier online to separate the object from the surrounding environment. Due to the high computational efficiency, as a discriminative tracking method, correlation filters which have been successfully applied to a variety of pattern recognition applications are introduced to the topic of visual tracking in recent years. The discriminative learning methods in visual tracking were introduced briefly firstly. Then the fundamental theory and some kinds of the typical methods of correlation filters were described. Finally, a detailed review of the applications of the correlation filters in visual tracking was provided, and the future applications and research trends were discussed.

Keywords Visual tracking, Discriminative learning, Correlation filter

1 引言

视觉跟踪是计算机视觉中的一个基础问题,跟踪的目的是用来确定我们感兴趣的目标在视频序列中连续的位置,即获取运动目标的运动参数,如位置、速度、加速度以及运动轨迹等,从而进行进一步的处理和分析,实现对运动目标的行为分析与理解,以完成更高级的任务。其可广泛应用于自动监控、人机交互、军事侦察、交通监视等领域。近年来,经过各国学者的努力和研究的广泛开展,一些先进的思想、算法和技术相继被提出,但由于部分遮挡、变形、运动模糊、快速运动、光照变化、背景杂乱、尺度变化等原因使得其在实际应用中仍然面临许多挑战,因此仍然是计算机视觉中最活跃的研究领域之一^[1]。

目前,将统计机器学习中丰富的学习算法引入到视觉跟踪中,特别是判别式跟踪方法的广泛应用是视觉跟踪研究中最大的突破之一^[2,3]。同时鉴于判别式相关滤波器定位性能好、在傅里叶域计算速度快等优点,它在目标检测、定位中已有广泛应用。Bolme^[4]将其引入跟踪应用中后,充分展现了

相关滤波器在视觉跟踪中的突出表现,后续也有许多改进算法相继被提出。本文首先介绍了判别式跟踪方法和相关滤波器的基础理论,然后综述并分析了相关滤波器在视觉跟踪中的应用方法及最近的研究成果,最后提出了可能的改进方向和发展趋势。

2 判别式跟踪方法

2.1 外观模型自适应

视觉跟踪一般是根据人工或自动标示出的目标区域,提取目标的特征信息,建立合适的外观模型,利用外观模型在新的帧中估计目标的新状态。由于跟踪过程中不可避免地存在光照、变形、运动、遮挡等各种因素,目标的外观会发生变化。对于有限的变化,选择适合特征和表示的静态模型可以胜任,但实际情况中,常常需要同时考虑多种变化,或者变化比较剧烈,静态模型很难满足实际需要,此时需要对模型进行在线自适应更新来提高跟踪的鲁棒性。模型主要框架如图1所示^[5]。 $I_k = \{I_m\}_{m=1}^k$ 表示图像序列,跟踪的目的就是估计一个隐变量序列 $X_k = \{x_m \in \mathcal{X} \subset \mathcal{R}_d\}_{m=1}^k$,其中 \mathcal{X} 是状态空间, x_m 是

到稿日期:2015-10-30 返修日期:2016-03-15

魏全禄 男,博士生,主要研究方向为计算机视觉、多媒体信息系统, E-mail: blessyou668@163.com; 老松杨 男,博士,教授,主要研究方向为多媒体信息系统与虚拟现实、人机交互、指挥决策分析等; 白亮 男,博士,副教授,主要研究方向为数字视频、音频处理和检索等。

估计的 m 时刻的目标状态, X_k 是从开始到时间 k 的目标的轨迹。 C_k 为信度图, A_k 表示 k 时刻目标的整个外观模型, a_{k+1} 为 I_k 帧估计的部分模型, 比较 I_k 帧中的观测值和模型 A_k 来估计信度图 C_k 和状态 x_k , 模型自适应就是根据当前帧状态 x_k 和信度 C_k 估计 I_k 帧部分模型 a_{k+1} , 对 A_k 进行更新, 得到新的外观模型 A_{k+1} [5]。

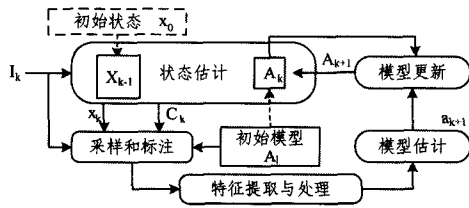


图1 模型更新框架

由于目标一般用矩形框或椭圆来描述, 这种分割很粗糙, 建立的模型必然会包含背景像素, 这种模型自身的误差以及跟踪过程中目标定位的误差会引入到模型的每一次更新中, 随着跟踪的进行, 背景像素逐渐累积, 最终可能导致模型漂移而使得跟踪失败。因此在模型设计和更新时需要减小误差对跟踪效果的影响。

在模型的更新过程中还存在精确性与稳定性的矛盾, 如模型更新过慢, 则不能跟上外观的变化, 从而不能满足精确性; 更新过快又会因为误差、噪声等造成模型的漂移而不能满足稳定性。因此需要在更新的过程中在线评价候选目标区域的质量, 以决定是否进行模型更新。不同的跟踪算法主要集中在样本如何选择、提取何种特征、如何建立模型、模型如何更新等方面。

2.2 跟踪方法

根据模型设计不同, 视觉跟踪方法主要可以分为两类: 生成式跟踪方法和判别式跟踪方法。其中判别式跟踪方法将目标跟踪问题看成一个二分类问题, 在跟踪过程中在线学习一个二值分类器来区分目标及其周围背景, 用学习到的分类器在当前图像帧中对图像块进行分类, 标注像素属于目标还是背景, 找到分类器信度最大的区域即为目标位置, 并使用跟踪结果作为样本更新分类器, 这种方法也称为 tracking-by-detection 法 [6]。

该方法最早由 Collins 和 Lin [7] 提出, 其自适应选择对当前背景和背景最有区分力的颜色特征。常用的 tracking-by-detection 方法包括基于 SVM、随机森林分类器或 boosting 及其变种 [2]。Avidan [8] 使用了一个自适应分类器组合, 每一个弱分类器是一个由 RGB 颜色和 HOG 组成的 11 维特征空间中的线性超平面, 加权组合弱分类器形成强分类器。Tian [9] 提出了一种在线组合线性 SVM 跟踪器, 从关键帧获取样本训练分类器, 将跟踪中的噪声干扰和变形等问题转换为训练样本的有效选择问题, 样本选择不准确容易造成错误学习, 导致跟踪失败。Grabner [10] 设计了一个在线 boosting 分类器, 从一个特征候选集中选择和保持最佳判别特征。Saffari [11] 提出了一种基于在线决策树生长过程的在线随机森林算法, 相比在线 boosting 方法, 该方法对标注误差更鲁棒。

为解决跟踪器每次更新引入误差而导致模型漂移的问题, Grabner [12] 提出了一种半监督的方法, 即只有第一帧中的样本是已标注样本, 后续帧中的都是未标注样本。这种方法尽管能适应目标完全离开视野的情况, 但在第一帧中很难确定目标精确位置。因此, Babenko [13] 提出了一种基于在线多实例学习的跟踪方法, 将多个正样本用样本包的形式代替单

个样本作为标签样本以提高标签样本质量, 解决跟踪中正样本的不确定性问题, 减少模型更新中的漂移。Aeschliman [14] 考虑到造成漂移问题的原因是背景像素的引入, 最直接的方法是获得目标的精确边界, 因此提出了一种联合分割和跟踪的概率框架, 在跟踪鲁棒性方面有了重要改进。Zhang [15] 考虑到跟踪开始时数据不足及跟踪过程中常会发生漂移的问题, 提出了一种有效的基于多尺度特征空间的特征提取方法来扩充样本, 可以得到更准确的样本信息, 且利用压缩感知的理论压缩扩充后的正负样本, 在压缩域在线更新一个朴素贝叶斯分类器, 计算速度快。Kalal [16] 将传统的跟踪算法和检测算法结合起来解决目标在跟踪过程中发生的形变、部分遮挡等问题, 同时通过一种改进的在线学习机制更新检测模块的目标模型及相关参数, 可提高跟踪的稳定性和鲁棒性, 并可解决目标的失联与重入问题。Hare [17] 认为二值分类器的使用是造成样本标注不准确的根本原因, 传统方法如半监督学习、多实例学习等虽可提高鲁棒性, 但不能从根本上解决问题, 因此提出了一种基于在线结构输出 SVM 学习的方法, 直接通过预测函数预测帧间目标位置的变化, 并将标注集成到学习中, 取得了较好的效果。

鉴于相关滤波器在检测、定位应用中的良好表现, 2009 年 Bolme [4] 提出一种在灰度图像上学习一个最小输出误差平方和的相关滤波器, 并将其应用到视觉跟踪中, 充分利用了相关滤波器在频域的快速计算和优良的定位性能, 其对目标的外观变化也有很好的适应性。

3 相关滤波器

3.1 相关滤波器基础

相关可用来测量两个信号的相似度 [18], 在空域表示为两个函数的内积:

$$g = h \otimes f \quad (1)$$

相关计算类似卷积, 根据卷积理论, 使用快速傅里叶变换, 在频域可表示为两个函数傅里叶变换的逐元素乘积:

$$G = H^* \cdot F \quad (2)$$

其大大提高了计算速度。二维图像信号可表示为:

$$g(x, y) = (h \otimes f)(x, y) = F^{-1}(H^*(\omega, \nu) \cdot F(\omega, \nu)) \quad (3)$$

两个信号 f, h 相似度越高, 则其相关输出值 g 越大, 由此可用于检测目标是否存在及其所在位置。但直接相关操作需要与参考图像精确匹配且不能有其他相似目标干扰, 并且对噪声敏感, 也不能处理外观、光照、旋转、尺度等变化, 以及相关输出的峰值锐利度与噪声容忍度之间存在矛盾 [19]。因此在实际应用中, 重点就是设计区分度高、能处理这些变化和影响的相关滤波器, 提高目标位置响应, 抑制其他部位的响应, 并且降低对噪声、旋转、尺度变化等的敏感。

相关滤波器通过将待检测图像与滤波器进行相关处理得到相应结果, 再根据获得的滤波输出来进行判别与定位。理想的滤波器期望是在相关输出值中目标位置产生强峰值而其他位置近似为 0, 但在实际应用中会有噪声, 对相关输出的度量有信噪比 (Signal-to-Noise Ratio, SNR)、峰值-旁瓣比 (Peak-to-Sidelobe Ratio, PSR) 和峰值-相关能量比 (Peak-to-Correlation Energy, PCE)。SNR 表示平均最大输出响应与噪声引起的方差之间的比值; PSR 表示最大峰值与峰值区域外旁瓣的比值; PCE 不考虑噪声影响, 主要反映峰值的锐利度 [19]。

相关滤波器有着诸如移不变性、噪声鲁棒性、良好退化性等特点, 在模式识别领域有着许多成功应用 [20]。实际应用

中,在检测时将设计好的滤波器与输入图像进行相关处理得到相关输出响应,根据响应结果进行分析,如图2所示。

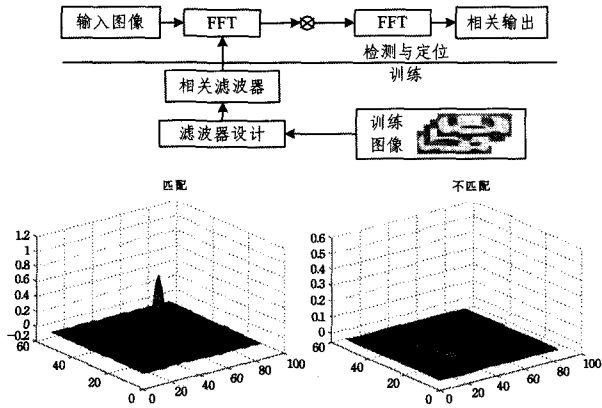


图2 相关滤波器的设计与应用示例

3.2 几种典型的相关滤波器设计

有许多不同的构造滤波器的方法^[18]来改进对不同输入的响应。主要包括合成判别函数(Synthetic Discriminant Functions, SDF)滤波器、无约束相关滤波器、最优相关输出滤波器(Optimized Correlation Output Filters, OCOF)、最大间隔相关滤波器(Maximum Margin Correlation Filters, MMCF)等。

(1) 合成判别函数滤波器

SDFs 包括一组对输出峰值进行硬约束的相关滤波器:最小方差合成判别函数(Minimum Variance Synthetic Discriminant Functions, MVSDF)、最小平均相关能量(Minimum Average Correlation Energy, MACE)、最优平衡滤波器(Optimal Tradeoff Filters, OTF)^[19]。该方法针对训练样本图像与滤波器的相关输出,约束正样本与滤波器的相关输出为1,负样本与滤波器的相关输出为0。其中 MVSDF 在设计时最小化噪声相关输出方差,增强低频,抑制了噪声响应,但同时也降低了响应峰值,降低了检测性能。MACE 最小化训练样本集的平均相关输出能量,增强高频,突出了响应峰值,但增加了对噪声和目标变化的敏感性。OTF 是 MVSDF 与 MACE 的折衷结果,通过一个调整参数平衡响应峰值与抗噪能力。

(2) 无约束相关滤波器

有约束滤波器的约束不一定满足非训练图像,去除约束可以增加解空间并可能提高滤波器性能。典型的无约束滤波器有最大平均相关高度滤波器(Maximum Average Correlation Height, MACH)和无约束 MACE(UMACE)^[20]。其中 MACH 同时最大化平均相关峰值强度并最小化平均相似度(Average Similarity Measure, ASM)和平均相关能量 ACE。UMACE 方法相比 MACH 忽略了最小化 ASM。无约束滤波器消除了有约束情况下的过拟合问题,提高了对目标外观变化的鲁棒性。

(3) 最优相关输出滤波器

OCOFs 考虑相关过程中图像到图像的映射,然后进行逆映射得到理想滤波器。它包括平均合成精确滤波器(Average Synthetic Exact Filter, ASEF)^[21]和最小输出误差平方和(Minimum Output Sum of Squared Error, MOSSE)滤波器^[4]。

使用相关滤波器时希望在目标位置产生强峰值,其余位置响应为0,滤波器根据一组训练图像构建。之前的滤波器训练是通过从训练图像中剪裁样本,OCOFs 采用不同的方法,首先创建一组在训练图像中目标位置为峰值的期望输出,

可用二维高斯函数定义为:

$$g_i(x, y) = \exp\{-\lfloor(x-x_i)^2 + (y-y_i)^2\rfloor/\sigma^2\} \quad (4)$$

然后学习一个映射训练图像到期望输出的滤波器。对于一幅图像样本,可以根据训练图像及其期望输出构建一个滤波器 h_i ,使其满足:

$$g_i = h_i \otimes f_i \quad (5)$$

在频域为:

$$G_i = H_i^* \circ F_i \quad (6)$$

可求得精确的单样本滤波器:

$$H_i^* = G_i / F_i \quad (7)$$

但是精确滤波器对于非训练图像性能不好,泛化能力差。

2009年 Bolme 等人提出的平均合成精确滤波器(Average Synthetic Exact Filter, ASEF)是对所有样本的精确输出进行平均来构造的滤波器:

$$H^* = \frac{1}{n} \sum_{i=1}^n H_i^* = \frac{1}{n} \sum_{i=1}^n \frac{G_i \circ F_i^*}{F_i \circ F_i^* + \epsilon} \quad (8)$$

ASEF 滤波器可以看作是由多个弱分类器组成的强分类器,单一图像样本的精确滤波器是弱分类器。ASEF 滤波器训练需要大量的训练样本,当样本不多时,性能下降很多,可以通过将训练图像随机缩放、旋转、平移来增加样本数量,提高鲁棒性。

2010年他们又提出了最小输出误差平方和(Minimum Output Sum Of Squared Error, MOSSE)滤波器^[4],利用较少的输入样本和合成输出组对,通过最小化所有训练图像的实际相关输出与期望相关输出之差的平方和:

$$\min_{H^*} \sum_{i=1}^n |F_i \circ H^* - G_i|^2 \quad (9)$$

构造滤波器:

$$H^* = \frac{\sum_{i=1}^n G_i \circ F_i^*}{\sum_{i=1}^n F_i \circ F_i^* + \epsilon} \quad (10)$$

MOSSE 滤波器计算速度更快并且更稳定。

(4) 最大间隔相关滤波器

2013年 Rodriguez 等人^[22]提出了最大间隔相关滤波器(Maximum Margin Correlation Filters, MMCF),联合了相关滤波器和 SVM。其多目标函数为:

$$\min_{h, b} (h^T h + C \sum_{i=1}^n \xi_i, \sum_{i=1}^n \|f_i \otimes h - g_i\|_2^2) \quad (11)$$

$$\text{s. t. } t_i (h^T f_i + b) \geq c_i - \xi_i$$

MMCF 的设计平衡了最大化分类间隔与最大化相关输出峰值,同时利用了 SVM 的泛化能力和相关滤波器的定位能力。

4 基于相关滤波器的跟踪

2010年 Bolme^[4]最早将相关滤波器引入到视觉跟踪。在跟踪应用中,使用样本图像训练滤波器建立目标外观模型。在跟踪开始阶段首先在第一帧中初始选择目标窗口并进行随机仿射变换,得到一组样本图像训练 MOSSE 滤波器:

$$H^* = \frac{\sum_{i=1}^n G_i \circ F_i^*}{\sum_{i=1}^n F_i \circ F_i^* + \epsilon} \quad (12)$$

然后在后续帧中,将滤波器与搜索窗口进行相关操作,找到相关输出的最大值位置来表示目标的当前帧位置,以此实现跟踪,并基于新位置图像在线更新滤波器:

$$H_i^* = A_i / B_i \quad (13)$$

$$A_i = \eta G_i \circ F_i^* + (1 - \eta) A_{i-1} \quad (14)$$

$$B_i = \eta G_i \circ F_i^* + (1 - \eta) B_{i-1} \quad (15)$$

相关滤波器最大的特点是可以利用快速傅里叶变换使得计算速度加快,使用 MOSSE 滤波器跟踪的速度可以达到数百帧每秒,且准确率比较高。文中只是最基本的应用,没有考虑一些复杂情况,且只使用了图像的亮度特征,后续许多方法在此基础上分别从提高计算速度、集成更有效多通道特征、尺度空间目标表示、利用上下文信息、处理遮挡等方面进行改进。如 Henriques 提出的 CSK^[23] 和 KCF^[2] 方法、Zhang^[24] 等对 KCF 方法的改进、Danelljan 的 CN^[25] 和 DSST^[26] 方法、Zhang 的 STC^[27]、Ma 的 LCT^[28],以及 Liu 的 PCF^[29] 等。

(1) CSK 和 KCF

Henriques^[23] 提出了一种密度采样方法。考虑到以前的基于检测的跟踪方法为了提高速度无法选取大量样本,只能采取稀疏采样方法在目标周围随机采样一些图像块作为样本训练分类器,密度采样方法对目标子窗口进行循环移位采样(图 3 为基本样本的垂直循环移位示例),并用其所有循环移位作为训练样本,可以利用所有的图像块,这个密度样本集具有循环结构 $C(u)$,如图 4 所示。

$$C(u) = \begin{bmatrix} u_0 & u_1 & u_2 & \cdots & u_{n-1} \\ u_{n-1} & u_0 & u_1 & \cdots & u_{n-2} \\ u_{n-2} & u_{n-1} & u_0 & \cdots & u_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_1 & u_2 & u_3 & \cdots & u_0 \end{bmatrix} \quad (16)$$



图 3 基本样本垂直循环移位示例

$$C(\leftarrow) = \begin{bmatrix} \text{基本样本} \\ \vdots \\ \text{样本移位} \end{bmatrix}$$

图 4 循环矩阵示意图

矩阵中第一行为向量 u ,下面每行为向量 u 的移位。在进行卷积运算时,向量 u 与向量 v 的卷积可表示为:

$$u * v = C(u)v = F^{-1}(F^*(u) \circ F(v)) \quad (17)$$

因此可在傅里叶域进行快速计算。

CSK 学习一个核化最小二乘分类器,使用目标灰度图像块周围的所有循环移位作为样本训练分类器,用高斯函数对样本进行标注,利用核变换扩展到多维空间,采用与 MOSSE 类似的方法实现跟踪。

KCF^[2] 将循环移位样本的岭回归与相关滤波器联系起来,利用核化相关滤波器将最初的像素值特征扩展到多通道非线性特征空间,利用 HOG (Histogram of Oriented Gradient) 等性能优良的特征,同时也保持了计算速度快的特性。

(2) CN 和 DSST

鉴于颜色在检测等方面有着成功应用,而在跟踪中常被忽略,考虑到颜色特征可以提高跟踪性能,CN^[25] 在 CSK 方法的基础上将灰度特征扩展到颜色特征,并改进了 CSK 方法的更新机制,使其更适应多通道颜色特征。由于高维的颜色特征会增加计算负担,采用自适应降维方法将初始的 11 维降低到 2 维。

文中使用了英文语言中最基本的 11 个颜色名特征,将 RGB 值映射到 11 维颜色表述。首先将 11 维颜色名特征投影到 10 维子空间的标准正交基上进行归一化,然后通过最小

化一个损失函数寻找一个合适的降维映射,将其自适应降维到 2 维以提高计算速度。

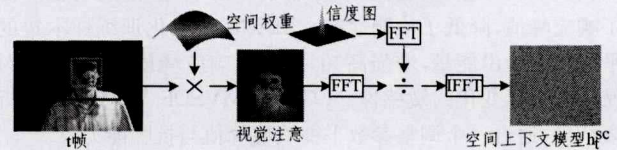
CSK 仅使用了简单的线性插值对分类器系数进行更新。CN 参考 MOSSE 方法,改进了 CSK 的更新机制,在更新模型时考虑了之前所有帧,并通过核化扩展到多维颜色特征。跟踪如同 CSK 方法用学到的模型在新帧中得到检测分数,最大值处为目标新的位置。

DSST^[26] 主要是解决跟踪过程中目标的尺度变化问题。使用 MOSSE 滤波器,并扩展到 HOG 特征以提高精度,在设计滤波器时,在 2 维位置平移的基础上增加 1 维尺度,最终设计一个 3 维位移-尺度相关滤波器。跟踪时在目标周围矩形区域计算一个特征金字塔,训练样本设置为特征金字塔的矩形立方体,中心为估计的目标当前位置和尺度,使用 3 维高斯函数作为相应的期望相关输出。为减少计算量,分别学习 2 维位移滤波器和 1 维尺度滤波器,考虑到跟踪中两帧间尺度的变化比位置变化小,在新帧中首先使用位移滤波器确定位置,然后在新的目标位置使用尺度滤波器确定尺度,最后根据新帧中检测的目标及其尺度分别更新两个滤波器。

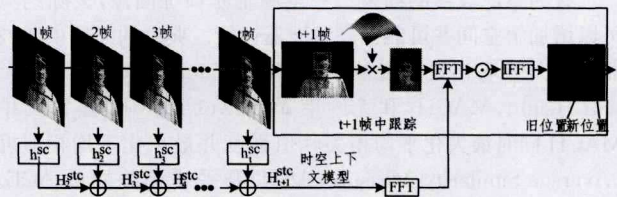
(3) STC 和 LCT

鉴于以往的跟踪方法(无论是生成式还是判别式外观模型)在跟踪过程中由于外观变化、遮挡等原因会使得模型逐渐发生退化,以至于不能提取足够的目标信息进行鲁棒的跟踪,Zhang^[27] 考虑目标周围区域的上下文时空信息,使其更准确地定位。时间信息指邻帧间目标外观和位置不能发生很大变化,空间信息指目标与周围场景存在关联。

STC 在 MOSSE 和 CSK 方法的启发下,在贝叶斯框架下对目标及其局部上下文之间的空间关系进行建模得到空间上下文模型 h^s (见图 5(a)),然后利用前面帧的空间上下文模型在线学习时空上下文模型 H^{sc} ,结合生物视觉上的注意力集中特性来评估新帧中目标出现位置的信度图,信度图中似然概率最大的位置就是目标新的位置(见图 5(b))。



(a) 学习第 t 帧的空间上下文模型



(b) 在 $t+1$ 帧确定目标位置

图 5 STC 方法空间上下文模型和跟踪流程

模型学习和目标检测用到了如 MOSSE 和 CSK 中的快速傅里叶变换来实现,提高了计算速度。在建立上下文模型时根据注意力集中特性考虑目标的尺度变化,在跟踪中通过更新尺度因子实现尺度的自适应。

LCT^[28] 方法将跟踪任务分解为对目标平移和尺度的估计,在跟踪中同样考虑上下文信息,同时使用 KNN 分类器选择信度最高的跟踪结果来训练一个在线随机分类器对丢失的目标进行再检测,实现长时间跟踪。对于跟踪中模型更新过慢不能适应变化而过快又容易导致漂移这个矛盾,文中建

立了两个基于相关滤波器的回归模型,一个是上下文模型,其能快速更新适应变形和遮挡;另一个是使用 HOG 特征构建的多尺度金字塔外观模型,其能慢速更新防止漂移,并适应尺度变化。训练样本使用了 CSK 中的循环移位方法,计算时用到 MOSSE 和 CSK 中的快速傅里叶变换以提高速度,处理多维特征时也使用了高斯核。在估计目标时与 DSST 方法一样先进行位置估计,确定位置后再进行尺度估计。跟踪过程如图 6 所示^[28]。

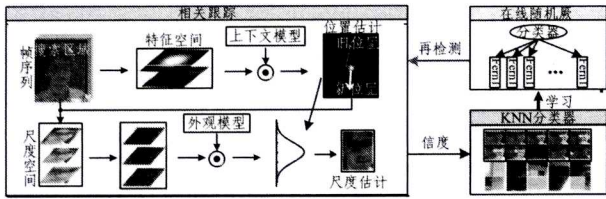


图 6 LCT 跟踪流程

为了保证模型的稳定性及实现长时间跟踪,对相关输出的信度值设置了两个阈值 T_a 和 T_r , $T_r < T_a$, 当最大信度值小于 T_r 时,说明信度很低,目标遮挡严重甚至跟踪失败,此时使用在线训练的随机森林分类器重新检测目标;当最大信度值大于 T_a 时,说明模型稳定,此时对目标的外观模型进行更新,防止模型漂移。

(4) PCF

PCF^[29]方法旨在解决跟踪中的部分遮挡问题。它将目标分成许多小块,对各个分块分别独立跟踪,当一些块被遮挡时,可利用其余块进行有效跟踪,但传统的基于分块的方法由于复杂的在线训练和更新过程,很难实现实时跟踪。相关滤波器跟踪方法计算效率高、跟踪速度快,但处理遮挡问题的能力较弱。因此文章将二者结合,跟踪时将目标分块,对每个分块分别使用一个 KCF 滤波器得到响应信度图,然后将信度图联合起来形成定位整个目标的单个信度图,并用到粒子滤波中推断目标候选的最大后验概率。

难点是如何联合各个分块的信度图,因为不同时刻不同分块的变化情况都不一样,不能简单地使用相同权重。滤波器输出的峰值-旁瓣比 PSR 大,说明两帧之间目标的匹配度高;同一个分块两帧之间的信度变化小,说明该分块在两帧之间外观变化小,因此联合使用相关输出的 PSR 和信度图平滑约束 SCCM 实时评价不同分块的跟踪性能,以赋予合适的权重:

$$\omega'_i = PSR_i + \eta \cdot \frac{1}{SCCM_i} \quad (18)$$

在分类器更新时也只更新权重值大的分块,且学习率与权重成正比,以保证跟踪的稳定性。

结束语 判别式跟踪方法同时利用目标和背景像素,可引入机器学习中丰富的学习算法对模型进行在线更新,可以较好地处理现实应用中的光照变化、外观变形、局部遮挡等问题,已成为近年来流行的跟踪方法。相关滤波器因其在傅里叶域计算速度快,可同时检测和定位、定位性能好等优点,在行人检测、目标定位、视觉跟踪等领域取得了许多成功的应用。

将基于相关滤波器的跟踪方法与传统改进方法相结合,可更好地应对跟踪问题中的各种挑战,提高跟踪的准确性、快速性、鲁棒性及实现长时间跟踪的能力。未来研究方向如下:

(1) 采用更好的特征提高目标模型的判别力。如将 Haar-like、LBP、颜色等特征进行有效结合,弥补相互的不足,提高复杂场景下跟踪的鲁棒性。

(2) 将目标分块适应局部遮挡。可以将目标进行分块描

述,或者对每个局部块分别使用一个相关滤波器。

(3) 改进模型的更新机制以提高模型对外观变化的适应性。

(4) 构建更优良的相关滤波器,提高定位、检测性能。

(5) 加入检测环节,构建一个在线检测器,在跟踪丢失后对目标进行重检测以实现长时间的跟踪。

参考文献

- [1] Smeulders A W M, Chu D M, Calderara S, et al. Visual Tracking: an Experimental Survey[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2013, 36(7): 1442-1468
- [2] Henriques J A F, Caseiro R, Martins P, et al. High-Speed Tracking with Kernelized Correlation Filters[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2014, 37(3): 583-596
- [3] Zhang Huan-long, Hu Shi-qiang, Yang Guo-sheng. Video Object Tracking Based on Appearance Models Learning[J]. Journal of Computer Research and Development, 2015, 52(1): 177-190 (in Chinese)
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[J]. 计算机研究与发展, 2015, 52(1): 177-190
- [4] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2010: 2544-2550
- [5] Salti S, Cavallaro A, Stefano L D. Adaptive Appearance Modeling for Video Tracking: Survey and Evaluation[J]. IEEE Transaction on Image Processing, 2012, 21(10): 4334-4348
- [6] Yang Han-xuan, Shao Ling, Zheng Feng, et al. Recent advances and trends in visual tracking: A review[J]. Neurocomputing, 2011, 74(18): 3823-3831
- [7] Collins R T, Liu Yan-xi, Leordeanu M. Online selection of discriminative tracking features[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1631-1643
- [8] Avidan S. Ensemble tracking [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2007, 29(2): 261-271
- [9] Tian M, Zhang W, Liu F. Online ensemble SVM for robust object tracking[C]// 8th Asian Conference on Computer Vision. Tokyo, 2007: 355-364
- [10] Grabner H, Bischof H. Real-time tracking via online boosting [C]// Proceedings of the British Machine Vision Conference. Edinburgh, 2006: 47-56
- [11] Saffari A, Leistner C, Santner J, et al. Online random forests[C]// IEEE 12th International Conference on Computer Vision Workshops. Kyoto, 2009: 1393-1400
- [12] Grabner H, Leistner C, Bischof H. Semi-supervised online boosting for robust tracking [C]// 10th European Conference on Computer Vision. Marseille, 2008: 234-247
- [13] Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning[C]// IEEE Conference on Computer Vision and Pattern Recognition. Miami, 2009: 983-990
- [14] Aeschliman C, Park J, Kak A. A probabilistic framework for joint segmentation and tracking[C]// IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, 2010: 1371-1378
- [15] Zhang Kai-hua, Zhang Lei, Yang M H. Fast Compressive Tracking[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2014, 36(10): 2002-2015

(下转第 18 页)

Images[D]. Beijing: Beijing Jiaotong University, 2010 (in Chinese)

鹿瑞. 自然图形的颜色恒常性计算研究[D]. 北京: 北京交通大学, 2010

- [15] Finlayson G D, Trezzi E. Shades of gray and color constancy[C]// Proceeding of IS&T/SID 12th Color Imaging Conference(CIC). 2004; 37-41
- [16] Weijer J V, Schmid C. Coloring Local Feature Extraction[C]// Proc. of European Conference on Computer Vision (ECCV). 2006; 334-348
- [17] Weijer V, Gevers T, Gijzen A. Edge-Based Color Constancy [J]. IEEE Trans. on Image Processing, 2007, 16(9): 2207-2214
- [18] Xiong W, Funt B. Estimating Illumination Chromaticity via Support Vector Regression [J]. Journal of Imaging Science and Technology, 2006, 50(4): 341-348
- [19] Li B, Xiong W, Hu W, et al. Multi-Cue Illumination Estimation via a Tree-Structured Group Joint Sparse Representation[J]. International Journal of Computer Vision, 2015, 117(1): 21-47
- [20] Huang G B, Zhu Q Y, Siew C K. Extreme Learning Machine: Theory and Applications[J]. Neuro Computing, 2006, 70: 489-501
- [21] Finlayson G D, Trezzi E. Shades of gray and color constancy [C] // Proceeding of 12th Conference on Color Imaging. 2004
- [22] Chakrabarti A, Hirakawa K, Zickler T. Color Constancy beyond Bags of Pixels[C]// Proc. of the Int. Conference on Computer Vision and Pattern Recognition(CVPR). 2008; 1-6
- [23] Finlayson G, Hordley S, Hubel P. Color by correlation: A simple, unifying framework for color constancy[J]. IEEE Trans. on Pattern Anal. Machine Intell. , 2001, 23(11): 1209-1221
- [24] Forsyth D A. A Novel Algorithm for Color Constancy[J]. International Journal of Computer Vision, 1990, 5(1): 5-36
- [25] Li B, Xiong W, Xu D, et al. A supervised combination strategy for illumination chromaticity estimation[J]. ACM Transactions on Applied Perception(TAP), 2010, 8(1): 885-900
- [26] Cardei V, Funt B, Barnard K. Estimating the Scene Illumination Chromaticity Using a Neural Network[J]. Journal of the Optical Society of America A, 2002, 19(12): 2374-2386
- [27] Gevers T, Smeulders A. Color based Object Recognition[J]. Pattern Recognition, 1999, 32(3): 453-464
- [28] Cardei V C, Funt B. Committee-based color constancy [C]// Proceeding of 7th Color Imaging Conference. 1999; 311-313
- [29] Russel S, Norvig P, Russel S, et al. Prentice Hall Series in Artificial Intelligence[M]. Prentice Hall, 1995
- [30] Barnard K, Martin L, Funt B. Color by correlation in a three dimensional color space[C]// Proc. of the 6th European Conference on Computer Vision(ECCV). 2000; 275-289
- [31] Funt B, Barnard K, Martin L. Is color constancy good enough [C]// Proc. of 5th European Conference on Computer Vision (ECCV). 1998; 445-459
- [32] Brainard D H, Freeman W T. Bayesian Color Constancy[J]. Journal of the Optical Society of America A(Optics, Image Science and Vision), 1997, 14(7): 1393-1411
- [33] Gershon R, Jepson A D, Tsotsos J K. From $[r, g, b]$ to Surface Reflectance: Computing Color Constant Descriptors in Images [C]// Perception. 1988; 755-758
- [34] Land E. The retinex theory of color version [J]. Scientific American, 1977, 237(6): 108-128
- [35] von Kries J. Chromatic adaptation[M]// Sources of Color Science, 1970
- [36] Khan R, Van de Weijer J, Shahbaz Khan F, et al. Discriminative color descriptors[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2013; 2866-2873
- [37] Land E H. The retinex theory of color vision [J]. Scientific American, 1977, 237(6): 108-128
- [38] Kecman V. Learning and Soft Computing: Support Vector Machine, Neural Networks, and Fuzzy Logic Models [M]. MIT Press, Cambridge, London, 2001
- [39] Funt B V, Finlayson G D. Color Constant Color Indexing[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1995, 17(5): 522-529

(上接第 5 页)

- [16] Kalal Z, Mikolajczyk K, Matas J. Tracking-Learning-Detection [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422
- [17] Hare S, Saffari A, Torr P H S. Struck: Structured Output Tracking with Kernels[C]// IEEE International Conference on Computer Vision. Barcelona, 2011; 263-270
- [18] Bolme D S. Theory and Application of Optimized Correlation Output[D]. Colorado State University, 2011
- [19] Kumar B V K V, Mahalanobis A, Juday R D. Correlation Pattern Recognition[M]. Cambridge University Press, 2005
- [20] Kumar B V K V, Fernandez J A, Rodriguez A, et al. Recent advances in correlation filter theory and application[J]. Optical Pattern Recognition, 2014, 9094, 909404-909404-13
- [21] Bolme D S, Draper B A, Beveridge J R. Average of synthetic exact filters[C]// IEEE Conference on Computer Vision and Pattern Recognition. Miami, 2009; 2105-2112
- [22] Rodriguez A, Kumar B V K V. Maximum Margin Correlation Filter: A New Approach for Localization and Classification[J]. IEEE Transaction on Image Processing, 2013, 22(2): 631-643
- [23] Henriques J A F, Caseiro R, Martins P, et al. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels[C]// 12th European Conference on Computer Vision, Florence, Italy, 2012; 702-715
- [24] Zhang Lei, Wang Yan-jie, Liu Yan-ying, et al. Visual object tracking algorithm based on correlation filters[J]. Journal of Optoelectronics Laser, 2015, 26(7): 1349-1357 (in Chinese)
- 张雷, 王延杰, 刘艳莹, 等. 基于相关滤波器的视觉目标跟踪方法[J]. 光电子激光, 2015, 26(7): 1349-1357
- [25] Danelljan M, Khan F S, Felsberg M, et al. Adaptive Color Attributes for Real-Time Visual Tracking[C]// IEEE Conference on Computer Vision and Pattern Recognition. Columbus, 2014; 1090-1097
- [26] Danelljan M, Häger G, Khan F S, et al. Accurate Scale Estimation for Robust Visual Tracking[C]// Proceedings of the British Machine Vision Conference. 2014; 1-11
- [27] Zhang K, Zhang L, Liu Q, et al. Fast visual tracking via dense spatio-temporal context learning [C] // 13th European Conference on Computer Vision. Zurich, 2014; 127-141
- [28] Ma Chao, Yang Xiao-kang, Zhang Chong-yang, et al. Long-term Correlation Tracking[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2015; 5388-5396
- [29] Liu Ting, Wang Gang, Yang Qing-xiong. Real-time part-based visual tracking via adaptive correlation filters[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2015; 4902-4912