

一种改进的视频兴趣区目标捕获算法研究

周 浩,周 平,胡耀武
(浙江理工大学信息学院,杭州 310018)

摘 要: 对于视频兴趣区的目标捕获技术提出了一种基于颜色梯度的背景差方法,通过结合阴影去除和连通域分析捕获兴趣目标。实验结果表明:该方法能有效解决光照突变等干扰,较好捕获兴趣区目标,减小误检率。对所实验的 6 个视频进行分析,在最多同时出现 5 个目标的视频检测中取得了 92.8% 的准确率、2% 的误检率,且基本满足实时性。

关键词: 目标捕获; 背景建模; 混合高斯模型; 颜色梯度

中图分类号: TP319 **文献标识码:** A

0 引 言

视频兴趣区目标的自动捕获技术是智能监控系统的前提。这里可以将兴趣目标捕获技术分为目标检测和目标识别两部分。每个视频跟踪系统都是从目标检测开始的,跟踪处理的前提和关键是如何正确地将运动目标从背景中准确的分离出来。所谓目标检测就是把目标物体对应的区域从视频序列中去除背景内容,并判断其是否为兴趣目标。在监控系统中,因为背景长时间固定不变,所以背景差分法效果很好。同时,目前许多研究都致力于背景模型的建立,用以减少背景对兴趣目标检测准确性的影响。常用的背景建模方法有:单高斯背景建模法^[1]、混合高斯背景建模^[2]、基于码本的背景差分法^[3]、非参数核密度估计法^[4]和基于贝叶斯模型的背景建模法^[5]等等。混合高斯模型假设场景中的每个像素点的变化的概率分布可以用 K 个高斯分布拟合,通常该方法能够很好地模拟周期性变化的像素点。但对于非周期变化或者复杂变化的像素点检测精度低。如果利用空域信息^[6]则能够在一定程度上改善由单个背景像素点模拟不准确造成的误检^[7-8]。由于场景的动态变化,如天气变化、光照变化、树叶摆动、水面波

浪、摄像机抖动等,使得运动目标捕获仍然是一项困难的工作。对此,针对高斯混合模型只对像素建模的缺陷,提出了一种基于颜色梯度的背景差分法,并结合阴影去除算法和连通域分析算法,该方法充分考虑了像素之间的相关性,通过实验看到即使在发生光照突变的情况,也能准确地捕获兴趣目标。

1 高斯混合模型

对于每个像素点使用 $k(k < K)$ 个独立的高斯模型来表征图像中各个像素点的特征, k 值越大,处理波动能力越强,所需处理时间越长。 K 中每个状态都可以用一个高斯函数来表示,这些状态一部分表示背景像素点,其余部分表示前景像素点。对于复杂的像素点,需要使用较多的背景模型,而对于帧间相似的像素点,使用一个背景模型就可以准确地反映真实的场景。混合高斯模型的概率密度函数为:

$$p(x_t(i, j)) = \sum_{m=1}^k w_{m,t}(i, j) \cdot \eta(x_t(i, j), \mu_{m,t}(i, j), \sum_{m,t}(i, j)) \quad (1)$$

式(1)中 $x_t(i, j)$ 为 t 时刻的某一个像素值, $w_{m,t}(i, j)$ ($\sum w_{m,t}(i, j) = 1, m = 1, 2, \dots, k$) 用来描

述每个像素点 (i, j) 的 k 个高斯分布分别具有不同权值,其中 $\eta(x_t(i, j), \mu_{m,t}(i, j), \sum_{m,t}(i, j)) = \frac{1}{(2\pi)^{n/2} \left| \sum_{m,t}(i, j) \right|^{1/2}} e^{-1/2(x_t - \mu)^T \sum_{m,t}(i, j)^{-1} (x_t - \mu)}$, $m = 1, 2, \dots, k$ 是 t 时刻的第 m 个高斯分布,其均值为 $\mu_{m,t}(i, j)$,协方差矩阵为 $\sum_{m,t}(i, j)$,由于 $\frac{\omega}{\sigma}$ 越大,与背景匹配的可能性越大,为了减少一些繁琐的计算,一般按照式(2)将 k 个高斯分布按照 $\frac{\omega}{\sigma}$ 降序排列,仅取前面的 B 个高斯分布表示背景,其他表示前景分布。

$$B = \operatorname{argmin}_b \left(\sum_{i=1}^b \omega_{i,t} > T \right) \quad (2)$$

对于该像素点的每个新样本点 X_{t+1} ,判断其与已存在的 K 个高斯分布中的哪一个匹配。如果 X_{t+1} 在某高斯分布的个标准差之内,则规定 X_{t+1} 与该高斯分布匹配。如果 X_{t+1} 与前 s 个高斯分布都不匹配,且与第 $s+1$ 个高斯分布匹配,则令 $M_{s+1,t+1} = 1$,且 $M_{i,t+1} = 0 (1 \leq i \leq K \text{ 且 } i \neq s+1)$,第 $s+1$ 个高斯分布按下式更新其权重、期望值和方差;其他高斯分布只是更新权重,期望值和方差不变:

$$\omega_{i,t+1} = (1 - \alpha)\omega_{i,t} + \alpha M_{i,t+1} \quad (3)$$

$$\mu_{i,t+1} = (1 - \rho)\mu_{i,t} + \rho X_{t+1} \quad (4)$$

$$\sigma_{i,t+1}^2 = (1 - \rho)\sigma_{i,t}^2 + \rho(X_{t+1} - \mu_{i,t+1})^T (X_{t+1} - \mu_{i,t+1}) \quad (5)$$

$$\rho = \frac{\alpha}{\omega_{i,t+1}} \quad (6)$$

如果 K 个高斯分布都不匹配,则 $\frac{\omega}{\sigma}$ 最小高斯分布 $\eta_{K,t+1}$ 的期望值用 X_{t+1} ,方差设为较大的值,权重设为较小的值。

采用式(4)~式(6)不断更新高斯分布的期望和方差。当背景自身缓慢变化时,大量背景点还会落入原背景模型的高斯分布,通过式(4)~式(6)不断修正,而趋于新的背景模型。最后将原始视频与背景更新出的视频进行差分,并选取阈值 TH 进行二值化,分离前景与背景。

2 基于颜色梯度的背景差

众所周知,基于颜色的背景差方法以及高斯混合模型均不能够处理光照突变的情况,针对这种情况,提出了一种基于梯度的背景差方法。对于图像中的每个像素,计算 $\Delta = (\Delta_m, \Delta_d)$ 作为特征向量,这里 Δ_m 是梯度的大小, $\Delta_m = \sqrt{f_x^2 + f_y^2}$, Δ_d 是梯度的方

向, $\Delta_d = \tan^{-1} \frac{f_y}{f_x}$,从灰度图像中得到梯度的信息。

因为 $p_{i,j}$ 的颜色值 $x_{i,j}$ 服从正态分布,其相应的灰度值 $g_{i,j}$ 也同样服从正态分布,其均值和标准差分别为 $\mu_{i,j}$ 和 $\sigma_{i,j}$ 。设定 $f_x = g_{i+1,j} - g_{i,j}$ 和 $f_y = g_{i,j+1} - g_{i,j}$,可以发现 f_x 服从正态分布。其均值为 $\mu_{f_x} = \mu_{i+1,j} - \mu_{i,j}$,方差为 $\sigma_{f_x}^2 = \sigma_{i+1,j}^2 + \sigma_{i,j}^2$ 。同样 f_y 也服从正态分布。其均值为 $\mu_{f_y} = \mu_{i,j+1} - \mu_{i,j}$,方差为 $\sigma_{f_y}^2 = \sigma_{i,j+1}^2 + \sigma_{i,j}^2$ 。确定了 f_x 和 f_y 的分布以及其标准分布转换方法,就可以设定特征向量 (Δ_m, Δ_d) 为:
 $F(\Delta_m, \Delta_d) =$

$$\frac{\Delta_m}{2\pi\sigma_{f_x}\sigma_{f_y}\sqrt{1-\rho^2}} \exp\left(-\frac{z}{2(1-\rho^2)}\right) \quad (7)$$

式(7)中,

$$z = \left(\frac{\Delta_m \cos \Delta_d - \mu_{f_x}}{\sigma_{f_x}} \right)^2 - 2\rho \left(\frac{\Delta_m \cos \Delta_d - \mu_{f_x}}{\sigma_{f_x}} \right) \left(\frac{\Delta_m \sin \Delta_d - \mu_{f_y}}{\sigma_{f_y}} \right) + \left(\frac{\Delta_m \sin \Delta_d - \mu_{f_y}}{\sigma_{f_y}} \right)^2,$$

$\rho = \frac{\sigma_{i,j}^2}{\sigma_{f_x}\sigma_{f_y}}$,式(7)中所涉及的参数都可以通过颜色分布的均值、方差求出。给出一个新的图像,同样可以求出其梯度和梯度方向值。通过背景梯度分布获取的梯度矢量概率,若小于公差,则判定该像素为前景像素。反之,则为背景像素。

3 目标阴影去除

在视频场景中,由于运动对象之间的遮挡以及光照不均匀等原因,视频图像中,大多伴有运动阴影,会导致目标丢失以及虚假目标出现等问题,因此阴影的检测和消除是兴趣目标识别的重要环节。相对于RGB颜色空间而言,HSV颜色空间更能直接对应人眼色彩视觉特征的三要素(亮度、色度、饱和度),三个分量相互独立且能反应出灰度和色彩信息,特别对于图像中极亮和极暗的区域也能很好地反映出相关信息。在HSV颜色空间模型中,同一目标在阴影区和不在阴影区的色相是相近的,阴影主要使该区域内色彩的明暗发生变化。阴影去除的基本思想如下:在进行HSV的阴影检测之前,首先需要进行颜色空间转换,然后将阴影的像素值与该点的像素值相比较,如果其包含的响应色度值、亮度值和饱和度值在一定的阈值之内,那么就认为该点是阴影点,判别方法为:

$$\begin{cases} T_1 \leq \frac{I_v(i, j)}{B_v(i, j)} \leq T_2 \\ |I_h(i, j) - B_h(i, j)| < T_h \\ |I_s(i, j) - B_s(i, j)| < T_s \end{cases} \quad (8)$$

式(8)中 $I_h(i,j)$ 、 $I_s(i,j)$ 、 $I_v(i,j)$ 分别表示当前帧的 H,S,V 信息量, $B_h(i,j)$ 、 $B_s(i,j)$ 、 $B_v(i,j)$ 表示 H,S,V 信息量的背景模型, $B_h(i,j)$ 、 $B_s(i,j)$ 、 $B_v(i,j)$ 的计算方法与灰度图像背景模型 $B(i,j)$ 的计算方法相同,而 T_1 则考虑了当前光线的强弱,通常光线越强, T_1 值就取得越小。因此阴影点的 $I_v(i,j)$ 值通常总是小于相应非阴影点的 $I_v(i,j)$ 值,所以 T_2 取值小于 1,用于去除噪声的影响, T_h 表示色度检测阈值, T_s 表示饱和度检测阈值。这种方法由于同时考虑了 H,S,V 分量信息,与基于 RGB 颜色空间或者仅利用灰度信息的方法相比,更加充分的利用了彩色信息。为了减少计算量,提高算法的实时性,阴影检测只对前景部分,在取值比较合理的情况下,可以得到鲁棒的效果,整个算法流程如图 1 所示。

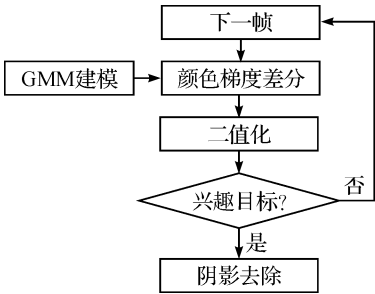


图 1 算法流程图

4 实验结果与分析

为了验证改进算法的有效性,在 PC 机上(CPU 为 Intel Core2 4 核 主频 2.66 GHz,内存 1G)利

用 C# 编写了代码进行实验。采用 10 余个不同场景下的目标检测视频,视频 1 一共包含 1 089 帧,帧大小为 320×240 ,帧率为 25fps,视频 2 一共包含 2 400 帧,帧大小为 640×480 ,帧率为 25fps,视频 3 一共包含 890 帧,帧率为 25fps,帧大小为 320×240 ,在捕获运动目标时,采用 125 帧(5s)进行背景建模设置目标的最小面积为 20 像素。实验参数如表 1 所示。

表 1 实验参数

参数	取值
T_h, T_s	1.2, 1.5
TH	35
T_1, T_2	0.2, 0.5

图 2 是改进方法与传统算法(即高斯混合模型,以下简称“传统算法”)对比较典型的 3 个视频进行目标捕获的实验结果对比图。在第一个视频序列中两辆骑车驶入视野时,图像发生细微变化,在图 2(b)中可以看到,在文献[7]中可以清楚地看到噪点相对减少(图 2(c)),而采用改进方法,在很大程度上较少了噪点,如图 2(d)所示。在第二个视频序列中,由于左方和中间的树叶一直在摆动,甚至由于大风,树的枝干也发生了较大的偏移,对目标的捕获造成了很大的麻烦,传统算法很难将摇摆的树叶完全键入背景,如图 2(f)所示,文献[7]中算法的结果中可以看到,虽然小的噪点能够很好地去除,但是对于晃动幅度过大的树叶而言,效果还是不好,如图 2(g)。而采用改进方法,可以充分考虑相关像素之间

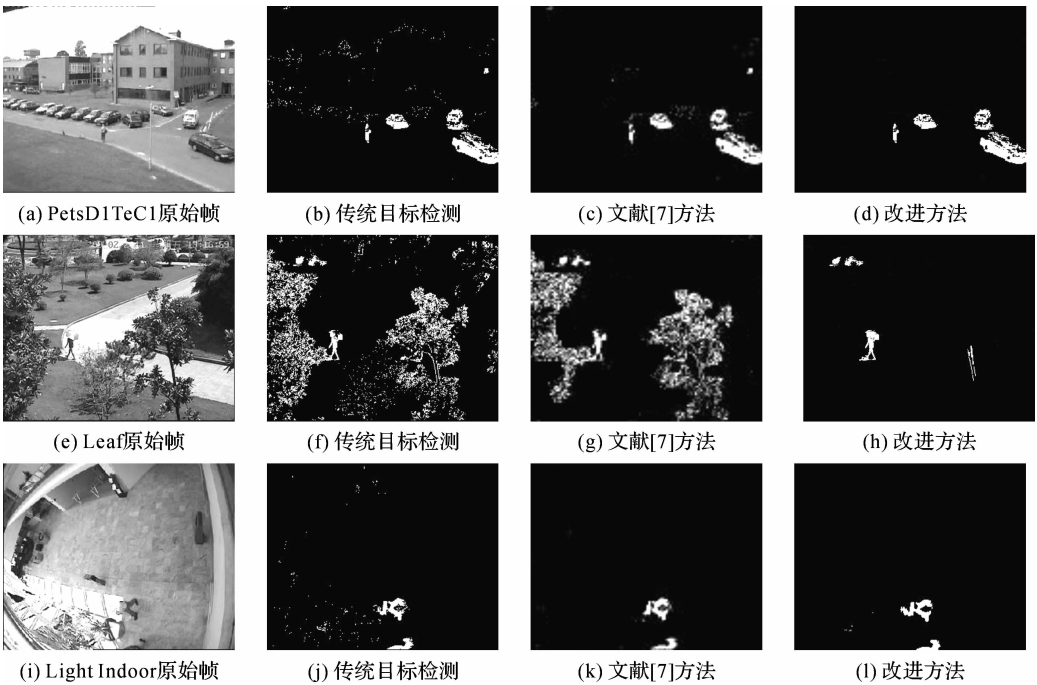


图 2 目标捕获结果

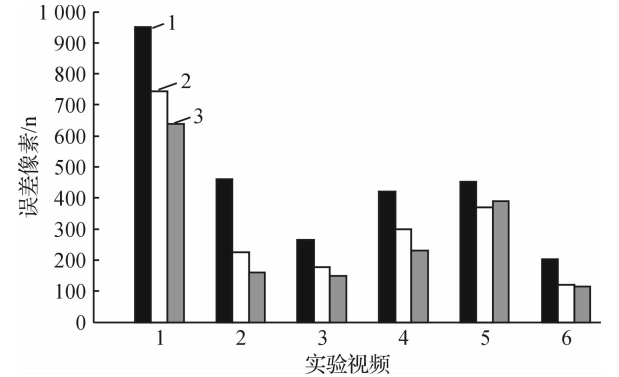
的关系,有效地避免了树叶摆动产生的噪点群,能够清晰地检测出撑伞的行人如图 2(h)。在第三个视频序列中,因为目标身体的部分走进了强光中,使得目标变得模糊,在图 2(j)中明显看到目标部分丢失,在文献[7]中算法的处理下,可以看到效果还是很好的,如图 2(k)。而通过颜色梯度分析进行差分,考虑像素的相关性,能够很大程度上减少了误检的发生,如图 2(l)所示。

表 2 为传统算法、文献[7]中算法与我们的方法对一组针对不同背景的视频的处理一帧平均用时(f/s)的对比结果。从文献[7]中了解到,该方法处理一帧的平均用时大概在 100 ms,通过与传统算法和文献[7]中算法的耗时比较,可以看出改进方法在处理用时上有一定的优越性,基本满足实时处理。

表 2 处理一帧平均用时 单位:s/f

视频名称	视频 1	视频 2	视频 3
传统算法	0.174 25	0.198 34	0.172 81
文献[7]算法	0.172 01	0.179 94	0.171 03
改进算法	0.163 31	0.177 58	0.161 25

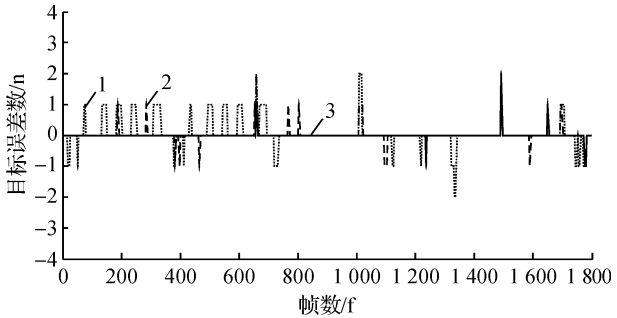
通过文献[7]中算法中给出的模型参数、误差像素算法,对上述 3 个视频以及额外的一组视频进行比较,实验 1,2,3 为上面实验视频,4,5,6 为额外进行的一组视频实验。图 3 可以看出改进算法、较传统算法和文献[7]中算法的精度,在各种不同的背景下,改进方法精度是最高的。



1. 传统方法, 2. 文献[7]方法, 3. 笔者方法
图 3 精度对比结果

图 4 中给出了改进的基于颜色梯度的目标检测算法在不同的视频中对于目标捕获个数的准确度的统计结果。横坐标表示帧数,纵坐标表示捕获的目标和实际目标个数的差,通过将传统方法、文献[7]方法和改进方法进行了以下比较,可以看出,传统算法在有背景干扰的条件下目标捕获上的效果不理想,容易产生误判和漏判,而文献[7]方法在有树叶晃动的场景中不能很好地捕获目标,而利用改进方

法在不同场景的视频序列中进行目标捕获则表现非常稳定,准确率较高,并且一直没有出现连续多帧的误检漏检现象,通过对 1800 帧的数据统计得出 3 种方法的平均准确捕获目标的概率为: 76. 5%, 84. 1%, 92. 8%, 误检率分别为: 11. 0%, 7. 3%, 4. 5%。



1. 传统方法, 2. 文献[7]方法, 3. 笔者方法
图 4 目标捕获准确度对比结果

5 结 语

在动态背景中进行兴趣目标捕获要保证有足够的检测精度,较少的误检是非常困难的。单纯的依靠高斯混合模型是很难满足需求的。在基于像素层面和相邻像素的梯度差上提出了一种基于颜色梯度的背景差算法,通过对运动区域进行特征分析,相邻像素分析,排除了大量噪点像素,在最多为 5 个目标同时出现的视频中仍取得了较高的检测精度和较低的误检,准确率达到 92. 8%,误检率为 2%,并且算法基本满足实时处理(23f/s)。

参考文献:

[1] Stauffer C, Grimson W E L. Learning patterns of activity using real-time tracking[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence. 2000, 22 (8): 747-757.

[2] Sundaraj K, Retnasamy V. Fast background subtraction for real time monitoring[C]//Proceedings of the Third Lasted International Conference. Phuket, Thailand, 2007: 382-387.

[3] Kim K, Chalidabhongse T H, Harwood D, et al. Background modeling and subtraction by codebook construction[C]//International Conference on Image Processing. Singapore: 2004(5): 3061-3064.

[4] Sigari M H, Mahmood Fathy. Real-time background modeling subtraction using two-layer codebook model [C]//Proceedings of International Multiconference of Engineers and Computer Scientists. Hongkong: 2008: 10-21.

[5] Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction[C]//European Conference of Computer Vision. Dublin: 2000: 751-767.

[6] 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.

[7] Valentine B, Apewokin S, Linda W. An efficient chro-

matic clustering-based background model for embedded vision platforms[J]. Computer Vision and Image Understanding, 2010(4): 1152-1163.

[8] Khan Z H, Irene Y, Andrew G. A robust particle filter-based method for tracking single visual object through complex scenes using dynamical object shape and appearance similarity [J]. Sign Process System, 2010: 9-26.

An Improved Target Capture Method of Video Interest Region

ZHOU Hao, ZHOU Ping, HU Yao-wu

(School of Informatics, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: A new background difference algorithm for interest region in video is proposed, which is based on color gradient by means of integrating shadow removal and connected domain analysis to capture the interesting target. Experimental results demonstrate that the method is capable of dealing light sudden changes as well as capturing target’s region of interest, in addition to its effectiveness for eliminating non-interest region’s error detection. The experiment of 6 videos with 5 targets at most, the authors obtain in the high precision and low by mistake examining, Correctly 92.8%, by mistake examining rate of 2%, and basically meet real-time processing.

Key words: target capture; background modeling; gaussian mixture model; color-gradient
(责任编辑: 陈和榜)