

基于多线索动态融合的交互式图像分割

耿晓伟^{1), 2)} 赵杰煜^{1), 3)}

¹⁾ (中国科学院计算技术研究所, 北京 100190) ²⁾ (中国科学院研究生院, 北京 100080)

³⁾ (宁波大学计算机科学技术研究所, 宁波 315211)

摘要 为了使图像分割效果稳定和提高分割算法的普适性, 提出了一种利用多线索动态融合进行图像分割的方法, 该方法首先将图像中的颜色信息、纹理信息、空间信息、边界信息融合到一个条件随机场模型中, 用于图像分割; 然后在对特定图像进行分割时, 由于该模型能够利用用户在分割前标定的前景和背景信息来建立一种衡量概率分布估计可靠性的标准, 并可通过该标准来组合条件随机场中的相关能量项, 使它们能够根据图像的内在属性进行动态融合, 从而提高了条件随机场图像分割的自适应性。实验结果表明, 该方法分割效果稳定、普适性强, 对一般的自然图像均能得到较好的分割效果。

关键词 交互式图像分割 条件随机场 动态融合 熵 香农散度 图分割

中图分类号: TP391.41 文献标志码: A 文章编号: 1006-8961(2010)01-0075-10

Interactive Image Segmentation Based on Multi-cue Dynamic Integration

GENG Xiaowei^{1), 2)}, ZHAO Jieyu^{1), 3)}

¹⁾ (Institute of Computing Technology, Chinese Academy of Science, Beijing 100190)

²⁾ (Graduate University of Chinese, Chinese Academy of Science, Beijing 100080)

³⁾ (Research Institute of Computer Science and Technology, Ningbo University, Ningbo 315211)

Abstract In order to improve the stability and generality of image segmentation, a new method based on the dynamic multi-cue integration is presented in this paper. It fuses color, texture, spatial and edge information by the construction of a conditional random field for image segmentation. Besides, in the course of interactive image segmentation, the proposed method is able to take full advantage of the user labeled information about the foreground and background, and builds a certain standard to measure the reliability of established probability distribution, with which to assemble the energy terms in the conditional random field. This makes the related energy terms fusing dynamically in accordance with the inner property of the image, and improves the segmentation power of the model. The experimental results demonstrate that the proposed method performs steadily and works well on various natural images.

Keywords interactive image segmentation, conditional random fields, dynamic integration, Jensen-Shannon divergence, graph cut

0 引言

图像分割是图像工程中一个非常重要的研究内

容, 在图像处理和机器视觉领域都有广泛的应用, 但是, 由于图像分割的特点和计算机自身的局限性, 直到现在, 该问题仍然未能得到彻底的解决。因为一方面要想获得良好的图像分割结果, 必然要求分割

基金项目: 国家自然科学基金项目 (60672071); 浙江省自然科学基金重大项目 (D1080807); 教育部新世纪优秀人才支持计划 (NCET-05-0534)

收稿日期: 2008-01-08 改回日期: 2008-12-05

第一作者简介: 耿晓伟 (1982—), 男, 中国科学院计算技术研究所计算机软件与理论专业硕士研究生。目前研究方向为图像处理、计算机视觉、模式识别。E-mail: gxw663@163.com

算法具备一定的图像分析和理解的能力,而这些技术是目前难以做到的,或者有些技术的解决与图像分割本来就具有相互依赖性;另一方面,在视觉领域,图像分割是一个介于底层视觉和中层视觉之间的问题,许多问题的解决需要借助图像分割的结果作为输入。基于这些原因,图像分割是目前图像处理领域研究的一个热点问题,同时也是一个难点问题。尽管如此,近几年来,随着图像处理领域相关技术的发展,图像分割技术已经获得了很大的进步,特别是许多分割算法已经能够给出一些令人鼓舞的结果^[1]。

在图像处理领域,如今已出现了很多用于图像分割的模型与算法,其中具有代表性的是基于偏微分方程的模型和基于概率的模型。前者是依据图像的相关特征首先构造相应的能量函数,然后把它转化为偏微分方程求解,虽然这种方法的一大优点是能量函数可以采用多种表示方式,但是由于目前缺少比较好的对相应能量函数进行优化的方法,仅仅能够保证得到局部最优解,因此分割的结果对初始值有较强的依赖性,一旦初始值选择不合理,那么分割的结果也就相应地比较差。后者是由 Geman 等人在文献 [2] 中首先提出的一种图像处理方法,它是以贝叶斯公式和马尔科夫随机场为理论基础,利用马尔科夫随机场和吉布斯场的等价性建立的一个模型,通常称为 MRF-MAP (Markov random field maximum a posterior) 框架,该模型最终也归结为一个能量优化问题。解决该能量优化问题最早提出的算法是模拟退火算法,虽然它可以收敛到最优解,但是优化时间比较长,难以在实际的图像处理中应用。为了提高运行效率, Besag 提出利用 ICM 算法^[3]来求解类似的能量函数的最优值,虽然速度有所提高,但是该算法求得的结果不是一个全局最优解,仅仅是一个局部最优解。由于缺少较好的解决该问题的方法,从而导致使用概率模型处理图像的技术一直停滞不前。近几年来,经过众人的努力,终于发现了几种新的算法和改进方法,大大增多了基于概率模型的能量函数优化的途径。例如,由 Boykov 等人改进的最大流最小割方法^[4]和基于信度传播理论的方法^[5]以及随后出现的改进算法^[6],不仅大大提高了运算效率,而且改进了分割效果。由于能量优化方面取得的进步改变了以往使用 MRF 模型遇到的优化时间太长的问題,从而使 MRF-MAP 成为目前最流行的图像处理模型之一。

1 本文的工作

如前所述,好的图像分割结果需依赖于对图像多种特征的把握。本文试图利用多种信息动态融合的方法来完成一般自然图像的分割。图像分割大体上可分为基于边界的分割和基于区域的分割两种方法。虽然这两种方法利用的原理是相近的,即图像特征(图像特征可以是图像的像素值、灰度值、纹理或者其他有意义的图像属性)在相同区域具有一定的连续性,在不同区域具有一定的跳跃性,但是这两种方法的分割效果并不完全相同,为了弥补两者之间的不足,最近又出现了将两种方法融合在一起对图像进行分割的方法。例如, lazy snapping^[7]就是其中一个代表性的例子。但由于这些方法都是针对图像中的颜色信息进行建模,因此难有太广的普适性。以 Rother 等人提出的 GrabCut 分割方法^[1]为例,当前景和背景的颜色对比度明显时,会取得比较好的分割结果,而当前景和背景的颜色对比度较差时,则分割效果就不很理想,从而在分割的后处理中,需要用户提供更多的交互信息才能得到满意的分割结果。众所周知,一般图像中,除了颜色信息之外,图像的纹理信息也很重要,因此,如果能把纹理信息融合到分割算法中,就有可能扩大它的适用范围。尽管将纹理信息与颜色信息融合在一起进行图像分割的方法已有不少^[8-9],但是这些方法仅仅将纹理信息和颜色信息进行静态融合,在对任意图像进行分割时,人为地认定这两种信息对分割结果的影响比重是固定的,由于未能充分考虑每幅图像的内在属性,因此模型往往难有普适性。本文在挖掘图像中的纹理信息以及颜色信息的同时,考虑到每幅图像中纹理信息与颜色信息对分割结果的贡献不同,提出了一种新的方法将两种信息动态地融合在一起,该方法一方面不仅克服了以前纹理信息和颜色信息以固定比例融合对图像特征进行刻画不准确的缺点,而且充分把握了图像中的目标信息;另一方面,由于一般的自然图像的边界复杂多变,因此本文在模型实现中,对边界信息进行了更强的刻画,并利用琴森-香农散度提取图像的边界信息。由于该方法对噪声不敏感,因此能更好地提取边界信息。实验结果表明,本文算法对一般的自然图像的分割结果较好,由于采用了动态融合技术,因此相应地普适性也较强,与以往算法相比,分割的结果也更稳定。

2 模型建立

条件随机场是一个 2 维离散网格上的随机场, 为了使条件随机场的输出能够表示图像分割的结果, 需要根据待分割图像来构造该 2 维网格, 通常的 2 维网格都是整数网格, 上面的每一个点既对应图像中一个像素, 同时又对应条件随机场中的一个随机变量。如果设构造的 2 维网格为 S , 2 维网格上每一点的位置为 s 那么有

$$S = \{s \mid s = (i, j), i \in \{1, 2, \dots, W\}, j \in \{1, 2, \dots, H\}\} \quad (1)$$

这里, W 为待分割图像的宽度, H 为图像的高度。这样待分割图像 X 就可以记为 $X = \{x_s, s \in S\}$, 其中, x_s 为彩色图像中点 s 的像素值, 在 RGB 颜色空间中, 它是一个 3 维向量, 每一维都是一个 0 到 255 之间的整数。相应地, 条件随机场 Y 可以记为 $Y = \{y_s, y_s \in \{F, B\}, s \in S\}$, 其中, y_s 为待分割图像对应像素点 x_s 的分割标记。对图像进行分割也就是对图像的每一个像素点进行分割, 一般分为两类, 即前景 F 和背景 B 。前景就是需要保留的部分, 背景为需要剔除的部分。根据条件随机场的性质, 在给定 X 的情况下, Y 满足如下的条件概率^[10]:

$$p(Y|X) \propto \exp \left\{ - \left[\sum_m \sum_s \lambda_{m,s} u_{m,s}(y_s, X, s) + \sum_n \sum_{\{s,k\} \in N} \mu_{n,\{s,k\}} v_{n,\{s,k\}}(y_s, y_k, X, s, k) \right] \right\} \quad (2)$$

其中, N 表示随机场的邻域系统, 即 $N = \{\{s, k\} \mid s$ 与 k 相邻, $s, k \in S\}$ 。本文中采用 8 邻域系统。不同上下标的 λ, μ 为待估参数, $u(\dots)$ 和 $v(\dots)$ 分别是随机场的 1 维势函数和 2 维势函数。从而要估计原始图像的最优分割标记 Y^* , 也就是最大化式 (2), 即

$$Y^* = \arg \max_Y p(Y|X) = \arg \min_X E(Y, X, \theta) \quad (3)$$

$$E(Y, X, \theta) = \sum_m \sum_s \lambda_{m,s} u_{m,s}(y_s, X, s) + \sum_n \sum_{\{s,k\} \in N} \mu_{n,\{s,k\}} v_{n,\{s,k\}}(y_s, y_k, X, s, k) \quad (4)$$

其中, θ 代表参数集 $\{\lambda_{m,s}, \mu_{n,\{s,k\}}, s, k \in S; m = 1, 2, \dots; n = 1, 2, \dots\}$ 。

本文希望模型能够融合图像中的纹理信息、颜色信息、空间信息和边界信息, 由于条件随机场的结构中已经蕴含了图像的空间信息和部分边界信息,

因此只需对其他两项进行建模就可以了。然而, 由于考虑到模型对边界信息刻画得较弱的特点, 因此为模型增加了一个边界先验项, 以便更好地把握图像的边界信息。根据上述分析并结合式 (4) 可知, 这里的 $m = 2, n = 3$ 。则能量函数可以写为

$$E(Y, X, \theta) = \sum_s \lambda_{c,s} U_c(y_s, X, s) + \sum_s \lambda_{t,s} U_t(y_s, X, s) + \sum_{\{s,k\} \in N} \mu_{cp,\{s,k\}} W_{cp}(y_s, y_k, X, s, k) + \sum_{\{s,k\} \in N} \mu_{tp,\{s,k\}} W_{tp}(y_s, y_k, X, s, k) + \sum_{\{s,k\} \in N} \tau_{ep,\{s,k\}} W_{ep}(y_s, y_k, X, s, k) \quad (5)$$

一般地, 根据 $E(Y, X, \theta)$ 中各个能量项的作用不同, 将它们分别称为颜色似然项 $U_c(y_s, X, s)$, 纹理似然项 $U_t(y_s, X, s)$, 颜色先验项 $W_{cp}(y_s, y_k, X, s, k)$, 纹理先验项 $W_{tp}(y_s, y_k, X, s, k)$ 以及边界先验项 $W_{ep}(y_s, y_k, X, s, k)$ 。

2.1 能量项构成

2.1.1 颜色似然项 ($U_c(y_s, X, s)$)

根据 MRFMAP 框架理论^[11], 颜色似然项可采用概率分布密度的负对数来估计。这里假设彩色图像的每一点的分割标记在颜色信息中只与对应的像素有关, 从而有

$$U_c(y_s, X, s) = - \ln(f(x_s | y_s)) \quad (6)$$

由于自然彩色图像的颜色概率分布一般比较复杂, 所以通常采用混合高斯的方法^[12]来进行估计, 但是由于这往往要求预先设定高斯核的个数, 从而在一定程度上限制了混合高斯函数模拟实际的概率分布的逼真程度, 而且如果采用 EM 算法进行参数的估计, 还需要花费大量的时间。本文采用核函数^[13]的方法来估计颜色在特征空间中的分布情况, 由于只需要设置一个带宽参量 h , 不需要其他待估参数, 从而大大减少了运算量。

为了获得前景和背景的颜色概率分布, 首先需要从交互提示的区域提取概率信息, 然后用核函数的方法来逼近颜色概率分布。通常的做法是首先从用户选定的区域进行抽样, 然后构造概率分布函数, 但是抽样也有缺点, 它难以保证所得到的数据能真实地代表选定区域的特征, 为此, 本文首先对彩色图像的选定区域进行均值漂移 (mean shift) 滤波, 这样, 就会有许多的像素点在均值漂移后具有相同的模式值 (mode), 由于模式值是一系列特征点的总体反映, 它

能够比较好地反映区域的特征,因此可以利用模式值作为该区域的特征值来刻画概率分布,这样以后在估计每个像素点分别属于前背景概率时就可以减少计算量。如果设选定的前(背)景区域有 M 个像素点,该区域滤波后有 N 个模式值,具有第 i 个模式值 P_i 的像素有 C_i 个,选定的颜色带宽参量为 h_c ,那么前(背)景的概率密度估计可以写为

$$\hat{f}(x) = \frac{1}{M(h_c)^d} \sum_{i=1}^N C_i K \left| \frac{x - P_i}{h_c} \right| \quad (7)$$

其中, $K(\cdot)$ 为选定的核函数,这里取为高斯核函数,颜色带宽参量选为 $h_c = 10$ d 为维数,由于彩色图像有 3 个通道,所以 d 取为 3。

2.1.2 纹理似然项 ($U_T(y_s, \mathbf{X}, s)$)

要得到纹理似然项的具体表示形式,首先要有图像的纹理特征信息。由于纹理不是像素级的特征,因此需要考虑像素及其周围像素的排列方式。纹理特征的提取方法,源于以下人类视觉系统对纹理的分辨方式:当两种未知纹理的图像投影到人眼视网膜上时,视觉系统会用一组滤波器对图像做卷

积滤波,如果这两种图像经滤波输出后得到的分布形式是相同的,那么视觉系统就认为这是两种相同的纹理,反之则是不同的。根据这一原理就可以设计一组滤波器,然后即可利用它们的输出来分析图像的纹理特征^[14]。这里采用 Gabor 滤波器来提取图像的纹理信息,Gabor 滤波器可以看作是由高斯函数调幅的复正弦波^[18],它的一般表达形式如下:

$$g(x, y) = \left| \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi jfx \right] \right| \quad (8)$$

式中 σ_x, σ_y 分别是滤波器在 y 轴和 x 轴上的尺度半径, f 是滤波器的中心频率。由于 Gabor 滤波器可以通过改变中心频率和方向来调节它在不同尺度和不同方向上的选择性,因此可以用于提取各种粒度、各种方向的纹理特征。图 1 是 Gabor 滤波器在不同尺度和不同方向上对图像进行滤波的一个例子。根据滤波器的输出,就可以把对应得到的特征图像集组成一个向量图像用来描述纹理的分布,这样,每一

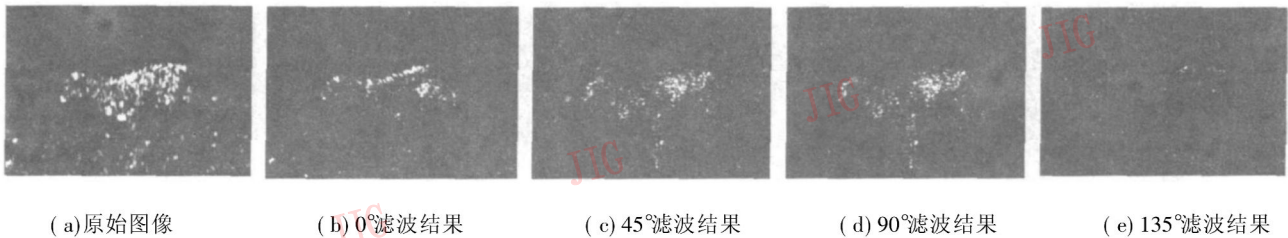


图 1 图像在不同尺度不同方向下的 Gabor 函数滤波的结果

Fig 1 Results of image filtered by different orientation and different scale gabor function

原始图像的像素就对应一个纹理向量,向量的维数即为特征图像的个数。对于纹理分布同样可采用核函数的方法进行估计,显然前景与背景的纹理特征的分布也是复杂的,采用与计算颜色似然项类似的方式,也可利用核函数的方法来近似这种分布。这样纹理似然项可定义如下:

$$U_T(y_s, \mathbf{X}, s) = -\ln(\hat{f}(T(\mathbf{x}_s) | y_s)) \quad (9)$$

式中, $T(\mathbf{x}_s)$ 为原始图像在点 s 处对应的特征图像的纹理向量,其中式 (9) 中的纹理概率分布估计可具体表示为

$$\hat{f}(x) = \frac{1}{M(h_T)^d} \sum_{i=1}^N C_i K \left| \frac{x - Q_i}{h_T} \right| \quad (10)$$

对于纹理概率分布的估计,由于原始图像为向量图像,难以直接滤波,这里选择对原始图像在 HSI 颜色空间下的 I 分量图像进行操作,即先用不同尺度和不同方向的滤波器对它进行滤波(这里选择 0° ,

45° , 90° , 135° 等 4 个方向,对应的中心频率的下界为 0.2 上界为 0.6 不同的尺度可以通过调节 Gabor 滤波器的中心频率得到^[15])就可以得到 4 幅滤波后的输出图像;然后再使用均值漂移来对原始图像的对应选定区域的纹理向量进行滤波,求取前景和背景的模式值,进而构造式 (10) 的纹理概率密度函数。这里 Q_i 为经纹理滤波后的第 i 个模式值,纹理带宽参量 h_T 取为 20 其他变量的意义与颜色似然项类似。

2.1.3 颜色先验项 ($W_{CP}(y_s, y_k, \mathbf{X}, s, k)$) 与纹理先验项 ($W_{TP}(y_s, y_k, \mathbf{X}, s, k)$)

先验项可以看作是根据图像特征的连续性和跳跃性的特点来指导能量函数进行图像分割的一种能量项,由于图像分割本身是一个不适定问题,其往往要加入一定的限制项,才能保证得到一个较为合理的分割结果,所以先验项通常又称为正则项。对于

一般的自然彩色图像来说, 其图像特征在相同区域都有一定的连续性, 在不同区域有一定的跳跃性, 这就是说, 在建模时可以加入类似的限制, 以便使分割结果更加符合图像特征的分布规律。这里采用类似文献 [4] 的方法将这两项定义为

$$W_{CP}(y_s, y_k, \mathbf{X}, \mathbf{s}, \mathbf{k}) = [y_s \neq y_k] \exp \left\{ - \frac{\|\mathbf{x}_s - \mathbf{x}_k\|^2}{2\sigma_c^2} \right\} \quad (11)$$

$$W_{TP}(y_s, y_k, \mathbf{X}, \mathbf{s}, \mathbf{k}) = [y_s \neq y_k] \exp \left\{ - \frac{\|\mathbf{T}_s - \mathbf{T}_k\|^2}{2\sigma_T^2} \right\} \quad (12)$$

式中, $[\cdot]$ 是布尔运算, σ_c , σ_T 分别定义为 $\langle \|\mathbf{x}_m - \mathbf{x}_n\|^2 \rangle$, $\langle \|\mathbf{T}_m - \mathbf{T}_n\|^2 \rangle$, 前者是所有相邻像素颜色距离的平均值, 后者是所有相邻像素纹理向量距离的平均值, $\langle \cdot \rangle$ 是对所有 \cdot 求平均值。对于式 (12), 原则上不同纹理滤波后, 其图像像素值具有不同的概率分布规律, 应该通过求取其不同概率分布函数之间的距离来描述纹理的差异^[10], 但是这样时间代价比较高, 为了方便, 本文采用纹理向量的欧氏距离来近似图像两点附近的纹理特征的差异, 效果也比较好。

2.1.4 边界先验项 $W_{EP}(y_s, y_k, \mathbf{X}, \mathbf{s}, \mathbf{k})$

边界先验项是用于更好地刻画图像的边界而引入的。事实上, 颜色先验项和纹理先验项已部分具有刻画图像边界的能力, 但是, 当边界带有噪声或者比较模糊时, 则模型对边界的把握就不是很准确。为此本文采用琴森-香农散度来进一步提取图像的边界特征。该方法的抗噪声能力较强, 即使图像的边界受到噪声的污染, 仍然能够较好地把握边界特征^[17]。琴森-香农散度是计算不同概率分布之间差异的一种度量。对于两个离散概率分布 P_1 和 P_2 , 琴森-香农散度的定义如下

$$J(P_1, P_2) = H(\alpha P_1 + (1-\alpha)P_2) - (\alpha H(P_1) + (1-\alpha)H(P_2)) \quad (13)$$

其中, $0 < \alpha < 1$ 是这两个概率分布的相对权重, 如果权重相同, 则一般设 $\alpha = 1/2$ 。 $H(P) = - \sum_{i=1}^n P_i \log P_i$ 是概率分布 P 的香农熵, P_i 是概率分布中 i 事件发生的概率, n 为事件的总数。

为了计算图像每一点处的琴森-香农散度, 首先要构造一个模板窗口, 并把该窗口分为左右相同的两个子窗口, 不妨记为 A 和 B ; 然后让该模板窗口按图像的行方向滑动, 每滑动一次, 分别计算子窗口 A 和 B 覆盖下的图像区域的灰度值 (图像 HSI 颜色空

间的 I 分量) 的概率分布; 最后依据式 (13) 计算琴森-香农散度, 并将其作为模板窗口覆盖下的图像区域中心像素点的散度值。当子窗口 A 和 B 的分隔边界和图像区域中的某一垂直于移动方向的边界重合时, 则计算得到的散度值将达到一个极大值, 利用这个特点, 就可以得到图像中的边界信息。一般地, 由于图像中的边界方向不会总是与图像的行方向垂直, 因此通常需要通过分别沿 0° , 45° , 90° , 135° 4 个方向移动窗口来计算像素在各个方向上的散度, 并把其中最大的散度值作为该点的最终散度值, 进而计算相应的方向, 具体计算方法见文献 [18]。这样就得到了图像的边界信息。如果记图像中点 s 处的琴森-香农散度为 J_s , 对应的方向为 α_s , 那么相应的边界先验项定义为

$$W_{EP}(y_s, y_k, \mathbf{X}, \mathbf{s}, \mathbf{k}) = [y_s \neq y_k] \times \exp \left\{ \frac{(J_s J_k \tilde{J}_s \tilde{J}_k)^2}{-2\sigma_J^2 (\varepsilon + (\text{dir}(\mathbf{s}, \mathbf{k}) - \alpha_s)^2 + (\text{dir}(\mathbf{s}, \mathbf{k}) - \alpha_k)^2)} \right\} \quad (14)$$

其中, $\text{dir}(\mathbf{s}, \mathbf{k})$ 表示图像中点 s 和点 k 的连线相对于图像行方向的倾角。 ε 是一个偏置参数, 选为 10^{-5} , σ_J 为一常量参数, 这里取为 $0.25 \tilde{J}_s$, \tilde{J}_s 是反映图像中点 s 的散度与邻域散度均值相对大小的一个量, 其定义为

$$\tilde{J}_s = \left| J_s - \frac{1}{n} \sum_{z \in N(s)} J_z \right| \geq 0 \left| J_s - \frac{1}{n} \sum_{z \in N(s)} J_z \right| \quad (15)$$

式中, $N(s)$ 表示点 s 的邻域像素集合, n 为邻域的大小, 这里取为 8, $[\cdot]$ 仍为布尔运算符。

2.2 参数估计

上一节已经详细介绍了各能量项的具体构造方式, 剩余的任务就是如何估计模型中的参数, 大多数的模型都采用了经验估计的方式, 并使用固定值作为这些参数的估计值。但这些经验值往往适用范围比较窄, 为此本文采用了一种“相对正确率”的方法来完成参数的估计, 得到了较好的结果, 实验表明, 本文方法具有更好的普适性。

一般而言, 颜色信息和纹理信息在图像分割中的贡献是不同的。在前景与背景颜色相似的图像中, 仅仅利用颜色信息进行分割很有可能失败; 同理, 当前景与背景纹理相似时, 只利用纹理信息进行分割同样会失败。因此, 如何有效融合这两种信息, 成为构造一个好的分割算法的关键。本文算法中的参数估计充分利用了交互输入的信息, 使它们的估计值能够尽可能较好地反映图像中的各种信息的构

成情况。由于在分割图像之前,用户都要选定图像中的前景和背景的代表区域,所以可以利用每个选定区域的颜色信息、纹理信息来构造相应的概率分布,进而构造对应的似然项。通常,都把构造的概率分布近似为前景和背景的概率分布,这样,在一般情况下,就默认属于前(背)景的像素点在构造的前(背)景概率分布中的概率比它在构造的背(前)景概率分布中的概率大。

然而实际情况并非总是如此,对于有些像素点,虽然它应该被分为前(背)景,但是在构造的背(前)景的概率表达式中所求得的概率却比在构造的前(背)景的概率表达式中求得的概率大。本文把这种错误分类的概率称为相对错误率,它反映了构造的前景概率分布函数的一种不可靠性,即属于前景的像素被误分为背景的概率。同样的道理,也可以得到构造的背景概率分布函数的不可靠性。通过比较图像中颜色概率分布与纹理概率分布的不可靠性,就可以合理地组合颜色和纹理的相关能量项,以便得到比较好的分割结果。

这里假设选定区域中的前景和背景像素点集分别为 $\{\mathbf{x}_i^F, i=1, 2, \dots, M\}$ 和 $\{\mathbf{x}_i^B, i=1, 2, \dots, N\}$, 由前景和背景分别构造的颜色分布函数为 $f_c^F(\mathbf{x})$ 和 $f_c^B(\mathbf{x})$, 选定的前景区域和背景区域中的纹理特征向量集为 $\{\mathbf{T}_i^F, i=1, 2, \dots, M\}$ 和 $\{\mathbf{T}_i^B, i=1, 2, \dots, N\}$, 由前景和背景分别构造的纹理分布函数为 $g_T^F(\mathbf{T})$ 和 $g_T^B(\mathbf{T})$, 那么根据上述分析,就可以将由前景构造的颜色概率分布函数的相对正确率定义为

$$\eta_c^F = 1 - \frac{1}{M} \sum_{i=1}^M f_c^B(\mathbf{x}_i^F) \quad (16)$$

由前景构造的纹理概率分布函数的相对正确率为

$$\eta_T^F = 1 - \frac{1}{M} \sum_{i=1}^M g_T^B(\mathbf{T}_i^F) \quad (17)$$

同理,定义由背景构造的颜色概率分布函数的相对正确率为

$$\eta_c^B = 1 - \frac{1}{N} \sum_{i=1}^N f_c^F(\mathbf{x}_i^B) \quad (18)$$

由背景构造的纹理概率分布函数的相对正确率为

$$\eta_T^B = 1 - \frac{1}{N} \sum_{i=1}^N g_T^F(\mathbf{T}_i^B) \quad (19)$$

这样,颜色项与纹理项在判断前景时的相对正确率的比例为 η_c^F / η_T^F , 在判断背景时的比例为 η_c^B / η_T^B 。这里,采用前景和背景进行几何平均的方法来计算似然项和先验项中颜色项与纹理项的组合比例,即

$$\sqrt{\frac{\eta_c^F \eta_c^B}{\eta_T^F \eta_T^B}} \quad (20)$$

考虑式(5)中的相关参数,本文假设构造的条件随机场为均匀各向同性的,即相关参数的具体值与像素点的位置和方向无关,那么就可以将式(5)改写为

$$E(\mathbf{Y}, \mathbf{X}, \theta) = \sum_s \lambda_C U_C(y_s, \mathbf{X}, s) + \sum_s \lambda_T U_T(y_s, \mathbf{X}, s) + \sum_{\{s, k\} \in N} \mu_{CP} W_{CP}(y_s, y_k, \mathbf{X}, s, k) + \sum_{\{s, k\} \in N} \mu_{TP} W_{TP}(y_s, y_k, \mathbf{X}, s, k) + \sum_{\{s, k\} \in N} \tau_{EP} W_{EP}(y_s, y_k, \mathbf{X}, s, k) \quad (21)$$

进一步地,根据上文对能量函数中相关项构成关系的分析,可将式(21)写成

$$E(\mathbf{Y}, \mathbf{X}, \theta) = \alpha \left| \sqrt{\eta_c^F \eta_c^B} \sum_s U_C(y_s, \mathbf{X}, s) + \sqrt{\eta_T^F \eta_T^B} \sum_s U_T(y_s, \mathbf{X}, s) \right| + \beta \left| \sqrt{\eta_c^F \eta_c^B} \sum_{\{s, k\} \in N} W_{CP}(y_s, y_k, \mathbf{X}, s, k) + \sqrt{\eta_T^F \eta_T^B} \sum_{\{s, k\} \in N} W_{TP}(y_s, y_k, \mathbf{X}, s, k) \right| + \tau \sum_{\{s, k\} \in N} W_{EP}(y_s, y_k, \mathbf{X}, s, k) \quad (22)$$

既然纹理和颜色相关项的参数是成比例的,那么通常需对它们进行归一化处理,从而得到能量函数的最终表示形式为

$$E(\mathbf{Y}, \mathbf{X}, \theta) = \mu \sum_s U_C(y_s, \mathbf{X}, s) + (1 - \mu) \sum_s U_T(y_s, \mathbf{X}, s) + \xi \left| \mu \sum_{\{s, k\} \in N} W_{CP}(y_s, y_k, \mathbf{X}, s, k) + (1 - \mu) \sum_{\{s, k\} \in N} W_{TP}(y_s, y_k, \mathbf{X}, s, k) \right| + \tau \sum_{\{s, k\} \in N} W_{EP}(y_s, y_k, \mathbf{X}, s, k) \quad (23)$$

其中, $\mu = \frac{\sqrt{\eta_c^F \eta_c^B}}{\sqrt{\eta_c^F \eta_c^B} + \sqrt{\eta_T^F \eta_T^B}}$, ξ 是似然项与先验项之间的一个比例关系,通常选为 17, τ 是边界项在能量函数中的一个比例常数,这里选为 5。

2.3 模型求解

前面已经提到,对能量函数进行优化已经提出了很多有效的算法,由于本文中要优化的能量函数满足文献[19]中提出的正则性条件,因此可以使用图分割算法进行求解,图分割算法的基础就是网络流理论中经常用到的最大流最小割算法。该算法首

先构造一个与原始图像对应的离散的 2 维网格; 然后分别设置相邻格点的边的权值, 同时, 增设源点 (前景) 和汇点 (背景) 两个特殊的点, 并且让网格中的每一个格点分别与这两点相连。根据前述的能量函数, 可以相应地先给网格中每一个点与源点和汇点相连的边分别赋一个权值, 然后对形成的带权网络求最大流最小割, 所得的最大流就是原来能量函数的最优值, 其对应的割集就是图像分割的结果。这里将根据图像构造的 2 维网格记为

$$\mathbf{R} = \{r \mid r = (i, j); i = 1, 2, \dots, W; j = 1, 2, \dots, H\} \quad (24)$$

增设的源点和汇点分别为 F, B 。让 \mathbf{R} 中的每一个点都与这两个点相连, 这样就可以给这个生成的网络的每一条边赋权值了。为了保证求得的最大流就是上述能量函数的最优值, 需要按照下述方式给每一条边赋以权值, 即首先给原来网格中形成的边赋权值

$$W_{p,q} = \left[\mu \exp \left| - \frac{\|x_p - x_q\|^2}{2\sigma_c^2} \right| + (1 - \mu) \exp \left| - \frac{\|T_p - T_q\|^2}{2\sigma_t^2} \right| + \tau \exp \left| - \frac{(J_p J_q \tilde{J}_p \tilde{J}_q)^2}{2\sigma_j^2 (\varepsilon + (\text{dir}(p, q) - \alpha_p)^2 + (\text{dir}(p, q) - \alpha_q)^2)} \right| \right] \quad (25)$$

其中, p, q 为 2 维网格 \mathbf{R} 中相邻的两个结点。

其次, 对于 \mathbf{R} 中任意的结点 r , 其与源点 F 和汇点 B 相连的边的权值则按照如下方式确定:

$$W_{r,F} = \mu U_C(F, X, r) + (1 - \mu) U_T(F, X, r) \quad (26)$$

$$W_{r,B} = \mu U_C(B, X, r) + (1 - \mu) U_T(B, X, r) \quad (27)$$

通过求解用上述方式构造的带权网络的最大流最小割, 所得的最大流就是原来能量函数的最优值, 形成的割集恰好是图像分割的结果, 更详细的分析见文献 [20]。

2.4 图像分割的后处理

一般的图像分割系统都要求是完备的, 也就是说, 要能够把图像中的任何需要分割的目标都分割出来, 当然通常的分割算法可能对于一些较为复杂的目标无法一次分割出来, 这就要求分割算法必须提供某种方法用来改进以前得到的分割结果, 直到得到满意的分割为止, 这就是所谓的图像分割的后处理。本文算法对一般的图像都能获得较好的分割效果, 对一些比较复杂的图像还可以通过后处理来改进初次得到的分割结果, 以便最终达到分割的要求。

如果在初次分割后, 有些细节没有得到满意的分割结果, 该分割算法还提供了一种交互的方法, 即允许用户选定不满意的区域, 并做进一步的提示, 那么分割系统就会在选定的区域上重新进行分割, 这样只需进行局部分割, 而不是进行整体分割, 就可以改进原来得到的分割结果, 这就大大提高了分割的效率, 而且对复杂的图像, 不需要花费太多的时间, 就可以得到很好的分割效果。

3 实验结果

为了验证算法的有效性, 本文收集了多种特征的图像进行了 3 组实验, 相关的分割结果见图 2。第 1 组是以颜色为主要特征的图像 (图 2(a) ~ 图 2(c)), 第 2 组是以纹理为主要特征的图像 (图 2(d) ~ 图 2(f)), 第 3 组是前景或背景都较为复杂的图像 (图 2(g) ~ 图 2(j))。为了说明模型的动态融合的有效性, 实验中分别记录了 3 组图像分割时对应的像素信息与纹理信息的组合比例 μ (见表 1)。从表 1 可以看出, 对于第 1 组图像, 相应的 μ 值较高, 都大于 0.5。通过分析原始图像和表 1 信息可以看出, 由于图 2(a) 的颜色信息和纹理信息都比较强, μ 值与 0.5 比较接近, 而图 2(b) 和图 2(c) 的颜色信息则比较突出, 所以 μ 值相应地也就较大, 这说明在分割中, 颜色信息比纹理信息的区分度更好。对于第 2 组图像, μ 的平均值低于 0.5, 由于这些图像大部分颜色都比较接近, 因此如果仅仅依靠颜色信息来进行分割, 已经无法得到较好的分割结果, 其 μ 值也反映出, 在分割中纹理信息占的比重更大一些。在第 3 组图像中, 挑选了一些稍微复杂的图像进行了分割实验, 效果基本令人满意, 根据本组的特点, 本文实验没有求取 μ 的平均值, 虽然 μ 值与 0.5 相差不大, 但在实际的图像处理中, 如果能够动态地改变 μ 值, 使模型可以根据图像的内在属性自动调节相关能量项的比例, 那么所得的分割效果将比利用固定比例进行分割的效果要好得多。其中, 文献 [8]、文献 [9] 都把相应颜色信息项跟纹理信息项的比重设为固定的经验值 0.5, 从这些简单的实验结果可以看出, 由于固定比例无法很好地刻画每幅图像内在的属性, 因而相应地普适性也较差, 而动态地融合颜色信息项和纹理信息项的方法则由于具有自动调节模型适应图像内在属性的能力, 因此具有较好的普适性, 并能够更好地把握图像的相关特征。

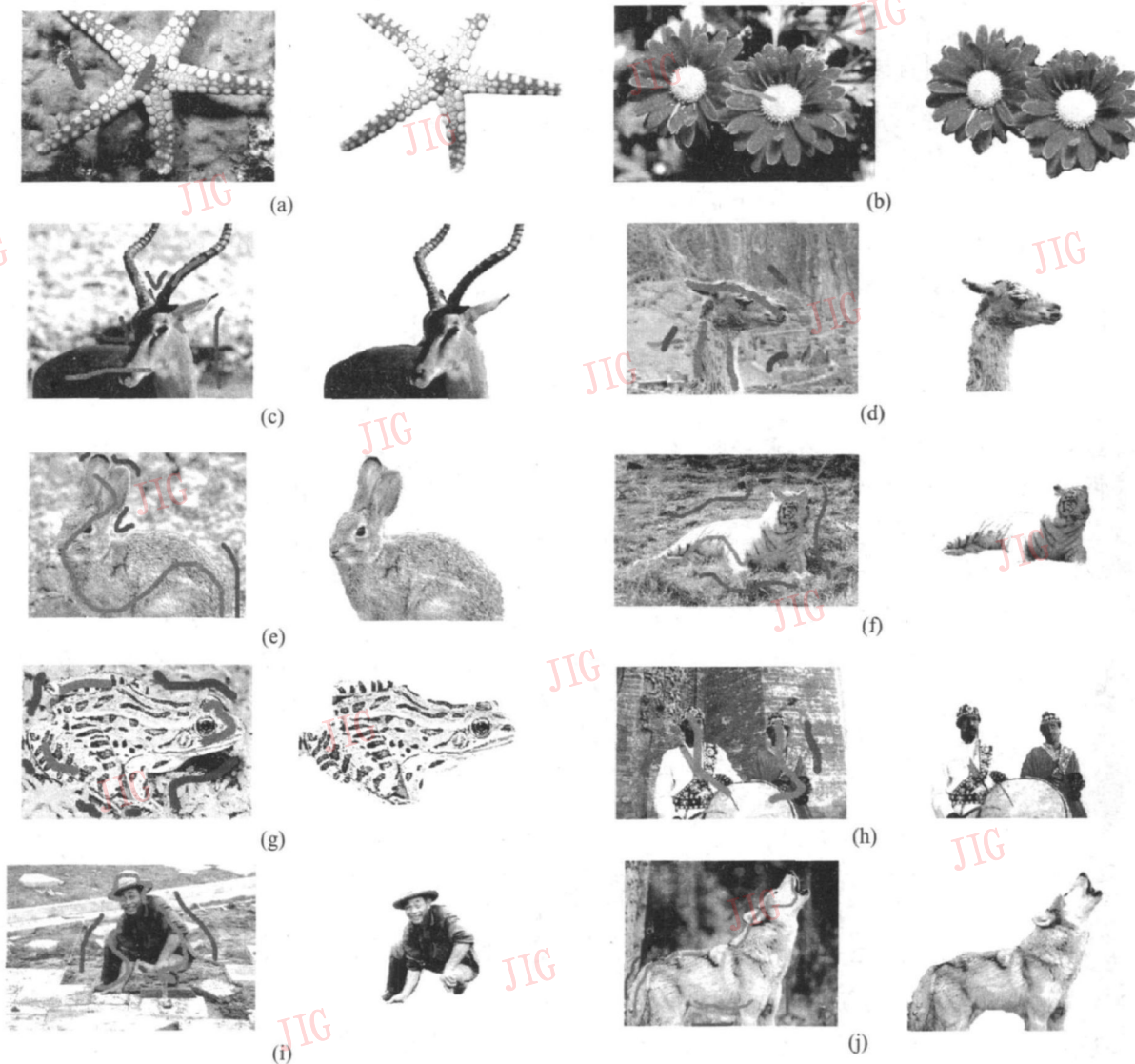


图 2 相关的分割结果

Fig 2 Related segmentation results

表 1 μ 值统计表Tab 1 Chart of μ statistical characteristics

图号	2-1-1	2-1-2	2-1-3	2-2-1	2-2-2	2-2-3	2-3-1	2-3-2	2-3-3	2-3-4
μ	0.513	0.639	0.597	0.446	0.341	0.479	0.485	0.546	0.523	0.552
$\bar{\mu}$	0.583	0.583	0.583	0.424	0.424	0.424	-	-	-	-

到目前为止, GrabCut算法是用于图像分割比较好的算法之一,它是一种基于迭代的算法,其对一般的自然图像的分割效果较好,而且它本身也带有后处理功能,便于对分割结果进行改进,但是该算法是不稳定的。如果是基于像素级的分割,那么对不同的图像,需要的迭代次数会有所不同,有的图像前景背景差异大,迭代结果比较稳定,执行多次迭代后,就会收敛到一个较好的分割结果,然而如果图像前景和背景颜色差异不大,甚至相似,那么就有可能

在多次迭代后得不到好的分割结果,而且还可能会随着迭代次数的增多,分割质量还会下降。后来文献[7]采用首先对图像使用分水岭变换(watershed transform),然后在变换后的过分割图像上,以每一个小区域为一个单位进行GrabCut,虽然这样得到的分割结果相对较为稳定,但是因为其不是基于像素级别的分割,边界就会变得粗糙,所以难以得到很好的分割结果。本文的分割算法虽然不是基于迭代的,但是由于模型中融合了图像的多种信息,一般初

次分割效果就不错, 因此能达到 GrabCut 算法迭代多次的分割效果 (图 3), 而且对于复杂的图像, 即使初次分割效果不佳, 但只需进行简单的后处理, 即进行局部修改就可以了。由图 3 的分割结果可以看

出, GrabCut 算法对于前景和背景颜色比较相近的图像, 分割能力有所下降, 由于本文分割算法融合了纹理等信息, 使模型能够进一步挖掘图像中的潜在信息, 所以取得了相对较好的分割效果。



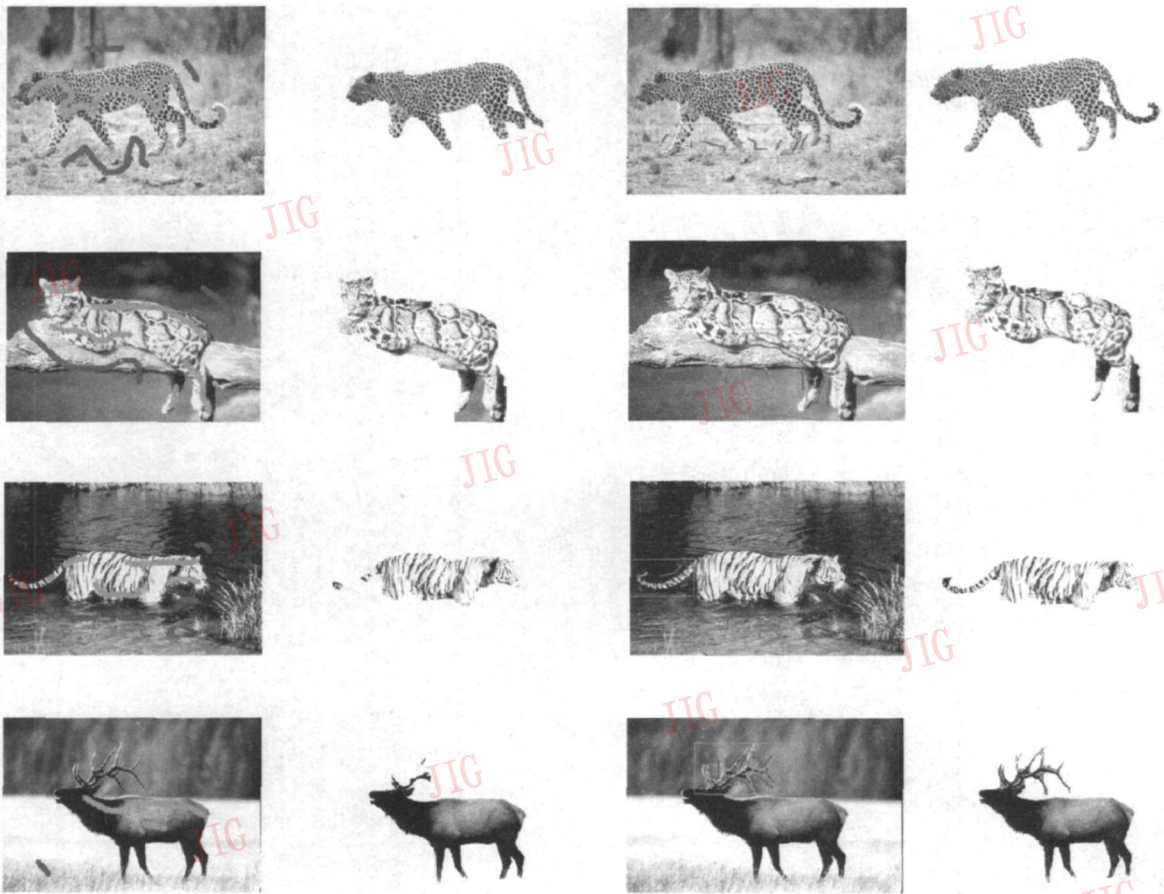
(a) GrabCut 算法的输入 (b) GrabCut 算法的分割结果 (c) 本文算法的输入 (d) 本文算法的分割结果

图 3 GrabCut 算法的迭代分割与多线索融合的分割

Fig 3 GrabCut segmentation and multicue integration segmentation

尽管如此, 对于一些较为复杂的图像, 尤其是边界比较细腻的图像, 本文算法的分割结果也并不是

很理想, 但可以选定不理想的区域做进一步的分割 (如图 4 所示), 最终可得到理想的分割效果。



(a) 初始的输入 (b) 初次分割的结果 (c) 后处理的输入 (d) 最终的分割结果

图 4 后处理的相关结果

Fig 4 Related results about postprocess

4 结 论

本文提出了一种利用多种信息的动态融合来进行图像分割的方法,该方法针对以前单一信息难以获得较好的分割结果的缺点,将纹理信息和颜色信息动态地融合到一个条件随机场中,并结合随机场对空间信息和边界信息的描述能力,对图像进行分割。对于颜色信息与纹理信息的融合,首先通过先验的信息确定纹理信息和颜色信息在分割算法中的决策比重,然后根据这种比重来组合条件随机场中的相关项。另外,由于条件随机场对边界刻画不很准确,为此本文利用琴森-香农散度来进一步提取图像中的边界信息,从而增强了条件随机场对边界的检测与分析能力。从实验结果中可以看出,一般情况下,对以颜色为主要属性的图像,本文算法的分割结果与 GrabCut 算法的分割结果相当;对于以纹理为主要属性的图像,当颜色对比度较小时,与 GrabCut 算法相比,本文的分割算法表现出更好的稳定性。另外,由于模型中的相关概率分布估计抛弃了以前比较流行的混合高斯模型估计的方法,而改用了核函数的估计方法,同样取得了比较好的分割效果。但是,也可以看到,该分割算法对边界比较细腻地方的分割效果还是不很理想,需要较多的交互操作方能得到较好的分割结果,这为以后进一步改进算法提供了方向。

参考文献 (References)

- [1] Rother C, Kolmogorov V, Blake A. Grabcut: interactive foreground extraction using iterated graph cuts [J]. *ACM Transactions on Graphics*, 2004, 23(3): 309-314.
- [2] Geman S, Geman D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, 6(6): 721-741.
- [3] Besag J. On the statistical analysis of dirty pictures [J]. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1986, 48(3): 259-302.
- [4] 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*, 2004, 26(9): 1124-1137.
- [5] Weiss Y, Freeman W T. On the optimality of solutions of the max-product belief propagation algorithm in arbitrary graphs [J]. *IEEE Transactions on Information Theory*, 2001, 47(2): 723-735.
- [6] Felzenszwalb P F, Huttenlocher D P. Efficient belief propagation for early vision [J]. *International Journal of Computer Vision*, 2006, 70(1): 41-54.
- [7] Li Y, Sun J, Tang C K, et al. Lazy snapping [J]. *ACM Transactions on Graphics*, 2004, 23(3): 303-308.
- [8] Jordi Freixenet, Xavier Muñoz, Joan Martí, et al. Colour texture segmentation by region-boundary cooperation [J]. *Lecture Notes in Computer Science*, 2004, 3022: 250-261.
- [9] Tsechpenakis G, Metaxas D N. CRF-driven implicit deformable model [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Minneapolis, MN, USA, 2007: 1-8.
- [10] Lafferty J, McCallum A, Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data [C] // *Proceedings of International Conference on Machine Learning*. San Francisco, CA, USA, 2001: 282-289.
- [11] Li S Z. Markov random field models in computer vision [C] // *Proceedings of European Conference on Computer Vision*. Stockholm, Sweden, 1994: 361-370.
- [12] Blake A, Rother C, Brown M, et al. Interactive image segmentation using an adaptive gmmrf model [C] // *Pajdla Tomas Matas Jiri*. *Proceedings of 8th European Conference on Computer Vision*. Heidelberg, Berlin, Germany: Springer, 2004: 428-441.
- [13] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 603-619.
- [14] Jain A K, Farokhnia F. Unsupervised texture segmentation using gabor filters [J]. *Pattern Recognition*, 1991, 24(12): 1167-1186.
- [15] Manjunath B S, Ma W Y. Texture features for browsing and retrieval of image data [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, 18(8): 837-842.
- [16] Huang X, Qian Z, Huang R, et al. Deformable model based textured object segmentation [C] // *Proceedings of 5th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*. Heidelberg, Berlin, Germany: Springer, 2005: 3757-119-135.
- [17] Juan Francisco Gómez-Lopera, José Martínez-Aroza, Aureliano M Robles-Pérez, et al. An analysis of edge detection by using the Jensen-Shannon divergence [J]. *Journal of Mathematical Imaging and Vision*, 2000, 13(1): 35-56.
- [18] Chakir A, Aït-Allah, Juan Francisco Gómez-Lopera, José Martínez-Torres, et al. Image segmentation by Jensen-Shannon divergence application to measurement of interfacial tension [C] // *Proceedings of IEEE 15th International Conference on Pattern Recognition*. Washington, DC, USA, 2000: 3383.
- [19] Kolmogorov V, Zabih R. What energy functions can be minimized via graph cuts? [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(2): 147-159.
- [20] 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.