

# 视频对象分割中基于 Gibbs 随机场模型的 时空分割结合方法

陈韩锋, 戚飞虎

(上海交通大学计算机科学与工程系, 上海 200030)

**摘 要:** 本文提出了一种基于 Gibbs 随机场模型的时空分割结合方法, 用于视频对象的分割. 该方法为每一帧图像的分割模板建立 Gibbs 随机场模型, 将时间域分割结果作为初始标记场, 空间域的分割结果作为一个图像观察场, 然后利用 Gibbs 模型的约束条件将二者结合起来, 得到该帧最后的分割标记场. 实验结果表明, 这种时空结合方法可以较好地避免以往的比重法过分依赖于空间域分割精度的问题.

**关键词:** 视频对象分割; Gibbs 随机场; 时空联合分割; 迭代条件模型

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112 (2004) 01-0034-04

## Integration of Spatial and Temporal Segmentation Based on Gibbs Random Field Model in Video Object Segmentation

CHEN Han-feng, QI Fei-hu

(Dept. of Computer Science & Engineer, Shanghai Jiaotong University, Shanghai 200030, China)

**Abstract:** An algorithm based on Gibbs random field model for the integration of spatial and temporal segmentation results is proposed in this paper for video object segmentation. The segmentation mask of each frame is modeled as a Gibbs random field. The field is initialized as the temporal segmentation result and the spatial segmentation is provided as an observed field of the image. Finally, the mask is gotten through integrating spatial and temporal segmentation results according to the restriction of Gibbs random field model. Experimental results show that the novel algorithm can avoid the overfull dependence on the accuracy of the spatial segmentation as the usual Ratio Method does.

**Key words:** video object segmentation; Gibbs random field (GRF); spatiotemporal segmentation; iterated conditional modes

### 1 引言

视频对象分割是指从一段视频中提取出所需的语义对象. 具体地说, 就是将每帧视频图像分割成若干个视频对象平面(VOPs). 就目前的应用需求和技术可行性而言, 大多数的研究对象都集中在从背景中分割出运动目标. 视频对象分割技术是新一代视频编码标准 MPEG-4 的一个重要支撑技术, 有着很广的应用范围(如多媒体检索、目标跟踪、基于内容的视频编码等).

视频对象分割的方法很多, 但是它们基本上都有一个共同的模式<sup>[1~3]</sup>, 即同时利用当前帧的颜色一致区域分割结果(本文中简称空间域分割)和相邻帧之间的运动检测结果(本文中简称时间域分割)进行联合分析, 得到最后的分割结果(如图 1). 本文把具体采用的联合分析方法称之为时空分割结合方法. 时间域分割利用相邻帧之间的运动信息确定哪些像素属于运动目标区域, 其结果可以大致地区分开运动目标

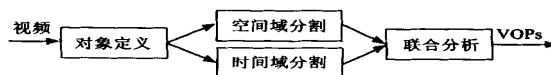


图 1 视频对象的时空分割框架

和背景区域. 空间域的分割将每帧图像分割成不同的颜色一致区域, 这样可以使目标分割边界更加精确. 这种分割框架被实践证明是有效的, 对于好的分割来说往往也是必需的. 因为简单的空间域分割在很多情况下无法引入人们对分割对象的语义定义信息, 也就很难确认人们所需的对象区域; 另一方面, 如果只是从时间域进行检测, 由于遮挡问题、孔径问题以及噪声的影响, 会造成分割结果的不准确和不完整. 具体可以采用的时间域分割方法很多, 如帧差法、光流法<sup>[4]</sup>以及半自动分割方法中的区域跟踪<sup>[5]</sup>等; 空间域的颜色一致区域分割方法也比较多<sup>[6]</sup>, 如分水岭算法、基于均值漂移算法的一致区域分割方法<sup>[7]</sup>等.

目前采用的时空分割结合方法大都采用比重法<sup>[1-3]</sup>. 假设  $F = \{f_{11}, f_{12}, f_{21}, \dots, f_{ij}, \dots, f_{MN}\}$  为时间域的目标区域检测结果, 其中  $f_{ij} \in \Gamma = \{0, 1, \dots, L-1\}$  标志了像素  $(i, j)$  属于背景或哪一个目标区域,  $\Gamma$  为标志集,  $L$  为目标区域个数; 用  $I = R_1 \cup R_2 \dots \cup R_k$  来表示空间域的颜色一致区域分割结果, 其中  $R_i$  为一个连通的区域, 所谓的比重法就是, 当某个区域  $R_i$  中有超过一定比例  $\rho$  的像素在  $F$  中被标记为  $j$  ( $j \in \Gamma$ ) 时, 则整个区域  $R_i$  的像素都被标记为  $j$ , 即属于第  $j$  个目标. 这种方法实现起来比较简单, 它是建立在这样一个假设之上: 每个目标区域都是由一些颜色一致区域拼接组成. 这个假设要求空间域的一致区域分割结果非常准确. 但是在有些情况下(如某个颜色一致区域同时跨越了目标区域和背景区域, 如图 2(a), 无法找出一个合适的  $\rho$  来获得好的分割, 如图 2(c)~(d).

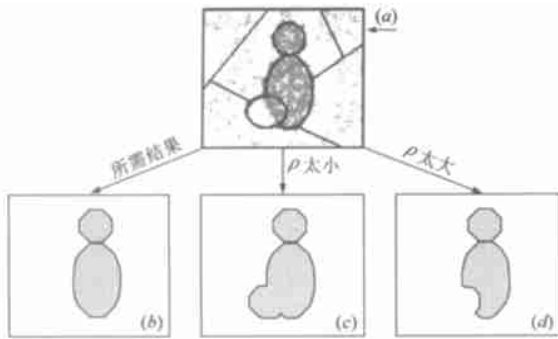


图 2 比重法的缺陷

本文提出一种基于 Gibbs 随机场模型<sup>[4]</sup>的时空分割结合方法, 可以在很大程度上弥补比重法的上述缺陷. 这种方法为图像帧的分割模板建立 Gibbs 随机场模型, 将时间域的分割结果作为初始标记场, 空间域的分割结果作为一个事先观察场, 然后利用 Gibbs 模型的约束条件将二者的有用信息结合起来, 得到最后的分割标记场(即分割模板).

## 2 Gibbs 随机场模型

本节首先简要介绍二维图像点阵的 Gibbs 随机场模型. 设  $Z = \{z(X), X \in I\}$  是一个定义在二维图像点阵  $I: M \times N$  上的随机场, 用  $\omega$  表示随机场  $Z$  的状态. 对于一个离散值的随机场来说, 随机变量  $z(X)$  的取值空间是一个离散值的集合, 即  $z(X) \in \Gamma = \{0, 1, \dots, L-1\}$ . 为了定义 Gibbs 随机场模型, 首先需要定义邻域系和基团的概念.

**定义 1**  $N_X$  是点阵  $I$  上点  $X$  的一个邻域, 是指  $N_X$  具有下列属性: (1)  $N_X \subset I$ ; (2)  $X \in N_X$ ; (3) 对于  $I$  上所有的点  $X_i$  和  $X_j$ ,  $X_i \in N_{X_j} \leftrightarrow X_j \in N_{X_i}$ .  $I$  上的邻域系则定义为:  $N_I = \{N_X, X \in I\}$ .

也就是说, 点  $X$  不属于它自己的相邻点集合; 如果  $X_i$  是  $X_j$  的相邻点, 那么  $X_j$  必是  $X_i$  的相邻点, 反之亦然. 此外, 根据邻域的半径大小, 可以定义不同阶数的邻域系.  $N_X$  是点阵  $I$  上点  $X$  的一个  $d$  阶邻域( $d = 1, 2, \dots$ ), 则:  $\forall X' \in N_X, \|X - X'\| \leq d$ , 其中  $\|\cdot\|$  表示点阵中的距离.

**定义 2**  $C$  称为定义在邻域系  $N_I$  上的一个基团, 如果  $C$

满足: (1)  $C \subset I$ ; (2)  $C$  由单个点组成, 或者  $C$  中所有点对都互为相邻. 点阵  $I$  的基团集合  $C_I$  指所有基团的集合.

随着邻域阶数  $d$  的增加, 基团数量会急剧增加. 在本文中, 选用 3 阶的邻域系, 并且只考虑包含两个像素的基团. 对于定义了邻域系  $N_I$  和与之对应的基团集合  $C_I$  的二维离散随机场  $Z$ , 其 Gibbs 分布可以描述为:

$$P(Z) = \frac{1}{Q} \sum_{\omega} e^{-U(Z)/T} \delta(Z - \omega) \quad (1)$$

其中,  $\delta(\cdot)$  表示狄拉克函数;  $T$  是所谓的温度, 用于控制分布的尖峰;  $Q$  是归一化因子, 又称配分函数, 由下式给出:

$$Q = \sum_{\omega} e^{-U(Z = \omega)/T} \quad (2)$$

$U(Z)$  表示 Gibbs 势能, 可以定义为:

$$U(Z) = \sum_{c \in C_I} V_c(Z) \quad (3)$$

$V_c(Z)$  称为基团势能, 它的值只依赖于  $z(X), X \in C$ .

可见, Gibbs 随机场的概率分布  $\rho(Z)$  是一个联合概率分布, 所有可能的状态都具有非零的概率, 且每个特定点  $X$  的局部条件概率  $z(X)$  仅仅依赖于该点的邻域内的随机场的值.  $U(Z)$  为随机场  $Z$  的势能函数,  $U(\omega)$  越小, 则随机场处于状态  $\omega$  的概率就越大.

## 3 基于 Gibbs 随机场的时空分割结合方法

Gibbs 随机场模型已经在图像处理的一些领域得到了很好的应用<sup>[8,9]</sup>. 基于 Gibbs 随机场模型的图像处理的基本方法是, 针对目标随机场提出一个势能函数, 不同的实际问题有不同的势能函数, 然后用适当的全局优化算法来求解使势能函数达到最小值的状态, 从而确定概率最大的随机场状态.

在视频对象分割中, 每帧图像  $I$  存在一个标记场  $Z = \{\omega_{11}, \omega_{12}, \omega_{21}, \dots, \omega_{ij}, \dots, \omega_{MN}\}$  和一个图像事先观察场  $D = \{d_{11}, d_{12}, d_{21}, \dots, d_{ij}, \dots, d_{MN}\}$  ( $\omega_{ij} \in \Gamma = \{0, 1, \dots, L-1\}$ ,  $L$  为待分割的目标数, 本文主要研究从背景中分割出运动目标, 此时  $L = 2$ , 即  $\omega_{ij} = 0$  或  $1$ ). 标记场用于标记每个像素属于哪个目标区域, 图像事先观察场则为当前帧的空间色彩分布信息. 如果为图像的标记场建立 Gibbs 随机场模型, 则视频分割的问题可以归结为: 在初始标记场、事先观察场和 Gibbs 模型约束条件下, 求解最大概率标记场的问题.

为了结合视频分割中时间域检测结果和空间域分割结果, 我们将时间域的检测结果作为初始标记场, 空间域分割结果作为事先观察场. 之所以将空间域分割结果而不是原图作为事先观察场, 是为了强化相似像素的空间联系. 空间域分割对时间域分割结果的约束作用将体现在 Gibbs 模型的势能函数中. 在基于 Gibbs 随机场的时空分割结合方法中, 势能函数定义如(3)式. 在本文中, 由于只考虑包含两个像素的基团, 因此(3)式又可以写成:

$$U(Z) = \sum_{c \in C_I} V_c(Z) = \sum_{(i,j) \in I} e_{ij} \quad (4)$$

其中  $e_{ij}$  为图像  $I$  中像素  $(i, j)$  处的势能,  $e_{ij}$  定义为:

$$e_{ij} = \frac{1}{2} \sum_{(m,n) \in N_{ij}} V_{ij}(m, n) \quad (5)$$

其中  $N_{\bar{y}}$  为像素  $(i, j)$  的 3 阶邻域,  $V_{\bar{y}}(m, n)$  为像素  $(i, j)$  和  $(m, n)$  间的相对势能. 为了结合时间域检测结果和空间域分割结果,  $V_{\bar{y}}(m, n)$  定义为:

$$V_{\bar{y}}(m, n) = V1(i, j, m, n) \cdot V2(i, j, m, n) + V3(i, j) \quad (6)$$

其中,

$$V1(i, j, m, n) = \begin{cases} 1 & \text{if } msa(i, j) \neq msa(m, n) \\ \alpha & \text{if } msa(i, j) = msa(m, n) \\ 1/\beta_l & \text{if } mask(i, j) = mask(m, n) = l \\ 1 & \text{if } mask(i, j) \neq mask(m, n) \end{cases} \quad (7)$$

$$V2(i, j, m, n) = \begin{cases} 1 & \text{if } mask(i, j) \neq mask(m, n) \\ 0 & \text{if } mask(i, j) = mask(m, n) \end{cases} \quad (8)$$

$$V3(i, j) = \begin{cases} 1 & \text{if } mask(i, j) \neq mask_0(i, j) \\ 0 & \text{if } mask(i, j) = mask_0(i, j) \end{cases} \quad (9)$$

其中  $msa(i, j)$  是空间域分割的结果(表明了每个像素点分别属于哪个颜色一致区域),  $mask(i, j)$  为当前标记场状态下像素  $(i, j)$  的标记值,  $mask_0(i, j)$  为初始标记场.  $V1$  通过  $\alpha$  的加权作用体现了空间域分割结果对时间域分割结果的约束作用, 参数  $\alpha$  总是大于 1, 这意味着颜色相似的相邻像素之间具有较大的相关性( $\alpha$  越大这种相关性越大). 这种约束作用是通过邻域系进行的, 它比比重法中的约束关系更为合理. 在比重法中, 所有与当前像素属于同一个一致区域的像素都对当前像素的标记发挥同等的约束作用.  $V2$  通过  $\beta_l$  体现对每个目标区域的偏重性.  $l \in \Gamma = \{0, 1, \dots, L-1\}$ , 本文的实验中只考虑  $L=2$  的情况, 因此式(8)中只需考虑  $\beta_0$  和  $\beta_1$ , 它们分别对应了背景和运动目标. 由于噪声等问题, 时间域分割结果中会将一些目标区域标记为背景区域, 这些区域通常表现为一些空洞或噪声点, 如图 3(c). 如果这些空洞或噪声点较多, 则  $\beta_1$  应选择大一些, 而  $\beta_0$  应小一些. 同样, 在背景中会有一些区域在初始标记场中会被标记为目标区域, 这种区域越多则  $\beta_0$  应选择大一些, 而  $\beta_1$  应小一些.  $V3$  体现了初始标记场(时间域分割结果)对最后分割模板的约束作用.

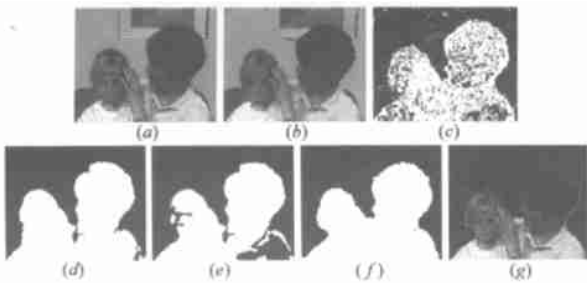


图 3 Mother&Daughter 序列的分割比较结果

#### 4 势能函数的全局优化

由式(1)可见, 当势能函数达到最小值时, 随机场处于最大概率状态. 因此, 基于 Gibbs 随机场的时空分割结合方法中的关键在于寻找使势能函数达到全局最小值的标记场状态. 势能函数的全局优化算法较多, 如模拟退火、迭代条件模型和最高置信度优化等<sup>[4]</sup>. 模拟退火算法能够找到全局最优解, 但是要使系统达到一个稳定状态, 必须模拟足够长的时间. 迭代条件模型和最高置信度优化收敛速度快, 但是容易陷入局部

最小值. 为了尽量避免陷入局部最小值, 须选择比较好的初始值. 在本文中, 由于初始标记场是时间域分割结果, 它本身已经比较接近正确的分割结果, 因此可以采用迭代条件模型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, 使势能增加的状态干扰被接受的概率总是为零.

这样, 基于 Gibbs 随机场的时空分割结合方法的过程可以归纳为以下几个步骤:

- (1) 将时间域分割结果作为初始标记场;
- (2) 对当前帧进行颜色一致区域分割, 作为事先观察场;
- (3) 根据式(5)–(9)计算像素点  $(i, j)$  被标记为不同区域  $l \in \Gamma$  时该点相应的势能  $e_{ij}$ , 使  $e_{ij}$  最小的标记得到保留;
- (4) 对整幅图像中的像素点都进行步骤 3 的操作, 并计算标记场总能量(势能函数的值)的减少量  $\Delta U$ ;
- (5) 重复步骤 3 和 4, 直到  $|\Delta U / U(Z)| < 0.1\%$ .

#### 5 实验结果与分析

针对本文提出的基于 Gibbs 随机场模型的时空分割结合方法, 我们对几段标准的视频数据进行了实验, 并与常用的比重法进行了比较. 图 3(a) 为 Mother&Daughter 视频序列的第 30 帧; 图 3(b) 为该帧的颜色一致区域分割结果; 图(c) 为时间域分割结果, 即目标区域的初始标记场; 图 3(d) 和 3(e) 是用比重法将时间域和空间域分割结果相结合得到的分割模板, 有关参数见表 1; 图 3(f) 和 3(g) 为用本文的时空分割结合方法得到的分割结果. 图 4 和 5 分别是对 Miss America 和 Hamburg Taxi 视频序列进行实验的对比结果, 每幅图的含义与图 3 相同. 通过图 3–5 中三组(c)–(f)图的比较不难看出, 利用 Gibbs 随机场模型来结合时间域和空间域的分割结果可以很好地利用两者的分割信息, 并能较好地克服比重法中由于一致区域分割不准确带来的负面影响.

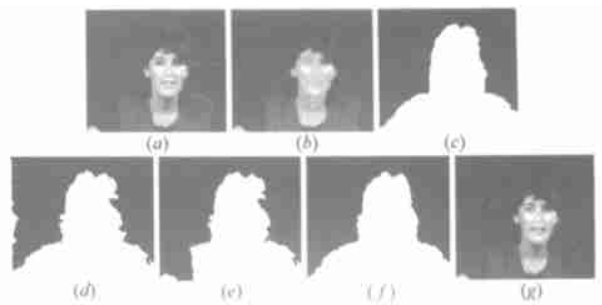


图 4 Miss America 序列的分割比较结果

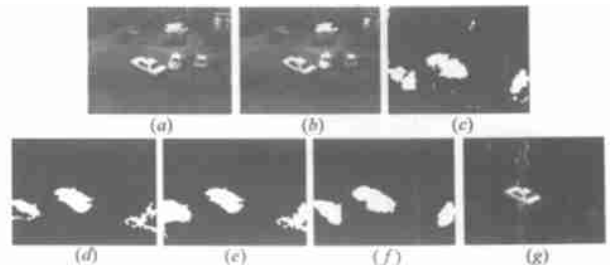


图 5 Hamburg Taxi 序列的分割比较结果

表 1 实验参数

	Gibbs 随机场模型法			比重法	
	$\alpha$	$\beta_0$	$\beta_1$	$\rho(\text{图}d)$	$\rho(\text{图}e)$
Mother& Daughter	3	1.5	3	0.3	0.4
Miss America	3	1	1	0.4	0.5
Hamburg Taxi	3	1.5	2	0.3	0.4

从图 3~ 5 可以看出, 由于空间域分割时有些一致区域有时跨越了目标和背景区域, 因此在比重法中, 当  $\rho$  太小时分割结果中目标区域中夹杂了许多背景区域, 当  $\rho$  太大时分割结果中目标区域不完整, 很难找到合适的值. 选用不同的空间域分割方法或分割参数可能可以减少这种同时跨越目标和背景的一致区域, 但是随之而来的是更多很小的小区域, 他们不利于发挥时空结合的优势. 而且正如第一节中所分析的那样, 空间域分割的不准确性在许多情况下是不可避免的.

本文中, 空间域的分割都采用了基于均值漂移算法的一致区域分割方法<sup>[7]</sup>, 而图 3、4 和 5 分别采用了三种不同的时间域分割方法: 图 3 为帧差法、图 4 为视频对象半自动分割中的区域跟踪、图 5 为运动矢量检测法(这几种方法的详细内容可以参见参考文献[4, 5]). 可见, 利用 Gibbs 随机场模型可将不同方法得到的时间域分割结果和空间域分割结果进行结合. 由于帧差法得到的时域分割结果中, 运动目标区域中掺杂的背景像素较多, 因此参数  $\beta_1$  与  $\beta_0$  的比例比较大(表 1); 而半自动分割得到的时域分割结果中, 运动目标区域最“纯净”, 因此参数  $\beta_1$  与  $\beta_0$  的比例最小; 运动矢量检测法的时域分割结果居前两者之间, 因此参数  $\beta_1$  与  $\beta_0$  的比例也居中. 这与本文第 3 节末尾分析的结果相符合.

6 结论

针对以往的比重法在结合时间域和空间域分割结果时的不足, 本文提出了一种基于 Gibbs 随机场模型的时空分割结合方法. 这种方法为图像帧的分割模板建立 Gibbs 随机场模型, 将时间域分割结果作为初始标记场, 空间域分割结果作为一个图像观察场, 然后利用 Gibbs 模型的约束条件将两者结合起来, 得到最后的分割标记场. 这种方法即能利用空间域分割的信息, 又可以较好地避免比重法在结合时间域和空间域分割结果时过分依赖于空间域分割精度的问题. Gibbs 随机场模型方法很大的一个难点在于模型中的若干参数的确定, 分割不同的视频可能需要不同的参数. 因此, 根据时间域分割结果自适应地确定参数是一个值得深入研究的内容.

参考文献:

[ 1 ] Chuang GU, Ming Chieh Lee. Semiautomatic segmentation and tracking of semantic video objects [ J ]. IEEE Trans. on Circuits And Systems For Video Technology, 1998, 8(5) : 572- 583.

[ 2 ] Jinhui PAN, Shipeng LI, Ya Qin ZHANG. Automatic extraction of moving objects using multiple features and multiple frames [ A ]. Proc. of 2000 IEEE Int. Symposium on Circuits and Systems [ C ]. Geneva, 2000, 1: 36- 39.

[ 3 ] Fur hui LONG, Da gan FENG, Har chuan PENG, Wai rchi SIU. Extracting semantic video objects [ J ]. IEEE Computer Graphics and Applications, 2001, 21( 1 ): 48- 55.

[ 4 ] A Murat Tekalp. Digital Video Processing [ M ]. Prentice Hall, 1998.

[ 5 ] 陈韩锋, 戚飞虎. 结合多种语义信息的半自动视频对象分割 [ J ]. 电子学报, 2002, 30(12A) : 2012- 2015.

[ 6 ] 章毓晋. 图像分割 [ M ]. 北京: 科学出版社, 2001.

[ 7 ] Comaniciu D, Ramesh V, Meer P. The variable bandwidth mean shift and data driven scale selection [ A ]. Proc. of 2001 IEEE Int' l Conf. Computer Vision [ C ]. Vancouver, 2001, 1. 438- 445.

[ 8 ] Jianqing LIU, Yee Hong YANG. Multiresolution color image segmentation [ J ]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1994, 16(7) : 689- 700.

[ 9 ] 赵建业, 王新, 余道衡. 基于 Gibbs 图像模型的细胞神经网络模板设计法 [ J ]. 电子学报, 2000, 28(1) : 48- 52.

作者简介:



陈韩锋 男, 1976 年生于浙江义乌, 1998 年和 2000 年在西北工业大学获得工学学士和硕士学位. 现为上海交通大学计算机科学与工程系博士研究生, 主要研究方向为图像分割, 视频分割, 运动估计和目标跟踪.



戚飞虎 男, 1938 年生于上海, 上海交通大学计算机科学与工程系博士生导师, 上海市图像图形学会常务副理事长, 目前主要研究方向为图像处理, 模式识别和视频监控等.