

基于图像理解视角的分割全局评价算法

钱晓华, 郭树旭, 李雪妍

(吉林大学电子科学与工程学院, 吉林长春 130012)

摘 要: 提出了一种具有全局属性的图像分割质量客观评价算法. 该方法是基于实际分割结果与参考分割之间的差异进行计算的. 从图像理解的角度出发, 根据空间信息给予误分像素不同的权重, 使得不同位置的误分像素具有不同的重要性; 在加权公式中引入了相对距离, 前景和背景搜索距离, 以获得评价算法对于分割图像具有缩放不变性的结论. 为了克服大比例过/欠分割造成的分割质量评价失真的现象, 提出一种失真惩罚项, 提高了评价算法的有效性和全局性. 实验结果表明, 本文提出的分割评价算法不仅能体现图像理解等后续处理的属性, 而且具有更强的全局性和鲁棒性.

关键词: 分割评价; 差异法; 理解权重; 缩放不变性; 惩罚项

中图分类号: TP301.6 **文献标识码:** A **文章编号:** 0372-2112 (2012)10-1989-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2012.10.015

Image Understanding Based Global Evaluation Algorithm for Segmentation

QIAN Xiao-hua, GUO Shu-xu, LI Xue-yan

(College of Electronic Science and Engineering, Jilin University, Changchun, Jilin 130012, China)

Abstract: An objective global evaluation algorithm of segmentation results is proposed, which is based on computing the deviation of the segmentation results from reference segmentation. By taking into account image understanding view, the discrepancy between two results is weighted based on spatial contextual information, thereby obtaining the different weightiness. Moreover, relative distance, foreground and background searching distance are introduced into the proposed metric, and the value of this metric is invariant to the segmentation images with varying scale. In order to overcome the irrational evaluation caused by the excessive over-segmentation and under-segmentation, a penalized term is proposed. Experiment results prove that our evaluation accords with the view of the image understanding and it is more global and robust.

Key words: segmentation evaluation; discrepancy metric; understanding weight; zoom invariant; penalized term

1 引言

图像分割是图像处理的重要分支之一, 是后续图像分析的关键步骤. 尽管图像分割算法已经研究了数十年^[1,2], 但仍然没有一个通用的分割理论和算法, 这是因为图像分割本身是一个不定问题: 对于同一个图像, 根据不同应用需求, 可以获得不同的最优分割. 因此, 一个有效的分割评价对选择符合特定应用的分割算法是非常重要的, 并且可以进一步优化参数, 以便获得最佳分割.

如图 1 所示, 图像分割评价可以分为主观评价和客观评价. 基于主观衡量的评价体系已经建立^[3], 并有学者称人是衡量分割算法的最佳判断者^[1], 但需要有足够数量具有代表性的观察者构成一个评价组来产生具有统计相关的结果, 这是个耗时耗力的过程. 客观评价又

可分为分析法和实验法^[2]. 分析法是直接分析算法本身的原理、性能和复杂度等, 而不需要实现算法, 但这类方法并没有受到过多关注, 因为这些属性通常独立于算法分割结果, 不能用来有效辨别不同分割算法间的差异. 实验法不是直接衡量分割算法, 而是通过测试分割结果的质量来间接判断分割算法. 实验法还能进一步分为两类: 优度法和差异法. 优度法是选择图像的某些特征作为优度参数来评价分割效果^[4]. 尽管这类方法不需要参考分割的先验知识, 并且是唯一的一类能够给予分割算法执行自我调整能力的评价方法, 但当评价不同算法时, 这类策略会变得不公正、不稳定. 特别当特征参数既被用来设计分割算法又被拿来评价它们的性能时, 这种评价几乎是无效的. 因此, 一个独立、可靠的评价算法应当需要参考分割.

差异法是基于实际分割和参考分割之间的不一致

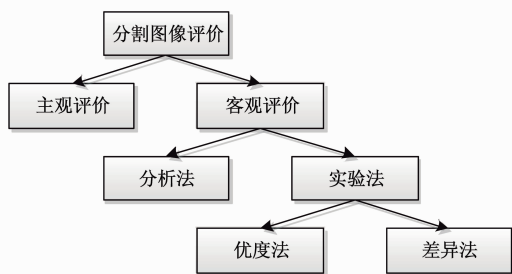


图1 分割质量评价方法分类

来评价算法性能.参考分割也被叫做金标准,可以通过手工或者可靠的方法产生^[5].差异法是分割评价中发展相对最快的一类方法^[6],可以基于误分像素数量和位置,分割目标的几何特征或者基于混合特征等^[7~12].其中,误分像素的数量是一种最基本也是最主要的图像质量评价指标,也是合测度的主要构成部分.基于像素区域的 Jaccord 测度(JSM)是实际分割和参考分割的交集与并集的比值,由于定义简单和可操作性强,近来得到了广泛的应用^[13,14].类似测度还有文献[7]提出的误分概率(PE),但是它们形式的简单也导致了评价性能的不全面.同样的误分像素,位置的不一样可能具有的重要性也不同.2000年 P Villegas^[8]首次以感知视角去分析误分像素,认为离正确分割远的误分像素相关性更大,也就是距离大的误分像素重要性大于距离小的误分像素,在此基础上提出了感知权重空间质量测度(PSQ).该测度将误分像素分为过分割区域和欠分割区域,不同区域内的误分像素的重要性也是不同的,同时,以空间距离为依据给予误分像素不同的权重.此后,以感知视角构建分割评价标准得到了广泛的关注^[9~11],特别是 E D Gelassca^[10],通过大量心理学实验来阐述感知属性的合理性,同时指出了只有位于参考目标边缘,并且影响目标形状的误分像素具有不同的感知权重,在此基础上提出了空间错误率(SAQ).此类感知评价测度存在一个通病,即存在大比重误分像素时容易造成评价失真.文献[12]在2009提出了分割评价有效性阈值的概念来解决这一失真问题(RTQ).当正确分割像素与实际分割像素比值小与某个阈值时,就认为此分割为无效分割,但该方法仍然没有完全解决这一问题,特别是存在大比重过分割情况时.

虽然上述方法提出了感知测度和有效性阈值概念,但普遍存在以下三类问题:(1)图像分割的主要目的是为后续图像理解和模式识别过程服务,因此,评价分割质量的好坏应该以后续处理的图像理解视角去分析;(2)评价测度应该对分割图像的尺寸不敏感,即具有缩放不变性;(3)针对存在大比重误分像素的情况,评价测度不仅应该能够避免评价失真,而且还能体现

分割本身无效性的趋势.本文从解决以上三个问题出发,提出基于理解视角的分割质量评价算法.本文首先给出分割质量评价算法的推理,然后提出分割质量评价算法,最后对量评价算法进行实验和分析.

2 评价算法推理

算法推理从评价的对象和原则,图像理解意义上的像素权重和全局属性的失真惩罚项三个方面进行讨论.

2.1 分割评价标准的对象和原则

提出分割标准的困难在于分割本身是不适应问题,那么对分割的评价也必然会随着任务的不同而变化.任务的不一样,会导致后续操作对分割的要求千差万别.有的处理(例如有些跟踪算法)只要能够分割目标的中心位置,并不在意目标的大小形状,即欠分割就足够了;而在视频压缩前的分割中,能够接受过分割,但对造成目标残缺的欠分割是不能容忍的.

因此建立对不同任务都适用的分割评价是一项极具挑战的任务.基于以上分析考虑,本文的研究对象不是针对所有的分割,而是为后续图像理解和模式识别服务的分割.该类分割不是一个孤立的过程,是图像理解系统的一个关键步骤.它将图像分成不同的前景和背景区域,是后续目标表达,参数测量的基础.分割的有效性应该以识别系统的整体表现来评价,后续操作的成功度是分割质量的一种体现,一个好的分割评价方法不仅应该能够区分各种算法的差别,更重要的是能够从识别系统的角度进行分割算法的选择或者参数的调整来提高目标的识别率.因此,一个分割的评价标准,应该能够体现图像理解系统好坏的趋势.

在图像理解系统中,图像分割的质量直接影响目标表达的好坏,也就是特征描述值的不一样.精确的图像分割,使得计算的特征描述符更能反应目标本身.目标表达的好坏又能影响到图像理解系统的优劣,从而图像分割质量能够间接影响目标识别.但是,在实际项目中,由于任务的目标不同,对图像目标识别率要求不一样.这样,就造成了不同任务对分割质量的依赖程度不一样.同样,即使在同一个项目中,由于先验知识的不一样,对目标分割质量的要求也是不一样的.因此,如果以任务的角度来分析分割质量,是有失公允的,甚至会得出分割评价标准无用的结论.

本文研究的分割评价标准将建立在后续操作的基础上,但是脱离具体任务的目标,摆脱先验知识的干扰,以有利于目标的精确表达为标准,以能够反应图像识别率的趋势为最高目标.

2.2 图像理解意义上的像素权重

基于像素的质量评价不仅与误分像素数量有关,也与误分像素的空间信息有关(位置).误分像素可以

分为过分割和欠分割. 过分割是将本来属性背景的像素误分为目标; 欠分割是将原属于参考分割的像素误分为背景. 过/欠分割距离是指过/欠分割像素离正确分割像素的最短距离; 外搜索距离是指背景像素离参考目标像素的最大距离; 内搜索距离是指目标像素离目标边界的最大距离. 误分像素的空间信息由这些距离来表征, 这个属性与后续操作紧密相关. 不同的误分像素具有不同的空间信息, 其理解的重要性也不同. 依照理解重要性分别给予误分像素不同的权重.

无论过分割还是欠分割, 当误分像素位于目标附近时, 对目标表达等后续操作的影响较大, 尤其是误分像素改变了目标的形状特征. 如图 2(a) 所示, 误分像素的存在改变了原来目标的形状特征, 例如重心、周长和面积等, 甚至会进一步影响目标的灰度类特征. 而图 2(b) 和 (c) 虽然误分像素数目和图 2(a) 一样, 但并不改变目标的形状, 仅是在二值图中起到干扰作用, 增加了

图像的目标数. 因此, 改变目标形状的误分像素重要性最大, 其空间属性表现为搜索距离最小.

关于过/欠分割像素的重要性, 当过/欠分割绝对距离一样时, 它们的重要性往往是不一样的: 当目标占整个图像的多数时, 过分割的重要性大于欠分割; 当目标仅占少数时, 欠分割的重要性大于过分割. 为了避免这种分类讨论, 引入相对距离, 定义为内(外)距离比上了内(外)搜索距离.

误分像素的重要性决定了像素的权重: 重要性越大, 权重越大. 重要性映射到空间信息中, 表现为相对距离越小, 权重越大.

对同一分割图像进行放缩, 如图 2(b) 和 (c), 由于采用非线性加权, PSQ 会得到不同的评价测度值. 引入相对距离的另一好处是可以避免分割图像放缩造成的评价测度的变化, 使得评价测度具有放缩不变性.

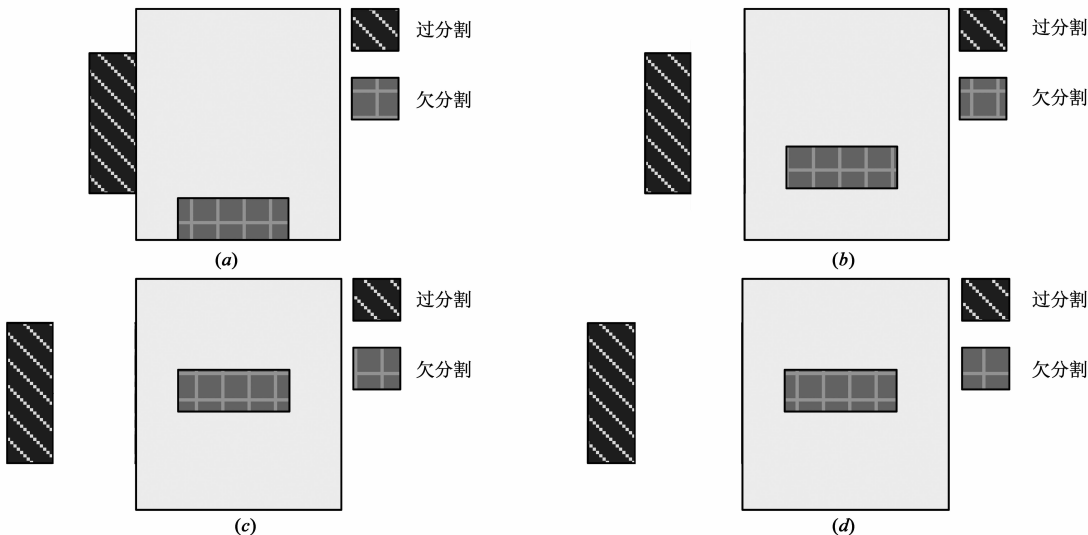


图2 不同分割情况:(a)(b)(c)的误分像素数量相同而位置不同;(c)(d)是同一分割图像的不同尺寸

2.3 全局属性的失真惩罚项

当过分割占背景比例过大, 或者欠分割占参考分割比例过大时, 现有的质量评价标准一般会得出失真的评价. 在参考分割区域内存在一个小面积分割, PE 等都会给出一个不错的评价, 但实际上这个分割毫无意义. 当参考目标在图像面积中比例很大, 并且过分割占背景比例也很大时, PE 等会给出一个较高的评价, 而这类分割实质上勾勒的是无用信息, 对后续操作毫无帮助. 因此, 出现上述情况时, 需要有一个失真惩罚的过程, 根据过/欠分割的比例, 给出失真惩罚权重, 使得评价测度能更全面的反映分割质量, 使得评价测度具有全局性.

3 分割质量评价算法

质量评价主要由理解误差和失真惩罚项构成. 假

设 R 是参考分割, S 是实际分割, 则正确分割 S_c 是 $R \cap S$, 过分割 S_o 是 $R \cap \bar{S}$, 欠分割 S_u 是 $R \cap S$.

3.1 理解误差

依照不同的理解重要性给予误分像素不同权重. 过分割和欠分割权重分别定义为

$$S_o^w = \sum_{i=1}^{N_o} w_o(x_i) S_o(x_i) \quad (1)$$

$$S_u^w = \sum_{j=1}^{N_u} w_u(x_j) S_u(x_j) \quad (2)$$

其中, N_o 和 N_u 分别是过分割和欠分割像素个数;

$$w_o(x_i) = 1 + \exp\left\{-\frac{d_o^i(x_i)^2}{2\sigma_w^2}\right\} \text{ 和}$$

$$w_u(x_j) = 1 + \exp\left\{-\frac{d_u^i(x_j)^2}{2\sigma_w^2}\right\} \text{ 分别是过分割和欠分}$$

割的加权函数,其中 σ_w 是标准方差,控制着加权值的变化趋势,主要根据具体应用来选择不同的值;

$$d'_0(x) = \frac{d_0(x)}{d_b} \text{ 和 } d'_u(x) = \frac{d_u(x)}{d_r}$$

分别是过分割和欠分割像素 x 的相对距离:像素距离和搜索距离的比值;

$$d_0(x) = \min\{dist(x, S_c)\} \text{ 和 } d_u(x) = \min\{dist(x, S_c)\}$$

分别是过分割和欠分割像素 x 到正确分割 S_c 的最短距离;

$d_b = \max\{d_x | x \in \bar{R}, d_x = \min\{dist(x, R)\}\}$ 是背景搜索距离,定义为背景内像素到前景的最大距离;

$d_r = \max\{d_x | x \in R, d_x = \min\{dist(x, \bar{R})\}\}$ 是前景搜索距离,定义为前景内像素到背景的最大距离;

$dist(x, \Lambda)$ 是像素 x 到集合 Λ 的欧式距离.

3.2 失真惩罚项

失真惩罚项表征为高斯权函数,定义为

$$\phi = 1 - \exp\left\{-\frac{\phi^2}{2\sigma^2}\right\} \quad (3)$$

其中, $\phi = \min\left(1 - \frac{S_u}{R}, \frac{S_c}{R}\right)$, $\phi \in [0, 1]$.

失真惩罚函数用来补偿评价失真.当 ϕ 小于某个阈值时,认为质量评价存在失真.阈值的选择主要依赖分割的具体应用,其经验值是 0.5.那么映射到加权高斯函数中的 σ 取 0.15.

3.3 基于图像理解的分割质量全局评价算法

分割质量评价分别定义分割准确率和错误率.分割准确率为

$$RSGQ = \chi \times \gamma \times \psi \quad (4)$$

其中, $\chi := \frac{S_c}{S_c + S_o^w}$, $\gamma := \frac{S_c}{S_c + S_u^w}$. S_c 是正确分割, S_o^w 是权重过分割, S_u^w 是权重欠分割, χ 表示存在过分割的情况下,正确分割的比率, γ 是存在欠分割的情况下,正确分割的比率,而 ψ 是全局失真处罚项.

分割错误率为

$$ESGQ = 1 - RSGQ \quad (5)$$

算法流程为

(1)先获得 S_o 过分割、 S_u 欠分割和 S_c 正确分割的面积;

(2)计算理解意义上的误分像素权重值 S_o^w 和 S_u^w ;

(3)考虑失真处罚项 ψ ;

(4)分别通过式(4)和式(5)计算分割准确率 RSGQ 和错误率 ESGQ.

4 实验与分析

针对评价失真惩罚、评价分割图像的放缩不变性和理解权重三个问题,本文分别设计了典型实验,通过计算 JSM, RTQ, PE, PSQ, SAQ, RSGQ 和 ESGQ 来比较评价测度的优劣.以下实验中 σ_w 均取 0.3, σ 取 0.15, 计算 RTQ 的阈值为 0.5, PSQ 的类 PSNR 表达计算中的 k 取 10.

图 3 是存在过/欠分割情况的图像测试序列.图 3(c)是欠分割,正确分割区域仅占参考分割的很小一部分,给后续操作提供的信息量很少;图 3(e)~(j)均是过分割,除了正确分割目标外,还将大量背景误分为目标,尤其是图 3(j)将整幅图像误认为是前景,没能提供任何目标的信息,属于完全无用分割.图 4 是 7 种测度对 8 种过/欠分割的评价. JSM, PE, PSQ 和 SAQ 均给予图 3(j)一定的评价值,没能判断出无用分割;RTQ 将比例较大的过分割直接判断为无用分割,即评价值为 0,但是,它不能体现不同程度大比例过分割的实质差别;而 RSGQ 由于引入了失真惩罚,不仅能够对无用分割给予 0 评价值,并且价值能随着欠分割像素的减少而增加,又随着过分割像素的增多而逐渐降低. RSGQ 和 ESGQ 的动态范围最大,而且是缓慢变换,能够符合视觉和图像理解属性;PE, SAQ 的动态范围较小,不能很好体现不同过/欠分割的差异;RTQ 变化过于剧烈,而 PSQ 没有归一化.

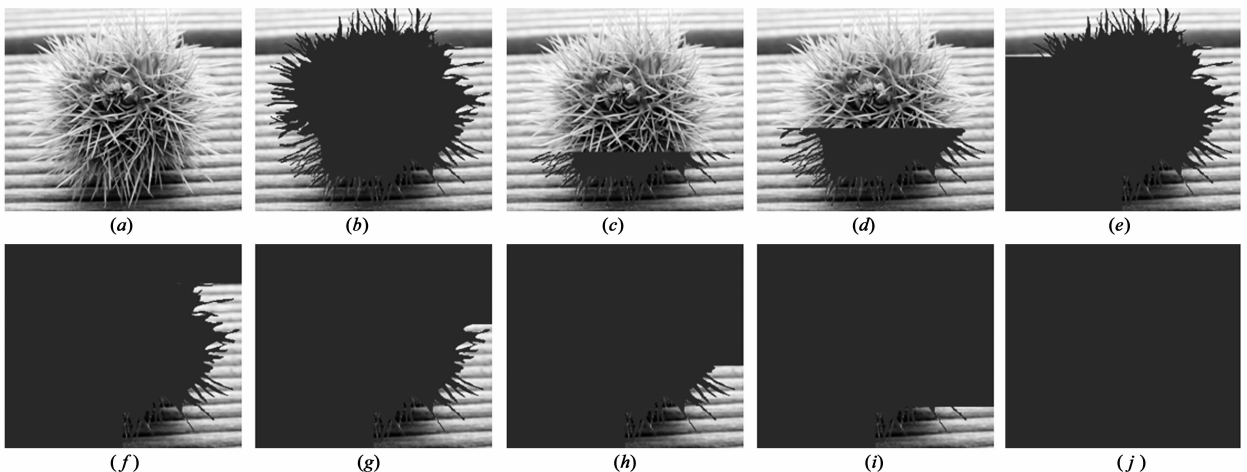


图3 不同分割结果:(a)是真实图像,(b)是参考分割,(c)-(j)是8种过/欠分割

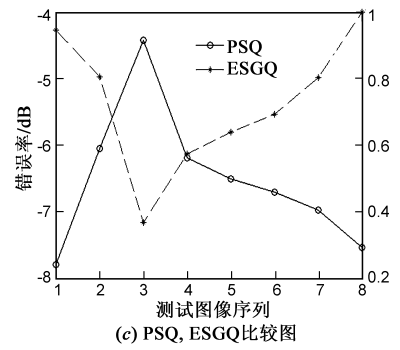
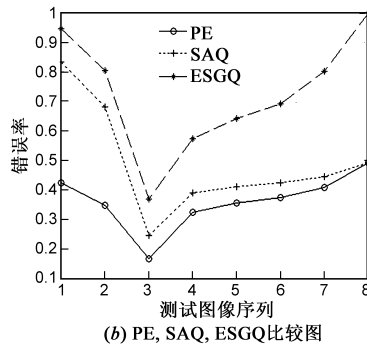
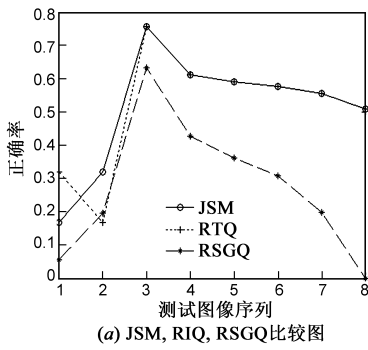


图4 7种评价测度的比较图

分别对图 5 分割图像进行放缩 0.5、0.8、1、1.6 和 2 倍,获得不同的测度值,如表 1. 为了消除变换影响,PSQ 没有进行 PSNR 表示.从图 6 中可以看出,不涉及非线性像素加权的 JSM,RTQ 和 PE 的比值较小,具有放缩不变性;而 PSQ 和 SAQ 的比值较大,表明对图像尺寸相当敏感.本文提出的 RSGQ 具有较小的比值,甚至小于没有进行非线性加权的 PE,说明 RSGQ 具有较强的放缩不变性,这主要归因于相对距离的引入.

表 1 不同尺寸分割图像的评价值

	0.5	0.8	1	1.6	2
JSM	0.7876	0.8297	0.8775	0.8450	0.8441
RTQ	0.7765	0.8227	0.8775	0.8391	0.8382
RSGQ	0.6450	0.7173	0.7888	0.7448	0.7513
ESGQ	0.3550	0.2827	0.2112	0.2552	0.2487
PE	0.0109	0.0097	0.0078	0.0091	0.0090
PSQ	0.7975	0.6489	0.4579	0.7528	0.8450
SAQ	0.2125	0.1703	0.1225	0.1551	0.1559

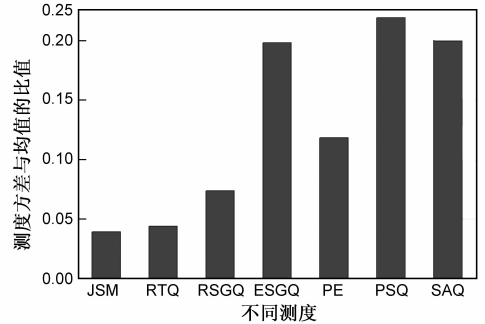


图6 不同缩放分割图像评价值的方差和均值比

图 7 是误分像素数量相同而位置不同的分割图像测试序列.从图 8(a)中可以看出,JSM 和 RTQ 在过分割情况下和位置没有关系,整个过程评价曲线几乎呈水平,而图 8(b)中的 PE 曲线在完全水平.SAQ 的变化平缓,但评价动态范围较小,不同位置的评价区别不大,相反 PSQ 变化剧烈,这两种评价均体现了误分像素离目标越近,错误的理解属性.

利用经典分割模型^[15]对 CT 肺癌结图像进行分割实验,分别选取方差 3.5、4、4.5、5、6、7、8、9、10 和 11 十个参数,代码网址是 <http://www.engr.uconn.edu/~cm-li/# Code>,分割结果如图 9.根据经验,选择方差参数为 5、6 和 7 时的分割为符合要求的最佳分割.那么,将这三种分割作为参考分割,分别对其它分割进行评价.图 10 中的分割正确率 RSGQ 客观的反映了过分割和欠分割的情况:欠分割逐渐变少,则正确率逐渐变高;过分割逐渐变多,则正确率逐渐变低.



图5 测试图像

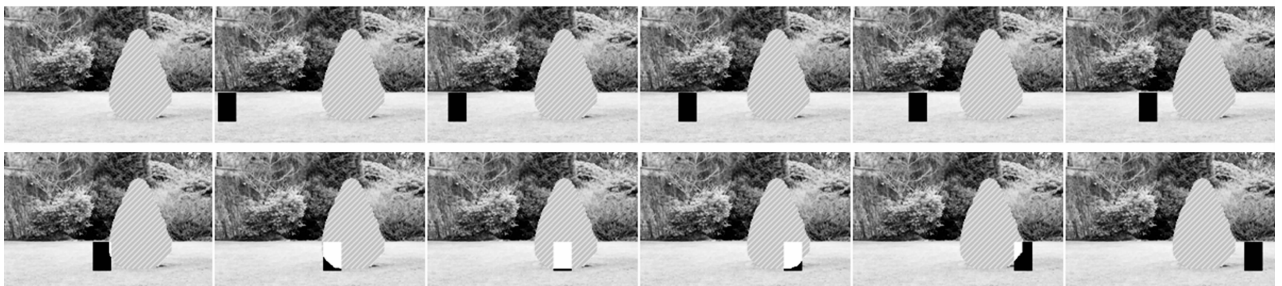


图7 误分像素数量相同而位置不同图像序列:斜线阴影部分是参考分割;黑色部分是过分割;而目标内白色部分是欠分割

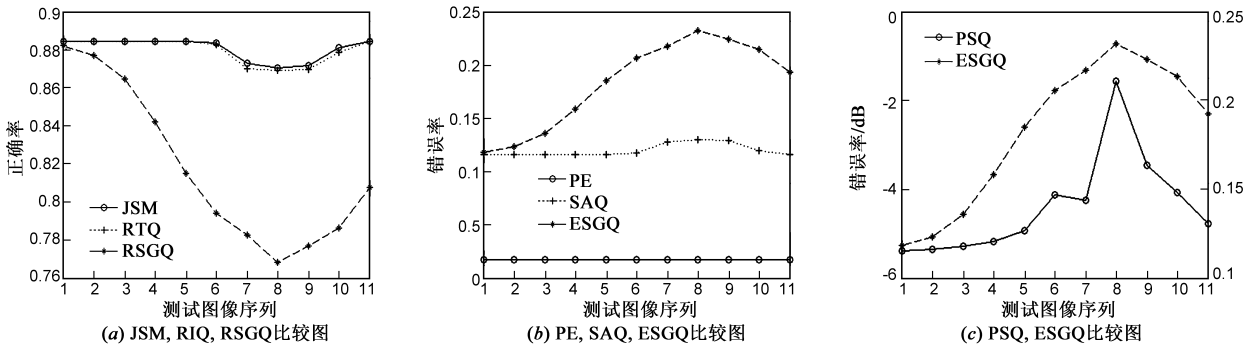


图8 7种测度对于不同位置分割图像的评价值的比较图

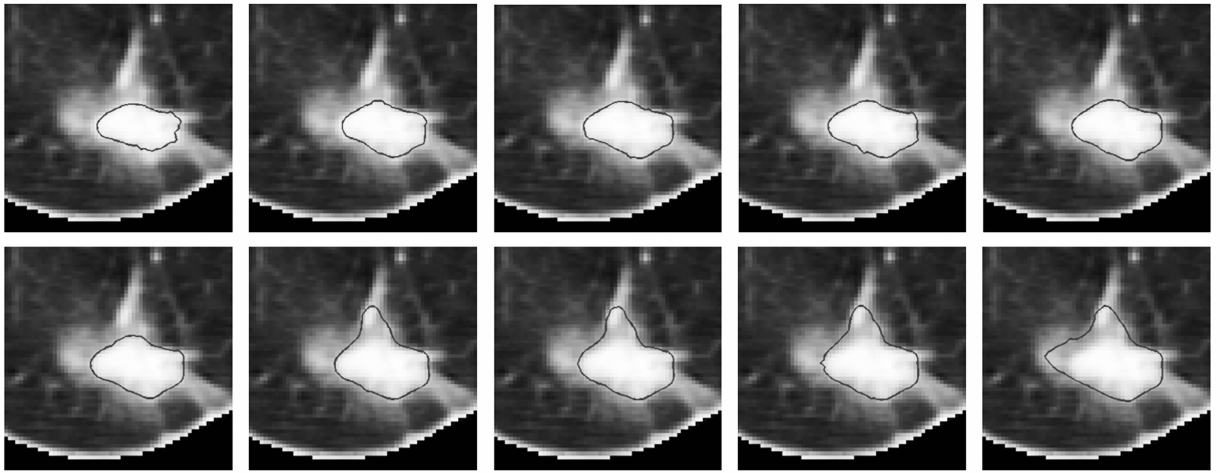


图9 CT肺癌图像的不同分割

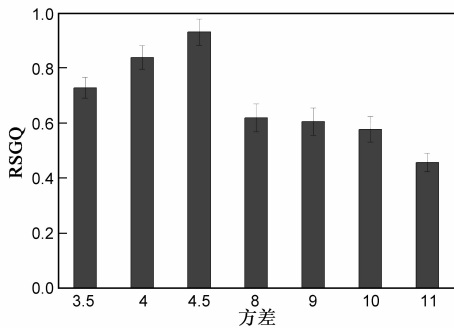


图10 不同方差情况下的分割准确率

5 总结

针对评价方法普遍存在评价失真、缺乏从图像理解的角度分析分割和对图像尺寸敏感的问题,提出了一个基于图像理解视角的全局分割图像评价标准算法.通过给予目标附近的误分像素大权重以获得具有图像理解属性的误分割像素,引入相对距离使得评价测度具有缩放不变性,并且提出了一个惩罚项来补偿失真.实验结果证明了本文所提评价算法的合理性和较强的鲁棒性.下一步的工作重点主要是提高测度的图像理解属性,使得分割评价测度更真实的反映目标

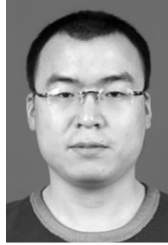
表达和识别的属性.并且单一测度仅能反映分割的一个或少数几个特性,因此可以进一步从图像评价测度融合的角度来提高测度的综合性和全面性.

参考文献

- [1] Pal NR, Pal SK. A review on image segmentation techniques [J]. Pattern Recognition, 1993, 26(9): 1277 - 1294.
- [2] Zhang Y. A survey on evaluation methods for image segmentation [J]. Pattern Recognition, 1996, 29(8): 1335 - 1346.
- [3] K McKoen, R NavarropPrieto et al. Evaluation of video segmentation methods for surveillance applications [A]. Proceedings of European Signal and Image Processing Conference [C]. Berlin: Springer-Verlag, 2000. 1045 - 1048.
- [4] H Zhang, J E Fritts, et al. Image segmentation evaluation: A survey of unsupervised methods [J]. Computer Vision and Image Understanding, 2008, 110(2): 260 - 280.
- [5] F Ge, S Wang, et al. New benchmark for image segmentation evaluation [J]. Journal of electronic imaging, 2007, 16(3): 033011-1 - 033011-16.
- [6] J K Udupa, V R LeBlanc, et al. A framework for evaluating image segmentation algorithms [J]. Computerized Medical

- Imaging and Graphics, 2006, 30(2): 75 – 87.
- [7] S U Lee, S Y Chung, et al. Performance study of several global thresholding techniques for segmentation[J]. Computer Vision, Graphics, and Image Processing, 1990, 52(2): 171 – 190.
- [8] P villegas, X Marichal, et al. Objective evaluation of segmentation masks in video sequences[A]. Proceedings of European Signal and Image Processing Conference[C]. Berlin: Springer-Verlag, 2000. 2193 – 2196.
- [9] P villegas, X Marichal. Perceptually-weighted evaluation criteria for segmentation masks in video sequences[J]. IEEE Transactions on image processing, 2004, 13(8): 1092 – 1103.
- [10] E D Gelasca, T Ebrahimi. On evaluating video object segmentation quality: A perceptually driven objective metric[J]. IEEE Journal of selected topics in signal processing, 2009, 3(2): 319 – 334.
- [11] A Cavallaro, E D Gelasca, et al. Objective evaluation of segmentation quality using spatio-temporal context[A]. Proceedings of IEEE International Conference on Image Processing [C]. New York, USA; IEEE press, 2002. 301 – 304.
- [12] 钱晓华, 郭树旭, 李雪研. 基于 Wasserstein 距离的局部能量分割模型[J]. 电子学报, 2010, 38(6): 1468 – 1472.
Qian Xiao-hua, Guo Shu-xu, Li Xue-yan. Wasserstein distance based local energy model of segmentation[J]. Acta Electronica Sinica, 2010, 38(6): 1468 – 1472. (in Chinese)
- [13] C Li, C Gatenby, L Wang, et al. A robust parametric method for bias field estimation and segmentation of mr images[A]. Proceedings of the Conferenc on Computer Vision and Pattern Recognition[C]. New York; IEEE press, 2009. 218 – 223.
- [14] C Li, C Xu, et al. MRI tissue classification and bias field estimation based on coherent local intensity clustering: A unified energy minimization framework[A]. Proceedings of Information Processing in Medical Imaging[C]. Berlin: Springer-Verlag, 2009. 288 – 299.
- [15] C Li, Chiu-Yen Kao, et al. Minimization of region-scalable fitting energy for image segmentation[J]. IEEE Transactions on Image processing, 2008, 17(10): 1940 – 1949.

作者简介



钱晓华 男, 1982 年出生于江苏省, 吉林大学电子科学与工程学院博士研究生. 主要研究方向为图像分割与医学图像处理.
E-mail: qianxh1982@gmail.com



郭树旭(通信作者) 男, 1959 年出生于黑龙江省, 吉林大学电子科学与工程学院教授、博士生导师. 主要研究方向为数字图像处理与分析.
E-mail: guosx@jlu.edu.cn