

一种新的基于时空马尔可夫随机场的运动目标分割技术

黄贤武 朱莉 仲兴荣 王加俊
(苏州大学电子信息学院 苏州 215021)

摘要 在图像处理领域, 视频图像序列中的运动目标分割技术是一个被广泛研究的热点课题。该文提出一种新的基于时空马尔可夫随机场的运动目标分割技术。首先, 对视频序列的前后 3 帧图像进行处理, 获得两帧初始标记场; 随后, 对两帧初始标记场进行“与”操作, 获得共同标记场; 最后, 以原始图像的色彩聚类图像作为先验知识, 重新定义 Gibbs 能量函数, 并利用迭代条件模型(ICM)实现最大后验概率(MAP)的估算问题, 获得优化标记场。实验结果表明: 该模型克服了传统时空马尔可夫随机场模型因运动产生的显露遮挡现象, 同时减弱了运动一致性造成的空洞现象并削弱了噪声的影响。

关键词 图像分割, 马尔可夫随机场, 迭代条件模型

中图分类号: TP391.4

文献标识码: A

文章编号: 1009-5896(2006)02-0367-05

A Novel Moving Object Segmentation Technology Based on Spatiotemporal Markov Random Field

Huang Xian-wu Zhu Li Zhong Xing-rong Wang Jia-jun
(School of Electronics & Information Engineering, Soochow University, Suzhou 215021, China)

Abstract In the field of image processing, the segmentation of moving object in video sequences is a hot research topic in recent years. In this paper, a novel method of moving object segmentation based on spatiotemporal Markov Random Field(MRE) is proposed. Firstly, two observations and two initial labels are derived from the three successive images with the same method in the first scheme. Secondly, the AND-label is obtained with the AND-operation on the two initial labels. Finally, the image segmented with the color clustering algorithm is regarded as prior knowledge, with which the corresponding Gibbs energy function is redefined, and the maximum *a posteriori* estimator, which is determined by using the iterated conditional mode algorithm, is employed to get optimized labels. The new MRF model contributes to the weakening of the noise and to the elimination of the covered-uncovered background and to the recovery of the uniform moving regions.

Key words Image segmentation, Markov Random Field (MRF), Iterated Conditional Mode (ICM)

1 引言

近年来, 国内外对于运动目标的分割技术的研究可以分为: 基于运动信息的运动目标检测方法和基于时空信息的运动目标检测方法^[1-3]。基于马尔可夫随机场的运动目标分割算法是一种基于时空信息的运动目标分割方法。

传统的时空马尔可夫随机场模型以图像的事先观察场作为先验知识, 获得优化标记场。由于传统模型不能对运动造成的显露遮挡现象进行处理, 往往导致对运动目标的过分割, 从而无法收敛到运动目标的真实轮廓上。本文提出一种新的基于时空马尔可夫随机场的运动目标分割技术, 以原始图像的色彩聚类图像作为先验知识, 获得图像标记场, 并在

此基础之上重新定义 Gibbs 能量函数。该模型克服了运动产生的显露遮挡现象, 减弱了运动一致性造成的空洞现象和噪声的影响。该模型可以自动获取运动目标准确且连续的轮廓。

2 传统的时空马尔可夫随机场模型

一幅 $m \times n$ 的图像 I 可以表示为一系列点的集合 $I = \{i_1, i_2, \dots, i_s\}$ $s = m \times n$ 。对于这些点, 存在一个对应的事先观察场 $D = \{d_1, d_2, \dots, d_s\}$ 和一个对应的标记场 $F = \{f_1, f_2, \dots, f_s\}$ 。其中, 图像的事先观察场 D , 即帧差图像, 可以根据连续时刻 $t, t+1$ 的两帧图像 $I(t), I(t+1)$ 经计算获得:

$$D(t+1) = |I(t+1) - I(t)| \quad (1)$$

对图像的事先观察场 $D(t+1)$ 进行二值阈值化处理, 便可得到图像的初始标记场 $F(t+1)$ 。文献[4]在确立了时空二阶邻域系统的基础之上定义了时空马尔可夫随机场模型, 将运动目标的分割问题归结为在事先观察场和其他一系列约束条件下对图像初始标记场进行优化, 从而得到优化标记场的问题。即找出一个分割 f^* , 使得到的后验概率分布 $P(f|d)$ 在全图像上能够有一个全局最大值, 即求

$$f^* = \arg \max_{f \in F} P(F = f | D = d) \quad (2)$$

然而, 对式(1)进行深入分析, 不难发现当运动物体从 t 时刻运动到 $t+1$ 时刻, $t+1$ 时刻的显露背景作为运动信息被划分到变化区域中去。此部分伪运动信息将同时存在于事先观察场与初始标记场中, 而传统模型将包含伪运动信息的事先观察场作为先验知识, 对同样包含伪运动信息的初始标记场进行优化, 必然无法消除显露遮挡现象而造成的运动目标过分割, 从而无法实现对运动目标的正确提取。

3 显露遮挡现象的产生机理及解决方法

图1详细给出了遮挡显露现象的产生机理。

定义 $L(t_1, t_2)$ 表示图像标记场 $F(t_1)$ 中标记为运动的像素点集合在视频图像 $I(t_2)$ 上的投影, 得到变化区域。

定义 $B(t_1, t_2)$ 表示图像标记场 $F(t_1)$ 中标记为静止的像素点集合在视频图像 $I(t_2)$ 上的投影, 得到静止区域。

定义 $U(t)$ 表示从 $t-1$ 到 t 时刻, 由物体运动造成的 t 时刻视频图像 $I(t)$ 中显露背景像素点集合。

定义 $C(t, t+1)$ 表示从 t 到 $t+1$ 时刻, 由物体运动造成的 t 时刻视频图像 $I(t)$ 中遮挡背景像素点集合。

因此, 可得

$$L(t-1, t) = M(t) \cup U(t) \quad (3)$$

$$L(t, t+1) = M(t) \cup C(t, t+1) \quad (4)$$

至此, 我们不难发现, 在假设物体的运动不是迂回的条件下, $t+1$ 帧视频图像的前序运动所形成的背景显露 $U(t)$ 与后序运动所形成的背景遮挡 $C(t, t+1)$ 不相重叠, 故而对 $L(t-1, t)$ 与 $L(t, t+1)$ 进行与操作, 可以获得去除伪运动信息而获得不包含遮挡显露部分的运动区域 $M(t)$:

$$\begin{aligned} L(t-1, t) \cap L(t, t+1) \\ = \{M(t) \cup U(t)\} \cap \{M(t) \cup C(t, t+1)\} = M(t) \end{aligned} \quad (5)$$

图1所示的网格部分是 $L(t-1, t)$ 与 $L(t, t+1)$ 的共同部分, 即去除了伪运动信息的真实运动区域 $M(t)$ 。

4 一种新的时空马尔可夫随机场模型

为了根除传统模型无法消除显露遮挡现象这一固有弊

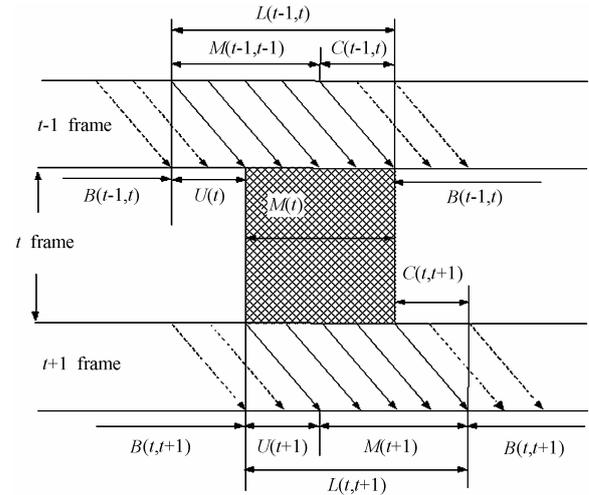


图1 显露遮挡现象的产生机理

Fig.1 The mechanism for engendering covered-uncovered background

端, 本节提出了一种新的基于时空马尔可夫随机场的运动目标分割思想: 以原始图像的色彩聚类结果图像作为先验知识, 对初始标记场进行优化, 获得图像标记场。在此基础上, 我们重新定义 Gibbs 能量函数, 并建立对应的马尔可夫随机场模型。检测运动目标步骤如下:

(1)采用区域生长法对 RGB 色彩空间的彩色图像进行色彩聚类, 获得色彩聚类图像。

(2)根据第2节的定义, $D = \{d_1, d_2, \dots, d_s\}$ 为序列图像的事先观察场。由式(1), 根据连续时刻 $t-1, t, t+1$ 的3帧图像 $I(t-1), I(t), I(t+1)$ 得到 $t, t+1$ 时刻的图像事先观察场 $D(t), D(t+1)$ 。

(3)对彩色图像的事先观察场进行二值阈值化处理, 得到图像的初始标记场。文献[4]提出一种自动获得阈值的方法, 通过寻找直方图最大斜率变化的谷点——直方图的极小值点, 自动确定阈值 z , 从而实现初始标记场的自动提取。

$$Z = \max_{i=1}^{254} \{h(i+1) + h(i-1) - 2 \times h(i)\} \quad (6)$$

式中 $h(i)$ 表示灰度差值取 i 的像素个数。由于图像中占数目较多的背景点对于光照的微弱变化响应较小, 式(6)右边在色彩差值取值较小的情况下产生极大值, 这样, 大部分的噪声点会作为运动像素点被划分到运动区域中, 从而对分割造成影响。因而, 我们需要对此方法进行改进。在光照条件几乎不变的条件下, 假设图像事先观察场中色差差值取值较小的图像点属于背景点是合理的。考虑到对于大多数的应用来说, 运动目标约占视频图像面积的30%, 通过提升搜索范围的下限, 使得小于该下限的色彩差值对应的像素总数最多占总像素的40%, 这样, 极有可能对应背景点的色彩差值被排除在阈值设定范围之外, 自动获得的初始标记场将更加接近真实情况。式(7)给出了阈值下限的设定方法, 式(8)则为提升搜索下限之后的阈值设定公式。

$$l = \left\{ \forall l \in (0, 254) \mid \sum_{i=0}^l h(i) \leq 0.4 \times N \text{ 且 } \sum_{i=0}^{l+1} h(i) > 0.4 \times N \right\} \quad (7)$$

$$Z = \max_{i=1}^{254} \{h(i+1) + h(i-1) - 2 \times h(i)\} \quad (8)$$

采用此方法对彩色图像各个色彩分量的色彩差值图像求得阈值 z_r, z_g 与 z_b , 可自动获得彩色图像的初始标记场。

$$f_i(t) = \begin{cases} 1, & |\Delta R_i(t)| \geq z_r \text{ 且 } |\Delta G_i(t)| \geq z_g, |\Delta B_i(t)| \geq z_b \\ 0, & \text{其他} \end{cases} \quad (9)$$

值得指出的是, 确定这 3 个阈值是一个比较复杂的问题, 直接关系到图像的分割质量及分割速率。本文仅仅对阈值设定提出了一种实验性的方法。依据经验, 利用马尔科夫随机场模型对初始模板处理之后, 所得的分割结果中运动像素的比例要比初始分割结果中的比例大。为了避免对运动目标的过分割, 我们将初始模板的计算公式定义为: 当且仅当 3 个色彩分量的差值均大于相应的色彩阈值, 方认为该点具有运动特性。

接着, 利用上文对显露遮挡现象的分析结论, 对前后标记场进行“与”操作以消除运动造成的显露遮挡现象:

$$\hat{F}(t+1) = F(t+1) \cap F(t) \quad (10)$$

最后, 采用空间二阶邻域系统(如图 2(a)所示)对获得的共同标记场进行优化, 进而获得优化标记场。需要说明的是, 由于该模型中被优化的共同标记场包含了时间域上连续两帧初始标记场的信息, 故而仍将该模型称为时空马尔可夫随机场模型。

在此模型中, 初始标记场的优化是将原始图像的色彩聚类图像 I 作为先验知识, 找出一个分割 f^* , 使得图像标记场后验概率分布 $P(f|i)$ 在全图像上能够有一个全局最大值, 即求

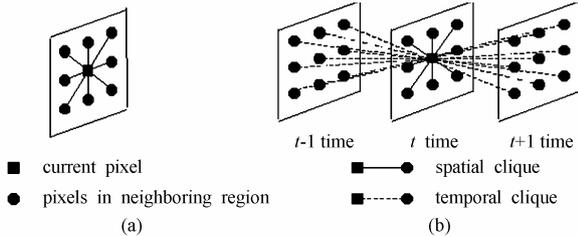


图 2 空间和时空二阶邻域系统

(a)本方案采用的空间邻域系统 (b)文献[4]采用的时空二阶邻域系统
 Fig.2 the neighborhood system with second order In spatiotemporal domain (a)the neighborhood system with second order in spatial domain in this scheme (b)the spatiotempral neighborhood system with second order in literature [4]

$$f^* = \arg \max_{f \in F} P(F = f | I = i) \quad (11)$$

根据贝叶斯理论和 Hammersley-Clifford 定理得

$$\begin{aligned} f^* &= \arg \max_{f \in F} P(F = f | I = i) \\ &= \arg \max_{f \in F} P(F = f)P(I = i | F = f) \\ &\propto \arg \min_{f \in F} U(f, i) \end{aligned} \quad (12)$$

其中 $U(f, i)$ 为系统能量函数, 该式表明: 运动目标分割的结果就是求取系统能量函数最小时的标记场。能量函数定义如下:

$$U(f, i) = U_1(f) + U_2(f, i) \quad (13)$$

式(13)中的 $U_1(f)$ 为先验模型能量, 是初始标记值的先验概率 $P(F = f)$ 对应的能量部分。由于假设初始标记场中各个像素的标记取值是独立的, 因此, 先验模型能量 $U_1(f)$ 可用图像中所有像素的先验模型能量的总和来表示:

$$U_1(f) = \sum_{j \in I} U_1(f_j) \quad (14)$$

像素的先验模型能量定义为该像素二阶邻域系统(如图 2(b)所示)中所有二元基团阈势能 $V_c(f_j, f_k)$ 的总和^[5,6]:

$$U_1(f_j) = \sum_{c \in C_j} V_c(f_j, f_k) \quad (15)$$

$\sum_{c \in C_j} V_c(f_j, f_k)$ 是点 j 的先验模型能量。其中, j 为当前像素

点, k 为像素点 j 的一个八邻域邻点。 C_j 是像素点 j 的二阶邻域系统 N_j 中所有二元基团的集合, $c = (j, k)$ 是基团集合中的一个二元基团, f_j, f_k 为二元基团中两像素的标记值, 二元基团阈势能定义如下

$$V_c(f_j, f_k) = \begin{cases} -\beta, & f_j = f_k \text{ 且 } j \text{ 与 } k \text{ 属于同一个色类} \\ -\beta_n, & f_j = f_k \text{ 且 } j \text{ 与 } k \text{ 不属于同一个色类} \\ -\beta_n, & f_j \neq f_k \text{ 且 } j \text{ 与 } k \text{ 属于同一个色类} \\ +\beta, & f_j \neq f_k \text{ 且 } j \text{ 与 } k \text{ 不属于同一个色类} \end{cases} \quad (16)$$

参数 β 及 β_n 的取值均为非负数。 β 与 β_n 取值大小反映了对先验模型与条件模型的信任度; β/β_n 的数值反映了对运动一致性与色彩一致性条件同时成立或任一条件成立的信任度。

至此, 我们可以得到完整的先验模型能量函数, 则式(14)改写如下:

$$U_1(f) = \sum_{j \in I} U_1(f_j) = \sum_{j \in I} \sum_{c \in C_j} V_c(f_j, f_k) \quad (17)$$

式(13)中 $U_2(f, i)$ 是反映色彩聚类图像和标记场之间相关性的条件概率模型所对应的能量函数。已知色彩聚类图像中各个色彩聚类之间相互独立, 在对某一个像素进行标记值测试时, 条件概率 $P(i|f)$ 仅仅与该像素点所属色彩聚类相关, 可以表示为

$$P(i|f) = \prod_{l=1}^L P(i_l|f) = P(i_l|f) \quad (18)$$

$$P(i_l|f) = [P(f|i_l) \times P(i_l)] / P(f) \quad (19)$$

在色彩聚类图像与初始标记场已知的条件下, $p(i_l), p(f)$ 与 $p(f|i_l)$ 均可计算求得

$$P(i_l) = \frac{\text{色类}l\text{像素点数}}{\text{图像像素点总数}} \quad (20)$$

$$P(f) = \frac{\text{初始标记场中标记取值为}f\text{的像素点数}}{\text{图像像素点总数}} \quad (21)$$

$$P(f|i_l) = \frac{\text{该色类中标记取值为}f\text{的像素点数}}{\text{该色类中像素点总数}} \quad (22)$$

考虑到概率的值域在 $[0,1]$ 之间,是一个小于1的数,故而条件概率模型能量函数定义为对条件概率函数取对数后乘以系数-1.0:

$$U_2(f, i) = -\lg P(i_l|f) = -\lg P(f|i_l) - \lg P(i_l) + \lg P(f) \quad (23)$$

GRF 能量耗费函数的构造表达了 GRF 模型对分割空间属性的期望。本文采用通用的最大后验概率估计的方法求解上述模型,标记场的后验概率分布最大化(最大后验概率)对应耗费函数最小化。Metropolis 算法和 Gibbs 采样器可以找到全局最小值,迭代条件模式(ICM)算法是这两种模拟退火算法的特例,其迭代结果是一种次优解。这样,式(17)改写为

$$U_1(f_j) = \sum_{c \in C_j} V_c(f_j, f_k) \quad (24)$$

ICM 算法的具体实现方法是:对于每一个像素点,采用运动标记“1”和静止标记“0”分别测试,使得在二阶邻域中能量较小的标记得到保留。当全局能量的减少量为整个能量的0.01%时,迭代过程结束。具体实现过程如下:

(1) 根据 $t-1, t, t+1$ 时刻的3帧连续图像 $I(t-1), I(t)$ 和 $I(t+1)$ 计算 $t, t+1$ 时刻的图像观察场 $D(t), D(t+1)$ 。

(2) 根据图像的事先观察场自动确定 RGB 各色彩分量的阈值,初始化 $t, t+1$ 时刻的标记场 $F(t), F(t+1)$ 。

(3) 对 $t, t+1$ 时刻的初始标记场 $F(t), F(t+1)$ 进行“与”操作,获得共同标记场 $\hat{F}(t)$ 。

(4) 对于共同标记场中每一个像素,分别用两个不同的标记0,1测试,使得系统局部能量耗费函数小的标记值得到保留,进而获得新的标记场。

(5) 重复步骤(4),直到满足算法终止标准。

5 实验结果与讨论

根据以上理论分析和算法描述,本文对 Ping 以及 Car 图像序列分别作了测试, Ping 图像序列是一个 176×144 像素大小的彩色图像序列,其图像复杂度较低(经色彩聚类算法之后,该序列图像中的色类数目在20个左右),该图像序列中每一帧图像的运动矢量均大于5像素/帧,图3(a),3(b)及3(c)分别为 Ping 的第000帧,003帧以及006帧的原图; Car 图像序列是一个 356×288 像素大小的彩色图像序列,其图像复杂度较高(经色彩聚类算法之后,该序列图像中的色彩数目在250个左右),该图像序列中的每一帧图像的运动矢量均大于

16像素/帧,图4(a),4(b)及4(c)分别为 Car 的第04帧,05帧以及06帧的原图。

实验中,参数 β 及 β_n 的取值对实验结果将产生重要影响。参数 β 及 β_n 的取值牵涉到两方面的问题。

(1) β 与 β_n 取值大小反映了对先验模型与条件模型的信任度;(2) β / β_n 的数值反映了对运动一致性与色彩一致性条件同时成立或任一条件成立的信任度。这样,参数 β 及 β_n 可以用这样的形式来表示: β 以及 $\beta \times \chi$ ($\beta > 0$ 且 $\chi > 0$),只要确定了参数 β 及 χ ,便可以相应地确定参数 β_n 。

对于参数 χ ,其取值大小反映了对运动一致性与色彩一致性条件同时成立或任一条件成立的信任度比。我们按经验值设定为0.22。实验结果证明,利用该经验值可以获得比较好的分割结果。

而对于参数 β ,其取值大小反映了对先验模型与条件模型的信任度比。我们采用参数训练的方法获得。首先,假定参数 β 取值为3.5,并将式(13)中的能量函数修改成如下形式:

$$U(f, i) = U_1(f) + \alpha U_2(f, i) \quad (25)$$

式(25)中的参数 α 为信息融合参数,表征条件模型提供的信

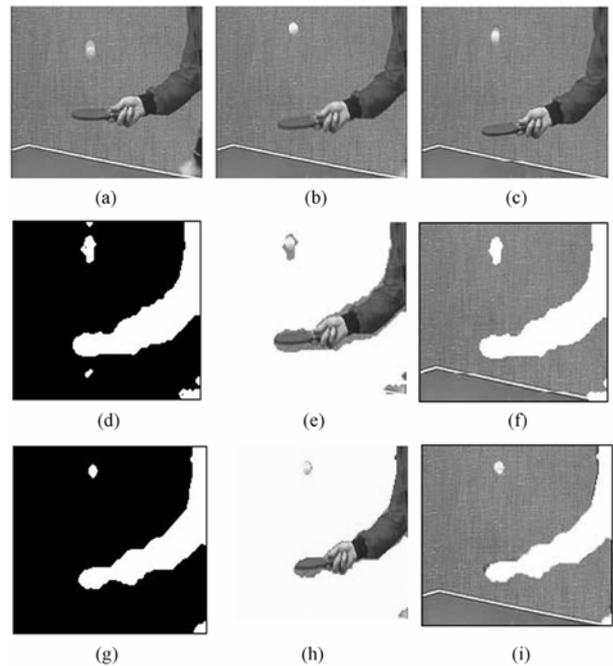


图3 Ping 图像测试序列的分割结果

(a)第000帧原始图像 (b)第003帧原始图像 (c)第006帧原始图像
(d)传统模型下的分割模板 (e)传统模型下的运动目标
(f)传统模型下的静止背景 (g)新模型下的分割模板
(h)新模型下的运动目标 (i)新模型下的静止背景

Fig.3 the segmentation result of Ping image sequence

(a)original image of frame 000, (b)original image of frame 003, (c)original image of frame 006, (d)the segment template by conventional MRF model, (e)the moving object by conventional MRF model, (f)the static background by conventional MRF model, (g)the segment template by novel MRF method (h)the moving object by novel MRF method, (i)the static background by novel MRF method

息量与先验模型提供的信息量之比。这样，先验模型与条件模型的信任度比值参数则以信息融合参数 α 的形式体现。 α 选取应以最小误分割作为判决标准。误分割的情况有两种：将背景区域划分到运动区域中被称为过分割；将运动区域划分到背景区域中称为欠分割。我们采用 Car 序列测试图像中的第 04 帧，第 05 帧以及第 06 帧作为测试图像，提取第 05 帧中运动目标的分割模板。并将运动分割模板与手工提取的标准分割模板作比较，得到图 5 所示的分割性能参数图。由图 5 不难发现：随着参数 α 的增大，条件概率模型在整个模型中的作用增大，先验模型在整个模型中的作用减小，导致分割模板中过分割现象与欠分割现象均显著减弱；而当参数 α 增加到一定程度时，欠分割的现象不再减弱，相反，过分割现象开始加剧。

通过以上分析，以最小错误分割为准则，取信息融合参数为 3。对信息融合参数 α 进行归一化，得参数 β 以及 β_n 的取值分别为 1.2，0.26。

利用上述参数，本文对两个测试序列图像进行了实验，分割结果如图 3(g)~3(i)，图 4(g)~4(i)所示。为了便于比较，同时给出了传统的时空马尔可夫随机场模型的分割结果，如图 3(d)~3(f)，图 4(d)~4(f)所示。需要说明的是：由于 Car 序

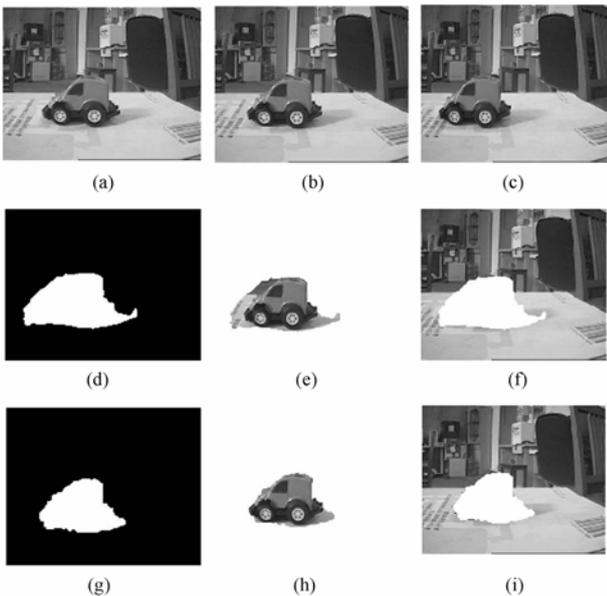


图 4 Car 图像测试序列的分割结果

(a)第 04 帧原始图像 (b)第 05 帧原始图像 (c)第 06 帧原始图像
(d)传统模型下的运动目标 (e)传统模型下的运动目标 (f)传统模型下的静止背景 (g)新模型下的分割模板 (h)新模型下的运动目标 (i)新模型下的静止背景

Fig.4 the segmentation result of Car image sequence

(a)original image of frame 04, (b)original image of frame 05, (c)original image of frame 06, (d)the segment template by conventional MRF model, (e)the moving object by conventional MRF model, (f)the static background by conventional MRF model, (g)the segment template by novel MRF method, (h)the moving object by novel MRF method, (i)the static background by novel MRF method

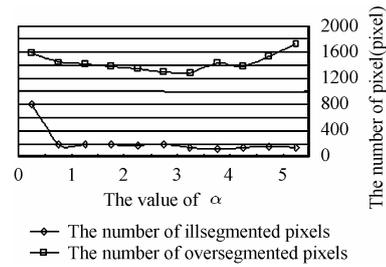


图 5 欠分割像素个数与过分割像素个数——信息融合参数关系图
Fig.5 the relationship picture for the parameter of information fusion ——the number of ill-segmented pixels and that of over-segmented pixels

列图像中的运动目标内部存在的色彩一致性区域使得初始分割模板存在一些空洞，本文采用形态滤波的方法对分割模板进行修正，填充了空洞。显然，本文提出的马尔可夫随机场模型有效地消除了运动导致的显露遮挡现象，从而获得更加逼近于真实运动目标的分割模板。然而， β 参数以及 β_n 的取值应对大量的图像进行处理，针对分割结果进行参数评估，并最终获得最佳的参数设定。在今后的实验中，我们将对参数设定的方法以及模式加以改进，以期获得更好的分割结果。

参 考 文 献

- [1] Kim Munchurl, Choi Jae Gark, Kim Daehee, et al.. A VOP generation tool: Automatic segmentation of moving objects in image sequences based on spatiotemporal information[J]. *IEEE Trans. on Circuits and Systems for Video Technology*, 1999, 9(8): 1216 – 1226.
- [2] Fan Jianping, Yu Jun, Gen Fujita, et al.. Spatiotemporal segmentation for compact video representation[J]. *Signal Processing: Image Communication*, 2001, 16(6): 553 – 566.
- [3] Meier T, Ngan K N. Automatic segmentation of moving objects for video object plane[J]. *IEEE Trans. on Circuits and Systems for Video Technology*, 1998, 8(5): 525 – 538.
- [4] 詹颂峰, 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. *通信学报*, 2000, 21(11): 120 – 126.
- [5] Luthon F, Caplier A, Lievin M. Spatiotemporal MRF approach to video segmentation: application to motion detection and lip segmentation, *Signal Processing*, 1999, 76(1): 61 – 80.
- [6] Park Sang Ho, Yun Dong, Lee Sang U K. Color image segmentation based on 3-D clustering: Morphological approach. *Pattern Recognition*, 1998, 31(8): 1061 – 1076.

黄贤武：男，1941 年生，教授，博士生导师，研究方向为图像处理，模式识别。
朱 莉：女，1978 年生，硕士生，研究方向为图像处理、模式识别。
仲兴荣：男，1970 年生，博士生，研究方向为图像处理、模式识别。
王加俊：男，1969 年生，副教授，博士，研究方向为图像处理、模式识别。