

基于立体匹配的改进 ViBe 算法

杨 硕, 吉爱萍

(沈阳化工大学 计算机科学与技术学院, 辽宁 沈阳 110142)

摘要:由于深度相机技术的不成熟,限制了融合深度信息的运动目标检测算法的应用范围。对此,提出一种高效的基于随机选择策略的立体匹配算法,用于代替深度相机为运动目标检测算法提供深度信息。首先,建立匹配点样本集合,并在每帧中随机选择匹配点计算其匹配代价和置信度;其次,更新集合并将置信度高的匹配点向邻域传播,随着时间的持续会由粗到精得到深度图;最后,用深度信息改进 ViBe 算法(visual background extractor,视觉背景提取)得到最终的检测效果。实验结果表明,改进算法能够在 400×300 的分辨率下以 8~10 帧/秒的速度运行,消除鬼影和抑制光照突变的速度要比经典 ViBe 算法快 3 倍以上,同时在普通场景中的运动目标的检测效果和消除阴影的能力也明显优于原算法。

关键词:视觉背景提取;鬼影;运动目标检测;深度信息;立体匹配;随机选择

中图分类号:TP391.41

文献标识码:A

文章编号:1673-629X(2018)03-0122-05

doi:10.3969/j.issn.1673-629X.2018.03.026

An Improved ViBe Algorithm Based on Stereo Matching

YANG Shuo, JI Ai-ping

(School of Computer Science and Technology, Shenyang University of Chemical Technology,
Shenyang 110142, China)

Abstract: Due to the immaturity of the depth camera technique, the application range of moving object detection algorithm with the fusion of depth information is restricted. For this, we propose an efficient stereo matching algorithm based on random selection strategy, which provides the depth information for moving object detection algorithm instead of the depth camera. First, the matching point sample sets are established, and their matching cost and confidence are computed by random selection of matching points in each frame. Then, the sets are updated and the matching points with high confidence are propagated to the neighborhood. As time goes on, the depth map will be obtained through a coarse-to-fine pattern. Lastly, the depth information is added to the conventional ViBe algorithm, which leads to the final test results. The experiments show that in the resolution of 400×300 the improved algorithm can achieve a nearly real-time frame rate of 8-10 frames per second and get the tripe speed of conventional ViBe on removing ghost and suppressing pseudo-target caused by light changing. Meanwhile, it obtains better detection and shadow elimination performances in common scene.

Key words: ViBe; ghost; motion object detection; depth information; stereo matching; random selection

0 引言

运动目标检测算法是计算机视觉的基础,应用广泛,因此受到了很多学者的关注并提出了大量的算法,其中具有代表性的有 Gaussian Mixture Model^[1]和 ViBe 算法^[2]等。特别是 ViBe 算法,作为一种通用场景的运动目标检测算法,采用了随机更新策略,计算过程简洁、高效,同时还具有占用资源少,能够在高分辨率下实时运行的优点。但原 ViBe 算法不能很好地处理运动目标检测算法中常见的问题,如鬼影、阴影、光照变化和周期

性运动目标等。因此,很多学者将经典 ViBe 算法作为一个框架,并在此基础上修改或添加自己的模块以实现不同的功能。例如,文献[3-4]通过扩大样本的取值范围或增加样本的比较次数改善 ViBe 的鬼影和阴影问题;文献[5-6]通过帧间差分或三帧差分改善 ViBe 的鬼影问题;文献[7]通过 MeanShift 迭代改善 ViBe 鬼影问题,等等。

同时,近年来,随着硬件设备性能的提高,以及新的、高效的算法的出现,越来越多的学者尝试将深度信

收稿日期:2017-03-05

修回日期:2017-07-11

网络出版时间:2017-11-15

基金项目:辽宁省教育基金项目(L2014171)

作者简介:杨 硕(1983-),男,讲师,博士,研究方向为图像处理与模式识别;吉爱萍(1991-),女,硕士研究生,研究方向为图像处理与模式识别。

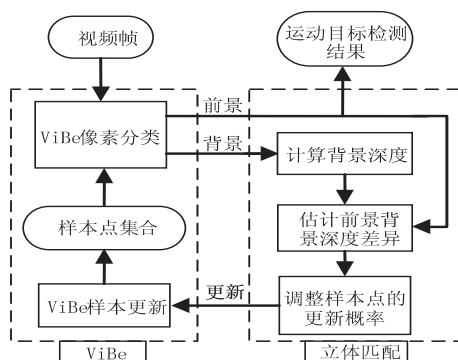
网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20171115.1436.046.html>

息融合到传统的基于单摄像机的运动目标检测算法中。深度信息的引入能够在一定程度上改善运动目标检测算法固有的难题^[8-11],如鬼影、光照变化和目标-背景颜色相近等。但多数学者使用深度相机获取深度信息,由于深度相机技术的不成熟,使得融合后的运动目标检测算法的效果受到了一定限制。具体地,市场上已经商品化的深度相机一般采用激光散斑和 TOF (time of flight, 飞行时间) 技术,存在的主要问题有:价格昂贵,除微软的 Kinect 外,动辄数万元;相机参数固定,对可测的场景深度有要求,基于激光散斑的难以在室外强光下使用;受到传感器和计算量的限制,获取的深度图像分辨率较低,如 TOF 深度相机获取的深度图像的水平分辨率和垂直分辨率仅在 200 ~ 300 之间。以上这些问题使得基于深度相机的运动目标检测算法很难适应通用的视频场景,如大场景、室外监控等,限制了算法的应用。

综上分析,考虑到使用非硬件的方式,即立体匹配算法^[12-14],也可以获得深度信息,且效果取决于算法本身而不受限于硬件平台,两个摄像机也可以随意搭配,具有很大的灵活性。因此,文中提出一种基于立体匹配的运动目标检测算法,并用深度信息改进 ViBe 算法实现运动目标的检测。首先,考虑到实时性,提出一种基于随机选择策略的局部立体匹配算法,局部算法的计算范围仅在有限的窗口内进行^[15-16],因此速度快,能够满足视频的实时处理。具体地:算法在背景建模期间随机选择匹配点,将原本计算量大的立体匹配算法分成若干部分进行,这样,随着时间的推移,背景图像和深度图像将由粗到精地求出。而对于前景,也采用类似的随机选择策略估计出运动目标与背景的深度差异(大于、小于或等于)。然后,根据前景和背景的深度差异调整 ViBe 前景样本点被更新的概率并采用 ViBe 更新算法更新样本点。最后,达到运动目标检测和抑制鬼影、光照变化的目的。

1 文中方法

文中提出算法的流程如图 1 所示。



万方数据 图 1 算法流程

1.1 ViBe 像素分类

ViBe 算法使用随机更新背景像素的策略,它的背景模型中的每一个像素由 N 个背景样本点组成。令图像 (x, y) 位置的像素为 $v(x, y)$, 则背景模型 M 可以表示为: $M(x, y) = \{v_1, v_2, \dots, v_N\}$ 。ViBe 算法的分类过程为:令 $D_s(x, y)$ 为像素距离度量函数,则满足式 (1) 的像素被分类为背景。

$$\#\{M(x, y) \mid D_s(v(x, y), v_i) < R\} \geq \#min, v_i \in M(x, y) \quad (1)$$

其中, R 为距离阈值。

当 $M(x, y)$ 中与 $v(x, y)$ 的距离小于 R 的样本个数不少于 $\#min$ 时, $v(x, y)$ 被分为背景,反之则为前景。

1.2 背景像素深度计算

令 v_L 和 v_R 为左、右视频帧。首先,定义匹配点样本集合 $W(x, y) = \{\varphi_1, \varphi_2, \dots, \varphi_{dm}\}$, 其中 dm (depth max) 为给定的视差最大值; $\varphi_i (1 \leq i \leq dm)$ 为样本, i 表示深度。 φ_i 由以下项目组成:

(1) ds 和 φ^{match} : ds 为 $v_L(x, y)$ 和 $v_R(x + i, y)$ 的匹配代价, φ^{match} 为由 ds 转换得到的分数值;

(2) φ^{belief} 和 φ^{b-adj} : 分别为置信度分数和由邻域置信传播得到的分数值;

(3) φ^{skip} : 用于抑制不稳定的或低置信度的匹配点所用的分数值。

综上, $\varphi_i(x, y) = \{ds_i, \varphi_i^{match}, \varphi_i^{belief}, \varphi_i^{b-adj}, \varphi_i^{skip}\}$ 。

对于分类为背景的像素,使用类似于 ViBe 的随机选择匹配策略获取该点深度,具体的算法思路和流程如下:

(1) 扫描左视频帧。对于像素 $v_L(x, y)$, 有 δ^{sel} 的概率被继续处理(转到步骤 2), 否则跳过处理下一个像素。 δ^{sel} 的取值决定了算法速度,典型值为 $1/10 \sim 1$ 。

(2) 计算匹配代价。对于满足 δ^{sel} 概率的像素 $v_L(x, y)$, 随机产生一个深度值 d , 并生成一个 d 深度样本 $\varphi_d(x, y)$, 按照式 (2) 计算和右帧像素 $v_R(x + d, y)$ 的匹配代价:

$$D(d) = \kappa \times |v_L(x, y) - v_R(x + d, y) + \bar{v}_R(x + d, y) - \bar{v}_L(x, y)| + (1 - \kappa) \times |\nabla v_L(x, y) - \nabla v_R(x + d, y)| \quad (2)$$

其中,第一项是灰度代价,第二项是梯度代价, κ 是权重,文中取 0.2, 令梯度权重更多一些。

第一项中的 $\bar{v}_R(x + d, y) - \bar{v}_L(x, y)$ 是补偿项,用于补偿局部亮度差异, \bar{v}_L 和 \bar{v}_R 是用大尺度均值滤波器得到的均值图像。引入补偿项的原因是:摄像机长期运行后会产生 CCD 老化或镜头落灰等问题,从而使采集到的两视频帧发生局部亮度不一致的问题,特别在

室外场景,该补偿项可以计算出两视频帧局部亮度的差异,用于灰度代价补偿。得到匹配代价后,令 $\varphi_d(x, y)$ 中的 $ds_d = D(d)$, 并更新 $W(x, y)$ 内所有样本的匹配分数:

$$\varphi_i^{\text{match}} = 100 - \kappa^{\text{match}} \times (ds_i - \min_{ds \in W(x, y)} (ds)), \quad 1 \leq i \leq dm \quad (3)$$

其中, κ^{match} 为系数, 让 φ_i^{match} 位于 0 ~ 100 之间, 超出截断。

(3) 继续计算该 d 深度样本的置信度分数:

$$\varphi_d^{\text{belief}} = \kappa^{\text{conf}} \times (2 \times \varphi_d^{\text{match}} - \varphi_{d-1}^{\text{match}} - \varphi_{d+1}^{\text{match}}) \quad (4)$$

其中, κ^{conf} 为系数, 让 $\varphi_d^{\text{belief}}$ 位于 0 ~ 100 之间, 超出截断。

置信度高说明匹配点附近有纹理, 反之意味平坦区域或目标边界。因此, 置信度分数只能由高向低传播, 即由 $\varphi_d^{\text{belief}}$ 向周围 8-邻域传播。具体地, 令 $\tilde{\varphi}_d^{\text{b-adj}}$ 为邻域, 则:

$$\tilde{\varphi}_d^{\text{b-adj}} = \begin{cases} \tilde{\varphi}_d^{\text{b-adj}} + \delta^{\text{adj}}, & \varphi_d^{\text{belief}} > \tilde{\varphi}_d^{\text{belief}} \\ \tilde{\varphi}_d^{\text{b-adj}}, & \varphi_d^{\text{belief}} \leq \tilde{\varphi}_d^{\text{belief}} \end{cases} \quad (5)$$

其中, δ^{adj} 为传播分数, 取 5 ~ 10。

式(5)中, 传播只能在同一深度进行, 即: 当前 d 深度样本的置信度如果大于邻域 d 深度样本的置信度, 则将邻域的 $\tilde{\varphi}_d^{\text{b-adj}}$ 分数增加 δ^{adj} 。因此, $\varphi^{\text{b-adj}}$ 项仅用于记录从邻域传播过来的置信度分数, 该值在每次迭代时不被清零。

(4) 计算 φ^{skip} , 文中算法没有采用双向匹配的措施, 为了抑制目标边界带来的匹配错误, 引入了该项。令 \hat{d}^{t-1} 和 \hat{d}^t 为点 (x, y) 上一帧和当前帧的最终深度值, 则:

$$\varphi_{\hat{d}^{t-1}}^{\text{skip}} = \begin{cases} \varphi_{\hat{d}^{t-1}}^{\text{skip}} + \delta^{\text{skip}}, & \hat{d}^{t-1} < \hat{d}^t \\ \varphi_{\hat{d}^{t-1}}^{\text{skip}}, & \hat{d}^{t-1} \geq \hat{d}^t \end{cases} \quad (6)$$

其中, δ^{skip} 为固定值, 一般取 10 ~ 20。该项的意义是: 当深度值不断跳变时, 说明该点置信度低或为目标边界, 按照惯例, 这样的点需要将其视差值置为 0, 因此, 当最终深度由低跳至高时, 将深度值低的采样的分值 $\varphi_{\hat{d}^{t-1}}^{\text{skip}}$ 增加一较小的固定值 (δ^{skip}), 这样在下一步处理中, 该低值深度会比高值深度更有可能被选为最终的深度。

(5) 确定最终深度。对于 $W(x, y)$ 中所有不同深度的样本, 将分值之和 ($\varphi_i^{\text{match}} + \varphi_i^{\text{belief}} + \varphi_i^{\text{b-adj}} + \varphi_i^{\text{skip}}$) 最大的样本所对应的下标索引作为最终深度, 并记作 \hat{d} 。这样, 由 $\hat{d}(x, y)$ 构成的图像就是最终的背景深度图。

(6) 分值更新, φ_i^{match} 和 $\varphi_i^{\text{belief}}$ 需要清零, 而 $\varphi_i^{\text{b-adj}}$ 和 φ_i^{skip} 要进行衰减, 避免堆积产生局部极大值, 具体为:

$$\begin{cases} \varphi_i^{\text{match}}(t) = 0 \\ \varphi_i^{\text{belief}}(t) = 0 \\ \varphi_i^{\text{b-adj}}(t) = \gamma \times \varphi_i^{\text{b-adj}}(t-1) \\ \varphi_i^{\text{skip}}(t) = \gamma \times \varphi_i^{\text{skip}}(t-1) \end{cases}, 1 \leq i \leq dm \quad (7)$$

其中, γ 为衰减系数, 取 0.8 ~ 0.95 效果较好。

衰减过程类似于运动目标检测算法中的背景建模过程, 从邻域传递过来的分数会随着距离和时间不断衰减, 以防止分数堆积导致所占比例过大。然后, 转步骤 1 继续扫描下一像素, 直到整帧处理完毕。

1.3 估计前景-背景深度差异

对于运动目标检测算法, 前景深度不需要精确求出, 只要判断出前景和背景的深度差异 (大于、小于或等于) 即可, 三种情况分为:

(1) 前景和背景深度相同: 如果背景深度置信度高, 则可能是阴影或光照变化等原因; 反之, 可能是由前景和背景距离摄像机较远或目标边界干扰等原因造成。

(2) 前景深度大于背景: 如果背景深度置信度高, 则可能是鬼影, 或者是匹配错误造成。

(3) 前景深度小于背景: 正常的运动目标。

具体的估计算法流程为:

(1) 处于前景点位置的背景深度已经得到, 根据该深度计算前景点的匹配代价 (见式(2)), 记作 ds_d 。

(2) 生成若干个 ($dm/10 \sim dm/4$) 小于背景深度 \hat{d} 的随机深度, 并计算前景匹配代价, 将代价最小的记作 ds_{Low} 。

(3) 同步骤 2, 生成若干个大于背景深度的随机深度, 将匹配代价最小的记作 ds_{High} 。

(4) 比较 ds_d , ds_{Low} 和 ds_{High} 的大小, 如果 ds_d 最小, 则将前景的深度记作“相等”; 如果 ds_{Low} 最小, 则记作“小于”; 其他则记作“大于”。

1.4 ViBe 样本更新

经典的 ViBe 算法对分类为背景的像素会以一定的概率将其更新为分类样本 (加入到 $M(x, y)$ 中), 并以一定概率向邻域传播; 对于分类为前景的像素, 则以很低的概率更新或者根据存在时间的长短进行更新。

文中算法在估计出前景-背景深度差异后, 也是通过调整经典 ViBe 算法前景样本的更新概率达到改进算法的目的。具体做法是: 对于分类为前景的像素, 如果前景的深度大于等于背景深度, 则该点被更新为背景样本的概率增加的数值为: $\varphi_d^{\text{belief}} / \text{MaxBelief}$ 。其中, $\varphi_d^{\text{belief}}$ 为背景像素置信度, \hat{d} 为背景深度值, MaxBelief 为置信度分数能取到的最大值。

2 实验和结果分析

算法在 Windows 10+Visual Studio 2012+OpenCV

2.4.6 环境下实现。计算机主要配置为 Intel Core I7 处理器,8 GB 内存。由于常用的运动目标检测算法视频库中缺少能够产生深度图的左右图像视频,因此使用两个海康网络摄像机获取实验视频,视频帧率为 15 帧/秒,将视频分辨率调整为 400×300。文中算法的主要参数设置为:

(1) 立体匹配部分:最大深度(视差)dm 为 30,更新概率 δ^{sel} 为 1/6,匹配代价权重为 0.2,置信度传播分数 δ^{adj} 为 8,跳变抑制分数 δ^{skip} 为 10,衰减系数 γ 为 0.85。

(2) ViBe 部分:样本个数为 15,距离阈值 R 为 20,像素分类阈值#min 为 4,更新概率为 1/16。

实验中使用的对比算法是经典 ViBe 算法,其参数设置同文中算法的 ViBe 部分。

2.1 鬼影抑制

鬼影是由静止的物体突然运动而产生的伪运动目标,鬼影抑制能力是衡量运动目标检测算法好坏的一个重要标准。图 2 为经典 ViBe 算法和文中改进算法对鬼影的抑制效果,其中产生鬼影的物体是一个遥控玩具机器人:视频第 212 帧时玩具开始向右运动,226 帧时运动到一个新的位置,从而在原位置产生了鬼影。图 2(c)、(d)分别是经典 ViBe 算法和文中算法在 237 帧得到的前景(运动目标)图像,图 2(d)中纯白色像素表示运动目标,背景是深度图。由于文中算法引入了深度信息,在目标突然移动时,鬼影处的深度信息由真实背景的深度代替,深度值会突然增加,算法对这样的像素加大更新概率,从而达到快速消除鬼影的目的。

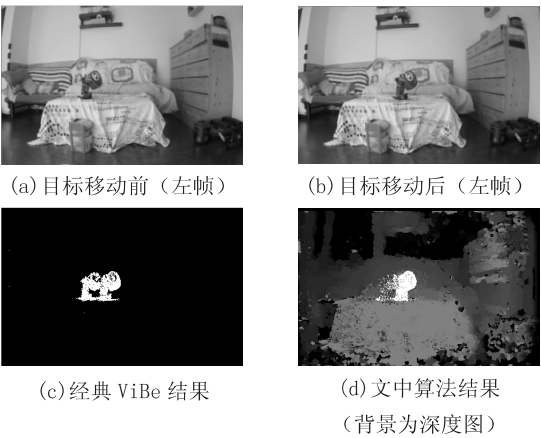


图 2 鬼影抑制效果对比

图 3 展示了两种算法的鬼影随时间变化的情况。从图中可见,文中算法在第 300 帧时鬼影已经基本消失,鬼影存在的时间约为 88 帧(300-212);而经典 ViBe 算法在第 516 帧时才基本消失,比改进算法慢了 3 倍多。

2.2 光照变化抑制

图 4 是两种算法抑制光照变化的结果,图 4(a)为初

始帧,此时场景左侧灯光未点亮,第 99 帧时,左侧灯光突然点亮产生图 4(b)。图 4(c)、(d)为产生光照突变 50 帧后经典 ViBe 算法和文中算法的效果。可以看出,虽然过了一段时间,但经典 ViBe 算法依然残余大量的伪目标。与处理鬼影相似,文中算法根据伪目标的深度与背景一致的性质快速消除了光照带来的影响。

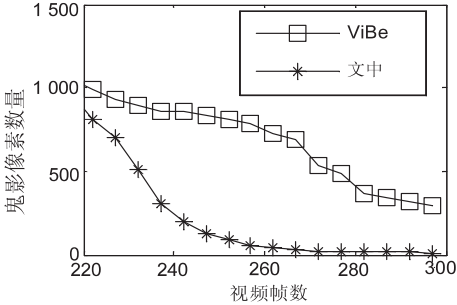


图 3 鬼影伪前景像素数量对比

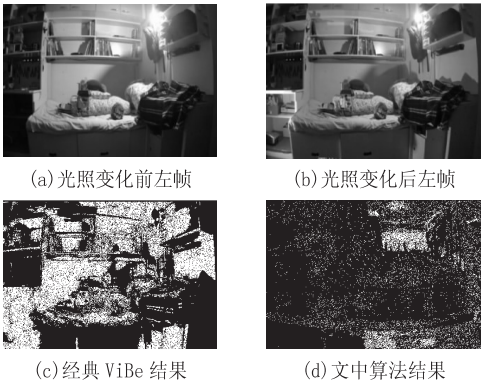


图 4 光照变化抑制效果对比

图 5 为光照变化的伪前景像素数量随时间逐渐减少的情况。由图可见,文中算法的下降速度远快于经典 ViBe 算法,并在 181 帧时伪前景几乎消失干净,伪前景像素的持续时间为 80 帧;而经典 ViBe 算法直到 350 帧时才消失干净,持续时间为文中算法的 3 倍多。

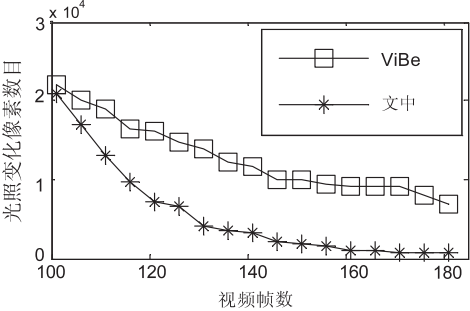


图 5 光照变化伪前景像素数量对比

2.3 阴影抑制

图 6 为算法对阴影的抑制效果。文中算法对阴影的处理与光照突变相同,都是利用前景和背景深度相同的特点增大前景点更新概率实现的,但是不同的是,阴影是随运动目标不断运动的,因此,仅通过深度信息完全去除阴影是不可能的,只能产生一定的抑制效果。

由图 6 可见,经典 ViBe 算法产生的阴影面积较大,而文中算法产生的阴影面积较小。

图 6 视频帧长度为 326 帧,目标一直处于运动状态,通过每隔 30 帧手工计算经典 ViBe 和文中算法阴影面积的大小并求平均值,得出文中算法阴影面积平均比经典 ViBe 算法小 29%,在抑制阴影方面比原算法有所提升。



(a) 原视频帧 (b) 经典 ViBe 结果 (c) 文中算法结果
(背景为深度图)

图 6 阴影抑制效果对比

2.4 算法效率对比

由于文中算法使用了随机更新策略,因此在计算量上相比经典 ViBe 算法并未增加太多。文中算法的平均处理时间为 109 ms/帧,帧速维持在 8~10 帧/s,平均为 9 帧/s,达到了准实时的效果。

3 结束语

文中提出的立体匹配算法和改进的 ViBe 运动目标检测算法是一个整体,而不是独立的两个部分,在硬件上只要求一台计算机和两台型号相同的普通摄像机即可,因此成本很低,易实现。同时,也克服了专用的激光散斑和 TOP 深度相机对环境要求高的问题,并可以通过调整两台摄像机的间距改变检测的深度范围,具有很大的灵活性和应用范围。不足之处在于,为了降低计算量,该算法仅用灰度信息作为特征,在像素分类效果上远差于颜色信息,同时深度图的精度还有待提高,这些问题将会在未来的研究中得以解决。

参考文献:

- [1] STAUFFER C, GRIMSON W. Adaptive background mixture models for real-time tracking[C]//Proceeding of IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 1999.
- [2] BARNICH O, DROOGENBRECK M V. ViBe: a powerful random technique to estimate the background in video sequences [C]//Proceeding of ICASSP 2009. Taipei: IEEE Computer Society Press, 2009: 945-948.
- [3] 余 烨, 曹明伟, 岳 峰. EVibe: 一种改进的 Vibe 运动目

标检测算法[J]. 仪器仪表学报, 2014, 35(4): 924-931.

- [4] 陈 亮, 陈晓竹, 范振涛. 基于 Vibe 的鬼影抑制算法[J]. 中国计量学院学报, 2013, 24(4): 425-429.
- [5] 赵光明, 韩 光, 李晓飞, 等. 基于融合帧间差的改进 Vibe 方法[J]. 计算机技术与发展, 2015, 25(3): 76-80.
- [6] 徐久强, 江萍萍, 朱宏博, 等. 面向运动目标检测的 ViBe 算法改进[J]. 东北大学学报: 自然科学版, 2015, 36(9): 1227-1231.
- [7] 张夏清, 茅耀斌. 一种改进的 ViBe 背景提取算法[J]. 计算机技术与发展, 2016, 26(7): 36-39.
- [8] 杨 磊, 任衍允, 蔡纪源. 一种基于深度数据的高斯模型运动目标检测方法[J]. 计算机技术与发展, 2015, 25(9): 27-30.
- [9] LEENS J, PIERARD S, BARNICH O, et al. Combining color, depth, and motion for video segmentation[C]//International conference on computer vision systems. [s. l.]: [s. n.], 2009: 104-113.
- [10] OTTONELI S, SPAGNOLO P, MAZZEO P L, et al. Improved video segmentation with color and depth using a stereo camera[C]//IEEE international conference on industrial technology. Western Cape, South Africa: IEEE, 2013: 1134-1139.
- [11] 胡良梅, 段琳琳, 张旭东, 等. 融合颜色信息与深度信息的运动目标检测方法[J]. 电子与信息学报, 2014, 36(9): 2047-2052.
- [12] SUN J, ZHENG N N, SHUM H Y. Stereo matching using belief propagation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(7): 787-800.
- [13] WANG L, YANG R G. Global stereo matching leveraged by sparse ground control points[C]//IEEE conference on computer vision and pattern recognition. Washington, DC, USA: IEEE Computer Society, 2011: 3033-3040.
- [14] HIRSCHMULLER H. Stereo processing by semiglobal matching and mutual information[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 328-341.
- [15] MIRCEA P M, SERGIU N, RADU D. Patch warping and local constraints for improved block matching stereo correspondence[C]//12th international conference on intelligent computer communication and processing. [s. l.]: IEEE, 2016: 321-327.
- [16] ZHANG K, LI J Y, LI Y J, et al. Binary stereo matching [C]//21st international conference on pattern recognition. Tsukuba, Japan: [s. n.], 2012: 356-359.