

中图法分类号: TP391 文献标志码: A 文章编号: 1006-8961(2011)05-0754-07

论文索引信息: 龚永义, 黄辉, 于继明, 关履泰. 基于熵的两区域图像分割 [J]. 中国图象图形学报, 2011, 16(5): 754-760

基于熵的两区域图像分割

龚永义¹⁾, 黄辉²⁾, 于继明⁴⁾, 关履泰³⁾

¹⁾(广东外语外贸大学信息学院, 广州 510006) ²⁾(广东商学院数计学院, 广州 510275)

³⁾(中山大学计算数学系, 广州 510275) ⁴⁾(浙江工业大学计算机学院, 杭州 310014)

摘要: 图像分割在多媒体, 图像处理, 计算机视觉领域扮演着重要角色。提出了基于图像分割熵的二区域图像分割方法。首先, 根据熵的特性: 单个随机变量所对应的熵越大, 所包含的信息量越大, 图像是单一区域时, 所含的信息量(熵)较小, 引入图像分割熵(ISE)测度, 用于度量两区域图像分割准确程度, 将两区域图像分割问题转化成 ISE 最小值问题。然后, 采用迭代图切(graph cut)算法给出 ISE 最小值问题的近似解, 实现二区域图像分割。实验结果表明, 基于图像分割熵的二区域图像分割方法是可行有效的。

关键词: 图像分割熵; 迭代图切

Entropy based bi-region image segmentation

Gong Yongyi¹⁾, Huang Hui²⁾, Yu Jiming⁴⁾, Guan Lütai³⁾

¹⁾(School of Informatics, Guangdong University of Foreign Studies, Guangzhou 510006 China)

²⁾(Computational Department, Guangdong College of C & B, Guangzhou 510275 China)

³⁾(Computational Department, Sun Yat-sen University, Guangzhou 510275 China)

⁴⁾(College of Computer & Technology, Zhejiang University of Technology, Hangzhou 310014 China)

Abstract: Image segmentation plays an important part in the areas of multimedia, image processing and computer vision. In the paper, the authors propose an image segmentation approach based on an entropy measurer. Specifically, the image segmentation entropy (ISE) is defined to describe the information of an image region. We further prove that the image after segmentation have the minimum ISE, if the image is correctly partitioned. Then, the image segmentation problem is cast into an optimization problem which minimizes ISE. Finally, we use the iterative graph cut approach (IGCA) to solve the optimization problem. The experiments provided in the paper show that our ISE based segmentation approach works well.

Keywords: image segmentation entropy; iterative graph cut approach

0 引言

近 20 年来, 图像分割一直是多媒体, 图像处理和计算机视觉领域的重要研究课题。所谓的图像分割, 就是将给定图像划分成若干有意义的区域, 同一区域内的图像像素拥有类似的图像特性, 不同区域

的像素拥有不同的图像特性。设计合适的图像特性相似性测度是图像分割成功的关键。到目前为止, 研究人员已设计出许多的相似性测度^[1-12]。

在众多相似性测度定义中, 有一部分相似性测度定义基于信息论^[7-9]。这类测度基于如下观点: 同一图像区域是服从某随机变量的样本集, 不同图像区域间的相似性可以通过所对应随机变量的差异

收稿日期: 2010-01-12; 修回日期: 2010-05-07

基金项目: 国家自然科学基金项目(10572154, 60873088); 浙江省自然科学基金(Y1090169)。

第一作者简介: 龚永义(1970—), 男, 2007 年于中山大学获计算机软件与理论专业博士学位, 主要研究方向为图像分割、图像配准、图像与视频放大。E-mail: gongyongyi@hotmail.com。

和信息量来衡量。

事实上,信息论很早就被运用于图像处理研究。例如, Collignon 等人^[13-15]首次在图像配准研究中,采用互信息(mutual information)度量两幅图像间的相似性。两图像间的互信息越大,说明图像相似度越高。Kapur 等人基于图像信号的随机性,提出最大熵图像分割方法^[11];随后 Pal 等人利用 2 维灰度直方图,对最大熵分割方法进行了改进^[12]。Unal 等人^[8]运用信息论测度, Jensen-Shannon 散度,区分同一图像不同区域所对应概率密度函数。Zhang 等人^[16]采用熵间格统计(entropy gap statistic)方法度量图像目标和背景数据熵之间的间格,进行图像分割研究。Kim 等人^[7]则将信息论与曲线演化方法结合起来。通过引入区域标签,结合曲线演化方法,最大化区域标签所对应随机变量与待分割图像所对应随机变量间的互信息,被转化成曲线演化问题。Zhu 等人^[10]基于滤波集提取纹理特性,设计最大熵概率模型进行纹理合成研究。

本文尝试将熵直接用于度量图像分割。根据熵的定义我们知道,单个随机变量所对应的熵越大,该随机变量所包含的信息量越大。图像是单一区域时,所含的信息量(熵)较小。因此我们有理由猜测,将图像作为整体包含的信息量应当多于图像正确分割后每个区域分别包含的信息量的某加权和。假设图像由多个不同区域组成(不同区域对应不同随机变量),经观察比较,确实如此,图像区域对应熵的某种加权和可以用于度量图像的分割准确程度。将这种加权熵称为图像分割熵(ISE)。如图 1,图像(c)由图像(a)和图像(b)组成,图像(c)的 ISE 明显小于它的熵。

依据这思路,引入图像分割熵测度(ISE),将图像分割问题转化成 ISE 最小值问题,并采用基于迭代图切(IGCA)算法进行优化求解。

1 图像分割熵测度

设 $I(x)$ 是定义在域 Ω 上的亮度图像。 $I(x)$ 由两个区域组成:目标区域 R_1 和背景区域 R_2 , 所对应分布分别是 $p_1(z)$ 和 $p_2(z)$, 其中 $R_1 \cup R_2 = \Omega$, $R_1 \cap R_2 = \emptyset, z \in \{0, \dots, L\}$ 。定义同质区域如下:

定义 1 给定区域 R , 所对应分布是 $p(z)$ 。若 R 的不小于给定大小的子区域也对应 $p(z)$, 则我们称区域 R 是同质区域。

如图 1(a)(b), 图 2(a) 是一同质区域。

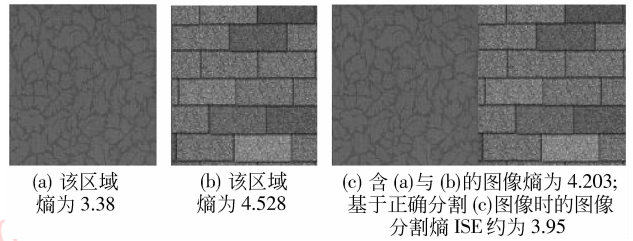


图 1 图像区域的熵

Fig. 1 Entropy of image region

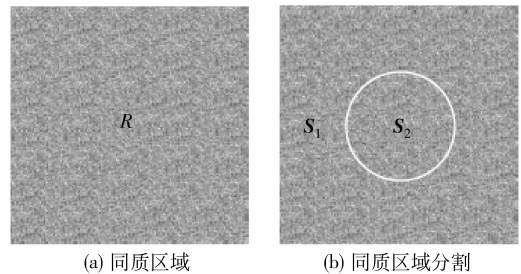


图 2 同质区域分割

Fig. 2 Segmentation of homogeneous

有许多方法可用于估计区域所对应的分布 $p(z)$, 如亮度直方图, parzen 窗口等。对于亮度图像, 采用亮度直方图方法估计概率分布。

定义 2 假设 $S(S_1, S_2)$ 是图像 $I(x)$ 的一个分割, 分割结果由区域 S_1 和 S_2 组成, 所对应的分布分别是 $p_1(z)$ 和 $p_2(z)$, 其中 $S_1 \cup S_2 = \Omega$, $S_1 \cap S_2 = \emptyset$ 。将区域 S_1 和 S_2 所对应熵的加权和称为关于分割 S 的图像分割熵(ISE), 记为 $S_H(S_1, S_2)$,

$$S_H(S_1, S_2) := \sum_{i=1}^2 \frac{|S_i|}{|\Omega|} H(S_i)$$

$$H(S_i) := - \sum_z p_i(z) \log p_i(z)$$

式中 $H(S_i)$ 表示区域 S_i 的熵, $|S_i|$ 表示区域 S_i 所含像素个数/面积。

在所有的分割中, 存在一类特殊的分割: $S(\Omega, \emptyset)$, 一个区域是整幅图像区域 Ω , 另一区域是空集 \emptyset 。对于这类情况, 我们可以得到如下等式:

$$S_H(\Omega, \emptyset) = H(\Omega)$$

上式说明 $S_H(\Omega, \emptyset)$ 和 $H(\Omega)$ 是等效的。方便起见, 将 $S_H(\Omega, \emptyset)$ 简记为 $S_H(\Omega)$ 。

首先考虑由同质区域组成的单区域图像。

定理 1 假设 1) 图像 $I(x)$ 由单一区域组成, 记为 Ω ; 2) 区域 Ω 是同质区域, 所对应分布是 $p(z)$;

3) $S(S_1, S_2)$ 是图像 $I(x)$ 的任一分割, 有如下结论

$$S_H(S_1, S_2) = S_H(\Omega)$$

下面我们进一步考虑由对应不同分布区域组成的两区域图像。

定理 2 假设 1) 图像 $I(x)$ 由两区域组成: 目标区域 R_1 和背景区域 R_2 , 所对应分布分别是 $p_1(z)$ 和 $p_2(z)$, 其中 $R_1 \cup R_2 = \Omega, R_1 \cap R_2 = \emptyset, z \in \{0, \dots, 255\}$; 2) $S(S_1, S_2)$ 是关于图像 $I(x)$ 的一个分割。 S_1 和 S_2 所对应的分布分别是 $p_{S_1}(z)$ 和 $p_{S_2}(z)$, 其中 $S_1 \cup S_2 = \Omega, S_1 \cap S_2 = \emptyset$ 。有如下不等式,

$$S_H(R_1, R_2) \leq S_H(\Omega)$$

$$S_H(R_1, R_2) \leq S_H(S_1, S_2)$$

定理 2 示例如图 3 所示。

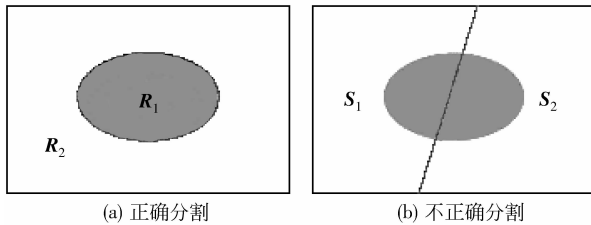


图 3 定理 2 示例

Fig. 3 An example of theory 2

根据定理 2, 图像分割熵 (ISE) 可以用于度量图像分割的准确性: 如果分割 $S(S_1, S_2)$ 是正确分割 $S(R_1, R_2)$, 则 $S_H(S_1, S_2)$ 达到最小值, 否则, $S_H(R_1, R_2) < S_H(S_1, S_2)$, 即 $S_H(S_1, S_2)$ 不是最小, 说明分割是不正确的。

由此, 依据定理 2, 将图像分割问题转换成图像分割熵 (ISE) $S_H(S_1, S_2)$ 的最小化问题。

我们的工作与 Kim 等人^[10]的工作比较接近。Kim 等人用轮廓线 C 将图像 $I(x)$ 分割成两区域 S_1 和 S_2 , 这时 $I(X)$ 有两不确定性, 像素属于区域 S_1 或 S_2 的不确定性, 给定位置像素亮度值的不确定性。同时, 由于区域标签 $L_C(X)$ 包含 $I(X)$ 的不确定性信息 (如可以通过区域标签知道 X 是位于区域 S_1 , 还是 S_2)。据此, Kim 等人进一步指出区域标签 $L_C(X)$ 所决定的 $X \in S_1$ 还是 $X \in S_2$, $I(X)$ 的不确定性越少, 区域标签 $L_C(X)$ 关于 $I(X)$ 的信息就越多。因此他们采用图像像素亮度 $I(X)$ 与区域标签 $L_C(X)$ 间的互信息作为图像分割标准。并将图像分割问题转换成轮廓线最短限制下, 图像像素亮度 $I(X)$ 与区域标签 $L_C(X)$ 间的互信息 $M_I(I(x), L_C(X))$ 最大化问题。

与 Kim 等人工作相比, 我们的方法简单直观的, 依据定理 1、2, 图像分割问题被直观地转化成图像分割熵 (ISE) 最小化问题。

事实上, 可以在 ISE 与 Kim 的互信息测度之间建立如下联系

$$M_I(I(X), L_C(X)) = H(I(X)) - S_H(S_1, S_2)$$

式中 $H(I(X))$ 是关于 $I(X)$ 的熵。

2 基于 ISE 的二区域图像分割

根据定理 2, 二区域图像分割问题可以转换成图像分割熵 (ISE) 最小化问题,

$$E(S^*) = \min_{S(S_1, S_2) \in \gamma} S_H(S_1, S_2) =$$

$$\min_{S(S_1, S_2) \in \gamma} - \sum_{i=1}^2 \sum_{x \in S_i} \log p_i(I(x))$$

式中 γ 表示所有可能的分割组成的集合。遗憾的是, 若不提供适当的先验知识, $E(S^*)$ 的最优化是比较困难的。Kim 等人采用主动轮廓方法处理他们的优化问题。在这里, 借助迭代 Graph Cut 方法 (IGCA) 处理优化问题。Boykov 等人系统地讨论了 Graph cut 技术^[1-5], 并将其成功地用于解决最大流/最小切问题。为更好地体现图像的区域特性, 在 ISE 能量函数中引入正则项

$$E(S^*) = \min_{S(S_1, S_2) \in \gamma} - \sum_{i=1}^2 \sum_{x \in S_i} \log p_i(I(x)) + \sum_{(x,y) \in N_{n-link}} B(x,y)$$

式中 N_{n-link} 是 n -连接边集合。由于是采用迭代算法, IGCA 不需要用户在目标和背景区域中预先指定种子目标区域和种子背景区域, 而只需用户给出一个随机初始分割即可。IGCA 由下列 5 步组成:

1) 随机给定初始图像分割 $S^0(S_1^0, S_2^0)$, 初始化迭代计数器 $i=0$, 设定域值 ε ;

2) 分别估计区域 S_1^i, S_2^i 所对应的分布, 并据此建立权图 G_i ;

3) 运用最大流/最小切算法: Graph Cut, 分割权图 G_i , 获取图像的新分割 $S^{i+1}(S_1^{i+1}, S_2^{i+1})$;

4) 比较分割 $S^i(S_1^i, S_2^i)$ 与 $S^{i+1}(S_1^{i+1}, S_2^{i+1})$ 的差异 $diff(S^i, S^{i+1})$;

5) 若 $diff(S^i, S^{i+1}) < \varepsilon$, IGCA 算法结束, 否则, $i=i+1$, 转至第 2) 步。

IGCA 算法的关键是定义权图 $G = (V, N)$ 。采用与 Boykov 等人^[4]相似的方法定义权图。如图 4, 除两附加端点 s 和 t 外, V 中每个节点 (图 4 中的黑

点)对应图像中相应像素。 V 中两附加端点: s 和 t ,分别是 source 端点和 sink 端点(图4中的空心圆点)。对每个节点 x ,存在两类无向边:与其邻域节点相连的 n -links边,与附加端点 s, t 相连的 t -links边。 t -links边 $\{x, s\}, \{x, t\}$ 的权分别定义为 $-\log p_2(I(x)), -\log p_1(I(x))$, n -links边 $\{x, y\}$ 的权定义为

$$B(x, y) := \lambda e^{-\frac{J(I(x)-I(y))^2+1}{2\sigma^2} \cdot J(x)} \cdot \frac{1}{dist(x, y)}$$

式中 λ 和 σ 是由用户指定的系数, $dist(x, y)$ 是欧氏距离函数, $J(x)$ 是 J -值函数^[6],用于度量图边位于边缘处的权重。若将要分割的图像是随机/纹理图像, $J(x)$ 能有效地用于度量随机/纹理区域, $J(x)$ 大说明说明 x 位于区域边缘, $J(x)$ 小,说明 x 位于区域内部。

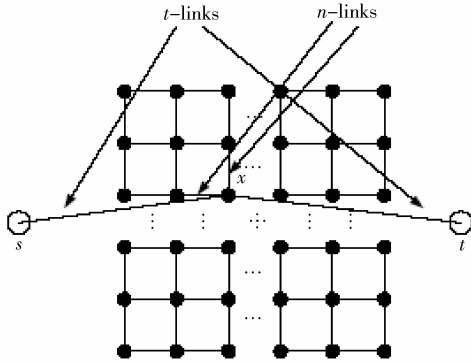


图4 权图G的定义
Fig.4 The overview of the weighted graph G

因此,基于 $J(x)$ 定义的 n -links边的权 $B(x, y)$ 具有与 $J(x)$ 类似的性质:

- 1) 当像素 x, y 位于图像区域内部时, $B(x, y)$ 通常较大,切(cut)经过边 $\{x, y\}$ 的概率较低;
- 2) 当像素 x, y 位于图像区域边缘时, $B(x, y)$ 通常较小,切(cut)经过边 $\{x, y\}$ 的概率较大。

传统的权函数 $B(x, y)$ ^[1]并不包含 $J(x)$,另外,当图像 x 和 y 处像素亮度相差较大时, x 和 y 可能位于区域边缘,根据传统的权函数定义,会认为 x 和 y 位于图像区域边缘,但事实并非一定如此。我们在权函数中引入 $J(x)$ 可以避免上述问题。

下面说明 IGCA 算法的停止条件。在这里,定义标示量 $L^i(x)$,在进行第 i 次迭代时,标记像素属于哪个区域,

$$L^i(x) := \begin{cases} 0 & x \in S_1^i \\ 1 & x \in S_2^i \end{cases}$$

分割 $S^i(S_1^i, S_2^i)$ 与 $S^{i+1}(S_1^{i+1}, S_2^{i+1})$ 之间的差异函数定义为

$$diff(S^i, S^{i+1}) := \sum_x |L^{i+1}(x) - L^i(x)|$$

当 $diff(S^i, S^{i+1}) < \varepsilon$ (ε 是用户指定的域值), IGCA 停止,最终的分割作为最优化问题的近似解。

3 实验与评价

为说明采用迭代 graph cut 方法优化 ISE,进行图像分割的有效性,提供了部分实验数据,实验在 CPU 是 Intel P3 1.9 GHz,内存是 540 M 的 Dell 台式机上进行,操作系统是 Windows XP,编程语言是 VC6.0。

表1列出实验所采用的图像大小,参数取值,所需时间等数据。实验结果如图5所示。

表1中变量 λ, σ 和 γ 是与权函数 $B(x, y)$ 及 J -图相关的参数。在实验中, $\sigma = 1, \varepsilon = 0.001$ 。 λ 和 γ 的取值与所处理图像的特性有关。当目标区域轮廓明显时, λ 可以取较小值,若图像是纹理图像, γ 应当取较大值。

正如表1第6栏所示,IGCA 算法的时间主要耗费在 J -图的计算上,并且 γ 越大,所耗费的时间越多。在我们的算法中,只需进行一次 J -图的计算。与之相比较,分割所需时间是较少的,事实上即使包含此项时间耗费,本文算法在时间上是可以接受的。

表1 采用 IGCA 时的参数取值及算法效率

Tab.1 The parameters and the performance of IGCA when performing on the images

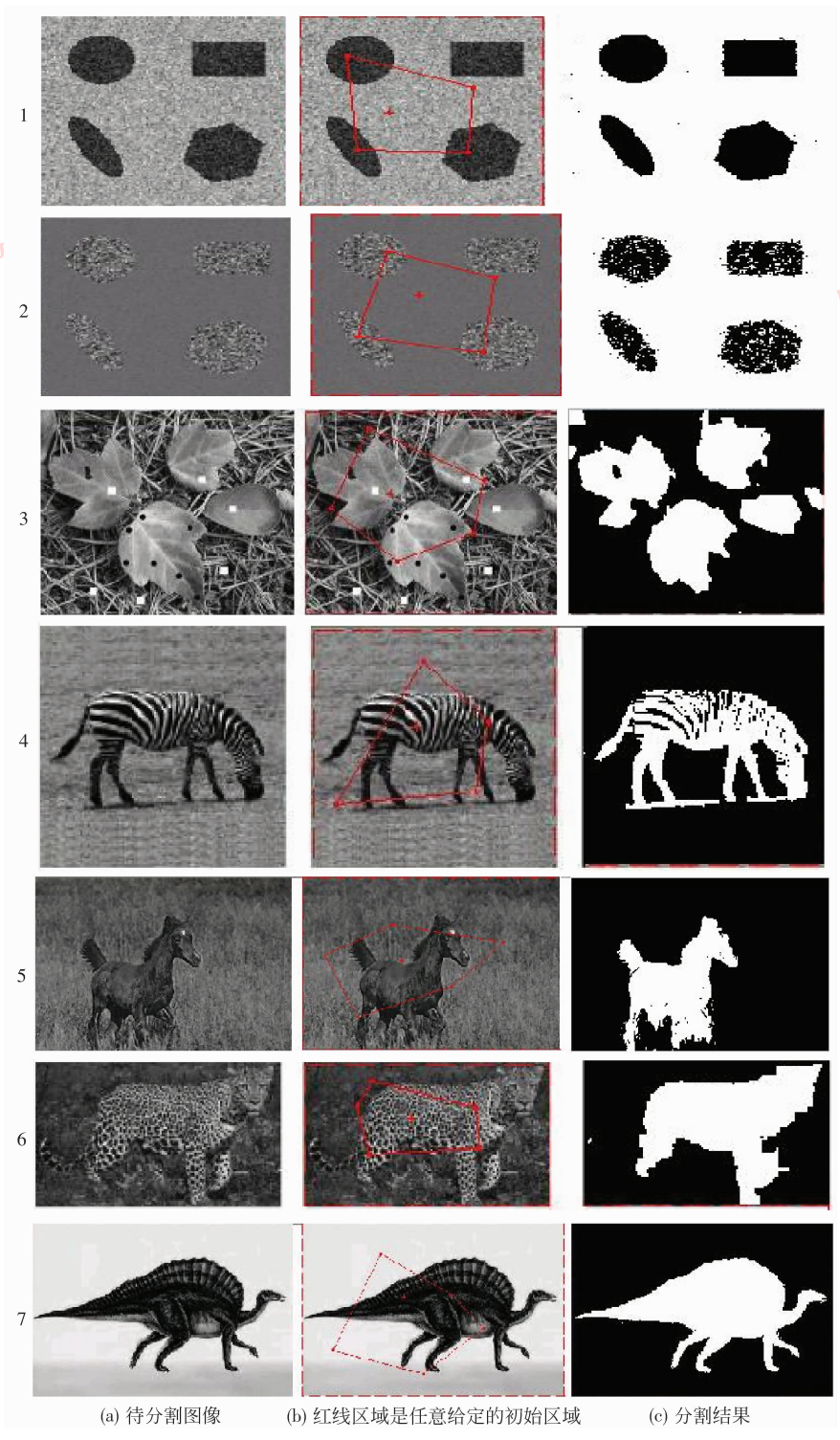
图5	图像大小	λ	σ	γ	计算 J -图 耗费时间/s	迭代 次数	时间/s
1	158 × 152	6	1	3	0.2	5	0.2
2	158 × 152	6	1	3	0.2	5	0.2
3	151 × 188	6	1	10	3	7	4
4	147 × 149	6	1	3	2	10	2
5	256 × 383	12	1	10	11	5	15
6	87 × 147	6	1	6	1	6	1
7	256 × 383	1	1	6	5	5	6

分别采用本文算法和 Zhang 等人的算法对 256 × 256 大小的 Cameraman 图像进行二区域分割如图6所示。实验结果表明两种算法各有不足,本文算法虽能较明确地分割出目标人物,但同时也将建筑物区域等分割为目标,而 Zhang 的方法显然克服

了上述不足,但该算法目标人物分割不够清晰,尤其是人物的双腿,几乎被划归背景。

采用 IGCA 优化 ISE,图像分割效果较好。特别由于 IGCA 由于引入 J -值项,通常即使是较复杂的

图像,也能取得较好分割效果,如图 5 中的第 3、5 实验。但对于包含多目标区域的图像,如图 6,本文算法分割效果亟待进一步改进,这正是我们目前正在努力的目标。



(a) 待分割图像 (b) 红线区域是任意给定的初始区域 (c) 分割结果

图 5 采用 IGCA 优化 ISE 实现图像分割

Fig. 5 The segmented results by IGCA



图6 本文算法与文献[16]算法比较

Fig.6 Comparing the segmented results by IGCA with those by Zhang's algorithm in [16]

4 结论

分析讨论了两区域图像分割准确程度的测度问题,工作体现在两方面:1)引入了图像分割测度:图像分割熵(ISE),将图像分割问题转换成关于图像分割熵的最小化问题;2)采用迭代 graph cut(IGCA)算法,给出图像分割熵的最小化问题的近似解。实验表明我们的方法是有效的。

参考文献(References)

- [1] Boykov Y, Kolmogorov V. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D Images [C]// International Conference on Computer Vision. Washington DC USA: IEEE Computer Society, 2001, 1: 105-112.
- [2] Rother C, Kolmogorov V, Blake A. Grabcut-interactive foreground extraction using iterated graph cuts [J]. ACM Transactions on Graphics (TOG), 2004, 23(3): 309-314.
- [3] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts [J]. IEEE Trans. on PAMI, 2001, 23(11): 1222-1239.
- [4] Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision [J]. IEEE Trans. on PAMI, 2004, 26(9): 1124-1237.
- [5] Kolmogorov V, Zabih R. What energy functions can be minimized via graph cuts? [J]. IEEE Trans. on PAMI, 2004, 26(2): 147-159.
- [6] Deng Y, Manjunath B S. Unsupervised segmentation of color-texture regions in images and Video [J]. IEEE Trans. on PAMI, 2001, 23(8): 800-810.
- [7] Kim J, Fisher J W, Yezzi A, et al. A nonparametric statistical method for image segmentation using information theory and curve evolution [J]. IEEE Trans. on Image Processing, 2005, 14(10): 1486-1502.
- [8] Unal G, Yezzi A, Krim H. Information-theoretic active polygons for unsupervised texture segmentation [J]. International Journal of Computer Vision, 2005, 34(3): 192-220.
- [9] Greig D M, Porteous B T, Seheult A H. Exact maximum a posteriori estimation for binary images [J]. J. R. Statist Soc. B, 1985, 51(2): 271-279.
- [10] Zhu S C, Wu Y, Mumford D. Filters, random fields and maximum entropy (FRAME): towards a unified theory for texture modeling [J]. International Journal of Computer Vision, 1998, 27(2): 107-126.
- [11] Kapur J N, Sahoo P K, Wong A K C. A new method for gray level picture thresholding using the entropy histogram [J]. Computer Vision, Graphics and Image, 1985, 29(3): 273-285.
- [12] Pal N R, Pal S K. Entropy thresholding [J]. Signal Processing, 1989, 16(2): 97-108.
- [13] Pluim J P W, Maintz J B A, Viergever M A. Mutual-information-based registration of medical images: a survey [J]. IEEE Trans. on Medical Imaging, 2003, 22(8): 986-1004.
- [14] Viola P, Wells W M. Alignment by maximization of mutual information [C]// International Conference on Computer Vision. Washington, DC, USA: IEEE Computer Society, 1995: 16-23.
- [15] Collignon A, Maes M, Delaere D, et al. Automated multi-modality image registration based on information theory [C]// Proc. of the Int. Conf. on Information Processing in Medical Imaging. Dordrecht: Kluwer Academic Publishers, 1995: 263-274.
- [16] Zhang Zhengjun, Zhu Yaoqin. Estimating the image segmentation number via the entropy gap statistic [C]// Proceedings of the 2009 Second International Conference on Information and Computing Science (ICIC). Washington, DC, USA: IEEE Computer Society, 2009: 14-16.

附录 A 定理 1 的证明

证明 由于 1) 图像 $I(x)$ 是单一区域 Ω , 且 2) 区域 Ω 是匀质区域, 因此, 区域 S_1 和 S_2 对应相同的分布 $p(z)$ 。有如下等式

$$S_H(S_1, S_2) = \sum_{i=1}^2 \frac{|S_i|}{|\Omega|} H(S_i)$$

从而,

$$\begin{aligned} S_H(S_1, S_2) &= - \sum_{i=1}^2 \frac{|S_i|}{|\Omega|} \sum_z p(z) \log p(z) = \\ &= - \sum_z p(z) \log p(z) = H(\Omega) = \\ &= S_H(\Omega) \end{aligned}$$

由定理 1, 我们知道, 如果图像由匀质区域组成, 则对任意关于该图像的分割, 所对应的图像分割熵 (ISE) 是常数。图 1(b) 是定理 1 的示例: 由分割 $S(S_1, S_2)$ 所得的子区域 S_1 和 S_2 与图像区域 Ω 拥有相同的分布, 则 $S_H(S_1, S_2) = S_H(\Omega)$ 。

附录 B 定理 2 的证明

证明 将所处理图像区域看成是相应随机变量的样本集。假设图像 $I(x)$ 所对应的分布是 $p_\Omega(z)$ 。 $p_\Omega(z)$ 表示在区域 Ω 中亮度值为 z 的概率。 $|\cdot|$ 表示区域内像素个数/区域面积。 $|\cdot|_z$ 表示区域内亮度值为 z 的像素个数/区域面积, 其中 $z \in \{0, \dots, 255\}$ 。有如下等式

$$\begin{aligned} p_1(z) &= \frac{|R_1|_z}{|R_1|}, p_2(z) = \frac{|R_2|_z}{|R_2|} \\ p_{S_1}(z) &= \frac{|S_1|_z}{|S_1|}, p_{S_2}(z) = \frac{|S_2|_z}{|S_2|}, p_\Omega(z) = \frac{|\Omega|_z}{|\Omega|} \end{aligned}$$

由于 $z \in \{0, \dots, 255\}$, $|\Omega| = |R_1| + |R_2| = |S_1| + |S_2|$, 我们有,

$$\begin{aligned} p_\Omega(z) &= \frac{|R_1|_z}{|\Omega|} + \frac{|R_2|_z}{|\Omega|} = \\ &= \frac{|R_1|}{|\Omega|} p_1(z) + \frac{|R_2|}{|\Omega|} p_2(z) \end{aligned}$$

为证明 $S_H(R_1, R_2) \leq S_H(\Omega)$, 我们仅需要证明 $S_H(\Omega) - S_H(R_1, R_2) \geq 0$,

$$\begin{aligned} S_H(\Omega) - S_H(R_1, R_2) &= - \sum_z p_\Omega(z) \log p_\Omega(z) - \\ &= \left[- \frac{|R_1|}{|\Omega|} \sum_z p_1(z) \log p_1(z) - \frac{|R_2|}{|\Omega|} \sum_z p_2(z) \log p_2(z) \right] - \\ &= - \sum_z \left(\frac{|R_1|}{|\Omega|} p_1(z) + \frac{|R_2|}{|\Omega|} p_2(z) \right) \log p_\Omega(z) - \\ &= \left[- \frac{|R_1|}{|\Omega|} \sum_z p_1(z) \log p_1(z) - \frac{|R_2|}{|\Omega|} \sum_z p_2(z) \log p_2(z) \right] = \end{aligned}$$

$$\begin{aligned} &= \frac{|R_1|}{|\Omega|} \sum_z p_1(z) \log \frac{p_1(z)}{p_\Omega(z)} + \frac{|R_2|}{|\Omega|} \sum_z p_2(z) \log \frac{p_2(z)}{p_\Omega(z)} = \\ &= - \frac{|R_1|}{|\Omega|} \sum_z p_1(z) \log \frac{p_\Omega(z)}{p_1(z)} - \frac{|R_2|}{|\Omega|} \sum_z p_2(z) \log \frac{p_\Omega(z)}{p_2(z)} \end{aligned}$$

根据不等式: $\log y \leq y - 1, y > 0$, 且仅当 $y = 1$ 时, 等号成立。可以得到,

$$\sum_z p_1(z) \log \frac{p_\Omega(z)}{p_1(z)} \leq \sum_z p_1(z) \left(\frac{p_\Omega(z)}{p_1(z)} - 1 \right) = 0$$

同理, 有

$$\sum_z p_2(z) \log \frac{p_\Omega(z)}{p_2(z)} \leq 0$$

因此,

$$S_H(\Omega) - S_H(R_1, R_2) \geq 0$$

当分布 $p_1(z), p_2(z)$ 和 $p_\Omega(z)$ 相同时, 等号成立。这时, 图像退化成为匀质区域图像。

下面我们证明 $S_H(R_1, R_2) \leq S_H(S_1, S_2)$ 。假设

$$S_1 = S_{1,R_1} \cup S_{1,R_2}, S_2 = S_{2,R_1} \cup S_{2,R_2}$$

$$R_1 = S_{1,R_1} \cup S_{2,R_1}, R_2 = S_{1,R_2} \cup S_{2,R_2}$$

可以得到下列等式,

$$p_{S_1}(z) = \frac{|S_{1,R_1}|_z}{|S_1|} p_1(z) + \frac{|S_{1,R_2}|_z}{|S_1|} p_2(z)$$

$$p_{S_2}(z) = \frac{|S_{2,R_1}|_z}{|S_2|} p_1(z) + \frac{|S_{2,R_2}|_z}{|S_2|} p_2(z)$$

$$p_{S_2}(z) = \frac{|S_2|_z}{|S_2|} p_\Omega(z) = \frac{|\Omega|_z}{|\Omega|}$$

根据不等式 $S_H(R_1, R_2) \leq S_H(\Omega)$, 有

$$S_H(S_{1,R_1}, S_{1,R_2}) \leq S_H(S_1)$$

$$S_H(S_{2,R_1}, S_{2,R_2}) \leq S_H(S_2)$$

$$S_H(R_1, R_2) \leq S_H(\Omega)$$

从而, 得到

$$\begin{aligned} S_H(R_1, R_2) - S_H(S_1, S_2) &= \\ &= \frac{|R_1|}{|\Omega|} H(R_1) + \frac{|R_2|}{|\Omega|} H(R_2) - \frac{|S_1|}{|\Omega|} H(S_1) - \frac{|S_2|}{|\Omega|} H(S_2) \leq \\ &= \frac{|R_1|}{|\Omega|} H(R_1) + \frac{|R_2|}{|\Omega|} H(R_2) - \\ &= \frac{|S_1|}{|\Omega|} S_H(S_{1,R_1}, S_{1,R_2}) - \frac{|S_2|}{|\Omega|} S_H(S_{2,R_1}, S_{2,R_2}) = \\ &= \frac{|R_1|}{|\Omega|} H(R_1) + \frac{|R_2|}{|\Omega|} H(R_2) - \\ &= \frac{|S_1|}{|\Omega|} \left[\frac{|S_{1,R_1}|}{|S_1|} H(S_{1,R_1}) + \frac{|S_{1,R_2}|}{|S_1|} H(S_{1,R_2}) \right] - \\ &= \frac{|S_2|}{|\Omega|} \left[\frac{|S_{2,R_1}|}{|S_2|} H(S_{2,R_1}) + \frac{|S_{2,R_2}|}{|S_2|} H(S_{2,R_2}) \right] = 0 \end{aligned}$$

并且仅当分割 $S(S_1, S_2)$ 是关于图像的正确分割时, 等号成立,

$$S_H(R_1, R_2) = S_H(S_1, S_2)$$