

基于改进的混合高斯模型的运动目标提取

杨宁, 杨敏

(南京邮电大学自动化学院, 江苏南京 210046)

摘要:背景提取技术是图像与视频处理中的关键技术。文中对静态背景下运动目标的提取算法进行了研究。混合高斯算法在近年得到了广泛的关注,但是算法使用固定个数的分布建模,在实际中不能满足最优模型,并且模型对学习率的调整比较敏感。文中提出改进的自适应算法提取前景运动目标,其中主要针对模型中的混合高斯分布的个数及学习判别准则进行了改进。实验证明,该改进算法相比传统算法有着较好的自适应性并且检测效率较高。

关键词:背景建模;运动目标提取;混合高斯模型;序列图像分析

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2012)07-0020-04

Moving Object Extraction Based on Improved Gaussian Mixture Model

YANG Ning, YANG Min

(College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210046, China)

Abstract: Background extraction is a key step for image and video processing technology. In this paper, the moving object extraction in the static background is studied. In recent years the Gaussian mixture algorithm received extensive attention. The traditional algorithm model each pixel a fixed number of components, which is not optimal in term of detection and computational time. And the algorithm is sensitive to the adjustment of the learning rate. In this paper, improved adaptive algorithm is put forward for moving object extraction. The major improvement is the number of mixture Gaussian components and the discriminant criterion. The experiment results show that the improved algorithm is better than traditional algorithm in both adaptability and computing speed.

Key words: background modeling; moving object extraction; mixture Gaussian model; sequential image analysis

0 引言

在许多视频监控系统及多媒体应用系统中,对视频处理的第一步就是从视频序列中分离出所要观察的目标,即运动目标提取。通常是对背景建模,然后从中检测运动物体。背景模型提取^[1]的准确与否,直接关系到后续目标处理与分析结果的准确性,因此研究出良好的背景提取算法和自适应的背景更新模型是非常重要的。

近年来,研究者针对实际的应用场景研究出许多不同的背景建模方法,例如帧间差分算法^[2]、线性预测法^[3]、核密度估计法^[4,5]及混合高斯模型法^[6]。帧间差分法实现简单、运算速度快,然而这种方法只在特定情况下(比如目标运动缓慢以及帧速率较低时)有效,并且该方法对阈值的选取特别敏感,不适用于动态复

杂场景中的目标提取;线性预测法使用采样协方差来估计参数,算法运算复杂;核密度估计算法本身并没有用到相关数据分布的先验知识,能够较好地适用于变化的场景中,然而算法需要占用大量内存;混合高斯模型法是 Chris Stauffer 和 W. E. L. Grimson^[7]提出的一种自适应的背景建模方法,该方法在近几年受到了广泛关注,它假设像素随时间变化的过程是服从高斯分布的,即对每一个像素点用高斯函数来建模(在 RGB 空间),然后将新的一帧图像中的像素点的高斯模型与背景像素的高斯模型进行比较。

混合高斯背景建模方法可以满足在场景中光照和亮度的变化时的目标检测以及动态的更新方程,并且能够克服一定的噪声影响。文中针对模型建立及方程更新时算法本身存在的几个问题给予改进,以便提高目标检测的准确性。

1 混合高斯背景建模

1.1 算法描述

混合高斯模型的基本思想是对于从视频中获取的

收稿日期:2011-12-20;修回日期:2012-03-24

基金项目:南京邮电大学攀登计划(NY208050)

作者简介:杨宁(1987-),女,硕士研究生,主要研究方向为计算机视觉;杨敏,博士,副教授,主要从事计算机视觉和图像理解的研究工作。

图像序列,用多个高斯分布对每一个像素点建模,也就是定义 K 个高斯分布来表示该像素点的状态,而不是像其他方法那样只选取一种特定的分布。这样,根据这些高斯分布的变化来确定哪些像素点属于前景,哪些是属于背景像素的。对于已定的 K 个高斯分布,若新的像素点不符合其中的任何一个分布,则该像素点被判为前景像素,直到有足够的证明能让其成为一个新的背景模型分布为止。

对图像中的每一像素 $\{x_0, y_0\}$, 其随时间的变化序列为 $\{y_1, \dots, y_2, \dots, y_i\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\}$, I 是图像序列。

在 t 时刻观察到该像素的概率:

$$p(y) = \sum_{j=1}^K \omega_j G(y, \mu_j, \Sigma_j) \quad (1)$$

K 是混合分量的个数, K 值越大, 越能表示更加复杂的场景, 但同时也增加了计算量, 影响实时效果; ω_j 是每一个分量的权重, $G(y, \mu_j, \Sigma_j)$ 是均值为 μ , 协方差为 Σ 的高斯分布。出于简化计算的考虑, 通常假设 RGB 颜色分量是相互独立的并且其方差相等。这样, 协方差有如下形式: $\Sigma = \sigma I$ 。

给定一个视频帧, GMM 模型通过在线近似 K 均值算法进行更新 ($k = \operatorname{argmin} d_j$)。

一个新的观测像素值通过两个步骤进行分类。首先, 计算每一个高斯分量的马氏距离

$$d_j = \|y' - \mu_j\|_{\Sigma_j} = \sqrt{(y' - \mu_j)^T \Sigma_j^{-1} (y' - \mu_j)} \quad (2)$$

从中找出距离最近的分量:

$$k = \operatorname{argmin} d_j \quad (3)$$

当距离小于给定阈值时就认为此像素与分量匹配: $d_k \leq \theta$ (文献[6]中 $\theta = 2.5$)

这样对每一个像素分类完成后就得到一个在线更新模型。如果没有与该像素匹配的分量, 权重最小的分量被一个新的高斯替换出, 这个高斯的均值为 y' , 并给予一个大的方差和小的权重。

如果 y' 与第 k 个分量匹配, 那么就更新参数:

$$\mu_k \leftarrow (1 - \alpha)\mu_k + \alpha y' \quad (4)$$

$$\sum_k \leftarrow (1 - \alpha) \sum_k + \alpha (y' - \mu_k) (y' - \mu_k)^T \quad (5)$$

$$\omega_{j,i+1} \leftarrow (1 - \beta)\omega_{j,i} + \beta \quad (6)$$

对于不匹配的高斯分量, 其均值和协方差不作调整, 而只是稍微减小权重:

$$\omega_{j,i+1} \leftarrow (1 - \beta)\omega_{j,i} \quad (7)$$

可以看出这个更新模型相当于一批观察值呈指数衰减权重的 K 均值估计, 这就允许混合模型根据背景变化自适应调整。模型中有两个重要的参数: 学习率

α 和 β 。 α 控制每一个分量随背景变化调整的速度, β 控制各个分量趋于相关的速度 (原始 GMM 算法中设置 $\alpha = \beta G(y', \mu_k, \Sigma_k)$)。

将混合分量以权重的降序排列, 选择前 B 个分量, 这些分量的权重满足 $\sum_{j=1}^B \omega_j > T$ (T 是背景所应包含的数据的最小比率)。最终, 若观测像素的似然估计低于背景参量, 则被标记为背景像素, 否则标记为前景像素。

1.2 算法优缺点

混合高斯模型^[8,9]采用 RGB 空间对图像建模, 这在后期的处理中有利于处理阴影的检测。

该模型在衡量目标时使用的参数为分布函数的均值和方差, 这样相当于对每个像素点给定其特有的阈值, 不像帧间差分算法对所有像素点采用同一阈值^[10], 使得背景的提取受局部影响。模型表示使用多模分布能够比较好地适应光照缓慢变化的场景, 学习率 α 的选取决定了适应光照变化的快慢。然而, 当有新的物体加入背景或背景中有物体移除时, 该学习率也影响了模型的更新速度, 两者是约束的, 需要折中考虑。

该算法在初始化时, 模型对所有的像素点选取固定个数的高斯分布, 这在实际应用中并不是最优模型, 因为其多模性是随时间空间变化的^[11]; 对均值、方差和权重进行初始化估计^[12]时, 要求训练帧序列最好没有移动目标, 这在实际情况中也不易实现; 另外, 参数估计还要求较大的内存来存储训练数据。

总之, 原始的像素级处理的混合高斯模型在白天及多模场景下能够比较好地提取出背景, 对于动态背景、有物体加入背景及前景目标短暂停留等情况下其性能一般, 不适应处理噪声图像、摄像机抖动、光照突变、物体自展、遮挡、前景目标有空洞及有阴影等情况。

2 改进的混合高斯模型

在混合高斯模型中, 权重 ω_k 描述了某数据属于第 k 个分量的程度, 可以看作是某样本来自该分量的概率。这些权重描述了一个隐含的多项式分布。假设有 t 个样本, 分别对应 GMM 中的每个分量, 属于第 k 个分量的样本数目是 $n_k = \sum_{i=1}^t M_i$, 由 n_k 的多项式分布得到似然函数 $\zeta = \prod_{k=1}^K \pi_k^{n_k}$, 引入拉格朗日乘数 λ , 最大似然估计为

$$\frac{\partial}{\partial \omega_k} (\log \zeta + \lambda (\sum_{k=1}^K \omega_k - 1)) = 0 \quad (8)$$

消掉 λ , 得到权重更新的递归形式

$$\omega_k^t = \omega_k^{t-1} + 1/t (M - \omega_k^{t-1}) \quad (9)$$

引入先验概率知识,混合权重的更新方程为

$$\omega_k \leftarrow (1 - \beta)\omega_k + \beta M - \beta c_T \quad (10)$$

其中, $c_T = c/T$, c 是共轭先验的相关系数^[13]。这样混合权重就可以自适应地调整。

针对不同情况下的学习率,可以给定一个时间限制,即若有移动缓慢的物体加入背景,则给定一个阈值,在此时间阈值内该物体一直符合背景分布时,才将其归为背景,否则就判定为前景。这样就可以解决学习率的选取的矛盾现象,即当有新的物体加入背景或背景中有物体移除时,不会因学习率选取稍大而影响模型的更新速度,也不会因学习率稍小而适时地更新背景模型。

针对高斯分布的个数,文中通过分析相关工作的研究现状,采用根据各个像素值出现频率的不同而更新 K 值的方法。在模型的更新阶段,设定一个参数加入自适应的更新方程。具体做法是:在开始阶段,将所有像素点的高斯分布的个数都设定为 $K = 1$,在后续帧中各个坐标点像素值出现的频率各不相同,采用模型自适应调整并且规定每个像素点坐标处的分布个数最大为 K_{max} 。其中高斯分布各参数的初始化是用第一帧的图像数据来完成的,显然初始阶段高斯分布的权重为 $1/K_{max}$,通常方差的取值比较大,以适应后面背景模型的加入。这样在一定程度上能够降低计算量,增

强实时性。

3 实验分析

图 1 是文中算法与原始的混合高斯背景提取方法进行实验仿真的结果图。实验中原始算法选取 k 值固定为 4,匹配阈值为 2.5,改进算法自适应地调整 k 值,视频每帧图像的大小为 360×240 。需要说明的是测试视频本身存在树叶的扰动及缓慢的光照变化等影响。

实验运行条件:MATLAB7.0,算法参数估计部分使用 C 语言的衍生程序 MEX 文件进行实现,CPU 型号为 Intel(R) Core(TM) i3,2.27GHz,2GB 内存。原始算法平均每帧的处理时间为 23ms,改进后处理时间约为 18ms,并且该仿真只是用于比较改进算法提取前景目标的效率,并没有使用任何预处理技术。

实验中,学习率因子 α 和 β 分别取 0.01 和 0.2,并且引入时间阈值来保证算法的稳定性及收敛速度。由实验结果分析可知,原始算法提取的前景目标存在一定的误检,比较序列图像第 80 帧的目标提取结果,原始方法对部分光照强烈的区域有漏提取现象,导致目标提取不完整;而序列第 14 帧及 24 帧则明显地将部分阴影也当成了前景目标。

综合实验结果,改进的方法在运算效率上有较大提高。



图 1 改进的混合高斯模型的运动目标提取

4 结束语

文中在背景建模过程中,考虑到算法的运算效率及模型的实时性要求,对每一个像素建模时采用自适应更新分布个数的方法;为了保证模型的稳定性、克服算法对学习率调整的敏感性,引入一个时间阈值,能够在一定程度上减少误判。经过实验分析,改进的方法对于背影扰动有较好的抑制效果,并且,通过参数配置的改进,实验的运行速度也有了较大提高,更适合应用于实时的目标检测场景的背景提取。

参考文献:

- [1] 王陈阳,周明全,耿国华. 基于自适应背景模型运动目标检测[J]. 计算机技术与发展,2007,17(4):61-63.
- [2] Long W, Yang Y. Stationary background generation: an alternative to the difference of two images[J]. Pattern Recognition, 1990, 23(12):1351-1359.
- [3] Halevy G, Weinshall D. Motion of disturbances: detection and tracking of multi-body non-rigid motion[J]. Machine Vision and Applications, 1999, 11(3):122-137.
- [4] Rosenblatt M. Remarks on some nonparametric estimates of a density function[J]. Annals of Mathematical Statistics, 1956, 27(3):832-837.
- [5] Parzen E. On estimation of a probability density function and model[J]. Annals of Mathematical Statistics, 1962, 33(3):

(上接第19页)

3 结束语

在整个实现过程中,充分考虑到二胡音乐的乐理特性。在基频识别的过程中,使用了离散小波变换和快速傅里叶变换联合处理操作,并通过功率谱函数的操作,最终准确地得到各个音符对应的基频,而在 MIDI 音乐合成的过程中,根据二胡的音域范围以及双门限的灵活设置则能十分方便准确地得到有效音符,并最终生成所需要的 MIDI 文件。

参考文献:

- [1] 徐国庆,杨丹. 小波变换与 FFT 联合识别乐音[J]. 重庆大学学报(自然科学版),2005,28(12):51-54.
- [2] Daubechies I. 小波十讲[M]. 李建平,杨万年译. 北京:国防工业出版社,2004:12-15.
- [3] 李晨,周明全. 音频检索技术研究[J]. 计算机技术与发展,2008,18(8):215-222.
- [4] 李海东,李青. 基于阈值法的小波去噪算法研究[J]. 计算机技术与发展,2009,19(7):56-58.
- [5] 陈峰,成新民. 基于小波变换的信号去噪技术及实现[J]. 现代电子技术,2005(3):12-15.

1065-1076.

- [6] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[C]//Proc of Computer Vision and Pattern Recognition. USA:IEEE,1999:252-258.
- [7] Stauffer C, Grimson W E L. Learning Patterns of Activity Using Real-time Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8):747-757.
- [8] 宋雪桦,陈瑜,耿剑锋,等. 基于改进的混合高斯背景模型的运动目标检测[J]. 计算机工程与设计,2010,31(21):4646-4649.
- [9] 黄鑫娟,周洁敏,刘伯扬. 自适应混合高斯背景模型的运动目标检测方法[J]. 计算机应用,2010,30(1):71-74.
- [10] 文灏,陈红涛. 基于减背景与对称差分的运动目标检测[J]. 微计算机信息,2007,23(9-1):99-101.
- [11] Zivkovic Z. Improved Adaptive Gaussian Mixture Model for Background Subtraction[C]//Proceedings of the 17th International Conference on Pattern Recognition. USA: IEEE, 2004:28-31.
- [12] Zivkovic Z, van der Heijden F. Efficient adaptive density estimation per image pixel for the task of background subtraction[J]. Pattern Recognition Letters, 2006, 27(7):773-780.
- [13] Zivkovic Z, van der Heijden F. Recursive unsupervised learning of finite mixture models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(5):651-656.

- [6] Ito N, Nishimoto K. Voice-to-MIDI System for Singing Melodies with Lyrics[C]//Proceedings of the international conference on advances in computer entertainment technology. [s. l.]:[s. n.],2007.
- [7] Xu J W, Principe J C. A novel pitch determination algorithm based on generalized correlation function[C]//Proc. of Machine Learning for Signal Processing. [s. l.]:[s. n.],2007.
- [8] Xu Jianwu, Principe J C. A Pitch Detector Based on a Generalized Correlation Function[J]. IEEE Trans on Audio, Speech and Language Processing, 2008, 16(8):1420-1432.
- [9] Bello J P, Daudet L, Abdullah S, et al. A tutorial on onset detection in music signals[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5):1035-1047.
- [10] Hu N, Dannenberg R. Bootstrap learning for accurate onset detection[J]. Machine Learning, 2006, 65(7):457-471.
- [11] Toh C C, Zhang B, Wang Y. Multiple-feature fusion based on onset detection for solo singing voice[J]. Proc. of ISMIR, 2008, 32(9):515-520.
- [12] Lacoste A, Eck D. Onset Detection with Artificial Neural Networks for MIREX 2005[C]//Extended Abstract of the 1st Annual Music Information Retrieval Evaluation Exchange (MIREX 2005). London:[s. n.],2005.