

关于图象分割性能评估的评述

狄宇春 邓雁萍

(上海交通大学图象处理与模式识别研究所, 上海 200030)

摘要 概述了图象分割性能评估的发展,总结了分割性能评估的基本理论框架:确定图象分割性能评估指标、构造评估测试图象集、评估模型与实验分析,以及分割性能评估的常用方法:统计法、分析法、基于AI的方法和混合法。对评估模型的设计作了一些尝试性的探讨。

关键词 自动目标识别 图象分割 测试图象集 图象性能指标 性能评估

0 引言

图象分割是自动目标识别系统(ATRS—Automatic Target Recognition System)的关键部分,其目的是将目标和背景分离,为目标识别、精确定位等后续处理提供依据,其结果直接影响到其后的信息处理过程。评估分割算法的性能是ATRS离线评估的重要指标之一。由于图象理解系统的开发、测试和认可必须依据该系统的性能而定,所以对ATRS的性能进行离线及模拟的评估有着重要的意义,因为只有经过正确评估方能保证ATRS有良好的在线性能,并在实时条件下提供性能的反馈信息。ATRS性能的评估十分复杂,评估既可以将其作为一个整体从软件的角度进行^[1],又可以从图象处理与模式识别的角度进行^[2],还可以进行实时评估^[3],根据种种ATRS评估经验,评估必须分阶段分步骤地进行,通常都是将ATRS分为分割、检测、特征提取以及分类等几个模块,并对它们分别加以研究和性能评估。作为识别过程关键部分的分割阶段所判断出的目标像素点与非目标像素点之间的关系直接影响着其后的信息处理过程,所以,分割算法是ATRS的重要组成部分,分割算法的性能评估可从实验和理论两方面为改进算法的设计提供定性和定量的依据。在图象分割性能评估的发展过程中不断有新的理论和方法出现,为了大致把握其发展脉络,了解国

内外常用的方法,本文首先对其发展情况作一个概略地描述,然后再作一些评述,最后对评估模型(方案)的设计作一些尝试性的探讨。

1 发展概况与评述

对ATRS算法及系统性能评估技术的研究始于70年代末,1990年,美国H. Nasr和M. Bazakos运用最优化和人工智能等技术,研制出第一套模块式的人机交互式的实时ATRS性能评估的软硬件系统Auto-I^[4]。L. G. Clark等^[5]从检测的角度对ATRS算法进行评估,着重处理了两方面的问题,第一,定义了图象的质量,包括目标对背景的度量和目标自身特性的度量两部分。目标对背景的度量所采用的指标有目标对背景的对比度、灰度域的K-S统计、边缘的 χ^2 连通性、目标对背景的熵等。第二,是定义目标自身特性的度量。它主要包括分割得出的目标灰度域中像素点个数及其灰度的标准方差、平均边缘强度、边缘标准方差以及目标的周长等。他们根据上述度量进行检测性能的评估,并将结果以检测—误警曲线表征。L. F. Bennet等^[6]在分割的性能评估方面提出了一种以分割精确性为评估指标的基于知识的评估方法,它指出了分割出的目标区域与标准参照区域的吻合程度。文献^[7]提出一种面向最终分析目的、定量地评价图象分割技术的方法。它利

用对目标特征的测量精度作为准则来判断分割图象的质量,并进而评价所使用的分割算法。用专门设计的合成图象来检验分割算法的性能,借助这些合成图可有选择地研究不同图象内容和各种干扰因素对分割的影响。文献[8]利用面向分析目的的最终测量精度思想、以最小错分概率、相对均匀测度及相对形状测度为评价准则函数,采用合成图象测试分割算法的性能。文献[9]提出了一个确定分割算法性能指标应该遵循的规则,即分割算法的性能评估不仅要考虑分割出的目标在图象中的面积元素,同时要需要考虑目标的形状因素,并据此建立了一种新的依赖目标在图象上的面积元素和形状因素的指标计算方法。文献[10]从像质评价的角度讨论了图象质量的统计描述方法,给出了12个与ATRS性能有关的图象质量指标。提出了一种基于响应函数模型的评价ATRS算法性能的系统化方法。

从评估过程的角度来看,图象分割的性能评估大致由确定图象性能指标,构造评估测试图象集,评估模型与实验分析3部分构成。

1.1 图象性能指标

在分割技术的评估中,确定分割评估性能指标应遵循的规则是最重要的,也是最困难的工作之一。性能评估指标的选取与研究决定着ATRS性能评估的成败。建立一套较为完善、合理的评估指标(如图象清晰度的度量以及该度量与误警率的关系曲线等)一直是人们普遍关注的课题,目前具有代表性且又较为成熟的观点如文献[7,8]等。他们认为在图象分析中对目标的特征测量和描述是进行分割和其它处理的最终目的,反过来说,特征测量的精度也主要取决于这些前期处理的结果。因此从分割后的图象中所测得的目标特征值的精度揭示了所用分割算法的性能。基于这种面向最终分析目的的思想,他们用这种测量精度作为准则来评判分割技术的性能,并把这些准则称为最终测量精度。由于在不同的应用中可以使用不同的目标特征,所以 UMA_i 代表一系列依赖于特征的从而面向分析目的的评价指标(准则)。在这种思想的指导下,文献[7]提出了绝对的最佳测量精度($AUMA_i$)和相对的最佳测量精度($RU-MA_i$);文献[8]选择相对形状测度、相对均匀测度和错分概率为评估准则。他们认为这些准则定量,客观地反映了目标本身的特性,而不是根据主观视觉效果来评价分割的性能。它们的值依赖于目标特征,是面向分析目的的,因而直接反映了人们对分析工作

的质量量度的需要。类似地文献[9]提出了基于图象上面积因素和形状因素的性能指标。M. D. Levine则提出一个名为均匀测度的评估准则^[11]。

1.2 构造评估测试图象集

图象分割方法的性能评估依赖其测试图象。测试图象既可以采用合成图象^[7,8]、又可以采用真实图象^[12]。文献[7]认为:作为对评估方法的基本要求,它应采用所有研究者都可获取的通用图象,同时这些图象应尽可能反映客观世界的真实情况和实际应用的共同特点。文献[8]认为图象分割性能评估依赖其测试图象,如果仅采用某一领域的图象,其测试结论必然有片面性,因此要求测试图象应尽可能反映客观世界。基于图象合成的思想,文献[7,8,10,13]等采用一种实景目标图象与背景图象合成的方法为评估系统提供输入测试信号。图象合成中重要的一点是产生的图象应能反映客观世界,这就需要它们把应用领域的知识结合进去,它们应可以调整以适应诸如图象内容的变化、各类获取图象的条件等实际情况。文献[7]的图象合成系统可以满足上述条件,它包括四个相关的部分:组成基本图、目标变换、模拟干扰因素^[8]、图象组合。

1.3 评估模型与实验分析

从评估方法的角度来看,图象分割的性能评估方法大致可分为4种方法:统计法、分析法、基于AI的方法、以及常用的混合法。

统计法 利用统计学中的参数估计、假设检验、回归分析与统计推断等方法建立评估模型^[5,10,14]。在文献[5,10]中通过Logistic回归模型、以图象质量指标为自变量、利用多元线性相关回归因素分析确定自变量与因变量之间的相关程度,选择那些相关性强的图象性能指标、舍去那些相关性弱的图象性能指标,建立Logistic回归响应函数,作响应曲线,观察分割算法关于目标与背景的对比度变化时的稳定性,作等概率曲线观察信噪比变化、目标与背景对比度变化对算法性能的影响。

分析法 针对各种算法的分割结果,计算性能指标值来比较其性能的优劣^[6,9]。

基于AI的方法 借助人工智能原理来建立评估指标(模型)进行评估^[4,6]。

混合法 在实际工作中人们常常将上述的方法综合起来用于评估分割算法的性能^[4~6]。

从评估过程的角度来考虑,目前在性能指标和

评估测试图象集以及实验中的参数选择等方面已完成的工作表现得较为成熟(例如人们根据面向最终分析目的的思想来建立分割评估性能指标,采用合成图象来构造评估测试图象集等),而在评估模型、特别是综合评估方面的工作显得相对单薄。结合当今运筹学与系统工程研究成果、对图象分割性能作全面的综合评估工作则极为罕见。

2 关于评估模型(方案)设计的一些探讨

目前对图象分割性能的评估大多数是从某个方面或少数几个方面(如某些形状特征、纹理特征,误警率,虚警率等)来进行的,而从多角度、全方位进行评估的少见。这其中有些以性能指标值为准则进行评估时,存在指标值奇异性和对分割性能不灵敏的缺陷^[6]。虽然,有的工作建立了依赖目标在图象上的面积因素和形状因素的性能指标,有的工作结果表明,分割性能与目标大小、目标和背景的灰度对比度、方差和噪声等因素有关^[8],但是从多角度、全方位的综合评估工作做得少。也有用回归分析作综合评判的^[5,10],但是,回归分析有其局限性(①要求大样本量,②要求样本有较好的分布规律,③计算工作量大,④可能出现量化结果与定性分析结果不符的现象);其次,目前获得的回归模型中的自变量只是少数的性能指标,而其它大部分性能指标对整体性能的影响则被忽略了。总之,目前从多角度、全方位(既考虑形状特征、又考虑纹理特征及误警率、虚警率等因素)进行综合评估的研究工作做得少。鉴于这种状况,我们就综合评估模型结合当今运筹学与系统工程的最新研究成果作一些探讨。

我们希望设计的评估模型既可以从不同角度层次进行性能评估,又可以在各层次评估的基础上作全方位、多角度的综合评估,即它是多功能的。事实上,对图象分割乃至整个 ATRS 进行评估,如同对宏观管理领域中的决策、项目、方案、计划、工程(以下简称项目)进行评估一样,一般都要求多角度、全方位地展开,即从每一个角度都可以对所有项目进行单指标(单准则)的评估,这时每项指标只能反映项目在某一侧面的相对优劣。在单一准则下评估较为容易。要全面地评价项目,就必须综合考虑项目的各项质量性能指标。但在多准则下的综合评估就较为困难了。首先项目在各指标下衡量的结果往往互有长短,比如有的分割算法相对受目标尺寸的影响最小^[7,8],有的分割算法相对受噪声影响程度最小;

其次出于不同的评估目的对准则的选取和偏重会对评估结果产生重大的影响;加之问题中既有定量信息(如计算给定的分割后的图象的形状特征值、纹理特征值的结果等)又有定性信息(对各个性能指标相对重要程度的评价)、既有精确信息又有模糊信息还有灰色信息(信息部分确知,部分不确定),从而使得多指标评估问题较为复杂。

为此人们在研究多准则(指标、属性)综合评估问题时,很容易想到的方法就是降维,减少准则的数量,期望构造出一个新准则以替代原准则进行评估。这种构造出来的新准则应能满足:(1)数目应大大地少于原准则数,最好是只有一个。(2)新构造的准则应尽量地反映出项目(在此即为分割算法)在原准则下的性态和各种信息。(3)若新准则多于一个则它们之间应互不相关,而且不同的新准则具有不同的重要程度。

如果能找到满足上述条件的替换准则,则原多准则下的综合评估问题就变成了一个在单一准则(或少数准则)下的简单比较问题。为实现这一目的,人们进行了各种不同的尝试。其中较有影响的方法有:AHP(层次分析)法,PCA(Principal Components Analysis)法,TOPSIS(Technique for Order Preference by Similarity to Idea Solution)法,DEA(Data Envelopment Analysis)法,基于相对地位不低于关系替代指标的 ELECTRE I、II 法,灰色系统决策评估法,模糊评价法,可拓决策评价方法,基于人工神经网络的综合评判模型以及将已有方法综合运用综合型方法等。这些实用决策方法对于解决复杂的决策(评判)问题带来极大的方便,较好地解决了决策思维的定量化描述,受到了决策者的欢迎,在工程设计方案评价技术、经济、社会、军事、政治等方面得到了广泛的应用^[15~17]。尽管这些决策方法在众多领域里得到广泛的应用,但是将它们应用于图象分割以及 ATRS 的性能评估方面的报道却极为罕见。鉴于基于熵权的模糊层次分析法的雷达情报质量评估系统已应用于防空指挥自动化系统、效果较好^[16],而且国外也有将 FUZZY AHP 法用于武器系统性能评估的报道(见文献^[16]后所列文献),他山之石、可以攻玉,那么作为尝试这些实用决策方法经适当改造后应用于图象分割以及 ATRS 的性能评估也是可行的。

目前较成功的模型有多层次灰色关联评估、灰色模糊综合评价、模糊层次分析模型等^[15~17]。作为示例,不妨以多层次灰色关联度评估模型为例简述

其方法如下:

系统层次结构模型与评估指标体系的建立

(1)层次结构模型 图象分割性能评估系统可分为三个层次:第1层次 目标层(最高层、决策层)即图象分割性能评估决策系统;第2层次 指标(准则)层(图象性能指标层),假设其包括3个子系统:形状特征、纹理特征、错分率;第3层次 具体的系统综合评估指标体系。

(2)评估指标体系 图象分割性能评估系统由3个子系统组成:①形状特征指标:如面积、周长、密度(圆度)等,可根据面向最终分析目的思想来确定^[7]。②纹理特征指标:如基于图象灰度直方图的均值、方差等,亦可根据面向最终分析目的思想来确定^[7]。③错分率:误警率、虚警率^[13]

指标权重的确定

确定各指标的权重是综合评估的难点之一。由 T. L. Saaty 教授于 70 年代中期提出的 AHP 法,是一种建立在专家咨询之上的优化方法,能把复杂系统中各因素(指标)划分为相互关联的有序层次,形成一个多层次的的分析结构模型,把多层次(指标)的权重赋值化为各指标重要性的比较,弥补了人的大脑难以在两维以上空间进行全方位扫描的弱点,便于专家对各层次各指标进行较客观的评价。AHP 作为一种对非定量事件作定量分析的一种实用决策方法、对于解决复杂决策(评判)问题带来极大的方便,深受决策者们的欢迎,被广泛应用于经济、技术、社会、军事、政治等方面。AHP 法的三大步骤如下:

- (1)建立递阶层次结构。
- (2)确定各层次中元素的优劣次序。
- (3)计算方案层元素对系统目标的合成权重,确定所列方案的优劣次序,并检验总体的一致性。

灰色关联度评估与分析

由我国学者华中理工大学邓聚龙教授于 80 年代初创立的(信息部分已知、部分未知的)灰色系统理论把一般系统论、信息论和控制论的观点和方法延伸到抽象系统,结合运用数学方法发展了一套解决信息不完备系统的理论和方法。以灰色关联度为测度的灰色关联度评估步骤^[15]如下:

- (1)单层次灰色关联评估。
- (2)多层次灰色关联评估。
- (3)评估结果分析。

小结

灰色多层次关联评估法是一种将灰色关联评估法与 AHP 法相结合的多层次评估法,较好地解决了

复杂系统多层次、多指标评估时确定权重这一难点,使评估结果更具有客观性。而且,该模型既能对复杂系统的各层次子系统进行评估,例如在各种条件(信噪比、对比度等)下从形状特征的角度评估,或从几何特征的角度评估,或从错分率的角度评估等;又能在子系统评估的基础上进行综合评估。参照文献[17]对综合评估向量作适当改进,可使模型尽可能多地吸取模式中的特征信息、使评估模型的结果分辨程度提高,数据离散性增强,判别能力更高,易于决策并提高决策的可信度,避免以往某些评估方法指标值中的奇异性和对分割性能的不灵敏性(失效性)。(在此通过定义并计算离散度^[17]来比较不同评估模型的数据离散性强弱。)

总之,这个多层次灰色关联度评估模型既可以从不同侧面对复杂系统的各层次子系统进行评估,又可以在各层次评估的基础上作全面的综合评估。因而这是一个多功能的评估模型。以上对分割评估模型的探讨仅是我们的一家之言,但愿它能在图象分割乃至整个 ATRS 性能评估中起到抛砖引玉的作用。

3 结束语

本文对图象分割性能评估进行了概要地评述,并对评估模型的设计作了一些探讨。目前在性能指标和评估测试图象集以及实验中的参数选择等方面的工作较为成熟,而在综合评估方面的工作相对薄弱。由此对其今后的发展趋势提出一些看法:(1)由于每种评估方法各有其特点,将继续在出于不同评估目的分割评估中应用,并在实践中提高和改进。(2)借鉴当今运筹学与系统工程最新研究成果的综合评判方法将引入图象分割以及整个 ATRS 的性能评估中,特别是人工神经网络因其优越的性能必将被引入图象分割及整个 ATRS 的性能综合评估中。(3)多种方法的综合将是一大重要趋势。神经网络与 FUZZY 理论、灰色系统理论与层次分析法等等的结合,将充分发挥各自的优势,使图象分割及整个 ATRS 性能综合评估日趋完善。

参考文献

- 1 Schneidewind N F. Validating Metrics for Ensuring Space Shuttle Flight Software Quality. Computer, 1994, (8): 50~57.
- 2 Sadjadi F A, Bazakos M. A perspective on automatic target recognition evaluation technology. Opt. Eng, 1991, 30(2): 141~

- 146.
- 3 Venkateswarlu R, et al. Real-time Evaluation Techniques for Real-time Image Processing Hardware. SPIE. Tracking, and Pointing VI, 1992; 510~519.
 - 4 Sadjadi F, Bazakos M. Automated Evaluation and Adaption of Automatic Target Recognition Systems. SPIE. 1990, 1310: 108~119.
 - 5 Clark L G. Image characterization for automatic target recognition algorithm evaluations Opt. Eng. 1991, 30(2): 147~153.
 - 6 Bennett L F. Knowledge-based evaluation of the segmentation component in automatic pattern recognition systems. Opt. Eng. 1991, 30(2): 154~165.
 - 7 章毓晋. 一种评价图象分割技术的新方法. 模式识别与人工智能, 1994, 12: 299~304.
 - 8 刘文萍, 吴立德. 图象分割中阈值选取方法比较研究. 模式识别与人工智能. 1997, 9: 271~277.
 - 9 邓雁萍, 李介谷. 图象分割的性能评估. 模式识别与人工智能, 1996, 6: 144~148.
 - 10 熊艳, 张桂林, 彭嘉雄. 自动目标识别算法性能评价的一种方法. 自动化学报, 1996, 3: 190~195.
 - 11 Ng W S, et al. Comment on Using the Uniformity Measure for Performance Measure in Image Segmentation. IEEE Trans. on PAMI, 1996, 18(9): 933~934.
 - 12 Heath M D, et al. A Robust Visual Method for Assessing the Relative Performance of Edge Detection Algorithms. IEEE Trans. on PAMI, 1997, 19(12): 1338~1359.
 - 13 Kanungo Tapas, et al. A Methodolgy for Quantitative Performance Evaluation of Detection Algorithms. IEEE Trans. on image processing, 1995, 4(12): 1667~1673.
 - 14 Cho K, et al. Performance Assessment Through Bootstrap. IEEE Trans. on PAMI, 1997, 19(11): 1185~1198.
 - 15 罗庆成, 何勇. 农业综合生产力的多层次灰关联评估. 系统工程理论与实践, 1994, 7: 75~80.
 - 16 王金根, 刘贺普. 基于熵权的模糊层次分析法在雷达情报质量评估系统中的应用. 系统工程与电子技术, 1996, 6: 9~14.
 - 17 苏洪潮, 王金根. 一种灰色模糊综合评判模型. 系统工程与电子技术, 1997, 7: 48~51.



狄宇春 讲师, 在职研究生. 研究方向为图象处理与模式识别、计算机视觉。



邓雁萍 博士, 副教授. 研究生导师, 研究方向为图象处理与模式识别、计算机视觉等。

Review of the Performance Evaluation of Image Segmentation

Di Yuchun, Deng Yanping

(Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai 200030)

Abstract The Progress of the research of evaluating image segmentation algorithms is discussed in this paper. First, three basic theoretical frameworks, which form the important research contents of evaluation are reviewed and classified as follows: Image metrics, Testing image set, Performance evaluation and analysis of experiments. Then, the evaluation methods of image segmentation are summarized: Statistical approaches, Analytical approaches, AI-based evaluation, Hybrid approaches. Finally, The comprehensive evaluation model, a new idea and way to comprehensive evaluating of image segmentation is discussed.

Keywords Automatic target recognition, Segmentation, Testing image set, Image metrics, Performance evaluation

中国图象图形学学会技术咨询服务部成立

中国图象图形学学会经讨论研究决定成立图象图形工程技术咨询服务部, 该部由吴章江研究员负责。其宗旨是发挥学会智力知识密集、人才资源雄厚的优势, 面向应用、面向生产实际、急广大科技工作者之所急, 为工程应用和生产实际中遇到的技术难题提供咨询服务。其主要业务范围包括:

- (1) 为各行业应用图象图形技术解决生产和工程疑难问题答疑解惑。
- (2) 为实施图象图形工程应用提供可行性研究、方案论证、技术分析、效益评估等, 直至全面解决方案的系列服务。
- (3) 图象图形新技术、新产品、新成果的推广、转让、展示、测评、推介以及营销策划、市场包装等服务。
- (4) 其他与图象图形技术相关的特殊需求服务。

该部联系地址: 北京海淀区花园路6号 邮编: 100088

电话/传真: (010)62378784

(吴章江)