

图象分割评价技术分类和比较

章毓晋

(清华大学电子工程系, 北京 100084) (汕头大学人工智能与模式识别研究所 汕头 515063)

摘要 本文对图象分割评价,特别是评价方法和评价准则的研究进展,作了一个综述。现已提出的多数分割评价方法可分为直接法和间接法。直接法研究分割算法本身,而间接法根据分割图象的质量来评价算法的性能。直接法和间接法都需要借助一定的评价准则来进行。本文对已提出的各种准则进行了分类与分析比较,并对常用的评价准则进行了实验比较,通过分析其性能表现排出了它们的优劣次序。本研究为在实际中正确地应用不同的评价方法提供了依据并指出了若干进一步研究的方向。

关键词 计算机视觉,图象分析,图象分割,分割评价,评价准则,图象质量

1 前言

在计算机视觉和图象分析的研究和应用中,图象分割是一种基本的和关键的技术。图象分割的过程是把图象分成各具特性的区域并把其中感兴趣的区域(也叫目标)提取出来的过程。图象分割对把输入图象转化为分割图象从而进一步提取目标特征、进行目标测量和分类以及其后的高层处理等都非常重要。从60年代开始人们就对图象分割进行了大量的研究,至今已提出了上千种针对各种具体应用的分割算法^[1-7]。

尽管人们在图象分割方面做了许多研究工作,但由于尚无通用的分割理论^[3],现已提出的分割算法大都是针对具体问题的,并没有一种适合于所有图象的通用的分割算法^[7]。另一方面,给定一个实际图象分割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研究对图象分割的评价问题。分割评价是改进和提高现有算法性能、改善分割质量和指导新算法研究的重要手段^[8]。

本文将对图象分割评价已有的研究成果和各种

评价方法作一个简单介绍。因为评价方法的性能主要取决于所采用的评价准则,所以本文将把重点放在对各种评价准则的分类讨论上。并在对各个准则概括介绍的基础上,对它们的特性进行了分析比较。还对其中一些常用准则设计方法进行了实验比较,得出了它们的优劣次序。最后对今后的工作方向提出了建议。

2 图象分割评价和评价方法

图象分割评价主要有两方面的内容:一是研究各个分割算法在不同分割情况中的表现,掌握如何选择和修正其参数的设置,以适应分割具有不同内容的在不同条件下获得的图象的需要;另一是分析比较多个分割算法在分割相同图象时的性能,确定各个算法的优劣,以帮助在具体分割应用中选取合适的算法。这两方面的内容是互相关连的。

人们在评价图象分割技术方面从70年代后期开始已进行了近20年的研究,提出了不少评价方法^[8,9]。从研究的角度来说,这些方法可分为两类:一

• 本工作得到教委留学回国人员科研资助费和教委博士点基金资助
收稿日期:1995.09.15

类可称为直接方法,也可叫分析法;另一类可称为间接方法,也可叫实验法。它们各自的特点及相互间的区别见图 1。在图 1 中我们给出一个简单的分割流程示意图(点划线框内),不同评价方法的作用点也标在了图上。由图 1 可见分割是利用一定的分割算法把输入初始图象转变为输出分割图象的过程。这个过程与分割算法,输入图和输出图都有关系。同样要评价分割技术,各种方法的作用点不同。直接评价法研究分割算法本身,分析它们的原理、性质、特点,从而推断和评判算法的优劣;而间接评价法去研究输出分割图的质量,或由输入图得到的参考图与输出图的差别从而通过归纳总结得到分割算法的性能。同样要评价分割技术,各种方法的难易程度也不同。用间接法评判算法需要用算法对图象进行分割实验以得到输出分割图(有时还需获得参考图),而用直接法则只需要对算法本身进行分析就可以了。

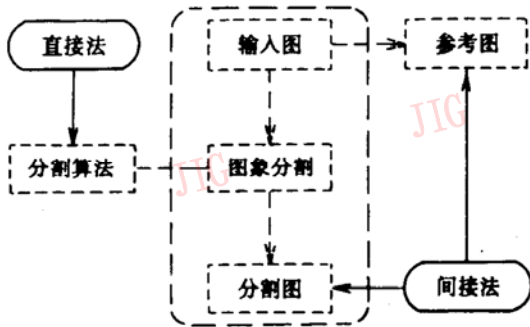


图 1 图象分割评价中不同方法示意图
Fig. 1 Diagram showing different methods for image segmentation evaluation

不管是直接法或间接法,它们基本上都有两个关键步骤或内容:一个是对分割算法进行分析或实验的机制和途径;另一个是用来评判算法特性的评价准则,也叫测度或指标。评价准则对直接法或间接法都很重要。对直接法来说要分析算法的某些特性,需要有衡量这种特性的测度。算法的特性要根据这种测度才能判别和比较。对间接法来说,因为它们是靠对图象分割后的结果进行研究来决定所用分割算法的性能的,因而定量的(可比的)和客观的评价指标是非常关键的,算法的优劣就是靠指标值来确定的。因此可见,要对图象分割进行评价,必须要采用合适有效的性能评价准则。

3 评价准则分类和介绍

目前对分割评价的研究与对分割算法的研究相比还差之甚远。所提出的评价准则归纳起来只几十

个,我们把它们分成 13 组。根据它们是用在直接法中或间接法中可分成分析的准则和实验的准则两种。另外根据准则本身的特性,每种准则又可分成定性的和定量的两种。综合起来分割评价的准则可分成各有其特点的 4 类(表 1)。下面按这个分类法对各种准则分别作一简单介绍。

表 1 分割评价准则分类

Table 1 Classification of segmentation evaluation criteria

	定性	定量
分析	L1	L2
实验	L3	L4

3.1 定性分析准则

3.1.1 所结合的先验信息(A priori information incorporated)

在设计分割算法时,一些有所要分析的图象的特性信息可以被用来提高分割的稳定性、可靠性和效率等。某些算法借助于对实际应用的先验知识有可能取得较好的分割结果^[10]。研究结果表明,先验信息的应用对分割算法的性能有较大影响。根据不同算法所结合的先验信息的种类和数量的不同,可以从一定程度上了解它们的优劣。需要注意的是并不是对所有算法都可有效地用这种准则来评判比较,这是因为目前还没有可以定量描述先验信息的方法。另外除了种类和数量,如何在分割算法中利用这些信息也很重要,对算法的性能也有很大的影响^[11]。这后一点是更难以定量描述的。

3.1.2 处理策略(Processing strategy)

图象分割和大多数图象分析技术一样,其处理过程可以串行、并行、迭代或混合实现。图象分割算法的性能常和这些处理策略紧密相联。并行工作的算法适合于用具有并行处理能力的计算机来快速实现。但是由于在并行处理中对所有象素是同时处理的,无法利用串行和迭代算法中的中间结果调整算法,所以不够灵活,受噪声的影响也较大^[12]。由此可见,根据算法的处理策略也可在一定程度上把握算法的特性。

3.2 定量分析准则

3.2.1 计算费用(Computation cost)

每个分割算法都由一系列的运算操作来实现。为完成这些操作所需的费用也是衡量算法的一个指标,尤其对快速应用更为重要。实际中完成各种操作的费用与许多因素(例如计算机硬件,图象内容等)

有关。为消除这些因素的影响,计算费用也可按不同处理操作的类型和数量来计算^[13]。在这种情况下,对某些算法的计算费用可由对算法的分析而定量得到。

3.2.2 检测概率比(Detection probability ratio)

它被定义为正确检测概率与错误检测概率之比。这个准则最初用来比较和研究各种边缘检测算子^[14]。给定一个边缘模式,一个算子在检测这类边缘时的正确检测概率 P_c 与错误检测概率 P_f 可以由下两式计算出来(T 为一给定的阈值):

$$P_c = \int_T^{\infty} P(t|\text{edge})dt \quad (1)$$

$$P_f = \int_T^{\infty} P(t|\text{no-edge})dt \quad (2)$$

它们的比值对简单的边缘检测算子可以分析得到。这个值越大则算子在检测这种边缘时的可靠性就越高。许多分割技术利用边缘检测算子以帮助分割图象^[7],所以这种准则可用来对这些算法的性能进行评价。

3.2.3 分辨率(Resolution)

利用不同的分割算法所得到的分割图象的分辨率可以有多种情况,例如可以是象素,若干个象素,或一个象素的若干分之一。实际中大多数算法的分辨率都以象素为单位,但子象素的分辨率在某些应用中是很重要的^[15]。

3.3 定性实验准则

要把一个原始图象分割成若干个区域常需要先进行一定的预处理,在分割后为了把目标干净的提取出来也常需要进行一定的后处理。这些处理可以是自动的或者是交互的。这些处理的类型和数量与所采用的分割算法有密切的联系,在一定程度上反映了算法的性能,所以也可用来衡量和比较分割算法^[10]。一般来说,如果用两个算法分割同一图象并要达到同样的分割效果,则预/后处理要求低的算法其性能可认为较高^[17]。然而对预/后处理进行定量描述是较困难的,特别是对交互式的处理来说^[16],所以该准则一般只能较定性描述算法性能。

3.4 定量实验准则

3.4.1 区域间对比度(Contrast across region)

图象分割要把一个原始图象分成若干个区域。

直观地考虑,这些区域的特性应有比较大的差距,或者说有明显的对比。根据区域之间特性对比度的大小可以判别分割图的质量,也可由此推出所用分割算法的优劣^[18]。对相邻接的两个区域来说,如果它们各自的平均灰度为 f_1 和 f_2 ,则它们之间的灰度对比度(Gray-leved Contrast)GC 可按式计算^[9]:

$$GC = |f_1 - f_2| / (f_1 + f_2) \quad (3)$$

事实上(3)式中的 f 也可代表除灰度外的其它特征量。这样就得到其它量的区域间对比度,当一幅图有多个区域时,可利用(3)式分别计算再求和^[9]。

3.4.2 目标计数一致性(Object count agreement)

对一幅图象进行分割所得到的目标个数 S_n 与图中实际存在的目标个数 T_n 由于分割结果的不完善有可能不同,这个差异在一定程度上反映了分割算法性能的一个方面。利用概率的方法可借助这个差异定义一个目标计数一致性的测度来评价分割算法^[19]。还有一种借助这个差异描述算法性能的指标叫做图象中的分块数(Fragmentation) F ^[20]:

$$F = 1 / \{1 + p|T_n - S_n|^q\} \quad (4)$$

式中, p 和 q 都是尺度参数。

3.4.3 象素距误差(Pixel distance error)

实际中的分割结果常不是完善的,在这种情况下总有一些象素被错分。这些被错分的象素与它们所应该属于的正确区域的距离(带有空间信息)反映了分割的好坏。人们已提出了若干个基于这个距离的测度来评价分割,其中常用的一个是优度函数 FOM^[14]:

$$FOM = \frac{1}{N} \sum_{i=1}^N \{1 / [1 + p \times d^2(i)]\} \quad (5)$$

式中, p 是一个比例系数, $d^2(i)$ 代表第 i 个错分象素与其正确位置的距离^[20]。另一个更早提出的测度叫做归一化的距离(Normalized Distance Measure) NDM^[21]:

$$NDM = 100 \times \sqrt{\sum_{i=1}^N d^2(i) / A} \quad (6)$$

式中, $d^2(i)$ 同上, A 是图象的面积。其它类似的测度还有象素空间分布(pixel spatial distribution)^[19],修正的(modified)FOM^[20]等。

3.4.4 象素数量误差(Pixel number error)

对图象分割结果来说,由于分割错误而产生的错分象素的个数是一个重要的图象质量衡量指标。据此,人们已提出了许多不同的加权方法借助这个误差以评判图象质量。常见的有面积错分率(per-

centage area mis-classified)^[21], 分类误差(classification error)^[22], 正确分割的百分数(fraction of correctly segmented pixels)^[23], 象素分类误差(pixel-class-partition error)^[19]等。最近提出的一种叫误差概率(Probability of Error)PE, 当图象是由目标和背景两部分构成时, PE 可用下式计算^[24]

$$PE = P(O) \times P(B/O) + P(B) \times P(O/B) \quad (7)$$

式中, $P(B/O)$ 是将目标错分为背景的概率, $P(O/B)$ 是将背景错分为目标的概率, $P(O)$ 和 $P(B)$ 分别是图象中目标和背景所占比例的先验概率。当图象中包含多个目标时, 可得到一个更加通用的 PE 定义, 有兴趣的读者可参见文献[25]。

3.4.5 形状测度(Shape measure)

它用来测量目标的外轮廓, 所以也叫轮廓指标^[26]。直观地讲, 一个好的分割结果应满足某些条件, 这个指标就是用来衡量轮廓的光滑程度的。如以 $f_{N(x,y)}$ 表示象素 (x,y) 的邻域 $N(x,y)$ 的平均灰度, $g(x,y)$ 表示象素 (x,y) 处的梯度, 对图象以 T 为阈值进行分割所得形状测度 SM 可用下式计算^[4]:

$$SM = \frac{1}{C} \left\{ \sum_{(x,y)} \text{Sgn}[f(x,y) - f_{N(x,y)}] g(x,y) \text{Sgn}[f(x,y) - T] \right\} \quad (8)$$

式中 C 是一个归一化系数, $\text{Sgn}(\cdot)$ 代表单位阶跃函数。

3.4.6 最终测量精度(Ultimate measurement accuracy)

图象分析中的一个基本问题, 就是要获得对图象目标特征值的精确测量^[27], 这也是图象分析中分割和其它处理的最终目标。因为特征的测量是基于分割结果的, 所以其精度直接取决于分割的质量。反过来, 这个精度(可称为最终测量精度 UMA)也反映了分割图象的质量, 并可以用来评判算法的性能^[8]。实践中, 为了描述目标的不同性质, 可以使用不同的目标特征(feature), 如几何特征, 灰度特征, 纹理特征等, 因此 UMA 可写成 UMA_r 。如用 R_r 代表从作为参考的图象(Reference image)中获得的原始特征量值, 而 S_r 代表从分割后的图象(Segmented image)中获得的实际特征量值, 它们的绝对差(Absolute UMA_r)可由下式算得^[8]:

$$AUMA_r = |R_r - S_r| \quad (9)$$

借助这个准则我们已评价研究了许多分割算法, 取得了预期的效果^[28,29]。

3.4.7 区域内部均匀性(Uniformity within region)

分割常被定义为要把一个原始图象 $f(x,y)$ 分成若干个内部具有相似特性的区域, 所以可用分割图中各区域内部特性均匀的程度来描述图象质量。如以 R_i 表示第 i 个区域, A_i 表示其面积, 则分割图的区域内部均匀性测度(Uniformity Measure)UM 可表示为^[4]:

$$UM = 1 - \frac{1}{C} \sum_i \left\{ \sum_{(x,y) \in R_i} [f(x,y) - \frac{1}{A_i} \sum_{(x,y) \in R_i} f(x,y)]^2 \right\} \quad (10)$$

式中, C 为归一化系数。与此类似的还有繁忙性指标(busyness)^[23]和高阶局部熵(higher order local entropy)^[30]等测度也是基于相同的想法的。

4 评价准则的比较

4.1 分析比较

首先看看各类准则的特点。在评价中利用定量准则(L2 和 L4)一般可获得对所评价的算法比利用定性准则(L1 和 L3)更精确的信息, 虽然一般也复杂一些。借助分析准则(L1 和 L2)进行评价与借助实验准则(L3 和 L4)进行评价相比有两个优点: 首先这时并不需要实际在计算机上实现被研究的算法去分割图象(见图 1), 其次评价结果也不会受评价实验的不同安排所带来的影响。其次再来看看各组准则, 尽管属于同一类时它们也各有其特点。下面我们根据对评价准则的要求, 在下列三个方面对它们进行一些分析比较。

4.1.1 通用性

分割算法的性能是靠评价准则来评判的。分割算法的种类很多, 所以用来评价它们的准则应具有通用性, 即应适用于研究各种不同的分割算法。有些准则由于自身的限制只适合于研究某些算法, 这样就会使评价工作受到限制。例如对形状测度的计算需要用到取阈值算法中的阈值^[4], 所以它只可用于研究取阈值算法。

4.1.2 复杂性

一种评价方法能否实用与它本身的复杂性, 或者说与为了进行评价所需的处理工作量有很大关系。这点对评价方法能否在线应用也非常重要。有些准则对每个象素都要进行一系列的计算(如区域内部均匀性), 由于一幅图中的象素数量很大, 所以评价所需的计算量就会很大; 还有些实验指标, 为了算得它们的数值除了需要对图象进行分割以外, 还需要其它额外的处理和辅助运算(如象素距离误

差), 这样评价就变得相当复杂了。

4.1.3 主客观性

对每个评价方法来说, 其背后常有某些主观或客观的考虑, 或者说所采用的评价准则是根据主观或客观因素而确定的。基于主观的准则常与人的直觉相吻合, 但并不一定反映了实际应用要求^[8]; 而基

于客观的准则则可以提供一致的和无偏的研究结果^[27], 在图象分析中客观的结果更为重要。

根据对各准则以上三方面情况的分析, 我们可综合得到表2, 从表中可看出各组准则的主要优点和缺点。

表2 分割评价准则的比较

Table 2 Comparison of segmentation evaluation criteria

序号	评价准则(组)	所属类	通用性	复杂性	主客观性
(1)	先验知识	L1	部分算法	低	主观
(2)	处理策略	L1	所有算法	低	客观
(3)	检测概率比	L2	部分简单算法	中等	客观
(4)	计算费用	L2	部分简单算法	低	客观
(5)	分辨率	L2	所有算法	低	客观
(6)	预/后处理要求	L3	部分算法	高中不等	主观
(7)	区域内的均匀性	L4	所有算法	高	主观
(8)	区域间的对比度	L4	所有算法	中等	主观
(9)	目标计数一致性	L4	所有算法	低	客观
(10)	象素数量误差	L4	所有算法	中等	客观
(11)	象素距离误差	L4	所有算法	高	客观
(12)	最终测量精度	L4	所有算法	中等	客观
(13)	形状测度	L4	取阈值算法	高	主观

4.2 实验比较

以上分析比较给出了一些定性的结果。本节对其中一些准则进行实验比较, 以获得定量的结果。在选取准则时我们考虑了以下因素: 首先各个直接方法所用的模型都不相同, 不具有可比性, 而定性的方法又不可能精确的比较, 所以应从定量实验(L4)类选择; 其次L4类准则共有7组, 各组通用情况也不同, 其中基于形状参数的准则只适合于研究取阈值算法, 不适用于其它算法, 而且它并不能描述目标外形^[8], 只是描述目标边界的光滑程度。基于目标计数一致性的准则虽可适用于各种算法, 但并不总是适合于各种实验场合。这是因为当分割结果比较接近最优时(现在着重研究的也都是这种情况), 实际分割得到的目标个数与图象中原有的目标个数常是吻合的, 存在的问题只是每个目标并不一定完全分割正确。在这种情况下, 目标计数一致性已不能用来判别分割的优劣了。考虑到以上因素, 我们在剩下的5组准则中各取一个进行了比较, 它们是

(1) 准则A: 区域间对比度, 按(3)式计算。

(2) 准则B: 象素距离误差, 按(6)式计算, 与(5)

式相比其好处是避免了主观选择(5)式中p系数所带来的不确定性。

(3) 准则C: 象素数量误差, 按(7)式计算。

(4) 准则D: 最终测量精度, 选择目标面积作为特征, 按(9)式计算。

(5) 准则E: 区域内均匀性, 可用(10)式计算, 但为了与其它准则相比较, 我们只计算(10)式后面的求和, 这样在分割结果较好时所有的指标值都较小, 反之则都较大。

对不同准则的比较可借助对一系列分割图的评价来进行。具体说, 就是将以上各个准则分别用来评判一系列相同的分割图, 算得各指标的一系列值, 从这些值的变化中可得到各准则不同的特性, 然后通过比较特性来比较准则的优劣。整个试验安排与研究目标特征的评价性能类似^[8], 只是这里用了5个准则, 而不是各种目标特征。表3给出了一组实验结果, 其中列的标号对应分割图象系列的标号, 由5个指标得到的5组值排成5行。表中数值已分别归一化到[0, 1]区间, 以便比较其相对大小。这些值在图2中以曲线的形式给出, 这样可以比较直观地研究它们的变化趋势。

表3 评价准则比较结果

Table 3 Comparison results of different evaluation criteria

	1	2	3	4	5	6	7	8	9	10	11
A	0.997	0.997	0.998	0.999	0.999	0.999	0.999	1.000	0.999	0.998	0.997
B	0.632	0.577	0.504	0.406	0.361	0.331	0.404	0.531	0.648	0.811	1.000
C	0.422	0.353	0.269	0.175	0.138	0.115	0.173	0.297	0.443	0.689	1.000
D	0.421	0.355	0.261	0.161	0.073	0.031	0.134	0.281	0.439	0.690	1.000
E	1.000	0.986	0.957	0.942	0.937	0.932	0.932	0.932	0.942	0.962	0.995

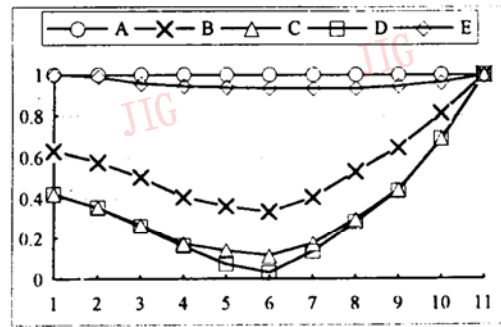


图2 评价准则比较曲线

Fig. 2 Plots of the comparison results of evaluation criteria

现在我们来分析一下表3和图2。一方面我们可看看各曲线的谷的深度。因为最大峰值都已归一化到1,所以曲线的动态范围就由谷的深度值来决定,谷越深则动态范围越大,而动态范围越大,则可表示的不同分割结果则越多越细。这5个准则按动态范围从大到小排列是D,C,B,E,A。最终测量精度最好,最差的是区域间对比度,其曲线上谷和峰的值几乎一样,很难分辨出不同的分割结果。另一方面我们可看看各曲线的形状,在这点上大部分曲线有相类似之处,都是先单调下降然后单调上升,这和分割图象排列的情况一致。注意,D曲线与C曲线在许多地方都重合,但在接近谷底时(即对应接近最优的分割效果时)D曲线比C曲线下降还多还快。这说明利用最终测量精度可以较好地评判接近最优的分割结果。这在实际中是很有意义的。

总的来说这些准则的性能很不一样。最终测量精度具有最大的动态范围,在评判接近最优的分割结果时也具有最好的分辨能力,接下来是象素数量误差,而象素距离误差又次之,最后是区域内均匀性和区域间对比度,它们在使用中虽然不需要参考图^[9],但评价能力很差。

5 结束语

本文对现已提出的各种评价准则进行了分析和

比较,据了解国内外尚缺乏对图象分割评价方法和准则的系统研究,本文仅是一个开头。事实上对图象分割本身还缺乏比较系统的精确的研究^[31],对其评价则更差一些。但是研究图象分割评价正是为了更好地研究图象分割本身。鉴于这个工作的重要性,因此需要更多更深入的研究。在对图象分割的评价研究中,我们以为有如下几个问题值得进一步讨论。

首先,目前有些评价研究,常借助用来分割的准则进行评价,这样评价结果就会出现偏差。如有人提出一种基于二阶局部熵的取阈值算法,然后又用二阶局部熵作为指标将这种算法与其它算法比较^[30]。显然比较结果对这种算法有利。再如区域内均匀性是一个常用的分割准则^[32],如用此作为评价准则(见文献[4],[24],[26]),评价结果就可能不客观。

其次,每种评价方法都是为一定的目的而提出来的^[8],各种不同的评价准则常常只能反映分割算法性能的一个方面,所以可认为它们具有互补性,而不是相斥性。对每个分割算法来说,它的性能是由多个因素决定的,所以也需要借助多个准则来综合评判。虽然这方面已有一些尝试,如[33],[30],[28],但还不够系统。因此,如何把各种准则结合起来全面地进行评价是一个重要的研究课题。

最后,评价的目的是为了能指导、改进和提高分割,如何把评价和分割应用联系起来尚有许多工作要做。一个可能的方法是结合人工智能技术,建立分

割专家系统,以有效的利用评价结果进行归纳推理,从而把对图象的分割由目前比较盲目的试验阶段推进到系统地实现的阶段。

参考文献

- [1] J. S. Weszka, "A Survey of Threshold Selection Techniques", *CVGIP*, 1978, 7, 259-265.
- [2] K. S. Fu & J. K. Mui, "A Survey on Image Segmentation", *Pattern Recognition*, 1981, 13, 3-16.
- [3] R. M. Haralick & L. G. Shapiro, "Image Segmentation Techniques", *CVGIP*, 1985, 29, 100-132.
- [4] P. K. Sahoo et al., "A Survey of Thresholding Techniques", *CVGIP*, 1988, 41, 233-260.
- [5] 吴一全, 朱兆达, "图象处理中阈值选取方法 30 年(1962-1992)的进展(一)", *数据采集与处理*, 1993, 8, 193-201.
- [6] 吴一全, 朱兆达, "图象处理中阈值选取方法 30 年(1962-1992)的进展(二)", *数据采集与处理*, 1993, 8, 268-282.
- [7] N. R. Pal & S. K. Pal, "A Review on Image Segmentation Techniques", *Pattern Recognition*, 1993, 26, 1277-1294.
- [8] Y. J. Zhang & J. J. Gerbrands, "Objective and Quantitative Segmentation Evaluation and Comparison", *Signal Processing*, 1994, 39, 43-54.
- [9] M. D. Levine & A. Nazif, "Dynamic Measurement of Computer Generated Image Segmentations", *IEEE Trans. PAMI*, 1985, 7, 155-164.
- [10] C. E. Liedtke et al., "Segmentation of Microscopic Cell Scenes", *AQCH*, 1987, 9, 197-211.
- [11] Y. J. Zhang & J. J. Gerbrands, "Transition Region Determination Based Thresholding", *Pattern Recognition Letters*, 1991, 12, 13-23.
- [12] J. J. Gerbrands, *Segmentation of Noisy Images*, Doctoral Thesis, Delft University of Technology, The Netherlands, 1988.
- [13] Y. J. Zhang, "Comparison of Segmentation Evaluation Criteria", *Proc 21CSP*, 1993, 870-873.
- [14] I. E. Abdou & W. Pratt, "Quantitative Design and Evaluation of Enhancement/Thresholding Edge Detectors", *Proc. IEEE*, 1979, 67, 753-763.
- [15] T. Sairanyi, "Subpixel Pattern Recognition by Image Histogram", *Pattern Recognition*, 1994, 27, 1079-1092.
- [16] C. N. Graaf et al., "Task-directed Evaluation of Image Segmentation Methods", *Proc. 11ICPR*, 1992, 3, 219-222.
- [17] C. N. Graaf et al., "Validation of the Interleaved Pyramid for the Segmentation of 3D Vector Images", *Pattern Recognition Letters*, 1994, 15, 467-475.
- [18] A. M. Nazif and M. D. Levine, "Low Level Image Segmentation, An Expert System", *IEEE Trans. PAMI*, 1984, 6, 555-577.
- [19] W. A. Yaanoff & J. W. Bacus, "Scene-Segmentation Algorithm Development Using Error Measure", *AQCH*, 1984, 6, 45-58.
- [20] K. Strasters & J. J. Gerbrands, "Three-dimensional Image Segmentation Using a Split, Merge and Group Approach", *Pattern Recognition Letters*, 1991, 12, 307-325.
- [21] W. A. Yaanoff et al., "Error Measures for Scene Segmentation", *Pattern Recognition*, 1977, 9, 217-231.
- [22] J. R. Fram & E. S. Deutsch, "On the Quantitative Evaluation of Edge Detection Schemes and Their comparison with Human Performance", *IEEE Trans. c*, 1975, 24, 616-628.
- [23] J. S. Weszka & A. Rosenfeld, "Threshold Evaluation Techniques", *IEEE Trans. SMC*, 1978, 8, 622-629.
- [24] S. U. Lee et al., "A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation", *CVGIP*, 1990, 52, 171-190.
- [25] Y. W. Lim and S. U. Lee, "On the Color Image Segmentation Algorithms Based on the Thresholding and Fuzzy C-means Techniques", *Pattern Recognition*, 1990, 23, 935-952.
- [26] 李立群, 陈维南, "一种强鲁棒的完全确定型的快速阈值化方法", *模式识别与人工智能*, 1993, 6, 235-241.
- [27] I. T. Young, "Sampling Density and Quantitative Microscopy", *AQCH*, 1988, 10, 269-275.
- [28] Y. J. Zhang, "Segmentation Evaluation and Comparison, A Study of Various Algorithms", *SPIE*, 1993, 2094, 801-812.
- [29] 章毓晋, "一种评价图象分割算法的新方法", *模式识别与人工智能*, 1994, 7, 299-304.
- [30] N. R. Pal and D. Bhandari, "Image Thresholding, Some New Techniques", *Signal Processing*, 1993, 33, 139-158.
- [31] 吴立德, "计算机视觉需要更扎实更艰苦的工作", *模式识别与人工智能*, 1992, 5, 262-265.
- [32] N. Otsu, "A Threshold Selection Method from Gray-level Histogram", *IEEE Trans. SMC*, 1979, 9, 62-66.
- [33] W. A. Yaanoff et al., "Error Measures for Objective Assessment of Segmentation Algorithms", *AQC*, 1979, 1, 107-121.



章毓晋 1989 年 3 月于比利时国立列日(Liège)大学获得应用科学博士学位,其后至 1993 年 6 月在荷兰德尔夫特(Delft)技术大学做博士后及研究工作,1993 年 8 月到清华大学任副教授。他的研究方向主要是图象工程(图象处理,图象分析,图象理解),感兴趣领域还包括计算机视觉,计算机图形学,人工智能,模式识别,多媒体应用与系统等。他已以第一作者身份在国际上发表研究论文 20 余篇,在国内发表研究论文 10 余篇。

A Classification and Comparison of Evaluation Techniques for Image Segmentation

Zhang Yujin

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

(The Institute of Artificial Intelligence & Pattern Recognition, Shantou University, Shantou 515063)

Abstract This paper reviews the research progress on image segmentation evaluation, especially on evaluation methods and evaluation criteria. Most evaluation methods proposed so far can be categorized into direct methods and indirect methods. The direct methods study segmentation algorithms themselves while indirect methods evaluate the performance of algorithms according to the quality of segmented images. Both direct methods and indirect methods need certain evaluation criteria for performing evaluation tasks. This paper makes a classification of different existing criteria and provides an analytical comparison about these criteria. In addition, for most commonly used evaluation criteria, an experimental comparison has been carried out. A rank of these criteria has been obtained through the analysis of their performance and behavior. This study provides a useful basis for correctly applying different evaluation methods in practical applications and indicates several further research directions.

Keywords Computer vision, Image analysis, Image segmentation, Segmentation evaluation, Evaluation criteria, Image quality

中美之间首度实现国际 ISDN 会议电视 北京电信部门支持中国奥运

美国视讯公司(CLI)北京办事处 1996年6月6日讯

北京电信部门今天与美国视讯公司(CLI)签约,引进 CLI 的会议电视系统 eclipse gold, 提供给中国奥委会,用于在奥运会期间提供中美之间的 ISDN 图象通信业务。

北京电信部门通过对多家会议电视厂商的技术、产品、服务进行考察调研,最终以 CLI 在会议电视领域中的领先地位,最佳的图象质量和良好的安装、维护服务以及在 ISDN 网上的丰富经验而决定选用 CLI 的会议电视产品。

此次引进的会议电视系统 eclipse gold 将分别安装于北京的中国奥委会和美国亚特兰大的运动员村。这是中美之间首次实现在 ISDN 网上以 384Kbps 通过会议电视进行异地通讯。奥运会期间,有关领导可以实现对运动员的异地接见;远在万里之外的运动员们可以通过会议电视与亲属们会面;新闻记者也可以由过去的电话采访而改为通过会议电视对运动员进行面对面的异地采访;还可以通过会议电视进行两地联欢、官员会谈、经验交流、信息传递。

这将是会议电视在应用领域的一大突破。

在此之前,CLI 已与北京电信部门有过多次合作。在今年 5 月 8 日,北京市电话号码升八位的仪式上,CLI 的会议电视系统 eclipse 以 384Kbps 成功地运行于北京市 ISDN 网。6 月 5 日,CLI 的 eclipse 还完成了通过 ISDN 网与日本联通的测试。在这几次合作中,CLI 的会议电视设备使用简便,图象清晰,声音逼真,给在场人员留下深刻印象。在此之前,CLI 的 eclipse 已完成在上海 ISDN 网上 128Kbps 和 384Kbps 的运行测试。CLI 在将会议电视应用于 ISDN 方面积累了丰富经验。

此次引进的会议电视系统 eclipse gold 是 CLI 今年推出的最新产品,也是当今世界上最先进的技术。可提供 384Kbps 下 30fps 最佳图象质量,与国际标准完全兼容,并且支持 T. 120 标准,首次可以实现“全面视像会议”,使多个地点的个人或集团会议电视可以实现电子白板书写、文件共享及数据交换等功能,是要求中等带宽下最佳图象质量会议电视用户的理想选择。