

基于像素块最大后验概率的视频对象分割方法

刘少华 张茂军 智利丁

(国防科学技术大学信息系统与管理学院,长沙 410073)

摘要 提出一种基于像素块最大后验概率的视频运动对象快速分割算法。首先根据连续3帧视频序列构建参考帧,计算基于像素块的差分图像;然后用Gibbs随机场模型描述基于像素块的差分图像,通过最大后验概率方法分割运动对象块;最后使用基于颜色梯度的snake模型提取运动对象轮廓。实验结果表明,该算法能避免视频对象的缺失,降低图像噪声的影响,减少误分割,同时具有较快的计算速度。

关键词 像素块 视频对象分割 最大后验概率 Gibbs随机场 snake模型

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2009)01-0125-06

Video Object Segmentation Based on MAP of Pixel Blocks

LIU Shao-hua, ZHANG Mao-jun, ZHI Li-ding

(School of Information System and Management, National University of Defense Technology, Changsha 410073)

Abstract A video object segmentation method based on MAP (maximum a posteriori) of pixel blocks is proposed. Firstly, a reference frame is built from three continuous frames and a pixel block-based difference image is calculated. Secondly, a pixel block-based difference image is described by Gibbs random field model, the moving object blocks are detected by MAP. Last, the contour of the moving object is extracted by the snake method. The experiments results show that the algorithm can avoid the losing and mistakes with a good speed.

Keywords pixel block, video object segmentation, MAP, Gibbs random field, snake model

1 引言

视频序列中运动对象分割对于视频监控、计算机视觉、面向对象的视频压缩编码等以视频对象为基础的应用有着重要的意义。运动对象分割是一种利用连续视频序列中的时间信息和单帧图像中的空间信息进行图像分割的过程。一般的分割方法处理的基本单位是像素,就是分别对每一个像素进行运算和判断,这样有较高的精度,同时带来较大的计算量,而且算法的鲁棒性较差,当运动对象面积较大且运动缓慢时,易产生空洞和缺失,当噪声较大时易产生误分割。

为了增加分割结果的完整性和鲁棒性,并减少计算时间,许多基于像素块的分割方法被提出来。像素块表现的是某个范围内像素点的整体特性,多数像素的共同特征可以得到继承,而少数像素的突出特征会被忽略,因此能够填补像素的缺失,并过滤噪声和干扰。文献[1]使用基于像素块的区域生长方法进行图像分割。文献[2]构造基于像素块的自适应背景模型,并与基于像素的方法相结合。文献[3]~[5]根据像素块的运动矢量分割运动对象。文献[6]根据分割出的边界块获得运动对象。文献[7]通过计算像素块的NVD(normalized vector distance)来分割运动块。以上方法中,文献[1]方法仅考虑了图像的空间特性,适合于静态图像。文

基金项目:国家自然科学基金项目(60273066,60473116)

收稿日期:2007-05-21; 改回日期:2007-08-31

第一作者简介:刘少华(1982~),男,2004年于国防科学技术大学获控制科学与工程专业硕士学位。研究方向为多媒体信息系统与虚拟现实技术。E-mail: lshh37@yahoo.com.cn

献[2]在粗糙分割后还要逐像素的处理,导致耗时大于单一的基于像素的方法。文献[3]~[5]、[7]只能获得粗糙的运动对象。文献[6]只适合具有小范围往复运动的头肩序列,否则会产生大量伪边界块。

针对以上问题,将最大概率后验概率(MAP)框架引入到基于像素块的对象分割中来,首先获得基于像素块的最优分割结果,然后使用 snake^[8]方法得到精确的对象轮廓,整体的计算时间远少于基于像素的方法,同时可获得较为精确的运动对象。整个过程如图 1 所示。

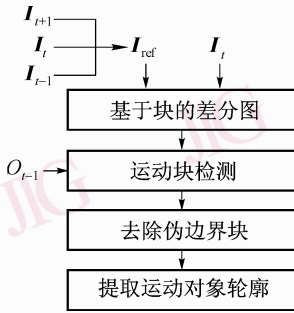


图 1 基于像素块的视频对象分割过程
Fig. 1 Process of video object segmentation based on MAP of pixel blocks

在构造基于像素块的差分图时,先利用连续的 3 帧视频序列 I_{t+1}, I_t, I_{t-1} 构建特殊的参考图像 I_{ref} , 突出像素变化部分,减弱像素未变化部分的噪声,使分割结果更加完整、准确。在使用 MAP 方法分割运动块时,利用马尔可夫场(MRF)与 Gibbs 分布的等价性^[9]定义系统势能函数,将先验知识、空间相关性和似然度结合起来,综合判断像素块的特性,以获得最优分割结果。在使用 snake 方法提取对象轮廓前,通过去除伪边界块减少迭代次数,进一步减少计

算时间。实验结果表明,该方法获得的视频运动对象有较好的完整度,而且可有效地减少误分割,同时具有较快的速度。

2 参考图像的构建

将当前图像与参考图像进行差运算,可以得到图像中的变化信息,在理想情况下,图像的变化部分也就对应着场景中的运动对象。

两个连续帧进行差运算得到图像变化的部分 C ,应当是两帧中运动对象的并集,即

$$C_{t,t+1} = O_t \cup O_{t+1} \quad (1)$$

O_t 是 t 时刻图像中属于运动对象像素点的集合,可以做出如下推导:

$$C_{t,t-1} \cap C_{t,t+1} = O_t \cup (O_{t-1} \cap O_{t+1}) = O_t \quad (2)$$

由式(2)可以得出,两个连续帧差图的交集就是我们所感兴趣的运动对象。然而这只是理想情况,现实中的图像噪声和各种干扰使得 $C_{t,t-1} \cap C_{t,t+1} \neq O_t$ 。

为了减小噪声干扰,对每一帧图像序列进行 3×3 的高斯滤波。令 $D_{t,t-1} = |I_t - I_{t-1}|$, $D_{t,t+1} = |I_t - I_{t+1}|$, I_t 表示 t 时刻高斯滤波后的图像, P 表示图像中的像素点集合,图像的宽度和高度分别为 W, H 。则

$$C_{t,t\pm 1} = \{P(i,j) \mid D_{t,t\pm 1}(i,j) > K, i = 1, \dots, W; j = 1, \dots, H\} \quad (3)$$

式(3)中 K 的取值视具体的应用情况在 5 ~ 10 之间选取^[10],则

$$C_{t,t-1} \cap C_{t,t+1} = O'_t \quad (4)$$

本文采用下式计算参考图像:

$$R_t(i,j) = \begin{cases} I_{t-1}(i,j) & \text{if } P(i,j) \in O'_t \text{ and } D_{t,t-1}(i,j) > D_{t,t+1}(i,j) \\ I_{t+1}(i,j) & \text{if } P(i,j) \in O'_t \text{ and } D_{t,t-1}(i,j) < D_{t,t+1}(i,j) \\ (I_{t-1}(i,j) + I_{t+1}(i,j))/2 & \text{if } P(i,j) \notin O'_t \end{cases}$$

式(5)的物理意义为当前帧对应的参考图像根据前一帧和后一帧获得。当前帧分别与前一帧和后一帧进行差运算,对完全满足式(3)的像素,可初步判定为发生变化的像素,此时为了更加突出变化信息,选取与当前像素差值较大的像素值为参考像素的值;对部分满足式(3)的像素,说明该像素可能只是帧间差产生的边缘重叠区域,此时应弱化其变化

$$\begin{aligned} & P(i,j) \in O'_t \text{ and } D_{t,t-1}(i,j) > D_{t,t+1}(i,j) \\ & \text{或 } P(i,j) \in C_{t,t+1} \text{ and } P(i,j) \notin C_{t,t-1} \\ & P(i,j) \in O'_t \text{ and } D_{t,t-1}(i,j) < D_{t,t+1}(i,j) \\ & \text{或 } P(i,j) \in C_{t,t-1} \text{ and } P(i,j) \notin C_{t,t+1} \\ & P(i,j) \notin C_{t,t-1} \text{ and } P(i,j) \notin C_{t,t+1} \end{aligned} \quad (5)$$

信息,选取与当前像素差值较小的像素值为参考像素的值;对于完全不满足式(3)的像素,可初步判定为未发生变化的像素,为减小噪声,参考像素的取值为前一帧与后一帧的均值。使用式(5)计算的参考图像 R_t 与当前帧进行差运算,可以在减小噪声干扰的同时获得更加“突出”的运动信息,有助于更加准确完整地分割出运动对象。

3 运动对象块的分割

3.1 基于块的差分图像构建

以块为单位进行运动对象分割,以每一个像素块的 SAD 为判断依据。令 $D_t = |I_t - R_t|$, 将 D_t 划分为 $M \times N$ 个 $BL \times BL$ 的像素块。则位于 (m, n) 位置像素块的 SAD 由下式计算:

$$SAD(m, n) = \sum_{x=1}^{BL} \sum_{y=1}^{BL} |I_t(m \times BL + x, n \times BL + y) - R_t(m \times BL + x, n \times BL + y)| \quad (6)$$

由每个像素块的 SAD 组成基于像素块的帧差图 D_{sad} , 其中的每一个点都表示 D_t 中一个 $BL \times BL$ 的像素块的整体特征, 对 D_{sad} 进行分割虽然减小运动对象的分割精度, 但可以利用空间相关性增加完整度并过滤噪声干扰。

3.2 差分图像的高斯模型

令 $D = \{d(x, y), (x, y) \in S\}$ 表示将 D_{sad} 在 $(0, 255)$ 间归一化后的灰度值。运动对象的分割可以看成是对 S 中每个像素点进行二值标记的过程。 S 可以看成是一个 2 维随机场, 其中包含两种事件即 $\Theta(x, y) = \text{static}$ (未发生变化), $\Theta(x, y) = \text{mobile}$ (发生变化)。令 $P(d|\text{static}), P(d|\text{mobile})$ 分别表示在 H_0, H_1 假设下像素灰度值的概率密度函数。这两个概率密度函数可以看作服从参数为 μ, σ 的高斯分布, 即

$$P(d(x, y) | \Theta(x, y)) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (7)$$

令 P_0, P_1 分别为 H_0, H_1 假设的先验概率, 则 S 中像素点的概率密度函数为

$$P(d) = P_0 P(d|\text{static}) + P_1 P(d|\text{mobile}) \quad (8)$$

该混合分布中的未知参数 $\{P_l, \mu_l, \sigma_l; l \in \{0, 1\}\}$, 可以通过 EM (期望最大) 算法求解。

3.3 基于最大后验概率的运动对象分割

运动对象的分割实际就是对 2 维随机场 S , 根据事先观察场 D 得到标记场 L 的过程。希望得到具有最大后验概率的标记场。根据贝叶斯公式, 最大后验概率标记场可表示为

$$\hat{L} = \arg \max(p(D|L)p(L)) \quad (9)$$

使用 Gibbs 模型描述 MRF, 根据式 (9) 定义 Gibbs 势能函数:

$$U(D, L) = U_1(D|L) + U_2(L) \quad (10)$$

根据 MAP 框架, 最大化后验概率等价于最小化

式 (10) 中的 Gibbs 势能函数。其中, $U_1(D|L)$ 反映了观察值与标记场之间的相关性, $U_2(L)$ 反映了时空邻域上的图像连续性, 分别定义如下^[11-12]:

$$U_1(D|L) = \sum_{s \in S} \left(\lambda_l \frac{\mu_{1-l}(d_s - \mu_l)^2}{2\sigma_l^2} \right) \quad (11)$$

$$U_2(L) = \sum_{s \in S} \sum_{r \in K} V(s, r) \quad (12)$$

$$V(s, r) = \begin{cases} -\beta & l_s = l_r \\ +\beta & l_s \neq l_r \end{cases} \quad (13)$$

式 (11) 中的 λ 为平衡参数, $\lambda_1 \lambda_0 = 1$, 目的是减小前景点和背景点间的能量差距。式 (12) 中的 R 表示某一标记点的时空二阶邻域^[11]。本文只考虑当前邻域和过去邻域两种情况。式 (13) 中的 $\beta > 0$, 目的是消除噪声并利用时空连续性填补分割区域空白。当 R 为过去邻域和当前邻域时, β 取值分别为 β_1 和 β_2 。本文采用迭代条件模型 (ICM) 来近似求解势能函数的最小值。

4 运动对象的轮廓提取

通过上一节的处理, 已经得到包含运动对象的像素块集合 O , 如图 2 所示为 OneManMoving 序列第 30 帧的分割结果。对于 O 中的某个元素 o , 如果其四连通邻域内存在某一点 $s \in S$, 且 $s \notin O$, 则 o 是运动对象的边界块。将每个边界块的中心作为运动对象轮廓的初始控制点, 然后使用基于彩色梯度^[10] 的 snake 方法就可得到运动对象的精确轮廓。

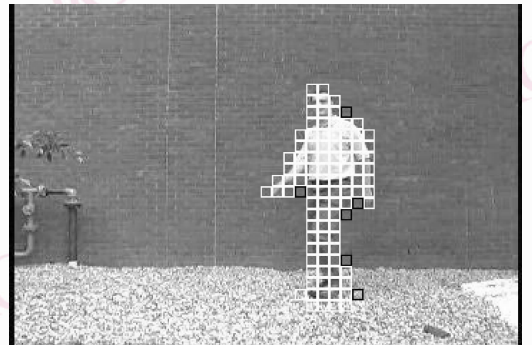


图2 视频对象块的分割及伪边界块的判断
Fig. 2 Segmentation of video object blocks and detection of false boundary blocks

边界块包含 3 种情况: 只包含运动对象, 运动对象与背景的混合, 只包含背景。最后一种边界块为伪边界块, 由其得到的初始控制点与对象的真实轮廓距离较大。在实际应用中, 伪边界块的数量很少,

约占全部边界块的 $1/10 \sim 1/30$, 但却直接导致整个 snake 算法迭代次数的增加, 严重时甚至无法收敛到轮廓处, 当运动对象周围背景比较复杂时, 还会导致收敛到错误的轮廓处。为避免这种情况, 要去除只包含背景的伪边界块。

如图 3 所示, 将边界块 O_b 划分为 5 个子块^[7], 分别为 $w_1 \sim w_5$ 。分别计算出 $w_1 \sim w_5$ 的 SAD 值 $c_1 \sim c_5$ 。计算 $c_1 \sim c_5$ 的方差 $V(c_i)$, 如果 $V(c_i) < th$, 可认为该块是伪边界块。检测出的伪边界块如图 2 中的黑色方框所示。

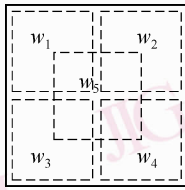


图 3 边界块的子块划分

Fig. 3 Partition of a boundary block

5 实验结果

分别测试 4 个视频序列, 分辨率分别为 360×240 , 320×240 , 352×240 , 768×576 。硬件环境为 CPU: Celeron 2.66 GHz, RAM: 512 M, 软件环境为 MATLAB7.1。BL 取值为 8。

5.1 耗时分析

表 1 显示了分别采用本文方法和文献[11]中基于像素的 MAP 方法对 4 个视频序列进行分割时平均每帧的耗时对比。表 2 显示的是本文方

法中各个步骤所耗的时间比例。文献[11]的方法可对应于本文方法的前两个步骤, 第 1 步二者的计算量基本相同, 因此时间的缩减主要来自第 2 步。本文方法的样本单位为像素块, 当 BL 取值为 8 时, 样本数量缩减为普通方法的 $1/64$, 由于结果的求取是一个迭代过程, 因此实际的计算时间会更少。如表 1 所示, 本文方法的平均计算时间仅相当于普通方法的约 $1/70$ 。如果要进一步提高速度可采用收敛速度更快的算法替换当前的 EM 算法和 ICM 算法, 但是会带来分割效果的下降。

表 1 本文方法与普通方法耗时对比

Tab. 1 Time used by the proposed algorithm and the normal algorithm

| 所耗时间(s) | 视频 1 | 视频 2 | 视频 3 | 视频 4 |
|---------|--------|--------|--------|--------|
| 本文方法 | 0.406 | 0.327 | 0.391 | 2.058 |
| 传统方法 | 32.402 | 20.282 | 29.727 | 94.621 |

表 2 基于块的视频对象分割算法耗时比例

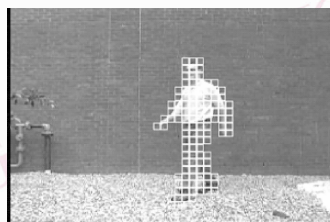
Tab. 2 Time used by every step of the algorithm

单位: %

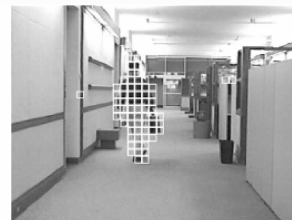
| | 对象分割步骤 | | |
|------|--------|-------|-------|
| | 参考帧构造 | 运动块分割 | 轮廓提取 |
| 视频 1 | 29.69 | 55.58 | 14.73 |
| 视频 2 | 31.88 | 58.84 | 9.28 |
| 视频 3 | 31.97 | 59.85 | 8.18 |
| 视频 4 | 34.38 | 63.22 | 2.50 |

5.2 结果分析

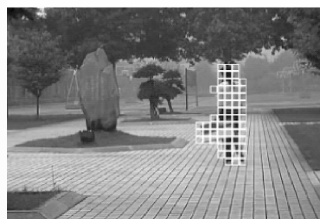
图 4 分别为 4 个视频序列的第 37、58、213 和 377 帧采用本文方法分割得到的运动对象, 图中用



(a) 第37帧



(b) 第58帧



(c) 第213帧



(d) 第377帧

图 4 基于像素块 MAP 方法分割出的运动对象

Fig. 4 Moving objects got by the pixel-block-based MAP algorithm

白色方框标记出了组成运动对象的像素块。图 5 为采用文献[11]中基于像素的 MAP 方法分割出的运动像素,为方便观看采用了掩模方式显示。图 6 为采用 snake 算法对图 4 中的运动对象块处理后得到的运动对象。

从图 5 可以看出,传统方法得到的运动像素并不完整,有缺失,连通性不好,在背景区域受噪声影响产生了许多误分割。这些问题不仅出现在运动对象面积较大的视频 1、2、3 中,即使对视频 4 中运动面积较小的汽车也仍然存在。尤其是视频 2,缺失面积很大,受噪声干扰严重,产生大量误分割像素

点。对这种分割结果必需再进行细致的后处理才能得到可用的运动对象,会增加计算量并且仍然不能保证运动对象的完整性。图 4 中本文方法对 4 个测试视频都未发现缺失,而且对噪声有很好的过滤作用,仅视频 2 有少量误分割块。

从图 4、图 5 的对比分析可以看出,相对于传统方法,本文的分割方法在完整性和抗噪性上都有较好的效果,缺点是获得的对象不够精细,但换来的是计算时间的大量减少。在许多实际应用中,如多对象分割、目标跟踪、对象编码等,视频对象的完整度以及计算速度才是最受关注的,对视频对象的细致

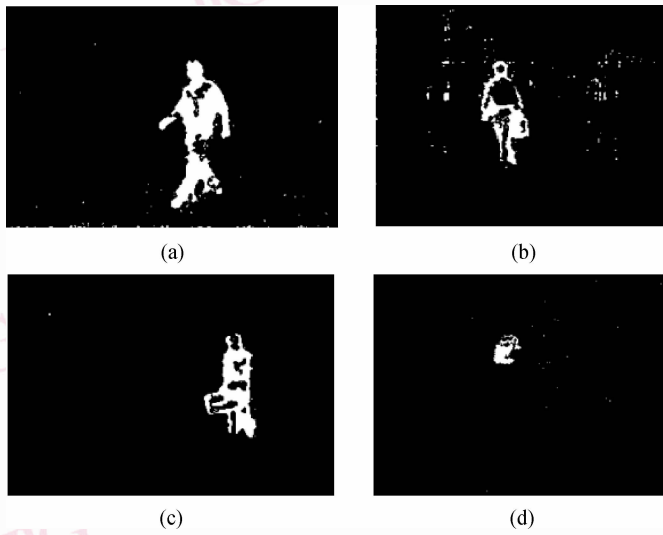


图 5 基于像素 MAP 方法分割得到的运动像素
Fig. 5 Moving pixels got by the pixel-based MAP algorithm

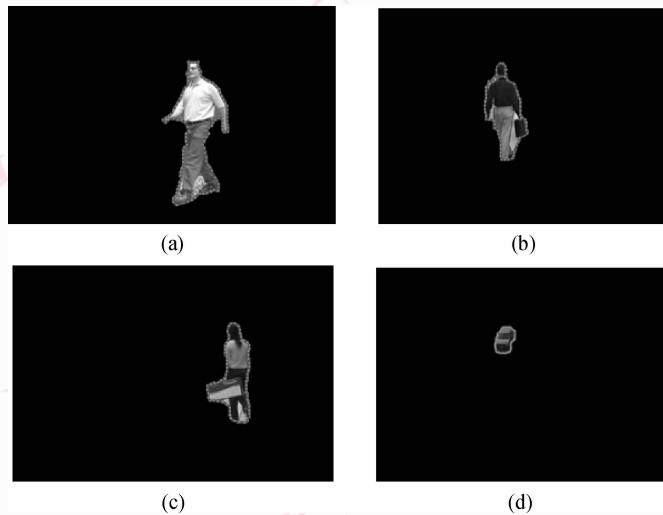


图 6 使用 snake 方法得到的运动对象
Fig. 6 Moving objects got by snake algorithm

程度要求在其次,而且进一步使用 snake 方法只需很小的计算代价就可得到细致的运动对象。存在的问题就是最终获取的运动对象在凹陷处不是十分准确,这是由 snake 算法无法收敛到凹陷处的特性所决定的。

其他的实验结果表明,本文分割方法所得到视频对象的完整度和计算时间与像素块以及视频对象本身的大小有关。视频对象的面积越大,则需要较大的 BL 值才能得到理想的效果, BL 值越大,则计算时间越短、得到的视频对象更加完整同时更加粗糙、snake 方法要更多的迭代次数,反之亦然。通常 BL 取值为 4、8、16,可根据视频对象面积的大小以及对速度的要求对不同的视频具体选取。

6 结 论

以像素块为单位,使用最大后验概率方法进行运动对象块的分割,使用基于颜色梯度的 snake 算法提取运动对象的轮廓。实验结果表明,该算法能较好地克服传统方法不可避免的像素缺失,降低噪声干扰,减少误分割,并具有较快的速度。同时也看到,由于 snake 算法的不足,对于不规则运动对象的凹陷区域,轮廓并不是非常精确。因此下一步可以考虑采用改进的 snake 算法来获取更为准确的轮廓。同时,如何利用摄像机运动参数估计来分割变化背景中的运动对象和多个对象的分割也是下一步要研究的问题。

参考文献 (References)

- 1 Tan W H, Coatrieux G, Solaiman B, *et al.* A region based segmentation using pixel block fuzzy similarity [A]. In: Proceedings of the 2nd IEEE International Conference on Information & Communication Technologies: from Theory to Applications [C], Damascus, Syria, 2006, **1**: 1516-1521.
- 2 Chen Y T, Chen C S, Huang C R, *et al.* Efficient hierarchical method for background subtraction [J]. Pattern Recognition, 2007, **40**(10): 2706-2715.
- 3 Zeng W, Du J, Gao W, *et al.* Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model [J]. Real-Time Imaging, 2005, **11**(4): 290-299.
- 4 Huang Y R, Kuo C M, Huang F C. Block-based motion field segmentation for video coding [J]. Journal of Visual Communication and Image Representation, 2005, **16**(6): 668-687.
- 5 Won C S. A block-based MAP segmentation for image compressions [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1998, **8**(5): 592-601.
- 6 Kim B C, Park R H. A Fast automatic VOP generation using boundary block segmentation [J]. Real-Time Imaging, 2004, **10**(2): 117-125.
- 7 Matsuyama T, Ohya T, Habe H. Background subtraction for non-stationary scenes [A]. In: Proceedings of the 4th Asian Conference on Computer Vision [C], Taipei, China, 2000: 627-630.
- 8 Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models [J]. International Journal of Computer Vision, 1988, **1**(4): 321-331.
- 9 Bao Hong-qiang, Zhang Zhao-yang. Multiple video object segmentation based on maximization of the a posteriori probability of spatio-temporal label field [J]. Journal of Electronics & Information Technology, 2006, **26**(2): 232-236. [包红墙, 张兆扬. 基于时空标记场最大后验概率的多视频目标分割算法 [J]. 电子与信息学报, 2006, **26**(2): 232-236.]
- 10 Wu Si, Lin Shou-xun, Zhang Yong-dong. Automatic segmentation of moving objects in video sequences based on dynamic background construction [J]. Chinese Journal of Computers, 2005, **28**(8): 1386-1392. [吴思, 林守勋, 张勇东. 基于动态背景构造的视频运动目标自动分割 [J]. 计算机学报, 2005, **28**(8): 1386-1392.]
- 11 Zhan Jing-fen, Qi Fei-hu, Wang Hai-long. Moving object segmentation technology based on spatial-temporal markov random field [J]. Journal of China Institute of Communications, 2000, **21**(11): 63-68. [詹劲峰, 戚飞虎, 王海龙. 基于时空马尔可夫随机场的运动目标分割技术 [J]. 通信学报, 2000, **21**(11): 63-68.]
- 12 Zhong Xing-rong, Huang Xian-wu, Wang Jia-jun, *et al.* Automatic segmentation of moving object based on markov random field models using fast EM algorithm [J]. Microelectronics & Computer, 2004, **21**(1): 102-105. [钟兴荣, 黄贤武, 王加俊等. 基于快速 EM 算法的马尔可夫随机场模型运动目标自动分割 [J]. 微电子学与计算机, 2004, **21**(1): 102-105.]