# 基于时空标记场最大后验概率的多视频对象分割算法

包红强[①] 张兆扬[②]
[①](上海大学通信与信息工程学院 上海 200072)
[②](新型显示技术及系统应用教育部重点实验室 上海 200072)

**摘 要** 该文提出了一种基于贝叶斯框架的时空标记场最大后验概率的多视频对象分割算法，根据视频序列帧间(时间域)和帧内(空间域)信息的不同特点，建立基于多个对象分割标记场的最大后验概率公式，并导出其最小能量函数，通过求解最小能量使其分割标记的后验概率达到最大。最小能量的优化求解用迭代条件模式(ICM) 方法，初始分割标记场用矢量直方图法得到。实验结果表明，该文提出的算法对存在局部遮挡的多运动对象分割是有效的。

**关键词** 图像处理，贝叶斯方法，多视频对象，时空分割
**中图分类号**: TN919.8 **文献标识码:** A **文章编号**: 1009-5896(2006)02-0232-05

# Multiple Video Object Segmentation Based on Maximization of the *A Posteriori* Probability of Spatio-Temporal Label Field

Bao Hong-qiang[①] Zhang Zhao-yang[②]
[①](*School of Communication and Information Engineering, Shanghai University, Shanghai* 200072,*China*)
[②](*Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai* 200072,*China*)

**Abstract** This paper presents a novel multiple object segmentation algorithm based on a Bayesian framework. According to the characteristic of the intra-frame and inter-frame (spatial and temporal) information, a representation of Maximization of the *A posteriori* Probability(MAP) of spatio-temporal label field is proposed. So a minimization of energy function is obtained. The optimization of solution is carried out by Iterated Conditional Mode(ICM) method. The initial segmentation label fields is gotten using vector histogram. The experimental results show that the algorithm is effective to multiple object segmentation with partial occlusion.

**Key words** Image processing, Bayesian methodS, Multiple video object, Spatio-temporal segmentation

## 1 引言

基于对象的技术在视频数据压缩(MPEG-4)、查询和检索(MPEG-7)、计算机视觉、智能监控等方面有着广泛的应用价值，但由机器自动判断并从复杂的视频场景中分离出具有语义的视频对象仍是一个引人注目的难题。近年来，视频对象分割技术已成为这一领域内科技工作者重点关注和研究的问题。目前，利用不同的处理技术提出了多种视频对象分割算法[1]，其中在某些场合下特殊的序列中提取视频对象已取得了一些成果。

在已经提出的方法中，大多是以单视频对象序列作为研究对象[2]。而在序列中有多个视频对象出现时，分割过程变得相当复杂而困难[3]。而现有的能有效分割多个复杂对象的算法多是半自动算法[4]，不利于向实用化方向发展。在本文中，我们提出了一种基于贝叶斯估计的多视频对象分割算法。算法流程图如图 1 所示。算法首先进行初始运动估计，并对运动矢量按照相似和相近的原则进行合并以得到初始的运动对象，对初始运动对象进行标记并用不同的时间和空间特征描述。由于同一对象的时空特征有一定的联系，可以通过建立贝叶斯分布模型，求得对象分割标记场的最大后验概率，为了方便地得到条件概率的解，利用马尔可夫场与吉布斯分布的等价性，代入吉布斯分布模型，用优化方法求取最小能量。下面详细给出该算法。本文第 2 节是多运动对象的基本描述；第 3 节是多个对象的初始分割；第 4 节给出贝

2004-08-02 收到, 2005-01-08 改回
国家自然科学基金(60172020)和上海市重点学科建设项目(2001－44)资助课题

叶斯模型及分割场的最大后验概率；第 5 节和第 6 节分别给出实验结果和结论。

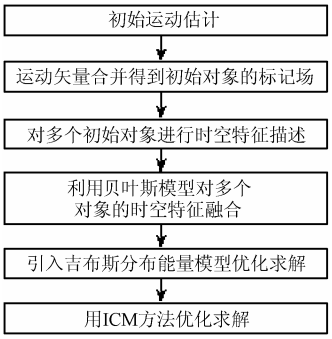


图 1　多个对象分割流程图

## 2　多运动对象的基本描述

多个运动对象在视频场景中出现时，具有不同的时空特征，通过这些特征，我们可以把它们区分开来。

### 2.1 多对象的分割标记

我们首先对不同的运动对象及背景都分配互不相同的标记。对整幅图像来说标记可视为一种基于对象时空特性的标记场，用 $L$ 表示，每一个像素点的标记值用 $l(m,n)$ 表示。从而多个对象的分割问题可转化为在一定的时空特征下，求取其最大的标记场概率分布。

### 2.2 对象的运动模型

对象的运动特征可以用参数模型来进行描述。不同的参数模型描述对象的准确度不同，参数越多，准确度越高，但相应的计算复杂度也越高。为了同时兼顾准确度和复杂度，我们用六参数仿射模型来描述对象的运动：

$$\left.\begin{aligned}\boldsymbol{u}(m,n)&=a_1\cdot m+a_2\cdot n+a_3\\ \boldsymbol{v}(m,n)&=a_4\cdot m+a_5\cdot n+a_6\end{aligned}\right\} \tag{1}$$

式中$(m,n)$为象素点的位置，$\{\boldsymbol{u}(m,n),\ \boldsymbol{v}(m,n)\}$为估计的运动矢量，$a_1,a_2,\cdots,a_6$为运动模型参数。从上式可以看出，只要得到运动模型的参数，就可以计算出运动矢量。理论上说，只需3个背景像素点的运动矢量就可求得式(1)中的6个参数$(a_1,a_2,\cdots,a_6)$。但一般的运动矢量估计方法只能得到精度有限的运动量，如果仅用很少几个像素的运动矢量来估计全局运动模型参数，得到的结果精度非常低，因此需要更多像素的运动矢量参与估计来提高准确性。这样，参数估计就成了一个解方程组的问题，可用最小二乘来求解。

如果已确定运动矢量场和参与估计的所有像素的集合$A$，就可计算出运动模型参数。运动矢量场可用块匹配法得到，块搜索过程用全搜索法或一些优化的搜索法。参与估计的像素集合 $A$ 可看作是某一个对象的标记区域。

### 2.3 对象的颜色特征

根据对象的定义，每一个对象都有相对固定的颜色，我们用对象所在的区域均值来描述其颜色特征。本文中，我们使用 *YUV* 分量代替 RGB 颜色空间。

$$\left.\begin{aligned}G_Y^A&=\sum_{(m,n)\in A}Y(m,n)\\ G_U^A&=\sum_{(m,n)\in A}U(m,n)\\ G_V^A&=\sum_{(m,n)\in A}V(m,n)\end{aligned}\right\} \tag{2}$$

对象区域仍用 $A$ 来描述，对象标记区域均值可用下式描述：

$$G(A)=\sum_{(m,n)\in A}\omega_1G_Y(m,n)+\omega_2G_U(m,n)+\omega_3G_V(m,n) \tag{3}$$

式中 $\omega_1,\omega_2,\omega_3$ 是 3 个颜色分量的加权系数，一般分别取为 1, 0.5, 0.5。为了与前面的运动矢量场和标量场对应，可定义一颜色均值场，用 $g(m,n)$ 表示，在区域 $A$ 内的颜色均值是相同的，可取某个适当的数值。

以上我们定义了描述多对象的 3 个场，标记场, 运动场，以及颜色均值场。标记场描述了多个对象的区域属性，而运动场和均值场描述对象的本身特征。从而使多对象的分割问题转化为通过运动场和均值场来求标记场的最大后验概率问题。

## 3　初始对象区域划分

初始对象区域的划分通常可使用运动聚类技术来实现，但由于无法自动确定类的数目，需要反复迭代试验才能实现。为此，我们提出了一种矢量合并的方法来确定初始的对象区域。首先对序列当前帧通过块匹配的方法进行运动估计，得到每个像素的运动矢量$\{u(m,n),v(m,n)\}$，具有相同矢量且相互连通的像素点被暂时合并同一类，每类的初始标记用 $l_i$ 表示，总像素数用 $N_i$ 表示。然后按下式进行类的合并：

$$l(m,n)=\begin{cases}l_i, & N_I\ge T\\ l_j, & \text{其他}\end{cases} \tag{4}$$

式中 $T$ 为预定义的门限(可取为 200)，$l_i$ 为当前像素初始类的标记，$l_j$ 为其相邻类的标记。上式表明，如果某一类像素的数量超过门限时，这些像素被确定为一个初始类，否则按相邻相近的原则划分到其它类中。通过以上合并，可以使类的数目显著减少，从而更为准确地反应视频场景中的初始区域。

## 4 多对象分割的贝叶斯模型

假设给定两帧 $f_k$ (当前帧)和 $f_{k-1}$ (搜索帧)，根据我们在上节定义的运动场($V$)，颜色场($G$)，以及它们与最后的分割标记场($L$)之间的关联性，我们希望在一定的运动矢量场和颜色均值场的条件下求取计算分割标记场的最大后验概率密度。利用贝叶斯公式，其后验概率密度函数可表示为

$$p(L/V,G,f_k,f_{k-1}) = \frac{p(V/G,L,f_k,f_{k-1})p(G/L,f_k,f_{k-1})p(L/f_k,f_{k-1})}{p(V,G/f_k,f_{k-1})} \quad (5)$$

上式中，分母 $p(V,G/f_k,f_{k-1})$ 相对于其它标记场来说是不变的常数，因此其最大后验概率的标记场可表示为

$$\hat{L}=\arg\max_{V,G,L} p(V/G,L,f_k,f_{k-1})p(G/L,f_k,f_{k-1})p(L/f_k,f_{k-1}) \quad (6)$$

上式右边第 1 个 $p$ 为条件概率密度函数提供了在第 $k-1$ 帧情况下，当前位移和分割估计与第 $k$ 帧相一致的程度。

在描述图像条件概率密度函数中，马尔可夫随机场(Markov Random Field，MRF)模型提供了一种有力的工具[5]。在MRF中，随机场中某个点的局部条件概率密度仅仅依赖于该点领域内的随机场分布，而且MRF可以将局部相关性加以传播，MRF模型还可有效地描述图像的许多属性。但是MRF一般都是通过条件分布或联合分布来表达的，这带来了计算上的不便。根据Gibbs分布与MFR之间等价性的理论，可以用Gibbs分布来描述MRF。因此上式中的条件概率可以用Gibbs分布模型可表示成

$$p(V/G,L,f_k f_{k-1}) = \frac{1}{Q_1}\exp\{-U_1(V/G,L,f_k,f_{k-1})\} \quad (7)$$

其中 $Q_1$ 是分割函数(常量)，且

$$U_1(V/L,f_k,f_{k-1}) = \mu\sum_{(m,n)}[f_k(m,n)-f_{k-1}(m+u(m,n),n+v(m,n))]^2 \quad (8)$$

称为 Gibbs 势能。式(6)中第 2 个 $p$ 是在给定运动分割和搜索帧的条件下，运动场的条件概率密度函数，忽略其对搜索帧的依赖性，用 Gibbs 分布作为模型，可表示为

$$p(G/L,f_k,f_{k-1}) = p(G/L,f_k) = \frac{1}{Q_2}\exp\{-U_2(G/L,f_k)\} \quad (9)$$

其中 $Q_2$ 是常量，$U_2(\cdot)$ 由下式给出:

$$U_2(G/L,f_k) = \alpha\sum_{(m,n)}[f_k(m,n)-g_k(m,n)]^2 + \beta\sum_{(m,n)}\sum_{(i,j)\in N_{m,n}}[f_k(m,n)-f_k(i,j)]^2\delta[l(m,n)-l(i,j)] \quad (10)$$

式中 $N_{m,n}$ 是像素点(m,n)的邻域，$\delta(\cdot)$ 是冲激函数。

式(6)中第 3 个 $p$ 是分割场的先验概率模型，忽略其对当前帧和搜索帧的依赖性，为了鼓励形成连续的区域，用 Gibbs 分布模型可表示为

$$p(L/f_k,f_{k-1}) = p(L) = \frac{1}{Q_3}\exp\{-U_3(L)\} \quad (11)$$

$Q_3$ 是常量，$U_3$ 由下式给出：

$$U_3(L) = \gamma\sum_{(m,n)}\sum_{(i,j)\in N_{m,n}} V_2(l(m,n),l(i,j)] \quad (12)$$

$\gamma$ 是能量控制项。上式中二阶邻域势能能量为

$$V_2[l(m,n),l(i,j)] = \begin{cases} -1, & l(m,n)=l(i,j) \\ 1, & \text{其他} \end{cases} \quad (13)$$

基于以上分析，其后验概率密度函数可重新表示为

$$p(L/V,G,f_k,f_{k-1}) \propto \exp\{-U_1(V/L,f_k,f_{k-1})-U_2(G/L,f_k)-U_3(L)\} \quad (14)$$

把式(8)，式(10)，式(12)代入上式，最大化后验概率密度等价于最小化如下的能量函数：

$$E_L = \mu\sum_{(m,n)}\Big\{[f_k(m,n)-f_{k-1}(m+\overline{u(m,n)},n+\overline{v(m,n)}]^2 + \alpha[f_k(m,n)-g_k(m,n)]^2 + \beta\sum_{(i,j)\in N_{m,n}}[f_k(m,n)-f_k(i,j)]\delta[l(m,n)-l(i,j)] + \gamma\sum_{(i,j)\in N_{m,n}}V_2[l(m,n),l(i,j)]\Big\} \quad (15)$$

可以看出，式(15)既考虑了帧间信息的作用，同时也考虑了帧间信息的约束作用和区域平滑作用，而时间信息和空间信息不同约束作用的权重可通过相关参数的改变来实现。

式(15)的求解可用全局优化算法，在本文中，采用迭代条件模型(ICM)来近似求解势能函数的最小值。在迭代条件模型中，随机场状态总是沿着势能减少的方向迭代，也就是说，使势能增加的状态被接受的概率总是为零。

## 5 实验结果

根据以上提出的时空能量模型的多对象分割方法，我们对一些常见的多对象视频测试序列进行了测试。根据式(15)计算能量时，需要对 $\mu,\alpha,\beta$ 和 $\gamma$ 四个参数进行估计。4 个参数的大小分别反映了时间能量、空间能量、空间约束在总能量中所占的比重。由于一般来说，空间分割比时间分割准确度要高，因此，当视频中多个对象的颜色均值对比明显时，我们以空间分割为主，$\alpha,\beta$ 的值可取得比较大，$\mu$ 取值较小。而各个对象的颜色均值对比不明显时，对象间的颜色差别不大，空间分割准确度较差，分割以运动特征为主，$\alpha,\beta$ 取较

小值，$\mu$ 取较大值。$\gamma$ 取值的大小反映相同标记像素之间的平滑约束作用，取值相对较为固定。实验中所用的参数如表 1 所示。

表 1　实验参数

| | $\mu$ | $\alpha$ | $\beta$ | $\gamma$ |
|---|---|---|---|---|
| Table tennis | 0.5 | 2 | 0.5 | 2 |
| Hamburg Taxi | 2 | 0.5 | 0.5 | 2 |
| Coastguard | 5 | 0.5 | 0.5 | 3 |

### 5.1 多视频对象分割的结果

当视频序列中存在多个视频对象时，利用本文方法能得到较好的分割结果。 图 2 显示了 Table tennis 序列第 1 帧、Hamburg Taxi 序列第 10 帧以及 Coastguard 序列第 45 帧的分割过程和结果。

从图中可以看出，提出的算法对多种情况下的视频对象分割是适用的。Table tennis 序列中的 2 个对象运动较为剧烈，而 Hamburg Taxi 序列中 3 个运动对象运动较为缓慢。Coastguard 序列是一个具有全局运动的序列，不仅序列中的对象在运动，而且背景也在运动，由于我们的分割方法把背景也当作一个对象来对待，故背景的运动对最后的分割结果没有影响。在实验中我们专门选用序列中两个对象发生连接并出现部分遮挡时进行分割，从结果来看，算法对这种复杂情况也是适用的，两个对象都能得到较好地分离分割，如图 2 的($a_3$)、 ($c_3^1$) 和($c_3^2$)。从实验中我们也可以看到，由于本算法过程尚较复杂，分割 1 帧约需 4 – 6s(PⅢ 800MHz的微机)，这与实时要求还有一定距离，这也是我们下一步努力的方向。

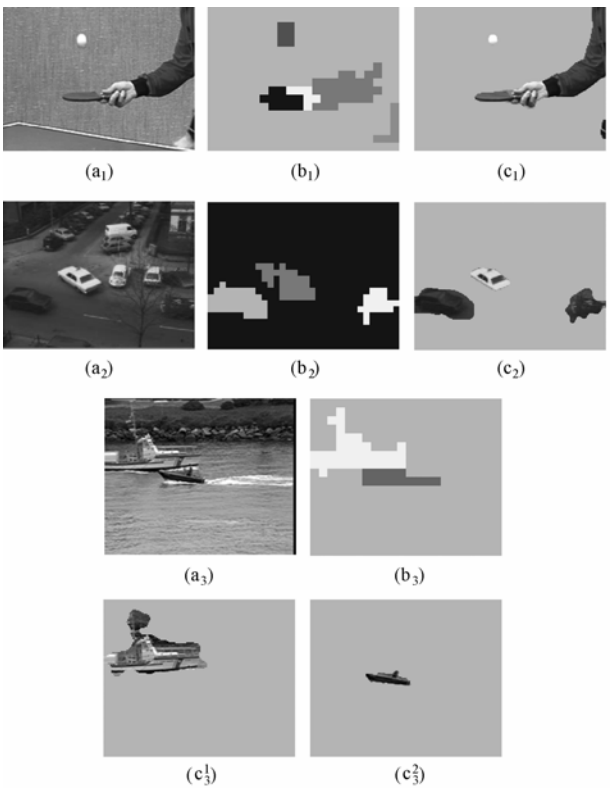


图 2　多个对象分割结果图示

### 5.2 分割结果的评价比较

为了评价视频分割算法的分割结果，需要一种客观的评价标准。在MPEG-4 核心实验中, Wollborn提出了一种存在参考对象模板的评价方法[6]，认为分割误差的产生有错分和漏分两种情况，如果事先已知一个准确的参考模板，这两种情况下的误差概率可统一定义为

$$d(A_t^{\text{est}}, A_t^{\text{ref}}) = \frac{\sum_{(x,y)} A_t^{\text{est}}(x,y) \oplus A_t^{\text{ref}}(x,y)}{\sum_{(x,y)} A_t^{\text{ref}}(x,y)} \tag{16}$$

其中 $A_t^{\text{est}}, A_t^{\text{ref}}$ 分别为第 $t$ 帧的分割对象模板和参考分割模板，$\oplus$ 为逻辑“异或”操作。可采用它来评价视频分割算法的空间准确度(Spatial Accuracy, SA), 即

$$\text{SA}(t) = 1 - d(A_t^{\text{est}}, A_t^{\text{ref}}) \tag{17}$$

视频分割算法评价的另一方面是时间一致性(Temporal Coherency, TC)。这实际上是衡量视频序列每一帧的空间准确度的变化程度。因此，可以通过计算连续帧间的空间准确度来评价。其定义如下：

$$\text{TC}(t) = 1 - \left|\text{SA}(t) - \text{SA}(t-1)\right|, \quad t = 2,3,\cdots,n \tag{18}$$

要进行这两种评价，必须先得到一个准确的参考模板，MPEG-4 并没有提供相关序列的参考模板，而国内外有关的学术组织也未见公开提供，因此，我们用常用的图像处理软件PHOTOSHOP 7.0 手动提取了一些帧的参考分割模板。COST211 是欧洲专门从事视频分析的论坛和研究机构，对视频对象的分割算法研究起步较早，在其公开网站上[7]，给出了一些分割结果，为了验证本文算法的有效性，我们对Table tennis序列的前 30 帧作了对比分析与其进行了比较，结果如图 3 所示：

图 3 (a)表示从式(17)得出的空间准确度，横轴表示序列帧的序号，从图中可以看出，提出的算法空间准确度较高，约为 80%，而 COST211 算法由于分割的对象不完整，准确度较低。图 3(b)表示分割结果的时间一致性，而横轴表示序列帧的序号，从图中可以看出，在时间一致性方面，提出的算法分割稳定性较好，而 COST211 算法受偶然因素影响较大，分割结果稳定性较差。

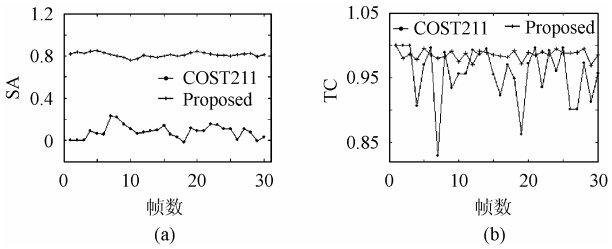


图 3 Table tennis 序列与 COST211 分割评价的对比

## 6 结束语

基于贝叶斯框架的分析方法在许多领域都有广泛的应用，在视频对象的分割提取技术中，也经常能见到有关的文献报告，但大多数方法只是利用了其部分特征，造成对多视频对象的分割效果有限。本文从多对象的基本特征出发，定义了分割标记场、矢量运动场、颜色均值场等描述多对象时空特征的量，通过贝叶斯公式推导出分割标记场的最大后验概率的数学表达式，根据其物理含义对表达式简化后，通过 Gibbs 场转化为求取最小能量函数问题，并采用 ICM 优化方法求解。最后可得到多视频对象的分割结果，从实验结果来看，本文提出的算法对视频序列中具有复杂运动的多对象分割是适用的。同时应指出，由于多对象分割比通常的单对象分割要复杂得多，至今未见有实时性分割的算法，本文的方法也离实时性有较大距离，对此，拟在下一步研究中作进一步探索。

## 参 考 文 献

[1] Zhang D S, Lu G J. Segmentation of moving object in image sequence: a review[J]. *Circuits Systems Signal Processing*, 2001－02, 20(2): 143－183.

[2] Michael M, Murat Tekalp. Simultaneous motion estimation and segmenation[J]. *IEEE Trans. on Circuits and Systems for Video Technology*, 1997, 6(9): 1326－1333.

[3] Tsaig Yaakov, Averbuch Amir. Automatic segmentation of moving objects in video sequences: A region labeling approach[J]. *IEEE Trans. on Circuits and System for Video Technology*, 2002, 12(7): 597－612.

[4] Masouri A－R, Konrad J. Multiple motion segmentation with level sets[J]. *IEEE Trans. on Image Processing*, 2002, 12(4): 1－19.

[5] Tekalp A M. Digital Video Processing[M]. Beijing: Tsinghua Press, 1998: 139－136.

[6] Wollborn M, Mech R. Refined procedure for objective evaluation of video object segmentation algorithms [R], Doc. ISO/IEC JTC1/SC29/WG11 M3448, March 1998.

[7] COST211 AM, Working site for sequences and algorithms exchange. http://www.tele.ucl.ac.be/exchange

包红强： 男, 1970 年生, 博士生，研究方向为数字图像压缩、视频处理等，已发表论文数篇.

张兆扬： 男, 1938 年生, 教授，博士生导师，研究领域包括数字图像处理、视频信息压缩编码、处理、传输和多媒体通信等，发表论文 100 余篇，著书 8 部.