

文章编号: 1002-0446(2001)03-0250-06

基于主运动分析的野外视觉侦察系统 —— 运动目标检测、跟踪及全景图的生成

刘 亚 艾海舟 徐光佑

(清华大学计算机科学与技术系 智能技术与系统国家重点实验室 北京 100084)

摘 要: 本文介绍了一个针对野外自然环境设计的视觉侦察系统,其主要功能包括运动目标的检测、跟踪及环境全景图的生成。摄像机安装在 360°水平旋转可控的车载云台上,视频信号及云台方位数据通过无线通讯传送给基地端的计算机。计算机自动地检测其中的运动目标,实时地拼接出镶嵌有运动目标的全景图,当用户指定某个运动目标后计算机将自动地控制云台跟踪该目标。我们根据二维仿射运动模型和鲁棒参数估计方法分析帧间主运动,对各帧图像进行运动补偿后求得差值图像,聚类分割变化点得到运动目标,再根据目标运动的连续性约束进行可靠的跟踪。为了满足实时处理的要求,程序进行了优化,该系统在 PC 机上,在 384×288 图像分辨率下达到 12.5 帧/秒的处理速度。

关键词: 视觉侦察;全景图生成;主运动估计;目标跟踪

中图分类号: TP24 文献标识码: B

A VISUAL SURVEILLANCE SYSTEM BASED ON PRINCIPAL MOTION ANALYSIS FOR OPEN COUNTRY —— MOTION TARGET DETECTION, TRACKING AND PANORAMA GENERATION

LIU Ya AI Haizhou XU Guang-you

(Computer Science and Technology Department, Tsinghua University State Key Laboratory of Intelligent Technology and Systems)

Abstract In this paper, a visual surveillance system for open country is introduced, which is composed of motion target detection, tracking and panorama generation. A camera is mounted on a 360° controllable platform for a vehicle with its video signal and orientation data transferred to the computer at remote base via wireless communication channels. The computer automatically detects motion targets, and generates panorama with motion target mosaics in realtime. When a target is selected by an user, the computer will control the camera platform to track it. The affine motion model and robust parameter estimation method are used to analyse principal inter-frame motion which will be used for motion compensation before calculating difference picture to locate motion target via clustering. Then target is tracked robustly based on motion smoothness constraint. Programs are optimized which can run on a PC for 384×288 size of processed image with a speed of 12.5 Hz.

Keywords: visual surveillance, panorama generation, principal motion estimation, target tracking

1 引言(Introduction)

视觉侦察或称视觉监控(Visual Surveillance)是计算机视觉研究的一个重要应用领域,随着计算机及摄像机性能成倍提高的同时而成本日益降低,这

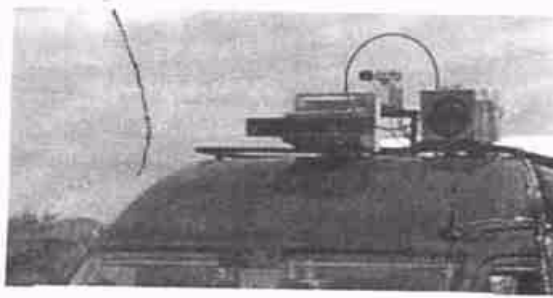
方面的研究逐渐面向实际问题展开,本文针对野外自然环境中的侦察问题,在综合了现有相关的计算机视觉技术的基础上,发展了一种具有实用价值的集成技术,所建立的实验系统具有示范性。

与目前大多数视觉监控系统采用多个摄像机、每个摄像机监控一个相对狭小的区域,通常应用于室内环境的情况^[1]不同,我们的问题是野外自然环境的远距离监控,所需监视的视场很宽,但受通讯问题等限制,相比于多摄像机方案,采用可控云台上的单个摄像机方案更切实可行.为了在使用单个摄像机的条件下,在大视场范围内检测并跟踪运动目标,就需要控制摄像机云台扫描整个视场,此时场景中原本静止的背景及运动物体在摄像机的视频中都是运动的,因而要监测环境中的运动目标,就需要区分静止背景的运动和运动目标的运动.基本的方法是首先依据某种运动模型^[2,3]估计背景运动(主运动),然后将运动目标作为相对于背景运动的出格点检测出来.运动模型的选取依赖于具体的问题,模型简单时,速度快且结果稳定,但适用范围受到限制;模型复杂时,适用于更一般的环境,但计算复杂度高,难以实时,稳定性也难以保证.由于本系统中运动目标和背景距离摄像机都比较远,而摄像机又作近似纯

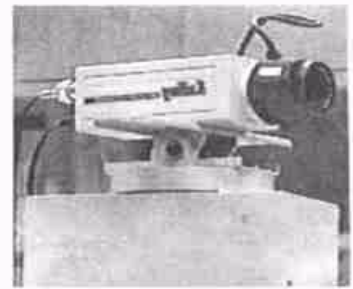
旋转运动,我们采用了 2D 仿射运动模型.考虑到大视野的全景图可以使人们对待监视的环境有身临其境的真实感,我们根据计算机视觉中的拼图技术依据全景图成像模型将摄取的图像序列拼接在一起形成全景图.在全景图上镶嵌上按照主运动补偿后检测出来的运动目标,再由用户指定待跟踪目标后,由计算机自动地控制云台跟踪该目标.

2 系统框架(System architecture)

如图 1 所示,侦察系统的摄像机固定于 360 度旋转可控的云台上,云台安装在侦察车的车顶前部,通过双向无线数据通讯与基地端的指挥中心相连.摄像机通过车顶的全向微波天线向基地发回现场的视讯信号^[4].指挥中心接收侦察车发来的视频和云台方位数据信号,经由基地中心计算机处理,向车载计算机发送控制命令控制云台转动.整个系统的数据通讯及处理流程如图 2 所示.



(a) 车载摄像机云台



(b) 摄像机云

图 1 旋转云台及摄像机

3 帧间主运动分析(Principal inter-frame motion analysis)

假设摄像机看到的是静止景物,图像的运动全部是由摄像机云台运动引起的.以摄像机为中心建立参考坐标系.设景物中某点在运动前和后的摄像机坐标系中的位置分别为 $(x, y, z)^T, (x', y', z')^T$,对应的图像坐标分别为 $(u, v)^T, (u', v')^T$.摄像机在 3 维空间的运动有 6 个自由度,即 $T = (T_x, T_y, T_z)^T$ 和 $(\alpha, \beta, \gamma)^T$,其中 T 为摄像机沿三坐标轴方向的平移量, α 是旋转角, β 是俯仰角, γ 是偏移角.对视频图像序列的连续两帧而言,它们之间的 α, β, γ 通常比较小($\leq 5^\circ$),因此可以忽略旋转矩阵 R 中二次以上的 Taylor 展开项,得到 R 的一阶近似

$$R \approx \begin{bmatrix} 1 & \alpha & -\gamma \\ -\alpha & 1 & \beta \\ \gamma & -\beta & 1 \end{bmatrix} \quad (1)$$

又记 $T' = (T'_x, T'_y, T'_z)^T = R T$,根据两个参考坐标系的变换关系及针孔成像模型:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} T'_x \\ T'_y \\ T'_z \end{bmatrix} \quad (2)$$

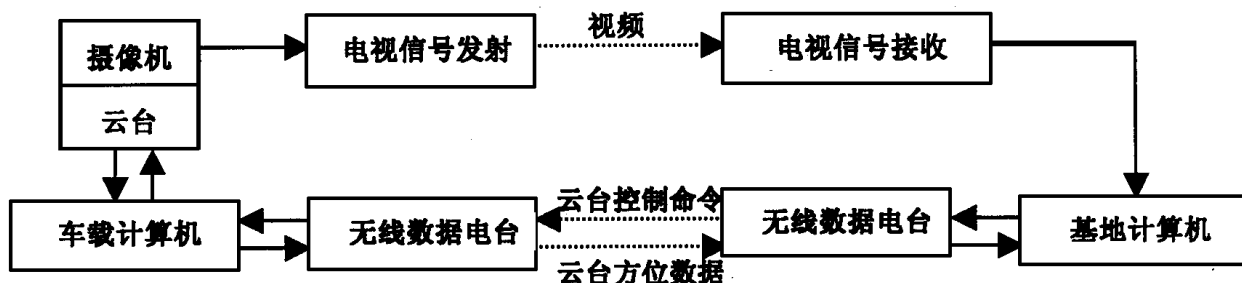
以 f 和 f' 分别表示运动前后摄像机的焦距,则由针孔成像公式

$$\begin{cases} u = fx/z \\ v = fy/z \end{cases} \quad \text{和} \quad \begin{cases} u' = f'x'/z' \\ v' = f'y'/z' \end{cases} \quad (3)$$

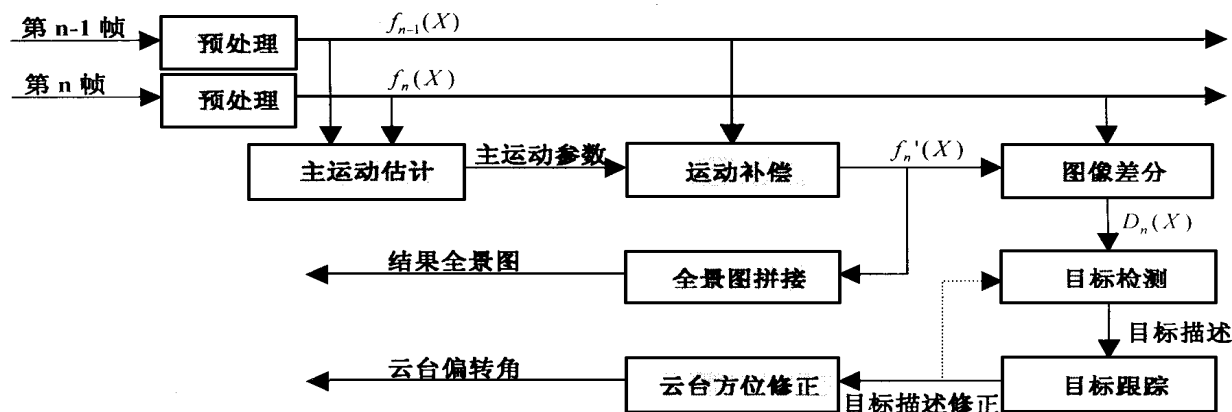
得到两帧图像对应点之间的关系:

$$\begin{cases} u' = f' \frac{u + \alpha v - \gamma f - fT'_x/z}{\gamma u - \beta v + f - fT'_z/z} \\ v' = f' \frac{-\alpha u + v - \beta f - fT'_y/z}{\gamma u - \beta v + f - fT'_z/z} \end{cases} \quad (4)$$

由(4)式可知,当摄像机没有任何平移运动时,图像的变化仅由摄像机的旋转量决定,特别是在旋转量较小的情况下可以看作是图像整体的平移和旋转.



(a) 系统数据通讯



(b) 系统数据处理

图 2 系统数据通讯及处理流程

在我们的应用环境中,摄像机基本在水平面内作纯旋转,旋转轴近似垂直于摄像机光轴和图像平面水平扫描线.由于该旋转轴并不一定准确地经过摄像机光心,所以这种旋转引起的帧间运动主要是绕摄像机垂直轴的旋转(γ)和水平方向的微小平移.因此(4)式可简化为

$$\begin{cases} u' = f' \frac{u + \alpha v - \gamma f - fT'_x/z}{\gamma u - \beta v + f} \\ v' = f' \frac{-\alpha u + v + \beta f}{\gamma u - \beta v + f} \end{cases} \quad (5)$$

令 $s = (\gamma u - \beta v + f)/f'$,并考虑到 β, γ 很小,而 u, v 也比焦距小一个量级左右(整帧图像的视角只有 5 度左右),可以认为 $s \approx f/f'$,对帧内各点近似为常数.这样式(5)就可写成如下形式

$$\begin{cases} s \cdot u = u + \alpha v - T_u \\ s \cdot v = -\alpha u + v - T_v \end{cases} \quad (6)$$

即背景运动(主运动)可以由 4 个独立的运动参数 s, α, T_u, T_v ,即比例、旋转、水平平移和垂直平移近似确定.这就是我们用于主运动分析的 2 维仿射运动模型.

在进行主运动估计前,为了减小野外自然条件下环境光照变化等因素对帧间主运动估计的不利影响,我们首先对图像序列的每一帧进行了直方图规范化预处理.

为估计主运动模型参数,必须求得两帧图像中属于主运动的对应点列.由于这一阶段还难以判断特征点从属于主运动或非主运动,所以我们避开特征对应而使用基于图像块匹配的办法获取整幅图像内的光流.在匹配时以两图像块的绝对差作为衡量相似性的标准.为了提高匹配速度,我们采用了多分辨率的计算框架^[5,6],将计算复杂度从 $O(n^2)$ 减少到

$O(k \log(n))$.

在根据光流估计主运动模型参数时, 考虑到光流估计本身有误差, 而且其中一小部分对应于非主运动的光流, 我们采用鲁棒参数估计算法^[7,8]. 常用的最小二乘法鲁棒性较差, 而鲁棒性很高的 LMS (Least Median of Square) 估计在多于 1 个参数后其计算复杂度很高. 综合考虑性能与速度的要求, 我们选择 M-估计 (Generalized Maximum Likelihood Estimation). 在算法实现时, 采用迭代的重新加权最小二乘法 (Iterated Re-weighted Least Square), 由最小二乘法给出模型参数的初始估计, 然后不断根据估计残差调整权值进行优化^[9].

4 运动目标检测 (Motion target detection)

设 $f_{n-1}(X)$, $f_n(X)$ 为相邻的两帧图像, $f'_n(X)$ 是 $f_{n-1}(X)$ 经主运动补偿后的结果, 则残差图 $D_n(X)$ 可表示为

$$D_n(X) = |f'_n(X) - f_n(X)| \quad (7)$$

由于匹配结果不可避免地带有误差, 所以残差图中存在很多噪声, 而且运动区域也是由比较密集的离散点组成, 因此必须进行预处理才能用于下一步的区域统计. 对残差图的预处理包括形态学开运算和平滑处理. 为加快处理速度, 预处理是在对残差图作阈值化处理后进行的, 处理过的残差图用于区域分割. 由于残差大的区域都有可能是运动区域, 可以先对残差大的点进行区域统计、聚类, 形成候选区域, 再使用一些启发式知识进行判断, 准确地分割出运动物体. 可用的启发式知识包括运动物体的大小和形状、运动物体之间的间隔、运动物体数目和所在位置等.

5 运动目标跟踪 (Motion target tracking)

运动目标跟踪的核心问题是在前后各帧的检测结果间建立对应关系, 对于指定的目标就是确定其运动模型及其轨迹的过程. 常见的运动模型包括一阶运动模型 (仅考虑速度因素)、二阶运动模型 (考虑了速度和加速度), 以及更为复杂的高阶模型如 Kalman 滤波器^[6]等. 出于实时处理的考虑, 我们在系统中使用了二阶运动模型.

算法实现时考虑到存在目标误检和漏检的情形, 在跟踪过程中维护一个运动目标缓冲区 (3-5 帧), 只有连续多次在根据运动模型预估的位置附近检测出的目标才认为是可靠的检测, 而偶尔出现的目标被当作误检测排除. 类似地, 当某目标连续消失

数帧以上时才认为它确实从视场中消失, 否则则认为这是漏检而加以填补.

摄像机云台的控制需要根据检测出的目标方位来确定. 由于摄像机本身的运动不利于对运动目标的准确检测, 所以我们采用有阈值的负反馈机制来控制云台的转动方位, 仅当运动目标偏移图像中心比较多时才控制云台转向目标方位, 以此减少云台的转动. 设当前图像帧上运动目标到图像中心的水平偏移是 dx , 超出了负反馈的控制阈值, 则云台反方向旋转的角度为

$$\alpha = \arctan\left(-\frac{dx}{f}\right) \quad (8)$$

其中 f 是摄像机的焦距, 其数值可以根据摄像机视角范围来估算, 假设摄像机水平视角为 θ , 图像帧宽度为 w (pixel), 则

$$f = \frac{w/2}{\tan(\theta/2)} \quad (9)$$

6 镶嵌运动目标的全景图拼接 (Panorama generation with motion target mosaics)

图像拼接实际上就是根据一定的全景图成像模型将各帧图像按照相应的运动参数进行变换后形成全景图的过程. 通常全景图拼接多采用柱面成像模型, 对应于摄像机光轴在水平面内绕垂直轴作纯旋转运动, 它具有操作简便和符合人的观看习惯等优点. 考虑到我们的实际情况, 摄像机光轴近似在与水平面成一很小的角度的锥面上绕垂直轴作纯旋转运动图 3, 各帧图像大致落在一圆锥面上, 因此我们采用锥面模型生成全景图.

相邻帧图像经变换后会有相当部分的重叠, 重叠部分的数据应由相关各帧根据一定的融合准则共同确定. 在我们的应用中, 由于摄像机沿某一确定方向旋转, 景深又非常大, 因此我们对每帧图像只选取它们变换后的一个竖直窄条镶嵌到最后的锥面图上. 这种方法在拼接时不需要保存图像序列, 而且各帧的运动补偿也可以限制在竖直窄条内进行, 大大节省了处理时间.

上面所述的拼接过程没有考虑存在运动目标的情况, 仅对静止背景有效. 若存在运动目标, 这种拼接方法可能会造成一运动目标出现多次或根本不出现的问题. 为解决这个问题, 我们采用了对背景和前景分层表示的全景图技术: 将根据主运动分析方法检测出的运动目标从各帧中去除后得到的静止背景图, 使用窄条拼接方法生成静止的背景全景图, 而将

各帧检测所得到的运动目标单独存储为前景图, 然后根据用户浏览全景图的时刻来将对应帧的运动目

标镶嵌到背景全景图上.

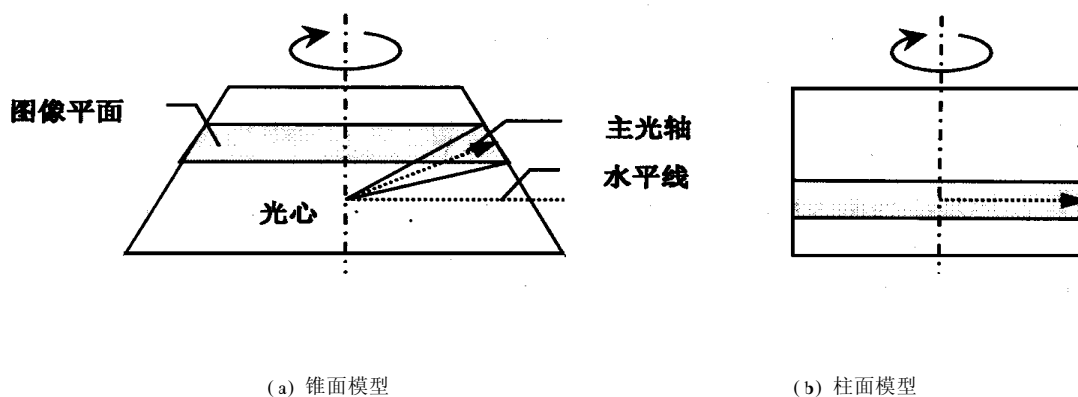


图 3 锥面模型与柱面模型示意图

一般情况下摄像机的主光轴不会与旋转轴完全垂直, 因此得到的全景图具有明显的锥面效果. 若摄像机转动超过一周, 同一个静止物体会在全景图两个不同位置出现, 可以利用它两次成像的位置和角度求出展开锥面的圆心、展开角和内外半径, 据此将锥面全景图校正为柱面全景图^[10].

7 实验结果及分析 (Experiment results and analysis)

我们的实验是在野外车辆试验场中进行的, 所用的基地端计算机为高档微机 (Pentium III 600 CPU, 256M 内存, Matrox Meteor II 图像采集卡), 软件在 MS Visual Studio 6.0 平台上开发. 摄取的彩色图像的分辨率为 384×288 , 为了提高速度在主运动估计和运动目标检测及跟踪阶段按灰度图像来处理; 摄像机云台的平稳转动速度约为 30 度/秒.

图 4 是在运动目标跟踪过程中某一帧的目标检

测结果, (a) 为摄像机摄取的原始图像, (b) 是经主运动补偿、差分和二值化、去噪等处理后的二值图像. 图中用矩形框标记出运动目标所在区域, 这里运动目标 (汽车) 在图像中大小为 88×45 (pixel).

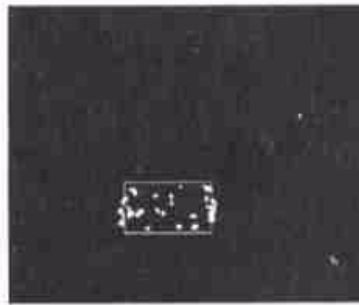
图 5 是对该汽车进行主动跟踪过程中的若干帧图像. 从图中可以看出, 跟踪过程中汽车曾受到树木遮挡 (f), 但由于使用了运动目标缓冲区, 有效地避免了跟踪目标丢失的情况.

图 6 是在试验场实时拼接所得全景图的局部 (由于整幅全景图太大 (4500×500), 为了突出其中的运动目标, 这里只显示了整幅全景图的一部分). 图中检测到的运动目标 (移动的人, 以矩形框标记) 用前文所描述的算法镶嵌到了静止的背景上.

由于系统对实时性能要求很高, 我们在各个处理环节都进行了代码优化. 系统的具体时间性能见表 1:



(a) 原始图像



(b) 经处理的差分图像

图 4 运动目标检测结果

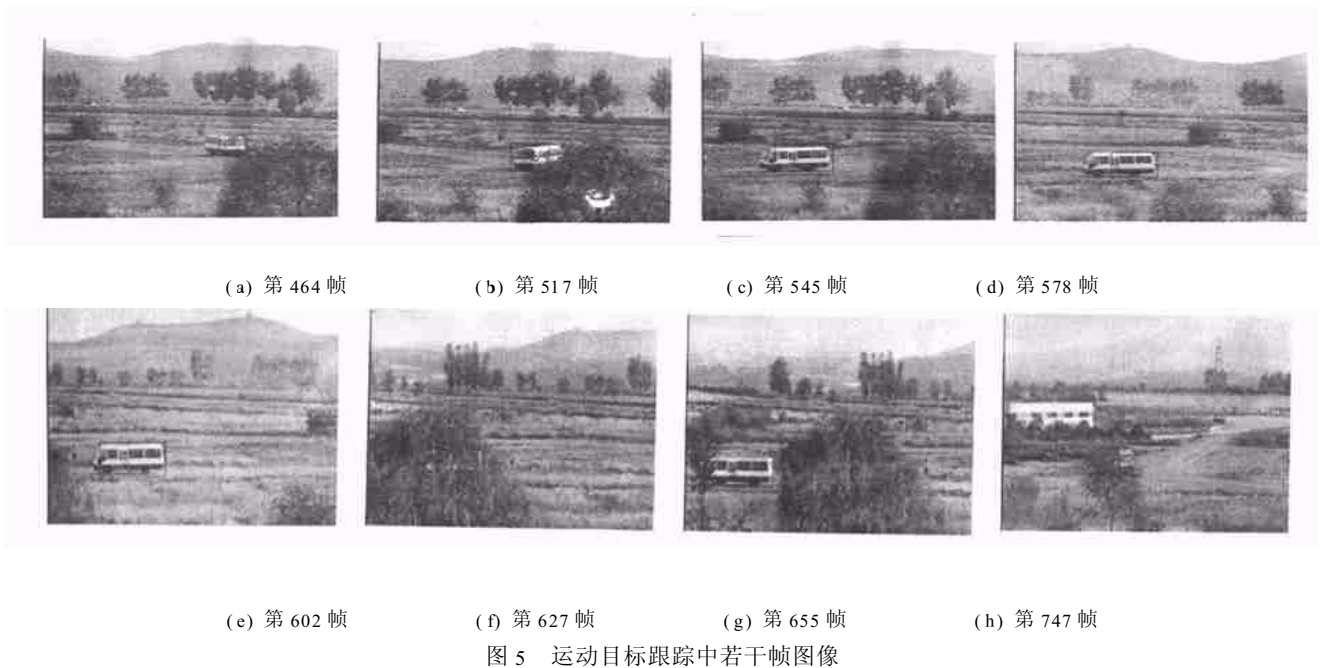


表 1 时间性能 (Table 1 Time performance)

帧间主运动估计	15 m s
运动补偿	8 m s
运动目标检测	17 m s
运动目标跟踪	5 m s
全景图拼接(一帧)	5 m s
系统处理速度	12.5 fps

需要说明的是,在摄像机本身运动的情况下跟踪目标比固定摄像机的情况困难得多,由于主运动估计存在误差,所以目标检测结果也存在偏差,但这种偏差比较小,对于目的在于使运动目标可见的跟踪来讲不会造成明显的影响。

8 结论(Conclusion)

本文介绍了一个野外自然环境中的视觉侦察系统,具有运动目标检测、跟踪和全景图生成等功能.我们采用 2D 仿射模型和鲁棒参数估计方法求解摄像机主运动参数,聚类残差图中的出格点得到运动目标区域,根据目标运动模型利用水平旋转可控的摄像机云台实现对目标的主动跟踪.用户可以在计算机实时地拼接出的镶嵌有运动目标的全景图上指定某个运动目标,让计算机自动地控制云台跟踪该目标.实验表明该系统能够在野外环境下比较可靠地工作,我们认为将该技术与适当的人机交互相结合是解决实际视觉侦察任务的有效途径。

(下转第 205 页)