基于长时间视频序列的背景建模方法研究

丁 洁 [1, 2] 肖江剑 [2] 况立群 [1] 宋康康 [2] 彭成斌 [2]

摘 要 针对现有背景建模算法难以处理场景非平稳变化的问题, 提出一种基于长时间视频序列的背景建模方法. 该方法包括训练、检索、更新三个主要步骤. 在训练部分, 首先将长时间视频分段剪辑并计算对应的背景图, 然后通过图像降采样和降维找到背景描述子, 并利用聚类算法对背景描述子进行分类, 生成背景记忆字典. 在检索部分, 利用前景像素比例设计非平稳状态判断机制, 如果发生非平稳变换, 则计算原图描述子与背景字典中描述子之间的距离, 距离最近的背景描述子对应的背景图片即为此时背景. 在更新部分, 利用前景像素比例设计更新判断机制, 如果前景比例始终过大, 则生成新背景, 并更新背景字典以及背景图库. 当出现非平稳变化时 (如光线突变), 本算法能够将背景模型恢复问题转化为背景检索问题, 确保背景模型的稳定获得. 将该框架与短时空域信息背景模型 (以 ViBe、MOG 为例) 融合, 重点测试非平稳变化场景下的背景估计和运动目标检测结果. 在多个视频序列上的测试结果表明, 该框架可有效处理非平稳变化, 有效改善目标检测效果, 显著降低误检率.

关键词 背景建模, 长周期视频, 背景图描述子, 背景检索

引用格式 丁洁, 肖江剑, 况立群, 宋康康, 彭成斌. 基于长时间视频序列的背景建模方法研究. 自动化学报, 2018, **44**(4): 707−718

DOI 10.16383/j.aas.2017.c160468

Background Modeling for Long-term Video Sequences

DING Jie[1, 2] XIAO Jiang-Jian[2] KUANG Li-Qun[1] SONG Kang-Kang[2] PENG Cheng-Bin[2]

Abstract Considering the difficulties to deal with scene non-stationary variation of proposed background modeling methods, we propose a method for moving targets by exploiting periodic spatial-temporal feature from a long-term video. We use three steps, training, retrieval and updating, to establish a background modeling framework for long-term video sequences. In the training step, we cut hours of video into a number of minute clips and compute the average background to generate a series of background images. After performing resize and dimension reduction on background images, a set of descriptors are obtained for the clustering process, where background descriptors are classified into different clusters and each cluster is represented by a typical background image in the background memory dictionary. In the retrieval step, we use foreground pixel ratio as a criterion to determine sudden change of background. For those scenarios, the current image is converted to a background descriptor and compared to the descriptors stored in retrieval database to find a suitable background frame. If no similar background descriptor is found in the database, a new background image is to be generated and added into our dictionary and background image database. Using this framework, the background modeling problem is converted to a background retrieval problem when non-stationary change happens especially for the indoor scene with quick illumination changes such as light on/off. Combining the popular ViBe or MOG algorithm with our framework, we test a number of long term video sequences and achieve better results in terms of tracking targets and the false detection rate.

Key words Background modeling, long-term video, background image descriptor, background retrieval

Citation Ding Jie, Xiao Jiang-Jian, Kuang Li-Qun, Song Kang-Kang, Peng Cheng-Bin. Background modeling for long-term video sequences. *Acta Automatica Sinica*, 2018, **44**(4): 707−718

收稿日期 2016-06-15 录用日期 2016-11-23
Manuscript received June 15, 2016; accepted November 23, 2016
国家自然科学基金 (61379080, 61273276), 浙江省杰出青年基金 (LR13F020004), 国家科技支撑计划 (2015BAF14B01), 钱江人才计划 (QJD1702031), 和中国博士后科学基金 (2017M612047) 资助
Supported by National Natural Science Foundation of China (61379080, 61273276), Excellent Youth Foundation of Zhejiang Scientific Committee (LR13F020004), National Key Technology R&D Program (2015BAF14B01), Qianjiang Talent Program (QJD1702031), China Postdoctoral Science Foundation (2017M612047)
本文责任编委 桑农
Recommended by Associate Editor SANG Nong
1. 中北大学计算机与控制工程学院 太原 030051 2. 中国科学院宁波工业技术研究院计算机视觉团队 宁波 315201
1. Computer and Control Engineering, North University of China, Taiyuan 030051 2. Computer Vision Group, Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo 315201

背景建模是计算机视觉的一个重要研究方法, 在智能视频监控、智能交通、人机交互等领域有广泛应用. 现有背景模型主要分为基于时域信息的模型和基于时空域信息融合的模型[1]. 基于时域信息的模型通常利用过去一小段时间内像素的统计特性来预测该像素短期未来的状态, 而基于时空域信息融合的模型在利用时域信息的同时也关注像素在空间域上的分布特性. 这些模型又可以分为参数化模型和非参数化模型. 参数化模型是利用含参模型对

每个像素点建模, 非参数化模型是使用已观察的像素值对该像素点建模[2].

Wren 等[3] 提出的单高斯背景模型是利用时域信息建立的参数化模型, 该方法对光照缓变适应性较强, 但在发生背景扰动时, 处理情况较差, 这主要是因为单高斯背景模型无法处理多模态变化. 此后, Stauffer 等[4] 提出混合高斯背景模型 (Mixture of Gaussian, MOG) 来处理多模态变化, 它也是一个只利用时域信息的参数化模型. 与单高斯模型不同的是, 它对图像每个像素点建立多个不同权重的高斯模型. 它可以有效地处理多模态场景, 但是如果背景中同时呈现高低频变换, 它的灵敏度调节困难, 会导致前景像素融入背景模型、丢失高频目标. 另外, 条件随机场[5]、码书[6] 等方法也被用于基于时间域信息的背景建模, 然而发生变化 (如风吹树枝) 时, 受模型更新速度的影响, 算法会产生大量的虚警数. 此后, Barnich 等[7] 通过利用像素的空间关系提高模型更新速度, 提出融合时空特性的非参数化模型—ViBe (Visual background extractor) 模型, 该模型利用像素点的邻居像素来对模型更新, 使其对变化场景可以较快适应. 然而在非平稳变化 (如光照突变) 下, 使用该模型仍然会产生大量的虚警数. 2014 年, St-Charles 等提出 SuBSENSE (Self-balanced sensitivity segmenter) 算法[8], 该算法对 ViBe 算法颜色空间以及距离公式进行改进, 可以有效填补 ViBe 算法目标内部空洞并提高更新速率, 然而该算法运算效率较低且容易出现大范围闪烁现象.

分析以上背景建模方法, 无论是只使用时间域信息的模型还是使用时空域信息融合的模型都只考量狭小时间段内的统计特性. 然而, 在整个背景建模的过程中, 场景背景的变化有周期性重现的特点 (如光线的变化情况等), 如果仅在小时段时空域上研究, 必定会丢失周期性信息, 使更新受限. 如果将周期性信息合理记录, 构成带记忆的模型. 在发生非平稳变化时, 直接在记忆字典中找到对应变化特点的背景作为此时背景, 并使用它更新模型, 必定能大幅度降低虚警数. 为了充分利用背景长时间周期性重现特点, 搭建一个合理融合大时空域信息的基于长时间视频序列的背景建模框架, 并在其上研究背景建模方法. 设计该框架时有以下几个难点: 1) 如何将大量长时间背景信息合理描述; 2) 如何训练生成简单并兼顾实时性的背景字典; 3) 如何在背景词典中查找所需背景; 4) 如何使背景字典长久的适用于场景; 5) 如何将长时间的时空域信息与短时间的时空域信息结合, 即如何将长时间记忆模型与短时间记忆模型融合.

针对 1), 本文通过对长时间视频剪辑、求平均背景生成背景图片, 并对图像降采样、降维[9], 产生有意义的背景描述子; 针对 2), 本文采用谱聚类[10] 对背景粗分类, 并使用 K-means[11] 对背景进一步细分类, 使用类别中典型图建立树形字典, 从而训练出简单可兼顾实时性的背景字典; 针对 3), 计算原图向量与背景词典向量之间的欧氏距离, 距离小的即为所需背景; 针对 4), 本文增加背景字典更新模块; 针对 5), 本文设计突发变化判断机制, 如果是平稳变化则使用现有短时空域信息模型, 如果是突发变化则利用带记忆的长时空域信息模型.

本文首先介绍该框架的建立方法, 然后介绍该框架与短时空域信息背景模型[12] 的融合方法, 重点测试突发变化发生时的运动目标检测结果. 实验结果表明: 该框架可显著提高背景模型 (如 ViBe 或 MOG 算法) 对突发变化 (主要测试光照突变) 的适应性和鲁棒性, 有效实现对前景目标的较准确检测.

1 长时间背景建模框架

本文以长时间定视角视频序列为研究对象, 给出同时满足运动目标检测实时性、准确性以及突发变化适应性 (如光照突变) 要求的长时间背景建模框架.

长时间背景建模框架如图 1 所示, 其内容可以分为三块: 背景字典训练模块、图像检索模块以及背景字典更新模块. 背景字典训练模块包括视频背景信息描述 (预处理与 PCA (Principal components analysis) 降维) 和生成背景字典 (谱聚类、K-means 再聚类以及字典生成); 图像检索模块包括非平稳变化判断、原图像合理描述与检索判断方法; 背景字典更新模块包括模型效果判断机制与更新方法. 下文将围绕以上三个模块展开.

2 训练背景字典

训练背景字典部分包括背景合理描述与生成背景字典两部分. 这部分将完成背景模型的记忆功能.

2.1 长时间视频合理描述

这部分本质为特征提取, 通过对长时间视频预处理以及降维, 生成背景描述子, 并以向量的组合描述长时间视频图像序列.

2.1.1 长时间视频预处理

根据长时间视频数据量大, 而每一分钟背景变化差异不大的特点, 对采集的定视角视频做预处理.

输入: 所采集的定视角视频 (本文采集 24 小时定视角视频).

输出: 预处理结果向量集 $\{\boldsymbol{x}_i\}$.

步骤 1. 将视频剪辑为一分钟短视频 ($24 \times 60 = 1\,440$ 个);

步骤 2. 依次对每一分钟的视频使用已有背景建模方法建立背景模型并求得背景 (本文使用高斯背景建模算法来建立背景模型);

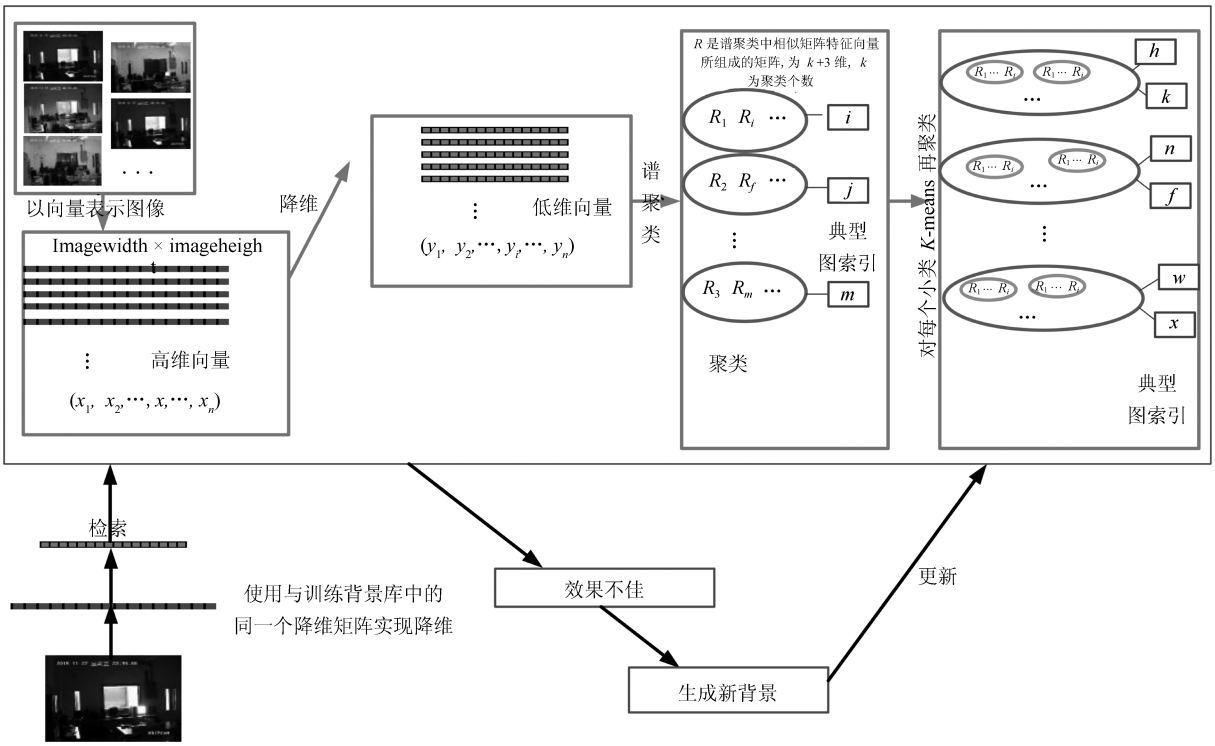


图 1　长视频背景建模框架

Fig. 1　Long time background modeling framework

步骤 3. 求每分钟的平均背景图 (共 1 440 张, 即背景记忆库);

步骤 4. 背景图像降采样, 主要目的是减小训练算法运算量. (将原图像 (Data1 分辨率 352×288) 变为分辨率 160×120 的图像);

步骤 5. 将图像转换为向量形式, 此后运算都以该向量集为基础. ($1 \times 160 \times 120$ 维的向量集 $\{\boldsymbol{x}_i\}$).

2.1.2　降维

通过预处理所得高维向量数据集为 $\{\boldsymbol{x}_i\}$, 在其之上直接处理, 会造成维数灾难[13], 因此使用降维算法对其降维. 本文采用被广泛使用的主成分分析法 (PCA) 对数据集降维, 主成分分析法的优点是概念简单、计算方便、重构误差小.

使用 PCA 算法, 计算合适的投影矩阵 U_d, 将图像数据集合 $\{\boldsymbol{x}_i\}$ 降到低维空间变为 $\{\boldsymbol{y}_i\}$, 计算公式:

$$\boldsymbol{y}_i = U_d(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \tag{1}$$

其中, $\boldsymbol{y}_i$ 是 $\boldsymbol{x}_i$ 降维后对应的向量, $\bar{\boldsymbol{x}}$ 是 $\{\boldsymbol{x}_i\}$ 的均值向量. 在降维后的空间, 背景数据集变为 $\{\boldsymbol{y}_i\}$, 它就是背景描述子. 降维的维数是通过保留信息量以及聚类结果确定的, 其确定方法在第 6 节阐述.

2.2　生成背景字典

这部分主要阐述背景记忆库中的向量分类方法、背景字典生成方法以及组织方式. 因为本文处理的定视角视频序列有如下特点: 1) 数据量较大; 2) 场景典型类别少. 本文利用聚类算法探索背景向量之间的关系并分类. 谱聚类算法对背景向量粗聚类, K-means 算法对背景向量细聚类. 与此同时, 使用类中典型图生成背景字典, 并根据粗细分类合理组织背景字典.

2.2.1　谱聚类

由于谱聚类算法有对不规则误差数据不敏感, 计算复杂度较小, 收敛于全局的优点, 本文使用该算法对数据聚类[14]. 2014 年, Zhu 等[15] 提出一种通过有效计算高维复杂数据之间相似度以改进相似度矩阵的方法, 大幅度提高高维谱聚类性能. 本文使用该方法计算相似度矩阵.

本文谱聚类流程:

输入: 背景描述向量集 $\{\boldsymbol{y}_i\}$.

输出: 聚类结果向量 $\boldsymbol{H}$ (指明每个向量的类别).

步骤 1. 计算这 n 个描述向量的相似度矩阵 $a_{n\times n}$, 其元素 a_{ij} 为数据 $\boldsymbol{y}_i$ 与 $\boldsymbol{y}_j$ 的相似度;

步骤 2. 计算矩阵 D, D 为对角矩阵, 除对角元素外都为 0, D 的对角元素为

$$D_{ij} = \sum_{i=1}^{n} a_{ij} \tag{2}$$

其中, D 的对角元素为 $a_{n\times n}$ 对应列的所有元素之和;

步骤 3. 计算规范拉普拉斯矩阵 L, 其中 I 是单

位矩阵;

$$L = I - D^{-\frac{1}{2}} a D^{-\frac{1}{2}} \tag{3}$$

步骤 4. 求 L 的特征值并按从小到大排列: $\gamma_1 \le \gamma_2 \le \cdots \le \gamma_n$ (对称矩阵有 n 个实值的特征值);

步骤 5. 对于 k 类聚类 (k 的选择由第 6 节阐述), 原算法选取前 k 个特征值所对应的特征向量, 按列组成新的矩阵 R, 它是 $n \times k$ 维矩阵, 本文算法根据经验选取前 $k+3$ 个特征值对应的特征向量, 按列组成新的矩阵 R, 它为 $n \times (k+3)$ 维矩阵;

步骤 6. 把矩阵 R 的每行元素作为新数据 (共 n 个, 每个数据 $k+3$ 维), 使用 K-means 聚类. 如果 R 的第 i 行元素被聚类到子类 K_j, 那么原 n 个数据中的第 i 个数据属于子类 j.

本文计算相似度矩阵 $a_{n\times n}$ 的方法 (由 Zhu 等[15] 提出) 如下:

如图 2 所示, γ 为根节点. 假如一对样本 $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ 从根节点开始直到到达它们各自的叶子节点 l^i 与 l^j. 最后由根节点、中间节点、叶子节点组成的一条路径 (如图粗体部分所示) 会被生成.

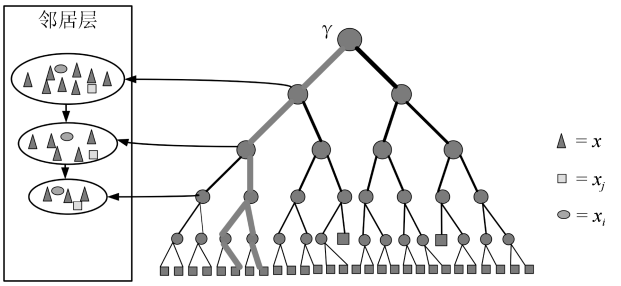


图 2 随机决策树

Fig. 2 Random decision tree

$$\begin{cases} p^i = \{\gamma, S_1^j, \cdots, S_k^j, \cdots, l^i\} \\ p^j = \{\gamma, S_1^j, \cdots, S_k^j, \cdots, l^j\} \end{cases} \tag{4}$$

S_k^i 和 S_k^j 分别表示 $\boldsymbol{x}_i$ 与 $\boldsymbol{x}_j$ 遍历的中间节点. 如果 p^i 和 p^j 经过相同的 λ 个节点, 则

$$\begin{cases} S_k^i = S_k^j, & k = 1, \cdots, \lambda \\ S_k^i \ne S_k^j, & k = \lambda+1, \cdots \\ l^i \ne l^j \end{cases} \tag{5}$$

$(\boldsymbol{x}_i, \boldsymbol{x}_j)$ 的相似度表示为

$$a_{ij} = \frac{\sum\limits_{k=1}^{\lambda} \frac{1}{|S_k|}}{\sum\limits_{m} \left(\frac{1}{|\frac{1}{S_m^b}|} \right) + \frac{1}{|\boldsymbol{A}^b|}} \tag{6}$$

其中, $b = \arg\max |p^b|$ 且 $b \in \{i, j\}$, $\boldsymbol{A}^b$ 表示到达叶子节点 l^b 的数据样本集, 分子表示 i, j 共同经过的权重和, 分母为整体权重. 这种表达方式可以有效表达数据点之间的相似性. 由 a_{ij} 构成的矩阵即为相似度矩阵 $a_{n\times n}$.

2.2.2 *K*-means 再聚类

根据上一部分的谱聚类算法, 背景图片可以分为 k 类, 类中的图片相似度较高. 由于背景图片量大, 假如直接使用新的视频图像向量与 k 类中每个背景向量比较则计算量太大. 而如果只与该类典型图片向量比较, 则比较向量太少, 会导致检索出的背景不够准确. 因而, 本文通过对每类向量 (由第 2.2.1 节可知, 该向量为 $k+3$ 维) K-means 再聚类, 聚为 10 个小类. 这样背景描述向量就又被分为 10 类.

2.2.3 背景字典的建立

建立的背景字典需满足两个要求: 第一, 能有效代表所有背景; 第二, 检索速度快. 针对这两个要求, 设计如图 3 所示的字典生成方法, 由第 2.2.1 节可知, 谱聚类将背景图聚为 k 类, 我们找到这 k 类的典型图, 之后再按第 2.2.2 节中 K-means 再聚类, 分别找到每部分 10 个类的典型图. 典型图是每类的载体, 背景字典由图 3 中浅色箭头虚线所连典型图构成. 由图 3 可知, 此背景字典为树形结构, 因而可加快检索速度.

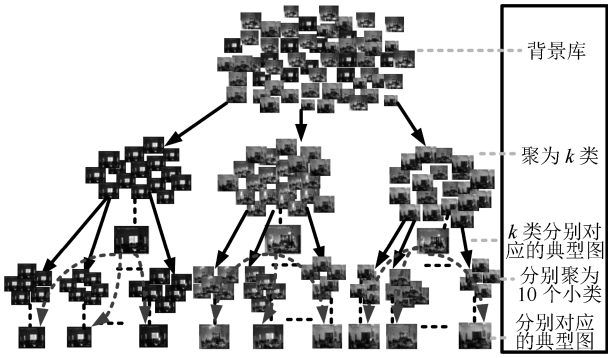


图 3 背景字典生成图

Fig. 3 Map of background dictionary

3 图像检索

图像检索部分主要讨论检索时机, 检索背景字典的方法. 本文使用非平稳变化判断机制确定检索时机, 通过计算欧氏距离检索背景字典.

3.1 非平稳变化判断机制

本文实验的非平稳变化是光线突变, 因而设计关于光线的非平稳变化判断机制.

根据光线突变时绝大多数像素点变为前景的特点, 本文通过关注前景像素点占总像素点的比例来统筹设计判断机制.

光线突变时, 前景像素比例迅速增大, 当大于临界值 T 时, 则认为发生了光线突变. 如式 (7) 其中

Ro 为前景像素比例, $flag_1 = 1$ 说明非平稳变化发生, 反之, 则不是.

$$\begin{cases} Ro \geq T, & flag_1 = 1 \\ Ro < T, & flag_1 = 0 \end{cases} \tag{7}$$

3.2 原图像描述

当判断结果为 $flag_1 = 1$ 时, 将此时原图经过第 2.1.1 节中步骤 4、步骤 5 两步变为与 $\boldsymbol{x}_i$ 维数一致的高维向量 $\boldsymbol{m}$, 将 $\boldsymbol{m}$ 映射到与 $\boldsymbol{y}_i$ 同一个空间中, 变为向量 $\boldsymbol{n}$, 计算公式:

$$\boldsymbol{n} = U_d^{\mathrm{T}}(\boldsymbol{m} - \bar{\boldsymbol{x}}) \tag{8}$$

其中, U_d^{T} 是第 2.1.2 节中的投影矩阵, 向量 $\boldsymbol{n}$ 就是原图像的合理描述.

3.3 检索判断方法

该处步骤与特征脸算法[16] 类似, 通过计算向量 $\boldsymbol{n}$ 与背景字典中 $\boldsymbol{y}_i$ 的欧氏距离, 距离最小的即为对应的背景向量. 将该向量的索引返回, 该索引背景就是此时背景, 并采用该背景初始化背景模型.

4 背景字典更新

这部分讨论模型效果判断机制以及更新方法.

4.1 更新判断机制

检索替换生成新模型之后, 前景像素比例 Ro 应该迅速下降. 如果下降较小, 说明场景与记忆场景差距较大, 此时需要更新背景字典. 根据这个特点设计判断机制如式 (9):

$$\begin{cases} Ro(M_p) - Ro(M_a) \geq T_u, & flag_2 = 0 \\ Ro(M_p) - Ro(M_a) < T_u, & flag_2 = 1 \end{cases} \tag{9}$$

M_p 表示原来模型, M_a 为新模型. T_u 为阈值, $flag_2 = 0$ 代表不需要更新背景字典, $flag_2 = 1$ 代表需要更新背景字典.

4.2 更新方法

结合本文第 2.1.1 节中背景的生成方法, 再考虑快速的背景字典更新, 最终从判断结果为 $flag_2 = 1$ 的当前帧开始累计 100 帧背景计算其平均背景, 将平均背景作为新的背景, 添加到背景库中. 同时, 与谱聚类典型图对应向量比较, 找到在背景字典合适的位置, 将该向量加入. 如果检索位置已满, 则根据被检索频率的高低来替换背景向量, 如果一个向量长时间没有被检索, 则被替换的概率高.

5 长短时空域背景建模融合

以上三部分就是本文框架的建立方法, 由于本文框架主要处理非平稳变化, 而非平稳变化并非常态, 因而设计将现有短时空域的背景建模与本文长时空域背景建模融合. 这样可以提高背景建模速度.

在第 3.1 节中, 我们谈到非平稳变化判断机制, 如果判断为 $flag_1 = 0$, 则使用现有短时空域背景建模算法实现运动目标检测. 如果判断为 $flag_1 = 1$, 则使用长时空域背景建模来建立背景模型. 使用该模型后, 当它转换为平稳变化后, 则继续使用短时空域背景建模算法, 这样既可以保证准确性又可以保障实时性.

当长时空域背景建立的背景模型要转换为短时空域背景模型时, 需要注意: 初始转换时, 增加更新速度可以达到更好的效果. 这主要是因为背景字典中图片与新的视频背景会有些许小差异, 这会引入一定的 ghost 区域.

6 长时间背景建模框架中参数值的确定

6.1 降维维数的预估

首先通过保留信息量的多少来选择一个预定维数, 再根据聚类结果对其做小范围调整. 降维中, 低维空间表达高维空间信息的程度是一个重要的衡量标准, 本文称为贡献率.

如图 4 所示 (以 Data1 为例), 背景图片降至 2 维就可表达 90 % 的信息量, 本文选取维数可以达到 99 % 以上的信息量. 由图可知, 在降至 30 维时其贡献率第一次大于 99 %, 因此, 预估降为 30 维.

图 4 中, 横坐标是降到的维数, 纵坐标是贡献率.

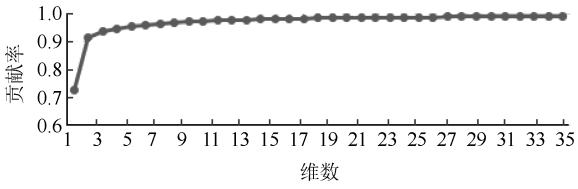


图 4　贡献率图

Fig. 4　Contribution rate

6.2 聚类个数的预估

首先通过谱聚类特点预估聚类个数, 再通过最终聚类结果对其调整. Ng 等提出的 NJW 谱聚类算法[17], 谱聚类的个数通过拉普拉斯特征值的特点来选取. 该算法认为: 对于存在 k 个理想的彼此分离簇的有限数据集, 可以证明拉普拉斯矩阵的前 k 个最大特征值为 1, 第 $k+1$ 个特征值则严格小于 1, 二者之间的差距取决于这 k 个聚类的分布情况. 当聚类内部分布得越密, 各聚类间分布得越开时, 第 $k+1$ 个特征值就越小.

然而, 本文中的聚类数据是图像的特征, 由图 5 可知, 如果直接按照上述方法来判断, 在第二个特征值时就严格小于 1, 那么图像只为一类, 这与聚类的目的相悖. 这也表明图像特征的区分特点没有普通

数据明显, 此时结合聚类的目标对 NJW 谱聚类算法中聚类个数的判断进行拓展. 首先, 此处聚类的目标是得到内部数据紧凑的几类, 而上述方法提到当聚类内部分布的越密, 各聚类间分布的越开, 特征值差异就越大, 就可以通过观察特征值拐点的方法来取合适的个数. 在图 5 中 (以 Data1 为例), 可以看到前 3 个特征值差距较大, 因而取 3 个较为合适.

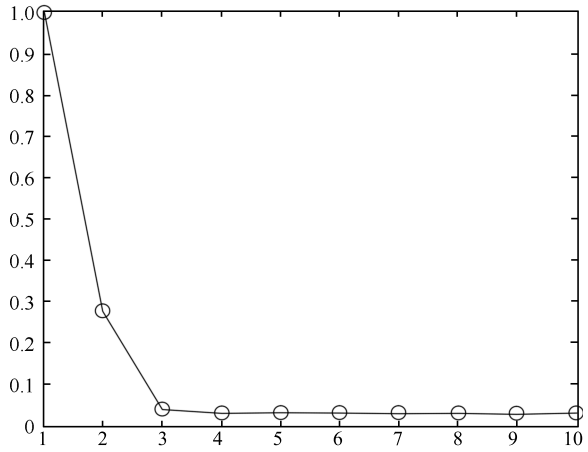


图 5 谱聚类中拉普拉斯矩阵特征值图

Fig. 5 Laplacian eigenvalues graph of spectral clustering

6.3 降维维数与聚类个数的确定

通过聚类结果图, 根据聚类的目标来调整降维维数与聚类个数. 本文中主要是判断开关灯影响, 经验理想值应该为夜晚、白天、开灯的三种情况 (经训练的背景图片信息按照时间顺序排列 (该数据集为晚上 0 点开始至第二天 0 点结束, 图 6 的横坐标即为按时间排列的图像)). 图 6 (以 Data1 为例) 为不同维数的聚类结果图, 观察该对比图: 发现在原来第 6.1 节所得维数的基础上再加两维可以达到聚类内部紧凑、类间分离的目标, 而在维数太大的情况下, 由于所展现的特征的不同, 出现过拟合, 反而达不到目标的效果. 图 7 是降维至 32 维时, 取不同的聚类个数的效果, 可以看到在聚为三类时, 它将上午、下午聚为一类, 中午以及晚上开灯情况分为一类, 其余一类是夜晚、晚上未开灯情况, 根据数据集本身特点

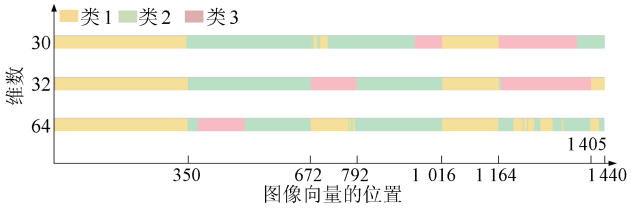


图 6 不同维数的聚类效果

Fig. 6 Cluster results of different dimension

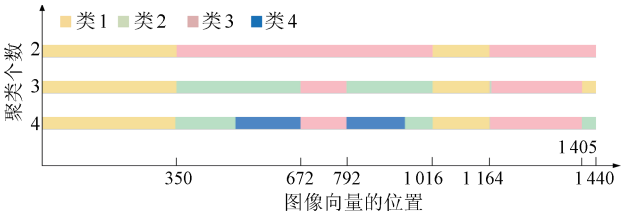


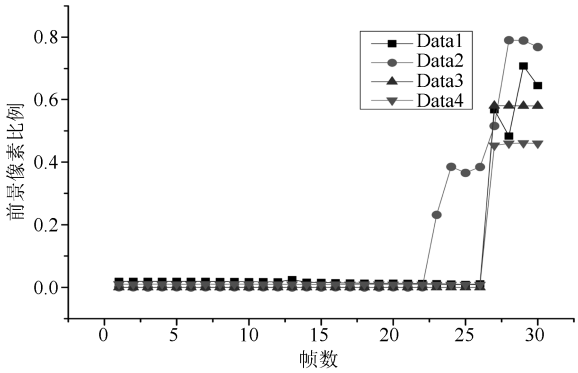
图 7 不同聚类个数效果图 (32 维)

Fig. 7 Cluster results of different cluster number (32)

显示, 该种情况符合研究特点. 而在聚为二类、四类时, 夜晚关灯后的背景的图像不符合研究特点. 因而, 最终取 32 维、聚为三类.

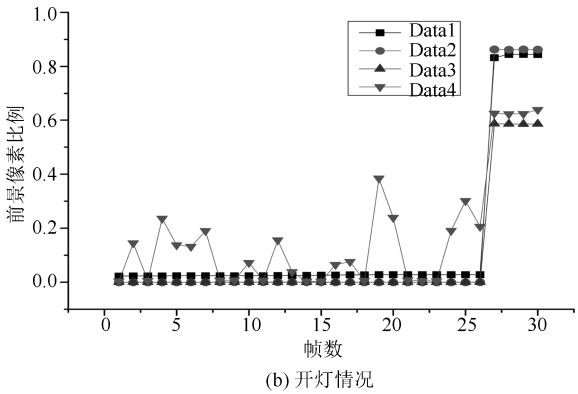
6.4 判断机制中参数的确定

通过统计短时空域背景建模算法 (以 ViBe 算法为例) 光线突变前、后前景像素点比例 (针对本文的四个数据集), 如图 8～10 所示, 对光线突变阈值 T 更新背景字典阈值 T_u 取值. 本文中 T 取 0.42, 取 T_u 为 0.35.



(a) 关灯情况

(a) Turn off



(b) 开灯情况

(b) Turn on

图 8 光线突变阈值 T 的确定

Fig. 8 Determination of sudden illumination change threshold T

图 8 是分别对关灯、开灯四个数据集光线突变前后 30 帧的前景像素统计图, 前 26 帧表示未发生光线突变, 后 4 帧表示已经发生光线突变. 在图 8 (a)

中, 未发生光线突变时, 除 Data2 数据集, 前景比例均很小接近 0, Data2 有波动是由于视频帧中有大目标出现, 而突变后, 前景比例最低的 Data4 接近 0.45; 在图 8 (b) 中, 除 Data4 数据集, 其余前景比例均很小接近 0, Data4 中波动主要是由于开灯时日光灯的闪烁造成的, 而突变之后, 前景比例最低的 Data3 接近 0.6. 综上所述, 结合开关灯突变像素比例变化, 开灯日光灯闪烁, 大目标出现三方面影响, 取 T 为 0.4 左右.

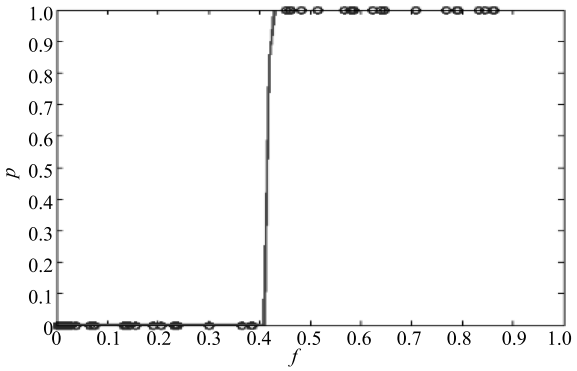
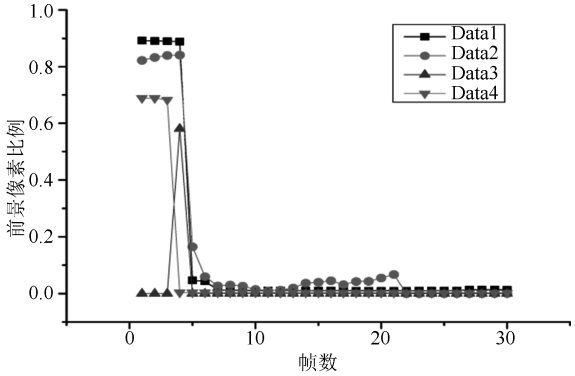


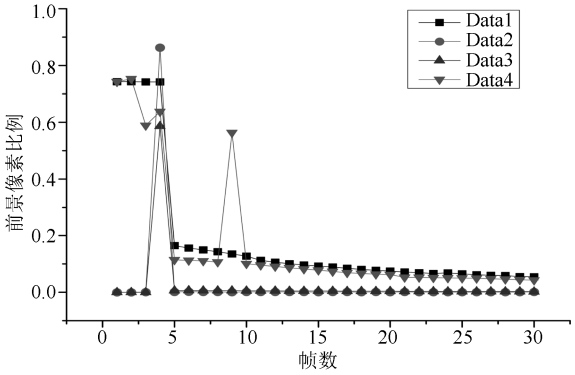
图 9　阈值 T 的逻辑回归分析

Fig. 9　Logistic regression analysis of threshold T



(a) 关灯情况

(a) Turn off



(b) 开灯情况

(b) Turn on

图 10　更新背景字典阈值 T_u 的确定

Fig. 10　Determination of threshold T_u for updating background dictionary

在实践中, 我们采用逻辑回归的方法最终确定阈值 T. 将生成图 8 的训练集 (像素点比例集) 作为样本, 是对应的二值随机变量的集合, 每个元素值为 0 或 1 (0 表示突变前状态, 1 表示突变后的状态); 如下式 (10)、(11) 所示:

$$h_i = \begin{cases} 1, & p \\ 0, & 1-p \end{cases} \tag{10}$$

$$p = \frac{1}{1+\mathrm{e}^{-z}} = \frac{1}{1+\mathrm{e}^{-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{f}_i}} \tag{11}$$

式 (11) 中, $\boldsymbol{f}_i$ 是输入训练样本向量, 其中每个样本都可以得到一个 h_i, $\boldsymbol{\beta}$ 是参数向量, p 表示 h_i 为 1 的可能性, 是 Sigmod 函数. 通过回归模型获得的 h_i 为 1 的概率与 $\boldsymbol{f}_i$ 的对应关系如图 9 所示. 若 $p = 0.5$, h_i 是 0 或 1 (当前状态是突变前或突变后) 的可能性是相同的. 因此, 我们取 $p = 0.5$ 对应的 f 的值作为临界点 T 的取值, 即 0.42.

图 10 是使用本文方法后, 前景像素比例的变化图. 图中展示不同数据集在第四帧处, 使用本文方法像素的变化情况, 也是背景适应程度的展现. 由图 8 可知, 正常情况下使用本文方法后, 前景像素比例降幅明显. 本文的衡量方法就是观察突变后, 不同算法的前景像素比例变化情况 (如图 12、14). 由图 10 (a)、图 10 (b) 展现的均为背景字典中背景能代表场景的情况 (实际场景变化如图 11、图 13), 其比例变化最小值接近 0.5. 而根据实验结果图观察, 被认为替换效果不佳的比例变化最大接近 0.3, 因此, 根据经验将 T_u 定为 0.35, 也可采用逻辑回归对其验证.

7　实验结果与分析

为了验证该框架的性能, 将该框架用于 ViBe 算法以及 MOG 算法, 在多个测试数据集上进行实验, 比较这两种算法与本文融合框架后算法在光线突变发生时的运动目标检测情况.

实验在 Intel(R)C @ 2.40 GHz 8.0 GB 的计算机上, VS2013、OpenCV2.4.9 和 MATLABR2013a 环境下实现, 在实验中 ViBe 维持原论文中参数, MOG 采用 Opencv 实现版. 本文算法未对视频做形态学等预处理以及后处理.

7.1　实验结果

本文讨论长时间视频背景建模方法, 数据集分为训练数据集、测试数据集. 训练数据集用于构建记忆背景字典, 测试数据集用于检索并实现运动目标检测.

本文训练数据集有四个, 第一个是由监控摄像头拍摄的实验室 2015 年 11 月 27 日整天视频数据 (后续称为 Data1, 分辨率为 352 × 288); 第二个是

通用数据集 WallFlower dataset[18] 中 LightSwitch 数据集 (称为 Data2, 分辨率为 160 × 120); 第三个是 WallFlower dataset 中 TimeOfDay 数据集 (称为 Data3, 分辨率为 160 × 120); 第四个是由焦距 2.6 mm 摄像头拍摄的室内 2016 年 4 月 20 日整天定视角视频 (称为 Data4, 分辨率为 640 × 320).

Data1 测试数据集为 2015 年 11 月 25 日的视频序列 (共 778 帧)、2015 年 11 月 26 日视频序列 (共 2 474 帧); Data2 测试数据集是 LightSwitch 数据集中未训练的开关灯图片序列 (开灯测试序列共 378 帧, 关灯测试序列共 1 625 帧); Data3 测试数据集是 TimeOfDay 数据集模拟的开关灯数据集 (开灯测试序列共 576 帧, 关灯测试序列共 132 帧); Data4 的测试数据集是 2016 年 4 月 20 日 (共 1 473 帧)、2016 年 4 月 19 日视频序列 (共 1 113 帧).

图 11、图 12、图 15、图 16 是四个不同数据集的实验结果对比图. 图 11、图 15 是关灯情况, 图 12、图 16 是开灯情况; 图 11、图 12、图 15、图 16

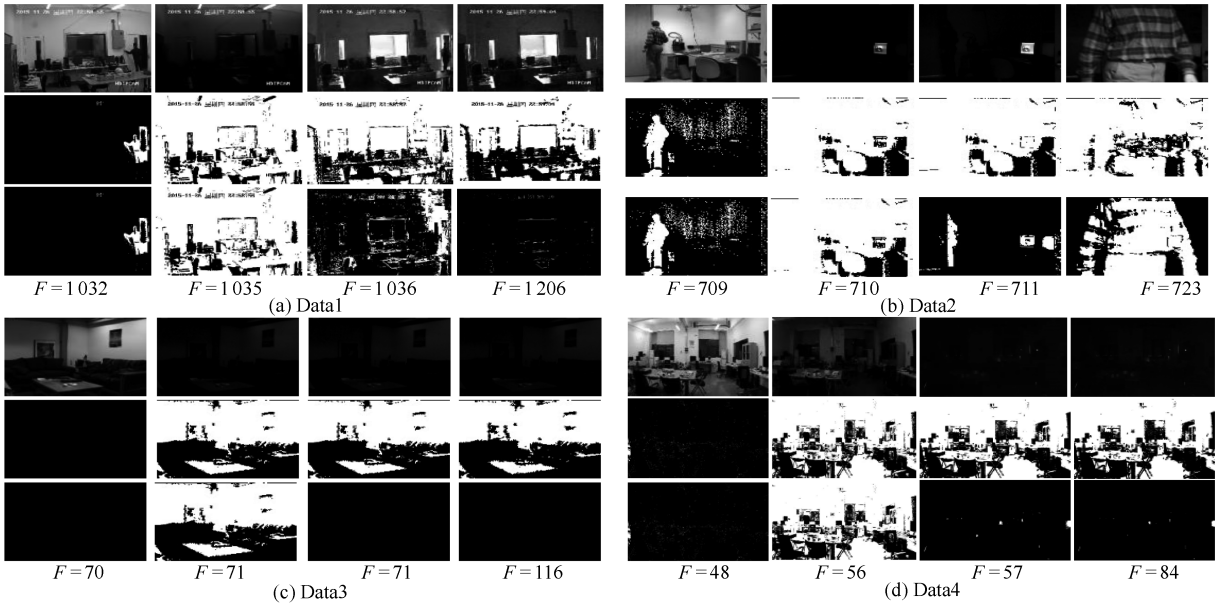


图 11 运动目标检测效果对比图 (ViBe 关灯)

Fig. 11 Moving object detection comparison charts (ViBe turns off the lights)

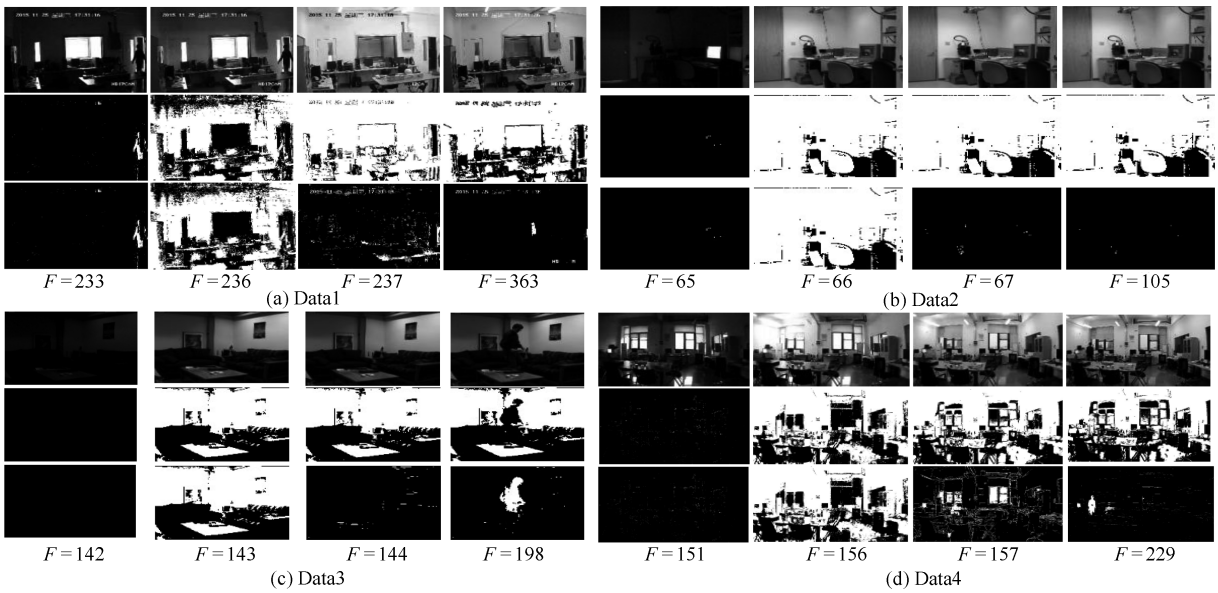


图 12 运动目标检测效果对比图 (ViBe 开灯)

Fig. 12 Moving object detection comparison charts (ViBe turns on the lights)

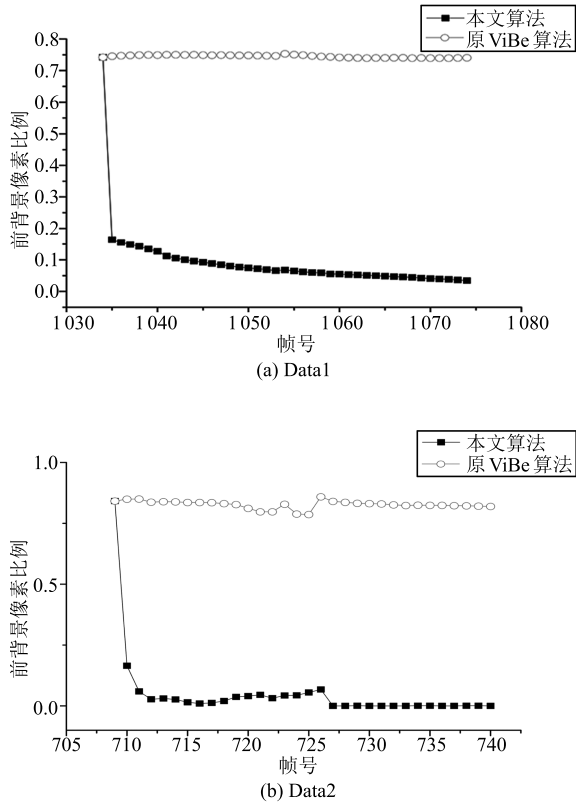


图 13　前景像素比例变化对比图 (对应图 11 (a) ∼ (b))

Fig. 13　Comparison chart of foreground pixel ratio (Corresponding to Fig. 11 (a) ∼ (b))

中 a、b、c、d 子图分别对应 Data1、Data2、Data3、Data4 实验结果, a、b、c、d 四个子图中每个子图第一排是原视频图像, 图 11、图 12 的第二排是 ViBe 运动目标检测前景, 图 15 图 16 的第二排是 MOG 运动目标检测前景, 第三排是本文提出算法的运动目标检测前景, F 表示视频序列的第几帧 (忽略日光灯闪烁帧、摄像头适应帧). 从图中可以看到: 在发生光线突变时, 大量的背景点被误判为前景点, 原 ViBe、MOG 算法恢复模型较慢, 使虚警数 (False positive, FP)[19] 在长时间内较高; 结合本文框架可以使它迅速适应光照突变, 大大降低虚警数, 从而更为准确地侦测运动目标.

图 13、图 14 分别是图 11 (a) ∼ (b)、图 13 (c) ∼ (d) 相应的前景点比例变化比较图, 图 15、图 16 的相应的前景变化比例可类似得到. 这两幅图通过前景像素点比例形象的表示光线突变后背景模型的适应情况.

图 17 为室外场景的测试效果, 首先使用 8 月 16 日的视频背景数据来建立背景字典, 测试数据集为 8 月 15 日傍晚室外开灯序列视频 (共 273 帧). 图 17 (a) 为 ViBe 算法与结合本文框架后算法的对比图, 其中第一排表示原图, 第二排表示 ViBe 算法目标检测效果图, 第三排为结合本文框架的目标检测效果图. 图 17 (b) 为混合高斯背景建模算法与结合本文框架后的算法的对比图, 其中第一排表示原图, 第二排表示混合高斯背景建模算法目标检测效果图, 第三排为结合本文框架的目标检测效果图. 由图可知, 对室外光线突变场景, 结合本文记忆字典模型可以显著提高模型适应能力, 有效降低虚警数.

总结图 11 ∼ 17 知, 本文算法有效提高短时空域算法光照突变适应能力, 降低原有算法误检率, 可以更好地侦测运动目标.

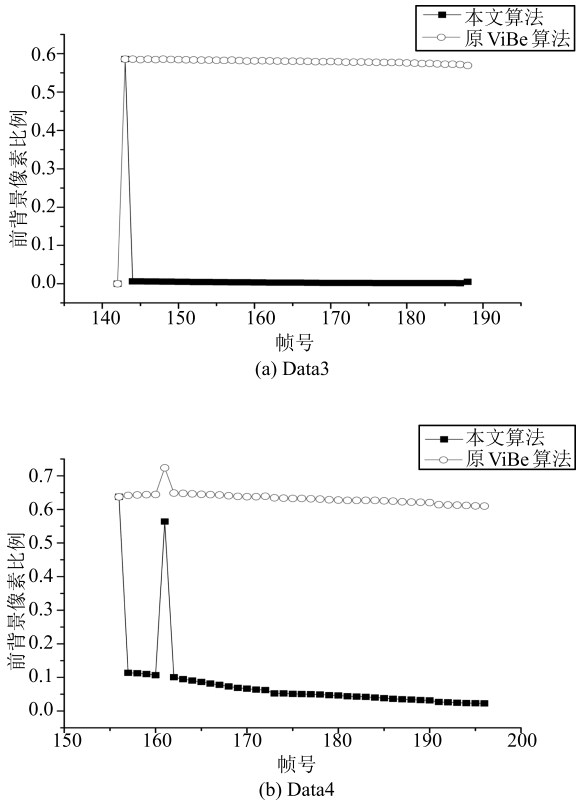


图 14　前景像素比例变化对比图 (对应图 12 (c) ∼ (d))

Fig. 14　Comparison chart of foreground pixel ratio (Corresponding to Fig. 12 (c) ∼ (d))

7.2　性能分析

本文使用虚警数 FP 以及漏检数 FN 来定量评估. 虚警数 FP 是本身为背景像素却被误判为前景的像素个数, 漏检数 FN 是本身为前景像素却被误判为背景的像素. 本文框架相当于是在原算法发生光照突变之后做的处理, 那么在未发生光照突变时, 本文算法与原算法的虚警数与漏检数一致; 而在发生光照突变后, 由图 13、图 14 知, 本文算法大大地降低了虚警数, 而漏检数与原算法未发生光照突变时一致, 也就是比此时原算法的漏检数少.

7.3　实时性分析

本文采取了训练、测试模式, 存在训练时间以及测试时间. 在运动目标检测时相当于处于测试阶段, 本文与之密切相关的为检索时间, 因而与原算法

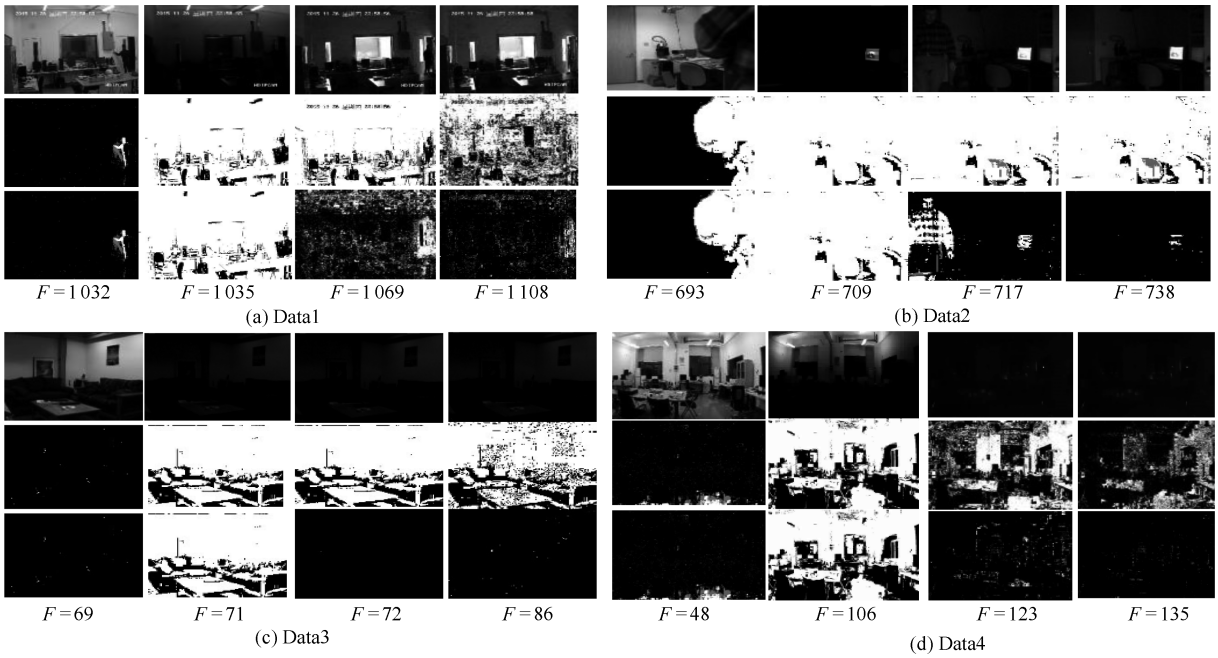


图 15 运动目标检测效果对比图 (MOG 关灯)

Fig. 15 Moving object detection comparison charts (MOG turns off the lights)

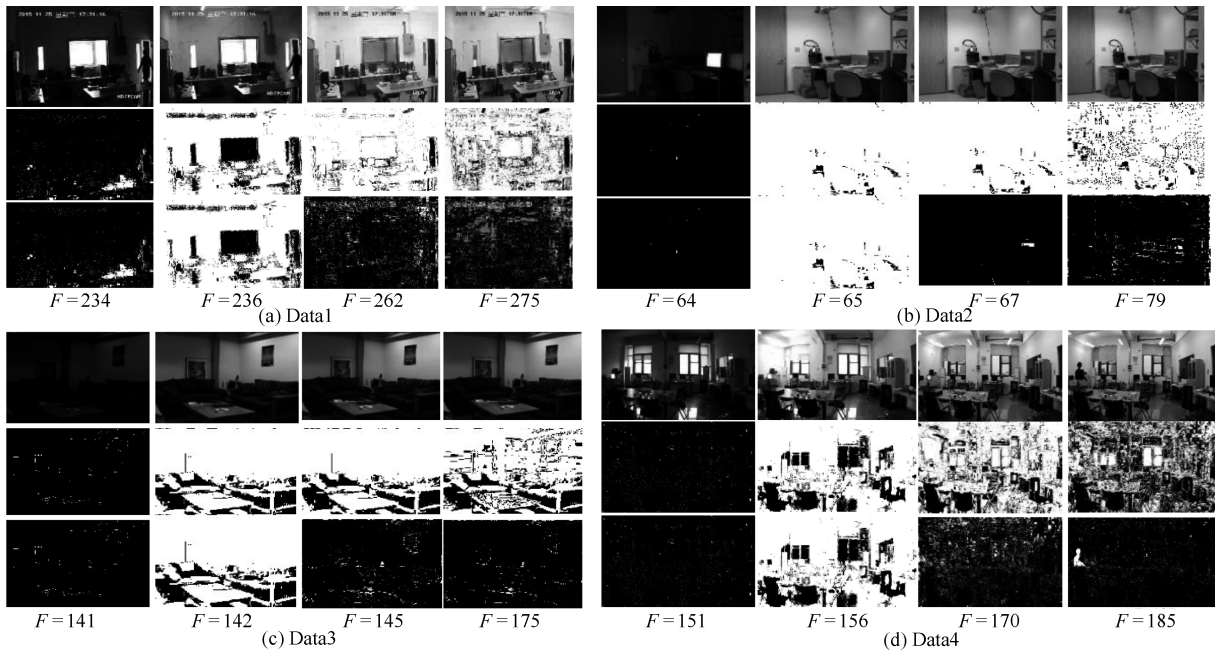
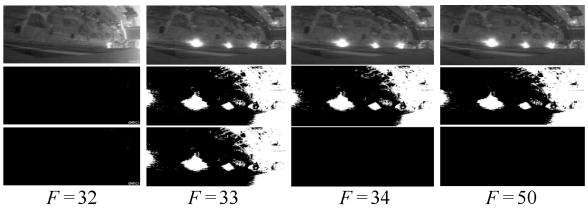


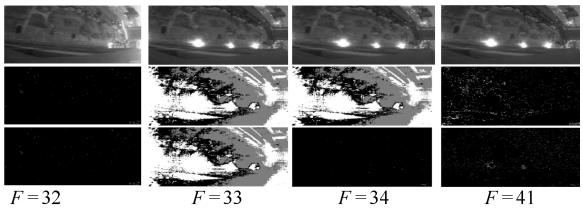
图 16 运动目标检测效果对比图 (MOG 开灯)

Fig. 16 Moving object detection comparison charts (MOG turns on the lights)

比较, 本文算法主要增加了额外的检索时间, 本文在检索部分的耗时运算为 13 次欧氏距离的计算 (参考图 3 结构, 其中 3 次为图像描述子与三类典型图描述子之间的欧氏距离, 求得最近的典型图之后, 计算原图描述子与该典型类中再聚类 10 个典型图描述子之间的欧氏距离). 在未采取优化机制的情况下, Data1、Data2、Data3、Data4 检索一次背景字典的时间分别为 0.137 s、0.051 s、0.105 s、0.123 s. 由此可以推想到, 视频中检索背景字典的频数对实时性有影响, 即突变越频繁, 检索背景字典次数越多, 对实时性影响越大. 然而在一般场景中, 开关灯情况并不频繁. 本文以每 300 帧 (大约 10 s) 发生一次检索来计算实时性, 与原算法的对比效果图如表 1 所示. 在更新背景字典时, 使用另外一个线程来生成新



(a) ViBe 与融合本文框架的 ViBe 效果对比
(a) Comparison between ViBe and ViBe fusion frame



(b) MOG 与融合本文框架的 MOG 效果对比
(b) Comparison between MOG and MOG fusion frame

图 17　室外情况运动目标检测情况
Fig. 17　Moving object detection of outdoor

的背景图, 不影响主线程的实时性.

由表 1 可知, 本文通过前期对视频集做背景训练, 在之后的目标检测中对实时性的影响很小, 而由图 11 ∼ 17 可知, 该方法可以有效降低误检率, 明显改善运动目标检测结果.

表 1　算法处理速度 (fps)
Table 1　Processing times of algorithm (fps)

算法	Data1	Data2	Data3	Data4
原 ViBe 算法	25.96	63.79	62.44	14.49
本文算法	25.65	63.13	59.48	14.40

8　结语

利用固定摄像头定视角视频背景周期性重现 (比如白天、夜晚周期性变换) 特点, 搭建基于长时间视频序列的背景建模框架并研究方法. 首先通过对长时间背景序列预处理、降维, 得到背景描述子; 然后, 通过聚类 (包括谱聚类、K-means 聚类) 来训练背景字典; 再设计非平稳变化下的检索替换机制, 并在效果差时对背景字典更新; 设计长短时空域模型的融合机制增强实时性, 可以有效改善运动目标检测. 通过搭建这样一个可以嵌入现有背景建模算法中的框架, 可以解决室内场景难题. 将 ViBe 或 MOG 算法与该框架融合, 测试非平稳变化 (本文主要测试光照突变), 实验结果表明, 该框架可以使 ViBe、MOG 算法迅速适应光线突变, 明显提高运动目标检测的准确性, 有效降低 ViBe、MOG 算法的误检率.

由于本文主要针对定视角室内场景, 仅对室内非平稳变换 (光照突变) 以及简单室外光照突变的情形进行测试. 如果是复杂室外场景, 则要考虑相机抖动、动态场景等情形, 未来将通过对相机抖动, 动态场景等训练学习, 探索更为通用的长时间域背景建模方法.

References

1 Chu Jun, Yang Fan, Zhang Gui-Mei, Wang Ling-Feng. A stepwise background subtraction by fusion spatio-temporal information. *Acta Automatica Sinica*, 2014, **40**(4): 731−743
(储珺, 杨樊, 张桂梅, 汪凌峰. 一种分步的融合时空信息的背景建模. 自动化学报, 2014, **40**(4): 731−743)

2 Niu Hua-Kang, He Xiao-Hai, Wang Xiao-Fei, Zhang Feng, Wu Xiao-Qiang. An improved ViBe object detection algorithm. *Journal of Sichuan University (Engineering Science Edition)*, 2014, **46**(S2): 104−108
(牛化康, 何小海, 汪晓飞, 张峰, 吴小强. 一种改进的 ViBe 目标检测算法. 四川大学学报 (工程科学版), 2014, **46**(S2): 104−108)

3 Wren C R, Azarbayejani A, Darrell T, Pentland A P. Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 780−785

4 Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. In: Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Fort Collins, Co, USA: IEEE, 1999, **2**: 252

5 Wang Y, Loe K F, Wu J K. A dynamic conditional random field model for foreground and shadow segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(2): 279−289

6 Kim K, Chalidabhongse T H, Harwood D, Davis L. Background modeling and subtraction by codebook construction. In: Proceedings of the 2004 International Conference on Image Processing. Singapore: IEEE, 2004, **5**: 3061−3064

7 Barnich O, Van Droogenbroeck M. ViBe: a universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, 2011, **20**(6): 1709−1724

8 St-Charles P L, Bilodeau G A, Bergevin R. Subsense: a universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, 2015, **24**(1): 359−373

9 van der Maaten L J P, Postma E O, van den Herik H J. Dimensionality reduction: a comparative review. *Journal of Machine Learning Research*, 2007, **10**(1): 66−71

10 Huang H C, Chuang Y Y, Chen C S. Affinity aggregation for spectral clustering. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012. 773−780

11 Arthur D, Vassilvitskii S. k-means++: the advantages of careful seeding. In: Proceedings of the 18th annual ACM-SIAM Symposium on Discrete Algorithms. Philadelphia, PA, USA: ACM, 2007. 1027−1035

12 Goyette N, Jodoin P M, Porikli F, Konrad J, Ishwar P. Changedetection. net: a new change detection benchmark dataset. In: Proceedings of the 2012 IEEE Computer Society Conference on Workshop on Computer Vision and Pattern Recognition Workshops. Providence, RI, USA: IEEE, 2012. 1−8

13 Su Ya-Ru. Research on Dimensionality Reduction of High-Dimensional Data [Ph. D. dissertation], University of Science and Technology of China, China, 2012
(苏雅茹. 高维数据的维数约简算法研究 [博士学位论文], 中国科学技术大学, 中国, 2012)

14 Cai Xiao-Yan, Dai Guan-Zhong, Yang Li-Bin. Survey on spectral clustering algorithms. *Computer Science*, 2008, **35**(7): 14−18
(蔡晓妍, 戴冠中, 杨黎斌. 谱聚类算法综述. 计算机科学, 2008, **35**(7): 14−18)

15 Zhu X T, Loy C C, Gong S G. Constructing robust affinity graphs for spectral clustering. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014. 1450−1457

16 Smiatacz M. Eigenfaces, Fisherfaces, Laplacianfaces, Marginfaces — how to face the face verification task. In: Proceedings of the 8th International Conference on Computer Recognition Systems CORES. Switzerland: Springer, 2013. 187−196

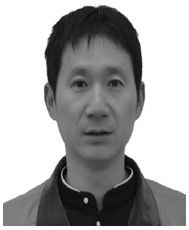
17 Ng A Y, Jordan M I, Weiss Y. On spectral clustering: analysis and an algorithm. In: Proceedings of Advances in Neural Information Processing Systems 14: Proceedings of the 2001 Conference. Vancouver, British Columbia, Canada: MIT Press, 2001, **14**: 849−856

18 Toyama K, Krumm J, Brumitt B, Meyers B. Wallflower: principles and practice of background maintenance. In: Proceedings of the 7th IEEE International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1991, **1**: 255−261

19 Chen Y T, Chen C S, Huang C R, Huang Y P. Efficient hierarchical method for background subtraction. *Pattern Recognition*, 2007, **40**(10): 2706−2715

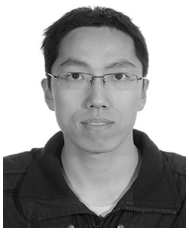
丁 洁 中北大学计算机与控制工程学院硕士研究生. 主要研究方向为计算机视觉, 虚拟仿真与可视化.
E-mail: jie__ding@163.com
(**DING Jie** Master student at the School of Computer and Control Engineering, North University of China. Her research interest covers computer vision, virtual simulation and visualization.)

肖江剑 中国科学院宁波工业技术研究院研究员. 主要研究方向为计算机视觉, 图像和视频处理, 车辆跟踪与智能交通, 模式识别. 本文通信作者.
E-mail: xiaojj@nimte.ac.cn
(**XIAO Jiang-Jian** Research fellow at Ningbo Institute of Industrial Technology, Chinese Academy of Sciences. His research interest covers computer vision, image and video processing, vehicle tracking and intelligent transportation and pattern recognition. Corresponding author of this paper.)



况立群 中北大学计算机与控制工程学院副教授. 主要研究方向为仿真与可视化, 图像处理, 虚拟现实.
E-mail: liqun_kuang@163.com
(**KUANG Li-Qun** Associate professor at the School of Computer and Control Engineering, North University of China. His research interest covers virtual simulation and visualization, image processing and virtual reality.)

宋康康 中国科学院宁波工业技术研究院工程师. 主要研究方向为图像处理, 计算机视觉.
E-mail: songkk@nimte.ac.cn
(**SONG Kang-Kang** Engineer at Ningbo Institute of Industrial Technology, Chinese Academy of Sciences. His research interest covers computer vision, image processing.)



彭成斌 中国科学院宁波工业技术研究院副研究员. 主要研究方向为数据挖掘, 模式识别和并行计算.
E-mail: pengchengbin@nimte.ac.cn
(**PENG Cheng-Bin** Associate researcher at Ningbo Institute of Industrial Technology, Chinese Academy of Sciences. His research interest covers data mining, pattern recognition, and parallel computing.)