

引文格式:张旭,郝向阳,李建胜,等.监控视频中动态目标与地理空间信息的融合与可视化方法[J].测绘学报,2019,48(11):1415-1423. DOI:10.11947/j. AGCS.2019.20180572.  
ZHANG Xu,HAO Xiangyang,LI Jiansheng,et al.Fusion and visualization method of dynamic targets in surveillance video with geospatial information[J]. Acta Geodaetica et Cartographica Sinica, 2019, 48 (11): 1415-1423. DOI: 10. 11947/j. AGCS. 2019.20180572.

## 监控视频中动态目标与地理空间信息的融合与可视化方法

张旭<sup>1</sup>,郝向阳<sup>2</sup>,李建胜<sup>2</sup>,李朋月<sup>2</sup>

1. 航天工程大学,北京 101416; 2. 信息工程大学地理空间信息学院,河南 郑州 450001

### Fusion and visualization method of dynamic targets in surveillance video with geospatial information

ZHANG Xu<sup>1</sup>,HAO Xiangyang<sup>2</sup>,LI Jiansheng<sup>2</sup>,LI Pengyue<sup>2</sup>

1. Space Engineering University, Beijing 101416, China; 2. Institute of Geographical Spatial Information, Information Engineering University, Zhengzhou 450001, China

**Abstract:** The dynamic foreground target analysis of surveillance video is an important basis for security construction in safe cities and smart campuses. The dynamic attributes can be assigned to static geographic data by integrating surveillance video with geospatial data. For the integration of traditional surveillance video and geographic information data, only video data is projected into geospatial, which causes storage issue and difficult in understanding video content. The model that integrates geographic information and foreground dynamic targets, and integrates surveillance video with geographic information is proposed in this paper. And the dynamic foreground target and the tracking trajectory in the image are projected into the geospatial by the derived mapping model. The multi-layer fusion mode is utilized to distribute the video dynamic summary content in geospatial space.

**Key words:** surveillance video; dynamic foreground target; geospatial mapping; geospatial information; information fusion; visualization

**Foundation support:** The National High-tech Research and Development Program of China (863 Program) (No. 2015AA7034057A); The National Natural Science Foundation of China(No. 61173077)

**摘 要:** 监控视频的动态前景目标智能分析是平安城市、智慧园区等安防建设的重要基础,将监控视频与地理空间数据融合可为静态的地理数据赋予动态属性。针对传统监控视频与地理信息数据集成仅仅将视频数据投射至地理空间,造成存储难、视频内容理解难度大等问题,本文提出了前景动态目标与地理空间信息的融合模型,通过推导出的映射模型将图像空间中的动态前景目标及跟踪轨迹映射至地理空间中,达到将监控视频与地理信息有机融合的目的。根据不同的应用需求,本文设计了4种多图层融合显示模式,实现了监控视频中的动态前景目标在地理空间的可视化。

**关键词:** 监控视频; 动态前景目标; 地理空间映射; 地理空间信息; 信息融合; 可视化

**中图分类号:** P208

**文献标识码:** A

**文章编号:** 1001-1595(2019)11-1415-09

**基金项目:** 国家863计划(2015AA7034057A); 国家自然科学基金(61173077)

地理信息的表达内容涉及地理实体及其空间关系、不确定性、地理动态及地理本体等方面<sup>[1]</sup>。地理数据包括空间位置、属性特征及时态特征3个部分,是对于不同的地理实体、地理要素、地理现象、地理事件、地理过程等的表达<sup>[2]</sup>。传统的地理信息表达以二维地图为主导,展现地理空间

的静态属性。为解决地理数据的多维表达,使以地理方式看待世界时更加贴近人的视角,同时附加所不具备的地理属性,将视频与地理信息融合的地理超媒体的表达方式成为研究的新方向<sup>[3-6]</sup>。

目前对于监控视频的智能分析,无论是目标检测<sup>[7]</sup>还是跟踪行为理解<sup>[8]</sup>等计算机视觉任务,

仅仅是基于视频影像本身,其中影像目标的检测与跟踪<sup>[9]</sup>等的精度是研究者们追求的主要目标,所得到的分析结果也仅仅是影像坐标。以监控视频的目标跟踪为例,监控者们更想得到的是实际地理位置和目标的动态方位、速度、运动轨迹等信息,而单纯的视频目标跟踪无法完成该任务,将地理空间信息与视频融合可有效解决这一问题。文献[10]描绘出摄像机的位置以及视图方向将视频影像置入“附近”视图,来进行跨摄像机的跟踪。文献[11]认为空间视频具有巨大的潜力,在GIS中使用基本数据类型来建模空间视频,使用Viewpoint数据结构表示视频帧来进行视频的地理空间分析。文献[12]提出了一个系统,用于从未标定的视频中利用地理空间数据进行相机姿态估计,通过GPS数据、序列影像和建筑物粗糙模型进行建筑物精细三维建模。文献[13]提出了基于视频运动物体和GIS的集成模型,通过空间定位和聚类运动物体的轨迹,构建运动物体的虚拟视野和表达模型,在虚拟场景中逐帧重建运动对象的子图。

传统的监控视频与地理信息的融合模型分为两类,即位置映射模型和视频影像映射模型。前者仅仅将监控相机的位置集成在地理信息的框架中,用统一的地理坐标参考系将处于该区域范围内的监控相机建立相关联系。监控视频与地理信息的集成仅仅处于松集成阶段,起到的作用更多是示意功能<sup>[14-15]</sup>。后者则是在此基础上将视频影像通过相机的内外参数等信息映射至地理空间中,与地理场景叠加进行处理分析<sup>[11,16-17]</sup>。虽然将更多信息映射至地理空间中,但是由于监控视频存在大量的信息冗余,无法使监控人员快速捕获真正感兴趣的目标信息。

## 1 图像空间向地理空间的映射关系模型

监控视频中含有静态环境背景信息和动态前景目标信息。首先对静态背景建模,将背景提取并映射至地理空间,然后建立前景目标和轨迹信息与地理空间的映射关系模型,通过该映射关系模型将其映射至地理空间,从而建立地理信息与前景目标相融合的可视化模型。通过提取监控视频中动态前景目标信息,既可减少视频的冗余信息,也可改善可视化效果。设某空间点 $P$ 的世界坐标为 $(X_w, Y_w, Z_w)$ ,通过旋转矩阵 $\mathbf{R}$ 和平移向量 $\mathbf{t}$ 可以转换为相机坐标系下的坐标 $(X_c, Y_c, Z_c)$ ,

坐标 $(X_c, Y_c, Z_c)$ 与其对应图像坐标 $(u, v)$ 存在如式(1)所示透视投影关系

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/d_x & 0 & u_0 \\ 0 & 1/d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_x & 0 & 0 & 0 \\ 0 & f_y & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{R} | \mathbf{t}] \mathbf{X}_w = \mathbf{P} \mathbf{X}_w \quad (1)$$

式中, $f_x, f_y$ 为相机焦距; $d_x, d_y$ 为相机传感器在水平与垂直方向的像元物理尺寸; $u_0, v_0$ 为图像像素主点坐标; $\mathbf{K}$ 是仅与相机内部结构相关的参数所决定的内参数矩阵; $[\mathbf{R} | \mathbf{t}]$ 是由相机相对世界坐标系的旋转矩阵 $\mathbf{R}$ 和平移向量 $\mathbf{t}$ 所决定的外参数矩阵; $\mathbf{P}$ 为相机投影矩阵。

不妨假设地面为一平面,将世界坐标系中的点映射至图像坐标系中,假设图像中一点 $m$ ,对应世界坐标点为 $\mathbf{M}$ ,则

$$\mathbf{m} = [x \ y \ 1]^T \quad (2)$$

$$\mathbf{M} = [X_w \ Y_w \ 0 \ 1]^T \quad (3)$$

$$\mathbf{m} = \mathbf{H} \mathbf{M} \quad (4)$$

即

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}] \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} = \mathbf{H} \mathbf{M} \quad (5)$$

式中

$$\mathbf{H} = \mathbf{K}[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}] = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & 1 \end{bmatrix} \quad (6)$$

上述所求解的 $\mathbf{H}$ 矩阵是将平面上的物方空间点透视变化至图像空间中的映射矩阵,为了求解图像空间点投射至物方空间中,需要对 $\mathbf{H}$ 矩阵求逆,即

$$\begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} = \mathbf{H}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (7)$$

$$\mathbf{H}^{-1} = (\mathbf{K}[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}])^{-1} \quad (8)$$

当假设世界坐标高程为0时,即将其看作平面时,通过计算相机内参矩阵 $\mathbf{K}$ 与外参矩阵 $[\mathbf{r}_1,$

$r_2, t]$  求解出  $H$  矩阵,内参矩阵的求解是利用张正友标定法<sup>[18]</sup>拍摄 10~20 张标定板图像进行标定,同时可获取相机畸变参数;外参矩阵可通过多点透视问题(perspective-n-points, PNP)进行求解,本文采用的是精度较高也较为流行的 EPNP+Iteration<sup>[19-20]</sup>的方法。

## 2 前景目标及其跟踪轨迹的提取与表达

在固定场景的监控视频中,背景信息并不引起人们关注,前景运动目标才是关注的重点,也是视频智能分析的关键信息,因此运动目标的提取尤为重要。本文采用 MOG2<sup>[21]</sup>算法进行前景目标的提取,并根据差分检测策略筛选出含前景目

标的视频帧,通过跨帧检测显著提高了检测效率,然后将提取出的前景运动目标的轮廓作为地图符号置入地理空间中进行可视化表达。在多目标跟踪任务中,则利用基于深度学习的 YOLOv3 算法进行目标检测并利用 DeepSort<sup>[22]</sup>算法实现多目标跟踪。将视频流输入差分筛选器剔除无需检测帧后,置入 YOLOv3 检测器中,输出检测框、类别与置信度,将该输出再次置于 DeepSort<sup>[22]</sup>多目标跟踪器中,通过改进的递归卡尔曼滤波<sup>[23]</sup>预测位置并跟踪,根据马氏距离与深度描述子的余弦距离作为融合后的度量,采用匈牙利算法<sup>[24]</sup>进行级联匹配,输出动态跟踪定位信息。具体流程见图 1。

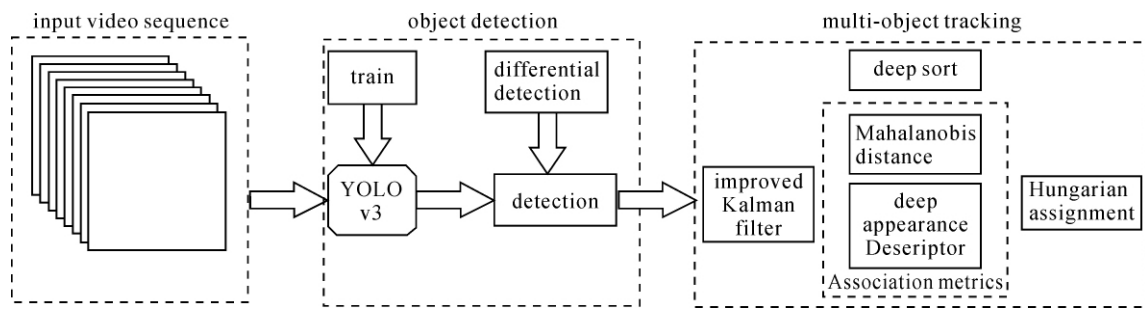


图 1 实时检测跟踪流程

Fig.1 Real-time detection and tracking flowchart

视频所输出的跟踪结果为所跟踪目标在视频流图像中的位置、大小、身份识别信息等组成,由于这种结果无法被人们很直观地感受到,因此在进行目标跟踪时需要同时绘制出目标的运动轨迹。目前通常以目标检测框的中心为轨迹节点,虽然这种表达方式能够显示目标的运动轨迹,但不能满足量测定位的精度要求。为此,本文以目标(以人为例)的双足中心作为轨迹节点的初值,然后根据相机相对地平面的位姿与目标在图像中所占比例大小进行轨迹校正。假设由多目标跟踪器中获得的当前帧  $t_i$  中某一目标  $O_m$  的检测框结果为  $(u, v, \gamma, h)$ , 分别对应检测框的左下点的横纵坐标、宽高比例以及高度,则  $t_i$  帧中目标  $O_m$  在图像中的轨迹节点  $T_j(u', v')$  可由式(9)求得

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} u + w/2 + \delta \\ v + \sigma \end{bmatrix} \quad (9)$$

式中,  $\delta, \sigma$  为校正值。由于由此获得的目标  $O_m$  的轨迹节点  $T_j$  存在误差,因此连接轨迹节点  $T_j$  所得到的轨迹  $Traj_n$  出现抖动现象,需要对所有轨迹节点进行拟合以取得光滑的跟踪轨迹。本文

采用式(10)所示的 3 次多项式进行轨迹拟合

$$y = a_0 + a_1x + \dots + a_kx^k \quad (k=3) \quad (10)$$

各节点到该曲线的偏差平方和为

$$R^2 = \sum_{i=1}^n [y_i - (a_0 + a_1x_i + \dots + a_kx_i^k)]^2 \quad (11)$$

经过求偏导化简后得到式(12)的矩阵表达式

$$\begin{bmatrix} 1 & x_1 & \dots & x_1^k \\ 1 & x_2 & \dots & x_2^k \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_k \end{bmatrix} \quad (12)$$

即

$$\mathbf{X} \times \mathbf{A} = \mathbf{Y} \quad (13)$$

$$\mathbf{A} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (14)$$

结合前文所计算得到的映射矩阵,当获得图像空间中的跟踪目标  $O_m$  的轨迹节点点集

$$T_j (j=1, \dots, j, \dots) = \begin{bmatrix} u'_1 & \dots & u'_j & \dots \\ v'_1 & \dots & v'_j & \dots \\ 1 & \dots & 1 & \dots \end{bmatrix} \quad (15)$$

通过映射矩阵可计算得到物方空间中该目标

$O_m$  的轨迹节点的对应点集,经过上述 3 次多项式拟合后得到的地理空间中平滑轨迹节点集为

$$T_w' = H^{-1} \times T_j \times A = (K[r_1, r_2 | t])^{-1} \cdot \begin{bmatrix} u_1' & \cdots & u_j' & \cdots \\ v_1' & \cdots & v_j' & \cdots \\ 1 & \cdots & 1 & \cdots \end{bmatrix} \times A \quad (16)$$

### 3 前景动态目标与地理信息的融合模式

在安防监控中,以井字格形式的监控视频最为常见,也广泛应用于公安系统、校园、小区等场所,然而这种以原始视频影像序列作为信息源的方式费时费力且未利用空间相关信息。同时,仅仅将监控视频映射至地理空间的模式也无法克服视频数据的冗余性缺点,难以突出视频的主要信息。与传统的监控视频相比,将视频的动态前景目标信息或者是管理者们感兴趣的信息提取出来,将处理分析后的结果,本文根据应用需求的差异性,共提出了 4 种融合模式,分别为:

融合模式 1: 轨迹要素层 + 前景动态目标图层 + 背景层 + 真实地图图层;

融合模式 2: 轨迹要素层 + 前景动态目标图层 + 真实地图图层;

融合模式 3: 轨迹要素层 + 前景动态目标图

层 + 背景层 + 矢量地图图层;

融合模式 4: 轨迹要素层 + 前景动态目标图层 + 矢量地图图层。

图 2 是以融合模式 1 为例的示意图,将多目标跟踪的轨迹信息映射至地理空间,利用前景目标提取算法提取目标与轨迹相关联,以真实的场景作为固定背景信息,实现地理信息与视频影像动态前景目标信息的融合。该模式包含要素最全,背景层的融合可利用视频所提供的背景对真实场景更新,可体现前景动态目标在真实场景下的定位结果以及轨迹的位置分布,提供了更多超媒体信息;模式 2 主要用于以遥感地图为参考底图的动态目标定位跟踪任务,当背景与真实地图场景差别较小时适用该模式,使得可视化效果更加真实;模式 3 与前两者相比将真实地图层更换为矢量地图层,对于相机可视范围外场景不被关注的情况下较为适用,同时使表达更为简洁;模式 4 适用于只关心地理信息表达,忽视场景中超媒体信息的任务,可视化效果也更为直观,突出前景动态目标的定位跟踪结果。同时模式 3 与模式 4 相比于前两者均更重视动态前景目标在地理空间中的数据分析、可量测、可查询统计等目的。

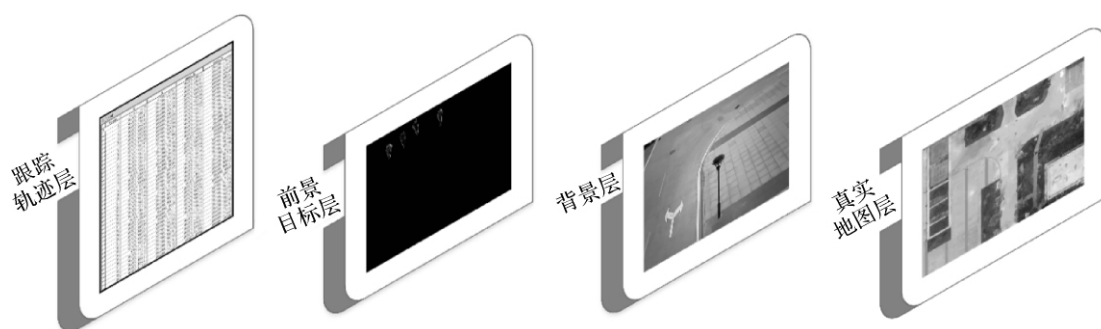


图 2 地理信息与视频动态前景目标信息的融合模式

Fig.2 Fusion mode of geographic information and video dynamic foreground summary information

### 4 试验

为验证所提出的模型及方法的可行性,笔者采集了部分监控视频影像进行试验,所采集视频场景为校园某一区域,所采用真实地图为无人机在 120 m 高度拍摄的遥感影像,像素分辨率为 0.08 m。硬件环境 Intel XEON CPU E5-1607 3.00 GHz, 16 GB 内存, NVIDIA GTX1060 6 GB 显卡。

试验数据是由 USB 外置相机采集,像素为  $640 \times 480$ , 相机内参矩阵为

$$K = \begin{bmatrix} 830.2155 & -2.0105 & 279.7148 \\ 0 & 834.8147 & 270.6643 \\ 0 & 0 & 1 \end{bmatrix}$$

畸变参数为

$$[-0.3855 \quad 0.2610 \quad 0.0037 \quad -0.0037]$$

由于该相机存在桶形畸变,一定程度上影响了映射结果,因此对视频影像首先进行畸变校正,再

将畸变校正后的结果选取对应点进行 PNP 的计算,获取相机外参数矩阵。对应点对如表 1 所示。

表 1 世界坐标与图像坐标对应点对  
Tab.1 Point pairs corresponding to world coordinates and image coordinates

世界坐标	图像坐标
(460 471.188 545,3 853 986.285 574,0)	(411.572 036,52.501 202)
(460 477.726 312,3 853 997.308 430,0)	(64.504 398,6.506 124)
(460 457.092 360,3 853 990.164 838,0)	(295.583 308,419.436 340)
(460 457.107 798,3 853 986.005 468,0)	(607.761 291,401.538 110)
(460 469.792 619,3 853 994.166 355,0)	(125.351 499,86.497 767)

标定出的相机在地理空间中的坐标为 (460 449.504 6,3 853 990.102,7.625 628 456), 3 个旋转角为 $[-178.157^{\circ} \ 55.467 \ 1^{\circ} \ 91.332 \ 6^{\circ}]$ , 外参数矩阵为

$$[r_1, r_2 | t] = \begin{bmatrix} -0.013 \ 37 & -0.998 \ 595 \ 84 & 3 \ 854 \ 649.154 \ 653 \ 25 \\ -0.566 \ 726 \ 39 & 0.049 \ 737 \ 35 & 69 \ 267.896 \ 528 \ 01 \\ 0.823 \ 800 \ 583 & -0.043 \ 403 \ 94 & -449 \ 594.599 \ 082 \ 83 \end{bmatrix}$$
$$H^{-1} = \begin{bmatrix} 100.687 \ 567 \ 97 & 1 \ 605.225 \ 750 \ 79 & 460 \ 476.810 \ 297 \ 99 \\ 842.732 \ 143 \ 58 & 13 \ 435.991 \ 124 \ 41 & 3 \ 853 \ 999.257 \ 770 \ 82 \\ 0.000 \ 218 \ 67 & 0.003 \ 486 \ 25 & 1 \end{bmatrix}$$

图 3 为监控视频背景通过地理映射模型投射前后的对比图,(a)为视频图像空间中的视角, (b)为地理物方空间中的视角,此时视频的背景已具有地理信息,可实现查询、量测等功能。



图 3 监控视频地理映射前后对比  
Fig.3 The comparison chart of surveillance video before and after geographical mapping

图 4 是在试验视频数据中截取的部分帧与所对应的跟踪结果,其中白色框是由递归卡尔曼滤波所得到的预测框,蓝色框是由差分 YOLOv3 算法所计算得出的检测框,绿色的 ID 号为通过匈牙利匹配所确定的跟踪结果。

视频的多目标跟踪量化评价见表 2,分别采用了 YOLOv2 与 YOLOv3 两种检测器作为多目标跟踪的目标检测算法。评估的度量标准为 MOT CHALLENGE 所提供的方法 CLEAR

MOT<sup>[25]</sup>,其中 MOTA 是结合了丢失目标,虚警率,ID 转换误配数,表示了多目标跟踪的准确度,而 MOTP 是所有跟踪目标的平均边框重叠率 IOU,表示了多目标跟踪的精确度。由表 2 可看出,当采用 YOLOv2 时多目标跟踪的 MOTA 与 MOTP 可达 78.4 与 79.8,采用 YOLOv3 时有一定增长,分别为 87.5 与 83.6,可以发现当目标检测的精度提高时,多目标跟踪的精度会随之提高,同时量化指标 MOTA 87.5 以及 MOTP83.5 体现

了本文算法对于所实验视频的适用性。

表 2 多目标跟踪量化评价结果

Tab.2 The quantitative evaluation result of multi-target tracking tasks

Detector	IDF1	Rcll	Prcn	FAR	MT	FP	FN	IDs	FM	MOTA	MOTP	MOTAL
YOLOv2	87.8	78.6	99.8	0.01	1	4	438	0	18	78.4	79.8	78.4
YOLOv3	92.1	89.3	98.1	0.05	4	35	218	3	22	87.5	83.5	87.6

试验中共选取了均匀分布的 18 个点作为图像空间与地理空间的对应点,其中表 1 中的对应点对作为映射模型计算的输入值,剩下的 13 个点对作为测试点对进行映射模型的精度评定,测试

的对应点对坐标及映射后的坐标见表 3,并计算其均方误差,其中最大误差为 0.117 4 m,最小为 0.017 7 m,平均均方差为 0.071 3 m。

表 3 映射坐标对应及误差分析

Tab.3 Mapping coordinate correspondence and error analysis

图像坐标系		世界坐标系(真实对应点)		世界坐标系(映射结果)		均方差/m
<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	
72.518 8	12.512 8	460 475.044 0	3 853 996.605 2	460 475.031 6	3 853 996.540 4	0.066 0
286.536 5	168.400 3	460 464.176 8	3 853 990.189 8	460 464.111 8	3 853 990.171 0	0.067 7
199.485 0	4.482 2	460 475.043 1	3 853 992.598 4	460 475.124 8	3 853 992.640 1	0.091 7
308.700 6	141.516 0	460 465.290 8	3 853 989.621 1	460 465.253 0	3 853 989.687 7	0.076 6
483.542 3	158.458 2	460 464.075 5	3 853 986.389 3	460 464.096 6	3 853 986.305 6	0.086 3
593.348 2	472.508 8	460 456.177 3	3 853 986.711 7	460 456.128 9	3 853 986.724 7	0.050 2
61.495 8	12.498 0	460 475.043 6	3 853 996.810 0	460 475.095 8	3 853 996.893 8	0.098 8
75.528 9	6.498 9	460 475.627 6	3 853 996.568 2	460 475.635 3	3 853 996.563 3	0.009 1
505.965 5	202.045 9	460 462.473 1	3 853 986.187 2	460 462.390 3	3 853 986.269 4	0.116 7
599.978 2	195.934 3	460 462.455 6	3 853 984.586 5	460 462.435 1	3 853 984.580 9	0.021 2
399.498 5	59.499 7	460 469.735 8	3 853 987.245 9	460 469.619 5	3 853 987.229 9	0.117 4
286.485 5	268.494 1	460 460.745 3	3 853 990.271 3	460 460.644 2	3 853 990.236 3	0.107 0
610.626 9	71.466 7	460 468.106 2	3 853 982.455 9	460 468.117 0	3 85 3982.470 0	0.017 7
平均均方差						0.071 3

图 5 中(a)、(b)分别对应了多目标跟踪轨迹在图像空间中与地理物方空间中的可视化表达;图 5(c)、(d)分别是原始跟踪轨迹节点与通过三次多项式拟合后的误差较小的平滑轨迹。

图 6 展示了融合模式中两种不同的底图,分别是(a)中的无人机遥感影像图,该底图可最大程度反映真实的场景,但由于更新的原因,在特殊情况下可与背景层叠加显示;(b)为矢量地图,以其作为底图层,可突出前景动态目标信息,使表达更加简明。红色框为相机可视域范围。

图 7 中(a)、(d)分别对应融合模式中的①到④,4 种融合模式分别适用于不同的任务需求,具有不同的可视化表达效果,在突出重点有差异的情况下,可根据各模式的特点灵活选择。

试验所采集视频时长 22 s,共 670 帧,为验证所提方法的实时性,进行了各部分耗时统计,结果如表 4 所示,其中目标检测部分利用监控视频冗

余特性,通过差分筛选剔除无须检测帧提高检测速度,共耗时 16.96 s,跟踪部分耗时 4.27 s,映射总耗时 0.13 s,其中检测每帧平均耗时 2.5 ms,跟踪每帧平均耗时 0.6 ms,总速度可达 31.36 fps,可知所提出方法在完成需求的同时满足实时性要求,同时当视频前景变化较小时,可改变筛选阈值,进一步提高检测速度以提高实时性。

表 4 实时性分析结果

Tab.4 The real-time analysis results

检测速度/(ms/帧)	跟踪速度/(ms/帧)	总速度/fps
2.5	0.6	31.36

## 5 结 论

本文在多目标跟踪的基础上,提出融合地理信息与动态前景目标的模型,设计了 4 种多图层融合模式将监控视频中动态前景目标与跟踪轨迹

通过地理映射模型投射至地理空间中,与传统视频与地理信息的结合方式相比,减少了视频数据传输中的冗余,极大程度上降低了数据的存储量,智能化提取了视频动态前景目标信息,减轻了监控人员的工作强度。实现了监控视频动态前景目标在真实地理空间中的表达,解决了传统目标跟踪任务仅仅处于图像空间中,无法实现真实地理空间

中可量测、可定位的问题。从试验结果来看,多目标跟踪的准确度 MOTA 可达 87.5,精确度 MOTP 可达 83.5,图像空间向地理空间的映射模型精度为 0.071 3 m,处理速度为 31.36 fps,在精度与实时性上可满足任务需求,监控视频的动态前景目标信息融合至地理空间的可视化效果良好,4 种融合模式也可为不同的需求提供相应映射方案。

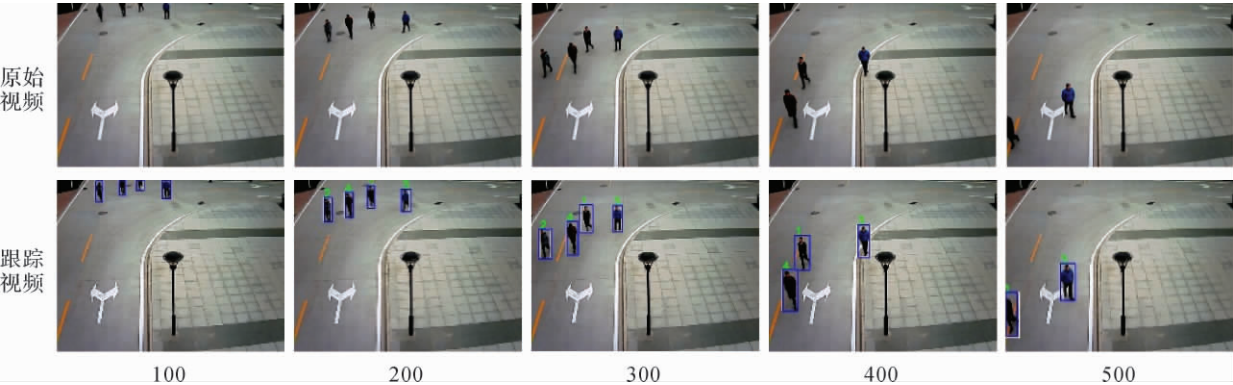


图 4 原始视频与跟踪视频对应帧

Fig.4 The corresponding frame of original video and tracking video

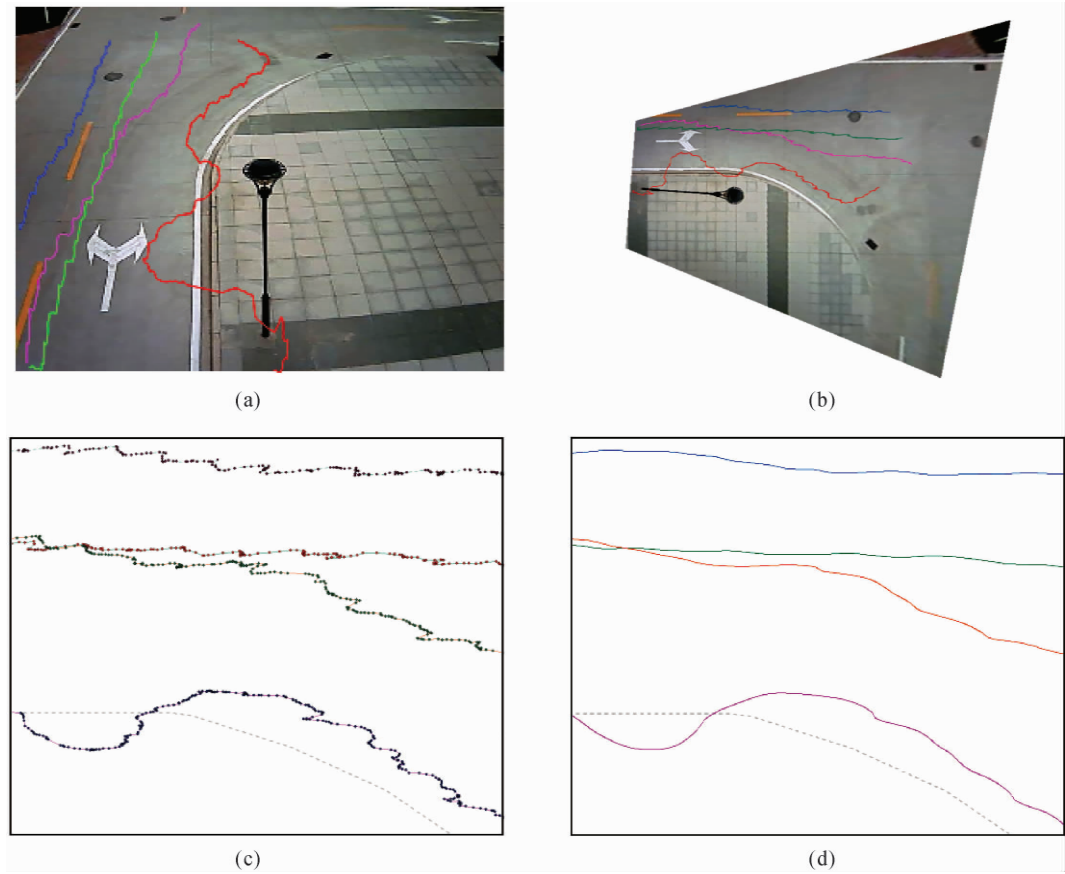


图 5 多目标跟踪轨迹结果

Fig.5 The results of multi-target tracking trajectory





图 6 试验区底图

Fig.6 The base map of experimental area

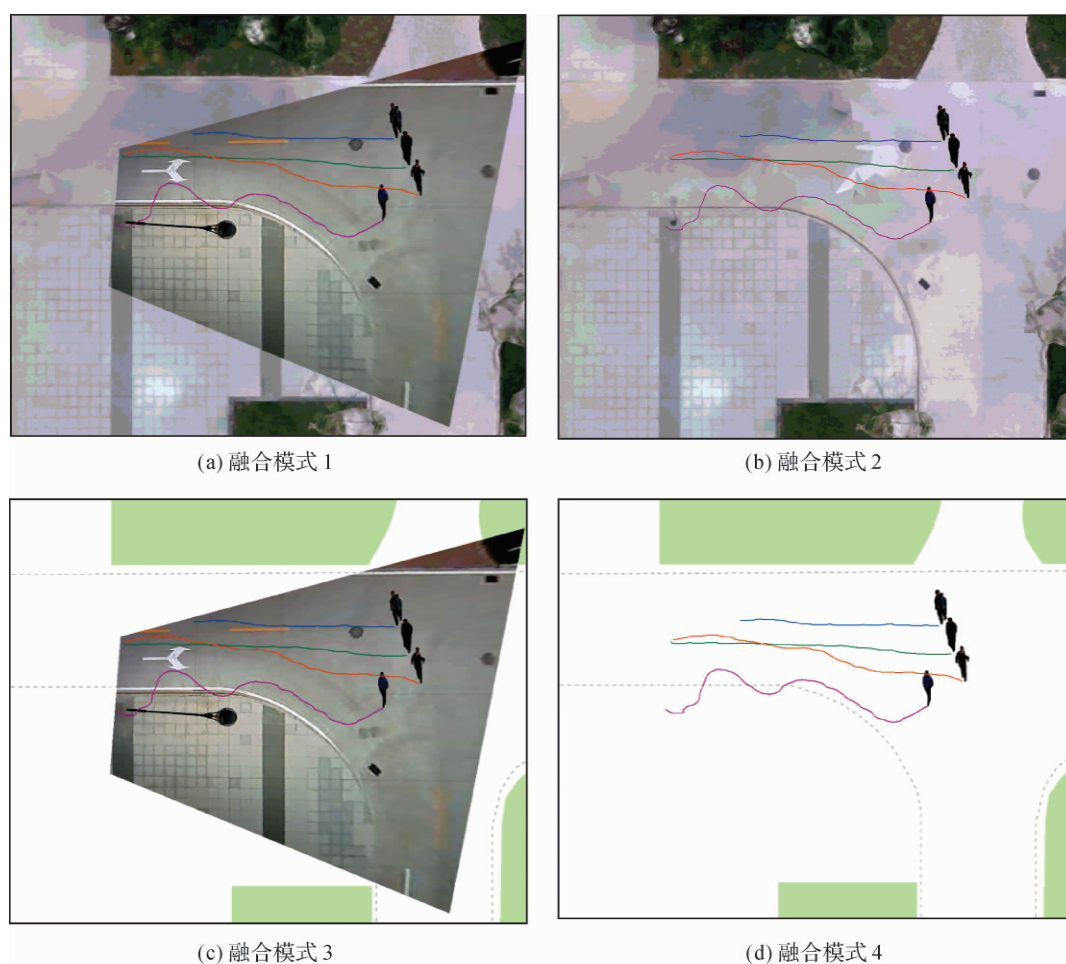


图 7 4 种融合模式对比

Fig.7 The comparison chart of four fusion modes

监控视频多用于以平面为主的场景,因此二维映射也可适用于大多情况,但未来的研究还可以从多方面展开,如利用不同平面高程约束实现三维映射;也可通过标注真值的数据集,来评估跟

踪及映射的精度,不断提高跟踪算法与映射模型的精度;实现多相机的多目标跟踪在统一的地理参考场景下的融合表达。



## 参考文献:

- [1] 韩志刚,孔云峰,秦耀辰.地理表达研究进展[J].地理科学进展,2011,30(2):141-148.  
HAN Zhigang, KONG Yunfeng, QIN Yaochen. Research on geographic representation: a review[J]. Progress in Geography, 2011, 30(2): 141-148.
- [2] YUAN M, MARK D M, EGENHOFER M, et al. Extensions to geographic representations[M]//MCMASER R, USERY L. A Research Agenda for Geographic Information Science. Boca Raton, FL: CRC Press, 2004: 129-156.
- [3] LIPPMAN A. Movie-maps: An application of the optical videodisc to computer graphics[J]. ACM SIGGRAPH Computer Graphics, 1980, 14(3): 32-42.
- [4] BILL R. Multimedia GIS-definition, requirements and applications[M]//The 1994 European GIS Yearbook. Oxford: Blackwell, 1994: 151-154.
- [5] BERRY J K. Capture "where" and "when" on video-based GIS[J]. GeoWorld, 2000, 13(9): 26-27.
- [6] 谢潇,朱庆,张叶廷,等.多层次地理视频语义模型[J].测绘学报,2015,44(5):555-562. DOI:10.11947/j. AGCS.2015.20140176.  
XIE Xiao, ZHU Qing, ZHANG Yeting, et al. Hierarchical semantic model of geovideo[J]. Acta Geodaetica et Cartographica Sinica, 2015, 44(5): 555-562. DOI:10.11947/j. AGCS.2015.20140176.
- [7] 尹宏鹏,陈波,柴毅,等.基于视觉的目标检测与跟踪综述[J].自动化学报,2016,42(10):1466-1489.  
YIN Hongpeng, CHEN Bo, CHAI Yi, et al. Vision-based object detection and tracking: a review[J]. Acta Automatica Sinica, 2016, 42(10): 1466-1489.
- [8] 徐光祐,曹媛媛.动作识别与行为理解综述[J].中国图象图形学报,2009,14(2):189-195.  
XU Guangyou, CAO Yuanyuan. Action recognition and activity understanding: a review[J]. Journal of Image and Graphics, 2009, 14(2): 189-195.
- [9] LUO Wenhao, XING Junjiang, MILAN A, et al. Multiple object tracking: a literature review[Z]. arXiv:1409.7618, 2015.
- [10] GIRGENSOHN A, SHIPMAN F, TURNER T, et al. Effects of presenting geographic context on tracking activity between cameras[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. San Jose, California, USA: ACM, 2007: 1167-1176.
- [11] LEWIS P, FOTHERINGHAM S, WINSTANLEY A. Spatial video and GIS[J]. International Journal of Geographical Information Science, 2011, 25(5): 697-716.
- [12] SOURIMANT G, MORIN L, BOUATOUCH K. GPS, GIS and video registration for building reconstruction[C]//Proceedings of 2007 IEEE International Conference on Image Processing. San Antonio, TX, USA: IEEE, 2007: VI-401-VI-404.
- [13] XIE Yujia, WANG Meizhen, LIU Xuejun, et al. Surveillance video synopsis in GIS[J]. ISPRS International Journal of Geo-Information, 2017, 6(11): 333.
- [14] 宋宏权,孔云峰. Flex 框架下网络视频 GIS 设计与实现[J].测绘科学,2010,35(5):208-210,151.  
SONG Hongquan, KONG Yunfeng. Design and implementation of a web-based VideoGIS using Adobe Flex[J]. Science of Surveying & Mapping, 2010, 35(5): 208-210, 151.
- [15] 韩志刚,曾明,孔云峰.校园地理视频监控 WebGIS 系统设计与实现[J].测绘科学,2012,37(1):195-197.  
HAN Zhigang, ZENG Ming, KONG Yunfeng. Design and implementation of the campus geovideo monitoring Web GIS system[J]. Science of Surveying & Mapping, 2012, 37(1): 195-197.
- [16] CHEN Kuanwen, LEE P J, HUNG Y P. Egocentric view transition for video monitoring in a distributed camera network [C] // Proceedings of International Conference on Advances in Multimedia Modeling. Berlin: Springer-Verlag, 2011: 171-181.
- [17] WANG Yi, BOWMAN D A. Effects of navigation design on contextualized video interfaces[C]//Proceedings of 2011 IEEE Symposium on 3D User Interfaces. Singapore: IEEE, 2011: 27-34.
- [18] ZHANG Z. A flexible new technique for camera calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.
- [19] LEPETIT V, MORENO-NOGUER F, FUA P. EPnP: An accurate O(n) solution to the PnP problem[J]. International Journal of Computer Vision, 2009, 81(2): 155-166.
- [20] PENATE-SANCHEZ A, ANDRADE-CETTO J, MORENO-NOGUER F. Exhaustive linearization for robust camera pose and focal length estimation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(10): 2387-2400.
- [21] ZIVKOVIC Z. Improved adaptive Gaussian mixture model for background subtraction[C]//Proceedings of the 17th International Conference on Pattern Recognition. Cambridge, UK: IEEE, 2004: 28-31.
- [22] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//Proceedings of 2017 IEEE International Conference on Image Processing. Beijing, China: IEEE, 2017: 3645-3649.
- [23] KALMAN R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960, 82(1): 35-45.
- [24] KUHNH W. The Hungarian method for the assignment problem[J]. Naval Research Logistics Quarterly, 1955, 2(1-2): 83-97.
- [25] BERNARDIN K, STIEFELHAGEN R. Evaluating multiple object tracking performance: the clear MOT metrics[J]. EURASIP Journal on Image & Video Processing, 2008(1): 246309.
- [26] 葛宝义,左宪章,胡永江.视觉目标跟踪方法研究综述[J].中国图像图形学报,2008,23(8):1091-1107.  
GE Baoyi, ZUO Xianzhang, HU Yongjiang. Review of visual object tracking technology[J]. Journal of Image and Graphics, 2018, 23(8): 1091-1107.

(责任编辑:张艳玲)

收稿日期:2018-12-10

修回日期:2019-05-18

第一作者简介:张旭(1994—),男,硕士生,研究方向为计算机视觉、导航定位与位置服务。

First author: ZHANG Xu(1994—), male, postgraduate, majors in navigation, positioning and location-based services.

E-mail: xuzhang\_2016@126.com

通信作者:郝向阳

Corresponding author: HAO Xiangyang

E-mail: xiangyanghao2004@163.com