

基于时空注意模型的视频分割算法

郑河荣¹⁾ 褚一平²⁾ 潘翔¹⁾

¹⁾(浙江工业大学计算机学院, 杭州 310014) ²⁾(杭州电子科技大学计算机学院, 杭州 310018)

摘要 针对已有视频分割算法对复杂动态背景下所出现的误分割问题,提出通过显著性映射构造时空注意特征,并采用分层条件随机场进行视频分割,提高分割准确率。算法首先根据视觉注意理论提取时域和空域特征,并建立加权混合模型。其次,采用该混合模型计算运动目标的显著性映射概率分布,有效地提取出运动目标区域。最后,在显著性映射概率分布基础上,采用高斯混合模型建立前景和背景的能量函数,构造分层条件随机场模型对这些特征能量函数进行分割建模,精确地提取出运动对象目标。实验结果表明,该算法即使对复杂动态背景下的视频也能够得到稳定的分割效果,有效地去除摄像机运动等所导致的误分割问题。

关键词 视频分割 时空信息注意模型 分层条件随机场

中图法分类号: TP391.6 文献标志码: A 文章编号: 1006-8961(2010)05-729-07

Video Segmentation Based on Spatial-temporal Attention Model

ZHENG Herong¹⁾, CHU Yiping²⁾, PAN Xiang¹⁾

¹⁾(College of Computer, Zhejiang University of Technology, Hangzhou 310014)

²⁾(College of Computer, Hangzhou Dianzi University, Hangzhou 310018)

Abstract To deal with the error segmentation problem of the existing video algorithms under complex and dynamic scenes, the proposed method extracts spatial-temporal attention features with salient maps, and adopts hierarchical conditional random field for video segmentation. Firstly, the algorithm constructs a weighted combination model based on spatial-temporal features by using information theory. Then, it uses the defined model to compute probability distribution of salient maps, which can locate region of moving object effectively. Finally, the Gaussian mixture model is adopted to construct energy functions with the above probability distribution, and the hierarchical conditional random field is used to constraint these feature energy functions to refine final segmentation. The experiment results showed that the algorithm can avoid the error segmentation problem induced by camera movement. So it is robust to handle the videos under complex and dynamic scenes.

Keywords video segmentation, spatial-temporal attention model, hierarchical conditional random field

0 引言

随着数字媒体技术的飞速发展,视频流技术已经被广泛应用于各种行业。从海底探测到航空航天、智能交通到体育竞技、数字娱乐到楼宇监控,数字视频遍及各个社会领域,已经成为社会生活不可或缺的一部分。因此,如何对这些视频进行分析和

理解成为一个当前迫切需要解决的问题。而视频分割研究从视频序列中准确地分割出运动目标,是保证后续视频分析和理解正确性的关键,也是目前国内学术界的研究热点。

近年来,研究人员为了能够提高分割算法的准确性,特别是针对复杂背景下的运动目标分割问题,提出了一系列改进算法。这些改进主要从以下3方面考虑:1)在像素级分割基础上,通过概率假设检验或

基金项目:国家自然科学基金项目(60703001);浙江省科技厅重大项目(2009C11G2020027);浙江省教育厅项目(Y200805048)

收稿日期:2009-10-28; **改回日期:**2010-01-25

第一作者简介:郑河荣(1971—),男,副教授。2009年于浙江工业大学机电学院获获化机专业智能识别方向博士学位。主要研究方向为图像处理及智能识别技术。E-mail: hailiang@zjut.edu.cn

通讯作者:潘翔。E-mail: panx@zjut.edu.cn

是状态机等理论减少像素误分割。统计推断^[1]是在高斯混合模型^[2]的基础上,提出采用对后续帧进行假设检验,进一步提高视频分割质量。分割建模可以转化为包含状态、决策、状态转移规律、代价函数和目标函数 5 个部分的动态规划问题^[3]。非参数估计方法采用高斯核函数估计视频序列上像素的概率密度^[4]。王林波等人提出一种利用贝叶斯学习方法来进行视频分割的算法,通过对每个像素点不断变化的背景建模,用贝叶斯学习方法判断在光线缓慢或者突然变化的时候,每个像素点是属于前景还是背景^[5]。高丽等人针对复杂背景设计了一种视频对象自动分割新算法,它利用随机信号的统计特性累计获得背景,并利用背景累积过程中分类讨论的观点,有效地将运动对象从视频序列中自动地分割出来^[6]。Carmona 等人通过颜色等信息构造多尺度 Blob 特征,用以提高分割准确率^[7]。刘少华等人提出根据连续 3 帧视频序列构建参考帧,并根据它们的差分图像构造像素块最大后验概率,用以提高分割准确率^[8]。2) 根据视频运动目标具有的时空相邻性,通过马尔可夫模型等提高分割准确率。Stenger 等人提出一个基于拓扑结构可变的隐马尔可夫模型的背景建模方法,它可以适应光线突然变化等情况^[9]。陈睿等人对视频分割邻域相关信息的约束建模,用以提高视频分割效果^[10]。Yaser 等人利用马尔可夫随机场模型对同一视频帧像素之间的邻域关系进行建模,通过求解马尔可夫随机场模型实现前景背景的分割^[11]。Migdal 等人利用 2 维马尔可夫随机场模型对像素之间的时空信息进行约束,通过求解最大后验概率实现对视频前景目标的分割^[12]。Wang 等人把条件随机场模型应用于视频目标分割中,该算法通过条件随机场模型对视频序列的时空邻域信息进行建模,构造了一个可以根据前一帧视频图像的条件随机场模型数据对当前帧视频图像的条件随机场模型进行迭代更新的滤波器^[13]。褚一平等人提出基于隐条件随机场模型的视频目标分割算法,通过使用隐条件随机场可以更直观地对视频序列的多种上下文特征进行建模^[14]。3) 采用运动目标的形状特征定义先验知识,建立基于识别的视频分割模型,对分割对象做前期学习训练提高分割结果。基于识别的视频分割技术由于考虑了特定目标的全局形状信息,因此对强光变化、焦距变化、摄像机位置变化等复杂动态背景也能得到比较理想的分割效果。ISM 模型^[15]是一种典型高级层次分割算法,该算法通过概

率的方法融合目标类型的识别信息,进而实现特定对象的自动分割,该算法主要应用于行人的分割和检测。针对场景中可能出现不同类型的物体,Mikolajczyk 等人提出了一种基于目标识别的多目标分割和检测方法^[16],并采用树状结构组织特征字典来加快分割过程中特征向量搜索的速度。Chum 等人在视频分割中融入目标样例信息,然后结合图切分算法实现视频分割^[17]。Yin 等人提出了一种对视频聊天类型的人物目标进行分割的有效方法^[18]。该方法针对视频聊天的特殊性,采用 Motons 作为视频分割的特征,通过对几段同类型的聊天视频来训练,得到视频运动特征,通过随机森林方法对这种特征进行学习,然后使用条件随机场模型对运动及颜色等多种特征进行建模来获得最终的分割结果,该方法对于背景非常复杂的视频,如光线强变化等,也能够得到理想的分割效果。褚一平等人提出采用随机森林的联合共享规则构造高精度分类器,建立双层识别分割模型^[19]。

尽管基于识别的分割能够有效地提高分割准确率,但是其需要大量的特定样本进行学习训练,而且只能针对包含特定目标的视频进行分割,从而限制了算法的适用性和普遍性。而其他两种改进方法,对于一些光线变化或是相机抖动等虽然具有比较好的稳定性。但是这两种改进方法需要假设相机相对固定,或者只有轻微晃动,而无法处理相机随运动物体移动所造成的复杂背景情况。因此,本文在已有研究基础上,提出采用视觉心理学中的视域注意理论,构造能够区分运动目标和背景的时空注意模型,用以有效地定位运动目标区域。然后通过分层条件随机场准确地提取出运动目标。这种分割方法,由于视域注意理论所定义的显著性映射在相机运动等情况下具有较好的稳定性,可以有效地去除摄像机运动等所导致的误分割问题,提高分割准确率。

1 时空信息论注意模型

视频分割的主要目标是如何有效地提取出运动目标。而在视觉心理学理论中,运动物体构成了显著性目标。在视频序列中,人眼往往会忽视一直出现的物体,如背景中的物体,而对偶尔(忽然)出现的物体往往会特别地关注,如运动目标。因此,本文根据视觉心理学的激励机制,采用视频时域运动特征和空域颜色特征等构造混合模型,计算用以区分

运动目标和背景的显著性映射概率分布。Liu 等人假设视频拍摄时摄像机是相对固定的情况下对视频序列的时空信息进行计算^[20]。而在实际的应用中,摄像机可能会出现不固定的情况。为了处理这个问题,在计算像素信息量之前先进行一致性匹配,利用单应性矩阵来构成视频像素块体。

视频序列的信息量由时域信息量和空域信息量两部分混合而成。假设视频帧大小为 $M \times N$,它可以划分成大小为 $m \times n$ 的像素块,连续 s 帧像素块可以组成 $m \times n \times s$ 的像素块体。第 t 帧上第 (i, j) 的像素块由 $B(i, j, t)$ 表示,它对应的像素体 $V(i, j, t)$ 则由前 $(t - W, t]$ 帧的对应位置上的像素块一起构成。

通过估计该像素块在邻域像素块中的概率计算得到空域上的像素块信息量。空域像素块信息量计算如下:

$$I_s(i, j, t) = -\log_2 P(B(i, j, t) | B_N(i, j, t)) \quad (1)$$

式中, $B_N(i, j, t)$ 表示像素块 $B(i, j, t)$ 对应的邻域像素块。计算式(1)的关键在于正确估计条件概率,采用类似于文献[11]中非参数化估计的方法计算:

$$P(B(i, j, t) | B_N(i, j, t)) = W^{-1} \sum_{l=0}^W K_H(B(i, j, t) - B_l(i, j, t)) \quad (2)$$

式中, $B_N(i, j, t) = \{B_0(i, j, t), B_1(i, j, t), \dots, B_W(i, j, t)\}$, $K_H(\cdot)$ 为核函数。由于 $B_N(i, j, t)$ 和 $B(i, j, t)$ 都为高维向量,为了加快计算,可以先进行 PCA 降维再进行非参数化估计。

通过估计该像素块在对应像素块体中的概率计算得到时域上的像素块信息量,具体公式如下:

$$I_t(i, j, t) = -\log_2 P(B(i, j, t) | V(i, j, t)) \quad (3)$$

计算式(3)的关键在于估计条件概率

$$P(B(i, j, t) | V(i, j, t)) = W^{-1} \sum_{l=0}^W K_H(B(i, j, t) - B'(i, j, t - l)) \quad (4)$$

式中, $V(i, j, t) = \{B'(i, j, t), B'(i, j, t - 1), \dots, B'(i, j, t - W)\}$ 。

由于需分割的视频中可能存在摄像机移动的情况,所以在构造 $V(i, j, t)$ 时,不是直接使用前 W 帧视频进行构造,而是使用单应性矩阵矫正之后的视频数据。通过对视频帧提取不变特征,并根据这些特征点计算单应性矩阵,再利用单应性矩阵计算得到各帧的像素块 $B'(i, j, t - l)$, $l = 0, \dots, W$ 。在理想情况下, $B'(i, j, t - l) = B(i, j, t - l)$ 。

时空注意特征采用如下的加权联合模型:

$$I(i, j, t) = k_t \times I_t(i, j, t) + k_s \times I_s(i, j, t) \quad (5)$$

k_t 和 k_s 分别为时域和空域信息量的权重,考虑到算法以分割运动目标为主,所以一般情况下时域权重相对取得比较大。根据时空像素块信息量,通过设定一个阈值,可以方便地得到比较粗糙的分割结果。

2 分层条件随机场建模

虽然通过时空注意特征能够得到一个比较粗糙的分割结果,但是这种显著性映射概率分布还无法构成一个精确的分割结果。因此,需要在显著性概率分布模型基础上,再做进一步地分割,通过视频帧间的相邻信息得到精确的分割结果。在这里采用分层条件随机场建模。对于分层条件随机场建模,首先以显著性概率分布模型为输入,采用高斯混合模型分别对前景和背景进行建模。然后采用该背景和前景模型为条件随机场的输入,设计特征函数,根据邻域相关性得到更为精确的分割结果。

2.1 分层条件随机场

不同于条件随机场,分层条件随机场在标签层和观察层之间定义了一个隐状态层。隐状态层由超像素定义。其中超像素计算采用多级图划分方法。超像素之内的各个像素之间的邻域形成了局部邻域关系,超像素与超像素之间形成了全局邻域关系。通过超像素描述,可以更为有效地利用邻域特征,从而能够更好地避免分割所形成的虫洞等问题。对于分层条件随机场,给定视频序列中的一帧 X ,分层条件随机场模型对基于视频帧 X 的分割标签 L 的条件概率建模:

$$P(L | X) = \sum_H P(L, H | X) = \frac{1}{Z(X)} \sum_H \exp(-E(L, H, X)) \quad (6)$$

式中, $Z(X)$ 为分配函数,它可以保证分层条件随机场模型有一个归一化的概率。它由公式 $Z(X) = \sum_{H, L} \exp(-E(L, H, X))$ 计算,其中 $E(L, H, X)$ 为能量函数,定义为

$$E(L, H, X) = \sum_k \lambda_k f_1(H_k, X_k) + \nu \sum_i \sum_{j \in N_i \cup M_i} f_2(H_i, L_j, H_j) + \mu \sum_i \sum_{j \in N_i \cup M_i} f_3(L_i, L_j, H_j) \quad (7)$$

式中, λ_k , ν 和 μ 为分层条件随机场模型的参数。特

征函数 $f_1(H_k, X_k)$ 表示依赖单个隐状态变量的强度。而特征函数 $f_2(H_i, L_j, H_j)$ 表示依赖于 2 个随机变量的强度函数,它要么依赖于 2 个隐状态变量 H_i 和 H_j ,要么依赖隐状态变量 H_i 和真实标签变量 H_j 的组合。特征函数 $f_3(L_i, L_j, H_j)$ 也表示依赖于 2 个随机变量的强度函数,它要么依赖于 2 个真实标签变量 L_i 和 L_j ,要么依赖真实标签变量 L_i 和隐状态变量 H_j 的组合。因此,对给定视频帧 X 和模型参数 λ_k, ν 和 μ ,可以计算模型的真实标签层:

$$\hat{L} = \arg \max_l P(L | X; \lambda_k, \nu, \mu) \quad (8)$$

使用复制函数 $\Gamma(g)$ 来实现对视频帧 X 的标签,复制函数 $\Gamma(g)$ 用来实现分层条件随机场模型中真实标签层中的随机变量与视频帧 X 进行映射关系。

2.2 特征函数设计

显然,对于式(7)分层条件随机场模型,在采用最小割 (Min-cut) 算法^[21] 进行分割结果求解时,需要定义相应的特征函数控制迭代收敛。

对于上述能量函数的第 1 部分,主要用于描述单个像素的概率分布,定义为

$$f_1(H_k, X_k) = \delta(H_k, H_{k,m}) \quad (9)$$

式中, $\delta(g)$ 为 Kronecker delta 函数。 $H_{k,m}$ 表示标签。 λ_k 被定义为

$$\lambda_k = \begin{cases} \log P(X_k | H_k = F) & \text{若 } H_k = F \\ \log P(X_k | H_k = B) & \text{其他} \end{cases} \quad (10)$$

式中, $P(X_k | H_k = F)$ 和 $P(X_k | H_k = B)$ 代表高斯混合模型得到的概率分布。

对于能量函数式(7)的第 2 部分,用于描述与隐状态变量相关的边能量函数的加权和,边能量函数由权重 ν 进行加权。边能量函数依赖于模型中的 2 个随机变量,它要么依赖于 2 个隐状态变量 H_i 和 H_j ,要么依赖隐状态变量 H_i 和真实标签变量 L_j 的组合。因此,该部分可以被拆分成如下 2 个部分:

$$\sum_i \sum_{j \in N_i \cup M_i} f_2(H_i, L_j, H_j) = \sum_i \sum_{j \in N_i} f_2(H_i, H_j) + \sum_i \sum_{j \in M_i} f(H_i, L_j) \quad (11)$$

而式(11)中的局部边特征函数 $f_2(H_i, H_j)$ 依赖于 2 个隐状态变量,它定义为

$$f_2(H_i, H_j) = \begin{cases} \beta_1 & \text{若 } H_i = H_j = B \\ \beta_2 & \text{若 } H_i = H_j = F_k \\ \beta_3 & \text{若 } H_i \neq H_j \end{cases} \quad (12)$$

式中,参数 β_1, β_2 和 β_3 为常量,表示隐状态层中局部邻域随机变量之间的强度。而局部边特征函数

$f(L_i, H_j)$ 依赖于一个隐状态变量和一个真实标签变量的组合,定义如下:

$$f(L_i, H_j) = \begin{cases} \phi_1 & \text{若 } L_i = H_j = B \\ \phi_2 & \text{若 } L_i = H_j = F_k \\ \phi_3 & \text{若 } L_i \neq H_j \end{cases} \quad (13)$$

式中,参数 ϕ_1, ϕ_2 和 ϕ_3 也为常量,表示隐状态层和真实标签层之间局部邻域随机变量之间的强度。由于分层条件随机场模型也具有马尔可夫属性,边特征函数 $f(L_i, H_j)$ 与 $f(H_j, L_i)$ 是相等的。

能量函数式(7)的第 3 部分,主要用于描述与真实标签变量相关的边能量函数的加权和,边能量函数由权重 μ 进行加权。边能量函数也表示依赖于 2 个随机变量的强度函数,它要么依赖于 2 个真实标签变量 L_i 和 L_j ,要么依赖真实标签变量 L_i 和隐状态变量 H_j 的组合。其中,模型参数 μ 用于控制真实标签层能量对整个能量的影响程度,用于用户交互设置来提高分割效果。本文主要研究注意模型对分割结果的影响,因此统一设该值为 1。类似地,该部分可以被拆分成 2 个部分:

$$\sum_i \sum_{j \in N_i \cup M_i} f_3(L_i, L_j, H_j) = \sum_i \sum_{j \in N_i} f_3(L_i, L_j) + \sum_i \sum_{j \in M_i} f(L_i, H_j) \quad (14)$$

全局边特征函数 $f_3(L_i, L_j)$ 依赖于 2 个真实标签变量,它定义为

$$f_3(L_i, L_j) = \begin{cases} \phi_1 & \text{若 } L_i = L_j = B \\ \phi_2 & \text{若 } L_i = L_j = F_k \\ \phi_3 & \text{若 } L_i \neq L_j \end{cases} \quad (15)$$

式中,参数 ϕ_1, ϕ_2 和 ϕ_3 也为常量,表示真实标签层中全局邻域随机变量之间的强度。

3 实验结果和分析

为了验证算法的有效性,对本文算法进行实验分析。算法由 VC++ 实现,机器配置为 P4 双核 1G 内存。主要是和典型的背景建模方法进行算法比较。对于背景建模算法,近年来虽然提出了一系列改进算法,但是其核心理论都是首先采用一段相对比较稳定的背景进行参数建模^[2]。因此,和文献[2]算法进行实验结果比较。实验数据包括两类视频,第 1 类是摄像机只有轻微抖动所拍摄的视频,第 2 类是摄像机一直在转动所拍摄得到的视频。

对于第 1 类视频,对于摄像机的轻微晃动,视频

背景像素的颜色变化呈多峰分布。因此,对于这类视频,无论是本文算法还是背景建模方法,都能够得到比较稳定的分割效果。在这里,只给出本文算法得到的分割结果(见图 1)。

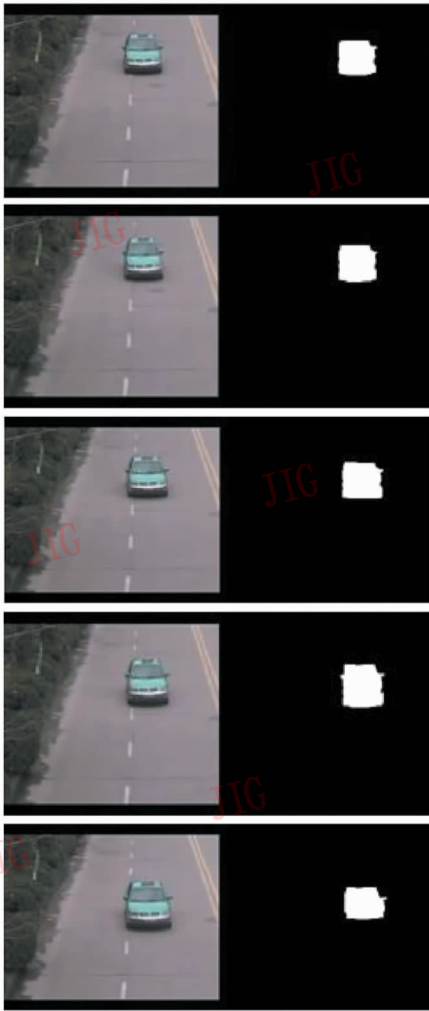


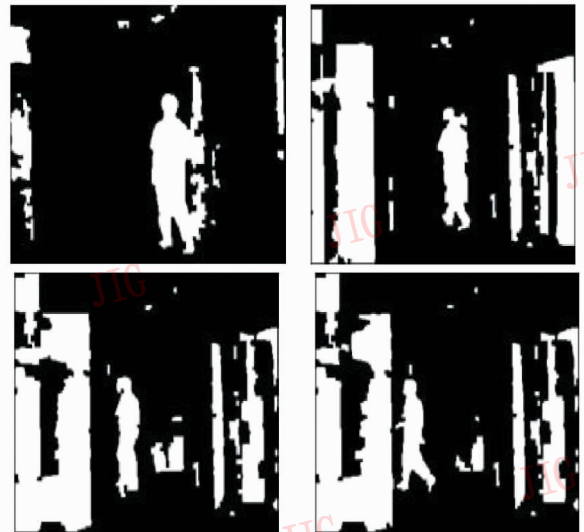
图 1 摄像机相对固定所得到的视频分割结果

Fig. 1 Video segmentation result for still cameras

但是,对于第 2 类视频,采用传统的背景建模方法,难以得到合理的分割效果,主要是背景建模方法需要一段相对比较稳定的背景进行模型初始化。但是,第 2 类视频由于摄像机一直在运动,背景建模方法无法有效地区分背景和运动目标。大面积的背景被错误地划分为运动目标,从而出现比较明显的分割错误。如图 2(b) 所示。而本文算法由于引入了显著性映射对时空特征进行建模,尽管背景在不断晃动和变化,但是整个视频中运动目标的显著特性并没有改变。因此视域注意模型能够有效地跟踪到运动目标,而忽略背景的变化,如图 2(c) 所示。



(a) 原始视频



(b) 使用背景建模方法的分割结果



(c) 本文算法结果

图 2 摄像机运动所得到的视频分割结果

Fig. 2 Video segmentation result for moving cameras

最后,对本文算法复杂度进行分析。本文算法主要由两部分构成,第 1 部分是计算时空注意模型,其复杂度为 $O(\theta)$,其中 $\theta = \frac{M \times N}{m \times n} \times \omega$,其中 ω 为降维得到的维数,约为 100。第 2 部分层条件随机场模型,其复杂度取决于 Boykov 提出的最小割算法,该算法最坏情况下复杂度为 $O(MN^2 |C|)$ 。因此,长度为 V 的视频序列的总体复杂度为 $O(V(MN^2 |C| + \theta))$ 。对于 Stauffer 等人的算法,需要对每一帧的每一个像素进行混合高斯建模,通常取 3~5 个高斯模型,则大小为 $M \times N$ 的一帧复杂度为 $O(MN)$,长度为 V 的视频序列的总体复杂度为 $O(VMN)$ 。从上述分析可以发现,本文算法复杂度有所增加。如何在保证分割质量的前提下,减少算法复杂度,提高算法实时性,也是后续研究需要考虑的问题。

4 结 论

如何从视频中准确地提取出运动目标是计算机视觉领域的难题。针对已有背景建模方法在处理复杂背景时的不足,提出根据视觉心理学理论,采用视频的运动特征、颜色特征等模拟人类视觉信号激励,然后通过条件随机场模型进行分割建模。实验结果表明,算法可以有效地提高分割准确率。

但是人类视觉感知是一个复杂的过程,如何建立能够有效地模拟其认知机制模型也是学术界的一大难题。本文只是采用运动特征、颜色特征建立视觉信号激励。但是对于人类视觉信号激励过程,需要进一步深入研究。如何能够像人的视觉一样,针对不同的动态背景,在不需要进行任何前期训练的情况下,快速准确地提取运动对象,是视频分割后续要研究的一个重要方向。

参考文献 (References)

- [1] Zhao Ming, Li Na, Chen Chun. Statistical inference for automatic video object segmentation[J]. Journal of Computer Aided Design & Computer Graphics, 2003, 15(3): 318-323. [赵明, 李娜, 陈纯. 采用统计推断的自动视频对象分割[J]. 计算机辅助设计与图形学学报, 2003, 15(3): 318-323.]
- [2] Stauffer C, Grimson W E. Adaptive background mixture models for real-time tracking[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Fort Collins, CO, USA: IEEE Computer Society, 1999: 246-252.
- [3] Wu Zhengping, Bu Jiajun, Chen Chun. Dynamic programming based video segmentation[J]. Journal of Computer Aided Design & Computer Graphics, 2002, 14(8): 743-746. [邬正平, 卜佳俊, 陈纯. 一种基于动态规划的视频分割方法[J]. 计算机辅助设计与图形学学报, 2002, 14(8): 743-746.]
- [4] Elgammal A, Harwood D, Davis L S. Non-parametric background model for background subtraction[J]. Lecture Notes in Computer Science, 2000, 1843: 751-767.
- [5] Wang Linbo, Zhao Jieyu. Video image segmentation based on bayesian learning[J]. Journal of Image and Graphics, 2005, 10(9): 1073-1078. [王林波, 赵杰煜. 基于贝叶斯学习的视频图像分割[J]. 中国图象图形学报, 2005, 10(9): 1073-1078.]
- [6] Gao Li, Yang Shuyuan, Li Haiqiang. New efficient automatic vop segmentation based on spatio-temporal information[J]. Journal of Image and Graphics, 2005, 9(10): 1096-1104. [高丽, 杨树元, 李海强. 一种有效的基于时空联合的视频对象自动分割新算法[J]. 中国图象图形学报, 2005, 9(10): 1096-1104.]
- [7] Carmona E J, Martínez-Cantos J, Mira J. A new video segmentation method of moving objects based on blob-level knowledge[J]. Pattern Recognition Letters, 2008, 29(3): 272-285.
- [8] Liu Shaohua, Zhang Maojun, Zhi Liding. Video object segmentation based on map of pixel blocks[J]. Journal of Image and Graphics, 2009, 14(1): 125-130. [刘少华, 张茂军, 智利丁. 基于像素块最大后验概率的视频对象分割方法[J]. 中国图象图形学报, 2009, 14(1): 125-130.]
- [9] Stenger B, Ramesh V, Paragios N, et al. Topology free hidden markov models: Application to background modeling [C]//Proceedings of International Conference on Computer Vision. New York, NY, USA: IEEE Computer Society, 2001: 294-301.
- [10] Chen Rui, Deng Yu, Xiang Shiming, et al. A non-parametric foreground/background segmentation method by fusion of intensity and edge Feature [J]. Journal of Computer Aided Design & Computer Graphics, 2005, 23(11): 1222-39. [陈睿, 邓宇, 向世明, 等. 结合强度和边界信息的非参数前景/背景分割方法[J]. 计算机辅助设计与图形学学报, 2005, 17(6): 1278-1284.]
- [11] Yaser S, Mubarak S. Bayesian modeling of dynamic scenes for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(11): 1778-1792.
- [12] Migdal J, Grimson E. Background subtraction using markov thresholds[C]//Proceedings of IEEE Workshop on Motion and Video Computing. Washington DC, USA: IEEE Computer Society, 2005: 58-65.
- [13] Wang Y, Ji Q. A Dynamic Conditional random field model for object segmentation in image sequences [C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE Computer Society, 2005: 264-270.

- [14] Chu Yiping. Study on Video Object Segmentation Algorithm Based Conditional Random Fields[D]. Hangzhou: Zhejiang University, 2008. [褚一平. 基于条件随机场模型的视频目标分割算法研究[D]. 杭州: 浙江大学, 2008.]
- [15] Seemann E, Fritz M, Schiele B. Towards robust Pedestrian detection in crowded image sequences [C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA: IEEE Computer Society, 2007: 1-8.
- [16] Mikolajczyk K, Leibe B, Schiele B. Multiple object class detection with a generative model [C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. New York NY, USA: IEEE Computer Society, 2006: 26-36.
- [17] Chum O, Zisserman A. An exemplar model for learning object classes[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA: IEEE Computer Society, 2007: 235-242.
- [18] Yin P, Criminisi, A, Winn J. Tree-based classifiers for bilayer video segmentation [C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA: IEEE Computer Society, 2007: 407-414.
- [19] Chu Yiping, Chen Qin, Huang Yejue, et al. Bilayer video segmentation based on random ferns[J]. Pattern Recognition and Artificial Intelligence, 2009, 22(3): 463-467. [褚一平, 陈勤, 黄叶珏, 等. 基于随机蕨丛的双层视频分割算法[J]. 模式识别与人工智能, 2009, 22(3): 463-467.]
- [20] Liu C, Yuen P C, Qiu G P. Object motion detection using information theoretic spatio-temporal saliency [J]. Pattern Recognition, 2009, 42(11): 2897-2906.
- [21] Boykov Y, Veksler O. Fast approximate energy minimization via graph cuts [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 23(11):1222-1239.