

# 基于 Graph Cuts 多特征选择的双目图像分割方法



金海燕<sup>1,2</sup> 彭晶<sup>1</sup> 周挺<sup>1</sup> 肖照林<sup>1,2</sup>

1 西安理工大学计算机科学与工程学院 西安 710048

2 陕西省网络计算与安全技术重点实验室 西安 710048

(jinhaiyan@xaut.edu.cn)

**摘要** 双目图像分割对后续立体目标合成与三维重建等应用至关重要。由于双目图像中包含场景深度信息,因此直接将单目图像分割方法应用于双目图像尚不能得到理想的分割结果。目前,大多数双目图像分割方法将双目图像的深度特征作为颜色特征的额外通道来使用,仅对颜色特征与深度特征做简单整合,未能充分利用图像的深度特征。文中基于多分类 Graph Cuts 框架,提出了一种交互式双目图像分割方法。该方法将颜色、深度和纹理等特征融合到一个图模型中,以更充分地利用不同特征信息。同时,在 Graph Cuts 框架中引入了特征空间邻域系统,增强了图像前景区域与背景区域内部像素点之间的关系,提高了分割目标的完整性。实验结果表明,所提方法有效提升了双目图像分割结果的精确度。

**关键词:** 双目立体视觉; 双目图像; Graph cuts; 图像分割

中图法分类号 TP391.4

## Binocular Image Segmentation Based on Graph Cuts Multi-feature Selection

JIN Hai-yan<sup>1,2</sup>, PENG Jing<sup>1</sup>, ZHOU Ting<sup>1</sup> and XIAO Zhao-lin<sup>1,2</sup>

1 School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China

2 Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an 710048, China

**Abstract** Binocular image segmentation is crucial for subsequent applications such as stereoscopic object synthesis and 3D reconstruction. Since binocular images contain scene depth information, it is difficult to obtain ideal segmentation results by applying monocular image segmentation methods to binocular images directly. At present, most binocular image segmentation methods use the depth feature of the binocular image as an additional channel for the color feature. Only the color feature and the depth feature are simply integrated, and the depth feature of the image cannot be fully utilized. Based on the multi-class Graph Cuts framework, this paper proposes an interactive binocular image segmentation method. Combining features such as color, depth and texture into a graph model can make full use of different feature information. At the same time, the feature space neighborhood system is introduced in the Graph Cuts framework, which enhances the relationship between the pixels in the foreground and background areas of the image, and improves the integrity of the segmentation target. Experimental results show that the proposed method improves the accuracy of binocular image segmentation results effectively.

**Keywords** Binocular stereo vision, Binocular image, Graph cuts, Image segmentation

## 1 引言

图像分割是计算机视觉领域的一个基础研究方向,它在目标识别、图像分析和图像理解等技术中扮演着至关重要的角色。近年来,双目立体视觉技术发展迅速,市场上的双目立体相机日益增多,双目立体图像数据也不断增加,双目立体视觉技术的实际应用也越来越多,如智能机器人导航、无人机避障等,这激发了研究人员对双目图像的目标分割技术的研究兴趣。但是,将传统单目图像的目标分割方法直接应用于双

目图像分割并不能得到理想的分割结果,这是由于双目图像中隐藏了场景的深度信息,而传统的单目图像分割方法不能有效地利用深度信息,因此,需要针对双目图像的特点设计相应的分割算法,才能有效利用双目图像中的深度信息,取得理想的分割结果。

传统的图像分割是一个二分类问题,即对图像中所有像素点分配标签 0 或 1(0 表示背景,1 表示前景),将图像分割为前景和背景两部分,从而分割出目标。目前,图像分割主要可以分为两类:自动分割和交互式分割。其中,自动分割算法

到稿日期:2020-08-30 返修日期:2020-09-27 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61871319);陕西省技术创新引导计划(2020CGXNG-026)

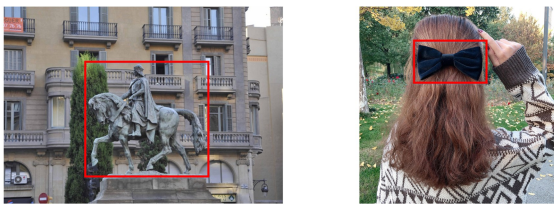
This work was supported by the National Natural Science Foundation of China(61871319) and Shaanxi Province Technical Innovation Guidance Special (2020CGXNG-026).

通信作者:肖照林(xiaozhaolin@xaut.edu.cn)

不需要用户参与,可以自动获得图像分割结果,但是当图像的前景或背景比较复杂时,这类算法很难得到满意的分割结果。而交互式分割在分割过程中需要用户参与,用户向算法提供图像的先验信息,算法通过先验信息计算图像分割结果,相比自动分割算法,这类算法更加灵活,用户可以不断对分割结果进行调整,直至得到理想的分割结果。因此,本文提出了一种交互式图像分割算法。

目前,大部分双目图像分割方法将双目图像的深度特征作为颜色特征的额外通道来使用,有些方法仅对颜色特征与深度特征做了简单的特征整合,再通过整合后的特征对图像中的像素点分类。但是,由于真实场景一般比较复杂,不同特征对分割结果的贡献也不同,这样简单地将颜色和深度特征整合在一起可能会导致整合后的特征对像素点的分类能力还没有单个特征的分类能力强。

图 1 给出了两种不同特点的场景,图 1(a)中前景和背景的颜色特征相似,深度特征对前景和背景的区分度大,此时深度特征对分割结果的贡献更大,而图 1(b)中颜色特征对前景和背景的区分度更大,此时颜色特征对分割结果的贡献更大,在这种情况下,若将深度特征作为一个额外通道整合进颜色特征中,对前景和背景区分度小的深度特征可能会干扰区分度大的颜色特征对像素点的分类,从而导致特征在整合后其分类能力比整合前单个特征的分类能力弱。



(a)前景背景深度特征差异大

(b)前景背景颜色差异大

图 1 两种不同特点场景

Fig. 1 Two different features of the scene

针对上述问题,本文通过引入特征标签将图像分割转化为多分类问题,利用多分类 Graph Cuts 框架<sup>[1-2]</sup>来求解,从而获得双目图像分割结果。本文将双目图像的颜色特征、深度特征和纹理特征融合在一个图模型中,让算法在分割过程中自动选择区分能力最强的特征来决定每一个像素点的标签。本文在双目图像的左视图和右视图之间构造跨视图邻域系统,以达到左右视图一致性约束的目的,同时,本文构造了特征空间邻域系统,通过该邻域系统加强了前景区域和背景区域内部像素点之间的关系,从而使分割出的目标更完整。

本文的主要工作集中在 3 个方面:1)提出了一种交互式双目图像分割方法,该算法在分割结果的精度上有明显优势;2)通过引入特征标签将图像分割转化为多分类问题,利用多分类 Graph Cuts 框架来求解,实现了在分割中自动选择区分能力最强的特征来决定每一个像素点的标签,为双目图像分割研究提供了一种新思路;3)构造了特征空间邻域系统,通过该邻域系统加强了前景区域和背景区域内部像素点之间的关系,从而使分割出的目标更完整。

## 2 相关工作

针对单目图像分割方法,经过多年的探索,研究人员已经

提出了一些优秀的分割方法,并将其应用到相关研究领域<sup>[3-11]</sup>。随着双目立体视觉技术的快速发展,双目图像目标分割逐渐引起了研究人员的关注,由于双目图像中包含拍摄场景的深度信息,因此不能直接将单目图像的目标分割方法应用于双目图像分割,故此类方法将双目图像的左右视图作为独立的两张图像进行分割,而未考虑到双目图像的特性,也没有充分利用双目图像中的深度信息。

双目图像分割是从两张相关的图像中提取出共同对象,而视频分割技术是从一系列相关的视频帧中提取出共同对象,这两种技术十分相似,因此可将视频分割方法用于分割双目图像。Price 等<sup>[12]</sup>提出的视频分割方法先通过用户交互分割出初始帧中的目标对象,然后提取当前帧的颜色、梯度、形状、运动信息以及相邻像素点之间的色彩关系等多种特征,再通过评估各个特征在分割时所起作用的大小来得到每个特征的权重,最后将这些特征加权组合构造出图模型,通过图割法优化得到最终的分割结果。Sun 等<sup>[13]</sup>提出了一种基于形状和颜色评估的视频分割方法,将形状和颜色模型集成到图切割优化框架中,获得目标对象的分割结果。然而,由于视频分割算法是针对视频的特点进行设计的,因此这些方法虽考虑到了物体运动和场景运动等特征,但没有利用到双目图像的对极几何特性和视差信息。

Zhou 等<sup>[14]</sup>于 2014 年提出了一种基于显著性检测和区域合并的协同分割方法,对一系列相似的图像进行联合分割,即从背景中自动提取相同或相似的物体,通过消除可能的误分类来进一步微调种子区域,合并过程从种子区域开始,并在单个图像上独立进行。协同分割这一技术也可以应用于双目图像分割。Li 等<sup>[15]</sup>提出了一种改进的抓取切割算法,该算法结合了抓取切割和基于图的图像分割,通过研究像素之间的关系来弥补抓取切割算法的缺陷。改进后的算法在背景复杂或背景目标非常相似的情况下可以获得更好的图像分割结果。Batra 等<sup>[16]</sup>提出了一种智能推荐方法,该方法可以根据用户当前交互输入的分割结果智能推荐接下来的用户输入,极大地减少了分割过程中的用户操作。然而,由于这些方法不是针对双目图像设计的,因此当这些方法应用于双目图像分割时,都无法都利用到双目图像的对极几何特性。

经过学者们对双目图像分割问题的不断研究,人们开始针对双目图像的特性来设计双目图像分割方法,这些方法考虑到了双目图像之间的关系。

双目图像的分割与其特征的选择有着重大关系,特征的选择属于降维的一项主要技术,选择最有用的特征对分析双目图像有着积极的作用。Li 等<sup>[17]</sup>于 2014 年提出了一种新的无监督特征选择算法,该算法被命名为聚类引导的稀疏矩阵结构学习(clustering-guided sparse structural learning),将其聚类分析和稀疏结构分析整合到一个联合框架中,并进行了实验评估。非负谱聚类是为了学习更准确的样本聚类标签,并指导特征选择。同时,利用不同特征共享的隐藏结构对聚类标签进行预测,可揭示特征相关性,使预测结果更加可靠。利用 Row-wiseparse 模型使所提出的模型适用于特征选择。Li 等<sup>[18]</sup>于 2015 年提出了一种新的无监督特征选择算法,即利用非负谱聚类和冗余分析相结合来约束冗余的非负谱分析。该方法可以直接识别最有用和冗余约束的特征的一个鉴

别子集,通过对聚类标签和特征矩阵的选择进行联合学习,来选择出最具鉴别性的特征。此外,其还明确利用特征间的冗余来控制所选子集的冗余。Li等<sup>[19]</sup>于2020年提出了一种带广义的不相关约束的正则化回归模型来进行特征选择,该模型具有3个优点:1)探索了低冗余和区分性特征;2)避免了琐碎的解决方案;3)简化优化。此外,通过一种新颖的无监督学习约束谱分析方法实现了局部聚类结构,将须链和不可链分别转化为内部图和惩罚图,而不是混合亲和图。该模型具有弹性梯度和全局可微性,收敛速度快。

Lo等<sup>[20]</sup>提出了一种类似于复制粘贴的针对双目图像的目标选择算法。该方法首先通过Mean-Shift方法对双目图像对中的一张图像进行聚类,然后用户通过交互输入合并聚类结果,从而得到双目图像对该视图的初始分割结果,最后通过Graph Cuts方法对初始分割结果进行迭代优化,以得到该视图的精细分割结果。该方法较为依赖用户的交互输入,并且对视差图的准确度要求较高。

Price等<sup>[21]</sup>提出了一种一致性约束的立体图像对目标分割方法。该方法首先使用左右视图构造马尔可夫随机场,然后将颜色先验信息、邻域平滑和立体匹配约束整合进马尔可夫随机场,最后使用全局优化方法进行优化,就会得到左右视图的目标分割结果。视差图用于立体图像对的左右视图之间建立平滑项,通过该平滑项促使左右视图中相匹配的像素点都被分割到前景或背景中,从而达到左右视图一致性约束的目的。接着,许多研究人员通过对Price等提出的分割框架进行改进提出了一些方法。例如,Ju等<sup>[22]</sup>将Grab Cut优化方法引入了这个框架,简化了用户交互。Ma等<sup>[23]</sup>通过对左右视图进行SIFT特征点匹配,来找到左右视图中的对应点,在这些点之间建立左右视图之间的平滑项,加快算法的求解速度。Ma等<sup>[24]</sup>通过RGB-D混合约束进一步约束双目图像的左右视图,让左右视图的分割结果更加一致。Peng等<sup>[25]</sup>提出了一种显著性感知的立体图像对目标分割方法。该方法基于图割法,其中提出的高阶能量项鼓励对应像素及其邻域内像素获得相同的标签,从而达到一致性约束。

Ju等<sup>[26]</sup>提出了一种基于轮廓的双目图像目标提取方法。该方法先通过单目图像分割方法分割立体图像对的其中一张图像,得到该视图中的前景目标及其轮廓信息,然后通过视差图得到左右视图中像素点的匹配信息,利用已分割视图的分割结果以及像素点的匹配信息在另一视图上构造一个能量函数,最后运用动态规划方法优化该能量函数,得到另一视图的前景目标。通过该方法可以将传统的单目图像分割方法引入到双目图像分割中。

由于双目图像包含颜色特征、深度特征和纹理特征,已有的研究基本上只选取了双目图像中的部分特征作为双目图像的分割依据,而忽视了其他特征对双目图像分割的影响,这将会影响其分割精度。本文将双目图像的颜色特征、深度特征和纹理特征融合在一个图模型中,让算法在分割过程中自动选择区分能力最强的特征来决定每一个像素点的标签。同时,本文构造了特征空间邻域系统,通过该邻域系统加强了前景区域和背景区域内部像素点之间的关系,从而使得分割出的目标更完整。

### 3 多特征选择的双目图像分割方法

定义双目图像 $I$ 的分类标签集合 $X = \{x_1, x_2, \dots, x_m\}$ ,其中, $m$ 表示 $I$ 的像素点数量, $x_p$ 表示像素点 $p$ 的分类标签,其形式为 $x_p = \langle l_p, \gamma_p \rangle$ , $l_p$ 表示像素点 $p$ 的分割标签,且 $l_p \in \{0, 1\}$ ,0表示该点为背景,1表示该点为前景; $\gamma_p$ 表示像素点 $p$ 的特征标签,若分割双目图像 $I$ 时使用了 $n$ 种特征,则 $\gamma_p \in \{0, 1, \dots, n-1\}$ ,此时像素点 $p$ 的分类标签 $x_p$ 有 $2n$ 种取值方式,即 $I$ 中的每一个像素点都会被分到两种类别中的一种。通过引入特征标签,本文将二分类问题转化为多分类问题,构造的能量函数如式(1)所示:

$$E(X) = E_d(X) + \lambda_1 E_s(X) + \lambda_2 E_c(X) + \lambda_f E_f(X) \quad (1)$$

$E_d(X)$ 是一次项,代表前景/背景模型; $E_s(X)$ , $E_c(X)$ 和 $E_f(X)$ 都是二次项,分别表示空间位置邻域系统、跨视图邻域系统和特征空间邻域系统,参数 $\lambda$ , $\lambda_1$ 和 $\lambda_f$ 则用于调整能量函数中各能量项的权重。下面将介绍能量函数中的各能量项。

#### 3.1 前景/背景概率模型构建

数据项 $E_d(X)$ 表示给双目图像 $I$ 中的像素点分配标签所需付出的代价,其表达式如式(2)所示:

$$E_d(X) = \sum_{p \in P} R(\phi_{\gamma_p}, l_p) \quad (2)$$

其中, $\gamma_p$ 表示像素点 $p$ 的特征标签, $\phi_{\gamma_p}$ 表示像素点 $p$ 的 $\gamma_p$ 特征值,函数 $R(\phi_{\gamma_p}, l_p)$ 用于计算点 $p$ 与通过 $\gamma_p$ 特征构造的前景/背景先验模型的匹配度,具体定义如下:

$$R(\phi_{\gamma_p}, l_p) = \frac{\delta(l_p, 1)P_{\gamma_p}(l_p | \phi_{\gamma_p}) + \delta(l_p, 0)P_{\gamma_p}(\bar{l}_p | \phi_{\gamma_p})}{P_{\gamma_p}(l_p | \phi_{\gamma_p}) + P_{\gamma_p}(\bar{l}_p | \phi_{\gamma_p})} \quad (3)$$

$$\delta(x, y) = \begin{cases} 1, & x=y \\ 0, & x \neq y \end{cases} \quad (4)$$

其中, $\bar{l}_p$ 表示与 $l_p$ 相反的分割标签(若 $l_p = 1$ ,则 $\bar{l}_p = 0$ ), $P_{\gamma_p}(\cdot)$ 表示通过 $\gamma_p$ 特征估计的前景/背景高斯混合模型的概率密度函数,通过 $P_{\gamma_p}(l_p | \phi_{\gamma_p})$ 计算该模型像素点 $p$ 属于被分配标签 $l_p$ 的概率。

在交互式图像分割中,用户通过交互标记出了前景和背景中的种子像素点,然后利用这些种子像素点的先验信息构造了高斯混合模型,最后通过高斯混合模型来判断其他像素点属于前景和背景的概率。然而如果没有用户输入的先验信息,那么一个像素点属于前景和背景的概率应该是相同的,即 $P(l_p) = P(\bar{l}_p)$ ,再结合贝叶斯公式可化简式(3),结果如式(5)所示:

$$R(\phi_{\gamma_p}, l_p) = \frac{P_{\gamma_p}(l_p | \phi_{\gamma_p})}{P_{\gamma_p}(\phi_{\gamma_p} | l_p) + P_{\gamma_p}(\phi_{\gamma_p} | \bar{l}_p)} \quad (5)$$

将上述步骤分别应用于双目图像的颜色、纹理和深度特征,以完成前景/背景模型的构造。

#### 3.2 邻域系统构建

本文方法含有传统邻域系统、跨视图邻域系统和特征空间邻域系统3种邻域系统,下面将详细介绍这3种邻域系统的构建过程以及计算过程。

##### 3.2.1 传统邻域系统

传统邻域系统主要判断两个空间位置相邻的像素点是否

处于前景和背景的边界处,针对该邻域系统的这一特点,本文对其进行修改。 $E_s(X)$ 表示传统邻域系统的能量项,其形式如式(6)所示:

$$E_s(X) = \sum_{(p,q) \in N} W_s(p,q) \quad (6)$$

其中,函数  $W_s(p,q)$  用于计算相邻像素点  $p$  和  $q$  的标签不连续惩罚,定义如下:

$$W_s(p,q) = \begin{cases} 0, & \gamma_p = \gamma_q \text{ 且 } l_p = l_q \\ F_{\min}(p,q), & \text{其他} \end{cases} \quad (7)$$

$$F_{\min}(p,q) = \min(\text{Sim}_{\gamma_p}(p,q), \text{Sim}_{\gamma_q}(p,q)) \quad (8)$$

其中,  $\text{Sim}_{\gamma_p}(p,q)$  和  $\text{Sim}_{\gamma_q}(p,q)$  分别表示像素点  $p$  和  $q$  的  $\gamma_p$  和  $\gamma_q$  特征的相似度,例如,若  $\gamma_p$  代表颜色特征,则  $\text{Sim}_{\gamma_p}$  表示  $(p,q)$   $p$  和  $q$  的颜色特征相似度;若  $\gamma_q$  代表深度特征,则  $\text{Sim}_{\gamma_q}(p,q)$  表示  $p$  和  $q$  的深度特征相似度。当像素点  $p$  和  $q$  的分割标签和特征标签相同时没有惩罚,在其他情况下,我们希望算法选择对  $p$  和  $q$  分类能力最强的特征来判断  $p$  和  $q$  是否处于目标与背景的边界,通过计算  $p$  和  $q$  在  $\gamma_p$  和  $\gamma_q$  特征的相似度来判断这两种特征对  $p$  和  $q$  的区分能力强弱。 $p$  和  $q$  在哪个特征的相似度越小,说明  $p$  和  $q$  在该特征中越不相似,即该特征可以更好地区分  $p$  和  $q$ ,将  $p$  和  $q$  在该特征的相似度作为  $p$  和  $q$  的标签不连续惩罚。

$$\text{Sim}_c(p,q) = \exp\left\{-\frac{\|c_p - c_q\|^2}{\theta}\right\} \quad (9)$$

$$\text{Sim}_d(p,q) = e^{-\|d_p - d_q\|^2}$$

$$\text{Sim}_t(p,q) = e^{-\|t_p - t_q\|^2}$$

本文使用了双目图像的颜色、纹理和深度特征,利用式(9)中的  $\text{Sim}_c(p,q)$ ,  $\text{Sim}_d(p,q)$  和  $\text{Sim}_t(p,q)$  分别计算两像素点的颜色特征相似度、深度特征相似度和纹理特征相似度。其中,  $c_p$ ,  $d_p$ ,  $t_p$  分别表示像素点  $p$  的颜色、深度和纹理特征。

### 3.2.2 跨视图邻域系统

为了使左右视图的分割结果更加一致,本文在左右视图中匹配正确的像素点对之间构造了跨视图邻域系统,通过该邻域系统来加强图模型中左右视图之间的关系。因此,在构建跨视图邻域系统前,需要先找出左右视图中正确匹配的像素点对。

根据双目立体视觉的对极几何原理,若像素点  $p$  和  $q$  匹配成功,则  $p$  和  $q$  应该处于同一极线,用  $(p_x, p_y)$  和  $(q_x, q_y)$  表示  $p$  和  $q$  的坐标,  $d^l$  和  $d^r$  表示双目图像的左右两幅视差图,  $d_p^l$  表示左视差图中像素点  $p$  的视差值,  $d_q^r$  表示右视差图中像素点  $q$  的视差值,则  $p$  和  $q$  应满足关系,如式(10)所示:

$$\begin{cases} p_y = q_y \\ p_x - q_x = d_p^l \\ d_p^l = d_q^r \end{cases} \quad (10)$$

如图2所示,左视差图中像素点  $p_l$  和  $q_l$  的坐标分别为(179,17)和(230,182),视差值分别为88和24,根据式(10)可知与像素点  $p_l$  和  $q_l$  匹配的像素点  $p_r$  和  $q_r$  在右视图中的坐标分别为(91,17)和(206,182),右视差图中  $p_r$  和  $q_r$  的视差值分别是16和24,则像素点  $p_l$  与  $p_r$  的视差值不相等,不满足式(10),因此是错误匹配,而像素点  $q_l$  和  $q_r$  的视差值相等,满足式(10),因此是正确匹配。在正确匹配的像素点对之间构建跨视图邻域系统,其能量项  $E_c(X)$  的表达式如式(11)所示:

$$E_c(X) = \sum_{(p,q) \in C} W_c(p,q) \quad (11)$$

其中,  $C$  表示双目图像中左右视图匹配成功的像素点的集合,函数  $W_c(p,q)$  用于计算正确匹配的像素点  $p$  和  $q$  的惩罚值,如式(12)所示:

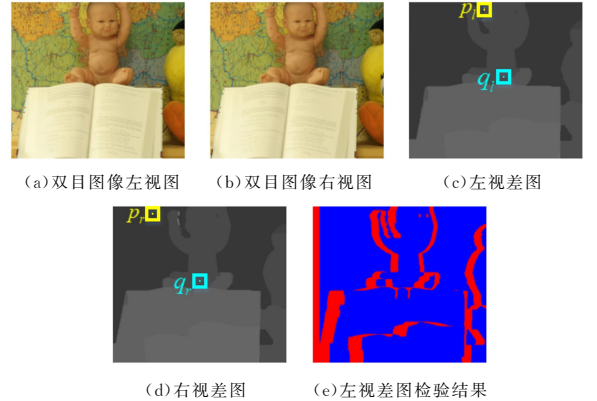
$$W_c(p,q) = \begin{cases} 0, & l_p = l_q \\ g(p,q), & l_p \neq l_q \end{cases} \quad (12)$$

其中,  $p$  和  $q$  是匹配正确的像素点,因此  $p$  和  $q$  应该都被分为前景或都被分为背景,应促使  $p$  和  $q$  的分割标签  $l_p$  和  $l_q$  获得相同的值,因此,当  $p$  和  $q$  的分割标签  $l_p = l_q$  时,惩罚值为0,当  $l_p \neq l_q$  时,通过函数  $g(p,q)$  计算惩罚值,其定义如下:

$$g(p,q) = \begin{cases} F_{\max}(p,q), & \gamma_p \neq d \text{ 且 } \gamma_q \neq d \\ F_{\min}(p,q), & \text{其他} \end{cases} \quad (13)$$

$$F_{\max}(p,q) = \max(\text{Sim}_{\gamma_p}(p,q), \text{Sim}_{\gamma_q}(p,q)) \quad (14)$$

其中,  $d$  表示深度特征的特征标签,当  $p$  和  $q$  的特征标签  $\gamma_p$  和  $\gamma_q$  都不等于  $d$  时,将  $\text{Sim}_{\gamma_p}(p,q)$  和  $\text{Sim}_{\gamma_q}(p,q)$  中的较大者作为标签不同的惩罚值,以此促使  $p$  和  $q$  获得相同标签。在其他情况下,将认为  $p$  和  $q$  的深度特征不可信,为避免不可信的深度特征影响分割结果,这种情况下我们不采用深度特征决定  $p$  和  $q$  的惩罚值。由于  $p$  和  $q$  的深度特征相同,此时  $p$  和  $q$  在深度特征的相似度达到最大,因此将  $\text{Sim}_{\gamma_p}(p,q)$  和  $\text{Sim}_{\gamma_q}(p,q)$  中的较小者作为标签不同的惩罚值。



注:红色是错误匹配像素点,蓝色是正确匹配像素点

图2 视差图一致性检验(电子版为彩色)

Fig. 2 Consistency test of parallax graph

### 3.2.3 特征空间邻域系统

传统的邻域系统是在以一个像素点为中心的  $3 \times 3$  大小的区域内构造该像素点的邻域,该邻域系统只考虑像素点在空间位置上的关系,属于空间域邻域系统。本文在特征空间构造了一个邻域系统,通过该邻域系统加强了前景区域和背景区域内部像素点之间的关系,使同一图像中空间位置相近、特征相似的像素点获得相同的分割标签,从而让分割的目标更完整。在构建特征空间邻域系统之前,需要先找到在特征空间中相似的像素点对。

通过函数  $\Phi(p,q)$  计算像素点  $p$  的特征与图像中其他像素点的特征相似度,然后找出前  $k$  个与  $p$  最相似的像素点,这  $k$  个像素点与  $p$  点构成的点对就是特征空间邻域系统像素的一部分,对图像中的所有像素点执行此过程,就得到了构成特征空间邻域系统的所有点对。 $\Phi(p,q)$  的定义如式(15)、式(16)所示:

$$\Phi(p,q) = \exp\{-\beta \text{Sim}_{\psi}(p,q)\} \quad (15)$$

$$Sim_{\psi}(p, q) = \lambda_c \omega(\psi_p^c, \psi_q^c) + \lambda_d \omega(\psi_p^d, \psi_q^d) + \lambda_t \omega(\psi_p^t, \psi_q^t) \quad (16)$$

其中,  $\psi_p^c$ ,  $\psi_p^d$  和  $\psi_p^t$  分别表示颜色、深度和纹理特征图像中以像素点  $p$  为中心的大小为  $v \times v$  的图像块, 参数  $\lambda_c$ ,  $\lambda_d$  和  $\lambda_t$  用于调整 3 种特征的权重, 函数  $\omega(\cdot)$  用于计算两个像素点的局部特征相似度, 如式(17)所示:

$$\omega(\psi_p, \psi_q) = \| H(\psi_p, \psi_q) \odot K \|_1 \quad (17)$$

其中,  $\psi_p$  和  $\psi_q$  分别表示像素点  $p$  和  $q$  的局部特征, 函数  $H(\psi_p, \psi_q)$  的计算结果是一个  $v \times v$  大小的矩阵, 矩阵中的每个元素是  $\psi_p$  和  $\psi_q$  中对应位置的元素之间的欧氏距离的平方,  $K$  是一个  $v \times v$  大小的高斯核, 符号  $\odot$  表示两个矩阵按元素相乘, 得到大小为  $v \times v$  的矩阵, 最后求出该矩阵的  $L1$  范数, 便可得到像素点  $p$  和  $q$  在该特征空间的相似度。

通过实验发现, 在某些情况下, 图像中的两个属于不同物体的像素点在空间中的距离很远, 但在特征空间中却十分相似, 若不对这种情况进行处理, 则可能导致将本不属于前景目标的区域划分到前景目标中, 造成分割质量下降。考虑到这种情况, 本文在特征空间搜索与像素点  $p$  最相似的  $k$  个像素点时, 只在以  $p$  为中心、大小为  $m \times m$  的矩形区域内进行搜索, 这样可以避免上述问题, 同时也降低了计算量。

图 3 给出了在颜色特征空间中寻找与像素点  $p$  相似像素点的过程。

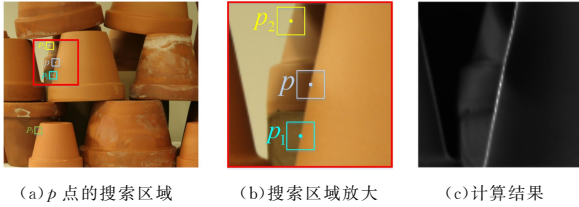


图 3 在颜色特征空间计算局部相似度(电子版为彩色)

Fig. 3 Calculation of local similarity in color feature space

如图 3(a)所示, 在以像素点  $p$  为中心、大小为  $m \times m$  的矩形区域内(见图中红色方框区域)搜索与  $p$  点在颜色特征空间中相似的点, 其中  $p_1$  和  $p_2$  是两个在  $p$  点的搜索区域内要与  $p$  计算相似度的点,  $p_3$  则是在搜索区域外与  $p$  点较相似的点; 图 3(b)是图 3(a)中红框区域的放大图, 将以  $p$ ,  $p_1$  和  $p_2$  为中心、大小为  $v \times v$  的图像块分别作为点  $p$ ,  $p_1$  和  $p_2$  在颜色空间的特征, 然后通过式(17)计算得到  $p$  点与  $p_1$  和  $p_2$  在颜色空间的特征相似度; 图 3(c)给出了  $p$  点搜索区域内的所有像素点与  $p$  点在颜色空间的相似度(越亮表示越相似)。

在找到的像素点对之间构造特征空间邻域系统, 其能量项  $E_f(X)$  的表达式如下:

$$E_f(X) = \sum_{(p, q) \in S} W_f(p, q) \quad (18)$$

其中,  $S$  是所有构成特征空间邻域系统的点对的集合, 函数  $W_f(p, q)$  用于计算像素点  $p$  和  $q$  获得标签时的惩罚值, 其定义如下:

$$W_f(p, q) = \begin{cases} 0, & l_p = l_q \text{ 且 } \gamma_p = \gamma_q \\ M(p, q), & \text{其他} \end{cases} \quad (19)$$

$$M(p, q) = \max(\omega(\psi_p^{\gamma_p}, \psi_q^{\gamma_p}), \omega(\psi_p^{\gamma_q}, \psi_q^{\gamma_q})) \quad (20)$$

其中,  $\psi_p^{\gamma_p}$  和  $\psi_q^{\gamma_p}$  分别表示在  $\gamma_p$  特征空间中以像素点  $p$  和  $q$  为中心的大小为  $v \times v$  的图像块。由于集合  $S$  中的像素点对  $(p, q)$  在特征空间中十分相似, 因此我们认为  $p$  和  $q$  应该都

属于前景或背景, 故鼓励算法赋给  $p$  和  $q$  相同的分割标签。分析式(19)可以看出, 当  $p$  和  $q$  的分割标签和特征标签相等时, 付出的代价为 0, 这种情况下代价最小, 在其他情况下, 总是选择对  $p$  和  $q$  的区分度最小的特征, 将  $p$  和  $q$  在该特征下的相似度作为赋给  $p$  和  $q$  不同分割标签的惩罚。

#### 4 仿真实验分析

本文的仿真实验中, 采用了 i5-8700 CPU 和 8G RAM, 使用 Adobe 立体目标分割数据集<sup>[15]</sup>来评估各种方法的性能, 该数据集包含 30 对双目图像。在 Matlab2018 环境下, 处理一幅图像大约需要 30ms。我们将本文方法与文献[25]和文献[27]中的方法进行比较。

仿真实验中, 使用交并比(IoU)、精度、和 F 值 3 项指标来评价本文方法。在各项指标的计算方法中, TP 为真正类, 样本类别为正, 预测结果也为正; FP 为假正类, 样本类别为负, 预测结果却为正, 属于误报; FN 为假负类, 样本类别为正, 预测结果却为负, 属于漏报; TN 为真负类, 样本类别为负, 预测结果也为负。

IoU 可以反映出正确预测的样本在真实值和预测样本的并集中所占的比例, 如式(21)所示:

$$IoU = \frac{TP}{FP + TP + FN} \quad (21)$$

精度可以反映出正确预测的正样本在预测样本中所占的比例, 如式(22)所示:

$$P = \frac{TP}{TP + FP} \quad (22)$$

R 表示召回率, 可以反映出正确预测的正 F 样本在真实样本中所占的比例, 如式(23)所示:

$$R = \frac{TP}{TP + FN} \quad (23)$$

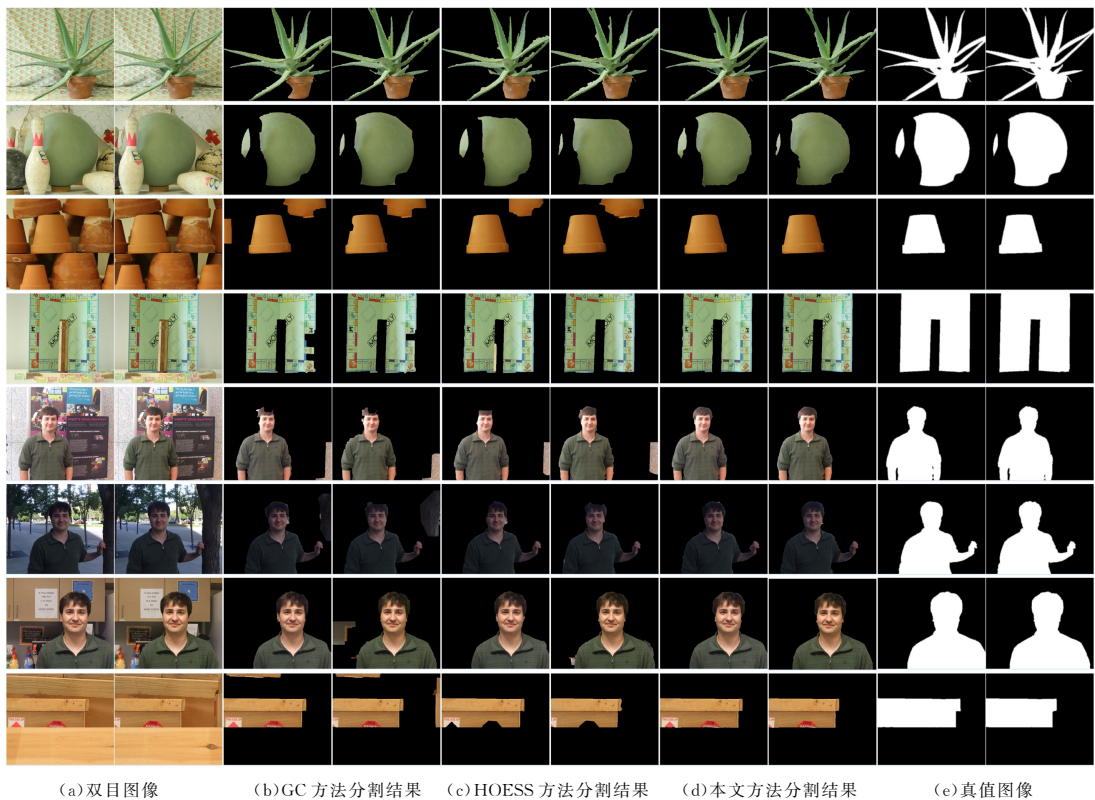
F 值是对精度和召回率的平衡, 可以反映出精度和召回率的平均值, 是一种常用的客观评价指标, 常用于评价算法的分类性能, 如式(24)所示:

$$F = \frac{2 * P * R}{P + R} \quad (24)$$

文献[27]提出了一种基于 Graph Cuts 的单目图像目标分割方法——GC。该方法是一种分割效果良好的交互式单目图像分割方法, 其在构建图模型的过程中没有使用双目图像中的深度信息, 通过与该方法进行比较可以看出双目图像中隐含的深度信息对分割结果的提升作用。

文献[25]提出了一种显著性感知的双目图像目标分割方法——HOESS。该方法将高阶能量项引入到 Graph Cuts 框架中, 又将深度信息用于高阶能量项的构造, 并且利用颜色特征构建前景/背景模型, 通过与该方法进行比较可以看出本文方法对双目图像中深度信息的使用效果更好。

为了保证实验的客观性, 本文在对比实验中使用相同的用户输入和视差图, 由于 GC 方法没有使用到深度信息, 因此在实验中将视差图像用于 HOESS 方法和本文方法。其中, 用户输入由数据集提供, 使用文献[28]中的立体匹配方法获得视差图, 使用文献[29]中的方法计算图像的纹理特征。图 4 给出了本文方法与 GC 方法和 HOESS 方法的分割结果。



(a) 双目图像 (b) GC 方法分割结果 (c) HOESS 方法分割结果 (d) 本文方法分割结果 (e) 真值图像

图 4 3 种方法分割结果对比

Fig. 4 Comparison of segmentation results of the three methods

从图 4 中第三对双目图像的分割结果可以看出, GC 和 HOESS 方法都错误地将背景区域划分到了前景目标中, 这是因为这两种方法构建的前景/背景模型都没有使用双目图像中的深度信息, 而这对双目图像中前景与背景的颜色和纹理差异较小, 此时, 使用颜色和纹理信息构造的前景/背景模型对像素点的区分能力会下降, 导致分割结果变差。本文方法构建的前景/背景模型会选择区分能力最强的特征为像素点分配标签, 当图像中前景与背景的颜色和纹理特征相似时, 算法会选择区分度最大的深度特征分配像素点的标签, 这使得本文方法的前景/背景模型更加完善。

分析图 4 中第五对双目图像的分割结果可以看出, 相比另外两种方法, 本文方法的分割结果更完整, 左右视图的分割结果更一致。这是因为本文方法构建的特征空间邻域系统加强了图像前景区域和背景区域的内部像素点之间的联系, 使得目标边界部分像素点的分割结果更加准确, 跨视图邻域系统则确保了左右视图的分割结果一致性。

表 1 列出了本文方法与另外两种方法在 IoU、精度和 F 值方面的比较结果。

表 1 分割结果评价指标

Table 1 Evaluation indexes of segmentation results

算法	<i>IoU</i>	精度	F 值
GC	0.828	0.849	0.889
HOESS	0.906	0.920	0.946
本文方法	0.971	0.986	0.985

从表 1 可以看出, GC 方法的 3 项指标都要明显低于 HOESS 方法和本文方法, 这说明深度信息对分割结果具有明显的提升作用, 本文方法则在 3 个指标上都优于 HOESS 方

法, 这说明本文方法对深度信息的利用更充分合理。

**结束语** 本文提出了多特征选择的双目图像分割方法, 通过引入特征标签将图像分割问题转化为多分类问题, 并利用多分类 Graph Cuts 框架来求解。本文将双目图像的颜色、深度和纹理特征融合在一个多分类图模型中, 让算法选择最具有区分度的特征为每个像素点分配标签, 同时, 本文在图模型中加入了跨视图邻域系统与特征空间邻域系统, 以保证左右视图的分割结果一致性和完整性。目前, 双目交互式图像数据集有限, 数据量较少, 且算法执行时间较长。为了使该领域研究工作的应用场景更加广泛, 在后续的工作中, 我们不仅应该对本文算法继续进行完善优化, 同时, 还应构建双目交互图像复杂场景数据集, 可以在后续的研究工作中为该领域的科研人员提供借鉴和参考。

## 参 考 文 献

- [1] KOLMOGOROV V, ZABIN R. What energy functions can be minimized via graph cuts[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(2): 147-159.
- [2] 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.
- [3] LIANG L M, HUANG C L, SHI F. Horizontal set of fusion shape priori for fundus image vascular segmentation[J]. *Journal of Computer Science*, 2018, 41(7): 1678-1692.
- [4] ZHU Z L, LIU Y J. Fusion chaos optimization and improved fuzzy clustering for image segmentation algorithm[J]. *Electronic*

- Journals, 2020, 48(5):975-984.
- [5] LI Y, CHEN X, ZHU Z, et al. Attention-Guided Unified Network for Panoptic Segmentation[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 7019-7028.
- [6] ZHANG Z L, LI A H, CHU W. Hyperpixel segmentation algorithm based on density peak search clustering[J]. Chinese Journal of Computers, 2020, 43(1):1-15.
- [7] QIN H F, LIU X. Finger Vein Image Segmentation Based on Sparse Auto-Encoder[J]. Journal of Chongqing Technology and Business University(Natural Science Edition), 2019, 36(4):1-8.
- [8] HE T, GUO Y G, XIN C H. ACM selection system combined with improved Chan-Vese model adaptive image segmentation method[J]. Journal of Chongqing University of Technology (Natural Science), 2020, 34(2):165-173.
- [9] YANG Z, WANG J, XIN C H. Automatic Segmentation and Classification of MR Images Using DCE-MRI Combined with Improved Convolutional Neural Network[J]. Journal of Chongqing University of Technology (Natural Science), 2020, 34(2):147-157.
- [10] WANG R H, LI B, TENG Q Z. Core FIB-SEM Image Segmentation Algorithm Based on Convolutional Neural Network[J]. Computer Engineering, 2021, 47(1):264-274.
- [11] LI P, CHEN J Q, MA W M, et al. Sonar Image Segmentation Based on Multiscale Features in Contourlet Domain[J]. Computer Engineering, 2019, 45(9):253-259.
- [12] PRICE B L, MORSE B S, COHEN S. LIVEcut: Learning-based interactive video segmentation by evaluation of multiple propagated cues[C]// Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, 2009:779-786.
- [13] SUN C, LU H. Interactive Video Segmentation via Local Appearance Model [J]. IEEE Transactions on Circuits and Systems for Video Technology. 2017, 27(7):1491-1501.
- [14] ZHOU C, LIU C. Co-segmentation of multiple similar images using saliency detection and region merging [J]. Iet Computer Vision, 2014, 8(3):254-261.
- [15] LI Y, ZHANG J, GAO P, et al. Grab Cut Image Segmentation Based on Image Region[C]// 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). 2018: 311-315.
- [16] BATRA D, KOWDLE A, PARIKH D, et al. iCoseg: Interactive co-segmentation with intelligent scribble guidance [C]// Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010:3169-3176.
- [17] LI Z, LIU J, YANG Y, et al. Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(9): 2138-2150.
- [18] LI Z, TANG J. Unsupervised Feature Selection via Nonnegative Spectral Analysis and Redundancy Control[J]. IEEE Transactions on Image Processing, 2015, 24(2):5343-5355.
- [19] LI X, ZHANG H, ZHANG R, et al. Discriminative and Uncorrelated Feature Selection with Constrained Spectral Analysis in Unsupervised Learning[J]. IEEE Transactions on Image Processing, 2020, 29:2139-2149.
- [20] LO W Y, BAAR J V, KNAUS C, et al. Stereoscopic 3D Copy & Paste[J]. Acm Transactions on Graphics. 2010, 29(6):147.
- [21] PRICE B L, COHEN S. StereoCut: Consistent interactive object selection in stereo image pairs[C]// Proceedings of the 2011 International Conference on Computer Vision, Barcelona, 2011: 1148-1155.
- [22] JU R, XU X, YANG Y, et al. Stereo GrabCut: Interactive and Consistent Object Extraction for Stereo Images[C]// Proceedings of the Advances in Multimedia Information Processing (PCM 2013). 2013:418-429.
- [23] MA W, YANG L, ZHANG Y, et al. Fast interactive stereo image segmentation [J]. Multimedia Tools and Applications, 2016, 75(18):10935-10948.
- [24] MA W, QIN Y, YANG L, et al. Interactive Stereo Image Segmentation with RGB-D Hybrid Constraints [J]. IEEE Signal Processing Letters, 2016, 23(11):1533-1537.
- [25] PENG J, SHEN J, LI X. High-Order Energies for Stereo Segmentation[J]. IEEE Transactions on Cybernetics, 2016, 46(7): 1616-1627.
- [26] JU R, REN T, WU G. StereoSnakes: Contour Based Consistent Object Extraction for Stereo Images[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). 2015:1724-1732.
- [27] BOYKOV Y Y, JOLLY M. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images[C]// Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV 2001). 2001:105-112.
- [28] ZHANG K, FANG Y, MIN D, et al. Cross-scale cost aggregation for stereo matching[J]. IEEE Transactions on Circuits and Systems for Video Technology. 2017, 27(5):965-976.
- [29] ARBELÁEZ P, MAIRE M, FOWLKES C, et al. Contour Detection and Hierarchical Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5):898-916.



**JIN Hai-yan**, born in 1976, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include computer vision, image processing, intelligent information processing and so on.



**XIAO Zhao-lin**, born in 1984, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include computer vision, computational photography and so on.