

一种基于主成分分析的 Codebook 背景建模算法

霍东海¹ 杨丹² 张小洪² 洪明坚²

摘要 混合高斯 (Mixture of Gaussian, MOG) 背景建模算法和 Codebook 背景建模算法被广泛应用于监控视频的运动目标检测问题, 但混合高斯的球体模型通常假设 RGB 三个分量是独立的, Codebook 的圆柱体模型假设背景像素值在圆柱体内均匀分布且背景亮度值变化方向指向坐标原点, 这些假设使得模型对背景的描述能力下降. 本文提出了一种椭球体背景模型, 该模型克服了混合高斯球体模型和 Codebook 圆柱体模型假设的局限性, 同时利用主成分分析 (Principal components analysis, PCA) 方法来刻画椭球体背景模型, 提出了一种基于主成分分析的 Codebook 背景建模算法. 实验表明, 本文算法不仅能够更准确地描述背景像素值在 RGB 空间中的分布特征, 而且具有良好的鲁棒性.

关键词 混合高斯模型, 运动目标检测, Codebook 算法, 主成分分析

引用格式 霍东海, 杨丹, 张小洪, 洪明坚. 一种基于主成分分析的 Codebook 背景建模算法. 自动化学报, 2012, 38(4): 591–600

DOI 10.3724/SP.J.1004.2012.00591

Principal Component Analysis Based Codebook Background Modeling Algorithm

HUO Dong-Hai¹ YANG Dan² ZHANG Xiao-Hong² HONG Ming-Jian²

Abstract The background modeling algorithm of mixture of Gaussian (MOG) and codebook is widely used in moving object detection of surveillance video. However, the ball model of MOG usually assumes that the three components of RGB are independent, while the cylinder model of codebook assumes that the value of background pixel is distributed uniformly within the cylinder and the changing direction of brightness points to the origin of the coordinate system. These assumptions reduce the descriptive capability for background modeling. Therefore, the paper proposes an ellipsoid-based background model, which overcomes the MOG and codebook's limitations. By using principal component analysis to depict the ellipsoid background model, a novel PCA-based codebook background modeling algorithm is proposed. Experiments show that this algorithm can not only give more accurate description of the distribution of background pixels but also have a better robustness.

Key words Mixture of Gaussian (MOG), motion detection, codebook, principal components analysis (PCA)

Citation Huo Dong-Hai, Yang Dan, Zhang Xiao-Hong, Hong Ming-Jian. Principal component analysis based codebook background modeling algorithm. *Acta Automatica Sinica*, 2012, 38(4): 591–600

随着视频监控摄像头在智能监控领域的广泛应用与智能视频分析技术的快速发展, 监控视频摘要技术逐渐进入人们的视线, 视频摘要技术采用一种将长时间的海量数据压缩到一个可控的时间段内的

方法, 提供给人们浏览监控事件. 其中如何设计一种高效实时的背景建模算法, 从而可以轻易进行运动前景的检测和运动物体的跟踪成为了视频摘要领域一个研究热点. 目前使用比较广泛的背景建模算法主要有背景差分法、核密度估计、混合高斯背景建模和 Codebook 背景建模等 4 种.

Heikkila 等^[1-2] 提出了背景差分法, 该算法使用了背景相减的方法, 即通过将待处理视频序列减去事先给定的一幅背景图像再应用二值化的方法来得到运动前景. 该算法的优点是容易实现, 算法复杂度低, 几乎不会耗费多少计算资源就可以快速地获取运动前景, 进而应用到实时视频摘要生成系统中去. 但是, 该算法需要事先给定背景图像, 且图像不易获取, 依赖外部输入, 且不能够随着时间推移进行更新, 导致视频处理后期的工作出现较大的误差.

Elgammal 等^[2-3] 提出了非参数背景建模方法, 估计背景像素值在时间序列上的概率密度, 通过一个长为 L 的窗口以及窗函数估计像素值属于前景或

收稿日期 2011-03-28 录用日期 2011-07-16
Manuscript received March 28, 2011; accepted July 16, 2011
国家自然科学基金 (60975015), 中央高校基本科研业务费专项资金 (CDJXS11181162), 中央高校基本科研业务费科研重点专项 (CDJZR11095501), 重庆市重点科技攻关项目 (CSTC2009AB2230), 重庆市自然科学基金 (CSTS2010BB2061) 资助
Supported by National Natural Science Foundation of China (60975015), Fundamental Research Funds for the Central Universities (CDJXS11181162), Fundamental Research Funds for the Central Universities (CDJZR11095501), Key Program for Science and Technology Development of Chongqing (CSTC2009AB2230), and Natural Science Foundation of Chongqing (CSTS2010BB2061)
本文责任编辑 周杰
Recommended by Associate Editor ZHOU Jie
1. 重庆大学计算机学院 重庆 400044 2. 重庆大学软件工程学院 重庆 400044
1. College of Computer Science, Chongqing University, Chongqing 400044 2. School of Software Engineering, Chongqing University, Chongqing 400044

者背景的概率, 从而判断该像素是否为背景像素. 该算法的优点是方便适应新的训练样本加入, 为密度估计的在线学习提供便利. 但是, 该算法计算复杂度太大, 不适合应用于实时运动检测系统中, 并且在动态背景和光线突变条件下的鲁棒性不高.

Wren 等^[4] 使用单高斯模型对背景进行建模, 克服了需要外部输入背景的限制, 在室内等单峰环境下具有较好的检测效果, 但是在复杂的多峰环境中, 如波动的湖面和摆动的树叶, 就很难使用该模型对背景环境进行准确地建模了. 为了解决这些问题, Stauffer 等^[2, 5-7] 提出了混合高斯 (Mixture of Gaussian, MOG) 背景建模算法, 该算法充分考虑了像素点在时间上的连续性, 将背景像素在时间序列上的分布假设为混合高斯模型, 并假设了 RGB 空间的三个分量之间是相互独立的, 给出了背景像素的分布特征, 即背景像素在 RGB 空间中呈现球状分布. 但是, RGB 空间中的三个分量并不是相互独立的, 因此该方法对背景像素的分布特征的描述不是很准确, 导致运动前景检测误差的增大.

Kim 等^[8-9] 提出了一种结构化的 Codebook 背景建模方法, 该方法利用圆柱体近似地表示了背景像素在 RGB 空间中的分布特征, 在多峰环境下得到了较好的实验效果; Wu 等^[10] 在 Kim 的基础上提出了基于时空上下文的 Codebook 改进方法, 避免了仅考虑一个像素的局限性, 通过添加时空像素关联提高了算法的准确度; Tu 等^[11] 提出了利用长方体 (Box-based) 近似表示背景像素值在 RGB 空间中的分布特征的方法, 降低了算法在圆柱匹配的计算中的复杂度, 提高了原始 Codebook 算法的运算速度. 为了消除前景检测中运动物体的阴影, Doshi 等^[12] 提出了圆柱圆锥混合 (Hybrid cone-cylinder) 的 Codebook 背景模型. 值得注意的是, 这些算法都基于如下假设: Codeword 圆柱体的中心轴指向 RGB 空间的原点, 并且背景像素在 RGB 空间中是均匀分布的. 本文通过实验观察到背景像素在 RGB 空间中的分布并不是 Codebook 算法假设的均匀分布, 而是呈现中间宽两头尖的多维高斯分布, 并且 RGB 三个分量也不相互独立.

因此, 本文针对混合高斯背景模型和 Codebook 背景模型假设存在的问题, 提出了一种基于主成分分析 (Principal components analysis, PCA) 的背景建模算法. 该算法克服了混合高斯背景建模方法对 RGB 空间三个分量独立和 Codebook 背景建模算法中圆柱体中心轴指向 RGB 空间原点的局限性, 利用椭球体近似表示背景像素分布特征, 应用主成分分析刻画背景像素的变化趋势, 从而得到更加精确的背景像素分布模型, 同时结合阈值法进行运动前景检测. 该模型不仅能更准确地描述背景像素在

RGB 空间的真实分布, 而且具有良好的实时性和鲁棒性, 在一系列的实验中都取得了较好的效果.

1 基于主成分分析的背景建模

1.1 Codebook 算法模型^[8]

视频序列图像中同一特定位置的像素点在时间轴上的观测值可以构成一个序列 $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t\} = \{I(R_i, G_i, B_i) | 1 \leq i \leq t\}$, 其中 \mathbf{X} 代表某一像素, t 和 i 代表时间. Codebook 算法为每一个像素数组建立一个称为 Codebook 的模型: $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_L\}$, 每个 Codebook 包含有 L 个称为 Codeword 的 \mathbf{c}_i . 在 RGB 颜色空间中, 每个 Codeword 包含一个三维向量 $\mathbf{V}_i = (\bar{R}_i, \bar{G}_i, \bar{B}_i)$ 和一个 6 元组 $\mathbf{aux}_i = (\hat{I}_i, \check{I}_i, f_i, \lambda_i, p_i, q_i)$. 其中:

- 1) $\bar{R}_i, \bar{G}_i, \bar{B}_i$ 为所有属于该类 Codeword 的像素的均值;
- 2) \hat{I}_i, \check{I}_i 为所有属于该类 Codeword 的像素中最小和最大亮度值;
- 3) f_i 为属于该类 Codeword 的像素个数, 也可理解为匹配成功的次数;
- 4) λ_i 为没有被匹配的最大时间间隔;
- 5) p_i, q_i 为第一次和最后一次匹配成功的时间.

Codebook 算法的思想是假设所有背景像素值都分布在 Codeword 圆柱体内部, 而前景像素值分布在外, Codeword 圆柱体的表面便构成了判别边界, 通过判别式 (1) 就能判断输入像素值是否为背景像素, 从而进行运动检测.

$$\text{brightness}(I, \langle \hat{I}_i, \check{I}_i \rangle) = \sqrt{R^2 + G^2 + B^2} \in \langle I_{\text{low}}, I_{\text{hi}} \rangle \quad (1)$$

$$\delta = \text{colordist}(\mathbf{X}_t, \mathbf{V}_i) \leq \varepsilon \quad (2)$$

其中, 各变量的计算方法如下:

$$I_{\text{low}} = \alpha \hat{I}_i \quad (3)$$

$$I_{\text{hi}} = \min \left\{ \beta \check{I}_i, \frac{\hat{I}_i}{\alpha} \right\} \quad (4)$$

$\varepsilon, \alpha, \beta$ 是先验参数.

$$\delta = \text{colordist} = \sqrt{\|\mathbf{X}_t\|^2 - \frac{\langle \mathbf{X}_t, \mathbf{V}_i \rangle^2}{\|\mathbf{V}_i\|^2}} \quad (5)$$

$$\|\mathbf{X}_t\|^2 = R^2 + G^2 + B^2 \quad (6)$$

$$\|\mathbf{V}_i\|^2 = \bar{R}_i^2 + \bar{G}_i^2 + \bar{B}_i^2 \quad (7)$$

$$\langle \mathbf{X}_t, \mathbf{V}_i \rangle^2 = (R\bar{R}_i + G\bar{G}_i + B\bar{B}_i)^2 \quad (8)$$

原始 Codebook 算法的优势是可以在不影响前景检测效果的基础上对背景模型进行压缩, 大大减少了

对内存的需求和计算量; 由于视频序列背景像素的变化主要集中在亮度的变化上, Codebook 算法通过一个简单而有效的方法规定了背景模型的亮度上下界, 从而降低了全局和局部光照变化对背景模型的影响; Codebook 算法通过引入 Codeword 被匹配的最大时间间隔很好地解决了训练阶段存在运动前景的问题。

由于原始 Codebook 算法是基于对背景像素值的统计观察而提出的, 算法模型对背景像素值分布的刻画不够准确, 比如亮度的计算 (式 (1)) 导致所有模型中的 Codeword 都是指向 RGB 颜色空间坐标原心的圆柱体, 并且在 Codebook 算法中背景像素值是均匀分布在圆柱体内部。

1.2 背景像素值分布

为了能够解决 Codebook 算法对背景分布刻画不准确的问题, 本文对 Toyama 等^[13] 视频序列进行了采样实验, 通过分别观察 RGB 空间三个分量的直方图统计信息, 了解背景像素值的分布情况。首先对输入的训练视频序列 $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_i\}$ 执行采样得到 RGB 三维向量 \mathbf{X}_i , 然后将其保存到样本集 $Sample[M \times N]$ 中, 其中 $M \times N$ 代表图像的分辨率, $Sample[k]$ 代表第 k 个像素序列数组 $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_i\}^k$ 。因此一段视频序列采样后可表示为

$$\begin{cases} Sample[M \times N] \\ Sample[k] = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_i\}^k, \quad (9) \\ 1 \leq k \leq M \times N \end{cases}$$

采样后将得到的像素样本序列值投影到 RGB 空间下观察背景像素值分布情况 (如图 1 所示)。为了进一步确定背景像素值的分布情况, 本文通过大量实验观察到不同位置的像素在 RGB 空间的分布情况不尽相同, 在图像序列 WavingTree^[13] 中主要分为三类情况 (如图 2 所示): 1) 对坐标为 (10, 10) 的点, 背景像素值分布在一个狭小的区域中 (如图 2(e)), 并且在 RGB 三个分量上都近似服从相同的高斯分布 (如图 2(i), 图 2(m) 和图 2(q)), 这样的像素值分布在 RGB 空间中呈现一个球状体; 2) 对坐标为 (24, 142) 和 (112, 50) 的点, 由于该点处在背景纹理较复杂的区域 (如图 2(b) 和图 2(d)), 摄像机在成像时由于外部光线变化的影响, 使得成像后的像素值在一个合适的范围内变化, 导致像素值分布在 RGB 空间中呈现一个椭球体 (如图 2(f) 和图 2(h)), 同时图 2(j), 图 2(l), 图 2(n), 图 2(p), 图 2(r) 和图 2(t) 所示的 RGB 三个分量上的直方图证明了该分布假设; 3) 对坐标为 (50, 50) 的点, 由于该点处于图像中间晃动的树枝上, 因此成像时像素值时而呈现较深的树枝像素, 时而呈现较浅的天

空像素, 有时又呈现这两种情况的中间状态值, 导致背景像素值呈现多模态的状态, 如图 2(g), 虽然背景像素值还是分布在一个狭长的椭球体区域中, 但是图 2(k), 图 2(o) 和图 2(s) 的直方图还是显示出多峰分布的本质。

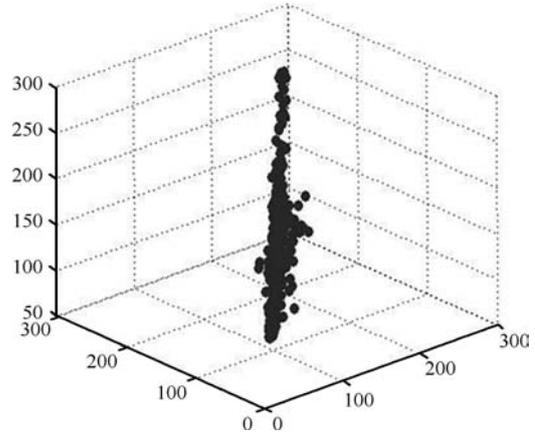


图 1 背景像素值分布

Fig. 1 Background pixel distribution

在对背景像素值分布进行建模时, Stauffer 等^[5] 提出的混合高斯 (MOG) 模型假设 RGB 三个分量相互独立, 且 RGB 每一维上像素的分布均服从相同的高斯分布, 因此其模型在 RGB 空间中呈现如图 3(c) 所示的球体分布; 而 Kim 等^[8] 假设背景像素值是分布在一个如图 1 中所示的狭长的圆柱区域内, 并假设圆柱的中轴是指向坐标系原心的位置 (如图 3(a)). 这两种算法都考虑到了背景像素值分布的一种情况, 球体分布或者是圆柱体分布, 都没有能够准确刻画出背景像素的分布情况。

1.3 基于主成分分析的背景建模

基于上述观察实验, 结合混合高斯算法和 Codebook 算法对背景像素值分布的假设, 本文提出了一种基于主成分分析的背景建模方法。其主要思想是在原始 Codebook 算法模型的基础上, 对训练过程中的所有背景像素值进行主成分分析运算, 得到背景像素变化的主方向和其他两个非主方向, 然后结合阈值法在各个成分方向上确定背景阈值, 从而精确刻画出背景像素值在 RGB 空间中的分布情况 (如图 3(b)). 该算法相比于混合高斯算法和原始 Codebook 算法的优点在于对背景分布的刻画上, 如图 3(d) 所示背景像素值的分布并不一定呈现球体状态, 同时也不一定指向坐标系原点, 并且 RGB 空间中三个分量也不相互独立。真实的像素值分布在如图 3(b) 所示的一个椭球体当中, 背景和前景的决策面就是椭球体的表面, 属于前景的像素分布在椭球体外部, 属于背景的像素分布在其内部。尤其是在纹

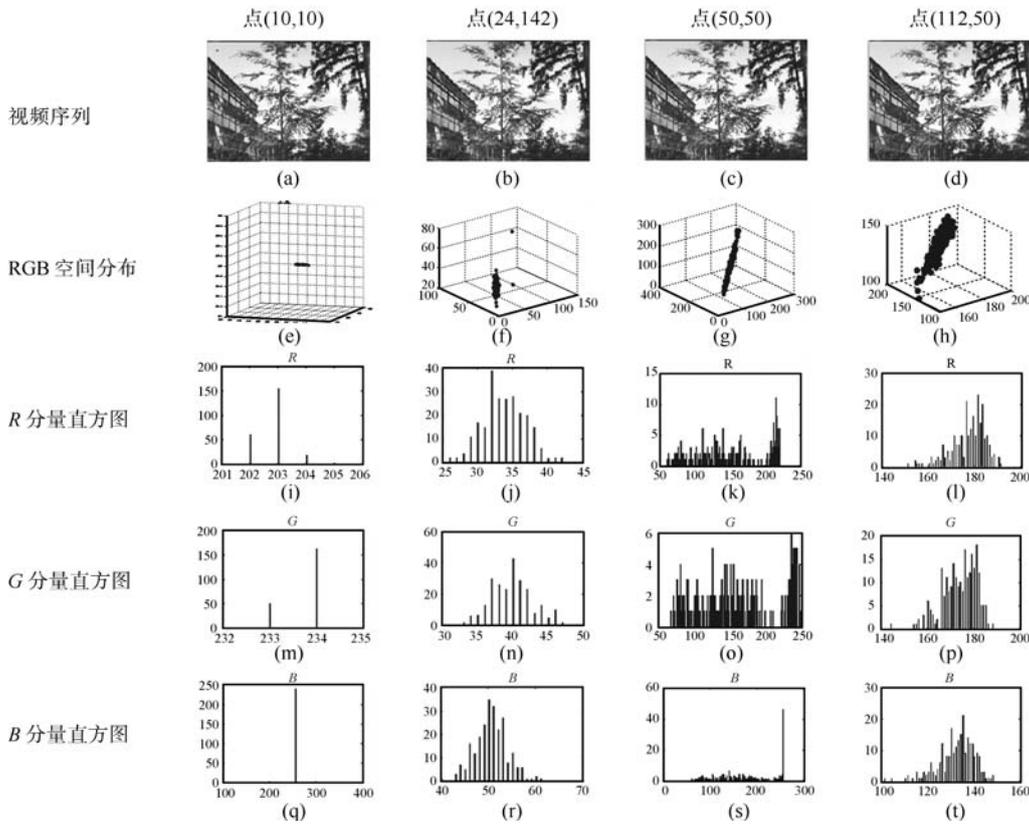


图2 图像序列采样分布和直方图

Fig.2 Sampling distribution and histogram

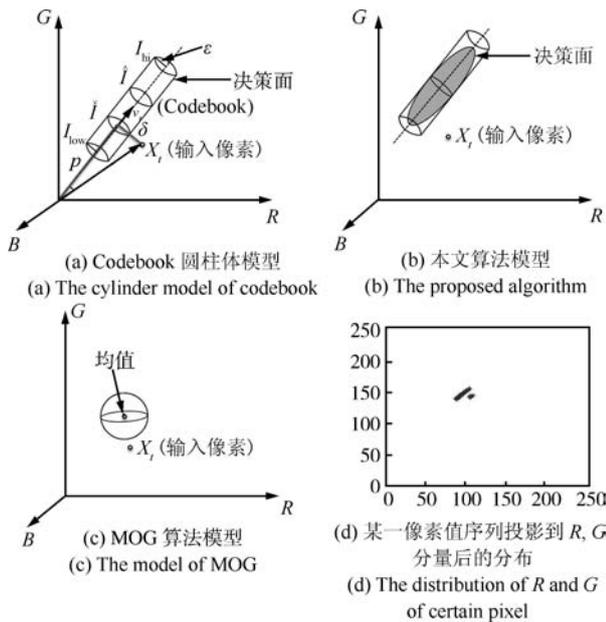


图3 算法模型

Fig.3 Model of the algorithm

理较复杂的背景条件下,背景像素并不一定分布在指向原心的中轴周围,而会偏向其他方向.本文算法通过引入主成分分析对呈现椭球体分布的背景像素

值进行描绘,更加精确地刻画了背景分布情况.

2 基于主成分分析的前景检测算法

本文提出的一种基于主成分分析的 Codebook 背景建模算法运用于视频序列的前景检测应用中,主要分为三个阶段的工作.第一阶段为背景模型训练阶段,主要是将训练阶段的所有背景像素值存储到一个数组中,然后对该数组进行主成分分析,计算出主成分量和其他两个次要量的特征向量,最后通过阈值确定方案获取前景检测阈值;第二阶段为前景检测,在该阶段中主要根据阈值判别输入的未知像素为前景还是背景,判别时根据不同的阈值选择不同的判别方法;第三阶段为背景更新,背景更新主要是为了能够适应视频序列中背景的变化,提高前景检测的鲁棒性.

2.1 背景模型训练

在背景模型训练阶段,首先对训练阶段视频序列中的像素进行采样,然后将采样得到的 RGB 颜色空间的像素值保存到式 (10) 所示的数组中,然后再通过主成分分析计算出主成分量和其他两个次要

分量, 最后计算出新线性空间下的前景检测阈值.

$$\begin{cases} \text{TrainingData}[M \times N] \\ \text{TrainingData}[k] = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_i\}^k, \quad (10) \\ 1 \leq k \leq M \times N \end{cases}$$

主成分分析算法是一个线性变换, 通过该变换可以把数据变换到一个新的坐标系统中, 使得任何数据投影的第一大方差在第一个坐标轴 (称为第一主成分, 用 PC1 表示) 上, 第二大方差在第二个坐标轴 (称为第二主成分, 用 PC2 表示) 上, 依次类推. 主成分分析经常用于减少数据集的维数, 同时保持数据集对方差贡献最大的特征, 在背景建模问题中该特征方向对应背景像素在复杂的纹理条件下由于摄像机成像或者树枝晃动等动态背景而导致的变化的主要方向, 即 RGB 颜色空间中椭球体的主轴方向 (图 4(a)).

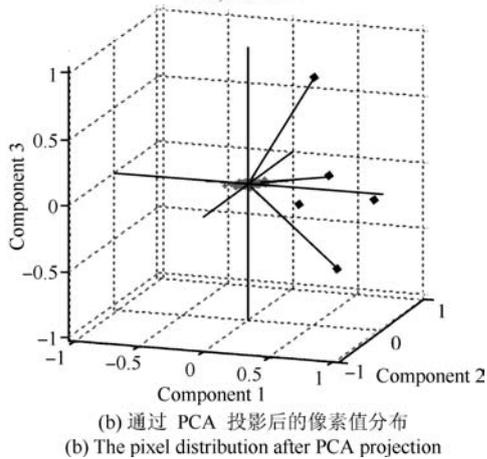
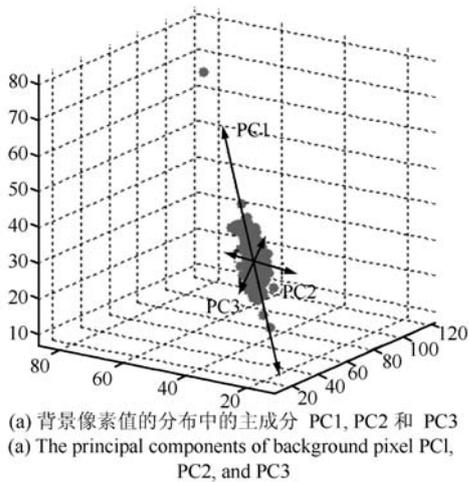


图 4 PCA 背景建模算法模型

Fig. 4 Model of the PCA background modeling algorithm

主成分分析步骤中有个样本数据, 每一个数据的维度相同, 即 RGB 三维. 算法的步骤如下:

步骤 1. 使用式 (11) 计算均值 $\bar{\mathbf{X}}$;

$$\bar{\mathbf{X}} = \frac{1}{M \times N} \sum_{i=1}^{M \times N} \mathbf{X}_i \quad (11)$$

步骤 2. 使用式 (12) 计算样本集的协方差矩阵;

$$\Sigma = \frac{1}{M \times N - 1} (\mathbf{X}_i - \bar{\mathbf{X}}) \times (\mathbf{X}_i - \bar{\mathbf{X}})^T \quad (12)$$

步骤 3. 如式 (13) 计算协方差矩阵的特征值 $\lambda_1, \lambda_2, \lambda_3$ 以及与特征值对应的特征向量 $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, 其中特征值按从大到小排列: $\lambda_1 > \lambda_2 > \lambda_3$, 较大的特征值对应的特征向量正好对于椭球体的较长轴;

$$\begin{cases} P = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3] \\ \Sigma = P^T \times \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \end{cases} \quad (13)$$

步骤 4. 使用式 (14) 按照调整过顺序的特征向量矩阵 P 投影训练样本集数据.

$$\mathbf{X}^{\text{projection}} = P^T \times \mathbf{X} \quad (14)$$

在步骤 3 中特征值从大到小排列的原因在于: 特征值越大, 背景像素在该特征方向的变化越大, 反之越小. 在步骤 4 中, 使用主成分分析计算出样本数据像素序列集的前三个相互正交的主成分方向后, 将数据集投影到相应的特征向量方向上, 投影后的数据如图 4(b) 所示.

计算出背景像素值的主成分方向后还需要给定前景检测时需要用到的阈值. 阈值确定可以选择训练样本数据到背景中心点的马氏距离的最大值或者 n 倍标准差. 两种阈值的区别是最大马氏距离容易获取, 但鲁棒适应性较差, 导致前景检测到的误差较大, 而 n 倍标准差需要多一步来计算标准差, 但得到的背景模型较稳定, 前景检测效果较好. 因此本文选择 n 倍标准差作为模型的阈值来进行前景检测. 同时为了能够提高算法前景检测的实时性, 本文还给出了一种利用长方体模型代替椭球体模型的阈值替代方法, 该方法通过一个外接在椭球体外部的长方体来代替原始模型, 从而可以直接在三个主成分分量上分别确定阈值来加快运算速度.

1) 最大马氏距离阈值法

如式 (15), 在背景建模训练过程中, 记录每个训练数据 \mathbf{X}_i 到背景中心的马氏距离的最大值作为前景检测时使用的阈值 (如图 5(a)). 使用该方法的优点是直观理解, 在 RGB 空间中分布在椭球体内部的像素为背景, 外部的则为前景; 缺点是在背景更

新过程中不能随着检测时间的推移更新该背景阈值, 导致长时间视频处理时背景模型的鲁棒性不好.

$$Threshold = \max\{Distance_i\}, \quad 1 \leq i \leq M \times N \quad (15)$$

其中, $Threshold$ 表示阈值.

2) n 倍标准差阈值法

根据切比雪夫不等式(式(16)), n 倍标准差范围内包含了 $1 - 1/n^2$ 的训练数据, 因此可以在式(17)中通过调整 n 来缩紧和放宽背景模型的范围, 进而适应不同场景下的应用(如图5(b)). 并且随着时间的推移, 阈值的大小可以通过标准差的更新而得到调整. 该方法的优点是可以随着背景的更新而更新阈值, 具有一定的鲁棒性; 缺点是需要计算标准差, 增加了前景检测时的计算量.

$$P\{|\mathbf{X} - E(\mathbf{X})| \geq \xi\} \leq \frac{D(\mathbf{X})}{\xi} \quad (16)$$

其中, P 表示概率, \mathbf{X} 表示训练数据, $E(\mathbf{X})$ 表示训练数据期望, $D(\mathbf{X})$ 表示训练数据方差, ξ 表示任意正数.

$$Threshold = n \times \sqrt{\text{tr}(\Sigma)} \quad (17)$$

其中, $\text{tr}(\Sigma)$ 表示协方差矩阵 Σ 的迹.

3) 长方体模型阈值法

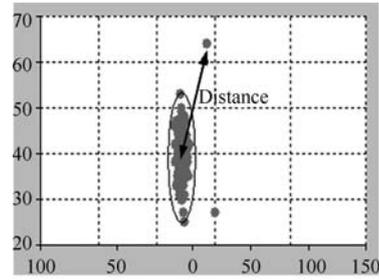
在本文算法中, 由于经过主成分分析投影后的数据的各维都是相互独立的, 因此如式(18)可以通过在投影后的三维中分别赋予上下界的方法来确定阈值. 这样就可以在前景检测时利用长方体模型近似地代替椭球体模型(如图5(c)), 减少了计算量, 提高了检测速度, 满足实时前景检测的需求.

$$Threshold_i = n \times \sqrt{\lambda_i}, \quad 1 \leq i \leq 3 \quad (18)$$

其中, $Threshold$ 表示第 i 个主成分的阈值, λ_i 表示第 i 个主成分的方差.

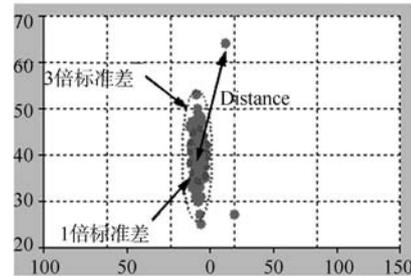
在前景检测时, 对于一个输入的测试数据, 利用式(19)分别计算其与中心点的距离, 只有在三个成分的距离都小于其对应的阈值的情况下才认定其为背景, 否则就是前景.

$$\begin{cases} Background, \forall i, 1 \leq i \leq 3, \\ \quad Distance_i \leq Threshold_i \\ Foreground, \exists i, 1 \leq i \leq 3, \\ \quad Distance_i > Threshold_i \end{cases} \quad (19)$$



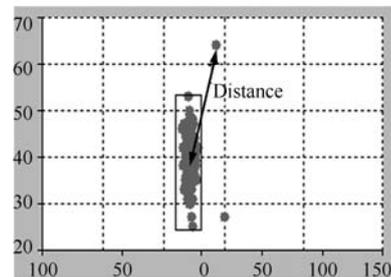
(a) 最大马氏距离阈值法

(a) The maximum Mahalanobis distance



(b) n 倍标准差阈值法

(b) The n times standard deviation



(c) 长方体模型阈值法

(c) The rectangle threshold

图5 阈值确定方法

Fig. 5 Methods of threshold determination

2.2 前景检测

对输入的测试样本同样按照背景建模中使用主成分分析得到的正交基进行投影变换, 利用式(20)计算变换之后的数据与背景模型中心点之间的马氏距离. 在前景检测中, 如式(21)所示若该距离小于用户输入的前景检测阈值, 则判定该测试样本为背景, 反之则为前景.

$$Distance_i = (\mathbf{X}_i - \bar{\mathbf{X}})^T \Sigma (\mathbf{X}_i - \bar{\mathbf{X}}), \quad 1 \leq i \leq T \quad (20)$$

其中, $Distance_i$ 表示第 i 个点与背景模型中心的马氏距离.

$$\begin{cases} Background, Distance \leq Threshold \\ Foreground, Distance > Threshold \end{cases} \quad (21)$$

前景检测关键的步骤之一就是阈值的确定, 阈值的好坏直接影响了检测效果的优劣, 鲁棒的、自适应的

阈值可以适应不同的场景下的前景检测任务。

2.3 背景更新

在背景更新步骤中, 为了能够适应长时间视频序列的处理需求, 本文设计了两个层次的背景更新方案, 第一层次是前景检测, 第二层次是长时间序列的背景更新方案。

1) 前景检测阶段的背景更新

在进行前景检测时, 对于检测为前景的像素使用 MOG^[5] 中的背景更新策略 (式 (22)), 这样可以在前景检测时实时更新背景模型来适应背景场景条件的变化。在 MOG 的背景更新方法中, 使用学习率 α 和 β 分别对背景模型的均值 \bar{x} 和方差 Σ 进行更新。

$$\begin{cases} \bar{\mathbf{X}} = (1 - \alpha)\bar{\mathbf{X}} + \alpha\mathbf{X}_i \\ \Sigma = (1 - \beta)\Sigma + \beta(\mathbf{X}_i - \bar{\mathbf{X}})^{M \times N}(\mathbf{X}_i - \bar{\mathbf{X}}) \end{cases} \quad (22)$$

其中, $\Sigma = \sigma_i^2 I, 1 \leq i \leq 3$ 。

2) 长时间序列的背景更新

长时间序列的背景更新策略如图 6 所示, 针对第一层次的背景更新不够快, 不能适应误差的积累等不适合长时间视频序列的问题, 本文提出在前景检测已经进行了一段时间后, 使用在前一阶段检测出来的背景像素值重新建立背景模型, 然后再进行背景更新的方法。如图 6 中所示, 该方法可以自定义训练时间和检测时间的关系, 通过调节 K 来适应不同场景的背景更新任务。在此更新策略的作用下, 本文算法能够胜任长时间的前景检测任务。

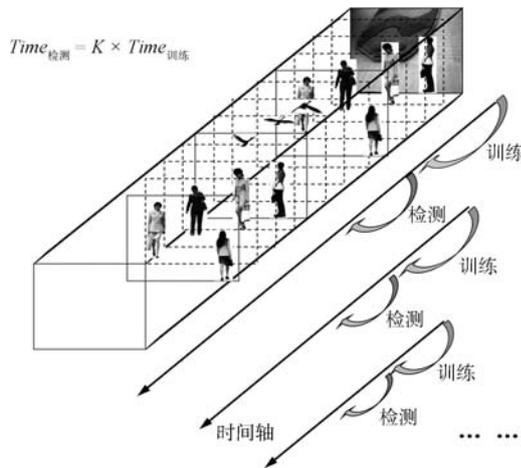


图 6 长时间序列背景更新

Fig. 6 Background update of long-time video

3 实验与讨论

为了验证本文算法的有效性和实用性, 在 Matlab 2010b 编程工具下实现了本文算法, 实验的硬件环境为 Intel Pentium(R) 4 3.0 GHz 处理器, 2.0 GB

内存。为了能够体现本文算法在复杂纹理背景条件下的优势, 本文选择了 Wallflower^[13] 中的 WavingTrees, MovedObject 和一个自行拍摄的带有大树和雾背景的视频序列作为测试数据集, 且图像分辨率均为 160 像素 \times 120 像素。其中 WavingTrees 序列中包含了空旷的天空、背景纹理复杂的固定大树和晃动的树枝, 共 287 张; MovedObject 序列中包含了一个本来静止后来被移动过的椅子的背景, 共 1744 张; 自行拍摄的序列包含了一棵茂密的大树和远处有雾的背景, 共 701 张。

在进行对比实验时, 本文实现了混合高斯 (MOG)^[5] 和 Codebook^[8] 背景建模算法。混合高斯算法应用混合高斯模型对背景像素进行建模, 假设背景像素在 RGB 空间分布在一个混合高斯分布模型中, 是目前应用最广泛的方法; Codebook 算法是假设背景像素值在 RGB 空间分布在若干个称为 Codeword 的圆柱体中, 该算法采用结构化的方法简化了背景模型, 得到了较好的效果。

3.1 ROC 曲线分析

ROC (Receiver operating characteristic) 曲线即接收者操作特征曲线是一种对于灵敏度进行描述的功能图像, ROC 曲线可以直观地展示分类算法的优劣。一个前景检测算法即一个前景背景分类器, 它是将一个实例映射到一个特定类的过程。分类器中类的边界是通过一个阈值检测实现的。在前景检测应用中, 一个像素的分类结果或者是真 (P) 或者是假 (N), 因此分类器有 4 类可能的输出, 如果某一像素输出的分类结果是前景 (P) 而该像素真实的也是前景 (P), 那么称之为真阳性 (TP); 然而如果像素真实的结果是背景 (N), 则称之为假阳性 (FP); 相反地, 一个真阴性发生在分类结果和实际情况都为背景 (N) 的时候; 而假阴性是当分类结果为背景 (N) 而实际值为前景 (P) 的时候。

由于本文算法所解决的前景检测问题是在像素级别上的, 因此对于一个单独的像素而言, 前景检测问题本质上是一个分类问题, 即判断一个输入像素是前景类还是背景类。所以在 ROC 曲线分析实验中, 本文在已实现的混合高斯和 Codebook 算法的基础上对 WavingTrees 视频序列的第 247 帧、自行拍摄视频序列的第 550 帧和 MovedObject 视频序列的第 985 帧进行分析, 首先手工分割出图像的真实前景 (Ground truth) 图像, 然后通过调节各种算法的参数获取了 ROC 曲线数据, 其实验结果如图 7 所示。

在图 7 中, 左上角为 WavingTrees 序列第 247 帧的结果, 右上角是自行拍摄序列的第 550 帧, 左下角是 MovedObject 序列第 985 帧, 右下角是前三个

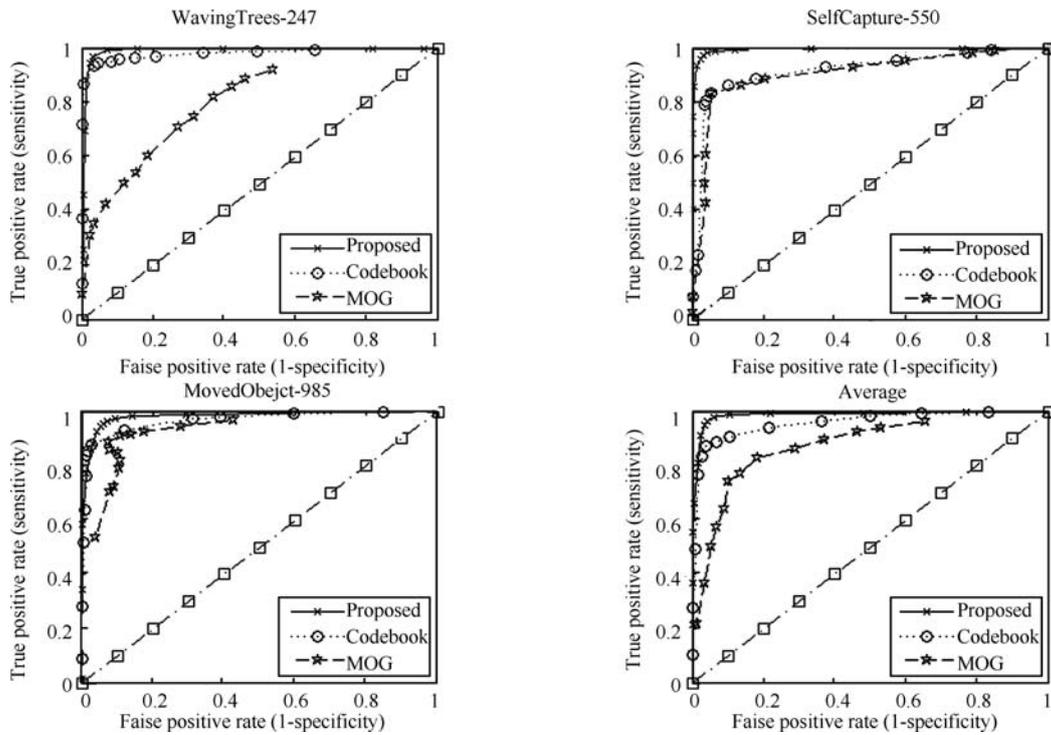


图7 不同算法 ROC 曲线

Fig. 7 ROC curves of the algorithms

结果的平均值. 从图7中可知, 本文算法在三个实验中的 ROC 曲线距离对角线都比其他两种算法远, 且曲线下方的面积明显大于其他两种算法的曲线, 因此, 本文算法在分类效果上明显优于 Codebook 和混合高斯算法.

3.2 实际检测效果

在实际检测效果实验中, 三种算法的参数选取如表1所示, 训练与测试时间相同, 如表2所示. 图8给出4组测试视频序列集的实际前景检测效果, 其中对于室外复杂场景的 WavingTrees 图像序列, 由于其背景包含了复杂多变的背景, 因此各种算法检测出来的前景都有很多误检情况, 但是对比三种算法的检测结果, 本文算法还是得到了比较完整的前景团块, 这是由于在室外复杂场景下背景像素值在 RGB 空间的分布呈现椭球体, 正好符合了本文算法的假设, 而其他两种算法在对这种背景特征的刻画上不够准确, 导致检测误差较大; 对于 MovedObject, 由于视频序列背景中经过了一次椅子的移动, 同时桌面上放置了一本书, 本文算法和 Codebook 算法都很好检测出来背景的变化, 而 MOG 算法只是检测出一部分的内容; 对于自拍序列的第550帧和631帧, 由于整个视频序列中背景存在着较复杂的大树和被雾遮盖的背景, MOG 算法在大树附

近的检测效果较差, Codebook 算法次之, 本文算法给出了三种算法中最好的效果. 综上所述, 本文算法对比于 MOG 算法和 Codebook 算法的优点是在视频序列中存在纹理复杂的背景时的检测效果更好, 而对于简单平稳纹理的背景各个算法的检测效果都差不多.

总体来说, 对于复杂的室外监控场景, 本文的算法具有较高的鲁棒性, 对前景的检测效果可以很好地满足后期跟踪算法的需求.

但是对于像 Camouflage 和 ForegroundAperture 这种多模态的视频序列, 本文算法的检测效果不够理想. 如图9所示, 上面两行是 Camouflage 视频序列, 其中第1行分别为原始帧、真实前景以及本文算法检测效果, 第2行是点 (60, 75) 的采样结果, 分别表示 RGB 三个分量上的直方图统计信息; 下面两行是 ForegroundAperture 视频序列, 其中第3行分别表示原始帧、真实前景以及本文算法检测效果, 第4行是点 (110, 70) 在 RGB 三个分量上的直方图统计信息. 从直方图中可以明显地看出在多模态的背景像素的情况下, 直方图中明显地出现两个峰值, 并且这两个峰值不连续, 所以背景像素值拥有多个模型.

由于本文算法并没有将这种多峰分布的特征考虑在内, 因此检测效果中出现了明显的空洞. 在以后

的工作中将继续考虑多峰分布的特征, 进一步修改算法来达到更好的效果, 尤其是在多模态背景下.

表 1 算法参数选取

Table 1 Parameters of the algorithms

算法	参数
Codebook	$\alpha = 0.4, \beta = 1.5, \xi = 100$
MOG	$k = 3, \alpha = 0.005$
Proposed	$n = 3$

表 2 图像序列训练与测试时间

Table 2 Training and testing time

图像序列	训练帧	测试帧
WavingTrees	1 ~ 200	247
MovedObject	1 ~ 200	985
SelfCapture	320 ~ 510	550
SelfCapture	320 ~ 510	631

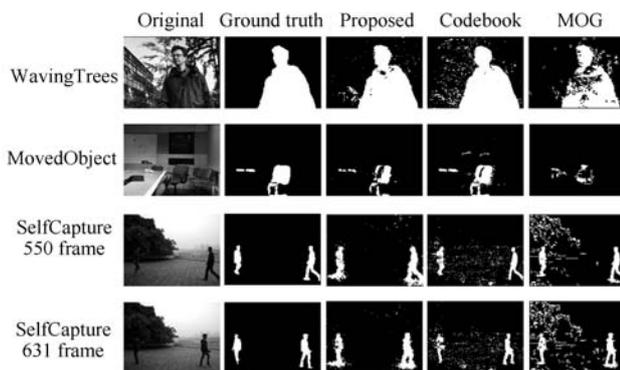


图 8 不同算法实验结果

Fig. 8 Experimental results of each algorithm

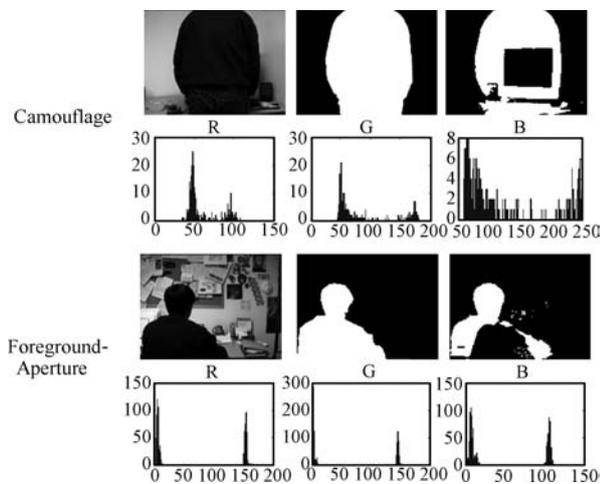


图 9 多模态背景下实验结果

Fig. 9 Experimental results of multi-model background

4 结论

运动前景检测是监控视频处理领域的关键问

题, 本文提出了一种新的基于主成分分析的 Codebook 背景建模算法. 该算法利用主成分分析计算背景像素在 RGB 颜色空间中分布的主轴, 然后给三个主成分分别确定阈值, 应用于前景检测当中. 椭球体模型准确刻画了背景像素在 RGB 颜色空间的分布特征, 提高了背景模型对像素变化的描述能力, 同时引入主成分分析, 既解决了不同环境下像素变化方向无法确定的难题, 又克服了传统背景建模算法对背景表征不足的问题. 通过在共同测试数据集上的实验表明, 对比其他背景建模算法, 本文算法可以得到较好的且满足后期跟踪任务需要的运动前景. 由于本文仅考虑了单模态背景环境的情况, 对于多模态的视频序列, 本文算法的检测效果不够理想, 对于多模态背景的背景建模问题将进一步考虑背景像素的混合分布特征, 建立一个由多个模型组合而成的背景模型来刻画分布特征.

References

- 1 Heikkila J, Silven O. A real-time system for monitoring of cyclists and pedestrians. In: Proceedings of the 2nd IEEE Workshop on Visual Surveillance. Fort Collins, USA: IEEE, 1999. 74–81
- 2 Piccardi M. Background subtraction techniques: a review. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics. The Hague, Netherlands: IEEE, 2004. 3099–3104
- 3 Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction. In: Proceedings of the 6th European Conference on Computer Vision. Dublin, Ireland: Springer, 2000. 751–767
- 4 Wren C, Azarbayejani A, Darrell T, Pentland A. Pfunder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 780–785
- 5 Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Fort Collins, USA: IEEE, 1999. 246–252
- 6 Bouwmans T, Baf F E, Vachon B. Background modeling using mixture of Gaussians for foreground detection — a survey. *Recent Patents on Computer Science*, 2008, **1**(3): 219–237
- 7 Wang Yong-Zhong, Liang Yan, Pan Quan, Cheng Yong-Mei, Zhao Chun-Hui. Spatiotemporal background modeling

based on adaptive mixture of Gaussians. *Acta Automatica Sinica*, 2009, **35**(4): 371–378

(王永忠, 梁彦, 潘泉, 程咏梅, 赵春晖. 基于自适应混合高斯模型的时空背景建模. *自动化学报*, 2009, **35**(4): 371–378)

- 8 Kim K, Chalidabhongse T H, Harwood D, Davis L. Real-time foreground-background segmentation using code book model. *Real-Time Imaging*, 2005, **11**(3): 172–185
- 9 Chalidabhongse T H, Kim K, Harwood D, Davis L. A perturbation method for evaluating background subtraction algorithms. In: *Proceedings of the Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*. Nice, France: IEEE, 2003. 110–116
- 10 Wu M, Peng X. Spatio-temporal context for codebook-based dynamic background subtraction. *AEU-International Journal of Electronics and Communications*, 2010, **64**(8): 739–747
- 11 Tu Q, Xu Y, Zhou M. Box-based codebook model for realtime objects detection. In: *Proceedings of the 7th World Congress on Intelligent Control and Automation*. Chongqing, China: IEEE, 2008. 7621–7625
- 12 Doshi A, Trivedi M. “Hybrid cone-cylinder” codebook model for foreground detection with shadow and highlight suppression. In: *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance*. Sydney, Australia: IEEE. 2006. 19–24
- 13 Toyama K, Krumm J, Brumitt B, Meyers B. Wallflower: principles and practice of background maintenance. In: *Proceedings of the 7th IEEE International Conference on Computer Vision*. Kerkyra, Greece: IEEE, 1999. 255–261



霍东海 重庆大学计算机学院硕士研究生. 主要研究方向为图像处理, 模式识别, 监控视频分析. 本文通信作者.

E-mail: Tunghai.huo@gmail.com

(**HUO Dong-Hai** Master student at the College of Computer Science, Chongqing University. His research interest covers image processing, pattern recognition, and surveillance video analysis. Corresponding author of this paper.)



杨丹 重庆大学软件学院教授. 主要研究方向为图像处理, 机器视觉, 人工智能, 软件工程.

E-mail: dyang@cqu.edu.cn

(**YANG Dan** Professor at the School of Software Engineering, Chongqing University. His research interest covers image processing, machine vision, artificial intelligence, and software engineering.)



张小洪 重庆大学软件学院教授. 主要研究方向为数字图像处理与计算机视觉.

E-mail: xhongz@yahoo.com.cn

(**ZHANG Xiao-Hong** Professor at the School of Software Engineering, Chongqing University. His research interest covers digital image processing and computer vision.)



洪明坚 重庆大学软件学院副教授. 主要研究方向为近红外光谱分析, 模式识别与人工智能. E-mail: hmj@cqu.edu.cn

(**HONG Ming-Jian** Associate professor at the School of Software Engineering, Chongqing University. His research interest covers spectroscopy, pattern recognition, and artificial intelligence.)