

客观的图象质量测度及其在分割评价中的应用¹

章毓晋

(清华大学电子工程系 北京 100084)

摘 要 图象分割是图象分析中的关键步骤。对图象分割的评价是研究分割技术性能的有效手段,其中分割质量测度起着重要作用。本文介绍一类新的客观质量测度并对其性能进行了比较研究。另外为了验证新测度的有效性,本文提出了一种比较合理全面的分割技术分类法。在此基础上选取各类技术中有代表性的算法借助新的质量测度进行了实际评价比较并取得了有意义的结果。

关键词 图象分割, 分割评价, 质量测度, 算法分类

中图分类号 TN911.73

1 引 言

图象分割是把图象分成各具特性的区域并提取出感兴趣的区域(目标)的技术。由于其结果对以后的特征测量和目标分类等过程有直接的影响所以是一种关键的图象分析技术。尽管人们已提出了上千种不同的算法^[1,2],但还没有一种适合于所有图象的通用的分割算法。另一方面,给定一个实际图象分析问题要选择合用的分割算法也还没有标准的方法。为解决这些问题,需要进行分割评价研究,这样既可掌握算法性能,以便在实际应用中根据需要进行选择有效算法;也可帮助改进和提高现有算法的性能,以适应各种不同的图象分割要求。另外分割评价对研究新的分割技术也具有指导意义。

多年前人们就开始研究分割评价方法。由于没有统一的分割理论所以大多数评价研究常采用实验的方法。具体就是将待研究的算法实际用于分割图象,借助一定的质量测度来判断分割结果的优劣,并根据分割结果得出所用分割算法的性能。从这个过程可以看出,客观的质量测度对分割评价是非常关键的,因为评价结果主要就是由这些测度所决定的。本文将讨论分割质量测度及其应用。第 2 节介绍我们新提出的分割质量测度并对其评价性能进行分析比较。为验证新的测度第 3 节将现有分割算法进行分类并在此基础上选取各类中的典型算法以进行评价实验。第 4 节介绍具体评价实验并分析所得到的结果。第 5 节归纳新测度的优点。

2 基于目标特征的新测度和不同特征评价性能的研究

近二十年来人们已提出了多个分割质量测度,但都存在一些问题而不能很好地满足实际应用要求^[3]。最近 Pal 等人^[4]又提出了两个质量测度。一个利用了图象内各区域的二阶熵。另一个则把参考分割图 and 实际分割图都看作一种概率分布,用这两个概率分布之间的偏差大小来评判分割结果。事实上这两个测度与文献^[3]中讨论过的测度在本质上是是一致的,只是在加权方法上有些不同,因而也存在类似的问题。

从理论上来说,一个子系统的性能应在其总系统的环境下检验才有实际意义。图象分割是图象分析中的一个关键步骤,图象分割算法的性能也应结合在图象分析这个总任务中进行评

¹ 1995-09-08 收到, 1996-02-14 定稿

国家教委留学回国人员科研资助费和教委博士点基金资助课题

判。对图象分析来说,其最终目的是从图象中测取目标的特征从而可进一步描述和识别。所以我们可以把对特征测量的精确程度作为质量测度来研究分割算法。因为这种测量在图象分析中是分割和其它处理的最终目的,所以我们把这种测度叫做最终测量精度(Ultimate Measurement Accuracy, 简称为 UMA)。具体来说,UMA 可根据对参考图 and 实际分割图中目标特征的测量值的差来计算^[3]。这个差越大说明测量精度越低,从分析的角度来说也就是分割的效果越坏。而分割效果的好坏正反映了所用分割算法的性能优劣。由于 UMA 充分考虑了图象分析的特点,直接以图象分析的结果来评判分割过程的优劣从而能定量地把握分割算法的性能。另外由于特征的测量是客观的从而消除了主观因素的影响。

这里要注意为了描述目标的不同性质在不同的实际应用中常使用不同的目标特征,所以最终测量精度 UMA 实际上是一组与目标特征有关的测度。当需要考虑目标各方面性质时,它们可以结合使用。从另一方面来说,各种特征不仅对目标的描述能力不同而且用于 UMA 中时,其对分割的评价性能也会不一样。为研究这个问题,我们根据图 1 所示的框图进行了一系列试验以确定不同目标特征的评价能力。首先利用图象合成获得具有参考目标的分割试验图,然后选取分割算法对它们进行分割。通过参数控制可逐渐改变算法的性能从而得到一系列效果不同的实际分割图。这些图如果按照算法参数的变化排列我们一般将发现其分割效果呈现从坏到好又从好到坏的趋势。这是因为当参数选取恰到好处时分割效果最好,参数值过小过大都将导致分割结果变差。利用这些实际分割图系列可对各个目标特征进行测量,再与参考目标的特征值比较就可以计算出 UMA 值从而得到相应的评价结果。

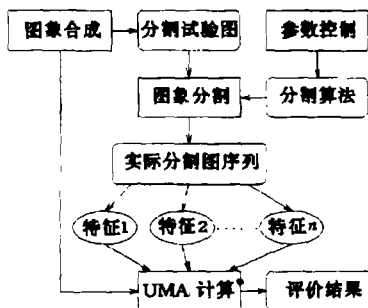


图 1 特征性能研究框图

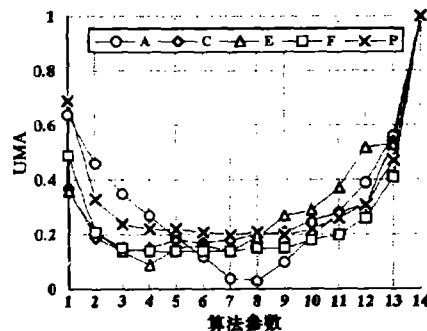


图 2 特征性能研究结果

如果算法和试验图给定,则对各个特征算得的 UMA 值就可以反映该特征的评价能力。图 2 给出对五个常用特征,即目标的面积(A),圆形性(C),离心率(E),形状参数(F)和周长(P),得到的一些试验结果,其中横轴对应算法参数的变化,纵轴对应被规一化到[0,1]间的 UMA 值。从图中可看到各特征的 UMA 值都随算法参数的改变而呈现从大到小又从小到大的变化,这说明这些特征都有表示不同分割质量从而用于评判分割算法性能的能力。不过这些 UMA 曲线也有差别,这反映了对应特征评判分割能力的不同。具体来说可从两方面进行分析。第一是看曲线的形状特别是谷的深度。因为这里特征值已规一化到了相同的范围,所以谷的深度反映了特征 UMA 值的动态范围。谷越深说明对好坏分割结果区别的能力越强。第二是看曲线的光滑程度。它反映了特征 UMA 值稳定跟踪分割结果微小变化的情况。一般说来曲线平滑表明特征反映相近分割结果比较一致。根据以上标准来看图 2 可以发现各曲线的不同特点,其中圆形性和离心率的曲线不够光滑而形状参数和周长的曲线在某些范围内几乎成为水平。相比之下面积曲线具有较明显的谷和较平滑的升降段,这表明在评价中用面积这个特征比用其它几个特征能更好地判断分割结果。

3 图象分割算法的分类和典型算法的选择

对一个质量测度来说其有效性要在实际评价工作中检验。一个重要的考虑是它应能适用于评价各种类型的分割算法。由于现有的分割算法非常多, 需要将它们分类并选取有代表性的加以研究。Fu 等人^[1]曾把分割算法分成三类: (1) 阈值分类; (2) 边缘检测; (3) 区域提取。事实上阈值分类法本质上也是一种区域提取方法, 所以 (3) 实际上包含了 (1)。Pal 等人^[2]在他们的综述中将分割算法分成六部分讨论: (1) 阈值分割; (2) 象素分类; (3) 深度图象分割; (4) 彩色图象分割; (5) 边缘检测; (6) 基于模糊集的方法。从算法的角度来看, 各部分内容是有重叠的。事实上对深度图和彩色图的分割仍需用 (1), (2), 或 (5) 中的方法进行, 至于 (6) 所讨论的只是把模糊集理论用于 (1), (2) 和 (5) 中的方法^[2]。另外 (1) 和 (2) 中的方法有许多相似之处^[1], 而常见的区域生长法却没有包含进这些类中去。

将算法分类实际上是把一个集合分成其子集的问题。一个合适的分类法应满足以下四个条件^[1]: (1) 每个算法都能被分进某一类中; (2) 各类的总和能包含所有算法; (3) 在同一类中的算法应该具有某些相同性质; (4) 在不同类中的算法应该具有某些不同性质。分类总是根据一些分类准则进行的。上述前两个条件说明分类准则应能把所有算法进行分类而后两个条件说明分类准则应能帮助确定各类算法有代表性的特性。参照这些条件我们建议用以下两个准则来对分割算法分类。

首先我们知道对灰度图象的分割常基于图象灰度值的两个性质: 不连续性和相似性。所以分割算法可据此分为利用区域间灰度不连续性的基于边界的算法和利用区域内灰度相似性的基于区域的算法两类。另一方面根据分割过程中处理策略的不同, 分割算法又可分为并行算法和串行算法两类。在并行算法中, 所有判断和决定都是独立地和同时地作出的。而在串行算法中, 早期处理的结果可被其后的处理过程所利用。串行算法一般所需计算时间比并行算法要长得多。上述这两个准则互不重合互为补充, 所以分割算法可根据这两个准则分成 4 类: PB: 并行边界类; SB: 串行边界类; PR: 并行区域类; SR: 串行区域类。这种分类法能满足上述的四个条件, 也可以包含文献 [1] 和 [2] 中所有的算法。例如基于边缘检测的方法可以是并行的或串行的, 这主要取决于边缘连接或跟踪时采用的策略, 阈值分割法和象素分类法都属于并行区域类, 而区域生长和区域分裂合并法则属于串行区域类。

目前已报道的分割评价工作大多只比较了一些阈值分割算法。本研究为使评价具有较大的广泛性和代表性, 也为了验证新测度的通用性和有效性, 我们从每一类算法中各选了一种典型的。它们是:

算法甲 (PB 类): 边缘检测边界闭合法^[5]; 算法乙 (SB 类): 动态规划轮廓搜索法^[6]; 算法丙 (PR 类): 改进的直方图凹面分析法^[7]; 算法丁 (SR 类): 分裂, 合并和聚集法^[8]。

4 实验结果和分析

运用基于目标特征的测度评价分割算法需要有参考分割图。为了保证评价研究的客观性, 我们采用合成的试验图作为参考分割图。这样不仅客观性好, 而且可重复性强, 结果稳定。而如用真实图则研究结果常受限于具体的应用, 并且由于需要人工分割以得到参考图从而在评价中引入了主观偏差。在本工作中我们根据需要合成了一组目标尺寸不同的试验图 (参见文献 [3])。图 3(a), 3(b), 3(c), 3(d) 分别给出对应图的信噪比为 64, 16, 4, 1 四种情况时的部分实验结果。图 3 中横坐标对应试验图中的目标尺寸, #1 最小, 随标号增加目标尺寸增加, #8 最

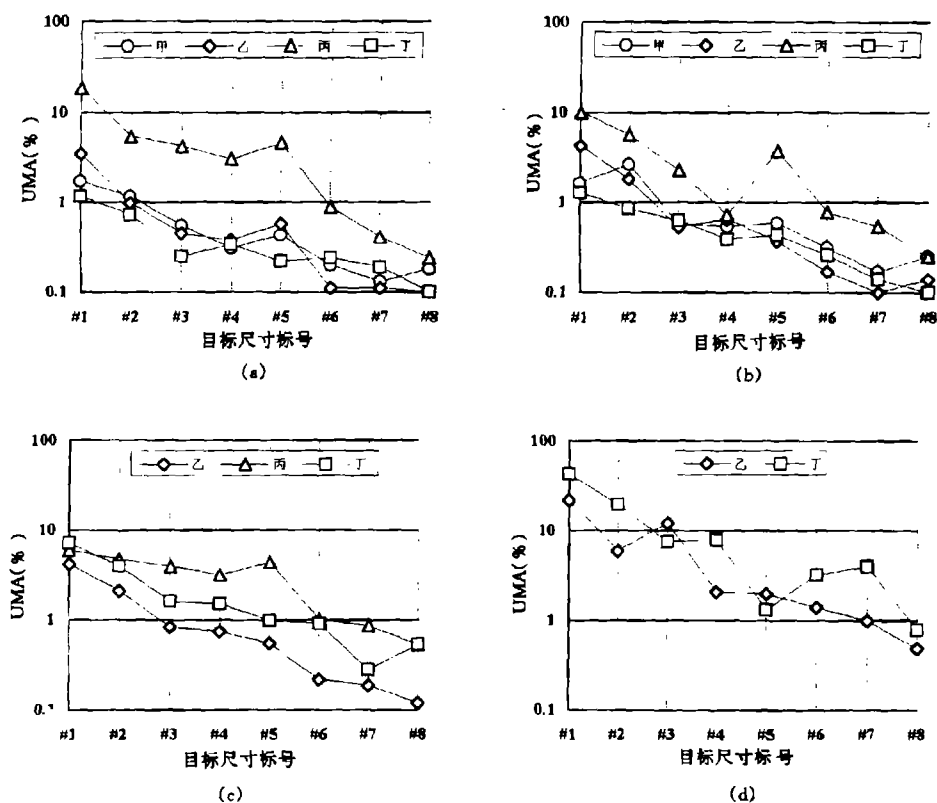


图 3 对各典型分割算法的评价结果

大。纵坐标对应根据特征量为目标面积而算出的相对 UMA 值，因为从第 2 节知道目标面积是一个具有较好评价能力的特征。

评价试验的目的是要从掌握各种算法的性能，通过对图 3 中曲线的分析可得如下结果：

(1) 所有曲线都有左高右低的趋势，且基本上单调变化。因为曲线下降表明精度提高所以可见四种算法的性能都随目标尺寸的增加而提高。这表明分割较小的目标一般要比分割较大的目标困难。

(2) 算法甲的性能当图的信噪比较大时与其它方法可以比拟。但当图象的信噪比较小时这种算法就不太合适了因为它对噪声相当敏感（在图 3(c) 和 3(d) 中因结果太差而没有标出）。

(3) 算法乙的曲线值在噪声较大时比其它几种算法都要小，即它比较抗噪声的影响。不过它对应的曲线倾斜较厉害，这说明它的性能受目标尺寸的影响比较严重。

(4) 算法丙的性能在多数情况下不如其它几种算法，尤其当目标尺寸小时差距更大。但它的性能随着图象信噪比的减小（直到信噪比为 4）变得越来越可以与其它算法相比拟（信噪比为 1 时性能很差，故略去）。这说明该算法的性能（在一定范围内）随噪声蜕化比其它算法慢一些。

(5) 算法丁在图象信噪比较大时对目标较小的图的分割效果最好。但当图象信噪比较小时性能下降较快。把它和算法乙相比两者各有所长。如果考虑抗噪声干扰的鲁棒性算法丁比不上算法乙。如果考虑目标变化时算法性能的一致性则算法丁比算法乙要好。

(6) 把图 3 各图综合起来看可以发现当图象中噪声比较大时串行类算法相对来说比并行类算法性能要好些。但这是否具有更广泛的普遍性尚需进一步的研究。

5 总 结

理论分析, 模拟试验和实际的评价结果都显示了新测度在图象分割技术的评价和比较中的适用性, 有效性和优越性。下面归纳为几点: (1) 新测度适合于评价各种分割算法, 这可由本文对各种不同类型的算法的评价实验证明; (2) 新测度适用于各种目标特征, 所以可根据不同的分析目的选择使用; (3) 新测度本身是客观的, 其评价结果不受观察者主观视觉质量的影响; (4) 新测度给出的评价结果是定量的, 直接反映了分割算法在实际应用中的性能; (5) 新测度的计算复杂度低, 而且对目标特征的测量原是图象分析中本来就要进行的工作; (6) 新测度与各种分割算法所依据的原理或准则独立, 所以在对各种算法进行比较时不会引入倾向性的偏差。

参 考 文 献

- [1] Fu K S, Mui J K. A survey on image segmentation. *Pattern Recognition*, 1981, 13(1): 3-16.
- [2] Pal N R, Pal S K. A review on image segmentation techniques. *Pattern Recognition*, 1993, 26(9): 1272-1294.
- [3] Zhang Y J, Gerbrands J J. Segmentation evaluation using ultimate measurement accuracy. *SPIE*, 1992, 1657(1): 449-460.
- [4] Pal N R, Bhandari D. Image thresholding: Some new techniques. *Signal Processing*, 1993, 33(1): 139-158.
- [5] Zhang Y J. Segmentation evaluation and comparison: A study of various algorithms. *SPIE*, 1993, 2094(2): 801-812.
- [6] Gerbrands J J. Segmentation of noisy images: [Doctoral thesis]. The Netherlands: Delft University, 1988.
- [7] Zhang Y J, Gerbrands J J, Back E. Thresholding three-dimensional image. *SPIE*, 1990, 1360(3): 1258-1269.
- [8] Strasters K, Gerbrands J J. Three-dimensional image segmentation using a split, merge and group approach. *Pattern Recognition Letters*, 1991, 12(5): 307-325.

OBJECTIVE IMAGE QUALITY MEASURES AND THEIR APPLICATIONS IN SEGMENTATION EVALUATION

Zhang Yujin

(*Department of Electronic Engineering, Tsinghua University, Beijing 100084*)

Abstract Image segmentation is a critical step of image analysis. Segmentation evaluation is an effective procedure for studying the performance of segmentation techniques, in which quality measure plays an important role. This paper presents a group of new objective quality measures for segmentation evaluation and compares their performances. In addition, to verify the effectiveness of these new measures, an appropriate classification of segmentation technique is proposed. According to this classification, several representative algorithms from different categories are selected for comparison with new quality measures. Some valuable results are obtained and presented.

Key words Image segmentation, Segmentation evaluation, Quality measures, Algorithm classification

章毓晋: 男, 1954 年生, 副教授, 目前主要从事图象工程 (包括图象处理, 图象分析和图象理解) 方面的教学和研究。