

一种基于码本模型的运动目标检测算法

郭春生 王 盼

(杭州电子科技大学通信工程学院, 杭州 310018)

摘 要 提出一种基于背景码本模型的视频图像中运动目标的检测算法。该算法利用归一化的 Mann-Whitney 秩和统计量自适应调整判决阈值,使用 Mean Shift 进行码本中码字和方差的更新,实现在检测过程中同时更新码本模型适应图像背景的变化。不同背景条件下的实验结果表明该算法利用 Mann-Whitney 秩和统计量的分布无关性,提高了运动目标检测精度。

关键词 运动目标检测 码本模型 秩和检验 均值移动

中图法分类号: TN911.37 **文献标志码:** A **文章编号:** 1006-8961(2010)07-1079-05

An Algorithm Based on Codebook Model to Moving Objects Detection

GUO Chunsheng, WANG Pan

(College of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018)

Abstract In this paper an algorithm of moving objects detection in video image was proposed, which is based on the background codebook model. In this algorithm, normalized Mann-Whitney rank-sum test was used to adjust the detection threshold of codebook model, and Mean Shift was also used to update the code and variance. So this algorithm can achieve the updating of codebook model and the detection of moving objects at the same time. The experimental results show that based on the distribution independence of Mann-Whitney rank-sum test this method can detect the moving objects efficiently under different conditions.

Keywords moving objects detection, codebook model, rank-sum test, Mean Shift

0 引 言

视频图像中的运动目标检测是计算机视觉信息提取的一个关键步骤,也是更高层次的视频图像分析如目标跟踪、目标分类和行为理解等的重要基础。在视频监控、目标跟踪和视频语义解释等系统中,运动目标检测的作用更是显得尤其重要。因此,从视频图像序列中快速准确地提取运动目标,是诸多研究人员所关心的一个问题。

运动目标检测的方法有很多种,其中帧差法、光流法、运动能量法和背景相减法最受关注。这几种方法各具利弊,如帧差法算法简单、计算量小,适用

于动态变化环境,但其检测精度不高,分割运动对象的完整性差;光流法具有不需要知道图像先验知识的优点,但其计算量过大,不适于实时检测;运动能量检测法适合于复杂变化的环境,能消除背景中振动的像素,使按某一方向运动的对象更加突出地显现出来,但运动能量检测法不能够准确地分割出运动目标;背景相减法是日前常用的一种运动目标检测方法,背景减法的原理是将当前帧与背景图像进行差分来得到运动目标区域,这种方法较其他方法能更好地检测和提取运动目标,但是需要构建背景模型,并不断地更新来适应当前背景的变化。如何准确构建背景图像模型是背景减除方法的研究重点和难点。

收稿日期:2008-11-19;改回日期:2009-04-02

第一作者简介:郭春生(1971—),男,副教授。2002年于南京航空航天大学获通信与信息系统专业博士学位。主要研究方向为统计信号处理。E-mail: guo.chsh@gmail.com

背景相减法根据背景图像模型特性又可分为参数化的和非参数化的、单模型的和多模型的、基于像素的和基于区域的。如单高斯模型^[1]属于基于像素的参数化单模型法,混合高斯模型^[2]与单高斯模型的差别仅在于模型数量的不同,K-均值聚类方法^[3]属于基于像素的非参数化单模型方法,自组织方法^[4]和码本模型算法^[5-7]则属于基于像素的非参数化多模型方法。

码本模型算法是 Kim 等人^[6]提出的一种非参数化的运动目标检测方法,该算法具有鲁棒性强,计算效率高的特点,可以通过迭代更新码本模型(包括码字,方差以及附加参数)来适应背景变化。本文针对码本模型与实际背景统计模型存在偏差而导致的虚检目标,以及局部运动目标颜色或灰度可能接近背景码本而引起的漏检错误,提出利用待检测像素的区域归一化 Mann-Whitney 秩和统计量自适应调整运动目标检测阈值,若背景区域的秩和统计量较小,增加判决阈值可以减少虚假目标数量;若运动目标区域的秩和统计量较大,降低检测阈值能够提高检出率。

1 背景码本模型的建模方法

图像序列背景像素的码本模型由码字,方差以及附加参数构成,即

$$C_{k,t} = [X_{k,t}, \sigma_{k,t}^2, f_{k,t}, \lambda_{k,t}, p_{k,t}, q_{k,t}] \quad k=1, \dots, L$$

其中, L 为码本容量,表示码本中的码字数量。设 N 为附加参数统计周期,当 $t = nN$ 时,则码本中附加参数分别记为 $[f_k, \lambda_k, p_k, q_k]$,其中, f_k 代表在一个统计周期内码字 $X_{k,t}$ 的出现次数(即作为当前码字的次数), p_k 为首次出现时间, q_k 为最后出现时间, λ_k 为最大未出现时间间隔。附加参数的更新如文献^[6]所述。下面主要讨论如何利用 Mean Shift 实现码字和方差的更新,适应图像序列的背景变化,以及如何通过计算 Mann-Whitney 秩和统计量,灵活调整检测阈值改善检测性能。

设图像序列第 t 帧中某像素值为 X_t ,码本中第 k 个码字 $X_{k,t}$ 为当前码字,相应方差为 $\sigma_{k,t}^2$,则基于 Mean Shift 算法当前码字和方差的更新公式为

$$\begin{aligned} X_{k,t+1} &= \alpha_k (X_t - X_{k,t}) + X_{k,t} \\ \sigma_{k,t+1}^2 &= \beta_k \delta_{k,t}^2 + \sigma_{k,t}^2 \end{aligned} \quad (1)$$

式中, α_k 为码字更新因子, β_k 为方差更新因子。在具体应用时,由于图像序列中的图像帧是依次串行输

入,根据平稳随机过程各态历经性,随着串行处理像素点数量的增加,当前码字和方差将各自收敛到其概率密度函数的局部极大值,因此,这里 $\alpha_k (X_t - X_{k,t})$ 和 $\beta_k \delta_{k,t}^2$ 的作用仍然等效于 Mean Shift。在不同色彩空间中 $\delta_{k,t}^2$ 的具体形式各异^[6-7],在 RGB 色彩空间中,令 $X_t = [R_t \quad G_t \quad B_t]$, $X_{k,t} = [R_{k,t} \quad G_{k,t} \quad B_{k,t}]$,则 $\delta_{k,t}^2 = R_t^2 + G_t^2 + B_t^2 - \frac{R_{k,t}R_t + G_{k,t}G_t + B_{k,t}B_t}{R_{k,t}^2 + G_{k,t}^2 + B_{k,t}^2}$;在 YCbCr 色彩空间中,令 $X_t = [Y_t \quad Cb_t \quad Cr_t]$, $X_{k,t} = [Y_{k,t} \quad Cb_{k,t} \quad Cr_{k,t}]$,则 $\delta_{k,t}^2 = [\max(|Cb_t - Cb_{k,t}|, |Cr_t - Cr_{k,t}|)]^2$;在灰度图像中,令 $X_t = I_t$, $X_{k,t} = I_{k,t}$,则 $\delta_{k,t}^2 = (I_t - I_{k,t})^2$ 。不失一般性,下面将在灰度图像上分析相关原理和具体实现方法。

在运动目标检测过程中阈值的选取至关重要,过大的阈值将导致运动目标检测的不完整甚至检测失败,过小的阈值则会增加虚检像素。本文通过待检测像素的归一化 Mann-Whitney 秩和统计量 $SMW_{n/2,n}$ 自适应调整检测阈值,提高运动目标检测率。

归一化的 Mann-Whitney 秩和统计量 $SMW_{n/2,n}$ 定义为

$$SMW_{n/2,n} = \frac{MW_{n/2,n} - E[MW_{n/2,n}]}{\sqrt{Var[MW_{n/2,n}]}} \quad (2)$$

式中, Mann-Whitney 秩和统计量

$$MW_{n/2,n} = \sum_{i=1}^{n/2} \sum_{j=n/2+1}^n M(I_j < I_i)$$

$$函数 M(I_j < I_i) = \begin{cases} 1 & I_j < I_i \\ 0 & I_j \geq I_i \end{cases}$$

Mann-Whitney 秩和统计量的均值

$$E[MW_{n/2,n}] = \frac{n^2}{8}$$

Mann-Whitney 秩和统计量的方差

$$Var[MW_{n/2,n}] = \frac{n^2(n+1)}{48}$$

当 $|SMW_{n/2,n}| < T_n$ 时,表示像素值的统计特性并没有发生变化,当 $|SMW_{n/2,n}| \geq T_n$ 时,表示像素值的统计特性发生变化。在具体应用中由于内存的限制不可能存储多幅图像,利用像素值统计特性的局部遍历性,由当前图像和上一幅图像像素点邻域中的所有像素点共同构成一个像素值序列,在该像素序列上进行像素值统计特性是否变化的判断。例如给定 $n=18$,令当前图像当前像素 3×3 邻域中的 9 个像素值,分别记为 I_i ,其中 $i=10, \dots, 18$,以及前一幅

图像当前像素 3×3 邻域中的 9 个像素值,分别记为 I_i ,其中 $i = 1, \dots, 9$,共同构成一个 18 点序列。秩和统计量的计算量较大,在具体应用中可以采取分块平滑策略以及利用前一帧图像的检测结果选择部分需要计算的数据块,降低算法处理的计算量。

假设当前像素点的码本模型有 L 个码字 $X_{i,t}$, $i = 1, \dots, L$ 和对应方差 $\sigma_{i,t}^2$, $i = 1, \dots, L$,当前像素值为 X_t ,则计算当前像素值与码本中所有码字的最小 Mahalanobis 距离

$$L_M = \min_i \frac{|\delta_{i,t}|}{\sigma_{i,t}} \quad (3)$$

码本模型的判决规则如下:若码本中的码字 $X_{i,t}$ 与当前像素值 X_t 的距离最小,且满足 $L_M < T_h$,其中 T_h 为阈值,则判决当前像素为背景像素;若 $L_M \geq T_h$,则需要在码本模型中添加新码字 $X_{L+1,t} = X_t$,以及相应的方差 $\sigma_{L+1,t}^2 = \sigma_0^2$,其中 σ_0^2 为确定常数。利用归一化 Mann-Whitney 秩和统计量调整固定阈值 T_h ,得

$$T'_h = T_h - \lambda U(SMW_{n/2,n}) \quad (4)$$

式中,函数 $U(t) = \begin{cases} T_n & |t| \geq T_n \\ |t| & |t| < T_n \end{cases}$, λ 为确定常数。

2 参数选择分析

背景码本模型中涉及两个更新因子 (α_k 和 β_k) 和两个阈值 (T_n 和 T_h),下面具体分析这些参数的确定。

在更新因子 α_k 和 β_k 的取值可以使用码字 C_k 作为当前码字的次数来确定,通常认为作为当前码字的次数越多,则码字对应像素值的波动越小,与真实背景也越接近,所需的更新也越小;反之,码字对应像素值的波动越大,需要较大的更新因子使码字更快地向真实背景靠近。已知码本中的附加参数 f_k 表示在 $t = nN$ 时刻统计的过去所有 N 次中码字 $X_{k,t}$ 作为当前码字的次数,由此可得更新因子:

$$\alpha_k = \frac{N - f_k}{N}, \beta_k = \frac{N - f_k}{N} \quad (5)$$

式中, N 为附加参数统计周期,即每隔 N 帧重新计算一次更新因子。

当样本数量足够多时,归一化的 Mann-Whitney 秩和统计量 $SMW_{n/2,n}$ 满足均值为 0,方差为 1 的高斯分布,因此可以近似认为

$$P(|SMW_{n/2,n}| < T_n) = \text{erf}(T_n / \sqrt{2})$$

其中, $\text{erf}(T_n) = \frac{2}{\sqrt{\pi}} \int_0^{T_n} e^{-y^2} dy$ 为误差函数。于是,取

$P(|SMW_{n/2,n}| < T_n) = 0.9$ 时,则得到 $T_n = 1.65$ 。

阈值 T_h 的确定是一个假设检验问题,即存在原假设 $H_0: |I_t - X_t| \leq T_h$,备择假设 $H_1: |I_t - X_t| > T_h$,判决区域分别为 D_0 和 D_1 。当 H_0 为真时, I_t 为背景像素,否则为前景像素。假设背景像素服从高斯分布,前景像素服从均匀分布,由高斯分布的对称性,可以得到整个灰度空间上假阳性率的计算公式:

$$P(D_1 | H_0) = \frac{2}{\sqrt{2\pi}} \int_{T_h}^{\infty} e^{-t^2/2} dt \quad (6)$$

取假阳性率为 0.1 时,则得到 $T_h = 1.65$ 。

3 算法的具体实现

运动目标检测算法可分为两个阶段,即训练阶段和检测阶段。训练阶段从输入图像序列中提取所有像素点的背景码本模型,检测阶段根据背景码本模型和阈值对输入图像的每个像素进行判别,检测出运动目标(即前景),同时不断更新背景码本模型适应背景变化。

将码本中附加参数矢量记为 $[f_{k,t}, \lambda_{k,t}, p_{k,t}, q_{k,t}]$,输入为 $1, \dots, S$ 帧图像序列,其中前 P 帧图像为训练图像,输出为二值化前景/背景图像序列 B_t ,则该算法的具体步骤如下:

1) 用第 1 帧图像初始化各像素码本模型中的码字和方差,令码字为 $X_{1,1} = I_1$,方差 $\sigma_{1,1}^2$ 皆为 σ_0^2 。

2) 当 $t = 2, \dots, S$ 时,分别执行第 3) 和第 4) 步。

3) 计算当前像素归一化的 Mann-Whitney 秩和统计量 $SMW_{n/2,n}$,并利用式(4)修正判决阈值 T'_h 。

4) 计算当前像素值与码本中所有码字的最小 Mahalanobis 距离 L_M ,设当前像素值与码本中的第 i 个码字具有最小 Mahalanobis 距离。

5) 当 $t > P$ 时,若 $L_M \geq T'_h$,则置当前码字为空,并在 B_t 中标记当前像素为前景像素;若 $L_M < T'_h$,在 B_t 中标记当前像素为背景像素,将码本中的第 i 个码字作为当前码字 $X_{k,t}$,并用式(1)更新当前码字和方差,对应附加参数更新为 $[f_m + 1 \quad \max(\lambda_m, t - q_m) \quad p_m \quad t]$ 。

当 $t \leq P$ 时,若 $L_M \geq T'_h$,则将当前像素灰度值 I_t 作为新的当前码字加入到码本模型,并初始化方差为 σ_0^2 、附加参数为 $[1 \quad t - 1 \quad t \quad t]$;若 $L_M < T'_h$,将码本中的第 i 个码字作为当前码字 $X_{k,t}$,并用式(1)

更新当前码字和方差, 对应附加参数更新为 $[f_m + 1 \quad \max(\lambda_m, t - q_m) \quad p_m \quad t]$ 。

6) 当 $t = nN$ 时, 调整码本中所有码字参数 $\lambda_i \leftarrow \max \{ \lambda_i, (N - q_i + p_i - 1) \}$, 并按照式(4)更新参数 α_k 和 β_k , 若 $t \leq P$, 且 $\lambda_i \geq N/2$, 将该码字从码本中删除。

7) 输出 B_i 。

4 实验结果与分析

为了验证本文算法的有效性, 下面使用 MSA 图像序列和 PETS2002 图像序列, 将其与单高斯模型算法、混合高斯模型算法、自组织算法和 Kim 码本模型算法的处理结果进行比较。

本文算法中码本中的码字数 n 从理论上讲范围为 $[1, +\infty)$, 在实际应用中根据序列背景的复杂程度具体选择, 由算法不难看出, n 的大小对处理速度和检测精度有影响, n 越大, 则需搜索的码字越多, 速度越慢, 但检测精度越高。通过实验可以得出当选取 $n > 3$ 时检测精度随 n 的增大提升的并不多, 因此选取码字数 $n = 3$ 。在训练阶段, 可以根据图像背景的复杂程度, 选取固定阈值 $T_h > 1.65$, 以便能够包含更多的背景信息; 在检测阶段, 选取固定阈值 $T_h < 1.65$, 包含更多的前景, 提高检出目标的完整性。由于使用秩和统计量计算确定数据时会出现错误, 在每次计算秩和统计量之前需要根据待检区域的均值及方差判断数据的随机性, 将近似确定数据无偏差的直接置为零。

4.1 MSA 图像序列的实验结果

该序列是室内环境图像大小为 240×320 像素的图像序列, 在 MSA 图像序列的实验中, 主要用于检验算法的检测遗弃物品的能力、阴影抑制能力以及背景变化很小情况下的运动目标检测精度。

由图 1 所示的实验结果可以看出, 所有的算法都能检测出遗弃物品(黑色的包)。在阴影不明显的环境下, 自组织方法和本文的算法均有阴影抑制效果, 但自组织方法的人像明显出现残缺。

4.2 PETS2002 图像序列实验结果

本图像序列图像大小为 576×768 像素, 主要用来检验算法的抗噪能力, 噪声在此图像序列中的具体表现为背景中的晃动树叶, 以及运动目标与背景存在较高的相似度。

PETS2002 图像序列的实验结果如图 2 所示,

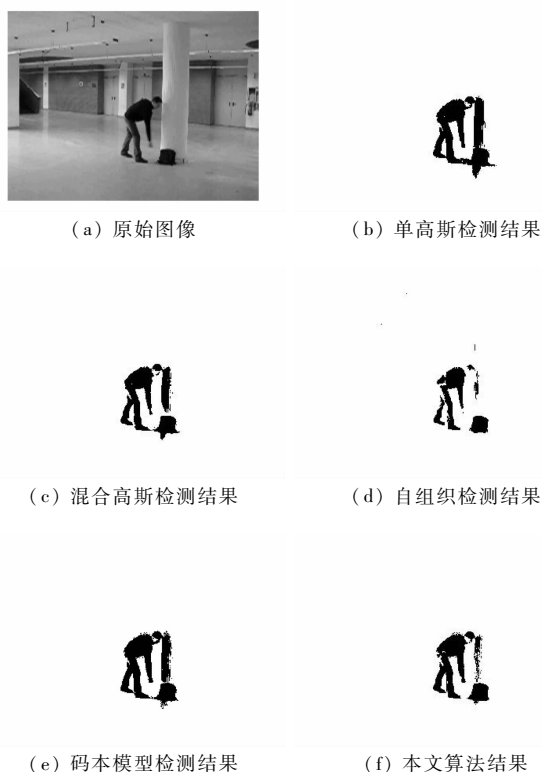


图 1 MSA 图像序列检测结果

Fig. 1 Detection results of MSA image sequence

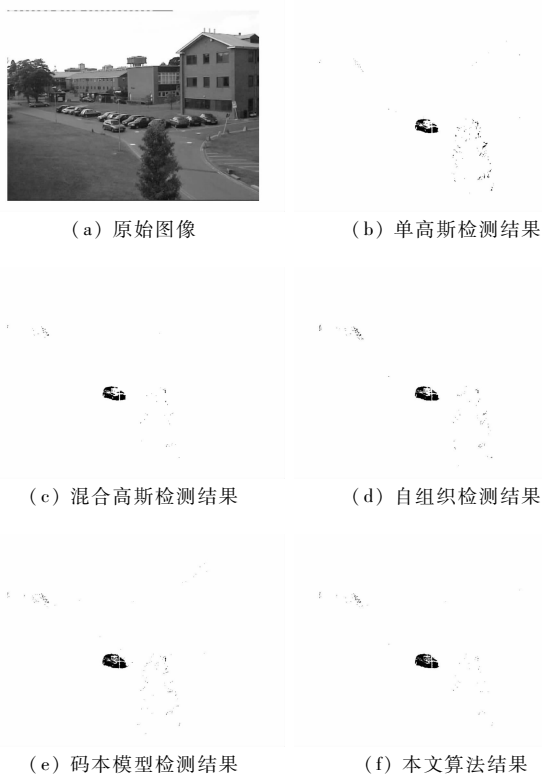


图 2 PETS2002 图像序列检测结果

Fig. 2 Detection results of PETS2002 image sequence

由于汽车挡风玻璃处与背景具有较高的相似度,其他算法在这一区域都表现不佳,虽然可以通过减小检测阈值的方法将该区域检出,但这样也同时导致了误检率的增加,本文算法由于根据秩和检验结果自适应地调整检测阈值,因而能提高该区域的检出率。

为了定量比较各种算法的性能差异,使用 F 测度衡量算法的检测性能。 F 测度是对检出率和检测精度的综合评价,计算公式为

$$F = \frac{2tp}{2tp + fn + fp}$$

其中, tp 是检测目标中的正确像素数, fn 是检测目标中的错误像素数, fp 是未检测出的动目标像素数。表 1 给出了各种算法对图 1 和图 2 所示图像帧处理结果的检测性能参数和处理速度,从 F 测度可以看出,本文算法的检测性能明显优于其他算法,由于性能改善是通过在码本模型中添加秩和统计量的计算,处理时间较码本模型有一定的增加。

表 1 各种算法检测性能参数和处理速度

Tab.1 Performance parameters and processing speeds of referred methods

图像序列	技术指标	单高斯	混合高斯	自组织检测	码本模型	本文算法
MSA	F 测度均值	0.79	0.83	0.82	0.84	0.88
	处理速度/(帧/s)	33.33	14.29	8.3	20	12.5
PETS2002	F 测度均值	0.66	0.71	0.64	0.74	0.78
	处理速度/(帧/s)	14.29	2.63	1.35	3.70	2.3

5 结 论

提出了一种基于背景码本模型的视频图像中运动目标的检测算法。该算法利用归一化的 Mann-Whitney 秩和统计量自适应调整判决阈值,有效改善算法的检测性能,使用 Mean Shift 进行码本中码字和方差的更新,提高了背景变化时算法的自适应能力,实现在检测过程中同时更新码本模型适应图像背景的变化。实验结果验证了该算法在抗噪声能力方面的优势,以及较好的运动目标检测效果。

参考文献 (References)

[1] Christopher R W, Ali A, Trevor D, et al. Pfnder: real-time tracking of the human body [J]. IEEE Transactions on PAMI, 1997, 19(7): 780-785.

[2] Nir F, Stuart R. Image segmentation in video sequences: a probabilistic approach [C]//Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence. San Fransisco, USA:

Morgan Kaufmann, 1997: 175-181.

[3] Zha Chengdong, Wang Changsong, Gong Xianfeng, et al. Moving target detection based on adaptive background model [J]. Opto-Electronic Engineering, 2008, 35(1): 26-30. [查成东, 王长松, 巩宪锋, 等. 基于自适应背景模型的运动目标检测 [J]. 光电工程, 2008, 35(1): 26-30.]

[4] Lucia M, Alfredo P. A self-organizing approach to background subtraction for visual surveillance applications [J]. IEEE Transactions on IP, 2008, 17(7): 1168-1177.

[5] Zhang Jun, Dai Kexue, Li Guohui. HSV color-space and codebook model based moving objects detection [J]. Systems Engineering and Electronics, 2008, 30(3): 423-427. [张军, 代科学, 李国辉. 基于 HSV 颜色空间和码本模型的运动目标检测 [J]. 系统工程与电子技术, 2008, 30(3): 423-427.]

[6] Kyungnam K, Thanarat H C, David H, et al. Real-time foreground-background segmentation using codebook model [J]. Real-Time Imaging, 2005, 11(3): 172-185.

[7] Louis S L, Donald P, Xavier M. Real-time object detection and background maintenance for uncontrolled environments [C]//Proceedings of the 5th International Conference on Signal Processing, Pattern Recognition, and Applications. Stevens Point, Wisconsin, USA: WSEAS, 2008: 599-604.