多媒體資訊檢索

邱志義

中央研究院資訊科學研究所

前言

- **多媒體**泛指多種媒體型態的融合，包括文字(text)、圖像(image)、影像(video)、語音(speech)、音樂(music)、動畫(animation)等等

- **資訊檢索**是擷取、組織和利用資訊的重要技術

Data Explosion!

One person will have
> 30 gigabytes in 4 years
(>100,000 e-mails and
16,000 photos)

– Gorden Bell
pioneer of MyLifeBits

Data is of no use

unless you can actually access it.

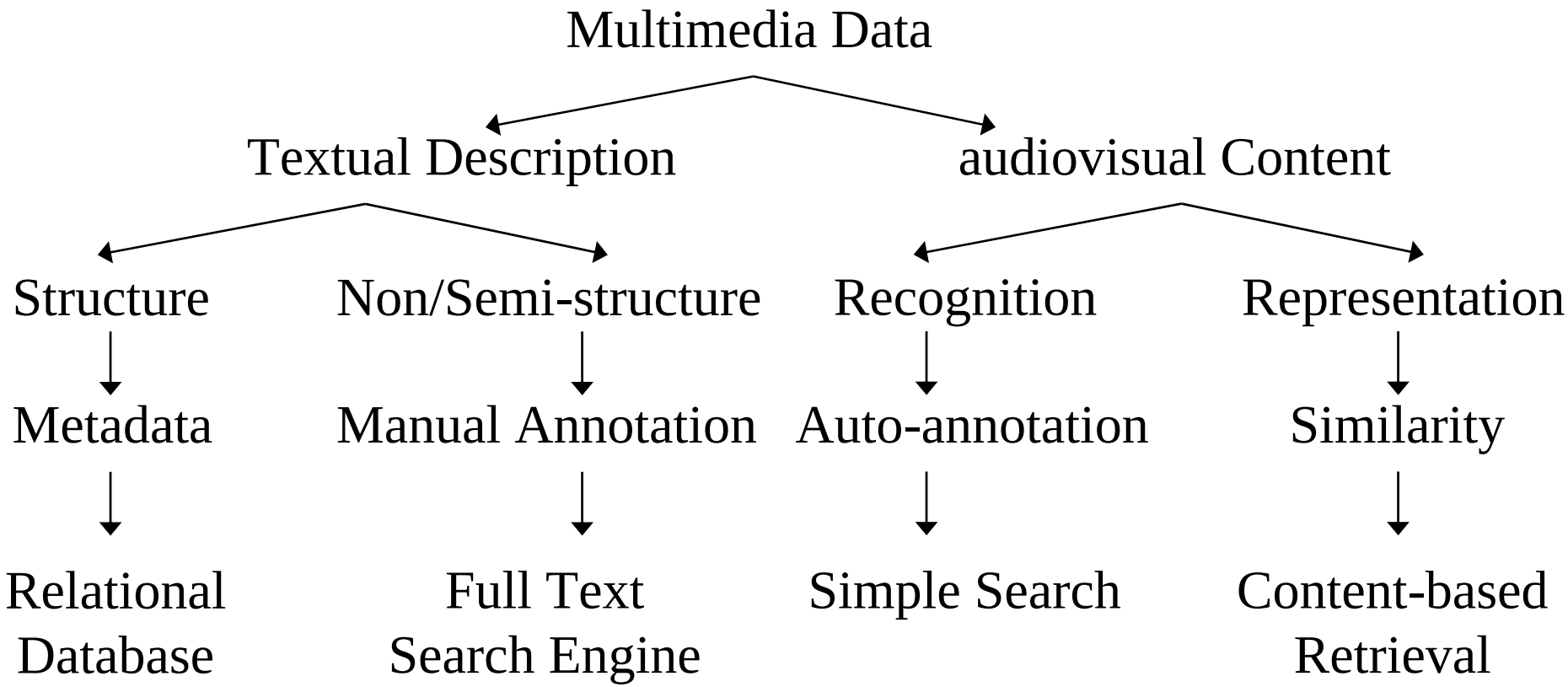
Multimedia Information Systems

Processing	**Meta-data**	**Real-time**
Recognition	**Indexing**	**Interaction**
Copyright	**Similarity**	**Adaptation**

Multi-disciplines

- Database
- Signal processing
- Pattern recognition
- Information retrieval
- Library science
- Cognitive psychology and science
- Human-computer interaction

Multimedia Information Retrieval - Roadmap



Outline

- 多媒體資料庫搜尋
- 多媒體全文搜尋引擎
- 以內容為基礎之多媒體檢索

Outline

- 多媒體資料庫搜尋
- 多媒體全文搜尋引擎
- 以內容為基礎之多媒體檢索

Multimedia Database Search

CalPhotos (Berkeley digital library project)

California Department of Water Resources

This form can be used to look for photographs from more than 17,000 images of California's natural resources from the Californ Department of Water Resources. Click here for more information about the DWR photos.

Instructions: To look for photos, choose **one or more** of the options below and click "Search".

Photo Caption	contains		
Subject	begins with	Computers (5)	
Category	equals	any	
Location	begins with	California (1)	
Date of Photo	equals		Example: 1995-01-31
Perspective	equals	any	

SELECT *
FROM img_dwr
WHERE subject like "Computers" and location like "California"

Multimedia Metadata

- Bibliographic attributes
- Sensory perception
  - Low-level audiovisual features (color, texture, shape, audio, motion, ...)
- Generic knowledge
  - High-level semantic features (subject, of/about, generic/specific, time, space, activity, object, …)
  - Perception subjectivity & labor-intensive annotation

Bibliographic Attributes

- DC (Dublin Core)
- CDWA (Categories for the Description of Works of Arts)
- EAD (Encoded Archival Description)
- CSDGM (Content Standards for Digital Geospatial Metadata)
- EXIF (Exchangeable Image File Format)

Example: EXIF

- 可交換影像檔案格式 (Exchangeable Image File Format)
  - 日本電子工業發展協會於1996年制定
  - 可以附加於JPEG、TIFF、RIFF等檔案類型的標頭檔(Header)裡，用來描述數位相機圖檔的相關資訊

EXIF

屬性名稱（英文）	屬性名稱（中文）	屬性值
主要資訊		
Make	製造商	FUJIFILM
Model	型號	FinePix4700 ZOOM
Orientation	影像方向	Left-hand side
X Resolution	水平解析度	72/1
Y Resolution	垂直解析度	72/1
Resolution Unit	解析度單位	Inch
Software	韌體版本	Digital Camera FinePix4700 ZOOM Ver. 1.00
Date Time	拍攝時間	2000:01:01 13:39:44
YCbCr Positioning	YCbCr位置	co-sited
Copy Right	版權	
Exif Info Offset	Exif資訊偏移量	262

主要資訊		
F Number	光圈(F)值	F2.8
Exposure Program	曝光程序	Program Normal
ISO Speed Ratings	ISO 值(感光度)	200
Exif Version	Exif 版本	0210
Date Time Original	原始成像時間	2000:01:01 13:39:44
Date Time Digitalized	相機存檔時間	2000:01:01 13:39:44
Component Configuration	Component Configuration	YCbCr
Compressed Bits Per Pixel	壓縮後每像素位元數	2/1 (bit/pixel)
Shutter Speed Value	快門速度	1/49Sec
Aperture Value	光圈值	F2.8
Brightness Value	亮度	EV2.8
Exposure Bias Value	曝光偏移	EV0.0
Max Aperture Value	最大光圈值	F2.8
Metering Mode	測光模式	Division
Flash	閃光燈	Not fired
Focal Length	焦距	8.30(mm)
Marker Note	Marker Note	FUJIFILM Format : 214Bytes (Offset:732)

Flash Pix Version	FlashPix 版本	0100
Color Space	色彩空間	sRGB
Exif Image Width	圖像寬度	1280
Exif Image Height	圖像高度	960
Exif Interoperability Offset	Exif Interoperability Offset	938
Focal Plane X Resolution	水平原始解析度	1270/1
Focal Plane Y Resolution	垂直原始解析度	1270/1
Focal Plane Resolution Unit	原始像素單位	Centimeter
Sensing Method	感光元件	One Chip Color Area sensor
File Source	檔案來源	DSC
Scene Type	取景方式	A directly photographed image

廠商原始資訊		
Version	版本	0130
Quality Mode	圖像品質	Normal
Sharpness	銳利度	Normal
White Balance	白平衡	Auto
Flash Mode	閃光燈模式	Off
Flash Strength	閃光時間	0/10
Macro Mode	近拍模式(?)	On
Focus Mode	聚焦模式	Auto Focus
Slow Sync	慢速同步	Off
Mode	模式	Auto
Sequence Mode	序列模式	Off
Blurring Warning	模糊警示	Yes
Focus Status	聚焦狀態	Nice
Exposure Status	曝光狀態	Nice

Sensory Perception

- Sensory perception is about the low-level feature representation of multimedia data

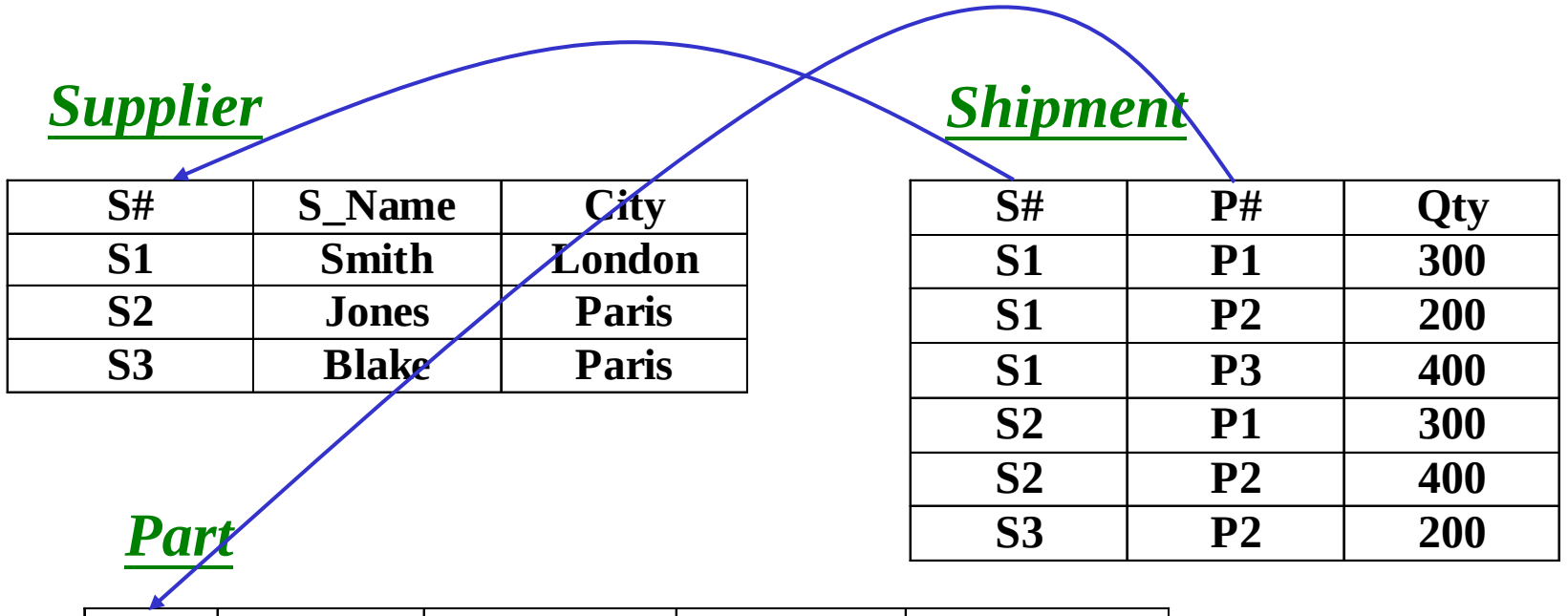
- We will discuss it in content-based retrieval

Generic Knowledge

Shatford Layne
Journal of the American Society for Information Science, 1994

層面（Facets）	屬於（Of）		關於（About）	非層面專指的相關意義
	狹義（Specific Of）	廣義（Generic Of）		
人物（Who）	有個別名稱的人、動物、事物	某一類型的人、動物、事物	某人或物所象徵的抽象意涵	
事件（What）	有個別名稱的事件	行動、情況	情緒或事件的抽象表達	
地點（Where）	有個別名稱的地理區	某一個地理區或建築地	含有抽象象徵意義的地區	
時間（When）	某個時間、日期或時期	某個季節或時日	某個時間所象徵的抽象情緒	

Relational Data Model



Supplier

S#	S_Name	City
S1	Smith	London
S2	Jones	Paris
S3	Blake	Paris

Shipment

S#	P#	Qty
S1	P1	300
S1	P2	200
S1	P3	400
S2	P1	300
S2	P2	400
S3	P2	200

Part

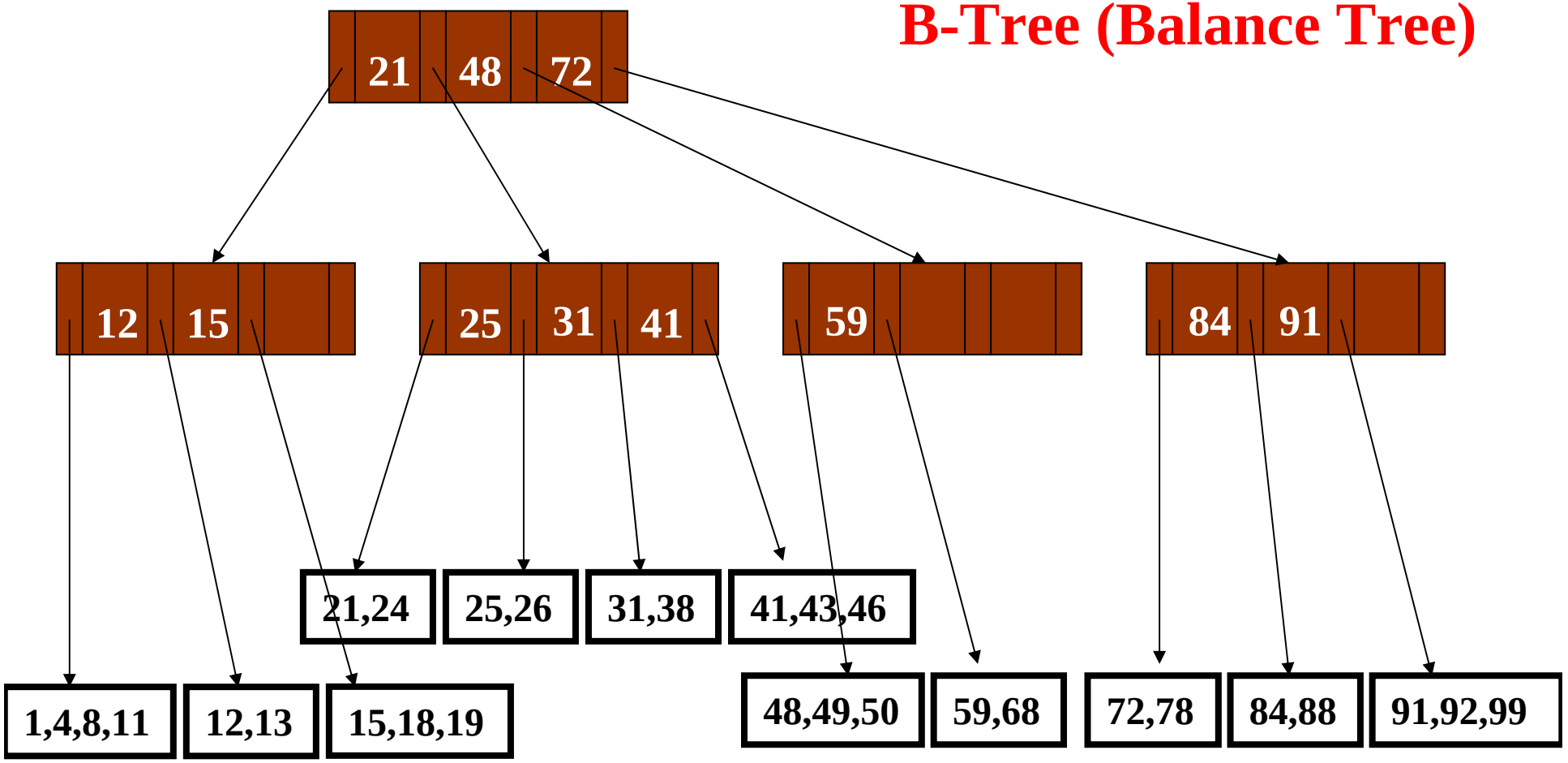
P#	Pname	Color	Weight	City
P1	Nut	Red	12	London
P2	Bolt	Green	17	Paris
P3	Screw	Blue	17	Rome
P4	Screw	Red	14	London

Querying Relational Databases

```
SELECT Supplier.S_Name
FROM Supplier, Shipment
WHERE Supplier.S# = Shipment.S# And Shipment.P# = P1
```

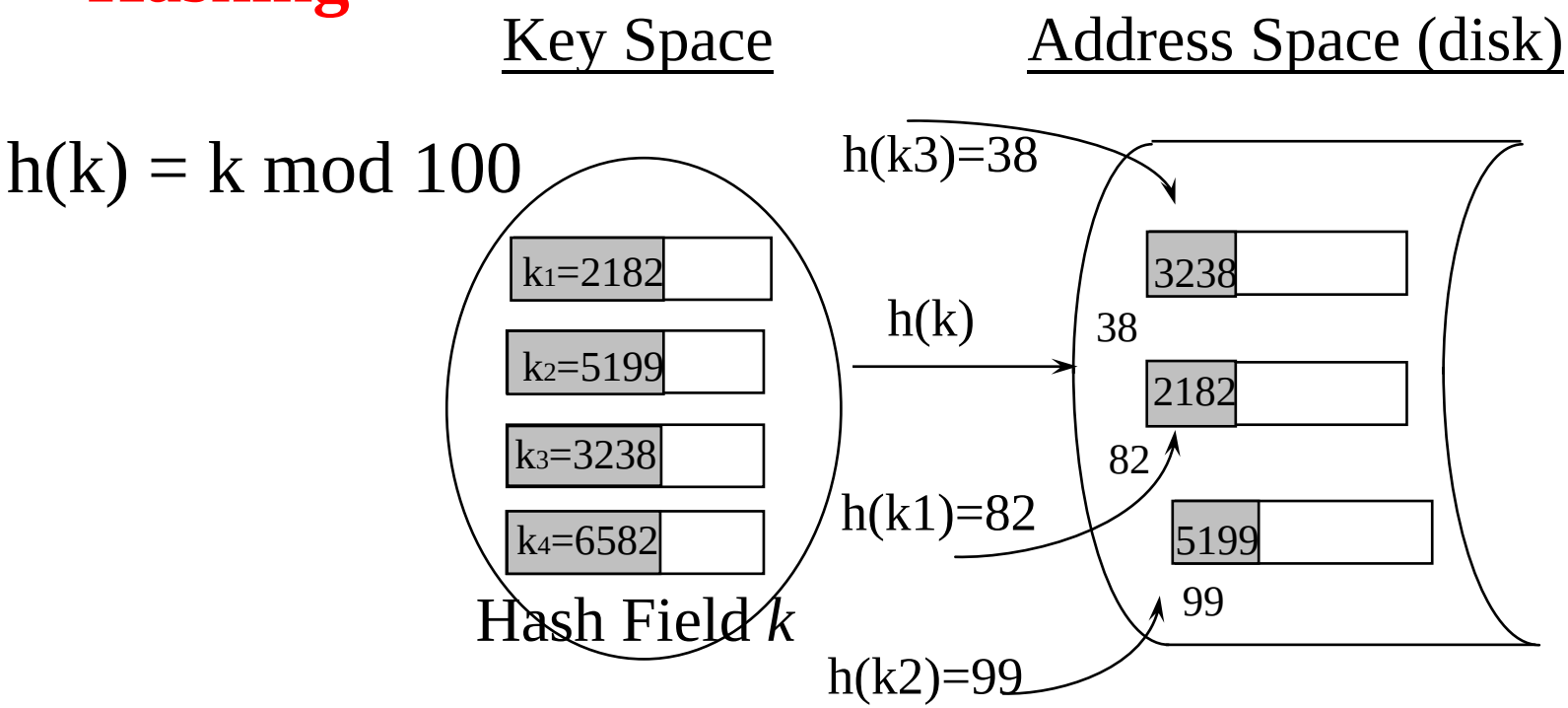
S#	S_Name	City	P#	Qty
S1	Smith	London	P1	300
S1	Smith	London	P2	200
S1	Smith	London	P3	400
S2	Jones	Paris	P1	300
S2	Jones	Paris	P2	400
S3	Blake	Paris	P2	200

Indexing for Relational Databases



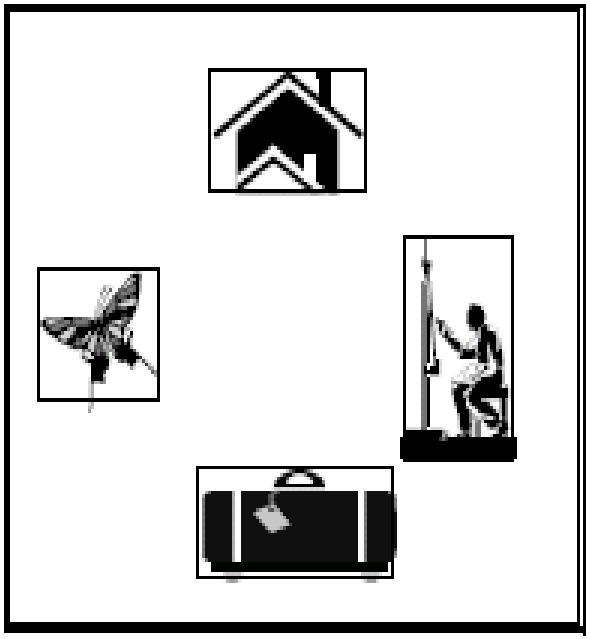
Indexing for Relational Databases

Hashing



h: a hashing function that map key set to address space

Signature Indexing



Objects: Butterfly (B), House (H), Luggage (L) and Artist (A)

Object Signatures:
(B): 001 000 110 010
(H): 010 001 100 010
(L): 001 000 110 010
(A): 001 010 110 000

Image Signature: 011 011 110 010

Queries	Signature	Result
1) House	010 001 100 010	Match
2) Desk	000 010 100 101	No Match
3) House and Artist	011 011 110 010	Match
4) Car	010 010 100 000	False Drop

Outline

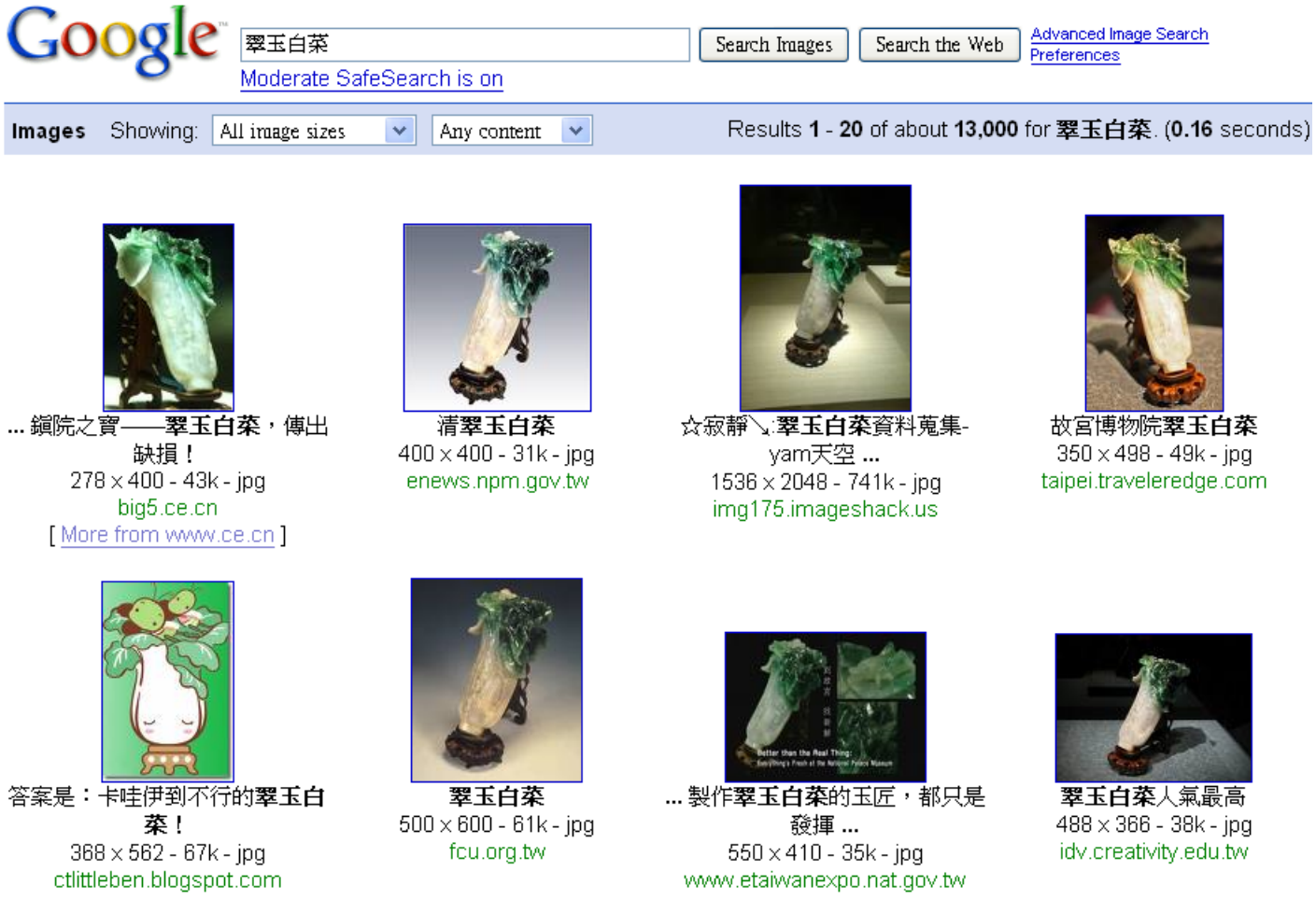
- 多媒體資料庫搜尋
- 多媒體全文搜尋引擎
- 以內容為基礎之多媒體檢索

Full-text Search Engine for Multimedia

- Just type:

  "*image/video/music search engine*"

  in Google for numerous search engine sites!



YAHOO! 奇摩 搜尋

網頁 | 知識+ | 圖片 | 影片 | 部落格 | 商家 | 新聞 | 商品 | 綜合 | 更多

翠玉白菜 搜尋

圖片搜尋

翠玉白菜 搜尋結果約305個，以下爲1 - 20個 ，共花0.04秒

顯示: 全部 桌布 - 大 - 中 - 小 | 彩色 - 黑白

翠玉白菜
hikaliwu 上傳至Flickr
www.flickr.com/photos/hikaliwu/438442917/

sub6.jpg
500 x 600 pixels - 60.8kB
www.npm.gov.tw/exh95/lateching/pic06_ja.html

kokyu06.jpg
600 x 450 pixels - 74.4kB
www2j.biglobe.ne.jp/~yamag/.../kokyu06.htm

10_K1C002103N00...AA.jpg
688 x 900 pixels - 182.9kB
www.npm.gov.tw/zh-tw/collection

翠玉白菜
TOPquark 上傳至Flickr
www.flickr.com/photos/65275041@N00/520635866/

W02007033017774...88.jpg
300 x 431 pixels - 80kB
www.yangtse.com/xwpd/gn/200703/t20070330_275148.htm

art3_2.jpg
384 x 512 pixels - 44.9kB
www.tabitabi-taipei.com/.../index3.html

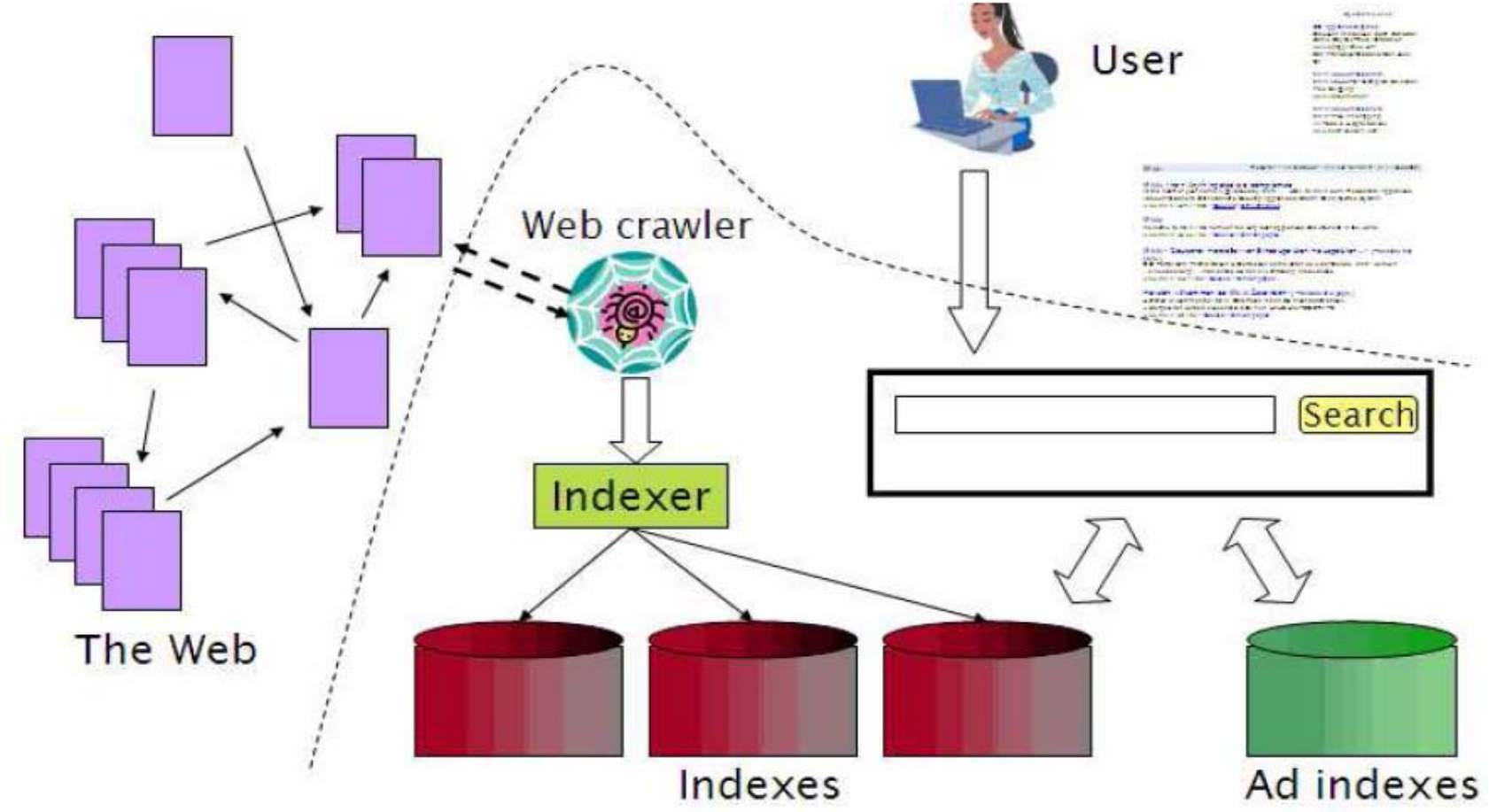
国立博物院（翠玉白菜）
benjaminraota... 上傳至Flickr
www.flickr.com/photos/77408240@N00/62054586/

Textual Image

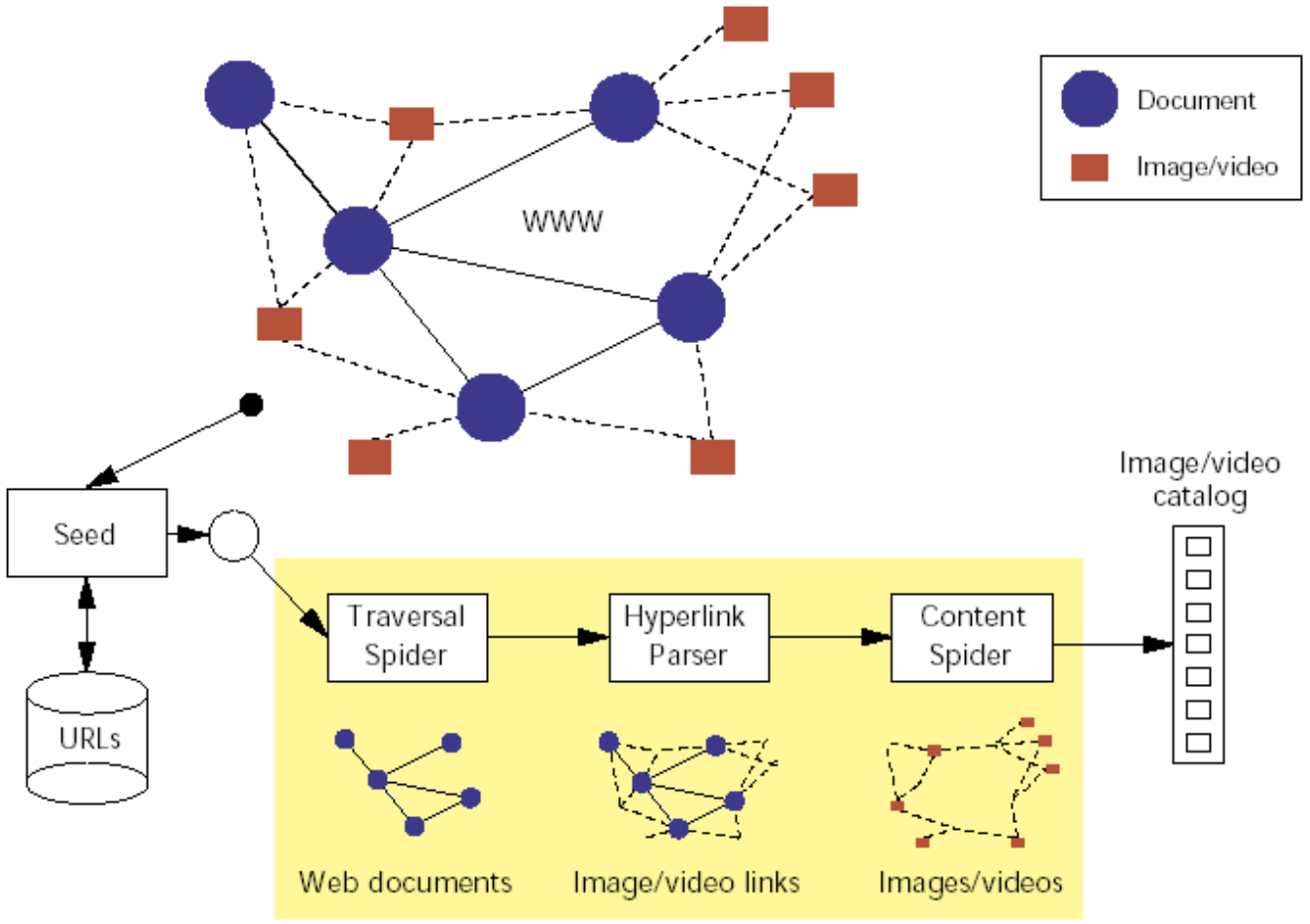
- Images of documents that contain mainly typed or typeset text

- How to obtain?
  - Use a set of keywords, i.e., **metadata**, associated with each image at creation time
  - Use **OCR** to recognize the words in the image
  - Use the **web crawler** to extract texts around the image

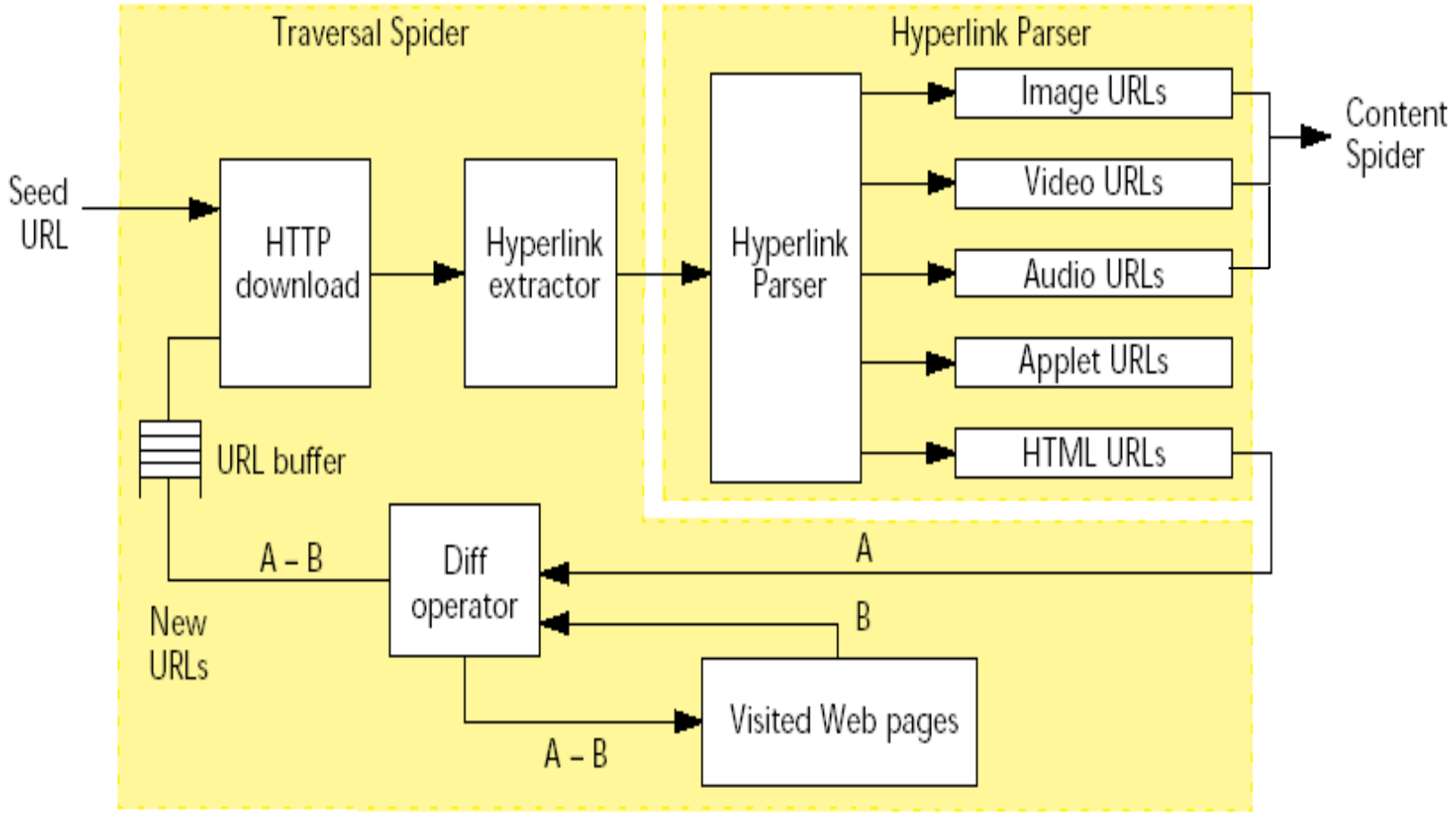
Web Search Overview



Gathering Web Multimedia Files



Gathering Web Multimedia Files



Semantic Extraction from Web

- From Web pages
  - Filename/URL, anchor text, image caption, page title, alternative text (ALT = ...), surrounding text
- From multimedia file headers
  - File type, file size, EXIF attributes, MPEG attributes, MP3 attributes

Image Purpose?

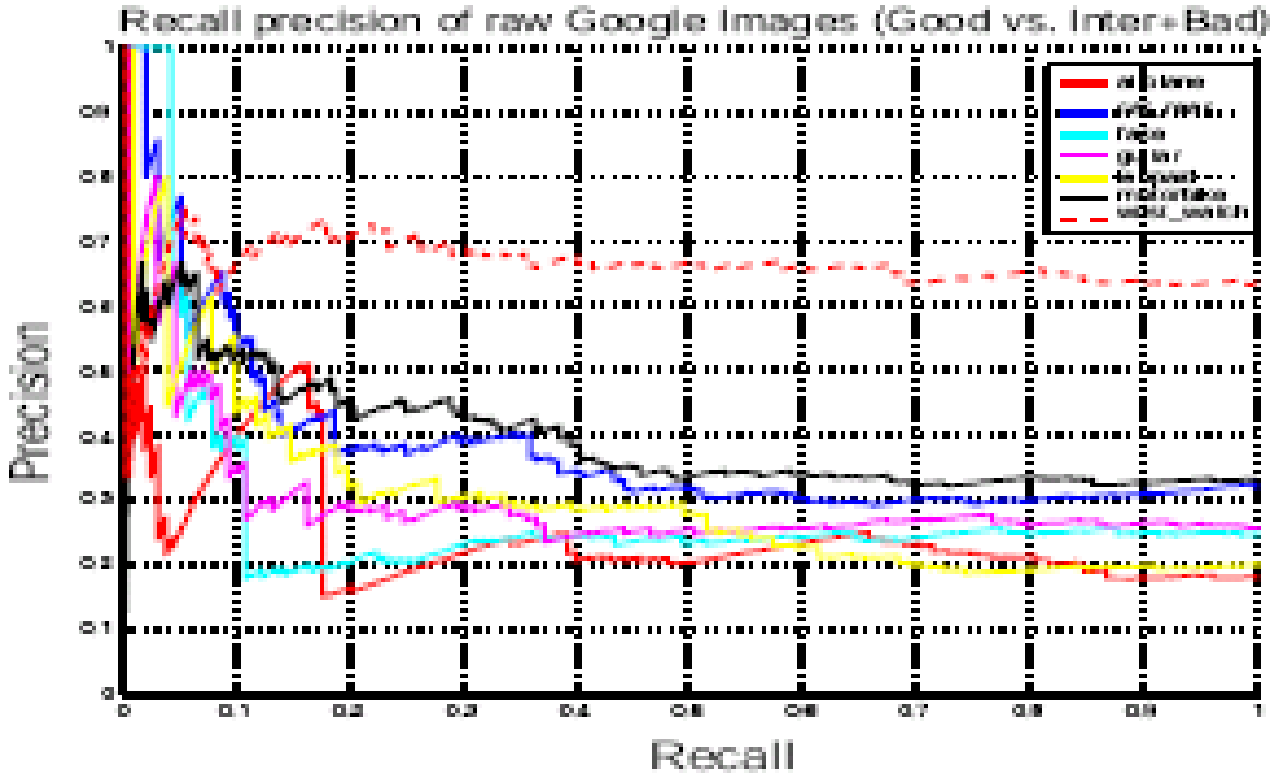
- Advertisement images
- Content images
- Decorative images (buttons, balls, …)
- Informational images (under construction, news, …)
- Logo
- Navigation images (arrows, back to home, image maps, …)

How to filter unwanted images?

Classic IR Models for Image Retrieval

- Set Theoretic Model
  - Boolean model
  - Fuzzy-set model
- Algebraic Model
  - Vector-space model
- Probabilistic Model
- Hybrid model

Recall vs. Precision of Google Image



Google Image Labeler

Outline

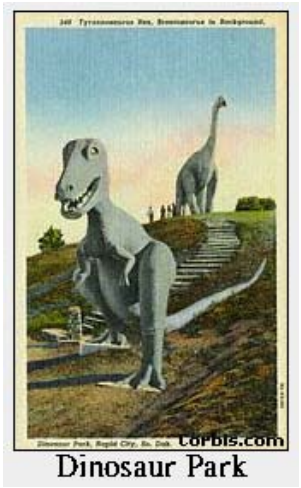
- 多媒體資料庫搜尋
- 多媒體全文搜尋引擎
- 以內容為基礎之多媒體檢索

Content-based Multimedia Retrieval

- Multimedia files that are indexing and retrieval based on audiovisual perceptual features, which can be extracted automatically
  - **Image:** color, texture, shape, spatial layout
  - **Audio:** MFCC, note frequency, rhythm, speech
  - **Video:** motion, image, audio

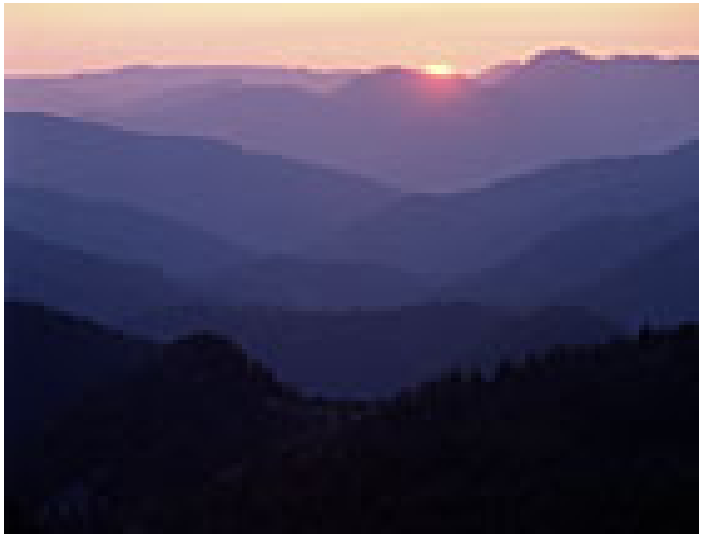
Why Content-based Approach?

- Limitation of **text-based retrieval**
  - Manual annotation is **labor sensitive** and **subjective interpretation**
  - An image is worth a thousand words?

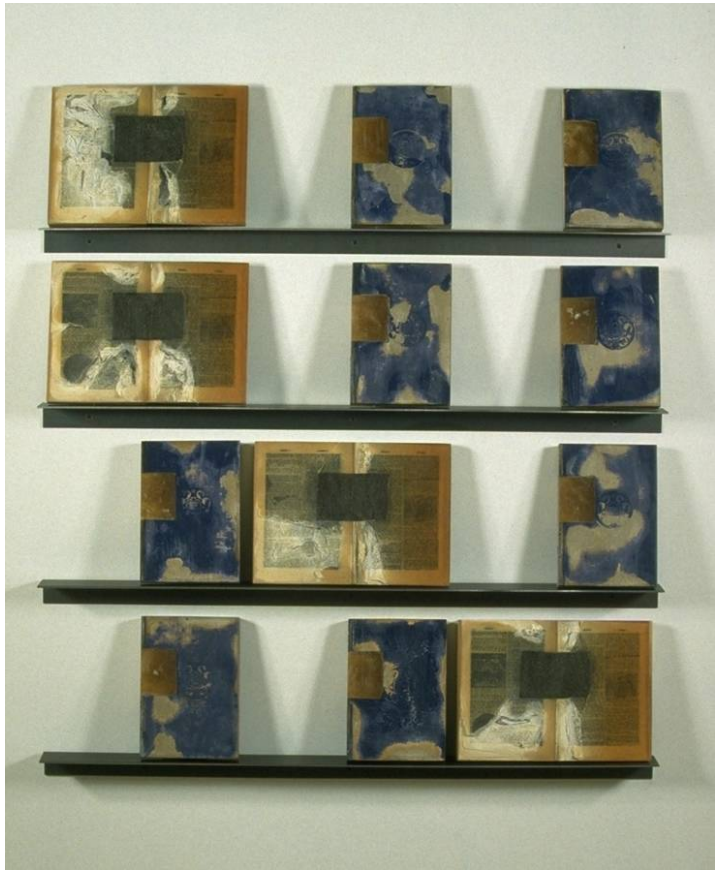


But the fact is a few words labeled!

Similar?



Similar?

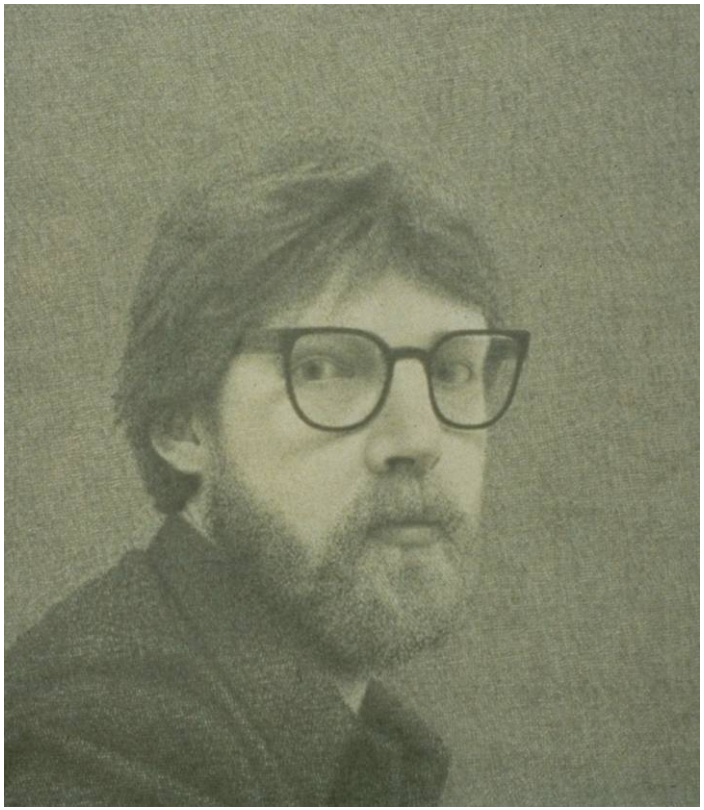


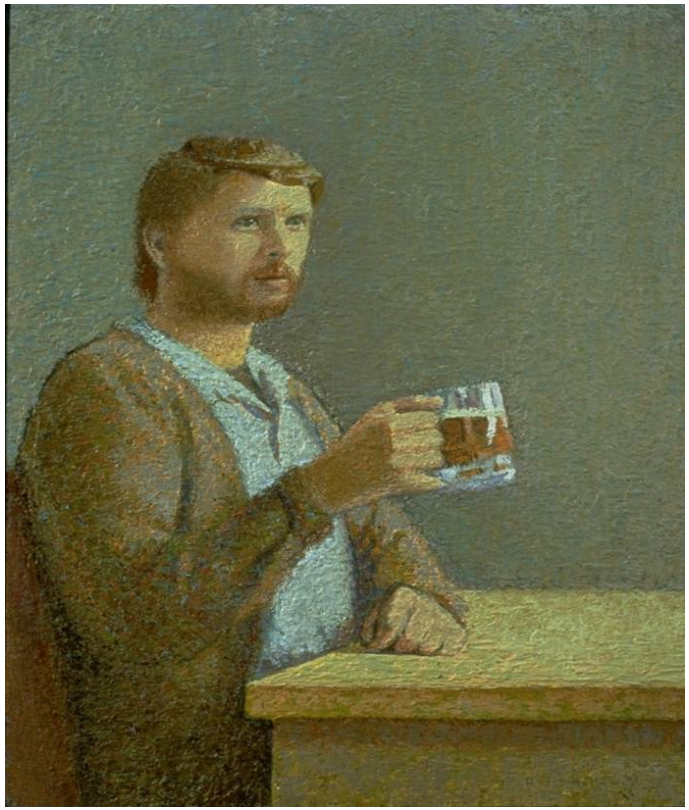
Similar?

Similar?

Similar?

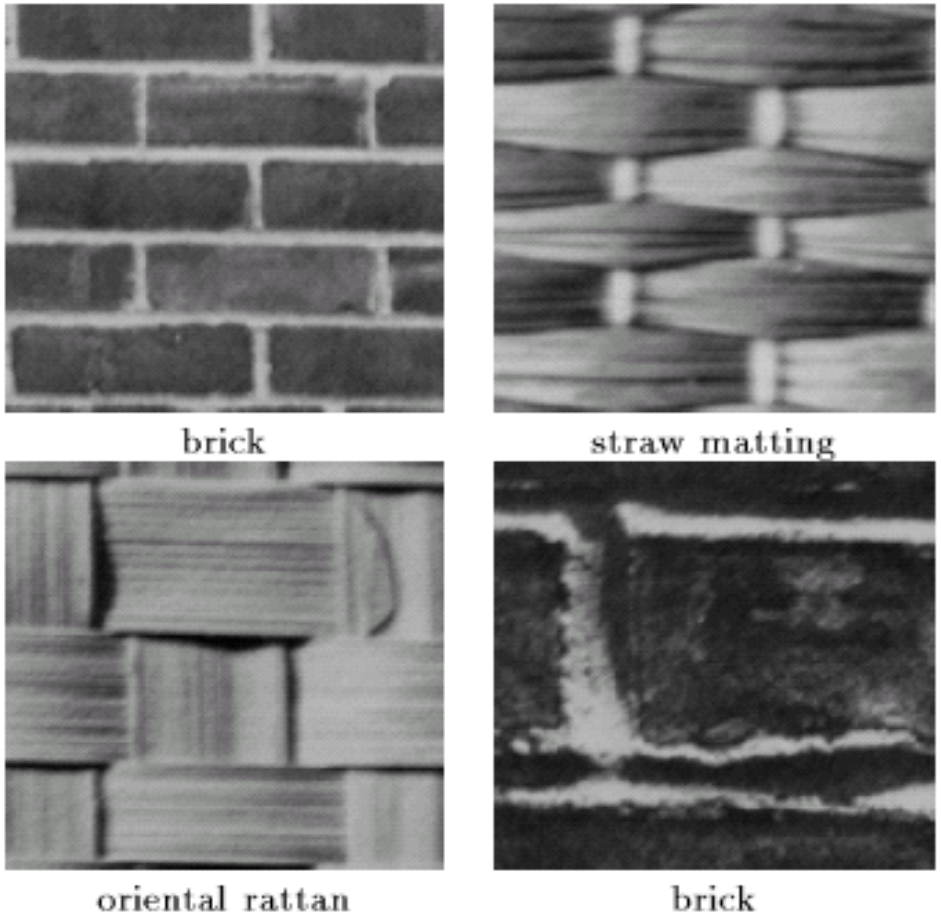
Similar?





Similar?

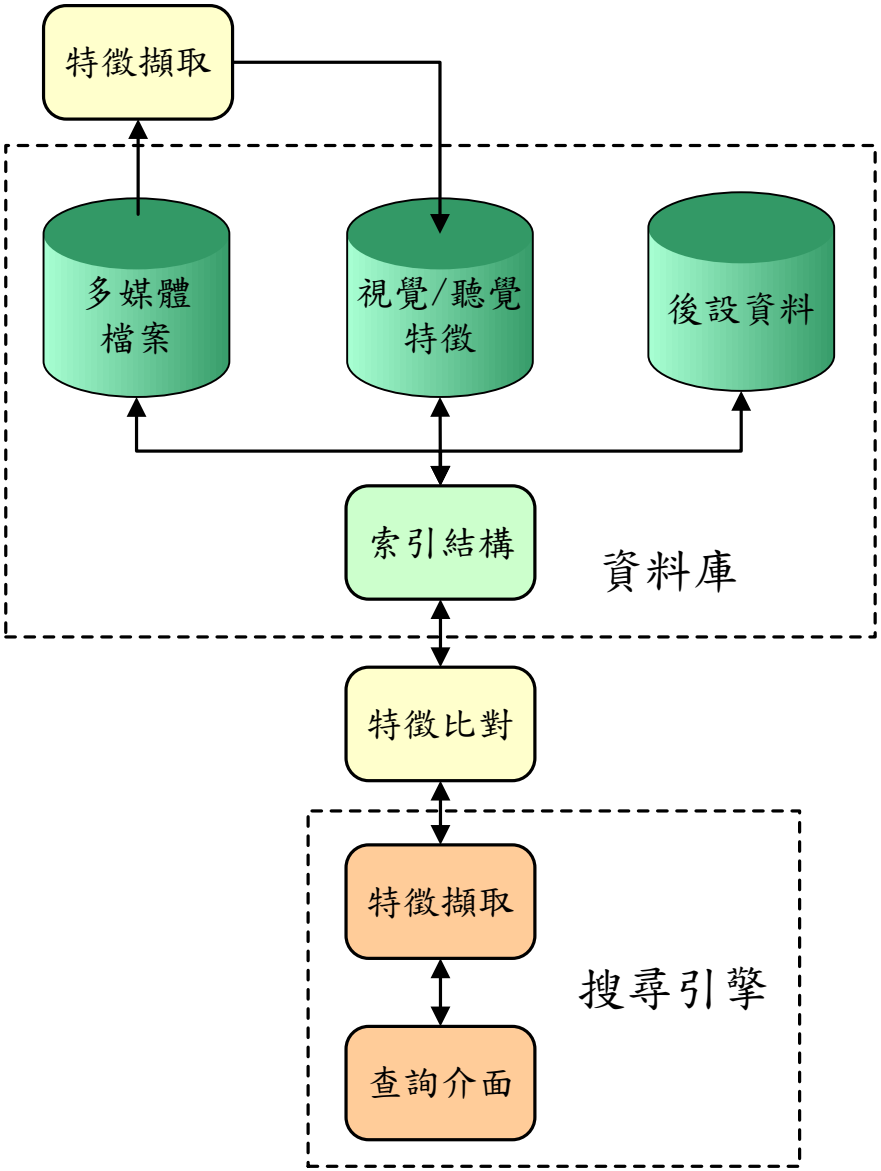
Similar?



[Picard95]

General Framework

- **Query interface**
  - Query type support
- **Browsing**
  - Data collection Navigation
- **Feature extraction**
  - Audiovisual perception features
- **Indexing**
  - Similar to the IR model
- **Similarity measurement**
  - Similar to the IR model
- **Relevance feedback**
  - Similar to the IR model

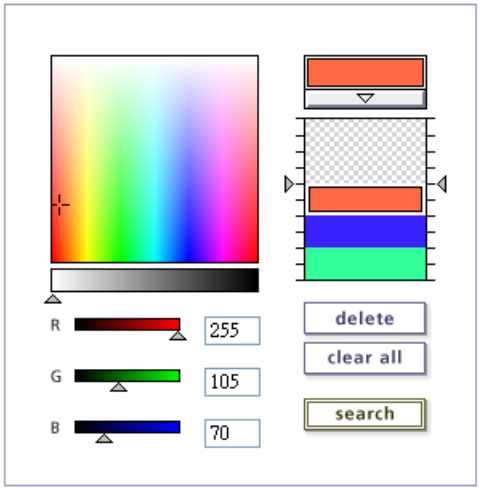


Query Interface in Content-based Retrieval

- Direct query (or concept query)
  - Use user interface to create a query (i.e., sketches, drawing, etc.)
- Query by example
  - Provide a multimedia document as a query
- Relevance feedback
  - Select multiple relevant or irrelevant examples from the search result

Direct Query – QBIC Color Search (IBM)

QBIC COLOUR SEARCH



1. Use your mouse to select a colour from the palette.

2. Click the arrow button to add the colour to the bucket.

3. Slide the triangular handles on the bucket to adjust the percentage of this colour.

4. You may repeat this process until the bucket is full. When you are ready, click Search.

You may also use the Colour Mixer to adjust RGB (red, green, blue) values to use in your search.

Click Delete to remove a colour from the bucket. Click Clear All to empty the bucket.

SEARCH RESULTS

modify search

new search

1) Girl with Tulips

Matisse, Henri
1910

2) Music

Matisse, Henri
1910

3) Portrait of Suzanne Dufy, the Artist's Sister

Dufy, Raoul
1904

4) Game of Bowls

Matisse, Henri
1908

5) Flower Study

Shaykh 1777

6) Boats at Saintes-Maries

Gogh, Vincent van 1888

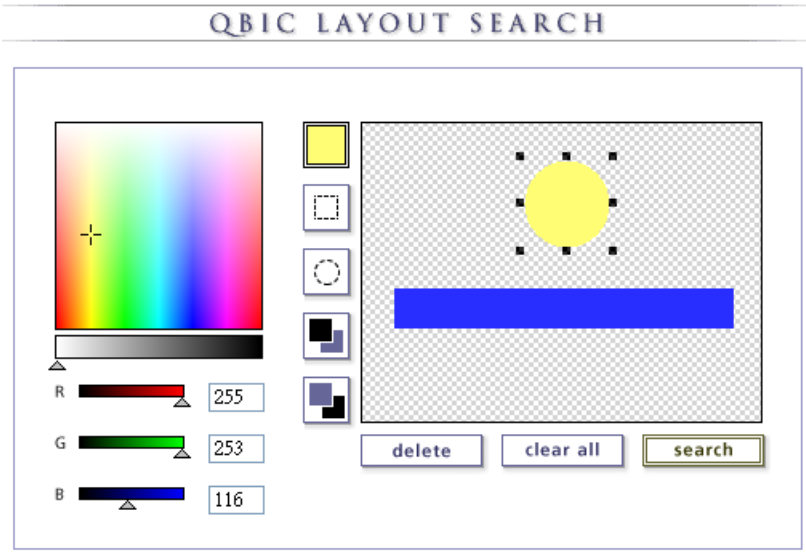
7) Ballerina

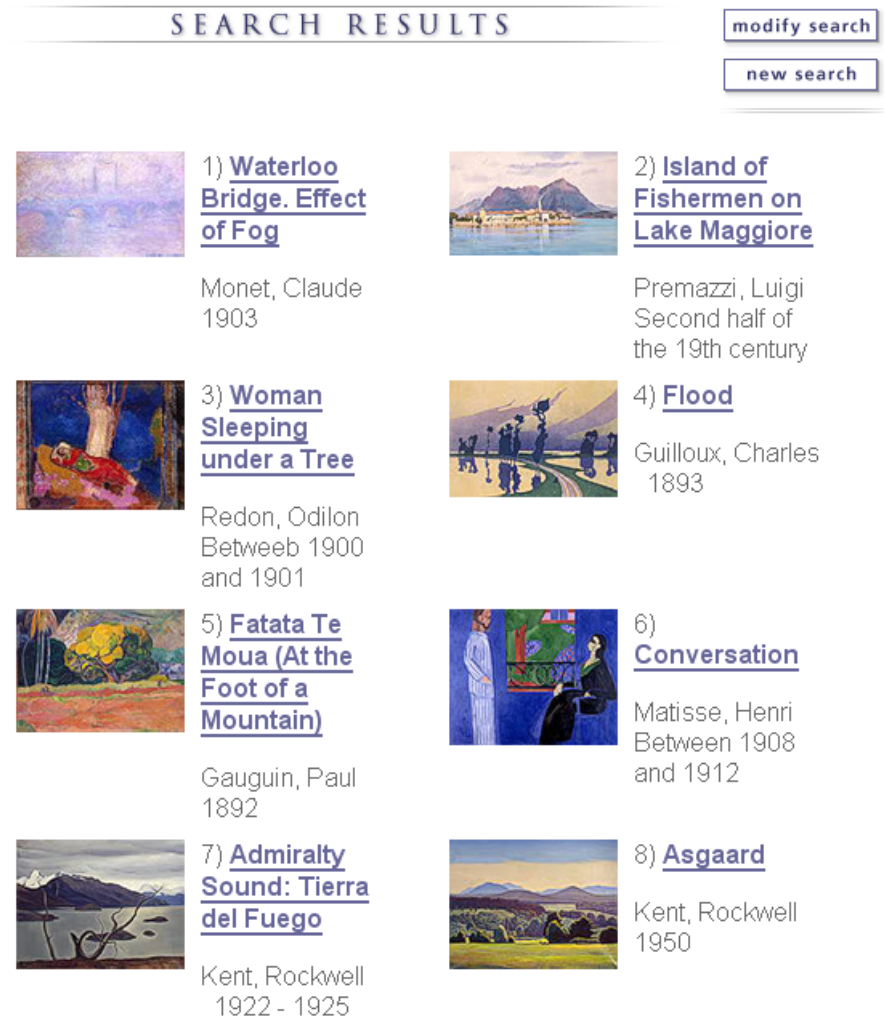
Matisse, Henri
Circa 1927

8) Portrait of a Man with a Newspaper (Chevalier X)

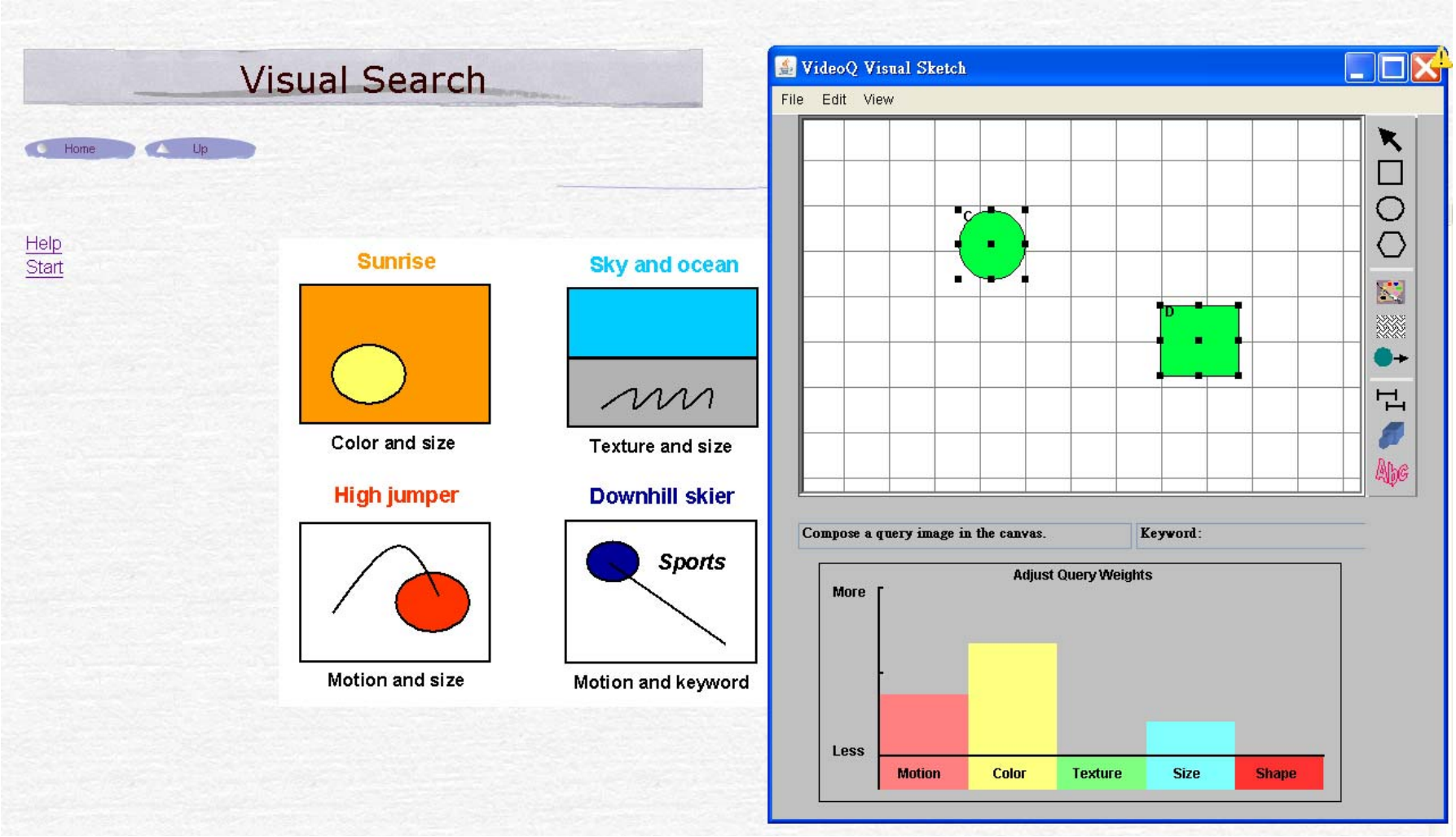
Derain, Andre
1911 - 1914

Direct Query – QBIC Layout Search (IBM)

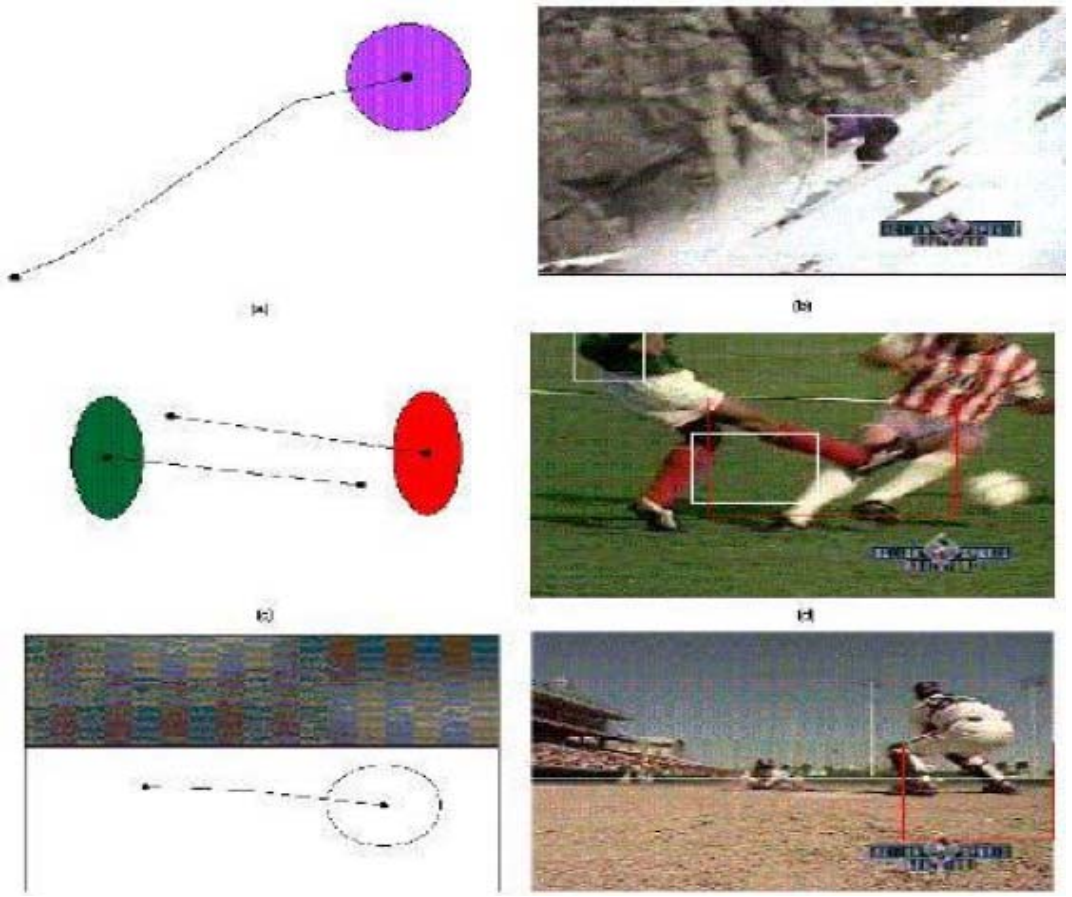




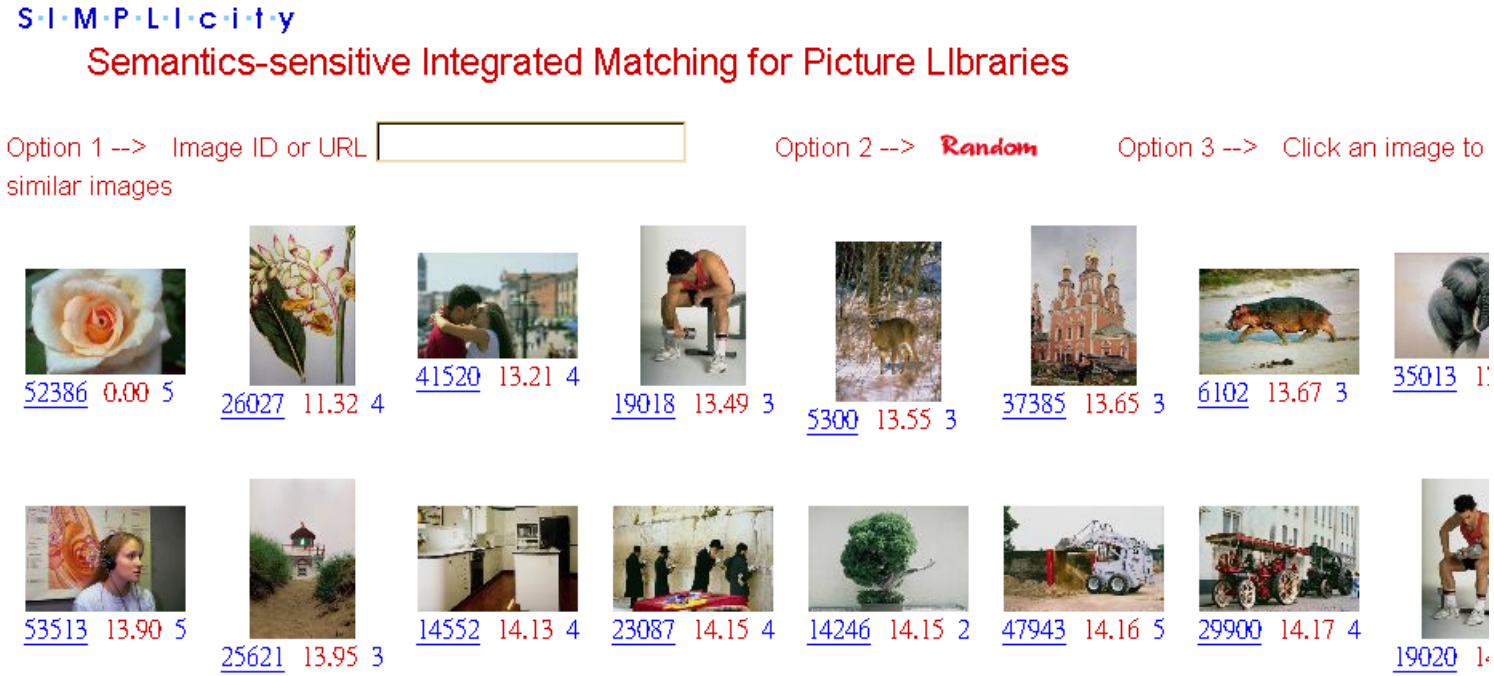
Direct Query – VideoQ (Columbia University)



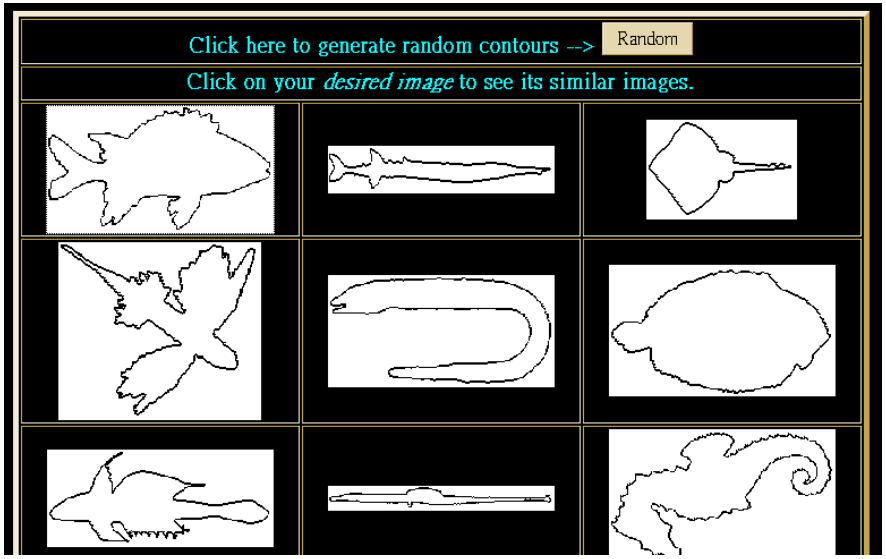
Direct Query – VideoQ (Columbia University)

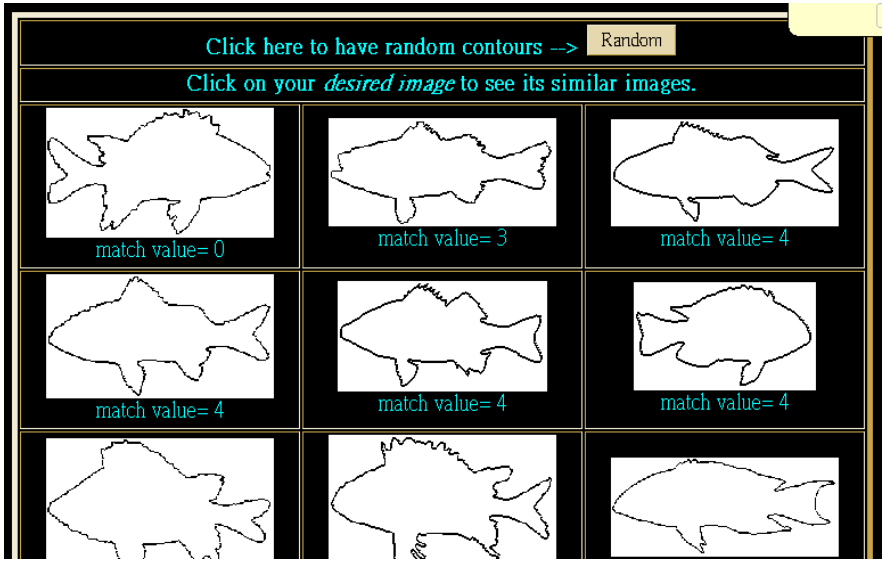


Query by Example – SIMPLICITY (Stanford University)

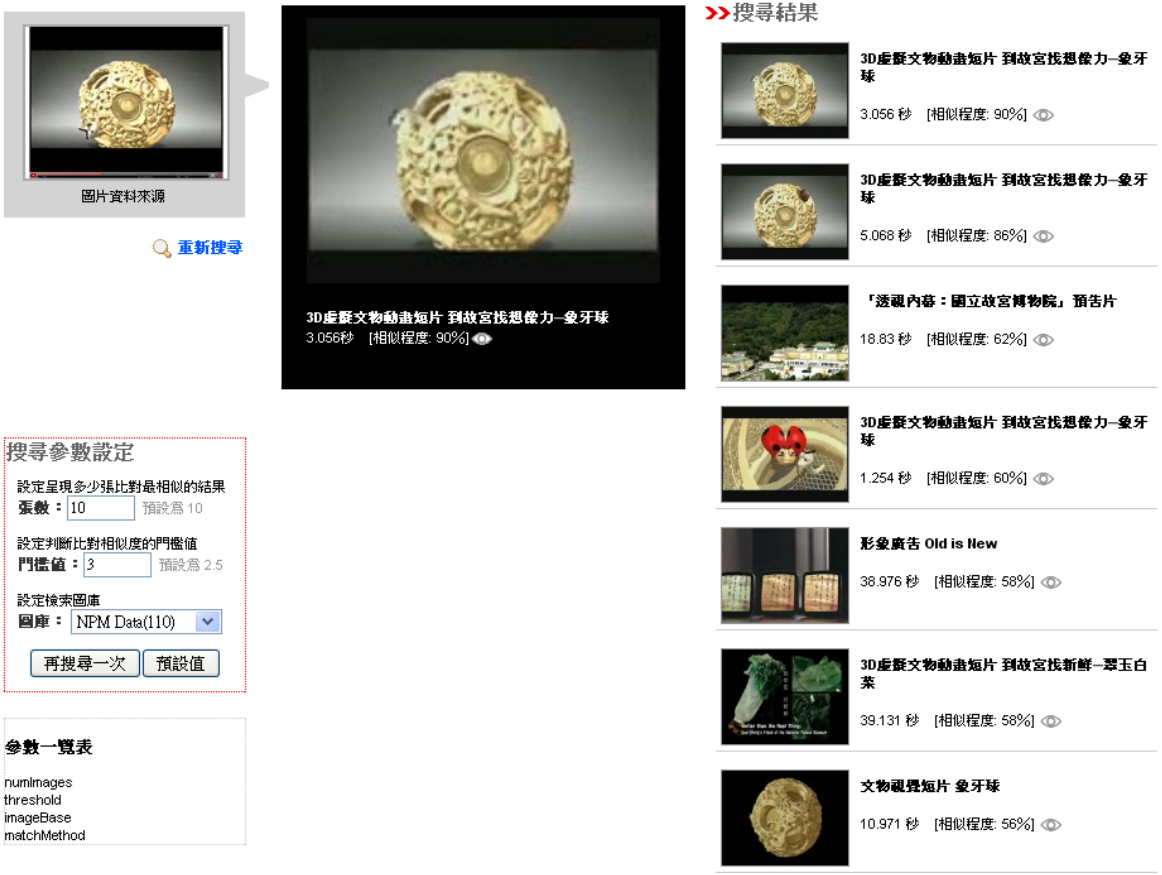


Query by Example – SQUID (Surry University)





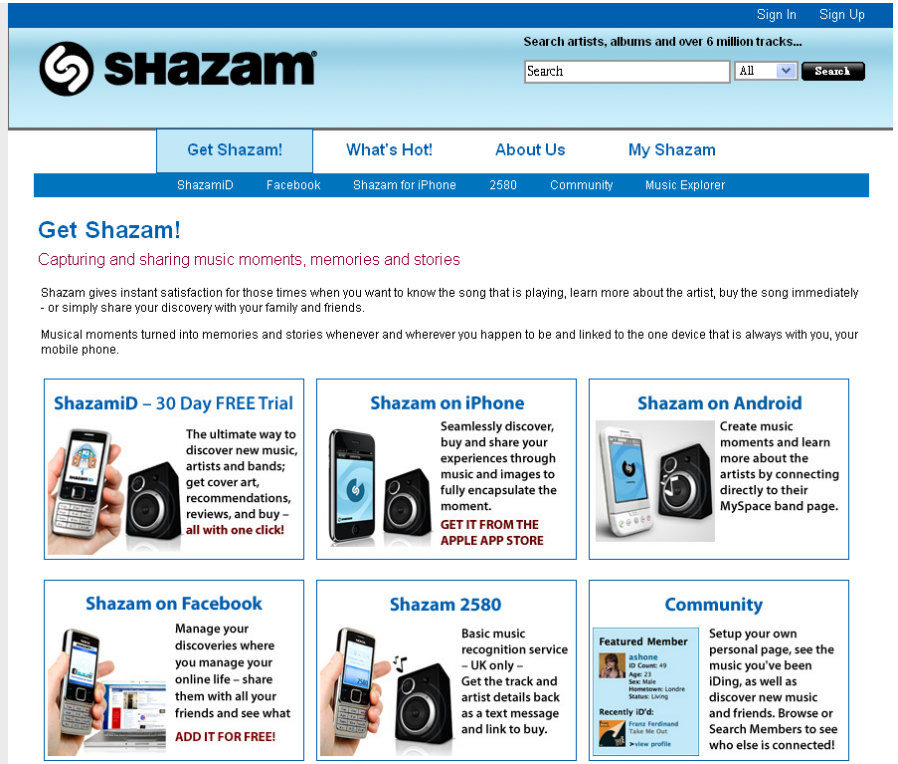
Query by Example – VCenter



http://vcenter.iis.sinica.edu.tw/lab/shotDetect/camera.html

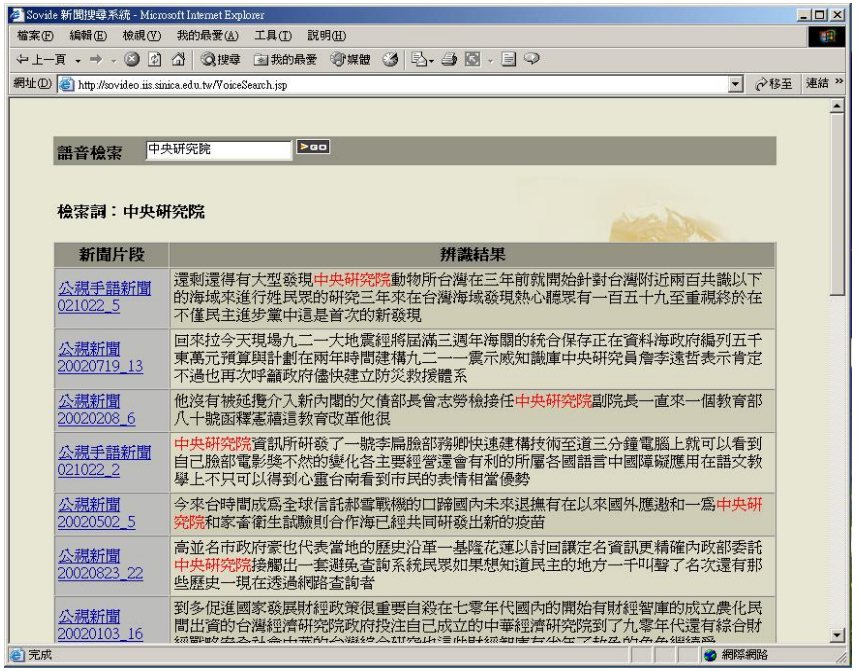
Query by Example - Shazam

- Find out what's playing right now by pointing your mobile at the music and uploading for search



Query by Spoken Content

- 中研院資訊所王新民副研究員發展的語音辨識技術，可以用來檢索影片中的語音內容



Relevance Feedback – 新視科技

Relevance Feedback – 新視科技

Relevance Feedback – 新視科技

Browse Multimedia Content

- Subject browsing is the most popular user operation in interactive multimedia retrieval

- Browsing is helpful when ...
  - Navigating the multimedia collection
  - Specifying an example as the query
  - Comparing the retrieval result

A Case Study: WebSEEk

- An image and video search engine for the Web
  - Information is collected by automated agents
  - Text and visual features are both processed
  - Media objects are classified
- Using URLs, alt tags and hyperlink text to create subject taxonomy (2128 classes)
- Using color histograms to classify image types
  - Color photo, color graphic, gray image, B/W image

Image Taxonomy in WebSEEk

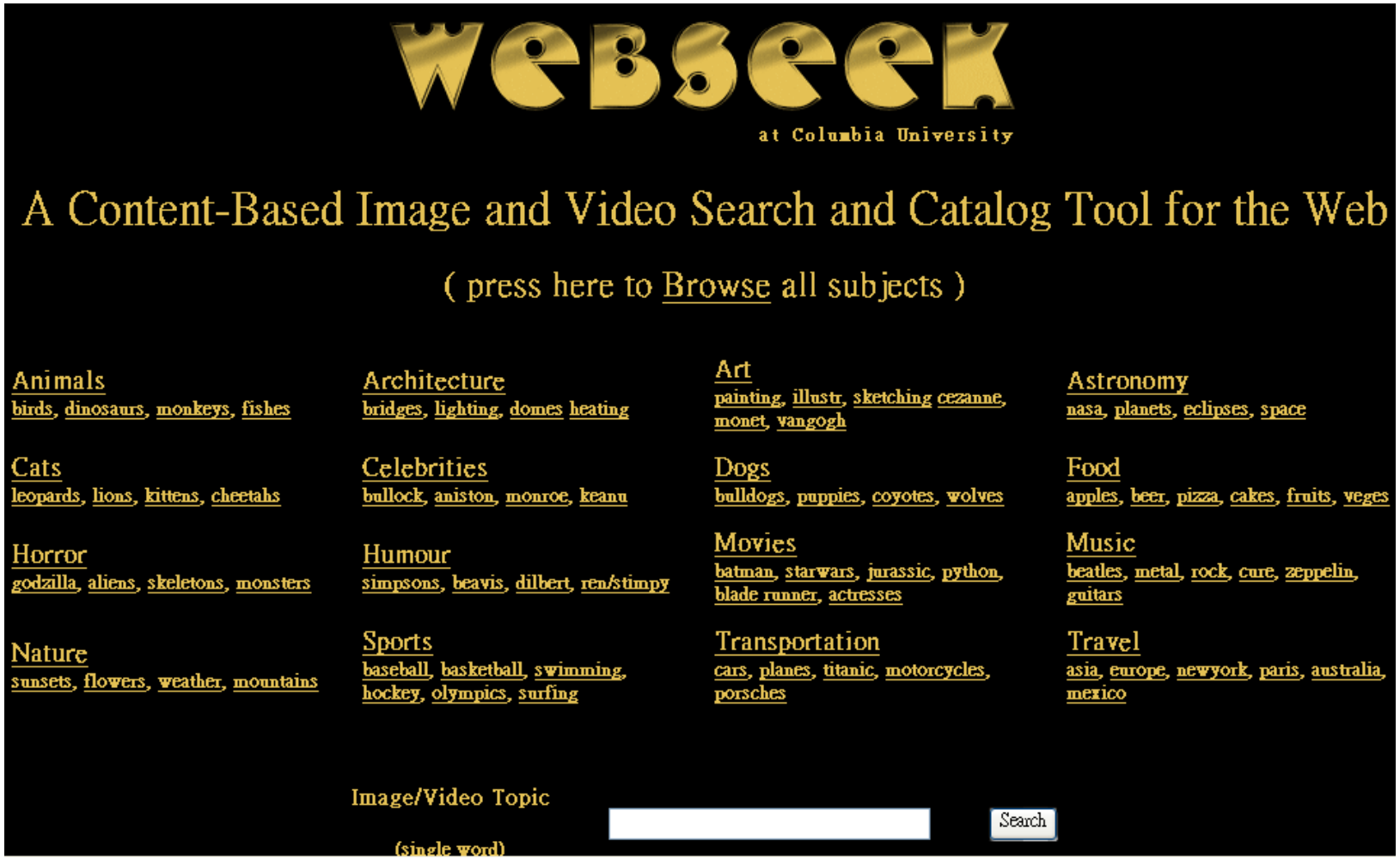
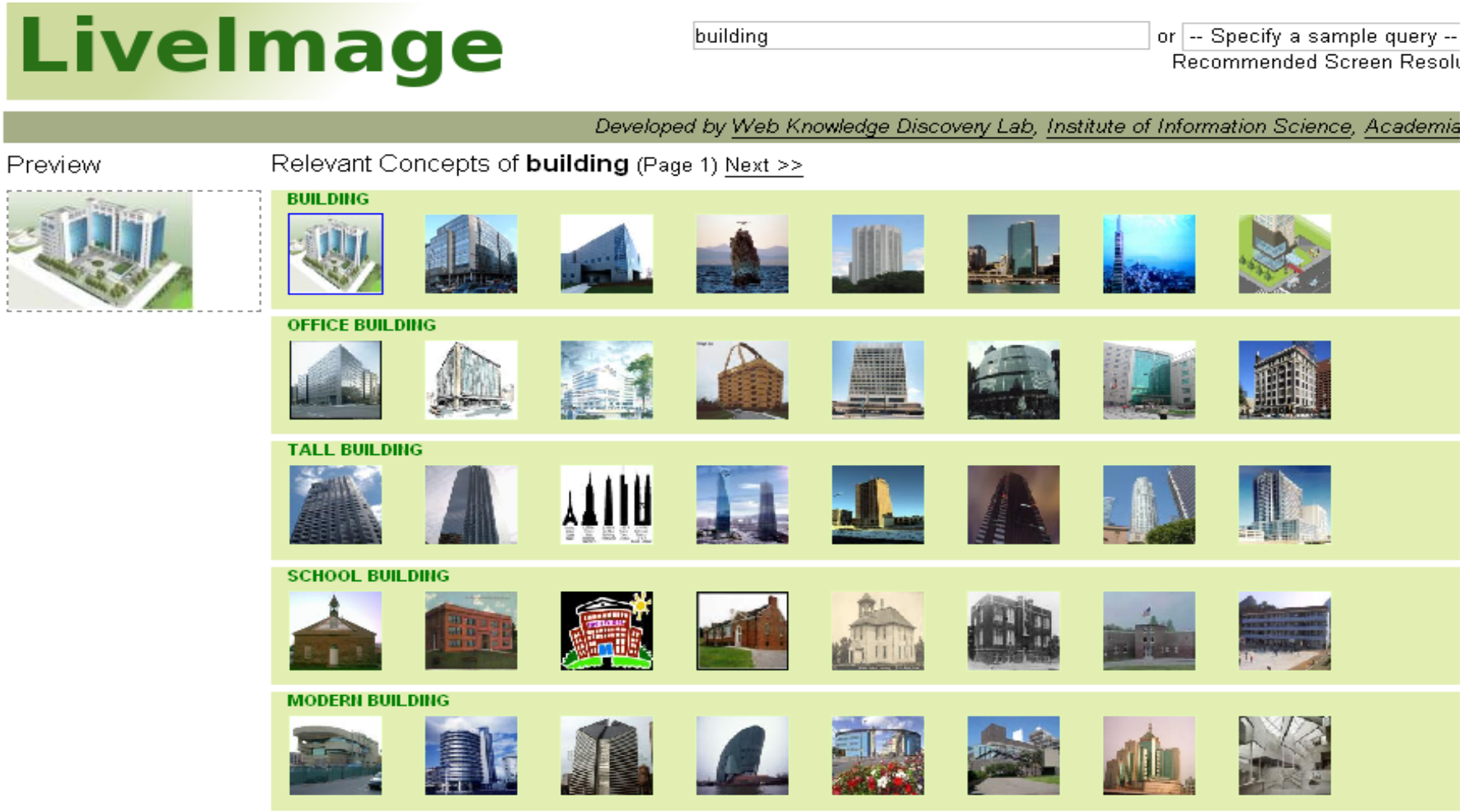
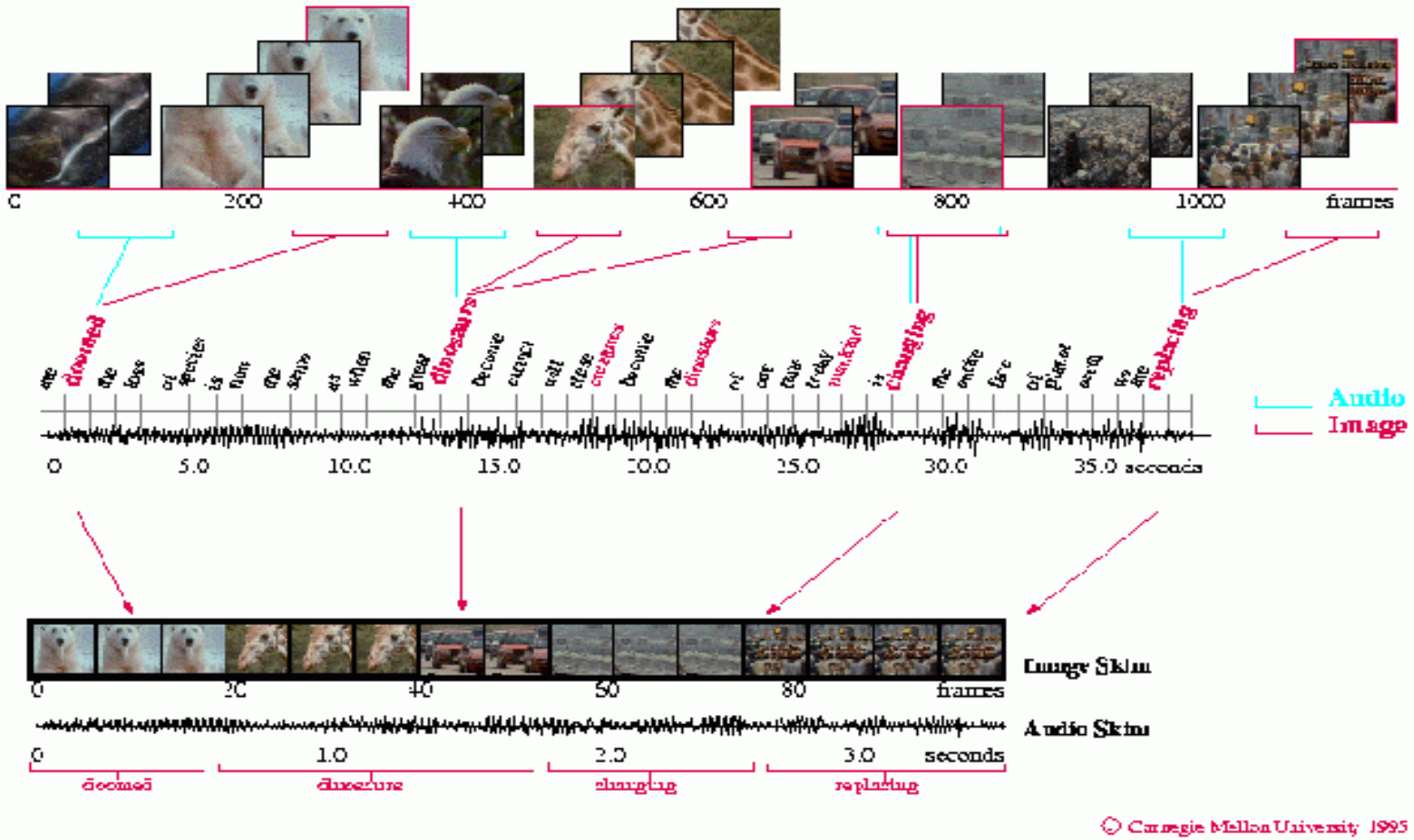


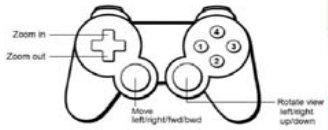
Image Classification

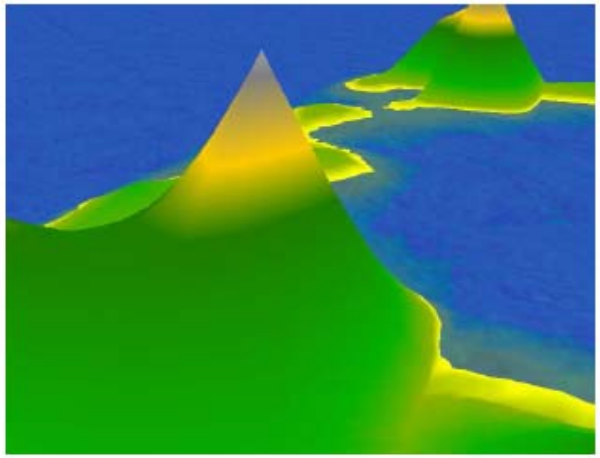


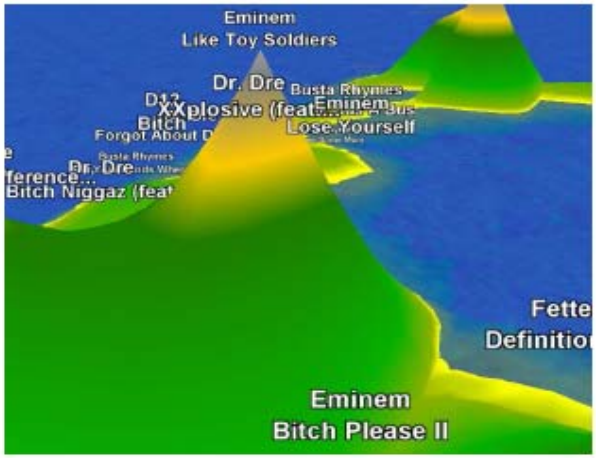
Video Summarization

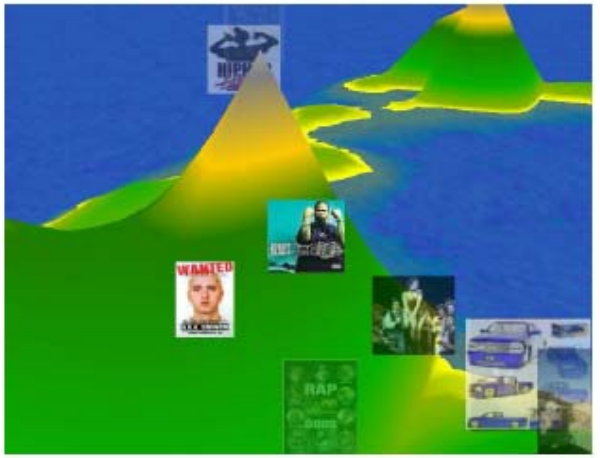


Music Exploration [Knees2006]









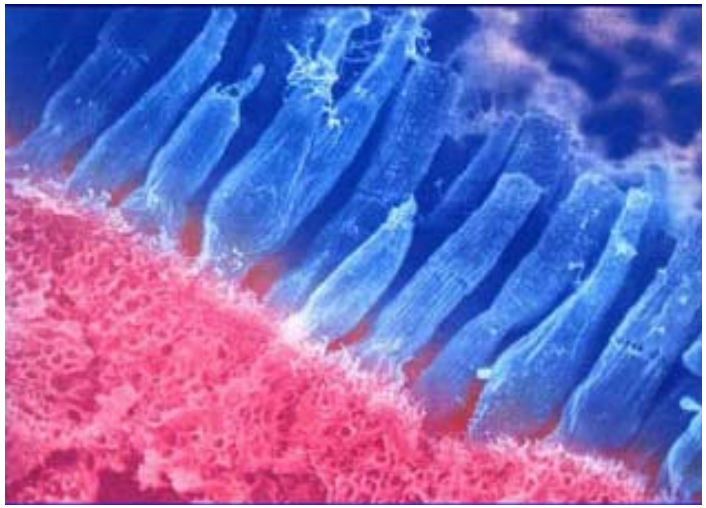
Feature Extraction

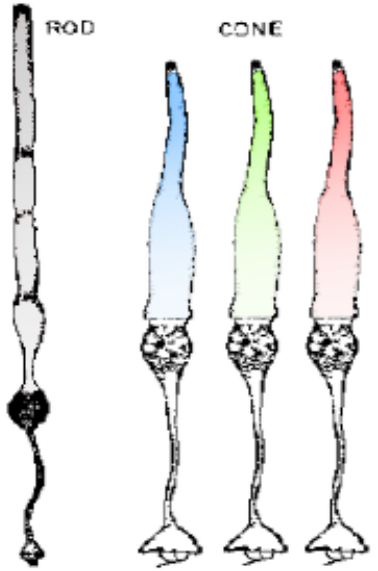
- **Image**
  - Color, texture, shape, spatial layout
- **Audio**
  - MFCC, note frequency, rhythm, speech
- **Video**
  - Motion, image, audio

MPEG-7

- **Multimedia content description interface**
- **Standardize tools** (descriptors, description schemes, and language) of audiovisual information
- Enable fast and efficient **content searching, filtering, and identification**
- Proposed by ISO/IEC

Visual Primitive





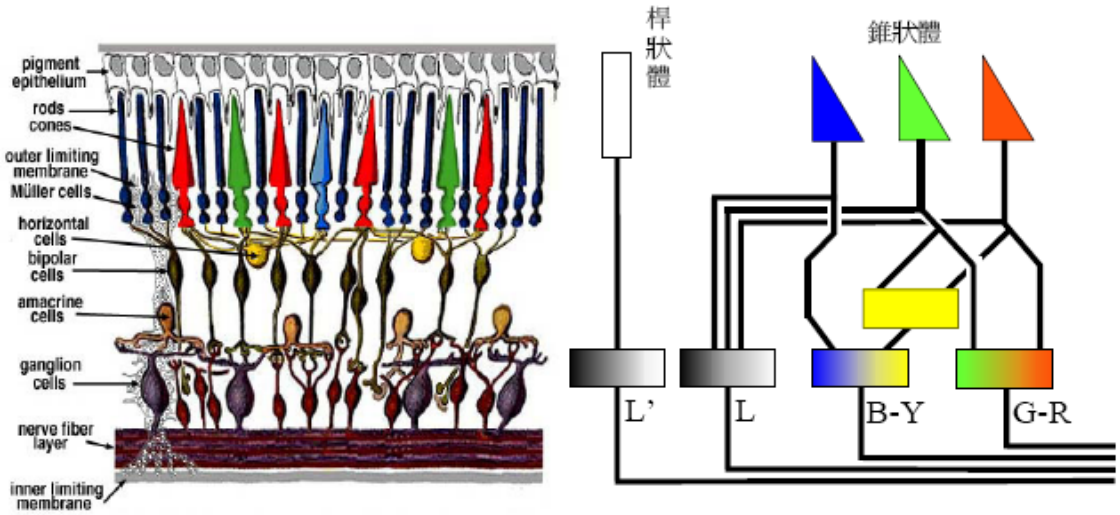
網膜下的桿狀細胞與椎狀細胞

Visual Primitive

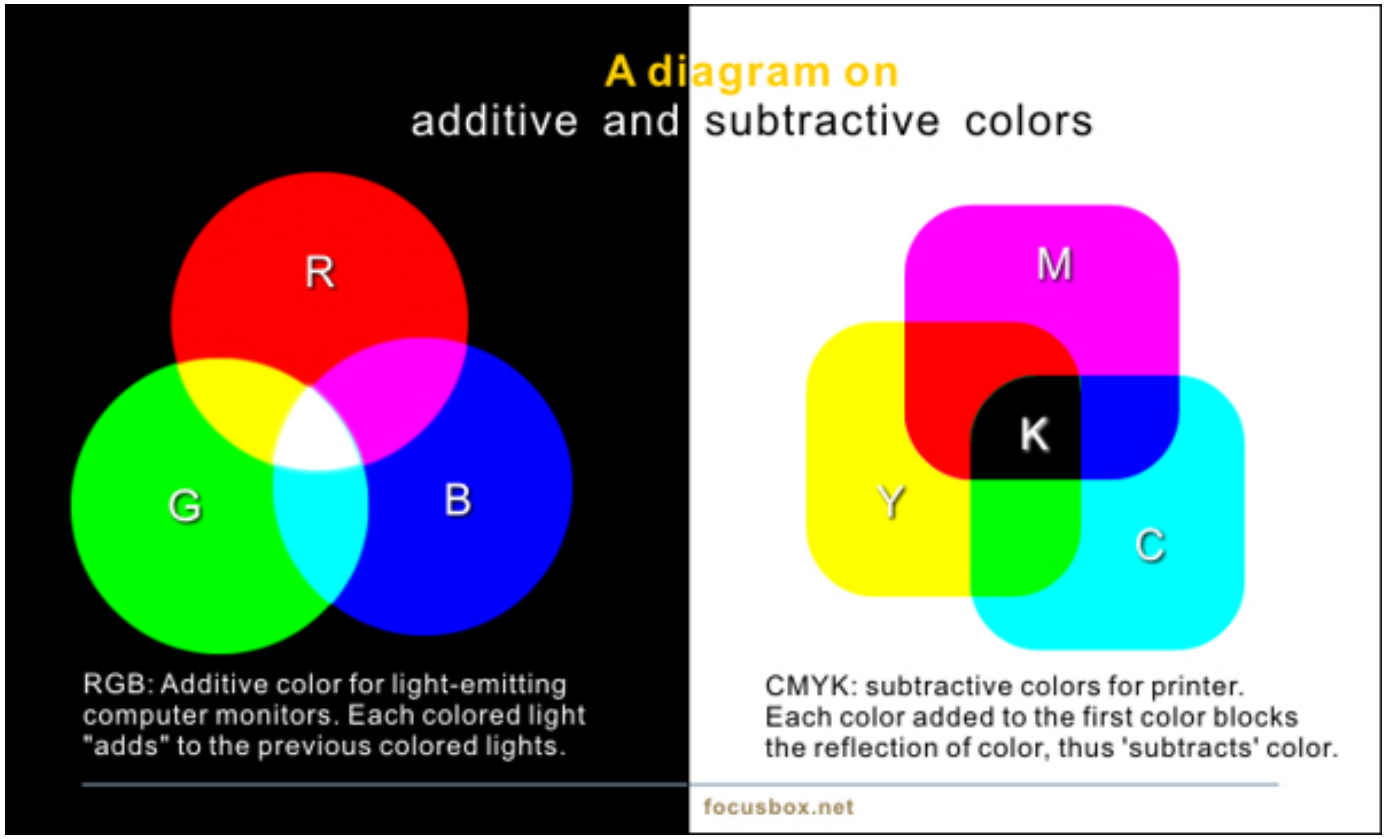
- **三原色說(trichromatic theory, by Young and Helmholtz)**
  - 人眼內含紅、綠、藍三種受光器，色覺是由個受光器對光產生的反應
- **對立色說(opponent-colors theory, by Hering)**
  - 人眼內含紅-綠、黃-藍、白-黑三種受光器

Visual Primitive

- **階段說(Stage theory, by Adam et al.)**
  - 初期階段為三原色，後期階段轉換為對立色

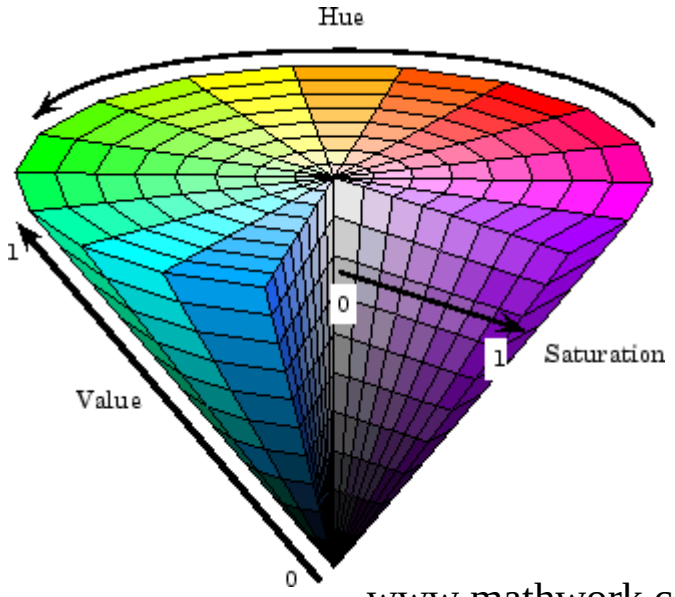


色光加法混色　　色料減法混色



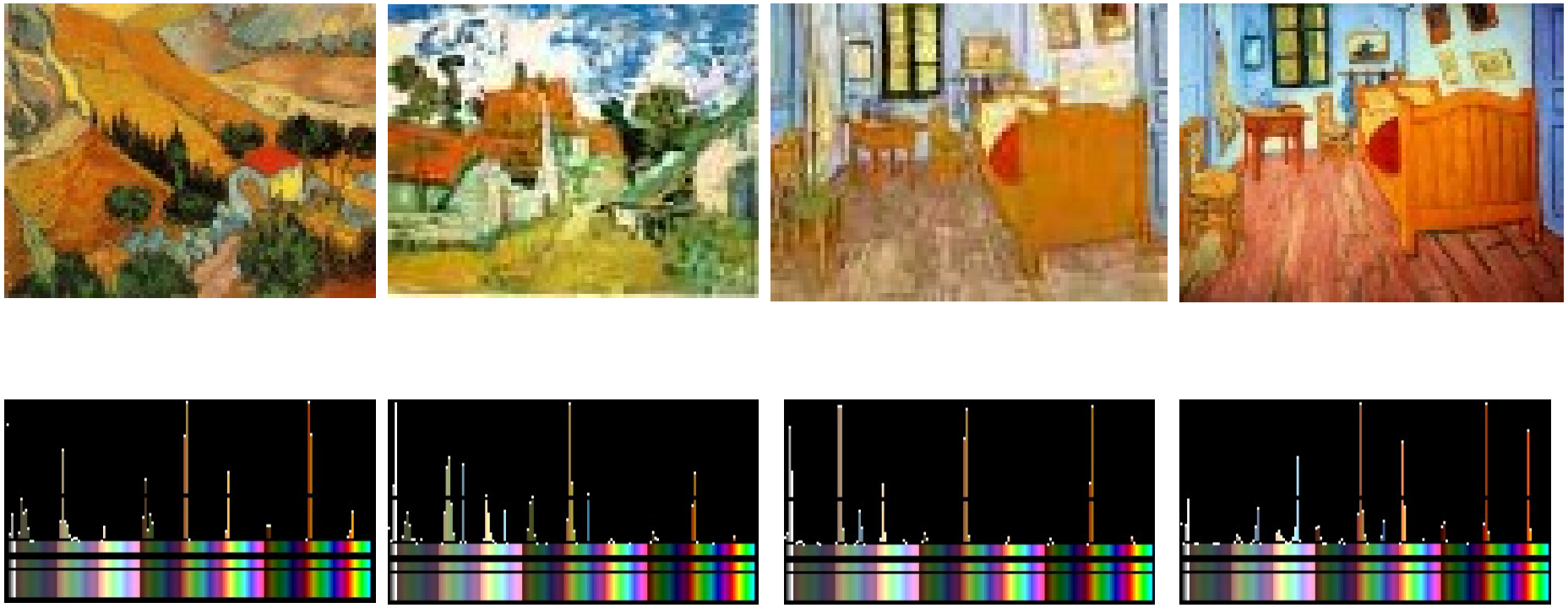
Color Space

- RGB
- YCbCr (digital system)
- YUV (analog system)
- YIQ (for NTSC)
- Lab
- HSV
- ...



www.mathwork.com

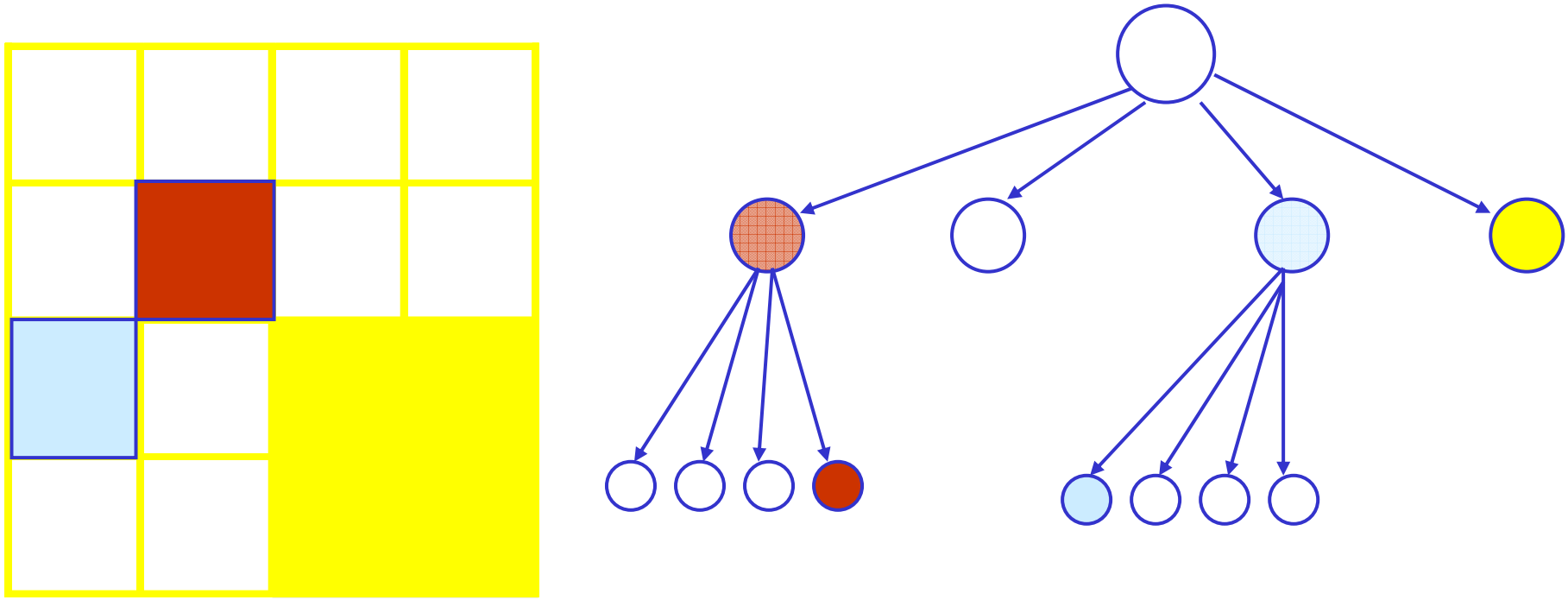
Color Histogram



Color Layout

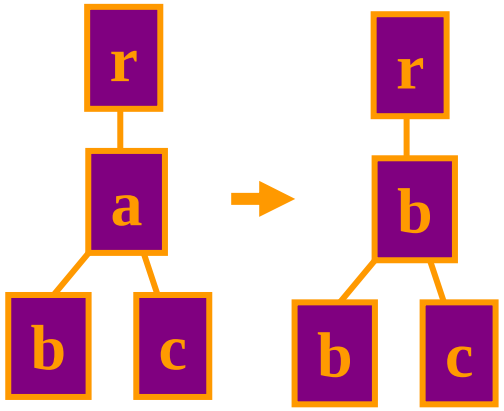
- More accurate color description
  - Histogram do not reflect the spatial information
- Quad-trees representations
  - Similarity computation between two quad-trees

Quad-tree Representation

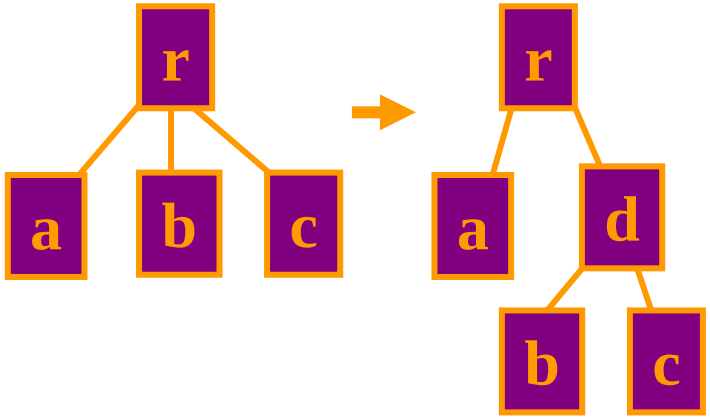


Quad-tree Similarity

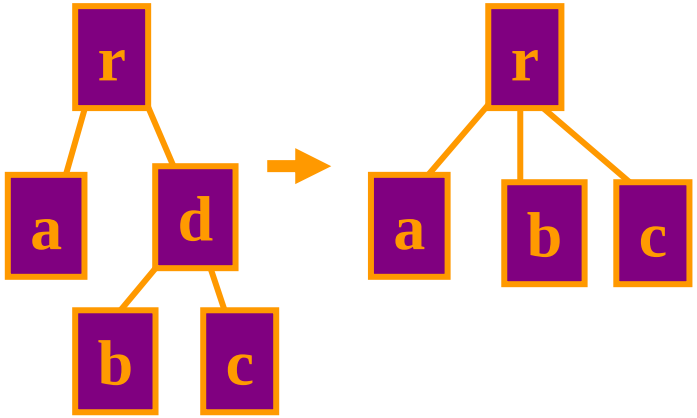
Replace



Insert

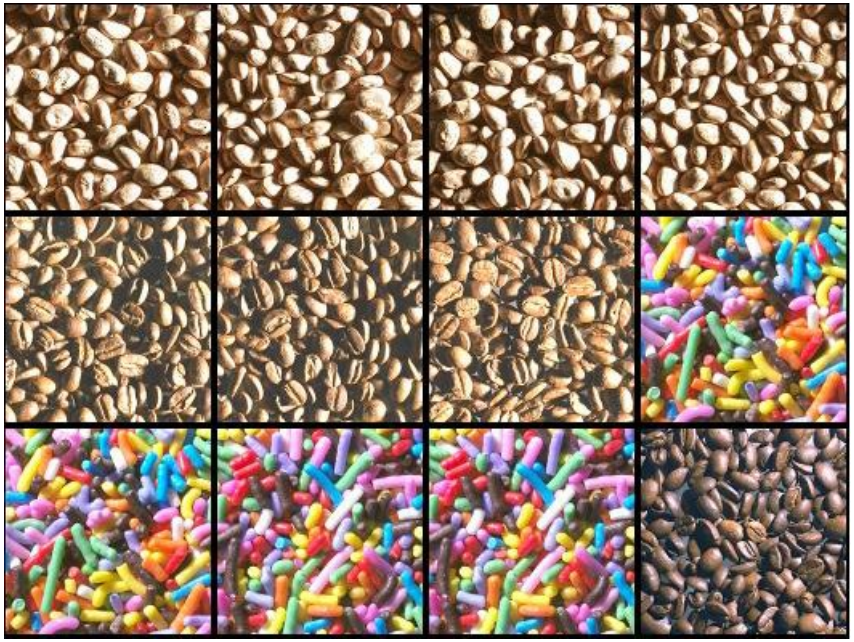


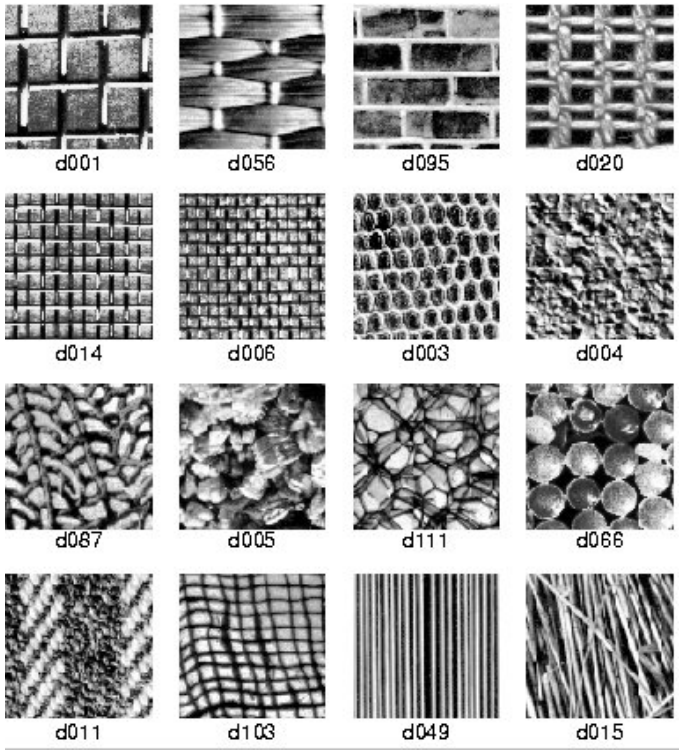
Delete



Texture

- The structural and statistical properties of brightness pattern





Texture Feature

- Statistical property
  - Co-occurrence matrix
    - Energy, entropy, contrast, homogeneity, tendency
  - Tamura texture (QBIC, MARS)
    - Coarseness, contrast, directionality, line-likeness, regularity, roughness
- Structure property
  - Wold feature (MIT Photobook)
    - Periodicity, directionality and randomness
  - Gabor wavelet (MPEG-7)
    - Regularity, directionality, scale

Shape

- **What do users want?**

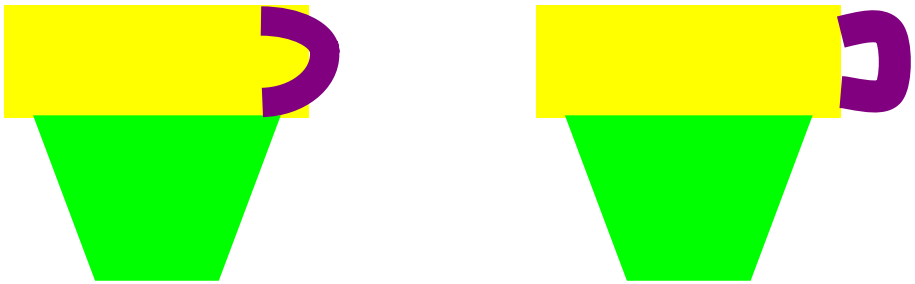
Question	*Yes*	*No*	**NA**
"Did you need an image ..."			
"...with a particular object on it?"	**122**	**41**	**7**
"...with a particular color on it?"	**25**	**137**	**8**
"...with a particular texture on it?"	**23**	**137**	**10**

(results of a WWW questionnaire, N=170 responses)

- A survey on WWW revealed that users are interested in objects (71%) and not in layout, texture or abstract features

Shape

- Objects are best recognized from **canonical views** [Blanz99]



Which is the canonical view?

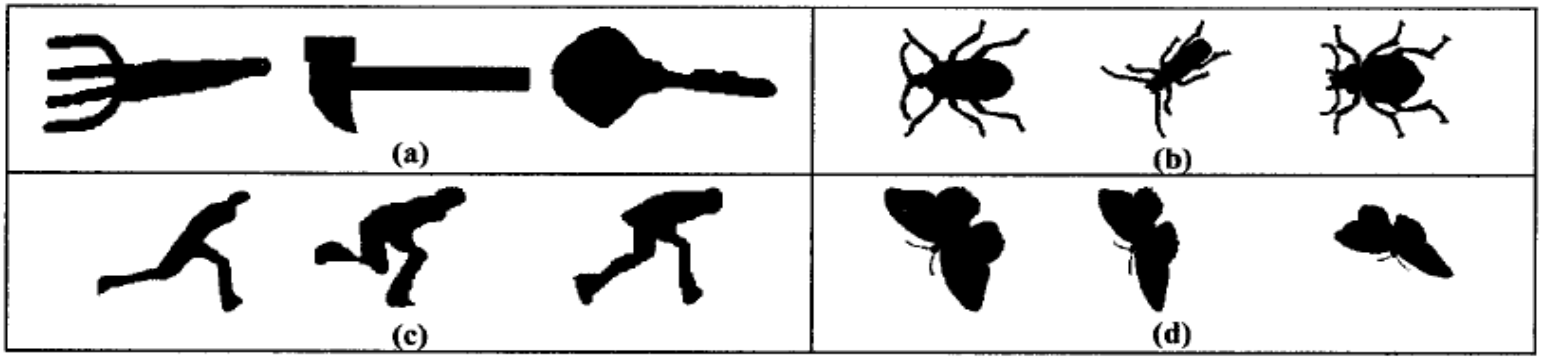
Shape Feature

- Preprocessing by image processing
  - color-based segmentation algorithm (dominate object)
  - Edge detection algorithm, outlining algorithm

- Representations (should be **affine invariant**)
  - Geometric attributes: area, perimeter, centroid, orientation (region-based)
  - Shape moments (region-based)
  - Chain codes, differential chain codes (contour-based)
  - B-splines (contour-based)
  - Curvature scale-space (contour-based)

Region-based

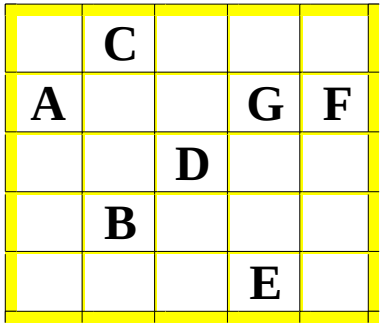
Contour-based



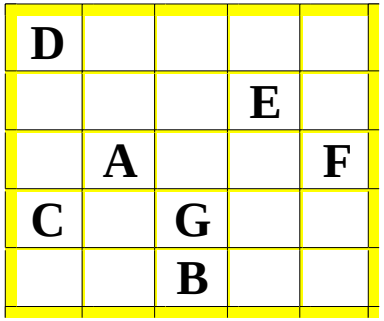
Spatial Layout

- Advantage
  - Spatial queries, spatial databases
- Representation
  - 2D string [Chang87], 2D C string [Lee90], …
  - Directional and topological relation

2D String Representation



(A<B=C<D<E=G<F, E<B<D<A=G<F<C)



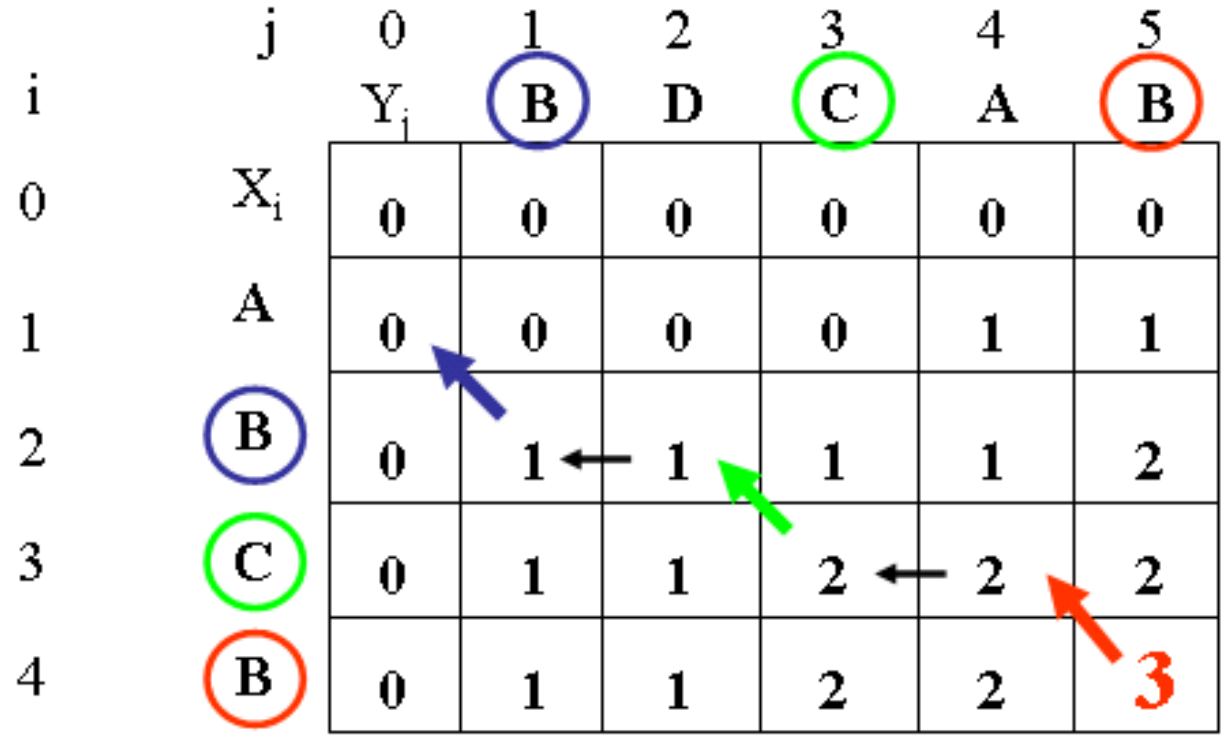
(C=D<A<B=G<E<F, B<C=G<A=F<E<D)

2D String Similarity

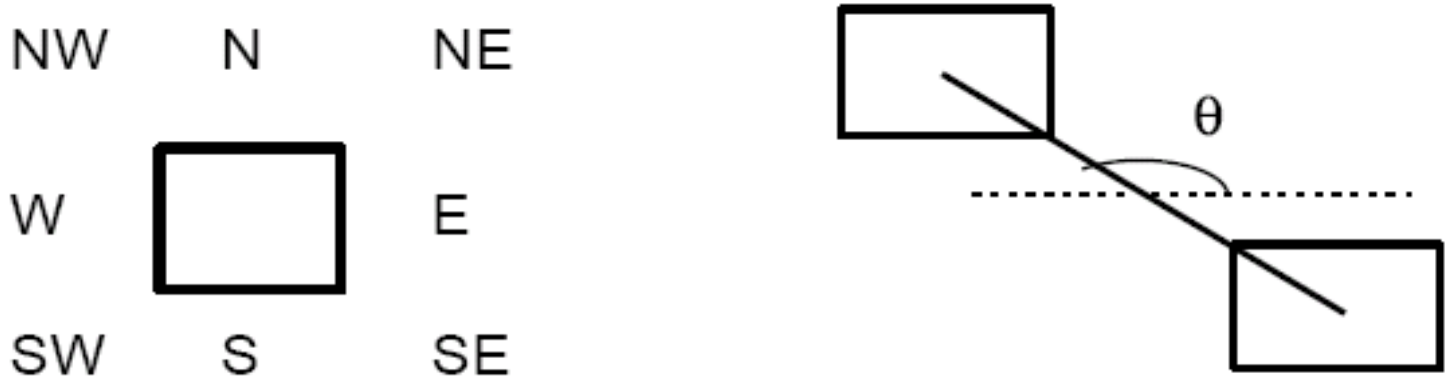
- Longest common subsequence (LCS)
  - To find the maximum possible length of a common subsequence of two strings
  - X=*abcdgh*, Y=*aedfhr*, LCS(X,Y)=*adh*
  - A dynamic programming technique

$$S[i, j] = \max \begin{cases} S[i-1, j-1]+1 & \text{if } x[i] = y[j], \\ \max(S[i, j-1], S[i-1, j]) & \text{otherwise} \end{cases}$$

i	j	0	1	2	3	4	5
		Y_j	B	D	C	A	B
0	X_i	0	0	0	0	0	0
1	A	0	0	0	0	1	1
2	B	0	1	1	1	1	2
3	C	0	1	1	2	2	2
4	B	0	1	1	2	2	**3**



Directional and Topological Relation



(a) Strict and positional directional relations (b) Slope directional relations

Fig. 2. Directional relations.

Fig. 3. Topological relations.

How People Organize Images in Computers [Rodden99]

• **Adding a title to an image**	**2.0**
• **Typing notes to associate with an image or group of images**	**1.8**
• Speaking notes to associate with an image or group of images	2.5
• Having spoken notes automatically recognized	2.2
• **Searching for images based on the text of your notes**	**1.5**
• Searching for images based on colors present in them	3.2
• Searching for images based on textures present in them	3.0
• Searching for images based on their layout/composition	2.3
• Searching for other images similar to a given one	2.3
• Searching for other images similar to a drawing	3.3
• **Seeing all of the images in a class at once, reduced in size**	**1.2**
• **Rapidly scanning through a group of images**	**1.6**
• Seeing a very large number of images at once, with similar images clustered together	2.1

(1: very useful 4: not at all useful)

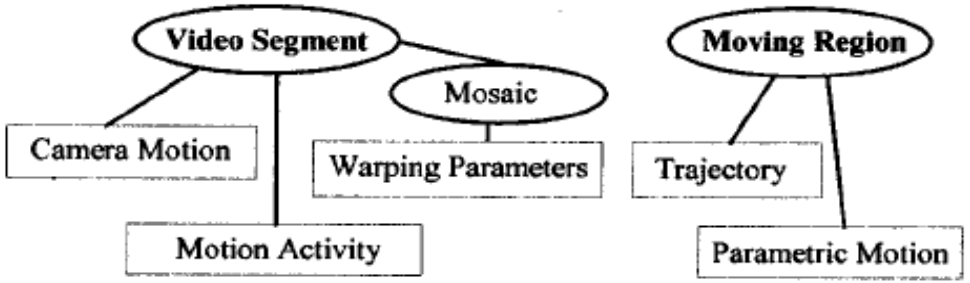
Motion

- The motion features of a video sequence provide the easiest access to its **temporal dimension**

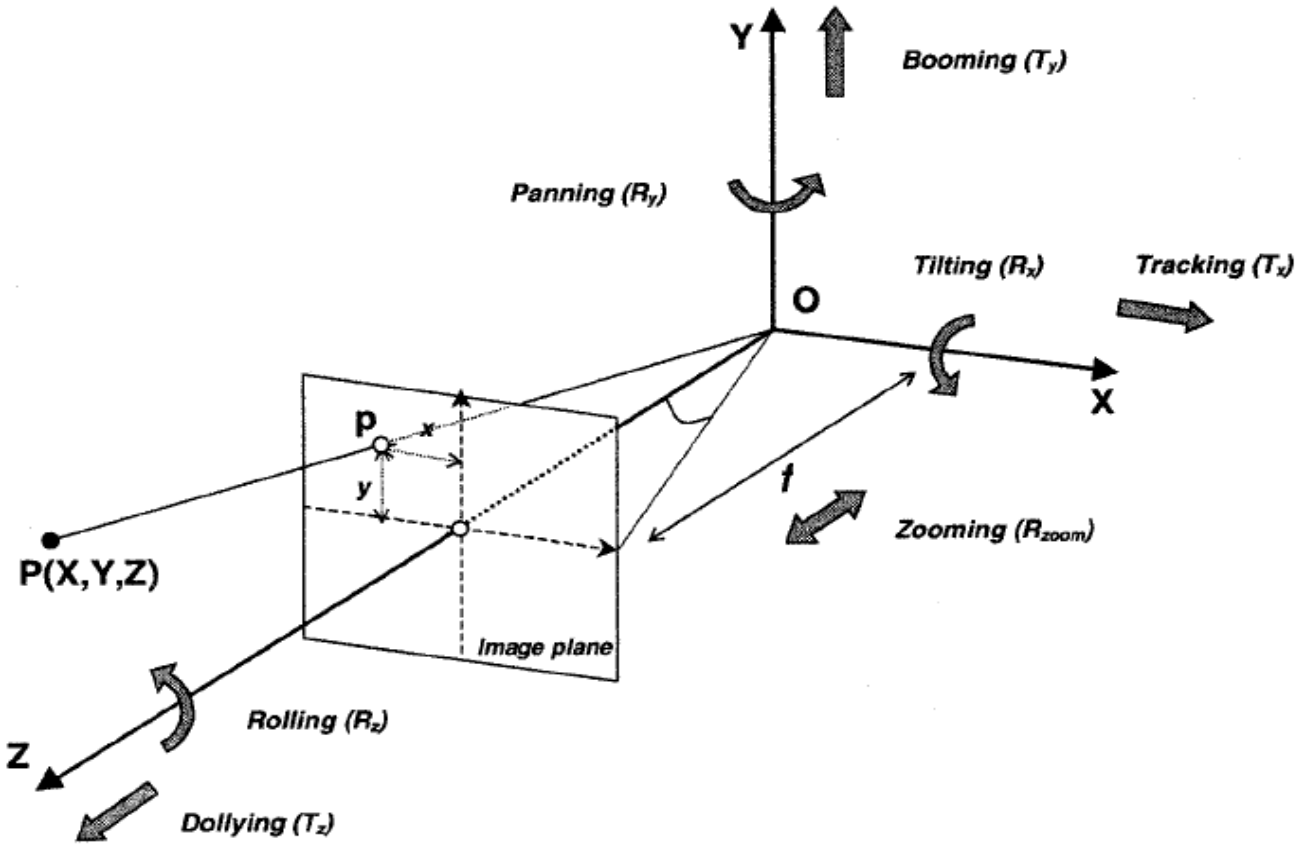
- In combination with indexing based on still-image features, motion-based indexing is useful in contexts where motion has a **higher-level semantic meaning**

Motion Feature

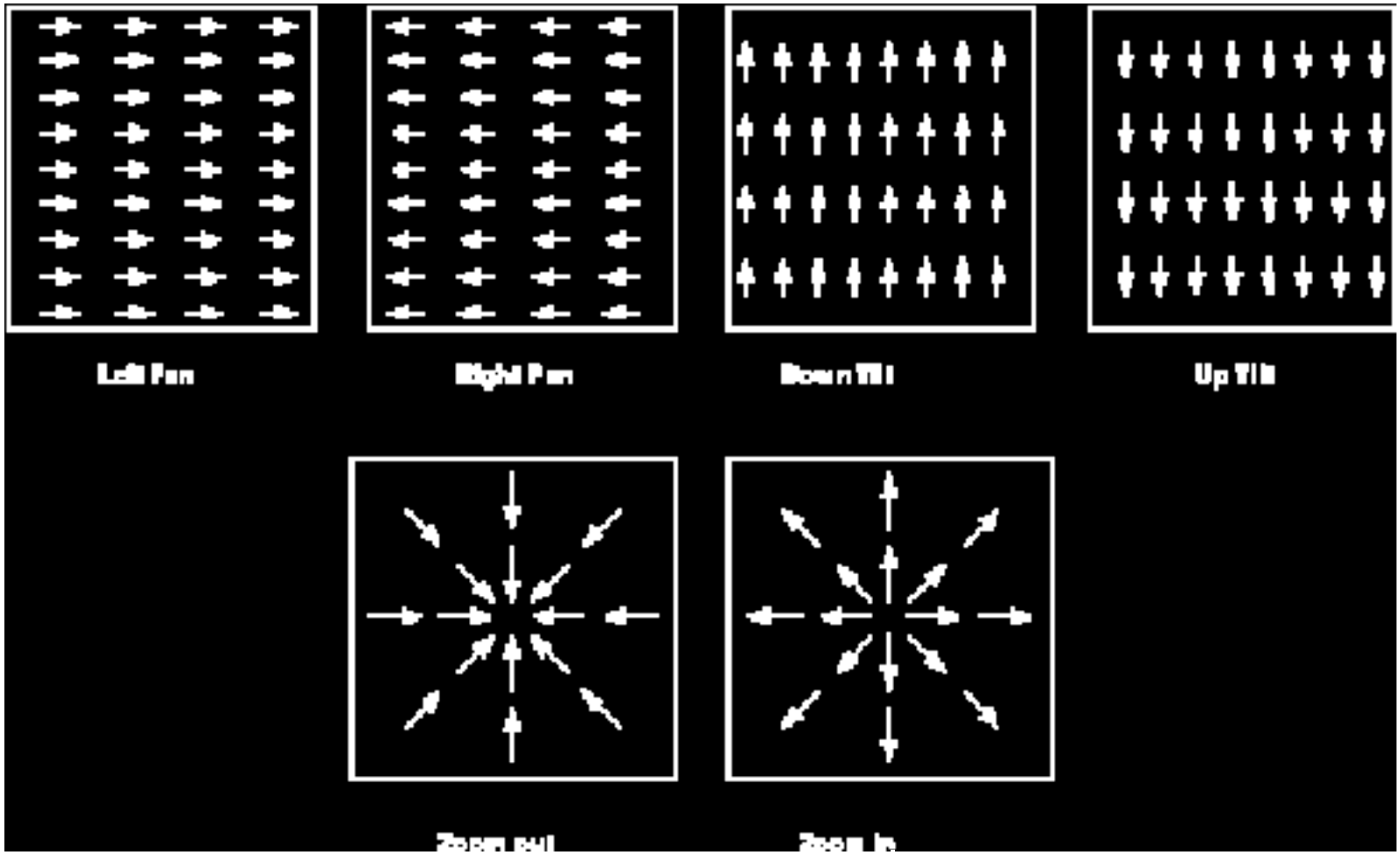
- Global motion
  - Motion activity
  - Camera motion
  - Warping parameters
- Local motion
  - Motion trajectory
  - Parametric motion



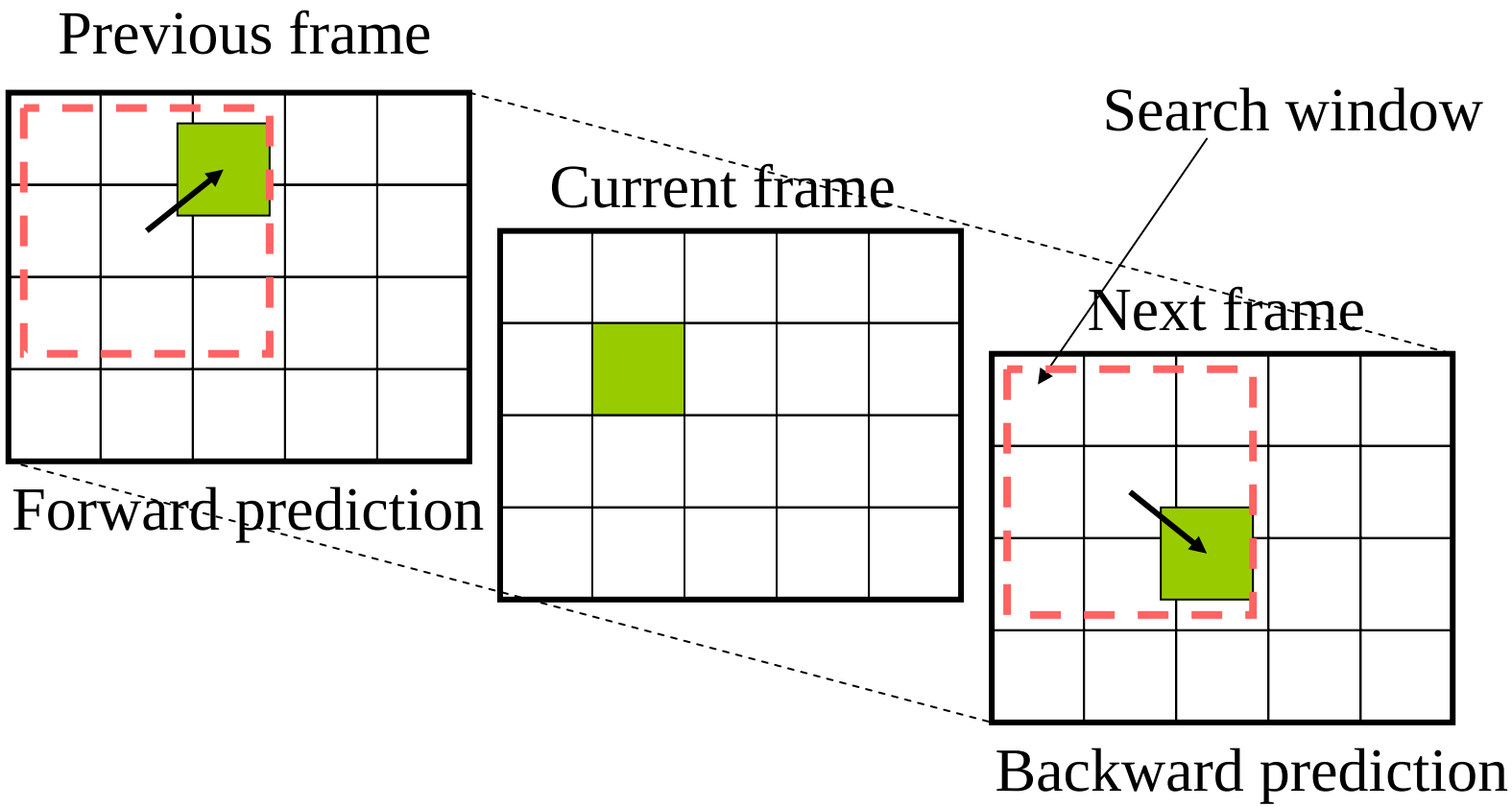
Camera Operation



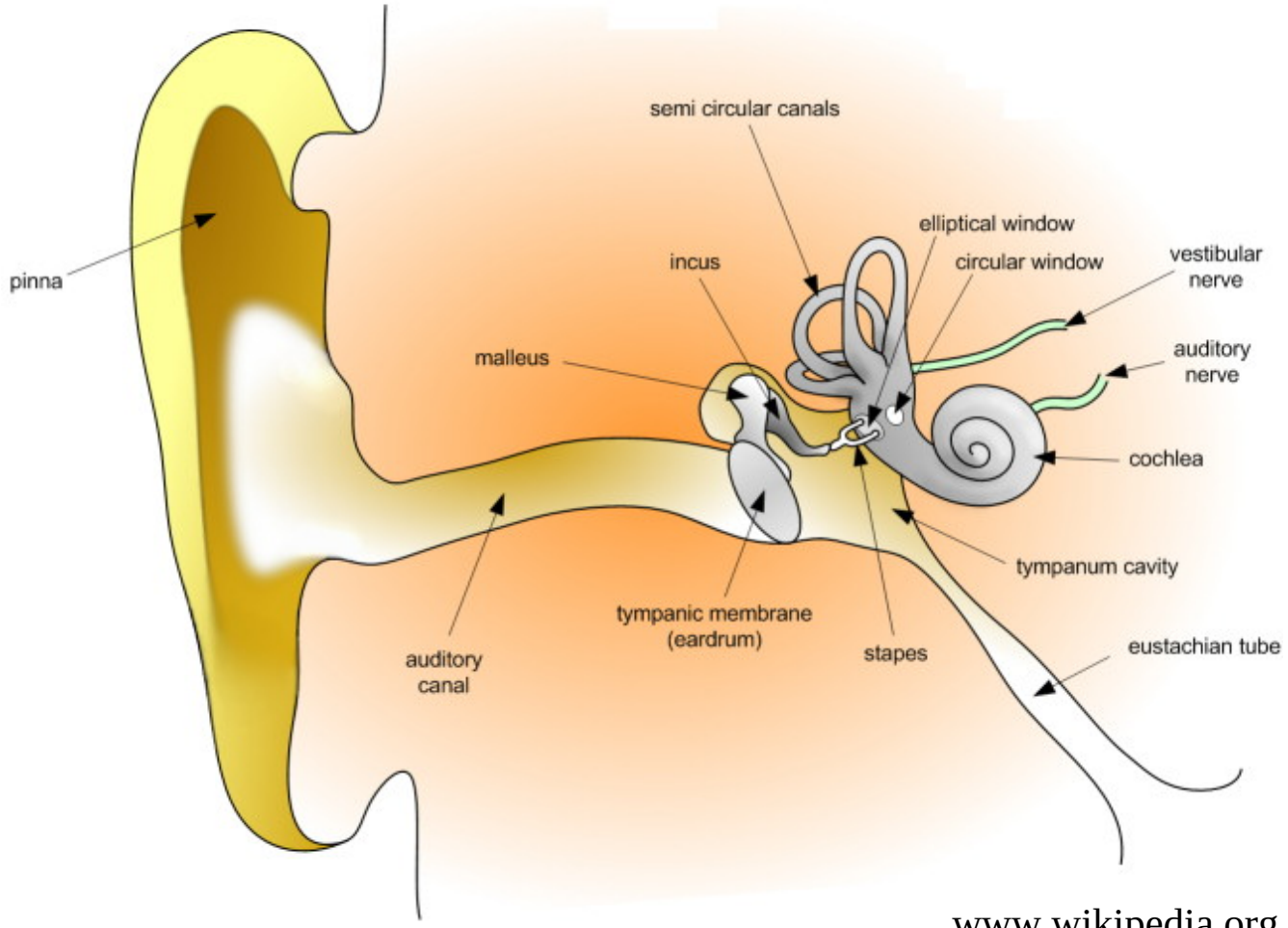
Optical Flow of Camera Motion



Motion Vector in Video Compression



Audio Primitive

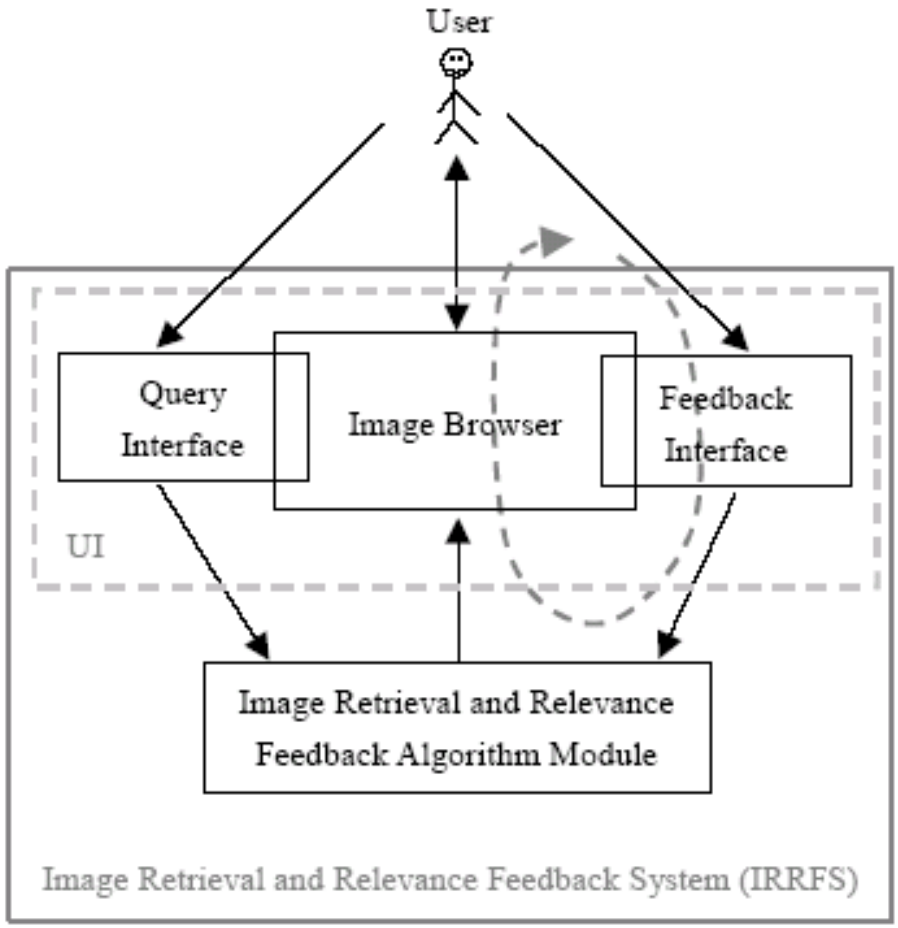


www.wikipedia.org

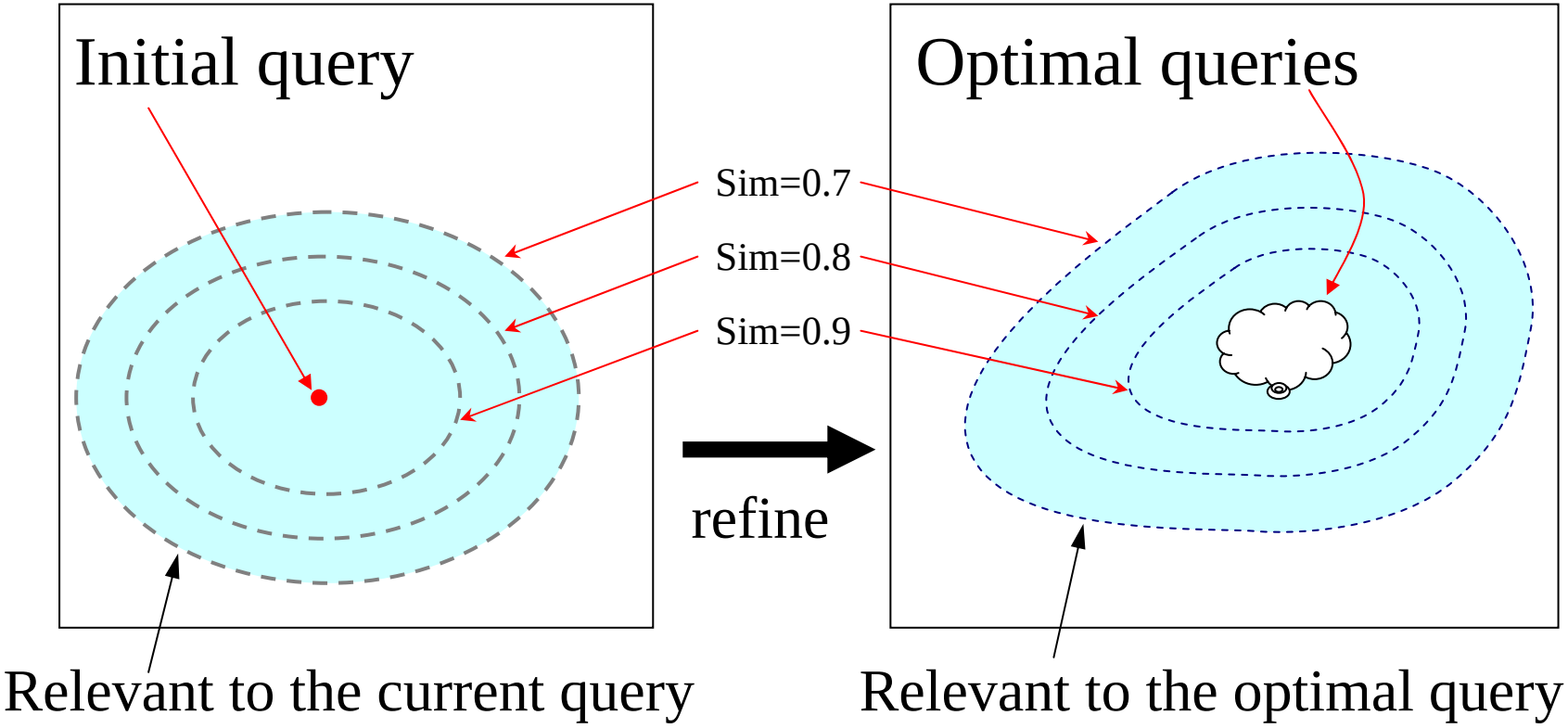
Audio Feature

- MFCCs (Mel-Frequency Cepstral Coefficients)
- Note frequency
- Rhythm
- Speech

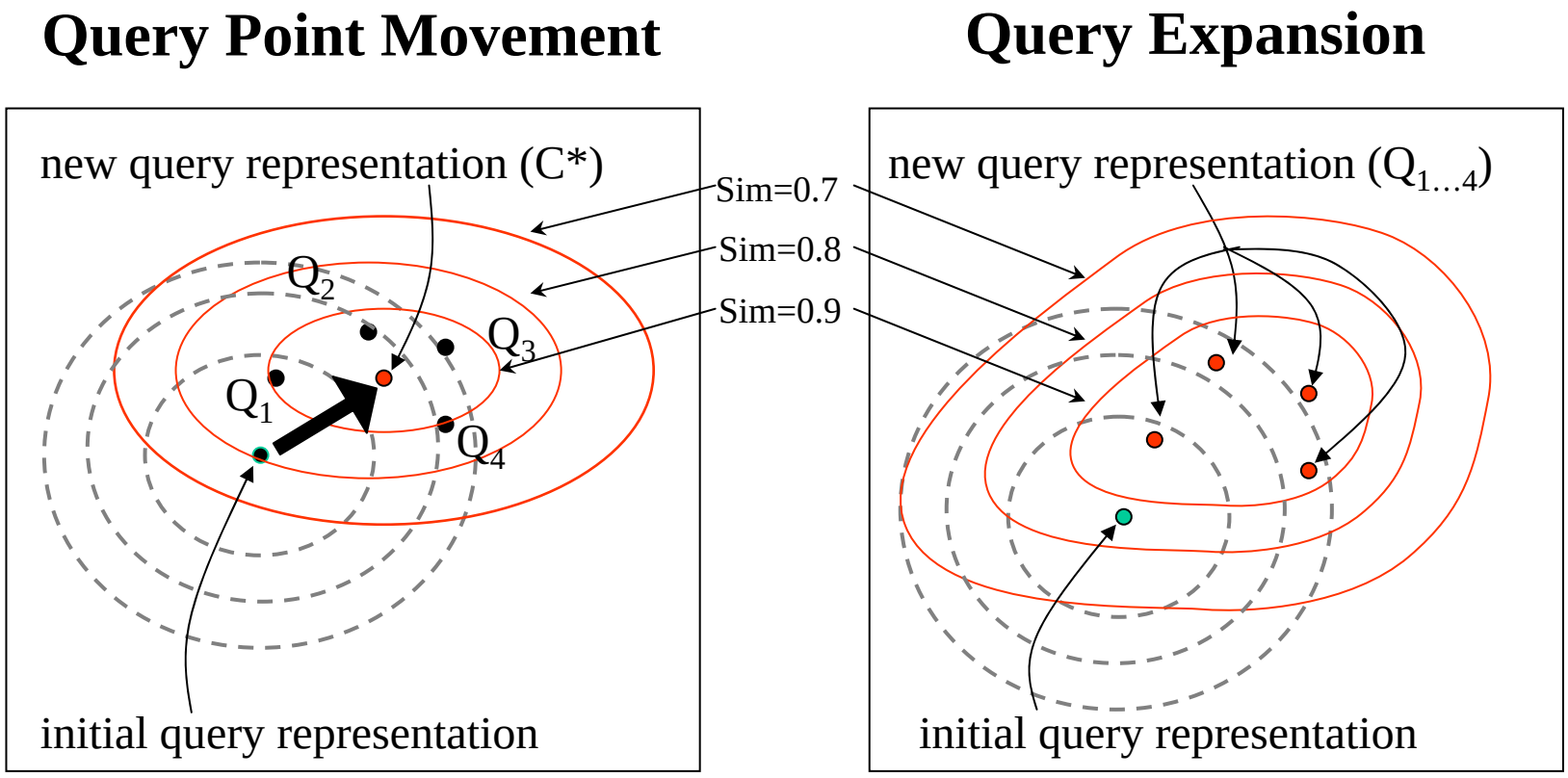
Relevance Feedback Interface



Relevance Feedback



Relevance Feedback



Benchmark

- ImageCLEF
  - Evaluating the retrieval performance of images with textual captions or metadata from multilingual document collections
  - http://ir.shef.ac.uk/imageclef/
- TRECVID
  - Contesting video retrieval tasks at NIST
  - http://www-nlpir.nist.gov/projects/trecvid/

The Problem in Content-based Retrieval

- The semantic gap
  - The difference between two descriptions of an object by different representations of human and computer
  - Traditional IR problem
  - The gap is getting larger in multimedia document retrieval

Thanks for Your Attention!

- Any question?

- Available resource
  - 數位典藏技術導論
    http://ebook.iis.sinica.edu.tw
  - The video retrieval showcase session, The ACM Conference on Image and Video Retrieval, 2008