

基于不完全分层 MRF 的非监督图象分割

汪西莉^{1,2}, 刘 芳³, 焦李成¹

(1. 西安电子科技大学雷达信号处理国家重点实验室, 陕西西安 710071;

2. 陕西师范大学计算机学院, 陕西西安 710062; 3. 西安电子科技大学计算机学院, 陕西西安 710071)

摘 要: 分层马尔可夫随机场(MRF)图象模型由于层间具有因果关系,且这种因果关系符合图象的性质,使基于该模型的图象处理时间比平面 MRF 模型所用的时间大为减少. 针对作者提出的一种新的分层马尔可夫图象模型——不完全分层模型,导出 EM 算法以估计模型参数. 算法继承了分层模型非迭代算法运算速度快的优点,并因为模型结构的简化进一步减少了计算量,算法在模型的最上层加入了平面节点间信息的交互,以较少的计算换来了更加精确的参数估计结果. 算法用于图象非监督分割的实验表明,和分层模型算法相比,其处理速度更快,由所估计的参数得到了更好或相当的分割结果,尤其适合大幅面图象的处理.

关键词: 不完全分层 MRF; EM 算法; 非监督图象分割

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2004) 07-1086-04

Unsupervised Image Segmentation Based on Incomplete Hierarchical MRF

WANG Xi-li^{1,2}, LIU Fang³, JIAO Li-cheng¹

(1. National Key Lab. of Radar Signal Processing, Xidian Univ., Xi'an, Shaanxi 710071, China;

2. School of Computer Science, Shanxi Normal Univ., Xi'an, Shaanxi 710062, China;

3. School of Computer Science, Xidian Univ., Xi'an, Shaanxi 710071, China)

Abstract: Hierarchical MRF image model has causality property between layers, and the causality is consistent with the characteristics of images. So the processing time of such models is much less than that of the plain MRF models. An expectation maximization (EM) algorithm estimating parameters of incomplete hierarchical MRF model, which is a new hierarchical MRF image model we presented, is deduced. The advantage of less time cost possessed by the non-iterative algorithm of hierarchical models is inherited. Time is further reduced due to simplified model structure. The interaction between neighbor nodes on the top layer is considered, which results in more accurate estimate values with less computing cost. The algorithm is used in unsupervised image segmentation. The experimental results demonstrate that it is characterized by high speed and better results compared with that of hierarchical models. It is more fit for large images.

Key words: incomplete hierarchical MRF; expectation maximization (EM) algorithm; unsupervised image segmentation

1 引言

马尔可夫随机场是图象分割中一个有效且常用的工具,它将分割标号建模为与 Gibbs 分布对应的 MRF,通过信息的局部交互,逐渐扩散到整个图象,这一过程带给这类随机图象模型的不仅有令人满意的分割结果,还有令人望而生畏的处理时间. 目前看来,解决这一问题最好的方法是采用分层马尔可夫图象模型,它将未知的分割标号建模为若干层 MRF,不考虑一个 MRF 内节点间的依赖关系,而在层与层的父子节点间建立因果关系,得到类似马尔可夫链的结构和算法^[1~3],极大地减少了处理时间.

我们提出了一种新的分层模型——不完全分层马尔可夫

图象模型^[4],不管图象有多大,该模型都只取较少的层数,使分层模型变得简洁,其算法综合了分层和非分层 MRF 的优势,方便了模型实现. 在时间开销更少的时候,减轻了分层模型用于图象分割时特有的块现象,得到了更好的分割结果.

使用 MRF 图象模型进行图象分割,一个重要的问题是模型参数的估计,估计的参数值是否符合实际直接关系到分割结果. 这里考虑观测图象只有一幅的情况,这也是很多实际问题所遇到的情况. 因为没有其它更多的图象用以事先估计模型参数,只能由一幅图象先估计参数,再利用该模型对这幅图象进行分割,实现非监督分割(假设类数已知). 在统计计算中,这种不完全数据的参数采用迭代的 EM 算法来估计. 下面在介绍了不完全分层 MRF 模型后,着重介绍其用于参数估计

收稿日期:2002-12-13;修回日期:2003-11-05

基金项目:国家自然科学基金重点项目(No. 60133010);教育部博士点基金资助项目

的 EM 算法,最后是实验结果及分析.

2 不完全分层 MRF 图象模型

该模型相对于一般的分层模型,其层数不完整. 分层模型

将图象的未知分类标号 X 建模为 n 个 MRF,如图 1. 第 0 层 X^0 只有一个节点,第 $n-1$ 层的大小和图象大小相同,每个像素对应一个节点. 除了叶子节点,每层的一个节点对应于其相邻下层的四个节点,形成

图 1 分层 MRF 模型

父子关系. 层间具有马尔可夫性质:每层只依赖于相邻的上一层,和其它上层无关,即 $P(x^i | x^{i-1} x^{i-2} \cdots x^0) = P(x^i | x^{i-1})$, $i \in \{1, 2, \cdots, n-1\}$. 同一层的各节点在其父节点已知的条件下相互独立,即 $P(x^i | x^{i-1}) = \prod_{x_s \in x^i} P(x_s | x_{\rho(s)})$, $i \in \{1, 2, \cdots, n-1\}$, $x_{\rho(s)}$ 为 x_s 的父节点. 在第 $n-1$ 层,每一个象素点 y_s ($y_s \in Y$, Y 为 $N \times N$ 图象) 依赖于一个类标号 x_s , 且假定已知 x_s 后

y_s 独立,即 $P(y | x^{n-1}) = \prod_{s=1}^{N \times N} P(y_s | x_s)$. 因此可以得到 x 、 y 的联合分布:

$$P(x, y) = P(x^0) P(y | x^{n-1}) \prod_{i=1}^{n-1} P(x^i | x^{i-1})$$

完整的层次结构繁琐,对大幅面图象尤为明显. 不完全分层马尔可夫图象模型只取整

个层次的下几层,如图 2 所示,其最下层仍和图象大小一致,而第 0 层则不止一个节点. 分层模型所具有的上述关系也存在于不完全分层模型

图 2 不完全分层 MRF 模型

中. 第 0 层的不同带来了两模型间最大的差异,联合分布中 $P(x^0)$ 在分层模型中是一个节点的概率,在不完全分层模型中 x^0 包含了多个节点,而它们的父节点被截去,使 x^0 中的节点彼此间不再条件独立, $P(x^0)$ 的计算变得不同于分层模型,这时必须考虑相邻节点间的影响.

3 EM 算法估计参数

参数估计要在 x 未知的情况下,根据不完全数据 y 求得使 $\log P(y | \theta)$ 最大的参数值 θ . EM 算法是求解这类问题的有效算法^[5,6],实现时用联合数据似然 $\log P(x, y | \theta)$ 代替 $\log P(y | \theta)$ 以利于计算. EM 算法主要有两步: E 步在当前参数 θ 下,计算联合似然的期望 $E[\log P(x, y | \theta) | y, \theta]$, M 步求新的参数值 $\theta = \arg \max_{\theta} E[\log P(x, y | \theta) | y, \theta]$, 两步交替执行直至算法收敛.

对于不完全分层模型, $P(x^0)$ 用伪似然 $P(x^0) = \prod_{x_s \in x^0} P(x_s | x_{\partial s})$ 代替, $x_{\partial s}$ 是 x_s 的邻域. E 步联合数据似然的期望为:

$$E[\log P(x, y | \theta) | y, \theta] = \sum_x P(x | y, \theta) \log P(x, y | \theta)$$

$$\begin{aligned} &= \sum_x P(x | y, \theta) \log [P(x^0 | \theta) P(y | x^{n-1}, \theta) \prod_{i=1}^{n-1} P(x^i | x^{i-1}, \theta)] \\ &= \sum_x P(x | y, \theta) \log [\prod_{x_s \in x^0} P(x_s | x_{\partial s}, \theta) \prod_{s=1}^{N \times N} P(y_s | x_s, \theta) \\ &\quad \cdot \prod_{i=1}^{n-1} P(x^i | x^{i-1}, \theta)] \\ &= \sum_i \sum_j \sum_{x_s \in x^0} P(x_s = i, x_{\partial s} = j | y, \theta) \log P(x_s = i | x_{\partial s} = j, \theta) \\ &\quad + \sum_i \sum_{x_s \in x^{n-1}} P(x_s = i | y, \theta) \log P(y_s | x_s = i, \theta) \\ &\quad + \sum_{i,j \in I} \sum_{x_s \in x^0} P(x_s = j, x_{\rho(s)} = i | y, \theta) \log P(x_s = j | x_{\rho(s)} = i, \theta) \end{aligned} \quad (1)$$

$\Lambda = \{1, 2, \cdots, K\}$, K 是类数, $\Omega = \{0, 1, 2, \cdots, M\}$, M 是邻域节点总数,假设图象各类服从高斯分布, $P(y_s | x_s) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp$

$(-\frac{(y_s - \mu_k)^2}{2\sigma_k^2})$, μ_k, σ_k^2 是第 k ($k \in \Lambda$) 类的均值和方差,则模型参数 $\theta = \{P(x_s | x_{\partial s}) (x_s \in x^0), P(x_s | x_{\rho(s)}) (x_s \in x^0), \mu_k, \sigma_k^2\}$.

M 步参数的更新是一个求条件极值的问题,在约束条件 $\sum P(x_s = i | x_{\rho(s)}) = 1$, $\sum P(x_s = i | x_{\partial s}) = 1$, 及 $\sum P(y_s = i | x_s) = 1$ 下,求使式(1)最大的参数 θ ,采用 Lagrange 乘法可以得到更新的参数:

$$\begin{aligned} P(x_s = i | x_{\partial s} = j) &= \frac{\sum_{x_s \in x^0} P(x_s = i, x_{\partial s} = j | y, \theta)}{\sum_{x_s \in x^0} \sum_k P(x_s = k, x_{\partial s} = j | y, \theta)} \\ P(x_s = j | x_{\rho(s)} = i) &= \frac{\sum_{x_s \in x^0} P(x_s = j, x_{\rho(s)} = i | y, \theta)}{\sum_{x_s \in x^0} \sum_k P(x_s = k, x_{\rho(s)} = i | y, \theta)} \\ \mu_i &= \frac{\sum_{x_s \in x^{n-1}} P(x_s = i | y, \theta) \cdot y_s}{\sum_{x_s \in x^{n-1}} P(x_s = i | y, \theta)} \\ \sigma_i^2 &= \frac{\sum_{x_s \in x^{n-1}} P(x_s = i | y, \theta) \cdot (y_s - \mu_i)^2}{\sum_{x_s \in x^{n-1}} P(x_s = i | y, \theta)} \end{aligned} \quad (2)$$

由于层间的因果关系,使 E 步可通过类似马尔可夫链的向上、向下算法来求得 $P(x_s, x_{\rho(s)} | y)$ 和 $P(x_s | y)$. 令 $d(s)$ 代表以 s 为根的子树的叶子节点,有:

$$\begin{aligned} P(x_s | y) &= \sum_{x_{\rho(s)}} P(x_s | x_{\rho(s)}, y) P(x_{\rho(s)} | y) \\ &= \sum_{x_{\rho(s)}} \frac{P(x_s, x_{\rho(s)} | y_{d(s)})}{\sum_{x_s} P(x_s, x_{\rho(s)} | y_{d(s)})} P(x_{\rho(s)} | y) \end{aligned} \quad (3)$$

$$P(x_s, x_{\rho(s)} | y_{d(s)}) \propto P(y_{d(s)} | x_s) P(x_s | x_{\rho(s)}) \quad (4)$$

$$P(y_{d(s)} | x_s) = \prod_{\rho(v)=s} \sum_{x_v} P(y_{d(v)} | x_v) P(x_v | x_{\rho(v)}) \quad (5)$$

对于最下层 $P(y_{d(s)} | x_s) = P(y_s | x_s)$, 根据式(4)、(5)经

过一遍自下而上的逐层计算可得到父子部分后验概率 $P(x_s, x_{p(s)} | y_{d(s)})$, 再经一遍自上而下的扫描由式(3)可得各层节点的后验边缘概率。

问题是第 0 层的 $P(x_s, x_{p(s)} | y)$ 和 $P(x_s | y)$ 如何求解, 如果层次是完整的, 该层只有一个节点, 则问题容易解决: $P(x_s | y) \propto P(y | x_s) P(x_s) = P(y_{d(s)} | x_s) \cdot P(x_s)$, $P(x_s, x_{p(s)} | y) = P(x_s | y)$, 但不完全层次模型无法采用前式。为此, 我们采用非分层 MRF 常用的 Gbbs 采样来计算, 根据 $P(x^0 | y)$ 对每个节点 $x_s \in X^0$ 采样 n 个样本 $x_s^1, \dots, x_s^n$, 由这些样本来估计概

率: $P(x_s = i, x_{p(s)} = j | y) = \frac{1}{n} \sum_{k=1}^n (x_s^k = i, x_{p(s)}^k = j)$, $P(x_s = i | y) = \frac{1}{n} \sum_{k=1}^n (x_s^k = i)$ 。完成了一次自下而上的计算同时也得到了 $P(x^0 | y) = \prod_{x_s \in X^0} P(y_{d(s)} | x_s) \cdot P(x^0)$, $P(x^0)$ 在这里采用

Potts 类型先验。因此 E 步要在向上、向下扫描过程之间增加一步, 经 Gbbs 采样求第 0 层后验边缘概率和相邻节点后验概率, 由于第 0 层节点数已比原图大小小了很多, 所以采样时间比直接对原图采样要少得多, 再加上层次的精简, 即使采样个数较多, 所用时间往往少于分层模型, 对多幅图象的实验也表明了这一点。由于第 0 层的这种处理加入了相邻节点间的交互, 在产生块现象上也比完全不考虑平面相邻节点间的影响, 只考虑父子节点间交互的分层模型要好。

Potts 类型先验为 $P(x^0) = \exp[-\beta \sum_{s, t \in \eta^0} (1 - \delta(x_s, x_t))]$,

η^0 是 x^0 的邻域系统。其中的参数 β 也需要事先确定, 我们采用了文献[7]中的估计方法来迭代估计 β 的值。此外还可以简单地指定该值, 这也是很多基于 MRF 的分割常用的方法, 从实验中也发现该参数值并不高度依赖于具体的图象, 当其值在较大的范围内变动时, 算法性能并没有什么变化。因此直接指定参数值是可行的, 后面所给的实验即将其值取为 1.5。

综上, 不完全分层模型的 EM 算法为:

(1) 置初始参数值

(2) E 步: 计算联合似然的期望

(a) 向上算法: 从下向上逐层计算 $P(y_{d(s)} | x_s)$ 、 $P(x_s, x_{p(s)} | y)$

(b) 利用 Gbbs 采样计算第 0 层的 $P(x_s | y)$ 、 $P(x_s, x_{p(s)} | y)$

(c) 向下算法: 对 $x_s \in X^1, \dots, X^{n-1}$ 计算 $P(x_s | y)$

(3) M 步: 根据式(2)更新参数

(4) 若收敛则停止, 否则转 2

4 实验结果

实验首先根据 EM 算法估计模型参数, 然后进行分割。一般通过求 x 的最大后验概率 (MAP) 或最大后验边缘概率 (MPM) 来得到分割结果, 由于后验边缘概率由 EM 算法的 E 步过程可以得到, 且和 MAP 相比它对误分类的惩罚不那么重, 更符合分割结果允许一些误分类存在的实际情况, 因此用 MPM 进行分割, 即 $\hat{x}_s = \arg \max_{x_s} P(x_s | y)$ 。

EM 算法的参数初值对收敛速度有较大影响, 这里采用 K -均值算法估计各类的均值、方差初值。第 0 层节点的初值采用极大似然根据 $P(y_{d(s)} | x_s)$ 确定。父子节点间的转换概率 $P(x_s | x_{p(s)})$ 若各不相同, 则大大增加了模型参数的数量, 也没有必要, 因此往往使两层间的父子转换概率相同, 甚至所有父子节点转换概率相同^[2], 我们采用后者:

$$P(x_s = j | x_{p(s)} = i) = \begin{cases} \alpha, & \text{若 } i = j \\ \frac{1-\alpha}{M-1}, & \text{若 } i \neq j \end{cases} \quad (6)$$

α 初值取为 0.95。

下面是均为灰度图象的人工合成图象和一幅 SAR 图象的实验结果。假设图象中各类服从高斯分布, 人工图象各类的均值和方差已知, 图象大小为 256×256 , 完整层次有 9 层。SAR 图象大小为 512×512 , 完整层次有 10 层, 按应用目的将其分为两类: 目标和背景。

不完全分层模型中, 用 Gbbs 采样求第 0 层后验边缘概率的 EM 算法记为 GEM, 采用伪似然的 EM 算法记为 PLEM, 利用分层模型进行参数估计的算法记为 CEM。分别用这三个算法对上述两幅图实现非监督分割。GEM 算法中层数取为 4, 采样 150 个。三个算法迭代的终止条件为 $\max_k (\mu_k^{t+1} - \mu_k^t)^2 + \max_k (\sigma_k^{t+1} - \sigma_k^t)^2 < \varepsilon$, ε 取 0.5。表 1、表 2 是参数估计结果和所用时间, 图 3 是人工合成图象及其实际分割, 图 4 是根据所得参数对 SAR 图象分割的结果。

表 1 人工合成图象参数估计结果

		μ				σ				时间(s)
实 验 1	真实值	80	178	105	150	8	20	12	28	—
	初值	81.15	180.50	106.30	148.27	7.94	20.62	12.22	26.09	—
	GEM	79.95	177.89	104.98	150.01	7.97	19.95	11.86	27.77	7
	PLEM	80.57	177.91	104.94	149.23	8.47	20.43	13.07	28.48	5
	CEM	79.82	177.34	104.79	149.46	7.90	19.73	11.59	26.82	12
实 验 2	初值	92	186	100	120	5	18	16	33	—
	GEM	79.95	177.90	104.97	149.98	7.97	19.96	11.87	27.76	9
	PLEM	81.36	177.86	104.27	147.75	8.96	19.49	15.06	30.18	22
	CEM	79.75	177.29	104.79	149.64	7.87	19.71	11.57	26.66	35

表 2 SAR 图象参数估计结果

		μ		σ		时间(s)
实 验 3	初值	98.66	24.62	27.25	9.43	—
	GEM	101.09	38.75	32.19	17.69	24
	PLEM	92.75	30.18	36.37	11.52	15
	CEM	107.27	41.16	30.89	18.99	91

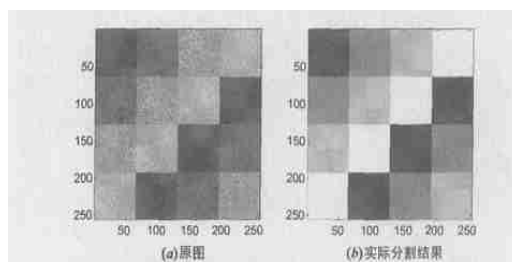


图 3 人工合成图象

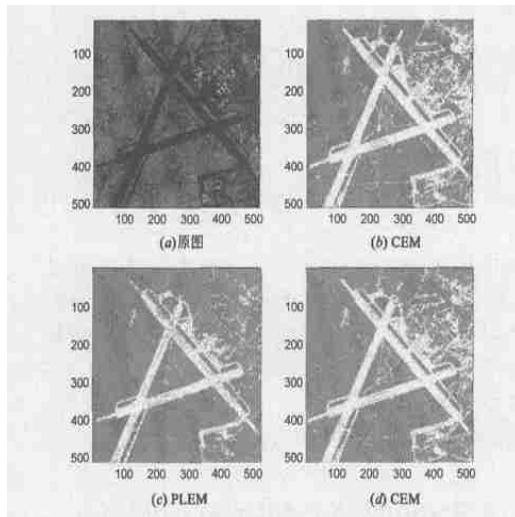


图 4 SAR 图象分割结果

人工合成图象的参数值已知,可将估计值和真实值进行对比.实验一采用 K -均值算法估计参数初值,它已很接近真实值,因此算法在较短时间内迭代即终止.而在同样的终止条件下,实验二的初值距离真实值较大,算法收敛变慢,但除了 PLEM 两个实验得到的参数估计值相差无几. PLEM 采用近似计算,虽然每一步迭代用时少于 GEM,但当初值距离真实值稍大时,它的迭代次数增加,因而所用时间比 GEM 要多,实验表明这时 PLEM 要用更长的时间才能达到和 GEM 一样的估计精度. SAR 图象的真实参数值未知,它的两个类别分别是目标——机场跑道、背景,背景的一部分和目标的灰度值相当,难以区分. PLEM 算法将属于背景的部分误分为目标的最少,CEM 则最多,但 PLEM 却将较多的属于目标的部分误分为背景,CEM、GEM 在这点上明显好于 PLEM. 实验表明,基于不完全分层模型的 GEM、PLEM 算法较分层模型的 CEM 速度快,不论参数初值如何,在同样的终止条件下, GEM 得到的结果最接近实际值. GEM 算法应该是使用中通常的选择,除非对参数初值的准确性有把握,才有理由采用 PLEM 以更短的时间得到同样的结果.

5 结束语

针对我们提出的不完全分层马尔可夫图象模型,给出了 EM 算法来估计模型参数,从而实现了非监督图象分割.该模型和分层模型一样解决了平面 MRF 模型的最大障碍——计算耗时,模型将分层结构由繁琐变简洁,进一步减少了计算量,结构的简化还使得实现中空间的开销大大降低.模型在最上层因果关系被截断,使我们引入了平面 MRF 模型中常用的方法来计算后验边缘概率,正是因为引入了平面相邻节点间

的交互,在只付出较少计算代价的情形下,得到了更好的估计、分割结果.

参考文献:

- [1] J M Laferte, P Perez, F Heitz. Discrete Markov image modeling and inference on the quadtree [J]. IEEE Transactions on Image Processing, 2000, 9(3): 390 - 404.
- [2] C A Bouman, M Shapiro. A multiscale random field model for Bayesian image segmentation [J]. IEEE Transactions on Image Processing, 1994, 3(2): 162 - 177.
- [3] M S Crouse, R D Nowak, R G Baraniuk. Wavelet-based statistical signal processing using hidden Markov models [J]. IEEE Transactions on Signal Processing, 1998, 46(4): 886 - 902.
- [4] 汪西莉, 刘芳, 焦李成. 一种新的分层马尔可夫图象模型及其推导算法 [J]. 软件学报, 2003, 14(9): 1558 - 1563.
- [5] A P Dempster, N M Laird, D B Rubin. Maximum likelihood from incomplete data via the EM algorithm [J]. Journal of the Royal Statistical Society Series B, 1977, 39(1): 1 - 38.
- [6] J Zhang, J W Modestino, D A Langan. Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation [J]. IEEE Transactions on Image Processing, 1994, 3(4): 404 - 419.
- [7] S Lakshmanan, H Derin. Simultaneous parameter estimation and segmentation of gaussian random fields using simulated annealing [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 1989, 11(8): 799 - 813.

作者简介:



汪西莉 女, 1969 年生于陕西西安, 陕西师范大学计算机学院副教授, 硕士生导师, 西安电子科技大学电子工程学院博士研究生, 主要研究领域为智能信息处理、模式识别.



刘 芳 女, 1963 年生于湖南华容, 西安电子科技大学计算机学院教授, 硕士生导师, 主要研究领域为网络智能计算及识别.

焦李成 男, 1959 年生于陕西白水, 教授, 博士生导师. 主要研究领域为智能信息处理.