

样本驱动的半自动图像集前背景分割

汪 粼波^{1,2)}, 郭延文^{1,2)}*, 夏天辰^{1,2)}, 金国平^{1,2)}

¹⁾(南京大学计算机软件新技术国家重点实验室 南京 210046)

²⁾(南京大学计算机科学与技术系 南京 210046)

(ywguo@nju.edu.cn)

摘 要: 图像集的前背景分割是近年来图像处理与图形学领域的一项热点研究工作. 针对图像集中的图像逐个进行交互分割会涉及大量的用户操作, 导致效率低下, 而联合分割方法通常局限于处理具有相似前景的图像集, 且因需求解大规模的优化问题较为耗时的问题, 提出一种样本驱动的半自动图像集分割方法. 首先选取若干图像作为样本进行手动交互分割, 训练基于样本图像超像素特征描述的支持向量机分类器; 对于其余待分割图像, 根据其超像素特征描述到支持向量机分隔超平面的距离计算基于双弯曲 Sigmoid 函数映射的前景置信度, 再采用图切割的算法实现目标图像的快速自动分割. 对于包含错误分割的个别图像, 进一步提出一种交互式局部修正方法修复错误分割区域, 并获得最终的精确分割结果. 在 2 个标准数据集上进行算法有效性验证和对比实验的结果表明, 与联合分割算法相比, 文中方法能更好、更快地实现在线分割; 与逐个交互分割算法相比, 文中方法能以相对较小的交互量实现对目标图像集的精确分割.

关键词: 图像分割; 联合分割; 超像素; 支持向量机; 双弯曲 Sigmoid 函数

中图法分类号: TP391

Example-Driven Semi-automatic Image Collection Segmentation

Wang Linbo^{1,2)}, Guo Yanwen^{1,2)}*, Xia Tianchen^{1,2)}, and Jin Guoping^{1,2)}

¹⁾(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210046)

²⁾(Department of Computer Science and Technology, Nanjing University, Nanjing 210046)

Abstract: Binary segmentation for image collection has received considerable attention in image processing and graphic communities recently. Interactively separating foregrounds from an image set one by one is time-consuming and requires tedious user guidance. Meanwhile, image co-segmentation techniques generally lack efficiency due to the complexity of solving large optimization problems and are only applicable to images sharing similar foreground appearance. In this paper, we propose an example-driven semi-automatic framework to tackle the image collection segmentation problem. First, we select few sample images from the given image collection and deliver them to user for hand-segmentation. Then, super-pixel features based support vector machine (SVM) classifier is trained. For each super-pixel of a given image, we estimate its foreground labeling confidence by applying Sigmoid function on the distance between its descriptor and SVM separation hyperplane. The confidence values are then encoded in a graph cut segmentation procedure to achieve automatic object cutout. For each image with incorrectly segmented regions, accurate result is further obtained by a new proposed local refinement process. Experiments on 2 standard datasets are presented, showing

修改稿收到日期: 2012-10-12. **基金项目:** 国家自然科学基金(61073098, 61021062); 国家“九七三”重点基础研究发展计划项目(2010CB327903); 江苏省自然科学基金(BK2009081). **汪 粼波**(1984—), 男, 博士研究生, 主要研究方向为计算机视觉、模式识别; **郭延文**(1980—), 男, 博士, 副教授, 论文通讯作者, CCF 会员, 主要研究方向为计算机图形学、图像视频等多媒体信息处理; **夏天辰**(1989—), 男, 硕士研究生, 主要研究方向为计算机视觉; **金国平**(1988—), 男, 硕士研究生, 主要研究方向为计算机视觉.

that the proposed algorithm not only greatly outperforms existing co-segmentation techniques, but also largely reduces users' efforts for cutting out object interactively.

Key words: image segmentation; co-segmentation; super-pixel; support vector machine; Sigmoid function

图像集的前背景分割是指将给定图像集中的所有图像前景从其对应的图像中分离出来. 近年来, 随着数码相机、摄像机, 特别是带有高分辨率摄像头的手机等数字图像获取设备的日益普及, 以及大容量存储设备和网络通信技术的不断发展, 数字图像资源正呈现爆炸性增长. 智能地分析和处理包含一定数量图像的图像集的研究需求也越来越迫切, 其中, 如何高效智能地对给定图像集进行前背景分割处理是一个重要的研究课题. 此外, 在理论研究方面, 图像的前背景分割也一直是图像理解、物体识别以及图像匹配与检索等方面工作的基础.

对包含一定数量的图像集进行前背景分割的一种思路是逐个对图像集中的图像进行交互式分割. 目前, 交互式分割算法已趋向成熟^[1-8], 其中有些算法更因操作方便、高效易用逐渐走向实际应用^[1,5,8], 因此这类算法是实现少量图像精确前背景分割的首选. 尽管如此, 当用户面对大量的待分割图像时仍须进行大量的交互工作, 这势必造成用户的操作疲劳, 进而降低用户操作的积极性和交互质量并影响分割效果. 从前背景分割任务所涉及的应用看, 待分割图像集所包含的图像往往呈现较强的相关性. 例如, 运动员参加比赛的一组照片或在动物园拍摄的一些小动物照片中, 所涉及的前景和背景虽不单一但具有较强的相关性, 如能在设计图像集分割算法时充分考虑并利用这种相关性, 可以在极大地减轻用户交互所需工作量的同时增强算法可用性并提高分割质量.

对图像集进行前背景分割的另一个思路是联合分割. 联合分割的概念由 Rother 等^[9] 首先提出, 其基本思想是给定一组前景相似但背景不同的图像, 以最大化图像之间相似性为原则建立能量方程并优化求解所有图像的前背景标号. 之后, 研究人员在此基础上做出了多方面的改进^[10-16]. 综合来说, 联合分割算法的缺陷在于因算法本身定位的限制, 使其不易拓展到对复杂图像集前背景分割的应用中. 由于联合分割算法着眼于针对具有相同前景和不同背景的图像集进行分割, 因此当给定图像集包含多个不同的前景时, 联合分割算法通常会失效. 另外, 目前的联合分割算法的目标方程多数是基于马尔科夫随

机场^[9-11,13], 其待优化方程中通常包含了所有待分割图像的像素. 当待分割图像数目增多时所需内存开销过大, 必将导致优化过程越来越慢, 当数目大于一定数量时甚至无法载入内存求解. 为了解决这个问题, 一些研究人员在运用联合分割算法前先采用过分割算法划分目标图像^[12,14,16], 再以得到的超像素代替原先像素点作为能量方程输入. 该做法虽在一定程度上缓解了内存和运行时间的压力, 但仍未从根本上解决问题, 且由于过分割算法本身的局限性, 其分割得到的区域在前背景边界处可能出现跨越边界的情形, 这种误差势必影响后期的前背景分割效果. iCoseg 算法^[16] 是一种基于用户交互的联合分割算法, 虽然该算法适合于分割包含一定数量图像的图像集, 但其所要求的图像集必须包含内容相似的前景和背景, 因此也无法应用于分割复杂图像集的前景和背景.

鉴于现有图像集前背景分割方法的局限性, 本文提出一种基于样本的半自动图像集前背景分割算法, 其目标是充分利用目前交互式图像分割研究的优势, 以尽可能少的交互需求完成对图像集的精确前背景分割. 本文算法首先从待分割图像集中挑选几幅具有代表性的待分割图像(样本), 并借助现有的交互式分割算法完成手动分割; 在此基础上, 充分挖掘样本图像与剩余待分割图像(目标图像)之间的关联性, 实现结合基于超像素块的支持向量机(support vector machine, SVM)分类器和基于图切割的前背景分割算法的目标图像集快速自动分割; 对包含错误分割的个别图像, 进一步提出一种适用于结果优化的交互式修正方法, 以获得最终的精确分割结果. 与基于交互的逐个单幅图像前背景分割方法相比, 本文算法将极大地减少用户的操作量, 减轻用户负担; 与联合分割算法相比, 该算法对图像集所包含图像的内容没有严格的限制. 此外, 本文算法的自动分割以及交互部分都是逐个图像进行的, 因此算法的时间和空间效率都相对较高.

1 本文算法概述

给定一个包含一定数量图像的图像集, 本文首

先将前背景分割问题视作一个有指导的二元分类问题,即对给定待分割图像中的任一像素点,判定其所属前背景类标号. 鉴于 SVM 特别适用于有指导的二元分类问题且通常可以取得良好的效果,本文选择其作为分类器. 此外,由于单个像素点的特征较弱,训练并使用基于单个像素点特征的 SVM 分类器判定目标图像中的像素标号既费时也不可靠,因此本文将样本图像划分成多个内部表征一致的超像素块,提取它们的特征描述并输入 SVM 进行训练,从而获得基于超像素特征描述的 SVM 分类器.

上述分类器可直接用于判定剩余各待分割图像的超像素块前背景标号,进而确定各图像的分割结果,但由于超像素块划分误差和特征提取的信息损失以及训练样本不足等因素,基于 SVM 分类器的超像素二元标号分割通常包含错误像素标号. 为解决该问题,本文首先基于变形的双弯曲 Sigmoid 函数将各超像素块到 SVM 分隔超平面的距离转化为各像素的前景标号置信度,然后定义融入所计算各像素的前景置信度概率和基于相邻像素点对比度的标号一致性约束的标号能量方程,最终通过最大流/最小割的图切割算法优化求解最佳的前背景标号.

对于包含错误分割区域的自动分割结果图像,本文将分别从前景和背景区域入手,结合用户交互构造局部的前背景颜色模型并实现基于图切割的局部修复分割,进而得到精确的分割结果.

2 基于样本超像素的 SVM 分类器训练

为利用样本图像的分割信息指导其他目标图像的分割,首先从待分割图像集中选取若干样本图像构造基于其超像素特征描述的 SVM 前背景分类器.

2.1 样本图像的选取与预分割

给定一个包含一定数量待分割图像的图像集,本文首先从中选取若干图像作为样本图像. 遴选的基本原则是在内容上涵盖图像库所有图像的特征信息,且不同的样本图像之间内容互补. 本文实验中初始的样本图像是自动筛选的,借鉴了文献[17]提出的基于多个图像特征学习选取代表性样本图像的基本思想,旨在减轻用户选择图像负担. 除自动选取之外,用户亦可手工指定样本.

本文实验采用了 Lazy Snapping 方法^[5]实现交互式前背景分割,首先把图像分成多个内部表征一致的超像素块^[18],以这些超像素块为结点构造无向图并基于图切割的方法进行自动前背景分割;然后

基于用户交互完成精确分割. 选择该方法是因为其一方面具有较高的时效性和较好的可操作性,另一方面在其中引入的超像素划分可进一步用于之后的基于 SVM 的超像素分类过程,实现样本分割与后期工作的无缝连接.

2.2 超像素特征描述的提取

本文为每个超像素块构建一个高维的特征描述向量,为提高描述向量的信息量,除了常规的颜色分布、尺度不变特征变换(scale-invariant feature transform, SIFT)分布等特征外,还特别融入了上下文信息. 具体来说,给定当前图像的一个超像素,其特征描述包括如下内容:

1) 基本特征. 提取当前超像素所含像素点的多种特征描述,逐个量化其分布,形成当前超像素的基本特征描述. 特征分布的量化方法采用词袋算法实现,其基本过程是先将所有样本像素的当前特征聚为类,再统计当前超像素的所有像素点属于各聚类的频数,进而形成维分布向量并进行 L_2 规范化. 所量化的特征包括 200 维的颜色分布向量,800 维的 Dense-SIFT(采用 Vifeat 实现^[19])特征分布向量及 300 维自相似特征^[20]分布向量. 为提高量化的效率,本文采用加速 k -means^[21]进行特征聚类.

2) 上下文特征. 将当前超像素同其周围的邻近超像素合并,形成拓展的超像素区域,提取该区域的基本特征描述过程同 1).

每个超像素的最终特征描述由其基本特征和上下文特征拼接得到,该特征描述不仅包含当前超像素的信息,也包含其周围区域像素信息. 此外,与空间金字塔一样^[22],该描述还包含一定的空间位置信息,因而更具鲁棒性.

2.3 SVM 分类器的构造

获得每个超像素块的特征描述后,本文将基于这些描述及其对应前背景标号训练 SVM 前背景分类器. 注意,由于图像内容的复杂性和超像素块划分算法的局限性,某些在图像前背景边界处的超像素块会出现跨越边界的情形,即这些超像素块既包含前景像素点也包含背景像素点,此时,本文设定这些超像素的标号为其所包含的多数像素点所对应的标号. 此外,基于文献[23]提出的 χ^2 核映射函数理论,本文将所有的超像素特征描述预先映射到高维空间中,再基于这些转化后的特征描述训练线性 SVM 分类器. 其优点是在保证分类器线性效率的基础上以线性核逼近 χ^2 核,从而提高分类器的性能. 本文

采用 Vlfcat^[19] 的相关实现完成超像素特征描述的转化。

假定当前所有样本的超像素共有 N 个,其映射后的描述为 $\{\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \dots, \mathbf{x}_{s_N}\}$, 对应的前背景标号为 $\{y_{s_1}, y_{s_2}, \dots, y_{s_N}\}$, 其中 $y_{s_i} = 1$ 对应于前景, $y_{s_i} = -1$ 对应于背景. 通过求解以下线性 SVM 优化问题确定其分隔超平面 (\mathbf{w}, b) ,

$$\begin{cases} \min_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \right\} \\ \text{s. t. } y_{s_i} (\mathbf{w}^T \mathbf{x}_{s_i} - b) \geq 1 - \xi_i \end{cases}$$

其中, ξ_i 为松弛变量, $\xi_i > 0, i = 1, \dots, N$; C 为惩罚因子(本文实验中设为 10). 本文采用快速线性分类器 LIBLINEAR^[24] 来求解分隔超平面 (\mathbf{w}, b) , 完成 SVM 分类器的建模。

3 目标图像的自动分割

给定一幅待分割目标图像, 本文首先对其进行超像素划分、特征描述提取及核函数特征映射; 得到各超像素的最终特征描述后, 原则上可以基于上述 SVM 分类器直接判定各超像素块的前背景标号, 进而实现目标图像的自动分割. 但是, 基于该方法得到的分割结果通常包含以下误差:

1) 超像素块划分所引起的分割误差. 鉴于图像内容的复杂性和算法本身的局限性, 超像素块的划分结果通常会出现一些跨越实际前背景边界的超像素块. 当采用基于 SVM 分类标号的分割方法对这些超像素进行直接统一标号时, 必将导致其中某些像素点的错误标号。

2) SVM 分类器的训练误差. 由于超像素块的特征提取存在一定的信息损失且参与 SVM 训练的样本有限, 如果对目标图像的超像素标号进行直接判定也可能出现误差。

为避免上述“硬”分割误差, 本文将各超像素基于 SVM 分类器的标号输出转化为其前景置信度概率, 再引入相邻像素点间基于对比度的标号一致性约束, 实现基于图切割算法的前背景“软”分割。

3.1 前景置信度的计算

给定一个超像素特征描述, 标准 SVM 分类器的输出是基于该描述到 SVM 分隔超平面距离判定的二元前背景标号. 通过分析 SVM 的分类特性, Platt^[25] 利用 Sigmoid 函数实现了 SVM 分类预测的概率化输出. 本文基于该理论计算各超像素的前景

置信度概率, 给定超像素描述 \mathbf{x} , 其前景置信度概率

$$P(\mathbf{x}) = \frac{1}{1 + \exp(\alpha\varphi(\mathbf{x}) + \beta)}$$

其中, $\varphi(\mathbf{x}) = \mathbf{w}^T \mathbf{x} - b$ 计算超像素特征描述到分隔超平面 (\mathbf{w}, b) 的距离; 参数 α, β 调节双弯曲 Sigmoid 函数的形状, 其值可基于样本超像素的标号分布采用最大似然法估算得到

$$\min_{\alpha, \beta} \left\{ - \sum_{i=1}^N (t_i \log(P_i) + (1 - t_i) \log(1 - P_i)) \right\},$$

$$P_i = P(\mathbf{x}_{s_i}), t_i = q_{s_i} \times \frac{N_+ + 1}{N_+ + 2} + (1 - q_{s_i}) \times \frac{1}{N_- + 2},$$

$$q_{s_i} = (y_{s_i} + 1) \times 0.5, N_+ / N_- \text{ 为样本前景/背景超像素总数, } \mathbf{x}_{s_i} \text{ 和 } y_{s_i} \text{ 为第 } i \text{ 个样本超像素特征描述和对应的前背景标号.}$$

3.2 基于图切割的自动分割

计算完当前目标图像各超像素的前景置信度后, 本文基于图切割方法^[3] 实现目标图像的自动分割. 首先定义基于目标图像像素点前背景标号的吉布斯能量方程, 并采用网络图中的最大流/最小割算法优化求解该能量的最小值, 从而确定各像素点的标号. 给定一幅待分割图像 I , 能量方程定义为

$$E(c_{p_1}, c_{p_2}, \dots, c_{p_n}) = \sum_{p_i \in v} E_1(c_{p_i}) + \lambda \sum_{(p_i, p_j) \in \epsilon} E_2(c_{p_i}, c_{p_j}), c_{p_i} \in \{0, 1\}.$$

其中, v 是目标图像的像素点集合, p_i 是其中任意一个像素点; ϵ 是像素点间 4 邻域的相邻关系集合; $c_{p_i} = 1$ 表示对应像素点 p_i 的标号是前景, $c_{p_i} = 0$ 表示对应像素点 p_i 的标号是背景; n 是图像的总像素点数; $E_1(c_{p_i})$ 是数据项, 表示对像素 p_i 进行前背景标号代价, 定义为

$$\begin{cases} E_1(c_{p_i} = 1) = -\log P(x_{p_i}) \\ E_1(c_{p_i} = 0) = -\log(1 - P(x_{p_i})) \end{cases},$$

x_{p_i} 是像素 p_i 所在超像素的特征描述, $P(x_{p_i})$ 为对应超像素的前景置信度; $E_2(c_{p_i}, c_{p_j})$ 是平滑项, 约束相邻像素点 p_i 与 p_j 标号的不一致性, 其定义为^[4]

$$E_2(c_{p_i}, c_{p_j}) = |c_{p_i} - c_{p_j}| \exp(-\gamma D_{ij}),$$

$$D_{ij} = \|I_{p_i} - I_{p_j}\|, \gamma = (2\langle D_{ij} \rangle)^{-1}, \langle \cdot \rangle \text{ 用于计算图像中相邻像素点的对比度期望; } \lambda \text{ 用来调节平滑项对标号过程的约束作用.}$$

本文将原图像映射到等价的带权无向图中, 并采用文献^[26] 的最大流算法优化求解能量方程的最小值, 确定各像素点的最佳标号, 进而实现目标图像的自动前背景分割。

4 基于交互的局部优化修复

上述基于 SVM 和图切割技术的前背景分割方法能够快速、有效地实现对图像集中部分图像的精确分割,但由于可能存在的处理误差,如超像素的划分不准确、特征提取的信息损失及 SVM 训练不足等,部分自动分割结果中仍包含错误分割区域.为解决该问题,本文采用一种基于 Grabcut^[4] 的交互式局部修正方法实现精确的前背景分割.由于修复的交互量通常较小,甚至只发生在前景或背景上,因此建立全局的前背景颜色模型相对困难.此外,对整幅待修复图像进行基于图切割的前背景分割通常会破坏现有的正确分割结果.为保证每次修复的局部性,首先在当前前景和背景上分别展开修复过程,所得结果经合并形成完整的结果,即当用户完成一次交互操作后,算法首先检查当前前景区域的用户标记,以前景点背景点分别建立前背景的颜色特征的高斯混合模型(Gaussian mixture model, GMM).基于 GMM 计算当前前景上各像素点的前背景概率,并采用 Grabcut 实现当前前景区域的修复分割;背景区域的修复过程与前景区域类似.该算法简单直接并具有以下优点:

1) 对前景/背景的交互修复只在前景/背景上展开,避免了对背景/前景区域的影响;

2) 当修复区域为当前前景/背景时,背景/前景的 GMM 建模只依赖用户标记像素,因此该背景/前景 GMM 对远离用户标记区域的像素影响较小,保

证了修复的局部性;

3) 虽然每次修复包含了 2 个子过程,但 GMM 建模和待优化问题的整体规模都变小了,因此算法的效率更优于基于全局的 Grabcut.

当一次交互修复不能取得满意的分割效果时,用户可进行再交互,直到获得精确的分割效果为止.

5 实验结果与分析

为了验证本文方法的有效性,我们选取多个标准图像集进行了分割测试,测试平台是 Intel Core i5 2.67 GHz CPU, 4 GB 内存的 PC 机.选用如下图像集:

1) 利兹蝴蝶数据集^[27].该数据集包含 832 幅蝴蝶图像,分为 10 类,每类包含 55~100 幅图像.本文从中随机选择了 3 类共 172 张图像进行测试,每类选取 1 幅图像作为样本,共 3 张样本.

2) VGG 的花朵数据集^[15].该数据集包含 8 189 幅花朵图像,分为 102 类,每类包含 40~250 幅图像.本文挑选了 4 类共 410 幅图像同时进行分割,每类同样选取 1 幅图像作为样本,共 4 张样本.

首先将算法的自动分割部分与文献[12,14]的联合分割算法进行比较,实验结果如图 1 所示,用线框标出的图像(第 1 个)为所选样本.可以看出,本文方法取得了较好的分割效果.虽然本文方法需要用户手动交互提前完成样本的分割,但少量的样本交互不会给用户造成太大的负担.特别是当图像集中图像的相关性较强时,例如所有的图像都包含同一

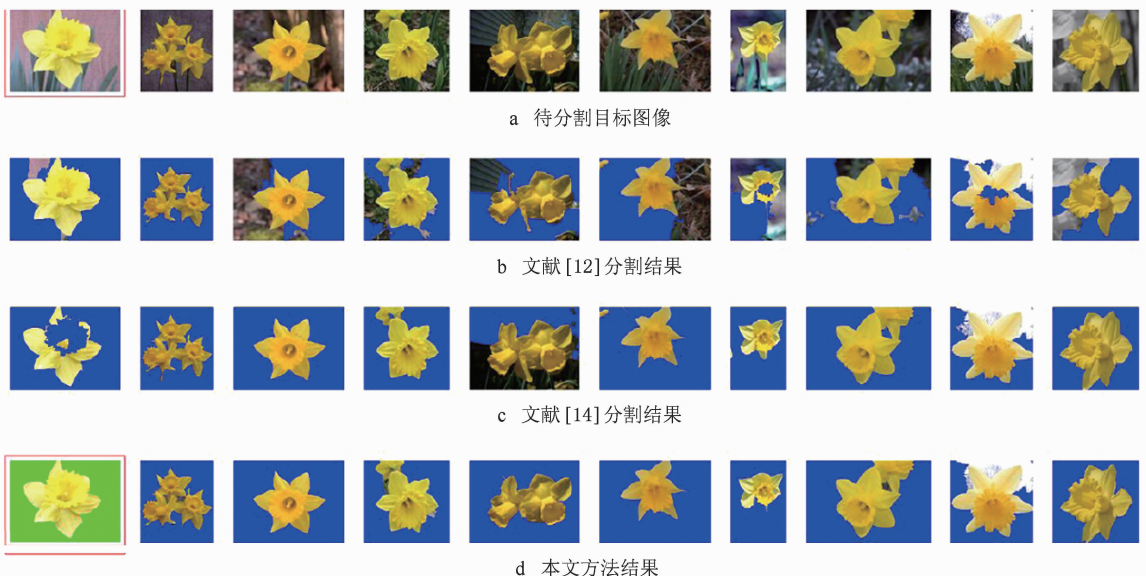


图 1 3 种方法的比较结果

对象时,只需给定若干甚至一张样本便可以取得较好的自动分割效果.此外,本文方法在上述 2 个标准数据集上的实验结果也进一步证实了其有效性,限于篇幅,文中并未给出全部分割效果,有兴趣的读者可参阅我们的网站^①.所有这些实验结果均表明,基于 SVM 分类方法很好地传递了样本图像的分割信息,而基于图切割的方法更优化了分割的结果.

为进一步验证本文方法各个环节的有效性,我们在蝴蝶图像集实验中额外训练了一个基于样本图像超像素基本特征描述的 SVM 分类器,利用该分类器对待分割图像进行基于其超像素基本特征描述的前背景标号;再以融入上下文特征的超像素特征描述的 SVM 分类器对目标图像集进行分割处理;

最后将这两部分结果与基于 SVM 分类的前景置信度和图切割算法的自动分割结果进行对比,其部分结果如图 2 所示.从图 2 b,2 c 的对比效果可以看出,融入上下文特征的超像素描述可以有效地提升 SVM 的分类精度,这种提升突出表现在当某个错误分类的超像素被正确分类的超像素“包围”时,上下文信息的引入将指导和实现其分类结果的自动修正.图 2 c,2 d 实现了 SVM 分类输出到图像前景置信度的转化,可以看出,转化后的概率图更清晰地反映了图像中各部分属于前景的置信度.最后,基于置信度图和图切割的方法实现了目标图像的自动分割.由于融入了基于像素间对比度的标号一致性约束,其分割结果边缘更加光滑,边界更加准确.



图 2 蝴蝶图像集的部分对比结果

本文方法的自动分割部分的运行时间分为训练时间和分割时间两部分,训练时间主要包括词袋模型中“字典”的生成和 SVM 训练时间.本文 SVM 训练采用了 LIBLINEAR^[24],其运行时间与待训练特征的个数成正比,由于采用超像素的特征描述作为 SVM 输入且平均每幅图像的超像素个数为 200 左右,训练时间极短,处于 0.01 s 级.本文采用加速 k -means^[21]生成“字典”,其优点是在不损失聚类精度的前提下处理效率约是传统 k -means 的 10 倍.实验

中,本文将 10 万个 128 维的 SIFT 特征聚为 800 类,其平均花费时间为 70 s;当“字典”生成后,使用 k -d tree 判定各像素点在“字典”中对应“单词”入口,加速计算各超像素的特征分布向量.实验中,给定一个新图像(分辨率为 500×375),基于给定“字典”完成超像素特征提取的平均时间为 5 s 左右. SVM 分类及基于图切割的分割算法的运行时间小于 1 s. 综

① <http://cs.nju.edu.cn/ywguo/ImageCollectionSeg.htm>

所述,本文方法的自动分割部分无论是训练或处理效率率都很高,训练的时间通常可控制在 10 min 以内.而对具体图像进行自动分割时,如果不预先进行离线超像素特征提取,其运行时间在 6 s 左右;反之,则可以控制在 1 s 以内,实现实时处理.

为验证本文方法对用户交互量的影响,表 1 所示为本次实验中花朵和蝴蝶图像集精确分割的用户交互量统计.可以看出,在完成基于 SVM 和图切割算法的初始自动分割后,大部分图像已不需要手动交互.其次,结合本文提出的局部优化交互修复算法,部分图像只需 1 次交互便可实现精确分割,通常这在单张图像交互式分割算法中是难以实现的.经典的交互式分割算法(如 Lazy Snapping^[5])至少需

要 2 次初始交互来实现前背景模型的建立,而基于下拉矩形框实现初始分割的 Grabcut^[4]也需要根据图像复杂度的不同进行数量不等的交互修正.最后,仍有少量图像需要进行 3 次甚至更多次的交互,排除 SVM 分类器不准确等因素,这部分图像通常由于内容比较复杂或前背景边缘比较模糊,致使图切割算法无法快速收敛进而需要多次交互实现精确分割,这也是基于图切割前背景分割算法中固有的缺陷.此外我们注意到,在进行交互分割操作过程中有些图像的出错区域是内容一致的,利用这种一致性实现相关区域的联合修复将进一步减少所需的交互量,减轻用户负担,这也将是我们未来工作的一个重要方向.

表 1 花朵与蝴蝶图像集精确交互量统计

图像集	样本平均交互次数	交互 N 次的图像数						图像总数	平均交互次数
		0	1	2	3	4	5		
花朵	3	228	70	59	32	13	8	410	0.94
蝴蝶	3	93	31	35	8	5	0	172	0.88

6 总结与展望

本文提出了一种基于交互的快速图像集分割算法.为充分利用样本图像与待分割图像之间的关联性,本文将图像分割问题视为一个二元分类标号问题,并基于样本图像建立基于 SVM 的超像素标号分类器;为提升自动分割的分割效果,采用双弯曲 Sigmoid 函数将 SVM 分类器输出转化为目标图像的前景置信度并基于图切割算法实现自动分割.对于不满意的分割结果,本文进一步提出一种基于交互的局部优化修复算法,实现了目标图像的精确分割.最后在多个标准图像集上进行了分割和对比实验,证明了本文算法的有效性.此外,在所消耗的时间和所需的用户交互方面,本文算法都具有明显的优势.

我们未来的工作将主要着眼于 SVM 分类器的更新及对分割结果的优化.在进行完一轮自动分割后,首先只对少量目标图像进行手动修复;在此基础上更新 SVM 分类器,并在不影响现有正确分割结果的前提下,实现对其他分割效果欠佳的图像的自动优化调整.在图像的手动修复方面,我们希望找到一种自动修复传播方法,即当用户完成一次手动修复后,自动传播该修复到其他具有类似错误分割区

域的图像上实现分割结果的自动优化,避免用户的重复交互操作,减轻用户负担.此外,我们将尽快完善原型系统,使之成为一个方便实用的图像集处理工具.

参考文献(References):

- [1] Mortensen E N, Barrett W A. Intelligent scissors for image composition [C] //Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH. New York: ACM Press, 1995: 191-198
- [2] Chuang Y Y, Curless B, Salesin D H, et al. A Bayesian approach to digital matting [C] //Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2001, 2: 264-271
- [3] Boykov Y, Jolly M. Interactive graph cuts for optimal boundary and region segmentation of objects in N -D images [C] //Proceedings of the 8th IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2001, 1: 105-112
- [4] Rother C, Kolmogorov V, Blake A. "GrabCut": interactive foreground extraction using iterated graph cuts [C] //Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH. New York: ACM Press, 2004: 309-314
- [5] Li Y, Sun J, Tang C K, et al. Lazy snapping [C] //Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH. New York: ACM Press, 2004: 303-308

- [6] Zheng Jiaming, Chen Zhaojiong. Connectivity constrained graph-cut for fast interactive image segmentation [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2011, 23(3): 399-405 (in Chinese)
(郑加明, 陈昭炯. 带连通性约束的快速交互式 Graph-Cut 算法 [J]. *计算机辅助设计与图形学学报*, 2011, 23(3): 399-405)
- [7] Grady L. Random walks for image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(11): 1768-1783
- [8] Adobe Photoshop 7.0 user guide [M]. San Jose: Adobe Systems Incorporated, 2002
- [9] Rother C, Minka T, Blake A, *et al.* Cosegmentation of image pairs by histogram matching-incorporating a global constraint into MRFs [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2006, 1: 993-1000
- [10] Mukherjee L, Singh V, Dyer C R. Half-integrality based algorithms for cosegmentation of images [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2009: 2028-2035
- [11] Hochbaum D S, Singh V. An efficient algorithm for co-segmentation [C] // *Proceedings of the 12th IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2009: 269-276
- [12] Joulin A, Bach F, Ponce J. Discriminative clustering for image co-segmentation [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2010: 1943-1950
- [13] Mukherjee L, Singh V, Peng J. Scale invariant cosegmentation for image groups [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2011: 1881-1888
- [14] Kim G, Xing E P, Li F, *et al.* Distributed cosegmentation via submodular optimization on anisotropic diffusion [C] // *Proceedings of the 13th IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2011: 169-176
- [15] Chai Y, Lempitsky V, Zisserman A. BiCoS: a bi-level co-segmentation method for image classification [C] // *Proceedings of the 13th IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2011: 2579-2586
- [16] Batra D, Kowdle A, Parikh D, *et al.* iCoseg: interactive co-segmentation with intelligent scribble guidance [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2010: 3169-3176
- [17] Batra D, Parikh D, Kowdle A, *et al.* Seed image selection in interactive cosegmentation [C] // *Proceedings of IEEE International Conference on Image Processing*. Los Alamitos: IEEE Computer Society Press, 2009: 2393-2396
- [18] Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation [J]. *International Journal of Computer Vision*, 2004, 59(2): 167-181
- [19] Vedaldi A, Fulkerson B. Vlfeat: an open and portable library of computer vision algorithms [C] // *Proceedings of the 18th International Conference on Multimedia*. New York: ACM Press, 2010: 1469-1472
- [20] Shechtman E, Irani M. Matching local self-similarities across images and videos [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2007: 1-8
- [21] Elkan C. Using the triangle inequality to accelerate k -means [C] // *Proceedings of the 20th IEEE International Conference on Machine Learning*. Palo Alto: AAAI Press, 2003: 147-153
- [22] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2006, 2: 2169-2178
- [23] Vedaldi A, Zisserman A. Efficient additive kernels via explicit feature maps [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(3): 480-492
- [24] Fan R E, Chang K W, Hsieh C J, *et al.* LIBLINEAR: a library for large linear classification [J]. *Journal of Machine Learning Research*, 2008, 9: 1871-1874
- [25] Platt J C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods [M] // *Advances in Large Margin Classifiers*. Cambridge: MIT Press, 2000: 61-73
- [26] 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
- [27] Wang J, Markert K, Everingham M. Learning models for object recognition from natural language descriptions [C] // *Proceedings of the 20th British Machine Vision Conference*. Manchester: BMVA Press, 2009: 2.1-2.11