

# 光学卫星视频中微小动目标检测 方法与实例

主题方向：中国科创

专业领域/方向：电子信息

适用课程：《人工智能》、《卫星遥感进展》等

作者姓名：徐丰 王海鹏 王峰 王岩 胡凤鸣

工作单位：复旦大学信息科学与工程学院

中国专业学位案例中心

2024 年 2 月 20 日

# 光学卫星视频中微小动目标检测方法 with 实例\*

徐丰 王海鹏 王峰 王岩 胡凤鸣

**摘要：**光学遥感视频卫星是一种新型的对地观测技术，目前国内视频卫星技术日趋成熟，目前形成视频卫星星座，在对地观测方面有着巨大潜力和经济价值。如何充分发挥该星座的优势，拓展该数据在不同场景的应用价值。本案例针对视频卫星技术在大尺度城市场景中交通监测的应用，聚焦视频中微小的车辆目标的检测与跟踪方法。通过介绍光学卫星视频与应用场景的特点，从而引入微小动目标检测与跟踪算法，引导学生深入理解光学视频图像特点和解决特定应用场景下目标智能识别关键问题的思路，训练学生的创造性思维能力。

**Abstract:** Optical remote sensing video satellite is a new type of earth observation technology. Nowadays, the video satellite technology is becoming more and more mature, and the video satellite constellation is currently formed, which has great potential and economic value in earth observation. How to make full use of such technology and expand its applications in different scenarios. This case investigates the application of video satellite technology in traffic monitoring in large-scale urban scenes, focusing on the detection and tracking methods of tiny vehicle targets. This case introduces the characteristics of optical satellite video and application scenarios, followed by the micro-east target detection and tracking algorithm. Then it guides students to deeply understand the characteristics of optical video images and solve the key problems of intelligent target recognition in specific application scenarios, and train students' creative thinking ability.

**关键词：**光学卫星视频；微小动目标检测；目标跟踪；人工智能

**Keywords:** optical satellite video; Tiny moving target detection; target tracking; artificial intelligence

**作者信息：**徐丰，复旦大学信息科学与工程学院教授，副院长；王海鹏，复旦大学信息科学与工程学院教授，通信系副主任；王峰，复旦大学信息科学与工程学院副教授；王岩，复旦大学信息科学与工程学院副研究员；胡凤鸣，复旦大学信息科学与工程学院助理研究员

---

<sup>\*</sup>(1)本案例系教育部学位与研究生教育发展中心 2022 年主题案例项目成果(项目名称：智慧天眼——卫星遥感图像智能解译技术及应用；项目编号：ZT-221024634；首席专家：徐丰)。

(2)本案例复制权、发表权、信息网络传播权等相关权益由教育部学位与研究生教育发展中心依法享有，如有相关需要，请取得教育部学位与研究生教育发展中心授权。

(3)本案例只供课堂讨论之用，并无意暗示或说明某种行为是否有效

# 1 引言

光学遥感卫星视频是一种利用搭载在运动卫星上面的大口径的光学摄像头对地面拍摄得到的视频信号。光学卫星视频技术的高分辨率和实时监测特点，能够获取地面目标的动态变化信息，在城市规划、交通管理、自然灾害监测和安全监控等领域均能发挥重要作用。例如，通过分析卫星视频图像中的动态变化，可以实现实时监测城市交通流量、车辆行驶轨迹、道路拥堵情况等信息，为城市交通管理和规划提供数据支持。在应急事件处理和灾害监测方面，同时，光学卫星视频技术也扮演了关键角色。例如，当自然灾害如地震、洪水发生时，光学卫星视频可以立即监测受灾区域的情况，提供实时数据支持，帮助相关部门快速做出应急响应并展开救援工作。在国家安全相关领域，卫星视频可以用于监测城市安全状况、边境巡逻等任务，提高国家安全防卫能力。

图 1(a)是采集于西班牙巴伦西亚区域的一个卫星视频中的一帧图像，(b)是该区域对应的谷歌光学影像。这个卫星视频覆盖区域大约为 $3 \times 4$ 平方公里。此外，为了优化交通流量和减少拥堵，洛杉矶交通管理局利用卫星视频技术监测各主要交通路段的车流量。通过卫星图像分析，城市规划者能够实时监控交通状况，调整信号灯控制系统，优化道路使用策略。例如，在洛杉矶市中心和圣塔莫尼卡之间的 405 号公路，卫星技术帮助交通管理中心通过实时数据调整交通灯周期，大幅缓解了高峰期的交通压力。其次，2017 年，加利福尼亚州遭遇了史上最严重的山火之一。在这场灾难中，卫星视频技术发挥了关键作用。通过从空中捕获的高分辨率视频，应急管理部门能够快速评估火势蔓延情况和影响范围。这些信息帮助决策者制定疏散路线，并优先部署资源到最需要的地区。卫星图像也用于监控火后的环境影响，为重建和赔偿提供了必要的技术支持。在 2012 年伦敦奥运会期间，组织者利用卫星视频技术进行全面的安保监控。卫星图像帮助安保团队监控奥运村和比赛场馆周围的人流动态和潜在的安全威胁。实时视频分析确保了快速响应，如有紧急情况，可以立即调派安保人员进行处理。因此，借助于卫星视频我们可以监测一个城市尺度范围内的目标的动态变化。以城市交通状态监测为例，采用卫星视频技术检测和跟踪运动的车辆目标，在实际应用中基于卫星视频的车辆目标监测存在技术难点，具体表现在以下三个方面。



图 1 光学卫星图像示例：(a) 一帧光学卫星视频（西班牙·巴伦西亚）； (b) 对应的光学影像(Google Earth))

(1) 在卫星视频中，车辆目标非常微小。由于目前卫星视频分辨率较低，一辆车辆目标往往只有几个或者几十个像素点，车辆目标尺寸非常微小。所以，卫星视频中车辆目标的检测与跟踪问题实质上是一个非常微小的运动目标的检测与跟踪问题。图 2 展示了全景图图 1 中的两个典型的道路区域，环岛交叉道路与笔直的道路。从图 2 可以看出，车辆目标是与道路背景有一定对比度的若干个运动的像素点，由于分辨率限制，这些车辆目标没有明显的纹理、形状和颜色信息。最鲁棒的特点是持续的运动，但是车辆目标的运动容易被卫星视频背景的运动掩盖。

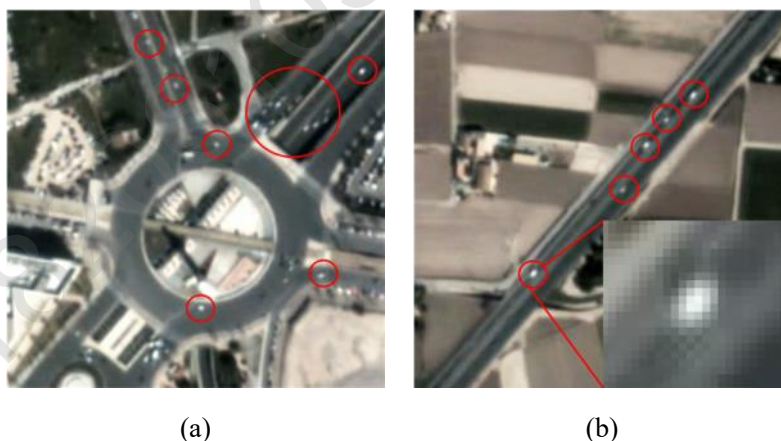


图 2 全景错误!未找到引用源。(a)中两个道路场景的放大图：(a) 环岛路口；(b) 笔直道路

(2) 卫星视频覆盖的范围巨大，环境复杂多变，如图 3。根据相机镜头和观测目标之前的距离，监控视频一般分为近场监控视频、中场监控视频和远场监控视频，卫星视频是一种新型的超远程监控视频，不仅能够提供目前最大的视频监控范围，而且能够覆盖一个复杂的城区环境。如下图所示，一个典型的卫星视频场景中包括了道路、建筑群、植被等地点，而且有笔直的道路、十字路口和环岛等多种复杂的交通场景。这给微小车辆目标的检测与跟踪带来了干扰和挑战。

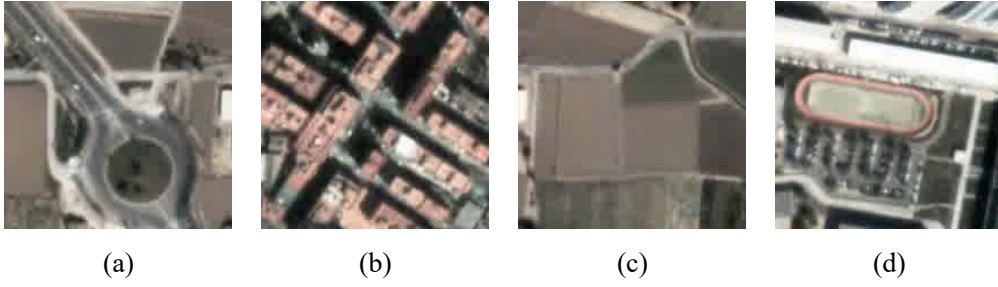


图3 卫星视频中的复杂城市环境：(a) 道路；(b) 建筑群；(c) 植被；(d) 体育设施

(3) 卫星视频的背景具有亚像素级和不均匀的运动。实验表明，卫星视频的背景有非常缓慢的移动，本研究中利用光流场来量化背景的运动。该卫星视频的第1帧与第100帧计算出的光流程如图4所示，(a)中的箭头的长度和方向表明了光流场的大小和方向，(b)是光流场幅度分布图。从图4可以看出，卫星视频的背景有不均匀的移动，这可能是由于背景在2维平面上的相对运动是卫星3维运动的投影。另外，由于卫星跟地面距离太远，卫星持续运动带来的视差依然很小，相邻两帧之间的平均移动低于1个像素距离。这样，不用对临近的帧进行图像配准，但是需要分离背景和车辆目标的运动。

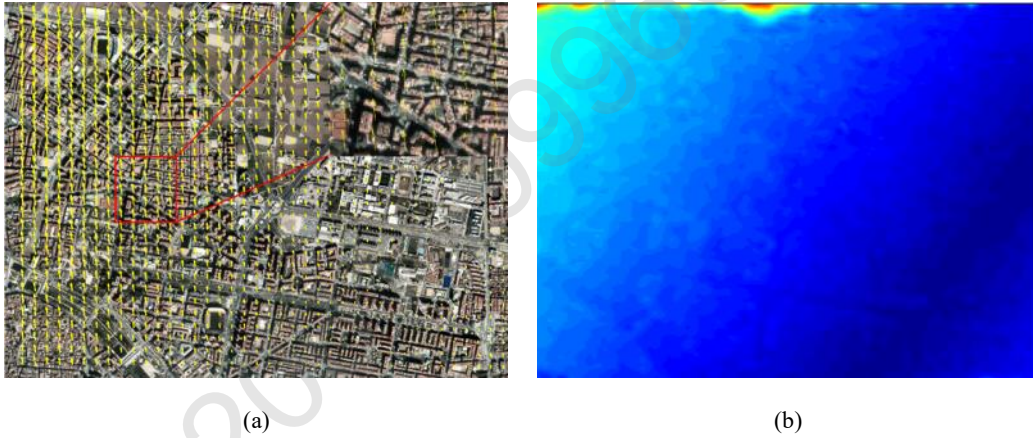


图4 卫星视频光流场：(a) 光流场强度与方向；(b) 光流场幅度图，颜色越深表示幅度越小

本案例关注城市环境中微小车辆目标的检测与跟踪，这些没有明显特征的车辆目标容易被噪声干扰。运动的背景会使得噪声信号出现虚假的但是有规律的移动，这会混淆车辆目标的检测与跟踪。为了解决上述的挑战，本研究提出了一个完整的卫星视频中微小车辆目标的检测与跟踪算法。特别地，本文提出了一种基于运动的目标检测算法，对局部噪声信号进行建模从而分离出真实的运动车辆目标。每一帧视频图像被当做一个二维信号，这个二维信号包括两个部分，一是原始的光学信号，二是一个加性的噪声信号。通过对噪声信号进行建模，从而识别可能存在的运动车辆目标。然后，设计了一个区域生长算法恢复完整的目标轮廓，并采用多种形态学特征对车辆目标和非车辆目标进行鉴别。上述的检测算法与整个基于卡尔曼滤波器的跟踪框架协作，从而实现对上百个车辆目标的实时跟踪。

## 2 微小运动车辆目标的检测与跟踪

### 2.1 算法框架简介

本案例提出的卫星视频中微小车辆目标的检测与跟踪框架如图 5 所示，该算法包括多个模块。本框架中用到的模块和术语有：

#### (1) 检测/检测器

检测是指通过局部噪声建模算法寻找可能是运动车辆目标的过程；

#### (2) 候选目标

检测器的输出称之为候选目标；

#### (3) 区域生长

区域生长是恢复检测到的目标的完整轮廓的过程，区域生长后的结果才是目标的完整轮廓，因此其输出也称之为候选目标；

#### (4) 鉴别/鉴别器

鉴别是指区分真实的车辆目标与噪声的过程，本研究中利用多种形态学特征来区分真实的车辆目标和其他噪声目标；

#### (5) 假设目标

鉴别器的输出称之为假设目标，因为此时依然可能存在一些噪声信号，所以假设目标不一定是真实的车辆目标，卡尔曼滤波器的跟踪过程会逐步淘汰掉随机的噪声目标；

#### (6) 状态

状态是一个 6 维向量，包括车辆目标的在 XY 方向上的位置、速度和加速度；

#### (7) 轨迹

一个车辆目标在时间维度上所有状态的集合叫这个车辆目标的轨迹，由此可知，车辆目标在视频中运动的路径只是轨迹的前两个维度；

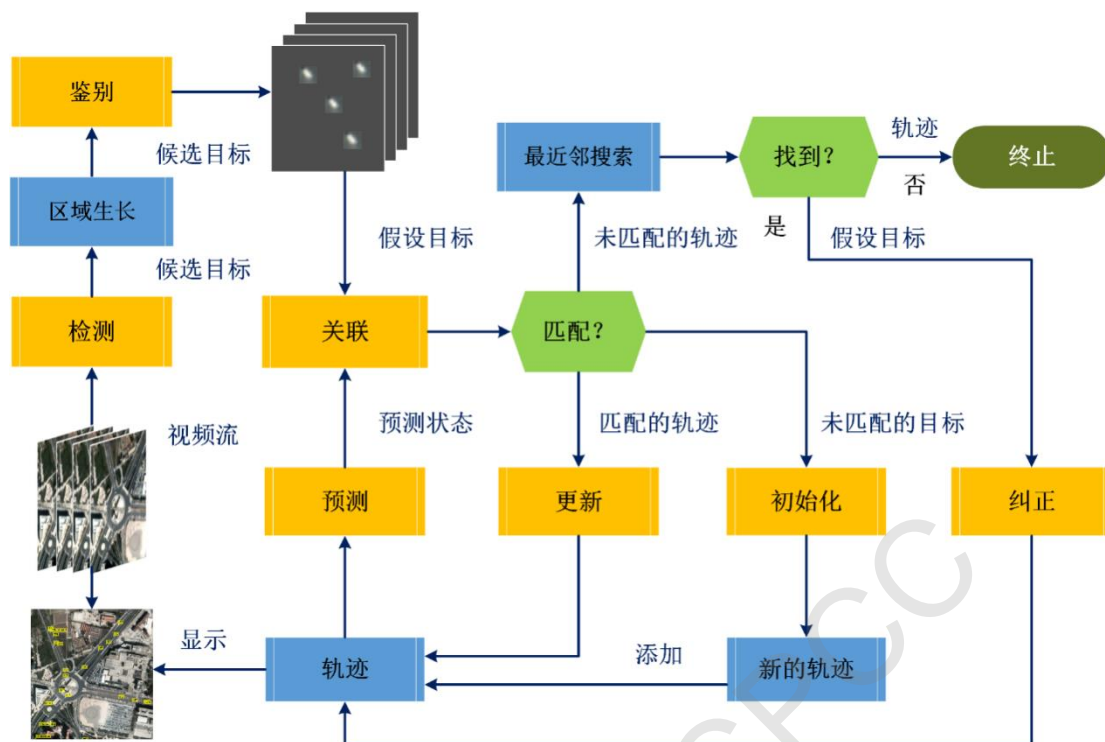
#### (8) 匹配

匹配是当有检测到多个假设目标时候，这些假设目标与所有轨迹最优的配对；

#### (9) 预测

预测是指卡尔曼滤波器的对跟踪的车辆目标下一个时刻状态的估计过程。





广泛采用的跟踪算法包括卡尔曼滤波器，粒子滤波（Particle Filter）和 Mean Shift。如图 5 所示，卡尔曼滤波器是卫星视频中微小车辆目标的检测与跟踪框架的核心部件，这个框架的运行包括几个关键的算法流程，如下：

### (1) 初始化

初始化是确定一条轨迹的初始状态,也就是确定这个车辆目标对应的卡尔曼滤波器的初始位置、初始速度和初始加速度。相邻两帧检测到的同一个假设目标通过重叠像素进行匹配,从而可以计算出该目标的初始速度。在本研究中,初始加速度设置为 0。

## (2) 预测

预测是指通过之前的观测值预测目标在当前的状态。

### (3) 假设目标与路径匹配

鉴别器会输出许多假设目标，跟踪的路径会预测出许多当前状态的预测值。假设目标与路径匹配是指将这些假设目标与多个轨迹进行匹配，整体满足最小的匹配代价，从而实现同时跟踪多个车辆目标。在本研究中，采用匈牙利算法实现多个目标与轨迹直接的最优匹配。匹配过程会产生不同的匹配情况，包括匹配的轨迹和目标，无法匹配的轨迹与无法匹配的目标。匹配的目标用来更新轨迹的卡尔曼滤波器的状态，即第（4）步；无法匹配的轨迹转到第（5）步；无法匹配的目标用来初始化新的轨迹。

#### （4）更新

更新是指用新的观测值来更新卡尔曼滤波器的状态。

#### （5）最近邻搜索，轨迹修正与轨迹终止

无法匹配的轨迹不能简单的丢弃或者终止跟踪，因为可能是在某些帧中没有成功检测到这些目标。所以，在跟踪框架中补充了一个最近邻搜索的策略提高对这些漏检目标的召回率。最近邻搜索是指在这个目标上一帧存在的位置的附近范围内搜索这个目标。匹配算法借助于 SSIM（Structural Similarity Index, SSIM）图像相似度评价指标。如果，最近邻搜索算法重新找到了这个目标，那么就用新的数据更新状态；否则就终止跟踪这个目标。基于 SSIM 的最近邻搜索算法结果如图 6 所示，该实验说明最近邻搜索策略可以有效地搜索到一部分检测识别的目标。

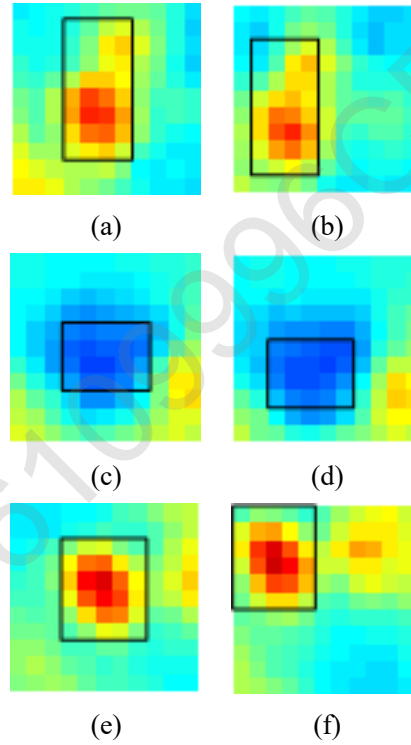


图 6 三个最近邻搜索的实例：图中黑色矩形框表示跟踪的车辆目标，第 1 列表示上一帧中目标的位置，第 2 列表示基于 SSIM 的最近邻搜索算法找到的车辆目标结果，为了视觉效果显著原始图像被转换成了伪彩色图像

在本研究中，卡尔曼滤波器被用来跟踪运动的车辆目标，它是在线性和高斯条件下的最优跟踪器。尽管在现实世界中车辆的运动比较复杂，但是在短时间内车辆的运动近似认为是匀加速直线运动是一个合理的假设。因此，卡尔曼滤波器对线性运动的车辆目标来说是一个简单但是有效的跟踪器。卡尔曼滤波器的演化方程可以写成：

$$\mathbf{X}_i = \mathbf{F}_i \cdot \mathbf{X}_{i-1} + \mathbf{V}_i \quad (2-1)$$



公式中， $\mathbf{X}_i$ 、 $\mathbf{F}_i$ 和 $\mathbf{V}_i$ 分别表示状态，演化矩阵和过程噪声，下标*i*表示第*i*帧图像。根据车辆目标状态的定义，状态向量 $\mathbf{X}_i$ 可以写为：

$$\mathbf{X}_i = [x, y, v_x, v_y, a_x, a_y]^T \quad (2-2)$$

公式中， $(x, y)$ 表示车辆目标的位置， $(v_x, v_y, a_x, a_y)$ 表示速度与加速度。不失一般性，相邻两帧之间车辆目标的运动可以认为是匀加速直线运动，所以演化矩阵 $\mathbf{F}_i$ 的定义为：

$$\mathbf{F}_i = \begin{bmatrix} 1 & 0 & \tau & 0 & \tau^2/2 & 0 \\ 0 & 1 & 0 & \tau & 0 & \tau^2/2 \\ 0 & 0 & 1 & 0 & \tau & 0 \\ 0 & 0 & 0 & 1 & 0 & \tau \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (2-3)$$

公式中， $\tau$ 表示两帧之间的时间间隔，也就是帧率的倒数。卡尔曼滤波器中观测方程的定义为：

$$\mathbf{Y}_i = \mathbf{H}_i \cdot \mathbf{X}_i + \mathbf{n}_i \quad (2-4)$$

公式中， $\mathbf{Y}_i$ 、 $\mathbf{H}_i$ 和 $\mathbf{n}_i$ 分别表示测量向量，测量矩阵和测量噪声。在本研究中，每次需要观测跟踪车辆目标的位置，所以测量矩阵 $\mathbf{H}_i$ 的定义为：

$$\mathbf{H}_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2-5)$$

卡尔曼滤波器的原理是不断用观测到的新的样本数据来更新状态模型，假设在第*i*帧时，对一个样本观测了*i*次，也就是 $\mathbf{Y}_{1:i} = \{\mathbf{y}_k | k = 1, 2, \dots, i\}$ 。卡尔曼滤波器可以得到此时状态 $\mathbf{X}_i$ 出现得到后验概率，根据贝叶斯定理可以得到：

$$p(\mathbf{X}_i | \mathbf{Y}_{1:i}) = \frac{p(\mathbf{Y}_i | \mathbf{X}_i) p(\mathbf{X}_i | \mathbf{Y}_{1:i-1})}{p(\mathbf{Y}_i | \mathbf{Y}_{1:i-1})} \quad (2-6)$$

## 2.2 基于本地噪声建模的运动目标检测

帧差法是一种传统的视频中运动目标检测的方法，但是传统的帧差法需要设置一个固定的阈值从而在差分图像上分割出运动的目标和静止的背景，这样可以得到一个与原图对应的前景背景分类图。上述的过程在本研究中称之为“二值化”，但是一个人工设置的分割阈值无法适应卫星视频中运动车辆目标的检测，因为卫星视频覆盖范围大、环境复杂。另外考虑到相邻像素相似的原理，本文拓展了传统的帧差法引入了本地噪声建模的策略，从而实现在对运动像素和缓慢运动卫星视频背景的有效分割。

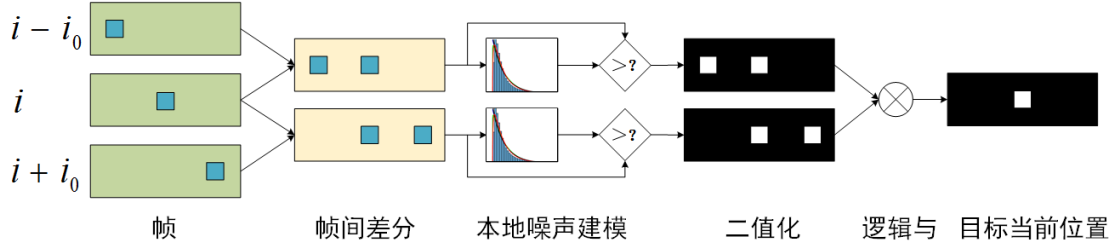


图 7 基于局部噪声建模的运动目标检测算法

普通的监控视频，背景单一稳定，覆盖范围有限。卫星视频覆盖范围大，环境复杂，本地策略就是将整个视频图像划分成小的区域分别处理，因为相邻像素具有相似性。而且，对小的区域进行建模可以解决背景的缓慢移动问题。本研究中，一个本地区域的大小为 $30 \times 30$ 像素。本文提出的卫星视频中微小车辆目标的检测算法流程如图 7 所示，包括 4 个步骤：（1）计算帧间差分图像；（2）本地噪声建模；（3）二值化；（4）确定目标当前位置，每个步骤的详细操作过程与原理如下。

#### （1）计算帧间差分图像

计算两帧之前的差分图像是一个常规操作，在本研究中我们提供了一种对每帧视频图像全新的看法，从而可以导出本地噪声建模与运动前景分割算法。一帧视频图像可以看做一个二维信号，但是接收到的二维信号往往会叠加一些噪声，所以一帧图像可以写为一个二维光学信号与一个二维加性噪声的和，也就是：

$$G_i(x, y) = g_i(x, y) + n_i(x, y) \quad (2-7)$$

公式中， $G_i(x, y)$ 表示像素 $(x, y)$ 在第 $i$ 帧图像上的灰度值， $g_i(x, y)$ 表示原始信号的灰度值， $n_i(x, y)$ 表示噪声信号的幅度。在本研究中，原始的 RGB 图像都转换成灰度图像处理。因为卫星视频分辨率低，车辆目标的颜色特征不太明显，转换成灰度图像可以简化处理过程。因此，完全配准的两帧图像的差分图像应该只保留噪声信号，也就是：

$$D_{i,i+k}(x, y) = |G_i(x, y) - G_{i+k}(x, y)| = |n_i(x, y) - n_{i+k}(x, y)| \quad (2-8)$$

公式中， $D(\cdot)$ 表示两帧图像差分的绝对值， $k$ 表示两帧图像的间隔。从公式(1-8)可以看出，帧间差分的结果只与噪声信号有关，噪声信号也包含了背景运动带来的噪声。所以，对噪声信号进行建模，拟合噪声信号的分布就能检测出可能的运动像素。

#### （2）本地噪声建模

微小车辆目标检测的挑战在于噪声会影响车辆目标的检测，运动目标检测的

问题就转换成噪声信号中局外点的分离问题。因为直观上来说，两帧图像的差分结果应该是 0，但是由于背景运动带来的噪声和其他噪声的影响使得差分结果是一个二维随机噪声图像。运动像素点就是在差分图像上不符合噪声图像分布的像素点。图 8 显示了噪声信号幅度分布直方图，可以看出，噪声信号的分布有显著的规律性，如平滑的衰减和长长的拖尾。

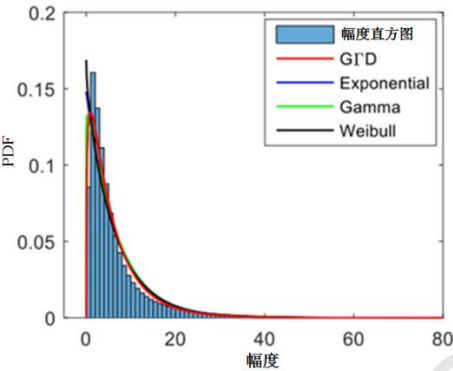


图 8 噪声信号幅度分布直方图和概率分布拟合效果

图 8 中噪声信号幅度分布重拖尾的特点可以用一些常用的概率分布函数进行拟合，如广义伽马分布（Generalized Gamma Distribution, GFD）、负指数分布（Exponential Distribution）、伽马分布（Gamma Distribution）和韦布尔分布（Weibull Distribution）。拟合实验的结果如图 8 所示，为了量化这些长拖尾分布对差分图像噪声的拟合效果，在本研究中采用 KL 距离和 KS 距离对拟合效果进行量化，结果如表 2-1 所示。

表 2-1 概率分布对噪声信号的拟合情况

拟合指标	指数分布	伽马分布	韦布尔分布	GFD	帧
KL	0.0959	0.0914	0.0919	0.0544	50
KS	0.0959	0.1018	0.0891	0.0813	
KL	0.0864	0.0812	0.0862	0.0579	100
KS	0.0875	0.0988	0.0896	0.0865	
KL	0.0846	0.0800	0.0845	0.0531	500
KS	0.0854	0.0964	0.0859	0.0816	

从定量和定性的拟合结果可以看出，这些具有长拖尾特征的概率分布都能很好的拟合差分图像中的随机噪声分布。特别地，自由度最高的 GFD 有最好的拟合效果，但是只有一个参数的指数分布的拟合效果也同样能够接受。因为概率分布参数越多，参数估计就越困难，所以本研究中采用指数分布拟合卫星视中差分

图像的随机噪声分布，其 CDF 为：

$$c_E(x; \lambda) \begin{cases} 1 - \exp(-\lambda x) & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (2-9)$$

### (3) 二值化

用指数分布拟合局部噪声信号的分布模型后，设置一个虚警率 $p_{fa}$ 来确定分割运动前景像素和静止背景像素的阈值，这个二值化的阈值为：

$$th = c_E^{-1}(1 - p_{fa}; \hat{\lambda}) \quad (2-10)$$

公式中， $c_E^{-1}(\cdot)$ 表示指数分布 CDF 的逆函数， $\hat{\lambda}$ 表示指数分布参数的估计值，

在本研究中 $p_{fa}$ 的值设置为 $5 \times 10^{-2}$ 。用指数分布拟合一个 $30 \times 30$ 像素大小区域的噪声分布后，根据式(1-10)推导出该区域运动像素和静止像素的分割阈值 $th$ ，如果在帧间差分图像上，这个像素的值大于 $th$ ，那么就认为这个像素是潜在的运动车辆目标像素。这样就实现了图像的二值化过程，运动像素为 0，静止像素为 1。可以看出，缓慢运动的背景带来的影响也被认为是一种噪声，所以本文提出的方面法可以适应背景缓慢运动的卫星视频。另外，利用局部处理的策略，就可以很好的解决图 4 中所示背景运动不均匀的问题，也就能适应这种大尺度复杂城区环境。

### (4) 确定目标当前位置

如图 4 所示，同一个运动目标在一幅二值化图像上有两个位置，因为根据 (3) 中的分析，一个运动目标在当前位置和上一时刻的位置都会被识别成运动像素，所以第 $i - i_0$ 帧和第 $i$ 帧的差分无法确定运动目标在第 $i$ 帧的位置。采用三帧差分，求出第 $i - i_0$ 帧和第 $i$ 帧差分图像的二值图像与第 $i + i_0$ 帧和第 $i$ 帧差分图像的二值图像的交集，就能确定一个运动目标在第 $i$ 帧图像上的位置，在本研究中 $i_0$ 的值设置为 10。同时，该操作也能很好的去除随机出现的噪声信号。

## 2.3 区域生长与基于多种形态学线索的虚警鉴别

因为观测角度的变换、地面目标反射强度的变换、背景的运动等原因，图像中一些“静止”的像素有时候会被误认为是“运动”的像素，这样的像素在本研究都任务认为是噪声。卫星视频中的车辆目标只有几十个甚至几个像素，这么小尺度的车辆目标非常容易受到噪声信号的干扰，因为一般的噪声信号或者非车辆目标也是相同的尺寸，噪声信号的会大大降低检测器的精度，同时给跟踪器带来不必要的计算量。根据 2.2 节的分析，只有同时出现两幅差分图像中的候选像素才会保留下来，这样不仅能确定运动车辆目标在当前帧位置也能有效地去除一些随机

出现的噪声信号。另外，噪声的出现一般具有随机性和存活寿命短的特征，卡尔曼滤波器在跟踪这些噪声运动的时候无法捕捉到规律的运动，这样跟踪器也能逐渐淘汰噪声信号。

根据对卫星视频特性的分析，卫星视频的背景有亚像素级不均匀的运动，这样会使得位于图像梯度变换大的像素在差分图像上的值很大，如一些建筑物的边缘区域、建筑物的角点。这些噪声形成规律的运动，如图 9 所示，这些噪声的运动轨迹符合光流场的大小和方向，它们的运动实际上就是背景运动的反映。这些噪声的出现具有规律性，它们的运动也具有规律性。在本研究中这样的噪声称之为“规律噪声”。

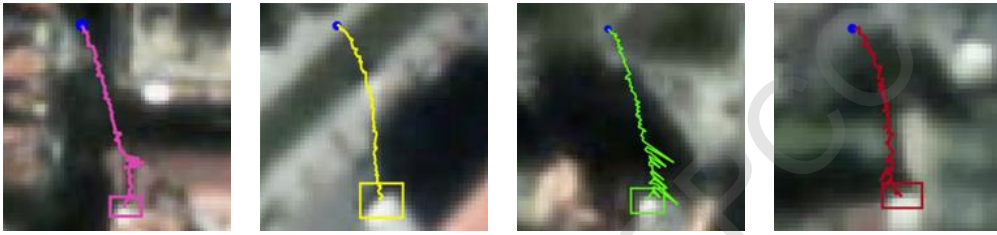


图 9 规则运动的噪声信号

为了鉴别规律噪声和真实的运动车辆目标，基于像素分布相邻相似的原理，本文提出了区域生长算法重构目标的完整轮廓，然后设计了一系列形态学鉴别算子区分这些规律噪声和真实的车辆目标。

如图 10 所示，(a)和(b)是车辆目标，(c)和(f)是检测器输出的检测结果，因为在相邻帧中车辆目标的位置有一定的重叠，所以检测器的输出是车辆目标的一部分像素而不是全部的轮廓。同样地，检测器错误识别的噪声信号(c)和(d)只是梯度变换较大的区别的部分像素，这些像素跟邻近的像素具有相似性。根据相邻相似原理，就能恢复出一个检测到的目标的完整轮廓，这样有利于后面对该目标的判断鉴别。为了便于观察目标空间分布的特点，图 10 中(a)-(d)转化成了伪彩色图像，可以发现相邻像素不仅具有相似的灰度值，而且一个目标的灰度值分布有逐渐变化的特点，这启发我们用高斯分布评价一个目标与周围像素的相似度。用检测到的目标的样本像素估计高斯分布的参数，高斯分布的 CDF 为：

$$c_G\{x|\mu, \sigma\} = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{x-\mu}{\sqrt{2}\sigma} \right) \right] \quad (2-11)$$

公式中， $\operatorname{erf}(\cdot)$ 、 $\mu$ 和 $\sigma$ 分布表示误差函数、均值和标准差。

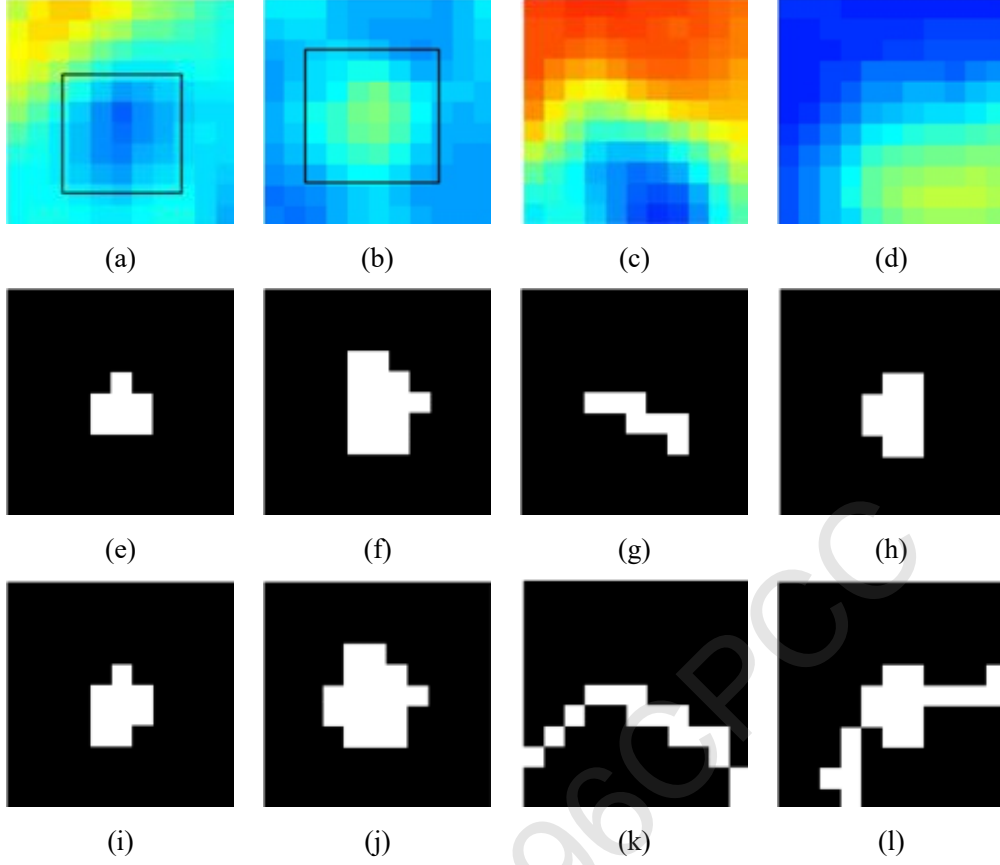


图 10 车辆和噪声目标检测与区域生长结果：(a)-(d) 目标在原图像上的分布，黑色矩形框标注了真实车辆目标，(c)和(d)的噪声目标，图像转换成了伪彩色图像便于观看；(e)-(h) 是运动目标检测器的输出；(i)-(l) 是区域生长后的结果

利用高斯分布拟合样本的分布，确定相似像素灰度值的上界 $th_G^+$ 和下界 $th_G^-$  分别为：

$$th_G^+ = c_G^{-1}(1 - p_G | \hat{\mu}, \hat{\sigma}) \quad (2-12)$$

$$th_G^- = c_G^{-1}(p_G | \hat{\mu}, \hat{\sigma}) \quad (2-13)$$

公式中， $c_G^{-1}(\cdot)$ 是高斯分布的 CDF 的逆函数， $p_G$ 设置为 $5 \times 10^{-3}$ 。相邻像素的定义为，在以目标中心为中心的 $11 \times 11$ 像素窗口大小区域内的像素是该目标的相邻像素。在这个窗口内的像素的值如果属于区间 $[th_G^-, th_G^+]$ ，这个像素就认为是相似像素。最后，所有与原目标连通的相似像素与原目标一起构成区域生长后的结果，如图 10(i)-(l)所示。从图 10 可以看出，区域生长的结果与原图中像素的分布吻合，完整回复了车辆目标和噪声目标的轮廓。区域生长前(e)-(h)，车辆目标与噪声目标的形态接近(i)-(l)，很难区分；区域生长后，车辆目标接近于矩形，噪声目标的形态具有随机性。所以，区域生长后的结果有助于从形态学上鉴别车辆目标和噪声目标，因此本文提出了一系列的形态学特征区分真实的车辆目标和噪声



目标。

因为视频卫星离地面距离远，车辆目标是刚体，所以运动车辆目标在卫星视频中的形态保持不变。而且卫星视频中的像素大小对应真实尺寸，视频中目标的大小具有物理意义。本文采用的形态学特征包括：面积，填充率，长度，离心率。

面积：目标的像素个数；

填充率：填充率是面积与 **bbox** (**Bounding Box**) 的比值；

长度：如果一个椭圆的归一化二阶中心矩与目标相同，那么这个椭圆的长轴长度就是这个目标的长度；

离心率：目标的离心率就是上述椭圆的离心率。

本研究中采用的卫星视频的分辨率是 1 米，所以面积和长度代表的是这个目标的尺寸特征。填充率和离心率描述的是目标的形状特征，也就是跟矩形的相似性。尽管卫星视频的分辨率低，如图 2 所示，无法观测到车辆目标细节的形状特征，但是在这个尺度下，车辆目标的形状近似矩形，填充率和离心率可以有效地描述车辆目标与矩形的相似程度。

### 3 检测与跟踪性能评价指标与评价方法

#### 3.1 评价指标

在目标检测领域常用的检测器性能评价指标包括精确率/召回率曲线，平均精度(AP)等。这些指标在许多视觉目标检测与识别数据集中得到了广泛的应用，如 PASCAL VOC 和 MOT。这些评价指标和评价方法不能直接应用于卫星视频中微小车辆目标的检测与跟踪。车辆目标只有几个至几十个像素，没有明显的颜色与纹理特征，识别困难，容易跟其他噪声混淆。这不仅给车辆目标的检测与跟踪带来很大的挑战，而且给运动车辆目标真值的标注带来很大的困难。为了全面合理的评价车辆目标检测和跟踪算法的性能指标，保证不同算法性能之间的公平比较，本研究引入了一系列的检测跟踪性能评价指标，设计了一套性能评价流程，而且提供了一套标注的视频卫星数据集。本研究开发的评价指标、评价方法和标注数据集会促进卫星视频处理技术的发展。

一般来说，单一指标不能全面的评价检测器和跟踪器的性能，本研究中引入了 6 个评价指标，包括精度、召回率、 $F_1$  - score、雅可比相似度、MOTA 和 MOTP，这些参数的定义如下。

##### (1) 精度

在目标检测领域中，人们往往关心检测到的目标是不是真实的目标或者研究感兴趣的目標。如果，一个检测到的目标对应一个真值 (Ground Truth)，那么这

个目标称之为 TP (True Positive), 否则这个目标称之为 FP (False Positive), FP 是指检测到的目标不是真实的目标。如果一个真实的目标没有检测到, 这个目标称之为 FN (False Negative)。精度 (Precision)是指检测到的目标中真实的目标所占的比率, 定义为:

$$Precision = \frac{TP}{TP+FP} \quad (3-1)$$

## (2) 召回率

召回率是指检测器能成功从所有的真实目标中检测出来的目标的概率, 定义为:

$$Recall = \frac{TP}{TP+FN} \quad (3-2)$$

## (3) $F_1 - score$

$F_1 - score$ 是一个传统的评价二分类的分类性能的指标, 它等于精度和召回率的调和平均数, 也就是:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3-3)$$

## (4) 雅可比相似度

雅可比相似对(Jaccard Similarity)是跟踪性能的评价指标, 雅可比相似度综合了 TP、FP 和 FN, 他的定义为:

$$J = \frac{TP}{TP+FP+FN} \quad (3-4)$$

## (5) MOTA

多目标跟踪准确度(Multiple Object Tracking Accuracy, MOTA)是评价多个目标跟踪系统性能的评价指标, 定义为:

$$MOTA = 1 - \frac{\sum_i (FN_i + FP_i + IDSW_i)}{\sum_i GT_i} \quad (3-5)$$

公式中,  $FN_i$ 、 $FP_i$ 、 $IDSW_i$ 和 $GT_i$ 分别表示第  $i$  帧图像中 FN、FP、IDSW 和真值的个数。IDSW 一条真实轨迹在检测轨迹中 ID (Identification)的变化。根据定义, MOTA 的取值范围是 $-\infty$ 到 1。MOTA 的值越大, 说明多目标跟踪器的性能越好。

## (6) MOTP

多目标跟踪精度(Multiple Object Tracking Precision, MOTP)是评价多目标跟

踪系统对多个目标定位精度的指标，它的定义为：

$$\text{MOTP} = \frac{\sum_i \text{IoU}_i}{\sum_i M_i} \quad (3-6)$$

公式中，IoU (Intersection over Union)表示交集并集比率， $\text{IoU}_i$ 表示第  $i$  帧图像中所有检测到目标与对应真值的 IoU 的和， $M_i$ 表示第  $i$  帧图像中目标与成功配对的真值的个数，也就是 TP (True Positive)的个数。从定义中可以看出，MOTP 的取值从 0 到 1，取值越大说明多目标的检测定位越精确。

### 3.2 评价方法

检测到的目标与真值的 bbox 的 IoU 在本研究当成目标与真值之间的相似度。类似于目标与轨迹的匹配，检测到的轨迹与真值轨迹之间的匹配在时空域上借助于匈牙利算法进行全局最优匹配。

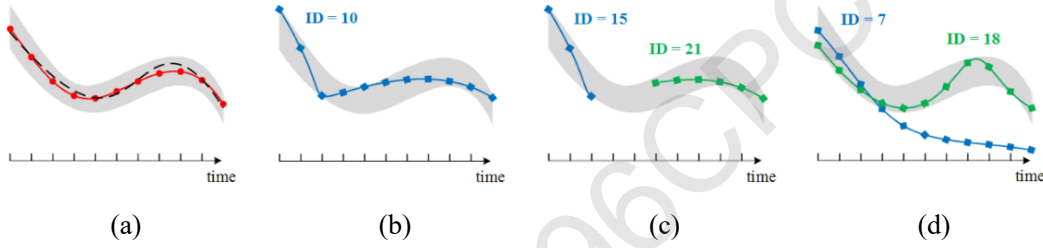


图 11 一些检测到的轨迹与真值轨迹匹配示意图：(a)中黑色虚线、红色实线和灰色区域分别表示人工标注的目标轨迹真值、检测到的目标轨迹、属于该真值轨迹的范围；(b)表示检测到的轨迹可以完全匹配上真值轨迹；(c)表示两个不同时间段的两条检测轨迹匹配上了同一条真值轨迹；(d)中，检测到的轨迹 7 在前几帧与真值轨迹配准的更好，但是在较长时间内检测轨迹 8 与真值匹配的更好

一些检测到的目标轨迹与人工标注的真值轨迹之间匹配情况如图 11(b)-(d)所示，图 11(a)是跟踪轨迹和真值理想匹配的情况。从图 11 可以看出，在多目标跟踪系统中，真值与检测值的匹配情况复杂，评价方法会影响到评价结果。如图 11(d)，如果采用逐帧匹配策略则不能很好反应检测器在前几帧产生的误差，因为在前几帧检测轨迹 7 会匹配上真值轨迹，之后轨迹 18 会匹配上同一条真值轨迹。显然，在这种情况下逐帧匹配会产生不合理的结果。在本研究中，真值和检测值直接的匹配在连续 10 帧的时空域上进行。多车辆检测与跟踪系统性能评价流程为：

计算每个检测到的车辆目标与真值的 IoU，然后将 IoU 的倒数作为车辆目标与标注真值之间的“距离”。显然距离越大说明车辆目标与真值的重叠越小，距离为 $\infty$ 时候目标与真值不重叠（所有的 IoU 会加上一个小值避免分母为 0）。假设在第  $i$  帧检测到  $M$  个车辆目标，人工标注  $N$  个真值，可以得到这  $M$  个车辆目标与  $N$  个真值匹配时候的距离矩阵为：

$$CM_i = \begin{bmatrix} \frac{1}{IoU_{1,1}} & \frac{1}{IoU_{1,2}} & \cdots & \frac{1}{IoU_{1,N}} \\ \frac{1}{IoU_{2,1}} & \frac{1}{IoU_{2,2}} & \cdots & \frac{1}{IoU_{2,N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{IoU_{M,1}} & \frac{1}{IoU_{M,2}} & \cdots & \frac{1}{IoU_{M,N}} \end{bmatrix} \quad (3-7)$$

检测到从车辆目标与真值匹配的最大距离在实验中设置为 50，也就是指 IoU 最小值为 0.02。这意味着，当一个检测到的车辆目标与一个真值的 IoU 大于 0.02 时，就认为这个车辆目标是真实的目标。通常在许多目标检测任务（如，行人检测、飞机检测、车辆检测等）中成功检测的 IoU 都设置为 0.5，因为这些目标一般占有很多像素，标注真值的误差小。在卫星视频中的车辆目标检测中设置较小的 IoU 阈值可以容忍人工标注的误差，同时也能保证实际应用的需求。

(1) 对连续  $K$  帧计算距离矩阵  $CM$ ，可以得到一个张量

$$CM_{1:K} = [CM_1, CM_2, \cdots, CM_K] \quad (3-8)$$

在本研究中  $K$  设置为 10，理论上应该计算全部帧的匹配结果然后求最佳的匹配结果。为了简化运算，本研究选取了  $K = 10$  的时间窗口。

(2) 根据匈牙利算法求的最佳的检测目标与真值的匹配结果。

(3) 重复(1)-(3)完成对这个视频全部帧检测结果和真值的匹配，最后计算 3.1 中所有的评价指标结果。

## 4 实验与讨论

### 4.1 实验设置：数据集，比较算法与真值

本实验采用的是长光卫星技术有限公司提供的卫星视频，该视频采集于 2017 年 3 月 7 日，该视频覆盖的区域是西班牙巴伦西亚市的一个区域，该视频的详细参数与地理位置信息见表 4-1。从表 4-1 可以看出，这个卫星视频覆盖范围大，帧率高。跟普通监控视频相比，卫星视频有分辨率低，视频尺寸大的特点，这给卫星视频中目标的检测和跟踪带来巨大的挑战，对处理算法的实时性要求较高。

本实验中采用的比较算法是在运动目标检测领域广泛采用的 GMM 和 ViBe，为了保证算法性能比较的公平性，GMM 和 ViBe 运动目标检测器同时连接到和本文提出的算法相同的卡尔曼滤波器的运动目标跟踪框架中。

表 4-1 实验使用的卫星视频的参数与地理位置信息

视频采集位置	帧率(fps)	分辨率(米)	时长(秒)	高×宽(像素)
--------	---------	--------	-------	---------

西班牙·巴伦西亚	20	1	29	3072 × 4096
图像位置	左上角	右上角	左下角	右下角
经纬度	39.4989N 0.3719W	39.4928N 0.3278W	29.4731N 0.3775W	39.4669N 0.3333W

由于目前世界上没有公开的相关的数据库,本文采用了自己手动标注的数据库。卫星视频中的运动车辆目标的标注是一个繁重而艰巨的任务,因为卫星视频的分辨率低,所以车辆目标本身没有特别的颜色、形状和纹理特征,如图 12 所示,这给肉眼识别人工标注车辆目标带来了很大的挑战。图 12(a)-(c)是原始卫星视频中 3 个运动的车辆目标, (d)是一个噪声目标(非运动车辆目标),可以看出从外观上来看车辆目标与噪声非常的类似,而且车辆目标与背景的对比度也并不强烈。图 12(e)-(h)是(a)-(d)进行直方图调整之后的图像,便于肉眼识别出目标的位置,可以看出目标只是一些相邻的像素点,它们唯一突出特征就是连续的有规律的运动。而且运动目标的标注与一般图像的中静止目标的标注不同在于,同一个运动目标在视频中应该标注同一个标签(Identification, ID),这样标注运动车辆目标的真值的时候需要同时检查连续若干帧图像,根据车辆目标的运动进行标注。

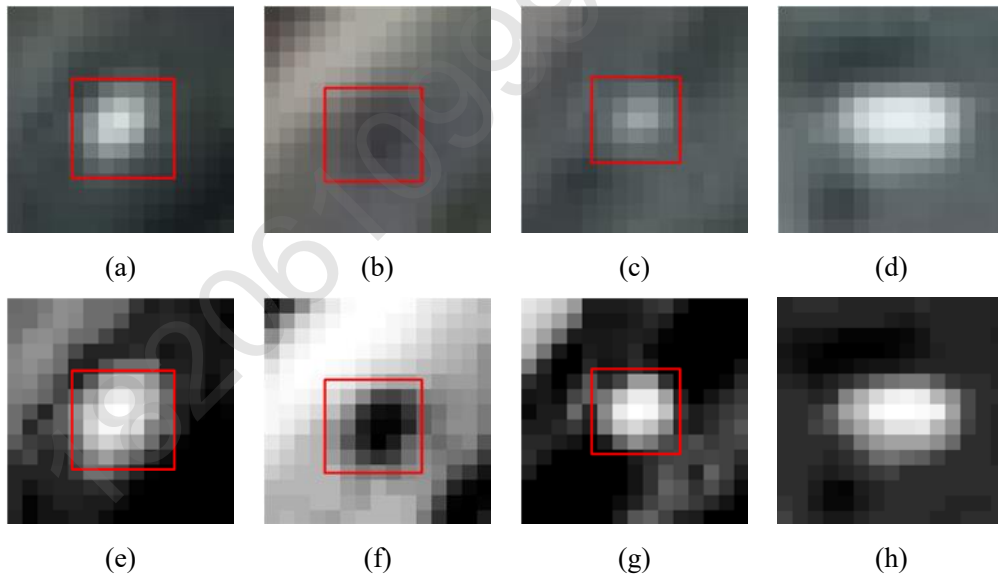


图 12 一些真实车辆目标的和一个噪声目标的样本: (a)-(c)是车辆目标, 图中红框标注了人眼确定的车辆的位置; (d)是一个与车辆目标相似的噪声目标; (e)-(h)是(a)-(d)对应的灰度调整后的图片, 通过增强对比度使得马路上的车辆目标在视觉上更加显著

为了减少人工标注的工作量,而且合理的评价算法的性能,本研究随机选择了实验卫星视频中的 3 个 $500 \times 500$ 像素大小的区域进行标注,如图 13 所示。本研究采用 MATLAB 2018a 软件自带的视频真值标注程序 Ground Truth Labeler 对视频进行标注,给每一个车辆目标分布一个 ID,用 bbox 对车辆目标进行标注,

如图 14 所示。在本研究中，对卫星视频每隔 10 帧标注一次，然后对标注结果在时间域上进行 10 插值，对目标的中心位置采用线性插值，对目标 bbox 的大小采用最近邻插值。在测试卫星视频中，标注区域 1、标注区域 2 和标注区域 3 分别有 49、41 和 29 个车辆目标。在实验中，我们对本文提出的算法和比较算法在整个视频全景中的性能进行定性比较，然后在这 3 个真值标注区域进行定量的比较。

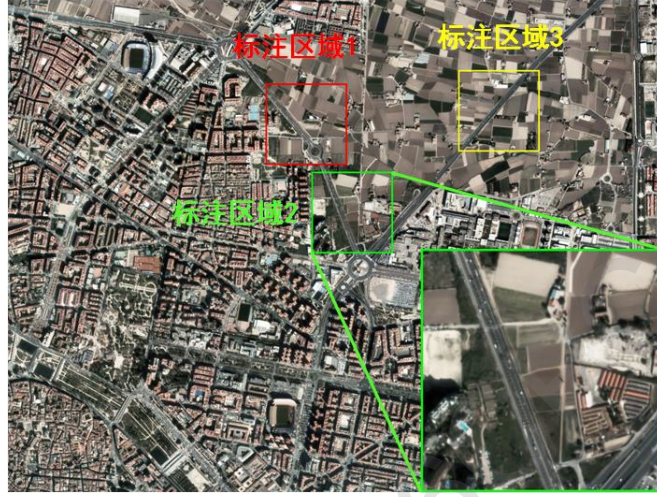


图 13 实验中标注的区域：随机选择 3 个区域人工标注真值

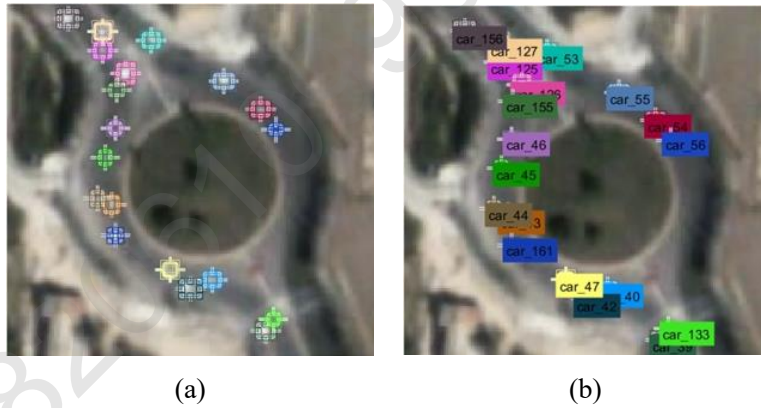


图 14 运动车辆真值标注：(a) 用 bbox 标注车辆目标；(b) 每个运动的车辆目标分配一个唯一的 ID

## 4.2 实验与讨论

定量评价比较实验中各个评价参数的结果见表 4-2，从表 4-2 可以看出，本文提出的算法在各个评价指标中的表现都由优于比较算法 ViBe 和 GMM。具体来说，传统的运动目标检测算法 ViBe 和 GMM 在保持 50%左右的召回率的同时，产生了 90%左右的 FP。低检测精度对 ViBe 和 GMM 的整体性能产生的不利的影响，这是因为 ViBe 和 GMM 都是在试图描述每一个像素点随时间变化的模型，它们只能适用于静止背景，无法适用于缓慢位移的卫星视频。也就是说，ViBe 和 GMM 算法不能区分背景的运动和车辆目标的运动。另外，ViBe 和 GMM 的多目



标跟踪准确率 MOTA 和精度 MOTP 都较低, 说明这两个算法不能很好的适应复杂城区场景中多车辆目标的跟踪。它们 MOTP 指标比本文提出的算法低了 10%, 这说明 ViBe 和 GMM 对真实运动车辆目标的定位精度也远远低于本文提出的算法, 本文提出的算法在该指标上的表现也得益于本文提出的区域生长算法和多形态学特征鉴别算法。本文提出算法的精度指标评价高于比较算法 70%, 这说明本算法对运动车辆目标有很高的识别能力, 可以很好的应对背景运动的卫星视频, 凭借本算的高精度, 所以本算法的 $F_1$ 参数、雅可比相似度指标都远远领先于 ViBe 和 GMM。ViBe 和 GMM 的召回率大概为 50%左右, 本文提出算法的召回率大于 60%。一方面说明本文算法对潜在车辆目标的识别能力高于比较算法, 另一方面也说明微小的车辆目标的检测依然是个巨大问题大调整。因为车辆目标只有连续运动的特征, 没有颜色、形状、纹理等特征, 与道路背景的对比度低, 如图 12 所示, 这些挑战共同造成了车辆目标较低的召回率。本文提出的算法已经具有了很大的实用价值, 在未来的算法开发中必需须进一步提高目标的召回率, 从而突破小目标识别的瓶颈。

表 4-2 本文提出的算法与比较算法检测和跟踪性能定量比较

测试区域	方法	Recall (%)	Precision (%)	$F_1$ -score	雅可比相似度	MOTA	MOTP
1	本文	64.15	81.71	0.72	0.56	0.46	0.50
	ViBe	51.72	15.10	0.23	0.13	-2.45	0.39
	GMM	43.82	12.29	0.19	0.11	-2.75	0.37
2	本文	62.80	82.23	0.71	0.55	0.47	0.52
	ViBe	61.70	9.14	0.16	0.09	-5.56	0.45
	GMM	61.83	7.5	0.13	0.07	-7.08	0.39
3	本文	60.42	77.26	0.68	0.51	0.41	0.56
	ViBe	41.53	6.76	0.12	0.06	-5.35	0.47
	GMM	46.10	6.34	0.11	0.06	-6.41	0.42
平均	本文	63.06	81.04	0.71	0.55	0.46	0.52
	ViBe	52.86	10.74	0.18	0.10	-3.92	0.43
	GMM	49.66	8.79	0.08	0.08	-4.72	0.39

图 15 展示了目前检测领域中最常用的 3 个参数召回率、精度和 $F_1$ -score 随着时间变化的曲线, 曲线显示了在每一帧中本文提出的算法与比较算法的评分。从图 15 可以看出, 本文提出的算法的召回率在绝大部分时刻都领先于 ViBe 和 GMM 算法, ViBe 和 GMM 算法的检测精度在检测跟踪一开始就开始出现剧烈的衰退, 这说明 ViBe 和 GMM 算法本身具有对像素建模和识别运动像素的能力, 但是无法区别背景的运动, 尽管卫星视频背景的运动是亚像素级, 这个缺陷造成了 ViBe 和 GMM 算法检测运动车辆目标的性能随着时间迅速衰退。整体而言,

本文提出的算法的性能随着时间的波动很小，说明本文算法性能非常稳健。

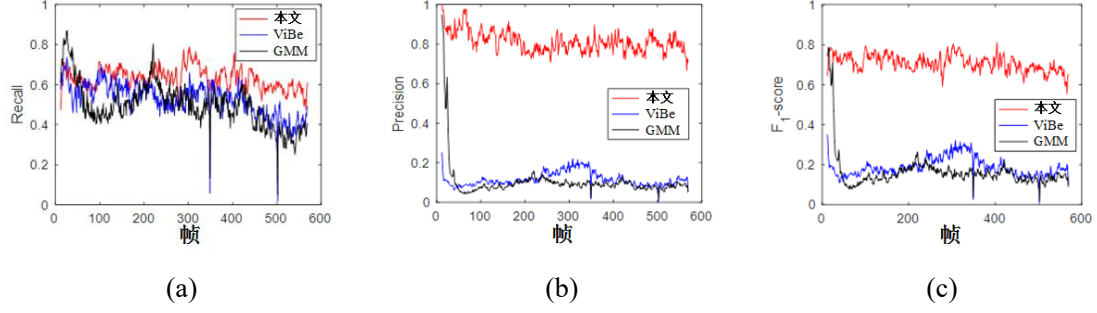
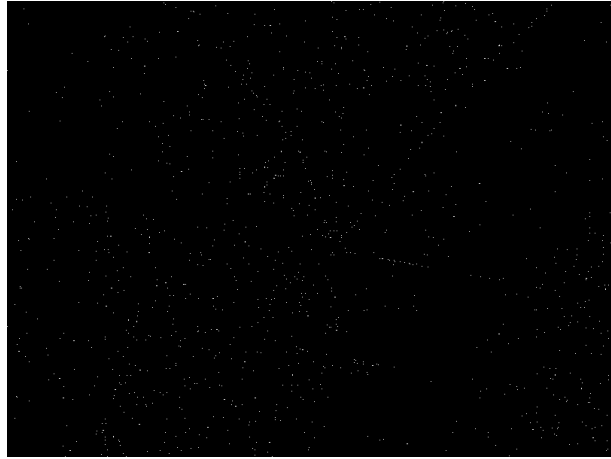


图 15 卫星视频中运动目标检测跟踪系统性能随时间变化曲线：(a) 召回率；(b) 精度；(c)  $F_1 - score$

本文提出的算法和 ViBe、GMM 对卫星视频中运动像素的检测结果如图 16 所示，图中白色像素表示运动前景，黑色像素表示静止背景。对比卫星视频的光流场图 4 和图 16(b)和(c)中 ViBe、GMM 检测到的运动像素可知，ViBe 和 GMM 检测到的运动像素与光流场有一定的相似性，这说明这些算法不能区分背景运动和前景的运动。图 16(a)显示本文提出的算法检测到的运动像素的分布与城市道路分布相似，这说明本文提出的基于本地噪声建模的运动像素检测算法可以有效地检测出微小的运动目标。在这一阶段，检测器的输出还有一些噪声信号，通过形态学鉴别和卡尔曼滤波器的跟踪可以逐渐滤除噪声信号。根据本研究中对车辆运动的定义是在相邻两帧之间进行匀加速直线运动，所以随机出现的噪声、存活时间端的噪声，会随着卡尔曼滤波器的跟踪而有效的滤除。图 17 进一步阐述了不同算法对运动像素检测的结果，本文提出的基于本地噪声建模的运动目标检测器虽然也产生了一些 FP，但是大部分分割出来的运动连通区域还是对于马路上的车辆目标。ViBe 和 GMM 尽管识别出了一些运动的车辆目标，但是产生了大量的 FP，而且集中在一些边缘区域，这是因为背景的运动使得边缘像素变化大，造成这些边缘像素容易被当成是本身运动的像素。这个对比实验重复说明了本文提出的基于局部噪声建模的运动目标检测算法的合理性。



(a)

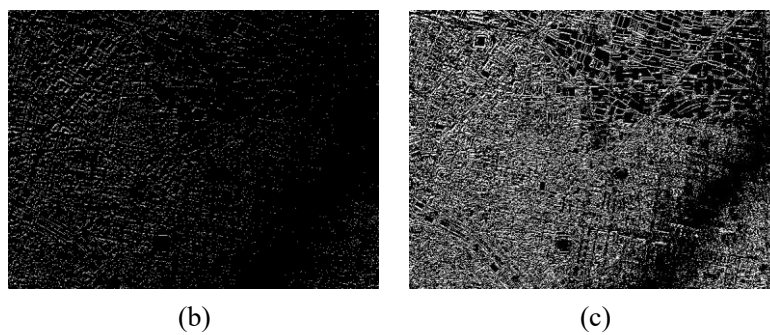


图 16 运动像素检测结果比较: (a) 本文提出的算法; (b) ViBe; (c) GMM

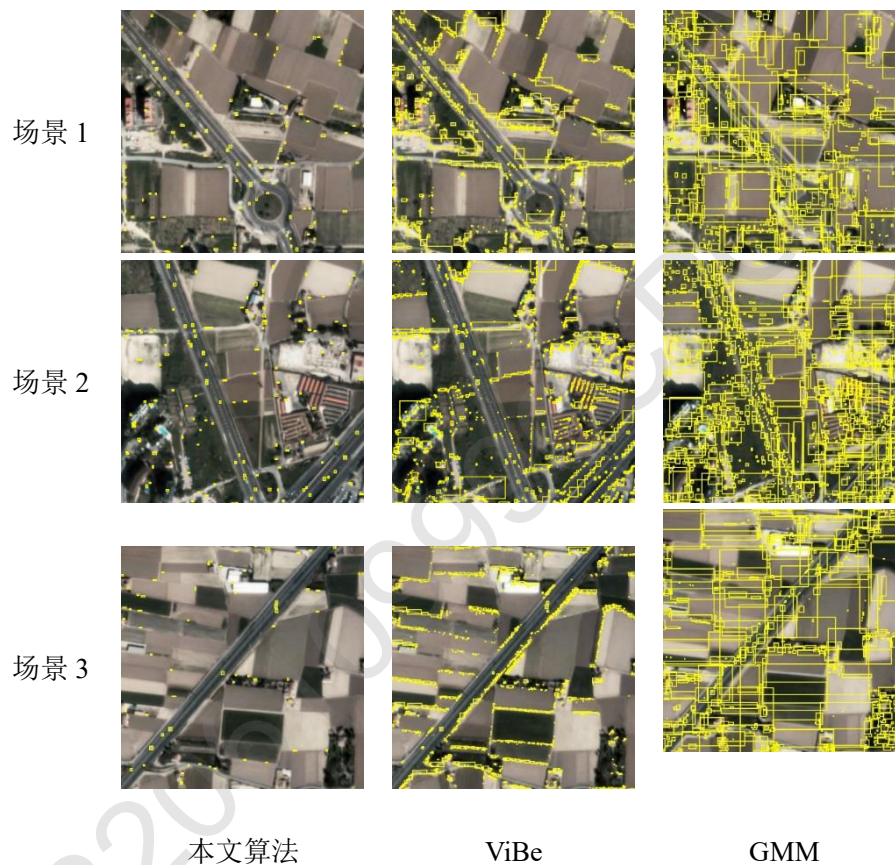


图 17 运动目标检测器运动前景分割结果



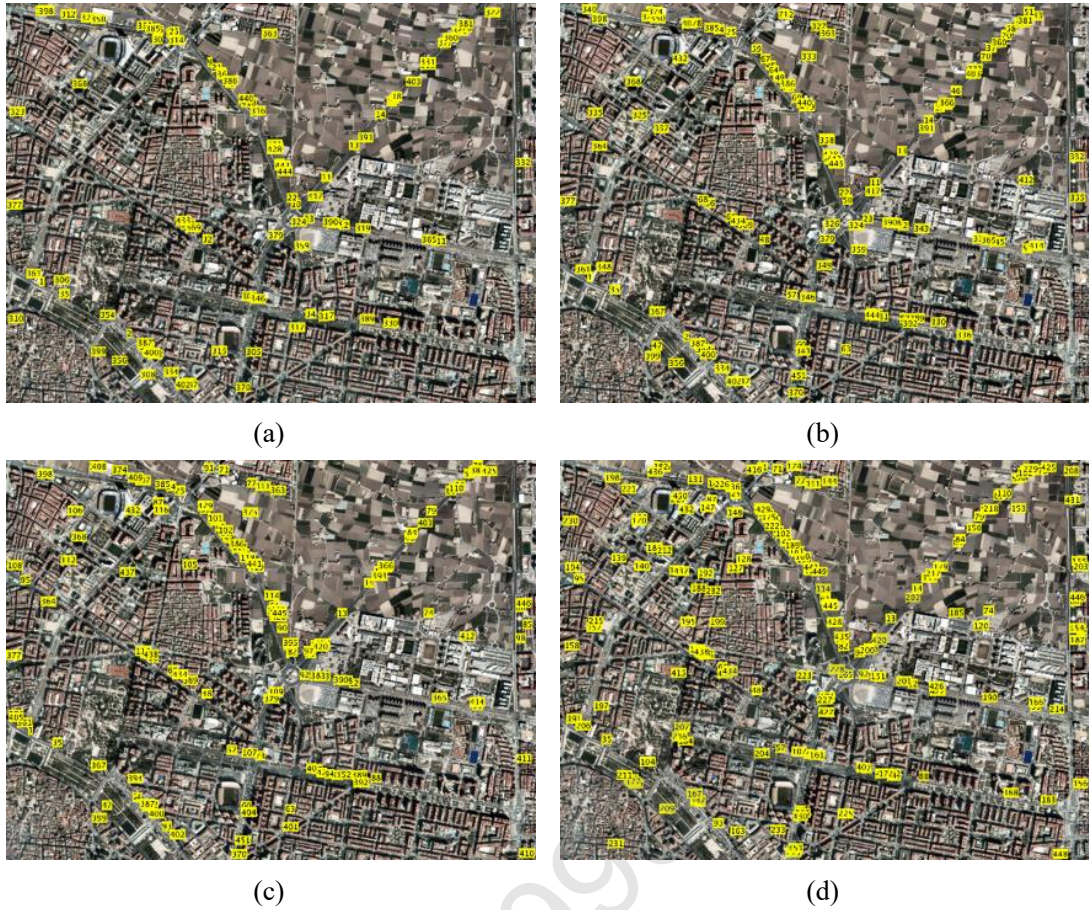


图 18 卫星视频中车辆目标检测结果：(a) 第 50 帧；(b) 第 100 帧；(c) 第 150 帧；(d) 第 200 帧

为了更清晰的展示微小车辆目标的检测与跟踪结果，图 19 展示了图 18 中四个典型道路场景的放大图，包括笔直的马路场景也有和复杂的环岛路。从图 18 可以看出，道路上的车辆目标都被有效地检测到，而且随着跟踪器的跟踪，车辆密集的道路上检测到的车辆目标逐渐增多，这说明跟踪器可以帮助检测器的跟踪。另外，道路 1 中车辆稀少，道路 2 和道路 4 中车辆密集，本文提出的算法都有很好的检测性能，说明该算法受车辆密集度的影响较小。

四个不同道路场景中车辆目标的运动轨迹如图 20 所示，(a)是车辆目标进行直线运动的情况，(b)是车辆目标转弯运动，(c)是车辆目标转弯驶出环岛，(d)是车辆目标转弯驶入环岛。图 20 几乎覆盖了车辆在城区环境中最典型的几种运动场景，这说明不论是直行还是转弯行驶，本文提出的方法都能有效的跟踪车辆目标的运动。该实验也进一步验证了本研究中对短时间内（相邻两帧之间）车辆目标进行匀加速直线运动的假设的合理性。

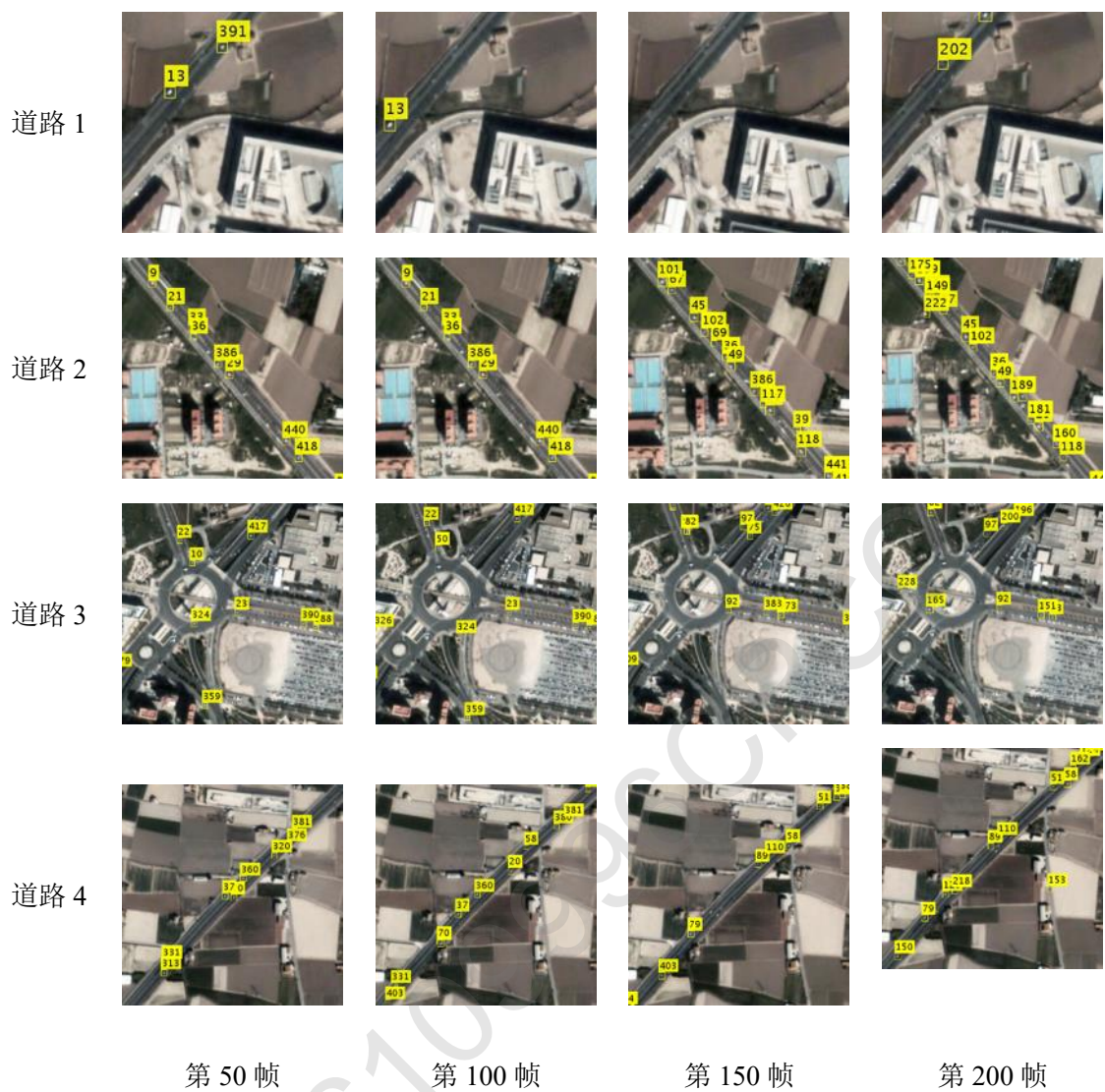


图 19 图 18 中四个典型场景车辆检测效果放大图

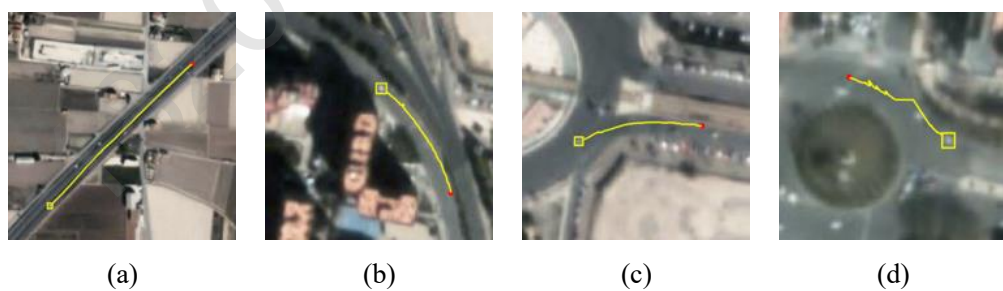


图 20 跟踪器确定的车辆目标的运动轨迹：黄色框是车辆目标的起始位置，红点是车辆目标在视频中最后出现的位置，黄色线段是运动轨迹

## 5 案例小结

卫星视频技术是目前唯一具备观测城市尺度动态变化的监控技术，本案例聚焦于卫星视频技术在城市区域动态目标监测中的难点，提出了面向微小动目标检



测算法,该算法拓展了传统的基于帧差的运动目标检测算法,在时空域上对噪声信号进行建模,从而识别出其中的运动像素。基于区域生长算法来恢复微小车辆目标的完整轮廓,并设计了一套基于卡尔曼滤波器的多目标跟踪系统,引入了基于多种形态学特征的 FP 鉴别算法和基于 SSIM 的最近邻搜索算法,最大程度保证跟踪系统能够“跟得上”运动的微小车辆目标。案例还设计了真值图像标注方法、算法性能评价指标和算法性能评价方法,这对未来处理卫星视频技术的进步提供了客观的评价体系。本研究提出的算法在真实的卫星视频数据上进行了广泛的测试,实验证明了算法的有效性。

## 6 案例思考

卫星视频技术是目前唯一具备观测城市尺度动态变化的监控技术,本案例根据卫星视频特性,对卫星视频中车辆目标的检测与跟踪、卫星视频中多个微小目标的检测与跟踪的算法设计论述了研究示例,针对卫星视频中车辆目标微小、卫星视频覆盖环境复杂多变,背景具有不均匀运动等难点开展了相关研究,本案例对卫星视频的处理做了全流程优化,对微小车辆目标的检测与跟踪框架由多个关键模块组成,首先利用帧差法结合噪声建模技术,有效地检测出卫星视频中的运动目标,作为候选目标。通过区域生长算法扩展检测到的像素区域,得到完整的目标轮廓。利用多种形态学特征,如形状、大小等排除噪声干扰。进一步构建卡尔曼滤波器用于目标跟踪,实现车辆目标位置、速度和加速度的估计。在卡尔曼滤波器的基础上,利用匈牙利算法进行最优匹配,确保了多个目标可以被同时跟踪。最后利用最近邻搜索处理无法匹配的轨迹,防止漏检。最终实现复杂城市环境目标的动态跟踪。

与传统的基于参数建模的运动目标检测算法 GMM 和不基于参数建模的运动目标检测算法 ViBe 相比,本案例提出的微小运动目标检测算法有非常优异的性能,可以很好的适用于卫星视频这种背景进行亚像素级运动的视频。在本案例的数据处理框架下,学生能够进一步融合新技术来提升目标识别的精度,如通过卷积神经网络(CNN)等深度学习技术,可以提高微小目标检测的精度和召回率。例如,使用 Faster R-CNN、YOLO、SSD 等目标检测模型来提升检测性能;采用多模态融合方法,结合光学图像和雷达图像,通过多模态数据融合提高检测和跟踪的准确性;采用时空信息融合:利用时空信息来增强目标检测和跟踪的效果,例如通过时序模型(如 LSTM)来处理视频数据中的时序信息。对于结果召回率提升可以通过在不同尺度上提取特征,可以更好地检测不同大小的目标;采用数据增强技术,增加训练数据的多样性,提高模型的泛化能力。



基于本研究提出的算法，未来的城市管理者可以通过卫星视频确定一个城市区域的交通状况，在整个城市尺度上优化地面交通网络。卫星视频技术未来在军事和民用领域都有巨大的应用前景，在智能交通管理系统中，可以结合车辆检测和轨迹预测，提高交通流量监测和管理的效率；在无人驾驶领域，通过精确的目标检测和跟踪，提高自动驾驶车辆的感知能力，确保行驶安全。在环境监测领域，通过检测和跟踪海洋垃圾、野生动物等目标，提升环境保护和管理的效果。其他还包括战场侦查、地震灾区调查等，本研究提出的面向卫星视频的微小目标跟踪技术具有重要的参考价值。

通过本案例的分析，我们了解了光学卫星视频数据的特点以及针对该类数据的动目标检测算法设计思路，同时学习了常用的检测与跟踪性能评价方法，主要思考的问题如下：

- (1) 目前光学卫星视频数据主要特点是什么？
- (2) 微小动目标检测的难点主要是什么？
- (3) 探讨本案例提出算法在微小动目标检测应用中的优点
- (4) 改进算法能够解决哪些关键问题，还有哪些问题尚待解决？

本案例从应用出发，通过对典型应用实例的分析引起技术存在的难点，从卫星视频在城市环境下车辆目标的识别与跟踪出发，分析目标识别算法存在的难点，给出了全流程的处理策略以及优化方案，实现基于卫星视频的动态目标检测。本案例对于教学方面的启发具有以下特点：

#### (1) 综合讨论技术的理论与实践：

通过对卫星视频数据特点的理论讲解与实际数据处理的结合，学生可以更好地理解卫星视频的时空分辨率、覆盖范围以及光照和天气对视频质量的影响。课程将通过案例研究，展示如何在实际情境中应用这些理论知识来提高动目标检测的准确性和效率。

#### (2) 算法设计与性能分析：

深入分析和讨论案例中提出的各种算法（如基于本地噪声建模的检测算法和基于卡尔曼滤波的跟踪算法），课程将通过编程实验和模拟项目，让学生实践这些算法，并评估它们在不同场景下的性能。这不仅帮助学生掌握算法原理，还能培养他们分析和解决实际问题的能力。

#### (3) 实际应用与案例分析：

结合卫星视频技术在城市交通管理、灾害响应等领域的应用案例，讨论这些技术如何转化为实际的解决方案。通过小组讨论和项目作业，学生可以探索如何将技术应用于解决具体的社会问题，提升其创新思维和应用能力。

#### （4）伦理和法律问题的探讨：

结合课程内容，深入探讨在使用卫星视频技术进行监控时可能遇到的伦理和法律问题，如隐私权的保护和数据的合法使用。通过辩论和案例分析，引导学生思考如何在技术创新和伦理法律之间找到平衡。

#### （5）跨学科的学习体验：

鼓励学生跨学科合作，结合数据科学、软件工程、遥感科技等领域的知识，共同开发和优化卫星视频数据的处理算法。通过团队合作，学生能够从不同角度了解问题，培养解决复杂问题的综合能力。

#### 参考文献

- [1] Moreira A, Prats-Iraola P, Younis M, et al. A tutorial on synthetic aperture radar[J]. IEEE Geoscience and remote sensing magazine, 2013, 1(1): 6-43.
- [2] Arikawa Y, Saruwatari H, Hatooka Y, et al. ALOS-2 launch and early orbit operation result[C]//2014 IEEE Geoscience and Remote Sensing Symposium. IEEE, 2014: 3406-3409.
- [3] Kankaku Y, Suzuki S, Osawa Y. ALOS-2 mission and development status[C]//2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS. IEEE, 2013: 2396-2399.
- [4] Sun J, Yu W, Deng Y. The SAR payload design and performance for the GF-3 mission[J]. Sensors, 2017, 17(10): 2419.
- [5] Séguin G, Ahmed S. RADARSAT constellation, project objectives and status[C]//2009 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2009, 2: II-894-II-897.
- [6] Wang R, Deng Y, Zhang Z, et al. Double-channel bistatic SAR system with spaceborne illuminator for 2-D and 3-D SAR remote sensing[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(8): 4496-4507.
- [7] Wang R, Wang W, Shao Y, et al. First bistatic demonstration of digital beamforming in elevation with TerraSAR-X as an illuminator[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(2): 842-849.
- [8] Xu F, Jin Y Q, Moreira A. A preliminary study on SAR advanced information retrieval and scene reconstruction[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(10): 1443-1447.

- [9] Dudgeon D E, Lacoss R T. An overview of automatic target recognition[J]. 1993, 6(1): 3-10.
- [10] Chen S, Wang H, Xu F, et al. Target classification using the deep convolutional networks for SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8): 4806-4817.
- [11] 王兆成, 李璐, 杜兰, 徐丰. 基于单极化 SAR 图像的舰船目标检测与分类方法[J]. 科技导报, 2017, 35(20): 86-93.
- [12] Yingzhi K, Yongfeng Z, Qiang F. Ship target detection for complicated inner harbor SAR images[J]. 2015.
- [13] Crisp D J. The state-of-the-art in ship detection in synthetic aperture radar imagery[R]. Defence Science And Technology Organisation Salisbury (Australia) Info Sciences Lab, 2004.
- [14] Vachon P W. Ship detection in synthetic aperture radar imagery[J]. Proc. OceanSAR, 2006: 1-5.
- [15] Greidanus H, Kourti N. Findings of the DECLIMS project—Detection and classification of marine traffic from space[J]. Proceedings of the SEASAR, 2006.
- [16] Wang Y, Liu H. A hierarchical ship detection scheme for high-resolution SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50(10): 4173-4184.
- [17] Kang M, Leng X, Lin Z, et al. A modified faster R-CNN based on CFAR algorithm for SAR ship detection[C]//2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP). IEEE, 2017: 1-4.
- [18] Meyer F J, Nicoll J B, Doulgeris A P. Correction and characterization of radio frequency interference signatures in L-band synthetic aperture radar data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(10): 4961-4972.
- [19] Velotto D, Soccorsi M, Lehner S. Azimuth ambiguities removal for ship detection using full polarimetric X-band SAR data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(1): 76-88.
- [20] Migliaccio M, Nunziata F, Montuori A, et al. Single-look complex COSMO-SkyMed SAR data to observe metallic targets at sea[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2012, 5(3): 893-901.

- [21]Lesage F, Gagnon L. Experimenting level set-based snakes for contour segmentation in radar imagery[C]//Visual Information Processing IX. International Society for Optics and Photonics, 2000, 4041: 154-163.
- [22]Otsu N. A threshold selection method from gray-level histograms[J]. IEEE transactions on systems, man, and cybernetics, 1979, 9(1): 62-66.
- [23]Liao P S, Chen T S, Chung P C. A fast algorithm for multilevel thresholding[J]. J. Inf. Sci. Eng., 2001, 17(5): 713-727.
- [24]Liu Y, Zhang M, Xu P, et al. SAR ship detection using sea-land segmentation-based convolutional neural network[C]//2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP). IEEE, 2017: 1-4.
- [25]Nunziata F, Buono A, Migliaccio M, et al. Dual-polarimetric C-and X-band SAR data for coastline extraction[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(11): 4921-4928.
- [26]Ding X, Nunziata F, Li X, et al. Performance analysis and validation of waterline extraction approaches using single-and dual-polarimetric SAR data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(3): 1019-1027.
- [27]Lang H, Wu S, Xu Y. Ship classification in SAR images improved by AIS knowledge transfer[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(3): 439-443.
- [28]Lang H, Wu S. Ship classification in moderate-resolution SAR image by naive geometric features-combined multiple kernel learning[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(10): 1765-1769.
- [29]Lang H, Zhang J, Zhang X, et al. Ship classification in SAR image by joint feature and classifier selection[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(2): 212-216.
- [30]Jiang M, Yang X, Dong Z, et al. Ship classification based on superstructure scattering features in SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(5): 616-620.
- [31]Bentes C, Velotto D, Tings B. Ship classification in terrasars-x images with convolutional neural networks[J]. IEEE Journal of Oceanic Engineering, 2018, 43(1): 258-266.

- [32]Huang L, Liu B, Li B, et al. OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(1): 195-208.
- [33]Xu F, Wang H, Song Q, et al. Intelligent Ship Recongnition from Synthetic Aperture Radar Images[C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018: 4387-4390.
- [34]Ao Wei, Xu F, Qian Yutong, Guo Q. Feature Clustering Based Discrimination of Ship Targets for SAR Images[C]// IET International Radar Conference 2018.
- [35]Li B, Liu B, Guo W, et al. Ship Size Extraction for Sentinel-1 Images Based on Dual-Polarization Fusion and Nonlinear Regression: Push Error Under One Pixel[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(8): 4887-4905.
- [36]Ao W, Xu F, Li Y, et al. Detection and discrimination of ship targets in complex background from spaceborne ALOS-2 SAR images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(2): 536-550.
- [37]Zhou Y, Wang H, Xu F, et al. Polarimetric SAR image classification using deep convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1935-1939.
- [38]Xu F, Jin Y Q, Moreira A. A preliminary study on SAR advanced information retrieval and scene reconstruction[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(10): 1443-1447.
- [39]Walha A, Wali A, Alimi A M. Video stabilization with moving object detecting and tracking for aerial video surveillance[J]. Multimedia Tools and Applications, 2015, 74(17): 6745-6767.
- [40]Jackson B P, Goshtasby A A. Registering aerial video images using the projective constraint[J]. IEEE Transactions on image processing, 2010, 19(3): 795-804.
- [41]Guo H, Liu S, He T, et al. Joint video stitching and stabilization from moving cameras[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5491-5503.
- [42]Molina E, Zhu Z. Persistent aerial video registration and fast multi-view mosaicing[J]. IEEE Transactions on Image Processing, 2014, 23(5): 2184-2192.
- [43]Jiang N, Liu W. Data-driven spatially-adaptive metric adjustment for visual tracking[J]. IEEE Transactions on Image Processing, 2014, 23(4): 1556-1568.

[44]KaewTraKulPong P, Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection[M]//Video-based surveillance systems. Springer, Boston, MA, 2002: 135-144.

[45]Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[C]//Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). IEEE, 1999, 2: 246-252.

[46]Barnich O, Van Droogenbroeck M. ViBe: a powerful random technique to estimate the background in video sequences[C]//2009 IEEE international conference on acoustics, speech and signal processing. IEEE, 2009: 945-948.

18206109996CPCC