

基于多模式均值时空模型的目标融合检测方法

赵高鹏 薄煜明

(南京理工大学自动化学院, 南京 210094)

摘要 针对复杂环境下的目标检测问题,提出了一种基于背景模型的融合检测方法。首先在多模式均值模型的基础上,构造多模式均值时空模型,结合像素在时空域上的分布信息,改善了模型对非平稳场景较为敏感的缺点,给出了模型更新方法和前景检测方法;然后利用该模型对可见光和红外图像序列分别进行建模和前景检测,给出了一种基于置信度的目标融合检测方法,利用双传感器信息提高检测精度和可靠性。实验结果验证了本文方法的有效性。

关键词 目标检测 多模式均值时空模型 融合检测

中图分类号: TP391 **文献标志码:** A **文章编号:** 1006-8961(2010)08-1254-06

Target Fusion Detection Method Based on Spatiotemporal Multimodal Mean Model

ZHAO Gaopeng, BO Yuming

(Department of Automation, Nanjing University of Science and Technology, Nanjing 210094)

Abstract Target detection is difficult to be realized in complex scenes when there are moving background objects such as trees. In this paper, a new target fusion detection method is proposed based on background model. Firstly, by combining temporal information of per-pixel and the spatial information in the local region, we introduce a variant of multimodal mean model called the spatiotemporal multimodal mean model that is well suited for the non-stationary scenes. Then, the proposed background model is separately used to extract foreground pixels in visible and infrared image sequences, and a fusion detection method based on the confidence map is proposed to get the target detection result. The multi-sensor information can improve the detection precision and handle different environmental conditions. Experiment results demonstrate the effectiveness of the proposed method.

Keywords target detection, spatiotemporal multimodal mean model, fusion detection

0 引言

在智能视频监控、运动分析等应用中,从视频序列中检测运动目标是一个基础而又关键的任务。在室外复杂环境下,由于各种环境因素的影响,单一的传感器序列图像不能全面地反映目标的特性,难以有效地处理场景变化及干扰,有一定的局限性,如可见光图像传感器通常在低亮度环境中成像效果差;

红外传感器图像通常信噪比和对比较低,且目标与背景、干扰物差异不明显。因此,研究人员利用多个图像传感器来进行目标检测^[1-4],提高检测精度和系统适应场景变化的能力。

背景减方法在静态摄像机应用中是一种有效的方法,对此研究人员提出了许多不同的方法^[5-6]。文献[6]综述了已有的背景减算法,其中典型的算法有:中值滤波法^[7]采用最近 n 帧图像的中值作为背景模型,该方法适用于场景目标不大且运动连续

收稿日期:2009-10-12;改回日期:2010-12-18

第一作者简介:赵高鹏(1983—),男。现为南京理工大学博士研究生。主要研究方向为图像融合与目标检测跟踪。E-mail: zhaogaopeng@sina.com

的情况,对大而慢的运动目标效果不好,且算法需要 n 倍帧大小的内存;混合高斯法^[8]利用当前帧图像和参数化的历史帧信息来区分背景和前景像素,将每个像素按照多个高斯分布混合建模,该方法的模型参数可以自适应更新,不需要缓存图像帧,但混合高斯法资源消耗大,计算复杂。文献[9]针对嵌入式实时视频监控应用,提出了多模式均值模型方法,该方法比混合高斯法计算快速简单、存储需求小、效率更高。

背景建模方法在实际场景中遇到的主要问题是在自然场景中背景的动态变化,如摇晃的树叶、光照的变化等,当背景发生这些非平稳变化时,现有的背景减方法依然会产生大量的误判,需要大量的后处理来消除错误分类的像素点。

在已有研究的基础上,针对可见光和红外双传感器系统,提出了一种基于背景减方法的目标融合检测方法。首先在多模式均值模型方法的基础上,引入像素的空间域信息,给出了多模式均值时空模型,该方法有效地融合了每个像素的时间域和空间域分布信息,提高了模型性能,采用自适应模型参数更新方法,计算快速简单;在此基础上,提出了一种基于像素置信度的融合检测方法,利用双传感器的冗余和互补特性,得到精确鲁棒的检测结果。算法流程如图 1 所示,实验结果表明了本文方法的有效性。

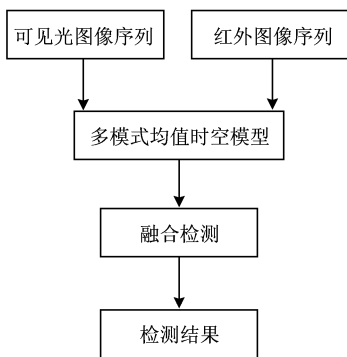


图 1 本文方法流程图

Fig. 1 Flowchart of the proposed method

1 多模式均值时空模型

背景减方法的基本思想是通过比较当前帧与背景模型直接的差别来实现对前景像素的检测,其关键是背景模型的建模和更新。背景建模是通过构建

某种模型来表示背景,背景更新是为了使背景适应场景的各种变化和干扰。由于多模式均值模型可以表示多模分布的背景,且具有存储量小、计算简单等优点,本文方法在此基础上,考虑像素的时间域和空间域信息,构造本文的多模式均值时空模型,给出了模型建模更新方法及前景检测方法。

设图像序列第 t 帧一个像素点表示为 I_t , 如果仅采用像素的灰度信息,则 I_t 表示该像素点的灰度值;如果采用像素的颜色信息,如 RGB(红、兰、绿)或 HIS(hue, saturation, intensity)颜色信息,则将该像素点的每个颜色分量表示为 $I_{t,x}$,如 $I_{t,r}$ 表示像素点 I_t 在 RGB 空间的红色分量值。以下采用颜色信息进行说明。

对图像序列帧的每个像素点逐一建立模型。设序列第 t 帧某一给定位置的像素点 I_t , 其相应的模型参数包括表示该像素点当前帧局部邻域内像素值的平均值 a ; 用来描述该像素点时域历史帧信息的均值分布模式 $m_i (1 \leq i \leq K)$, K 的值通常取 3 ~ 7。

参数 a 的值为向量,分别对应像素局部邻域均值的每个颜色分量,计算如下:

$$a_{t,x}(u,v) = \frac{1}{N} \sum_{n \in N} I_{t,x}(u+n, v+n) \quad (1)$$

式中, (u, v) 表示像素点 I_t 的位置, $x \in (r, g, b)$, $N = n \times n$ 表示以像素点 I_t 为中心的 $n \times n$ 邻域像素个数; $I_{t,x}(u+n, v+n)$ 表示相应位置处像素点的各颜色分量值; $a_{t,x}$ 表示像素点 I_t 局部邻域内各颜色分量的均值。与单个像素相比,像素的局部邻域能够更多地反映空间分布信息,使得基于局部邻域均值的空间域模型对场景的局部变化不敏感。

每个均值分布模式 m_i 定义为下列参数值的集合: c, s, r_1, r_2 。其中, c 表示像素值与模式 m_i 匹配的次数量; s 为向量,值对应像素值每个颜色分量值的和; r_1 和 r_2 表示在最近的历史帧中像素与模式 m_i 匹配的次数量,用于表示模式的优先级。因此,对于第 t 帧,均值分布模式各颜色分量的均值定义为

$$\mu_{i,t,x} = s_{i,t,x} / c_{i,t} \quad (2)$$

式中, $s_{i,t,x}$ 表示模式 m_i 在 t 帧中每一个颜色分量值的和; $c_{i,t}$ 表示在 t 帧中像素值与模型 m_i 匹配的次数。

1.1 模型参数更新

随着场景的变化,需要对模型进行不断地学习

更新。对于新的图像帧,设为第 t 帧 I_t ,本文模型的更新过程如下所述。

对空间域模型参数 \mathbf{a} 进行更新,为了避免参数对场景变化过于敏感,采用加权平均方法,如下式:

$$a_{t,x} = (1 - \beta) \cdot a_{t-1,x} + \beta \cdot a_{t,x} \quad (3)$$

式中, β 为空间域模型参数更新系数,其取值范围为 $\beta \in [0, 1]$ 。

对均值分布模式 \mathbf{m}_i 的参数更新采用以下方法。将像素值与该像素点已有的 $k(k \leq K)$ 个模式逐个进行匹配,如果像素点值满足下式的匹配判断条件:

$$\bigcap_x |I_{t,x} - \mu_{i,t-1,x}| \leq E_x \quad (4)$$

式中, E_x 为预定义的距离阈值,表明该像素点与模式 \mathbf{m}_i 匹配,则调整模式 \mathbf{m}_i 的参数,如下式所示:

$$\begin{cases} s_{i,t,x} = s_{i,t-1,x} + I_{t,x} \\ c_{i,t} = c_{i,t-1} + 1 \end{cases} \quad (5)$$

逐个模式进行匹配判断,如果已有的 k 个分布模式均不满足式(4)条件,表明该像素点与已有的分布模式不匹配,如果 $k < K$,则增加一个新的分布模式;如果 $k = K$,则用一个新的分布模式代替已有分布模式中优先级最低的一个模式。新的分布模式参数初始化为 $s_{i,t,x} = I_{t,x}, c_{i,t} = 1, r_1 = r_2 = 0$ 。

均值分布模式的优先级采用模式 \mathbf{m}_i 的参数 r_1 和 r_2 来度量。定义一个时域滑动窗口,包括 w 帧图像。 r_1 表示当前窗口像素与模式 \mathbf{m}_i 匹配的次数,在滑动窗口开始初始化为 0,当像素与模式 \mathbf{m}_i 匹配时, r_1 的值增加 1; r_2 表示在前一个窗口中 r_1 的最大值;参数 r_1 和 r_2 更新过程如下:

$$r_{1,i,t} = \begin{cases} 0 & t \bmod w = 0 \\ r_{1,i,t-1} + 1 & \text{式(4)条件满足且} \\ & t \bmod w \neq 0 \\ r_{1,i,t-1} & \text{其他} \end{cases} \quad (6)$$

$$r_{2,i,t} = \begin{cases} r_{1,i,t-1} & t \bmod w = 0 \\ r_{2,i,t-1} & \text{其他} \end{cases} \quad (7)$$

定义变量 $R_{i,t} = r_{1,i,t} + r_{2,i,t}$,它表示在最近的历史帧中模式 \mathbf{m}_i 匹配的一种度量,值越大,表明该模式的优先级越高。当用一个新的分布模式代替一个已有的分布模式时,选择满足 $R_{i,t} < w/K$ 条件的子集中 $c_{i,t}$ 最小值对应的模式进行替换。当满足 $R_{i,t} < w/K$ 条件的子集为空时,选择所有 $c_{i,t}$ 值中最小值对应的模式进行替换。

对于长时间背景模型更新,为了避免数据溢出,

预定义值 d 每隔 d 帧将每个模式的参数进行如下更新:

$$\begin{cases} s_{i,t,x} = s_{i,t-1,x}/2 \\ c_{i,t} = c_{i,t-1}/2 \end{cases} \quad t \bmod d = 0 \quad (8)$$

1.2 前景检测

对于新的图像帧 I_t ,利用已建立的背景模型逐个像素点进行比较来区分前景像素和背景像素。判断一个像素点为背景像素点或是前景像素点,需要满足式(4)及以下两个条件:

$$c_{i,t-1} > T_{FG} \quad (9)$$

$$\bigcap_x |a_{t,x} - a_{t-1,x}| < T_{avr} \quad (10)$$

式中, T_{FG} 为阈值,如果 $c_{i,t-1}$ 的值大于该阈值,则表明该像素点的第 i 个分布模式 \mathbf{m}_i 表示背景;否则表明第 i 个分布模式 \mathbf{m}_i 表示前景; $x \in (r, g, b)$; T_{avr} 为阈值,表示空间域模型对背景局部变化的容忍程度,如果变化量小于阈值,表明该像素点属于背景,否则表明该像素点属于前景。

式(4)和式(9)反映了时间域像素值信息的变化,根据式(4)和式(9)逐个比较该像素点已有的 $k(k \leq K)$ 个分布模式;式(10)反映了空间域当前帧局部邻域均值与历史帧信息的变化量,当判断条件式(4)、(9)和(10)同时不成立时,认为该像素点是前景像素点,否则,该像素点被认为是背景像素点。

2 目标融合检测

多传感器融合检测是指根据多传感器检测的结果进行进一步的处理以提高对目标的检测性能和系统对环境变化的适应能力,特别是在某一个传感器受到强干扰时,多传感器融合检测可以维持较好的检测性能。常见的像素级融合检测方法为对单个传感器的检测结果进行二值化,然后进行逻辑与或者逻辑或判断。与、或逻辑各有优缺点,与逻辑有效地降低了虚警数,但漏检数随之提高;或逻辑能够有效地降低漏检数,但提高了虚警数。

本文在前述多模式时空模型的基础上提出了一种基于置信度的融合检测方法。对于新的图像帧 I_t ,逐个像素按照 1.2 节前景检测方法进行判断,因此可得到前景图像。对前景图像计算置信度,然后对可见光和红外图像进行基于置信度测量的融合检测,得到最终的检测结果,具体如下。

设前景像素点的位置为 (u, v) , 定义变量 $D_i(u, v)$, 表示当前帧像素值与模型背景均值间距离,

$$D_i(u, v) = \frac{1}{M} \sum_{j=1}^M \left(\sum_x |I_{i,t,x} - \mu_{i,t-1,x}| \right) \quad (11)$$

式中, $x \in (r, g, b)$, M 表示该像素点模式 m_i 中表示背景的模式个数, 模式 m_i 属于背景的判断条件如式(9)所示。 $D_i(u, v)$ 值越小, 表明当前像素属于背景的可能性越高, 其值越大, 表明当前像素属于前景的可能性越高。

对可见光序列和红外序列的当前帧图像分别计算 D_i 值图像, 然后分别将其归一化, 得到目标置信度图像 C_i , 如下式:

$$C_i^{ch}(u, v) = \frac{D_i^{ch}(u, v)}{\max(D_i^{ch}(u, v))} \quad (12)$$

式中, $ch \in \{TV, IR\}$, TV 为可见光序列, IR 为红外序列, C_i 值的范围为 $[0, 1]$ 。

融合检测运算即对两个图像序列当前帧得到的置信度图像 C_i 进行融合, 融合规则如下式:

$$C_i^{fus}(u, v) = \max\{C_i^{ch}(u, v)\} \quad (13)$$

$$ch \in \{TV, IR\}$$

C_i^{fus} 为融合后的置信度图像, 满足 $C_i^{fus} > T_{fus}$ 的

认为是最终双传感器融合检测的结果, T_{fus} 为阈值, 本文实验取为 0.1。单一传感器的检测精度和鲁棒性受传感器特性、输出图像质量和环境因素的影响, 因此, 利用可见光和红外双传感器进行融合检测, 能够提高检测精度和系统对环境的适应能力, 扩大系统的工作时间和范围。

3 实验结果及分析

为了验证本文方法的有效性, 在 Matlab 环境下进行了仿真实验。机器配置为 Pentium 1.6 GHz, 1 GB 内存。

实验 1 对本文多模式均值时空背景模型 (STMM) 进行了测试, 并与文献[8]提出的经典混合高斯背景模型方法 (MoG)、文献[9]提出的多模式均值背景模型方法 (MM) 进行了比较, 如图 2 所示。视频序列采用标准测试视频 Wavetrees 序列, 长 287 帧, 大小为 160×120 像素, 包含了动态变化的背景, 如摇晃的树。测试帧为序列第 247 帧, 所有检测结果均未作后处理。

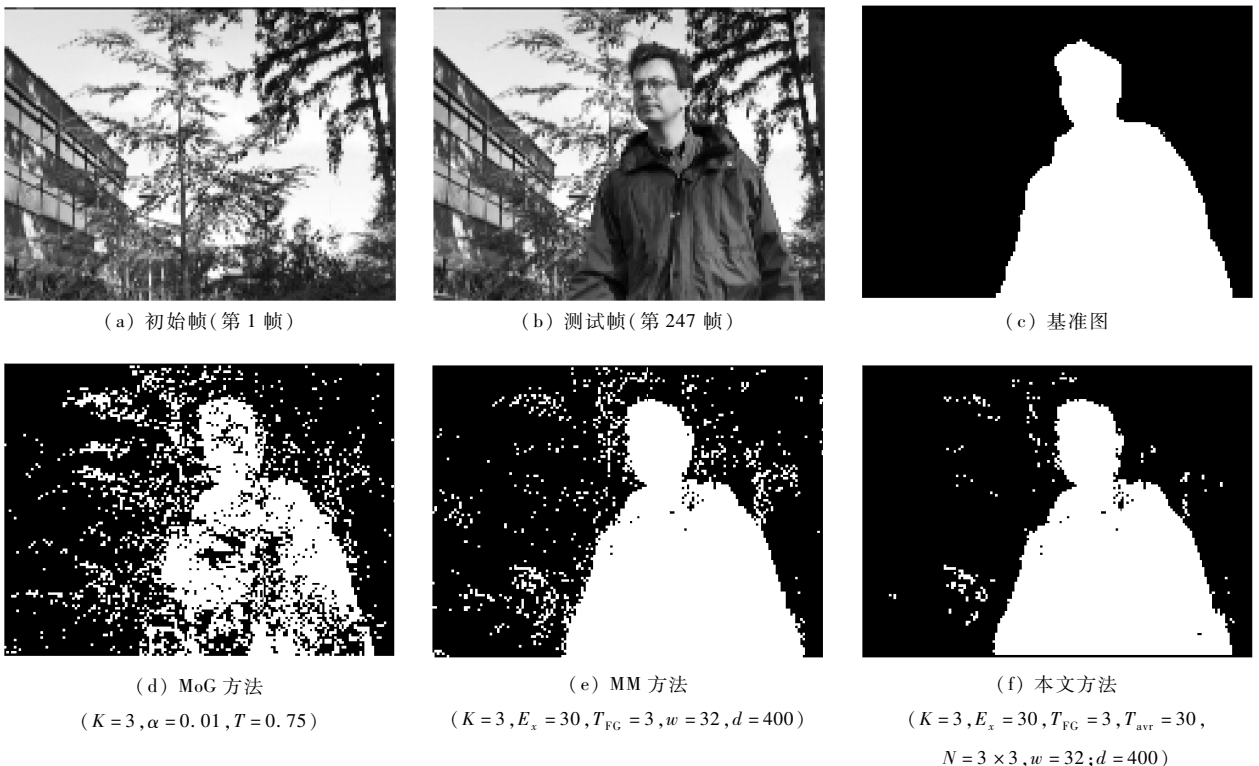


图 2 运动目标检测结果

Fig. 2 Detection results of different methods

从图 2 可以看出,以上 3 种算法对背景运动都有一定的自适应性,但由于背景运动的不规则,依然产生了错误分类,从目视结果来看,MoG 方法产生的虚警最多,MM 方法次之,本文方法最少;图 3 给出了 3 种算法检测结果的定量评价,评价指标采用文献[10]的方法,分别统计了 3 种算法的虚警数 FP 和漏检数 FN,虚警数指背景被检测为前景像素个数,漏检数是指前景被检测为背景的像素个数。图 3 横坐标依次表示 MoG、MM 和本文方法,从图 3 可以看出,本文方法在保持较低漏检数的同时,有效地降低了虚警数,准确地提取到前景目标。

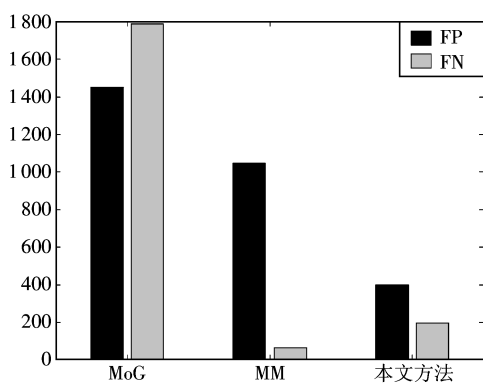


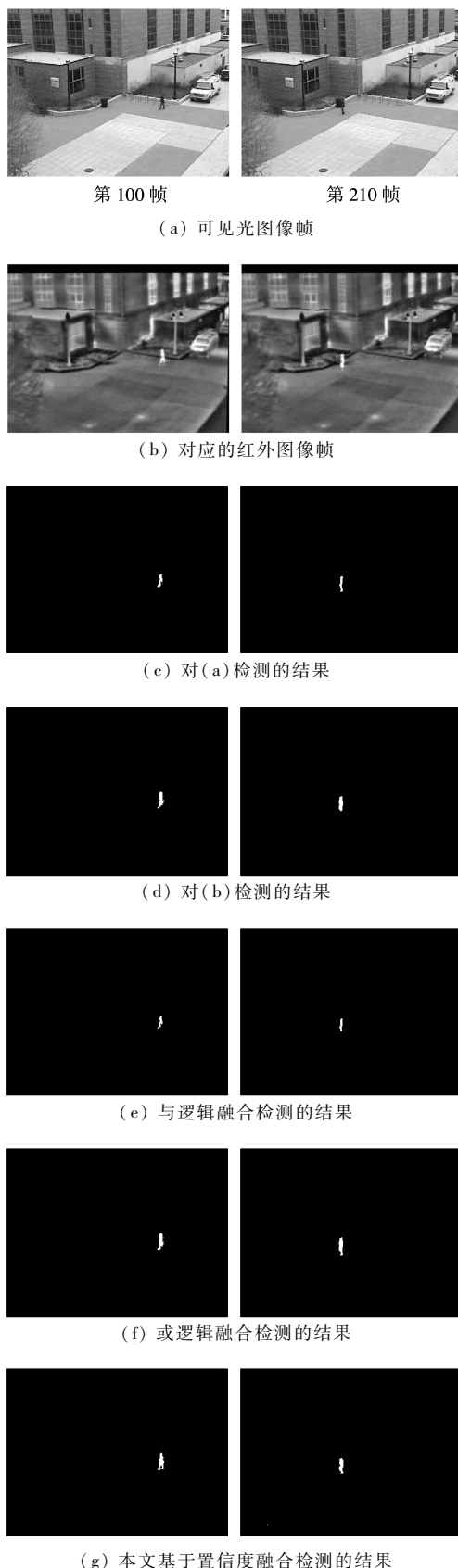
图 3 检测结果的定量比较

Fig. 3 Comparison of different methods

从算法的运行速度来看,MM 方法的速度最快,MoG 方法最慢,文献[9]的研究结论表明 MM 方法处理速度是 MoG 方法的 6 倍,本文 STMM 方法由于增加了空间域信息处理,使得处理速度要低于 MM 方法,但远高于 MoG 方法。在 Matlab 环境下的仿真结果表明,本文方法的处理速度约为 15 帧/s。如果采用 DSP 或 FPGA 等硬件实现,则可以进一步提高处理速度。

实验 2 对本文提出的在 STMM 方法基础上基于置信度的融合检测进行测试,视频序列来源于 OTCBVS 数据库中 Dataset 03 图像序列,大小为 320×240 像素,序列中目标为行人。实验与常用的与、或逻辑融合方法进行了比较,图 4 给出了序列中两帧图像及检测结果。

从图 4 可以看出,对可见光和红外图像序列,本文 STMM 方法均有效地检测到了运动目标;第 210 帧,可见光图像中目标部分与背景相似,可见光图像背景减的检测结果不够完整,而红外图像背景减的检测结果比较完整,利用融合的方法能够提高检测的精度和可靠性,相比于与、或逻辑,本文基于置信



(g) 本文基于置信度融合检测的结果

图 4 融合检测结果

Fig. 4 Fusion detection results

度的融合检测效果更好。需要说明的是,本文方法要求可见光和红外图像序列经过配准。图像配准属于另一个研究方向,在此不做讨论,本文实验序列已经过手动配准。

4 结 论

本文针对可见光和红外双传感器系统应用,提出了一种基于多模式均值时空模型的目标融合检测方法。多模式均值时空模型保留了多模式均值模型计算简单、快速等优点,同时利用了像素在时间域和空间域的分布信息,改善了模型对非平稳场景背景变化较为敏感的缺点。与仅利用单一传感器检测相比,利用双传感器信息,基于置信度的像素级融合检测算法取得了更好的检测效果,提高了检测精度和可靠性。实验结果显示了本文多模式均值时空模型和融合检测方法的有效性。

参考文献 (References)

- [1] Bondzolic B, Petrovic V. Multisensor background extraction and updating for moving target detection [C] // Proceedings of 2008 11th International Conference on Information Fusion. Piscataway, NJ, USA: IEEE Press, 2008: 1-8.
- [2] Han J, Bhanu B. Fusion of color and infrared video for moving human detection [J]. Pattern Recognition, 2007, 40(6): 1771-1784.
- [3] Davis J W, Sharma V. Background-subtraction using contour-based fusion of thermal and visible imagery [J]. Computer Vision and Image Understanding, 2007, 106: 162-182.
- [4] Chen Y X, Han C Z. Night-time pedestrian detection by visual-infrared video fusion [C] // Proceedings of 2008 7th World Congress on Intelligent Control and Automation. Piscataway, NJ, USA: IEEE Press, 2008: 5079-5084.
- [5] Radke R J, Andra S, Al-Kofahi O, et al. Image change detection algorithms: a systemic survey [J]. IEEE Transactions on Image Processing, 2005, 14(3): 294-307.
- [6] Dai Kexue, Li Guohui, Tu Dan, et al. Prospects and current studies on background subtraction techniques for moving objects detection from surveillance video [J]. Journal of Image and Graphics, 2006, 11(7): 919-927. [代科学, 李国辉, 涂丹, 等. 监控视频运动目标检测减背景技术的研究现状和展望 [J]. 中国图象图形学报, 2006, 11(7): 919-927.]
- [7] Lo B P, Velastin S A. Automatic congestion detection system for underground platforms [C] // Proceedings of International Symposium on Intelligent Multimedia, Video, and Speech Processing. Piscataway, NJ, USA: IEEE Press, 2001: 158-161.
- [8] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking [C] // Proceedings of the Computer Society on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE Press, 1999: 246-252.
- [9] Apewokin S, Valentine B, Wills L, et al. Multimodal mean adaptive background for embedded real-time video surveillance [C] // IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, NJ, USA: IEEE Press, 2007: 1-6.
- [10] Heikkila M, Pietikainen M. A texture-based method for modeling the background and detecting moving objects [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4): 657-662.