

DOI: 10.11992/tis.201612040

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20170702.0425.018.html>

高斯核函数卷积神经网络跟踪算法

汪鸿翔¹, 柳培忠¹, 骆炎民², 杜永兆¹, 陈智¹

(1. 华侨大学 工学院, 福建 泉州 362021; 2. 华侨大学 计算机科学与技术学院, 福建 厦门 361021)

摘要: 针对深度学习跟踪算法训练样本缺少、训练费时、算法复杂度高等问题, 引入高斯核函数进行加速, 提出一种无需训练的简化卷积神经网络跟踪算法。首先, 对初始帧目标进行归一化处理并聚类提取一系列初始滤波器组, 跟踪过程中结合目标背景信息与前景候选目标进行卷积; 然后, 提取目标简单抽象特征; 最后, 将简单层的卷积结果进行叠加得到目标的深层次特征表达。通过高斯核函数加速来提高算法中全部卷积运算的速度, 利用目标的局部结构特征信息, 对网络各阶段滤波器进行更新, 结合粒子滤波跟踪框架实现跟踪。在 CVPR2013 跟踪数据集上的实验表明, 本文方法脱离了繁琐深度学习运行环境, 能克服低分辨率下目标局部遮挡与形变等问题, 提高复杂背景下的跟踪效率。

关键词: 视觉跟踪; 深度学习; 卷积神经网络; 高斯核函数; 前景目标; 背景信息; 模板匹配; 粒子滤波

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2018)03-0388-07

中文引用格式: 汪鸿翔, 柳培忠, 骆炎民, 等. 高斯核函数卷积神经网络跟踪算法[J]. 智能系统学报, 2018, 13(3): 388-394.

英文引用格式: WANG Hongxiang, LIU Peizhong, LUO Yanmin, et al. Convolutional neural network tracking algorithm accelerated by Gaussian kernel function[J]. CAAI transactions on intelligent systems, 2018, 13(3): 388-394.

Convolutional neural network tracking algorithm accelerated by Gaussian kernel function

WANG Hongxiang¹, LIU Peizhong¹, LUO Yanmin², DU Yongzhao¹, CHEN Zhi¹

(1. College of Engineering, Huaqiao University, Quanzhou 362021, China; 2. College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract: In view of such defects existing in the depth learning tracking algorithm as lack of training samples, large time consumption, and high complexity, this paper proposed a simplified convolutional neural network tracking algorithm in which training is unnecessary. Moreover, the Gaussian kernel function can be applied to this algorithm to significantly lower the computing time. Firstly, the initial frame target was normalized and clustered to extract a series of initial filter banks; in the tracking process, the background information of the target and the candidate target for the foreground were convoluted; then the simple and abstract features of the target were extracted; finally, all the convolutions of a simple layer were superposed to form a deep-level feature representation. The Gaussian kernel function was used to speed-up the convolution operations; also, the local structural feature information of the target was used to update the filters in every stage of the network; in addition, the tracking was realized by combining the particle filter tracking framework. The experimental results on the CVPR2013 tracking datasets show that the method used in this paper can help avoid the typically cumbersome operational environment of deep learning, overcome local object occlusion and deformation at low resolution, and improve tracking efficiency under a complex background.

Keywords: visual tracking; deep learning; convolutional neural network (CNN); gauss kernel function; foreground object; background information; template matching; particle filter

收稿日期: 2016-12-31. 网络出版日期: 2017-07-02.

基金项目: 国家自然科学基金项目 (61203242, 61605048); 福建省自然科学基金项目 (2016J01300, 2015J01256); 华侨大学研究生科研创新能力培育计划资助项目 (1511422004).

通信作者: 柳培忠. E-mail: pzliu@hqu.edu.cn.

视觉跟踪是计算机视觉领域的研究热点, 在虚拟现实、人机交互、智能监控、增强现实、机器感知等场景中有着重要的研究与应用价值。视觉跟踪通

过分析视频图片序列,对检测出的各个候选目标区域实施匹配,定位跟踪目标在视频序列中的位置。目前跟踪算法已经取得很多研究成果,但在实际中应对各类复杂场景时仍面临很大挑战,例如面对遮挡、形变、视频序列分辨率低等影响因素时,如何实现更加鲁棒和准确的跟踪仍然是目前研究的核心^[1]。

视觉跟踪算法一般分为两部分:目标表观建模和跟踪策略。

1) 目标表观建模

根据对目标表观的建模方式可分为判别式模型和生成式模型两类^[2-3]。判别式模型将跟踪问题建模为一个二分类问题,用以区分前景和背景。B. Babenko 等^[4]提出多示例学习算法(MIL),针对跟踪中训练样本不足的问题,引入多示例学习机制,有效抑制跟踪过程中跟踪器的漂移问题;文献^[5]提出具有元认知能力的粒子滤波(MCPF)目标跟踪算法,通过监控到突变,快速调节决策机制,实现稳定的目标跟踪。生成式模型不考虑背景信息直接为目标进行建模。文献^[6]提出了L1跟踪系统,用稀疏方法表示跟踪目标,但算法运算复杂度高;K. Zhang 等^[7]提出了压缩跟踪(CT),用一个稀疏的测量矩阵提取特征以建立一个稀疏、健壮的目标表观模型,取得快速有效、鲁棒性好的跟踪效果;文献^[8]引入小波纹理特征,改善单纯依靠颜色特征不能很好适应环境变化的情况,与单一特征相比能够实现更加稳健的跟踪。

2) 跟踪策略

采用运动模型来估计目标可能的位置,通过先验知识来缩小搜索范围。代表性方法有隐马尔可夫模型^[9]、卡尔曼滤波^[10]、均值漂移算法^[11]和粒子滤波^[12]等。其中,粒子滤波算法因为对局部极小值相对不太敏感且计算非常有效而被广泛应用。另外,近几年相关滤波跟踪算法在目标领域也取得不错的成绩。D.S. Bolme 等^[13]首次将相关滤波引入跟踪领域,通过设计一个误差最小平方和滤波器(MOSSE),在跟踪过程中寻找目标最大响应值来实现跟踪。J.F. Henriques 等^[14]提出的CSK算法使用循环矩阵结构进行相邻帧的相关性检测,利用灰度特征空间提高了算法的准确性。文献^[15]在CSK的基础上,通过循环偏移构建分类器的训练样本,使数据矩阵变成一个循环矩阵,同时引入HOG、颜色、灰度多通道特征,提高了算法的速度和准确性。

传统跟踪算法大多数直接使用视频图像序列中的像素值特征进行建模,当跟踪过程中出现复杂场景等较大挑战时,浅层的像素级特征无法很好应

对。针对卷积神经网络具有强大的特征提取功能,设计一种无需训练的卷积神经网络特征提取方法,在粒子滤波框架下,利用核函数加速卷积运算,实现了一种快速卷积神经网络跟踪算法,通过与其他算法的对比分析,最终验证了所提出算法的有效性。

1 相关工作

2013年以来,深度学习算法在跟踪领域已经取得了很大进展。如深度神经网络、卷积神经网络等深度学习方法能够挖掘出数据的多层表征,而高层级的表征更能够反映数据更深层的本质,相比传统浅层学习特征,基于高层次特征的跟踪算法可以提高目标的跟踪效率^[16]。

1.1 CNN 特征提取结构

卷积神经网络(convolutional neural network, CNN)的网络结构类似于生物神经网络,采用局部连接、权值共享和时空下采样的思想降低了网络复杂度,减少了权值数量,使得CNN在处理高维图像时更具优势。

卷积神经网络具有多层性,在传统神经网络的基础上增加了特征提取的卷积层和保证位移不变的下采样层。每层由多个二维平面特征映射层组成,每个映射层由多个独立的神经元组成。卷积特征的提取首先对原始输入图像进行取片操作获取大量小的局部块,然后应用卷积网络模型对局部块进行训练,得到结果为各个卷积层的神经元卷积滤波器,与新输入的样本图像进行卷积滤波,提取样本图的抽象卷积特征从而得到原始图像的深度特征。如图1所示,给出了一个简单卷积特征提取结构,输入图像卷积后在 C_1 层产生若干特征映射图,然后特征映射图中每组的若干像素再进行求和,加权值与偏置,通过一个激活函数(Sigmoid、Relu)得到 S_2 层的特征映射图,这些映射图再经过滤波得到 C_3 层,这个层级结构再和 S_2 一样产生 S_4 。最终,这些像素值全连接成一个向量输入到传统的神经网络,得到输出。

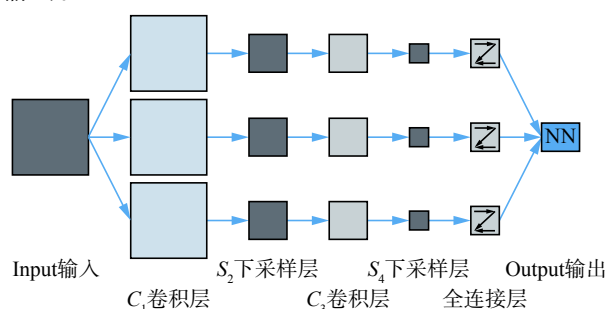


图1 卷积特征提取结构

Fig. 1 Convolution feature extraction structure

1.2 基于深度学习的跟踪算法

深度学习在跟踪领域面临两个重要问题: 1) 深度学习网络在训练阶段需要大量的样本, 而跟踪领域仅能够提供第一帧的目标进行训练, 面临训练样本缺少问题; 2) 采用深度学习的网络模型运算时间复杂度大, 跟踪实时性低。

针对以上问题, N.Y. Wang 等^[17]提出 DLT 算法首次将深度学习算法应用到跟踪领域, 在 ImageNet 数据集上使用栈式降噪自编码器离线预训练来获得通用的物体表征能力, 并在跟踪过程中更新自编码器实现跟踪; K. Zhang 等^[18]结合人脑视觉处理系统, 简化了卷积网络结构, 使用目标区域中随机提取的归一化图像块作为卷积神经网络的滤波器, 从而实现了不用训练卷积神经网络的快速特征提取。文献[19]中的 MDNet 算法提出将训练方法和训练数据交叉运用的思路, 在检测任务中重点关注背景中的难点样本, 显著减轻了跟踪器漂移问题。

深度学习算法需要搭建专用的深度学习硬件平台, 通过大量的前期预训练来训练神经网络提取深度特征, 面临样本缺少, 算法时间复杂度高, 硬件运行平台要求高, 跟踪实时性低等显著缺点。本文结合卷积神经网络在特征提取时能够获得一定的平移、尺度和旋转不变性且能够大幅降低神经网络规模的特性, 结合文献[18]提出的无需训练的卷积特征提取方法, 采用两层前馈处理方式简化卷积网络结构, 通过分层滤波器卷积来抽取目标的高维抽象特征, 利用高斯核函数进行运算加速, 提出一种快速卷积神经网络跟踪算法。

2 高斯核函数卷积神经网络跟踪算法

针对卷积计算时间复杂度过高问题, 本文引入高斯核函数进行变换, 对算法的卷积运算实现加速; 针对深度学习算法训练样本缺少的问题, 本文采用简单两层前馈处理网络实现一种无需训练的特征提取方法。

2.1 核函数卷积

本文的卷积运算采用高斯核函数进行变换加速运算, 文献[15]采用子窗口高斯核函数 $k(\mathbf{x}, \mathbf{x}')$ 的快速计算方法, 其向量表达式为

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{\sigma^2}(\|\mathbf{x}\|^2 + \|\mathbf{x}'\|^2 - 2F^{-1}(\sum_d \hat{\mathbf{x}}^* \odot \hat{\mathbf{x}}'))\right) \quad (1)$$

式中: “*”表示复共轭, d 表示滤波器的个数。

假设 $\varphi(\mathbf{x})$ 是一个高维核希尔伯特空间的映射, 则核函数 $k(\mathbf{x}, \mathbf{x}') = \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle$, 权重可表示为 $\mathbf{v} = \sum_i a_i \varphi(\mathbf{x}_i)$, 系数向量为 α , 其元素为 a_i , 最终要求解的

参数由 \mathbf{v} 变为 α , 因此, 核正则化最小二乘分类器(kernelized regularized least square, KRLS)具有一个简单的闭式解, 即

$$\alpha = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y} \quad (2)$$

式中: \mathbf{K} 是核函数矩阵, 矩阵元素为 $K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$; \mathbf{I} 是单位矩阵; 向量 \mathbf{y} 的元素为 y_i 。由于 \mathbf{K} 为循环矩阵, 则式(2)可转换到 DFT 域, 即

$$\hat{\alpha}^* = \hat{\mathbf{y}} \times (\hat{\mathbf{K}}^{xx'} + \lambda)^{-1} \quad (3)$$

式中: $\hat{\mathbf{K}}^{xx'}$ 是核函数矩阵 \mathbf{K} 的第一行元素组成的向量, 符号 \wedge 表示傅里叶变换。KRLS 分类器的闭式解可利用 FFT 快速得到。

2.2 特征提取

本文利用卷积网络设计一个分层的目标表示结构。在第 1 帧中, 将目标归一化到 $n \times n$ 大小, 从中随机选取一部分重叠的小图像块组成一系列滤波器, 用于同第 t 帧候选目标框中所有抽取的图像块进行卷积, 得到一系列局部特征的映射, 提高候选目标全局特性的鲁棒性。

简单层特征, 通过预处理将图像归一化到 $n \times n$ 大小, 对目标区域利用 $w \times w$ 大小的滑动窗口进行采样, 得到长度为 L 的图像块组 X , 然后用 K-means 聚类的方法从 $L = (n - w + 1)^2$ 个图像块中聚类出 d 个图像块滤波器作为卷积核, 记作 $\mathbf{F}_d = \{\mathbf{F}_1^o, \mathbf{F}_2^o, \dots, \mathbf{F}_d^o\} \subset X$ 。输入图像 \mathbf{I} 对应的响应为

$$\mathbf{S}_i^o = \mathbf{F}_i^o \otimes \mathbf{I}, \mathbf{S}_i^o \in R^{(n-w+1)^2} \quad (4)$$

跟踪目标周围的上下文可为区分目标和背景提供大量有用信息, 对目标周围的区域随机采样出 l 个样本, 同样进行 K-means 聚类获得一组背景模板:

$$\mathbf{F}_l = \{\mathbf{F}_1^b, \mathbf{F}_2^b, \dots, \mathbf{F}_l^b\} \quad (5)$$

最后, 由目标卷积核减去背景卷积核与输入图像 \mathbf{I} 做卷积得到简单层的特征表达:

$$\mathbf{S}_i = \mathbf{S}_i^o - \mathbf{S}_i^b = (\mathbf{F}_i^o - \mathbf{F}_i^b) \otimes \mathbf{I}, i \in \{1, 2, \dots, d\} \quad (6)$$

复杂层特征, 为了加强对目标的特征表达, 本文将 d 个简单层的特征进行堆叠, 构成一个 3 维张量来表示目标的复杂层特征, 记作:

$$\mathbf{C} \in R^{(n-w+1) \times (n-w+1) \times d} \quad (7)$$

这种特征具有平移不变特性, 由于图像归一化后, 使得特征对目标的尺度具有鲁棒性, 且复杂层特征保留不同尺度目标的局部几何信息。文献[20]表明可以通过一个浅层的神经机制实现跟踪, 因此本文没有使用高层次的对象模型而是利用一个简单的模板匹配方案, 结合粒子滤波实现跟踪。

2.3 粒子滤波

本文基于粒子滤波框架, 设第 t 帧时总的观测序列为 $\mathbf{Z}_t = \{z_1, z_2, \dots, z_t\}$, 则跟踪主要是求出后验概率 p 的最大值, 根据贝叶斯理论, 可知:

$$p(\mathbf{S}_t|\mathbf{Z}_t) \propto \frac{p(\mathbf{S}_t|\mathbf{Z}_t)}{\int p(\mathbf{S}_t|\mathbf{S}_{t-1})p(\mathbf{S}_{t-1}|\mathbf{Z}_{t-1})d\mathbf{S}_{t-1}} \quad (8)$$

式中: $\mathbf{S}_t = [x_t, y_t, s_t]^T$, x_t, y_t 为目标的位置, s_t 为尺度参数; $p(\mathbf{S}_t|\mathbf{S}_{t-1})$ 为运动模型, 用于根据第 $t-1$ 帧的位置预测第 t 帧的位置, 假设目标状态参数是相互独立的, 可用 3 个高斯分布来描述, 从而运动模型即为布朗运动, 于是

$$p(\mathbf{S}_t|\mathbf{S}_{t-1}) = N(\mathbf{S}_t|\mathbf{S}_{t-1}, \Sigma) \quad (9)$$

式中: $\Sigma = \text{diag}(\sigma_x, \sigma_y, \sigma_s)$ 为对角协方差矩阵; $p(\mathbf{S}_t|\mathbf{Z}_t)$ 为观测模型, 用于评估观测结果 \mathbf{Z}_t 与目标的相似性。观测模型通过测量样本与目标之间的相似度来计算:

$$p(\mathbf{Z}_t|\mathbf{S}_t) \propto e^{-\|\text{vec}(\mathbf{C}_t) - \text{vec}(\mathbf{C}_t^i)\|_2^2} \quad (10)$$

于是, 整个跟踪过程就是求最大响应:

$$\hat{\mathbf{S}}_t = \arg \max_{\{\mathbf{S}_t^i\}_{i=1}^N} p(\mathbf{Z}_t|\mathbf{S}_t^i)p(\mathbf{S}_t^i) \quad (11)$$

2.4 跟踪算法

前文给出了简单前馈卷积网络的特征提取方式, 并采用高斯核函数对卷积计算进行加速处理, 获取目标的深层次复杂表示。基于这种卷积特征, 结合粒子滤波, 提出的跟踪算法流程如图 2 所示。

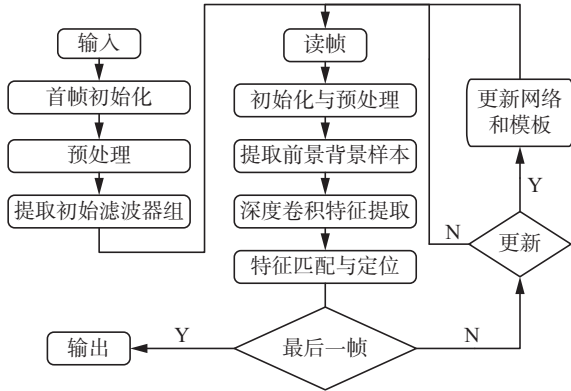


图 2 跟踪算法流程图

Fig. 2 Tracking flow chart

跟踪算法的主要步骤:

- 1) 输入: 输入视频序列, 并给定跟踪目标。
- 2) 初始化: 归一化, 粒子滤波, 网络规模, 样本容量等参数设置。
- 3) 初始滤波器提取: 利用第一帧的目标, 通过滑动窗口和 K-means 聚类提取一个初始滤波器组用作后续网络的滤波器使用。
- 4) 卷积特征提取: 利用上文的卷积网络结构提取出各候选样本的深层抽象特征, 并使用高斯核函数进行加速。
- 5) 粒子滤波: 按照粒子滤波算法, 归一化后生成规定尺寸大小的候选图片样本集, 并进行目标识

别与匹配。

6) 网络更新: 采取限定阈值的方式, 即当所有粒子中最高置信值低于阈值时, 认为目标特征发生较大表观变化, 当前网络已无法适应, 需要进行更新。利用初始滤波器组, 结合跟踪过程中得到前景滤波器组, 进行加权平均, 得到全新的卷积网络滤波器。

7) 模板更新: 以第一帧中目标的中心点为中心, 偏移量为 ± 1 个像素点范围内进行等尺寸采样, 构成正样本集合。以当前帧目标的远近两类距离采样, 构成负样本集合。跟踪过程中为了减轻漂移现象, 预设一个更新阈值 $f=5$, 目标模板每 5 帧更新一次。

3 实验结果与分析

本文利用 MATLAB2014a 编程环境, PC 配置为 Inter Core i3-3220, 3.3 GHz, 8 GB 内存, 根据 Database OTB2013^[3,21]中提供的测试视频序列对算法进行了仿真分析。本文仿真参数设置为: 滑动窗口取片尺寸为 6×6 , 滤波器个数为 100, 归一化尺寸为 32×32 , 学习因子设置为 0.95, 粒子滤波器的目标状态的标准偏差设置为: $\sigma_x = 4, \sigma_y = 4, \sigma_s = 0.01$, 使用 $N = 300$ 个粒子。

3.1 定性分析

限于论文篇幅, 本文仅给出几组代表性的跟踪实验结果。如图 3(a)、(b)、(c)、(d) 所示, 比较的算法有 CT^[7]、KCF^[15]、CNT^[18]与本文算法。图示给出 Crossing、Football、Walking、Walking2 四组序列, 均存在目标形变问题, 其中 Crossing、Walking、Walking2 均是在低分辨率场景下的跟踪, Football、Walking、Walking2 均存在局部遮挡问题。对于 Crossing 序列, 随着目标的运动导致目标本身的尺度变化, 在低分辨率监控场景中, 在第 45 帧, 目标在行进过程中面临光线的干扰, 并且出现运动车辆导致的背景干扰, 在所有比较的算法中, 同样跟踪成功, 本算法性能都能达到最优。对于 Football 序列, 目标在运动过程中, 一直伴随着大量的形变问题, 整个视频序列中大量的相似目标导致背景干扰问题, 在第 150 帧目标进入人群中导致局部遮挡问题, 本文算法在所有算法中表现最优。对于 Walking 序列, 目标在监控的低分辨率场景中, 目标运动过程中伴随着一定的尺度变换, 并在第 90 帧出现柱子遮挡目标的情况, 本文算法在所有算法中表现最优。对于 Walking2 序列, 在监控的低分辨率场景中, 目标运动途中伴随着尺度变换、遮挡、背景干扰。在第 190 帧与第 360 帧, 目标均面临相似目标的背景干扰与遮挡, 本文算法在所有算法中表现最优。

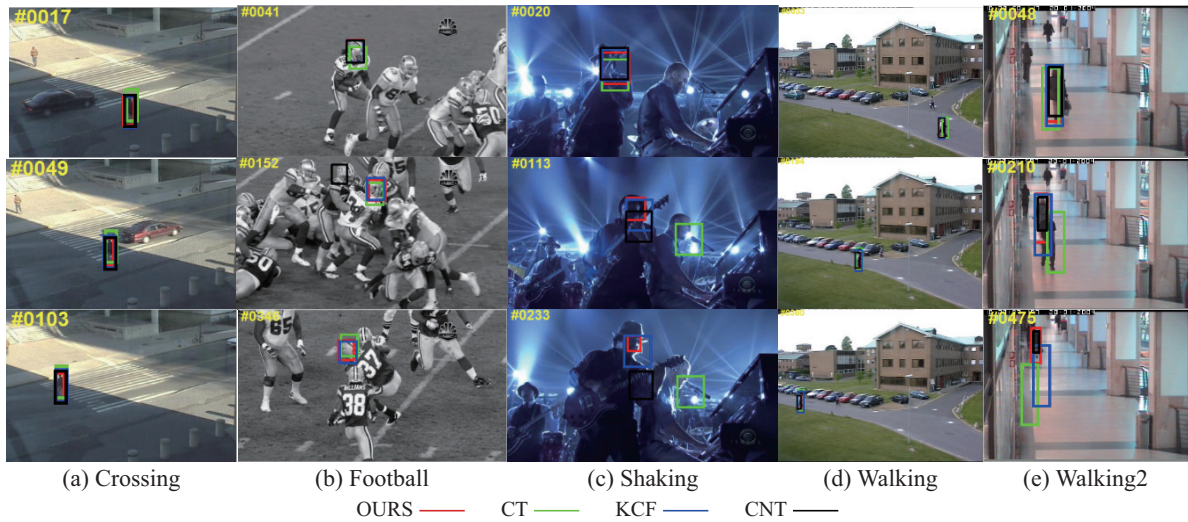


图3 视频序列跟踪结果示例

Fig. 3 Examples of the tracking results on video sequences

可见, 本文算法在形变、遮挡、低分辨率等复杂背景干扰下均能取得有效的跟踪效果。

3.2 定量分析

为了测试算法性能, 给出了部分序列的中心位置误差与距离精度的具体数据^[3,21]。中心位置误差 (center location error, CLE) 表示目标的中心位置与标准中心位置的欧氏距离的误差, 表达式为 $\varepsilon = \frac{1}{n} \sum_{i=1}^n \|C_i - C_i^*\|$, n 是图片序列的帧数, C_i 为目标中心位置, C_i^* 为标准中心位置; 距离精度 (distance precision, DP) 表示中心误差小于一个给定阈值的帧的相对数量, 表达式为 $DP = \frac{m}{n}$, m 为 CLE 小于某个固定阈值 (实验中均选为 20 像素) 的图片序列帧数。表 1 给出了算法的平均中心位置误差的对比数据, 表 2 给出了算法的距离精度 DP 的对比数据。为保证数据真实性, 本文所示数据均为 5 次实验后取平均值。

表 1 中心位置误差 (像素)

Table 1 Center location error(pixels)

| 视频序列 | OURS | CT | KCF | CNT |
|----------|------|-------|-------|-------|
| Crossing | 1.46 | 5.31 | 2.25 | 3.79 |
| Dancer2 | 6.11 | 9.51 | 6.41 | 7.14 |
| David2 | 2.24 | 15.4 | 2.08 | 2.56 |
| Football | 16.2 | 21.9 | 14.6 | 17 |
| Shaking | 18.8 | 110 | 113 | 45.9 |
| Skater | 11 | 7.24 | 10.7 | 11.3 |
| Walking | 2.11 | 4.3 | 3.97 | 2.21 |
| Walking2 | 3.44 | 61.3 | 29 | 3.78 |
| Average | 7.67 | 29.37 | 22.75 | 11.71 |

表 2 距离精度 DP

Table 2 Distance Precision

%

| 视频序列 | OURS | CT | KCF | CNT |
|----------|------|-------|-------|-------|
| Crossing | 100 | 100 | 100 | 100 |
| Dancer2 | 100 | 100 | 100 | 100 |
| David2 | 100 | 100 | 100 | 100 |
| Football | 92.8 | 99.2 | 91.4 | 82.6 |
| Shaking | 100 | 12.3 | 12.9 | 57.5 |
| Skater | 100 | 100 | 100 | 100 |
| Walking | 100 | 100 | 100 | 100 |
| Walking2 | 100 | 48.6 | 75.6 | 100 |
| Average | 99.1 | 82.51 | 84.98 | 92.51 |

对于算法速度, 同样采用卷积网络结构提取特征的 CNT 算法, 与本文算法在相同实验环境下进行速度对比, CNT 没有采用高斯核函数进行加速, 算法速度为 1~2 f/s, 本文算法采用高斯核函数进行加速, 算法平均速度为 5 f/s。由实验可知, 采用高斯核函数加速, 在不影响跟踪精度的同时能够提升算法的速度。

实验中, 绘制了 4 种算法跟踪精度曲线图^[4], 跟踪精度曲线图首先设定一个目标估计位置与真实位置的阈值距离, 在跟踪过程中, 统计跟踪算法估计的目标位置与真实位置的距离小于阈值范围的帧数, 并计算帧数占整个视频帧的百分比。图 4 给出了以上 4 种算法对应的 8 个视频序列的跟踪精度曲线, 横坐标为阈值, 纵坐标为精度, 阈值越低而精度值越高的跟踪器性能越好。由曲线图可清晰地看到, 本文算法具有较高的跟踪精度。

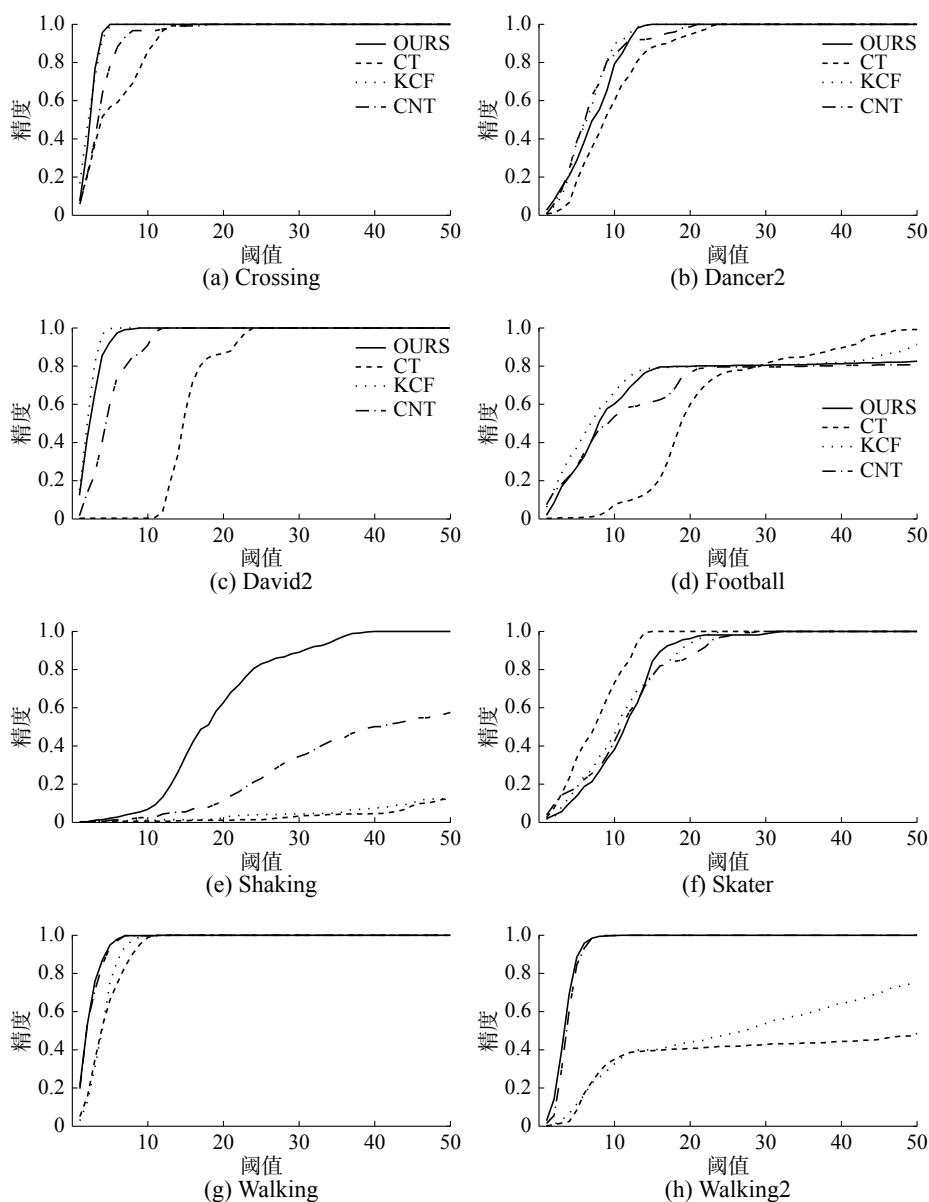


图4 跟踪精度曲线图

Fig. 4 Tracking accuracy curve

4 结束语

本文针对深度学习跟踪算法训练费时, 硬件要求高等问题, 采用高斯核函数加速计算, 采用简单两层前馈卷积网络提取目标鲁棒性特征, 基于简化的卷积神经网络提出跟踪算法, 第一层利用 K-means 在第一帧中提取归一化图像块作为滤波器组提取目标的简单层特征, 第二层将简单的单元特征图堆叠形成一个复杂的特征映射, 并编码目标的局部结构位置信息, 在粒子滤波框架下, 在目标形变、遮挡、低分辨等场景下, 脱离深度学习复杂的硬件环境, 仍能取得较好跟踪效果。因为本文的特征提取方式采用卷积神经网络特征, 所以本文算法在快速运动、目标出界等场景下仍面临很大挑战, 在今后的工作中, 将主要致力于解决此类场景的跟踪问题。

参考文献:

- [1] 杨戈, 刘宏. 视觉跟踪算法综述[J]. 智能系统学报, 2010, 5(2): 95-105.
YANG Ge, LIU Hong. Survey of visual tracking algorithms [J]. CAAI transactions on intelligent systems, 2010, 5(2): 95-105.
- [2] 黄凯奇, 陈晓棠, 康运锋, 等. 智能视频监控技术综述[J]. 计算机学报, 2015, 38(6): 1093-1118.
HUANG Kaiqi, CHEN Xiaotang, KANG Yunfeng, et al. Intelligent visual surveillance: a review[J]. Chinese journal of computers, 2015, 38(6): 1093-1118.
- [3] WU Yi, LIM J, YANG M H. Online object tracking: a benchmark[C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA, 2013: 2411-2418.

- [4] BABENKO B, YANG M H, BELONGIE S. Robust object tracking with online multiple instance learning[J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 33(8): 1619–1632.
- [5] 陈真, 王钊. 元认知粒子滤波目标跟踪算法[J]. 智能系统学报, 2015, 10(3): 387–392.
- CHEN Zhen, WANG Zhao. Object tracking algorithm with metacognitive model-based particle filters[J]. CAAI transactions on intelligent systems, 2015, 10(3): 387–392.
- [6] MEI Xue, LING Haibin. Robust visual tracking using ℓ_1 minimization[C]//Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan, 2009: 1436–1443.
- [7] ZHANG Kaihua, ZHANG Lei, YANG M H. Real-time compressive tracking[C]//Proceedings of the 12th European Conference on Computer Vision. Berlin, Germany, 2012: 864–877.
- [8] 韩华, 丁永生, 郝矿荣. 综合颜色和小波纹理特征的免疫粒子滤波视觉跟踪[J]. 智能系统学报, 2011, 6(4): 289–294.
- HAN Hua, DING Yongsheng, HAO Kuangrong. An immune particle filter video tracking method based on color and wavelet texture[J]. CAAI transactions on intelligent systems, 2011, 6(4): 289–294.
- [9] RABINER L R. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceedings of the IEEE, 1989, 77(2): 257–286.
- [10] BAR-SHALOM Y, FORTMANN T E, CABLE P G. Tracking and data association[J]. The journal of the acoustical society of America, 1990, 87(2): 918–919.
- [11] COMANICIU D, RAMESH V, MEER P. Real-time tracking of non-rigid objects using mean shift[C]//Proceedings of 2000 IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head Island, SC, USA, 2000: 142–149.
- [12] ISARD M, BLAKE A. CONDENSATION—conditional density propagation for visual tracking[J]. International journal of computer vision, 1998, 29(1): 5–28.
- [13] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters [C]//Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA, 2010: 2544–2550.
- [14] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]//Proceedings of the 12th European Conference on Computer Vision. Berlin, Germany, 2012: 702–715.
- [15] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(3): 583–596.
- [16] 余凯, 贾磊, 陈雨强, 等. 深度学习的昨天、今天和明天 [J]. 计算机研究与发展, 2013, 50(9): 1799–1804.
- YU Kai, JIA Lei, CHEN Yuqiang, et al. Deep learning: yesterday, today, and tomorrow[J]. Journal of computer research and development, 2013, 50(9): 1799–1804.
- [17] WANG Naiyan, YEUNG D Y. Learning a deep compact image representation for visual tracking[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2013: 809–817.
- [18] ZHANG Kaihua, LIU Qingshan, WU Yi, et al. Robust visual tracking via convolutional networks without training[J]. IEEE transactions on image processing, 2016, 25(4): 1779–1792.
- [19] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, CA, USA, 2016: 4293–4302.
- [20] ROSS D A, LIM J, LIN R S, et al. Incremental learning for robust visual tracking[J]. International journal of computer vision, 2008, 77(1/2/3): 125–141.
- [21] WU Yi, LIM J, YANG M H. Object tracking benchmark [J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1834–1848.

作者简介:



汪鸿翔, 男, 1992 年生, 硕士研究生, 主要研究方向为视频、图像处理、视觉跟踪、深度学习相关算法。



柳培忠, 男, 1976 年生, 副教授, 美国杜克大学高级访问学者, 博士, 主要研究方向为仿生智能计算、仿生图像处理技术、多维空间仿生信息学。



骆炎民, 男, 1975 年生, 副教授, 博士, 主要研究方向为智能图像处理、机器学习。