

# 基于贝叶斯网络的运动对象提取

黄向生<sup>1</sup> 王阳生<sup>1</sup> 杨小帆<sup>2</sup>

(中国科学院自动化所 北京100080)<sup>1</sup> (重庆大学计算机学院 重庆400044)<sup>2</sup>

**摘要** 本文分析利用颜色和运动信息做运动对象提取所存在的两个问题。针对这两个问题,本文引入场景上下文知识和经验知识来辅助提取运动对象。而贝叶斯网络(Bayesian Network, BN)是不确定性知识表达和推理的重要工具,我们利用它来表达运动对象的场景上下文知识和人们对运动对象的经验知识,并用它来推理运动对象的可能大小、位置和形状。

**关键词** 运动对象提取,贝叶斯网络,不确定性知识

## Extraction of Moving Objects Based on Bayesian Network

HUANG Xiang-Sheng<sup>1</sup> WANG Yang-Sheng<sup>1</sup> YANG Xiao-Fan<sup>2</sup>

(Institute of Automation, Chinese Academy of Sciences, Beijing 100080)<sup>1</sup>

(College of Compute Science, Chongqing University, Chongqing 400044)<sup>2</sup>

**Abstract** Two problems of moving object extraction using color and motion information are discussed in this paper. In order to solve these problems, scene context knowledge and prior knowledge of moving object are used. Bayesian network (BN) is an important tool to depict uncertainty knowledge and infer it. In this paper, BN is used to describe the scene context knowledge and the prior knowledge of moving object, and to compute the possibility of moving object's size, location, and shape.

**Keywords** Extraction of moving object, Bayesian network, Uncertainty knowledge

## 1 引言

无论是在识别领域,还是在基于对象的视频编码领域,视频运动对象的分割提取都是其中一个很重要的技术环节。已经有许多基于多特征的交互式对象提取方法,虽然这些方法比较灵活和相对精确,但是需要人来参与分割和提取,这无形中给用户增加负担,这样的方法不适合实时处理,例如视频会议系统。另外也有些自动提取的方法,而且是可以提取语义上的对象,但算法的结果却过于粗糙,以至很难适用于一般的情况。

文[1]采用基于颜色的分割,然后根据各个颜色区域的运动情况判断是否是运动对象的一部分。目前大多数运动对象提取的算法<sup>[1,2,7~9]</sup>都基于这样三个假设:1、运动对象时时刻刻都处于运动的状态;2、对象的各个部位都处于运动的状态;3、根据颜色分割能做到过度分割(Over Segmentation)。但这些假设在实际应用中很难成立。



(a) 视频



(b) 提取结果

图1 运动对象提取

图1是根据文[1]实现的结果,从图1可看出两个问题,一是主运动对象的运动是相对的,在“头-肩”这样的视频序列中,对象的某些部位经常出现短暂的静止,例如肩膀经常出现相对比较长时间的静止,导致被认为是静止的背景;二是对象的某些部位与部分背景颜色、纹理非常接近,而且这两个区域连通,导致出现欠分割(Under Segmentation),而不是过分割,例如头发与黑色的衣服。对计算机来说,很难将头发和衣服区分开,但对人来说,这是一件很容易的事情,因为人具有经验知识(常识),人知道头发与人脸的相对关系,头发不应该是图1(b)分割的那种形状,人也知道人脸下面应该有肩膀,即使肩膀是静止不动的。因此,计算机如果不根据场景的上下文知识和对象的常识就很难将与运动对象相连的背景区域从运动对象上分离出去,也很难将运动对象的静止部位(例如肩膀)提取出来,导致无法提取一个语义上完整的对象。

贝叶斯网络(Bayesian Network, BN)是不确定性知识表达和推理的有力工具,具有关系运算和数值运算的能力,它是概率论和图论相结合的产物。著名图论学家 Erdős 对概率图论做了很多开创性工作。而将贝叶斯网络作为概率型专家系统的一种图模型,首先是由 Pearl 于1986年提出的<sup>[10]</sup>。它将不确定性知识表示为变量之间的因果关系或概率依赖关系,然后进行因果推理、诊断推理、支持推理。这种做法具有语义清晰、易于理解等优点。十多年来,贝叶斯网络的理论研究一直受到国际上人工智能专家们的广泛关注,一些基于贝叶斯网络的专家系统已经被开发出来,成功地应用于故障诊断、实时决策、信息检索等领域。

贝叶斯网络的许多特点使得其在语音识别方面得到广泛的应用。最近也有不少关于贝叶斯网络在计算机视觉上的应用,例如用于解决交通视觉监控的技术问题。一方面,贝叶斯网络作为专家系统模型,可用于构建视觉系统的高层子系统;另一方面,贝叶斯网络作为一种概率模型,其完备的、易懂的语义可用于处理重叠对象的区分与跟踪。尤其是具有时间参数的动态贝叶斯网络(Dynamic Bayesian Network, DBN)可看作隐马尔可夫模型(HMM)和 Kalman 滤波器(Kalman Filters)的推广。而隐马尔可夫模型和 Kalman 滤波器正是许多视觉系统处理图像运动分割和对对象提取的主要工具。文[13]将贝叶斯网络和神经网络相结合用于静态图像的分割。文[14]将贝叶斯网络用于空间信息的集成和推理。目前,对贝叶斯网络应用的研究大多数集中于对图像的理解和解释方面,即在高层子系统的应用[15]。

本文分析基于颜色和运动信息的运动对象提取存在的问题,即欠分割问题和局部不运动问题。针对这两个问题,本文提出一种基于场景上下文知识和经验知识的运动对象提取算法,该算法是通过贝叶斯网络来组织场景知识和经验知识的。实验效果表明,基于贝叶斯网络的运动对象提取算法能较精确地提取运动对象。

## 2 贝叶斯网络

### 2.1 网络结构

贝叶斯网络由网络结构和条件概率分布两部分组成。 $B=(V, E)$ 是有向无环图,即贝叶斯网络结构; $V$ 是节点集,节点 $v_i$ 的值域记作 $val(v_i)$ ;  $E$ 是有向边集;每一条边表示节点之间的直接依赖关系,而依赖程度由条件概率决定;每一个节点 $v_i$ 都有相应于该节点的父节点集 $parent(v_i)$ 和条件概率分布表。在具体应用时,贝叶斯网络结构的构造是要根据具体问题的知识原理,将直接相互依赖的变量之间用有向线段连结,形成网络结构;贝叶斯网络的条件概率分布表可以通过专家指定或通过学习得到。

$B_i$ 是满足马尔可夫条件的,即若节点的父节点取值状态给定,那么该节点的取值将独立于所有非子孙节点。对于节点总数为 $N$ 的贝叶斯网络,网络结构决定于如下一组条件独立性假设:

$$P(v_i | v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_N) = p(v_i | parent(v_i)) \quad i=1, 2, \dots, N \quad (1)$$

由此条件独立性假设及概率的链式法则可知联合概率分布为:

$$P(V) = \prod_{i=1}^N p(v_i | v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_N) = \prod_{i=1}^N p(v_i | parent(v_i)) \quad (2)$$

若给定条件独立性假设和各节点在其父节点所有可能取值下的条件概率分布,则可以求出联合概率分布。

贝叶斯网络根据问题自身特点和变量之间的直接影响的局部性,使用条件独立性的假设,把对联合概率测度的计算或确定,分解为对若干局部条件概率测度的确定。若限定每个节点最多有 $k$ 个可能的父节点,而且每个节点最多有 $m$ 个取值状态,则确定联合概率最多只要计算或指派 $N(m^k - 1)$ 个概率,这样大大减少计算量。

贝叶斯网络的每个节点 $v_i$ 可以看成属性变量,每条边表示两个节点间直接的单向概率依赖关系,依赖程度决定于条

件概率的参数,两个节点没有边连结则表示两个节点没有直接的单向依赖关系。因此,贝叶斯网络不仅是一个图模型(Graphical Model),而且还是一种知识表达模型,基于该模型的知识表达语义清晰、易于理解。利用贝叶斯网络所确定的联合概率分布,人们可以计算出任何感兴趣事件发生的概率,进行因果推理、诊断推理和支持推理。

### 2.2 推理算法

当贝叶斯网络结构和条件概率给定时,基于贝叶斯网络的概率推理算法就是在某些网络节点 $V_i$ 的取值状态为 $e$ 情况下,计算出其余节点各取值状态的概率分布,即要计算后验概率分布 $p(V | V_i = e)$ 。此时,称 $V_i$ 为证据节点, $e$ 是证据。贝叶斯网络的精确推理的计算复杂度是很高的[12],尤其对于节点个数多的情况,不适合于精确推理。有很多近似推理算法来解决这个问题。本文采用一种基于概率传播和更新的贝叶斯网络推理算法,该算法由 Pearl 在文[10]中提出,其正确性也得到证明。为使本文内容完整,下面简要介绍该算法。

概率传播和更新是通过信息传播机制来实现的,传播证据对有关网络节点的影响,生成每个点的后验概率分布 $p(v | V_i = e)$ , $v$ 表示网络中任一节点。来自其某父节点的信息称为 $v$ 的 $\pi$ 信息,记为 $\pi_v(v_p)$ ;来自其某子节点的信息称为 $v$ 的 $\lambda$ 信息,记为 $\lambda_v(v_c)$ 。当证据产生时,节点首先接收并综合来自所有父节点和所有子节点的信息,得到该节点的总 $\pi$ 信息和总 $\lambda$ 信息。然后,将这些总信息分别传播给节点的所有父节点和所有子节点。节点 $v$ 关于证据的后验概率分布 $p(v | V_i = e)$ 就是该节点的总信息和总信息的乘积的常数倍,即 $p(v | V_i = e) = c \cdot \pi(v) \cdot \lambda(v)$ 。该算法是个递归运算过程,具体运算步骤如下:

① 初始化。定义根节点的总的 $\pi$ 信息为该节点的先验概率,每个叶节点总的 $\lambda$ 信息为1。在证据传播到来之前,对于网络中其他非证据节点,其总的 $\pi$ 信息初始化成先验概率,总的 $\lambda$ 信息初始化为1。

② 消息生成。当证据 $e$ 产生时, $e$ 所对应的节点变成证据节点,证据节点集 $V_e$ 中的每个节点的总的 $\pi$ 信息和总的 $\lambda$ 信息分别变成:当 $V_i = e$ 时, $\pi(V_i) = 1$ ;否则, $\pi(V_i) = 0$ ;当 $V_i = e$ 时, $\lambda(V_i) = 1$ ;否则, $\lambda(V_i) = 0$ 。在这轮概率传播与更新中, $V_i$ 不接收任何信息。

③ 计算各非证据节点的消息。设 $v$ 是网络中的任一节点,它有 $k$ 个父节点的和 $l$ 个子节点,如果每个父节点发送一个消息给 $v$ ,每个子节点发送一个消息给 $v$ ,则总的 $\pi$ 信息和总的 $\lambda$ 信息分别为:

$$\pi(v) = p(v | parent(v)) \prod_{i=1}^k \pi_v(v_{p_i}) \text{ 和 } \lambda(v) = \prod_{i=1}^l \lambda_v(v_{c_i})$$

④ 传播新的 $\lambda$ 消息 $\lambda_v(v_{p_i})$ 给 $v$ 的父节点 $v_{p_i}$ 。由于 $v_{p_i}$ 可能有多个子节点,因此其 $\lambda$ 计算如下:

$$\lambda_{v_{p_i}}(v) = \sum_v \lambda(v) p(v | parent(v)) \prod_{n \neq i} \pi_v(v_{p_n}) \quad i=1, 2, \dots, k$$

⑤ 传播新的 $\pi$ 消息 $\pi_v(v)$ 给 $v$ 的子节点 $v_{c_i}$ 。

$$\pi_{v_{c_i}}(v) = \pi(v) \prod_{n \neq i} \lambda_v(v_{c_n}) \quad i=1, 2, \dots, l$$

⑥ 网络中每个节点生成的后验概率分布。

$$p(v | V_i = e) = c \cdot \pi(v) \cdot \lambda(v) \quad \text{其中 } c \text{ 是归一化系数}$$

## 3 基于贝叶斯网络运动对象提取算法

### 3.1 贝叶斯网络模型

运动对象的提取就是要从视频序列中分割出语义上完整的对象,以便对象的跟踪或基于内容的视频编码。目前大多数运动对象提取算法是基于光流场计算和颜色空间的,即基于运动信息、颜色信息、纹理信息等低级特征(Low Features)。即使不考虑光照影响的情况下,基于这些信息的运动对象提取算法也很难从复杂的背景中提取语义上完整运动对象。造成运动对象提取困难的主要原因在于这些算法没有利用到语义上的知识,即场景上下文知识和经验知识。

下面以“头-肩”视频序列为例子,说明人对“头-肩”的经验知识或常识。图2是表达“头-肩”常识的贝叶斯网络。图2(a)的每个节点都有三个参数:位置、形状、大小。从图中的网络可以看出,当运行运动对象提取算法时,可以通过颜色和运动信

息获得一些证据数据,例如人脸前额或整个人脸的位置和大小,然后根据这些证据数据推理出其他的部位(如肩膀)的可能位置、形状、大小。通过贝叶斯网络推理得到的其他各个部位可能位置、形状、大小不能直接就被认为是运动对象的位置、形状、大小,还要根据图像的颜色和纹理等信息做一个综合考虑。如果根据颜色和运动信息分割得到的部位的大小、形状与推理得到的大小、形状差不多,那么就用颜色和纹理分割得到的大小、形状作为提取的结果;否则就采用通过推理得到的大小、形状。另外,在很多情况下,运动对象提取要求实时性,因此不能用太复杂的贝叶斯网络;否则,将无法实时提取运动对象。本文在做“肩-头”视频提取的时,采用图2(b)的简化贝叶斯网络。

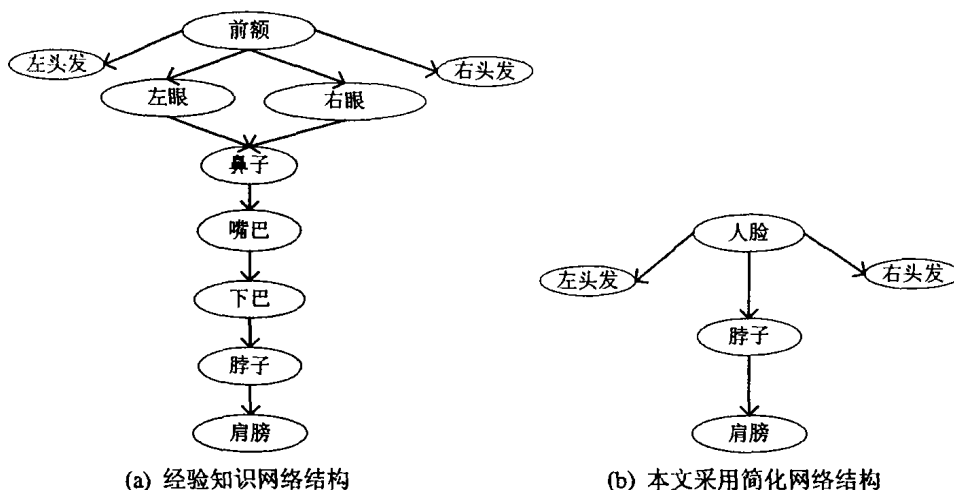


图2 “头-肩”序列经验知识的贝叶斯网络

图2的贝叶斯网络模型的节点参数和条件概率分布表可以使用统计中参数估计的方法从以往的历史数据中得到,当然条件概率表也可以直接由专家给出。

### 3.2 证据数据的获取

目前,有许多利用运动信息进行目标检测。文[9]采用绝对差分运算的方法,即采用连续三帧视频数据,相邻两帧分别求差分,然后将得到的差分灰度图像进行二值化,并将这两个二值化后的图像求“与”运算,最后得到运动对象的形状轮廓。这个方法的实时性好,但该方法要求对象连续运动,即要求每相邻两帧都有运动。这实际上是一个很强的假设,即使是对象的局部都很难每相邻两帧都有运动存在。另外,由于随机噪声的影响,很难得到一个比较完整的区域。

本文采用10帧视频数据来判断当前帧的各个像素是否属于运动对象,即采用10帧视频数据,两两分别求绝对差分,得到9帧灰度图像,选择一个适当的阈值,对这9帧灰度图像进行二值化,然后,根据这9帧二值化的图像判断当前帧内各个像素点是否属于运动对象。方法如下:

$$c(i, j) = \sum_{s=1}^9 MB_s(i, j) \quad (3)$$

$$M(i, j) = \begin{cases} 1, & \text{当 } c(i, j)/S > T_1 \\ 0, & \text{其他} \end{cases} \quad (4)$$

式中  $MB_s(i, j)$  表示第  $s$  个二值图像的像素值;  $S$  是二值图像帧数,本文取9;  $T_1$  是一个阈值,可以取50%。 $M(i, j)$  就是求得的运动对象的掩码图像,即当  $M(i, j) = 1$  时,表示像素点属于运动对象;否则属于背景。

由于各方面的原因,根据运动信息分割只能得到运动对象的一个粗糙的形状轮廓(实验结果),这种粗糙的形状轮廓

很难确认为一个运动对象的证据数据块;而颜色分割则可以提供精确的分割。因此,可以综合利用这两种分割来提高获取运动对象的证据数据,即考虑利用颜色分割各个区域,如果区域内有一半以上的像素点被认为是运动的,则这个区域可以认为是运动对象。其形式化方法如下:

$$B(R) = \sum_{(i, j) \in R} M(i, j) \quad (5)$$

$$J(R) = \begin{cases} 1, & \text{当 } B(R)/A(R) > T_2 \\ 0, & \text{其他} \end{cases} \quad (6)$$

式中  $R$  表示颜色分割得到的区域,  $A(R)$  表示颜色分割得到区域的大小,  $T_2$  是一个阈值,可以取50%。当  $J(R) = 1$  时,表示颜色区域属于运动对象。

值得注意的是,本文在这里采用基于颜色和运动信息的方法提取运动对象的证据数据与文[1]的方法类似。本文利用该方法是用来找一些运动对象的证据数据,只要能找到一些证据数据就可以,不需要提取完整的对象,因此使用这个方法的前提是运动对象有些局部运动存在,颜色分割可以是超分割,也可以是欠分割。而文[1]是想利用该方法提取完整的对象,它基于的两个前提是:运动对象是连续运动的;根据颜色分割是超分割的。

根据运动信息和颜色信息分割得到一个语义上不完整的对象,根据这些证据数据,通过2.2节的对象常识(经验知识)贝叶斯网络推理出一个合理的语义上完整的对象作为提取结果。

### 3.3 实验结果

在做颜色分割时,本文采用的是归一化颜色空间,即  $R/(R+G+B)$ ,  $G/(R+G+B)$ ,  $B/(R+G+B)$ ; 计算每个像素点的两个值(因为第三个值冗余的信息),然后,用量化这两个

值,得到如图3(b)的结果。由于噪声和纹理的影响,颜色分割也是粗糙的。因此,经过颜色分割后的图像还需要腐蚀形态处理,将各个区域内的小斑点去除掉,得到图3(c)的结果。根据2.2节提到的方法,计算运动信息,得到图3(d)的结果,结合

颜色分割和运动信息,可以得到图3(e)的运动对象的证据数据。将颜色和运动信息与对象经验知识相结合,通过贝叶斯网络推理和判断,可以得到如图3(f)相对比较精确的运动对象提取结果。



图3 基于贝叶斯网络运动对象提取实验

**总结与展望** 基于对象的视频编码将有利于视频数据的管理和消费,而运动对象提取却是基于对象的视频编码的关键技术和难点技术。目前,大多数运动对象提取算法都是基于运动信息、颜色信息、纹理信息,而这些算法很难从复杂的场景中提取语义上完整的对象。因为,这些算法都只用到图像处理的一些低级特征(Low Features),没有用图像理解或场景上下文知识和经验知识。本文针对在光照条件较好情况下的欠分割和对象局部不运动的问题,提出基于贝叶斯网络运动对象提取算法,该算法的思想是利用对象的常识和场景的上下文知识来辅助提取运动对象,即通过一些证据数据,根据常识来推理对象的其他部位的可能位置、形状、大小,从而获得语义上相对完整的对象。实验证明,该方法能够简单有效地进行对象提取。但是,由于各种对象的经验知识(常识或专家知识)不一样,因此,今后的工作在于构造各种对象的经验知识的贝叶斯网络,以便运动对象提取使用,而可能的难点在于各种不同的对象的经验知识的完备表达和条件概率表的确定。当然,由于大多数应用场合关心的运动对象是人,因此,我们的下一步工作将是构造一个更完整、更合理的贝叶斯网络来表达人这个对象的常识。另外,由于贝叶斯网络推理需要一些时间,而目前的计算机经不起贝叶斯网络的复杂的“思考”,如何提高算法的运算效率也是今后工作的方向。

### 参考文献

- Pan J, Li S, Zhang Y. Automatic Extraction of Moving Objects Using Multiple Features and Multiple Frames. Available at: <http://research.microsoft.com/asia>
- Gu C, Lee M C. Semiautomatic Segmentation and Tracking of Semantic Video Objects[J]. IEEE Transaction on circuits and systems for video technology, 1998, 8(5)
- Wang R, Zhang H, Zhang Y. A Confidence Measure Based Moving Object Extraction System Built for Compressed Domain. Available at: <http://research.microsoft.com/asia>
- Castagno R, Ebrahimi T, Kunt M. Video Segmentation Based on Multiple Features for Interactive Multimedia Applications [J]. IEEE Transaction on circuits and system for video technology, 1999, 8(5)
- Wang H, Chang S. A Highly Efficient System for Automatic Face Region Detection in MPEG Video [J]. IEEE Transaction on circuits and systems for video technology, 1997, 7(4)
- Meier T, Ngan K N. Automatic Segmentation of Moving Objects for Video Object Plane Generation [J]. IEEE Transaction on circuits and systems for video technology, 1998, 8(5)
- 罗涛. 头肩视频图像的运动物体自动提取[J]. 北京大学学报(自然科学版), 2000, 36(5)
- 贾振堂, 贺贵明, 韩艳芳. 运动视频对象分割的一种快速算法[J]. 中国图像图形学报, 2002, 7A(11): 1123~1127
- 李凌娟, 贾振堂, 贺贵明. 一种鲁棒的视频分割算法[J]. 中国图像图形学报, 2002, 7A(11): 1128~1133
- Pearl J. Fusion, Propagation, and Structuring in Belief Networks [J]. Artificial Intelligence, 1986, 29: 241~288
- Pearl J. Probabilistic Reasoning in Intelligent System: Networks of Plausible Inference [C]. Morgan Kaufman, San Mateo, California, 1988
- Cooper C. Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks (Research note) [J]. Artificial Intelligence, 1990, 42: 393~405
- Feng X, et al. Combining Belief Network and Neural Network for Scene Segmentation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2002, 24(4): 467~483
- Sarkar K L S. Integration, Inference, and Management of Spatial Information Using Bayesian Networks: Perceptual Organization [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1993, 15(3)
- Huang T, Koller D, et al. Automatic symbolic traffic scene analysis using belief networks [J]. In: Proc. of the 12<sup>th</sup> National conf. on Artificial Intelligence, Seattle, WA, 1994. 966~972