

针对运动摄像机的快速低存储开销运动目标检测算法

吴明军^{1,2}, 许建铮¹, 周 楨¹, 张亚涛¹

- (1. 中国空空导弹研究院凯迈测控有限公司, 河南 洛阳 471009;
2. 中国空空导弹研究院, 河南 洛阳 471009)

摘 要: 摄像机的运动会导致整幅图像的运动, 使得此情形下的目标检测极具挑战性。针对该问题提出一种快速低存储开销检测算法。首先, 利用一种快速低存储开销配准方法计算相邻两帧的单应变换矩阵。而后, 使用单应变换矩阵进行相邻两帧之间的配准, 并由帧间差分获取帧间运动信息。最后, 采用积累运动信息的方式构造不断更新的运动图像, 通过对此运动图像进行阈值分割分离出最终的运动目标。在多个不同视频序列下的实验表明该算法能够有效地从嘈杂的场景中检测出运动目标。此外, 与先前算法相比, 该算法检测性能更好, 且显著地降低了存储开销与计算时间开销。对于 480×360 的序列而言, 该算法需要的存储开销仅为 825 kByte, 且运算速度达到 16 帧/m。

关键词: 运动摄像机; 目标检测; 图像配准; 快速算法; 低存储开销

中图分类号: TP391 **文献标志码:** A **文章编号:** 1007-2276(2013)08-2275-06

Fast and memory-saving algorithm for moving object detection from a moving camera

Wu Mingjun^{1,2}, Xu Jianzheng¹, Zhou Zhen¹, Zhang Yatao¹

- (1. Measurement and Control Co. Ltd. China Airborne Missile Academy, Luoyang 471009, China;
2. China Airborne Missile Academy, Luoyang 471009, China)

Abstract: It is challenging to detect moving objects from a moving camera as a motion field in the entire image can be induced by the camera motion. A fast and memory-saving detection method is proposed to resolve this problem. First, a fast and memory-saving registration scheme is used to estimate the homography transform between two neighboring frames. Then, neighboring frames is registered with the estimated transform, and frame-to-frame difference is performed to capture the motion cue. Finally, the motion cues are aggregated to construct a constantly updated motion image. After thresholding the motion image, separation of moving objects from the background is achieved. The effectiveness of the proposed method in detecting moving objects from cluttered scenes is validated via experiments on several different video sequences. In addition, this method performs better than previous techniques, while using a fraction of the computation time and a fraction of the memory as well. Specifically, with a memory usage of 825 kByte only, this method runs at 16 frames per second for a sequence with an image resolution of 480×360.

Key words: moving camera; object detection; image registration; fast algorithm; memory-saving

收稿日期: 2012-12-13; 修订日期: 2013-01-15

基金项目: 中国空空导弹研究院青年创新基金(CQKJ00)

作者简介: 吴明军(1984-), 男, 工程师, 博士, 主要从事目标检测、跟踪与模式识别方面的研究。Email: wumingjun@gmail.com

0 引言

作为计算机视觉的一项基本任务,运动目标的检测在多个领域有着广泛的应用(如视频监控、目标跟踪等),其目的是为了实现在背景与运动目标的分离。针对静止摄像机情形下的运动目标检测问题,已存在大量的有效算法^[1-3],如混合高斯背景建模^[4]。但对于摄像机自由运动的情形而言,背景像素位置保持不变的假设不再成立,从而导致先前的背景建模技术^[1-4]不再适用。为了弥补这一缺憾,相关研究人员针对如何将背景建模从静止摄像机推广到运动摄像机的情形作了很多工作^[5-10]。参考文献[5]通过将图像进行 Homography 矫正,然后为每个像素点匹配对应点并建立多高斯模型,从而得到全景图的背景模型。参考文献[6-7]以一幅图像为参考帧,并将大量的图像相对于该帧进行背景补偿,在此基础上为该帧的各像素建立起统计模型。为了使该算法能够快速运行以满足应用需求,参考文献[5]给出了该算法的 GPU 实现。参考文献[9]针对低帧率图像序列,通过在摄像机运动补偿后的每个像素位置采用中值滤波来获得全帧的背景图像,而后使用背景相减进行运动检测,并利用背景图像的梯度信息对配准误差导致的虚警进行抑制。以上算法在原作者给出的实验中都取得了良好的效果,但是针对运动摄像机下的目标检测问题,背景建模仍旧面临着一些困难,体现为:(1)背景建模往往需要大量的训练图像,而当摄像机运动时,每个像素在视场中的停留时间通常有限,于是导致训练样本的不足;(2)由于获取背景模型需要精确的像素对应,这对图像配准的精度提出了非常高的要求,如果其中某两帧之间存在较大配准误差,那么将影响整个时间窗口内的检测结果;(3)大量的训练图像使得所需的存储开销非常高,此外,建模时需要为整个时间窗口内的图像相对参考帧进行配准,其所需的大量运算导致算法运行速度缓慢。与背景建模技术不同,参考文献[11]通过补偿摄像机运动得到帧间差分图像,如同摄像机静止下的帧差检测一样,此差分图像检测到的目标区域支离破碎,于是,该文献接着利用复杂的滤波技术对差分图像进行分析。该方法只能对目标进行定位,而不能分离出目标区域。

针对摄像机自由运动下的运动目标检测问题,文中提出一种快速低存储开销算法,同时确保优异的检测性能。首先,以文中作者先前的工作--针对图像序列的快速低存储开销配准算法^[12]为基础获取稳定的特征点对,并与归一化直接线性变换^[13]相结合计算出视频序列中相邻两帧的单应变换矩阵。然后,利用单应变换矩阵进行相邻两帧之间的配准,由此得到帧间运动信息。最后,为了更加完整地检测出目标,采用积累运动信息的方式获得随着时间不断更新的运动图像并分离出最终的运动目标。将该算法与先前的检测算法进行对比,可以看出该算法不仅能够有效地检测出运动目标,还显著地降低了存储开销与计算时间开销。

1 图像配准

由于文中处理的对象是视频序列,相邻帧间运动往往较为平缓,因此采用针对图像序列的快速低存储开销配准算法^[12]获取相邻两帧图像 I_{t-1} 与 I_t 之间的稳定特征点对 $X_i \leftrightarrow X_i'$, $i=1, \dots, m$, 其中,在齐次坐标系下, $X_i=(x_i, y_i, 1)^T$ 表示位于图像 I_{t-1} 中 (x_i, y_i) 位置上的特征点, $X_i'=(x_i', y_i', 1)^T$ 表示位于图像 I_t 中 (x_i', y_i') 位置上的特征点,而 X_i 与 X_i' 相对应、构成一组特征点对,表示稳定特征点对的数量。如图 1(a)所示,稳定特征点对由直线连接而成。有了这些稳定特征点对,接下来,将通过归一化直接线性变换^[10]计算得到可用于两帧图像配准的单应变换矩阵。

在以上稳定特征点对的基础上,实现两帧图像的配准也就等同于寻找一个 3×3 的单应变换矩阵

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} = \begin{pmatrix} h_1^T \\ h_2^T \\ h_3^T \end{pmatrix} \quad \text{使得公式得以成立:}$$

$$X_i' = HX_i, i=1, \dots, m \quad (1)$$

由公式(1)可以看出,对任意一组特征点对 $X_i \leftrightarrow X_i'$, 都可以整理得到以下方程:

$$\begin{pmatrix} 0 & -X_i^T & y_i' X_i^T \\ X_i^T & 0 & -x_i' X_i^T \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = A_i h = 0 \quad (2)$$

式中: A_i 为 2×9 的系数矩阵, h 是由 H 的 9 个参数形成的 9×1 的向量。虽然 H 有 9 个参数,但在齐次坐标系下只有 8 个需要求解的未知参数。这样,利用所有

稳定的特征点对 $X_i \leftrightarrow X_i', i=1, \dots, m$ 可以按照公式(2)所示的形式构成一个 $2m \times 9$ 的系数矩阵 A , 并且对 h 的求解也即为:

$$\min \|Ah\| \text{ s.t. } \|h\| = 1 \quad (3)$$

公式(3)表示在满足约束条件 $\|h\| = 1$ 的同时求取使得 $\|Ah\|$ 最小化的参数 h , 其可进一步等价成:

$$\min h^T A^T A h \text{ s.t. } \|h\| = 1 \quad (4)$$

由此, $A^T A$ 的最小特征值所对应的特征向量便是所求解, 也就求得所需的单应变换矩阵 H 。

以上求解单应变换矩阵的过程即为直接线性变换, 但是由于数值计算问题容易导致一个非常不精确的解。为了减少数值误差, 需要预先进行一个归一化的过程。也即, 对所有稳定特征点对 $X_i \leftrightarrow X_i', i=1, \dots, m$, 利用相似性变换 T_1 得到 $\hat{X}_i = T_1 X_i, i=1, \dots, m$, 使得这组新的点集质心处于原点且它们到原点的平均距离为 $\sqrt{2}$ 。于是, $T_1 = \begin{pmatrix} s & 0 & t_x \\ 0 & s & t_y \\ 0 & 0 & 1 \end{pmatrix}$, 其中, $s = \frac{\sqrt{2} m}{\sum_{i=1}^m \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}$ ($\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i, \bar{y} = \frac{1}{m} \sum_{i=1}^m y_i$), $t_x = -s\bar{x}$ 以及 $t_y = -s\bar{y}$ 。类似地, 对 $X_i', i=1, \dots, m$ 进行相似性变换 T_2 使得 $\hat{X}_i' = T_2 X_i', i=1, \dots, m$ 。针对新的特征点对 $\hat{X}_i \leftrightarrow \hat{X}_i', i=1, \dots, m$, 利用直接线性变换可计算得到相应的单应变换矩阵 \hat{H} 。因为存在如下恒等变换公式:

$$X_i' = H X_i = T_2^{-1} \hat{X}_i = T_2^{-1} \hat{H} \hat{X}_i = T_2^{-1} \hat{H} T_1 X_i \quad (5)$$

所以, 可以得出对应于所有稳定的特征点对 $X_i \leftrightarrow X_i', i=1, \dots, m$ 的单应变换矩阵 $H = T_2^{-1} \hat{H} T_1$ 。以上即为归一化直接线性变换过程, 并由其计算得到最终可用于两帧图像配准的单应变换矩阵。

2 运动目标检测

对于复杂背景下的目标检测而言, 运动信息是一种有效的实现目标与背景分离的特征, 因为它能够及时捕捉到场景的变化。运动信息可通过配准以后相邻两帧图像之间的差分来快速获取:

$$D_t(x, y) = \begin{cases} 1, & \text{if } |I_t(x, y) - H_{t-1,t}^{-1} I_{t-1}(x, y)| > T_{\text{diff}} \\ 0, & \text{else} \end{cases} \quad (6)$$

式中: D_t 为指示运动区域的二值图像; $H_{t-1,t}$ 为一节阐述的思路计算得到的用于配准 I_{t-1} 与 I_t 的单应

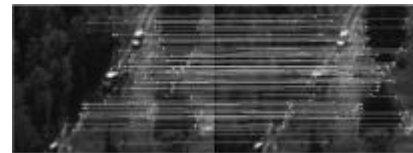
变换矩阵; T_{diff} 为差分阈值。可是, 仅仅利用帧间差分通常难以给出满意的检测结果, 这在图1(b)中也得到了验证, 由于目标灰度均匀、纹理较少且相对摄像机运动速度较慢, 导致检测到的运动区域支离破碎。

为了克服上述困难, 文中提出采用积累运动信息的方式获得随着时间不断更新的运动图像, 其积累过程如下:

$$E_t(x, y) = \alpha H_{t-1,t} E_{t-1}(x, y) + (1 - \alpha) \tau \delta(D_t(x, y) - 1) \quad (7)$$

式中: τ 为赋予当前存在运动的像素的最大值; α 为衰减系数; E_t 为积累到第 t 帧的运动图像; $H_{t-1,t} E_{t-1}$ 表示通过 $H_{t-1,t}$ 将积累到第 $t-1$ 帧的运动图像 E_{t-1} 变换到第 t 帧的坐标系下, 而 $\delta(D_t(x, y) - 1)$ 为指示函数, 当 $D_t(x, y) = 1$ 时, 该函数值为 1, 否则为 0。由于公式(7)中引入了衰减项, 因此 E_t 可对以前的运动信息进行逐步的积累, 并且离当前帧越近的运动反映在 E_t 上的像素亮度越大。如图 1(c)所示, 图像亮度的变化合理地反映出目标的运动方向, 且对应于真实目标运动的区域亮度较大, 从而有效地弥补了因帧间差分获取的运动信息支离破碎导致检测不完整的缺陷。

通过观察运动图像, 目标区域与背景区域的亮度被明显地分离。于是, 对运动图像 E_t 采用阈值 T_{sep} 分割, 得到目标的二值检测结果(见图 1(d))。



(a) 稳定特征点对

(a) Reliable feature point pairs



(b) 帧间差分

(c) 运动图像

(d) 二值检测结果

(b) Frame-to-frame difference

(c) Motion image

(d) Binary detection result

图 1 文中算法概览

Fig.1 Overview of proposed method

3 实验与分析

3.1 实验比较

为了验证文中提出的算法的有效性, 在此采用

两个地物场景下的视频序列进行运动目标的检测实验。在实验中,参数设置为: $T_{diff}=25$, $\tau=10$, $\alpha=0.7$, $T_{sep}=\tau/3$,其中,进行参数设置时主要考虑如下因素:(1) T_{diff} 是差分阈值,其影响运动信息检测的敏感程度,其值越高,运动信息检测的越完整,但因图像噪声带来的错误运动信息也越多,对于 256 级的灰度图像而言,设置为 25 可以较好地保证运动信息检测的完整并同时抑制噪声带来的错误检测;(2) τ 为赋予当前存在运动的像素的最大值,其值的设置只会影响到运动图像分割阈值的选择,于是在实验中设为 10;(3) α 为衰减系数,其值越大,某帧检测到

的运动信息对运动图像影响的时间跨度越长,其值越小,则运动图像对最新检测到的运动信息依赖程度越大, α 设为 0.7 可较好地平衡先前检测到的运动信息与最新的运动信息对运动图像的影响;(4) T_{sep} 为运动图像分割阈值,其值依赖于 τ ,设置为 $\tau/3$ 可以保证有效地分离出目标并同时减少虚假的目标检测。

图 2 给出的是针对序列一的比较实验。序列一为 480×360 的丛林场景下的视频序列,该序列给运动目标检测带来的困难体现在:图像的对比较低;车辆在丛林中穿行,树木的遮挡频繁发生,以及存在阴影和光照的明显变化。图 2(a1~a12)显示了文中算

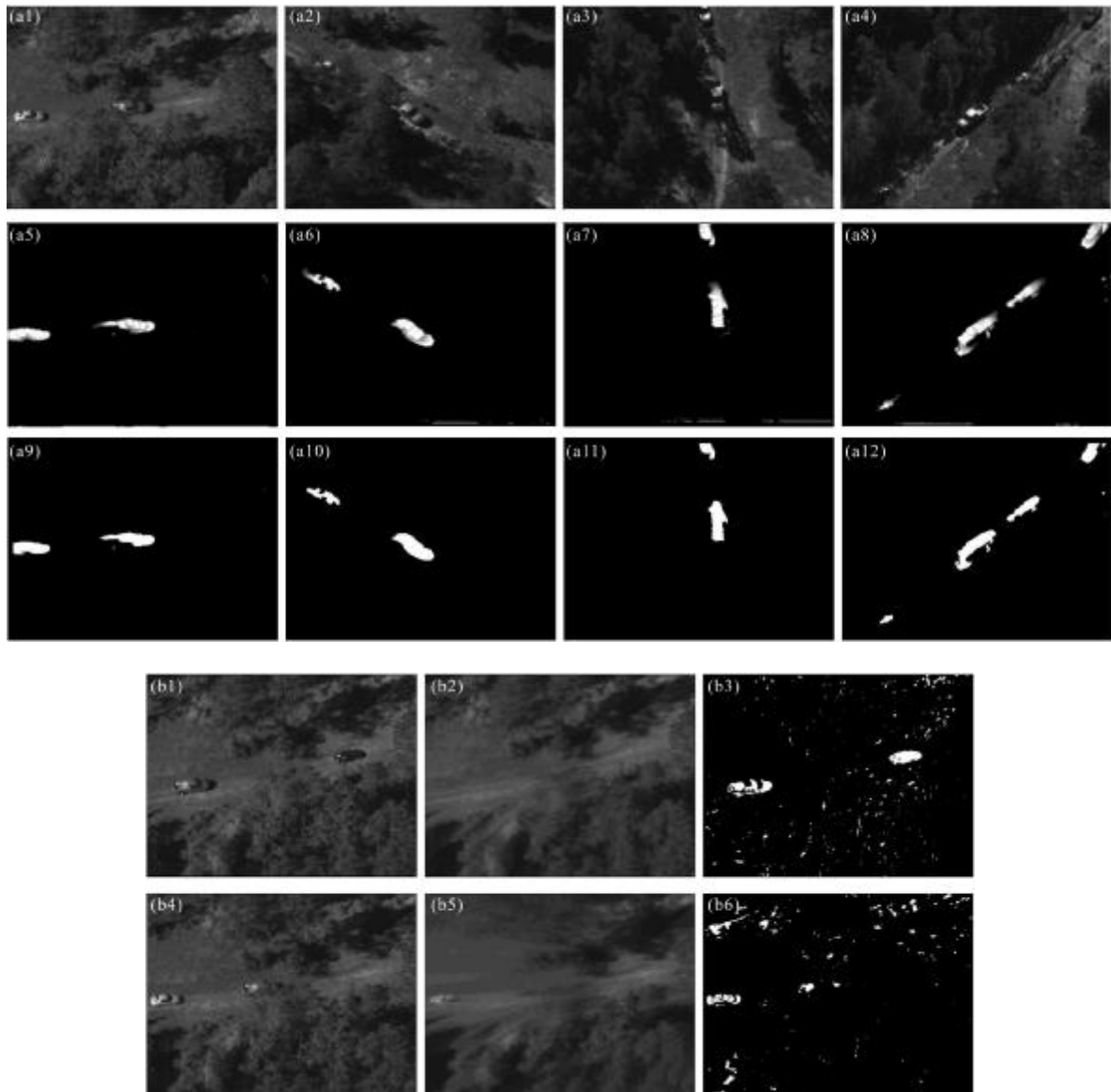


图 2 针对序列一的检测实验

Fig.2 Experiments on the first sequence

法的检测结果。图(a1~a4)给出了原始图像。图(a5~a8)为运动图像,与原始图像对比可看出,运动图像中的亮度大的区域与原始图像中运动目标实际所在的区域是一致的,如在第 160 帧中,其中一辆车的尾部被树木遮挡,这一事实也在运动图像中被合理的反映出来,同时,其他区域的亮度低表明目标与背景得到很好的分离。图(a9~a12)给出的是对运动图像进行阈值分割得到的二值检测结果。为了便于比较,实验中所有的二值检测结果都没有进行形态学滤波等后续处理。图 2(b1~b6)给了参考文献[8]中的背景建模算法得到的检测结果,背景建模算法的运行过程大致为:首先,将整个滑动时间窗口(设置为 90 帧)内的所有帧配准到参考帧;然后,沿着时间方向、采用中值滤

波为参考帧的各像素构建背景像素值,如图 2(b2~b5)中的即为背景图像;最后,将参考帧与其背景图像相减并阈值化,得到二值的检测结果,如图 2(b2~b6)所示。由图 2 可以发现背景建模算法得到的检测结果存在较多的噪声区域,原因在于摄像机的运动导致训练样本不足、以及某两帧的配准误差会影响整个时间窗口内的检测结果,与之相比,文中检测算法得到的检测结果则完整、平滑。

图 3 给出的是针对序列二的检测实验。序列二为 480×360 的道路场景下的视频序列,与上一个序列相比,该序列背景较简单,难点为运动目标较小。图(a)~(d)为原始图像。相应地,图(e)~(h)为由文中算法积累得到的运动图像,图(i)~(l)是文中算法的二值

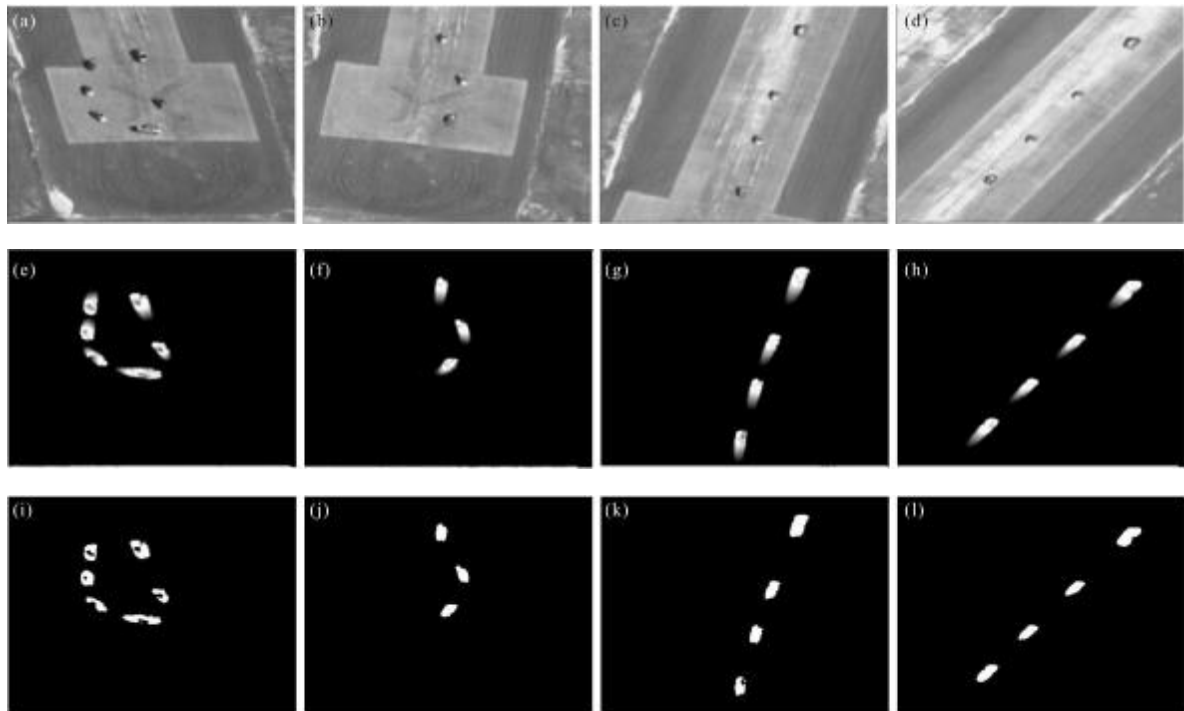


图 3 针对序列二的检测实验

Fig.3 Experiments on the second sequence

检测结果。尽管目标较小,但检测结果仍旧准确。

3.2 存储开销与计算时间分析

假设视频序列中图像的分辨率为 $M \times N$, 那么,对于文中算法而言,其主要存储开销在于:(1)快速图像配准需要约 $2MN + 1/2 MN$ Byte 的存储空间;(2)在积累运动信息的过程中,需要 Byte 用于存储运动图像,另外还需要 $2 MN$ Byte 用于运算过程中缓存中间

结果。由此,文中算法主要存储开销约为 $5.5 MN$ Byte。而参考文献[8]中的背景建模算法所需存储开销主要由缓存在滑动时间窗口内的图像帧数来决定,但是,要构建出较为满意的背景图像,滑动时间窗口至少应该设置为 90 帧。于是背景建模算法主要存储开销不小于 $90 MN$ Byte。以 480×360 的视频序列为例,文中算法的存储开销约为 $825 k$ Byte,其远低于背景建模算法

所需的 13 500 kByte。

在计算时间方面,文中算法主要开销在于 1 次图像配准,2 次图像坐标变换,1 次图像求差,以及 1 次图像求和;而背景建模算法主要开销在于 1 次图像配准,90 次图像坐标变换,此外,在构建背景图像的过程中,需要 MN 次中值滤波,以及背景相减过程中需要一次图像求差。由此可见,文中算法的运行速度远快于背景建模算法。以针对序列一的检测实验为例,在通用 PC 机(双核 2.0 G CPU,1 G RAM)上实现的结果表明,若需完成一帧的检测过程,文中算法仅需 60 ms,而背景建模算法则需要约 10 s。

4 结 论

针对摄像机运动情形,文中提出一种快速低存储开销的运动目标检测算法。该算法以快速低存储开销配准方法为基础来获取帧间运动信息,而后采用积累运动信息的方式得到运动图像,实现目标与背景的明显分离。将该算法应用到多个视频序列中,实验结果表明其能够完整、平滑地从复杂的地物场景中检测出运动目标。此外,与先前检测算法相比,文中算法不仅能够提供更好的检测性能,而且能够有效降低存储开销和计算时间开销。

参考文献:

- [1] Elhabian S Y, El-Sayed K M, Ahmed S H. Moving object detection in spatial domain using background removal techniques-state-of-art [J]. *Recent Patents on Computer Science*, 2008, 1(1): 32-54.
- [2] Lu Zhimao, Liu Minghua, Liu Chen. Method for moving object detection based on construction of feature frames [J]. *Infrared and Laser Engineering*, 2012, 41(7): 1959-1963. (in Chinese)
- [3] Zhang Haojun, Chang Yong, Li Fanming, et al. Real-time embedded implementation of foreground/background segmentation algorithm [J]. *Infrared and Laser Engineering*, 2012, 41(2): 523-530. (in Chinese)
- [4] Stauffer C, Grimson W E L. Learning patterns of activity using real-time tracking [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 747-757.
- [5] Rao N, Di H, Xu G. Joint correspondence and background modeling based on tree dynamic programming [C]//*Proceedings of International Conference on Pattern Recognition*, 2006: 425-428.
- [6] Kang J, Cohen I, Medioni G. Continuous tracking within and across camera streams [C]//*IEEE Conference on Computer Vision and Pattern Recognition*, 2003: 267-272.
- [7] Yuan C, Medioni G, Kang J, Cohen I. Detecting motion regions in presence of strong parallax from a moving camera by multi-view geometric constraints [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2007, 29(9): 1627-1641.
- [8] Yu Q, Medioni G. A GPU-based implementation of motion detection from a moving platform [C]//*IEEE Workshop on Computer Vision on GPU*, 2008: 1-6.
- [9] Reilly V, Idrees H, Shah M. Detection and tracking of large number of targets in wide area surveillance [C]//*European Conference on Computer Vision*, 2010.
- [10] Jin Ting, Zhou Fugen, Bai Xiangzhi. Moving object detection in airborne video using kernel density estimation [J]. *Infrared and Laser Engineering*, 2011, 40(1): 153-158. (in Chinese)
- [11] Yao F, Shao G, Sekmen A, et al. Real-time multiple moving targets detection from airborne IR imagery by dynamic gabor filter and dynamic gaussian detector [J]. *EURASIP Journal on Image and Video Processing*, 2010: 1-22.
- [12] Wu M, Zhou Z, Zhao L, et al. A fast and memory-saving algorithm for sequence-based image registration [J]. *High Power Laser and Particle Beams*, 2012, 24(5): 1038-1042. (in Chinese)
- [13] Hartley R, Zisserman A. *Multiple View Geometry in Computer Vision* [M]. 2nd ed. UK: Cambridge University Press, 2004.