DOI:10.16136/j.joel.2023.06.0446

# 基于显著性检测的星载嵌入式跟踪系统

方歆楠,李 奇\*,冯华君,徐之海

(浙江大学现代光学仪器国家重点实验室,浙江杭州 310007)

摘要:随着航天科技的不断发展,计算机视觉算法在卫星上的应用方兴未支,为了实现更多的功能需求和应对可能的威胁,视觉目标跟踪作为其中基础但具有挑战性的任务更是至关重要。然而,目前已有的目标跟踪算法大多数算法只限于对图像序列进行跟踪。另一方面,受到硬件条件制约,很多优秀的算法因为复杂度较高很少被应用到星载嵌入式系统中。这些目标跟踪算法运行时,通常需要人为地给出目标的边界框。为了自动得到边界框,需要目标检测算法对输入图像进行运动目标检测。本文提出了一种基于显著性检测和相关滤波的单目标检测与跟踪一体化算法,并与嵌入式系统相结合,在搭载的 TMS320C6678 芯片上达到了 2048 pixel×2048 pixel 分辨率下 24 fps 的帧率。具体地,检测算法负责对图像进行预处理并获得边界框,然后目标跟踪算法给出目标在后续帧中的位置。为了验证算法在实际跟踪中的有效性,本研究搭建了一个由相机、DSP 和 云台组成的光学平台并进行了实验验证。在该系统中,DSP 自动完成检测、跟踪、驱动云台和再检测任务,达到了很好的检测跟踪效果。

关键词:星载嵌入式系统;显著性检测;运动目标检测;相关滤波;一体化算法 中图分类号:TP391.4;TP311.1 文献标识码:A 文章编号:1005-0086(2023)06-0576-10

# Satellite-borne embedded tracking system based on salient detection

FANG Xinnan, LI Qi\*, FENG Huajun, XU Zhihai

(State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, Zhejiang 310027, China)

Abstract: With the development of aerospace science and technology, the application of computer vision algorithms on satellites is in the ascendant. In order to achieve more functional requirements and deal with possible threats, visual target tracking is a fundamental but challenging task. However, most existing algorithms only operate on pure image sequences. On the other hand, restricted by hardware conditions, few excellent algorithms are applied to satellite-borne embedded systems because of their high complexity. These pure object tracking algorithms usually need to artificially give the bounding box of the object. To get the bounding box automatically, an object detection algorithm is necessary. We propose an integrated detection and tracking algorithm for single target based on saliency detection and correlation filtering, which achieves 24 fps on DSP TMS320C6678 for the resolution of 2048 pixel×2048 pixel. The detection algorithm is responsible for preprocessing the image and obtaining the bounding box, then the object tracking algorithm during the actual tracking, in this research, we build an optical platform consisting of a camera, DSP and Pan/Tilt/Zoom(PTZ) and test it. In this system, the DSP automatically completes the tasks of detection, tracking, driving the PTZ and re-detection, and achieves good results of detection and tracking.

Key words: satellite-borne embedded system; salient detection; moving object detection; correlation filtering; integrated algorithm

# 0 引 言

随着商用航天的迅猛发展,在轨航天器数量 将呈爆炸式增长。以星链为例,近几年已发射了 近2000颗,2019年以来,已数次对其他航天器造 成风险,而未来星链还将构建多达四万多颗卫星 的庞大系统。考虑到这些航天器多为非合作空间 目标,为保障我国航天器的安全,尤其是高价值空 间基础设施的安全,迫切需要实现在轨的空间目 标探测与跟踪算法及功能,提升空间环境态势感 知能力。

在软硬件条件的不断升级下,目标跟踪算法 不再仅仅局限于图像跟踪,嵌入式的目标跟踪已 经得到了实际应用。光轴跟踪,指的是通过搭载 于云台上的相机获取物理空间图像,将图像传输 给主机后,由主机计算给出视场内的运动目标坐 标,并将信号发送给云台,驱动云台将相机光轴对 准运动目标的物理跟踪过程,从而在进行图像跟 踪时将跟踪目标始终对准在视场中央,达到更好 的观察效果,在监控领域有很大的应用空间。光 轴跟踪任务主要由目标检测和目标跟踪两部分组 成。目标检测用以在视场中发现目标并给出位置 信息,跟踪算法则计算目标在像面中的位置变化 以及控制运动平台以对准目标达到持续的跟踪 效果。

运动目标检测算法主要有帧差法、背景差分 法、光流法等。帧差法速度快,通过图像差分可以 检测出运动目标的轮廓,但是精度较低,且容易受 到背景运动的影响;背景差分法通过对场景背景 进行建模,算法精度受到背景复杂程度影响;光流 法精度高,但是计算复杂度高,难以做到实时。

目前的跟踪算法主要分为基于相关滤波<sup>[1]</sup>的 传统方法和基于深度学习的网络跟踪方法。相关 滤波的原理在于,两个信号越相似,其相关性越 高。在跟踪过程中,目标的变化是在相邻帧之间 的变化很小,根据以上先验,就可以通过在相邻帧 之间寻找与目标模板相关响应最大的峰值位置来 得到目标的位置,BOLME<sup>[1]</sup>等通过将相关运算转 换到频域上的点乘而大大提升了运算速度。为了 进一步提高跟踪精度,HENRIQUES等提出了利 用核函数映射和循环矩阵加速的 CSK 算法,使得 相关滤波算法效果有了很大的提升。后续的 KCF<sup>[2]</sup>、Staple<sup>[3]</sup>、DSST<sup>[4]</sup>等算法通过引入更鲁棒 的特征、尺度池等方式提高了跟踪的准确度,在图 像跟踪上取得了很好的效果。

深度学习的跟踪方法得益于硬件条件的不断 提升,在跟踪领域获得了极大发展。基于孪生网 络的方法将新一帧图像与目标模板输入到分支相同的网络,通过特征匹配实现跟踪,通过更复杂更鲁棒的特征和语义信息可以获取更好的效果。SiameseRPN<sup>[5]</sup>、Dasiamrpn<sup>[6]</sup>、Siamrpn++<sup>[7]</sup>、AT-OM<sup>[8]</sup>等是其代表。另一方面,基于分类卷积神经网络(convolutional neural network, CNN)的算法则将跟踪过程作为目标和背景的分类问题。多域网络<sup>[9]</sup>算法根据目标位置来建立正负样本进行分类,通过在线更新样本集达到了很高的精度,但缺点是复杂度很高。此外结合了 CNN 的特征提取能力以及相关滤波算法快速跟踪的特点提出的ECO、MCCT等在保证算法准确度的同时提高了速度。

很多基于深度学习的方法在跟踪领域的精度 和速度都达到了很高的水平,在具有图形处理器 (graphics processing unit,GPU)的嵌入式系统上 可以达到很好的效果。在目前主流的星载嵌入式 系统中,通常使用数字信号处理器(digital signal processor,DSP)和现场可编程逻辑门阵列(field programmable gate array,FPGA)的架构,由于其硬 件条件的限制,基于相关滤波的 KCF 等算法在实 际部署中得到了普遍的应用。

在图像跟踪算法中,都需要人为地给出目标 的边界框,为了能自动得到目标边界框并进行后 续跟踪任务,本文开发了适用于嵌入式系统的检 测跟踪一体化的算法技术并搭建了系统验证了算 法的可行性,采用多核心并行检测方式提升算法 效率。达到了检测、跟踪、机械控制一体化的效 果,对于接近的非合作目标和需要观察的合作目 标有很好的观察效果,推进了图像检测、跟踪技术 在实际场景中的应用。

# 1 理论推导

#### 1.1 运动目标检测算法

对于高轨运行的卫星,成像方式通常可以分 为常规曝光成像和凝视曝光成像。凝视曝光方式 可以在弱光照条件下,通过卫星姿态控制和微位 移补偿的方式获得更好的成像质量,但是对于运 动目标,会导致目标像移量大于像元尺寸,成像质 量下降。常规曝光成像曝光时间短,在多帧曝光 时由于高轨卫星轨道稳定,短时间内通常可以认 为背景静止,更适应对运动目标的观察。本文研 究基于常规曝光成像设计的系统和实验。

靠近的运动目标相对于背景来说,通常具有静态显著性,静态显著性指的是目标与背景相比 在颜色、亮度以及纹理信息等方面存在较大差异。 基于以上先验知识,本文提出了一种基于帧间差 分和目标颜色显著性图谱的运动目标识别算法。

对于一个序列中相邻的两帧图像  $I_1$ 、 $I_2$ ,得 到它们的差分图像公式如下:

$$I_{\rm diff} = \Phi(I_1) - \Phi(I_2) , \qquad (1)$$

$$I_{\rm tho} = f(I_{\rm diff}, I_{\rm tho}) , \qquad (2$$

式中,  $I_1$ 和  $I_2$ 分别表示输入的两张图像,为 RGB 三通道格式,若为灰度图像,则将其复制三通道格 式;  $\phi$ 表示对输入图像进行特征提取;  $\phi(\cdot)$ 可以 使用不同的特征<sup>[10]</sup>适应具有不同性质的目标,如 目标的颜色特征、卷积特征等,当仅使用目标的亮 度特征时,  $\phi(\cdot)$ 为高斯模糊函数,采用的模糊核 大小为 5×5;  $I_{diff}$ 表示差分结果;  $f(\cdot)$ 表示特征 二值化操作,为二值化阈值;  $I_{tho}$ 表示二值化结果; 将二值化结果与输入图像点乘得到显著运动区域 对应的图像:

 $I_{\mathrm{mot}}=I_{\mathrm{tho}}$  •  $I_{\mathrm{2}}$  ,

简单的差分会导致提取到的运动区域包含较 多的背景噪声。JIANG等<sup>[11]</sup>提出,池化层的添加 与否对检测结果的噪声干扰很大。尤其在背景有 较大移动时,不添加池化层,通常会在背景处检测 到较强的运动。由于 maxpooling 操作选取区域内 的最大值代表区域的特征,因此可以有效抑制场 景中微小的形变和背景的微小抖动。通过对输入 图像高斯模糊后再进行相减,设置 3 层网络,每层 有一个侧输出来提取相减结果。将高斯模糊后的 图像池化后再传入下一层,再次进行高斯模糊。 将多层侧输出上采样到原图大小,进行累加来获 得最终的运动区域,如图 1 所示。



Fig. 1 Pooling difference process

对输入图像 I<sub>2</sub> 和 I<sub>mot</sub> 进行颜色直方图统计,并 用最大后验概率公式计算得到颜色直方图中各种颜 色属于运动目标的概率:

$$H_{\text{target}} = P(H_{\text{mot}}, H_2) , \qquad (4)$$

$$P(H_{\text{mot}}, H_2) = \frac{P(c \mid mot) \times P(mot)}{P(c)}, \quad (5)$$

式中, $H_{mot}$ 和 $H_2$ 分别代表运动区域和输入图像 $I_2$ 的颜色直方图统计,c表示像素在颜色直方图中的颜 色索引, $P(c \mid mot)$ 表示当前颜色出现在显著运动区 域内的概率, $P_{mot}$ 表示像素属于运动区域的概率,可 以通过运动区域 $I_{mot}$ 与 $I_2$ 的像素数目之比得到, P(c)代表当前颜色在输入图像 I<sub>2</sub>中出现的概率。 具体地,在颜色直方图统计时,将每个通道均匀划分为16份,所以得到的颜色直方图大小为1×4096的向量,得到 H<sub>target</sub> 之后,逐像素判断输入图像 I<sub>2</sub> 的颜 色索引并根据 H<sub>target</sub> 赋值得到原图的目标概率图 I<sub>target</sub>。为了防止大量像素 RGB 值被归类于(0,0,0) 影响其他颜色的直方图归一化,需要单独对向量的 第一个元素进行置0。

根据得到的  $I_{mot}$  与  $I_{target}$  建立初始的目标检测 框。具体地,对于检测框中心的确定,通过循环腐蚀 的方法确定目标的中心,使用直径为1的模板在显 著运动区域  $I_{mot}$  上进行形态学滤波的腐蚀,直至最后 仅余一个像素,根据当前像素的位置定位检测框的 中心位置(当最后一次腐蚀之后剩余像素数目为0 时,使用前一次腐蚀结果的重心):

$$P_{\rm cen}(x,y) = \frac{\sum_{p \in I_{\rm mot}} p(x,y)}{num} , \qquad (6)$$

式中,  $P_{cen}(x,y)$  表示重心的坐标, p(x,y) 表示二值 化后判别为1的像素坐标, num 为当前 $I_{mot}$ 中值为1 像素的数目。为了进一步提升腐蚀效率:

$$I_{\text{candidate}} = Rect(I_{\text{mot}}) , \qquad (7)$$

式中, *I*<sub>candidate</sub> 为循环腐蚀的作用区域,即当前的 *I*<sub>mot</sub> 的外接矩,*Rect*(•)表示取外接矩操作。目标区域往 往在整个像面中的大小占比不大,因此对全像面的 循环腐蚀往往会增加不必要的开销,将腐蚀的作用 区域限定在运动目标区域的外接矩上可以很好地解 决这个问题,只需记录当前腐蚀操作后 *I*<sub>mot</sub> 像素位置 (*x*,*y*)的最大和最小值就可以得到。

目标检测框的宽高 w、h 由初始 I<sub>mot</sub> 像素值不为 零的区域外接矩大小来表示,由于 I<sub>mot</sub> 由相邻图像特 征差分得到,对于纹理信息不强、帧间差别不大的情 况会导致 I<sub>mot</sub> 中值为1的像素数目偏小,使用外接矩 能够减小因差分导致的误差。通过特征差分图像的 行投影和列投影来快速得到初始检测框:

$$Rect(I_{mot}) = (x_0, y_0, x_1, y_1)$$

$$x_0 = \min(hist_{col}(I_{diff}) \neq 0)$$

$$y_0 = \min(hist_{row}(I_{diff}) \neq 0)$$

$$x_1 = \max(hist_{col}(I_{diff}) \neq 0)$$

$$y_1 = \max(hist_{row}(I_{diff}) \neq 0)$$
(8)

为了得到最优的检测框,使用梯度下降法对当 前检测框进行迭代优化:

$$x = lr \cdot \frac{S(x, y, w, h) - S(x + \Delta x, y, w, h)}{\Delta x},$$

$$y = lr \cdot \frac{S(x, y, w, h) - S(x, y + \Delta y, w, h)}{\Delta y},$$
$$w = lr \cdot \frac{S(x, y, w, h) - S(x, y, w + \Delta w, h)}{\Delta w},$$
$$h = lr \cdot \frac{S(x, y, w, h) - S(x, y, w, h + \Delta h)}{\Delta h}, \quad (9)$$

 $S(x,y,w,h) = \alpha * sum(I_{target}) + \beta * IOU,$  (10) 式中, br 表示学习率,  $\Delta x \cdot \Delta y \cdot \Delta w \cdot \Delta h$  分别表示 对矩形框  $x \cdot y \cdot w \cdot h$  的随机扰动,在 1—10 之间随机 取值, S(x,y,w,h) 表示矩形框得分计算式,主要由 颜色置信图谱  $I_{target}$  在当前边界框 (x,y,w,h) 内的 得分以及当前边界框与初始边界框  $(x_0,y_0,w_0,h_0)$ 之间的交并比(intersection over union, IOU)决定,  $\alpha = \beta$  为权重系数,可以认为两项分别代表了颜色显 著性和运动置信度的影响。迭代循环上述优化过程 直至边界框 (x,y,w,h) 不变,或者迭代次数达到设 定值时停止循环,得到的结果即为最终求得的运动 目标检测框。为了得到更稳定的检测结果,对连续 多帧检测框进行比较<sup>[12]</sup>,只有相邻差分框之间位置 变化持续 5 帧大于框长度的 10%,才认为视场中存 在运动目标。

#### 1.2 运动目标跟踪算法

本文提出了改进的 DSST<sup>[4]</sup>算法,在保证了跟踪 精度的情况下,提升了跟踪算法的速度。DSST 基于 相关滤波的原理,通过提取图像的梯度方向直方图 (histogram of oriented gradients, HOG)特征,利用 前一帧模板和当前帧的最大相关响应来得到目标在 当前帧中的位置。通过最小化损失函数来得到最佳 滤波器:

$$\boldsymbol{\varepsilon} = \| \sum_{l=1}^{d} h_{1}^{l} * f - g \| + \lambda \sum_{l=1}^{d} \| h_{1}^{l} \|^{2}, \quad (11)$$

式中, *f* 为*d* 维特征, *h*<sub>1</sub> 为对应的*d* 维度滤波器, *l* 为 第 *l* 维度特征, \* 为相关操作, *g* 为对应的相关输出, 此处为峰值在 *f* 目标中心的高斯模板, λ 为正则化系 数, 由上式可以得到最佳相关滤波器的解为:

$$H^{l} = \frac{\overline{G}F^{2}}{\sum_{k=1}^{d} \overline{F^{k}}F^{k} + \lambda},$$
(12)

式中, H、F 分别为 h<sub>1</sub>、f 的傅里叶变换结果, 上划线 代表其复共轭。上式中的分子分母分别为 A, B。可以 给出 A, B更新策略:

$$A_{t}^{l} = (1 - \eta)A_{t-1}^{l} + \eta \,\overline{G}_{t}F_{t}^{l}, \qquad (13)$$

$$B_{t} = (1 - \eta) B_{t-1}^{l} + \eta \sum_{k=1}^{a} \overline{F}_{t}^{k} F_{t}^{k}, \qquad (14)$$

式中,t为当前帧,t-1为前一帧, $\eta$ 为学习率,由此可以得到相关响应图 y:

$$y = \mathscr{F}^{-1} \{ \overline{H}_{t} Z \} =$$
$$\mathscr{F}^{-1} \left\{ \frac{\sum_{l=1}^{d} \overline{A}^{l} Z^{l}}{B + \lambda} \right\}, \tag{15}$$

式中, *F*<sup>1</sup>为傅里叶逆变换, *Z*为当前计算的图像块, *y*最大的位置即为当前帧中目标所在的位置。

对于具有尺度变化的运动目标,跟踪过程中,首 先提取模板位置处的特征信息,通过式(15)计算当 前帧中的位置响应图,在得到最大响应位置后 P,以 P为目标中心位置,计算不同尺度的目标大小与模板 的相关响应,获取响应最大的空间尺度参数,从而确 定目标在当前帧中的位置和大小。原文中设置了 S =33个尺度,尺度参数为1.02。本文对 DSST 算法 的尺度池进行了简化,将33个尺度缩小为9个,尺度 参数为 1.08;将目标模板大小重新采样为 32 pixel× 32 pixel 或 64 pixel×64 pixel,从而提高在 DSP 中进 行快速傅里叶变换的效率。通常相关滤波跟踪算法 的复杂度要远低于高分辨率下的检测算法。针对在 跟踪过程中可能出现的目标形变和遮挡,由文献 [13]可以得到两者对于相关响应图的影响是不同 的,目标被遮挡时,响应图峰值变小但波动很大;当 目标产生形变时,响应图峰值变小但波动较小,通过 计算相关响应图的梯度和与峰值响应的关系来 判断。

 $RPV = 0.0092 \cdot \ln(NAG - 104)$ , (16) 式中, RPV表示峰值响应, NAG表示归一化的梯度 和。当实际峰值响应高于式(16)的 RPV时,认为此 时处于形变状态;反之,则认为为遮挡状态。对目标 遮挡情况采用较低的更新率,对于目标形变时采用 较高的更新率以适应不同情况。

#### 1.3 重检测模块

在跟踪进行的过程中,由于目标运动方式的不确定性,以及可能存在的目标形变、遮挡、光照变化等<sup>[14]</sup>,会导致跟踪目标的模板和实际目标的相关响应度下降。当跟踪置信度低于设定的阈值时,应当认定为跟踪失败。此时不应该继续驱动光学平台转动进行跟踪,而是重新开始检测流程。在长时间的跟踪中,模板更新会将复杂的背景信息混入目标跟踪框中,同样会导致跟踪模板和实际目标的相关响应度下降,从而影响跟踪效果,因此在多帧的跟踪之后,需要对目标进行重检测,获取最新的目标模板。

# 2 硬件设计

#### 2.1 系统架构和原理

实时检测跟踪系统主要由光学相机、信号处理

模块、运动控制模块 3 部分组成,其中信号处理模块 分为 FPGA 模块与 DSP 模块。其主要架构如图 2 所示,感光芯片通过光学相机采集光强信号,通过 SPI 总线将数据传输给 FPGA 模块,在 FPGA 中对 数据进行基本的图像处理(伽马矫正、直方图统计 等)之后交由 DSP 芯片进行处理,也可以通过 Camlink 将图像上传到上位机进行查看。DSP 芯片对图 像进行检测、跟踪之后输出结果,通过串口控制云台 转过相应角度,并实时获取云台当前位置。将目标 对准在像面中间,达到光轴跟踪的效果。



图 2 系统结构图 Fig. 2 System structure diagram

#### 2.2 信号处理模块与运动控制模块原理

TMS320C6678 最大工作频率为 1.25 GHz。内 置 8 个核心,可以同时进行不同的任务,从而提高计 算效率。在本系统中,算法主要分为检测与跟踪两 部分,上一节中已经指出,检测算法的计算量要远高 于跟踪算法。为了提高 DSP 的处理效率,达到实时 的光轴跟踪,需要对 DSP 的 8 个核心进行任务分配。

检测与跟踪过程具有非常强的时序性,即必须 先检测,获取跟踪框后开始跟踪。在应用中,每个核 心性能存在差异,处理相同任务所需的时间是不同 的,因此,给每个核心分配同样的检测、跟踪任务是 不现实的。基于上述分析,本研究将 6 个核心分配 进行检测任务,一个核心负责跟踪任务。多核工作 原理见图 3。

CORE0 master	
CORE 1 CORE 2 CORE 3 CORE 4 CORE 5 CORE 6	CORE 7
Detection	Tracking

# 图 3 多核工作原理 Fig. 3 Multi core working principle

其中,CORE0 作为主控核,对当前系统状态进行判断,例如,若当前状态为检测,则通知 CORE1— CORE6 读取图像并进行处理,CORE7 初始化后不进行运算;若当前状态为跟踪过程,则 CORE7 读取 图像并进行运算,CORE1—CORE6 处于等待状态。

CORE0 以向其他核心发送核间中断的方式进 行激活,当从核任务完成后,通过改变共享内存中变 量的值通知 CORE0,并由 CORE0 通知下一个核心。 感光芯片得到的信号首先由 FPGA 进行处理。图像 差分任务由 FPGA 完成,并将差分结果的最后两行 改为当前差分图像的横向和纵向直方图统计。FP-GA将当前帧原图、原图与上一帧的差分图以及差分 图的横纵向直方图统计传递给 DSP。由于图像大小 为 2048 pixel×2048 pixel,在 DSP 中进行遍历的代 价是相当高昂的,此方法的优点在于不需要在 DSP 中进行全图遍历得到目标所在区域,具体的实现方 式为开辟两块内存地址,FPGA 通过给定的帧率在 两个地址中进行交替储存,DSP 在主控核 CORE0 的 控制下,各个核心按顺序读取当前 FPGA 传输数据 所在的内存地址,在当前核心读取完毕之后向主控 核发送信号,主控核心通知下一个核进行数据读取。 由于信号的读取需要时间,在 DSP 核心进行读取时 原地址不可进行改写,否则会导致图像出现撕裂等 情况,因此需要协调 DSP 处理和 FPGA 数据传输速 率。实际上 FPGA 在进行每一帧图像时耗费的时间 很小,图像传输速率主要受制于图像曝光时间,在正 确曝光的情况下,传输速率仍然高于 DSP 处理图像 的速率,因此需要限制 FPGA 传输图像的速率。由 于 DSP 的运算复杂度主要集中在检测任务,如图 3 所示,用6个核心并行处理检测任务。

运动控制模块主要由 DSP 芯片与双轴云台组 成。DSP 进行信号处理之后将需要转动的角度以串 口的形式发送给云台,同时从云台反传当前云台坐 标系中相机的角度,由于当前相机位置是由云台反 传得到,从而消除了位置的累计误差,实现精确的连 续的光轴跟踪。

# 3 实验与结果分析

系统中采用的相机的水平和竖直视场角均为 40°,在边缘视场存在较大的畸变。由于实时光轴跟 踪的性质,在跟踪过程中可以认为目标始终位于相 机视场的中心部分约 300 pixel×300 pixel 大小的区 域内,目标在中心视场内的畸变可以忽略。在检测 过程中,只有在最初检测到目标时目标可能处于视 场的边缘,由于在驱动云台对准目标的过程中会有 多次修正,因此可以忽略边缘视场的畸变,从而通过 线性变换获得像面像素移动量和云台转动角度的对 应关系。

#### 3.1 检测算法

#### 3.1.1 数据集验证

在 OTB50 和实拍数据集上分别对本文算法进 行了验证。由于数据集的帧率不同,将图像序列的 前 5 或 10 张作为输入,在相邻帧之间利用检测算法 给出初始的跟踪框,并取得分最高的作为最后的检 测结果。针对单目标跟踪场景,实拍数据集通常由 运动目标前景和背景构成,背景中通常有干扰信息, 对 OTB50 与实拍数据集共 33 个图像序列上进行了 验证。OTB50 数据集中选取背景稳定的序列作为检 测输入,对序列的真值进行了重新标定。检测结果 如图 4 所示。从上到下依次为序列中相邻的原始帧 1,2 和显著性图谱局部放大图,图谱的灰度值代表目 标的显著性,灰度值越大对应的显著性得分越高。 实线边界框代表真实的目标边界框,虚线边界框为 差分给出的初始检测框,点划线边界框为检测算法 优化后的边界框。其中图像序列(a)、(b)来自 OTB50数据集,图像序列(c)、(d)来自实拍数据集。

序列(a)中,检测算法能够将自行车及车手较为 完整地框选出来,轮胎部分的显著性较弱,从而在梯 度下降优化时选框收缩;序列(b)、(c)中运动目标与 背景显著性图谱得分差别较大,算法能够很好地将 目标框选出来。序列(d)中,腿部部分黑色是由于为 了避免直方图归一化时压缩其他颜色的值,对 RGB 值为(0,0,0)元素进行了扣除。

检测算法能够有效地排除由背景帧差引进的干扰,在运动目标与背景颜色相差较大的情况下可以取得较好的效果,在目标颜色与背景相近时,优化函数式(10)中的 IOU 部分起主要作用。

当检测算法给出的边界框与真实边界框 IOU 大 于阈值 50% 时,可以认定为检测成功。本文在 OTB50 数据集和实拍数据集共 33 个序列上验证了 算法,检测结果表明算法在运动目标检测中对目标 的检出率达到 91%。

3.1.2 卫星目标检测

在实验室环境下对卫星目标进行了检测,检测 效果如图 5 所示。

图5中,上图为卫星实物,对应的下图为显著性



图 4 检测算法结果 Fig. 4 Result of detection algorithm



图 5 卫星检测结果 Fig. 5 Satellite detection results

检测图谱,虚线框为初始的检测框,实线为优化后的 检测框。初始检测框往往受到噪声等干扰而偏离较 大,相较于简单帧差的结果,本文检测算法对于姿态 变换的卫星目标检测效果更贴近于实际情况,能够 很好地对目标整体进行检测。

### 3.2 跟踪算法

为了验证跟踪算法的有效性,在 OTB50 数据集 上将本文算法与其他相关滤波 CSK、KCF<sup>[2]</sup>、DCF、 DSST、SAMF 等相关滤波器进行了测试和比较。 CSK 算法利用了循环矩阵和核函数来计算运动目标 在下一帧中的位置。KCF/DCF 采用 HOG 特征取 代 CSK 中的灰度特征,利用循环矩阵对角化来提高 运算效率;DSST 算法利用尺度池的方法进一步解决 了 KCF 中检测框形状不变的问题,提升了算法对尺 度变化目标的跟踪效果。SAMF 采用灰度、颜色和 HOG 特征的融合特征来进行跟踪,进一步提高了算 法的精度。跟踪算法效果如图 6 所示。

图 6(a)展示了图像序列中,成功跟踪帧数与总 序列长度的比值随跟踪边界框中心与 groundtruth 中心位置像素差阈值变化曲线图,图例中括号内数 字为阈值设定为 20 pixel 时的精度。图 6(b)展示了 图像序列中,成功跟踪帧数与总序列长度的比值随 跟踪边界框与 groundtruth 框的 *IOU* 阈值变化曲线 图,图例中括号内数字为曲线下面积(area under curve,AUC)。

表 1 为不同跟踪算法在同一环境下的运行速度 比较,设备处理器为Intel Core i 5 - 9400 F CPU @ 2.90 GHz,操作系统 WIN10 64 bit。

在经过特征裁剪和尺度裁剪后,算法相较于 DSST效果有一定的下降,在算法精度阈值为







Fig. 6 Precision and success rate diagram

of tracking algorithm

表1 跟踪算法速度比较

Tab. 1 Speed comparison of tracking algorithm

Name	SAMF	DSST	KCF	DCF	CSK	Ours
Value /fps	32.0	63.4	426.9	682.6	525.3	101.1

20 pixel时,DSST 精度为 0.700,本文算法为 0.633, 成功率曲线下面积 DSST 为 0.514,本文算法为 0.461。整体效果与 KCF 算法相近,但由于 KCF 目 标模板固定,本文算法在目标尺度变化的情况下具 有优势。在速度方面,优化后的算法相较于 DSST 提升约 60%。

#### 3.3 嵌入式跟踪系统

由图 3 可知,共有 6 个核心用于目标的检测任务。出于稳定性考虑将 FPGA 传输速率设置为 6 的 倍数,达到 24 fps。

嵌入式跟踪系统效果受到目标大小、目标运动 速度、目标显著性等因素影响,本文在实验室条件下 对上述因素进行了实验验证。

嵌入式跟踪系统实验将目标分为实物目标和视频目标,实物目标架设在一维运动导轨上,通过程序 设置目标的速度和方向,用以研究跟踪成功率与目标运动速度(像面速度)的关系。由于在相机对焦景 深内,物体大小变化通常很小,且在同一时间由光照 变化引起的显著性变化不明显,所以本实验利用大 屏幕投影视频运动目标的方式来实现运动目标大小 和显著性变化以及运动过程中可能出现的旋转形变 等。嵌入式系统云台参数见表 2。

表 2 云台运动参数

Tab. 2 Basic parameters of the PZT

Parameters	Value
Horizontal range	0°—360°
Horizontal maximum speed	$25^{\circ}/s$
Vertical range	$-90^{\circ}$ 90°
Vertical maximum speed	$20^{\circ}/s$
Precision	0.01°

在表 3-表 6 中,除对应研究的因素之外,其余 实验中目标大小(外接矩)为 50 pixel × 50 pixel,目 标运动速度为 50%/s,即每秒运动长度为本身长度 的 50%。运动目标显著性图谱得分为 1 左右,默认 目标大小不变,其余因素保持不变。其中在研究跟 踪成功率与目标显著性关系时,通过调整目标颜色 和背景颜色来实现。当单次跟踪达到 500 帧及以上 时认定为跟踪成功。

# 表 3 跟踪成功率与目标大小关系

Tab. 3 Relationship between tracking success rate and target size

			8			
Pixel	<10	20	50	100	200	400
Success rate	50	95	95	100	95	90

#### 表 4 跟踪成功率与目标速度关系

Tab. 4 Relationship between tracking success

rate and target speed

Speed /(%/s)	10	20	50	100	200
Success rate/%	95	90	85	80	0

#### 表 5 跟踪成功率与目标显著性关系

 Tab. 5
 Relationship between tracking success

rate a	and	target	salient
--------	-----	--------	---------

Salient map score	0.5	1	2
Success rate/ $\%$	55	85	95

#### 表 6 跟踪成功率与目标尺寸变化速度关系

Tab. 6	Relationship between tracking success rate and	
	target size change speed	

	8		81		
Change rate /( %/s)	0	5	10	15	20
Success rate	90	90	85	80	50

表 3 中,当目标外接矩过小时:1) 检测算法高斯

模糊的滤波操作会导致无法检测出运动目标从而检 测失败,无法进入跟踪流程;2)在跟踪过程中,由于 模板中的目标较小,混入大量的背景信息,在背景相 似的情况下,相关响应最大值会偏离目标所在位置, 从而导致跟踪失败; 3) 由于云台运动引起的运动模 糊也会影响跟踪效果。表 4 中,随着目标运动速度 的加快,跟踪成功率逐步降低,原因是从算法计算到 云台对准需要一定的时间。当运动速度过快时,随 着跟踪的进行,会在模板中积累更多背景信息,导致 模板漂移从而导致跟踪失败。表 5 中,随着目标颜 色与背景的差异变大,运动目标显著性图谱得分变 大,跟踪的效果也随之提升。表6中,由于视频目标 的最大和最小尺寸有限,采用目标大小循环变化的 方式进行实验,变化速率由相对于初始尺寸得到,另 外,运动速度也会随目标大小而变化。可以看到在 目标大小变化速度较小时,算法可以保持较好的跟 踪效果,当变化速度进一步增大时,跟踪效果显著下 降。一方面是因为算法为了提升速度削减了 DSST 的尺度池,使得最佳尺度的选择效果有所降低,另一 方面受到系统帧率的影响,模板更新速度不能与变 化速度相匹配导致跟踪效果迅速下降。实验得到, 跟踪成功率与视频目标的运动方式和轨迹无关(上 下、左右、单向、往返、曲线等不产生形变或旋转的方 式)。在每次跟踪结束(无论跟踪成功 500 帧或响应 度过低),重检测模块都能检测到目标并驱动云台对 准目标,继续跟踪的效果同样计入表 3-表 6 中。在 目标有效范围参数下,系统在实验场景下的跟踪平 均成功率达到了85%。

#### 3.4 卫星实验

本文对视频卫星目标进行了系统的验证,跟踪 场景见图 7。图 7(a)与图 7(b)对应于目标在不同位 置时的系统跟踪情况。验证结果如图 8 所示,视频 目标在处于暗室的显示器上,相对于不同对比度的 背景进行移动。可以看到:在跟踪过程中,从图 (1)—图(4),当目标逐渐偏离相机中心位置(模板中 心)时,系统判断目标相对于模板向左偏移,驱动云 台转动对准目标中心。在图(5)处将目标重新置于 视场正中央。在随后的图(6)—图(8)重复以上过 程,达到了在整个跟踪过程中,物体中心相对于像面 中心偏移量始终不大于物体尺寸的 50%(始终在像 面中央 300 pixel×300 pixel 的范围内),可以认为物 体能够一直保持在像面中心。



图 7 视频目标场景 Fig. 7 Video target scene

# 4 结 论

本文提出了一种针对单目标的一体化算法,并 与嵌入式系统相结合,获得了较好的检测跟踪效果。 检测算法利用运动目标的运动特征和静态显著性, 能够有效检测前景中的运动物体,跟踪算法与其他 相关滤波算法相比,在保证了跟踪精度的前提下,提 升了算法的运行速度。针对 DSP 芯片进行了算法优 化,实现了 2048 pixel×2048 pixel 大像面上 24 fps 的运行速度,并在实验室条件下测试达到了 85%的 跟踪成功率。但由于仅采用了 hand-craft 特征,表达 能力不强,因此在测试集上的效果确实不如很多应



图 8 视频目标跟踪局部图 Fig. 8 Partial view of video target tracking

用了 CNN 特征、以及 CNN 跟踪框架的算法。但限 于本系统的实验条件,提出的算法已经比较好地完 成了跟踪任务,并且相比于其他朴素算法,在准确度 和实时性上都有良好的表现。

嵌入式算法的跟踪精度和速度主要受到硬件条件制约,后续工作在条件允许的情况下,检测算法可以加入更多的差分特征用以获取更准确的运动区域,相关滤波跟踪算法可以有更稳定的多维度特征<sup>[15]</sup>来提高跟踪精度。

# 参考文献:

- BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters [C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, USA. Piscataway, NJ:IEEE, 2010;2544-2550.
- [2] HENRIQUES J F, CASEIRO R, MARTINS P, et al, Highspeed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(3):583-596.

- BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: Complementary learners for real-time tracking[C]//
   IEEE Conference on Computer Vision and Pattern Recognition, June 27-July 1, 2016, Las Vegas, USA. Piscataway, NJ:IEEE, 2016;1401-1409.
- [4] DANELLJAN M, HAGER G, KHAN F, et al. Accurate scale estimation for robust visual tracking[C]//British Machine Vision Conference, September 1-5, 2014, Nottingham, United Kingdom. Cambridge: BMVA Press, 2014.
- [5] LI B,YAN J,WU W, et al. High performance visual tracking with siamese region proposal network [C]//IEEE Conference on Computer Vision and Pattern Recognition, June 18-22, 2018, Salt Lake City, Utah, USA. Piscataway,NJ:IEEE,2018;8971-8980.
- [6] ZHU Z, WANG Q, LI B, et al. Distractor-aware siamese networks for visual object tracking [C]//European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. Berlin: Springer, 2018:101-117.
- [7] LI B, WU W, WANG Q, et al. Siamrpn + + : Evolution of siamese visual tracking with very deep networks[C]// IEEE/CVF Conference on Computer Vision and Pattern

Recognition, June 16-20, 2019, Long Beach, CA, USA. Piscataway, NJ: IEEE, 2019: 4282-4291.

- [8] DANELLJAN M, BHAT G, KHAN F S, et al. Atom: Accurate tracking by overlap maximization [C]//IEEE Conference on Computer Vision and Pattern Recognition, June 16-20, 2019, Long Beach, CA, USA. Piscataway, NJ: IEEE 2019: 4660-4669.
- [9] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 27-July 1, 2016, Las Vegas, USA. Piscataway, NJ: IEEE, 2016: 4293-4302.
- [10] LIU H.HE Y.HE B X.et al. Infrared target tracking algorithm based on multiple feature fusion and region of interest prediction [J]. Acta Photonica Sinica, 2019, 48(7): 0710004.

刘辉,何勇,何博侠,等基于多特征融合与 ROI 预测的 红外目标跟踪算法 [J]. 光子学报,2019,48(7): 0710004.

[11] JIANG X, LI P, ZHEN X, et al. Model-free tracking with deep appearance and motion features integratio [C]// 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), January 7-11, 2019, Waikoloa Village, HI, USA. Piscataway, NJ: IEEE, 2019:101-110.

- [12] AN Y F, PAN H P. A new method for moving bubbles detection based on frame difference[J]. Journal of Optoelectronics Laser, 2016, 27(9): 963-972.
  安云凤,潘海鹏. 一种新的三帧差运动气泡检测方法.
  [J].光电子 激光, 2016, 27(9): 963-972.
- [13] HE Z,LI Q,CHANG M,et al. Adaptive model update strategy for correlation filter trackers[J]. IEEE Access,2019, 7:151493-151505.
- [14] HUO Y L,LI M, ZHENG H L, et al. Context-aware target tracking algorithm fused with redetection mechanism[J]. Journal of Optoelectronics • Laser, 2021, 32(9): 992-999. 火元莲,李明,郑海亮,等.融合重检测机制的上下文

感知目标跟踪算法[J].光电子·激光,2021,32(9): 992-999.

[15] ZHANG H Y, WANG H S, HE P Y. Correlation filter tracking based on superpixel and multifeature fusion[J]. Optoelectronics Letters, 2021, 17(1):47-52.

#### 作者简介:

**李** 奇 (1973-),男,博士,教授,博士生导师,主要从事光学工程、 成像技术、图像获取与处理等领域的研究工作.