

光学学报

基于光谱匹配降维和特征融合的高光谱目标跟踪

郭业才^{1,2}, 曹佳露^{1,2}, 韩莹莹³, 张恬梦⁴, 赵东^{1,2*}, 陶旭^{1,2}

¹南京信息工程大学电子与信息工程学院, 江苏 南京 210044;

²无锡学院电子信息工程学院, 江苏 无锡 214105;

³中国船舶集团有限公司第七〇三研究所, 黑龙江 哈尔滨 150000;

⁴哈尔滨工程大学航天与建筑工程学院, 黑龙江 哈尔滨 150000

摘要 针对现有高光谱视频目标跟踪算法在目标尺度发生变化时容易出现跟踪精度下降的问题, 提出一种基于光谱匹配降维和特征融合的高光谱目标跟踪算法。首先, 利用目标局部光谱和阈值来估计目标光谱, 并利用目标光谱与高光谱图像进行朴素相关, 实现高光谱图像降维, 从而提取目标的深度特征。然后, 利用局部方差判断目标区域, 提取目标的 3D 方向梯度直方图(HOG)特征。为保留高光谱图像的光谱信息以及深度特征的语义信息, 利用通道卷积融合的方法, 得到更具辨别力的融合特征。最后, 将融合特征送入相关滤波器, 通过尺度池思想提高算法在目标尺度变化挑战下的跟踪鲁棒性。实验结果表明, 所提跟踪算法在目标尺度变化挑战下具有更好的性能。

关键词 测量; 目标跟踪; 高光谱视频; 光谱匹配降维; 通道卷积融合

中图分类号 TP751.1; TP79

文献标志码 A

DOI: 10.3788/AOS230776

1 引言

在智能交通监控^[1]、汽车自动驾驶^[2]、人工智能^[3]等领域中, 目标跟踪^[4]一直是计算机视觉研究的热门问题。近年来, 可见光视频目标跟踪算法取得了显著进展, 从相关滤波^[5]向基于深度学习^[6]的技术方向发展。然而, 可见光视频在描述物体的物理特性方面有一定的局限性, 如当目标遭遇背景杂乱、遮挡、形变等干扰时, 基于可见光视频的跟踪算法很难保持跟踪鲁棒性。

高光谱图像^[7-9]包含物体的光谱信息, 可增强物体之间的识别能力^[10-12]。因此, 高光谱视频在识别材料方面比传统彩色视频具有更多可利用的光谱信息。然而, 高光谱视频维度高、数据量大, 给特征提取带来了较大的困难和较高的计算成本, 因此可见光视频目标跟踪技术很难直接应用于高光谱视频。近年来, 快照高光谱技术的发展使得获取高光谱视频成为可能, 许多研究人员将重心转向高光谱视频目标跟踪^[13-14]技术: Qian 等^[15]提出一种基于卷积网络的高光谱跟踪(CNHT)方法, 该方法在每个波段提取一组小立方体作为卷积核, 但忽略了高光谱图像中波段之间的相关性; Xiong 等^[16]提出一种基于材料的高光谱视频跟踪

器(MHT), 该跟踪器对多维定向梯度直方图和丰度直方图分别提取局部光谱空间信息; Uzkent 等^[17]提出一种深度核化相关滤波(DeepHKCF)方法, 将高光谱图像转换为伪彩色图像, 使用 VGGNet^[18] (visual geometry group net) 学习判别卷积神经网络(CNN)特征^[19], 但是转换后的伪彩色图像丢失了光谱信息, 导致跟踪精度降低。在目标跟踪过程中, 目标经常会发生尺度变化, 导致跟踪失败。上述算法均没有针对目标尺度变化挑战而提出相应的解决策略。

Danelljan 等^[20]提出了 DSST (discriminative scale space tracking) 算法, 采用分离位置估计和尺度估计的思想来应对目标尺度变化。但是, DSST 使用第一步的定位为基准, 后续容易造成误差累加, 影响了尺度的准确性。Li 等^[21]提出的 SAMF (scale adaptive kernel correlation filter tracker) 算法是利用尺度池的思想解决目标尺度变化。然而, 上述算法是针对可见光视频目标跟踪提出的解决方法, 不能直接应用于高光谱视频目标跟踪。为了更好地解决高光谱视频目标跟踪中目标尺度变化的问题, 本文提出一种基于光谱匹配降维和特征融合的高光谱视频目标跟踪(DF-HVT)算法。首先, 提出一种光谱匹配的降维方法, 该方法利用高光谱图像的光谱特性分割目标和背景, 利用目标光

收稿日期: 2023-04-04; 修回日期: 2023-05-01; 录用日期: 2023-05-19; 网络首发日期: 2023-06-28

基金项目: 国家自然科学基金(62001443, 62105258)、江苏省自然科学基金(BK20210064)、山东省自然科学基金(ZR2020QE294)、江苏省高等学校基础科学(自然科学)研究面上项目(22KJB140015)、无锡市创新创业资金“太湖之光”科技攻关计划(基础研究)项目(K20221046, K20221043)、无锡学院人才启动基金(2021r007, 2021r008)

通信作者: *dzhao@cwxu.edu.cn

谱曲线与图像进行朴素相关实现降维;其次,针对目标跟踪过程中的尺度变化问题,提出一种利用局部方差来增强目标 3D 方向梯度直方图(HOG)特征的方法,该方法使目标特征更加明显,提高了跟踪的鲁棒性;最后,提出通道卷积融合方法,将深度特征和增强的 3D HOG 特征逐通道卷积,充分利用特征之间的相关性。

2 研究方法

图 1 所示为 DF-HVT 算法的总体流程。首先,对第一帧高光谱图像手动选择的搜索区域进行特征提

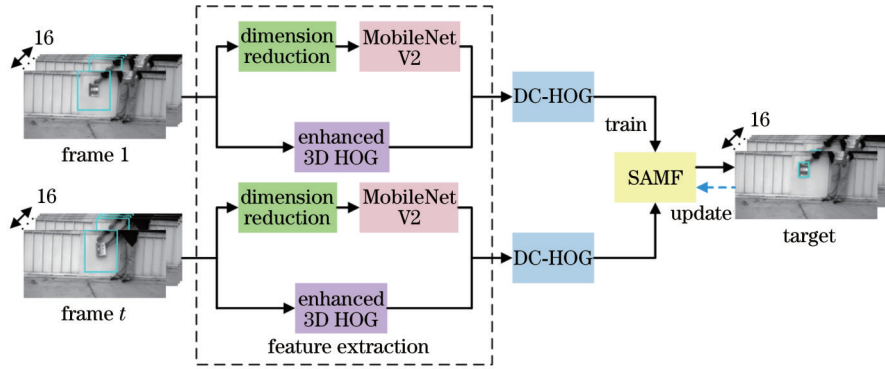


图 1 DF-HVT 的总体流程

Fig. 1 Overall flow chart of DF-HVT

2.1 降维

MobileNet V2^[22]的输入要求一般是三波段的可见光图或单波段的灰度图,但是实验使用的高光谱视频序列是由 16 个波段的高光谱图像组成的。因此,为了达到网络的输入要求,必须对高光谱视频进行降维。

根据不同物体的光谱信息差异,提出一种基于光谱匹配的降维方法。目标局部光谱曲线只能代表目标局部特征,不能很好地反映目标整体特征。因此利用每个像素的光谱曲线和目标局部光谱曲线作差,统计误差值来分割目标像素和背景像素。最终得到具有代表性的目标光谱曲线,利用目标光谱曲线与图像朴素相关实现高光谱图像降维。

用 H_s 表示高光谱图像搜索区域, $H_s \in \mathbf{R}^{M_s \times N_s \times L}$ 包含 $M_s \times N_s$ 个像素点和 L 个波段;用 P 表示目标局部区域, $P \in \mathbf{R}^{3 \times 3 \times L}$ 包含 3×3 个像素点和 L 个波段。首先确定目标局部区域 P 在第 i 个波段的光谱平均值 C_{pi} , 即

$$C_{pi} = \frac{\sum_{u=1}^9 P_{ui}}{9}, i \in [1, L], u \in [1, 9], \quad (1)$$

式中: P_{ui} 为第 i 个波段目标局部区域第 u 个像素点的灰度值。实验中使用 16 波段的高光谱视频序列, 所以 L 取 16。

同一类别的像素在每个波段的灰度值相似, 灰度值差异越大, 表示两个像素光谱差异越大。将搜索区域内像素的灰度值 C_{sij} 和目标局部光谱平均值 C_{pi} 作

差。特征提取分为两个模块, 一个模块通过 MobileNet V2 网络^[22]提取降维图像的深度特征, 另一个模块直接对高光谱图像提取增强目标的 3D HOG 特征。其次, 将两个模块提取的特征进行通道卷积融合, 得到分辨率更高的 DC-HOG (deep convolution HOG) 特征。然后, 将 DC-HOG 特征输入相关滤波器后获得模板, 对后续帧的搜索区域进行特征提取并将提取结果输入相关滤波器。最后, 通过响应图中最大的响应值来确定目标位置, 并对模型的参数进行实时更新。

差, 获得第 i 个波段上第 j 个像素的误差值 Q_{ij} :

$$Q_{ij} = |C_{sij} - C_{pi}|, j \in [1, M_s \times N_s]. \quad (2)$$

将每个像素在每个波段上的误差值 Q_{ij} 进行累加, 得到累加值 A_j :

$$A_j = \sum_{i=1}^L Q_{ij}, j \in [1, M_s \times N_s]. \quad (3)$$

像素的累加值 A_j 越高, 说明该像素和目标像素的差异越大; 累加值 A_j 越低, 说明该像素与目标像素越相似。设定阈值 ρ , 将累加值 A_j 超过阈值 ρ 的像素定义为背景像素, 没有超过阈值 ρ 的像素定义为目标像素。

$$\begin{cases} C_j \in H_o, A_j < \rho \\ C_j \in H_b, A_j \geq \rho \end{cases}, j \in [1, M_s \times N_s], \quad (4)$$

式中: C_j 表示搜索区域的像素; H_o 表示目标像素的集合; H_b 表示背景像素的集合。本实验中 ρ 取 0.2。 H_o 由目标像素组成, 能更好地代表目标的光谱曲线, 因此将 H_o 内像素的灰度值 C_{Oij} 求平均, 得到目标光谱曲线 C_o 。

$$C_o = \left\{ C_{oi} \mid C_{oi} = \frac{\sum_{j=1}^{N_o} C_{Oij}}{N_o} \right\}, i \in [1, L], j \in [1, N_o], \quad (5)$$

式中: C_{oi} 表示 H_o 在第 i 个波段上的像素平均值; N_o 表示 H_o 内像素数。

为了将高光谱图像降维成一波段的图像, 通过目标光谱曲线和第 $t+1$ 帧的搜索区域 $H_{s,t+1}$ 进行朴素相

关操作,得到降维后的图像

$$H_d = H_{s,t+1} \otimes C_o, \quad (6)$$

式中: H_d 为降维后的结果; \otimes 表示朴素相关,即计算每个像素的光谱曲线与目标光谱曲线的相似度。图

2所示为降维结果,其中图2(a)所示为原始高光谱图像,图2(b)所示为局部光谱曲线和平均光谱曲线,图2(c)所示为降维后的图像,可以看到,背景得到了抑制。

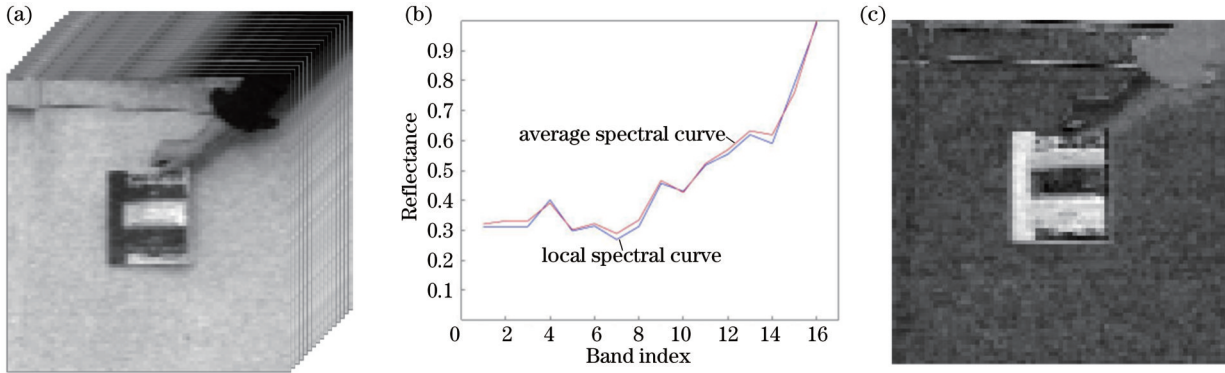


图2 降维结果图。(a)原始高光谱图像;(b)局部光谱曲线和平均光谱曲线;(c)降维后的图像

Fig. 2 Dimensionality reduction results. (a) Original hyperspectral image; (b) local spectral curve and average spectral curve; (c) image after dimension reduction

2.2 深度特征提取

MobileNet系列^[22-23]的轻量网络被广泛应用于图像分类。MobileNet V1^[20]利用可分离卷积结构建立了一个轻量级模型。MobileNet V2^[22]在MobileNet V1的基础上增加倒残差结构和瓶颈结构,从而降低对内存的消耗。完整的MobileNet V2包括一个常规的 1×1 卷积、17个瓶颈结构、一个全局平均池化层和一个Softmax分类层。

本实验中,使用预训练好的MobileNet V2网络提

取深度特征。浅层的特征包含空间信息,便于定位目标^[24]。深层的特征包含语义信息,有利于区分目标和背景。为确保深度特征能够保留一定的空间信息,选取的输出特征在空间上不能过小,本实验选取bottleneck_13层来输出特征。将降维后的图像 H_d 输入MobileNet V2后,得到大小为 $14 \times 14 \times 96$ 的深度特征 D ,其中 14×14 表示空间分辨率,96表示特征通道数。图3所示为MobileNet V2提取的前32通道的深度特征。从图3可以看出,深度特征包含着丰富的语义信息。

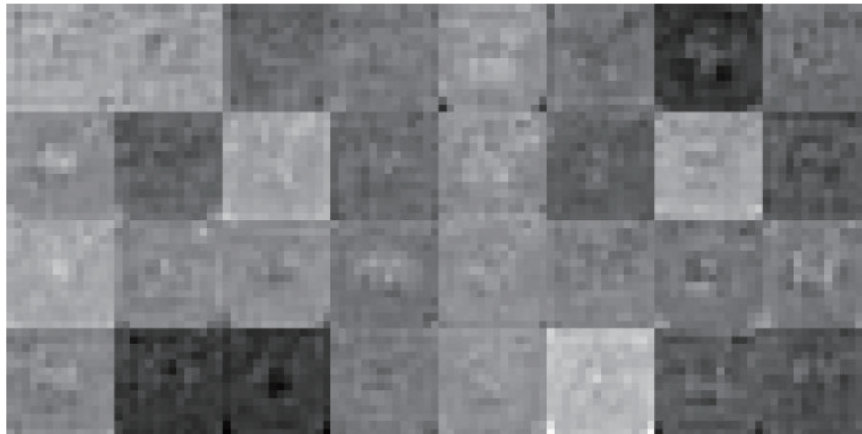


图3 bottleneck_13层前32通道的深度特征

Fig. 3 Depth features of the first 32 channels of the bottleneck_13 layer

2.3 目标增强

针对高光谱图像的三维数据,采用传统特征提取手段会浪费高光谱图像的光谱信息。Xiong等^[16]通过建立三维局部光谱空间直方图来描述物体特征。将搜索区域 H_s 中的每个像素用 (χ, θ, ϕ) 表示,其中 χ 表示光谱-空间梯度幅值, θ 表示空间梯度方向, ϕ 表示光谱梯度方向,定义如下

$$\begin{cases} \chi(x, y, l) = \sqrt{\nabla T_x^2 + \nabla T_y^2 + \nabla T_l^2} \\ \theta(x, y, l) = \arctan(\nabla T_y / \nabla T_x) \\ \phi(x, y, l) = \arctan(\nabla T_l / \sqrt{\nabla T_x^2 + \nabla T_y^2}) \end{cases}, \quad (7)$$

式中: ∇T_x 和 ∇T_y 表示二维空间内的水平梯度和垂直梯度; ∇T_l 表示光谱梯度。将高光谱图像分成若干大小为 $z \times z \times z$ 的细胞单元,统计每个细胞单元在空

间方向和光谱方向上的梯度直方图,以捕获细胞单元内的空间和光谱信息。将相邻的细胞单元组合成若干个小立方体,对每个小立方体内的梯度直方图进行拼接,形成特征向量。

方差既能反映各像素与其相邻像素的相关性,又能反映各像素灰度值的变化程度^[25]。在高光谱图像中,不同材料的物体具有不同的光谱,灰度值的一致性

不高,因此可以获得更大的局部方差。为此,提出一种利用局部方差来增强目标的方法。在所提局部目标增强方法中:对于高方差区域内的点,将其特征向量的权重增大;对于低方差区域内的点,将其特征向量的权重减小。图 4 为局部目标增强示意图,可以看出,目标部分的特征向量占比变大,而背景部分的特征向量占比变小。

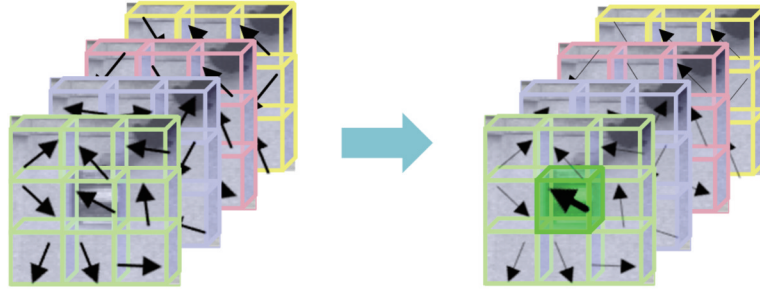


图 4 局部目标增强示意图

Fig. 4 Schematic of local target enhancement

将搜索区域划分成 W 个同等大小的小立方体,大小为 $m \times n \times l$,对于 H_s 中第 k 个 $m \times n \times l$ 大小的小立方体,其方差 V_k 的计算公式为

$$V_k = \sum_{i=1}^l \sum_{j=1}^{m \times n} \frac{(C_{cij} - M_c)^2}{m \times n \times l}, \quad (8)$$

式中: $1 \leq k \leq W$; C_{cij} 表示小立方体内像素灰度值; M_c 表示小立方体内像素平均灰度值。

小立方体内像素灰度值的方差越大,表示该部分的信息量越多,有较大的可能是目标,因此赋予小立方体的权重就越大,则第 k 个小立方体的权重为

$$\zeta_k = \frac{V_k}{\sum_{k=1}^W V_k}. \quad (9)$$

提取小立方体内特征向量 G_k ,对特征向量 G_k 进行加权得到增强目标的 3D HOG 特征 U :

$$U = \{U_k | U_k = G_k \times \zeta_k\}. \quad (10)$$

图 5 所示为前 32 通道增强目标后的 3D HOG 特征,可以明显看出目标。本实验中, z 取 4,空间方向的 bin 为 18,光谱方向的 bin 为 6,将细胞单元内像素点的特征向量投给不同直方图区间;将相邻的 2×2 细胞单元组合在一起,形成 $8 \times 8 \times 4$ 的块,然后连接相应细胞单元的特征得到块级特征,将块级特征按照光谱方向连接得到增强目标后的 3D HOG 特征 U , $U \in \mathbb{R}^{\frac{M_s}{z} \times \frac{N_s}{z} \times 96}$,其中 $\frac{M_s}{z} \times \frac{N_s}{z}$ 为空间大小,96 为通道数。

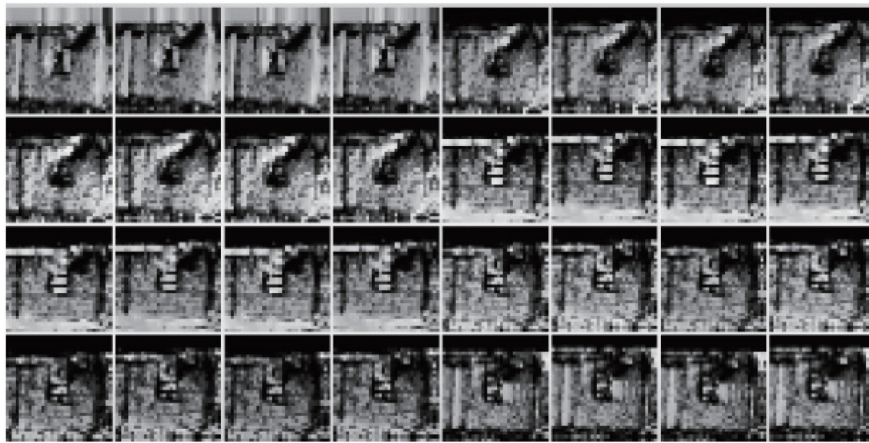


图 5 前 32 通道增强目标后的 3D HOG 特征

Fig. 5 3D HOG features of the first 32 channels after enhancing the target

2.4 通道卷积融合

特征融合是将多种特征合并为单一的特征集。深度特征 E 包含更多的语义信息,有利于目标定位,而

3D HOG 特征包含空间光谱信息,有利于区分目标和背景。将特征进行简单的平均加权融合并不能充分利用特征包含的信息。为融合语义信息和空间光谱信

息,充分利用特征之间的相关性,提出一种逐通道卷积融合的方法,将增强目标的 3D HOG 特征和深度特征按通道数划分为 96 个二维矩阵,每个通道对应卷积得到 DC-HOG 特征 Z

$$Z = \{Z_n | Z_n = D_n * U_n\}, n \in [1, 96], \quad (11)$$

式中: n 表示通道数; $*$ 表示卷积。经过融合后得到 DC-HOG 特征,图 6 所示为前 32 通道的 DC-HOG 特征。

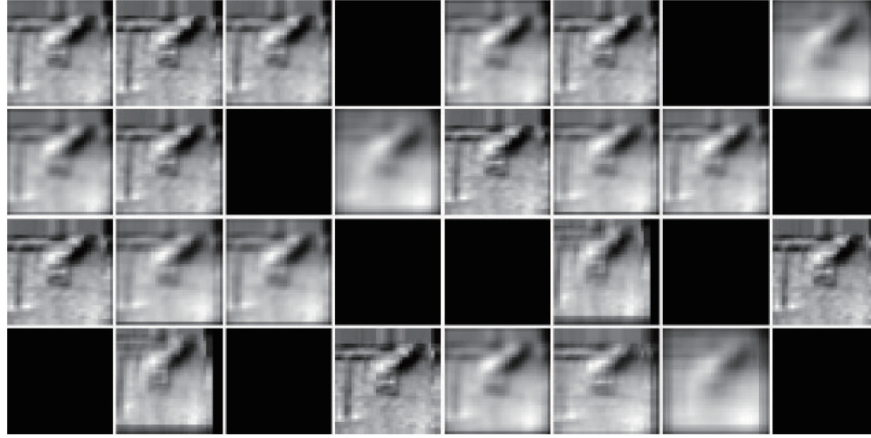


图 6 前 32 通道的 DC-HOG 特征

Fig. 6 DC-HOG features of the first 32 channels

2.5 SAMF 算法

SAMF 算法^[21]利用不同尺度的目标框对目标进行采样,与每帧的模板进行比较,通过搜索最大响应值来获得合适的尺度。SAMF 主要分为训练滤波器、尺度估计和更新模型 3 个部分。和 KCF^[5]类似,SAMF 利用循环矩阵的分解来求解岭回归问题,即

$$\min_{\omega} \sum_m [f(x_m) - y_m]^2 + \lambda \|\omega\|^2, \quad (12)$$

式中: ω 表示权重系数; x_m 表示第 m 个训练样本; y_m 表示对应的标签函数; λ 表示正则化参数。为防止过拟合,实验中 λ 设置为 0.0001。通过核函数,得到对偶空间系数 α

$$\hat{\alpha}^* = \frac{\hat{y}}{\hat{k}^{xx} + \lambda}, \quad (13)$$

式中: $\hat{y}(\eta = \alpha, y, k)$ 表示向量傅里叶变换; $*$ 表示复共轭; \hat{k}^{xx} 表示 x 的核相关。

在跟踪时, SAMF 把上一帧的目标位置当作中心,基于预定义的尺度池,获得 7 个不同的尺度样本 $\{0.985, 0.990, 0.995, 1.000, 1.005, 1.010, 1.015\}$, 找到最大响应值 F_{\max}

$$F_{\max} = \arg \max [\mathcal{F}^{-1}(\hat{k}^{xx} \odot \hat{\alpha})], \quad (14)$$

式中: \mathcal{F}^{-1} 表示傅里叶逆变换; \hat{k}^{xx} 为由训练样本和测试样本之间高斯核互相关矩阵导出的元素; \odot 表示元素点乘。在更新过程中,通过将新滤波器与旧滤波器的线性组合来更新对偶空间系数 α 和基础数据模板 \tilde{x} , 即

$$M_{\text{new}} = \gamma M + (1 - \gamma) M_{\text{pre}}, \quad (15)$$

式中: $M = [\alpha^T, \alpha \tilde{x}^T]^T$; M_{pre} 是前一帧滤波器的参数; M 表示当前帧滤波器的参数; γ 表示学习率,学习率越

大,表示滤波器更新速度越快,本实验中将 γ 设置为 0.01。

3 实验结果分析

3.1 实验配置

为了验证 DF-HVT 的有效性,以 MATLAB R2021b 软件为实验平台,使用配置为 Intel(R) Core (TM) i7-12700H CPU@2.30 GHz, 16 GB RAM, RTX3060 GPU 的计算机进行测试,并使用 Matconvnet 工具包从 MobileNet V2 网络中提取深度特征。该算法处理视频序列的速度为 2.9 frame/s。

本实验选取的高光谱视频序列来自文献[16]的数据集,该数据集可在网站(www.hsitrack.com)中下载。每个视频序列都由对应的高光谱视频和可见光视频组成。所有视频序列都由波长为 470~620 nm 的 16 波段高光谱相机采集。为了测试该算法对抗尺度变化的能力,有 3 个实验序列带有尺度变化挑战。除此之外,所选的实验序列还涵盖其他挑战,如形变、快速移动、面内旋转等。

3.2 定性分析

为验证所提算法的有效性,实验选取 3 种高光谱目标跟踪算法进行对比,分别是 CNHT^[15]、DeepHKCF^[17]、MHT^[16]。CNHT 算法使用简单的卷积来提取特征,并且从初始帧的目标区域中选取归一化的三维小立方体和随机固定卷积核来提取后续帧的特征。DeepHKCF 算法使用伪彩色视频来提取特征。MHT 算法利用解混来提取高光谱图像的材料信息,区分目标和背景。为了可视化所提算法的性能,对实验序列进行了结果展示。图 7 所示为 4 种算法在所选实验序列上的定性结果。

从图 7 可以看到,图像中有 5 种颜色的矩形框,不同的颜色代表不同的算法。其中,红色代表所提算法,粉色代表 CNHT,绿色代表 MHT,蓝色代表

DeepHKCF,白色代表地面真实值。不同颜色的矩形框与白色矩形框重合越多,表示该颜色所代表算法的跟踪效果越好。

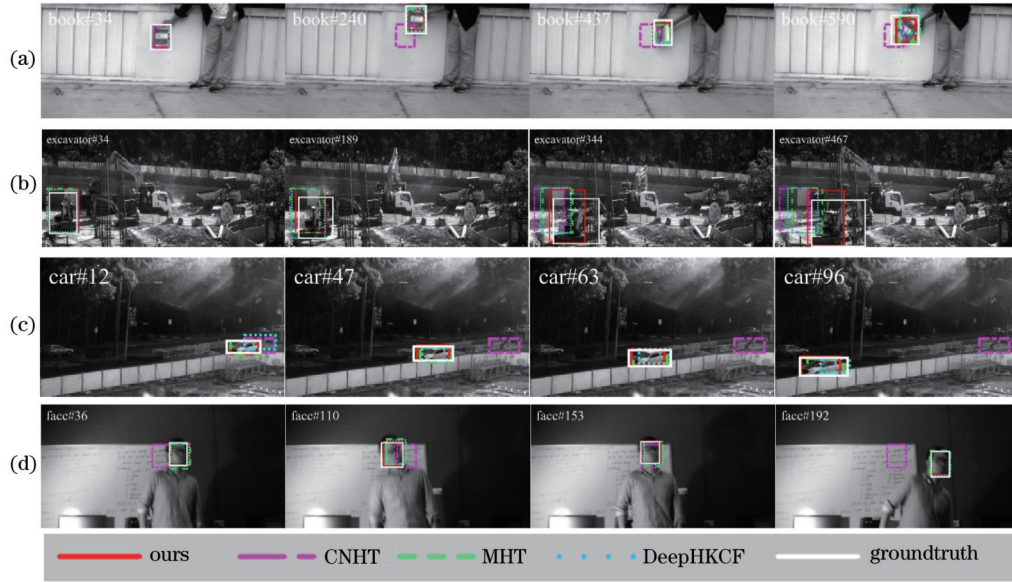


图 7 在所选序列上的定性分析结果图。(a)书序列;(b)挖掘机序列;(c)小车序列;(d)脸序列

Fig. 7 Qualitative analysis results on the selected sequences. (a) Book sequence; (b) excavator sequence; (c) car sequence; (d) face sequence

如图 7(a)所示,CNHT算法在训练卷积滤波器时只考虑固定的正样本,导致它在第 240 帧就丢失了目标。在书本被翻动或者转换方向时,目标发生形变和尺度变化,所提 DF-HVT 算法由于利用了尺度池的思想,能够估计目标的尺度,实现稳定跟踪。

如图 7(b)所示,挖掘机在杂乱的工地上工作,DeepHKCF算法使用VGGNet提取的单一深度特征,导致其在背景杂乱的情况下很难定位目标,结果显示DeepHKCF算法在第 344 帧完全丢失了目标。所提 DF-HVT 算法利用光谱匹配降维和增强目标的 3D HOG 特征提高了目标与背景的辨别性,在跟踪过程中的鲁棒性更强。

如图 7(c)所示,小车快速驶来,CNHT算法从一开始就跟踪错了目标。CNHT算法的随机固定卷积滤波器产生的特征没有足够强的辨别力,使得目标位置预测失败。DeepHKCF算法使用预先训练的模型获得比CNHT算法更好的性能,但是DeepHKCF和MHT算法没有尺度估计策略,使得目标框在追踪过程中产生漂移。所提 DF-HVT 算法采用增强目标的 HOG 特征和 SAMF 滤波器,对尺度变化挑战表现出更好的适应性。

如图 7(d)所示,人在白色的背景前跳跃,虽然光照对目标产生了影响,但是大部分跟踪器依旧保持良好的跟踪性能。在 110 帧目标发生形变时,其他 3 个跟踪器的目标框都产生漂移,只有 DF-HVT 算法的跟踪目标框依旧保持良好的鲁棒性。

3.3 定量分析

从准确率(accuracy)和成功率(success rate)两个方面定量评估算法的性能。准确率是指预测值的中心位置和地面真实值的中心位置之间的误差小于给定阈值的百分比。成功率表示预测输出框和地面真实值边界框的交集超过给定阈值的百分比。表 1 和表 2 分别展示了 4 种算法的准确率和成功率,其中粗体表示每个类别的最高得分,SV 表示尺度变化,OPR 表示平面外旋转。图 8 所示为 4 种算法在所选测试序列上的准确率和成功率曲线,图 9 和图 10 所示分别为与尺度变化挑战和平面外旋转挑战有关的准确率和成功率曲线,图例中括号内的数字表示该算法相应指标的平均值。

表 1 4 种算法在不同挑战下的准确率
Table 1 Precision of four algorithms under different challenges

Algorithm	Precision	Precision with SV	Precision with OPR
Ours	0.820	0.905	0.916
MHT	0.787	0.707	0.927
DeepHKCF	0.786	0.693	0.779
CNHT	0.647	0.627	0.418

准确率和成功率结果表明,在测试序列上,所提算法表现出良好的性能。如图 8 所示,所提算法在总的测试序列上准确率和成功率都排名第一。具体而言,与 MHT 相比,所提算法的准确率提高了 3.3%,成功

表 2 4 种算法在不同挑战下的成功率

Table 2 Success rate of four algorithms under different challenges

Algorithm	Success rate	Success rate with SV	Successrate with OPR
Ours	0.641	0.671	0.763
MHT	0.619	0.531	0.736
DeepHKCF	0.617	0.483	0.536
CNHT	0.356	0.440	0.386

率提高了 2.2%。由于所提算法使用通道卷积融合,其特征辨别性增强,跟踪鲁棒性提升。如图 9 所示,在

尺度变化挑战下,所提算法的准确率比第二名提高了 19.8%,成功率提高了 14.0%,表现出优异的适应性。其余 3 种算法由于没有对应的尺度估计模块,在尺度变化挑战下的表现略逊一筹。如图 10 所示,由于没有相应的坐标仿射变化策略,在平面外旋转挑战下所提算法的准确率比 MHT 低 1.31%,但在成功率上排名第一。

3.4 消融实验

为检验所提算法的有效性,进行相关消融实验。具体来说,采用另外 3 种算法与所提 DF-HVT 算法进行对比,这 3 种算法分别是 3D HOG、average 和 PCA。

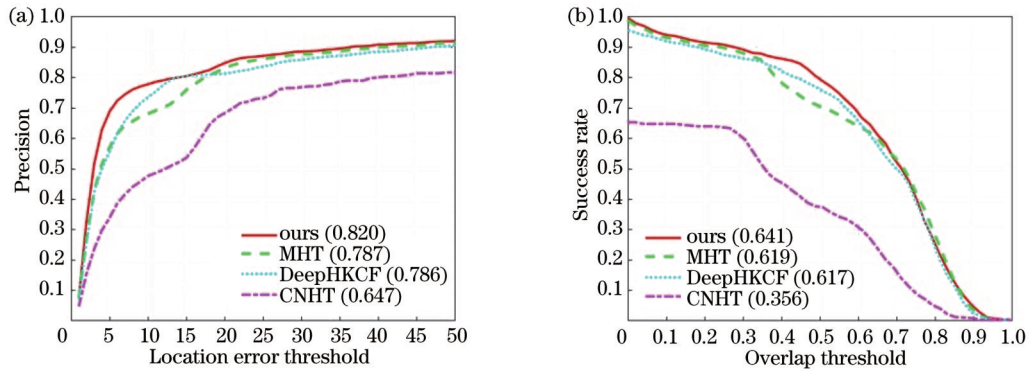


图 8 4 种算法在测试序列上的跟踪准确率和准确率。(a)准确率;(b)成功率

Fig. 8 Tracking precision and success rate of four algorithms on the test sequences. (a) Precision; (b) success rate

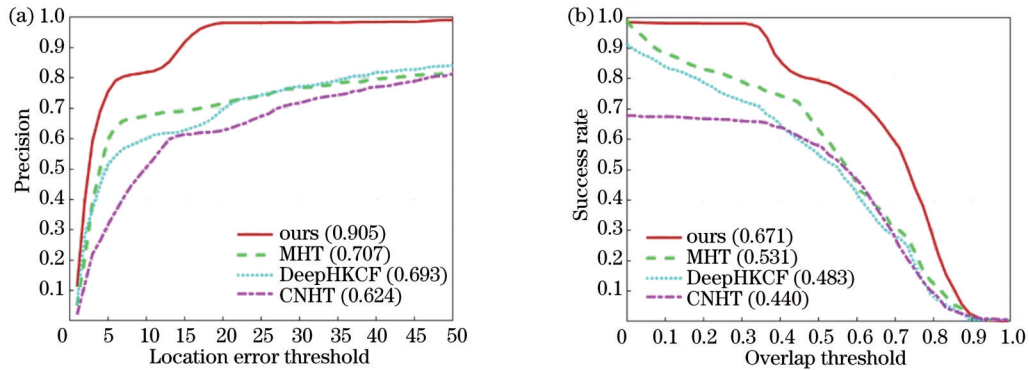


图 9 4 种算法在尺度变化挑战上的跟踪准确率和成功率。(a)准确率;(b)成功率

Fig. 9 Tracking precision and success rate of four algorithms on the scale variation challenge. (a) Precision; (b) success rate

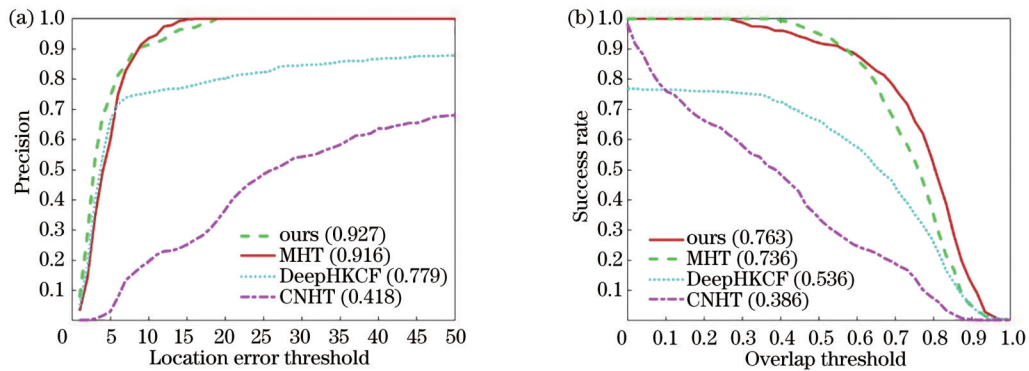


图 10 4 种算法在平面外挑战上的跟踪准确率和成功率。(a)准确率;(b)成功率

Fig. 10 Tracking precision and success rate of four algorithms on out-of-plane challenge. (a) Precision; (b) success rate

这些对比算法的框架与 DF-HVT 算法相同,但缺少特定的改进模块:3D HOG 算法没有进行特征增强;average 算法没有进行通道卷积融合,而是采用普通的平均融合;PCA 算法采用主成分分析进行降维,并没有进行光谱匹配降维。图 11 所示为 4 种算法在实验序列上的准确率曲线和成功率曲线,图例中括号内的数字表示该算法相应指标的平均值。表 3 所示的数据充分说明了 DF-HVT 算法的有效性,提高了跟踪效果。

从图 11 可以看出,DF-HVT 算法利用多特征融合取得更好的性能。具体来说,相比第二名的 3D HOG 算法,DF-HVT 算法的准确率提高了 4.4%,成功率提高了 5.2%。PCA 算法由于没有采用光谱匹配降维,

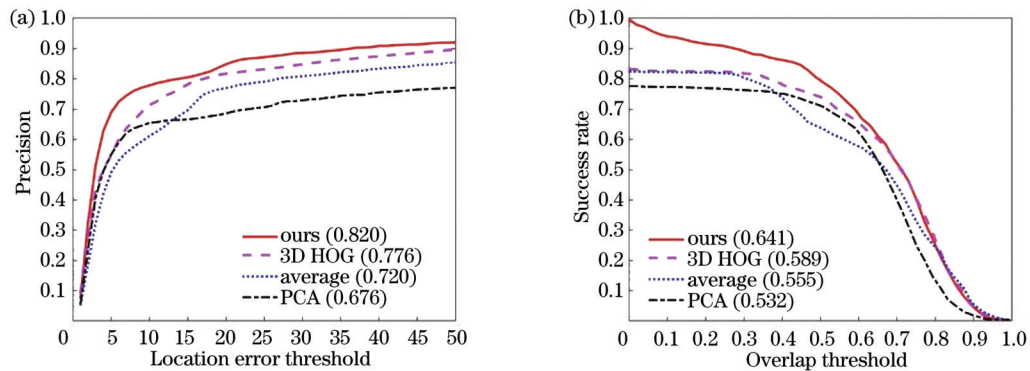


图 11 所有测试序列的消融实验结果。(a)准确率;(b)成功率

Fig. 11 Results of ablation experiments for all test sequences. (a) Precision; (b) success rate

4 结 论

针对高光谱目标跟踪任务中的尺度变化挑战,提出一种基于光谱匹配降维和特征融合的高光谱目标跟踪算法。光谱匹配降维通过像素光谱的差异性分割目标与背景,为网络提供低维特征输入,降低计算复杂性。DC-HOG 特征提高了目标的可辨别性。实验结果表明,所提算法对目标跟踪的鲁棒性优于其他算法,能很好地应对尺度变化挑战。未来的工作将探索性能更好的高光谱视频目标跟踪算法来应对背景杂波、平面内旋转等挑战,进一步提高算法的普适性。

参 考 文 献

- [1] Yuan Y, Lu Y W, Wang Q. Tracking as a whole: multi-target tracking by modeling group behavior with sequential detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(12): 3339-3349.
- [2] Zhang C L, Siranosian A, Krstić M. Extremum seeking for moderately unstable systems and for autonomous vehicle target tracking without position measurements[J]. Automatica, 2007, 43(10): 1832-1839.
- [3] Ye J, Yang X P, Kang S, et al. A robust MTMC tracking system for AI-city challenge 2021[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 19-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 4039-4048.
- [4] Pan Z, Liu S, Fu W N. A review of visual moving target tracking[J]. Multimedia Tools and Applications, 2017, 76(16):

表 3 消融实验的准确率和成功率

Methods	Ours	3D HOG	average	PCA
Precision	0.820	0.776	0.720	0.676
Success rate	0.641	0.589	0.555	0.532

准确率和成功率均排在最后一名,可见光谱匹配降维模块对实验结果的影响最大。与 average 算法相比,DF-HVT 算法的准确率提高了 10%,成功率提高了 8.6%,表明使用通道卷积来融合特征可以提高特征的可分辨性。实验结果表明,所提 DF-HVT 算法可增强对目标跟踪的鲁棒性。

- 16989-17018.
- [5] 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.
- [6] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking[M]//Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 850-865.
- [7] 李非燕, 霍宏涛, 李静, 等. 基于多特征和改进稀疏表示的高光谱图像分类[J]. 光学学报, 2019, 39(5): 0528004. Li F Y, Huo H T, Li J, et al. Hyperspectral image classification via multiple-feature-based improved sparse representation[J]. Acta Optica Sinica, 2019, 39(5): 0528004.
- [8] 颜铭靖, 苏喜友. 基于三维空洞卷积残差神经网络的高光谱影像分类方法[J]. 光学学报, 2020, 40(16): 1628002. Yan M J, Su X Y. Hyperspectral image classification based on three-dimensional dilated convolutional residual neural network [J]. Acta Optica Sinica, 2020, 40(16): 1628002.
- [9] 国强, 彭龙. 基于三维卷积神经网络与超像素分割的高光谱分类[J]. 光学学报, 2021, 41(22): 2210001. Guo Q, Peng L. Hyperspectral classification based on 3D convolutional neural network and super pixel segmentation[J]. Acta Optica Sinica, 2021, 41(22): 2210001.
- [10] 白鹤轩, 杨峰, 李丹阳, 等. 基于表面增强拉曼光谱的多组分物质分类识别[J]. 光学学报, 2021, 41(20): 2024001. Bai H X, Yang F, Li D Y, et al. Multi-component substance classification and recognition based on surface-enhanced Raman spectroscopy[J]. Acta Optica Sinica, 2021, 41(20): 2024001.
- [11] 王健, 崔天翔, 王一, 等. 高分五号可见短波红外高光谱影像云检测研究[J]. 光学学报, 2021, 41(9): 0928003. Wang J, Cui T X, Wang Y, et al. Cloud detection for GF-5

- visible-shortwave infrared advanced hyperspectral image[J]. *Acta Optica Sinica*, 2021, 41(9): 0928003.
- [12] 饶利波, 庞涛, 纪然仕, 等. 基于高光谱成像技术结合堆栈自动编码器-极限学习机方法的苹果硬度检测[J]. *激光与光电子学进展*, 2019, 56(11): 113001.
- Rao L B, Pang T, Ji R S, et al. Firmness detection for apples based on hyperspectral imaging technology combined with stack autoencoder-extreme learning machine method[J]. *Laser & Optoelectronics Progress*, 2019, 56(11): 113001.
- [13] Zhao D, Cao J L, Zhu X G, et al. Hyperspectral video target tracking based on deep edge convolution feature and improved context filter[J]. *Remote Sensing*, 2022, 14(24): 6219.
- [14] Zhao D, Zhu X G, Zhang Z, et al. Hyperspectral video target tracking based on pixel-wise spectral matching reduction and deep spectral cascading texture features[J]. *Signal Processing*, 2023, 209: 109033.
- [15] Qian K, Zhou J, Xiong F C, et al. Object tracking in hyperspectral videos with convolutional features and kernelized correlation filter[M]//Basu A, Berretti S. *Smart multimedia. Lecture notes in computer science*. Cham: Springer, 2018, 11010: 308-319.
- [16] Xiong F C, Zhou J, Qian Y T. Material based object tracking in hyperspectral videos[J]. *IEEE Transactions on Image Processing*, 2020, 29: 3719-3733.
- [17] Uzkenet B, Rangnekar A, Hoffman M J. Tracking in aerial hyperspectral videos using deep kernelized correlation filters[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(1): 449-461.
- [18] Conneau A, Schwenk H, Barrault L, et al. Very deep convolutional networks for text classification[EB/OL]. (2016-06-06)[2023-02-05]. <https://arxiv.org/abs/1606.01781>.
- [19] Gu J X, Wang Z H, Kuen J, et al. Recent advances in convolutional neural networks[J]. *Pattern Recognition*, 2018, 77: 354-377.
- [20] Danelljan M, Häger G, Khan F S, et al. Discriminative scale space tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(8): 1561-1575.
- [21] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration[M]// Agapito L, Bronstein M M, Rother C. *Computer vision-ECCV 2014 workshops. Lecture notes in computer science*. Cham: Springer, 2015, 8926: 254-265.
- [22] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [23] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2023-02-05]. <https://arxiv.org/abs/1704.04861>.
- [24] Ma C, Huang J B, Yang X K, et al. Hierarchical convolutional features for visual tracking[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2016: 3074-3082.
- [25] Rout R, Parida P, Alotaibi Y, et al. Skin lesion extraction using multiscale morphological local variance reconstruction based watershed transform and fast fuzzy C-means clustering[J]. *Symmetry*, 2021, 13(11): 2085.

Hyperspectral Target Tracking Based on Spectral Matching Dimensionality Reduction and Feature Fusion

Guo Yecai^{1,2}, Cao Jialu^{1,2}, Han Yingying³, Zhang Tianmeng⁴, Zhao Dong^{1,2*}, Tao Xu^{1,2}

¹*School of Electronics & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China;*

²*School of Electronics and Information Engineering, Wuxi University, Wuxi 214105, Jiangsu, China;*

³*No. 703 Research Institute of China State Shipbuilding Corporation Limited, Harbin 150000, Heilongjiang, China;*

⁴*College of Aerospace and Civil Engineering, Harbin Engineering University, Harbin 150000, Heilongjiang, China*

Abstract

Objective Spectral features in hyperspectral video (HSV) enhance the ability to identify similar targets. However, HSV has high dimensions and a large amount of data, which causes great difficulties and high computing costs for feature extraction, and thus it is difficult to apply target tracking technology to HSV. In recent years, the development of snapshot hyperspectral technology has made it possible to acquire HSV. Many researchers have also turned their focus to HSV target tracking technology. In many target tracking processes, the target scale often changes to result in failed algorithm tracking. How to track targets robustly under the scale variations is an urgent problem to be solved.

Methods The algorithm is based on the correlation filtering framework and the scale-adaptive kernel correlation filter tracker. We employ the difference between the spectral curve of each pixel and the local spectral curve of the target and count the error value to segment the target pixel and the background pixel. The target spectral curve is obtained by averaging the target pixels, and the dimensionality reduction is realized by adopting the simple correlation between the target spectral curve and the image. Meanwhile, the dimensionally reduced image is input into the MobileNet V2 to extract deep features. The target area is judged by the local variance, and the 3D histogram of oriented gradient (HOG)

features of the target are enhanced. To preserve the unique spectral information of hyperspectral images and the semantic information of deep features, we utilize the method of channel convolution fusion to obtain more discriminative deep convolution HOG features which are fed into the filter to adapt to scale variations through the scale pooling idea.

Results and Discussions Three hyperspectral target tracking algorithms are selected for comparison in the experiment to verify the effectiveness of the proposed algorithm. Additionally, the results are presented in the experimental sequence for visualizing the performance of the algorithm. Fig. 7 presents the qualitative results of the algorithm on selected experimental sequences. In the book sequence, since the proposed algorithm adopts the scale pooling idea, it can estimate the scale of the target and track it stably. In the excavator sequence, the proposed algorithm is more robust by leveraging multi-feature fusion. The algorithm shows better adaptability when car sequence and face sequence are challenged by scale variations. We quantitatively evaluate the algorithm performance from two aspects of precision and success rate. Tables 1 and 2 present the values of the precision and success rates of the four algorithms respectively. Fig. 8 indicates the precision and success rate curves of each algorithm on the selected test sequences. Figs. 9 and 10 display the precision and success rate curves related to the scale variation challenge and the out-of-plane rotation challenge respectively. The precision and success rate results indicate that the proposed algorithm shows sound performance on the test sequence. As shown in Fig. 10, the proposed algorithm ranks first in the precision and success rate of the total test sequence. Specifically, the precision is improved by 3.3% and the success rate is increased by 2.2% compared with material based object tracking in hyperspectral video (MHT). Due to the utilization of fused features, the proposed algorithm is more robust. The precision and success rate of the proposed algorithm under scale variations are 19.8% and 14.0% higher than those of the second place, showing excellent adaptability (Fig. 11). The other three algorithms perform slightly worse under the scale variation challenge because they do not have corresponding scale estimation modules. Additionally, as shown in Fig. 12, the precision of the proposed algorithm under the out-of-plane rotation challenge is 1.31% lower than that of MHT, but it ranks first in the success rate due to the absence of a corresponding coordinate affine transformation strategy. Table 3 reveals the precision and success rate of the ablation experiment, and the proposed methods have improved the target tracking robustness.

Conclusions To solve the tracking failure caused by scale variations in hyperspectral target tracking tasks, we propose a hyperspectral target tracking algorithm based on spectral matching dimensionality reduction and feature fusion. Spectral matching dimensionality reduction provides low-dimensional feature input for the network and reduces computational complexity, and DC-HOG features improve target discriminability. Experimental results demonstrate that the proposed algorithm is better than other algorithms and can handle the scale variations well. Future research will explore an HSV target tracking algorithm with better performance to deal with challenges such as background clutter and out-of-plane.

Key words measurements; target tracking; hyperspectral video; spectral matching dimensionality reduction; channel convolution fusion