

# 图像检索中的两层描述和非对称区域匹配

戴声扬, 章毓晋

(清华大学电子工程系, 北京 100084)

**摘 要:** 在基于内容的图像检索中, 需要描述图像中的空间信息从而克服仅基于全局特征的传统方法的局限. 图像分割是得到图像空间描述信息的重要手段, 但对于所有图像利用同一种分割结果的检索算法(单层描述方法)会受到图像分割算法精确度的影响而使性能受到限制. 本文提出了基于图像的两层描述(包括粗略描述和精细描述)和非对称区域匹配的算法以减少不精确分割带来的不利影响. 利用从 7000 张通用图片库中随机选取的 700 幅查询图像而进行的统计实验结果表明此算法可以有效的提高检索效果.

**关键词:** 基于内容的图像检索; 两层描述; 非对称区域匹配

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112 (2005) 04-0725-05

## Two-Level Description and Unbalanced Region Matching in Image Retrieval

DAI Sheng-yang, ZHANG Yu-jin

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

**Abstract:** Research on integrating spatial information into content-based image retrieval is aimed at solving the problem caused by global feature based algorithm. Most systems derive the spatial information from image segmentation. However, the description of images based on one-level description (OLD) and the inevitable inaccuracy of segmentation results limit the performance. The proposed two-level description (TLD) describes images by a rough description and a detailed description to avoid improper spatial constrain caused by OLD. Similarity measurement based on unbalanced region matching (URM) is introduced in taking the advantage of TLD to reduce the influence of inaccurate segmentation. The performance of the system is illustrated by experimental results with 700 query images randomly selected from a database of 7000 general-purpose images.

**Key words:** content-based image retrieval; two-level description; unbalanced region matching

## 1 引言

近年来, 基于内容的图像检索成为一个研究热点, 其中如何描述空间信息得到广泛重视. 图像分割可以在目标级别上帮助获取图像的空间信息, 所以基于区域的检索算法都利用了分割的结果. 例如 Blobworld 系统通过用户交互来得到空间信息, 用户需要指定感兴趣的分割区域作为检索目标<sup>[1]</sup>. 但如果检索图像不包含明显物体(如风景图片)或者分割结果不够理想使用户无法选择需要的目标, 检索将难于进行.

通用图像的准确分割非常困难, 如何克服分割算法的不准确性对检索系统性能的不利影响值得研究. 例如 Wang 等提出 Integrated Region Matching (IRM) 的算法计算由若干区域组成的图像的相似度<sup>[2]</sup>. 这一算法通过在两幅图像的各区域之间建立多对多的映射来计算图像之间的相似度. 但它属于贪婪算法, 而且没有将空间布局信息考虑在内.

本文工作的目的是为了减轻分割算法对检索系统带来的不利影响, 并在检索系统中合理引入空间信息. 主要特点如下:

- (1) 对每幅图像采用两层的描述方法;
- (2) 用非对称区域匹配进行相似度度量.

## 2 图像的两层描述

### 2.1 单层描述存在的问题

空间信息是描述图像内容的一个重要特征, 但过多的空间信息反而会引入过多的约束, 使得图像之间难于进行匹配. 如图 1(a), 两幅图像都包含一个椭圆形状的目标和一个背景区域, 其中目标和背景都分别包含相同的颜色成分, 但颜色的位置分布有所不同. 如果过分割, 区域之间由于颜色成分位置的不同, 会得到过低的相似度. 图 1(b) 用以说明欠分割的结果引起的问题, 这时如果在第二幅图像中区域 A 和 B 没有被分割开, 两幅图像会由于区域形状的不同得到过低的相似度.

上例说明, 匹配时如仅利用同一种分割结果, 即采用单层描述进行检索会受到分割结果的严重影响. 这时即使采用 IRM 方法也解决不了问题<sup>[2]</sup>. 图 1(c) 描述了 IRM 的基本思想, 两幅图像间的相似度是区域对 (AW, BW, AX, AY, DV, CV

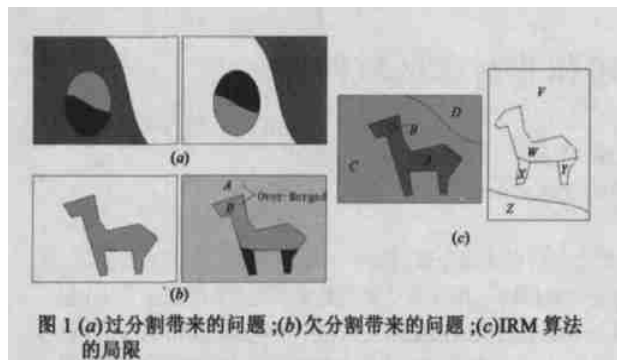


图 1 (a) 过分割带来的问题 ;(b) 欠分割带来的问题 ;(c) IRM 算法的局限

和  $CZ$  之间相似度的一个加权和. 然而, 正如 Wang 所指出, 一个合理的相似度量应该是匹配区域融合后比较的结果, 而不是区域之间直接进行比较<sup>[2]</sup>. 例如, 应该用  $A$  和  $B$  组合而成的区域和  $W, X$  和  $Y$  组合成的区域进行比较.

## 2.2 两层描述方案

本文提出两层描述的方案用以减轻上述过分割或者欠分割带来的不利影响. 具体就是对查询图像采用较为粗略的分割结果进行描述, 而对其他候选图像采用相对精细的分割结果进行描述. 查询图像和候选图像之间的相似度量并不是直接对这样的区域对进行, 而是将查询图像中的每一个粗略分割的区域和所有与之匹配的候选图像的精细区域组合而成的区域进行比较.

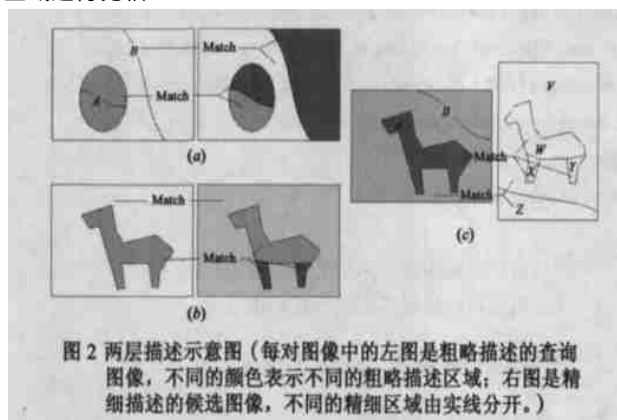


图 2 两层描述示意图 (每对图像中的左图是粗略描述的查询图像, 不同的颜色表示不同的粗略描述区域; 右图是精细描述的候选图像, 不同的精细区域由实线分开.)

对查询图像的粗略描述可以避免引入过多的空间限制. 如图 2(a), 左边图像采用了粗略描述后, 和右边图像中的区域可以得到很好的匹配. 候选图像的精细描述可以避免区域欠分割带来的不利影响. 再如图 2(b), 右边图像采用了精细描述后, 和左边图像中的区域可以得到很好的匹配. 图 2(c) 说明了两层描述如何避免 IRM 算法的局限. 左边图像中的区域  $A$  和右边图像中的区域  $W, X$  和  $Y$  的组合进行比较, 左边图像中的区域  $B$  和右边图像中的区域  $V$  和  $Z$  的组合进行比较.

图像的精细描述基于精细分割结果<sup>[5]</sup>, 而粗略描述所需的分割结果通过对精细描述区域进行合并而得到 (一些示例见后). 每一个粗略描述的区域都是由若干精细描述的区域组成的, 所以不会影响最终结果. 在精细描述中, 每个区域的颜色均值被存储下来作为特征. 在粗略描述中, 每个组成此粗略描述区域的精细区域的颜色均值和此精细区域所占权重被存

储下来作为特征. 在精细描述和粗略描述中均使用了空间描述. 空间描述的算子采用矩阵描述的方法, 它将整个图像平均分成  $M \times N$  个小块, 矩阵的每个元素  $M(i, j)$  取值为对应小块中被描述区域所占的面积比例. 空间描述算子之间的距离定义为

$$S_{shape}(M_1, M_2) = \frac{\min(M_1(i, j), M_2(i, j))}{M \times N} \quad (1)$$

两层描述的数学表达式可定义如下:

(1) 精细描述 ( $D^d$ ):

$$D^d = \{R_i^d \mid i = 1, 2, \dots, m\} \quad (2)$$

其中  $R_i^d = (C_i^d, M_i^d, w_i^d)$  是第  $i$  个精细区域的特征矢量;  $C_i^d = (L_i^d, a_i^d, b_i^d)$  是此区域的颜色均值;  $M_i^d$  是此区域的空间描述;  $w_i^d$  是此区域占整个图像面积的比例.

(2) 粗略描述 ( $D^r$ ):

$$D^r = \{R_i^r \mid i = 1, 2, \dots, n\} \quad (3)$$

其中  $R_i^r = (\{C_{i,t}^r, C_{i,2}^r, \dots, C_{i,n_t}^r\}, M_i^r, w_i^r)$  是第  $i$  个粗略区域的特征矢量,  $C_{i,t}^r = (L_{i,t}^r, a_{i,t}^r, b_{i,t}^r, w_{i,t}^r)$  用于描述第  $i$  个粗略区域中的第  $t$  个精细描述区域 ( $t = 1, 2, \dots, n_t$ );  $M_i^r$  是此区域的空间描述;  $w_i^r$  是此区域占整个图像面积的比例.

## 3 非对称区域匹配

### 3.1 对称相似度量问题

许多研究者把图像的相似度量问题归结为如下被称为运输问题的线性规划问题<sup>[2-4]</sup>.

用两个带权重的集合  $A$  和  $B$  描述两幅图像  $A = \{r_1, w_1; r_2, w_2; \dots; r_m, w_m\}$  和  $B = \{r_1, w_1; r_2, w_2; \dots; r_n, w_n\}$ , 其中  $r_i$  和  $r_j$  用来描述图像中一个区域 (或者颜色成分) 的特征矢量,  $w_i$  和  $w_j$  是对应矢量的权重. 定义  $S_{i,j}$  为矢量  $r_i$  和  $r_j$  的相似度 ( $1 \leq i \leq m, 1 \leq j \leq n$ ), 进一步定义集合  $A$  和  $B$  的相似度如下:

$$S(A, B) = \max_{\{w_{i,j} \mid 1 \leq i \leq m, 1 \leq j \leq n\}} \sum_{i,j} w_{i,j} S_{i,j} \quad (4)$$

且满足如下的限制条件:

$$w_{i,j} \geq 0, \quad \sum_{j=1}^n w_{i,j} = w_i, \quad \sum_{i=1}^m w_{i,j} = w_j, \quad 1 \leq i \leq m, 1 \leq j \leq n \quad (5)$$

Rubner 等在其中加入了一个权重项后定义了旅行者距离 (Earth Mover's Distance, EMD) 以避免两个集合权重之和不同引起的问题<sup>[3]</sup>. 求极值利用了 Transportation simplex 方法, 但这一算法的计算复杂度较高.

Mojšilovic 等定义了最优颜色成分距离 (Optimal Color Composition Distance, OCCD)<sup>[4]</sup>, 它与式 (4) 本质上是相似的. 运输问题通过对权重的量化被转化为最小代价图匹配问题. 它的运算的复杂度为  $O(n^3)$ , 其中  $n$  是权重总和除以量化步长.

Wang 等利用迭代的 IRM 算法得到上述问题的次优解<sup>[2]</sup>. 在每一步运算中, 选出相似值矩阵中剩余元素的最大者, 赋予其最大可能的权重, 同时去掉对应的此元素所在的行或列. 经过最多  $m + n$  步运算之后, 可以得到最后的权重分配结果. 这一算法的复杂度很低, 但属于贪婪算法, 可能会陷入局部极

值,仅得到次优解。

上述几种方法均假设匹配图像的地位是平等的,在区域之间采用多对多的映射来计算相似度,所以性能受到限制。另外在空间信息的引入和特征矢量的描述能力之间也存在矛盾。如果  $m$  和  $n$  相对较大,将会引入过多的空间约束,这会降低系统的性能。另一方面,如果  $m$  和  $n$  相对较小,区域描述的表达力将会受到限制,无法充分描述当前的区域。比如有些方法中每个区域只用一种颜色描述,这种描述方法对相对较大的区域是不够的。

### 3.2 从精细区域到粗略区域的映射

将图像间的相似度问题归结为运输问题会导致在不同图像的区域之间寻找一个多对多的映射。本文采用两层描述,查询图像采用粗略描述,候选图像采用精细描述。两层描述打破了查询图像和候选图像之间的平等关系,允许在图像的区域之间建立多对一的映射,从而避免解决复杂的运输问题。所以这里采用两层描述反而比采用单层描述的计算复杂度要小。

在多对一的映射中,需将  $D^d$  中的每一个区域都唯一映射到  $D^r$  中的一个区域上。映射函数  $F: \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, m\}$  中综合了颜色和空间信息,

$$F(i) = \arg \max_j S(R_i^d, R_j^r) \quad (6)$$

其中  $i \in \{1, 2, \dots, n\}$  是精细描述区域的序号,  $j \in \{1, 2, \dots, m\}$  是粗略描述区域的序号,

$$S(R_i^d, R_j^r) = S_{\text{space}}(M_i^d, M_j^r) \times \max_{1 \leq t \leq n_j} \{S_{\text{color}}(C_i^d, C_{j,t}^r)\} \quad (7)$$

$$S_{\text{color}}(C_1, C_2) = \begin{cases} 1 - D_{\text{color}}(C_1, C_2) / T & \text{if } D_{\text{color}}(C_1, C_2) < T \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

式(7)和(8)中  $D_{\text{color}}(C_1, C_2)$  是颜色  $C_1$  和  $C_2$  在  $\text{CIEL}^* a^* b^*$  颜色空间中带权重的欧式距离。 $T$  是一个常数(实验中,取  $T=30.0$ )。 $S(R_i^d, R_j^r)$  考虑了颜色和空间信息的相似性,因为每个精细区域只用一种颜色来描述,所以  $C_i^d$  和粗略区域中与之最接近的颜色  $C_{j,t}^r$  之间的相似度可以反映把  $R_i^d$  映射到  $R_j^r$  上的合理程度。

对于一个特定的粗略区域  $R_i^d$ ,假设  $D^d$  中的  $l$  个区域  $R_{i1}^d, R_{i2}^d, \dots, R_{il}^d$  映射到它,  $w_1, w_2, \dots, w_l$  是对应区域的权重,定义为  $w_j = S_{\text{space}}(M_i^r, M_{ij}^r)$ 。所有映射到这一粗略区域的精细区域将可以构成一个带权重的颜色集合  $\text{MAP}(R_i^r) = \{C_{i1}^d, w_1; C_{i2}^d, w_2; \dots; C_{il}^d, w_l\}$ 。

### 3.3 图像相似度度量

先定义区域的重要性权重  $W(R)$  如下

$$W(R) = [1 - E(R)] \times \min\left\{\frac{T}{w_R}, 1\right\} \quad (9)$$

其中  $E(R)$  定义如下

$$E(R) = \frac{\max_{i \in R} \left\{ \frac{|2 \times i - \text{width}|}{\text{width}}, \frac{|2 \times i - \text{height}|}{\text{height}} \right\}}{N(R)} \quad (10)$$

用以度量区域  $R$  与图像边界的距离。如果区域  $R$  中的像素大

多在靠近图像边界的位置,  $E(R) \approx 1$ ; 如果区域  $R$  中的像素大多在靠近图像中心的位置,  $E(R) \approx 0$ 。一般说来,图像的中心部分会受到更多的注意,所以这可以在一定程度上反映区域的重要程度。式(9)中右边的第二部分用来减小较大区域带来的影响(特别是单一颜色的背景区域),  $w_R$  是 3.2 中定义的区域权重,本文实验中取定  $T=0.4$ ,  $N(R)$  是区域  $R$  中的像素个数。这样,  $W(R)$  可以根据区域大小和位置来度量它的重要程度。

查询图像的粗略描述和候选图像的精细描述之间的相似度定义如下:

$$S_{\text{image}}(D^r, D^d) = \sum_{R^r \in D^r} [S_{\text{colorset}}(R^r, \text{MAP}(R^r)) \times W(R^r)] \quad (11)$$

其中  $S_{\text{colorset}}(A, B)$  用来度量两个带权重的颜色集合  $A$  和  $B$  之间的相似度。这里采用了类似 IRM 的贪婪算法来计算这一相似度以保持较低的运算复杂度。每个粗略描述的区域  $R^r$  和所有映射到它之上的精细描述区域进行比较,  $S_{\text{colorset}}(R^r, \text{MAP}(R^r))$  表示这一比较结果,图像之间的相似度定义为查询图像的每个粗略描述区域对相似度的贡献之和。加入  $W(R^r)$  以结合区域的重要性信息。

## 4 实验结果和讨论

### 4.1 数据库和对比算法

选取了 Corel 图像库中 70 类图像,每类包括 100 幅图像,共 7000 幅图像。总体说来,每个类中的图像具有相似的语义含义和视觉特性。这一分类结果被作为统计检索结果的参考标准(ground truth)。实验中随机选取了 700 幅图像(每类 10 幅)用作查询来计算统计结果。待检索的图像库包含了全部的 7000 幅图像。

将本文提出的算法与另外四种算法进行了比较,它们是:

- (1) 基于全局颜色直方图<sup>[6]</sup>;
- (2) 基于单层精细描述的 IRM(IRM+OLD(D));
- (3) 基于单层粗略描述的 IRM(IRM+OLD(R));
- (4) 基于单层精细描述的非对称区域匹配(URM+OLD(D))。

在实验中,算法(1)也是在  $\text{CIEL}^* a^* b^*$  颜色空间中实现的,三个颜色分量分别被量化成 5, 10, 10 个区间,用二次距离度量的方法来计算相似度<sup>[7]</sup>。算法(2)和(3)均用下式计算区域之间的相似度

$$S(R_1, R_2) = \frac{S_{\text{shape}}(M_1, M_2) \times S_{\text{colorset}}(R_1, R_2)}{w_{R_1} \times w_{R_2}} \quad (12)$$

而图像间相似度通过 IRM 算法得到。算法(4)和本文算法的唯一不同是算法(4)中的查询图像使用了精细描述,而本文算法的查询图像使用了粗略描述。

### 4.2 基于统计的性能比较

本文使用了两种指标(平均查准率和查准率-查全率曲线)进行了统计性能比较。查准率  $P(n)$  定义为前  $n$  个返回结果中正确结果所占的比例。不同算法返回的前  $n$  ( $n=10, 20, 50, 100$ ) 个结果的平均查准率  $\bar{P}(n)$  列在表 1 中。查准率-查全

率曲线是最常用的衡量 CBIR 系统性能的方法之一。不同算法的查准率-查全率曲线见图 3, 其中横轴表示查全率(返回正确结果占正确结果总数的比例), 纵轴表示平均查准率。

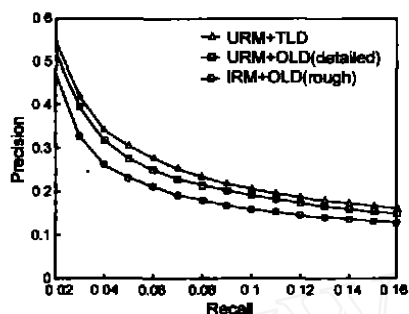


图 3 查准率-查全率曲线比较(算法(1)和算法(2)的曲线与算法(3)的曲线非常接近, 没有显示)

表 1 前  $n$  ( $n=10, 20, 50, 100$ ) 个返回结果的平均查准率

$\bar{P}(n)$ (%)	$n=10$	$n=20$	$n=50$	$n=100$
本文算法	32.90	25.04	17.50	12.99
算法(1)	27.29	20.49	14.37	11.03
算法(2)	27.11	20.39	14.52	11.04
算法(3)	27.93	20.86	14.87	11.12
算法(4)	30.81	23.36	16.65	12.69

从表 1 和图 3 中可以看到, 从统计意义上说, 非对称区域匹配可以有效提高检索性能, 基于两层描述的非对称区域匹配的算法比所有对比算法性能均要好, 与单层描述的 IRM 方法比较, 性能提高约 20%。

从表 1 和图 3 中可以推断出, 平均来说, 前 20 个返回结果中有大约 5 个 ( $20 \times 25.04\%$ ) 是正确结果。如果需要看到 8 个正确的返回结果, 平均需要察看前 35 个 ( $8/0.23$ ) 返回结果。由于每个类中的图像可能具有相似的语义特征但完全不同的底层特征, 而且不同类别的图像也可能具有类似的语义特征, 所以上述的统计结果只是对算法性能的一个保守估计。另外, 考虑到是基于较大的通用图像库的平均性能, 这样的结果仍然是令人鼓舞的。

#### 4.3 不同算法对同一幅查询图像的具体结果比较

图 4 中显示了本文算法和算法(1)检索结果的比较, 本文



图 4 (a) 本文算法的检索结果 ;(b) 算法(1)的检索结果

算法利用了空间信息, 特别是不同位置区域的不同重要性信息, 检索效果比较满意。算法(1)受大面积绿色背景区域的影响, 而且没有引入空间信息, 所以检索效果较差。

下面结合图 5 给出本文算法和其他算法对几幅检索返回结果图像的具体讨论。



图 5 几幅检索结果图像(每组图像中第一张是原图, 第二, 三张分别是精细分割和粗略分割的结果)。

图 5(a) 中的图像是用于查询的图像。本文算法对它进行了两层描述, 在保证区域描述能力的基础上没有引入过多的空间约束, 从而获得了满意的检索结果。

图 5(b) 中的图像在本文算法的返回结果中排在第 10 位, 但在算法(2)的返回结果中排在 100 位之后, 这是由于候选图像的精细描述包含了一个大背景区域, 它和查询图像的精细描述中的背景区域无法得到很好的匹配。

图 5(c) 中的图像在本文算法的返回结果中排在第 3 位, 但在算法(3)的返回结果中排在 100 位之后, 这是由于此图在粗略描述中背景包括了两个区域(暗绿区域和浅绿区域), 所以它和查询图像的背景区域由于形状的不同而只得到了较小的相似度值, 这反映了 IRM 方法的局限。

图 5(d) 中的图像在本文算法的返回结果中排在第 9 位, 但在算法(4)的返回结果中排在 50 位之后, 这是由于此图的精细描述中有一个单色的大区域, 这一区域只能映射到查询图像中的一个精细区域上, 所以降低了它们之间的相似度值。这说明确对查询图像采用粗略描述的有效性。

## 5 结论

本文提出了一种基于两层描述的图像检索算法。检索图像采用粗略描述, 而候选图像采用精细描述, 这样可以避免单层描述带来的空间约束不理想的影响。本文还提出了基于非对称区域匹配的图像相似度度量算法, 以减轻分割不精确带来的不利影响。大数据量的实验说明本文提出的算法可以提高对通用图像的检索效率。

## 参考文献:

- [ 1 ] Carson C, Belongie S, Greenspan H. Blobworld: Image segmentation using expectation-maximization and its application to image querying[J]. IEEE Trans PAMI, 2002, 24(8): 1026 - 1038.
- [ 2 ] Wang J Z, Wiederhold G, Fischel O, et al. Content-based image indexing and searching using Daubechies' wavelets[J]. Int. Journal Digital Libraries, 1998, 1(4): 311 - 328.
- [ 3 ] Rubner Y, Tomasi C, Guibas L. Metric for distributions with applications to image databases[A]. Proc Int Conf on Computer Vision[C]. Bombay, India: IEEE, 1998. 59 - 66.
- [ 4 ] Mjasilovic A, Hu J, Soljanin E. Extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis[J]. IEEE Trans on Image Processing, 2002, 11(11): 1238 - 1248.
- [ 5 ] Dai S Y, Zhang Y J. Color image segmentation with watershed on color histogram and Markov random fields[A]. Proc ICICS-PCM[C/CD]. 2003, Singapore: IEEE, p0240.
- [ 6 ] Swain M, Ballard D. Color indexing[J]. Int Journal on Computer Vision, 1991, 7(1): 11 - 32.
- [ 7 ] Hafner J, Sawhney H, Equitz W. Efficient color histogram indexing for quadratic form distance function[J]. IEEE Trans PAMI, 1995, 17(7): 729 - 736.

## 作者简介:

**戴声扬** 男, 1978 年出生于山东, 2001 年和 2004 年分别于清华大学电子工程系获学士和硕士学位, 现在美加州大学洛杉矶分校计算机系攻读博士学位, 研究方向主要是图像处理, 基于内容的图像检索等, 已在国内外发表了 10 篇研究论文。



**章毓晋** 男, 1954 年出生于山西, 现为中国图像图形学学会副理事长, 学术委员会主任, IEEE 高级会员, 河南省特聘教授; 《中国图像图形学报》副主编, 《电子与信息学报》, 《计算机辅助设计与图形学学报》, 《Pattern Recognition Letters》, 《International Journal of Image and Graphics》编委. 主要研究领域为图像工程(图像处理、图像分析、图像理解及其技术应用)及相关教学手段和方法, 已在国内外发表了 200 多篇研究论文, 出版了《图像分割》, 《基于内容的视觉信息检索》等书 8 本。