



南京邮电大学
Nanjing University of Posts and Telecommunications

数字图像与视频处理

第10章 基于内容的图像 和视频检索





第10章 基于内容的图像和视频检索

- 10.1 基于内容检索技术概述
- 10.2 基于内容的图像检索
- 10.3 基于内容的视频检索



10.1 基于内容检索技术概述

所谓**信息检索**，就是根据用户的信息需求，从信息集合中检索出与用户信息需求相关的信息子集。

多媒体信息数据量极其庞大，并且多媒体数据在组织结构、表达形式等方面也不同于传统的文字数据，如何对它们进行组织、表达、存储、管理、查询和检索是对传统数据库技术提出的一个严峻挑战。特别是视频信息检索问题，尤其是基于内容的图像和视频检索技术已经成为国内外研究的热点问题。



10.1 基于内容检索技术概述

多媒体数据库的早期检索方法

- 基于属性的检索(**Attribute-Based Retrieval**)
- 基于文本的检索(**Text-Based Retrieval**)
 - 先利用文本标注媒体内容，将对媒体内容的检索转换为基于文本描述的检索。检索时，系统根据用户输入的关键字按照相似度大小排序返回部分匹配的结果。



基于文本的多媒体信息检索

- 人工建立元数据标注，本质类似于纯文本信息检索
- **检索方式**：信息分类、扩展名、关键词
- **普及程度**：是现行网络检索系统中的主流方式，如Google 图片搜索、音乐搜索与视频搜索，Yahoo 音乐搜索与图片搜索，Yahoo Video Search，百度mp3、图片与视频搜索等
- **优势**：实现原理相对简单，技术成熟
- **劣势**：目前的技术无法对多媒体信息进行自动文本描述，人工标注费时费力。而且，文字描述受观察者的主观因素影响很大，缺乏统一的描述标准，如，同一幅图像在不同人眼中会有不同的描述。



基于文本的多媒体信息检索

- 基于文本的检索方式，适合于处理**结构化的文本类和数值类**的记录信息检索。在数据库中，每个记录都有固定数目的域，每个域也都具有明确的意义（定性描述），这样每个记录仅具有特定和有限的语义信息，通过对不同域的（关键字）是否比较和（数值）大小比较可进行检索。
- 但这种方法用于多媒体信息检索则具有相当大的局限性。

基于文本的图像检索



[新闻](#) [网页](#) [贴吧](#) [知道](#) [MP3](#) [图片](#)

爱晚亭

百度搜索

[帮助](#) | [高级搜索](#)

☐ 新闻图片 ☒ 全部图片 ☐ 大图 ☐ 中图 ☐ 小图 ☐ 壁纸

[百度设为首页](#)

[乐吗? 官方手机音乐下载](#)

[让你的祝福全世界都知道](#)



600x432 48k jpg

[旅游图片](#) [爱晚亭](#) [爱晚亭](#) [照片](#)



759x503 84k jpg

[旅游图片](#) [爱晚亭](#) [爱晚亭](#) [照片](#)



800x600 70k jpg

[爱晚亭](#) [长沙市](#) [湖南省](#) [景点导游](#) [芒](#)



200x289 28k jpg

[爱晚亭](#)



1920x2560 1820k jpg

[爱晚亭](#)



300x451 34k jpg

[爱晚亭](#)



10.1 基于内容检索技术概述

基于文本检索的缺陷：

- 文本描述难于充分表达媒体的丰富内容
- 文本描述具有一定的主观性
- 处理文本描述涉及到自然语言理解问题
- 手工输入文本描述效率低，难以满足大容量媒体库的要求



10.1 基于内容检索技术概述

- 基于内容的检索(Content-Based Retrieval, CBR)在传统的检索框架中融合了对媒体内容的理解，是对多媒体对象的内容及上下文语义环境进行检索。
- 原理：提取特征并进行量化，表示成向量空间，建立索引库，将用户提问转化成向量，并与已有信息的向量空间进行相似度匹配计算。
- 常用的媒体特征：
 - ①图像：颜色、纹理、形状及空间关系等
 - ②视频：镜头、场景、镜头的运动等
 - ③音频：时域、频域、时频特征等



10.1 基于内容检索技术概述

- 基于内容的检索突破了传统的基于文本检索技术的局限，直接对图像、视频、音频内容进行分析，抽取特征和语义，利用这些内容特征建立索引并进行检索。
- 在这一检索过程中，它主要以图像处理、模式识别、计算机视觉、图像理解等学科中的一些方法为基础技术，是多种技术的合成。

多媒体信息检索分类

根据所检索媒体对象的不同，基于内容的多媒体信息检索又可分为基于内容的图像检索、基于内容的视频检索和基于内容的音频检索等。





第10章 基于内容的图像和视频检索

- 10.1 基于内容检索技术概述
- **10.2 基于内容的图像检索**
- 10.3 基于内容的视频检索



10.2 基于内容的图像检索

图像的特征是人们对图像视觉感受的量化描述。图像特征从各个方面描述了图像的内在语义，从而可以作为图像的抽象表示，然后根据对每个不同特征定义的相似度来表示不同图像之间的相似程度，从而实现对图像的检索。这就是所谓的**基于内容的图像检索**（**Content-based Image Retrieval, CBIR**）。

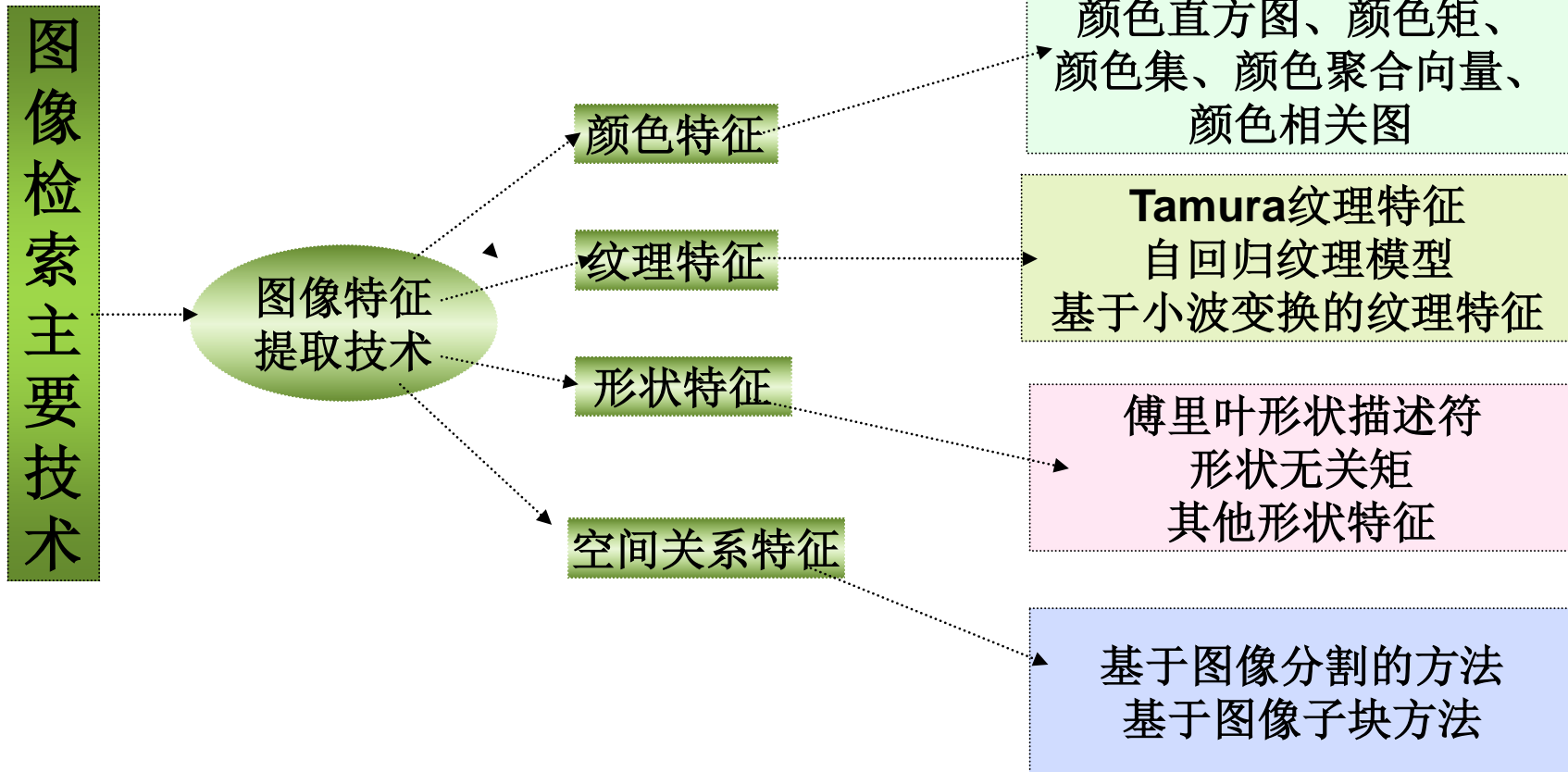


10.2.1 基于内容的图像检索概述

基于内容的图像检索（CBIR）技术涉及图像特征的提取、分析及匹配三个模块。

- **特征提取**：提取各种特征，如**颜色**、**纹理**、**形状**以及**对象空间关系**等。根据提取的特征不同，采取不同的处理，比如提取形状特征，就需要先进行图像分割和边缘提取等步骤。选择合适的算法，并在效率和精确性方面加以改进，以适应检索的需要，实现特征提取模块。
- **特征分析**：对图像的各种特征进行分析，建立图像的特征索引。
- **特征匹配**：选择何种模型来衡量图像特征间的相似度。

10.2.1 基于内容的图像检索概述





10.2.2 图像颜色特征的提取与表示

- 颜色特征是在图像检索中应用最为广泛的视觉特征，主要原因在于颜色往往和图像中包含的物体或场景十分相关。此外，与其他特征相比，颜色特征计算简单，同时对图像本身的尺寸、方向、视角的依赖性较小，具有较好的紧致性。
- 颜色具有一定的稳定性。在许多情况下，颜色是描述一幅图像最简便而有效的特征。



颜色特征的表达

- 颜色直方图
- 颜色矩
- 颜色集
- 颜色聚合向量
- 颜色相关图
-



颜色特征——颜色直方图

- 定义如下：

$$h(i) = \frac{n_i}{N}, i = 0, 1, \dots, K$$

- 其中 n_i 为图像中颜色取值为 i 的像素个数， N 为像素总数， K 为可能的颜色取值范围。
- 这样计算得到的颜色直方图就是一个 K 维的特征向量。颜色直方图所描述的是不同色彩在整幅图像中所占的比例，而并不关心每种色彩所处的空间位置，所以特别适合描述那些不需要考虑特定物体空间位置的图像内容。



颜色特征——颜色矩（Color Moments）

- 图像中任何的颜色分布均可以用它的矩来表示；
- 矩也是图形处理中常用概念；
- 由于颜色分布信息主要集中在低阶矩中，所以只采用颜色的一阶矩、二阶矩和三阶矩就可以表达图像的颜色分布。
- 与颜色直方图比较，该方法的一个好处就是无需对于特征进行量化。

优点：可以反映颜色分布

缺点：计算量大



颜色特征——颜色矩 (Color Moments)

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

一阶矩 (mean)

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^2 \right)^{\frac{1}{2}}$$

二阶矩 (variance)

$$s_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^3 \right)^{\frac{1}{3}}$$

三阶矩 (skewness)



颜色特征——颜色聚合向量

- 针对颜色直方图和颜色矩无法表达图像色彩的空间位置的缺点，Pass提出了图像的颜色聚合向量（color coherence vector）。
- 它是颜色直方图的一种演变，其核心思想是将属于直方图每一个bin的像素分为两部分：如果该bin内的某些像素所占据的连续区域的面积大于给定的阈值，则该区域内的像素作为聚合像素，否则作为非聚合像素。
- 由于包含了颜色分布的空间信息，颜色聚合向量相比颜色直方图可以达到更好的检索效果。



颜色特征——颜色相关图

- 颜色相关图（color correlogram）是图像颜色分布的另一种表达方式。这种特征不但刻画了某一种颜色的像素数量占整个图像的比例，还反映了不同颜色对之间的空间相关性。实验表明，颜色相关图比颜色直方图和颜色聚合向量具有更高的检索效率，特别是查询空间关系一致的图像。
- 如果考虑到任何颜色之间的相关性，颜色相关图会变得非常复杂和庞大。一种简化的变种是颜色自动相关图（color auto-correlogram），它仅仅考察具有相同颜色的像素间的空间关系。

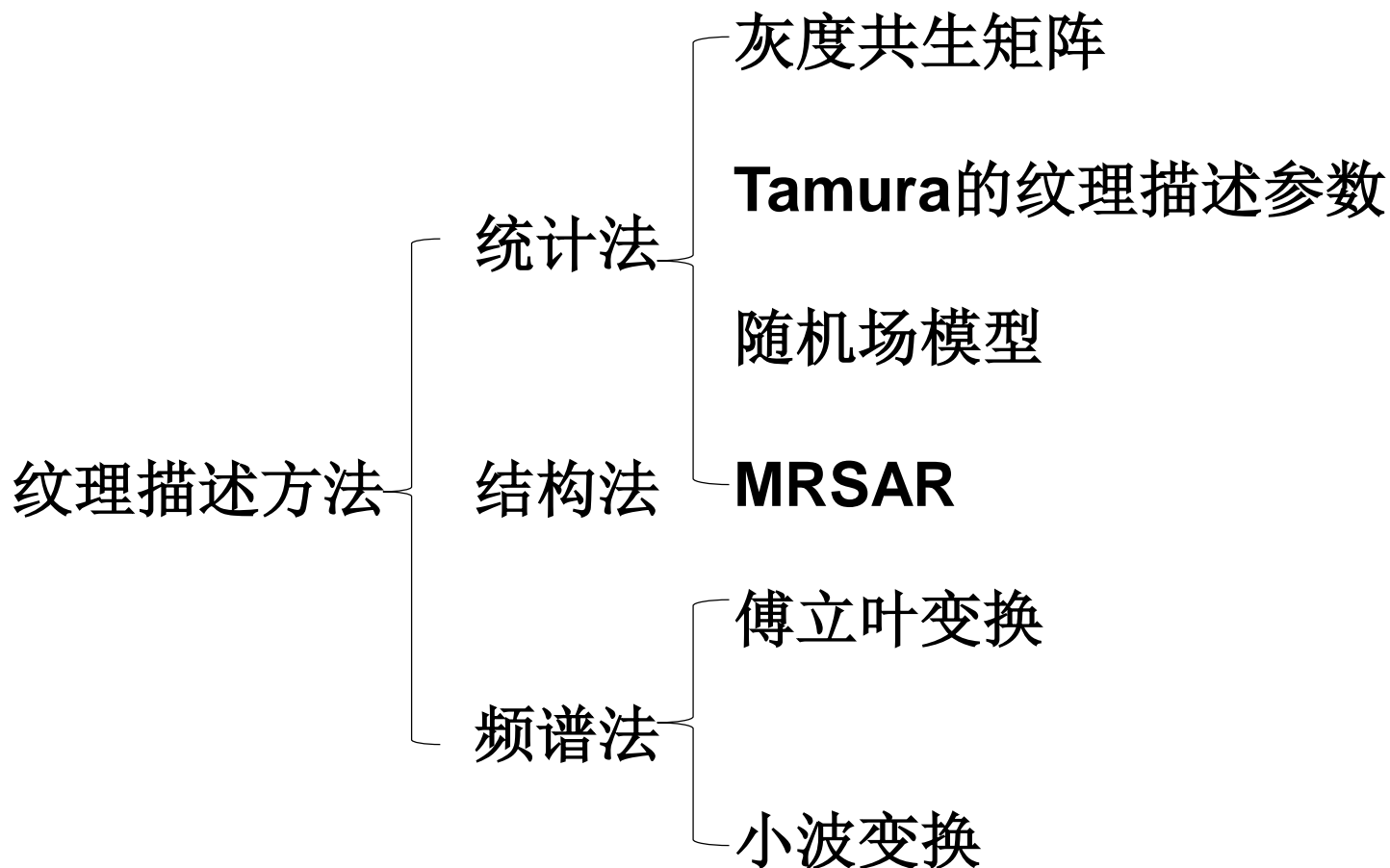
10.2.3 图像纹理特征的提取与表示

- 纹理是图像中颜色或灰度按一定的规律变化或分布形成的图案。
- 具有一定的重复性、均匀性和方向性等。
- 规律变化或分布的局部区域常称作纹理元，纹理就是纹理元规律性分布的结果。





纹理描述方法





纹理特征——Tamura纹理特征

- 基于对纹理的视觉感知心理学研究，Tamura等人提出了纹理特征的表达方法。
- Tamura纹理特征的6个分量：
 1. 粗糙度 (coarseness)
 2. 对比度 (contrast)
 3. 方向度 (directionality)
 4. 线像度 (linelikeness)
 5. 规整度 (regularity)
 6. 粗糙度 (roughness)
- 在Tamura表示中的所有纹理性质都是有意义的，与人的主观感受比较吻合，这使得Tamura纹理表示在图象检索中非常具有吸引力，而且可提供一个更有友好的用户界面。



10.2.4 图像形状特征的提取与表示

- 形状是描述图像内容的一个重要特征。它常与目标联系在一起，又有一定的语义含义，因而可以看作是比颜色或纹理**更高层次**的特征。
- 但另一方面，对形状的表达比对颜色或纹理的表达从本质上要复杂得多，常需要先**对图像进行分割**。由于当前的技术无法做到准确和通用的自动图像分割，图像检索中的形状特征只能在特定应用场合使用。在这些应用中，利用特定领域知识可以从图像中分割出目标（物体或区域）。



➤ 形状描述

形状描述方法

边缘描述

关键在于边缘检测的研究

区域描述

关键在于图像分割的研究

傅立叶描述符、小波描述符、不变矩等



基于形状特征的图像检索

■ 基本原理

- 形状是刻画物体的本质特征之一。
- 可以针对面积（可用像素点的个数计算）、主轴方向、偏心率、圆形率、连通性、正切角等形状特征进行匹配。



10.2.5 图像空间关系特征的提取与表示

- 颜色、纹理特征反映的是图像的整体特征，而无法体现图像中所包含的对象或物体。
- 复杂物体或物体间的关系有时成为图像检索的兴趣点。



▶ 图像的目标关系

- 空间关系(如绝对位置和相对位置)
- 朝向关系(如上、下、左、右等)
- 拓扑关系(如相邻、覆盖、包含等)
- 结构关系



提取图像空间关系特征的方法

- 基于区域的图像检索

对图像进行自动分割



划分出图像中所包含
的对象或颜色区域



索引

- 基于分块的图像检索

将图像划分为若干子块



对每个图像子块提取特征



索引



➤ 空间位置关系的讨论

- 空间关系描述包括基于目标的描述和基于关系的描述。
- 空间关系检索要解决空间关系的建模、表征、相似性量度等问题。
- 由于空间关系特征的模糊性、不确定性以及对自然语言描述的依赖性，因此基于空间关系的检索更复杂、更困难。



➤ 小结

- 视觉特征的提取是**CBIR**的基础。
- 仅仅靠单一特征很难完成图像内容的描述。
- 趋势：综合多个特征进行检索。

综合特性检索



图 8.3.2 仅用颜色特征的检索结果



图 8.3.3 仅用纹理特征的检索结果



图 8.3.4 综合颜色和纹理特征的检索结果



基于内容的图像检索——典型系统介绍

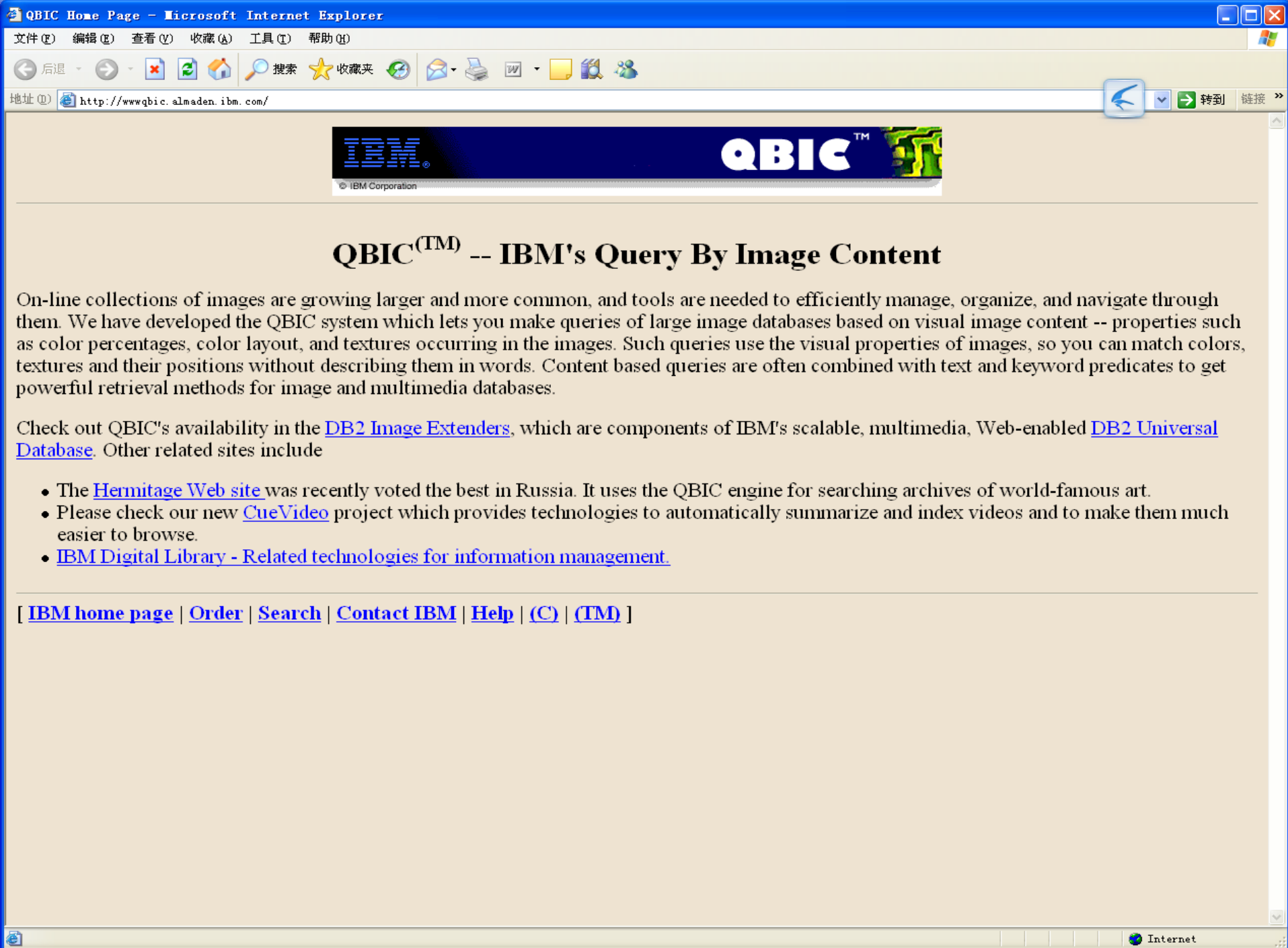
目前，比较成功的基于内容的图像检索系统有：

- **IBM公司的QBIC系统**
- **MIT媒体实验室的Photobook系统**
- **美国哥伦比亚大学的VisualSEEK系统**
- **Virage公司开发的virage检索系统**
- **香港公共图书馆的多媒体信息系统（MMIS）**



IBM公司的QBIC系统

- **Query By Image Content**
- **<http://www.qbic.almaden.ibm.com/>**
- **IBM Almaden研究中心开发的第一个商用的基于内容的图像及视频检索系统**
- **提供对静止图像及视频信息基于内容的检索手段**
- **系统结构及所用技术对后来的视频检索有深远的影响**





MIT媒体实验室的Photobook系统

- <http://vismod.media.mit.edu/vismod/demos/photobook/>
- 由MIT的媒体实验室开发研制
- 图像在存储时按人脸、形状或纹理特征自动分类
- 图像根据类别通过显著语义特征压缩编码

Photobook



Photobook is a tool for performing queries on image databases based on image content. It works by comparing **features** associated with images, not the images themselves. These features are in turn the parameter values of particular **models** fitted to each image. These models are commonly color, texture, and shape, though Photobook will work with features from any model. Features are compared using one out of a library of matching algorithms that Photobook provides. In [version 5](#), these include euclidean, mahalanobis, divergence, vector space angle, histogram, Fourier peak, and wavelet tree distances, as well as any linear combination of these. [Version 6](#) allows user-defined matching algorithms via dynamic code loading.

Since there are no image models which are optimal for all tasks and it is rarely clear which models are appropriate for a task, Photobook includes [FourEyes](#), an interactive learning agent which selects and combines models based on examples from the user. This makes Photobook different from tools like QBIC, Virage, SWIM, and CORE, which all support search on various features but offer little assistance in actually choosing one for a given task. FourEyes, by contrast, allows users to directly address their intent.

Example uses of Photobook:

- [Texture modeling](#)
- [Face recognition](#)
- [Shape matching](#)
- [Brain matching](#)
- [Interactive segmentation and annotation](#)
- [Other possible domains](#)

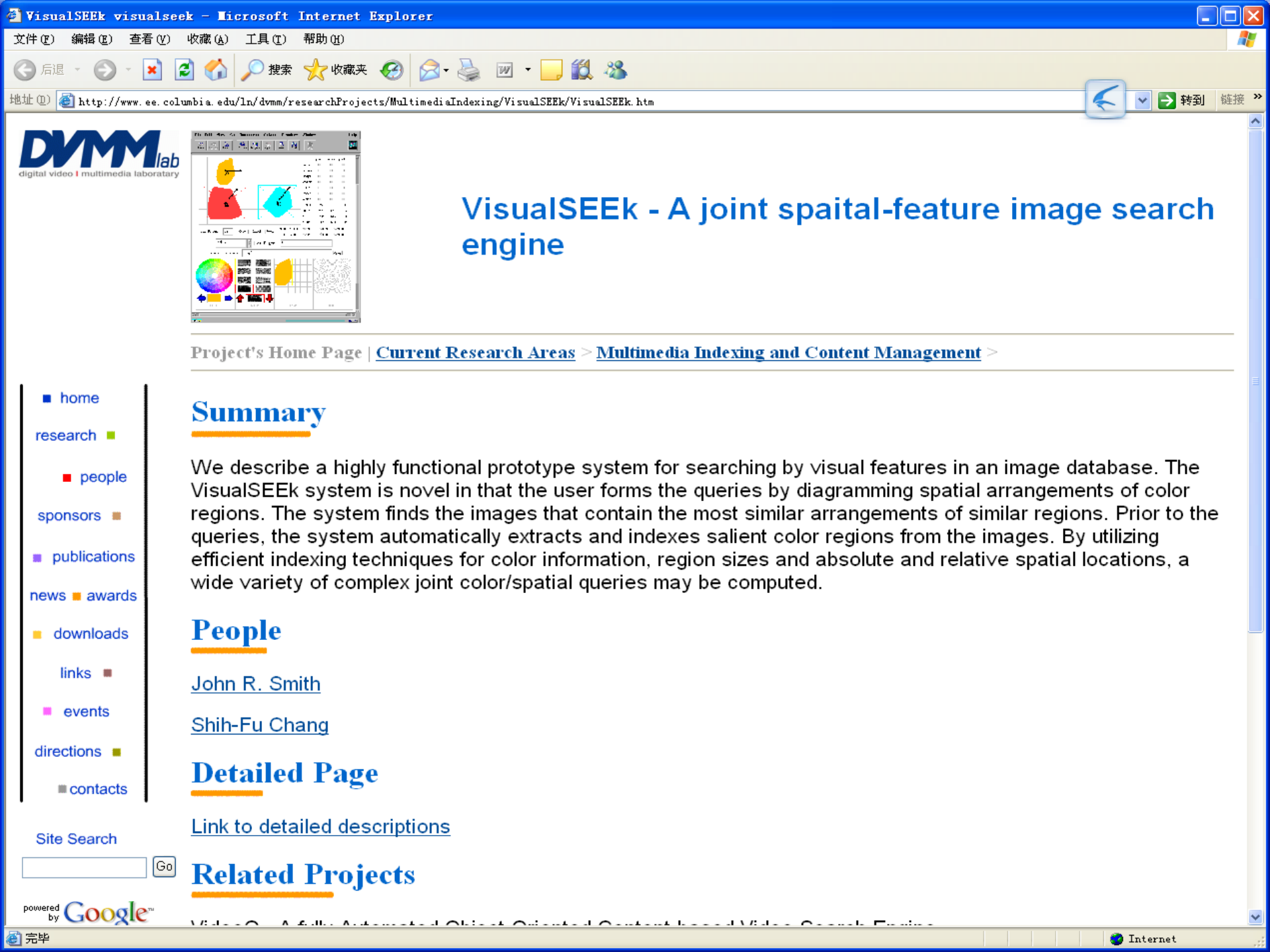
Photobook is available [free by FTP](#). It runs under the UNIX/Linux operating system. The distribution contains very little feature extraction code; you have to provide that yourself. In particular, no face recognition code is provided. For face recognition code, you should look in [/pub/eigenfaces/](#) or get [face-recognition.tar.Z](#). Send mail to [Alex Pentland](#) for more face recognition info. If you are looking for heavy-duty image retrieval software that runs out of the box, check out the commercial products listed in [the Northumbria report](#).

Publications (also see [FourEyes](#)):

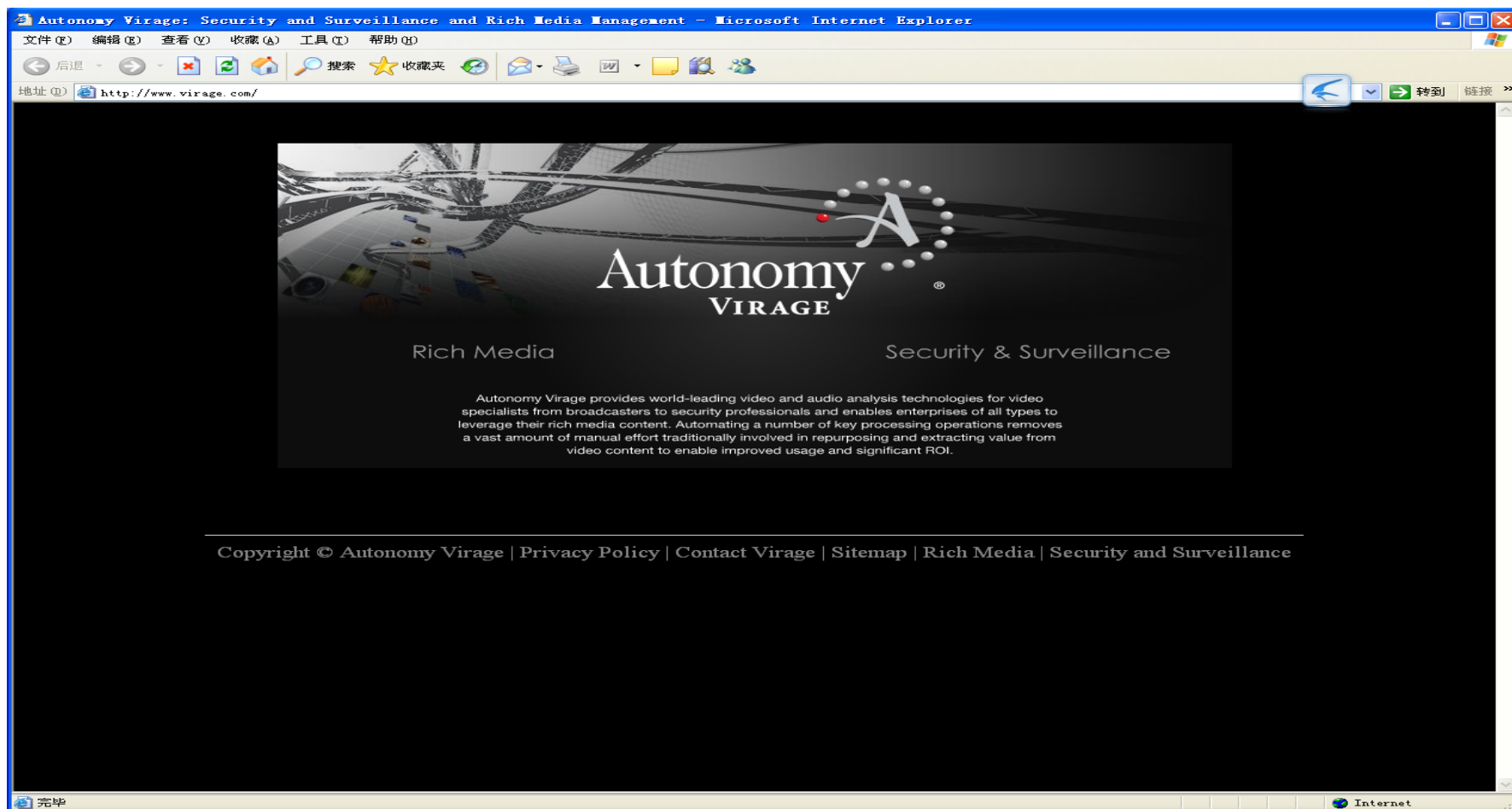


VisualSEEk系统

- <http://www.ee.columbia.edu/ln/dvmm/researchProjects/MultimediaIndexing/VisualSEEk/VisualSEEk.htm>
- 美国哥伦比亚大学开发的图像查询系统
- 主要特点是用到了图像区域的空间关系查询和直接从压缩数据中提取视觉特征



Virage公司开发的virage检索系统





香港公共图书馆的多媒体信息系统（MMIS）

- **Multimedia Information System**
- **<http://hkclweb.hkpl.gov.hk/hkclr2/internet/eng/html/welcome.html>**
- **IBM和分包商ICO于1999年底开始承建190万美元的数字图书馆项目**
- **被认为是世界上最大且最复杂的“中文/英文”双语图书馆服务之一**
- **采用的DB2 Text 和Image Extenders既支持文本检索，也支持图片检索**



HONG KONG PUBLIC LIBRARIES

MULTIMEDIA INFORMATION SYSTEM

SKIP

繁體 | 简体 | ENGLISH



多
媒
體

mmis

[What's New](#) [Workstation Booking](#) [Booking List](#) [Basic Search](#) [Advanced Search](#) [Image Search](#) [Online e-Resources](#)
[Library Catalogue](#) [eBook](#) [Web Links](#) [Electronic Directory](#) [Internet Service Hours](#) [Disclaimer](#) [E-mail to us](#) [Help](#)

Best viewed with Internet Explorer 5.5 at 1024 x 768 pixels with 16-bit high colour resolution.
Please install Media Player 7.0, QuickTime 5.0 and Acrobat Reader 5.1 or above for viewing multimedia content.
Please use Traditional Chinese (Big 5 encoded) or English to perform searching.
56k or above connection speed is recommended.
© Copyright Leisure & Cultural Services Department. All right reserved.

Hong Kong Public Libraries



第10章 基于内容的图像和视频检索

- 10.1 基于内容检索技术概述
- 10.2 基于内容的图像检索
- **10.3 基于内容的视频检索**



10.3 基于内容的视频检索

视频信息是一类复杂的信息，它不仅包含静止图像的内容，还包含场景中目标运动的信息和时间变化的信息。视频包含的数据量远远超过其他媒体，因此视频数据一直是多媒体处理和检索中最难解决的问题。



10.3.2 视频内容的结构化

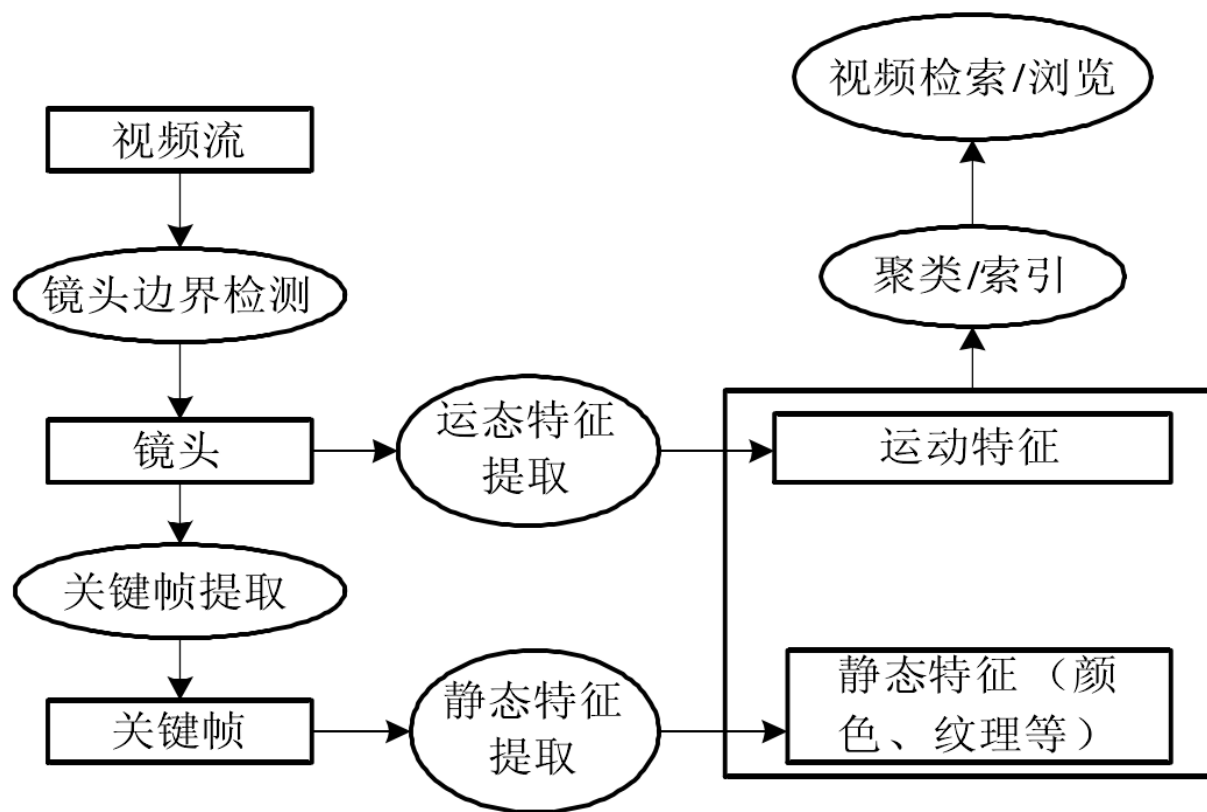
- **视频**又称动态图像，是一组图像按时间顺序连续表现，它的表示与图像序列、时间关系有关。
- 视频序列可用视频片段、场景、镜头、帧等描述。
视频片段由一系列相关的场景组成，表达一个完整的事件；**场景**由一些语义相关的镜头组成；**镜头**由一系列连续的帧组成；**帧**是一幅静态的图像，是组成视频的最小单位。
- 视频检索的实质就是在大量的视频数据中找到所需要的视频片段。



10.3.3 基于内容的视频检索工作流程

- 视频检索的首要任务是进行镜头的检测，将不同的镜头互相分开，即视频的时域分割。
- 在镜头检测的基础上，针对视频数据中由大量的冗余信息，可以采用提取镜头**关键帧**的方法来表达镜头的主要内容。
- **关键帧**是一幅能描述镜头主要内容的帧。
- 得到关键帧以后，就可以使用基于内容的图像检索技术对关键帧进行检索，于是视频检索问题就转化为图像检索问题。
- 镜头检测（镜头分割）、关键帧提取和动态特征提取是基于内容的视频检索的关键技术。

10.3.3 基于内容的视频检索工作流程



基于内容的视频检索工作流程



1.镜头检测（镜头分割）

镜头实际上是一段时间的视频信号，在这段时间内，摄像机可以由各种运动及变焦等操作，但没有摄像机信号的中断，因此，一段镜头内的图像不会有大的变换。

镜头检测算法有模板匹配法、直方图法、基于边缘的方法等。



2.关键帧提取

关键帧 (Key Frame) 也称代表帧，是用于描述一个镜头的关键图像帧，它通常反映一个镜头的主要内容，可以是一帧图像或多帧图像。在视频信息中，关键帧起着与关键词类似的作用。

关键帧提取的算法主要有固定间隔抽取法、基于图像特征变化法、视频帧聚类法等。



哥伦比亚大学的VideoQ

- <http://www.videoq.com/>
- 允许用户使用视觉特征和时空关系来检索视频
- 集成文本和视觉搜索方法
- 自动的视频对象分割和追踪
- 丰富的视觉特征库，包括颜色、纹理、形状和运动

