

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 1 页

一、(20 分) 概述一个典型的图像检索系统的基本框架。

QBIC(Query By Image Content)图像检索系统是 IBM 公司 90 年代开发制作的图像和动态景象检索系统, 是第一个基于内容的商业化的图像检索系统。QBIC 系统提供了多种的查询方式, 包括: 利用标准范图(系统自身提供)检索, 用户绘制简图或扫描输入图像进行检索, 选择色彩或结构查询方式, 用户输入动态影像片段和前景中运动的对象检索。在用户输入图像、简图或影像片段时, QBIC 对输入的查询图像进行颜色、纹理、形状等特征进行分析和抽取, 然后根据用户选择的查询方式分别进行不同的处理。QBIC 中使用的颜色特征有色彩百分比、色彩位置分布等; 使用的纹理特征是根据 Tamura 提出的纹理表示的一种改进, 即结合了粗糙度、对比度和方向性的特性; 使用的形状特征有面积、圆形度、偏心度、主轴偏向和一组代数矩不变量。QBIC 还是少数几个考虑了高维特征索引的系统之一。QBIC 除了上面的基于内容特性的检索, 还辅以文本查询手段。例如为旧金山现代艺术博物馆的每幅作品给予标准描述信息: 作者、标题、日期, 许多作品还有内容的自然描述。QBIC 系统的结构图如下图 1-1 所示:

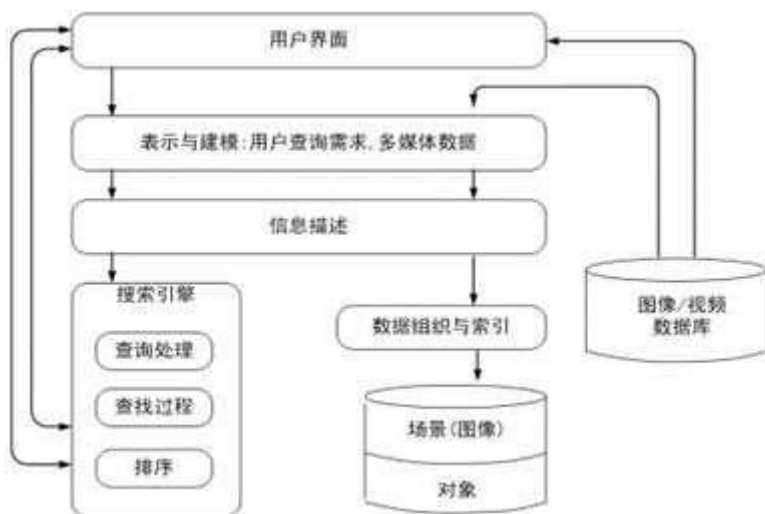


图 1-1 QBIC 系统基本框架

在 QBIC 中, 其数据模型具有以下成分包含对象(图像的子集)的静态图像和场景; 包含运动对象的由一组连续的帧组成的视频镜头。QBIC 的数据模型分成两个层次场景(或图像)和对象。所谓场景就是一幅图像或视频中的单个代表帧。对象是场景的一部分, 或者是视频中的一个运动的实体。特征将从图像和对象中提取出来并存入数据库中。按摄像机的场景变换将视频序列分割成若干片断就是镜头。系统为每一个镜头生成一个代表帧也 R 帧。R 帧将按静态图像进行处理, 其内容特征也要存入数据库中。对镜头的进一步处理可得到运动对象, 例如驶过整个画面的汽车。运动对象是对一系列视频帧中指定对象处理的结果。这样, 查询可以基于对象(寻找红色圆形对象的图像)、场景(寻找大

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 2 页

约有 30% 红色、15% 蓝色的图像)、镜头(寻找所有缓慢转动取全景的镜头)等进行,也可以基于以上三者的任意组合(寻找有 30% 红色、并具有蓝色纹理对象的图像”)进行。在 QBIC 中,相似性是用特征之间的距离函数来表示的。设计这些距离函数时考虑了能够模仿人类的视觉的方法,使之逼近数据库中各个对象的特征。图 1 中的匹配器就是所有距离函数的集合。匹配器与过滤检索模块交互以支持快速搜索方法和检索。

二、(20 分) 概述一个典型的视频检索系统的基本框架。

图像是静态的,视频是动态的,且视频流可以看做是图像帧的序列,典型的视频网站有 YouTube、优酷、土豆等。为了对视频数据库进行基于内容的查询,首要构造便于检索的视频结构,视频数据可分为场景(Scene)、镜头组(Group)、镜头(Shot)、关键帧(Key frame)、帧(Frame)。

一个视频序列可以是多个视频场景来构成,一个视频场景又是由多个镜头构成,而镜头是有帧构成。镜头是指摄像机从打开到关闭的过程中记录下来的一组连续图像帧。镜头边界是客观存在的,可以采用一定的方法自动检测镜头边界。在实际应用中,用户浏览一个镜头中所有图像帧是非常耗时的,因此常用关键帧技术实现快速浏览。关键帧是指代表镜头中最重要的、有代表性的一幅或多幅图像。依据镜头内容的复杂程度,可以从一个镜头中提取一个或多个关键帧或构造一个关键帧。为了在语义层建立视频结构模型,需要对视频进行场景划分。场景定义为语义上相关、时间上相邻的一组镜头,它们能够表达视频的高层次概念或故事等。镜头是组成视频的基本物理单位,而场景(又称故事)则是视频在语义层的单位,通常只有场景才能向观看者传达相对完整的语义。镜头组是一组在时间上相邻并在内容上相似的一组镜头,它是介于镜头和场景之间的一组连续的物理实体,是联系镜头和场景的桥梁。节目则是由时间上有序的场景组成,例如新闻节目、娱乐节目、体育节目、天气预报等。

视频结构化分析是指对视频流进行镜头分割、关键帧提取和场景分割等处理,从而得到视频的结构化信息。镜头分割的关键在于确定镜头的边界,现有镜头分割方法多以视频内容的不连续性为划分镜头的依据。研究者们通常选取视频的某种特征来度量视频内容的不连续性,如颜色特征、运动矢量特征、边缘特征等。渐变镜头边界的检测目前仍然是一个具有挑战性的课题。由于同一个镜头中的各帧图像之间的内容有相当程度的冗余,因此可以选取反映镜头中主要信息内容的帧图像作为关键帧。镜头分割后,对每个镜头可提取若干关键帧,并用关键帧来简洁地表示镜头。

场景分割目标在于获取视频的最小语义结构单元——场景。一般而言,场景是由一组连续的、同属于一个故事单元的多个镜头组成。通过融合视频的文本、声音等信息对已分割出的镜头进行聚类,将内容相近的连续镜头合并为一个单元组,从而得到场景信息,为进一步进行视频内容分析提供基础。图 2-1 给出了典型的 CBVR 系统的基本框架。

在基于内容的视频检索中,一般首先将视频序列分割为镜头,并在镜头内选择关

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 3 页

键帧，然后提取镜头的动态特征与关键帧的静态特征并存入视频数据库，根据这些特征进行视频的索引。最终，用户可以通过一种简单方便的方法浏览和检索视频。下图内容的视频检索系统框图。

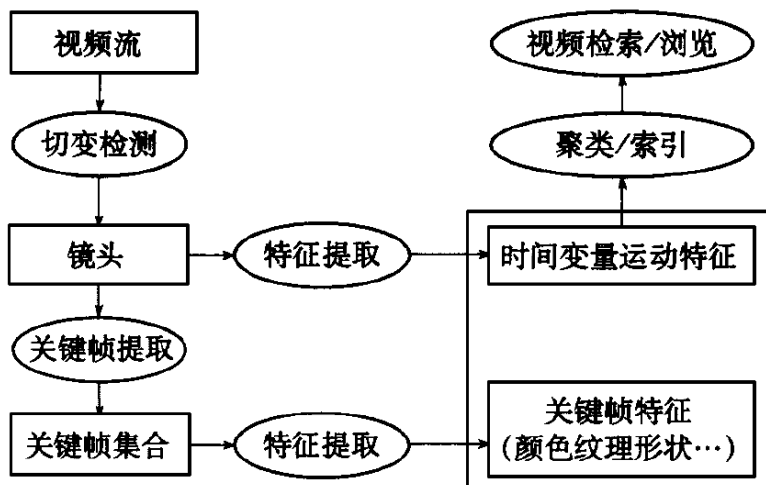


图 2-1 CBVR 系统基本框架

基于内容的视频检索（CBVR）是根据视频的内容和上下文关系,对大规模视频数据库中的视频数据进行检索。它提供一种算法在没有人工参与的情况下,自动提取并描述视频的特征和内容。目前基于内容的视频检索研究,除了识别和描述图像的颜色、纹理形状和空间关系外,主要的研究集中在视频分割,特征提取和描述(包括,视觉特征、颜色纹理和形状及运动信息和对象信息等)关键帧提取和结构分析等方面。

在视频检索系统中,需要采用一些具有代表性的关键帧(或关键帧的序列)来索引场景或者整个视频,而这种自动索引是建立在识别、区分有意义的视频段的基础上,也即视频结构化。简单的说,视频结构化就是从视频中检测出拍摄时所遵循的“场景”和镜头等单位,从而把视频中的帧序列按照“视频——场景——镜头——关键帧”的层次结构进行组织。可以分别计算两个视频流在关键帧、镜头和场景定不同层次的相似性,然后总体判断这两个视频之间的相似性。一种最通用的视频层次结构包括帧(关键帧)、镜头、镜头组、场景等一些结构单位。

(1) 镜头的边界检测。镜头是视频数据的基本单元,视频处理首先就需要把视频自动地分割为镜头,以作为基本的索引单元,这一过程就称为镜头边界的检测。它是实现基于内容的视频检索的第一步,其核心处理是识别镜头的切换。镜头切换即一个镜头到另一个镜头的转换。镜头切换时,视频数据会发生一系列的变化,主要表现在颜色差异突然增大、新旧边缘的远离、对象形状的改变和运动的不连续性等方面。现在镜头检测的方法有很多,主要有:像素差法、数值差法、颜色直方图法等等。

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 4 页

(2) 关键帧的提取。关键帧的提取是基于内容的视频检索中一个很重要的环节, 一般来说视频数据流中的图像帧之间存在时间和空间的冗余。所以在表示一个镜头时, 如果把这一个镜头中的每个时刻的图像帧都用上, 那么就使用了太多冗余的图像帧。因此人们需要从一个镜头中提取出“关键图像帧”来代表这个镜头。这对于视频浏览、检索都极为有利。

(3) 特征提取。特征提取是视频检索系统中的关键, 在视频分割成镜头之后需要对各个镜头进行特征抽取, 得到一个尽可能充分反映镜头内容的特征空间, 这个特征空间将作为视频聚类 and 检索的依据。特征提取有镜头层次上的运动特征的提取以及关键帧层次上的静态视觉特征的提取等。

(4) 聚类/索引。特征提取完, 需要对特征进行聚类, 很多种方法都可以用于这里的特征聚类, 例如 K-均值法、ISODATA 法、松弛迭代法、基于关联规则的算法、基于模糊图论聚类法等。

(5) 视频索引/浏览。视频索引是将视频内容分析, 视频结构分解和摘要的过程中提取的结构和内容属性归入小同类别或进行索引结构的聚类建立的属性表。视频浏览是视频检索系统中交互查询的一个组成部分。为了有效地浏览, 视频文档的内容应表示成用户易于理解的静态画面的形式。并且必须提供非线性的访问, 即应组织成能反映视频语义的结构。

三、(20 分) 解释颜色特征、纹理特征、形状特征, 如何使用这三种特征实现图像检索?

颜色特征:

颜色特征是人类感知和区分不同物体的一种基本视觉特征, 世界上每一物体都有其固有的颜色特征, 颜色特征非常稳定, 对于旋转、平移、尺度变化, 甚至各种变形都不敏感, 表现出相当强的鲁棒性。在图像的自动分析中, 颜色是一种能简化目标提取和分类的重要描述符。常见的颜色特征提取方法有颜色直方图、颜色索引、颜色矩等。颜色直方图颜色特征的基础, 是对图像中不同颜色像素数量的统计。颜色索引是一种以颜色直方图和直方图相交算法为基础的颜色特征表示方法, 该方法通过将颜色空间划分为多个固定的子空间, 然后统计每个固定的子空间中像素数目, 使用与直方图相交的策略得到特征描述。该方法除了具有颜色特征通用的优点外, 还包含突出优点是计算简单, 但它存在和颜色直方图一样没办法反映出颜色的空间分布信息的缺点。颜色矩方法适用于图像中有较大颜色差别的图像库。由于颜色分布信息主要集中在低阶颜色矩中, 因此仅采用颜色的一阶中心矩、二阶中心矩和三阶中心矩就可以表达图像的颜色特征, 它们分别表示图像的平均颜色、标准方差和三次根非对称性。

纹理特征

纹理定义成对图像的像素灰度级在空间上的分布模式的描述, 反映物体的质地, 如粗糙度、光滑性、颗粒度、随机性和规范性纹理可认为是灰度在空间以一定的形式变化而产生的图案, 是真实图像区域固有的特征之一。纹理的描述技术大体上可以分为统计方法、结构方法、频谱方法三类。

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 5 页

统计方法是最早的纹理描述方法之一。它又分为空间域方法和变换域方法, 包括自相关函数、纹理边缘、结构元素、灰度共生矩阵、灰度行程和自回归模型。统计方法将纹理描述为光滑、粗糙、粒状等特征。

结构方法的基本思想是认为复杂的纹理可以由一些简单的纹理基元以一定的有规律的形式重复排列组合而成。结构法研究的主要内容是纹理基元及其空间关系。基元一般定义为具有某种属性而彼此相连的单元的集合, 属性包括灰度、连同区域的形状、局部一致性等。空间关系包括基元的相邻性、在一定角度范围内的最近距离等等。根据基元间的空间关系, 纹理可以分为弱纹理或强纹理。

频谱法。

频谱方法的典型是对图像进行傅立叶变换, 从傅立叶频谱成分的分布中来求得纹理特征。频谱分析技术是用区域自相关函数或傅立叶变换域的能量分布来检测纹理的周期, 包括计算峰值处的面积、峰值处的相位、峰值与原点的距离平方、两个峰值间的相角差等手段。

形状特征

形状特征通常与图像中的特定目标对象有关, 是人们的视觉系统对目标的最初认识, 被认为是比颜色特征和纹理特征更高一层的特征。形状描述的准确与否决定图像检索算法优劣, 一个好的形状描述符应具备独特性、完备性、几何不变性、灵活性以及抽象性。形状的描述符大体可以分为两大类: 第一类是描述形状目标区域边界轮廓的像素集合, 称为基于轮廓的形状描述符, 如链码、傅里叶描述子、小波变换、边界长度等; 第二类称为基于区域的形状描述符, 是对形状目标区域内所有像素集合的描述如区域统计, 偏心率、几何矩、形状描述矩等;

一般的图像检索系统多采用多特征进行检索任务, 每增加一个特征信息就会缩减待检索空间。使用这三类特征进行检索, 遵循如下过程: 首先是系统先根据数据库内所有图像提取对应的颜色、纹理、形状特征, 并与图像相连接存储于数据库中, 当用户输入一张待检索图像时, 即对其进行特征提取, 然后按不同权重因子进行特征融合, 并于数据库进行比对, 系统按特征向量之间的间距大小对相似度进行判断, 然后将一系列结果按相似顺序掉出。一般图像检索系统都遵循这一模式。

四、(10 分) 什么是相关反馈? 什么是基于语义的检索?

相关反馈是指通过一种人际交互机制能够使计算机不断了解用户对查询结果的满意程度, 并通过不断学习把输出调整到符合用户期望的状态。用户主观感知和语义间存在隔阂, 并没有一种固定的模式的高层语义映射到图像底层来表达用户的查询概念。于是人们试图让用户参与到检索过程中来, 这是弥合高层语义和低层特征之间语义间隔的一个重要途径。相关反馈一般使用示例查询。示例查询有两种方式, 一是示例图像查询, 二是示例草图查询。前者运用图像的全部特征, 而后者运用图像的局部特征。实际上, 一幅示例图像所提取出来的特征很难包含人们所期望的全部特征, 且同时包含了一些未

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 6 页

被期望的特征。所以在系统进行初次检索后, 用户可从返回结果中选择接近期望的图像作为正例样本, 也可以选出与期望不符的反例样本, 这些信息反馈给系统使系统可以调整搜索方向, 检索出符合用户要求的图像。

基于语义的图像检索方法是人们为了克服基于简单视觉特征的图像检索方法的不足而提出的, 与基于低层物理特征查询不同“语义特征查询是基于文字的查询”包含了自然语言处理和传统图像检索技术。这种检索方法的目标是最大限度地减小图像简单视觉特征与丰富的语义之间的语义鸿沟, 缩小语义鸿沟的办法有(种)由高层语义导出低层特征和由低层特征向高层语义的转换图像语义具有模糊性、复杂性、抽象性, 一般包括 3 个语义层次: 特征语义、目标和空间关系语义、高层语义。特征语义就是图像的颜色、形状、纹理等低级视觉特征, 与视觉感知直接相连; 目标语义和空间关系语义需要识别和提取图像中的目标类别、目标之间的空间位置等关系, 涉及到模式识别和逻辑推理的相关技术; 高层语义主要涉及图像的场景语义(如海滨、街道、室内等)、行为语义(如表演、超越、进攻等)和情感语义(如平静、和谐、振奋等)。一般而言, 高层的图像语义往往建立在较低层次的语义获得的基础上, 并且层次越高, 语义越复杂, 涉及的领域知识越多。

基于语义的图像检索一般指的是基于目标和高级语义的图像检索方法。在原有检索系统中加入高级语义到低层特征的转化, 可以在不改变现有的图像特征库和匹配方式的情况下, 实现基于语义的图像检索。

### 五、(20 分) 如何评价一个图像检索系统或视频检索系统的性能?

图像检索策略性能评估的比较通用的几个准则是: 有效性 (effectiveness)、效率 (efficiency)、灵活性 (flexibility)。它们分别代表查找与示例图像相似的图像的成功率、检索的速度及对不同应用的适应性。目前对图像检索效果的评价则更多地放在检索结果的正确与否上, 即检索的有效性上。可以从以下 3 个方面来综述基于内容的图像检索系统的性能评价问题: 测试数据集的建立、量化的评价方法和相关性评价的获取。

1. 测试数据集的建立。对于图像检索领域, 一个具有相当规模的, 平衡的测试数据集尚有待建立。建立这样的测试数据集将有助于更好地评价系统所提取图像特征的有效性和系统的整体性能。这种测试库中的图像应该可以免费获取, 且没有版权限制, 不仅库中图像应该具有充分的多样性, 而且图像集要代表图像检索的实际应用。目前, 一些研究小组虽做了这方面的尝试, 但还没有一个大家公认测试图库, 其中使用最多的图像集有美国加州水资源图像集和 Corel 公司的 22000 幅专业收藏照片、Brodatz 纹理图像以及 Vistex 纹理数据库等。还有一些研究机构也自己建立了数据库, 但图像数量还不够。由于统一标准的测试数据集的建立有利于对新的图像检索策略进行统一的评价, 因而会加速新的更有效的检索策略的产生, 同时更便于研究者之间的交流与合作。

2. 量化的评价方法。有效性评价可以从以下两个方面来考虑: 一是人的主观感受; 二是有一个量化的评价标准, 但是人的主观感受由于受其主观性和个体性影响, 很不容

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 7 页

易把握, 因此量化的评价标准是一种较直观, 且具有通用性的检索有效性的评价方法。量化评价方法有多种, 我们简述其中两种:

第一种: 匹配百分数。有一种实用的判别方法是从  $512 \times 512$  大小的中选取两块不重叠的  $256 \times 256$  大小的子图像, 并取其中一幅作为查询图像, 另一幅作为目标图像。理想的检索情况下, 目标图像应该在检索结果中排位第一。由于在实际检索中, 很难要求检索算法每次都能把目标图像提取在第 1 个位置, 所以衡量检索结果优劣的匹配百分数  $M$  定义为

$$M = \frac{N - S}{N - 1} \times 100\%$$

其中,  $N$  是数据库中目标图像的总数,  $S$  代表另一幅图像在检索结果中的排位。如果进行多次试验, 并取平均匹配百分数, 则  $M$  反映了检索结果的优劣。该方法虽简便易于计算, 但是由于仅考虑了一幅相关图像的检索情况, 而未顾及更多相关图像, 因此对于需要较高查全率的应用并不适应。

第二种: 相似性排序百分比。相似性排序百分比(percentag e of similarity ranking, PSR)是由 Bimbo 和 Pala 提出的。该方法首先得到相似性评判矩  $Q_j^{(i,k)}$ , 其中  $Q_j^{(i,k)}$  代表对于查询图像  $j$ , 图像  $i$  排在第  $k$  位的人数; 然后根据该矩阵计算图像  $i$  每一行的均值  $\overline{P_j^{(i)}}$  和方差  $\sigma_j^{(i)}$  设对于查询图像  $j$ , 图像  $i$  的返回位置是  $p_j^{(i)}$ , 则相似性排序百分比  $E_j^{(i)}$  的计算如下:

$$E_j^{(i)} = \sum_{k = p_j^{(i)} - \frac{\sigma_j^{(i)}}{2}}^{p_j^{(i)} + \frac{\sigma_j^{(i)}}{2}} Q_j^{(i,k)}$$

$E_j^{(i)}$  值越高, 代表检索算法的准确率越高。

3. 相关性评价的获取。借助其内容已经标注的图像数据库是帮助评价的一种方法。如借助 Corel 图像数据库, 该数据库已将图像分为不同的类别, 如斑马类、天空类等, 这样就可以认定具有相同主题的图像是相关的, 不同主题的是不相关的。由于此类评价往往不是根据整幅图像的全局特征, 而是根据图像中的对象来确定相关性, 因此用户还可以在已知分类的基础上将视觉上并不相似的图像从某分类中去除, 这样更有助于性能评价的准确性。这种评价方法也存在如下一些困难: 首先, 根据标注来确定相关性可能并不真实; 其次, 为了得到有用评价, 需要对图像内容进行全面的标注, 这就带来工作量大和复杂性高的问题。用户确定相关性的方法还有以下 3 种, 分别适不同的量化评价方法:

(1) 预先制定一个阈值, 对于每一测试查询图像, 由若干用户确定图库中的图像与

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 8 页

示例图像相关与否, 若评判该图为相关图像的用户人数大于给定阈值, 则确定该图为相关图像。最终, 图库被分为相关图像和不相关图像两部分。

(2) 类似于以上方法, 用户依旧对图库中图像判其与测试查询图像是否相关。对于每一幅图, 则首先统计评判该图为相关图像的用户个数, 然后将归一化的该统计值作为该图像相关性权值。

(3) 由用户对每个测试查询图像依据相关性给图库中图像进行排序, 其中最相似的赋以序号 1, 次相似的为 2, 依此类推。这种评判方法的结果将产生一个矩阵  $Q$ 。 $Q(i, k)$  代表将第  $i$  幅图像排在第  $k$  位的评判者的个数。获取图像之间的相关性是进行性能评价的一个难点。如果图像库中有  $N$  幅图像和  $M$  个基准查询图像, 则需要  $M \times N$  次观察才能得到它们的相关性, 可见这种观察的工作量是巨大的。

四、(40 分) 不少于 2000 字的图像/视频检索系统的实验报告。

### 一. 实验背景

近几十年来, 随着多媒体技术的快速发展, 数字图像数据呈现出几何式的增长趋势, 如何从海量图像中快速定位出具备特定信息的图像受到了越来越多研究者的关注。从 20 世纪 70 年代开始, 有关图像检索的研究就已开始, 当时主要是基于文本的图像检索技术 (Text-based Image Retrieval, 简称 TBIR)。到 90 年代以后, 出现了对图像的内容语义, 如图像的颜色、纹理、布局等进行分析和检索的图像检索技术, 即基于内容的图像检索 (Content-based Image Retrieval, 简称 CBIR) 技术。之后随着计算机和网络技术得到广泛应用之后, CBIR 技术取才得了较大的突破, 从当前的 CBIR 的研究的热点和未来的发展趋势的来看, 可大致将其研究分为三层:

1. 最低层次的检索, 在这个层次主要是利用全局特征 (例如: 颜色、纹理、形状等底层特征) 去描述图像的内容, 进行检索, 这也是最直接的层面;

2. 较高层次的检索, 该层次主要是利用图像对象语义来描述图像的内容, 进行检索, 这种技术是底层特征的基础上加入了图像语义特征;

3. 最高层次的检索, 该层次是利用图像概念级语义来描述图像的内容, 进行检索, 该技术是建立在图像对象语义特征提取的基础上加入了图像的概念级语义 (例如: 用户对图像的理解、情感等高层语义)。

尽管对 CBIR 技术的研究已经经过了很长时间, 但目前较为常用的 CBIR 技术基本上都是基于底层特征的检索。现有的 CBIR 底层特征的检索系统主要包括基于颜色的图像检索系统、基于纹理的图像检索系统和基于形状的图像检索系统, 下面就对上述三种检索系统进行一下简单的概述。

### (1). 基于颜色特征的图像检索

颜色特征表示方法主要包括颜色索引、颜色矩等。颜色索引是一种以颜色直方图和直方图相交算法为基础的颜色特征表示方法, 该方法在上述基础上, 首先, 将颜色空间划分为多个固定的子空间; 其次, 统计每个固定的子空间中像素的数目; 最后, 用直



# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 9 页

方图相交算法计算图像之间的相似度，按照相似度的大小对检索结果进行输出。该方法除了具有颜色特征通用的优点外，还有个突出优点是计算简单。但也存在两个不足之处：一是它没有办法反映出颜色的空间分布信息，例如两幅图像中对象空间分布的位置不同时，这两幅图像可能具有几乎完全一致的颜色直方图。二是量化方法的不一样也可能对检索结果也可能产生很大的影响，比如原来两幅图像的颜色直方图很相似，但是由于量化过大，使得某些颜色消失，最终得到的两幅图像的颜色直方图大相径庭，这样造成了图像检索的准确度大大折扣。

### (2) 基于纹理特征的图像检索

和颜色一样，纹理是也是图像中一种不可或缺的视觉特性。纹理没有统一的定义，一般可以用图像的某种局部性质来对纹理下定义，纹理描述了局部区域中像素之间的关系，同时也描述了图像中的空间分布信息。对纹理可以通过粗糙度、方向性、对比度和规则性四个方面来描述。应用较为广泛的基于纹理特征的图像检索技术包括采用共生矩阵，和基于数学模型如小波变换的纹理特征检索。

### (3) 基于形状特征的图像检索

形状描述的是图像中各种物体的外在特征，所以基于形状的检索系统中最关键的技术是如何表示图像中物体的特有的外在特征和如何对提取出的形状特征进行特征匹配，目前常用的描述形状特征的方法主要包括：矩描述法、边界描述法和几何参数法。以前表示形状特征的常用方法有链码、曲线、傅里叶描述子和 B 样条曲线等。后来又出现了小波描述逼近法和超二次曲线法。形状特征的提取需要人工对图像进行分割、提取目标，就目前的发展状况，还没有实现对形状特征的自动提取。而人眼对颜色和纹理特征比较敏感，提取方法也相对比较简单，所以在目前的检索系统中使用最多的底层特征是颜色和纹理特征。

## 二. 实验方法

为了结合各个底层特征的优点，本实验同时提取图像的颜色、形状和纹理三种特征结合为一个综合特征，并实现结合权重的自定义。

在颜色特征方面，为减小运算量，采用颜色矩作为检索的特征，分别提取被检测图像和数据库中图像的一、二、三阶矩。由于 HSV 颜色空间的色调、饱和度、明度与人眼对颜色的主观认识相对比较符合，与其他颜色空间相比 HSV 空间能更好的反映人类对颜色的感知，所以本文采用 HSV 颜色空间下的颜色矩作为颜色特征。使用曼哈顿距离进行颜色矩相似度度量。颜色信息集中在图像颜色的低阶矩中，主要是每种颜色分量的一阶矩(均值)、二阶矩(方差)、三阶矩(斜度)，可以有效地表示图像中的颜色分布。三个颜色矩的数学定义如下：

$$\mu_i = \frac{1}{N} \sum_{j=1}^N f_{ij}$$

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 10 页

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2}$$
$$\xi = \sqrt{\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^3}$$

其中  $f_{ij}$  表示像素  $j$  的颜色值为  $i$  的概率,  $N$  为图像中像素点的个数。若用  $Q, I$  表示两幅图像, 则它们的矩的距离公式为:

$$D[Q, I] = \sum_{i=1}^3 (\omega_{i1} |\mu_i^Q - \mu_i^I| + \omega_{i2} |\sigma_i^Q - \sigma_i^I| + \omega_{i3} |\xi_i^Q - \xi_i^I|)$$

其中,  $\omega_{ij} \geq 0 (1 \leq i, j \leq 3)$  是指定的加权, 一般要设置为  $\omega_{11} \geq \omega_{12}$ ,  $\omega_{11} \geq \omega_{13}$ , 提高了平均颜色的比重。提取到待匹配图和数据库图的颜色矩后, 利用曼哈顿特征进行度量, 即可得到颜色特征的相似性向量。

在形状特征方面, 利用目标区域所占的区域的矩作为形状描述参数是常用的形状描述方法。 $Hu$  矩不变量由于具有不随图像的位置、大小和方向而变化的特点, 对于提取

图像中的形态特征来说, 是一个非常有效的工具。区域  $f(x, y)$  的  $(p+q)$  阶矩定义为:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad p, q = 0, 1, 2, \dots$$

其相应的中心矩定义为:

$$u_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad p, q = 0, 1, 2, \dots$$

其中,  $\bar{x} = \frac{m_{10}}{m_{00}}$ ,  $\bar{y} = \frac{m_{01}}{m_{00}}$ , 即重心坐标。 $(\bar{x}, \bar{y})$  是目标区域灰度质心。 $f(x, y)$  的归一化

$(p+q)$  阶中心矩定义为:

$$\eta_{pq} = \frac{u_{pq}}{u_{00}^\gamma} \quad p, q = 0, 1, 2, \dots$$

其中

$$\gamma = \frac{p+q}{2} + 1 \quad p, q = 2, 3, 4, \dots$$

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上，否则无效。

共 1 页 第 11 页

下列 7 个二维不变  $Hu$  矩是由归一化的二阶和三阶中心矩得到的。它们对平移、旋转、镜面以及尺度变换具有不变性：

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \left[ (\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right] \\ + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \left[ 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right]$$

$$\phi_6 = (\eta_{20} - \eta_{02}) \left[ (\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] \\ + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} - \eta_{02})(\eta_{30} + \eta_{12}) \left[ (\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right] \\ + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03}) \left[ 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right]$$

此组不变矩不受平移、旋转以及大小比例改变的影响。提取待匹配图片和数据库图片的  $Hu$  不变矩特征后，再利用曼哈顿距离进行相似性度量，即可得到形状特征的相似性向量。

在纹理方面，由于图像检索的数据库一般比较大，如果直接对整幅图的纹理特征进行提取和分析，运算量将会非常大。所以本实验先对图像进行分割，然后再对分割后的图像进行基于傅立叶描述子的纹理检测和纹理特征提取，以减少运算量。通常认为，图像特征主要体现在大连通域处，我们基于这个原理进行分割。首先将 RGB 图像二值化，进行形态学处理包括开运算和闭运算后，进行连通域分析，取最大的连通域作为目标区域进行纹理特征的检测。傅立叶描述子将目标放在频域进行分析，得到的目标频谱中低频分量代表物体的宏观形状属性，而高频分量则可以代表物体的细节特征，某些情况下即纹理特征。我们提取最大连通域的边界，只对边界的有序复坐标做傅立叶描述子分析，得到图像的纹理特征。同样利用度量，即得到纹理特征的相似性向量。

我们综合颜色特征、形状特征、纹理特征三种特征进行检索，设三种特征的相似性向量分别为：c，s和t，且在检索时分别赋予权重 $w_c$ 、 $w_s$ 和 $w_t$ ，综合相似性向量M为：

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 12 页

$$M = w_c C + w_s S + w_T T$$

利用  $M$  作为相似度的判据，即可得到图像相似度检索的结果。

### 三．实验结果及分析

利用上面的实验方法，设计了一个 GUI 界面，便于检索。GUI 界面如下：



图 1 检索系统界面

我选择颜色，纹理，形状百分比均为 33.333%，然后进行检索，得到如下检索结果：

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 13 页



图 2 检索结果界面

在界面右下角，我们可以选择显示颜色，形状，纹理等分析的界面，根据原图，我们把分析的结果均显示出来：

基于颜色特性，图片的 H、S、V 分量分别如下：

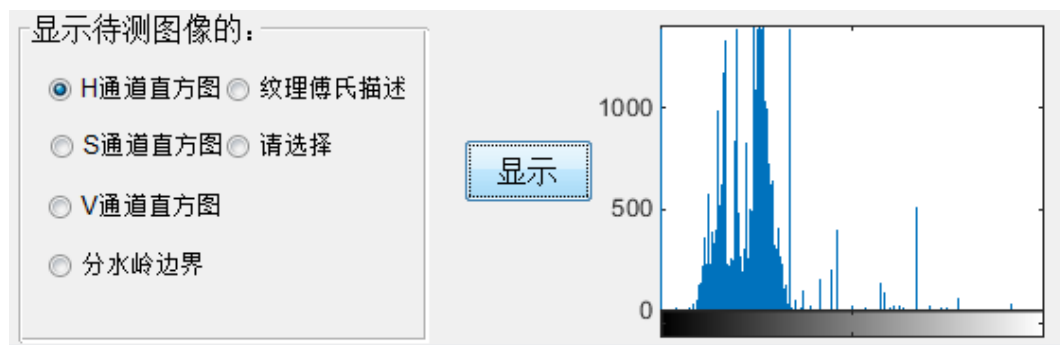


图 3 H 通道直方图

西北工业大学研究生院  
学位研究生课程考试试题

考试科目：基于内容的视觉信息检索  
开课学期：2  
说明：所有答案必须写在答题册上, 否则无效。

课程编号：085012  
考试时间：  
共 1 页 第 14页

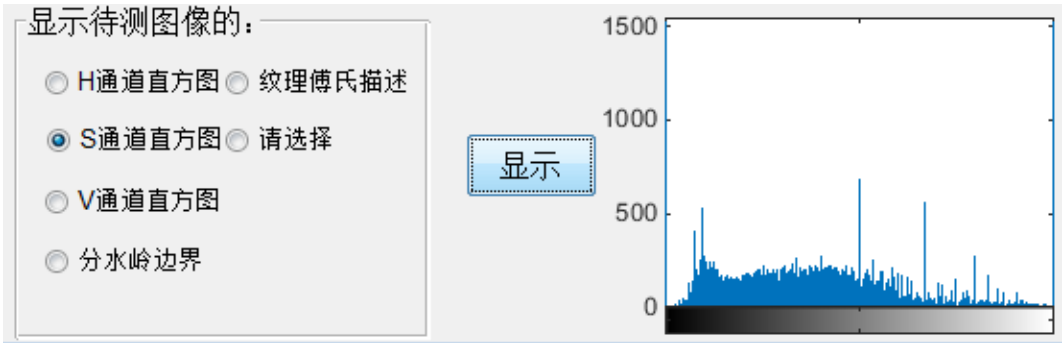


图 4 S 通道直方图

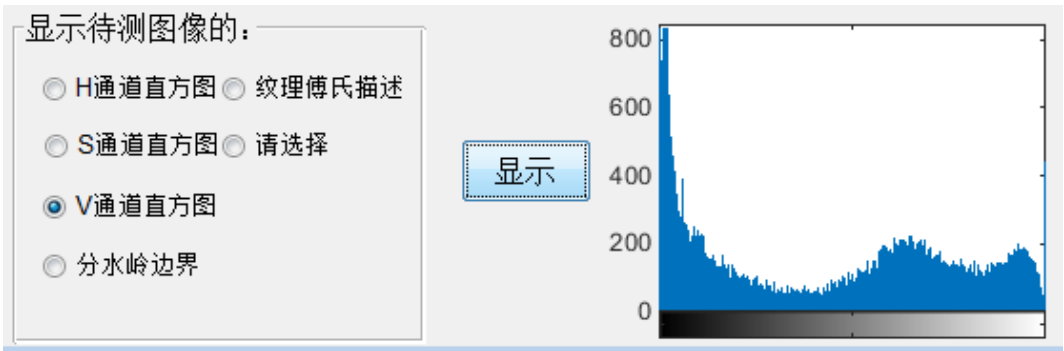


图 5 V 通道直方图

基于形状特性，提取得到分水岭边界图如下：



图 6 分水岭边界图

基于纹理方面，计算得到傅式变换频域的高频部分如下：

# 西北工业大学研究生院

## 学位研究生课程考试试题

考试科目：基于内容的视觉信息检索

课程编号：085012

开课学期：2

考试时间：

说明：所有答案必须写在答题册上, 否则无效。

共 1 页 第 15 页



图 7 纹理傅氏描述图

经过以上检索，我们可以看到检索得到的前两张基本没有差别，但是最后一张颜色相近，形状相差甚远，出现这种情况的原因主要是因为用于检索的数据库图片太少，数据量小，如果数据量多的情况下，检索的效果应该会更好。