



数字媒体技术基础 ——多媒体数据检索

王晗

wanghan@bjfu.edu.cn



专题提纲

- 基于内容的多媒体数据检索
 - 基于内容的图像检索
 - 基于内容的视频检索
 - 基于内容的音频检索
- 视频存储与检索

基于内容的多媒体数据检索

- 从媒体数据中提取特定的信息线索，根据这些线索从存储在数据库的大量媒体中进行查找，检索出具有相似特征的媒体数据。
- 基于内容的检索（Content-Based Retrieval, CBR），根据媒体对象的语义和上下文联系进行检索：
 - 从多媒体内容中提取信息线索
 - 近似匹配
 - 大数据库的快速检索
 - 多种检索手段



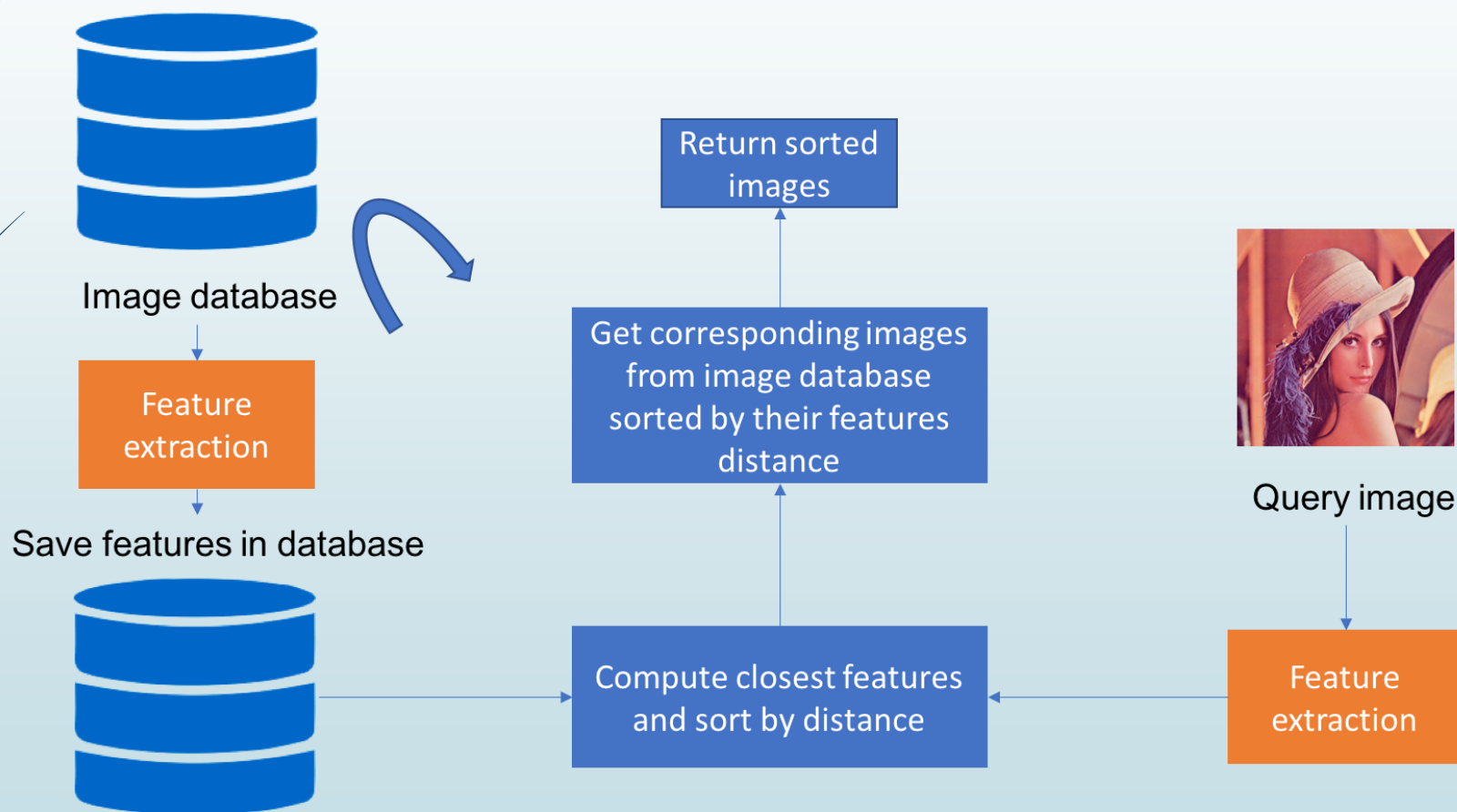
基于内容的多媒体数据检索

- 基于内容的图像检索
(Content-Based Image Retrieval)
- 基于内容的视频检索
(Content-Based Video Retrieval)
- 基于内容的音频检索

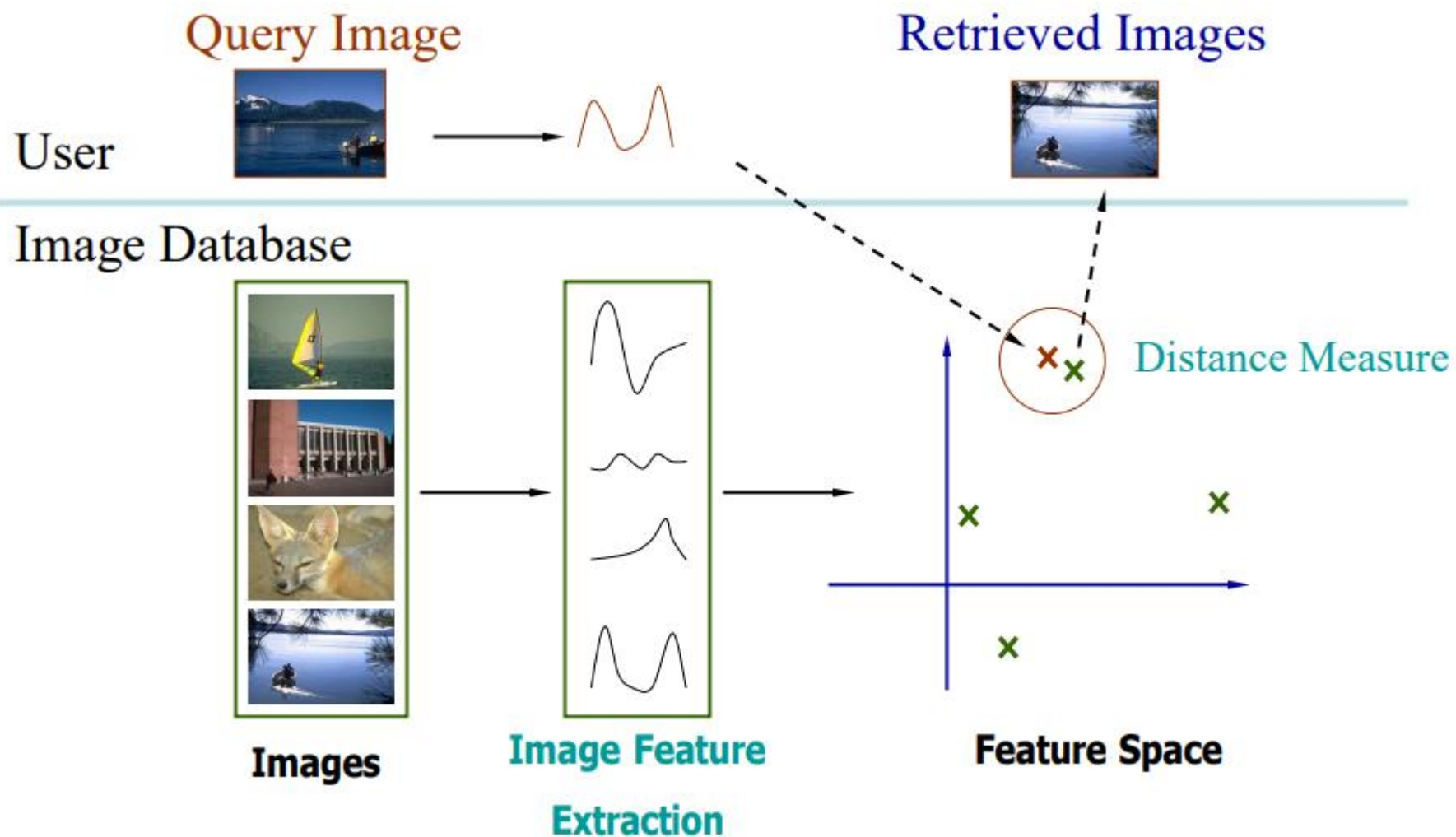
基于内容的图像检索 (CBIR)

- 传统的基于文本的图像检索系统是通过图片的名称、文字信息和索引关系来实现查询。
- 典型的CBIR系统，允许用户输入一张图片，以查找具有相同或相似内容的其他图片。
- 基于内容图像检索这一概念于1992年T.Kato提出，
- 搜索引擎（Google、百度）的以图搜图功能，电商网站（淘宝、Amazon、ebay）相似商品检索，社交平台（Pinterest）的相似内容推荐等。
- 基本思想：根据图像、图像的内容语义以及上下文联系进行查找，以图像语义特征为线索从图像数据库中检出具有相似特性的其它图像，如图像的颜色、纹理、布局等进行分析 and 检索的图像检索技术。

基于内容的图像检索 (CBIR)



基于内容的图像检索 (CBIR)



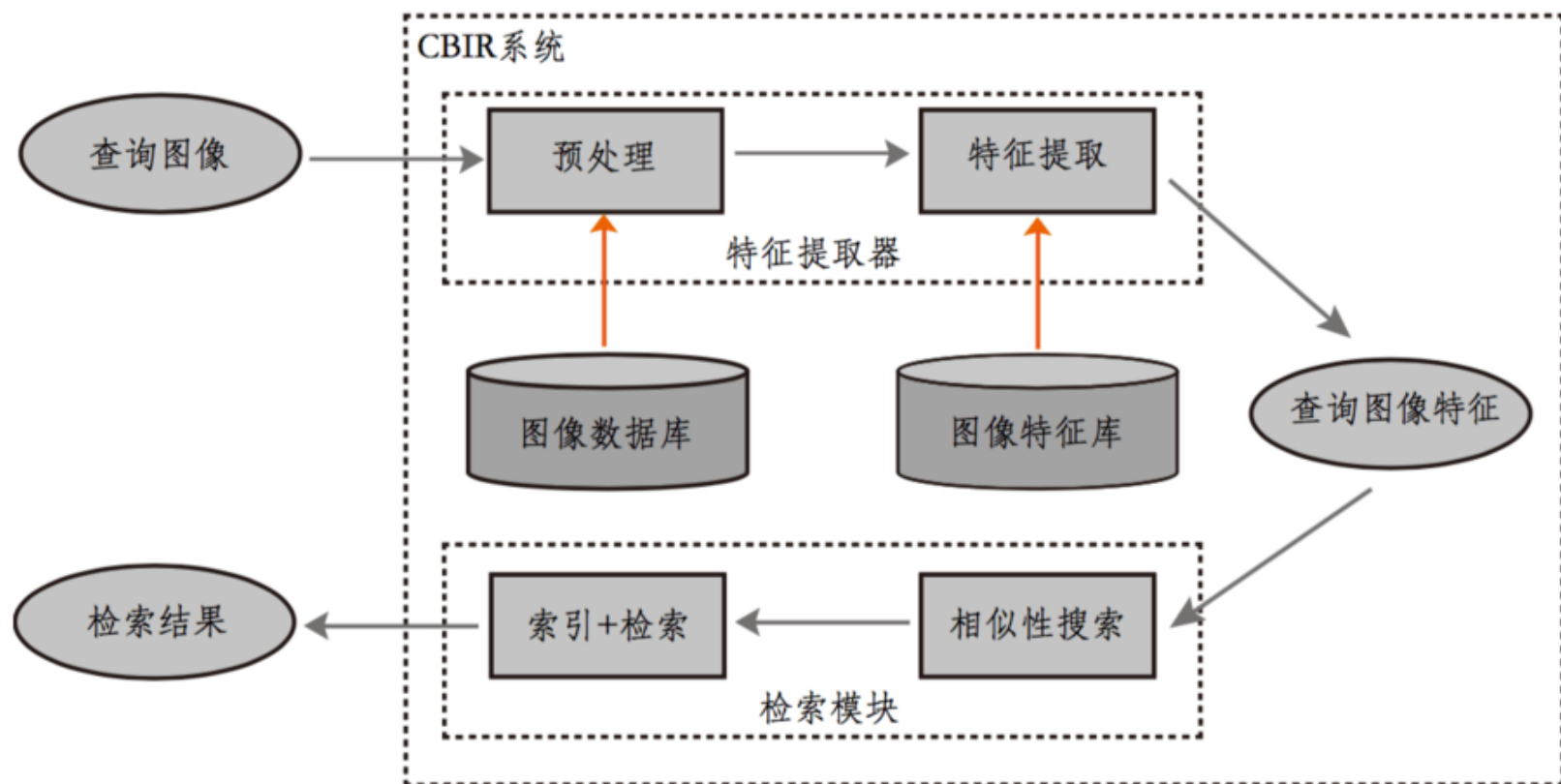


基于内容的图像检索 (CBIR)

- Query By Example
- Semantic retrieval
- Relevance feedback (human interaction)
- Iterative/machine learning
- Other query methods

基于内容的图像检索

以图搜图



基于内容的图像检索

- 相同物体图像检索和相同类别图像检索面临的挑战



(a)光照变化

(b)尺度变化

(c)视角变化

(d)遮挡

(d)背景杂乱

相同物体图像检索面临的挑战



(a)类内变化巨大(湖泊)



(b)类间相似性干扰

相似类别图像检索面临的挑战

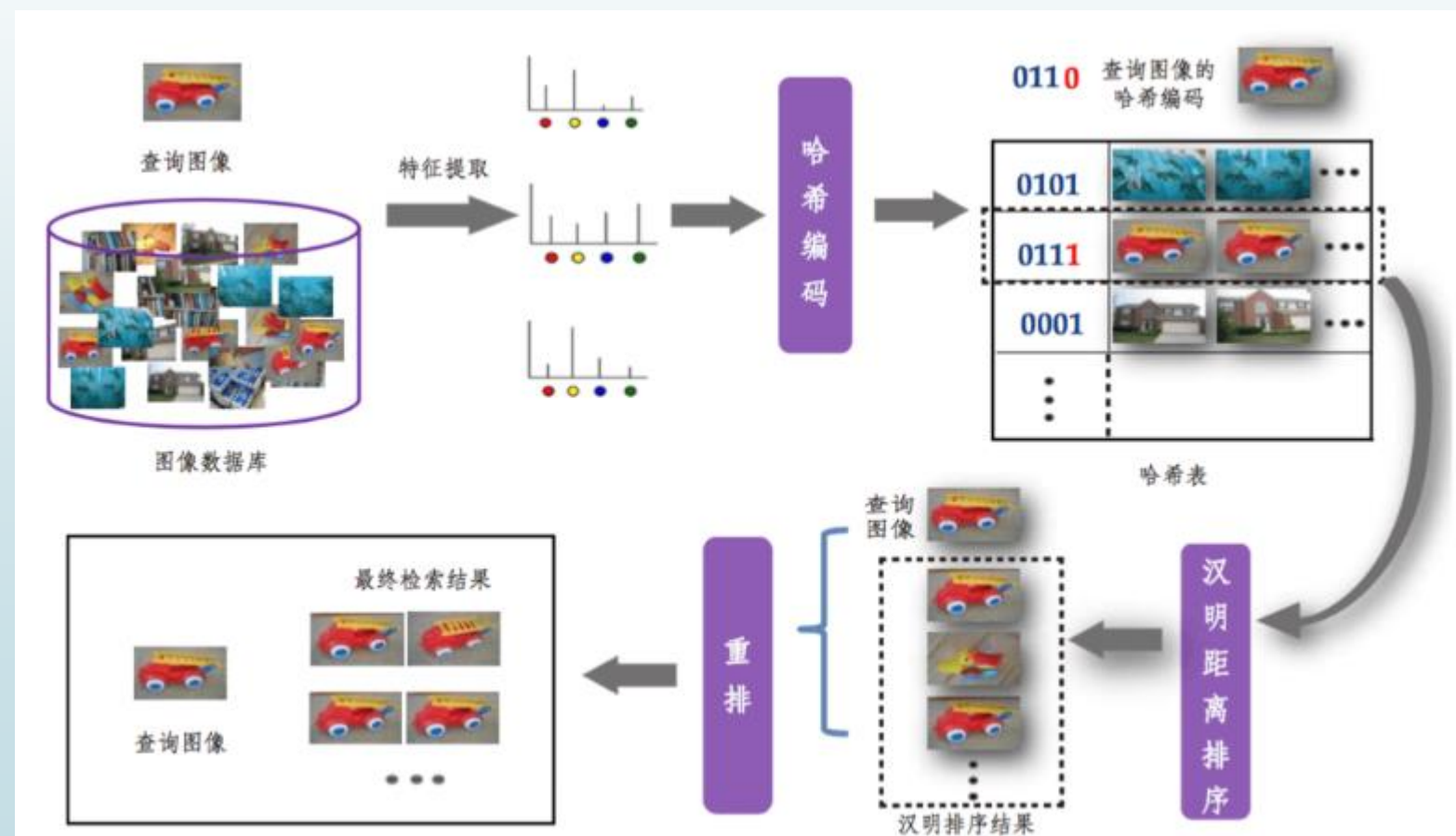
基于内容的图像检索

——Hash

- 什么是Hash
 - Hash也称散列、哈希
 - 基本原理就是把任意程度的输入，通过Hash函数映射成固定长度的输出。
 - 原始数据映射后的二进制串就是哈希值
 - 通过对图像进行Hash编码，不仅能实现时间上的高效（基于XOR操作的快速计算）还能实现存储上的高效（基于位存储的紧致表达）
 - 好的Hash函数一般要求原始相近的数据其哈希编码也要尽可能的相近，而当原始数据发生变化时其Hash编码的差异也要尽可能的大，同时还能抵御各种攻击以保护原始数据。

基于内容的图像检索

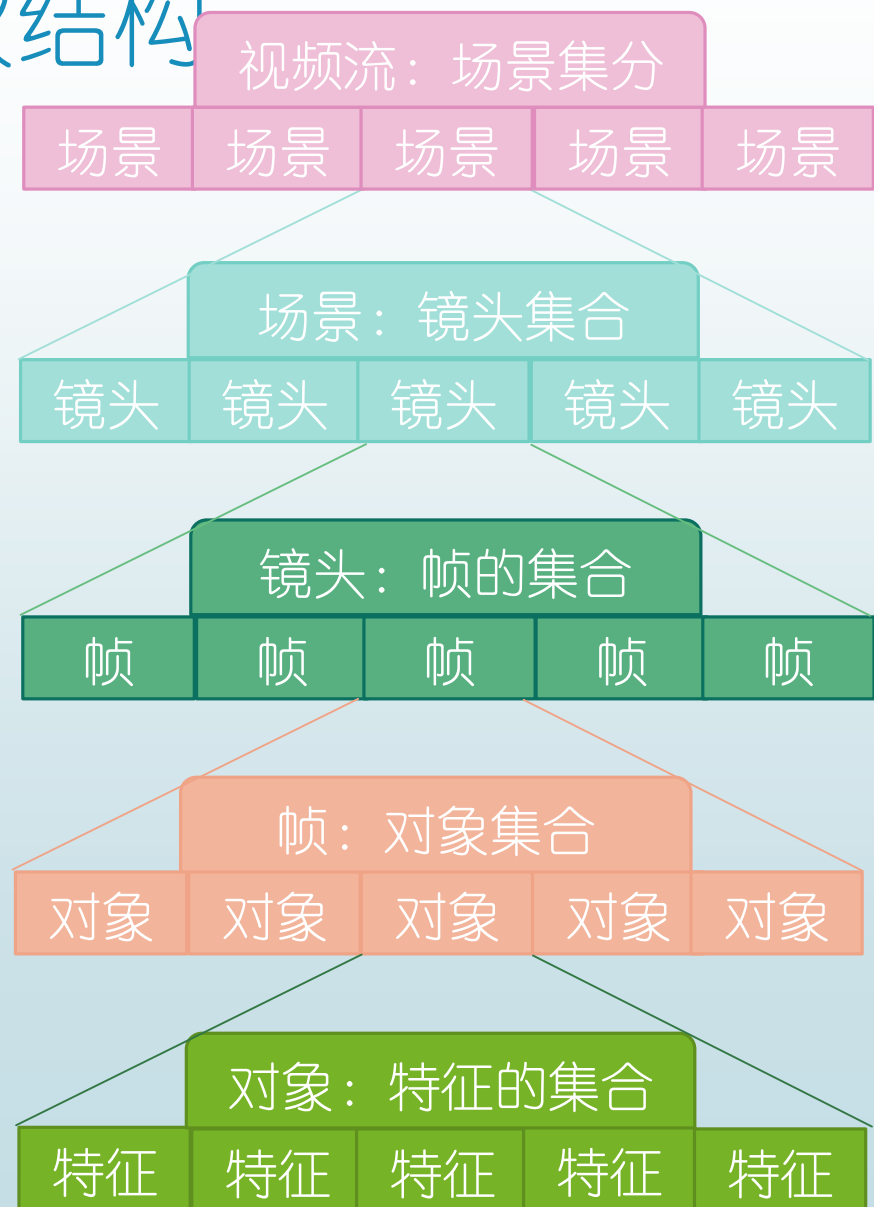
基于哈希的图像检索框架



基于内容的视频检索

——视频的层次结构

- ▶ 右图是一个包含了视频的语义内容，并且可以提供视频内容的时空分析和描述的框架。
- ▶ 视频可由场景，镜头，帧，对象等语义信息进行描述



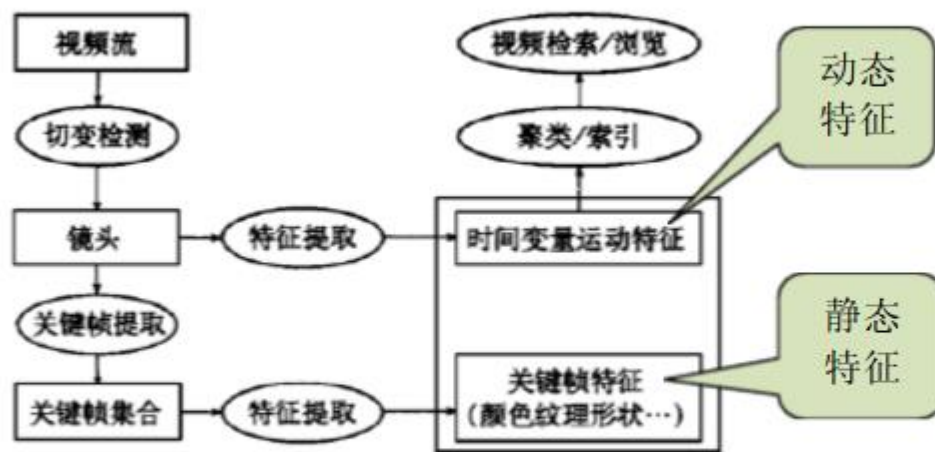
基于内容的视频检索

——视频的层次结构

► 镜头/场景切边和过渡检测

视频检索的基本流程:

结构化分析 → 特征提取 → 语义提取 → 高维索引 → 检索反馈 → 浏览应用

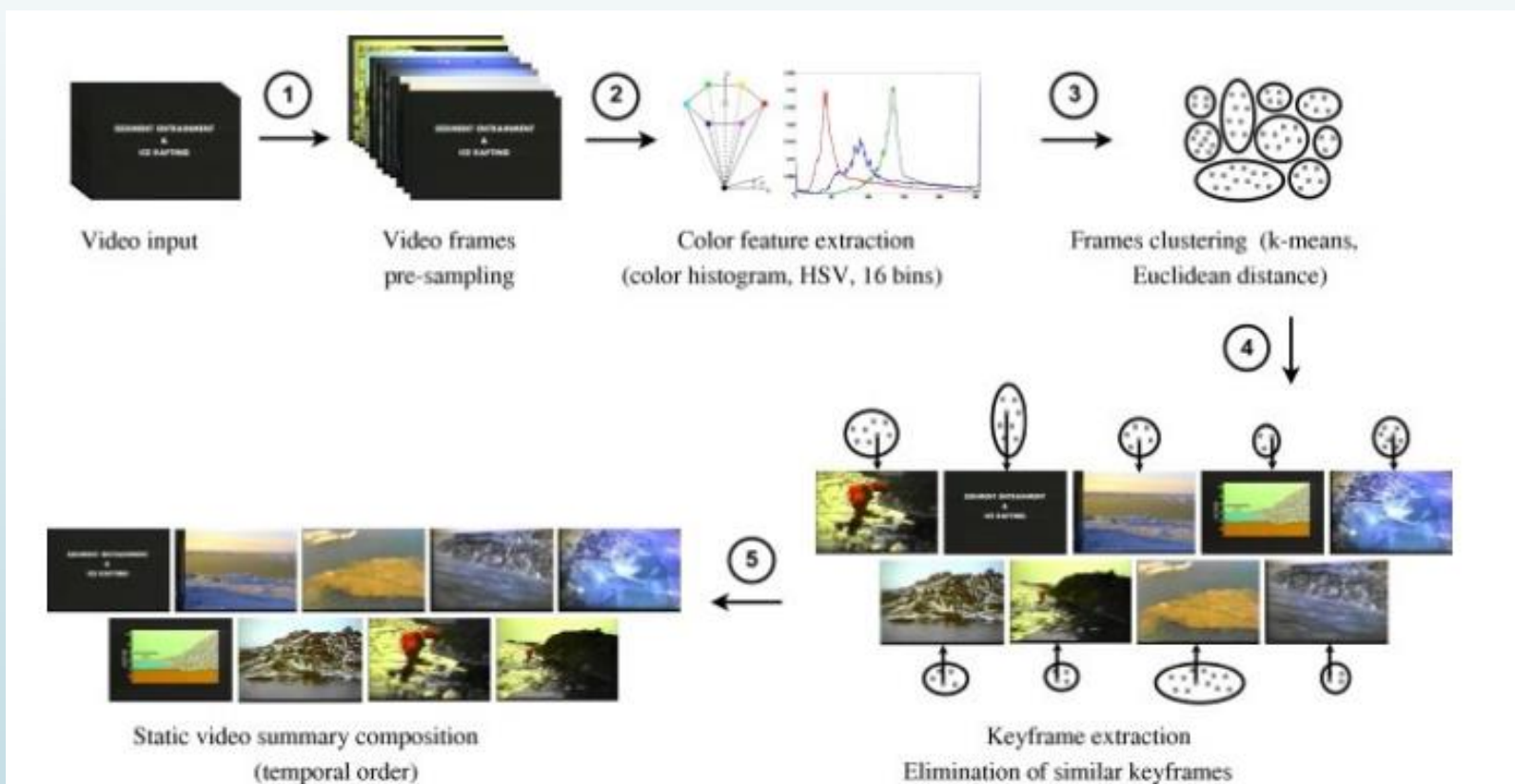


提取镜头的特征及关键帧的视觉特征存入视频数据库。在建库后,利用相似度的测量实现基于内容的检索。

基于内容的视频检索

——视频的层次结构

➤ 关键帧选取和视频摘要

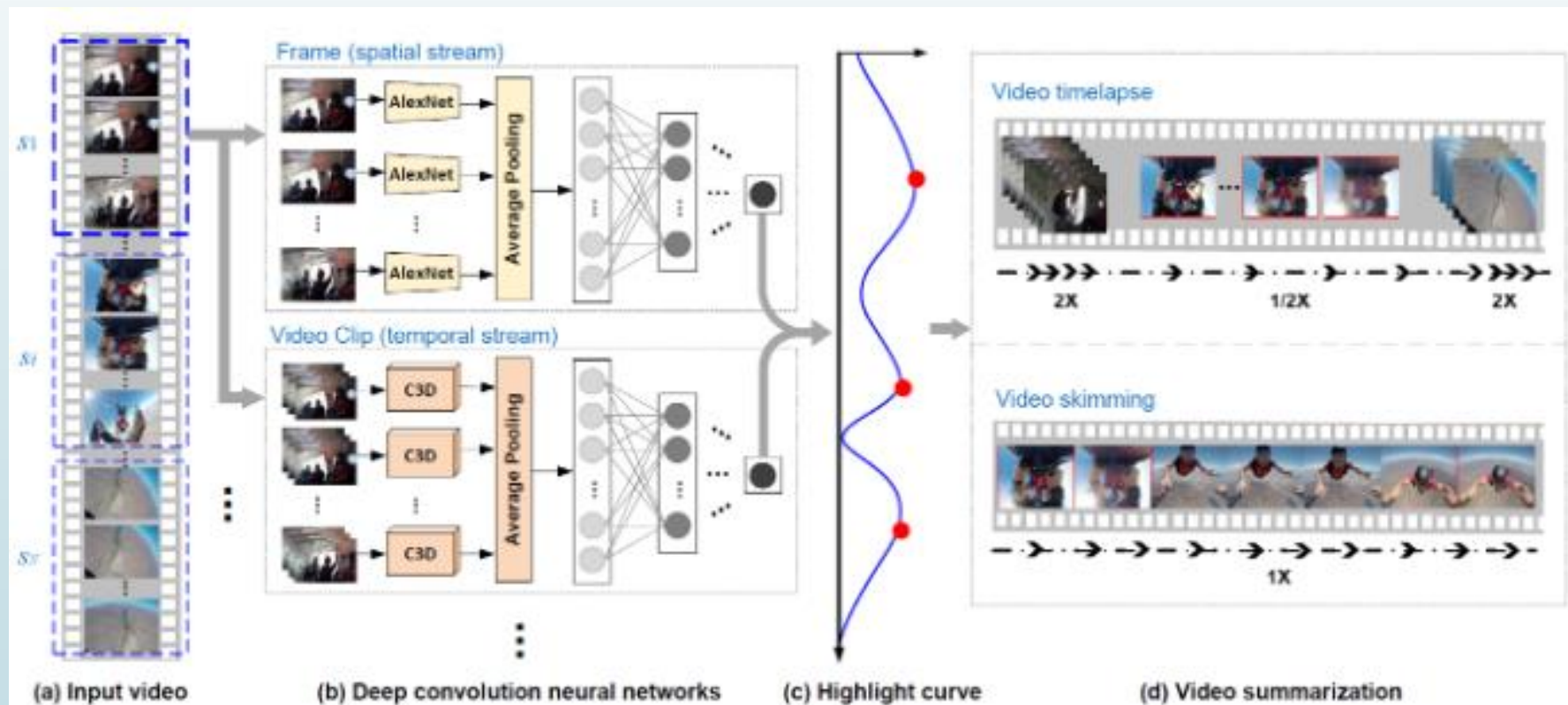


The Process of Static Video Summary Composition

基于内容的视频检索

——视频的层次结构

- 关键帧选取和视频摘要

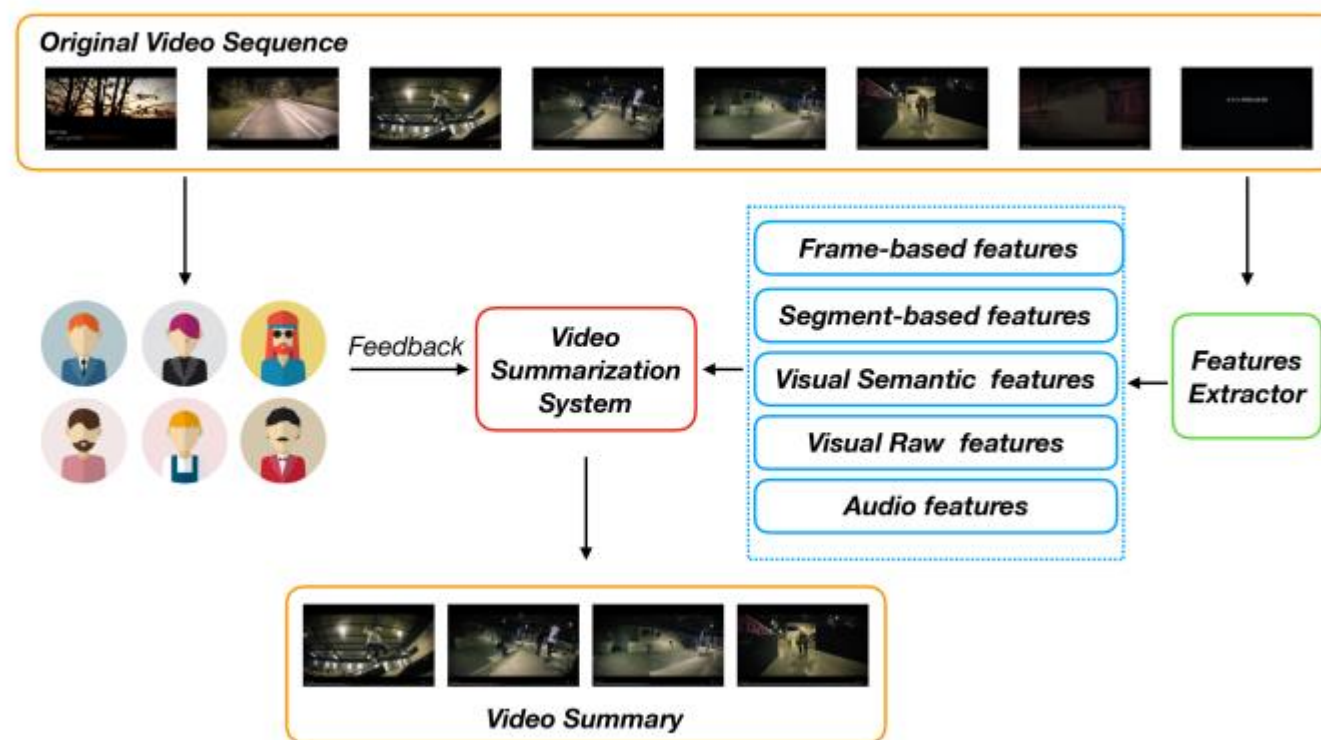


The Process of Dynamic Video Summarization

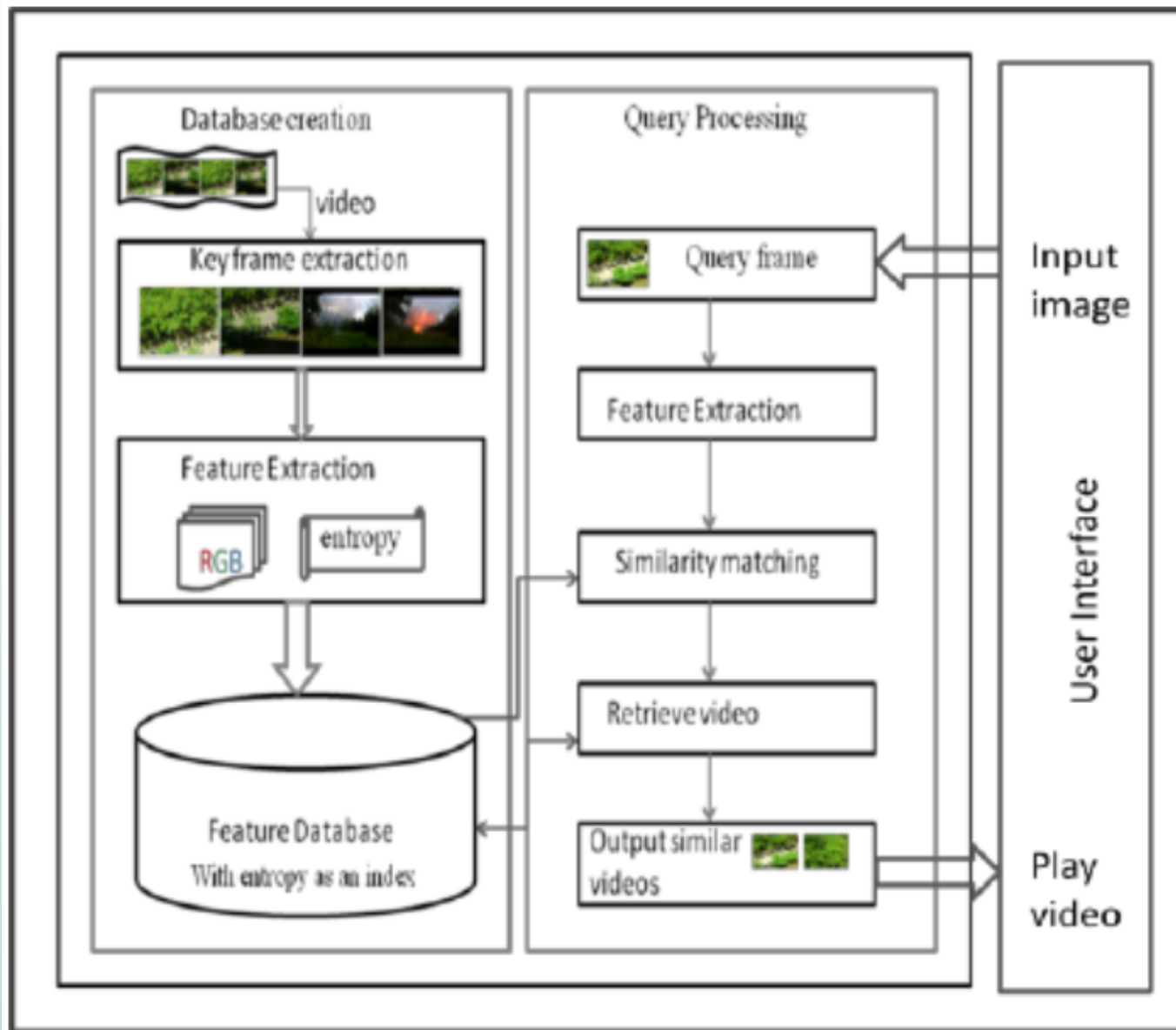
基于内容的视频检索

——视频的层次结构

► 关键帧选取和视频摘要



基于内容的视频检索



基于内容的视频检索

——MPEG-7视频内容标准

■ MPEG-7概述

- MPEG-7 的产生是为了解决上述问题，其重点在于影音内容的描述和定义，以有弹性、具延伸性、多层次及明确的数据结构和语法来定义影音数据的内容，经由 MPEG-7 的定义格式，使用者可以有效率地搜寻、过滤和定义想要的影音数据。

■ MPEG-7: 基于信息内容的搜索引擎

- MPEG-7是属于信息方面的检索和搜寻，而所谓信息，则可以是影像或音乐；换句话说，MPEG-7在本质上来说就是我们常常在网上使用的搜索引擎，只不过它提供的是多媒体的信息查询服务。

基于内容的视频检索

——MPEG-7视频内容标准

► MPEG-7中的主要概念

- 数据(Data) 是用MPEG-7描述的视听资料，不考虑它们的存储、编码、显示、传输、媒介或技术。
- 特征(Feature) 指数据的特性。特征本身不能比较，而要用有意义的特征表示(描述子)和它的实例(描述值)。
- 描述子(Descriptor, D) 是特征的表示。它定义特征表示的句法和语义，可以赋予描述值。
- 描述值(Descriptor Value) 是描述子的实例。
- 描述模式(Description Scheme, DS) 说明其成员之间的关系结构和语义。



基于内容的视频检索

——MPEG-7视频内容标准

■ MPEG-7中的主要概念

- 描述 (Description) 一个描述由一个描述模式 (结构) 和一组描述值组成。
- 编码的描述 (Coded Description) 是对已完成编码的描述, 满足诸如压缩效率、差错恢复和随机存取的相关要求。
- 描述定义语言 (Description Definition Language, DDL) 是一种允许产生新的描述模式和描述子的语言, 允许扩展和修改现有的描述机制。

基于内容的视频检索

——MPEG-7视频内容标准

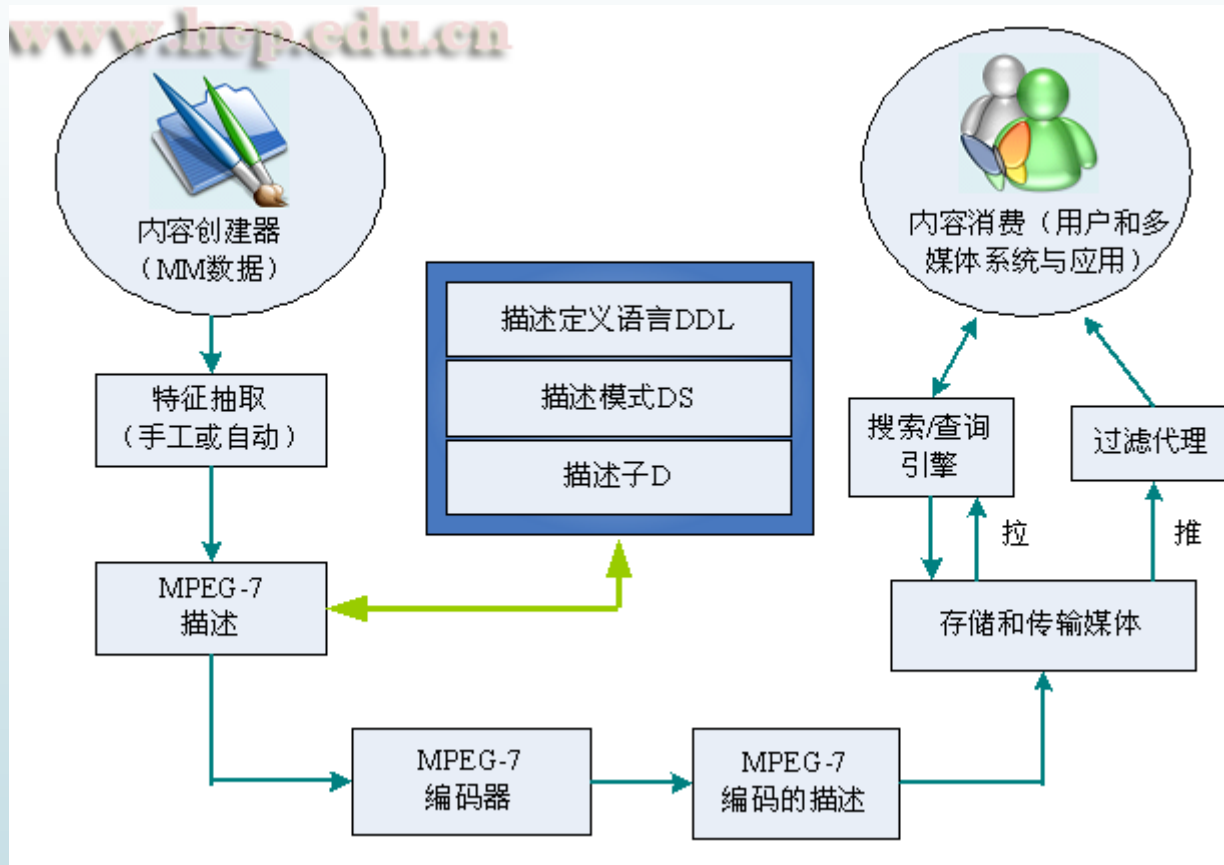


图 18 MPEG-7应用的表示

基于内容的视频检索

——MPEG-7视频内容标准

► MPEG-7的范围

- MPEG-7是针对存储形式(在线、脱机)或流形式(如 Internet上的广播、推拉模型)的应用而制定的, 并且可以在实时和非实时环境中操作。一个实时环境意味着当采集资料时, 信息是与内容相关的。
- 下图是MPEG-7处理过程的一种高度抽象示意图, 用于解释MPEG-7的范围。它包括特征抽取(分析)、描述本身和搜索引擎(应用)。

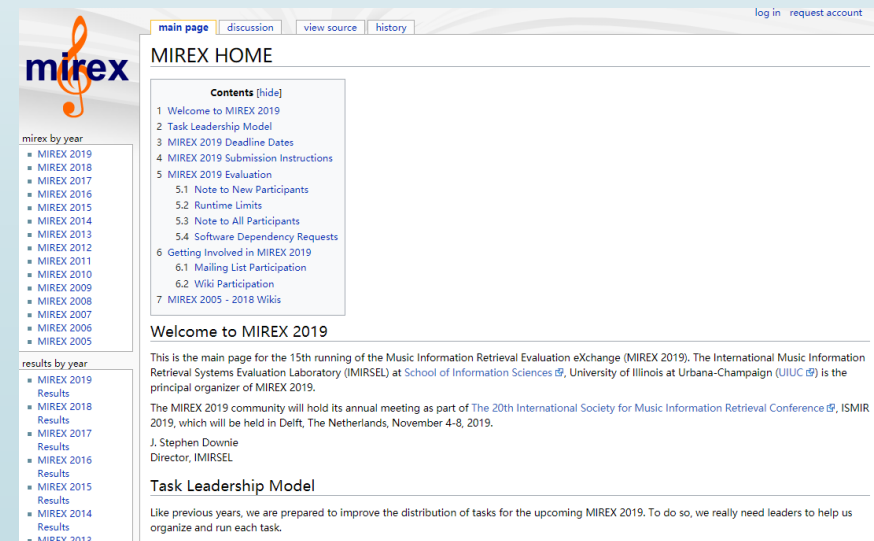


基于内容的音频检索

- ▶ 音频检索系统分为两种类型：原音检索和[模糊检索](#)；
- ▶ 根据对音频媒体的划分，音频信息检索分为以下三种：
 - ▶ 基于语音技术的检索：以语音为中心的检索，主要采用语音识别等技术，如电台节目、电话交流、会议录音等。
 - ▶ 音频检索：以波形声音为对象的检索。
 - ▶ [音乐检索](#)：以音乐为中心的检索，利用音乐的音符和旋律等音乐特性来检索，如检索乐器、声乐作品等

基于内容的音频检索

- [国际音频检索评测大赛 \(MIREX\)](#)
- MIREX比赛，是由美国伊利诺依大学厄本那-香槟分校(UIUC)国际音乐信息检索系统评估实验(ISMIRSEL)主办，是音乐信息检索领域顶级研究型竞赛，受到美国国家自然科学基金和梅隆-安德鲁基金的共同资助。MIREX大赛旨在为音乐信息检索及音乐信号处理领域中的各种前沿技术提供公正、可信的评估平台，每届大赛，都吸引了世界各国在音乐信息检索领域最为知名的大学、研究所和企业研究部门的广泛参与，从而使得MIREX成为每年中音乐信息检索领域的一件盛事。



基于内容的音频检索

——听歌识曲

小程序

智能听歌识曲

图片/音频/视频

听歌识曲小程序，QQ音乐官方出品，根据听到的声音识别出歌曲，支持歌曲识别模式和哼。

扫码 下载

ACRCloud

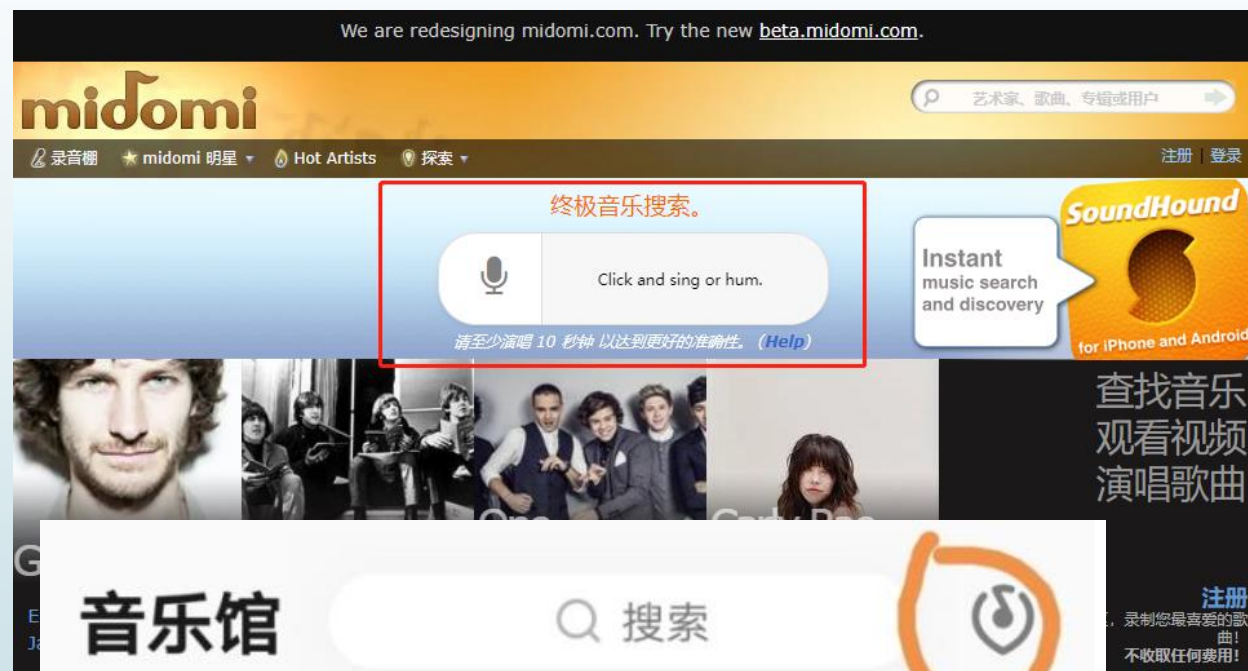
歌曲鉴权 音乐识别 车载影音 接唱抢麦 电视互动 开发者平台 新闻动态 联系我们

音频歌曲检索，支持识别用户哼唱

让用户不再错过好音乐



ACRCloud Music 可为移动应用、版权机构、艺人和音乐数据研究部门提供识别歌曲、歌手和专辑等信息的服务。通过ACRCloud高效的识别算



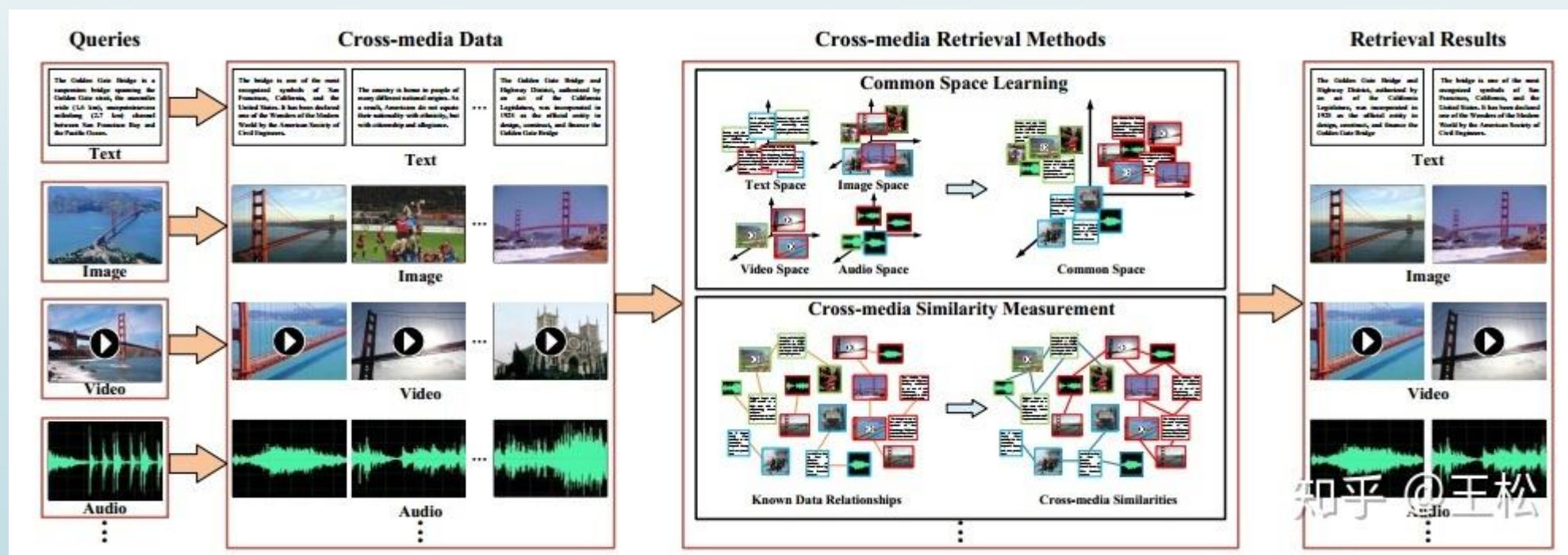
音乐馆

搜索























跨媒体检索

- 跨媒体检索的目标是计算不同媒体数据间的相似度，对于给定的查询样例，检索出与查询样例相关的不同媒体数据。关键挑战在于不同媒体的表示形式不一致，难以进行直接的相似性度量，即“媒体鸿沟”问题。两种主要的跨媒体检索方法：共同空间学习方法和跨媒体相似性度量方法。



跨媒体检索

➔ [demo1](#)

	Image	Text	Audio	Video	3D
violin		<p>one that commonly available string. By converting a standard violin to the picture range are represented by frequency under the label. The range, however, on the bottom edge commonly use amplification to overcome the limited projection and usually interact with a lower frequency, which is a small instrument.</p>			
flute		<p>The alto flute is a type of woodwind instrument. It is the most common of the C flute after the flute. It is characterized by its distinct, mellow sound in the lower portion of its range. It is a transposing instrument in G and, like the piccolo and alto flute, and the same fingering as the C flute.</p>			
airplane		<p>Airbus began as a consortium of aerospace manufacturers. In 1993, the European Commission of European defense and aerospace companies in 1994 and 1995 along the establishment of a unified, joint-stock company in 2000, merged by the European Aerospace Defense and Space Company (EADS) and the system (2006).</p>			
helicopter		<p>The Apache originally started as the Model 77 developed by Hughes Helicopters for the United States Army's Advanced Attack Helicopter program to replace the UH-1H. The prototype Model 77 was first flown on 30 September 1975. The U.S. Army ordered the first 100 over the last 100 in 1976, and later approved full production in 1982.</p>			
camera		<p>A camera is an optical instrument for recording or capturing images which may be stored locally, transmitted to another location, or both. The images may be recorded on photographic or electronic image capturing devices or recorded on a storage device. A camera is a remote sensing device as it senses subjects without physical contact.</p>			

知乎 @王松