

一种基于图割的交互式图像分割方法

刘嘉^{①②③} 王宏琦^①

^①(中国科学院电子学研究所 北京 100190)

^②(微波成像技术国家重点实验室 北京 100190)

^③(中国科学院研究生院 北京 100039)

摘要: 交互式的图像分割方法目前已经受到越来越多的重视。该文提出一种新的基于图割(graph cuts)的交互式图像分割方法。该方法将图像的纹理、色彩、边缘等多种特征通过一个概率模型结合在一起。其中纹理和色彩用以 Texton 为基的直方图来建模,并用 Fisher 判别准则来对特征空间进行降维。利用图割方法,可以快速求解该模型下的最优分割。在遥感图像分割等问题中的应用结果表明该方法在速度和精确性方面都有很好的表现。

关键词: 交互式图像分割; Texton; Fisher 判别准则; 图割

中图分类号: TP753

文献标识码: A

文章编号: 1009-5896(2008)08-1973-04

A Graph Cuts Based Interactive Image Segmentation Method

Liu Jia^{①②③} Wang Hong-qi^①

^①(Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China)

^②(National Key Laboratory of Microwave Imaging Technology, Beijing 100190, China)

^③(Graduate University, of Chinese Academy of Sciences, Beijing 100039, China)

Abstract: Interactive image segmentation methods have recently gained more and more attentions. A new interactive segmentation method is proposed based on the graph cuts. It combines several image features like texture, color and edge together through a probabilistic model. Texture and color features are modeled with histograms. Dimensionality reduction in feature space is achieved with a fisher discriminant criterion based on texton. The global optimal segmentation can be efficiently computed via graph cuts. Efficiency and accuracy of the method is demonstrated on aerial image segmentation and some other applications.

Key words: Interactive image segmentation; Texton; Fisher discriminant criterion; Graph cuts

1 引言

图像分割是图像解译中非常重要的一项任务。传统的人工手动分割方式费时费力,已无法满足目前在遥感等领域日益增长的图像解译需求。计算机自动分割方式可以大大提高分割的效率,但往往难以取得令人满意的分割结果。因此,结合两者优势的交互式图像分割方法已经开始越来越受到人们的重视。对交互式分割方法的一个比较全面的回顾可参见文献[1]。

本文以文献[2]工作为基础,提出了一种新的交互式图像分割算法,主要在以下几个方面做出了改进:首先,在贝叶斯框架下重新建立了模型,加入了纹理和色彩特征,克服了文献[2]中方法不能用于彩色图像和纹理图像分割的问题。其次,提出了用 Fisher 判别准则来对 Texton^[3-5]直方图进行降维,以使其适用于交互式应用。最后,提出了在再编辑阶段用新加种子点重新估计分割模型,以弥补初始模型的不准确性。实验结果表明,本文方法虽在速度上有所降低,却大

大提高了其应用的范围及精确性。

2 基本原理

在交互式分割应用中,用户通过交互在待分割图像 I 中标记出目标和背景种子像素点集合 I_{seed_obj} 和 I_{seed_bck} 。根据 I_{seed_obj} 和 I_{seed_bck} ,分别建立目标和背景的统计模型,这里将它们表示为 $M(obj)$ 和 $M(bck)$,其中 obj 为目标, bck 为背景。则分割问题可以用一个标记过程来表述:给定种子像素点集合 I_{seed_obj} 和 I_{seed_bck} ,给予每个像素点一个标记 $W_i \in \{obj, bck\}$,使得全局标记场 W^* 最大化一个后验概率:

$$W^* = \arg \max_w P(W | I) \quad (1)$$

本文将 $P(W | I)$ 表示为如下形式:

$$P(W | I) = \frac{1}{Z} \exp \left[- \sum_{i \in s} D_i(W_i, I) - \lambda \sum_{i \in s} \sum_{j \in N(i)} V_{ij}(W_i, W_j, I) \right] \quad (2)$$

其中 s 表示图像网格上的像素位置, $N(i)$ 表示任意位置 i 的一个邻域, λ 是相对权重, Z 是归一化常数, D_i 是单元势能(unary potential), V_{ij} 是邻域势能(pairwise potential)。

本文对式(2)做如下简化： D_i 的计算只在 $N(i)$ 内进行； V_{ij} 的计算只涉及到相邻位置 i, j 。因此最大化 $P(W|I)$ 可表示为最小化以下能量的形式：

$$E = \sum_{i \in S} D_i(W_i, N(i), M(W_i)) + \lambda \sum_{i \in S} \sum_{j \in N(i)} V_{ij}(W_i, W_j, I_i, I_j) \quad (3)$$

以下各节将分别详述对 D_i 和 V_{ij} 的建模以及对能量 E 的最小化。

3 纹理的建模

纹理是图像理解中的一个非常重要的概念。有一类基于特征原型的纹理建模方法被广泛应用于纹理分类和检索中^[3-5]，并得到了很好的结果，本文将其统称为基于Texton的方法。在这一节中，将首先介绍基于Texton的纹理建模方法，并讨论如何对其进行完善，以获得交互式方法所需的高效性。

3.1 滤波器组

在基于Texton的方法中，通常采用一组不同类型、方向和尺度的滤波器与图像做卷积，来获得纹理的特征。本文的滤波器组以LM滤波器组^[4]为基础构建。这里使用24个DooG滤波器(3个尺度，4个方向，2个相位)，3个尺度的LOG滤波器，3个尺度的Gauss滤波器，并增加了1个灰度滤波器(二维冲击函数)。对于彩色图像，灰度滤波器被应用到HSV 3个通道。然后将彩色图像转换为灰度图像，再与其它滤波器进行卷积。每个点的滤波响应将被连接在一起组成一个特征向量。不难得出，对于灰度图像，每个点的特征向量是31维；而对于彩色图像，每个点的特征向量是33维。

在交互式分割中，训练样本(种子像素)的数量通常较少，在如此高维的空间中训练将会导致所谓的“维数灾难”。其次，在高维特征空间中进行计算复杂度过高，无法达到交互式应用对速度的要求。本文采用最大响应准则和Fisher判别准则来对特征空间进行降维。

3.2 Fisher判别准则

对高维特征空间进行降维，通常会用到PCA方法。但PCA寻找的是有效表示所有样本的主轴方向，而不是有效的分类方向。本文采用如下Fisher判别函数作为启发式准则，来寻找有利于分类的方向：

$$J(v^i) = |\bar{m}_{\text{obj}}(v^i) - \bar{m}_{\text{bck}}(v^i)|^2 / [s_{\text{obj}}^2(v^i) + s_{\text{bck}}^2(v^i)] \quad (4)$$

其中 $J(v^i)$ 为第 i 个特征判别能力的度量， $\bar{m}_{\text{obj}}(v^i)$ 和 $\bar{m}_{\text{bck}}(v^i)$ 为目标和背景种子像素在该特征上的响应均值， $s_{\text{obj}}^2(v^i)$ 和 $s_{\text{bck}}^2(v^i)$ 为目标和背景种子像素在该特征上的响应方差。Fisher判别准则的意义是很直观的：若两类样本在该特征上的响应均值差别较大，而散布程度较小，则该特征更适合用来分类。本文采用的特征选择方法如下：由于灰度和色彩通常是非常有效的特征，因此直接保留灰度滤波器的响应；对LoG, Gauss滤波器响应进行排序，保留响应最强的2个滤波器；对DooG滤波器响应进行排序，保留其中最大响应的12个滤波器，并用Fisher判别准则从中选择 M 个判

别能力最强的滤波器。需要说明的是，Fisher准则通常用来选择降维的最佳投影方向。这里并不计算最佳投影方向，而仅以其作为启发式准则，采用贪心法保留 $J(v^i)$ 较大的 M 个DooG特征，丢弃其他的特征。在本文稍后给出的实验结果中， M 值均取为4。因此，对于灰度图像，降维后的特征是7维；而对于彩色图像，降维后的特征是9维。

3.3 目标和背景统计模型

为获得目标和背景的纹理统计模型，先对降维后的所有特征向量进行 K 均值聚类。 K 个聚类中心代表典型的响应向量，通常被称为Texton。为了将目标和背景区分开，一个好的 K 值应尽可能地将 $I_{\text{seed_obj}}$ 和 $I_{\text{seed_bck}}$ 中的点分到不同的聚类中。一般将 K 值取为30~50，实验证明这个取值范围在保证速度的同时，对多数图像均能获得较好的效果。对于每个聚类依次赋予编号1~ K ，并将图像 I 的每个像素标记为它的特征向量对应的聚类号，则 I 被映射到一个标记图 I_{texton} 上。 I_{texton} 的每个像素在1~ K 之间取值，该值等于该像素对应的聚类号。本文将 I_{texton} 中目标和背景种子点的直方图作为目标和背景的统计模型：

$$M(W) = \text{hist}(z; W) = \frac{1}{|I_w|} \sum_{i \in I_w} \delta(z - I_{\text{texton}}(i)), \quad z \in 1, 2, \dots, K \quad (5)$$

其中 z 为直方图的位(bins)， δ 是Dirac函数。

4 模型的建立和求解

4.1 单元势能 D_i

对于待分割图像中任意一个未确定类别的位置 i (即非种子点)，单元势能 D_i 定义为局部邻域 $N(i)$ 内的统计特征与 $M(W_i)$ 之间的距离，即将该位置标记为 W_i 的损失。用 I_{texton} 在 $N(i)$ 内的直方图 $\text{hist}(N(i))$ 表示 $N(i)$ 的统计特征，则 D_i 就是 $M(W_i)$ 和 $\text{hist}(N(i))$ 之间的直方图距离。直方图距离度量方式有很多种，这里采用 χ^2 距离。 χ^2 距离不仅考虑对应位之间的平方差异，而且用该位的值来规范化该差异，这样可以在一定程度上平衡不同位之间相对作用的大小。则 D_i 定义如下：

$$D_i(W_i, N(i), M(W_i)) = \text{dist}(M(W_i), \text{hist}(N(i))) = \chi^2(M(W_i), \text{hist}(N(i))) \quad (6)$$

其中两个直方图 h_1 和 h_2 之间的 χ^2 距离定义为

$$\chi^2(h_1, h_2) = \sum_{n=1}^{\#\text{bins}} \frac{(h_{1n} - h_{2n})^2}{h_{1n} + h_{2n}} \quad (7)$$

$\#\text{bins}$ 为直方图的位数。对于种子点位置，将其硬性赋为用户交互时给定的标号：

$$\left. \begin{aligned} D_i(W_i = \text{obj}; i \in I_{\text{seed_obj}}) &= 0, \\ D_i(W_i = \text{bck}; i \in I_{\text{seed_obj}}) &= +\infty \end{aligned} \right\} \quad (8)$$

$$\left. \begin{aligned} D_i(W_i = \text{obj}; i \in I_{\text{seed_bck}}) &= +\infty \\ D_i(W_i = \text{bck}; i \in I_{\text{seed_bck}}) &= 0 \end{aligned} \right\} \quad (9)$$

4.2 邻域势能 V_{ij}

邻域势能通常是一个平滑项。为了保留边缘，这里采用文献[1]的邻域势能形式，对边缘处的平滑进行惩罚。该势能在 Ising 模型^[6]的基础上，度量邻域像素点色彩的差异，并避免对色彩对比度较高的区域进行平滑：

$$V_{ij}(W_i, W_j, I_i, I_j) = \delta(W_i \neq W_j) \cdot \exp(-\beta \|I_i - I_j\|^2) / \text{dist}(i, j) \quad (10)$$

其中 I_i 和 I_j 是相邻两点 i, j (这里采用 8 邻域系统) 的像素值 (在彩色图像中是三维矢量)。 $\text{dist}(i, j)$ 是 i, j 的欧氏距离。 β 设置为

$$\beta = (2 \langle (I_i - I_j)^2 \rangle)^{-1} \quad (11)$$

$\langle \rangle$ 表示在整幅图像上求均值。 β 的作用类似于一个归一化因子，用来适当调整式(10)中指数项在高对比度和低对比度处的差异。邻域势能的引入，起到了平滑分割结果，保留强边缘的作用，解决了 Texton 方法应用到分割中所引起的分割边界偏移问题^[3]。

4.3 模型的求解

为获得最优分割，需要最小化全局能量式(3)。文献[7]中证明了式(3)这种形式的能量最小化问题，可以转化为图割问题来求解。为使论述清楚，这里首先对流网络(flow network)的基本概念做一些介绍。

一个图(graph) G 通常由顶点集合 V 和边的集合 E 构成。流网络是一个有向图，其中给出了两个特殊的顶点，分别称为源(source)和汇(sink)。网络中的每条有向边 $(u, v) \in E$ 均有一非负容量 $c(u, v) \geq 0$ 。一个典型的流网络见图 1，其中 s, t 分别为源点和汇点。未标记出方向的边为双向边。

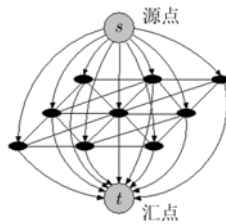


图 1 流网络

流出 s (或流进 t) 的总流量被称为网络流量的值，记为 $|f|$ 。网络 G 中的一个割(cut) (S, T) 将 V 划分为不相连的两部分 S 和 $T = V - S$ ，使得 $s \in S$ ， $t \in T$ 。割的容量定义为

$$c(S, T) = \sum_{u \in S} \sum_{v \in T} c(u, v) \quad (12)$$

最大流/最小割定理^[8]证明了：一个流网络的最大流量值等于最小割的容量，即

$$\max(|f|) = \min(c(S, T)) \quad (13)$$

除 s, t 外，用网络中的每个结点来代表图像的像素点。像素结点之间的连接称为 n -连接，像素结点与 s 和 t 的连接称为 t -连接。 n -连接代表了图像的邻域系统，而 t -连接代表

了像素与标记的关系。定义 s 为 obj 标记， t 为 bck 标记， λV_{ij} 为任意两相邻位置 i, j 之间 n -连接的容量 (本文实验中均将 λ 取为 1.25)， $D_i(W_i = \text{obj})$ 为 i 与 t 之间 t -连接的容量， $D_i(W_i = \text{bck})$ 为 i 与 s 之间 t -连接的容量。则该网络的最小割 (S, T) 定义了最小化式(3)能量的一个分割，证明见文献[2]。根据最大流/最小割的等价性定理，可以采用求解网络最大流的算法来获得最小割。这里采用文献[9]中的最大流算法。

最小割所定义的分割赋予了每个位置唯一的标记。种子点标记将与用户给定值相同，这是通过式(8)和式(9)中的定义，将 $+\infty$ 容量赋给与种子点标记相同的 t -连接实现的。 $+\infty$ 容量保证了该 t -连接不会出现在最小割中。实际上，只需 t -连接值稍大于任意一点处所有 n -连接值之和，即可保证不会被割去。因此，在程序编制中，我们将 $+\infty$ 值重设为 K ：

$$K = 1 + \max_i \sum_{(i, j) \in N} \lambda V_{ij} \quad (14)$$

在获得初始分割结果后，用户可能会进行再编辑。文献[2]建议仅仅将新种子点做标记，重新计算最小割。本文采取不同的做法，将新种子点加入到目标/背景种子点集合中，重新计算式(9)中的目标/背景直方图，并按式(6)，式(8)，式(9)更新 D_i ，重设网络所有连接的权值，再次求解最小割。这种方法在速度上会稍有损失，却充分利用了新种子点提供的信息来更新目标和背景模型，从而使简单的再编辑可以显著改善分割结果。

5 实验结果与分析

本文方法可广泛应用于多种类型的图像分割问题中。图 2 给出了将其应用到遥感图像分割和自然图像编辑中获得的部分结果，图像实际大小从 256×256 到 400×400 不等。图 2(b)，2(d)，2(g) 中的结果均是进行二次编辑后得到的，这是由于平滑项通常会造目标凹凸部分分割不准确。从图 2(a) 中可以看到，在用户标记为“目标”的两块水体区域被分割出来的同时，另外两块不联通的水体区域也被成功分割出来了。图 2(b) 中的水体部分被云层遮挡，对于这类有噪声的图像，采用一般的区域生长等方法进行分割通常会失败，而用本文方法则获得了很好的分割结果。图 2(c) 是在高分辨率遥感图像中对城市中的植被进行提取，图 2(d) 是对复杂边界的水体进行提取，这几个结果都表明了本方法在分割复杂边界目标时所提供的高效性和精确性。本文方法同样能有效的应用到灰度图像中。图 2(e)，2(f) 是在全色波段遥感图像中对田地和不同植被的分割结果。除遥感图像分割应用外，在图 2(g)，2(h) 中还给出了两个对自然图像进行分割的结果。

本文算法用 C++ 实现，在 VC 6.0 下编译通过，未做任何优化。在 P42.8GHz 的计算机上进行测试，对一幅 256×256 的彩色图像，初始分割需要约 1~2s；对于 400×400 的彩色图像，初始分割需要约 4~5s。如果要求更好的速度性能，可以在训练和分割过程中对图像进行 1/2 降采样，这样能使

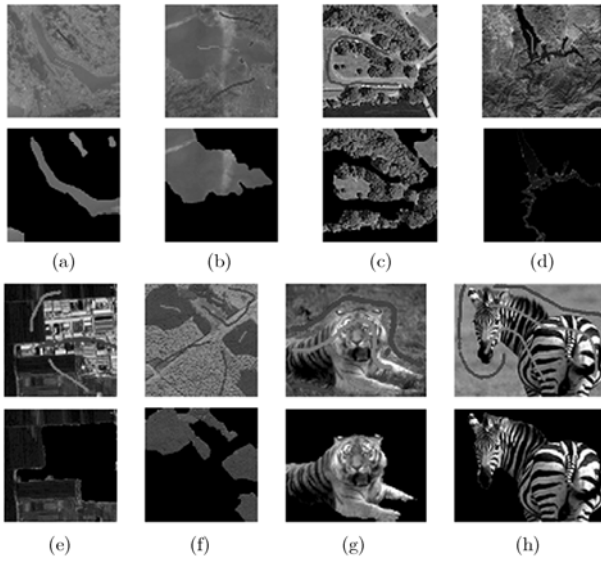


图2 用户交互及分割结果

速度获得3倍左右的提高,而不严重降低分割结果的精确性。进一步的代码级程序优化也可以提高性能。对于以上所有图像,用户对初始分割结果编辑后进行再分割,只需要不到1s的时间。

由于文献[2]方法并未公开源码,为便于比较,本文根据原文论述重新实现了其算法。首先给出了两种方法均能适用的图像及结果,见图3(a)。其次,在综合的纹理图像上对两种方法进行了比较,结果见图3(b)。从图3(b)中可以看出,文献[2]方法由于只考虑了灰度统计特性,因此对于两种有相似灰度而具有不同纹理模式的区域无法进行有效的分割,而本文的方法很好地解决了这一问题。此外,文献[2]的方法只适用于灰度图像,而利用彩色图像中更丰富的色彩信息往往能获得更好的分割结果,在这一点上本文方法也具有明显的优越性。在速度方面,文献[2]中没有给出具体的测试值。由于算法实现方式不同,测试平台也不同,因此这里仅做定性的评估,即本文算法在速度上有所降低,但仍然可以满足交互式应用的需要。

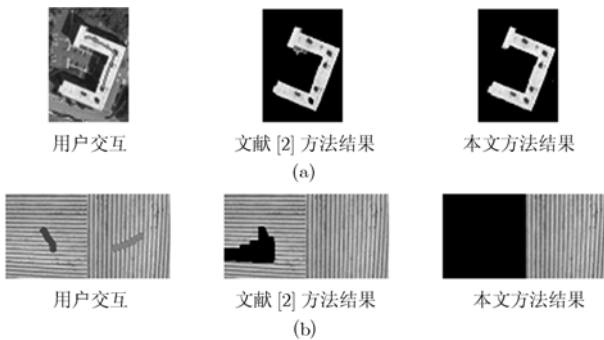


图3 两种方法结果比较

6 结束语

本文提出了一种新的基于图割的交互式分割方法。在交互式应用中,为了满足速度上的要求,通常不得不损失模型的精确度,其代价往往是不能获得好的分割结果。本文方法在保证速度的同时,依然保证了分割结果的准确性,从而证明了该方法的有效性。

参考文献

- [1] Rother C, Kolmogorov V, and Blake A. GrabCut — interactive foreground extraction using iterated graph cuts. *ACM Trans. Graphics, Los Angeles*, 2004, 23(3): 309-314.
- [2] Boykov Y and Jolly M P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. *International Conference on Computer Vision*, Vancouver, BC, Canada, 2001, 1: 105-112.
- [3] Malik J, Belongie S, Shi J, and Leung T. Textons, contours and regions: cue integration in image segmentation. *International Conference on Computer Vision*, Kerkyra, Corfu, Greece, 1999, 2: 918-925.
- [4] Leung T and Malik J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 2001, 43(1): 29-44.
- [5] Varma M and Zisserman A. Unifying statistical texture classification frameworks. *Image and Vision Computing*, 2004, 22(14): 1175-1183.
- [6] Stan Z L. *Markov Random Field Modeling in Image Analysis*. Inc. Secaucus, NJ, USA, Springer-Verlag New York, 2001, 18.
- [7] Kolmogorov V and Zabih R. What energy functions can be minimized via graph cuts? *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2004, 26(2): 147-159.
- [8] Thomas H C, Charles E L, Ronald L R, and Clifford S. *Introduction to Algorithm*, 2nd Edition. Cambridge, MA, USA, The MIT Press, 2001, 656.
- [9] Boykov Y and Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2004, 26(9): 1124-1137.

刘 嘉: 男, 1981年生, 博士生, 研究兴趣为遥感图像解译、计算机视觉、模式识别等。
王宏琦: 男, 1964年生, 研究员, 博士生导师, 目前主要研究领域为遥感图像处理及目标识别。