

视觉信息处理中的马尔可夫随机场

陶霖密 王奇凡 邱慧军

(清华大学计算机科学与技术系, 北京 100086)

摘要 由于视觉信息处理中存在大量的不确定性, 概率图模型在计算机视觉领域有着广泛的应用, 备受广大学者的关注。许多视觉问题都可以通过建立概率图模型进行求解, 随着高效求解算法的提出和发展, 马尔可夫随机场在解决计算机视觉领域的大规模数据问题中具有很大的优势。首先简要地介绍了概率图模型的概念, 然后对马尔可夫随机场模型的定义、特性和推导求解进行了分析和讨论, 在此基础上, 以马尔可夫随机场在视觉信息的应用为线索, 对目前基于马尔可夫随机场的计算机视觉信息处理的主要技术进行了概述和比较研究。

关键词 马尔可夫随机场 概率图模型 贝叶斯网络 信息传播 图像切割 能量最小化

中图法分类号: TP391

文献标识码: A

文章编号: 1006-8961(2009)09-1705-07

Markov Random Field in Visual Information Processing

TAO Lin-mi, WANG Qi-fan, DI Hui-jun

(Department of Computer Science and Technology, Tsinghua University, Beijing 100086)

Abstract Probabilistic graphical models (PGM) is widely applied in visual information processing for the intrinsic uncertainty in visual information, and followed by a group of researchers recently. PGM offers a number of advantages for resolving variety problems in visual information processing, in which Markov Random Field (MRF) can be used to model pixel level information processing based on the development of high efficiency inference algorithms. In this paper, we shortly introduced concepts of PGM, and gave detailed analysis and discussion on the definition, features and inference of MRF followed by typical examples of its application in computer vision.

Keywords markov random field (MRF), probabilistic graphical model (PGM), bayesian network, belief propagation (BP), graph cut (GC), energy minimization

1 引言

由于视觉信息处理中的各种不确定性, 概率方法是目前解决视觉问题的主要手段。概率方法通过刻画联合分布概率 $P(\text{隐含变量} | \text{观测变量})$ 或后验分布概率 $P(\text{隐含变量} | \text{观测变量})$ 来描述待解决的视觉问题, 包含两个主要方面: 概率模型 (即目标问题的概率描述) 和相关的推理或学习算法 (即问题求解)。

在通过概率方法描述目标问题时, 比较关心随机变量间的 (条件) 独立关系。通过有效利用这些

独立关系可以设计高效的推理算法, 从而大大简化问题的求解。概率图模型 (probabilistic graphical model, PGM) 是一类用图的形式表达随机变量之间条件独立关系的概率模型总称, 是概率论与图论的结合。视觉中常用的概率图模型有 3 类: 贝叶斯网 (bayesian network, BN)、马尔可夫随机场 (Markov random field, MRF) 和因子图。不同的图模型在分布函数的表达方式上有所不同:

(1) 贝叶斯网是由随机变量构成的有向无环图 (如图 1(a) 所示), 其中每个节点的条件概率分布为 $P(\text{当前节点} | \text{它的父节点})$, 最终的联合分布为所有节点的条件概率分布的乘积。如图 1(a) 中的

联合分布函数为 $P(X)P(Y)P(Z|X,Y)$ 。

(2) 马尔可夫随机场是由随机变量构成的无向图(如图 1(b)、图 1(c)所示),其中每个最大全连通子图的势函数定义为 g (子图的节点),最终的联合分布正比于所有势函数的乘积(即差一个常数因子)。如图 1(b)中的联合分布函数正比于 $g(X,Y,Z)$,图 1(c)中的联合分布函数正比于 $g(X,W) \times g(Y,W)g(X,Z)g(Y,Z)$ 。

(3) 因子图是由一组局部函数节点和随机变量构成的双向图(如图 1(d)所示),其中每个局部函数定义为 g (与当前函数节点相连的节点),而最终的联合分布正比于所有局部函数的乘积。如图 1(d)的联合分布函数正比于 $g(X,Y)g(X,Z)g(Y,Z)$ 。

对比图 1(a)和图 1(b),图 1(b)给出了唯一一个能表达图 1(a)中三元函数 $P(Z|X,Y)$ 的马尔可夫随机场,但是图 1(a)表达的 $X \perp Y$ (“ \perp ”表示随机变量间的独立关系)在图 1(b)中丢失了。因此,某些贝叶斯网所表达的随机变量间的独立关系是不能够通过马尔可夫随机场来实现的^[1]。反过来,图 1(c)中马尔可夫随机场所表达的 $X \perp Y | \{W,Z\}$ 和 $W \perp Z | \{X,Y\}$ 也不能够通过贝叶斯网来实现^[1]。这样,马尔可夫随机场和贝叶斯网在随机变量间独立关系的表达能力上各有千秋。对于因子图,它具有其他两种模型的所有表达能力,是 3 个模型里表达能力最强的。如图 1(a)、图 1(b)和图 1(c)中的模型都可以用因子图来表示,而图 1(d)的因子图中给出的 $g(X,Y)g(X,Z)g(Y,Z)$ 的联合分布函数形式是没有办法用任何贝叶斯网或马尔可夫随机场来表达的^[2]。

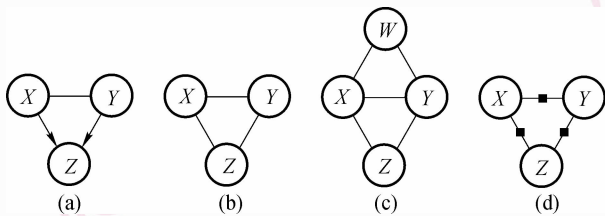


图 1 不同概率图模型的表达能力

Fig. 1 The representation of probability graphical models

这 3 种概率图模型在视觉信息处理领域具有各自不同的应用范围。贝叶斯网最适合于表达因果关系或前一时间对后一时刻的影响,所以广泛地应用于识别理解、目标跟踪等各个方面。马尔可夫随机场可用于表达相邻像素之间的关系,构成节点数量

巨大的马尔可夫随机场,其求解则依赖于 GC (graph cut)算法和 BP (belief propagation)算法。因子图是三者中表达能力最强的,许多视觉问题都可以转化为因子图来求解。但由于引入了局部函数节点,因子图的求解比较困难。

本文主要关注马尔可夫随机场,其原因是 MRF 具有相应的高效算法,是目前求解像素层次视觉问题的重要方法。和一般综述文献不同,本文主要是一种关于 MRF 在视觉信息处理中应用的简单的教程,首先给出了马尔可夫随机场的定义及其特性,并对其推导求解算法进行了介绍;然后以马尔可夫随机场模型在视觉信息应用领域为线索对其主要技术进行了概述和比较研究;最后对马尔可夫模型所存在的问题及进一步的发展进行了总结和展望。

2 MRF 模型与定义

马尔可夫随机场,又称马尔可夫网或者无向图模型,是概率图模型的一种。马尔可夫随机场模型中包含了一组具有马尔可夫性质的随机变量,这些变量之间的关系可以用无向图来表示。马尔可夫随机场模型在变量关系的依赖形式上和贝叶斯网相似,但是它可以用来表示一些具有特定依赖关系的变量集合,而贝叶斯网却无法做到(比如带环的依赖关系);另一方面,有些特殊的变量依赖关系贝叶斯网能表示而它却不能(比如节点的父子关系)。

马尔可夫随机场称为马尔可夫网,是由 Pearl 提出^[3-4]。和贝叶斯网不一样,马尔可夫随机场是一个无向图,可以用无向图 $G(V,E)$ 来表示。在马尔可夫随机场模型中,对应于每个节点 i 都有一个隐状态 X_i 和一个观测量 Y_i , X_i 可以是连续随机变量也可以是离散的。为了更加直观地说明 MRF 模型,本文将针对状态空间为离散的 MRF 来进行讨论。所谓离散是指每个节点的状态取值于一组离散候选状态之一,所以求解马尔可夫随机场模型往往也被看作是状态标记问题,即在候选状态中对每个节点标记一个状态,使得整个模型在给定观测下取得最大后验概率 (MAP)。

马尔可夫随机场的定义如下:首先给定无向图 $G(V,E)$,指明变量间的连接关系。再给定随机变量集合 $X = \{X_v, v \in V\}$,一组离散的状态空间 $A = \{1,2,3,\dots,L\}$, $X_v \in A$,所有变量组成的状态解空间: $\Omega = \{\omega = (X_{v_1}, \dots, X_{v_n}) : X_{v_i} \in A, 1 \leq i \leq n\}$,那么

称 X 是一个马尔可夫随机场,当且仅当:

- (1) $P(X = \omega) > 0$ 时,对所有 $\omega \in \Omega$
- (2) $P(X_i | X_j, j \neq i) = P(X_i | X_j, j \in N(X_i))$

其中, $N(X_i)$ 表示节点 i 的邻域节点的集合,也就是说当给定其邻域节点后,变量 X_i 的概率分布只依赖于和它相邻的节点,条件独立于其他节点。

根据马尔可夫随机场模型的定义可以得出以下3条性质:

- (1) 成对的马尔可夫性 当给定所有其他变量后,任意两个不相邻的变量条件独立。记做: $X_u \perp X_v | X_{V \setminus \{u, v\}}$ 当 $\{u, v\} \notin E$
- (2) 局部马尔可夫性 当给定一个变量的邻域变量后,它条件独立于其他所有变量。记做: $X_v \perp X_{V \setminus \{v, ne(v)\}} | X_{ne(v)}$, 其中, $ne(v)$ 表示节点 v 的邻域节点集合。
- (3) 整体马尔可夫性 当给定变量的割集,任意两个变量的子集合条件独立。记做: $X_A \perp X_B | X_S$, 其中对于任意一条由集合 A 中的节点到集合 B 中的节点的路径都要经过集合 S 。

一般来说,对于马尔可夫随机场模型,假设对节点 i 的观测量为 Y_i , 这个观测量和隐状态 X_i 之间具有一定的统计相关性,可以把这种相关性用一个相容性函数 $\phi_i(X_i, Y_i)$ 来表示,这个函数通常又被称作变量 X_i 的“证据”。又根据马尔可夫随机场的定义,一个变量 X_i 的概率分布和它的邻域节点有关,也就是说在无向图 G 中,每条边都蕴含了它所连接的相邻两个变量 X_i 和 X_j 之间的相关性,把这层相邻节点之间的相关性用另一个相容性函数 $\varphi_{ij}(X_i, X_j)$ 来表示。那么所有变量的联合概率分布可以表示如下:

$$P(X, Y) = \frac{1}{Z} \prod_{(i, j) \in E} \varphi_{ij}(X_i, X_j) \prod_i \phi_i(X_i, Y_i) \quad (3)$$

式中, Z 是一个归一化常量, i, j 为所有相邻的节点。由式(3)可以看出,通过 X 取遍状态空间 Ω 中所有的可能而使得 $P(X, Y)$ 取得最大后验概率时的 ω 即为马尔可夫随机场模型的最优解。

马尔可夫模型以其刻画数据的规律及模拟结构能力强、表述问题清晰、直观,在物理、化学、生物、医学、多媒体应用等诸多领域得到广泛的应用,尤其在计算机视觉领域中更是备受关注和研究。然而由于利用马尔可夫随机场处理的往往是视觉图像相关的问题,而基于视觉图像的马尔可夫随机场模型通常

包含了几百万个变量,这样给问题求解的计算量上带来了很大的负担,遍历所有解空间的算法复杂度更是随着变量增加成指数级增长从而无法实现。所以对马尔可夫随机场模型的高效求解成为了人们关注的热点。

3 MRF 推理算法

在式(3)中,可以把观测量 Y_i 认为是固定的量,所以式(3)可以表示为

$$P(X) = \frac{1}{Z} \prod_{(i, j) \in E} \varphi_{ij}(X_i, X_j) \prod_i \phi_i(X_i) \quad (4)$$

对式(4)两边求对数并且忽略常数项可以得到:

$$\ln P(X) = \sum_{(i, j) \in E} \ln \varphi_{ij}(X_i, X_j) + \sum_i \ln \phi_i(X_i) \quad (5)$$

令 $E(X) = -\ln P(X)$, 根据统计热力学中的波尔兹曼定律和 Potts Model^[5]:

$$P(X) = \frac{1}{Z} e^{-E(X)/T} \quad (6)$$

可以看出,当 $T=1$ 时马尔可夫随机场模型就是一个 Potts Model。如果每个节点 X 的状态数只有两个,那么这个模型称作 Ising Model。而式(6)中 X 的分布也就是著名的吉布斯分布 (gibbs distribution)。

对马尔可夫随机场模型的求解方法一直是人们关注的问题,也提出了很多算法,有一些局部的算法,如:松弛算法,动态规划(DP)等,也有一些基于整体的算法,如:MFA(mean field annealing),模拟退火(simulated annealing)等。但这些算法的复杂度都很高,处理问题的速度不令人满意,尤其在处理大规模的视觉数据上。

近年来,随着人们对问题的深入研究,提出了两种全新的高效算法来求解马尔可夫随机场模型,他们便是著名的 BP 算法^[6]和 GC 算法^[7]。

3.1 BP 算法

BP 算法是通过相邻节点间的统计相关性并且通过相邻节点间的“消息”传递来求得每个节点的边缘概率分布,使得式(4)取最大值也就是达到最大联合后验概率的 ω 便是 MRF 问题的最优解。

在 BP 算法中,定义变量 $m_{ij}(x_j)$ 代表一个从节点 i 到节点 j 的消息,这个消息表达了节点 i 认为它的相邻节点 j 应该处于哪个状态。这个变量是一个

向量,其维度和 X_j 一样(也就是状态数 L),这个向量的每一维表示节点 i 认为节点 j 应该为这个相应状态的概率。另外定义变量 $b_i(x_i)$ 表示节点 i 处于状态 X_i 的概率,它正比于自己的局部“证据”($\phi_i(x_i)$)和所有来自于其相邻节点的消息的乘积:

$$b_i(x_i) = k\phi_i(x_i) \prod_{j \in Ne(i)} m_{ji}(x_i) \quad (7)$$

式中, k 是一个归一化因子(因为概率和应等于 1)。那么每个消息 $m_{ij}(x_j)$ 可以计算如下:

$$m_{ij}(x_j) = \sum_{x_i} \varphi_{ij}(x_i, x_j) b_i(x_i) \quad (8)$$

把式(7)带入式(8),可以得到消息更新的规则:

$$m_{ij}(x_j) = \sum_{x_i} \varphi_{ij}(x_i, x_j) \phi_i(x_i) \prod_{k \in Ne(i) \setminus j} m_{ki}(x_i) \quad (9)$$

式(9)右边取所有到达节点 i 的消息的乘积,除了来自节点 j 的消息。有了消息更新规则(式(9))以及概率计算公式(7),就可以先任意初始化每个 $b_i(x_i)$,然后迭代的求解 $m_{ij}(x_j)$ 和 $b_i(x_i)$ 直至收敛,也就是说 $m_{ij}(x_j)$ 不再发生变化。图 2 表示了几个节点之间消息的传递方式。

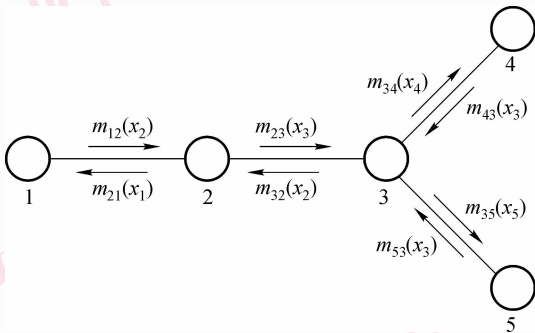


图 2 BP 算法中的消息传递示意图

Fig. 2 A illustration of the messages passed in BP

为了进一步说明 BP 算法是如何高效求解 MRF 模型的,下面就具体的几个节点来讨论 BP 算法过程并说明 BP 算法如何精确收敛到真实的边缘分布。给出如图 3 中所示的 4 个节点,根据式(4)可以得到节点 2 的边缘概率为(忽略常数项):

$$P(X_2) = \phi_2(X_2) \sum_{x_j} \prod_{j \in \{1,3,4\}} \varphi_{2j}(X_2, X_j) \phi_j(X_j) \quad (10)$$

即节点 1,3,4 取遍所有的可能状态。而另一方面,根据 BP 算法中的式(7),有:

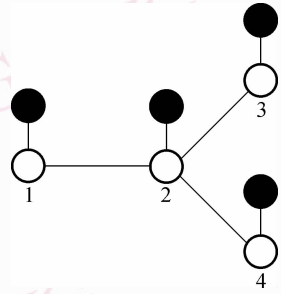


图 3 BP 算法求解示意图

Fig. 3 An illustration of the messages passed in BP

$$b_2(x_2) = k\phi_2(x_2) \prod_{j \in Ne(2)} m_{j2}(x_2) \quad (11)$$

再将消息计算公式(8)带入式(11)可得(忽略常数项):

$$b_2(x_2) = \phi_2(x_2) \prod_{j \in \{1,3,4\}} \sum_{x_j} \varphi_{j2}(x_2, x_j) b_j(x_j) \quad (12)$$

把式(7)带入式(12),根据更新规则(式(9))可得:

$$b_2(x_2) = \phi_2(x_2) \prod_{j \in \{1,3,4\}} \sum_{x_j} \varphi_{j2}(x_2, x_j) \phi_j(x_j) \quad (13)$$

从式(13)中可以看出,根据 BP 算法求得的节点 2 的边缘概率正是其真实的边缘概率。可以证明,对于任意简单连接的马尔可夫随机场模型,也就是其中没有环路,对所有的节点,BP 算法最后都将精确收敛到其真实的边缘分布。

BP 算法利用模型中所有节点的整体信息来计算模型中每个节点的边缘分布,而不仅仅是利用节点间的局部相关性。在模型中,传递的消息就像是河流一般由每个节点的邻域节点流入该节点中,同时每个节点也有消息流向其相邻的节点中,其中每条消息都蕴含了来自网络中各个分支的所有信息。而对于一个存在环路的马尔可夫随机场模型来说,在一定的参数下,确实可以找到使 BP 算法不收敛或者收敛到不准确的边缘概率的情况^[8]。另一方面,BP 算法在处理很多具有回路的“Tanner Graph”^[9]以及计算机视觉中广泛存在回路的 MRF^[10]中取得了巨大的成功,另外文献[6]中提出了基于有环路的“Generalized Loopy BP”算法对存在环路的 MRF 取得了很好的效果,有兴趣的读者可以进一步阅读相关的文献。

BP 算法有效解决了巨大计算量的问题,通过计算消息在图像网络中的传递来求解每个节点的边缘

概率分布。文献[11]提出了一种全新的消息更新方法,进一步加快了BP收敛的速度,被称为“加速”的消息更新算法。文献[12]详细的讨论了针对不同的相容性函数采取的高效动态规划算法,大大减少了计算量,同时也针对计算机视觉中典型的网格MRF进行优化,使得BP的效率为原来的一倍。

3.2 GC 算法

GC算法把求解MRF模型问题看做是一个最优标记问题,正如前面所说,节点状态的取值是一组有限的集合 \mathbf{A} ,其中的状态数为 L ,又称为 L 个级别。而对解空间中任意一个解 ω 也就对应着一组标记,记作 f ,每个节点 X_i 的状态标记为 L 个级别中的一个,记作 f_i 。使得式(4)取得最大值的那组标记即为最优标记 \hat{f} 。

令 $E(f) = -\ln P(X)$, $V_{ij}(f_i, f_j) = -\ln \varphi_{ij}(X_i, X_j)$

$D_i(f_i) = -\ln \phi_i(X_i)$ 则式(5)可以写为

$$E(f) = \sum_{(i,j) \in E} V_{ij}(f_i, f_j) + \sum_i D_i(f_i) \quad (10)$$

式中, $E(f)$ 可以看做标记 f 的能量,要使式(4)取最大值即使式(10)取极小值。(式(10)是对式(4)的两边取负对数而得)

GC算法可以高效地求得使能量取极小值时对应的最优标记 \hat{f} ,其核心思想是逐步调整标记 f ,使得调整后的新标记 f' 的能量 $E(f') \leq E(f)$,在GC算法中包含两个基本的步骤, $\alpha - \beta$ 交换和 α 扩展。 $\alpha - \beta$ 交换指的是在标记 f 中,标记为 α 的节点和标记为 β 的节点可以互相任意交换其标记,而对其他标记不为 α 或 β 的节点其标记保持不变。 α 扩展指的是对于一个标记 f ,其中标记不为 α 的节点可以修改其标记为 α ,而标记为 α 的节点保持不变。可以证明,一个最优的 $\alpha - \beta$ 交换或者 α 扩展都与一个最小分割对应,文献[13]、[14]对最小分割有详细的论述。

下面给出GC算法中 $\alpha - \beta$ 交换的流程:

- (1)任意给定一个标记 f
- (2)定义 success = 0
- (3)对状态集中的每一对标记 $\{\alpha, \beta\}$
 - ① 在所有的 $\alpha - \beta$ 交换后得到的 f' 中搜索最优 $\hat{f} = \operatorname{argmin} E(f')$ 。
 - ② 如果 $E(\hat{f}) < E(f)$ 则令 $f = \hat{f}$ 且 success = 1
 - (4)如果 success = 1 则转流程(2)
 - (5)返回 f

以下是GC算法中 α 扩展的流程:

- (1)任意给定一个标记 f
- (2)定义 success = 0
- (3)对每一对标记 α
 - ① 在所有的 α 扩展后得到的 f' 中搜索最优 $\hat{f} = \operatorname{argmin} E(f')$ 。
 - ② 如果 $E(\hat{f}) < E(f)$ 则令 $f = \hat{f}$ 且 success = 1
 - (3)如果 success = 1 则转流程(2)
 - (5)返回 f

GC算法将对 $\alpha - \beta$ 交换和 α 扩展进行迭代直至得到的标记 f 不再发生变化。可以证明,通过有限次迭代算法必将收敛于使得能量取得极小值的标记 \hat{f} 。文献[7]中对 $\alpha - \beta$ 交换以及 α 扩展与最小分割的对应性进行了完整详细的证明。

4 基于MRF的视觉信息处理

近年来,基于MRF的视觉信息处理受到了很大的关注,在图像分割、图像还原以及目标跟踪、目标检测,立体视觉等方面提出了很多有效的算法,得到了广泛地应用,同时也为进一步拓宽MRF在视觉信息处理的应用领域奠定了基础。下面通过MRF在目标检测和立体视觉上的经典例子来说明MRF的具体建模和求解方法。其中立体视觉需要考虑左右两帧图像之间的关系,而目标检测则需要考虑当前帧图像和其前后帧图像之间的关系。通过对这两个典型问题及其建模、求解过程的仔细分析,可以理解更多的视觉问题的基于MRF的求解方法。

4.1 基于MRF的目标检测

在复杂的图像场景中进行背景的动态建模,准确地反映出背景和前景的变化是许多视觉信息处理的基础,同时也是视觉信息处理的一个难点问题。近年来,很多方法被提出用于解决物体运动检测,包括背景差分、运动历史图、确定性松弛算法等,然而这些算法都没能完全利用图像的整体相关性以及图像序列在时间上的关系,导致目标检测的不精确以及不准确。文献[15]提出一种基于MRF模型的目标检测算法,该算法通过建立图像的时间和空间相关信息对变化的背景建模,同时利用图像颜色一致性和运动的平滑性以及混合模型中的BP算法来解决模型的学习和估计问题,下面具体讨论MRF模型如何应用于视频中的运动检测。

- (1)问题建模 图4给出了视频中图像序列的

3 维网格图表示,视频图像序列中的所有像素构成了 MRF 的状态节点集合(图 4 网格中的圆点),节点状态的取值空间为 $\{0,1\}$,0 表示该节点没有运动,1 则表示该节点属于运动目标。在 t 时刻下,结点的状态由其观察值(像素的颜色信息,图 4 网格中的方点)和所有邻域节点隐含状态来共同决定,其中包括空间上的相关性(图 4 中红色表示)即与该节点相连的 4 个邻域节点,以及时间上的相关性,即每一个隐状态都与它相邻时刻的各隐状态相互联系(图 4 中绿色表示)。这样空间上的 2 维图像关系和时间上的 1 维关系便组成了一个如图 3 所示的 3 维网格,也就是一个 3 维的 MRF 模型。

(2)模型求解 建立好 3 维 MRF 之后便要定义出相容性函数,即定义一个函数来描述相邻节点之间的相关性。相容性函数 ϕ 是表示节点状态值与观察的相容性,也就是说,如果观测到的像素的颜色变化越大,则该节点状态取 1 的概率就越高,也就越倾向于运动节点。对表示相邻节点间的相容性函数 φ_{ij} ,当一个节点相邻节点的取值为 1,则该节点也越倾向于运动节点,反之,当该节点相邻节点取值为 0,则该节点越倾向于静止节点。定义好相容性函数后便可以根据前一部分所叙述的 BP 算法来对该 MRF 模型求解。此算法不但可以有效地从复杂的背景中检测出运动物体,而且还克服了以往模型在视觉算法中的缺点。

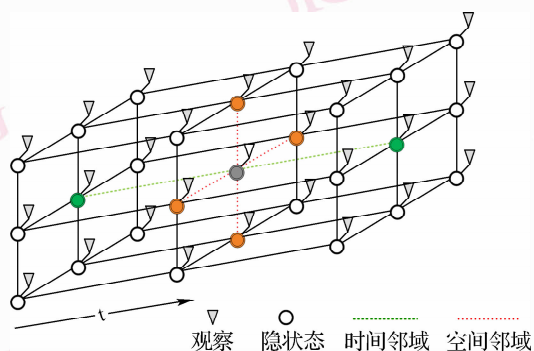


图 4 3 维网格的示意图

Fig. 4 3D lattice defining the neighbors of a pixel

此外,文献[14]讨论了不同的 MRF 模型参数的选取对目标检测结果的影响,同时通过对基于 BP 算法下的 MRF 和之前的经典运动历史图(MHI)算法的详细比较说明了 MRF 算法在目标跟踪上的高效性和准确性。文献[11]则利用光流算法对目标进行跟踪,该算法的核心同样是基于 MRF 模型对目标进行建模,然后通过高效的 BP 算法对 MRF 模型求解。

4.2 基于 MRF 的体视匹配

立体视觉作为计算机视觉中非常重要的一个分支受到了广泛的关注,每年都有很多算法被提出用于解决立体视觉问题。

立体视觉问题又可以叫做立体重构问题,其核心在于视觉图像的匹配,也就是对两路或者多路不同视角同一时刻拍摄的同一个场景图像进行点与点之间的匹配对应。下面就具体如何建模进行介绍,这里假定只对两张同一时间不同视角拍摄的图片进行图像匹配。

(1)问题建模 首先对两张图片做外极线校正,这样可以使得两张图片中对应的节点处于同一水平线上。然后把其中一幅图的所有像素作为 MRF 中的状态节点,其状态的取值也就是该节点相对于其在另一幅图中对应节点的视差。图 5 给出了经典的用于立体视觉匹配的 MRF 网格模型,其中白色的圈表示隐状态节点;黑色的圈表示观测节点,也就是根据估计的视差所观测另一幅图像的像素节点。

(2)模型求解 相容性函数 ϕ 是表示节点状态值与观察的相容性,当对应节点之间的像素颜色差异越大,则 ϕ 得取值越小,也就是说该节点处于该视差的概率越小,反之亦然。而表示相邻节点间的相容性函数 φ_{ij} 可以这样来取,相邻节点的视差值越接近则概率越大,其直观意义就是要使得视差场趋于平滑。定义好了相容性函数后便可利用 BP 算法或者 GC 算法来求解。

文献[15]同时利用 3 个 MRF 模型对立体视觉问题进行建模,包括视差 MRF 模型,线性 MRF 模型和二值 MRF 模型。然后通过对这个混合 MRF 模型的推理,把这 3 个模型融合在一起,进而利用 BP 算法对该模型进行优化求解,取得了很好的效果。该文献还 TP 将图像分割技术与 MRF 模型融合在一起,利用图像分割作为一个强有力的支持,增强算法的鲁棒性同时也提高了算法的精确度。

虽然文献[16]在立体视觉匹配问题上取得了不错的效果,但是对于有遮挡的情况(如有些像素只在一幅图像上可见,而在另一幅图像上不可见)并没有做特殊的处理,文献[17]则通过两幅立体视觉图像的内在对称性建立了基于块的 MRF 模型,该算法利用图像分割作为块模型的基本预测,有效地把图像的遮挡问题加入了该模型中,再利用图像的对称性迭代地求解两幅立体视觉图像的视差和遮挡,该算法利用 GC 算法来求解建立的对称 MRF 模

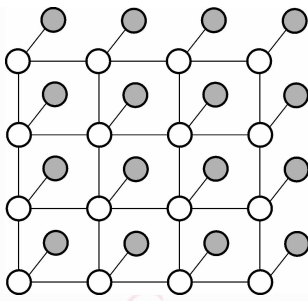


图5 经典的立体视觉 MRF 网格模型

Fig. 5 Classic grid-based MRF for stereo matching

型,其结果在处理遮挡的问题上取得了突破性的进展,当然其核心思想来自于更早的一篇文献[18]。

文献[19]对基于 MRF 建模的立体视觉算法进行了总结和比较,对不同参数以及不同的函数选取对匹配结果的影响做了详细的分析。一篇最近的文献[20]提出了更加高效的立体视觉匹配算法,该算法的核心思想是将像素的视差值进行二进位化,然后对每一位都进行 MRF 建模再利用 GC 算法求解,该模型大大加速了立体视觉匹配的速度,因为对一个 GC 算法来说,其标记集只有 $\{0,1\}$ 2 个状态,因为每一位都是一个二进制数。同时该文也利用立体视觉匹配算法对场景进行了前景和背景的联合建模,从而实现了图像分割。

5 结 语

概率图模型是处理视觉信息中的不确定性问题的一个重要方法。本文通过对 MRF 的深入分析,介绍了 MRF 和 PGM 理论、方法和具体的视觉应用。希望能够抛砖引玉,促进视觉信息处理理论和方法的发展。

参考文献 (References)

- Jordan M, Bishop C. An introduction to Graphical Models[EB/OL]. <http://www.cs.pitt.edu/~milos/courses/cs2750-Spring03/Readings/chapter10.ps>,2001.
- Loeliger H A. An introduction to factor graphs [J]. IEEE Signal Processing Magazine,2004, **21**(1):28-41.
- Pearl J. Bayes and Markov Networks: a Comparison of Two Graphical Representations of Probabilistic Knowledge [R]. 860024 (R-46), Los Angeles, California, USA; University of California, Los Angeles Computer Science Department, 1986.
- Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference [M]. San Mateo, California, USA: Morgan Kaufmann Publishers, 1988.
- Baxter R J. Exactly Solved Models in Statistical Mechanics [M].

- London, UK: Academic Press, 1982.
- Yedidia J S, Freeman W T, Weiss Y. Understanding belief propagation and its generalizations [A]. In: Exploring Artificial Intelligence in the New Millennium [C], San Francisco, CA, USA: Morgan Kaufmann Publishers, 2003:239-269.
- Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, **23**(11): 1222-1239.
- Murphy K P, Weiss Y, Jordan M I. Loopy belief propagation for approximate inference: An empirical study [A]. In: Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence [C], San Francisco, CA, USA: Morgan Kaufmann Publishers, 1999:467-475.
- Frey B J, Mackay D J C. A revolution: belief propagation in graphs with cycles [A]. In: Advances in Neural Information Processing Systems 10 [C], Cambridge, MA, USA: MIT Press, 1998: 479-485.
- Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision [J]. International Journal of Computer Vision, 2000, **40**(1), 25-47.
- Tappen M F, Freeman W T. Comparison of graph cuts with belief propagation for stereo, using Identical MRF parameters [A]. In: Proceedings of the 9th IEEE International Conference on Computer Vision [C], Nice, France, 2003, **2**:900-906.
- Felzenszwalb P F, Huttenlocher D P. Efficient belief propagation for early vision [J]. International Journal of Computer Vision, 2006, **70**(1):41-54.
- Ahuja R K, Magnanti T L, Orlin B. Network Flows: Theory, Algorithm and Applications [M]. Upper Saddle River, New Jersey, USA: Prentice Hall, 1993.
- Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, **26**(9):1124-1137.
- Yin Z S, Collins R. Belief propagation in a 3D spatio-temporal MRF for moving object detection [A]. In: Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition [C], Minneapolis, MN, USA, 2007:1-8.
- Sun J, Shum H Y, Zheng N N. Stereo matching using belief propagation [J]. Lecture Notes in Computer Science, 2002, **2351**:510-524.
- Deng Y, Yang Q, Lin X Y, et al. Stereo correspondence with occlusion handling in a symmetric patch-based graph cuts model [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, **29**(6):1068-1079.
- Sun J, Li Y, Kang S B, et al. Symmetric Stereo Matching for Occlusion Handling [A]. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C], San Diego, CA, USA, 2005, **2**:399-406.
- Szeliski R, Zabih R, Scharstein D. A comparative study of energy minimization methods for markov random fields [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, **30**(6):1068-1080.
- Chai D F, Peng Q S. Bilayer Stereo Matching [A]. In: Proceedings of IEEE 11th International Conference on Computer Vision [C], Rio de Janeiro, Brazil, 2007:1-8.