

基于高斯超像素的快速 Graph Cuts 图像分割方法

韩守东¹ 赵勇¹ 陶文兵^{2,3} 桑农⁴

摘要 提出了一种交互式的快速图像分割方法. 该方法通过使用高斯超像素来构建 Graph cuts 模型以实现加速. 首先, 利用融合了边缘置信度的快速均值漂移算法, 将原始图像高效地预分割为多个具有准确边界的同质区域, 并将这些区域描述为超像素, 用于构建精简的加权图. 然后, 使用区域的彩色高斯统计对超像素进行特征描述, 并在信息论空间中对高斯距离度量进行设计. 另外, 为了准确而精炼地对先验知识进行参数化学习, 本文还使用了分量形式的期望最大化混合高斯 (Component-wise expectation-maximization for Gaussian mixtures, CEMGM) 算法对用户交互进行聚类. 最后, 在改进的加权图模型中应用 Graph cuts 方法, 获得最终的分割结果. 通过使用不同的彩色图像进行分割实验比较, 仿真结果表明本文的方法在准确性和高效性方面都具有很好的性能.

关键词 图像分割, 图切分, 超像素, 高斯模型, 均值漂移, 期望最大化算法

DOI 10.3724/SP.J.1004.2011.00011

Gaussian Super-pixel Based Fast Image Segmentation Using Graph Cuts

HAN Shou-Dong¹ ZHAO Yong¹ TAO Wen-Bing^{2,3} SANG Nong⁴

Abstract This paper proposes a fast interactive image segmentation method. To achieve acceleration, the method constructs the graph cuts model using Gaussian super-pixels. The fast mean shift algorithm embedded with edge confidence is first applied to efficiently pre-segment the original image into homogenous regions with precise boundary, and these regions are described as super-pixels to construct the compact weighted graph. The feature of super-pixel is then represented by the Gaussian statistics of color information in the corresponding region, and the dissimilarity measure of Gaussians is designed in the space of information theory. Additionally, in order to learn the parameters of priori knowledge accurately and compactly, the component-wise expectation-maximization for Gaussian mixtures (CEMGM) algorithm is used to cluster the user interactions in this paper. Finally, the graph cuts algorithm is applied to the improved weighted graph model to achieve the final segmentation. Through the comparison of different color image segmentation experiments, simulation results demonstrate the superior performance of the proposed method in terms of segmentation accuracy and computation efficiency.

Key words Image segmentation, graph cuts, super-pixel, Gaussian model, mean shift, expectation-maximization algorithm

图像分割是指按照一定的相似性准则将图像划分成具有特殊语义的不同区域, 从而在复杂的背景环境中把感兴趣的目标分离出来. 其应用领域非常广泛, 是图像分析、模式识别和计算机视觉中的一个

基本问题, 同时也是一个关键问题. 根据是否提供先验知识, 现有的分割算法可以分为有监督和无监督两类. 由于自然图像类型的多样性和内容的复杂性, 再加上理想分割目标对于人类主观视觉感知的依赖性, 所以有监督分割 (特别是交互式分割) 相比全自动的无监督分割具有更强的实用性. 而由 Boykov 等提出的 Graph cuts 模型^[1], 是目前研究和应用最为广泛的交互式图像分割方法之一. 它具有很多优良的特性, 比如多特征/约束融合能力, 全局最优, 数值鲁棒性强, 执行效率高, 分割加权图的拓扑结构自由和 N-D 图像分割能力等.

传统的 Graph cuts 方法在分割低分辨率简单图像时具有很好的实时交互性. 但是, 对于高分辨图像 (如大尺度的卫星遥感图像)、N-D 图像 (如三维医学图像) 或者图像序列 (如视频), 特别是当融合了多种视觉特征信息并使用迭代求解模式时 (如文献 [2] 中的方法), 无论是内存开销还是计算复杂度, 都严重限制了这些方法的可用性. 因此, 研究如何在短时间内获得高质量分割结果的快速 Graph cuts

收稿日期 2010-04-16 录用日期 2010-09-10
Manuscript received April 16, 2010; accepted September 10, 2010

国家高技术研究发展计划 (863 计划) (2009AA042107, 2007AA01Z166), 国家自然科学基金 (61073093), 国防自主创新基金资助
Supported by National High Technology Research and Development Program of China (863 Program) (2009AA042107, 2007AA01Z166), National Natural Science Foundation of China (61073093), and National Defense Independent Innovation Foundation of China

1. 华中科技大学系统工程研究所 武汉 430074 2. 华中科技大学计算机科学与技术学院 武汉 430074 3. 南京大学计算机软件新技术国家重点实验室 南京 210093 4. 华中科技大学图像识别与人工智能研究所 武汉 430074

1. Institute of Systems Engineering, Huazhong University of Science and Technology, Wuhan 430074 2. School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074 3. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093 4. Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074

方法成为了本文的研究内容。

为了能够对 Graph cuts 模型进行简化和加速,常用的思路主要有三个: 1) 提高最大流/最小割算法的运行效率; 2) 使用多层窄带技术压缩加权图的拓扑结构; 3) 基于预分割结果构建加权图模型. 比如, 文献 [3] 对应于第一种思路, 它针对巨型的规则图结构提出了一种可伸缩的图切分算法, 该算法在预流推进思想的基础上, 引入了内存分配控制和并行计算机制. 但是, 这种策略的加速效果和实用性非常有限, 因为它需要大量的并行处理器, 而且算法对物理内存空间的大小非常敏感, 最重要的是它只能应用于规则图结构. 另外, 文献 [4] 对应于第二种思路, 首先运用金字塔技术缩减图像的分辨率, 进而在压缩的加权图中使用 Graph cuts 模型进行分割, 最后通过多层窄带机制和某种还原技术将分割结果逐层传递到原始分辨率. 但是, 当使用这种策略分割具有细小结构的前景目标时, 由于降采样后局部信息的大量丢失和 Graph cuts 方法本身对细小目标分割的局限性, 这种加速策略的应用受到了限制.

与前两种策略相比, 第三种思路在加速效果和应用范围方面都具有良好的特性. 这种策略首先使用某种过分割算法将原始图像预分割为许多小的区域, 然后使用这些区域作为加权图中的节点来构建 Graph cuts 模型, 进而完成图像的分割. 相对传统 Graph cuts 方法把像素视为节点的做法, 这种策略极大地精简了加权图的拓扑结构, 自然也就缩减了相应的计算负担. Lazy snapping^[5] 就是基于这种思路的一个典型应用. 但是, Lazy snapping 在设计上存在很多问题: 其预分割算法使用的是分水岭变换, 这种算法基于灰度图像的梯度信息而忽略了彩色图像所包含的丰富信息, 造成预分割结果的区域一致性不够强, 而且分水岭算法的过分割现象非常严重, 势必会导致图像突出特征的丢失, 严重影响后续 Graph cuts 的分割性能和计算效率; 另外, Lazy snapping 仅仅使用区域的彩色均值来描述对应加权图节点的特征信息, 其描述方式过于简单, 描述力不够, 没有充分考虑区域中各采样点之间的协变关系, 所以在对种子标记进行先验聚类时需要使用大量的簇 (Lazy snapping 中为前景和背景设定的簇个数都为 64) 来提高模型的区分能力; 而且, Lazy snapping 中所采用的聚类算法是最简单的 K-means 方法, 这种算法通过最小化均值的方差函数来实现聚类, 而非最大似然估计, 而且这种算法对初始聚类中心严重敏感, 聚类效果不够理想. 虽然有较多的论文对 Lazy snapping 方法进行了改进, 但都主要集中于对其能量模型的设计和视频分割的扩展, 而将其作为加速策略进行改进和实验比较的文献却很少. 为了解决以上的问题, 本文

提出了一种基于高斯超像素的快速交互式图像分割方法, 主要思路是: 使用融合了边缘置信度的快速均值漂移 (Mean shift, MS) 模型代替分水岭变换, 因为这种 MS 算法可以对彩色特征空间进行有效而快速的分析, 实现图像的不连续保持性滤波, 而且其预分割结果中实体的一致性更强, 区域个数更少; 将区域看作超像素来构建加权图, 使用区域的彩色高斯统计对其进行特征描述, 并在信息论空间中对高斯距离度量进行定义, 用以对 Graph cuts 模型进行一些改进设计; 使用分量形式的期望最大化混合高斯 (Component-wise expectation-maximization for Gaussian mixtures, CEMGM) 聚类算法代替 K-means 聚类, 因为 CEMGM 算法对初始聚类中心的鲁棒性更强, 并提供最大似然估计, 而且模型的分量可以自适应地精简, 这样 CEMGM 就可以在保证高聚类精度的同时减小后续计算的复杂度.

1 Graph cuts 算法分析

图像分割问题可以描述为对图像中各像素进行典型二元 (前景和背景) 标号的组合优化过程. 而 Graph cuts 模型的核心思想在于构造一个能量泛函, 然后通过加权图的映射和网络流理论的运用, 将标号问题的全局最优化解转化为对应加权图的最大流/最小割问题. 为了有机地融合图像的区域信息和边缘信息, 我们定义如下形式的马尔可夫随机场能量泛函:

$$E(\alpha) = \sum_{i \in I} D_i(\alpha) + \gamma \sum_{\substack{(m,n) \in C \\ \alpha_m \neq \alpha_n}} S_{m,n}(\alpha) \quad (1)$$

其中, α 表示分割标号值, 0 代表背景, 而 1 代表前景; I 表示图像中所有的像素点; C 为图像中所有邻接像素对所组成的集合. 上式右边的第一项为数据项, 可以描述图像的区域信息, 而 $D_i(\alpha)$ 用于判断当前像素点 i 相对标号 α_i (前景或背景) 的特征相似度; 第二项为平滑项, 可以描述图像的边缘信息, 而 $S_{m,n}(\alpha)$ 用于估计邻接像素点 m 和 n 分属不同标号集时的惩罚值; 而 γ 为非负实数, 是用来控制数据项和平滑项各自权重的平衡系数, 此参数越大则分割结果的区域整体一致性越好, 比较适合形状单一而相对集中的分割目标, 而此参数越小则分割结果的局部细节可分性越强, 比较适合形状复杂或相对分散的分割目标.

然后, 需要将定义好的能量泛函 (1) 映射成为对应的加权图. 如图 1 所示, 我们给出了一个使用 Graph cuts 对 3×3 图像进行分割的简单实例. 原始图像 $I(x, y)$ 被映射为一个具有双极点的加权图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, 它包括一系列的节点 \mathcal{V} 和连接它们的权值边 \mathcal{E} . 其中, 节点对应于图像中的像素点和称作

极点的附加点. 这些极点与标号值集 α 相对应, 比如源极点 S 表示前景标号, 而汇极点 T 表示背景标号. 另外, \mathcal{E} 中有两种类型的权值边, 一种是 n -link, 另一种是 t -link. 其中, n -link 连接相邻的像素点, 其权值描述了相邻像素点的不连续性, 对应于能量泛函 (1) 中的平滑项, 如图 1 (b) 中各椭圆节点间互连的边所示; 而 t -link 连接像素点与极点, 其权值描述了对像素点进行对应标号的合理性, 对应于能量泛函 (1) 中的数据项, 如图 1 (b) 中 S 或 T 节点与各椭圆节点相连的边所示. 为了能够交互式地提供关于前景和背景的先验知识, 如图 1 (a) 所示, 我们需要标记一些前景和背景的种子点 (f_1 和 b_1), 进而影响 t -link 的权值计算. 如图 1 (b) 所示, t -link 的粗细反映了交互后的权值分配. 通过以上的方式, 能量泛函 (1) 就被映射到了一个 S - T 结构的加权图中. 于是, 如图 1 (b) 所示, 我们可以使用网络流的最大流/最小割理论, 通过对加权图 \mathcal{G} 进行全局 S - T 最小割运算实现能量泛函的最小化, 进而得到如图 1 (a) 所示的最优图像分割.

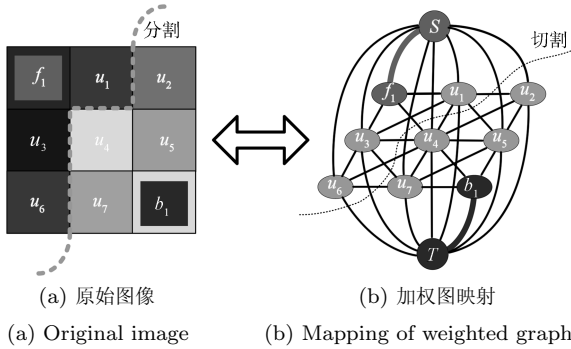


图 1 一个对 3×3 图像进行 Graph cuts 分割的简单实例
Fig. 1 A simple instance of segmenting the 3×3 image using graph cuts

2 快速 Graph cuts 分割模型

在本节中, 将介绍基于高斯超像素的快速 Graph cuts 图像分割方案. 本文使用 Lazy snapping 的直观交互方式, 采用 Graph cuts 框架作为分割模型. 为了最大限度地提高分割的高效性、准确性和鲁棒性, 本文针对 Graph cuts 模型的各个环节进行了改进设计. 下面将具体地讨论这些设计方案的改进思路.

2.1 融合边缘置信度的快速均值漂移算法

在基于预分割的 Graph cuts 分割模型中, 初始区域划分的准确性对后续分割的性能起着至关重要的作用. 均值漂移算法是一种用于特征空间分析的有效工具, 在计算机视觉的很多领域中均有广泛的应用, 表现出了良好的鲁棒性和较高的处理效率^[6].

为了使得预分割结果不仅在色彩上相似, 而且在空间上也保持连续, 本文采用了文献 [6] 中所描述的方法, 首先在五维特征空间上进行均值漂移滤波, 即三维的 $L^*u^*v^*$ 彩色空间和两维的空间坐标. 假定 $f(x)$ 为这个五维特征空间中的概率密度函数, 而 x_i 为实际的特征采样点. 于是, 我们可以将 $f(x)$ 的梯度估计为

$$\widehat{\nabla f(x)} \sim (\text{avg}_{x_i \in W_{h,x}} [x_i] - x) \quad (2)$$

这里, “ \sim ” 表示正比于; “avg” 表示求平均; $W_{h,x}$ 表示以 x 为中心而 $h = \{h_s, h_r\}$ 为带宽的五维超球体, 其中, h_s 和 h_r 分别表示空间和彩色的带宽. 传统的均值漂移算法需要对图像中每个像素点使用梯度上升法, 通过在特征空间中反复迭代, 自适应地寻找概率密度的极大值. 但是, 这种计算方式耗时较多. 所以, 为了提高滤波速度, 我们对那些之前处于其他像素均值漂移轨迹上的像素点不执行均值漂移过程, 而这些像素点将直接归于对应的模式吸引域. 通过这种方式的加速处理, 不会对滤波结果产生很明显的影 响. 滤波之后, 再按照区域邻接图算法、彩色带宽和区域大小准则等对模式吸引域进行递归合并. 为了控制分割的质量, 我们需要手动设定 h_s , h_r 和 M 的值. 其中, h_s 代表均值统计的空间窗口, 此参数越大则分割的整体性越好、预分割区域数越少, 但同时微弱边缘的识别也越模糊, 而此参数越小则分割结果的过分割现象越严重; h_r 代表模式吸引域的合并阈值, 此参数对分割结果的影响与 h_s 类似; M 代表分割结果中允许的最小区域大小, 此参数太大则会丢失大量的细节信息, 而太小则会引入大量的噪点区域. 本文通过实验取 $h_s = 7$, $h_r = 6$, $M = 50$.

另外, 为了得到更加准确的预分割结果, 本文还参考了文献 [7] 融合边缘信息的方式, 在均值漂移算法中融入了边缘置信度的概念. 本文使用了文献 [8] 提出的广义梯度边缘检测算法, 因为它对于微弱而尖锐的边缘信息具有很好的描述能力. 于是, 在均值漂移算法的滤波过程中, 对应式 (2) 的 “avg” 操作符本文采用了加权平均运算, 而每个像素 i 的权重则设定为 $\omega_i = 1 - \alpha_i \rho_i - (1 - \alpha_i) \eta_i$, 其中, ρ_i 和 η_i 分别为图像的梯度幅值和局部模式, 而 α_i 则为混合比例系数. 由于越靠近边缘像素的权重越小, 所以这样就可以进一步加强图像的不连续保持性滤波, 不仅可以滤除图像中的噪声影响, 而且能够对局部结构中的显著边缘信息自适应地减弱平滑. 而当进行吸引域合并时, 本文还附加了一个边缘阈值 t_e 的约束条件. 这样, 我们就需要手动设定 α , t_e 和 n 的值. 其中, α 对分割结果的影响并不大, 但一般情况下此参数越大则整体边缘提取越准确, 而越小则局部微

弱边缘识别能力越强; t_e 对分割结果的影响与 h_r 类似; n 代表梯度检测带宽, 此参数越大则局部细节边缘越丰富、预分割区域数越多, 而此参数太小则不能准确地识别局部微弱边缘信息. 本文通过实验取 $\alpha = 0.3$, $t_e = 0.9$, $n = 5$.

2.2 基于高斯超像素的 Graph cuts 模型

通过第 2.1 节所描述的均值漂移算法, 我们可以将原始图像预分割为多个区域一致性和边缘描述性较好的区域, 而这些区域在本文中被称为超像素. 基于这些超像素构建 Graph cuts 模型, 相对传统基于原始像素的构建模式, 可以大大精简加权图拓扑结构中的节点和权值边的数目, 进而可以节省大量的内存开销和计算负担.

当提取超像素的特征信息时, 传统的计算方式都是简单地使用区域中各采样点的特征均值来描述. 但是, 这种方式没有充分地考虑区域中各采样点在特征空间的相互协变关系, 所以描述不够准确, 为后续的计算带来了关于准确性、鲁棒性和可区分性方面的问题. 如图 2 所示, 区域 A 和区域 B 中都包含有大量的采样点, 分别以圆点和方点表示. 通过统计运算, 区域 A 和区域 B 具有相同的均值, 正如图 2 中心的圆球所示; 而它们的协方差却有显著的差异, 也就是说这两个区域的采样点具有完全不同的空间分布. 对于这种极端的情况, 传统基于区域特征均值的描述方式显然不能对区域 A 和区域 B 进行区分, 导致区域的区分能力不够. 当基于这些超像素进行 Graph cuts 分割时, 传统的均值特征描述将严重影响微弱边缘的分割. 而且, 为了提高均值特征超像素的区分能力, 势必要大量增加先验聚类中簇的个数, 这样做会显著加大分割模型的计算负担.

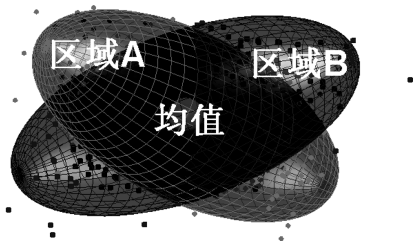


图 2 高斯超像素的特征示意图

Fig. 2 The diagram of features of Gaussian super-pixels

所以, 本文提出了使用高斯统计特征对超像素进行描述, 每个超像素 i 都可以表示为 $G_i = \{\mu_i, \Sigma_i\}$, 其中, μ_i 和 Σ_i 分别对应于区域 i 中彩色特征的均值和协方差. 其可视化描述如图 2 所示, 在本文中各预分割区域都可以描述为一个椭球, 而不是传统均值特征所描述的一个点. 由于椭球的大小和方向与区域的协方差密切相关, 所以高斯特征

可以更加准确地估计采样点在特征空间的概率分布. 所以, 即使是图 2 中所示的情况, 区域 A 和区域 B 一样可以区分开来. 下面我们将基于高斯超像素的特征描述, 在信息论空间中定义特征间的距离度量, 并给出 Graph cuts 能量模型 (1) 的详细设计.

基于高斯超像素的 Graph cuts 能量模型与泛函 (1) 一样, 只是其中的 I 表示图像中预分割得到的所有超像素; C 为预分割结果中所有邻接超像素对所组成的集合. 当计算 $D_i(\alpha)$ 时, 首先使用下一节中将要介绍的 CEMGM 算法对前景种子和背景种子进行先验聚类, 分别得到对应的高斯集 $\{G_1^F, \dots, G_N^F\}$ 和 $\{G_1^B, \dots, G_M^B\}$, 其中, N 和 M 分别为前景和背景的聚类簇数, 一般情况下这两个参数设置得越大, 则 CEMGM 算法对富含多种颜色分布的前景和背景描述越准确, 但同时也会加重 CEMGM 算法和后续 Graph cuts 的计算负担. 然后, 我们将超像素 i 的高斯特征 G_i 到前景高斯集的最小距离计算为 $d_i^F = \min_{n \in [1, N]} \text{dis}(G_i, G_n^F)$. 类似的, G_i 到背景高斯集的最小距离计算为 $d_i^B = \min_{m \in [1, M]} \text{dis}(G_i, G_m^B)$. 其中, 对于高斯特征间的距离度量 $\text{dis}(\cdot)$, 我们可以借助信息论中的 KL (Kullback-Leibler) 散度来定义. 比如, 对于高斯概率密度函数 $G_m(x)$ 和 $G_n(x)$, 其 KL 散度可以定义为 $\text{KL}(G_m \| G_n) = \int G_m(x) \log \frac{G_m(x)}{G_n(x)} dx$. 但是, KL 散度是非对称的, 所以可以使用 J 散度 $J(G_m, G_n) = \frac{1}{2} (\text{KL}(G_m \| G_n) + \text{KL}(G_n \| G_m))$ 对其进行对称化处理. 于是, 我们可以参考文献 [9] 中对结构张量距离度量的定义方式, 将 $\text{dis}(\cdot)$ 定义为 J 散度的平方根:

$$\text{dis}(G_m, G_n) = \sqrt{J(G_m, G_n)} \quad (3)$$

假定 $G_m(x)$ 和 $G_n(x)$ 分别具有高斯参数集 $\{\mu_m, \Sigma_m\}$ 和 $\{\mu_n, \Sigma_n\}$, 则根据文献 [10] 给出的计算方式, 式 (3) 具有如下的简单形式:

$$\text{dis}(G_m, G_n) = \frac{1}{2} (\text{tr}(\Sigma_m^{-1} \Sigma_n + \Sigma_n^{-1} \Sigma_m) + (\mu_m - \mu_n)^T (\Sigma_m^{-1} + \Sigma_n^{-1}) (\mu_m - \mu_n) - 2d)^{\frac{1}{2}} \quad (4)$$

这里, $\text{tr}(\cdot)$ 为矩阵的迹运算; 而 d 为高斯特征的空间维数, 这里对应于彩色空间的通道数. 于是, $D_i(\alpha)$ 可以按照非参数邻近模型定义为:

$$D_i(\alpha \equiv 1) = \begin{cases} 0, & \forall i \in F \\ \infty, & \forall i \in B \\ d_i^F (d_i^F + d_i^B)^{-1}, & \forall i \in U \end{cases} \quad (5)$$

$$D_i(\alpha \equiv 0) = \begin{cases} \infty, & \forall i \in F \\ 0, & \forall i \in B \\ d_i^B (d_i^F + d_i^B)^{-1}, & \forall i \in U \end{cases}$$

其中, F 和 B 分别代表预分割结果中被标记为前景和背景的超像素集, 而 U 代表其他未被标注的超像素集.

另外, 可以将 $S_{m,n}(\alpha)$ 改进设计为

$$S_{m,n}(\alpha) = \tau + \exp(-\xi \text{dis}^2(G_m, G_n)) \quad (6)$$

其中, τ 为滤噪常数, 但此参数设置得太大会导致分割结果出现区域过平滑, 而太小又无法避免强噪声信息影响分割区域的一致性; 参数 ξ 可以用如下的方式自适应计算为

$$\xi = \left(\frac{2}{|C|} \sum_{(m,n) \in C} \text{dis}^2(G_m, G_n) \right)^{-1}$$

其中, $|C|$ 表示集合 C 中元素的总数.

2.3 分量形式的期望最大化混合高斯聚类

为了对交互式先验知识进行混合模型的参数化学习, 期望最大化 (Expectation-maximization, EM) 算法通常是最好的选择, 因为它不但对初始参数的鲁棒性强, 而且还提供极大似然估计. 与 K-means 算法相比, EM 算法更加适合 Graph cuts 这种基于似然最大化的分割模型. 但是, 在某些情形下, 传统 EM 算法存在收敛慢的问题, 其原因主要是因为混合模型的所有参数都是同时更新的. 而且, 当分量数目很大的时候, 传统 EM 算法可能会出现分量无法全部初始化和混合系数无法确定等问题. 为了解决以上的问题, 本文参考了文献 [11] 的做法, 使用分量形式的期望最大化混合高斯 (Component-wise expectation-maximization for Gaussian mixtures, CEMGM) 算法来进行先验聚类. 通过串行地更新各个分量的统计参数, CEMGM 算法可以大大提高收敛速度. 而且, 一旦某个分量的混合系数为零, 它所对应的概率分布将会立刻分配给其他的分量, 以增加它们的生存机会. 这种特性允许我们使用任意的分量数进行聚类, 而且还可以对给定的分量进行结构精简. 另外, CEMGM 算法可以轻松地为自适应分量选择版本^[11], 这样将会进一步释放用户的手工参数设定. 如下所示, 我们给出了 CEMGM 算法的完整描述.

输入: 采样点 $\{I_s\}_{s=1,2,\dots,|S|}$, 簇分量数 k , 迭代终止条件 ε , 初始高斯混合参数 $\theta(0)$

输出: CEMGM 参数 $\hat{\theta}$

$$Z = d + \frac{d \times (d+1)}{2}, \quad d = 3;$$

$t = 0$, 初始迭代

$$\rho_s^{(k)}(t) = P(I_s | \theta_k(t)), \quad k = 1, \dots, K \text{ 和 } s = 1, \dots, |S|$$

$$L(t) = \sum_{s=1}^{|S|} \log \sum_{k=1}^K \omega_k(t) \rho_s^{(k)}(t);$$

repeat

$$t = t + 1, \quad k = 0;$$

Do

$$\varpi_s^{(k)}(t) = \omega_k(t-1) \rho_s^{(k)}(t-1) \left(\sum_{j=1}^K \omega_j(t-1) \rho_s^{(j)}(t-1) \right)^{-1},$$

$s = 1, \dots, |S|;$

$$\omega_k(t) = |S|^{-1} \max \left\{ 0, \left(\sum_{s=1}^{|S|} \varpi_s^{(k)}(t) \right) - \frac{Z}{2} \right\};$$

$$\omega_j(t) = \omega_j(t-1), \quad j = 1, \dots, k-1, k+1, \dots, K;$$

$$\{\omega_1(t), \dots, \omega_K(t)\} = \{\omega_1(t-1), \dots, \omega_K(t-1)\} \left(\sum_{j=1}^K \omega_j(t-1) \right)^{-1};$$

if $\omega_k(t) > 0$ then

M-step:

$$\mu_k = \left(\sum_{s=1}^{|S|} \varpi_s^{(k)} \right)^{-1} \sum_{s=1}^{|S|} \varpi_s^{(k)} I_s$$

$$\Sigma_k = \left(\sum_{s=1}^{|S|} \varpi_s^{(k)} \right)^{-1} \sum_{s=1}^{|S|} \varpi_s^{(k)} (I_s - \mu_k)(I_s - \mu_k)^T$$

$$\rho_s^{(k)}(t) = P(I_s | \theta_k(t)), \quad s = 1, \dots, |S|;$$

$$k = k + 1;$$

else

去除第 k 个分量 θ_k 并重新计算其他分量;

$$K = K - 1;$$

end if

while $k < K$

$$L(t) = \sum_{s=1}^{|S|} \log \sum_{k=1}^K \omega_k(t) \rho_s^{(k)}(t);$$

until $|L(t) - L(t-1)| < \varepsilon |L(t-1)|$

$$\hat{\theta} = \theta(t)$$

在以上算法中, Z 为每个分量所包含的参数总数, $P(\cdot)$ 为标准的多元高斯概率密度函数, 而 ω 为高斯混合模型中的分量混合系数. 阈值 ε 用于平衡 CEMGM 算法的收敛速度和聚类精度, 在我们的实验中设定为 10^{-5} . 当估计初始高斯混合参数 $\theta(0)$ 时, 我们将所有的混合系数 $\omega_k(0)$ 设定为 $1/K$, 均值 $\mu_k(0)$ 使用文献 [12] 中所描述的启发式初始化策略, 而所有分量的初始协方差都设为 $\Sigma_k(0) = \delta \Lambda$, 其中 Λ 为单位矩阵, 而 δ 设置为全局协方差中的最大对角元素乘以某个系数 (由于 CEMGM 算法对初始参数不太敏感, 这里我们根据经验将其设定为 0.08).

3 实验结果与分析

本节使用了一系列的自然彩色图像进行实验, 比较分析本文方法在各个方面相对 Lazy snapping 的性能优势. 所有实验图像来源于 Berkeley 自然图像库^[13] 和 GrabCut 分割图像库^[14]. 由于第 2.1 节中均值漂移算法的预分割结果不但准确而且区域数少, 再加上本文所提高斯超像素对区域特征的准确性描述, 所以此精简 Graph cuts 分割模型对参数并不敏感. 而就各参数的一般影响, 本文已在对应章节的介绍中进行了讨论. 下面我们直接给出分割模型中几个核心参数的经验参考值: 式 (1) 中的

权重 γ 设为 50; 式 (5) 中当计算 d_i^F 和 d_i^B 时, 前景和背景高斯集的分量数 N 和 M 都设定为 5; 式 (6) 中的滤噪常数 τ 设为 2.5. 本文中对基于边缘置信度的快速均值漂移算法的实现参考了文献 [7] 中的 EDISON 系统, 而最大流/最小割运算则使用了文献 [15] 给出的经典实现. 本文方法采用 C++ 进行系统实现, 并使用了 OpenCV 图像开发库. 实验平台为 Intel Core 2 Duo CPU T5750 2.00 GHz 的 CPU 和 2 GB 的内存.

首先, 通过对本文使用的均值漂移算法和 Lazy snapping 中的分水岭算法进行了预分割实验比较, 图 3 显示了对应的预分割结果, 而表 1 则显示了对应的定量化描述: 区域数和计算时间. 从实验结果来看, 分水岭算法的过分割现象非常严重, 即使如图 3(a) 所示简单图像, 也会产生上万的区域. 由于分水岭算法严重的过分割现象, 势必会增加后续 Graph cuts 建模的复杂度和计算效率. 相比之下, 本文中使用的融合了边缘置信度的快速均值漂移算法则能够更好地控制过分割现象, 区域数平均不到分水岭的 1%. 而从图 3 的预分割结果还可以看出, 本文的方法可以在保证充分滤波的情况下提取出图像中的重要边缘信息, 无论是目标还是背景, 语义结构都非常完整. 另外, 从表 1 我们还可以看出, 本文的方法虽然没有分水岭算法的运行效率高, 但是对于像图 3(c) 中 640 像素 \times 480 像素的分辨率来说, 运行时

间也能基本控制在半秒以内, 所以本文的方法绝对可以胜任实时应用中的分割任务.

为了验证本文中所采用的高斯超像素描述及其信息论距离度量的准确性和高分辨能力, 图 4 在图 3 中均值漂移预分割结果的基础上, 使用 n -link 图与 Lazy snapping 中的均值超像素和欧氏距离进行了可视化比较. 由于快速 Graph cuts 模型中的 n -link 主要由超像素特征间的距离度量来决定, 所以我们对每个超像素的所有邻域 n -link 进行求和以对其进行可视化处理. 从比较结果来看, 本文的描述方式相对 Lazy snapping 来说, 前景实体和背景之间的 n -link 值差异更大, 对微弱边缘的描述特性更好, 而且各语义区域内的 n -link 值更加平滑, 一致性更好. 所以本文使用的高斯超像素及其信息论距离度量具有更好的区域区分能力和语义一致性描述能力.

另外, 基于图 3 中均值漂移预分割结果, 图 5 进一步比较分析了本文提出的 CEMGM 算法和 Lazy snapping 的 K-means 算法在特征聚类中的差异. 初始的簇数量都设定为 5, 而采样空间为均值漂移滤波后得到的平滑图像. 由于本文中采用了文献 [12] 描述的启发式初始化策略, 所以 CEMGM 算法和 K-means 算法对初始聚类中心都具有很强的鲁棒性, 聚类结果不会有太大的随机波动. 所有实验结果都表明, 本文提出的 CEMGM 算法相对 Lazy snapping 的 K-means 算法在准确性和鲁棒性方面

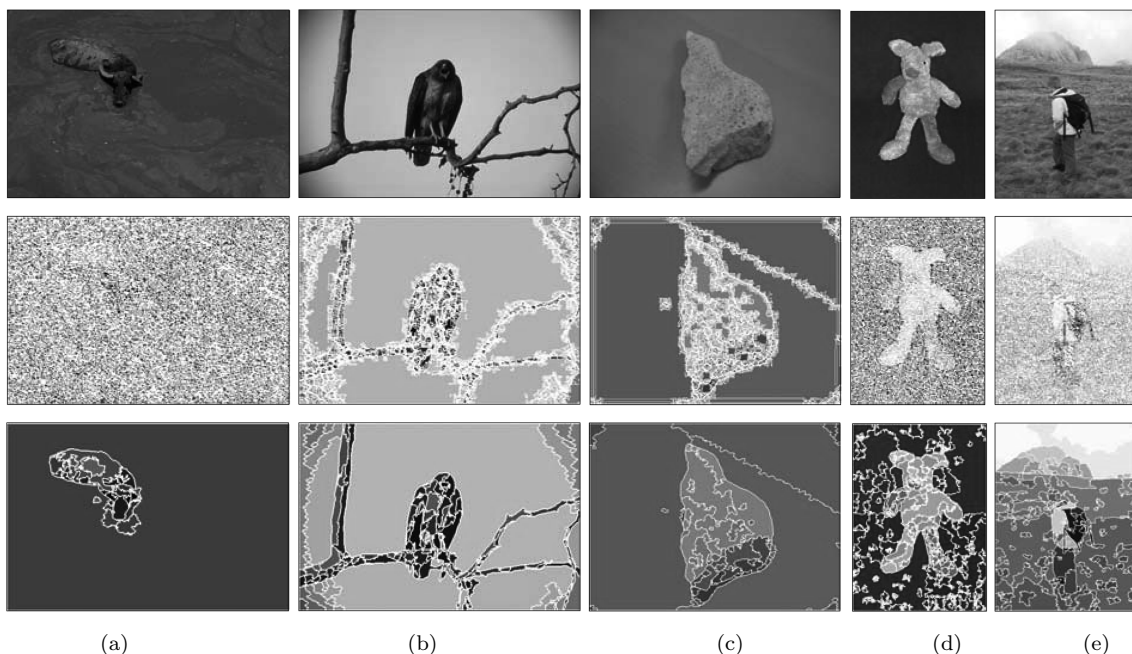


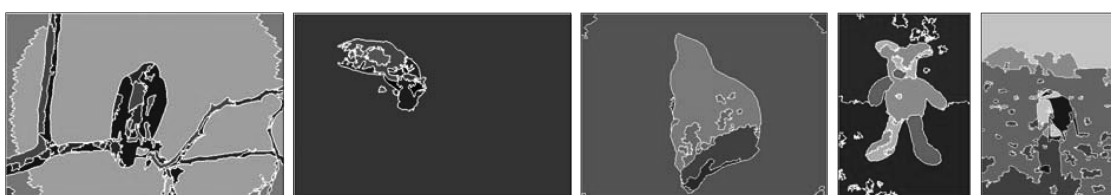
图 3 使用分水岭算法和均值漂移算法对彩色图像进行预分割的结果比较 (第一行为测试图像, 第二行和第三行分别为对应的分水岭算法和均值漂移算法的预分割结果.)

Fig. 3 The comparison of color image pre-segmentations between watershed algorithm and mean shift algorithm (The first row shows the test images; the second and third rows show the corresponding pre-segmentations of watershed algorithm and mean shift algorithm, respectively.)

表 1 图 3 实验中关于预分割区域数和计算时间的比较

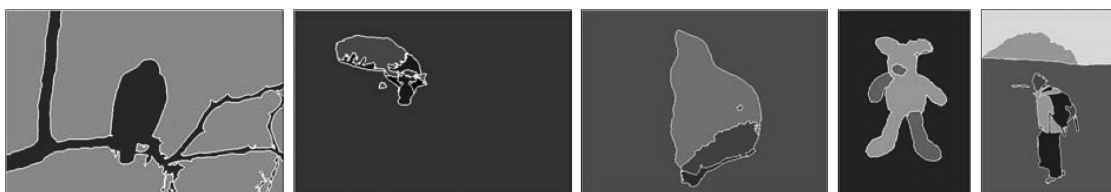
Table 1 The comparison of pre-segmentation regions' numbers and computation times for the experiments in Fig. 3

图像	分辨率 (像素)	预分割区域数 (个)		预分割耗时 (s)	
		分水岭	均值漂移	分水岭	均值漂移
图 3 (a)	481 × 321	18 513	32	0.031	0.219
图 3 (b)	481 × 321	6 132	158	0.032	0.234
图 3 (c)	640 × 480	8 340	62	0.031	0.375
图 3 (d)	284 × 398	15 588	179	0.032	0.422
图 3 (e)	450 × 600	25 671	209	0.047	0.375

(a) 基于均值超像素和欧氏距离的 n -link 可视化图(a) The visualizations of n -link images based on mean super-pixel and Euclidean distance(b) 基于高斯超像素和信息论距离的 n -link 可视化图(b) The visualizations of n -link images based on Gaussian super-pixel and information theoretic distance图 4 使用 n -link 图对不同超像素特征描述方式和距离度量的可视化比较. 所有实验全部基于均值漂移的预分割结果, 而且实验结果的布局与图 3 相对应Fig. 4 Visualization comparison of n -link images between different representations of super-pixels' features and dissimilarity measures (All the experiments are executed based on the pre-segmentation results of mean shift algorithm and the arrangement of experiments corresponds to that of Fig. 3.)

(a) 基于均值漂移滤波的 K-means 聚类结果

(a) Clustering results using K-means algorithm based on mean shift filtering



(b) 基于均值漂移滤波的 CEMGM 聚类结果

(b) Clustering results using CEMGM algorithm based on mean shift filtering

图 5 使用 K-means 和 CEMGM 对均值漂移滤波后的预分割结果进行聚类的实验比较. 实验结果的布局与图 3 相对应

Fig. 5 Comparing results between K-means and CEMGM for clustering the pre-segmentation results of mean shift algorithm (The arrangement of experiments corresponds to that of Fig. 3.)

都存在优势. 而且, 从图 5(b) 我们还可以看出, CEMGM 算法的确具有簇分量精简特性, 可以更加准确地还原图像中的模式类别, 比如通过 CEMGM 聚类后, 最终的聚类簇数从左到右分别为: 2、4、4、3、5.

更多的实验比较如图 6 所示. 实验比较了在使用相同的用户交互时, 本文方法与 Lazy snapping 在分割结果上的差异. 为了对分割质量进行定量的比较, 我们以人工分割的结果作为理想情况, 并参考文献 [16] 的做法使用 F 度量 (查准率, 查全率) 进行评估. 其中, 查准率代表当前分割结果中准确分割部分所占的比例, 查全率代表此准确分割部分在理想分割结果中所占的比例, 而 F 度量则为查准率和查全率的线性加权 (加权系数参考文献 [16], 设定为 0.5). F 度量所表示的综合指标值越大, 则分割

结果越符合人类视觉的主观目标判断准则. 从图 6 的实验结果可以看出, 无论是定性还是定量比较, 都进一步验证了本文方法的有效性. 在分割复杂自然图像时, 通过简单的用户交互, 本文方法相对 Lazy snapping 具有更好的分割准确性. 另外, 表 2 显示了对应于图 6 中关于预分割和后续快速 Graph cuts 在内存需求和计算时间上的比较. 通过实验比较我们发现, 本文方法相对 Lazy snapping 的内存开销, 在预分割阶段平均节省了 1.6 倍, 而快速 Graph cuts 阶段则更是节省了 19.5 倍; 当比较运算效率时, 虽然本文方法相对 Lazy snapping 在预分割阶段多消耗了将近 1.8 倍的计算时间, 但是在后续的快速 Graph cuts 阶段本文的方法却平均加快了 2.7 倍. 由于本文所使用的均值漂移算法可以很容易地进行并行实现, 所以预分割阶段还有进一步提速的空间.

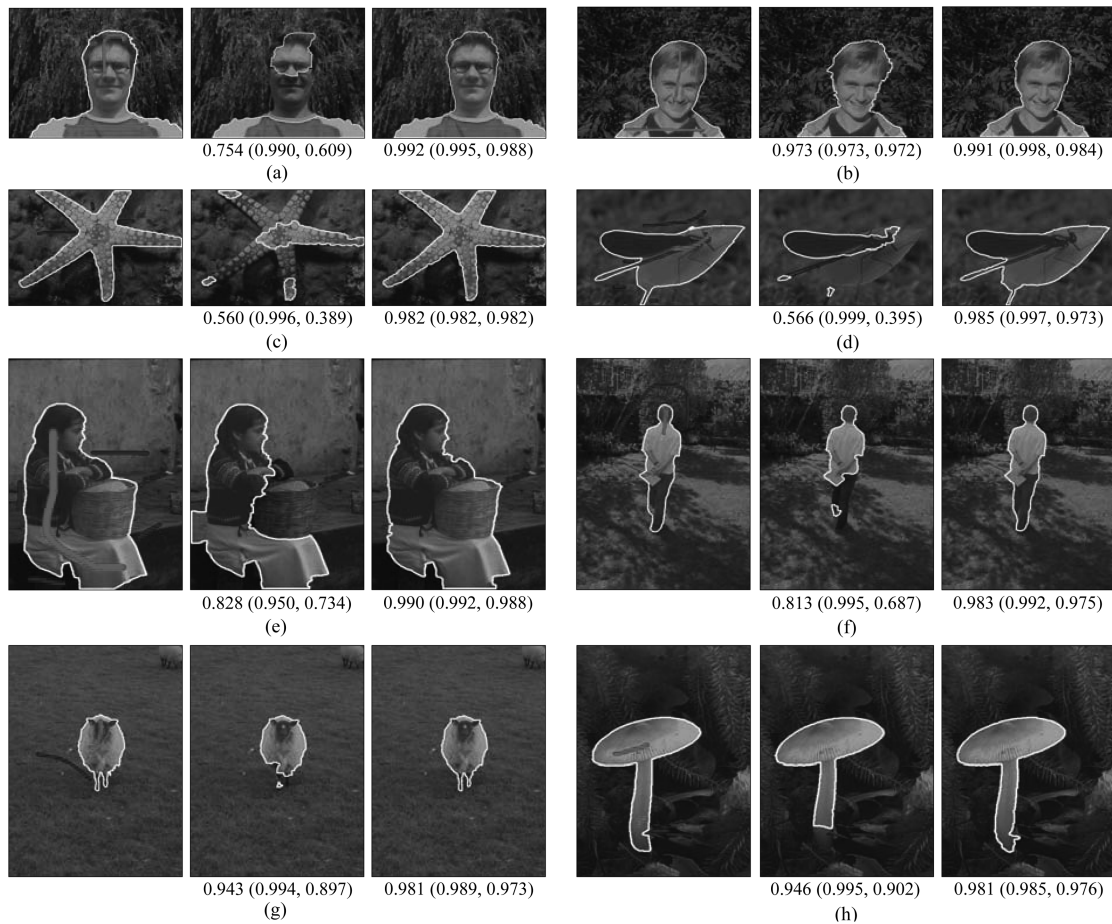


图 6 本文方法与 Lazy snapping 方法的分割结果比较 (每组实验从左到右分别为带用户标记的理想分割结果、Lazy snapping 分割结果和本文方法的分割结果. 分割质量的定量评估按照 F 度量 (查准率, 查全率) 的形式显示在对应分割结果下方.)

Fig. 6 Comparing segmentations between the proposed method and lazy snapping (The experiments are executed in groups, and every group shows the ground truth with user interactions, segmentation results of lazy snapping, and segmentation result of the proposed method, respectively, from left to right. The quantitative evaluations of accuracy are displayed below the corresponding segmentation results in the form of F-measure (precision, recall).)

表 2 图 6 实验中关于内存消耗和计算时间的比较

Table 2 Comparison of memory usage and computation times for the experiments in Fig. 6

图像	分辨率	内存需求 (MB)				计算时间 (s)			
		预分割		快速 Graph cuts		预分割		快速 Graph cuts	
		Lazy snapping	本文方法	Lazy snapping	本文方法	Lazy snapping	本文方法	Lazy snapping	本文方法
图 6 (a)	800 × 600	60.261	36.676	10.997	1.215	1.891	3.906	0.359	0.109
图 6 (b)	600 × 450	32.957	18.321	5.968	0.633	1.078	2.000	0.219	0.047
图 6 (c)	481 × 321	16.137	10.043	4.020	0.199	0.500	1.000	0.063	0.031
图 6 (d)	481 × 321	15.840	10.824	3.176	0.082	0.422	0.859	0.063	0.031
图 6 (e)	321 × 481	19.321	11.832	3.390	0.230	0.422	0.891	0.094	0.078
图 6 (f)	450 × 600	34.609	19.208	6.356	0.167	1.157	1.750	0.140	0.047
图 6 (g)	450 × 600	36.097	24.500	7.012	0.715	1.109	1.656	0.109	0.046
图 6 (h)	321 × 481	20.215	14.472	4.649	0.286	0.453	0.781	0.047	0.016

而且, 前期的预分割处理只是一次性的, 而后续的快速 Graph cuts 才是影响交互式响应速度的关键. 因为, 为了获得满意的分割结果, 往往需要进行多次的交互, 而每次交互都需要完整的运行快速 Graph cuts 分割, 所以本文方法在实际的图像分割应用中具有更强的实用性.

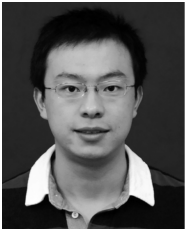
4 结论

本文基于高斯超像素和 Graph cuts 模型, 提出了一种相对 Lazy snapping 更加有效的交互式图像分割方法. 该方法使用融合了边缘置信度的快速均值漂移算法对原始图像进行高效而准确的预分割, 而且使用高斯超像素及其信息论距离度量来对预分割区域进行特征描述和建模. 为了准确而精炼地对先验知识进行参数化学习, 本文还使用了分量形式的期望最大化混合高斯算法对用户交互进行先验聚类. 通过使用大量的自然图像实例与 Lazy snapping 进行全面的比较, 实验结果表明本文方法在高效性、准确性和鲁棒性等方面具有更好的性能.

References

- 1 Boykov Y, Funka-Lea G. Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision*, 2006, **70**(2): 109–131
- 2 Han S D, Tao W B, Wang D S, Tai X C, Wu X L. Image segmentation based on grabcut framework integrating multiscale nonlinear structure tensor. *IEEE Transactions on Image Processing*, 2009, **18**(10): 2289–2302
- 3 Delong A, Boykov Y. A scalable graph-cut algorithm for N-D grids. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008. 1–8
- 4 Han S D, Tao W B, Wu X L, Tai X C, Wang T J. Fast image segmentation based on multilevel banded closed-form method. *Pattern Recognition Letters*, 2010, **31**(3): 216–225
- 5 Li Y, Sun J, Tang C K, Shum H Y. Lazy snapping. *ACM Transactions on Graphics*, 2004, **23**(3): 303–308
- 6 Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(5): 603–619
- 7 Christoudias C M, Georgescu B, Meer P. Synergism in low level vision. In: Proceedings of the 16th International Conference on Pattern Recognition. Washington D.C., USA: IEEE, 2002. 150–155
- 8 Meer P, Georgescu B. Edge detection with embedded confidence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(12): 1351–1365
- 9 Wang Z Z, Vemuri B C. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Transactions on Medical Imaging*, 2005, **24**(10): 1267–1277
- 10 Falk T H, Yuan H, Chan W Y. Single-ended quality measurement of noise suppressed speech based on Kullback-Leibler distances. *Journal of Multimedia*, 2007, **2**(5): 19–26
- 11 Figueiredo M A T, Jain A K. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(3): 381–396
- 12 Arthur D, Vassilvitskii S. K-means++: the advantages of careful seeding. In: Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms. Philadelphia, USA: SIAM, 2007. 1027–1035
- 13 Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the 8th International Conference on Computer Vision. Vancouver, Canada: IEEE, 2001. 416–423
- 14 Rother C, Kolmogorov V, Blake A. “GrabCut”: interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 2004, **23**(3): 309–314

- 15 Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(9): 1124–1137
- 16 Han S D, Tao W B, Wu X L. Texture segmentation using independent-scale component-wise Riemannian-covariance Gaussian mixture model in KL measure based multi-scale nonlinear structure tensor space. *Pattern Recognition*, 2011, **44**(3): 503–518



韩守东 华中科技大学系统工程研究所讲师, 博士研究生. 主要研究方向为图像分割、纹理分类、模式识别、计算机视觉、系统分析和随机模型预测控制. 本文通信作者.

E-mail: shoudonghan@gmail.com

(HAN Shou-Dong Lecturer and Ph. D. candidate at the Institute of Systems Engineering, Huazhong University of Science and Technology. His research interest covers image segmentation, texture classification, pattern recognition, computer vision, systems analysis, and stochastic model predictive control. Corresponding author of this paper.)



赵勇 华中科技大学系统工程研究所教授. 主要研究方向为决策分析、复杂系统建模与优化、图像处理.

E-mail: zhiwei98530@sohu.com

(ZHAO Yong Professor at the Institute of Systems Engineering, Huazhong University of Science and Technology. His research interest covers decision analyzing, complex system modeling and optimization, and

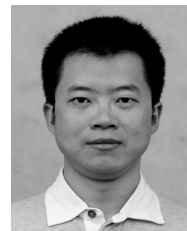
image processing.)



陶文兵 华中科技大学计算机科学与技术学院副教授. 2008 年至 2009 年在新加坡南洋理工大学数学系担任 Research Fellow 职位. 主要研究方向为计算机视觉、图像分割、目标识别及跟踪.

E-mail: wenbingtao@hust.edu.cn

(TAO Wen-Bing Associate professor at the School of Computer Science and Technology, Huazhong University of Science and Technology. He was a research fellow in the Division of Mathematical Sciences, Nanyang Technology University from 2008 to 2009. His research interest covers computer vision, image segmentation, and object recognition and tracking.)



桑农 华中科技大学图像识别与人工智能研究所教授. 主要研究方向为生物视觉感知模型及其在计算机视觉中的应用、基于统计学习的图像分析与目标识别、医学图像处理与分析、遥感影像解译、智能视频监控.

E-mail: nsang@hust.edu.cn

(SANG Nong Professor at the Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology. His research interest covers computational modeling of biological vision perception and applications in computer vision, image analysis and object recognition based on statistical learning, medical image processing and analysis, interpretation of remote sensing images, and intelligent video surveillance.)