

中图分类号: TP391 文献标志码: A 文章编号: 1006-8961(2011)05-0829-06

论文索引信息: 常晓夫, 张文生, 董维山. 基于多种类视觉特征的混合高斯背景模型[J]. 中国图象图形学报, 2011, 16(5): 829-834

基于多种类视觉特征的混合高斯背景模型

常晓夫, 张文生, 董维山

(中国科学院自动化研究所复杂系统与智能科学重点实验室, 北京 100190)

摘要: Stauffer 等人提出的混合高斯背景减除建模技术及其改进算法在真实场景的运动目标检测系统中取得了较好的检测效果且被人们广泛应用。然而, 此类方法通常采用单一的颜色视觉特征进行建模。当运动目标的表观颜色和背景场景的表观颜色相近时, 检测准确度会大大降低。对于场景亮度条件的突变而引起的前景噪声, 即使采用模型更新机制, 也不能有效及时的去除。针对这些不足, 提出一种基于颜色、边缘和纹理视觉特征的混合高斯建模技术。新的建模特征能够很好的描述背景区域的本质, 对前景目标有着非常好的区分力, 并且采用准确率和召回率对实验结果进行定量分析。实验分析表明, 新算法有效地解决了传统算法存在的问题。同时也为后继的高层视觉分析任务打下了良好的基础。

关键词: 多种类视觉特征; 混合高斯; 离散余弦变换; 准确率; 召回率

Mixture of Gaussian background modeling method based on multi-category visual features

Chang Xiaofu, Zhang Wensheng, Dong Weishan

(Key Laboratory of Complex System and Intelligent Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 China)

Abstract: The mixture of Gaussian method (GMM) proposed by Stauffer et al and its improved versions achieve better performance and have been widely used in smart vision systems. Traditionally, GMM only uses a single color visual feature to model scenes, but it will fail when moving objects have the same color appearance with scenes' color appearance. In addition, traditional GMM can't effectively solve problems caused by lighting changes and especially lighting sudden changes. To overcome these weaknesses, in this paper, a new modeling method based on multi-category visual features (i. e. color, texture and edge) is proposed. Experiment results show that our new method solves the above problems effectively, which will be a baseline for high-level vision analysis.

Keywords: multi-category visual features; DCT; mixture of Gaussian model; precision ratio; recall ratio

0 引言

如何精确且实时地检测场景中的运动目标是众多智能视觉系统, 如安全监控^[1]、交通流量监控^[2]、手势识别^[3]等系统中的一个非常基础和关键的技术环节。

然而在实际的应用环境中, 背景建模技术经常面临很多技术难题。其中之一来自自然环境因素的挑战, 例如, 太阳光照的改变, 场景中树枝的摆动, 降雨等。另外, 来自非自然因素的挑战, 例如, 被监控的房间内, 由于开关灯造成的屋内光亮条件的剧烈变化; 场景中包裹之类的遗留物^[4]或者滞留的

收稿日期: 2010-02-05; 修回日期: 2010-05-31

基金项目: 国家高技术研究发展(863)计划项目(2008AA01Z121, 2007AA01Z338); 国家自然科学基金项目(909Z4026)。

第一作者简介: 常晓夫(1984—), 男。现为中国科学院自动化研究所图像处理、模式识别专业硕士研究生, 主要研究方向为视频监控、物体识别、视频检索。E-mail: changxiaofu123@163.com。

行人和车辆;闪动的电脑屏幕等。因此,为了精确地检测出场景中的运动目标,理想的背景减除建模技术应该能够较好地解决上述所有的挑战难题。同时算法的实时处理性能也要给予考虑。

为了解决上述挑战难题,学者们提出了很多种背景减除建模技术。早期的背景减除建模技术有均值滤波^[5]和中值滤波^[6]。其主要思想是将一个时间窗内的历史像素亮度值取平均或取中值构建背景模型。此外,Mason 等人^[7]提出了基于图像块的颜色分布直方图或边缘方向分布直方图构造背景模型。由于传统的直方图技术缺乏空间信息,Philippe 等人提出了带有空间信息的核颜色直方图^[8]来克服这一技术缺陷。文献[9]中的 Pfinder 系统较早地采用基于统计学的单高斯建模技术来描述场景中亮度特征随时间变化的分布情形。文献[10]中的 W4 系统则采用根据历史像素亮度的最大值和最小值以及相邻帧的像素亮度差值构建双峰背景模型。不过实际复杂场景中的亮度信息受自然条件等不确定因素影响经常呈现多模态分布。为了解决这一问题,能够刻画背景场景多模态分布的非参数密度估计模型^[11]以及混合高斯模型^[12]分别由 Elgammal 等人^[11]和 Stauffer 等人^[12]提出。前者计算复杂度较高并且需要耗费大量的内存存取历史数据;后者计算简单,模型描述较为精准,在实际中获得了较好的检测效果和实时效果。文献[13]是 Zivkovic 等人提出的改进版本,改进版本侧重于对 GMM 算法的描述规则加以改进来提高检测精度和速度。

虽然 Stauffer 等人提出的混合高斯背景建模方法较好地解决了背景场景的多模态分布情形,但是由于其采用基于灰度空间的颜色特征来构建背景场景模型,因此当场景中的入侵目标的表现颜色与背景场景表现颜色相近时,会导致这一方法不能精确完整地检测出运动目标。另外,当场景中的光照条件改变时,也会导致虚警前景区域的产生。针对这些不足,本文采用了基于离散余弦变换后的系数组合得到的多种类视觉特征(即颜色、纹理、边缘)来构造新的建模向量。新的建模视觉特征向量力求更加真实的反映所构建背景场景的本质。同时采用 Stauffer 等人提出的混合高斯技术对新的特征向量随时间变化的多模态分布情形加以刻画描述。最后,3 组不同场景测试数据的定量实验分析结果表明,新算法比传统的混合高斯方法获得了更好的检测效果,这为后继的高层视觉任务提供了较好的分

析基础。

1 新的背景模型

离散余弦变换是与傅里叶相关的一种变换,由于它的正交变换特性以及对变换后的信号具有在低频分量聚集且保持原能量大小的特性,使其成为当前图像、视频压缩技术的基础,例如, JPEG, MPEG, H. 2621 等。此外,这一技术也促进了图像、视频检索领域的发展。本文提出的新的建模特征受启发于一些学者在图像检索领域的研究工作^[14-16]。本文算法的大致流程如下:首先,将采集到的图像变换到亮度空间,再以 8×8 建模尺度大小的方格对其进行划分;其次,对变换后的所有 8×8 尺度大小的图像块进行离散余弦变换,基于这些变换后的频率域系数组合提取 4 维特征向量;最后,采用混合高斯技术对提出的 4 维多种类视觉特征向量进行描述。具体的细节可见下面部分。

1.1 8×8 离散余弦变换(DCT)

8×8 尺度大小的图像的 2 维离散余弦变换可以用下式表示

$$F(u, v) = \frac{1}{4} \alpha(u) \alpha(v) \left[\sum_{i=0}^7 \sum_{j=0}^7 f(i, j) \times \cos\left(\frac{(2i+1)u\pi}{16}\right) \times \cos\left(\frac{(2j+1)v\pi}{16}\right) \right] \quad (1)$$

式中, $f(i, j)$ ($i, j = 0, 1, \dots, 7$) 表示原变换尺度中图像的亮度像素值, $F(u, v)$ ($u, v = 0, 1, \dots, 7$) 是变换后的系数。当 $w = 0$ 时, $a(w) = 1/\sqrt{2}$ 。其余情况, $a(w) = 1$ 。 $F(0, 0)$ 称为 DC 系数,即直流分量。其余的 63 个系数称为 AC 系数,即交流分量。DC 系数代表被变换图像的平均亮度,AC 系数和被变换图像内容的细节相关。

1.2 多种类视觉特征向量

除了上面介绍的特性,文献[16]证明了变换后的系数具有相应的物理意义。以变换后的系数 $F(0, 1)$ 为例

$$F(0, 1) = \frac{a(0)a(1)}{4} \left\{ \cos\frac{\pi}{16} \left[\sum_{i=0}^7 f(i, 0) - \sum_{i=0}^7 f(i, 7) \right] + \cos\frac{3\pi}{16} \left[\sum_{i=0}^7 f(i, 1) - \sum_{i=0}^7 f(i, 6) \right] \right\}$$

$$\left. \begin{aligned} & \sum_{i=0}^7 f(i,6) \Big] + \cos \frac{5\pi}{16} \left[\sum_{i=0}^7 f(i,2) - \right. \\ & \left. \sum_{i=0}^7 f(i,5) \Big] + \cos \frac{7\pi}{16} \left[\sum_{i=0}^7 f(i,3) - \right. \\ & \left. \sum_{i=0}^7 f(i,4) \Big] \right\} \quad (2) \end{aligned}$$

从上式可以看到,变换后 $F(0,1)$ 的值等于 8×8 图像块水平方向左右两部分的亮度差分。这个水平方向的差分可以近似等价于水平方向的梯度或垂直方向边缘的近似。其余 63 个系数的物理意义表示可参见图 1。

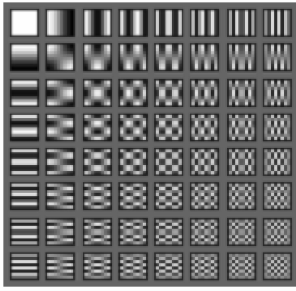


图 1 64 个系数的物理意义表示

Fig. 1 Physical representations of 64 coefficients

从图 1(原图来自于文献[14])可以看到,变换后的系数具有丰富的视觉信息。新的建模特征向量的第 1 维是 DC 系数,即图 2 中左上角最亮的格子。该系数代表了该区域平均亮度信息,因此第 1 维特征分量体现了颜色视觉信息。接下来的两个特征分量是和图像的边缘特征相关的。它们分别是水平边缘特征 f_2 和垂直边缘特征 f_3 。如图 2 所示,特征 f_2 等于图中蓝色线覆盖的系数之和; f_3 等于图中红色线覆盖的系数之和;最后 1 维的特征分量 f_4 是和纹理相关的视觉特征,它等于图 2 中绿色线覆盖的系数之和。由于聚集能量的特性,右下角部分的系数非常小,甚至为 0,因此这些系数被忽略,没有被采

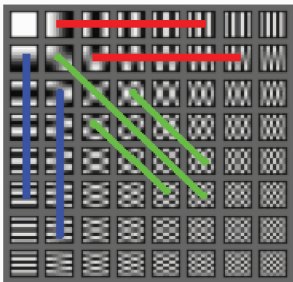


图 2 用于构建特征向量的系数

Fig. 2 Coefficients used for modeling features

用。最后,用于建模的新特征向量表示为 $F = \{f_1, f_2, f_3, f_4\}$ 。

1.3 基于多种类视觉特征的混合高斯背景建模

与 Stauffer 等人提出的混合高斯建模技术不同,本文的建模尺度是基于 8×8 大小的一个图像块而不是单个像素点,并且建模的特征向量不是由基于单个像素点的 RGB 空间的 3 个颜色分量构成,而是按照 2.2 节中的方法提取多种类视觉特征向量。则新的基于图像块的多种类视觉特征的 K 元混合高斯方法,可以用下式表示

$$P(F_t) = \sum_{i=1}^K W_{t,i} \times \eta(F_t, \mu_{t,i}, \Sigma_{t,i})$$

$$\eta(F_t, \mu_{t,i}, \Sigma_{t,i}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \times \exp\left[-\frac{1}{2} (F_t - \mu_{t,i})^T \Sigma^{-1} (F_t - \mu_{t,i})\right] \quad (3)$$

式(3)表示当前图像块属于背景的概率。式中的 F_t 代表 t 时刻在 8×8 大小的图像块上提取的由亮度、水平边缘、垂直边缘以及纹理视觉特征构成的 4 维的特征向量。 $\eta(F_t, \mu_{t,i}, \Sigma_{t,i})$ 代表第 i 个高斯概率密度函数。 $W_{t,i}$ 为其相应高斯密度函数的权重,亦等价于每个高斯分布 η 作为背景的先验概率,且有 $\sum_{i=1}^K W_{t,i} = 1$ 。 $\mu_{t,i}$ 是 t 时刻第 i 个高斯均值向量, $\Sigma_{t,i}$ 是 t 时刻第 i 个高斯协方差矩阵。根据 DCT 的正交特性可知,变换后的 64 个系数(以随机变量的形式考虑)相互独立。根据概率论的知识,容易得到变换后的上述 4 维特征分量彼此相互独立。而这与传统的 GMM 算法不同,传统的 GMM 算法为了方便计算协方差矩阵的逆矩阵,假设 RGB 构成的 3 维颜色特征分量彼此相互独立。不过这是以牺牲模型的准确性为代价的。假设特征向量 F 的每一维有相同的方差 σ^2 ,则上面的协方差矩阵可以表示为

$$\Sigma = \sigma^2 I \quad (4)$$

I 是单位矩阵。实际上判断每一个图像块的 F_t 是否属于背景模型,可以用 $|F_t - \mu_{t-1}| < 2.5\sigma$ 来判断。为了适应场景的变化,需要对模型的参数(如均值、方差、权重)进行自适应更新。权值更新如下

$$w_{i,t} = (1 - \alpha)w_{i,t-1} + \alpha(M_{i,t}) \quad (5)$$

匹配时 $M_{i,t} = 1$,不匹配时 $M_{i,t} = 0$ 。采用下式对匹配的高斯分布的均值和方差进行更新,不匹配的保持不变。

$$\mu_{j,t} = (1 - \beta)\mu_{j,t-1} + \beta F_t$$

$$\sigma_{j,t}^2 = (1 - \beta)\sigma_{j,t-1}^2 + \beta(\mathbf{F}_t - \boldsymbol{\mu}_{j,t})^T(\mathbf{F}_t - \boldsymbol{\mu}_{j,t})$$

$$\beta = \alpha\eta(\mathbf{F}_t, \boldsymbol{\mu}_{t,i}, \boldsymbol{\Sigma}_{t,i}) \quad (6)$$

式(5)(6)中 α 和 β 为学习率。

2 实验结果和分析

本文实验平台配置为酷睿双核(2.10 GHz)处理器,1 GB 内存。处理图像的分辨率大小为 320×240 ,平均处理速度达到 18 帧/s。新方法与经典的基于 RGB 色彩空间的混合高斯方法在 3 组不同的测试序列上进行了对比实验。二者对比的实验结果可参见图 3—5。每个图的第 1 列为原始的测试序列,第 2 列为手工标注的理想测试结果,第 3 列为传统的基于颜色特征的 GMM 算法的结果,第 4 列为

本文所提方法的实验结果。除了第 4 列以灰度化显示检测到的前景运动目标以外,其余都以白色表示前景运动区域,背景则以黑色表示。实验中,两种方法的初始化方差设为 30,学习率设为 0.005。

第 1 组测试序列(图 3)选自于文献[8]。在这组测试序列中,初始时房间无人且屋内亮度较低。随后有人进入,将灯打开,此时场景亮度发生了骤变。从图中可以看到,传统的基于颜色的 GMM 技术面对场景亮度的突变情形,检测完全失效,误将场景中本属于背景的大部分区域当为前景运动区域。而本文的基于多种类视觉信息的建模方法,则较为完整的将场景中的运动目标区域检测出来。其原因在于,本文的建模特征向量不仅包含颜色信息,而且还包含纹理、边缘两个重要的视觉特征分量。而这

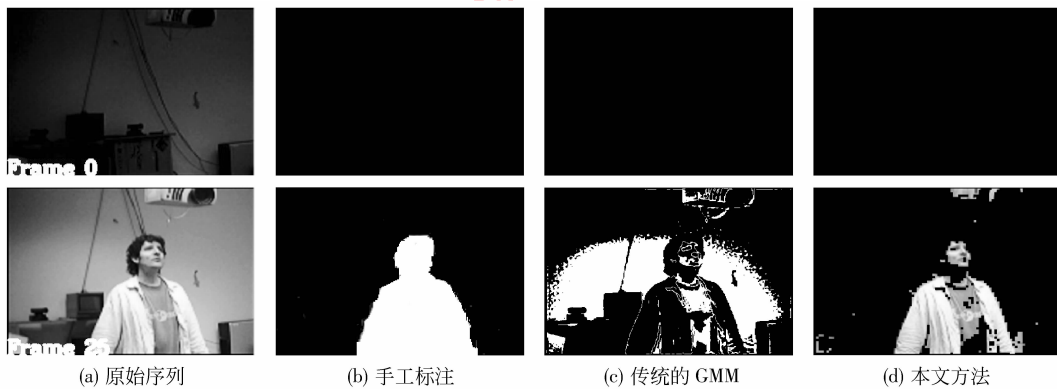


图 3 场景亮度突变测试

Fig. 3 Lighting sudden changes in scenes

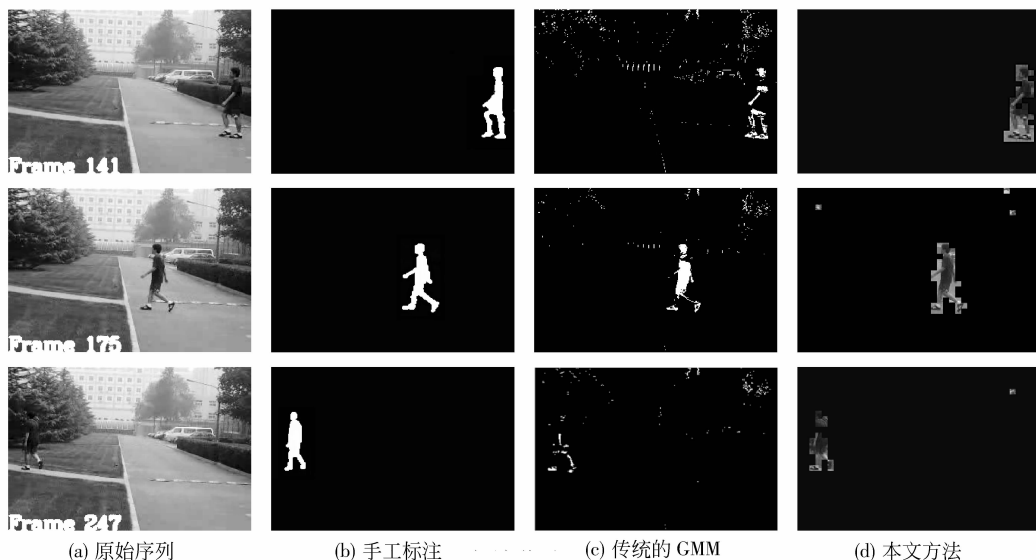


图 4 穿越草地

Fig. 4 Cross grass

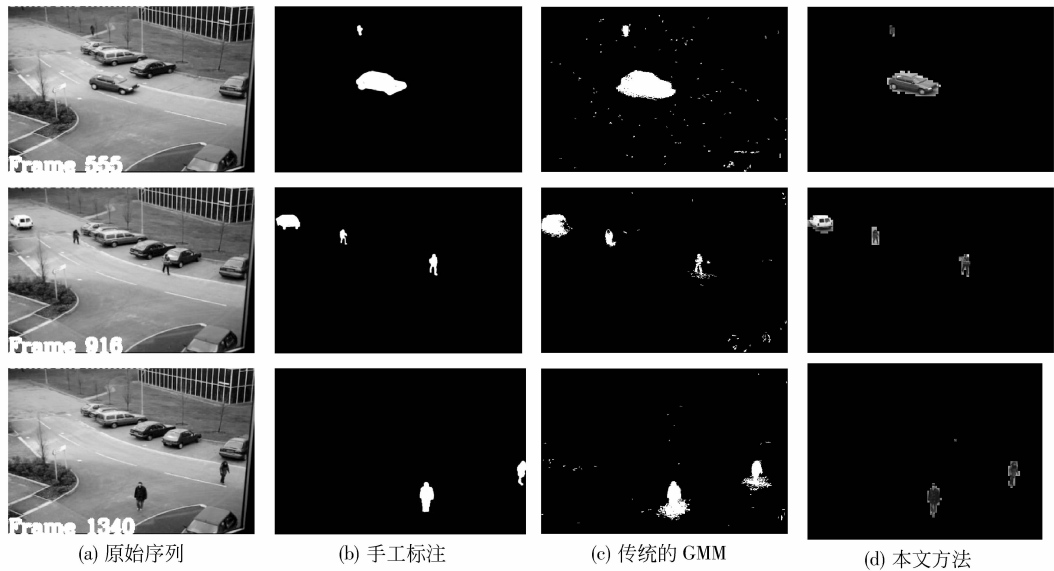


图 5 PETS2000 数据集

Fig. 5 PETS2000 datasets

两维视觉信息,在运动目标没有覆盖的区域,几乎不受光亮度变化的影响。因此,新的建模方法能够对 GMM 所误判的前景区域起到较好的抑制作用,而对真正的前景运动目标区域则起到较好的区分作用。在第 2 组题为穿越草地的测试序列中,由于行人上半身的衣着颜色和背景场景的树丛颜色十分接近,因此,在第 141 和 247 两帧中,传统的 GMM 算法只能将运动目标区域与背景颜色不一致的部分检测出来,而本文的方法则较完整的将运动目标检测出来。原因在于,尽管运动目标衣着的颜色和其身后场景颜色相近,但是运动目标视觉上的纹理和边缘样式与其身后场景的纹理和边缘样式大不相同,因此新方法能够正确地将颜色相近的运动目标检测出来。最后一组测试序列选自 PETS2000 测试数据集。从图 5 可见,尽管两种方法有效地将前景运动目标检测出来,但是相比于本文方法,传统的 GMM 算法误检测出较多的前景像素点。原因在于,基于图像块区域的新的建模方法,充分利用了图像块临域的空间信息,因此对于零散的虚警像素点起到了一定的抑制作用。本文同时也采用准确率 (PR) 和查全率 (RR) 对两种算法在上述 3 组不同场景的测试序列的实验结果进行了定量的对比分析。

$$PR = \frac{\#TruePositivePixelsDetected}{\#TotalPositivePixelsDetected} \quad (7)$$

$$RR = \frac{\#TruePositivePixelsDetected}{\#TotalTruePositivePixels} \quad (8)$$

#代表个数。PR 和 RR 的分析结果如图 6 和图 7 所示。

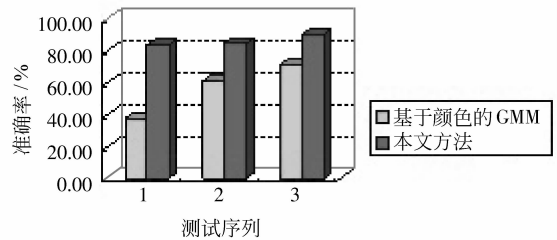


图 6 准确率对比结果

Fig. 6 Comparative results on precision

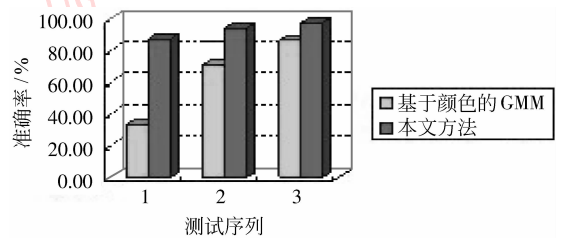


图 7 查全率对比结果

Fig. 7 Comparative results on recall

3 结 论

提出基于多种类视觉特征的混合高斯建模技术。新的建模特征通过对图像进行离散余弦变换后的系数组合得到。新的特征向量不仅包含颜色视觉

信息,而且还包含水平边缘、垂直边缘和纹理视觉信息。这克服了传统的 Stauffer 等人提出的仅利用颜色信息不能准确刻画场景模型的缺点。3 组不同场景下的测试实验结果表明,新的基于多种类视觉特征的混合高斯建模方法的正确性,且有效地解决了传统的 GMM 方法面对场景亮度突变以及运动目标和场景颜色相近时检测失效的难题。

参考文献 (References)

- [1] Hampapur A, Brown L, Connell J, et al. Smart video surveillance [J]. IEEE Signal Processing Magazine, 2005, 22(2):38-51.
- [2] Kamijo S, Matsushita Y. Traffic monitoring and accident detection at intersections[J]. IEEE Transactions on Intelligent Transportation Systems, 2000,1(2):108-118.
- [3] Zhan Fengyue, Chellappa R. Synthesis of silhouettes and visual hull reconstruction for articulated humans[J]. IEEE Transactions on Multimedia, 2008,10(8):1565-1577.
- [4] Yigit A, Temizel A. Abandoned object detection using thermal and visible band image fusion[C]//Proceedings of the IEEE 18th Conference on Signal Processing and Communications Applications. New York, America; IEEE, 2010:617-620.
- [5] Friedman N, Russell S. Image segmentation in video sequences: probabilistic approach [C]//Proceedings of Thirteenth Conference on Uncertainty in Artificial Intelligence. San Francisco, America; Morgan Kaufmann, 1997:175-182.
- [6] Cutler R, Davis L. View-based detection and analysis of periodic motion[C]//Proceedings of International Conference on Pattern Recognition. New York, America; IEEE, 1998: 495-500.
- [7] Mason M, Duric Z. Using histograms to detect and track objects in color video [C]//Proceedings of the Applied Imagery Pattern Recognition Workshop, New York, America: IEEE, 2001:154-159.
- [8] Philippe N, Benedict B, Olivier B. Local kernel color histograms for background subtraction [C]//Proceedings of International Conference on Computer Vision Theory and Applications. America: INSTICC, 2006:213-219.
- [9] Wren C, Azarbayejani A, Darrell T, et al. Pfister: real-time tracking of the human body [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7):780-785.
- [10] Haritaoglu I, Harwood D, Davis L S. W4: real-time surveillance of people and their activities [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8):809-830.
- [11] Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction[J]. Lecture Notes in Computer Science, 2000, 1843(1): 751-767.
- [12] Stauffer C, Grimson W. Learning pattern of activity using real-time tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8):747-757.
- [13] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction [C]//Proceedings of the 17th International Conference on Pattern Recognition. New York, America; IEEE, 2004:28-31.
- [14] Tsai T, Huang Y L, Chiang T. Image retrieval based on dominant texture features [C]//Proceedings of the IEEE International Symposium on Industrial Electronics. New York, America; IEEE, 2006:441-446.
- [15] Huang Y L, Chang R L. Texture features for DCT-coded image retrieval and classification [C]//Proceedings of the IEEE International Conference on the Acoustics, Speech, and Signal Processing. New York, America; IEEE, 1999:3013-3016.
- [16] Shen B, Sethi I K. Direct feature extraction from compressed images[C]//Proceedings of the Storage and Retrieval for Image and video Databases IV. New York, America; SPIE, 1996: 404-414.