

敦煌壁画数字图像语义描述方法研究*

王晓光 徐 雷 李 纲

摘 要 数字图像的语义描述与标注是解决图像检索中语义鸿沟问题的关键。由于缺乏面向领域的有效的数字图像描述方法规范,基于图像底层视觉特征的机器标注和基于专家知识的人为标注的标注结果都存在标注信息质量不高和结果不统一的问题。针对这一现实问题,本文基于图像元数据和信息需求理论,针对敦煌壁画数字图像这一特定文化遗产领域,提出了语义描述框架和领域主题词表相结合的数字图像内容语义描述方法,详细阐述了语义层次及其相互关系。同时,从图像语义描述粒度的角度讨论了语义粒度大小对标注成本的影响,以及该语义描述框架的可移植性问题。图 4。表 5。参考文献 25。

关键词 数字图像 语义标注 语义描述框架 语义层次模型 敦煌壁画

分类号 G250

Semantic Description Framework Research on Dunhuang Fresco Digital Images

Wang Xiaoguang,Xu Lei & Li Gang

ABSTRACT Semantic description and annotation of digital images are the key to solve the semantic gap problem in image retrieval. Due to the lack of an effective and domain-oriented description framework for digital images, the annotation results that produced by current methods, such as machine annotation based on low-level visual features of images and human annotation based on expert knowledge are of low quality and no uniform. To solve the problem, this paper proposes a semantic description method based on image metadata and information need theory for Dunhuang fresco digital images content, a specific field of cultural heritages, which combines semantic description framework and domain thesaurus and elaborates each layer's content of the semantic framework and their relationship. What's more, we discuss the effect of the semantic granularity on annotation cost from the perspective of semantic description granularity and the portability of the semantic description framework proposed here. 4 figs. 5 tabs. 25 refs.

KEY WORDS Digital images. Semantic annotation. Semantic description framework. Semantic hierarchy model. Dunhuang fresco.

1 引言

信息时代数字图像的数量增长十分迅速,人们已由读文时代逐步进入读图时代。面对海量的

数字图像,如何快速找到所需的内容已经成为一个棘手的难题。为了有效组织数字图像,图书情报领域普遍采用了元数据的方法。然而,元数据主要是对图像外部特征信息的描述,在图像内容描述上缺乏必要的规范,再加上图像底层特征与高层语义之

* 本文系科技部 973 项目“文化遗产数字化保护的理论与方法”(编号:904171200)和中组部“青年拔尖人才支持计划”项目“语义出版关键理论与技术”的研究成果之一。

通讯作者:王晓光,Email:whu_wxg@126.com

间存在明显的“语义鸿沟”,由此导致图像内容信息的语义标注工作困难重重,基于语义的图像检索和图像内容的语义计算不能达到理想效果,大量数字图像数据库被闲置,无法得到充分和高效的利用。

对图像进行有效的描述和标注是填补图像检索中“语义鸿沟”的关键,也是利用计算机实现数字图像知识发现的必要环节。图像的描述与标注的内容主要是图像的高层语义信息。目前,不管是图像语义检索领域还是图情领域,都缺乏领域性的图像内容语义描述规范,由此导致图像的语义标注缺乏框架性指导,标注结果因人而异,漏标和错标较为常见。基于机器学习的计算机自动标注结果更是不理想,远远达不到行业的要求。

敦煌是中国文化遗产中一个特别的领域,敦煌壁画更是人类文化遗产中的瑰宝,具有极高的艺术观赏和研究价值。随着文化遗产数字化和数字人文^[1]研究的兴起,敦煌研究者拍摄积累了大量的数字壁画图像,为敦煌研究和敦煌壁画的传播奠定了基础。然而囿于语义标注的不足,这些图像并没有得到充分利用。为了发掘敦煌壁画数字图像中蕴含的语义信息,并对其进行规范表达和有效组织,以实现图像的自动标注和语义出版^[2,3],有必要提出一套针对敦煌壁画的数字图像语义描述规范,以解决当前数字图像元数据在图像内容细粒度描述上的不足。

我们首先对数字图像语义标注相关的研究进行了综述,然后针对敦煌壁画提出一套数字图像语义描述框架(Semantic description framework of digital image, SDFDI)。该框架目的是实现数字图像标注内容的规范化,为机器自动标注提供参照系。该框架虽然只针对敦煌壁画领域,但可为相同领域数字图像的组织和管理提供借鉴和参考。

2 相关研究

数字图像描述与标注及其相关研究可从图像元数据标准、用户图像检索需求描述以及图像内容描述等角度来考察。图像元数据是常见的图像描

述框架,广泛应用在图像保存和共享领域。目前图像元数据标准主要有 EXIF、VRA、MPEG-7、SVC 等^[4],然而,图像元数据的描述标准只适合于图像外部特征的浅层描述,对图像深层语义内容揭示不足。从用户图像检索需求来研究图像的描述与标注是个很好的角度,用户的图像检索需求反映了用户对图像的理解,对应着图像的深层语义内容,由于该内容来自于大众用户,其检索需求的描述范围相对比较全面。目前,国内外关于用户图像检索需求的研究成果已较为丰富^[5-13],这些研究主要通过用户图像检索的行为及心理来对用户图像检索需求进行分类,以更好地辅助图像标注及检索过程。

在图像内容描述方面,Eakins 在 1996 年从图像检索需求的角度将图像语义内容分为三个层次^[14],分别是图像底层视觉特征层、对象层(对象和对象空间)和语义概念层(场景、行为和情感)。该研究主要从图像检索的角度对图像内容进行分类,而非专门研究图像内容的描述问题。Badr^[15]等从医学图像关联的诊断报告中抽取图像的描述信息,并将描述信息分为元数据层和内容层两部分,元数据层又分为面向上下文、面向领域和面向图像的子层,内容层分为物理特征、关系特征和语义特征三个层次。由于该描述框架严重依赖于医学图像领域,因此通用性较差。Gupta^[16]等人从图像的标注信息中生成该图像的自然语言描述的句子,这种方法产生的图像描述内容存在内容简单、不够全面等问题。对诸如壁画等艺术作品的数字图像来讲,简单的几句话描述是远远不够的。Li^[17]等人也曾提出一个基于自然语言表达的图像内容描述框架,该框架使用一个语义向量来表示图像的语义内容,使用模糊集理论来确定各分量的权重,但是该语义向量只包含诸如“粗糙”、“精细”、“不规则”等抽象的语义表示,对图像内容的揭示还不够准确。

通过文献分析发现,目前真正从图像内容本身进行语义描述的研究很少。由于图像标注是图像描述的具体过程,对图像进行标注也就是对图像进行描述。从这个意义上讲,当前研究更多的是关于

图像内容的“粗描述”,如图像分类、图像显著区域检测、图像中的对象识别、图像中对象语义类型标注等。这些“粗描述”一般借助机器学习等技术自动或半自动完成。由于“语义鸿沟”的存在,这些方法对图像中更高层的语义内容的标注效果很不理想。

3 敦煌壁画数字图像的语义描述

3.1 壁画数字图像的语义特征

壁画作为现实世界的实体对象,需要和对其进行拍摄得到的数字图像相区别。数字图像的内容可以反映现实壁画的颜色、形状等视觉信息,但对于壁画中对象本身所表达的行为、情感等高层语义信息,数字图像是不能量化的,数字图像本身也是一个对象,有自己的属性。而壁画的内容是由人

工绘制的,它不同于数字图像的内容是现实对象的反映,它还包括非现实世界的艺术抽象成分。

该区别可以通过图 1 来反映。数字图像和壁画都存在于现实世界中,数字图像的内容依赖于数字图像本身,壁画数字图像是数字图像的子集,壁画数字图像内容是数字图像内容的子集,其依赖于壁画数字图像。这里需要注意的是壁画数字图像内容和壁画内容是不同的,壁画对应的数字图像和壁画本身是两个不同的对象,其内容是有差别的,壁画数字图像内容依赖于壁画的内容,而壁画的内容一方面来自壁画对象本身,可通过其数字图像来反映,另一方面壁画的内容包含了艺术家的艺术创作成分,该部分一般是抽象的,不能和现实世界完全对应或不存在于现实世界中。因此,对敦煌壁画数字图像的描述内容应包括对壁画对象本身的描述和对壁画反映的内容的描述。

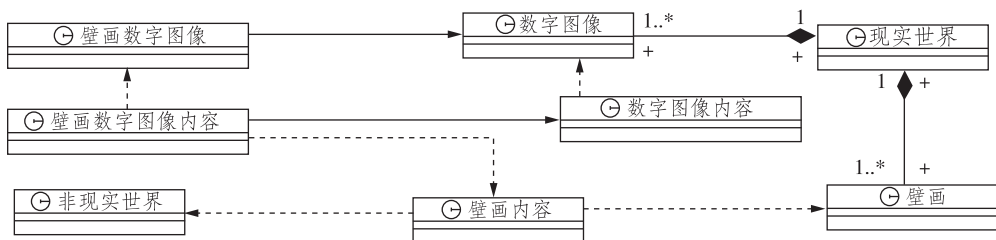


图1 壁画与数字图像的区别与联系

3.2 数字图像的语义描述层次模型

本文根据用户的图像检索需求,将相应的术语词表和图像元数据加入到图像的高层语义表

达中,提出了一个数字图像语义描述层次模型(见图2)。

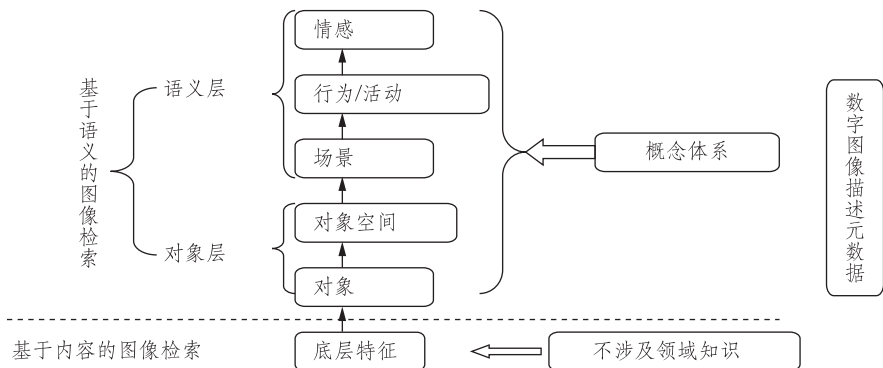


图2 数字图像语义描述层次模型

在数字图像语义描述层次模型中,最底层是图像的视觉特征层,包括图像的颜色、形状以及纹理等特征,采用统计学习、模式识别等技术手段可以实现基于内容的图像检索(CBIR),该层次不涉及图像的领域知识,因此存在图像底层特征与高层语义之间的“语义鸿沟”。图像底层视觉特征层之上可统称为图像的高层语义层。

第二层和第三层合称为对象层,包括对象和对象空间层。对象是指图像中包含的实体,如图像中的人物、动植物以及其他物品等;对象空间是指对象之间的方向、拓扑等空间关系。对象层之上是更为抽象的语义层,依次是场景、行为和情感层。场景是一组特定对象的特定空间关系的抽象,图像中除主要对象之外的部分都可称作场景。行为的主体一般是指人类,是指人类在一定场景下的特定动作或状态,人类的集体活动也在行为范畴之内。情感是指图像中人物对象所流露的情绪感受,如高兴、悲哀等。图像的语义层次模型由底向上是逐层抽象的,对图像检索也更加困难,当然在情感层之上还可以增加更为抽象的图像语义层次,但这样做需要更为抽象的语义表达。由于图像的高层语义内容的描述需要用户的经验以及领域术语的支持,利用机器学习等方法来处理图像视觉特征以获取图像高层语义信息将变得很困难。将用户知识和领域术语抽象为概念体系融入到图像语义层次当中,可满足用户各个角度的图像检索需求。

相对于基于视觉特征的图像检索,利用图像的高层语义信息可提供基于语义的图像检索,以发掘用户深层次的检索需求。同时,对于图像本身而言,图像的元数据描述了图像的外部技术特征,对图像资源的获取、共享与保存有重要的意义。图像的名称、主题等元数据需要高层语义的支持,图像的大小、格式、EXIF 信息等元数据则直接来自于图像固有的特征,因此图像的元数据信息贯穿于整个

语义层次模型。按照图 2 的层次模型,下文将分别从是否可自动描述、是否需要受控术语支持等角度对敦煌壁画数字图像的各个层次进行阐述。

(1)底层视觉特征。图像的颜色、纹理、形状等特征已经存在于图像中,对其进行描述就是对这些特征进行提取,这种提取可通过机器自动完成,可使用不同的提取方法如模式识别、图学习等技术来获取。对于敦煌壁画而言,由于其作为艺术作品具有特殊性,不同于现代生活中一般的数字图像,如壁画中的对象可能是抽象的或现实世界不存在的,壁画中的场景、情感等也与现实世界中的场景、情感信息有很大的差别。梁广显^[18]曾通过实验发现有些情况下对壁画数字图像采用这种底层特征来识别高层语义是不完全合适的。由于本文主要关注图像的高层语义信息的描述,对于图像底层特征的提取,本文不做深入探讨。

(2)对象描述。对象在整个语义层次模型中是最重要的。本文将对象的描述包括三种内容:对象的分类、属性及关联。除了图像元数据及视觉特征外,图像中的对象作为现实世界的一个反映有其自身属性,如图像中的人物有名称、性别、姿态等属性。由于使用模式识别等方法来确定这些高层语义信息存在一定的困难,而这些属性又不能完全由底层特征识别得到,所以需要相应的领域受控术语或本体,如《敦煌学大辞典》^[19]等来对其进行描述。除了对象本身的描述之外,对象之间还存在丰富的语义关联,本文将在下文进行阐述。对象的描述可通过对象识别技术辅助完成一部分工作,如图像中对象的分割、提取、区域范围表示等,都是实际标注工作需要考虑的。

(3)对象空间。对象空间的描述也是整个层次模型中较为重要的部分。MPEG-7 中关于空间关系的描述可分为三类:数字图像中各个对象所在区域的方向关系,壁画中对象本身的拓扑关系和对象间的语义空间关系(见表 1)。

表 1 对象空间关系

关系类型	示例	是否可自动获取
空间方向关系	top of、bottom of、left of、right of、right top of、middle of	是
拓扑关系	nearby、within、contain、adjacent to	需区分情况
空间语义关系	belongs to、part of、related to、consist of	否

空间语义关系即图像对象间的逻辑关系如从属、依赖等关系,这种描述一般不能自动提取。空间方向关系可自动提取,如根据对象的区域位置来自动确定,但这种方向的确定依赖于图像中对象区域范围的划分方法和方向的计算方法,不同的区域划分和计算方式产生的空间关系类型也会有些差异。拓扑关系可根据空间方向关系来辅助确定,如对象间 middle of 关系可能是 within 的拓扑关系,但这需要区分情况,如桌子上的茶杯,不能说茶杯在桌子里面,但可以表述为茶杯在桌子中间,对于房间中的桌子则可以表示为 middle of 的方向关系和 within 的拓扑关系。

(4)场景描述。图像中除主要对象外其余部分都可以叫做场景,场景包含环境,如雨雪等,场景可以解决一些图像语义上的歧义问题。对于主要对象的确定,不同的人有不同的认识,当然可以采用视觉注意力模型^[20]判断图像的显著区域。由于壁画中场景与现实世界的场景存在差别,现实世界中的场景可通过数字图像背景的颜色、纹理等视觉特征来识别,如大片的绿色区域可能代表绿草地,但壁画中的场景是对现实世界的描绘,艺术成分各异,不能用相同的手段来获取,而且有的场景可能是融合的,如室内与室外场景的交替。由于敦煌壁画描绘的不是实景,有些壁画包含场景,如故事画、经变画。有些壁画不包含场景,如尊像画、图案画。更多的壁画则不能有效地区分出具体的场景。目前大致将敦煌壁画中可辨别出的场景分为:地理场景,如天空、室内、室外、山河湖海等;时间场景,如春夏秋冬、清晨、傍晚等;天气场景,如风雨雷电等。需要注意的是地理场景中一些词汇也可以作为对象来描述,不过在敦煌壁画中,主要以人物及故事画为主题,一般将这些词汇视为场景,而不纳入对象描述范畴。

(5)行为描述。这里的行主要是指人的行为。在敦煌壁画中主要指佛教神祇、世俗人物及供养人等,其含义比较广泛,包括事件、活动等集体行为。在描述一个行为时必不可少的要素是对象(who)、内容(what)、地点/场景(when)。因此对行为的描述至少应包含这四个内

容。有些行为是人与物、人与人之间的作用与被作用的关系,这些行为的描述和对象的空间关系有关,行为描述需要人类大量的背景知识和推理能力。其中地点的描述较为特殊,它本身可以作为对象进行描述,同时地点也可以充当行为发生的场景。由于敦煌壁画中行为的发生地点与时间大部分是抽象不明确的,所以只能使用场景中的地理场景和时间场景以及朝代信息来代替行为发生的地点和时间。而且,敦煌壁画中有大量约定俗成的行为描述术语可以用来对图像中的行为名称进行规范描述,如“阿弥陀经变”等。

(6)情感描述。图像情感描述可从两个角度考察:整个图像给人的主观感受,这个角度不是本文的考察范围;另外一个图像中的人物自身所流露出的情感。对于后者,心理学家将人类的基本情感分为高兴、惊奇、恐惧、悲伤、厌恶、愤怒等六类^[21]。本文将之应用于壁画中人物情感的分类描述。对于不同的人来,不同的壁画对象所表现的情感可能是不同的,或壁画中人物的情感是抽象模糊不可辨识的,因此人物情感的识别需要结合人物所在的场景进行描述。

3.3 图像语义关联

语义层次模型和壁画领域术语只是浅层描述,语义层次之间还存在复杂的关联,描述对象本身也存在固有的属性。为了避免混淆,本文将对象的属性分为两类:反映对象元数据特征的属性称为数据属性,反映对象间关联关系的属性称为对象属性。

由于人物对象在术语词汇中出现的最多,在语义层次模型中人物对象占的比例也很大。因此,本文按敦煌专家的需求给出了人物属性的详细描述内容。表2给出了人物的数据属性,对于集合性质的对象增加了如下属性描述:对象数量、对象形状(团、圈、列、排)。

对象间的关联关系可分为三种,即对象空间层次上的关系、行为层次中的关系、其他可作为对象属性的关系。由于对象空间和行为层次已有描述,这里主要对对象属性进行阐述。但是要注意行为层次和对象属性之间的区别。敦煌壁画的行为描

表2 人物数据属性

描述属性	取值范围	描述属性	取值范围
性别	男、女	眼颜色	白、黑等
姿势/动作	走、坐、跪、跑、跳等	头部朝向	上下左右等
方向	上下左右等	双手动作	举起、交于胸等
体态	丰腴、细弱等	头部背光颜色	黄、白等

述大量的专业术语,如“射箭”行为蕴含着人物和弓箭之间的“持有”关系,但对于如“舞蹈”之类的行为则不存在类似的情况。人物对象的“服饰”属性蕴含着人物与服饰对象的“穿着”行为。这两者互为补充,丰富了对象间语义关联关系的描述。由于“吹

奏、穿着”之类的词汇作为对象属性进行设计将十分庞杂,因此对于可用行为术语或空间关系来描述的内容,将不作为对象属性描述,只有通过行为层次不易描述时再考虑通过对象属性的方式进行描述。表3总结了对象间通用的对象属性。

表3 通用对象属性

对象属性	说明	范围
持有	人物与其他对象之间的关系	举、托、抱等
服饰	人物的头饰/衣饰/配饰等	—
社会关系	人物之间的关系	弟子、大臣等
图案	对象上的图案	服饰上的图案、器物上的图案、建筑上的图案等

可通过图3来说明图像语义层次之间的关联关系。语义层次模型以对象为中心,其他高层语义层次都依赖于对象层,对象的聚合是数字图像的主体,对象有自身属性,并通过对象属性、空间关系、

行为产生对象间的互联。行为依赖于对象并和场景关联,数字图像本身具有元数据描述属性,数字图像的视觉特征也有不同的属性描述方式。

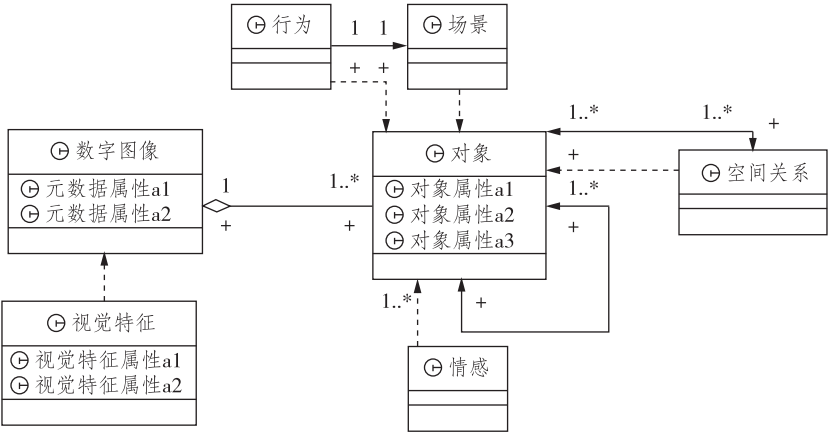


图3 语义层次间的关联图

3.4 敦煌壁画数字图像内容的概念体系

领域术语是对该领域的良好描述,蕴含着描述对象的某些特征,是获取领域知识的最佳途径。

例如敦煌壁画对尊像画的分类就涉及很多的壁画术语,如飞天、菩萨等。我们综合参考《敦煌学大辞典》^[19],以及其他对敦煌壁画分类的研究文

献^[22,23],整理出该领域的专业术语,并对这些词汇进行有效的分类,形成了敦煌壁画领域的分类主题词表。敦煌壁画内容分类体系如表4所示,该体系

按图2的模型进行分类设计,只显示了语义层次模型各层的最高层级的分类术语。

表4 敦煌壁画顶层分类术语

语义层次	分类术语
情感	高兴、惊奇、恐惧、悲伤、厌恶、愤怒、安详
行为/活动	文体活动、生产生活、佛教活动、一般动作
场景	地理场景、时间场景、天气环境
对象空间	空间方向关系、拓扑关系、空间语义关系
对象	人物、动物、植物、器具、交通工具、建筑、服饰、装饰图案

敦煌壁画的术语主要分布在对象和行为语义层次。因为对象和行为更能突出敦煌壁画的领域特色,而对象空间、场景、情感等语义层次具有普遍性,不仅局限在壁画领域,在其他数字图像的描述中,对象空间、场景、情感的分类可能大同小异。

4 敦煌壁画数字图像描述方法及示例

4.1 分层描述方法

综合考虑用户对图像的检索需求和图像内容的语义层次模型,本文提出了一个融合敦煌术语词汇的分层描述方法。借鉴图形信息可视化中 context + focus, zoom^[24]的思想,先对数字图像所能表达的能够容纳大部分内容的整体进行描述,即全局性的描述,这一阶段主要使用敦煌壁画的最高层分类术语以及语义层次模型中的行为和场景信息。在视域不变的基础上,逐层放大得到局部图像继续对该局部的整体进行描述,包括对象、对象空间关系、行为、情感等,需要敦煌壁画次级的或更低层的术语词汇,经过这样的过程直至整个图像中没有可再分的对象为止。该方法实质上也是一种粒度分析法,图像的逐级放大就是对图像进行更细粒度的观察,高层的术语对应泛化的概念,底层的术语对应具体的概念,分别代表用户对图像的不同理解程度,用户的检索结果也就对应不同大小的结果集。分层描述的关键就在于运用粒度原理将各层级的描述信息进行融合。

通过分层描述的好处是越低层的图像所携带

的描述信息越多,因为它继承了其周围上下图形环境的描述信息,这样在对图像的某一个区域进行提取后,直接包含了其继承的上层信息,不用再单独进行描述,只需增减少量因提取操作带来的元数据内容层面的变化信息。同时,分层描述可以满足用户不同角度的图像检索需求。需要注意的是,随着描述的深入,对象将有越来越多的细节,对这些细节的描述将越来越不受术语体系的控制。

4.2 描述框架的应用示例

敦煌壁画数字图像的描述规范,可应用于敦煌壁画的图像标注系统、检索系统,壁画的分类整理以及为其它壁画艺术作品的语义描述提供参考。图像描述和标注的根本目的是为了满足不同用户图像查询检索的需求,填补用户需求与图像描述之间的鸿沟,解决“查询非所得”的困境。下面将给出使用上面的壁画语义层次模型对一个实际的图像进行描述,表5是对图4的语义描述结果。



图4 散花飞天^[25]

表 5 对图 4 的语义描述

语义层次	描述			
底层视觉特征	颜色直方图、纹理、形状等			
对象	第一层次	散花飞天、莲花、祥云		
	第二层次	对象的组成：飞天的躯干、四肢；人物的服饰：饰物（头冠、项圈、臂钏、手镯）、长裙、帛带		
对象空间	空间方向	飞天位于图像中上部、莲花位于图像下半部分		
	拓扑	飞天位于祥云上方		
	空间语义	飞天对象包括头、躯干、四肢等		
场景	室外、天空、山峦			
行为/活动	对象	飞天、莲花、祥云		
	内容	散花		
	地点	天空	时间	—
情感	安详			

使用分层描述方法可以穷尽图像中所有的对象,保证了图像描述的全面性,在图 4 的描述中,对于人物对象的描述分为两个层次,第一层次是个体层次,第二层次是个体的组成部分,当然对象仍可再分,如对于绘制详尽的尊像画,其头部的描述可再划分为发饰、背光、眼眉、鼻、耳、口等,这些描述是有必要的,因为敦煌壁画的研究少不了通过观察人物的面部分布、体态、神情等来总结归纳整个人物在该时期的风格特征。对象空间和场景容易处理,如对象空间中人物的头部与躯干部分是上下的方向关系,但采用空间语义关系更为合理,用飞天包括头部、躯干等来描述;对于行为的描述,需要参考敦煌术语来规范,行为内容使用规范术语或统一的行为动词组成。情感层次的描述较为简单,但识别情感却较为困难,根据场景及行为信息并结合人物实际表情,可以给该图中的人物赋予安详的情感描述。

对象的属性描述需要根据图像描绘的精细程度、图像的可辨识度来进行取舍,对于图 4 来说,人物的服饰颜色为棕褐色、体态丰腴、双手托举莲花、脖子戴项圈等都可以通过属性来反映,并和其他对象产生语义关联。

5 讨论与结论

5.1 图像描述的粒度与成本

为了保证图像描述的全面性,势必要细化图

像描述的粒度。图像描述的粒度包括描述术语的粒度和图像分层描述的粒度。描述术语粒度依赖于领域术语的层次划分方式,分层描述粒度则取决于在实际的图像分层描述过程中对图像层级大小的控制。一般来讲,描述术语的粒度和图像分层的粒度是一致的。在实际工作中,由于数据量与具体应用的差异以及人力、资本的不同投入,导致图像描述的全面性、粒度等方面都会受到制约。粒度太细标注人员的工作量大,标注的复杂性高,粒度太粗则标注的效果不好。

本文认为对敦煌壁画的描述应遵循以下原则:对于抽象对象的描述粒度宜粗,对于绘制详细的对象描述粒度要细;有些壁画有外框、自然景物、榜题、花纹等作为装饰,可以根据这些装饰图案来确定图像所包含的完整主体,以防止语义描述信息的遗漏;由于敦煌壁画距今年代久远,一些壁画的内容已不可辨识,对于这些无法辨识的部分或通过文献不能确认的部分可以不予以描述或给予注释性说明。

5.2 语义描述信息的质量控制

为了控制图像语义描述信息的质量,本文提出了领域主题词表的方法,并选用《敦煌学大辞典》和敦煌相关研究成果作为词汇来源,经过专家筛选最终确定,由此保证了主题词的规范性。本文提出了敦煌壁画的语义描述框架,后期欲采用 RDF 格

式对图像描述信息进行组织,相关研究将在另外的文章中详细探讨。

5.3 图像语义描述框架的可移植性

图像语义描述框架的可移植性直接关系到其应用范围。本文提出的图像语义描述框架只依赖于具体的领域,和具体的应用无关,不涉及软硬件环境,也不依赖于系统环境,所以该图像语义描述框架结构可直接应用于任何种类的数字图像。需要注意的是不同领域的概念体系不同,由此会导致语义描述内容不同,同时图像元数据也会有一定差别。通过领域概念体系的更换以及对象间关系的重定义可实现该描述框架的跨域使用。

5.4 未来的工作

本文基于图像元数据和信息需求理论,针对敦煌壁画数字图像提出了一套较为完整的图像语义描述方法,但尚未应用该方法开展大规模应用。下一阶段,我们将进行三个方面的工作:一是按照本文提出的数字图像语义描述框架(SDFDI),开发一款功能完善操作简便的数字图像语义标注软件;二是与敦煌图像研究专家合作完善敦煌壁画标注术语表;三是与敦煌研究院合作,针对特定专题或洞窟的壁画数字图像进行语义标注实验,依此检验和优化本文提出的框架与方法。

参考文献

- [1] 王晓光.“数字人文”的产生、发展与前沿[C]//2009中国高校哲学社会科学发论坛论文集,重庆,2009。(Wang Xiaoguang. The generation, development and research of “Digital Humanities” [C]//2009 Chinese University Philosophy Social Science Development Forum, Chongqing, 2009.)
- [2] 王晓光,陈孝禹.语义出版的概念与形式[J].出版发行研究,2011(11):54-58。(Wang Xiaoguang, Chen Xiaoyu. The concept and forms of semantic publishing[J]. Publishing Research, 2011(11): 54-58.)
- [3] 张晓林.颠覆数字图书馆的大趋势[J].中国图书馆学报,2011(37):4-11。(Zhang Xiaolin. The trends that will disrupt digital libraries [J]. Journal of Library Science in China, 2011(37): 4-11.)
- [4] Vassilis Tzouvaras. Multimedia annotation interoperability framework[EB/OL]. [2013-01-12]. <http://www.w3.org/2005/Incubator/mmsm/XGR-interoperability/>.
- [5] Jorgensen C. Attributes of images in describing tasks[J]. Information Processing and Management, 1998, 34(2-3): 161-174.
- [6] Cunningham S J, Bainbridge D, Masoodian M. How people describe their image information needs: A grounded theory analysis of visual arts queries[C]//Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries, Tucson, AZ, USA. New York, USA: ACM Press, 2004: 47-66.
- [7] Shatford S L. Analyzing the subject of a picture: A theoretical approach[J]. Cataloging & Classification Quarterly, 1986, 6(3): 39-62.
- [8] Jaims A, Jaims R, Chang S F. A conceptual framework for indexing visual information at multiple levels[C]//Proceedings of IS & T/SPIE Internet Imaging. San Jose, CA, 2000.
- [9] Batley S. Visual information retrieval browsing strategies in pictorial databases[C]//Proceedings of the 12th International Online Information Meeting. London, 1988: 373-381.
- [10] Hollink L, Schreiber A T, Wielinga B J, et al. Classification of user image descriptions[J]. International Journal of Human-Computer Studies, 2004, 61(5): 601-626.
- [11] 曹梅.网络图像检索行为与心理研究[J].中国图书馆学报,2011(05):53-60。(Cao Mei. Users' information behavior and psychology in image retrieval on the internet[J]. Journal of Library Science in China, 2011(05): 53-60.)
- [12] 武人杰.图像层次语义描述的初步研究[J].电脑开发与应用,2011,24(5):12-14。(Wu Renjie. Research on image hierarchical semantic description[J]. Computer Development & Applications, 2011, 24(5): 12-14.)

- [13] 周卫东, 冯其波, 匡萃方, 等. 图像描述方法的研究[J]. 应用光学, 2005, 26(3): 27-31. (Zhou Weidong, Feng Qibo, Kuang Cuifang. Study on image description methods[J]. Journal of Applied Optics, 2005, 26(3): 27-31.)
- [14] John P E. Automatic image content retrieval—Are we getting anywhere[C]//Proceedings of Third International Conference on Electronic Library and Visual Information Research. De Mont fort University, Milton Keynes: Aslib, 1996: 123-135.
- [15] Badr Y, Chbeir R. Automatic image description based on textual data[M]//Journal on Data Semantics VII. Springer Berlin Heidelberg, 2006: 196-218.
- [16] Gupta A, Mannem P. From image annotation to image description[C]//Neural Information Processing. Springer Berlin Heidelberg, 2012: 196-204.
- [17] Li Q, Shi Z, Shi Z. Linguistic expression based image description framework and its application to image retrieval[M]//Soft Computing in Image Processing. Springer Berlin Heidelberg, 2007: 97-120.
- [18] 梁广昱. 莫高窟档案数据及壁画图像的存储与检索[D]. 中国科学院研究生院(计算技术研究所), 2000. (Liang Guanyu. Storage and retrieval for the archive data and fresco image of MoGao Grotto[D]. Institute of Computing Technology of the Chinese Academy of Sciences, 2000.)
- [19] 季羨林. 敦煌学大辞典[M]. 上海: 上海辞书出版社, 1998. (Ji Xianlin. Dunhuang studies dictionary[M]. Shanghai: Shanghai Dictionaries Press, 1998.)
- [20] Bulthoff H H, Lee S W, Poggio T, et al. Biologically motivated computer vision [M]. New York: Springer Publishing Company, 2003: 150-159.
- [21] Ekman P, Friesen W V, Ellsworth P. Emotion in the human face: Guidelines for research and an integration of findings [M]. New York: Pergamon Press, 1972.
- [22] 杨雄. 论敦煌艺术的概念和分类[J]. 前沿, 2011, (23): 230-233. (Yang Xiong. Theory of Dunhuang art concept and classification[J]. Frontier, 2011, (23): 230-233.)
- [23] 郑炳林, 沙武田. 敦煌石窟艺术概论[M]. 甘肃: 甘肃文化出版社, 2005. (Zheng Binglin, Sha Wutian. Introduction to Dunhuang grottoes art[M]. Gansu: Gansu Culture Press, 2005.)
- [24] Cockburn A, Karlson A, Bederson B B. A review of overview + detail, zooming, and focus + context interfaces[J]. ACM CSUR, 2008, 41(1): 1-31.
- [25] 郑汝中, 台建群. 敦煌石窟全集: 飞天画卷[M]. 香港: 商务印书馆, 2002. (Zheng Ruzhong, Tai Jianqun. Dunhuang grottoes corpora: Apsaras picture scroll[M]. Hongkong: The Commercial Press, 2002.)

王晓光 武汉大学信息管理学院教授、博士生导师。通讯地址: 湖北省武汉市珞珈山。邮编: 430072。

徐 雷 武汉大学信息管理学院博士研究生。通讯地址同上。

李 纲 武汉大学信息管理学院教授、博士生导师。通讯地址同上。

(收稿日期: 2013-04-16; 修回日期: 2013-06-29)