

在无人机视频监控场景中,当前用于动态背景的运动目标检测方法主要有两类,分别为背景运动补偿和光流法^[2].

背景运动补偿的目的是将摄像机运动导致的背景运动去除.利用相邻帧图像中背景对应的特征点匹配来估计背景运动模型的参数^[3],并根据计算得到的全局运动参数计算出当前帧中每个像素的移动速度,预测该像素在下一帧图像中的位置,将目标的检测转化到静止背景条件下.文献[4-7]通过建模从前一帧到当前帧的背景模型,来适应由相机运动造成的背景动态变化,分别取得了不错的效果,但是它们都没有考虑所需的计算资源,很难在无人机平台上达到实时的效果.在此基础上,文献[8,9]提出了改进的方法,减少了检测时间,但仍无法满足移动计算平台的实时性要求.文献[10]进一步通过利用单高斯模型对图像块进行背景建模的方法来降低时间开销,有效地提高了运动目标检测的效率,但是对光照变化较为敏感,当光照条件发生变化时,检测效果大打折扣.

光流法通过估计并分割光流场,实现运动目标检测.文献[11,12]通过稠密光流场的方法来检测运动相机所拍视频中的运动目标,但由于需要计算的光流场的像素点太多,性能受到很大影响.文献[13]采用了一种结合运动边界和稀疏光流的运动目标检测方法,达到了实时的效果,但是没有充分考虑相机运动的情况.

针对以上问题,在文献[10]改进的单高斯模型的基础上,本文提出了一种结合单高斯与光流法的无人机视频运动目标检测方法.该方法首先通过改进的单高斯模型来进行背景建模,获得前景图像,然后将此前景图像作为掩模,提取稀疏特征点,并进行光流检测;最后对检测跟踪到的特征点进行层次聚类,并将运动目标框出.

1 总体流程框架

改进的单高斯模型能有效应对运动相机场景下背景模型被前景像素干扰的问题,并具有实时的建模效率,但经过实验发现,该方法对光照变化比较敏感.相对地,光流法对光照变化有一定的鲁棒性,但由于机载相机是运动的,仅使用光流轨迹聚类进行目标检测时,会出现大量背景像素被误检为前景目标的情况,并且航拍视频往往对比度不高,检测到的特征点在前景目标上分布较少.本文考虑将改进的单高斯模型与光流法相结合,使其在一定程度上实现优势互补.

总体的流程框架如图1所示.针对每一帧视频图

像,首先通过空间高斯滤波器和均值滤波器进行预处理;随后,将图像划分为若干块,对每块图像通过改进的单高斯模型来进行背景建模,以降低计算开销;接着,将计算得到的前景图像作为掩模,执行稀疏特征点的提取与光流检测;最后,对检测跟踪到的特征点进行层次聚类,勾画出运动目标的外边框.

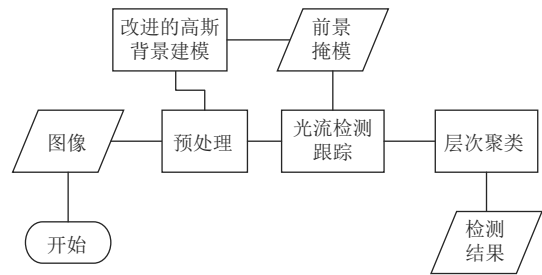


图1 总体流程框架

2 改进的单高斯模型

2.1 单高斯模型的生命值

单高斯 (Single Gaussian Model, SGM) 背景模型基于一个基本假设,在一段时间内,图像序列中的背景像素值应服从高斯概率分布.为了判断 t 时刻对应图像中某个点是否为前景点,需要基于前 $t-1$ 帧图像计算高斯背景模型的各个参数,包括均值和方差.静止相机的目标检测往往具有固定的学习率,也即将一个像素的第一个观测值作为无穷更新模型的均值,这对于需要运动补偿的运动相机而言是不合理的.为此,引入生命值 l 来逐帧更新 SGM 参数,以应对随着时间的推移背景像素产生的变化.

为了减少计算负载以达到实时效果,将每帧图像划分为若干 $N \times N$ 大小的块,以块为单位进行 SGM 背景建模.定义第 i 块图像在 t 时刻的像素集合为 S_i^t , 则其包含像素的数目为 $|S_i^t|$, 将其中第 j 个像素在 t 时刻的强度表示为 I_j^t , 此块对应的均值表示为 μ_i^t , 方差为 σ_i^t , 生命值为 l_i^t , 则对应 SGM 的参数更新公式如下:

$$\mu_i^t = \frac{\tilde{l}_i^{t-1}}{\tilde{l}_i^{t-1} + 1} \tilde{\mu}_i^{t-1} + \frac{1}{\tilde{l}_i^{t-1} + 1} U_i^t \quad (1)$$

$$\sigma_i^t = \frac{\tilde{l}_i^{t-1}}{\tilde{l}_i^{t-1} + 1} \tilde{\sigma}_i^{t-1} + \frac{1}{\tilde{l}_i^{t-1} + 1} V_i^t \quad (2)$$

$$l_i^t = \tilde{l}_i^{t-1} + 1 \quad (3)$$

其中, U_i^t 和 V_i^t 的定义分别如下:

$$U_i^t = \frac{1}{|S_i^t|} \sum_{j \in S_i^t} I_j^t \quad (4)$$

$$V_i^t = \max_{j \in S_i^t} (\mu_i^t - I_j^t)^2 \quad (5)$$

2.2 单高斯模型的选择

采用上述单高斯模型对背景建模容易造成背景模型被前景像素污染的问题, 当有大物体缓缓略过视场时, 现象较为明显. 由式 (1)~(5) 可知, 背景建模从 $t=1$ 时开始, 在初始阶段, 学习率较大, 如第二帧的学习率为 0.5, 随后每一帧的学习率逐渐下降, 初始阶段学习率的快速变化加剧了上述污染问题的严重性. 为此, 采用了两个单高斯模型来建模每帧图像, 一个是当前背景模型 (Current SGM, 简称 CSGM), 一个是备选背景模型 (Optional SGM, 简称 OSGM), 然后按照参数更新的频率确定每帧图像真实的当前背景模型.

将一帧图像第 i 块在 t 时刻的 CSGM 和 OSGM 模型的参数, 包括均值、方差和生命值, 分别表示为 $\mu_{C,i}^t, \sigma_{C,i}^t, l_{C,i}^t$ 和 $\mu_{O,i}^t, \sigma_{O,i}^t, l_{O,i}^t$, 对应的观测均值表示为 U_i^t , 则如果满足式 (6), 则更新 CSGM 的各项参数, 如果满足式 (7), 则更新 OSGM 的各项参数, 并且当满足式 (8), 即 OSGM 的更新频率大于 CSGM 的更新频率时, 则将 OSGM 模型作为新的 CSGM, 并重新初始化 OSGM.

$$(U_i^t - \mu_{C,i}^t)^2 < \theta_s \sigma_{C,i}^t \quad (6)$$

$$(U_i^t - \mu_{O,i}^t)^2 < \theta_s \sigma_{O,i}^t \quad (7)$$

$$l_{C,i}^t < l_{O,i}^t \quad (8)$$

上述方法能够从两个 SGM 中自动选择合适的背景模型, 即便其中几帧图像的前景像素对背景模型造成了影响, 随着时间的推移, 更贴近真实背景的 SGM 模型更新次数变多, 又会变成当前背景模型, 有效避免了前景的不利影响.

2.3 运动补偿

对于运动相机而言, $t-1$ 时刻的 SGM 无法直接用于 t 时刻的目标检测, 需要进行运动补偿. 由于每个 SGM 对应于一个图像块, 直接运动插值的方法在此处不适用, 为此, 采用融合 $t-1$ 时刻 SGM 模型的方法来补偿 t 时刻的模型.

将每帧图像划分成大小为 24×18 的多个网格, 针对每个网格中的角点, 先通过 KLT^[14] 算法进行 $t-1$ 时刻到 t 时刻的跟踪, 然后根据跟踪的结果采用 RANSAC^[15]

算法计算得到单应矩阵 $H_{t,t-1}$, 通过单应矩阵就可以计算得到 t 和 $t-1$ 时刻两帧图像之间各点的对应关系.

假设 t 时刻第 i 块图像的中心点坐标为 C_i^t , 则可通过 $H_{t,t-1}$ 映射得到其在 $t-1$ 时刻的位置坐标 \widetilde{C}_i^{t-1} , 以 \widetilde{C}_i^{t-1} 为中心的虚拟块 \widetilde{S}_i^{t-1} 与 $t-1$ 时刻的多个真实块相重叠, 通过融合此 $t-1$ 时刻各块的 SGM 的各项参数, 可得到 $t-1$ 时刻的各参数 $\widetilde{\mu}_i^{t-1}, \widetilde{\sigma}_i^{t-1}$ 和 \widetilde{l}_i^{t-1} , 以实现运动补偿.

重叠块的集合用 G_i^t 表示, 融合所用到的权重用 $\omega_k (k \in G_i^t)$ 表示, 则模型融合的相关公式如下:

$$\widetilde{\mu}_i^{t-1} = \sum_{k \in G_i^t} \omega_k \mu_k^{t-1} \quad (9)$$

$$\widetilde{\sigma}_i^{t-1} = \sum_{k \in G_i^t} \omega_k [\sigma_k^{t-1} + (\mu_k^{t-1})^2 - (\widetilde{\mu}_i^{t-1})^2] \quad (10)$$

$$\widetilde{l}_i^{t-1} = \sum_{k \in G_i^t} \omega_k l_k^{t-1} \quad (11)$$

其中, 权重 ω_k 满足以下条件:

$$\omega_k \propto \iint_{\widetilde{S}_i^{t-1} \cap G_i^t} dx dy \text{ 并且 } \sum_k \omega_k = 1 \quad (12)$$

当得到了 t 时刻的 SGM 模型后, 通过计算某个像素点属于背景模型的概率, 即可分辨此像素点是前景还是背景.

3 稀疏光流分析

上述改进的单高斯模型检测运动目标具有较好的效果, 但容易受到光照变化的影响, 当前后帧图像的光照条件发生变化时, 容易导致检测出“假前景”. 为此在检测的前景基础上, 通过光流法做进一步分析, 排除因光照变化造成的前景误判. 为了降低光流运算量, 此处采用稀疏光流分析法, 它是基于特征点的跟踪方法, 在提取特征点时, 采用上述检测的前景图像作为掩模.

为了进一步节省计算时间, 采用毫秒级别的 FAST 特征提取方法. FAST (Features from Accelerated Segment Test) 角点探测算法^[15]于 2006 年被提出, 该算法检测的角点满足以下特点: 在像素点的周围邻域内有足够多的像素点与该点处于不同的灰度区域. 领域通常选择以角点为中心、半径为 3 的离散化圆形区域, 如图 2 所示. 一个候选点是否为角点需通过如下角点响应函数来判断, 即

$$N = \sum_p |I(p) - I(o)| < \xi \quad (13)$$

式中, o 为中心像素点 (即候选角点), p 为以 o 为中心的圆周上的任意一点, $I(o)$ 为中心像素点的图像灰度值, $I(p)$ 为圆周上点的图像灰度值, ξ 为给定的阈值, 提取的角点数目与 ξ 负相关, ξ 越小, 角点数目越多. 根据上述角点响应函数, 统计圆周上满足式 (13) 的像素点个数 N . 若 N 大于给定的阈值, 就可以确定该候选点为角点. 通常阈值取 12, 此阈值可以很快排除伪角点.

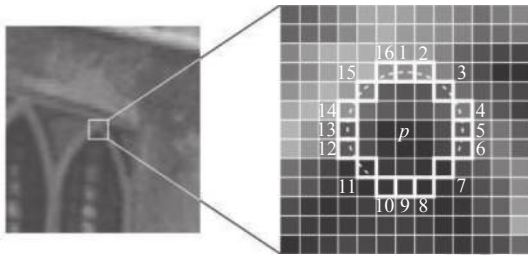


图2 角点探测模板示意图

在检测到特征点以后就可以运用 LK 光流法匹配两帧图像之间的特征点得到光流图. LK 光流算法是由 Lucas 和 Kanade 于 1981 年提出的, 由于算法易于应用在输入图像中的一组点上, 成为了求稀疏光流的一种重要方法. LK 光流算法能够求解出两帧图像之间精确的稀疏光流, 最初运用于图像配准领域并逐渐应用于目标跟踪.

由于 LK 光流法的跟踪过程需要一个大的窗口来捕捉大的运动, 而大窗口往往会违背运动连贯的假设, 导致检测效果不太理想. 金字塔 LK 光流法最初在较大的空间尺度上进行跟踪, 然后通过对图像金字塔自顶向下直到图像像素的处理来修正初始运动速度的假定, 有效解决了这个问题. 也即, 在图像金字塔的最高层计算光流, 用得到的运动估计结果作为下一层金字塔的起始点, 重复这个过程直到金字塔最底层.

稀疏光流分析法的总体算法流程如下:

- 1) 读入一帧图像, 对其进行灰度化操作, 若当前处理的是第一帧图像, 只初始化相应参数, 否则转步骤 2);
- 2) 采用 FAST 算法计算得到前一帧图像的特征点集合 $PreFeaturePtSet$, 获取原跟踪点的运动轨迹在前一帧的位置集合 $PreTrackPtSet$;
- 3) 计算 $PreFeaturePtSet$ 集合中的特征点与 $PreTrackPtSet$ 集合中的各跟踪点之间是否存在欧式距离小于阈值的情况, 如果不存在, 则认为该特征点是新出现的需要跟踪的点, 将其加入 $PreTrackPtSet$ 集合中;
- 4) 对 $PreTrackPtSet$ 中的跟踪点在前一帧和当前帧图像间进行 LK 金字塔光流检测, 得到在当前帧的

位置集合 $CurTrackPtSet$;

5) 根据检测的结果依次处理各类跟踪点. 对已存在运动轨迹的原跟踪点, 如果检测到光流, 则更新跟踪点在当前帧的位置为 $CurTrackPtSet$ 中的对应点, 并更新其 $last_update_index$ 为当前帧序号; 如果没有检测到光流, 则继承上一帧的跟踪结果, 不更新其 $last_update_index$. 对检测到光流的新跟踪点, 为其建立运动轨迹对应的数据结构, 并按上述方法更新;

6) 判断每个跟踪点的 $last_update_index$ 与当前帧序号之间的差值大小, 若大于阈值, 意味着该点对应的运动轨迹长时间没有被更新, 则删除;

7) 统计各跟踪点运动轨迹对应每帧图像的位置变动, 若随着时间推移, 该跟踪点在每帧图像间的位置变动不大, 则删除. 继续执行步骤 1).

4 层次聚类

检测到的每一帧图像的特征点是稀疏的局部运动像素点, 这些点可能属于一个或多个运动目标. 为了有效检测并区分出不同的运动目标, 需要对其做进一步聚类处理. 聚类的算法种类繁多, 常用的有 k -means^[16,17] 等. 但 k -means 方法需要人为设定 k 值和初始化聚类中心点, 不适用于多运动目标检测的场景. 为了避免这些问题, 此处选择一种比较实用的聚类算法——层次聚类.

层次聚类算法有凝聚和分裂两种方式, 取决于层次分解是自底向上 (合并) 还是以自顶向下 (分裂) 的. 凝聚的层次聚类方法使用自底向上的策略, 开始时每个对象都是一个独立的类 (N 个), 然后逐次找出两个距离最近的类进行合并, 所以最多进行 N 次迭代就将所有对象合并到一起了. 分裂的层次聚类方法使用自顶向下的策略, 开始时所有对象都在同一个类中, 然后不断的划分成多个更小的类, 直到最小的类都足够凝聚或者只包含一个对象.

本文采用凝聚的层次聚类算法, 将属于同一运动目标的稀疏特征点聚成一类, 并在对应帧图像上画出运动目标的外边框. 凝聚的层次聚类算法原理:

输入: 给定要聚类的 N 个特征点以及 $N \times N$ 的距离矩阵 (或者是相似性矩阵)

步骤:

- 1) 将每个特征点归为一类, 共得到 N 类, 每类仅包含一个特征点, 类与类之间的距离就是它们所包含的特征点之间的距离;
- 2) 找到距离最近的两个类并合并成一类, 于是类

别个数减一;

3) 重新计算合并的新类与所有旧类之间的距离;

4) 重复步骤 2) 和步骤 3), 直到两类间的距离不小于距离阈值;

根据步骤 3) 的不同, 可将层次聚类方法分为: 单连接算法、全连接算法以及平均连接算法. 其中单连接采用的是最小距离, 全连接采用的是最大距离, 平均连接采用的是平均距离, 各距离公式的定义如下, 其中 $|p-p'|$ 是两个特征点 p 和 p' 之间的距离.

$$D_{\min}(c_i, c_j) = \min_{p \in c_i, p' \in c_j} \{|p-p'|\} \quad (14)$$

$$D_{\max}(c_i, c_j) = \max_{p \in c_i, p' \in c_j} \{|p-p'|\} \quad (15)$$

$$D_{\text{avg}}(c_i, c_j) = \frac{1}{n_i n_j} \sum_{p \in c_i, p' \in c_j} \{|p-p'|\} \quad (16)$$

5 实验结果与分析

为验证本文提出算法的有效性, 主要采用了两个无人机视频数据集, 分别是 VIVID-PETS 2005^[18] 和 Stanford Drone Dataset^[19]. 前者由运动的无人机采集得到, 相机的运动速度较快, 所以背景也存在快速地运动, 并包含有光线变化的样本集; 后者是无人机在校园高空拍摄所得, 无人机处于盘旋的状态, 背景的运动较小, 目标也很小. 下述实验的运行平台是处理器为 Intel Core i7-6700 3.4 GHz, 内存为 DDR4 16 G 的 PC 机.

图 3 所示为金字塔 LK 光流法、改进的单高斯方法以及本文方法在 VIVID-PETS 2005 的 egtest05 数据集上的对比效果. 从数据集第 257 帧原图中可以看到, 无人机在运动过程中, 拍摄到地面上有两辆汽车正在行驶 (其中一辆被树遮挡, 仅漏出一小块表面). 由于树的遮挡、背景的运动等原因, 在不使用掩模的情况下, 光流法检测到的前景有大量的误检, 如图 3(b) 所示; 通过改进的高斯方法对运动背景进行建模, 并分离出前景目标图像, 得到图 3(c) 所示结果, 可看到由于光照的变化, 该方法也检测出了一些“假前景”; 当将上述两种方法结合起来, 即将后者分离得到的前景图像作为前者特征提取的掩模后, 较好的检测出了两个运动目标, 并排除了大量的虚警, 如图 3(d) 所示.

为了考察光照变化以及大目标缓慢运动对检测效果的影响, 采用了光线剧烈变化的 LIMU^[20] 室内部分数据集. 在此数据集上, 先通过改进的单高斯模型进行前景的提取 (如图 4(c)~(d)), 然后通过本文提出的方法检测

图中的运动目标 (如图 4(e)~(f)), 并对两种方法进行比较.

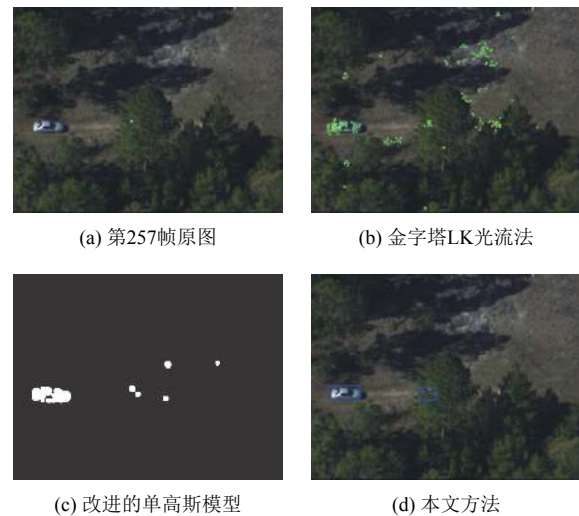


图 3 三种方法运动目标检测效果对比

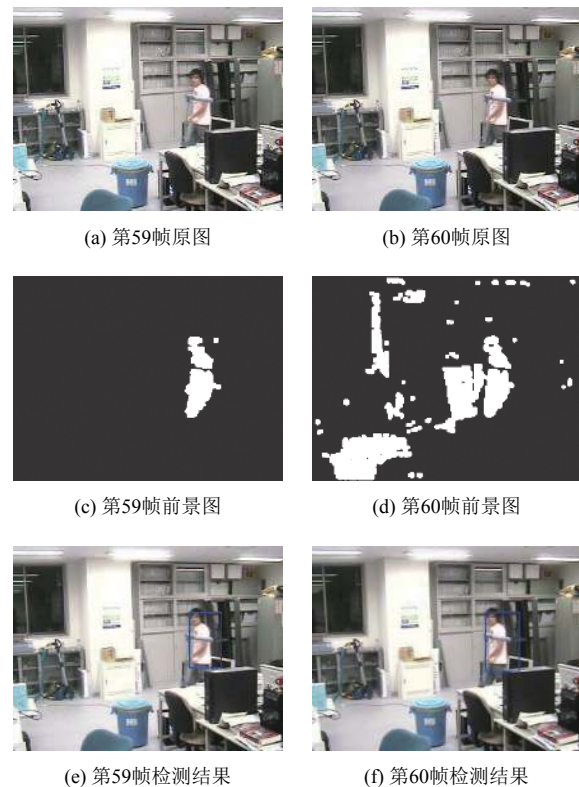


图 4 光线变化对目标检测效果的影响

从图 4 中可以看到, 从数据集的第 59 帧到第 60 帧, 房间的光线变暗了, 由此导致了直接采用改进的单高斯模型提取的前景图像效果变差, 出现了大量的误检测点, 如图 4(d) 所示. 但是采用本文提出的方法, 仍然得到了较为理想的检测结果, 如图 4(f) 所示.

表1中是在VIVID-PETS 2005和Stanford Drone Dataset的部分数据集上,执行本文的运动目标检测算法后,每帧图像的计算耗时,以及与运动相机场景下两种近似实时的运动目标检测算法(文献[9]和文献[10])的性能比较,三者的测试结果分别对应表中的方法1、2、3。其中,本文方法图像块的大小 N 设置为4,表中时间的单位为毫秒(ms)。

表1 各运动目标检测算法耗时情况对比

数据集	分辨率	方法1	方法2	方法3
VIVID01	640×480	50.94	39.16	116.13
coupa	640×480	49.20	37.82	112.15
gates	860×1000	86.61	66.58	197.45
hyang	860×1000	93.62	71.97	213.44

从表中可以看出,由于本文的方法整合了改进的单高斯模型和金字塔LK光流法,因此耗时比文献[10]略高,但是比文献[9]的方法仍然有很大优势,能够在存在运动背景的场景下快速准确地进行运动目标检测,满足无人机平台的实时性要求。

6 总结

本文设计了一种结合了改进单高斯与光流分析的算法,以进行无人机航拍视频图像的运动目标检测。该方法使用改进的单高斯模型来进行背景建模和运动补偿,并计算得到前景图像;然后将其作为掩模,采用FAST算法提取稀疏特征点,并采用金字塔LK光流法进行光流检测和跟踪;最后,通过层次聚类将属于同一个运动目标的前景图像归为一类,勾画出运动目标的外边框。通过实验验证与分析,该方法能有效地处理运动相机造成的前景对背景模型的干扰,背景建模速度快,对光照变化具有一定的容忍度,检测出的目标接近真实目标。

参考文献

- 全小敏. 航拍视频运动目标检测与跟踪方法研究[博士学位论文]. 西安: 西北工业大学, 2015.
- 徐诚, 黄大庆. 基于鲁棒M估计和Mean Shift聚类的动态背景下运动目标检测. 光子学报, 2014, 43(1): 136–141.
- 齐美彬, 汪巍, 蒋建国, 等. 动态场景下的快速目标检测算法. 电子测量与仪器学报, 2011, 25(9): 756–761.
- Zamalieva D, Yilmaz A, Davis JW. A multi-transformational model for background subtraction with moving cameras. 13th European Conference on Computer Vision. Zurich, Switzerland. 2014. 803–817.
- Lim J, Han B. Generalized background subtraction using superpixels with label integrated motion estimation. In: Fleet D, Pajdla T, Schiele B, et al. Computer Vision–ECCV 2014. Cham: Springer, 2014, 8693: 173–187.
- Kwak S, Lim T, Nam W, et al. Generalized background subtraction based on hybrid inference by belief propagation and Bayesian filtering. International Conference on Computer Vision. Barcelona, Spain. 2011. 2174–2181.
- Minematsu T, Uchiyama H, Shimada A, et al. Adaptive search of background models for object detection in images taken by moving cameras. 2015 IEEE International Conference on Image Processing. Quebec City, QC, Canada. 2015. 2626–2630.
- Kim J, Wang XF, Wang H, et al. Fast moving object detection with non-stationary background. Multimedia Tools and Applications, 2013, 67(1): 311–335. [doi: 10.1007/s11042-012-1075-3]
- Kim SW, Yun K, Yi KM, et al. Detection of moving objects with a moving camera using non-panoramic background model. Machine Vision and Applications, 2013, 24(5): 1015–1028. [doi: 10.1007/s00138-012-0448-y]
- Yi KM, Yun K, Kim SW, et al. Detection of moving objects with non-stationary cameras in 5.8 ms: Bringing motion detection to your mobile device. 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Portland, OR, USA. 2013. 27–34.
- Shah AJ. Optical flow for motion detection with moving background. <https://www.researchgate.net/publication/282913643>. 2015.
- Kurnianguro L, Shahbaz A, Jo KH. Dense optical flow in stabilized scenes for moving object detection from a moving camera. 2016 16th International Conference on Control, Automation and Systems. Gyeongju, South Korea. 2016. 704–708.
- 何伟, 张国云, 吴健辉, 等. 结合运动边界和稀疏光流的运动目标检测方法. 小型微型计算机系统, 2017, 38(3): 635–639.
- Tomasi C, Kanade T. Detection and tracking of point features. Technical Report, 1991, 91(21): 9795–9802.
- Rosten E, Drummond T. Machine learning for high-speed corner detection. 9th European Conference on Computer Vision. Graz, Austria. 2006. 430–443.
- 王慧贤, 勒惠佳, 王娇龙, 等. K均值聚类引导的遥感影像多尺度分割优化方法. 测绘学报, 2015, 44(5): 526–532.
- 李果家, 李显凯. 基于稀疏光流场分割的运动目标实时检测算法. 计算机工程与设计, 2017, 38(11): 3029–3035.
- VIVID-PETS 2005. <http://vision.cse.psu.edu/data/vividEval/datasets/datasets.html>.
- Robicquet A, Sadeghian A, Alahi A, et al. Learning social etiquette: human trajectory understanding in crowded scenes. 14th European Conference on Computer Vision. Amsterdam, The Netherlands. 2016. 549–565.
- LIMU Dataset. <http://limu.ait.kyushu-u.ac.jp/dataset/en>.