

基于目标分割和融合的快速视频浏览算法^①

周荣华¹, 战荫伟¹, 梅振顺¹, 邹利华²

¹(广东工业大学 计算机学院, 广州 510006)

²(东莞职业技术学院, 东莞 523000)

摘 要: 针对稀疏对象的监控视频的快速浏览, 提出了一种基于目标分割和融合的快速视频浏览的方法. 首先, 利用背景差分法得到前景目标, 并对背景模型不断更新; 然后, 滤除无目标出现的视频帧, 并以一定的密度对目标对象进行融合播放. 实验结果表明, 该方法鲁棒性好, 能取得较好的效果.

关键词: 混合高斯模型; 目标分割; 目标融合; 快速浏览

Quick Video Browsing Algorithm Based on Object Segmentation and Fusion

ZHOU Rong-Hua¹, ZHAN Yin-Wei¹, MEI Zhen-Shun¹, ZOU Li-Hua²

¹(Faculty of Computer, Guangdong University of Technology, Guangzhou 510006, China)

²(Dongguan Polytechnic, Dongguan 523000, China)

Abstract: A quick video browsing algorithm is proposed for videos of sparse objects. Background subtraction is applied to get foreground object and the background is updated constantly. Those frames without objects are first discarded and the remaining frames are further fused. Experiment shows our proposed algorithm is robust and has good performance.

Key words: GMM; object segmentation; object fusion; quick video browsing

社会信息化促进视频监控领域的发展, 诸多行业对视频监控的需求大量增加. 监控系统通常记录长时间的视频数据, 这些数据往往对事件的事后追溯和调查具有重要意义. 由于调阅和回放视频需要花费大量的时间和精力, 其过程枯燥乏味, 不利于在海量的视频数据中查到感兴趣的信息. 而且, 传统的拖拉浏览的方法对长时间的数据进行搜索时, 非常容易忽视短时间的突发异常事件. 因此, 如何快速浏览大量视频快速的搜索、定位感兴趣的信息, 一直是研究的热点.

在很多监控区域, 只有少量的运动对象间断的出现, 或者, 不同时间段运动对象的密度差别较大. 拥有这些特点的监控视频我们称之为稀疏运动对象的监控视频. 现实中, 摄像机位置固定、现场背景基本不变、运动目标比较稀少这三种条件并存的情况是相当普遍的. 在这种情况下, 相应的视频表现为相对固定的背景与稀疏的运动物体的融合.

针对这类视频, 本文提出了一种将多个时间段内

的运动目标融合到同一背景上, 以缩短视频播放时间的方法. 由于用户感兴趣的内容集中在运动物体上, 这种融合的方法能够在保持内容信息不变的前提下, 有效而快速地浏览视频、定位目标.



图 1 空间稀疏的监控场景

1 背景建模、初始化及其更新

1.1 背景建模

实际的监控环境随时间推移是不断变化的, 考虑

① 基金项目: 东莞市科技计划(201010825606917); 广州市科技计划(2010Y1-C611)

收稿时间: 2012-02-22; 收到修改稿时间: 2012-04-02

到背景像素值在一段时间内分布是多峰的, 目标区域的像素变化很难用单一分布的模型来描述, 须采用多个单高斯模型^[1]的集合来描述像素值在一段时间内的状态, 这就是混合高斯背景模型. 其基本思想是: 对每个像素, 定义 K 个高斯分布来表示其状态, K 值为像素值多峰分布的峰的个数, 在检测过程中, 只要该像素符合 K 个高斯分布中的一个, 那么就认为该像素是具有背景特征的; 反之, 该像素则被认定为前景^[2]. 时刻 t 像素值 I_t 的第 k 个高斯分布的概率密度函数为

$$f(I_t | \mu_{k,t}, \sum_{k,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\sum_{k,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(I_t - \mu_{k,t})^T \sum_{k,t}^{-1} (I_t - \mu_{k,t})} \quad (1)$$

式中 $\mu_{k,t}$ 和 $\sum_{k,t}$ 为均值和协方差矩阵. 像素点当前的特征与该像素点过去时刻的特征有关, 像素 $I(x,y)$ 的概率为

$$P(I_t) = P(x_{k,t} | x_{k,1}, \dots, x_{k,t-1}) = \sum_{k=1}^K \omega_{k,t} f(I_t | \mu_{k,t}, \sum_{k,t}) \quad (2)$$

式中 $\omega_{k,t}$ 为第 k 个高斯分布的权重. 根据 $\omega_{k,t}$ 的值从大到小排序, 满足

$$\frac{\sum_{k=1}^b \omega_k}{\sum_{k=1}^K \omega_k} > T \quad (3)$$

的最小 b 个高斯分布作为背景模型.

1.2 背景模型初始化

计算一段时间内视频序列图像中每一个像素点平均灰度值 μ_0 及像素方差 σ_0^2 , 来初始化混合高斯模型参数

$$\mu_0 = \frac{1}{N} \sum_{t=0}^{N-1} I_t \quad (4)$$

$$\sigma_0^2 = \frac{1}{N} \sum_{t=0}^{N-1} (I_t - \mu_0)^2 \quad (5)$$

1.3 背景模型更新

由于背景随时间的推移是不断变化的, 所以背景模型必须能自适应地变化. 背景模型的更新策略是背景模型设计中最关键的技术^[3]. 在模型更新时应注意如下原则: 对背景变化的响应速度要足够快; 对运动目标要有较强的抗干扰能力.

首先, 本文采用 Stauffer 等人^[4-6]的方法来近似计算, 将每一个新捕获的像素值 I_t 与已知混合模型里的

K 个高斯分布进行比较, 如果满足

$$|I_t - \mu_{k,t-1}| \leq \tau \sigma_{k,t-1} \quad k = 1, 2, \dots, K \quad (6)$$

则 I_t 和该高斯分布匹配, 其中 τ 是确定匹配相似度的一个常数, 采用 $\tau = 2.5$.

其次, 匹配可能出现如下两种情况:

第一种情况: I_t 与混合高斯模型中的某一个或者多个高斯分布匹配. 当 I_t 与某个高斯分布相匹配时就会影响原先的概率分布, 因此要更新与 I_t 匹配的高斯分布的均值和方差. 另外, 当 I_t 与这个分布中某一个或几个单高斯模型相匹配时, 说明该单模型较为符合当前像素值的分布, 所以需要适当增加其权值.

$$\begin{aligned} \mu_{k,t} &= (1 - \alpha) \mu_{k,t-1} + \alpha I_t \\ \sigma_{k,t}^2 &= (1 - \alpha) \sigma_{k,t-1}^2 + \alpha (I_t - \mu_{k,t})^2 \\ \omega_{k,t} &= (1 - \beta) \omega_{k,t-1} + \beta \theta \quad k = 1, 2, \dots, K \end{aligned} \quad (7)$$

式中 α 为参数更新速率, β 为学习速率, 通常为了减少背景噪声, β 一般被固定在比较小得数值水平上, 经过多次实验, 采用 $\beta = 0.05$ 时适应性更佳. 当新像素值 I_t 符合第 k 个高斯分布时, $\theta = 1$; 若不符合, 则 $\theta = 0$.

第二种情况: I_t 未能与混合高斯模型中的任何一个高斯分布匹配. 当没有一个高斯分布与 I_t 匹配时, 说明出现了新的分布形式, 则需增加一个新的单高斯分布. 而增加一个高斯分布后, 高斯分布的个数超过最大值 K 时, 则去除一个当前混合高斯模型中权值最小的高斯分布. 而根据 I_t 引入新的高斯分布, 其均值为 I_t , 同时给一个较大方差、较小的权值.

最后, 根据更新之后的权值对高斯分布重新排序, 并对更新之后的权值做归一化处理

$$\omega_{k,t}^* = \frac{\omega_{k,t}}{\sum_{k=1}^K \omega_{k,t}} \quad k = 1, 2, \dots, K \quad (8)$$

2 目标分割及融合

2.1 目标分割

背景差分法是常用的运动目标检测方法之一, 该方法实现简单, 并能完整地分割出运动对象. 设当前帧图像为 $C_t(x,y)$, 生成的背景图像为 $B_t(x,y)$, 检测到的前景图像为 $D_t(x,y)$. 计算当前帧图像 $C_t(x,y)$ 像素与上一帧背景图像 $B_{t-1}(x,y)$ 像素的差值, 若差值大于阈值 ε 则认为该像素是前景像素点, 否则被认为是背景

像素点

$$D_t(i, j) = \begin{cases} 1 & |C_t(i, j) - B_t(i, j)| \geq \varepsilon \\ 0 & |C_t(i, j) - B_t(i, j)| < \varepsilon \end{cases} \quad (9)$$

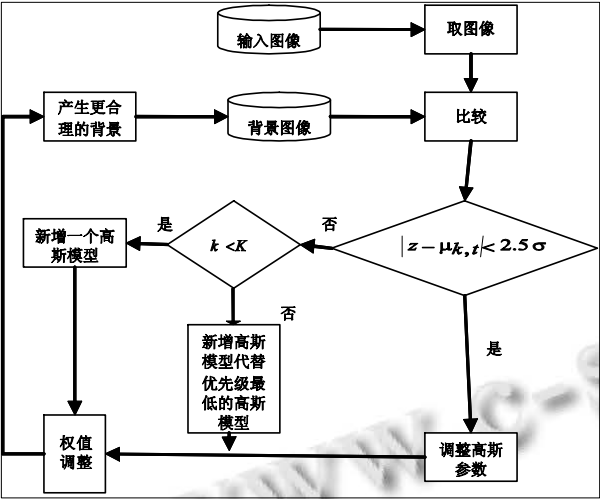


图 2 背景建模及更新流程

其中 ε 是设定的阈值, 本文实验设定 ε 为 60. 当 $D_t(i, j)=0$ 时为背景像素, $D_t(i, j)=1$ 时为前景像素. 由于噪声的影响, 背景差分提取前景产生的结果不是非常理想. 运用数学形态学对差分结果进行处理, 滤除伪目标, 连通目标.

2.2 目标融合

无目标出现的帧图像并不是我们所关注的视频图像, 而且我们也希望去除这些时间分布上的冗余信息, 因此, 在目标融合之前滤除这些无目标的视频帧.

$$F_t = \begin{cases} 1 & D_t = 1 \\ 0 & D_t = 0 \end{cases} \quad (10)$$

式中 D_t 为前景图像标记, 当 $D_t=1$ 时, 表示该帧有目标; 相反, 该帧无目标; F_t 为视频图像标记, 当 $F_t=1$ 时保留该帧图像, 否则去除该帧. 至此, 整段视频可以分割成数段至数十段, 统计每段目标的个数 N_k .

目标融合是针对视频每一帧来操作的^[7-9], 为了进一步压缩视频的浏览时间, 把不同时间段出现的目标融合到同一时间段播放. 因为同一场景中, 目标在空间上的分布是随机的且是稀疏的, 因此, 这也给不同时间出现的目标融合到同一时间段提供了可能. 设 $D_{k1,t}$ 为第 $k1$ 段视频 t 时刻的前景图像标记, $D_{k1,t}$ 类似 D_t , 只是 $D_{k1,t}$ 表示第 $k1$ 段视频 t 时刻的前景标记.

$F_{k1,t}(x, y)$ 为其对应的目标图像(除了目标区域像素保持原来值外, 其他区域像素置为 0); $D_{k2,t}$ 为第 $k2$ 段视频 t 时刻的前景图像标记, $F_{k2,t}(x, y)$ 为其对应的目标图像; $B_k(x, y)$ 为第 k 段视频背景, $F'_{k,t}(x, y)$ 为融合后的帧图像; O_n 为用户指定的密度. 若 $O_n \leq N_k$ 时, 此段视频图像保持不变; 若 $O_n > N_k$ 时, 将下一时间段的目标按融合到此段视频中.

$$F'_{k,t}(x, y) = \begin{cases} F_{k1,t}(x, y) & (x, y) \in Object_{k1}, D_{k1,t} = 1 \\ F_{k2,t}(x, y) & (x, y) \in Object_{k2}, D_{k2,t} = 1 \\ B_k(x, y) & (x, y) \in Background \end{cases} \quad (11)$$

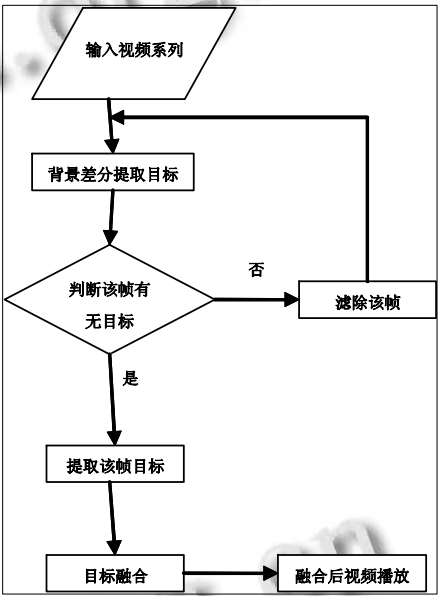


图 3 算法处理流程

3 实验结果和分析

实验环境: CPU AMD Athlon(tm)64x22.01 GHz 1.0GB 内存; WindowsXP 操作系统, VS2008+Open CV2.1. 测试视频概况如表 1 所示.

表 1 测试视频概况

视频格式:	AVI
帧率:	7fps
视频宽度:	532
视频高度:	288

图 4 为处理原始视频后的最终结果, 通过运动目标检测及分割提取出运动目标. 最后通过视频融合对提取到的运动目标进行融合. 根据需要控制运动目标的疏密程度, 即控制每一帧中有几个运动目标.

由于融合密度不同, 因此处理后的视频的时长也

不同,融合密度越小,处理后的视频时长越长;融合密度越大,处理后的视频时长越小,剔除的时间与原视频时间的比值按

$$R = \frac{T_{\text{视频时长}} - T_{\text{算法耗时}} - T_{\text{融合后视频时长}}}{T_{\text{视频时长}}} \quad (12)$$

计算所得.

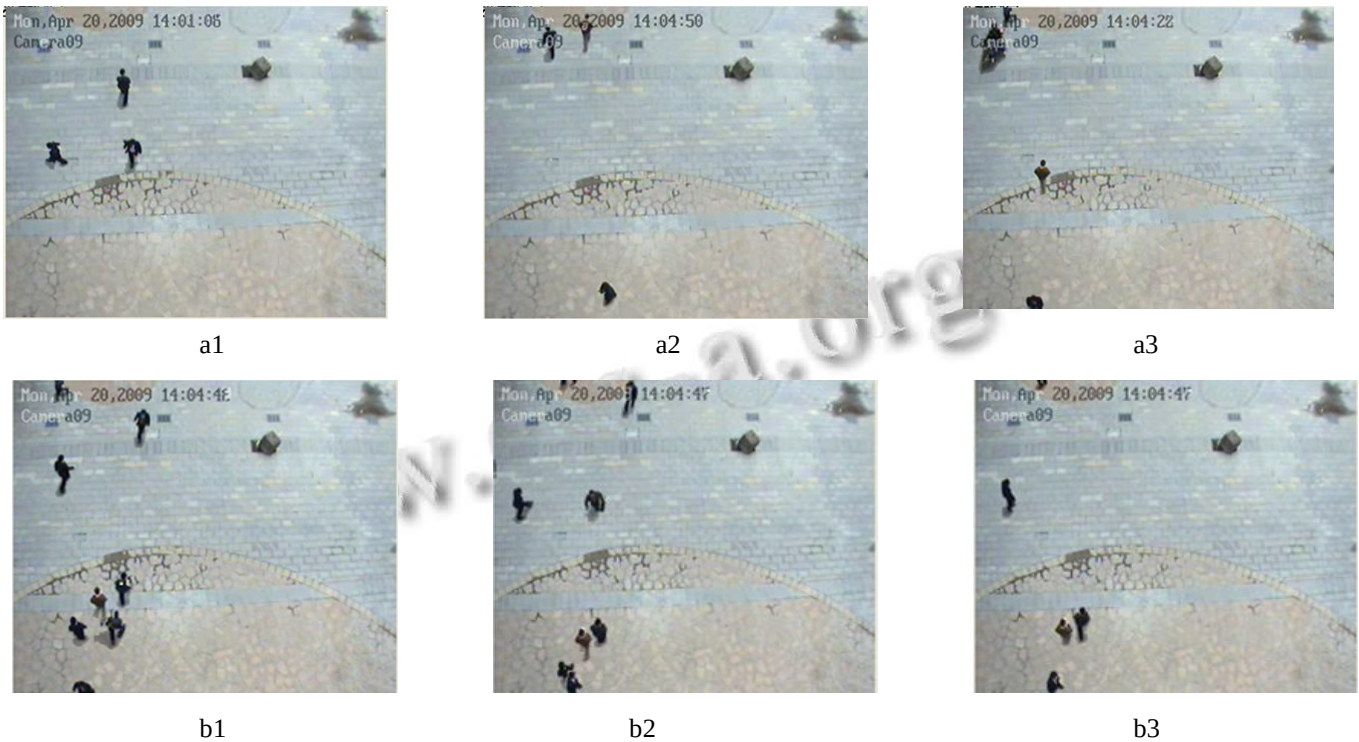


图 4 视频融合结果, a1-a3 是密度为 3 的融合结果, b1-b3 是密度为 7 的融合结果

- 1)从算法耗时上看,每秒可以处理约 35 帧,该算法满足视频实时处理要求;
- 2)从融合后帧数看,该算法能很好的滤除无目标的帧,保留关键信息的帧;
- 3)从剔除的时间与原视频时间的比值上看,该算法能够起到快速视频浏览的效果,压缩的效果也非常显著.
- 从表 2 表 3 对比中可以看出,由于融合过程需要处理每一帧图像,而且,有对象存在的帧的总数是不变的,因此,融合密度的增加,算法耗时变化并不大,但融合后视频帧数下降还是比较明显的.表明,该方法对稀疏对象的快速浏览是适合的.

4 结 论

实验结果表明基于目标分割和融合的快速视频浏览的方法对稀疏对象的视频是有效的,能快速地剔除无目标的帧,提取感兴趣的目 标.在基本保持有用信息的前提下能快速的浏览稀疏对象的监控视频.该方法还引入了背景更新机制,对光照缓慢变化的场景有

很好的适应性.而且,使用了目标检测及密度控制,能很好的滤除无目标的帧,保留目标的关键信息.

表 2 融合密度为 3

视频总 帧数	视频总时间 (s)	算法耗时 (s)	融合后 帧数	压缩时间比
2926	423.14	84.67	309	0.696
3261	465.86	93.81	481	0.650
3209	458.43	92.03	512	0.640

表 3 融合密度为 7

视频总 帧数	视频总时间 (s)	算法耗时 (s)	融合后帧数	压缩时间比
2926	423.14	86.34	212	0.724
3261	465.86	93.42	387	0.681
3209	458.43	95.19	465	0.647

该方法不足之处是在运动对象相互遮挡的情况下提取的目标不够完整,融合后会影 响观看者对目标的
(下转第 129 页)