基于多级描述模型的渐进式图像内容理解

高永英, 章毓晋
(清华大学电子工程系, 北京 100084)

摘 要: 针对目前基于内容的图像检索技术中低级特征无法准确全面地描述高级语义的问题, 本文提出了一种基于多级图像描述模型的渐进式图像内容理解. 该图像描述模型在不同层次上对图像内容进行分析和提取, 实现了图像内容的全方位描述, 从底层向高层的过渡是渐进式的图像理解过程. 特别是从视觉感知层到目标层, 体现了图像低级特征与高级语义之间的过渡. 本文给出了一种基于先验知识的上下文驱动的目标理解算法, 实现了图像语义的提取. 作为一个应用实例, 本文给出了以上方法在基于内容的图像检索技术中的具体应用.

关键词: 图像描述模型; 语义; 目标理解; 图像检索
中图分类号: TN919.85 文献标识码: A 文章编号: 0372-2112 (2001) 10-1376-05

Progressive Image Content Understanding Based on Multi-Level Image Description Model

GAO Yong-ying, ZHANG Yu-jin
(*Department of Electronic Engineering, Tsinghua University, Beijing* 100084, *China*)

Abstract: In this paper, an new method for progressive image content understanding based on multi-level image description model is proposed, aiming at overcoming the considerable gap between low-level image features and high-level image semantics in the field of image retrieval. In the proposed image content description model, image contents are analyzed and extracted in different levels, reaching at omnidirectional image content description. In addition, the transition from low level to high level is exactly a progressive image understanding. In this paper, a new algorithm for object understanding is proposed, which is based on pre-knowledge and is context-driven, in order to extract image semantics. As a practical instance, discussion about combining the proposed method into content-based image retrieval is also given.

Key words: image description model; semantics; object understanding; image retrieval

1 引言

基于内容的图像检索已成为当前图像检索技术的主流, 一些著名的图像检索系统, 如 QBIC[1], Photobook[2], VisualSEEK[3] 和 CAFIIR[4] 等均采用了基于内容的检索技术. 然而, 现有的图像检索系统对图像内容的描述大多直接采用了传统的低级图像特征, 如颜色, 纹理, 形状等. 由于这些特征与人对图像的内容理解之间存在相当大的差异, 因此很多情况下以图像低级特征为检索依据的检索结果不尽人意. 如何将图像语义特征结合到检索中是提高检索系统性能的关键所在, 已得到了越来越多的关注. 其中通过在不同层次上对图像内容进行分析和提取是目前被广泛接受的一种图像语义的获得方法[5,6]. 然而, 目前还没有一种完整的基于语义的图像内容描述方案来支持实际的检索过程.

本文提出了一种多级图像描述模型, 该模型在不同层次上对图像内容进行分析和提取, 实现了渐进式图像语义理解. 在此基础上, 我们还将该图像内容描述方法应用于实际的图像检索系统, 使系统的交互性以及检索效率大大提高.

2 多级图像描述模型

如何描述图像内容, 使其尽可能与人对图像内容的理解一致, 是图像检索的关键所在, 也是其难点所在. 传统的图像描述模型建立在低级特征的基础之上, 一般以统计数据的形式出现. 实际上, 这些统计数据与人对图像的内容理解存在很大差异, 首先, 人对图像的理解是建立在人类已有知识的基础之上的, 而这些低级特征无法反映这些经验知识. 其次, 图像内容具有"模糊"的特性, 无法用简单的特征矢量来描述. 针对以上两点, 我们提出了一种多级的图像内容描述模型, 在不同层次上提取图像的内容信息, 实现了一种渐进式的图像内容描述. 该模型可以用图 1 来说明.

从图 1 可以看出, 该模型包含五层: 原始图像层, 有效区域层, 视觉感知层, 目标层以及场景层. 虚线以右给出了每一层相应的公式描述. 整个模型可以用式(1)来表示.

收稿日期: 2000-05-20; 修回日期: 2001-06-20

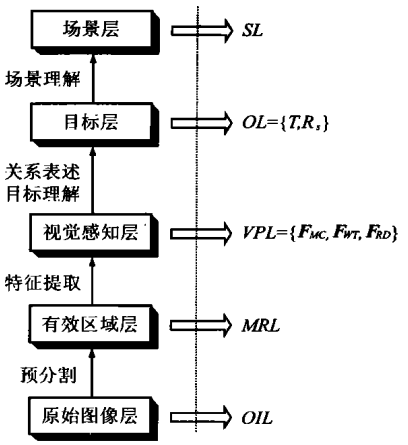

图 1　多级图像描述模型

$$IDM = \{OIL, MRL, VPL, OL, SL\} \quad (1)$$

图 1 还反映了各层之间的关系. 在相邻两层之间, 上一层的描述直接来源于下一层, 也就是说, 在下一层描述的基础上, 通过一定的操作来获得上一层的描述. 其中, 目标层是关键的一层, 将在第三节详细讨论. 这里先对有效区域层, 视觉感知层以及场景层的描述给予简单介绍.

2.1　预分割

预分割是图 1 流程中的第一步, 其目的是获得有效区域层的表述. 所谓有效区域是指一幅图像中相对于人类视觉有一定意义的区域. 与通常意义下的分割不同, 这里无须对图像进行像素级的精细分割, 关键是提取相对完整的区域. 具体算法见文[7].

2.2　特征提取

获得有效区域层的描述之后, 需要对图像中每一有效区域进行特征提取以获得视觉感知层的描述. 该层的描述直接影响到上一层目标层的描述, 因此需要选择适宜的图像特征. 在具体应用中, 要根据实际需要选择特征. 考虑到我们所选择的图像库是风景图像(见下), 其颜色和纹理信息均比较丰富, 另外对区域的描述也是视觉感知层描述的一部分, 所以选择了以下三个特征

(1) 混合颜色特征(Mixed Color Feature)

该特征定义在 HIV 颜色空间的一个子空间, 综合考虑了色度、亮度和饱和度对人视觉感知的影响. 详细的讨论请参见文献[8]. 在图 1 中, 混合颜色特征用 $\boldsymbol{F}_{MC}$ 来表示.

(2) 小波包纹理特征

多尺度描述目前已广泛应用于图像纹理分析. 这里引入小波包变换, 提取各分解层的 L_1 范数和方差构成纹理特征矢量[9]. 在图 1 中, 小波包纹理特征用 $\boldsymbol{F}_{WT}$ 来表示.

(3) 区域描述特征

考虑到目标层描述的需要, 我们引入了区域描述特征, 包括归一化面积, 区域重心坐标, 形状因子等. 在图 1 中, 区域描述子用 $\boldsymbol{F}_{RD}$ 来表示.

2.3　场景理解

场景层是多级图像描述模型的最高层, 它主要考虑的是一幅图像作为一个整体所体现的语义, 对场景层的描述需建立在对目标层的描述基础之上, 本文不加详细讨论.

3　目标层描述

在图 1 所给出的多级图像描述模型中, 目标层是至关重要的一个部分, 它是从图像低级特征到高层语义的过渡. 具体来说, 目标层利用视觉感知层提供的信息, 通过目标理解和关系表述形成本层的描述.

3.1　目标理解

目标理解是获得目标层描述的关键环节, 也是整个图像描述模型的关键所在, 它的目的是根据人类的知识"识别"出图像中每一个有效区域"是什么". 自动目标理解填补了图像低级特征对图像内容描述的不足, 与人工识别相比又克服了主观性影响, 提高了效率. 这里我们采用了迭代方法对训练样本集和测试样本进行逐次比较, 最终完成目标理解. 具体流程如图 2 所示.

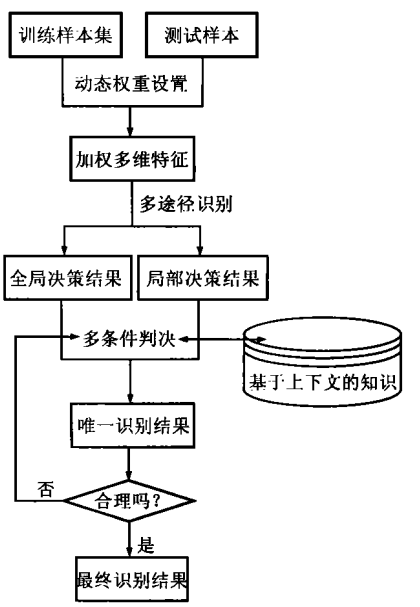

图 2　目标理解流程

从图 2 可以看出, 在目标理解的过程中有三个关键环节: 动态权重设置, 多途径识别以及多条件判决. 下面分别加以介绍.

3.1.1　动态权重设置　与单一的图像低级特征相比, 多特征包含了更多的信息. 在实际应用中, 如何确定各特征之间的比例关系直接影响到最后结果. 这里针对我们的实际需要提出了一种动态权重设置方案, 主要思想是考虑测试样本与整个训练样本集之间的关系, 使权重与其相符合. 给定特征集 $F = \{\boldsymbol{F}^1, \boldsymbol{F}^2, \cdots, \boldsymbol{F}^K\}$, 则其中第 k 个特征的权重可由式(2)~4 得到:

$$\bar{d}^k = \frac{1}{N}\sum_{i=1}^{N} Dis(\boldsymbol{F}_{test}^k, \boldsymbol{F}_i^k) \quad (2)$$

$$\sigma^k = \frac{1}{N}\sum_{i=1}^{N} \left| Dis(\boldsymbol{F}_{test}^k, \boldsymbol{F}_i^k) - \bar{d}^k \right| \quad (3)$$

$$w^k = \sigma^k / \bar{d}^k \quad (4)$$

在式(2), (3)中, N 表示训练样本集中的样本数, $\boldsymbol{F}_{test}^k$ 表示测试

样本的第 k 个特征, F_i^k 表示训练集中第 i 个样本的相应特征. 在式(4)中, 权重与距离的均值成反比, 而与距离的方差成正比. 前者可消除各特征绝对数值带来的影响; 后者, 由于方差反映了特征距离的离散程度, 方差的数值越大, 表示该特征越能够反映出测试样本在训练样本集中的倾向性. 因此, 我们对于具有较大方差值的特征适当加大其权重.

为测试动态权重设置的性能, 我们进行了一组对比实验来比较动态权重设置与固定权重设置. 在每一次实验中, 除了权重的设置方法不同以外其他操作完全按照图 2 的流程进行. 训练样本集共包含 163 个样本, 8 测试样本共有 292 个. 实验结果如表 1 所示, 其中 w_{MC} 是对混合颜色特征的权重, w_{WT} 是对小波包纹理特征的权重.

表 1 不同权重设置下的误识率

权重设置方案	误识率
动态权重设置	11.0%
$w_{MC}=w_{WT}=0.5$	24.3%
$w_{MC}=1.0, w_{WT}=0$	29.5%
$w_{MC}=0, w_{WT}=1.0$	30.5%

由表 1 的实验结果可见, 使用动态权重的误识率比使用不同固定权重的误识率要低.

3.1.2 多途径识别 图像的目标具有复杂和多样的特点, 在同一目标范畴之下可能存在多个具有不同的低级特征的子类. 比如说, 在“天空”这一目标范围下存在“晴朗无云的天空”, “乌云密布的天空”以及“夕照的天空”等等. 显然这些子类的低级特征描述存在很大差异, 因此简单的判决方法——例如常用的最小距离法——无法给出较高的正确识别率. 这里我们通过多途径识别综合考虑每一目标范畴下不同子类的差异. 具体来说提供了两种识别方案: 全局决策和局部决策.

所谓局部决策, 简单说就是在训练集的每一目标子集中寻找与测试样本的最近点, 然后再确定最终的识别结果. 具体说, 首先将训练样本集按照其与测试样本的特征距离升序排列: $Dis(F_{test}, F_1) < Dis(F_{test}, F_2) < \cdots < Dis(F_{test}, F_N)$, 其中局部识别结果可由式(5)得到:

$$R_{local} = \mathrm{Arg}\max_{i\in I}\left\{\sum_{k=1}^{N_{top}}\frac{x_i(k)}{Dis(F_{test}, F_k)}\right\} \tag{5}$$

其中 $N_{top} \ll N$, i 表示训练样本的目标集中第 i 个元素, x_i 为特征函数:

$$x_i(k) = \begin{cases} 1, & \text{if } F_k \in i \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

在式(5)中, 我们选取训练样本集中与测试样本距离最近的 N_{top} 个样本, 主要是为了获得每一目标范畴中与测试样本最“接近”的子集, 而对于与测试样本存在差异的训练样本则可不考虑. 在具体应用中, N_{top} 的值应根据具体需要来确定. 在我们的实验中选择了 $N_{top}=0.3N$.

全局决策可以如下定义:

$$R_{global} = \mathrm{Arg}\min_{1\le i\le N}\{Dis(F_{test}, F_i)\} \tag{7}$$

需要说明的一点是, 全局决策在本质上就是最小距离判决.

3.1.3 多条件判决 多条件判决是目标理解的最后一个重要步骤. 在多条件判决中需要完成下述三个功能: 第一, 在多途径识别给出的多个(这里是两个)识别结果中确定唯一的结果; 第二, 根据先验知识确定该结果是否合理, 以决定是否需要迭代; 最后, 根据本次判决情况给出相应的上下文知识, 以服务于下一次迭代. 多条件判决是一个复杂的过程, 然而对此并没有统一的公式化表达, 应根据实际需要设计多条件判决的方案. 这里我们引入了状态矩阵, 对每一种可能的多识别结果给出相应的判决调整以及上下文知识提取的方案. 多途径识别与多条件判决相结合提高了正确识别率. 表 2 给出了一组对比实验的结果, 其实验条件与表 1 的实验条件完全相同.

表 2 不同识别方案下的误识率

识别方案	误识率
多途径识别与多条件判决相结合	11.1%
单一的全局决策	24.8%
单一的局部决策	30.8%

由表 2 的实验结果可见, 使用多途径识别与多条件判决相结合的误识率比使用单一全局决策或单一局部决策的误识率都要低.

3.2 关系表述

关系表述是目标层的另一个组成部分. 在目标识别的基础上对图像中各个“目标”之间的关系加以描述是获得图像语义的一个重要部分. 特别是对于图像理解来说, 有了关系表述则可以对同一幅图像给予不同的描述, 体现了图像内容的“模糊”特点. 尤其需要强调的是, 将关系表述应用到图像检索系统中, 用户可以更灵活地选择所希望的图像内容, 大大提高检索的效率(见后).

关系表述包含很多方面, 最常用的是空间关系表述. 给定目标 A 和目标 B 的区域重心坐标(由视觉感知层的区域描述特征提供) $P_A(x_A, y_A)$ 和 $P_B(x_B, y_B)$, 它们之间的空间关系表述(Spatial Relationship Representation)可以由式(8)得到:

$$SRR(A, B) = \mathrm{tg}^{-1}\left(\frac{y_B - y_A}{x_B - x_A}\right) \tag{8}$$

以上我们详细讨论了目标层描述的获得, 其中重点放在目标理解算法. 与诸如通过模糊 C-均值分类器对目标进行聚类分析以方便机器操作的方法不同, 我们的目标理解过程采用了基于先验知识的上下文驱动方式, 在此过程中, 系统对测试对象的“认识”不断深入, 为准确判断目标的属性提供了条件. 实验结果也证实了这一点.

4 渐进式图像理解在图像检索中的应用

前面提到如何描述图像内容是图像检索技术的关键之一. 这里我们将渐进式图像理解应用到图像检索中去, 以实现结合语义的检索过程.

在目标层, 整幅图像被分成若干个表达一定语义且存在一定关系的区域. 这就使得用户可以根据自身需求向系统提出希望图像中包含什么样的语义区域(也就是目标), 同时还可以进一步要求这些目标之间应满足什么关系. 这就将用户从提供样例图像的不便中彻底解放出来, 更重要的一点是便利人机在“语义”层次上交流. 而对于图像的低级特征是完全

透明的.

下面介绍一种基于渐进式图像理解的检索系统, 该系统支持基于目标及基于空间关系的查询. 利用这个检索系统, 对一个用于专业美术设计的图片库进行了检索试验, 验证了渐进式图像理解在图形检索中的有效性. 该图片库中, 全部为自然景物图片, 共有约 800 幅图像. 各图像包含的目标可分为七类: 山、树、地、水、天空、建筑物以及花卉, 含有较复杂的视觉信息.

4.1　基于目标的查询

所谓目标查询, 即用户提供目标名称进行查询. 系统首先在图像库中搜索包含用户所选目标的候选图像, 然后对每幅图像计算其"相似度", 以用来确定系统提取图像的排序. 相似度的计算应根据实际需要来确定, 这里我们选择目标的归一化面积(由视觉感知层的区域描述特征提供)作为相似度的标准. 例如用户选择了图像库中第 k 种目标, 则候选图像的相似度可由下式得到:

$$S = \sum_{i=1}^{N} A_i^k \tag{9}$$

在式(9)中, N 表示候选图像的总数, A_i 表示第 i 幅候选图像中隶属于第 k 种目标的区域归一化面积. 式(9)是建立在这样一种假设之上的: 用户希望所选择的目标在整幅图像中所占的比例越大越好. 这一假设在实际应用中是合理的, 相似度值越大的候选图像在检索系统中的结果位置越靠前. 对于用户选择了多个目标的情况, 相似度应定义为候选图像中所有隶属于用户选择目标区域的归一化面积的总和. 设用户选择了 $k_1, k_2, \cdots, k_L$ 共 L 个目标, 则候选图像的相似度可由下式得到:

$$S = \sum_{i=1}^{N} \sum_{j=1}^{L} A_{iL}^{k} \tag{10}$$

图 3 给出了一个目标查询的实例. 用户选择了"山"和"树"两个目标, 所得到的查询结果的前十幅图像见图 3, 这十幅图像从低级图像特征的角度来看存在很大差异, 用颜色、纹理等方法很难得到这样的检索结果. 但从图像语义的角度来看他们都包含了相同的目标范畴, 因此出现在同一检索输出结果中.

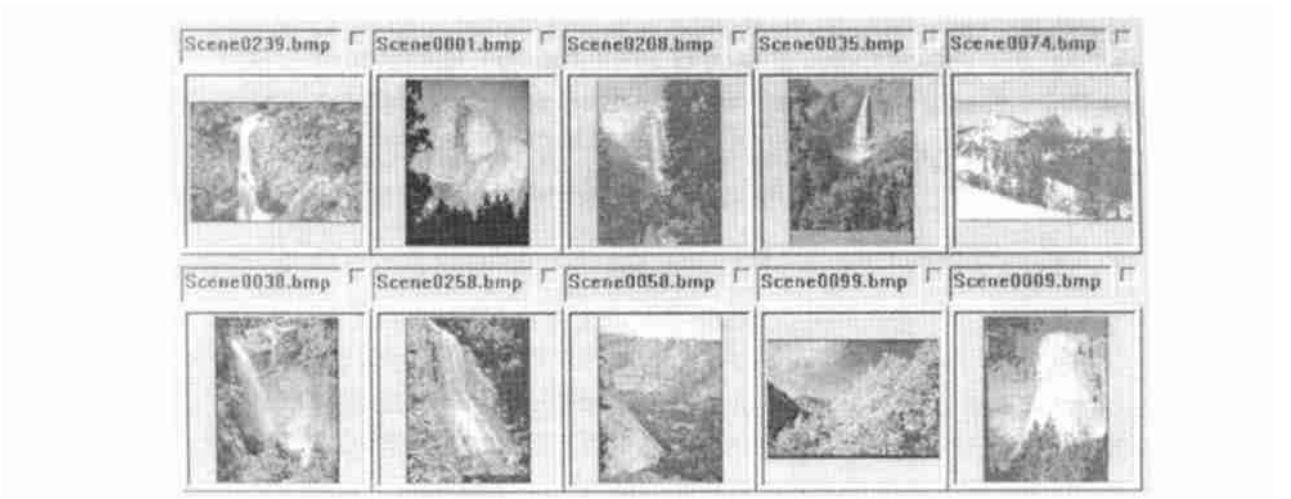

图 3　目标查询实例

4.2　基于空间关系的查询

在用户选择了一个以上目标的情况下, 用户还可以对所选目标之间的关系提出要求. 这一点是多级图像描述模型中目标层的关系表述的具体应用, 它提供给用户更多的查询条件, 使检索过程更能反映用户的愿望. 目前主要考虑的目标之间的空间关系. 例如可在式(8)的基础上, 将关系查询简化为每两个目标之间离散化的空间关系. 表 3 给出八种位置关系的表述.

表 3　$A \overrightarrow{\ } B$ 空间关系表述

$A \overrightarrow{\ } B$ 空间关系(SR_k)	$SRR(A, B)$ 所在范围
上$\overrightarrow{\ }$下(SR_1)	[67.5°, 112.5°]
下$\overrightarrow{\ }$上(SR_2)	[247.5°, 292.5°]
左$\overrightarrow{\ }$右(SR_3)	[-22.5°, 22.5°]
右$\overrightarrow{\ }$左(SR_4)	[157.5°, 202.5°]
左上$\overrightarrow{\ }$右下(SR_5)	[22.5°, 67.5°]
右下$\overrightarrow{\ }$左上(SR_6)	[202.5°, 247.5°]
左下$\overrightarrow{\ }$右上(SR_7)	[292.5°, 337.5°]
右上$\overrightarrow{\ }$左下(SR_8)	[112.5°, 157.5°]

在实际应用中, 对应已选择了多个目标的情况, 用户可以指定其中任意两个目标的空间关系, 从而缩小搜索范围, 力求找到更符合要求的图像. 假定目标 A 和 B 的空间关系为 SR_k, 则候选图像的相似度可由式(11)计算得到:

$$S = \frac{1}{|SRR(A, B) - V_k| + \varepsilon} \tag{11}$$

其中 ε 为一个小正数, V_k 为 $SRR(A, B)$ 所在范围的中值.

在图 3 目标查询的基础之上可以通过空间关系查询进一步优化查询结果. 如前所述, 本系统提供了八种空间关系, 用户可以自由选择. 这里选择了"左(山)-右(树)"的关系, 查询结果如图 4 所示.

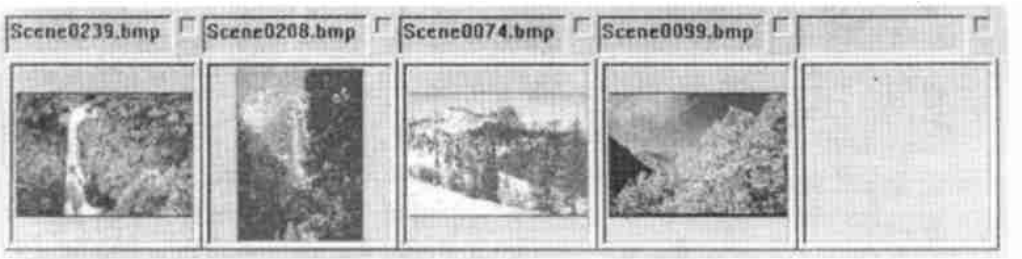

图4 空间关系查询实例

5 结论

本文提出了一种多级图像描述模型, 在不同层次上对图像内容进行分析和提取, 实现了渐进式的图像语义理解, 并获得了目标层的描述. 与以往基于相关反馈或基于分类的图像语义提取算法不同, 本文给出的是一种基于先验知识的上下文驱动的目标理解算法. 将上述方法应用到图像检索中去, 可以使得用户与计算机在“语义”层进行交互, 使检索过程更方便于用户.

参考文献:

[1] Lee D, Barber R, NiBlack W, et al. Indexing for complex queries on a query-by-content image database [C]. Proc. 12ICPR, 1994: 142-146.

[2] Pentland A, Picard R W, Sclaroff S. Photobook: content-based manipulation of image database [J]. International Journal of Computer Vision, 1996, 18(3): 233-254.

[3] Smith J R, Chang S F. Tools and techniques for color image retrieval [C]. SPIE, 1996, 2670.

[4] Wu J K. Content-based indexing of multimedia database [J]. IEEE Trans. KDE, 1997, 9(6): 978-989.

[5] Jaimes A, Chang S F. Model-based classification of visual information for content-based retrieval [C]. SPIE 1999, 3656: 402-414.

[6] Hong D Z, Wu J K, Singh S S. Refining image retrieval based on context-driven method [C]. SPIE 1999, 3656: 581-593.

[7] 罗 , 章毓晋, 高永英. 基于分析的图象有意义区域提取. 计算机学报, 2000, 23(12): 1313-1319.

[8] Y Y Gao, Y J Zhang. Object Classification Using Mixed Color Feature [C]. Proc. of ICASSP 2000, 4: 2003-2006.

[9] 吴高洪, 章毓晋, 林行刚. 利用特征加权进行基于小波变换的纹理分类 [J]. 模式识别与人工智能, 1999, 12(3): 262-267.

作者简介:

高永英 1998 年和 2000 年分别于清华大学电子工程系获学士和硕士学位, 现在美国密西根攻读博士学位. 感兴趣研究领域是多媒体信息处理, 图像分析和理解技术.

章毓晋 1989 年获比利时列日大学应用科学博士学位, 从 1989 年至 1993 年在荷兰德尔夫特大学作博士后及研究工作. 1993 年到清华大学工作, 现为图象图形研究所副所长, 教授, 博士生导师. IEEE 高级会员,《中国图象图形学报》副主编,“Pattern Recognition Letters”,“International Journal of Image and Graphics”,《电子与信息学报》编委. 主要研究领域是图象工程(图像处理, 图像分析, 图像理解及其技术应用), 已发表了 160 多篇研究论文, 著有《图象分割》等书 4 本.