

基于 SACON 背景模型的人体检测与跟踪

高山¹, 毕笃彦¹, 魏娜²

(1. 空军工程大学 工程学院, 西安 710038; 2. 空军工程大学 教育技术中心, 西安 710038)

(gaoshan1114@126.com)

摘要:提出了一种基于抽样一致性(SACON)的背景模型。采用七种典型视频序列进行了实验, SACON 背景模型较目前常用五种背景模型具有较高的准确性, 适用于复杂场景下的目标检测与跟踪。应用于人体检测与跟踪时, 将目标作为非刚性物体进行处理, 结合颜色和空域信息建立一种新的目标外观模型。实验证明, 该方法能较准确的描述人体特征, 即使在有遮挡、颜色相近及小目标等情况下也均能准确的对人体目标进行检测和跟踪。

关键词:运动目标检测与跟踪; 随机抽样一致性; 背景模型

中图分类号: TN929.12 **文献标志码:** A

Pedestrian detection and tracking based on SACON background model

GAO Shan¹, BI Du-yan¹, WEI Na²

(1. Engineering College, Air Force Engineering University, Xi'an Shaanxi 710038, China;

2. Teaching Technology Center, Air Force Engineering University, Xi'an Shaanxi 710038, China)

Abstract: A new background modeling method based on Sample Consensus (SACON) was proposed for detecting foreground objects. The method, which has higher accuracy than other five popular methods in the experiment when seven typical video sequences were adopted, is suit for detecting and tracking the moving target under complex circumstances. On pedestrian detection and tracking, considering the target to be non-rigid, a new target appearance model containing spatial domain and color information was built based on SACON. The experimental results show that the method can describe pedestrian feature exactly and can detect or track targets with barriers and similar color and small targets with high accuracy.

Key words: moving target detection and tracking; sample consensus; background model

0 引言

人体检测就是根据人体的某些特征在静态图像中检测人体目标。目前大多数人体跟踪算法^[1-2]都需要已知目标的初始位置(如通过人机交互方式), 而实用的跟踪系统需要自动定位初始目标, 这就需要人体检测算法的支持。由于人体目标是典型的非刚性目标, 可能呈现出各种不同形状, 单一的模型几乎不可能捕获各种形状的人体, 其次, 人的不同衣着, 也使得基于区域特征的算法不能有效地工作。因此, 检测人体目标比检测其他目标更具挑战性。

目前工作主要集中在如何对人体轮廓建模^[3-6], 文献[7-8]对这些工作做了总结。文献[9-10]利用高斯混合模型对人体的颜色分布进行估计, 尽管这种方法能获得较好的效果, 但是选择合适的高斯模型数目很困难; 文献[11]采用人体的颜色直方图进行跟踪, 但两个人即使穿着不同的衣服也可能有相近的颜色分布, 而且采用颜色直方图需要大量的抽样才能保证其准确性; 文献[6]采用概率模板对人体每个像素建模, 但该方法在计算中对于内存的要求很高; 文献[3-4]采用非参数的高斯核密度的方法, 改进了文献[5]的方法, 但这种非参数的方法计算量太大, 尽管文献[12]提出了一种快速核密度估计的方法, 但还是不能满足实时的要求。

本文将基于抽样一致性(Sample Consensus, SACON)背景模型进行推广, 结合颜色和空域的信息, 建立了一种简单有

效的轮廓模型, 能够准确检测跟踪有遮挡情况下的人体运动。

1 SACON 的概念

SACON 的概念与随机抽样一致性(Random Sample Consensus, RANSAC)的概念^[13]类似。每个像素点过去的 N 个抽样值记作 $\{x_i(m) \mid i = 1, \dots, N; N < t\}$, 其中 $x_i(m) = \{x_i^{c_1}(m), \dots, x_i^{c_k}(m)\}$ 表示像素 m 在 t 时刻的观测值, 每个观测值有 k 个颜色通道, k 值由图像的颜色模型决定, 如 RGB 颜色模型有三个通道 (R, G, B) , 则 $k = 3$ 。定义一个二值函数来表示当前抽样与过去抽样值的一致性:

$$\Gamma_i^c(m, t) = \begin{cases} 1, & |x_i^c(m) - x_i^c(m)| \leq T_r \\ 0, & \text{其他} \end{cases} \quad (1)$$

其中 T_r 是与偏差 $|x_i^c(m) - x_i^c(m)|$ 相关的阈值。SACON 实际表示了当前像素的抽样值与过去 N 次抽样相一致的次数:

$$B_i(m) = \begin{cases} 1, & \sum_{i=1}^N \Gamma_i^c(m, t) \geq T_n, \forall c \in \{C_1, \dots, C_k\} \\ 0, & \text{其他} \end{cases} \quad (2)$$

其中 T_n 是一个阈值, 定义了偏差 $|x_i^c(m) - x_i^c(m)|$ 小于 T_r 的个数, B_i 为 1 表示像素点为背景像素, 0 表示前景像素。

实际上, T_n 的值与 N 的值有关, N 取值越大, 则 T_n 越大, 同样, T_n 的取值也反映阈值 T_r 的大小, T_r 取值越大, T_n 则越

收稿日期: 2008-12-08; 修回日期: 2009-02-25。 基金项目: 国家 863 计划项目(2007AA701206)。

作者简介: 高山(1983-), 女, 山东济宁人, 博士研究生, 主要研究方向: 数字图像处理; 毕笃彦(1964-), 男, 陕西扶风人, 教授, 博士生导师, 主要研究方向: 模式识别、图像处理; 魏娜(1981-), 女, 陕西西安人, 讲师, 主要研究方向: 网络安全。

大,一般取 T_n 为 $\tau T_r N$, τ 是一个经验常数。

2 基于 SACON 的背景模型

2.1 阴影移除

RGB 颜色空间对亮度的变化很敏感,常用归一化的颜色空间来减小亮度的影响^[6]

$$\begin{cases} r = R/(R+G+B) \\ g = G/(R+G+B) \\ b = B/(R+G+B) \end{cases} \quad (3)$$

对于 8 比特彩色图像,往往将 r, g, b 通道的值限定在 $[0, 255]$ 。采用 rgb 颜色模型损失了图像的亮度,所以文献[14]中采用 rgI 颜色模型进行建模。 (r_b, g_b, I_b) 为背景像素 x_b 当前的抽样值, (r_t, g_t, I_t) 为 t 时刻某个像素的抽样值,假设背景为静态,则阴影像素较背景像素暗一些, $\beta \leq I_t \leq 1, \beta \leq I_b \leq 1$ 相反,若像素点为高亮点则 $1 \leq I_t \leq \gamma, 1 \leq I_b \leq \gamma$ 。当像素点为阴影时,对于 rgI 颜色空间 I 的通道,式(1)可改写为:

$$\Gamma_i^c(m, t) = \begin{cases} 1, & |x_i^c(m) - x_b^c(m)| \leq T_r, \forall c \in \{r, g\}, \\ & \beta \leq I_t \leq 1, \beta \leq I_b \leq 1, c \in \{I\} \\ 0, & \text{其他} \end{cases} \quad (4)$$

其中 β 和 γ 为经验常数。在本文实验中 $\beta = 0.6, \gamma = 1.5$ 。

2.2 rgb 空间噪声抑制

从 RGB 空间变换到 rgb 空间是非线性变换,如果亮度值 I 很小,则归一化颜色空间 (r, g, b) 中的噪声变得很大。为解决这个问题,当亮度值 I 小于某个阈值 I_{td} 时(根据经验 $I_{td} = 7$), r 通道和 g 通道的值不可靠,仅根据亮度 I 值计算 $\Gamma_i^c(m, t)$, 所以 x 的颜色模型可表示为:

$$x = \begin{cases} (r, g, b), & I \geq I_{td} \\ (I), & I < I_{td} \end{cases} \quad (5)$$

2.3 空洞像素验证

当前景与背景的色度很相似时,前景目标区域内部往往会出现“空洞”,即部分前景像素被错分为背景像素。尽管可以由式(4)的条件 $\beta \leq x'_t \leq \gamma, \beta \leq x'_b \leq \gamma$ 抑制阴影,但从某种程度上损失了图像中的亮度信息。当背景和前景的亮度差别较大但色度接近时,也可能满足条件 $\beta \leq x'_t \leq \gamma, \beta \leq x'_b \leq \gamma$,尤其在 x'_b 大于 x'_t 时,前景目标像素就容易被错分为背景像素,从而出现“空洞”现象。直接采取填充空洞的方法,如形态学中的闭运算,容易导致前景目标的一些细节丢失,因为前景目标区域中本身可能有一些空洞结构,为此,我们采用一个验证过程验证“空洞”中的像素的“有效性”。验证的过程是:对“空洞”中的每个像素,若其亮度 x'_t 满足 $|x'_t - x'_b| \leq T_l$, 则判定为像素前景目标像素,反之则为背景像素。 T_l 为阈值,经验值取为 7。尽管验证过程在前景和背景很相近的时候也不能正确分类前景和背景像素,但能在很大程度上提高算法的准确性。

2.4 T_r 的取值讨论

有两种方法设置 T_r : 一是,根据经验将 T_r 设置为一个全局阈值,但要获得适当的全局阈值是很困难的;二是,对每个

像素计算其标准差 σ_i , 取 T_r 为 $\eta \sigma_i$, η 一般取 2.5 或 3, 不足之处是当模型为多模分布时,会出现对 σ_i 的过估计现象。

本文综合两种方法的优点,采用如下方式获取每个像素的阈值 T_{ri} :

$$T_{ri} = \min(T_1, \eta \sigma_i) \quad (6)$$

2.5 背景更新

背景更新是背景建模中一个非常中要的步骤,因为当场景发生变化后,需要对背景重新抽样来更新模型参数,以适应场景中的变化。本文提出了一种兼顾单个像素和整块区域的选择性背景更新方法。

2.5.1 像素级背景更新

为处理背景中有目标移入/移出的情况,类似文献[7]的做法,本文也采用了一个“超时映射”(Time Out Map, TOM), 对于第 t 帧中的像素 m , 有对应的计数量 $TOM_t(m)$, 如果某个像素被判定为前景像素, 则其 TOM 值加 1。

$$\begin{cases} TOM_t(m) = TOM_{t-1}(m) + 1, & B_t(m) = 1 \\ TOM_t(m) = 0, & \text{其他} \end{cases} \quad (7)$$

TOM 值记录了每个像素连续被分类为前景像素的次数,一旦该像素被分类为背景像素, 则其 TOM 值重置为 0。如果 TOM 的累加值已经大于阈值 T_{TM} (在本文的实验中取 45), 则将该像素分类为背景像素, 前提是假定前景目标不能在场景中长时间静止。但是, 有些情况下, 仅仅根据像素被判定为前景像素的次数判断目标是否在场景中是否静止有时也无法反映真实的情况, 因而, 需要结合其他的判别准则。

2.5.2 区域级背景更新

根据目标中心的位置和区域中像素的数量判断某个区域是否静止; 若两个量有一个发生了显著变化, 就判定目标有运动, 否则判定目标静止。对于一个区域 Ω , 内部像素的 TOM 值根据式(8)更新:

$$\begin{cases} TOM_t(m') = TOM_{t-1}(m') + 1, & \Omega \text{ 静止时} \\ TOM_t(m') = 0, & \text{其他} \end{cases} \quad (8)$$

与单个像素的 TOM 值相似, 如果区域静止, 则区域的 TOM 值加 1, 如果判定区域有运动, 则将区域中所有像素的 TOM 值置 0。也即当区域的 TOM 值高于阈值 T_{TM} , 就将该区域判定为背景区域采用区域级更新的准则可消除像素级更新中的“空洞”。

2.6 SACON 背景模型框架

SACON 背景模型包括三个主要的步骤: 提取候选前景像素、SACON 验证和“空洞”像素验证, 如图 1 所示。第一步根据相邻两帧的差分获得可能的候选前景目标区域^[15], 这样在后续的计算中仅需要对候选前景目标区域进行计算和验证, 从而提高了计算的效率。如果背景中包含局部运动, 则候选前景区域也会包含这些动态的背景区域, 这个问题在算法的第二步解决; 第二步将所有候选区域根据式(4)进行 SACON 验证, TOM 值超过预定阈值的像素点也要进行 SACON 验证, SACON 验证的输出就是检测到的前景目标; 第三步对检测出来的前景区域的“空洞”进行验证, 消除前景目标内部的“空洞”; 最后根据 3.5 节的讨论对背景进行更新, 准备处理下一帧。

2.7 算法仿真实验

我们采用文献[15]中的7个实验视频 Moved object (MO), Time of day (TOD), Light switch (LS), Waving trees (WT), Camouflage (C), Bootstrapping (B), Foreground aperture (FA), 这些视频具有很强的代表性, 包含了各种复杂场景, 如背景变化, 光照的变化, 背景中包含摇摆的树枝, 前景目标与背景的色调相似等。采用 SACON 背景模型和当前讨论得较多的五种背景模型对视频进行处理, 对实验结果进行对比分析。为研究算法的性能, 引入 total error (te)——背景像素被错分为前景像素的数目及前景像素被错分为背景的数目之和 (即总误差)。

结果如图2所示, 对于 LS 序列, Wallflower 算法和基于特征值的方法的总误差 (te) 最小; 对于 FA 序列, Wallflower 算法的 te 最小; 对于 MO 序列, 除了基于特征值的方法外, 都能获得较为精确的检测结果; 本文提出的 SACON 背景模型在 MO、TOD、WT、C 和 B 序列的检测中都获得了最小的 te, 在 LS、FA 序列的检测中也获得了较小的 te, 性能优越。

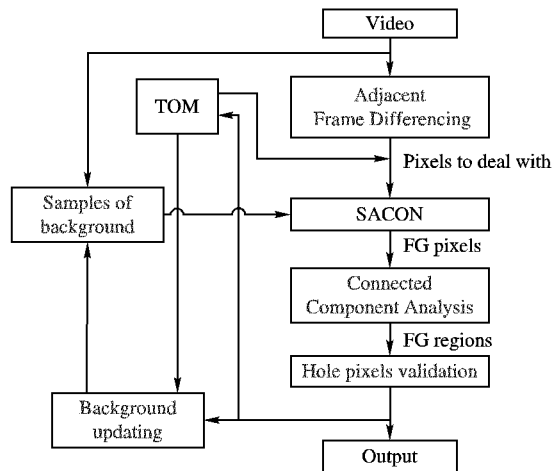


图1 SACON 背景模型框图

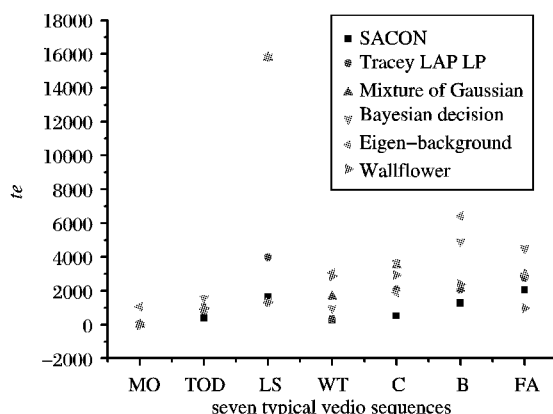


图2 6种背景模型对7种典型视频序列的处理结果对比

3 在人体检测与跟踪中的应用

本文在 SACON 背景模型的基础上结合颜色和空域的信息, 建立一种简单有效的轮廓模型, 能够准确跟踪有遮挡情况下的人体运动。算法采用了两个跟踪模块: 基于矩形区域的跟踪以及基于轮廓的跟踪。当目标 (人体) 没有发生互相遮挡的情况时, 采用类似于文献[16]的基于矩形区域的跟踪方法, 有遮挡发生的情况下, 采用轮廓模型进行跟踪。

3.1 基于轮廓的跟踪

为描述人体的轮廓特征, 首先假定人的姿态是直立的, 即使在发生重叠的时候, 人体轮廓不会发生特别显著的变化。

3.1.1 前景目标的轮廓模型

本文将人体轮廓的形状定义为包含在具有一定大小矩形框内的斑点, 在多个人站在一起的时候, 用矩形结构表示人体轮廓是很困难的。假设 k 个人 $P_j (j = 1, \dots, k)$ 在一起, 还有互相遮挡的情况, 我们需要对光斑中的每个像素 p_i 进行分类, 确定其属于哪个人 P_j 。用 A_j 表示 P_j 的光斑中像素点数目, $\{S_j(m)\}_{m=1, \dots, A_j}$ 表示第 j 个人在被遮挡前的轮廓抽样, 像素 p_i 的分类准则如下:

$$p_i \in P_k; k = \operatorname{argmax}_k \prod_{c=r, g, I} \frac{\sum_m \Gamma_k^c(p_i, m)}{A_k} \quad (9)$$

其中:

$$\Gamma_k^c(p_i, m) = \begin{cases} 1, & |p_i^c - S_k^c(m)| \leq T_r' \\ 0, & \text{其他} \end{cases} \quad (10)$$

其中: T_r' 为阈值, $S_k^c(m)$ 是 $c (c = r, g, I)$ 颜色通道的抽样值, 即先计数像素在前几帧中属于第 k 个人轮廓的次数, 然后将该值归一化, 这样轮廓形状的大小就不会影响到最后的结果, 根据这个归一化的值判断当前该像素应该属于哪个人的轮廓。

3.1.2 前景目标的空域信息

空域信息, 即目标 (人) 的位置, 是很重要的一个信息, 但在式 (9) 中没有体现出来, 下面讨论如何在检测中引入这种空域信息。

本文分垂直方向和水平方向对位置进行讨论。对于纵坐标而言 (垂直方向), 引入空域 (位置) 信息后, 式 (9) 可重写为:

$$p_i \in P_k; k = \operatorname{argmax}_k \prod_{c=r, g, I} \frac{\sum_m \Gamma_k^c(p_i, m')}{A'_k} \quad (11)$$

其中: m' 表示 P_k 的一组抽样, 其中 m' 的 y 坐标与 p_i 邻近 (本文中如果满足 m' 的 y 坐标与 p_i 的 y 坐标差值的绝对值小于 7 即认为两者邻近), A'_k 表示第 k 个人的抽样数量。

在水平方向的检测中, 若 P_k 被遮挡住, P_k 的中心可能发生变化, 仍采用 $t-1$ 时刻的中心点可能出现错误结果, 此时需更新 P_k 的参数, 更新的方法是将所有抽样的 x 坐标值的均值作为 P_k 的新的 c_x 。

为充分利用横坐标上的位置信息, 定义如下两个步骤:

1) 通过式 (11) 确定像素点的初始分类; 2) 计算类 P_k 的中心 $c_x(P_k)$, 然后通过式 (12) 再次对每个像素分类:

$$p_i \in P_k; k = \operatorname{argmax}_k \frac{\prod_{c=r, g, I} \sum_m \Gamma_k^c(p_i, m') / A'_k}{|x(p_i) - c_x(P_k)|} \quad (12)$$

其中 $x(p_i)$ 是 p_i 的 x 坐标, $x(p_i) - c_x(P_k)$ 的值可能为 0, 出现这种情况时, 采用一个预先定义好的很小的值来作分母。

3.2 有遮挡情况下算法验证

首先 SACON 背景模型检测出前景目标的初始位置, 然后将这些区域作为跟踪的区域, 初始化目标的轮廓 $\{P_j\}_{j=1, \dots, k}$, 然后在每帧图像都更新轮廓参数。

图3显示了本文提出的算法在室外场景的检测分割结

果,尽管有些像素出现了分类错误,但大多数像素还是保证了正确的分类,即使出现遮挡,也能准确的跟踪目标。

图4中,人的衣服颜色与背景很相近,人的轮廓形状也很接近,人在图像中所占比例也很小,但本文提出的检测跟踪算法仍然能正确检测和跟踪目标。

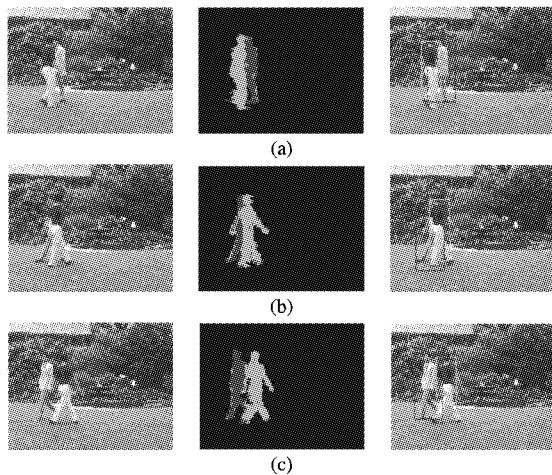


图3 有遮挡情况下算法验证

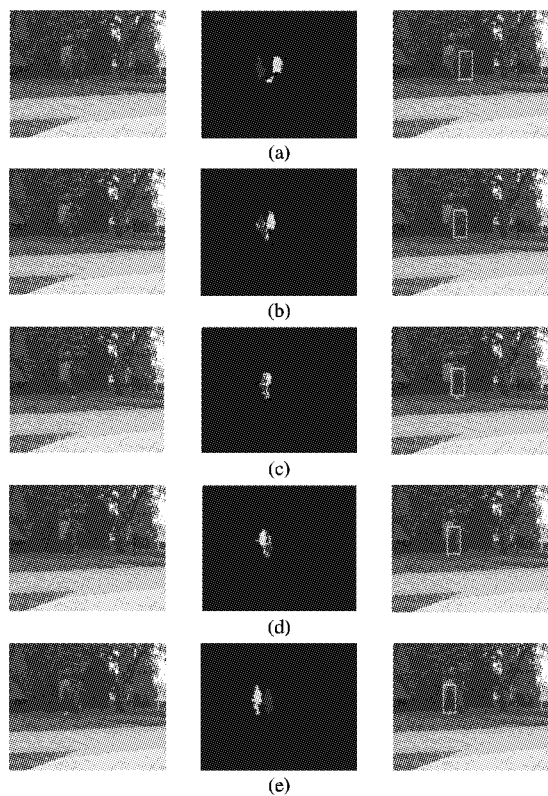


图4 颜色相近和小目标时算法验证

4 结语

本文提出了一种基于SACON的背景模型,适用于复杂场景下的目标检测与跟踪。将SACON背景模型应用到人体检测与跟踪,结合颜色和空域位置信息,提出了一种新的人体轮廓建模方法,能较准确描述人体的特征,即使在有遮挡的情况下也能正确分割和跟踪人体。

参考文献:

[1] PARAGIOS N K, DERICHE R. A-PDE-based level-set approach for detection and tracking of moving objects [C]// Proceedings of

the 6th International Conference on Computer Vision. Washington, DC: IEEE Press, 1998.

- [2] COMANICIU D, RAMESH V, MEER P. Real-time tracking of non-rigid objects using Mean-Shift [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Press, 2000.
- [3] ELGAMMAL A, DURAI SWAMI R, HARWOOD D, *et al.* Background and foreground modeling using nonparametric kernel density estimation for visual surveillance [C]// Proceedings of the IEEE. Washington, DC: IEEE Press, 2002: 1151–1163.
- [4] ELGAMMAL A, DAVIS L S. Probabilistic framework for segmenting people under occlusion [C]// Proceedings of the IEEE Eighth International Conference on Computer Vision. Washington, DC: IEEE Press, 2001: 145–152.
- [5] KHAN S, SHAH M. Tracking people in presence of occlusion [C]// Asian Conference on Computer Vision. Taiwan: [s. n.], 2000: 263–266.
- [6] SENIOR A W. Tracking with probabilistic appearance models [EB/OL]. [2008-10-10]. <http://www.research.ibm.com/people/levision/PETS2002.pdf>.
- [7] HU WEI-MING, TAN TIE-NIU, WANG LIANG, *et al.* A survey on visual surveillance of object motion and behaviors [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2004, 34(3): 334–352.
- [8] WANG L, HU W, TAN T. Recent developments in human motion analysis [J]. Pattern Recognition Letters, 2003, 36(3): 585–601.
- [9] KHAN S, SHAH M. Tracking people in presence of occlusion [C]// Asian Conference on Computer Vision. Taiwan: [s. n.], 2000: 263–266.
- [10] McKENNA S J, RAJA Y, GONG S. Tracking color objects using adaptive mixture models [J]. Image Vision Computing, 1999, 17(3): 225–231.
- [11] McKENNA S J, JABRI S, DURIC Z, *et al.* Tracking groups of people [J]. Computer Vision Image Understanding, 2000, 80(1): 42–56.
- [12] ELGAMMAL A, DURAI SWAMI R, DAVIS L S. Efficient kernel density estimation using the fast Gauss transform for computer vision [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(11): 1499–1504.
- [13] FISCHLER M A, ROLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. Communications of ACM, 1981, 24(6): 381–395.
- [14] ELGAMMAL A, HARWOOD D, DAVIS L S. Non-parametric model for background subtraction [C]// 6th European Conference on Computer Vision. Dublin, Ireland: [s. n.], 2000: 751–767.
- [15] TOYAMA K, KRUMM J, BRUMITT B, MEYERS B. Wallflower: principles and practice of background maintenance [C]// Proceedings of the IEEE International Conference on Computer Vision. Washington, DC: IEEE Press, 1999: 255–261.
- [16] GREER D R, FUNG I, SHAPIRO J H. Maximum-likelihood multiresolution laser radar range imaging [J]. IEEE Transactions on Image Processing, 1997, 6(1): 36–46.