

融合 SORM 背景模型和 DTCNN 阈值模型的运动目标检测

贾丽娟

(西北民族大学数学与计算机科学学院, 兰州 730030)

摘 要: 为降低动态背景对运动目标检测性能的影响, 提出一种运动目标检测方法。该方法融合自组织视网膜映射图(SORM)背景模型和离散时间卷积神经网络(DTCNN)阈值模型。依据 SORM 构建背景模型, 结合运动目标检测要求, 改进 DTCNN 算法, 构建阈值模型, 在检测运动目标的过程中降低噪声干扰, 自适应更新 SORM 和 DTCNN 模型中的相关参数, 以适应场景变化。实验结果表明, 与经典的高斯混合模型、自组织背景差分 and 增长自组织映射图方法相比, 该方法的目标检测性能更好, 尤其是对动态背景的适应能力更强。

关键词: 目标检测; 背景模型; 阈值模型; 自组织视网膜映射图; 神经网络; 动态背景

中文引用格式: 贾丽娟. 融合 SORM 背景模型和 DTCNN 阈值模型的运动目标检测[J]. 计算机工程, 2016, 42(1): 220-224.

英文引用格式: Jia Lijuan. Moving Object Detection Fused with SORM Background Model and DTCNN Threshold Model[J]. Computer Engineering, 2016, 42(1): 220-224.

Moving Object Detection Fused with SORM Background Model and DTCNN Threshold Model

JIA Lijuan

(Mathematics and Computer Institute, Northwest University for Nationalities, Lanzhou 730030, China)

[Abstract] For reducing the impact on the performance of moving object detection by dynamic background, a moving object detection method is proposed. This method is fused with Self-organizing Retinotopic Maps(SORM) background model and Discrete Time Cellular Neural Network(DTCNN) thresholding model, builds background model according to SORM, ameliorates DTCNN in terms of requirement of moving object detection, constructs thresholding model to remove noise during the process of moving object detection, and updates the parameters of SORM and DTCNN models to adapt scene change. Experimental results show that this method has better performance of object detection, especially has stronger adaptation capacity for dynamic background, compared with traditional methods such as Gaussian mixture model, self-organizing background subtraction and growing self-organizing maps.

[Key words] object detection; background model; threshold model; Self-organizing Retinotopic Maps(SORM); neural network; dynamic background

DOI: 10.3969/j.issn.1000-3428.2016.01.039

1 概述

运动目标检测是自动目标识别、行为理解和人工智能领域的研究基础, 现有运动目标检测方法主要分为 3 类, 即光流法^[1]、帧差法^[2]和背景差法^[3]。光流法尽管检测精度高, 但检测速度很慢; 帧差法运算效率高, 但经常存在漏检现象; 背景差法可以兼顾目标变化与背景变化, 且检测速度快, 是目前研究最多的运动目标检测方法。通过构建不同的背景模型, 目前已有多种实用的基于背景差的运动目标检

测方法。

文献[4]提出一种高斯混合模型(Gaussian Mixture Model, GMM), 可以构建适应光照突变和阴影的背景模型, 提高运动目标检测方法对光照和阴影变化的适应能力, 是目前最为经典的背景模型。但该模型对动态背景(如水波纹、摇摆的树叶等)的建模效果不够理想。为提高运动目标检测方法对动态背景的适应能力, 文献[5]提出了一种自组织背景差分(Self-organizing Background Subtraction, SOBS)方法, 采用一个自组织模型表示输入模式, 提高对动

基金项目: 国家自然科学基金资助项目(61163066, 60902074)。

作者简介: 贾丽娟(1979-), 女, 副教授、硕士, 主研方向为图形图像处理、人工智能。

收稿日期: 2015-06-02 **修回日期:** 2015-06-29 **E-mail:** jialijuanxz@163.com

态背景的适应能力,但是 SOBS 对每一个视频序列都需要人工调整一些参数,方法的自适应能力一般;文献[6]为了将图像的颜色特征和纹理特征相融合,研究了基于 Choquet 模糊积分的运动目标检测方法和自适应阈值的 Choquet 积分方法。

文献[7]采用4种观测向量生成方式结合多通道数据,提取运动目标的运动特性和有效信息,研究了能够有效应对缓慢移动和低区分性目标的运动目标检测方法。文献[8]为了降低传统方法的复杂度,研究一种基于分块速度域的迭代红外运动目标检测方法。文献[9]以加快对象检测的一部分为基础模型,研究一种快速变形物体检测方法。文献[10]研究了目标检测中的尺度不变轮廓分割上下文问题。文献[11]根据亮度特征自相关参数判断,结合亮度特征自相关参数的拟合,基于 GMM 和亮度差值,研究一种基于亮度特征自相关的运动目标检测方法。文献[12]基于文法描述特征和扩展实现通用目标类识别,并将目标检测和姿态估计相结合研究了一种判别化视觉文法模型。文献[13]在研究部件选取特征学习策略和弱部件模型基础上,构建了一种自适应部件学习方法。文献[14]基于马尔可夫随机场研究一种运动目标检测与分布差异演化方法。

为进一步提高运动目标检测方法对动态背景的适应能力,本文提出一种融合自组织视网膜映射图(Self-organizing Retinotopic Maps, SORM)背景模型和离散时间卷积神经网络(Discrete Time Cellular Neural Network, DTCNN)阈值模型的运动目标检测方法,并在 Changedetection.net 测试数据库中的动态背景测试子集下进行定性和定量评价。

2 本文方法

本文方法主要包括4个步骤,即颜色空间转换、背景模型构建、阈值模型构建和参数自适应更新。

2.1 颜色空间转换

一般地,输入视频图像序列的原始颜色空间为 RGB,该颜色空间易受光照干扰,不利于面向动态背景的背景模型构建。为此,本文先将 RGB 颜色空间转换到 HSV 颜色空间,转换公式为:

$$\begin{cases} V = \max(R, G, B) \\ S = 1 - \min(R, G, B)/V \\ H = \pi h/3 \end{cases} \quad (1)$$

其中:

$$h = \begin{cases} 5 + (V - B)/[V - V_{\min}] & R = V \text{ 且 } G = V_{\min} \\ 1 - (V - G)/[V - V_{\min}] & R = V \text{ 且 } B = V_{\min} \\ 1 + (V - R)/[V - V_{\min}] & G = V \text{ 且 } B = V_{\min} \\ 3 - (V - B)/[V - V_{\min}] & G = V \text{ 且 } R = V_{\min} \\ 3 + (V - G)/[V - V_{\min}] & B = V \text{ 且 } R = V_{\min} \\ 5 - (V - R)/[V - V_{\min}] & B = V \text{ 且 } G = V_{\min} \end{cases} \quad (2)$$

其中, $V_{\min} = \min(R, G, B)$ 。为便于描述,本文采用笛卡尔坐标表示 HSV 颜色空间,具体为:

$$z = \{X_s, Y_s, V\} \quad (3)$$

其中:

$$\begin{cases} X_s = V \cdot S \cdot \cos(H) \\ Y_s = V \cdot S \cdot \sin(H) \end{cases} \quad (4)$$

2.2 背景模型构建

在面向动态背景的目标检测系统中,经常要采用背景模型来适应不同时刻背景的变化。本文以文献[7]提出的 SORM 方法为基础构建背景模型。

仿照视觉皮层的底层交互机理, SORM 对每一个神经元分配一个权重图,神经元竞争表示为:

$$\eta(m)z, t = \sum_k I(x, y, z)^t w(x, y, m)^{z, t} \quad (5)$$

其中, $I(x, y, z)^t$ 表示第 t 帧像素点 (x, y) 处颜色通道 z 的颜色值; $w(x, y, m)^{z, t}$ 表示对应的神经元 m 的权重向量, $m = 1, 2, \dots, M$; M 为神经元总数,本文取 $M = 4$ 。

胜出的神经元向量 $w(m)^{z, t}$ 在每一个颜色邻域中都拥有最大的 $\eta(m)^{z, t}$ 。邻域权重依据 Hebbian 准则在每一次学习迭代过程中进行更新,表示为:

$$w(x, y, m)^{z, t+1} = w(x, y, m)^{z, t} + \alpha [I(x, y, z)^t - w(x, y, m)^{z, t}] \beta(m)^z \quad (6)$$

其中, α 表示学习速率; $\beta(m)^z$ 表示邻域函数,具体为:

$$\beta(m)^z = \exp(- (m_w^z - m)^2 / \sigma_\beta^2) \quad (7)$$

其中, m_w^z 表示胜出神经元在颜色通道 z 处的坐标; σ_β 表示邻域半径,定义为:

$$\sigma_\beta = \max[13.5 \exp(-\frac{5t_a}{T_f}), \sigma_s] \quad (8)$$

为避免 $t \rightarrow \infty$ 时 $\beta(m)^z \rightarrow \infty$, σ_s 应大于 0。 T_f 决定了使 σ_β 落到 σ_s 所需的帧数, t_a 为帧序号,初始值为 0,后续逐帧加 1。

学习速率可以通过下式计算:

$$\alpha = \max[\exp(-\frac{6t_a}{T_f}), \alpha_s] \quad (9)$$

其中, $\alpha_s \in (0, 1]$, 用于避免 $t \rightarrow \infty$ 时 $w(x, y, m)^{z, t+1} \rightarrow w(x, y, m)^{z, t}$ 。

神经元的输出为:

$$\bar{w}(x, y, z)^m = w(x, y, m)^{z, t} \quad (10)$$

这样,对于所有神经元,其输出结果的平均值即为像素的期望值,也即背景模型,表示为:

$$I_B(x, y, z)^t = \frac{1}{M} \sum_{m=1}^M \bar{w}(x, y, z)^m \quad (11)$$

对于动态背景视频, $I_B(x, y, z)^t$ 可能存在噪声。为了降低噪声干扰, $\bar{w}(x, y, z)^m$ 也需要实时更新,本文采用的更新策略时:

$$\bar{w}(x, y, z)^m = \begin{cases} w(x, y, z)^{z, t} & m_r = m \\ \bar{w}(x, y, z)^m & \text{其他} \end{cases} \quad (12)$$

其中, m_r 表示 t/M 的余数。

2.3 阈值模型构建

在建立背景模型之后,本文采用背景差分法检测运动目标。当前帧与背景的差分图像为:

$$I_E(x,y)^t = \|I_B(x,y,z)^t - I_B(x,y,z)^{t-1}\|_2 \quad (13)$$

对差分图像进行分割,可以得到运动掩膜。这里,阈值的选择对运动目标的提取结果影响很大,阈值太大,运动目标点易丢失,反之易将背景点误判为目标点。

本文以 DTCNN 模型为基础构建阈值模型。DTCNN 模型对每一像素点分配一个阈值,通用模型为:

$$\begin{aligned} X(x,y)^p &= \sum_{u=-r}^r \sum_{v=-r}^r A(u,v) y(u+x,v+y)^p \\ &+ \sum_{u=-r}^r \sum_{v=-r}^r B(u,v) I(u+x,v+y)^p \\ &+ Z(x,y) \end{aligned} \quad (14)$$

其中, $X(x,y)$ 表示 DTCNN 的状态; p 为迭代次数; A 和 B 为权重矩阵; Z 为偏移量。迭代输出为:

$$\begin{aligned} y(x,y)^{p+1} &= \frac{1}{2} (|X(x,y)^p + 1| - |X(x,y)^p - 1|) \\ &= f[X(x,y)] \end{aligned} \quad (15)$$

上述模型在后续权重矩阵下会引发较多噪声:

$$\begin{cases} A = \begin{bmatrix} a_2^p & a_2^p & a_2^p & a_2^p & a_2^p \\ a_2^p & a_1^p & a_1^p & a_1^p & a_2^p \\ a_2^p & a_1^p & 2 & a_1^p & a_2^p \\ a_2^p & a_1^p & a_1^p & a_1^p & a_2^p \\ a_2^p & a_2^p & a_2^p & a_2^p & a_2^p \end{bmatrix} \\ Z = -z_0 \\ B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{cases} \quad (16)$$

其中, $a_1 < 1$; $a_2 < 1$; $z_0 \in [-1, 1]$ 。

为降低噪声干扰,提高运算效率,本文对 DTCNN 模型进行改进,将式(14)改为:

$$\begin{aligned} X(x,y)^{p+1} &= 2f[X(x,y)^p] + a_1^p \sum_{(u,v) \in N_8} f[X(u,v)^p] \\ &+ a_2^p \sum_{(u,v) \in N_{16}} f[X(u,v)^p] - z_0 \end{aligned} \quad (17)$$

其中, N_8 表示像素点的 8 邻域; N_{16} 表示像素点的 5×5 邻域窗口内除去 8 邻域和该像素点的其他像素点集合。

该阈值模型在每次迭代过程中进行阈值分割,每一次迭代的分割阈值为:

$$\begin{aligned} T(x,y)^p &= a_1^p \sum_{(u,v) \in N_8} f[X(u,v)^p] \\ &+ a_2^p \sum_{(u,v) \in N_{16}} f[X(u,v)^p] - z_0 \end{aligned} \quad (18)$$

在差分图像 $I_E(x,y)^t$ 中,对于一个像素点 $(x,$

$y)$, 如果其值接近 1, 且其邻域内像素点的值接近 0, 则认为该像素点为噪声。于是,在阈值模型中,令:

$$I(x,y) = 1 - 2I_E(x,y)^t \quad (19)$$

本文取迭代次数为 8, 迭代后的最终结果为:

$$I_D(x,y)^t = [y(x,y)^8 - 1]/2 \quad (20)$$

当 $I_D(x,y)^t = 1$ 时,表示像素点 (x,y) 为运动目标点,否则为背景。

由式(18)可以看出,由于 $a_1 < 1$ 且 $a_2 < 1$, 即使 $z_0 = 1$ 的情况下噪声仍可以在迭代过程中下降。而当 $z_0 = 1$ 时,差分图像中所有像素值不为 0 的点在 $I_D(x,y)^t$ 中均为 1。因此,本文取 z_0 为常数 1, 这样可以避免目标丢失。在本文中,模型的初始值为 $X(x,y)^0 = I(x,y)$, 参数 a_1 和 a_2 依据目标检测结果自适应调整。

2.4 参数自适应更新

参数的自适应更新有助于提高运动目标检测方法对复杂环境的鲁棒性。本文方法包含两部分参数:(1)背景模型部分,相关参数有: $\alpha, \alpha_s, \beta(m)^z, T_f, \sigma_s$ 和 σ_β ; (2)阈值模型部分,相关参数有: a_1 和 a_2 。由式(7)~式(9)可见, $\beta(m)^z, \sigma_\beta$ 和 α 由 T_f, α_s 和 σ_s 确定,故需要自适应调整的参数有 $T_f, \alpha_s, \sigma_s, a_1$ 和 a_2 , 详细讨论如下:

(1) T_f 的更新

当视频序列的第 1 帧就存在运动目标时,背景模型会在更新过程中将其更新为背景。为了避免目标被更新, SORM 必须在不影响后续帧的背景模型的基础上,快速检测和学习第 1 帧中的变化,这可以通过自适应调整 T_f 来实现,通过调整 T_f , 改变 α 和 σ_β 的衰减速度,从而产生不同的学习速度。

记第 1 帧图像与第 2 帧图像之间的亮度差异的平均值为 μ_v , 表示为:

$$\mu_v = \sum_y \sum_x [I(x,y,V)^2 - I(x,y,V)^1] \quad (21)$$

当 μ_v 不为 0 时,认为第 1 帧图像中存在运动目标,否则认为第 1 帧图像中无运动目标。依据 μ_v 的不同,设置不同的 T_f , 具体为:

$$T_f = \begin{cases} 15 & \mu_v = 0 \\ 50 & \mu_v \neq 0 \end{cases} \quad (22)$$

(2) $a_1, a_2, \alpha_s, \sigma_s$ 自适应更新部分:

由于运动目标在运动过程中可能会停下来,当其停下时,背景模型易将其更新为背景。但如果不更新,又存在动态背景被误识别为运动目标的问题。为解决这一问题,需要自适应调整参数 a_1, a_2, α_s 和 σ_s 。

由于动态背景和运动目标需要采用不同的参数更新策略,因此本文先区分视频图像中的动态背景和运动目标。记前后 2 帧图像的差分图像为:

$$I_{VD}(x,y,V)^t = |I(x,y,V)^t - I(x,y,V)^{t-1}| \quad (23)$$

令 μ_d, μ_t 和 μ_{t-1} 分别表示 $I_{VD}(x,y,V)^t, I(x,y,V)^t$ 和 $I(x,y,V)^{t-1}$ 的期望值,记:

$$\mu_a = |\mu_t - \mu_{t-1}| \quad (24)$$

μ_a 和 μ_d 可用于检测动态目标和动态背景,两者的值越小,视频帧中含运动目标的概率越小。参数 a_1, a_2, α_s 和 σ_s 依据 μ_a 和 μ_d 进行自适应更新,具体为:

$$\begin{cases} \alpha_s = 0.21 + S_d(0.13, 0.10, \mu_d) - S_a(0.31, 0.15, \mu_a) \\ \sigma_s = 0.49 + S_d(2.38, 0.15, \mu_d) - S_a(2.38, 0.15, \mu_a) \\ \alpha_s = 0.07 + S_d(0.07, 0.13, \mu_d) - S_a(0.07, 0.13, \mu_a) \\ \alpha_s = 0.02 + S_d(0.05, 0.08, \mu_d) - S_a(0.05, 0.08, \mu_a) \end{cases} \quad (25)$$

其中:

$$\begin{cases} S_a(s_1, s_2, \mu) = \frac{s_1}{1 + \exp[-40(10\mu + s_2)]} \\ S_d(s_1, s_2, \mu) = \frac{s_1}{1 + \exp[-30(10\mu + s_2)]} \end{cases} \quad (26)$$

在本文中,参数的初始值为: $a_1 = 0.04; a_2 = 0.01; \alpha_s = 0.26; \sigma_s = 0.52$ 。

3 实验结果与分析

为验证本文方法的目标提取性能,选用国际上公认的 Changedetection. net 测试数据库下的 dynamic background 测试子集作为实验数据库,并与目前性能较优的 GMM, SOBS 和 GSOM 方法进行实验对比,从定性和定量2个层面评价本文方法的跟踪性能。

3.1 定性评价

限于篇幅,本文仅列举在动态背景测试子集下的 boats 图像序列中的部分运动目标检测结果,如图1所示。其中,第1列~第3列分别为第1950帧、第1960帧、第1970帧。由图可见,湖面背景一直在变化,这种情况下前3种方法检测的运动目标存在明显的噪声,而本文方法的运动目标检测结果明显优于前3种方法,可见本文方法对动态背景的适应能力较强,误检率较低。

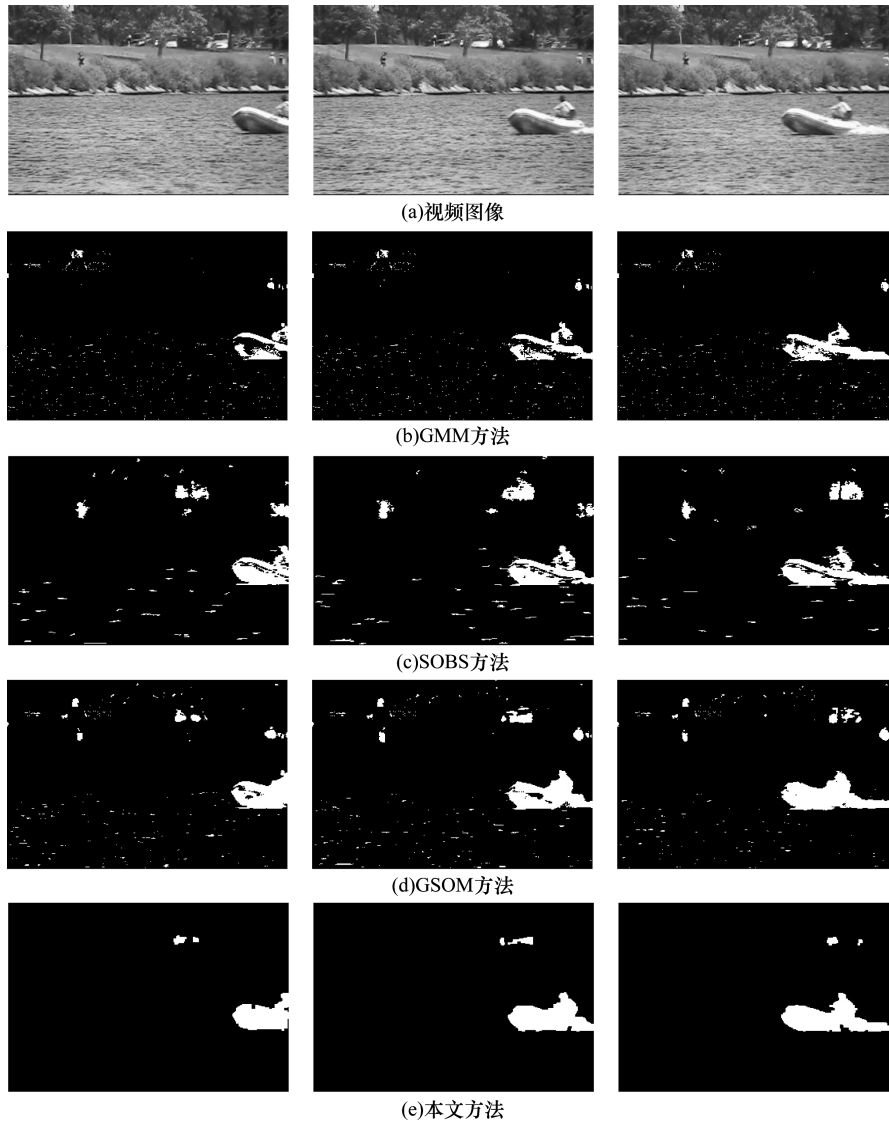


图1 动态背景下不同方法的运动目标检测结果

3.2 定量评价

本文采用运动目标检测领域国际上通用的 3 个评价指标^[15]来定量评价目标检测性能:

(1) 检索率 (Recall, RE), 正确检测到的目标点数量与目标点总数的比值, 值越大性能越好。

(2) 精确度 (Precision, PR), 正确检测到的目标点数量与检测到的目标点总数的比值, 值越大性能越好。

(3) 处理帧率 (Processing Frames Per Second, PFPS), 平均 1 s 内处理的帧数, 值越大越好。

表 1 给出了不同方法在动态背景测试子集下的目标检测性能指标。很明显, 本文方法的检索率和精确度指标都明显高于其他 3 种方法, 尤其是精确度指标比其他方法高 20% 以上, 说明本文方法对动态背景的适应能力强, 检测效果好。另外, 本文方法的处理帧率与 GSOM 方法相当, 尽管比 GMM 方法稍低, 但检索率和精确度指标要明显高于 GMM 方法, 因此, 总体评价本文方法的目标检测性能最优。

表 1 不同方法目标检测性能的定量评价

方法	RE	PR	PFPS
GMM 方法	0.71	0.52	152
SOBS 方法	0.79	0.59	124
GSOM 方法	0.82	0.61	139
本文方法	0.86	0.82	137

4 结束语

本文提出一种融合 SORM 背景模型和 DTCNN 阈值模型的运动目标检测方法。在背景模型构建方面, 采用 SORM 模型适应背景变化, 尤其是动态背景下的背景杂乱运动变化; 在运动目标分割方面, 改进了 DTCNN 模型, 用其构建阈值模型, 降低噪声对运动目标提取的干扰; 在参数自适应更新方面, 对背景模型和阈值模型中所用到的参数, 依据背景状态 (如动态背景或静态背景) 和目标状态 (如是否存在走走停停现象) 进行自适应更新, 以提高方法对不同场景的适应能力。实验结果证明了该方法的有效性。后续将针对摄像机抖动情况下的运动目标检测难题, 进一步完善背景模型和阈值模型, 拓展方法的应用范围。

参考文献

- [1] 崔智高, 李艾华, 冯国彦. 采用多组单应约束和马尔可夫随机场的运动目标检测算法[J]. 计算机辅助设计与图形学学报, 2015, 27(4): 621-632.
- [2] 甘明刚, 陈杰, 刘劲, 等. 一种基于三帧差分 and 边缘信息的运动目标检测方法[J]. 电子与信息学报, 2010, 32(4): 894-897.
- [3] 陈明生, 梁光明, 孙即祥, 等. 复杂背景下 H. 264 压缩域运动目标检测算法[J]. 通信学报, 2011, 32(3): 91-97.
- [4] Mukherjee D, Wu Q M J, Nguyen T M. Multiresolution Based Gaussian Mixture Model for Background Suppression [J]. IEEE Transactions on Image Processing, 2013, 22(12): 5022-5035.
- [5] Maddalena L, Petrosino A. A Self-organizing Approach to Background Subtraction for Visual Surveillance Applications [J]. IEEE Transactions on Image Processing, 2008, 17(7): 1168-1177.
- [6] 丁莹, 李文辉, 范静涛, 等. 基于 Choquet 模糊积分的运动目标检测算法[J]. 电子学报, 2010, 38(2): 263-268.
- [7] 张超, 吴小培, 吕钊. 基于独立分量分析的运动目标检测算法中对通道数选择和观测向量生成方式的实验和分析[J]. 电子与信息学报, 2015, 37(1): 137-142.
- [8] 侯旺, 于起峰, 雷志辉, 等. 基于分块速度域改进迭代运动目标检测算法的红外弱小目标检测[J]. 物理学报, 2014, 7(7): 149-161.
- [9] Marco P, Andrea V, Jordi G, et al. A Coarse-to-fine Approach for Fast Deformable Object Detection [J]. Pattern Recognition, 2015, 48(5): 1844-1853.
- [10] Xiao Jinwen, Wei Hui. Scale-invariant Contour Segment Context in Object Detection [J]. Image and Vision Computing, 2014, 32(12): 1055-1066.
- [11] 王思明, 赵伟. 亮度特征自相关和 GMM 相结合的目标检测[J]. 计算机工程, 2014, 40(5): 219-223.
- [12] 陈耀东, 李仁发, 李实英, 等. 面向目标检测与姿态估计的联合文法模型[J]. 计算机学报, 2014, 37(10): 2206-2217.
- [13] 陈耀东, 李仁发. 一种面向目标检测的部件学习方法[J]. 计算机研究与发展, 2013, 50(9): 1902-1913.
- [14] Ashish G, Ajoy M, Susmita G. Moving Object Detection Using Markov Random Field and Distributed Differential Evolution [J]. Applied Soft Computing, 2014, 15(2): 121-136.
- [15] Prasad R, Murthy C R, Rao B D. Joint Approximately Sparse Channel Estimation and Data Detection in OFDM Systems Using Sparse Bayesian Learning [J]. IEEE Transactions on Signal Processing, 2014, 62(14): 3591-3603.

编辑 刘冰