

一种基于高斯混合模型的视频监控运动物体提取系统

房 胜¹ 钟玉琢²

(山东科技大学计算机科学与技术系 青岛 266510)¹ (清华大学计算机科学与技术系 北京 100084)²

摘 要 现有的视频监控系统必须要依靠人对监控图像的实时观察才能发挥作用。设计并实现了一种基于高斯混合模型的自适应的视频监控运动物体提取系统,使用高斯混合模型对监控场景进行建模,利用对象形状的空间连通性和最小像素尺寸约束去除噪声影响,从而实时地分离出前景中的运动物体并对其进行追踪和计数,仿真实验的结果证明了系统的有效性。

关键词 运动物体提取,前景分割,混合高斯模型,视频监控

A Moving Object Extracting System for Surveillance Based on Mixer Gaussian Probability Model

FANG Sheng¹ ZHANG Yu-Zhuo²

(Department of Computer Science and Technology, Shandong University of Science and Technology, Qingdao 266510)¹

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)²

Abstract Now the working of video surveillance must depend on the man's real time observation of video. This paper designs and implements an adaptive moving object extracting system for surveillance, which explores mixer Gaussian probability process to model the video pixels, uses object spatial connectivity and mini size pixels constraints to reducing noise. The system can effectively isolate the moving object from background and counts the moving object number. The good results of simulate experiments prove that the system is effective for surveillance.

Keywords Moving object extraction, Foreground segmentation, Mixer Gaussian models, Video surveillance

1 引言

视频监控系统在生活中具有重要的意义,它能实时拍摄各个监控地点的图像并传送到监控中心,监控者在监控中心对多个地点的情况进行观察,任何运动物体都会在监控图像中得到反映。但是现有的视频监控系统仍然存在一些需要解决的问题。首先,视频监控系统有效工作的前提是必须有人在监控中心实时观察图像,一旦监控者离去或者注意力从监视器上转移,那么监控系统也就失去了实时监控能力,由于一个监控中心可能会支持几个或者几十个摄像头的工作,这种注意力的转移是不可避免的,甚至入侵者会故意转移监控者的注意力。另外视频图像必须实时的从摄像头传输到监控室,并且保留一段时间,这就要求具有较高的传输带宽和存储空间。

在视频监控中任何入侵都是以运动物体的形式呈现的。只要能够对视频场景中的运动物体进行跟踪并计数,那么监控场景中任何新的运动物体的出现系统都可以自动给出警告,引起监控者的注意。同时新一代的国际视频编码标准,如 MPEG-4,是面向对象的(Video Object)进行编码的,如果能够把视频监控中的运动物体提取出来,编码的效率就可以大大提高,从而节省传输带宽和存储空间。

视频监控中运动物体的提取,从本质上讲包含视频场景中前景/背景的分割(Foreground/Background Segmentation)以及前景运动物体的提取和追踪等步骤。视频场景中运动目标的提取是现在视频编码研究的一个关键问题^[1]。运动目标

的提取主要包含两大类方法,借助人工方式的交互式提取(非自适应提取)和自动提取(自适应提取)。交互式提取必须依靠人工给出初始化的分割轮廓,不适合实时的实现。自动提取是现在研究的主流方向,然而至今没有明确的实现标准,传统的实现技术包括自动边界检测,傅立叶分析等,但是由于运动场中固有的遮挡与显露,孔径,匹配误差等,实现的分割效果并不理想。文[2]提出了使用彩色分割与运动检测的方法,彩色分割采用区域生长与合并算法,使用单一 Gauss 模型的统计特性实现运动检测,获得精确的运动边界,适用于可视电话等运动较小的图像序列。文[3]采用图像序列高阶统计量进行假设检验,三帧序列图像的帧差解决遮挡问题等方法相结合,获得了较好运动物体提取效果,但是计算量大,不适合实时应用。近年来一种新的方法是把视频序列根据运动的共同特征分割成不同的层^[4],这些层包括灰度信息,各层之间的关系,运动信息等,但不仅仅局限于这些信息,所有的信息都以特定的运动模型的各种参数形式表现。文[5]采用高斯混合模型对视频场景的背景区域自适应地进行建模,能够适用于户外光照变化的场景,并且计算量小,但是提取运动物体的边缘不如文[3]的效果好。

在视频监控中视频中背景往往是固定的,运动物体的进入形成了前景,因此自然地可以把视频场景中分层进行运动物体提取的技术应用到其中。但是视频监控运动物体提取系统有自己的特殊要求。首先,视频监控运动物体提取系统中运动物体的提取必须使用自动提取的方式,只有自动提取才能够在监控者不在的情况下捕捉到任何运动物体的进入。另

房 胜 副教授,主要研究领域为视频编解码技术、信息家电等。钟玉琢 教授,博士生导师,主要研究领域是多媒体数据压缩、分布式多媒体体系结构和支撑技术。

外运动物体的提取必须要能够实时进行,这就要求算法复杂度较低,虽然文[2,3]中的方法获得运动物体边缘较为精确,但是在实时性上不能满足要求。文[5]中的高斯混合模型背景建模方法虽然边缘较为粗糙,但是实时性较好。另外由于监控时间长,系统必须适用于长程图像序列,而且系统往往应用于户外,必须适用于光照变化的场景。

本文设计并实现了一个基于高斯混合模型的视频监控运动物体提取系统,可以对视频监控场景中的运动物体进行自适应的实时提取和计数。系统首先对的高斯混合模型进行调整和改进,根据视频监控应用场景的要求,对混合模型的参数进行设定和调整,然后对视频场景中的背景进行建模,利用对象形状的空间连通性和最小像素尺寸的约束去除噪声影响,从而分离出前景中的运动物体并对其进行追踪,同时对根据对象空间连通性获得掩码模板中的运动对象进行计数。实验结果表明在通常的监控场景中系统取得了较好的效果。本文第2节介绍高斯混合模型对背景建模的原理,第3节介绍系统结构及其实现,第4节是试验结果,最后是结论。

2 基于高斯混合模型的视频背景建模

视频背景自适应建模技术中一个经常使用的方法就是对图像序列中的多帧取平均值,这样就可以获得一个与当前背景近似的背景模型。对于包含快速移动物体的图像序列以及背景在图像序列中所占比例较高的场景这种方法是适当的,但是如果图像序列中的运动物体较多并且移动缓慢,那么得到的背景模型误差较大。对于一副图像彩色 M , 含有 R 行, C 列, 对其中的任一像素其色彩分量的值表示为 $I(r, c, k)$, $0 \leq r < R, 0 \leq c < C, k$ 表示色彩分量, 如果场景是固定的, 那么其值也将是固定的。如果在采样过程中引入了一个独立的高斯噪声, 那么在时间序列上, 其值 $I(r, c, k, t)$ 可以用一个基于像素平均值的独立高斯过程进行描述。然而在视频场景中普遍的情况是, 光线会经常地变化, 并不可避免地会出现运动物体, 这种情况下使用一个独立的高斯过程进行描述, 显然难以胜任。

光线的变化意味着图像中所有像素值的整体变化, 而运动物体的出现意味着局部像素方差值的增大, 因此可以考虑采用多个高斯过程的混合模型对色彩值的变化进行描述。假设有 K 个高斯过程, 其中部分是用于描述背景过程的, 部分是描述前景过程的。光线的变化导致像素值的变化, 但是对各个高斯过程之间的方差关系影响不大, 而运动物体则产生较大方差, 影响高斯过程之间方差关系。对每个像素的每个色彩分量值在时间序列上的分布 $I(t)$ 用 K 个高斯过程建模, 当前像素的观测值的概率分布为:

$$P(I(t)) = \sum_{i=1}^K w_{i,t} * G(I(t), u_{i,t}, \Sigma_{i,t}) \quad (1)$$

式(1)中, K 为高斯分布过程数, $w_{i,t}, u_{i,t}$ 和 $\Sigma_{i,t}$ 分别为时间 t 下第 i 个高斯过程的估计权重, 均值和协方差矩阵, G 为高斯概率密度函数:

$$G(I(t), u_i, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(I(t)-u_i)^T \Sigma^{-1}(I(t)-u_i)} \quad (2)$$

为了简化计算, 协方差矩阵采用如下形式进行计算:

$$\Sigma_{i,t} = \sigma_i^2 I \quad (3)$$

即假设色彩分量 R, G, B 的值是独立的, 但是其高斯分布的方差是相同的, 式(3)中 I 为单位矩阵。

式(1)~(3)用来表示视频场景中最近观测到的像素值采用的混合高斯模型特性。一个新的像素值可以用混合模型中的一个或者两个主要分量进行表示, 然后使用该值对分布过程参数进行在线更新。其原理是: 对当前像素值, 在 K 个高斯过程中寻找一个最佳匹配, 或者两个匹配过程。匹配的定义为: 当前像素值在某分布过程标准偏差的一定距离之内。如果没有分布过程与当前像素值匹配, 那么将最不可能的分布过程的平均值以当前像素值代替, 其方差升为一个指定的较高数值, 权重降为初始权重的一半。如果找到匹配的过程, 相应的参数需要调整。显然在 K 个高斯过程中匹配过程 i 的权重需要上升, 而其他的过程的权重则要降低, 公式如下:

$$w_{i,t} = (1-\alpha)w_{i,t-1} + \alpha(\beta_{i,t}) \quad (4)$$

其中, $\beta_{i,t}$ 对匹配过程为 1, 其他过程为 0, α 是一个学习因子, 它反映了分布过程各参数变化的速度。匹配过程 i 的均值和方差利用学习因子, 高斯概率函数 G 和前一时刻的参数值进行在线更新, 公式如下:

$$g = \alpha G(I(t) | u_i, \sigma_i) \quad (5)$$

$$u_i = (1-g)u_{i,t-1} + gI(t) \quad (6)$$

$$\sigma_i^2 = (1-g)\sigma_{i,t-1}^2 + g(I(t)-u_i)^T(I(t)-u_i) \quad (7)$$

式(4)~(7)表明了, 在时间序列上, 与当前时间越接近的值受到的重视程度越高, 对高斯过程参数的影响越大。

3 视频监控运动物体提取系统

3.1 视频监控运动物体提取系统结构

本文实现的视频监控运动物体提取系统的结构如图 1 所示。对于输入的视频首先进行彩色空间的转换, 系统接受两种彩色空间格式输入: RGB 或者 YUV(4:2:0)。如果输入的是 YUV 格式将其转化为 RGB 格式。对 YUV 格式中的两个色彩分量使用双线性差值方式进行尺寸的调整, 然后利用 matlab 中相应的函数转化为 RGB 格式。

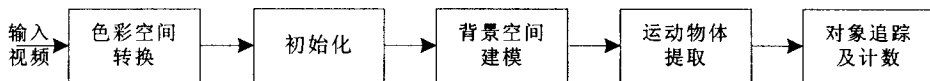


图 1 视频监控运动物体提取系统结构图

初始化工作对于系统的运行具有重要的意义。对于视频监控运动物体提取系统来说, 每一个摄像头监控的范围是大体确定的, 也就是背景确定, 而对于运动物体的出现方式以及运动物体在场景中所占比例的估计有助于确定运动参数。式(4)中的 α 学习因子反映了对当前像素值的重视程度, 该值越大高斯混合模型各过程参数变化的速度越快。如果运动场景中的运动物体速度较快, 可以选用较高的 α 值, 这样高斯过程

的参数调整速度加快。考虑到计算量和运动物体提取的效果, 选用 $K=5$ 个高斯过程描述每个色彩分量。对于高斯过程的初始均值和方差通过随机数生成, 但是采用相对较高的方差值, 这样图像序列中的开始几帧被认为全部都是背景。各个过程的初始权重均为 0.2。

在初始化工作完成后, 进行对背景空间的建模。对于实时输入的每一帧根据第 2 节所述的混合高斯模型逐帧逐像素

进行统计。在判断匹配过程时,使用 2.5 倍的标准偏差作为匹配标准。在对背景空间建模时,必须利用第 2 节各公式得到的参数确定混合模型中哪些高斯过程可以用于对背景建模,即确定能够有效表示当前图像背景的分量过程。方法是根据得到的权重和方差两个数值对各个过程进行排序。根据 w/σ 的数值各高斯过程降序排列。该值的大小与权重成正比,与方差成反比。如果权重较大,表明过程匹配较好,如果方差较低表明运动物体出现的可能较小。这样最可能是背景过程的高斯过程排在队列之前,而最不可能是背景过程的高斯过程排在队列最后,并随不匹配现象的出现被新的分布过程所取代。据此,可以从排序后的 K 个高斯过程中选择 N 个分布过程作为背景模型:

$$N = \arg \min_n \left(\sum_{i=1}^n w_i > T \right) \quad (8)$$

T 代表了背景在图像中所占的比例。

在具体实现时,并不对背景进行直接的建模,而是在运动物体提取过程中,利用式(8)的原理,从排序过程的队列中从后向前选取 2~3 个过程,如果其所占权重超过一定值,认为是运动物体,这是基于运动物体与背景过程不匹配,从而更新不可能匹配过程的参数。同时,由于运动物体与背景的交互作用,在提取的运动物体中不可避免地会出现噪声,同时获得的运动物体的二值掩码模板中也会出现空洞,对此通过 3×3 的中值滤波以及物体空间连通性进行噪声去除。在这个过程中需要考虑视频监控运动物体应用的场景,因为摄像头的视角和距离问题,运动物体的尺寸在各个场景中变化较大。根据各个摄像头的工作方式,选择一个最小像素尺寸,当对二值掩码模板的空间连通处理完毕后,如果对象的像素尺寸小于最小像素尺寸,那么认为该对象为噪声并去除该对象。

物体的追踪在获得运动物体的掩码模板后进行,采用传统的视频编码标准,如 h. 263+ 的运动估计技术。考虑到实时性的要求以及前面各处理步骤以及获得的信息,不必对所有的像素进行运动估计,只要针对最终掩码模板中获得的连通物体进行即可。将连通对象划分为 8×8 的块,针对每一块的中心计算当前帧和前一帧之间的运动,而对运动物体的计数通过对连通对象的个数进行计数即可完成。

3.2 视频监控运动物体提取系统的算法流程

对视频监控运动物体提取系统采用 matlab 进行了模拟



图 2 Hall 序列检测结果(从左至右依次为第 18,19,86,246,247 帧)



图 3 Hall 序列应用掩码模板得到的运动物体结果(从左至右依次为第 18,19,86,246,247 帧)

Hall 测试序列长度为 300 帧,考虑监控场景的特点人物较小,所以空域形状连通性设置为 8 连通,最小连通对象尺寸为 16×7 。学习因子 α 设为 0.3。实验结果如图 2,3 所示,其

实现,其算法流程如下:

```

算法流程 MovObjExtract {
    输入 视频信号,彩色空间转换;
    初始化 { %描述 RGB 三颜色分量的 K 个高斯过程参数的初
        始化
        %R,C,K 分别为帧的行列数,以及高斯分布过程的个数
        K=5;
        mu = zeros(3, K, R, C); % R,G,B 三分量的均值
        sig = zeros(1, K, R, C); % 方差,假设 R,G,B 三分
        量相同
        wgt = zeros(1, K, R, C); % 权重,假设 R,G,B 三分
        量相同
        其他参数如  $\alpha$  等值的设定;
    }
    For 每一输入帧 {
        For 每一像素 { %借助 matlab 的矩阵功能,三色彩分量
            一起处理
            for klass=1:K, % 每一个高斯过程差值,偏差,协方
                差,概率
                vec_diff = currpix - mu(:, klass, r, c); % cur-
                rpix 包含三色分量数值
                dist(klass) = (vec_diff' * vec_diff) / (6.25 * sig
                (1, klass, r, c));
                cov = diag(ones(1,3)/sig(1, klass, r, c));
                prob(klass) = exp(-0.5 * vec_diff' * cov *
                vec_diff) / sqrt((sig(1, klass, r, c)^3) * (2 * pi)^
                3);
            end
        } % For 每一像素 end
        if 过程匹配
            更新匹配过程权重,所有分布过程的权重归一化;
            更新匹配过程均值,方差; %公式(4)~(7)
        else % 没有匹配过程
            更新最不可能匹配过程的权重,均值,方差;
            所有过程的权重归一化;
        end
        确定描述运动物体的高斯过程并计算在各像素上所占
        的权重;
        根据权重是否大于 0.1 获得初始的掩码模板;
        % 利用 matlab filter2, bwfill, bwlabel 等函数实现
        掩码模板滤波,去洞,连通性,最小尺寸约束等处理;
        运动物体追踪,计数
    } % For 每一帧 end
} % MovObjExtract end
    
```

4 试验结果

对实现的视频监控运动物体提取系统进行了模拟测试。测试对象为图像序列 Hall 和 Singer。Hall 是实际的视频监控场景,包含两个人物的运动;Singer 序列中背景不变,符合视频监控图像的特点,同时运动对象较多,而且包含膨胀,收缩,选装等更复杂的运动,更可以检测系统提取运动物体的效果。

中第 18,19 帧为第一个人物出现,246 和 247 为第一个人物消失的关键帧,第 86 帧为中间两个人物同时出现的场景。从图 2 可以看出运动的人物均被检测出来,并且计数正确,计数

显示为图2中白色的方框。从图3中可以看出,基于K个高斯混合模型的视频监控系统的一些特点,第86帧和247帧的掩码图像中人物只有部分被显示出来,这是由于当一个对象在场景中停留的时间较长时,该部分就变为背景中的一部分,86帧和247帧中人物由于是沿摄像头光轴方向运动,部分区域长时间为运动物体的同一部分,从而被集成到背景模型中。从图2可以看出第18帧人物已经显露出一小部分身体,在第19帧中出现的人体被检测出来,这是由于运动对象出现时,混合高斯分布模型对参数的调整需要一个缓冲时间,在本测试中延迟为2帧间距。

Singer测试序列为90帧,由于运动物体的尺寸较大,所以空域形状连通性设置为8连通,最小连通对象尺寸为20*

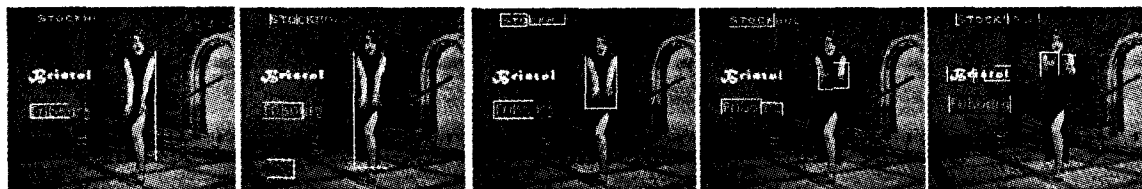


图4 Singer序列检测结果(从左至右依次为第13,15,17,20,23帧)

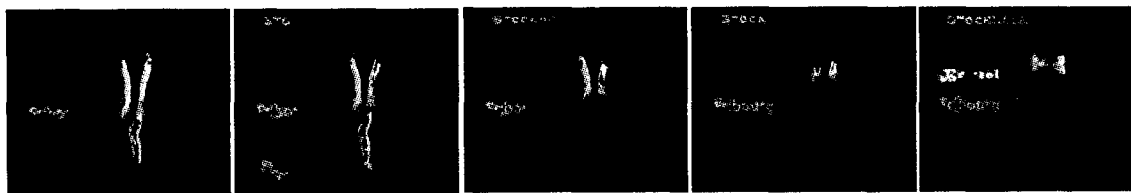


图5 Singer序列应用掩码模板得到的运动物体结果(从左至右依次为第13,15,17,20,23帧)

从以上结果可以看出,实现的基于高斯混合模型的视频监控运动物体提取系统,能够对运动物体实现较准确的提取,计数也基本正确。同时通过滤波,连通性,最小像素约束等处理操作,得到的掩码图像噪声相当小。

结论 本文设计并实现了一个基于高斯混合模型的视频监控运动物体提取系统,在此系统中利用K个高斯分布过程的混合模型对视频场景的R,G,B彩色空间进行建模,通过对当前像素值与各个分布过程的对比确定是否存在匹配的分布过程,并进行相应的各个过程的参数更新。在对各分布过程进行基于权重和方差比值的排序后,使用队列中的后几个分布过程确定运动物体的掩码模板,然后使用滤波,连通性和最小尺寸约束消除噪声影响。试验结果表明了设计和实现的系统能够有效的分离出监控场景中的运动物体,并进行较正确的计数。当然由于实时性的要求,得到的运动物体边缘并不精确,当两个运动物体比较接近时会合并为一个物体。今后系统的重点是在保持实时性的基础上,提高运动物体分割的

精度,并进一步研究各参数对提取结果的影响。

20。学习因子 α 设为0.5。图4,5显示了部分连续视频帧中运动物体的提取情况。如果从视频编码的Video object角度看,唱歌的女歌手和4个不断随机膨胀收缩并运动的单词组成了运动前景,但是从图5可以看出,在本系统中由于歌手的下半身从缓慢运动变为静止,所以在某些帧中高斯混合模型在构建背景模型时,把歌手的下半身集成到了背景中,显示了本系统对运动物体处理的有效性,即只有运动物体才会在本系统中得到反映。从图4,5中可以看出,对人体运动的检测与实际运动部分切合,从较大范围到较小范围逐渐变化,定位较准确。另外对于正在发生膨胀,旋转等运动的单词,系统也较为准确地提取出来,表明系统不仅使用与简单的刚体运动,还适用于复杂运动。

精度,并进一步研究各参数对提取结果的影响。

参考文献

- 1 Haskell B G, Howard P G, Lecun Y A, et al. Image and video coding-emerging standards and beyond [J]. IEEE Transaction On Circuits and Systems for Video Technology, 1998, 8(7):814~836
- 2 潘锦辉,廖庆敏,林行刚. 视频序列中运动目标的自动提取[J]. 清华大学学报(自然科学版),2001,41(4/5):190~193
- 3 杨莉,张弘,李玉山. 视频运动对象的自动分割[J]. 计算机辅助设计与图形学学报,2004,16(3)
- 4 Wang J Y A, Adelson E H. Representing Moving Images with Layers [J]. IEEE Transactions on Image Processing Special Issue; Image Sequence Compression, 1994,3(5):625~638
- 5 Stauffer C, Grimson W E L. Adaptive Background Mixture Models for Real-Time Tracking [A]. In:CVPR '99 [C], Fort Collins, CO, June 1999,2:246~252