

基于隐条件随机场的自适应视频分割算法

褚一平^{1,2} 张引¹ 叶修梓^{1,3} 张三元¹

摘要 视频目标分割是视频监视与视频目标跟踪、视频目标识别以及视频编辑的基础. 本文提出了一种基于隐条件随机场 (Hidden conditional random fields, HCRF) 的自适应视频分割算法, 利用 HCRF 模型对视频序列中的时空邻域关系建模. 使用在线学习的方式对相应的参数进行调整, 实现对时空邻域约束关系的权重调整, 提高视频目标分割细节上的效果. 大量的数据测试表明, 与高斯混合模型 (Gaussian mixture model, GMM) 和联合时空的马尔可夫随机场 (Markov random fields, MRF) 等算法相比, 该算法的分割错误率分别降低了 23% 和 19%.

关键词 视频分割, 隐条件随机场, 在线学习
中图分类号 TP391

Adaptive Video Segmentation Algorithm Using Hidden Conditional Random Fields

CHU Yi-Ping^{1,2} ZHANG Yin¹ YE Xiu-Zi^{1,3} ZHANG San-Yuan¹

Abstract Video object segmentation is important for video surveillance and video object tracking, video object recognition and video editing. An adaptive video segmentation algorithm based on hidden conditional random fields (HCRFs) is proposed, which models spatio-temporal constraints of video sequence. In order to improve the segmentation quality, the weights of spatio-temporal constraints are adaptively updated by on-line learning of HCRFs. The experimental results have demonstrated that the error ratio of our algorithm is reduced by 23% and 19%, respectively, compared with Gaussian mixture model (GMM) and spatio-temporal Markov random fields (MRF).

Key words Video segmentation, hidden conditional random fields, on-line learning

1 引言

视频目标分割是计算机视觉中的一个热点问题^[1]. 在视频监视与视频目标跟踪、视频目标识别以及视频编辑中, 视频目标分割算法的优劣直接关系到这些后继处理的质量. 由于应用环境的复杂多样性, 视频目标分割算法必须在保证实时的情况下, 能够在摄像机抖动、光线变化以及户外的动态复杂的背景下仍有效地对前景目标进行分割. 在视频序列中, 各帧的每个像素点与同一帧中的相邻像素点以及前一帧的相邻像素点都存在着相互影响的关系, 合理地利用这些相互关系有利于提高视频目标分割的效果.

背景累积相减法是一种比较简单有效的视频目

标分割方法^[2], 它通过累积的方法把像素值变化比较小的点归类为背景, 通过视频序列的当前帧与背景相减获得前景目标. 背景累积相减法的优点是原理简单、实现方便, 而且算法的运行很快, 但是它无法很好地处理光线变化快、有树叶杂草晃动等复杂动态变化的背景. 高斯混合模型 (Gaussian mixture model, GMM)^[3,4] 是一个非常著名的背景建模方法, 它可以有效地对复杂背景进行建模. 利用 GMM 和背景相减相结合的方法可以较好地实现对视频目标的分割. Stenger^[5] 提出一种拓扑结构可变的隐马尔可夫模型 (Hidden Markov model, HMM) 方法对背景进行建模, 它可以处理光线的突然变化. 这两种方法都是基于像素一级的分割, 没有利用像素之间的相关信息.

相同帧的像素点之间以及相邻帧的像素点之间存在相互影响的关系, 合理地利用这些信息有助于提高视频目标分割的质量. 马尔可夫随机场 (Markov random fields, MRF) 可以较好地处理像素点之间的邻域约束关系, 文献 [6] 利用 MRF 对像素点之间的时空信息进行约束, 使用 Gibbs 采样算法求解最大后验概率实现了对视频前景目标的分割. Zhou^[7] 使用多分辨 MRF 模型, 进一步对像素点之间的时空信息进行约束, 使得算法可以较好地抵抗户外环境中的图像噪声以及摄像机抖动. 文献 [8]

收稿日期 2006-8-10 收修改稿日期 2006-12-20
Received August 10, 2006; in revised form December 20, 2006
国家自然科学基金 (60473106, 60333010), 国家教育部博士点基金 (20060335114), 浙江省教育厅资助项目 (G20030433) 资助
Supported by National Natural Science Foundation of China (60473106, 60333010), National Research Foundation for the Doctoral Program of Higher Education of China (20060335114), and Education Office of Zhejiang Province (G20030433)

1. 浙江大学计算机学院 杭州 310027 2. 杭州电子科技大学计算机学院 杭州 310018 3. 浙江大学 CAD & CG 国家重点实验室 杭州 310027

1. College of Computer Science, Zhejiang University, Hangzhou 310027 2. College of Computer Science, Hangzhou Dianzi University, Hangzhou 310018 3. State Key Laboratory of CAD & CG, Zhejiang University, Hangzhou 310027
DOI: 10.1360/aas-007-1252

提出了一种基于分层 MRF 模型的视频目标分割方法, 用于交通流量视频统计.

Elgammal^[9] 提出了一种非参数方法对背景进行建模, 采用高斯核函数估计视频序列上的像素点的概率密度, 用于处理复杂环境中对背景的建模问题. 陈^[10] 在文献 [9] 的基础上结合边界信息用于提高运动目标检测的健壮性. Yaser^[11] 对 Elgammal 的算法进行了扩展, 增加了对前景目标的建模能力, 同时利用 MRF 对同一帧图像的邻域关系建模, 通过求解 MRF 实现前景背景的分割.

条件随机场 (Conditional random fields, CRF) 是由 Lafferty^[12] 首先提出来的, 它相对于 HMM 等模型而言, 可以选择上下文相关特征, 解决了标记偏见的问题, 通过对所有特征进行全局归一化, 获得全局的最优值. 它的各种改进模型被大量应用于自然语言处理以及图像目标识别等领域^[13~15], 效果明显优于使用 HMM 以及 MRF 模型的结果. Wang^[16, 17] 把 CRF 模型引入到视频目标分割, 该算法通过 CRF 模型对视频像素点时空邻域信息进行建模, 构造了一个滤波器可以根据前一帧视频图像的 CRF 模型数据对当前帧视频图像的 CRF 模型进行迭代更新. Quattoni^[18] 把隐变量引入到 CRF 中, 替代 HMM 进行图像中的目标识别, 这就是隐条件随机场 (Hidden conditional random fields, HCRF), 之后 HCRF 分别被应用于语音分类^[19] 和手势识别^[20] 中.

MRF 模型是一种生成模型 (Generative model), 它对视频目标的分割最终需借助于计算后验概率 $P(L|X) = P(X|L)P(L)/P(X)$ 来完成, 其中 X 为观察值, L 为对观察值的标签. 这需要对视频序列图像的各种概率分布作更多的假设. 而 CRF 是判别模型 (Discriminative model), 它直接对 $P(L|X)$ 进行建模, 相对 MRF 来说可以减少概率分布的假设. HCRF 模型引入了隐变量, 增强了对随机变量 X 为结构型数据时的处理能力. 视频序列的各像素点在相同帧以及相邻帧之间存在着时空邻域约束关系, 利用 HCRF 对结构型数据的处理能力可以更加自然地在这种时空邻域关系进行建模.

本文提出了一种自适应的视频目标分割算法, 算法利用 HCRF 模型对视频序列中时空邻域信息进行建模, 实现对每个像素点的分类 (前景和背景). 通过在线学习方法, 根据场景的变化自动地调整 HCRF 的参数, 从而调整当前帧的时空邻域约束关系的权重, 提高视频分割的质量.

2 隐条件随机场视频分割模型

HCRF 是由 Quattoni 首先提出来的, 并应用于图像目标识别^[18]. 假设 $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ 是

视频序列中像素点的观察集合, 则 X 的每个元素 \mathbf{x}_j 为局部观察, 它对应一个特征函数 $f(\mathbf{x}_j)$, 存在一个隐随机变量集 $H = \{h_1, h_2, \dots, h_m\}$ 与 X 对应. 假设存在 $h_i \in H$, 则它的邻域记为 $h_k (k \in N_i \cup M_i)$, 其中 N_i 表示点 i 的所有同一视频帧的邻域, 即空域邻域; M_i 表示点 i 在上一帧对应的邻域, 称为时域邻域. 随机变量集 H 及其相应的邻域关系组成一个无向图 G , 用 E 表示无向图的边, $(i, k) \in E$ 表示隐随机变量 h_i 和 h_k 存在邻域关系, 在本文中它是一种树结构. $L \in \{0, 1\}$ 表示分类标签, 其中 0 表示背景, 1 表示前景, 在本文中 $h_i \in H$ 取值范围与 L 相同.

2.1 像素级前景背景分割

GMM 算法^[3, 4] 是一种流行的像素级别的前景背景视频分割方法, 它利用 3 ~ 5 个高斯函数混合对视频序列中的各个像素进行估计. 视频序列中的每一个像素点, GMM 都有 K 个固定的状态与之对应, 每个状态 $i = 1, 2, \dots, K$ 都有一个高斯密度函数 $p_i(\mathbf{x}^t)$ 和一个先验概率 ω_i^t 与之关联, 其中 \mathbf{x}^t 表示第 t 帧中像素点的 RGB 值. 高斯密度函数 $p_i(\mathbf{x}^t)$ 定义为

$$p_i(\mathbf{x}^t) = \frac{1}{(2\pi)^{d/2} |\Sigma_i^t|^{1/2}} \times \exp\left(-\frac{1}{2}(\mathbf{x}^t - \boldsymbol{\mu}_i^t)^T (\Sigma_i^t)^{-1} (\mathbf{x}^t - \boldsymbol{\mu}_i^t)\right) \quad (1)$$

式中 $\boldsymbol{\mu}_i^t$ 为均值, Σ_i^t 为方差矩阵, 一般假设 RGB 各分量是相互独立的, 所以 Σ_i^t 为对角矩阵; d 表示 RGB 颜色空间的维数.

每个新的 \mathbf{x}^t 将与各个状态对应的 $\boldsymbol{\mu}_i^t$ 和 Σ_i^t 计算相应的马氏距离 (Mahalanobias distance) $\lambda = (\mathbf{x}^t - \boldsymbol{\mu}_i^t)^T (\Sigma_i^t)^{-1} (\mathbf{x}^t - \boldsymbol{\mu}_i^t)$. 如果 λ 在阈值范围内^[3], 则认为该 \mathbf{x}^t 与该状态对应的高斯分布相匹配, 它的参数作如下更新

$$\omega_i^{t+1} = \omega_i^t + \alpha M^t$$

$$\boldsymbol{\mu}_i^{t+1} = (1 - \rho)\boldsymbol{\mu}_i^t + \rho\mathbf{x}^{t+1}$$

$$\Sigma_i^{t+1} = (1 - \rho)\Sigma_i^t + \rho(\mathbf{x}^{t+1} - \boldsymbol{\mu}_i^{t+1})^T (\mathbf{x}^{t+1} - \boldsymbol{\mu}_i^{t+1})$$

式中 α 和 ρ 为学习速率, 均为常数. M^t 表示当前的状态是否被匹配, 如果是则为 1, 否则为 0. ω_i^t 将被归一化, 并且根据 $\omega_i^t / \|\Sigma_i^t\|$ 的值以降序的方式排列相应的状态. 使用公式 $\hat{B} = \arg \min_b \sum_{i=1}^b \omega_i^t > th$ 估计 GMM 中的前 \hat{B} 个状态, 其中 th 为阈值. 如果在前 \hat{B} 个状态中存在匹配的高斯分布, 则认为该像素为背景, 并保存相应的马氏距离供后继处理使用, 否则为前景. 如果前 \hat{B} 个状态中没有与 \mathbf{x}^t 匹配的

状态, 则选出一个先验概率最小的状态, 对其进行重新初始化.

2.2 HCRF 模型

HCRF 模型对一系列的局部观察进行建模, 计算其条件概率

$$P(L|X; \theta) = \sum_H P(L, H|X; \theta) = \frac{1}{Z(X; \theta)} \sum_H \exp(\Psi(L, H, X; \theta)) \quad (2)$$

其中 $Z(X; \theta)$ 是分配函数, 它保证模型有适合的归一化概率, 它由式 $Z(X; \theta) = \sum_{L', H} \exp(\Psi(L', H, X; \theta))$ 计算得到, 其中 H 是隐状态变量, 它无法直接观察得到, θ 是模型的参数. $\Psi(L, H, X; \theta)$ 是特征函数, 在本文中它被定义为

$$\Psi(L, H, X; \theta) = \sum_i \sum_k \theta_k^1 f_k^1(L, h_i, \mathbf{x}) + \sum_{(i,j) \in E} \sum_k \theta_k^2 f_k^2(L, h_i, h_j, \mathbf{x}) \quad (3)$$

其中 θ_k^1 和 θ_k^2 是需要学习的参数, 特征函数 $f_k^1(L, h_i, \mathbf{x})$ 和 $f_k^2(L, h_i, h_j, \mathbf{x})$ 分别代表单个隐变量对应的强度以及与相邻的隐变量之间形成的强度.

在视频目标分割中, 邻域能够提供一些非常有用的信息, 利用这些信息可以提高分割的准确性. 图 1 为同一帧中各结点之间形成的 8 空域邻域关系, 以及相邻两帧之间形成的 9 时域邻域关系. 这些时空邻域关系被称为团 (Clique), 团的各个结点对应 HCRF 中的隐变量. 隐变量邻域强度特征函数 $f_k^2(L, h_i, h_j, \mathbf{x})$ 表示团中各结点的分类与其邻域结点的相互影响程度. 邻域强度特征函数 $f_k^2(L, h_i, h_j, \mathbf{x})$ 定义如下

$$f_k^2(L, h_i, h_j, \mathbf{x}) = \delta(h_i, h_j) \quad (4)$$

其中 $\delta(\cdot)$ 是 Kronecker Delta 函数.

本文使用 GMM 模型对背景建模, 单个隐变量被分类为背景时, 它对应的强度特征值将由其对应的背景马氏距离表示. 背景对应的马氏距离值在 2.1 节中已讨论过. 假设前景的像素是均匀分布的, 那么对于以 RGB 组成的视频颜色空间来说, 前景的概率为 $\exp(-\ln 2^{24})$ ^[6]. 特征函数 $f_k^1(L, h_i, \mathbf{x})$ 被定义为

$$f_k^1(L, h_i, \mathbf{x}) = \begin{cases} \lambda/2 & \text{if } h_i = 0 \\ \ln 2^{24} & \text{if } h_i = 1 \end{cases} \quad (5)$$

式中 λ 为背景对应的马氏距离.

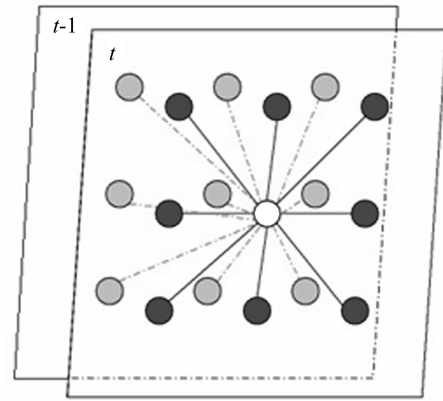


图 1 视频序列中的时空邻域关系 (第 t 帧的黑色结点为同一帧中白色结点的空域邻域, 第 $t-1$ 帧的灰色结点与第 t 帧的白色结点形成了时域邻域关系)

Fig. 1 Spatio-temporal neighboring relationship in video sequence (The sites colored by black are spatial neighboring relationship for white site in the same frame, while gray sites in previous frame and white site in current frame form temporal neighboring relationship.)

团中隐变量之间的欧氏距离的计算需分两种情况. 空域邻域关系的隐变量之间的欧氏距离可以根据它们各自对应的帧图像的像素点位置来计算; 对于时域邻域关系的隐变量之间的欧氏距离计算, 我们假设当前帧的结点与前一帧的各个邻域结点在 x 和 y 坐标上相差一个位置, 所以时域邻域关系的隐变量之间的欧氏距离被固定为 $\sqrt{2}$.

2.3 模型的估计与参数学习

HCRF 的估计可以通过下式来计算

$$\hat{L} = \arg \max_L P(L|X; \theta) \quad (6)$$

下面介绍如何快速、有效地计算条件概率. 我们可以推导出下式

$$P(L|H, X; \theta) = \frac{1}{Z(L, X; \theta)} \exp(\Psi(L, H, X; \theta))$$

其中 $Z(L, X; \theta) = \sum_H \exp(\Psi(L, H, X; \theta))$. $Z(X; \theta)$ 可以使用置信度传播算法 (Belief propagation, BP)^[21] 计算得到. $Z(L, X; \theta)$ 也可以使用与 $Z(X; \theta)$ 相同的推导算法计算. 获得 $Z(L, X; \theta)$ 之后, 我们可以使用下式来计算边缘似然

$$P(L|X; \theta) = \frac{1}{Z(X; \theta)} \sum_H \exp(\Psi(L, H, X; \theta)) = \frac{Z(L, X; \theta)}{Z(X; \theta)} \quad (7)$$

把式 (7) 代入式 (6) 中, 即可得到估计的结果.

在文献 [20] 中, 参数学习通过梯度增长法搜索 $\hat{\theta} = \arg \max_{\theta} l(\theta)$, 并使用了 Quasi-Newton 优化技术, 其中 $l(\theta)$ 是参数学习时的目标函数, 它是使用批量训练方法来进行参数学习的. 在视频分割中, 使用在线训练方法进行参数学习更加适合. 在实验中, 我们使用视频序列中先前分割结果对后继参数进行训练, 对模型的参数进行迭代更新使得参数可以根据视频序列场景的变化而变化.

算法使用随机梯度增长法 (Stochastic gradient descent, SGD)^[19] 更新模型的参数. 假设 $(L^{(i)}, X^{(i)})$, $i = 1, 2, \dots$ 是训练数据对的集合, 参数根据下式进行更新

$$\theta_j^{(n+1)} = \theta_j^{(n)} + \eta^{(n)} \nabla_{\theta_j} \ln P(L^{(i)}|X^{(i)}; \theta) \quad (8)$$

其中 $\eta^{(n)}$ 是学习速率, 本文选择一个常量作为学习速率, $\nabla_{\theta_j} \log P(L^{(i)}|X^{(i)}; \theta)$ 是对单个训练数据的梯度, 对 θ_k^1 和 θ_k^2 计算分别如下

$$\begin{aligned} \nabla_{\theta_k^1} \ln P(L^{(i)}|X^{(i)}; \theta) = & \\ & \sum_{h_i \in H} P(h_i|L^{(i)}, X^{(i)}) f_k^1(L, h_i, \mathbf{x}) - \\ & \sum_{h_i \in H, L'} P(h_i, L'|X^{(i)}) f_k^1(L', h_i, \mathbf{x}) \quad (9) \end{aligned}$$

同理可得

$$\begin{aligned} \nabla_{\theta_k^2} \ln P(L^{(i)}|X^{(i)}; \theta) = & \\ & \sum_{(i,j) \in E} P(h_i|L^{(i)}, X^{(i)}) f_k^2(L, h_i, h_j, \mathbf{x}) - \\ & \sum_{(i,j) \in E, L'} P(h_i, L'|X^{(i)}) f_k^2(L', h_i, h_j, \mathbf{x}) \quad (10) \end{aligned}$$

式中 $P(h_i|L^{(i)}, X^{(i)})$ 和 $P(h_i, L'|X^{(i)})$ 也使用 BP 算法计算.

2.4 实现

使用 GMM 对每个局部观察 \mathbf{x}_i 进行分类, 在分类的同时计算出对应的马氏距离. 本文将 θ_k^1 赋值为 1, $(i, j) \in E$, $j \in N_i$ 邻域对应的参数 θ_k^2 , 使用 $1/\|h_i - h_j\|$ 来初始化, 其中 $\|h_i - h_j\|$ 表示两个隐变量对应图像中的欧氏距离, 它可以在程序开始时预计算; $(i, j) \in E$, $j \in M_i$ 邻域对应的参数 θ_k^2 , 赋初始值为 $\sqrt{2}/(2|M_i|)$, 其中 $|M_i|$ 表示 h_i 对应前一帧的邻域像数, 由于本文采用固定的时空邻域数目, 它的值也是预先计算的. 每个参数 θ_k^2 都设有最大和最小的阈值, 如果参数经过不断的学习之后超出了这个范围就会被重新初始化.

HCRF 视频分割算法流程为:

- 1) 初始化参数 θ , 学习间隔 $LStep$ 初始化为 0;
- 2) 获取一帧图像, 利用 GMM 算法计算帧中各像素点对应的分类, 以及对应的特征函数值;
- 3) 根据参数 θ , 利用 BP 算法计算 $\hat{L} = \arg \max_L P(L|X, \theta)$;
- 4) 如果 $LStep$ 达到学习间隔, 则把当前帧的分类结果作为训练数据 $(L^{(i)}, X^{(i)})$ 进行增量学习, 并把 $LStep$ 设置为 0. 如果 θ 超出了设定的阈值, 则重新初始化该参数;
- 5) 学习间隔 $LStep$ 加 1;
- 6) 重复步骤 2), 直到结束.

3 实验结果

实验使用 CPU 为 2.4 GHz、内存为 512 M 的计算机, 算法使用 Visual C++ 和 OpenCV 实现. 测试的视频序列帧大小都为 320×240 , 程序的处理速度大约为每秒 10 帧. 采用的实验方式是通过摄像机拍摄各种情况的视频 (包括室内户外、不同光线、不同背景), 再到计算机上测试. 还有一部分测试视频是来自于网络上的一些标准测试视频数据. 大约测试了 12000 多帧的视频图像, 通过与 GMM^[3] 和联合时空信息的 MRF^[6] 等视频分割算法的比较, 本文算法在细节的分割上准确性更高. 下面给出了在不同环境、不同光照和不同的背景条件下, 本文算法与其他算法的分割效果的比较. 本文的算法和联合时空信息的 MRF 都采用 8 空域和 9 时域邻域关系, 算法的学习间隔为 12 帧, 三种算法分割后的结果均未使用形态学滤波器处理.

图 2 为街道行人监视视频, 这段视频的背景相对比较简单. GMM 由于每个像素是单独分割的, 没有考虑像素之间的时空信息, 在前景与背景着色比较接近的边界上分割的效果不是特别好. 联合时空信息的 MRF 由于分割过程中使用了时空的信息, 分割的效果要好于 GMM, 但是由于 MRF 没有加入动态学习势函数参数的功能, 对环境的适应性不如本文算法. 图 2(d) 为本文算法分割的效果, 从图中可以看出, 在细节的分割上明显优于前两种算法. 这是由于本文的算法具有自适应能力, 它可以根据不同的环境变化, 通过学习 HCRF 的参数来调节时空领域信息的强度.

图 3 是商店中顾客经过一个店铺的视频数据, 它的背景比前一个例子要复杂一些. 由于这段视频夹杂着较多由摄像机产生的噪音, 在 GMM 分割的结果中分布着较多的噪音点. 联合时空信息的 MRF 以及本文算法由于利用了时空邻域的信息, 这些噪音点基本上被消除了. 对比图 3(c) 和 (d) 中视频目标的头发部分的分割, 可以看出本文算法比联合时

空信息的 MRF 效果更好.

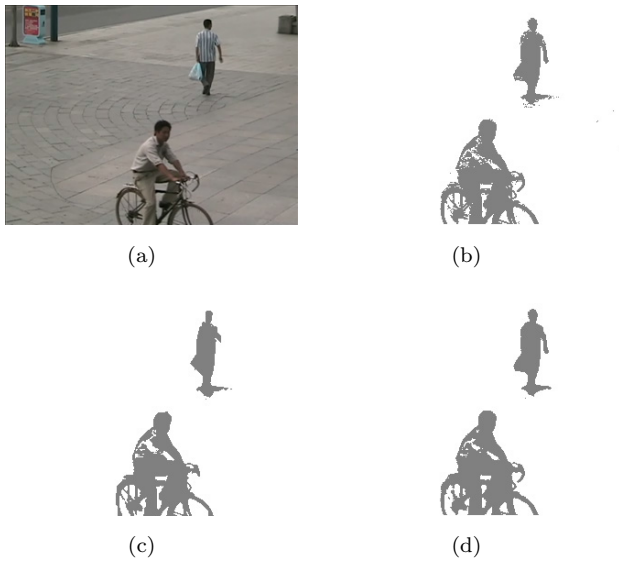


图 2 街道行人的分割效果 ((a) 第 477 帧的原图像; (b) GMM 方法分割的结果; (c) 联合时空信息的 MRF 分割的结果; (d) 本文算法分割的结果)

Fig. 2 Segmentation results for pedestrian ((a) Original image of frame 477; (b) Result segmented by GMM; (c) Segmented by using spatio-temporal MRF; (d) Result using proposed method)

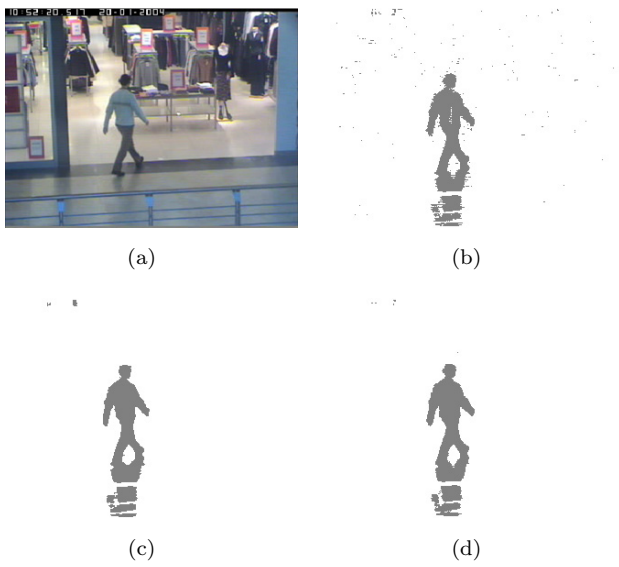


图 3 商店顾客的分割效果 ((a) 第 490 帧的原图像; (b) GMM 方法分割的结果; (c) 联合时空信息的 MRF 分割的结果; (d) 本文算法分割的结果)

Fig. 3 Segmentation results for a customer at a mall ((a) Original image of frame 490; (b) Result segmented by GMM; (c) Segmented by using spatio-temporal MRF; (d) Result using the proposed method)

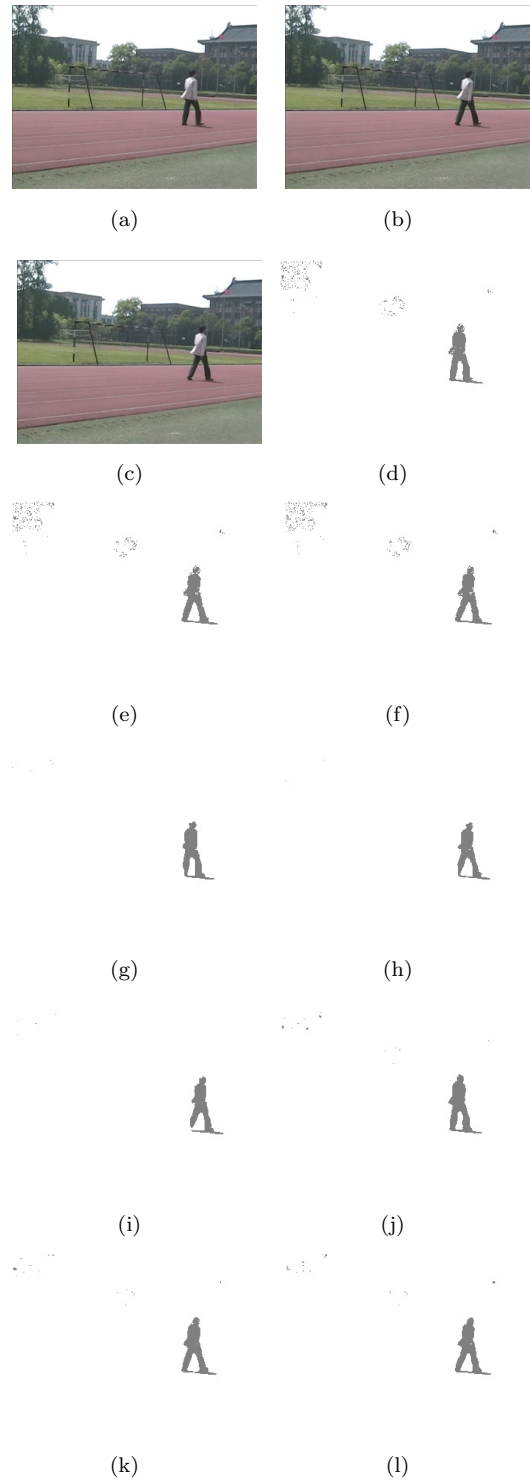


图 4 校园跑道中行人的连续三帧分割结果 ((a)~(c) 第 201~203 帧的原图像; (d)~(f) GMM 方法分割的结果; (g)~(i) 联合时空信息的 MRF 分割的结果; (j)~(l) 本文算法分割的结果)

Fig. 4 Segmentation results for the walker at campus ((a)~(c) Original images of frames 201~203; (d)~(f) Results segmented by GMM; (g)~(i) Segmented by using spatio-temporal MRF; (j)~(l) Result using the proposed method)

图 4 是在校园操场上拍摄的一段视频, 我们提供了三种不同的方法对这段视频连续 3 帧的分割结果. 这段视频背景很复杂, 它既有大量晃动的树叶又有不断飘动的红旗, 目标头部的颜色与背景也比较接近. 由于背景是动态的, 使用 GMM 算法分割时, 少量的背景点被误分为前景点. 由于这些点比较散, 使用时空信息之后, 大部分点又重新被归为背景. 图 4(j)~(l) 为本文方法分割的效果, 它减少了前景背景以及边界部分的误分类.

为了估计各种算法的分割错误率, 我们通过手工分割视频序列的各帧图像作为目标分割的基准图像. 使用 GMM、联合时空的 MRF 以及本文算法分割的结果与基准图像进行比较, 从而计算各个算法的分割错误率. 使用 FP 表示误报率, FN 表示漏报率, 总错误率为两者之和. 表 1 列出了 GMM、联合时空的 MRF 以及本文算法的分割错误率. 从表中可以看出 MRF 的分割错误率要低于 GMM, 而本文方法的分割错误率最低, 只有 0.86%, 比 GMM 和 MRF 的分割错误率分别降低了 23% 和 19%.

表 1 GMM、MRF 以及本文方法的分割错误率对比
Table 1 Error ratios for GMM, MRF, and the proposed method

	FP(%)	FN(%)	Total(%)
GMM	0.53	0.58	1.11
MRF	0.32	0.73	1.05
HCRF	0.37	0.49	0.86

4 结论

本文提出了自适应视频目标分割算法, 该算法利用 HCRF 模型融合了视频序列中的时空邻域信息, 通过在线学习的方法对场景变化进行学习, 调整了时空约束关系的权重, 提高了后继分割的质量. 通过实验对比, 使用本文的方法分割的质量要好于其他算法. 下一步的工作将解决阴影的处理问题.

References

- Bao Hong-Qiang, Zhang Zhao-Yang, Chen You-Ming. Multiple video object segmentation based on spatio-temporal curve evolution. *Acta Electronica Sinica*, 2005, **33**(1): 181~185
(包红强, 张兆杨, 陈右铭. 基于时空曲线演化的多视频运动对象分割算法. *电子学报*, 2005, **33**(1): 181~185)
- Yang T, Li S Z, Pan Q, Li J. Real-time and accurate segmentation of moving objects in dynamic scene. In: Proceedings of the ACM 2nd International Workshop on Video Surveillance and Sensor Networks. New York, USA: IEEE, 2004. 136~143
- Stauffer C, Grimson W. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 747~757
- Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition. USA: IEEE, 2004. 28~31
- Stenger B, Ramesh V, Paragios N, Coetzee F, Buhmann J M. Topology free hidden Markov models: application to background modeling. In: Proceedings of the 8th International Conference of Computer Vision. USA: IEEE, 2001. 294~301
- Migdal J, Grimson E. Background subtraction using Markov thresholds. In: Proceedings of Workshop on Motion and Video Computing. USA: IEEE, 2005. 58~65
- Zhou Y, Xu W, Tao H, Gong Y H. Background segmentation using spatial-temporal multi-resolution MRF. In: Proceedings of Workshop on Motion and Video Computing. USA: IEEE, 2005. 8~13
- Chu Yi-Ping, Ye Xiu-Zi, Zhang Yin, Zhang San-Yuan. Algorithm for vibration-resisted video segmentation via hierarchical MRF. *Journal of Zhejiang University(Engineering Science)*, 2007, **41**(11): 47~50
(褚一平, 叶修梓, 张引, 张三元. 基于分层 MRF 模型的抗抖动视频分割算法. *浙江大学学报(工学版)*, 2007, **41**(11): 47~50)
- Elgammal A, Duraiswami R, Harwood D, Davis L S. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of IEEE*, 2002, **90**(7): 1151~1163
- Chen Rui, Deng Yu, Xiang Shi-Ming, Li Hua. A non-parametric foreground/background segmentation method by fusion of intensity and edge feature. *Journal of Computer-Aided Design and Computer Graphics*, 2005, **17**(6): 1278~1284
(陈睿, 邓宇, 向世明, 李华. 结合强度和边界信息的非参数前景/背景分割方法. *计算机辅助设计与图形学学报*, 2005, **17**(6): 1278~1284)
- Yaser S, Mubarak S. Bayesian modeling of dynamic scenes for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(11): 1778~1792
- Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning. San Francisco, USA: IEEE, 2001. 282~289
- Sha F, Pereira F. Shallow parsing with conditional random fields, Technical Report CIS TR MS-CIS-02-35, University of Pennsylvania, USA, 2003
- Kumar S, Hebert M. Discriminative random fields: a discriminative framework for contextual interaction in classification. In: Proceedings of the 9th International Conference on Computer Vision. USA: IEEE, 2003. 1150~1157
- Kumar S, Hebert M. A hierarchical field framework for unified context-based classification. In: Proceedings of the 10th International Conference on Computer Vision. Pittsburgh, USA: IEEE, 2005. 1284~1291
- Wang Y, Ji Q. A dynamic conditional random field model for object segmentation in image sequences. In: Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2005. 264~270

- 17 Wang Y, Loe K F, Wu J K. A dynamic conditional random field model for foreground and shadow segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(2): 279~289
- 18 Quattoni A, Collins M, Darrel T. Conditional random fields for object recognition. In: *Proceedings of Conference on Neural Information Processing Systems*. USA: 2004
- 19 Gunawardana A, Mahajan M, Acero A, Platt J C. *Hidden Conditional Random Fields for Phone Classification*. USA: International Speech Communication Association, 2005
- 20 Wang S B, Quattoni A, Morency L P, Demirdjian D, Darrell T. Hidden conditional random fields for gesture recognition. In: *Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition*. USA: IEEE, 2006. 1521~1527
- 21 Yedidia J S, Freeman W T, Weiss Y. Understanding belief propagation and its generalizations. *Exploring Artificial Intelligence in the New Millennium*. USA: Morgan Kaufmann, 2003. 239~269



褚一平 浙江大学计算机学院博士研究生. 2005 年获得杭州电子科技大学软件学院硕士学位, 主要研究方向为计算机视觉与机器学习.

E-mail: hzcyp@yahoo.com.cn

(**CHU Yi-Ping** Ph.D. candidate at College of Computer Science, Zhejiang University. He received his master degree from Hangzhou Dianzi University in 2005. His research interest covers computer vision and machine learning.)

His research interest covers computer vision and machine learning.)



张引 博士, 副教授. 主要研究方向为计算机图形图像处理.

E-mail: Yinzh@cs.zju.edu.cn

(**ZHANG Yin** Associate professor at College of Computer Science, Zhejiang University. Her research interest covers computer graphics and imaging.)



叶修梓 博士, 教育部“长江学者奖励计划”特聘教授. 主要研究方向为 CAD/CAM、计算机图形图像处理、计算机视觉. 本文通信作者.

E-mail: yxz@cs.zju.edu.cn

(**YE Xiu-Zi** He is currently a Cheung Kong Chair Professor under China Education Ministry's Cheung Kong

Scholar's Program at College of Computer Science, Zhejiang University. His research interest covers CAD/CAM, computer graphics and imaging, and computer vision. Corresponding author of this paper.)



张三元 教授. 主要研究方向为计算机图形图像处理.

E-mail: syzhang@cs.zju.edu.cn

(**ZHANG San-Yuan** Professor at College of Computer Science, Zhejiang University. His research interest covers computer graphics and imaging.)