

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!



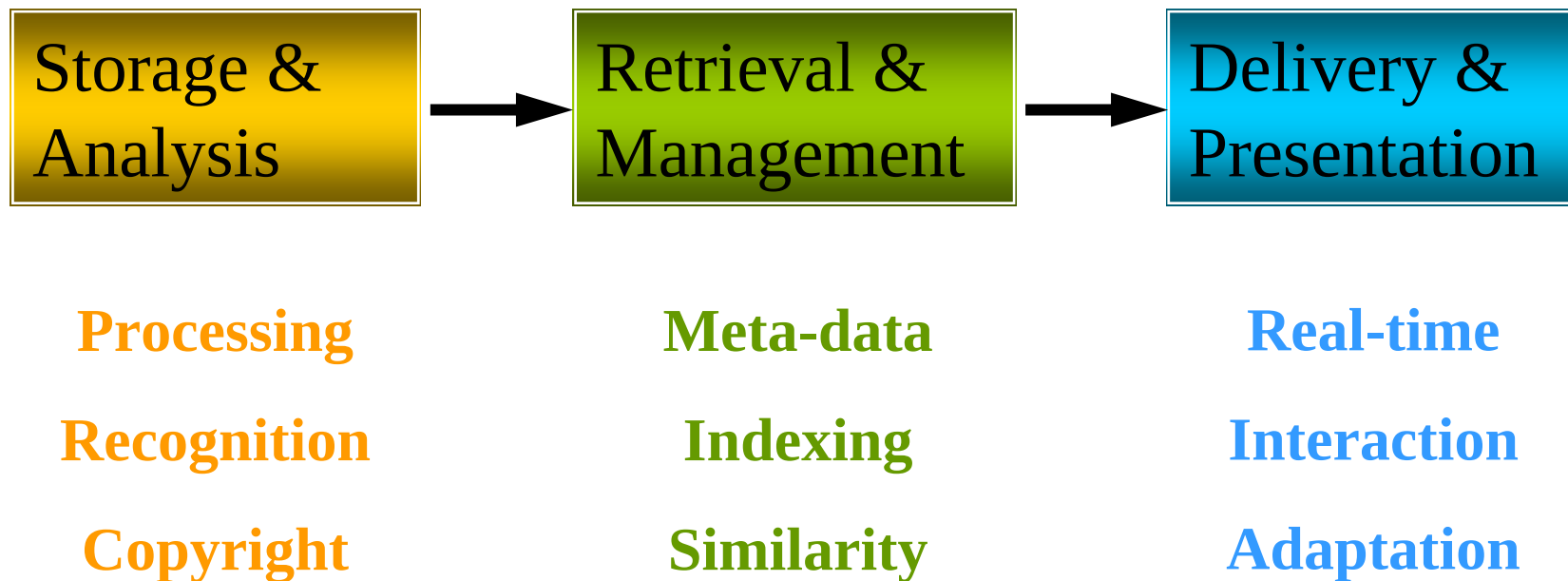
One person will have
> 30 gigabytes in 4 years
(>100,000 e-mails and
16,000 photos)

– Gorden Bell
pioneer of MyLifeBits

*Data is of no use
unless you can actually access it.*



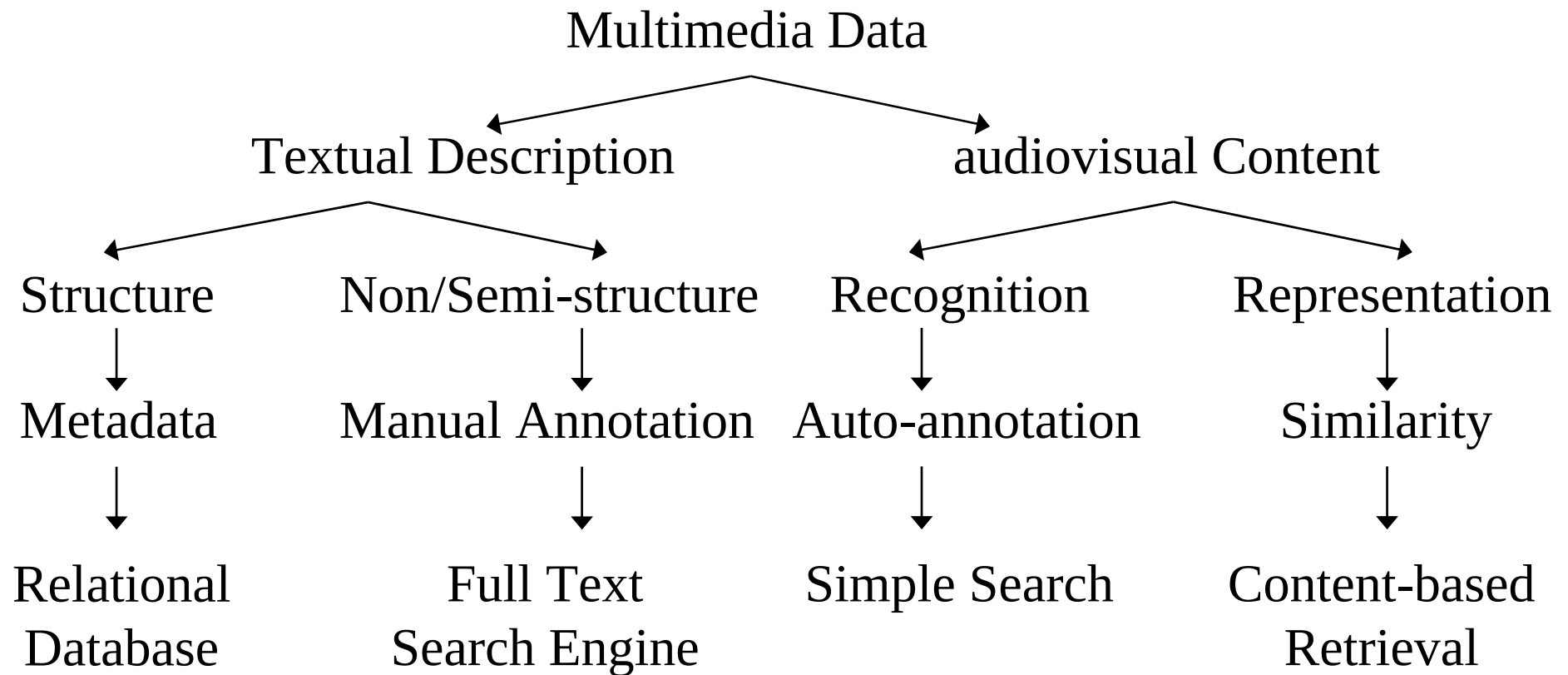
Multimedia Information Systems



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 California 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	軟體版本	Digital Camera FinePix4700 ZOOM Ver. 1.00
Date Time	拍攝時間	2000:01:01 13:39:44
<u>YCbCr Positioning</u>	<u>YCbCr 位置</u>	co-sited
<u>Copy Right</u>	版權	
<u>Exif Info Offset</u>	<u>Exif 資訊偏移量</u>	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
<u>Exif</u> Version	<u>Exif</u> 版本	0210
Date Time Original	原始成像時間	2000:01:01 13:39:44
Date Time Digitized	相機存檔時間	2000:01:01 13:39:44
Component Configuration	Component Configuration	<u>YCbCr</u>
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	<u>FlashPix</u> 版本	0100
Color Space	色彩空間	<u>sRGB</u>
<u>Exif</u> Image Width	圖像寬度	1280
<u>Exif</u> Image Height	圖像高度	960
<u>Exif</u> Interoperability Offset	<u>Exif</u> 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

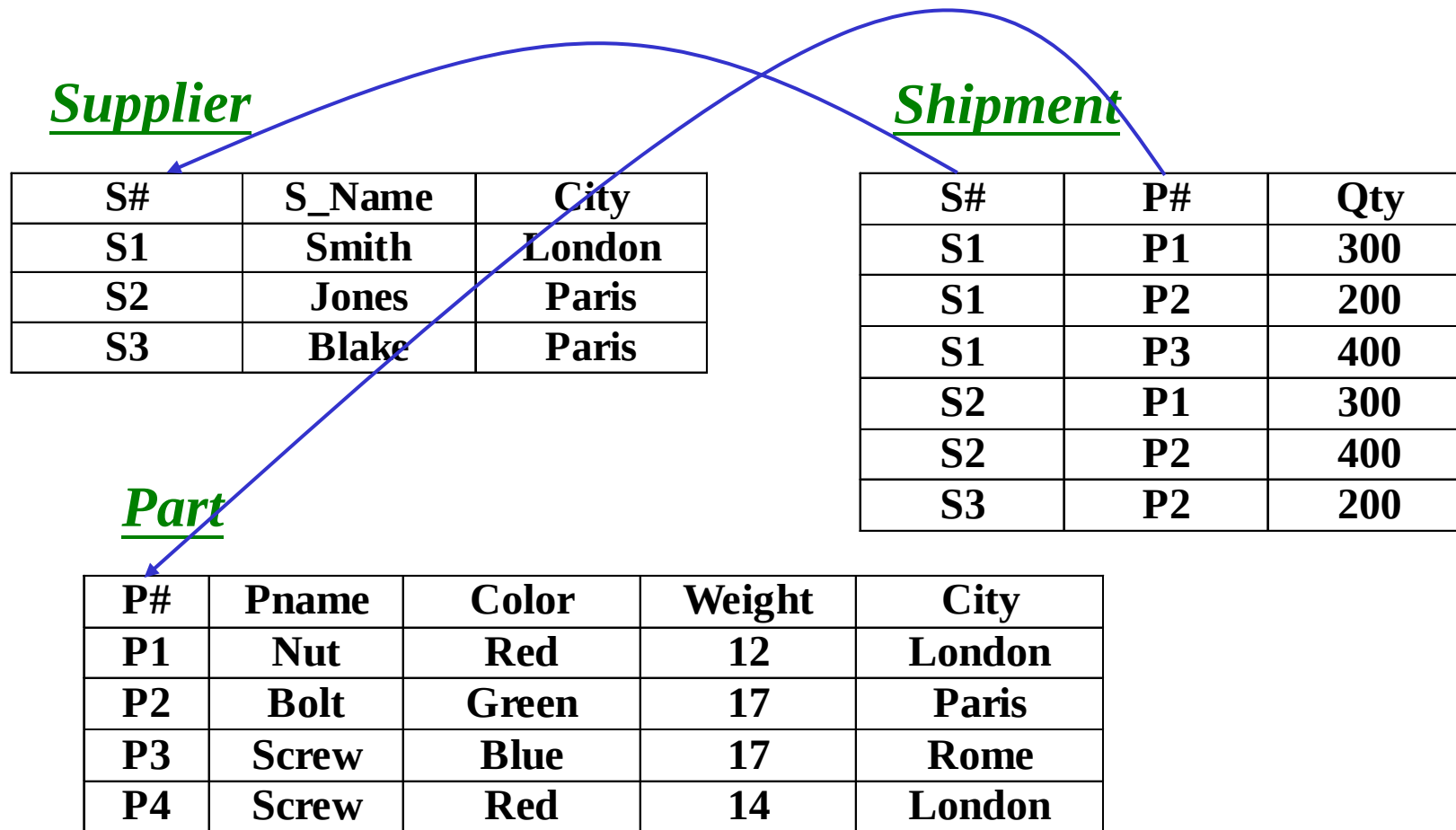
Generic Knowledge

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



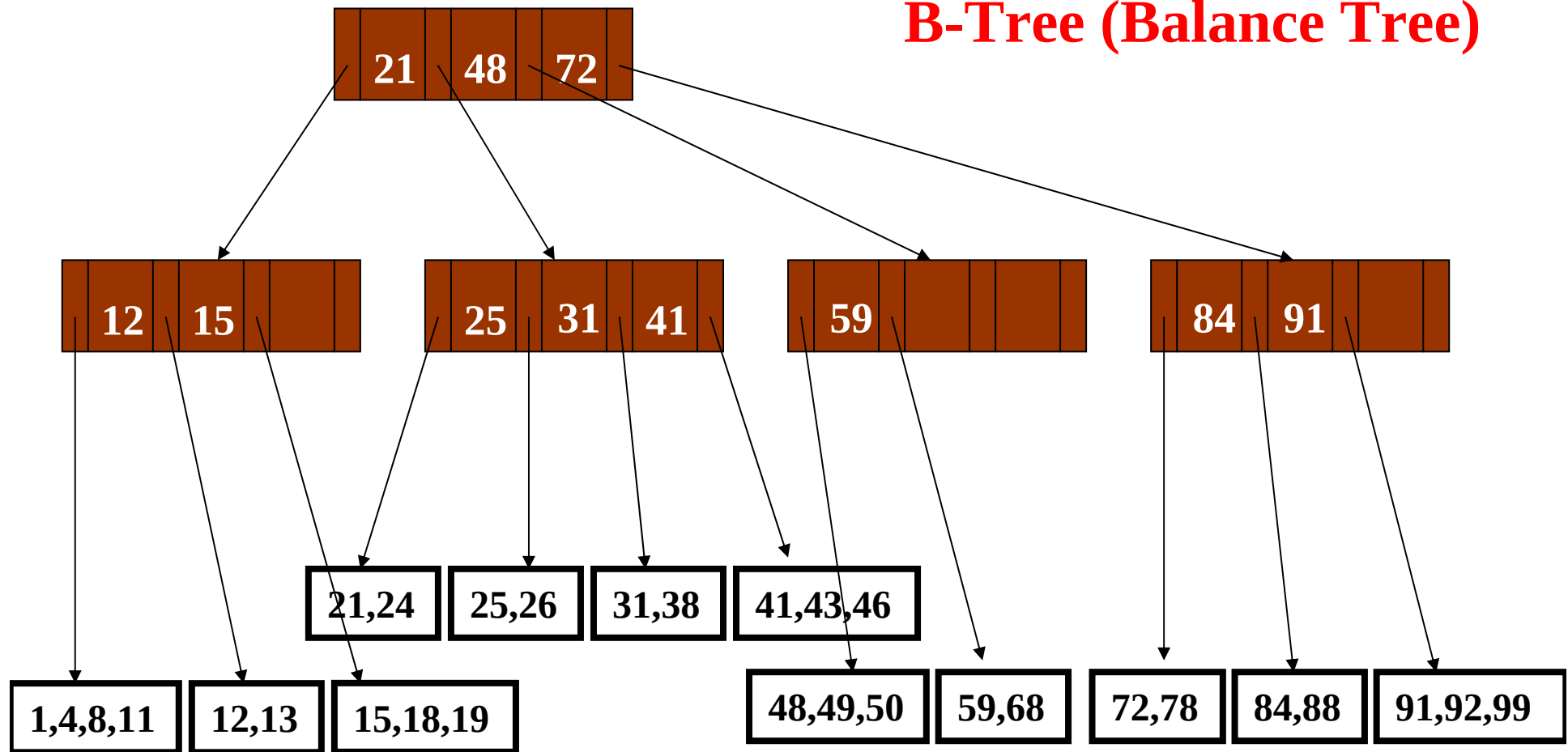
Querying Relational Databases

```
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

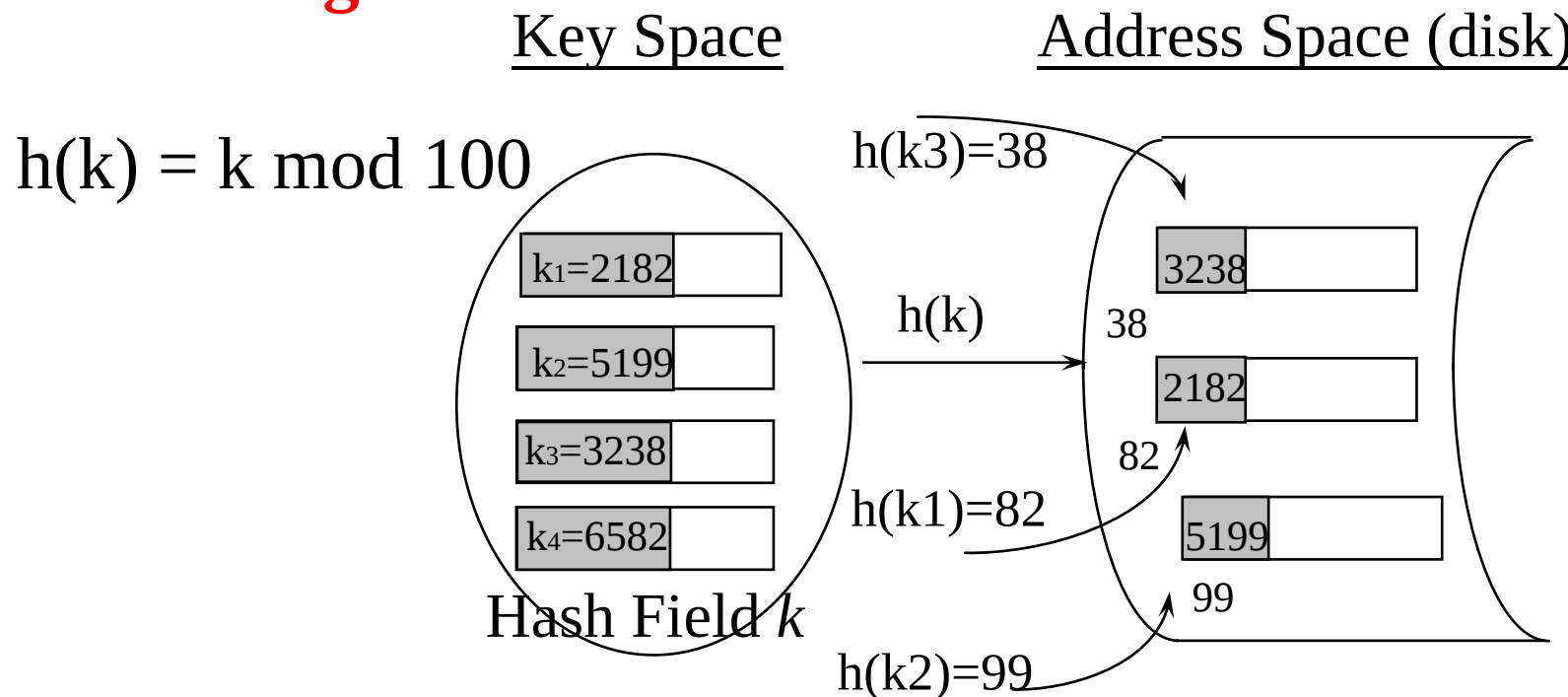
Indexing for Relational Databases

B-Tree (Balance Tree)



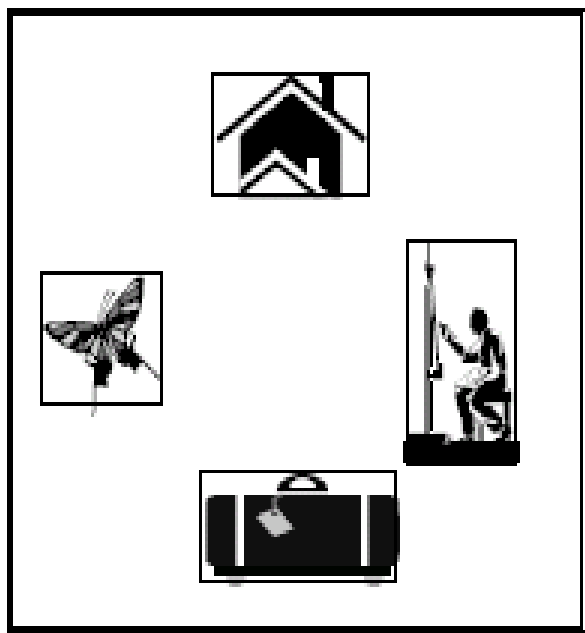
Indexing for Relational Databases

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!



翠玉白菜

Search Images

Search the Web

[Advanced Image Search](#)
[Preferences](#)

[Moderate SafeSearch is on](#)

Images

Showing:

All image sizes

Any content

Results 1 - 20 of about 13,000 for 翠玉白菜 (0.16 seconds)



... 鎮院之寶——翠玉白菜，傳出
缺損！

278 x 400 - 43k - jpg
big5.ce.cn

[[More from www.ce.cn](http://www.ce.cn)]



清翠玉白菜
400 x 400 - 31k - jpg
enews.npm.gov.tw



☆寂靜\翠玉白菜資料蒐集-
yam天空 ...
1536 x 2048 - 741k - jpg
img175.imageshack.us



故宮博物院翠玉白菜
350 x 498 - 49k - jpg
taipei.traveleredge.com



答案是：卡哇伊到不行的翠玉白
菜！

368 x 562 - 67k - jpg
ctlittleben.blogspot.com



翠玉白菜
500 x 600 - 61k - jpg
fcu.org.tw



... 製作翠玉白菜的玉匠，都只是
發揮 ...
550 x 410 - 35k - jpg
www.etaiwanexpo.nat.gov.tw



翠玉白菜人氣最高
488 x 366 - 38k - jpg
idv.creativity.edu.tw



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/

首頁

特色藏品

快速導覽

主題分類

典藏機構

進階搜尋

►搜尋結果:翠玉白菜

搜尋字詞



由搜尋結果繼續



重新搜尋

搜尋

您搜尋 翠玉白菜

分類中符合的項目

共找到 5 筆符合的資料

- 書畫 (1)
- 新聞 (3)
- 影音 (1)
- 器物 (1)



清 翠玉白菜

玉石器、清 **翠玉白菜**、故玉002103N0000000000、1644 A.D.-1911 A.D.、長18.7公分 厚5.07公分 寬9.1公分、此器利用玉之原色施以巧琢，以綠色部份琢為菜葉，白色處施作菜身，酷似實物，葉上並琢二螭斯。為故宮所藏「巧作玉器」中最受人矚目者。、《國之重寶》、《國立故宮博物院藏品選目》、國立故宮博物院

1/5

關於翠玉白菜與象牙球的一些事

關於**翠玉白菜**與象牙球的一些事、施靜菲、**翠玉白菜** 象牙球、來源期刊:故宮文物月刊、卷期:288 2007.03[民96.03]、頁次:頁4-10、出版時間:2007.03[民96.03]、期刊論文、系統識別號:A07031879、國家圖書館

2/5

翠玉白菜--從材料到作品 -1-

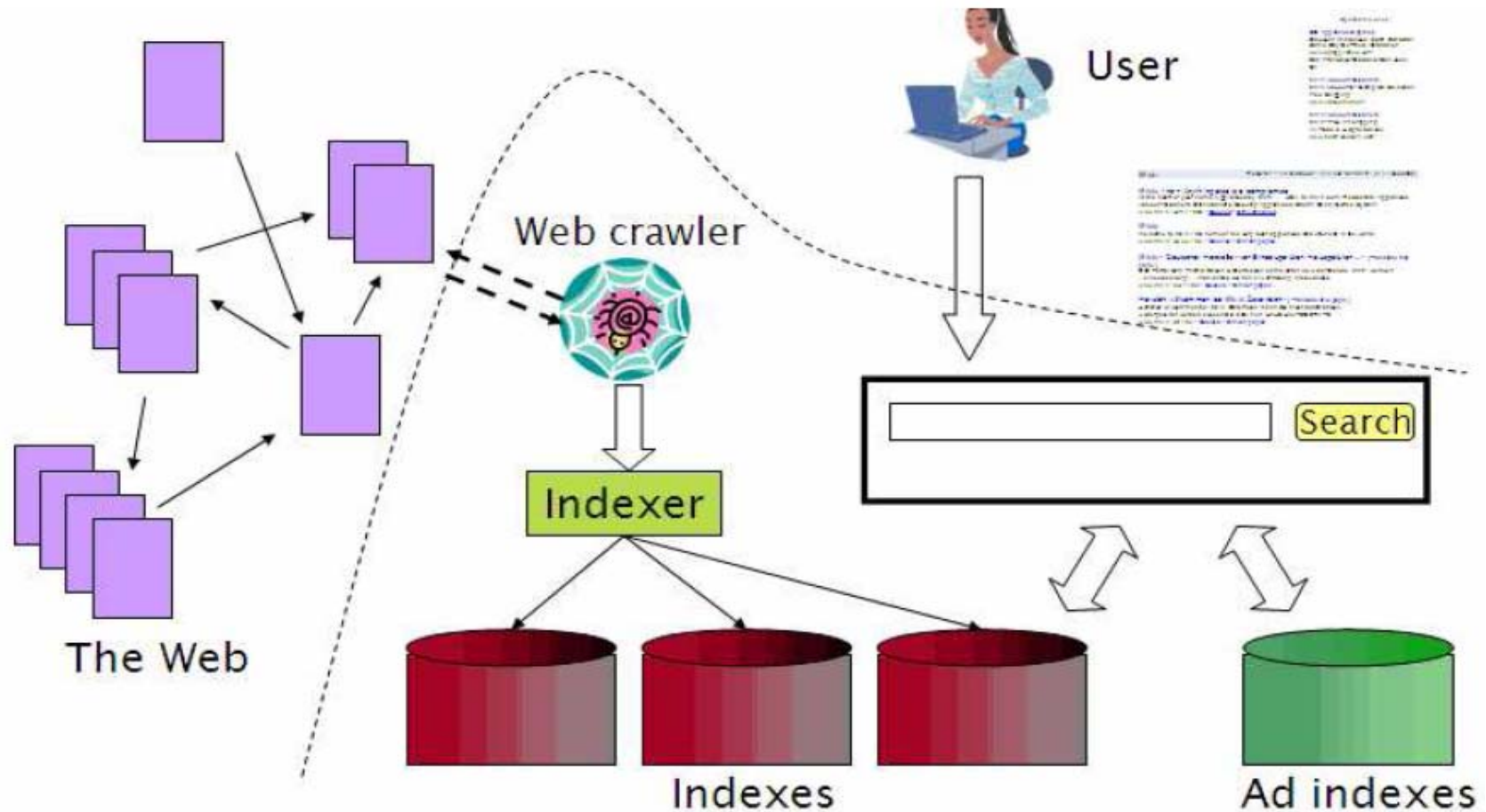
翠玉白菜--從材料到作品 -1-、張麗端、來源期刊:故宮文物月刊、卷期:21:11=251 民93.02、頁次:頁54-62、期刊論文、A0400391、國家圖書館遠距圖書服務系統、國家圖書館

3/5

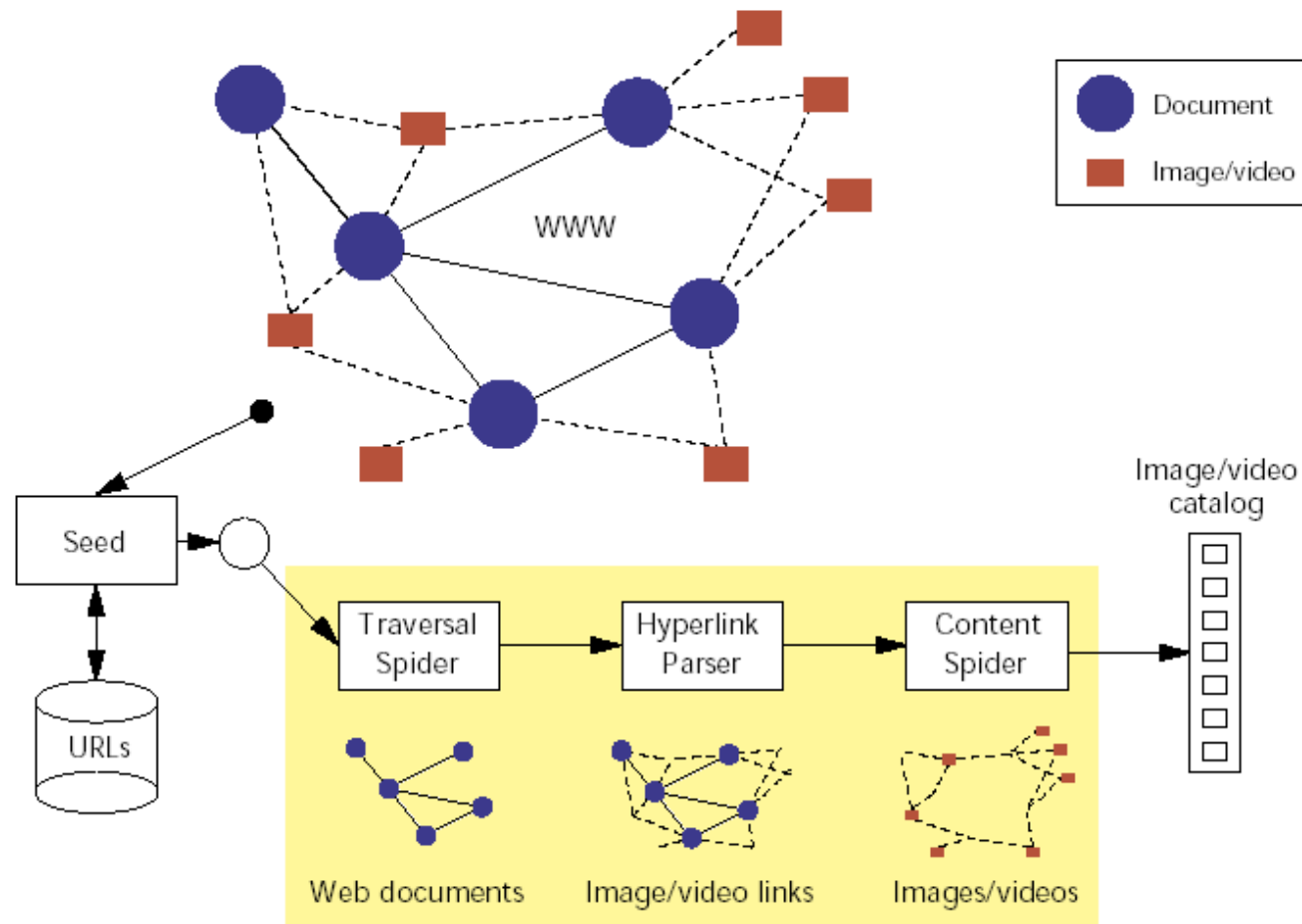
Textual Image

- 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

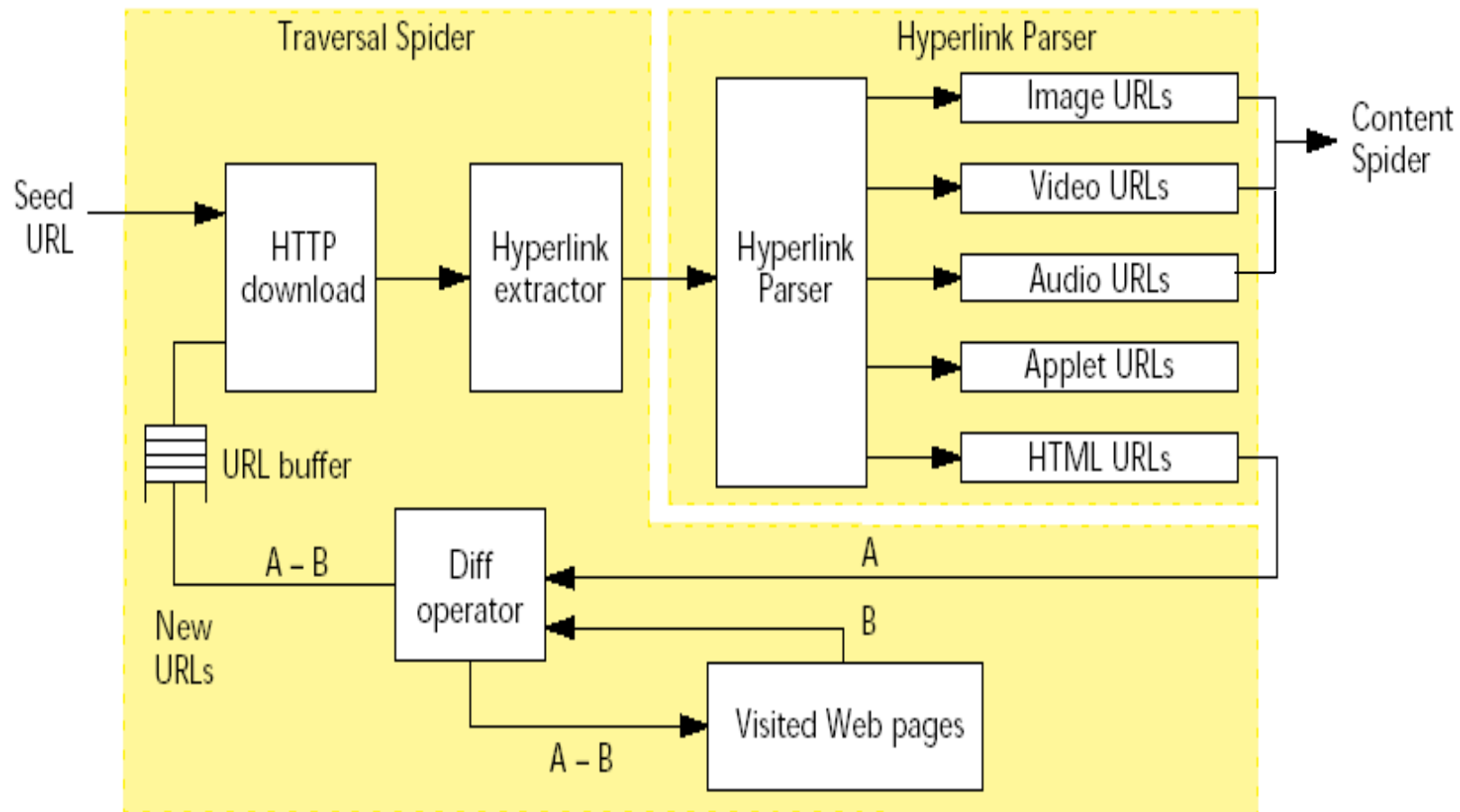
Web Search Overview



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?

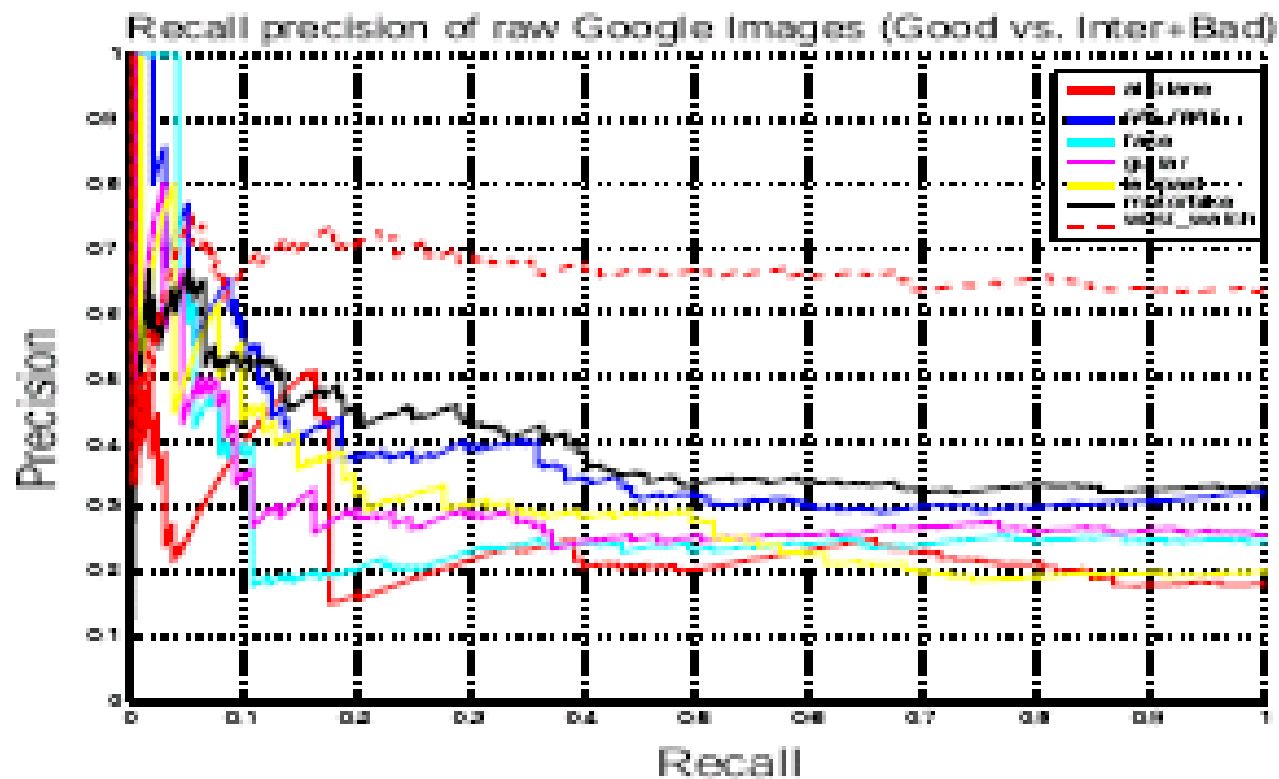
- Advertisement images
- Content images
- Decorative images (buttons, balls, ...)
- Informational images (under construction, news, ...)
- Logo
- Navigation images (arrows, back to home, image maps, ...)

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

Google Image Labeler BETA

chihyi.chiu@gmail.com | [Help](#) | [Sign Out](#)

time left
01:00

score
220

passes
0

Your partner has suggested 4 labels.



off-limits
woman
hair
lady
lips
model

my labels
beauty
white

[Privacy Policy](#) - [Terms of Use](#) - [Return to Google Image Search](#)
© 2007 Google

Outline

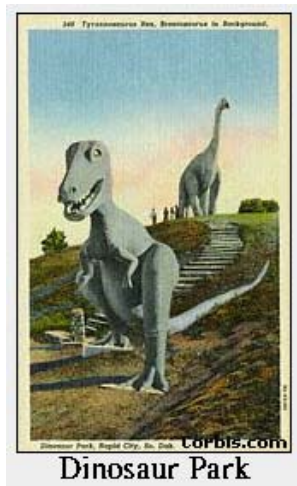
- 多媒體資料庫搜尋
- 多媒體全文搜尋引擎
- 以內容為基礎之多媒體檢索

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

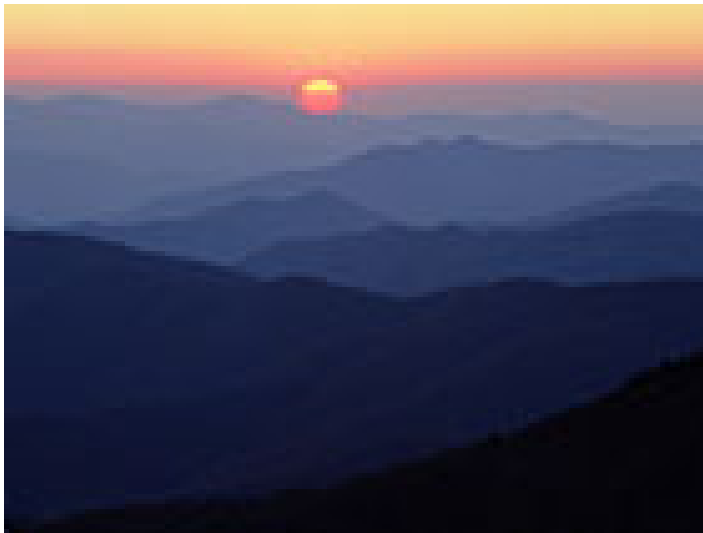
Why Content-based Approach?

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

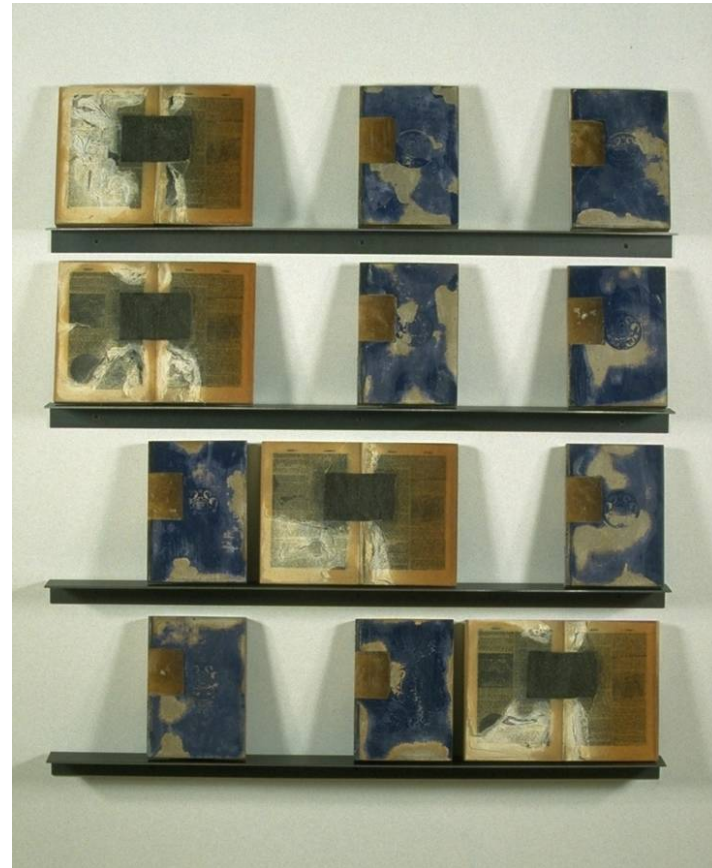


But the fact is a few words labeled!

Similar?



Similar?



Similar?



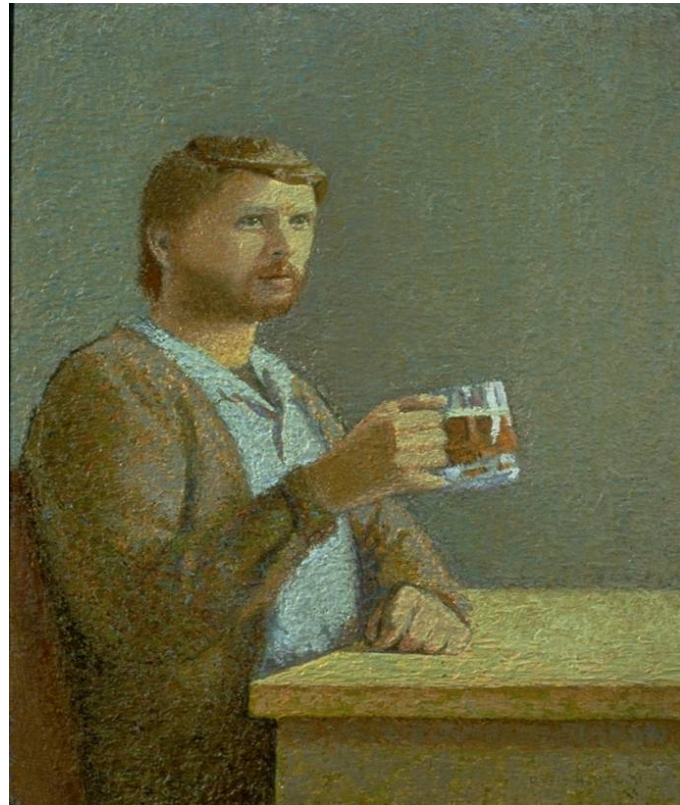
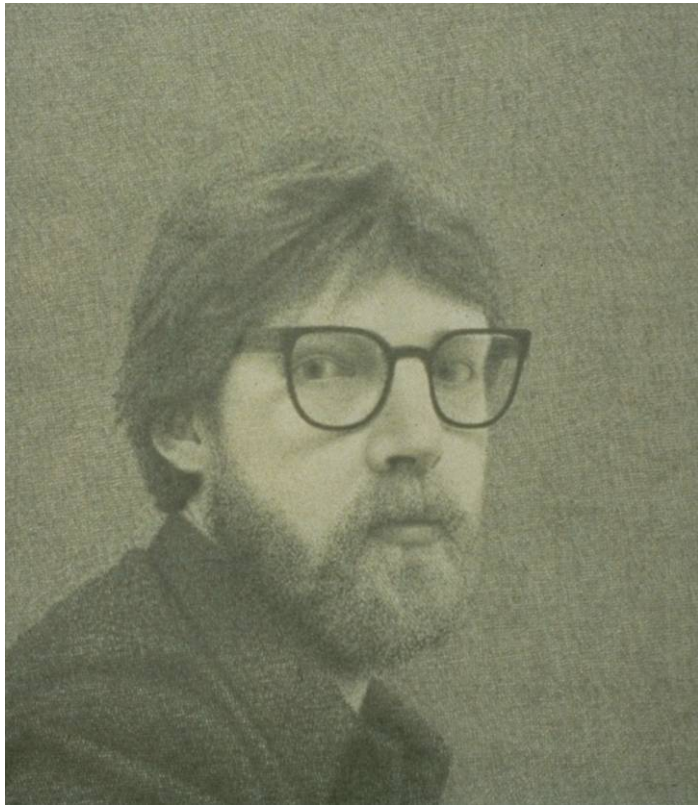
Similar?



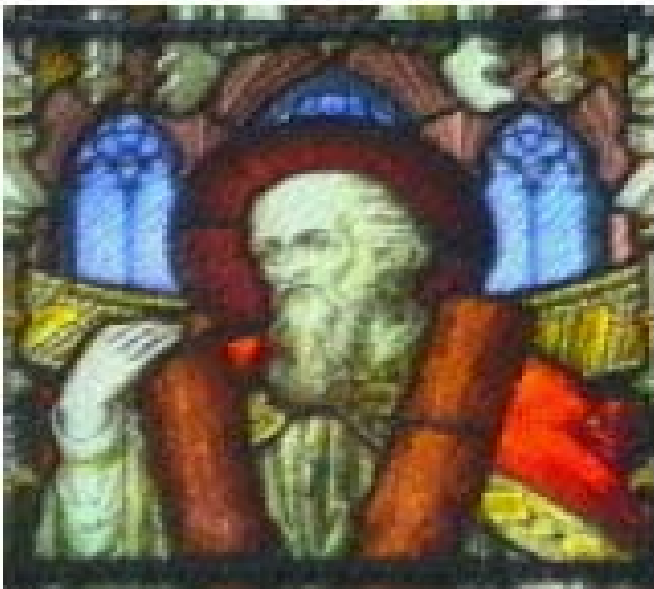
Similar?



Similar?



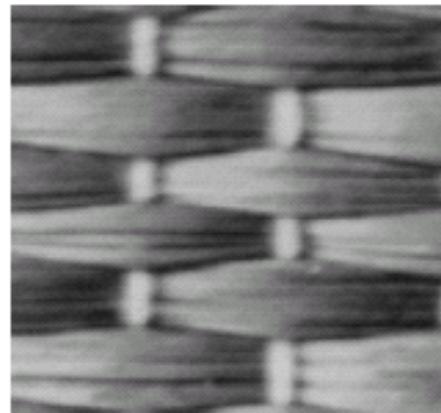
Similar?



Similar?



brick



straw matting



oriental rattan

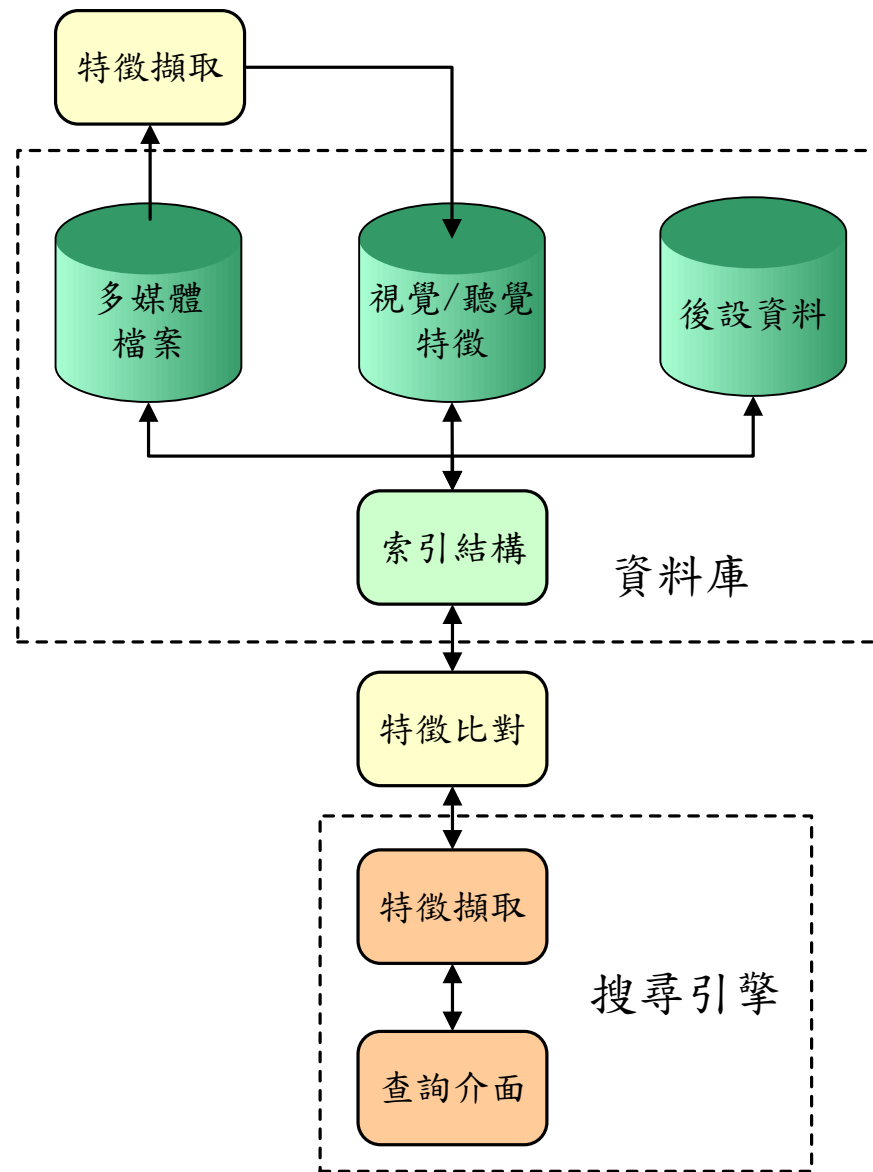


brick

[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

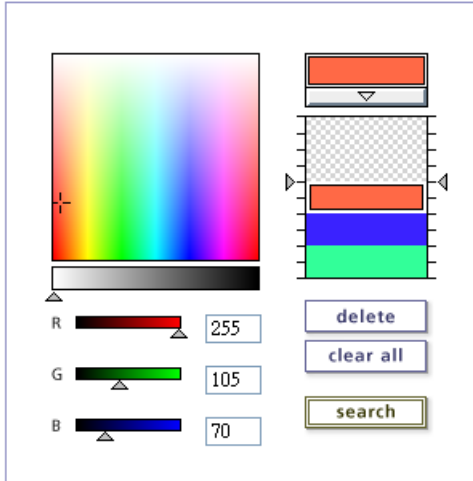


Query Interface in Content-based Retrieval

- 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



The interface includes a color palette, a bucket with a triangular handle, and RGB sliders. The RGB values are: R=255, G=105, B=70. Buttons for 'delete', 'clear all', and 'search' are also present.

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.

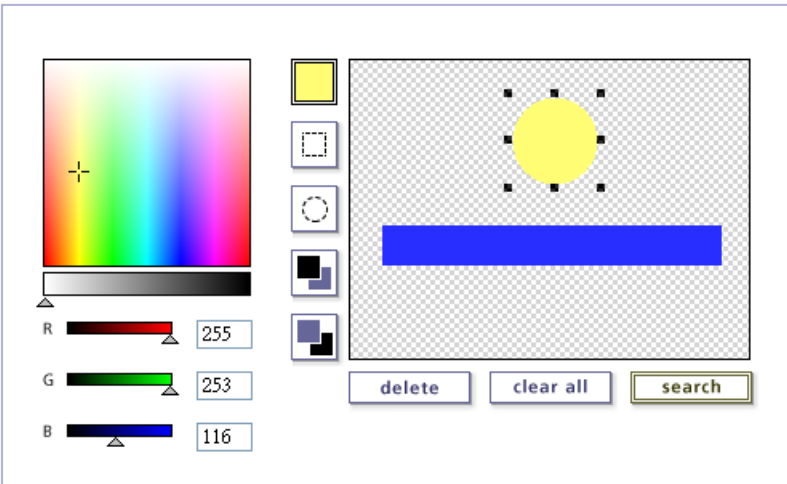
SEARCH RESULTS

[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)

QBIC LAYOUT SEARCH



The QBIC Layout Search interface features a color selection tool on the left with a rainbow gradient and RGB sliders (R: 255, G: 253, B: 116). A central canvas shows a yellow circle and a blue rectangle with selection handles. To the right of the canvas are icons for different shapes and a 'search' button.

SEARCH RESULTS

[modify search](#)
[new search](#)

 1) <u>Waterloo Bridge. Effect of Fog</u> Monet, Claude 1903	 2) <u>Island of Fishermen on Lake Maggiore</u> Premazzi, Luigi Second half of the 19th century
 3) <u>Woman Sleeping under a Tree</u> Redon, Odilon Between 1900 and 1901	 4) <u>Flood</u> Guilloux, Charles 1893
 5) <u>Fatata Te Moua (At the Foot of a Mountain)</u> Gauguin, Paul 1892	 6) <u>Conversation</u> Matisse, Henri Between 1908 and 1912
 7) <u>Admiralty Sound: Tierra del Fuego</u> Kent, Rockwell 1922 - 1925	 8) <u>Asgaard</u> Kent, Rockwell 1950

Direct Query – VideoQ (Columbia University)

The screenshot displays the VideoQ Visual Sketch interface, which is used for visual search. The interface is divided into two main sections: a visual search grid on the left and a query canvas on the right.

Visual Search Grid:

- Header:** "Visual Search" with "Home" and "Up" buttons.
- Help/Start:** Links for "Help" and "Start".
- Search Results:** A grid of four search results, each with a title, a visual representation, and a description:

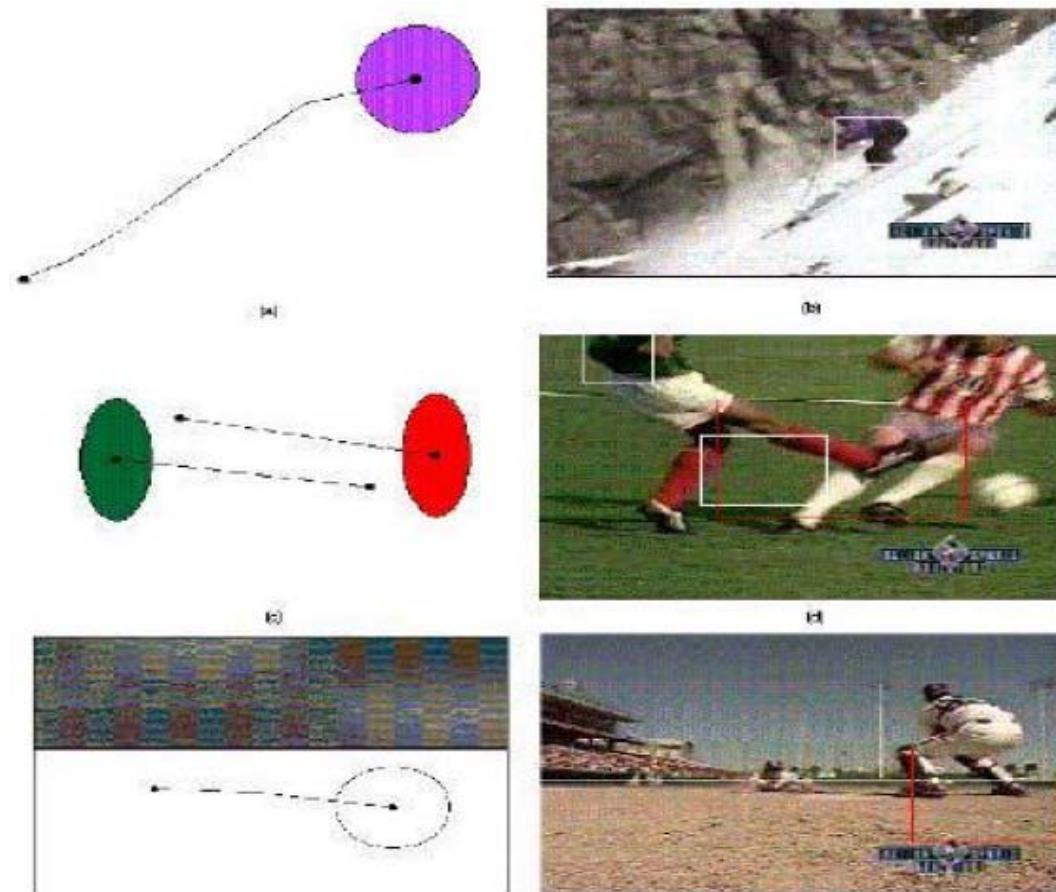
Search Result	Visual Representation	Description
Sunrise	Orange square with a yellow circle	Color and size
Sky and ocean	Blue rectangle with a grey wavy line	Texture and size
High jumper	Red circle with a black arc	Motion and size
Downhill skier	Blue circle with a black line and the word "Sports"	Motion and keyword

VideoQ Visual Sketch Canvas:

- Header:** "VideoQ Visual Sketch" with "File", "Edit", and "View" menus.
- Canvas:** A grid where two shapes are being sketched: a green circle labeled "C" and a green square labeled "D".
- Tools:** A vertical toolbar on the right with various drawing tools (arrow, rectangle, circle, ellipse, line, text, etc.).
- Query Input:** A text box labeled "Compose a query image in the canvas." and a "Keyword:" field.
- Adjust Query Weights:** A bar chart showing the relative weights assigned to different features for the query image. The y-axis ranges from "Less" to "More".

Feature	Weight (Less to More)
Motion	Low
Color	High
Texture	Low
Size	Medium
Shape	Low

Direct Query – VideoQ (Columbia University)



Query by Example – SIMPLICITY (Stanford University)

S·I·M·P·L·I·c·i·t·y

Semantics-sensitive Integrated Matching for Picture Libraries

Option 1 --> Image ID or URL
similar images

Option 2 --> **Random**

Option 3 --> Click an image to



[52386](#) 0.00 5



[26027](#) 11.32 4



[41520](#) 13.21 4



[19018](#) 13.49 3



[5300](#) 13.55 3



[37385](#) 13.65 3



[6102](#) 13.67 3



[35013](#) 13.67 3



[53513](#) 13.90 5



[25621](#) 13.95 3



[14552](#) 14.13 4



[23087](#) 14.15 4



[14246](#) 14.15 2



[47943](#) 14.16 5

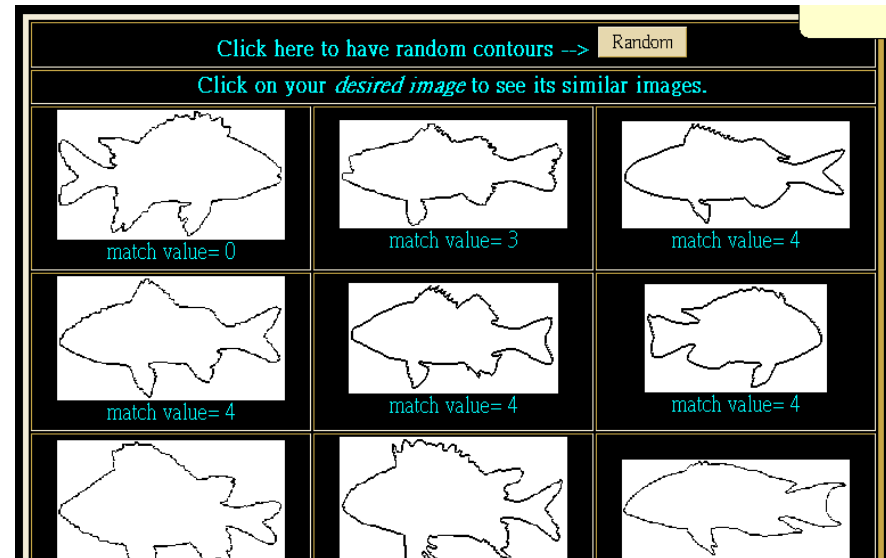
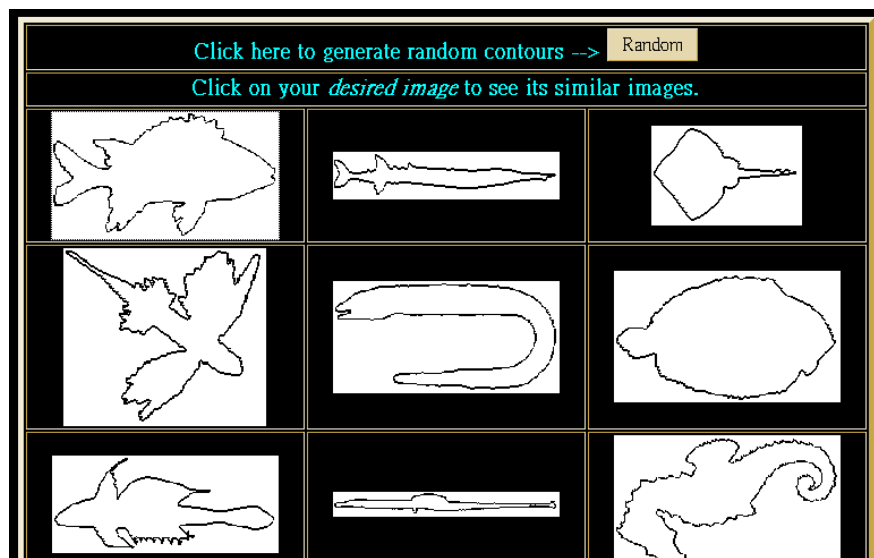


[29900](#) 14.17 4



[19020](#) 14.17 4

Query by Example – SQUID (Surry University)



Query by Example – VCenter



圖片資料來源

重新搜尋



3D 虛擬文物動畫短片 到故宮找想像力-象牙球
3.056秒 [相似程度: 90%]

>> 搜尋結果



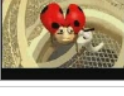
3D 虛擬文物動畫短片 到故宮找想像力-象牙球
3.056 秒 [相似程度: 90%]



3D 虛擬文物動畫短片 到故宮找想像力-象牙球
5.068 秒 [相似程度: 86%]



「透視內容：國立故宮博物院」預告片
18.83 秒 [相似程度: 62%]



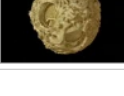
3D 虛擬文物動畫短片 到故宮找想像力-象牙球
1.254 秒 [相似程度: 60%]



形象廣告 Old is New
38.976 秒 [相似程度: 58%]



3D 虛擬文物動畫短片 到故宮找新鮮-翠玉白菜
39.131 秒 [相似程度: 58%]



文物視聽短片 象牙球
10.971 秒 [相似程度: 56%]

搜尋參數設定

設定呈現多少張比對最相似的结果
張數: 10 預設為 10

設定判斷比對相似度的門檻值
門檻值: 3 預設為 2.5

設定檢索圖庫
圖庫: NPM Data(110)

再搜尋一次 預設值

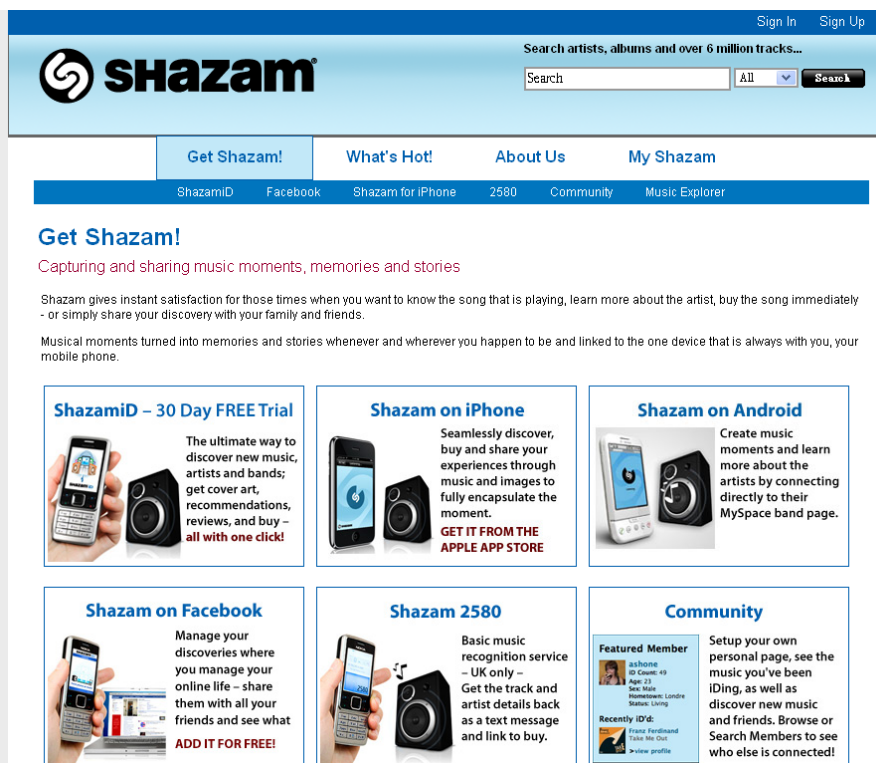
參數一覽表

numImages
threshold
imageBase
matchMethod

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

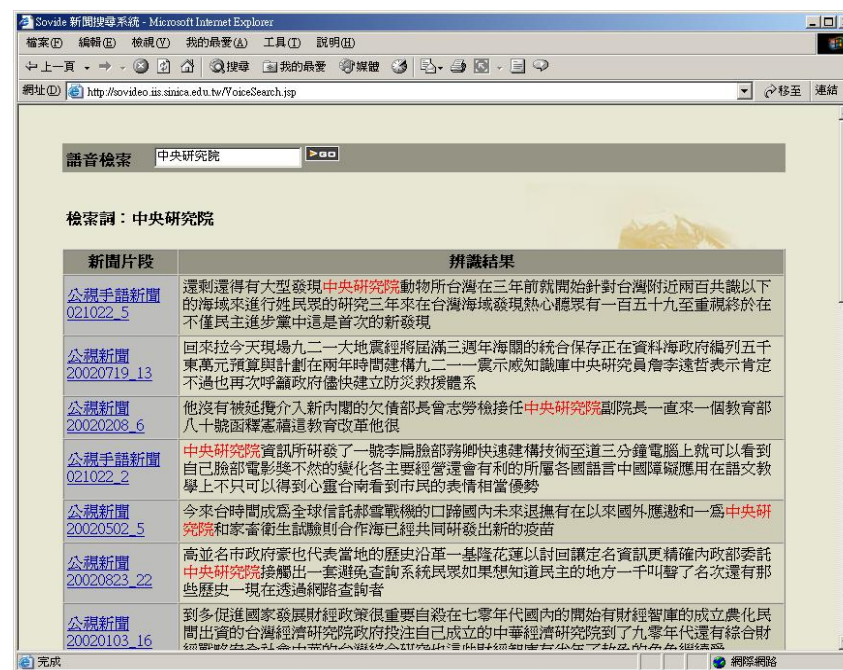
Query by Example - Shazam

- 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

圖片內容相似檢索 (30000張推薦圖庫)

 ☒ 相似 ☐ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似
 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似
 ☒ 相似 ☐ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似	 ☐ 相似 ☒ 不相似

共找到 100 圖片 (1-18) 頁: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

相似程度 傳回圖數

請至少選擇一個相似圖片，按 開始查詢，或按 重新選擇

Relevance Feedback – 新視科技

圖片內容相似檢索 (30000張推薦圖庫)

					
☒ 相似 ☐ 不相似	☐ 相似 ☒ 不相似	☐ 相似 ☒ 不相似	☐ 相似 ☒ 不相似	☐ 相似 ☐ 不相似	☐ 相似 ☒ 不相似
					
☐ 相似 ☐ 不相似	☐ 相似 ☐ 不相似	☐ 相似 ☐ 不相似	☐ 相似 ☐ 不相似	☐ 相似 ☐ 不相似	☐ 相似 ☐ 不相似
					
☐ 相似 ☐ 不相似	☐ 相似 ☐ 不相似	☐ 相似 ☐ 不相似	☐ 相似 ☒ 不相似	☐ 相似 ☒ 不相似	☐ 相似 ☒ 不相似

共找到 100 圖片 (1-18) 頁: [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#)

相似程度 傳回圖數

請至少選擇一個相似圖片, 按 開始查詢, 或按 重新選擇

Relevance Feedback – 新視科技

圖片內容相似檢索 (30000張推薦圖庫)

 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似
 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似
 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似	 ☐ 相似 ☐ 不相似

共找到 100 圖片 (1-18) 頁: [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#)

相似程度 傳回圖數

請至少選擇一個相似圖片，按 開始查詢，或按 重新選擇

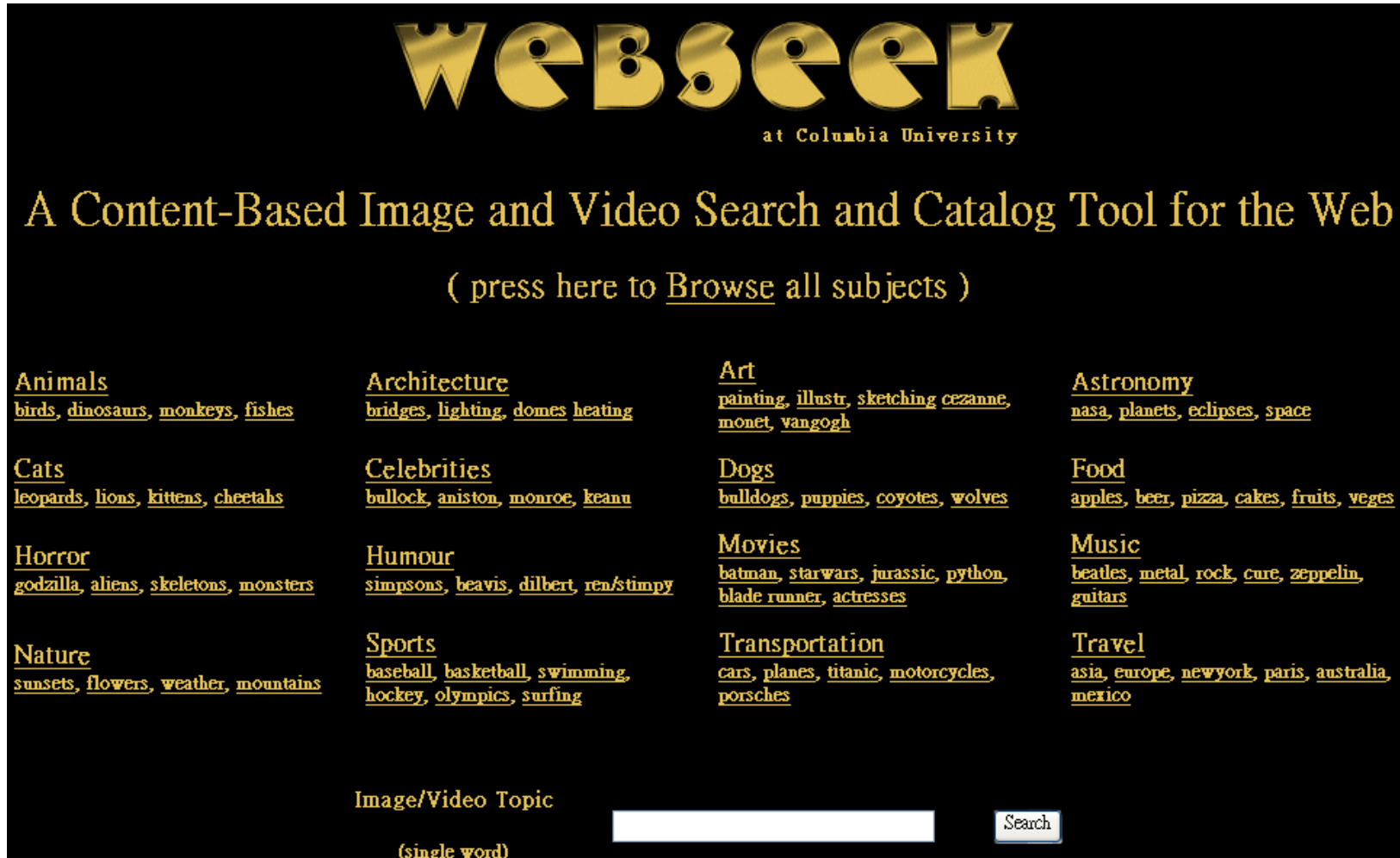
Browse Multimedia Content

- 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



WEBSEEK
at Columbia University

A Content-Based Image and Video Search and Catalog Tool for the Web
(press here to Browse all subjects)

<u>Animals</u> <u>birds</u> , <u>dinosaurs</u> , <u>monkeys</u> , <u>fishes</u>	<u>Architecture</u> <u>bridges</u> , <u>lighting</u> , <u>domes</u> , <u>heating</u>	<u>Art</u> <u>painting</u> , <u>illustr</u> , <u>sketching</u> <u>cezanne</u> , <u>monet</u> , <u>vangogh</u>	<u>Astronomy</u> <u>nasa</u> , <u>planets</u> , <u>eclipses</u> , <u>space</u>
<u>Cats</u> <u>leopards</u> , <u>lions</u> , <u>kittens</u> , <u>cheetahs</u>	<u>Celebrities</u> <u>bullock</u> , <u>aniston</u> , <u>monroe</u> , <u>keanu</u>	<u>Dogs</u> <u>bulldogs</u> , <u>puppies</u> , <u>coyotes</u> , <u>wolves</u>	<u>Food</u> <u>apples</u> , <u>beer</u> , <u>pizza</u> , <u>cakes</u> , <u>fruits</u> , <u>veges</u>
<u>Horror</u> <u>godzilla</u> , <u>aliens</u> , <u>skeletons</u> , <u>monsters</u>	<u>Humour</u> <u>simpsons</u> , <u>beavis</u> , <u>dilbert</u> , <u>ren/stimpy</u>	<u>Movies</u> <u>batman</u> , <u>starwars</u> , <u>jurassic</u> , <u>python</u> , <u>blade runner</u> , <u>actresses</u>	<u>Music</u> <u>beatles</u> , <u>metal</u> , <u>rock</u> , <u>cure</u> , <u>zeppelin</u> , <u>guitars</u>
<u>Nature</u> <u>sunsets</u> , <u>flowers</u> , <u>weather</u> , <u>mountains</u>	<u>Sports</u> <u>baseball</u> , <u>basketball</u> , <u>swimming</u> , <u>hockey</u> , <u>olympics</u> , <u>surfing</u>	<u>Transportation</u> <u>cars</u> , <u>planes</u> , <u>titanic</u> , <u>motorcycles</u> , <u>porsches</u>	<u>Travel</u> <u>asia</u> , <u>europa</u> , <u>newyork</u> , <u>paris</u> , <u>australia</u> , <u>mexico</u>

Image/Video Topic
(single word)

Image Classification

LivImage

or -- Specify a sample query --
Recommended Screen Resol

Developed by *Web Knowledge Discovery Lab, Institute of Information Science, Academia*

Preview

Relevant Concepts of **building** (Page 1) [Next >>](#)



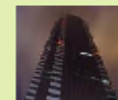
BUILDING



OFFICE BUILDING



TALL BUILDING



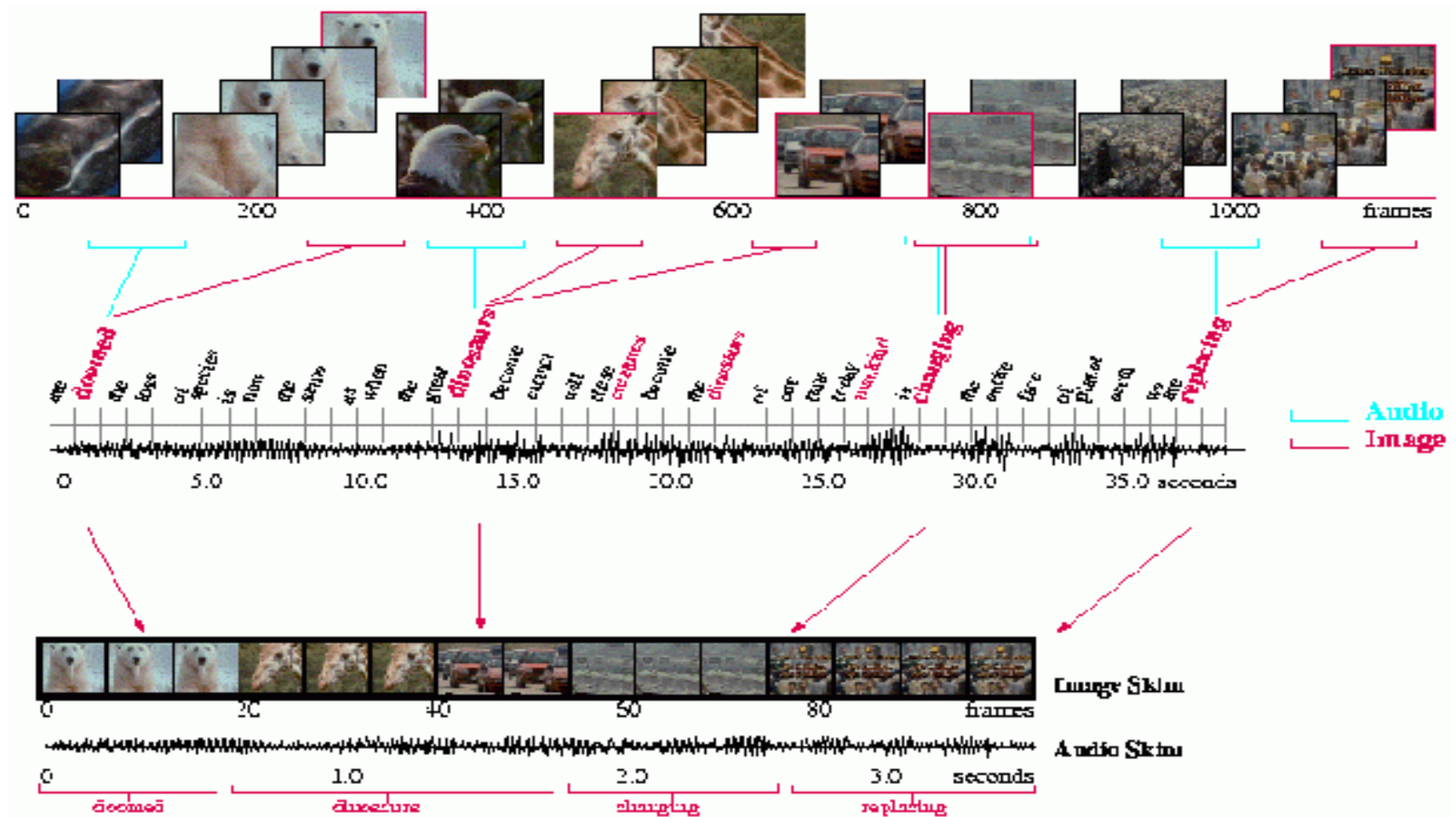
SCHOOL BUILDING



MODERN BUILDING

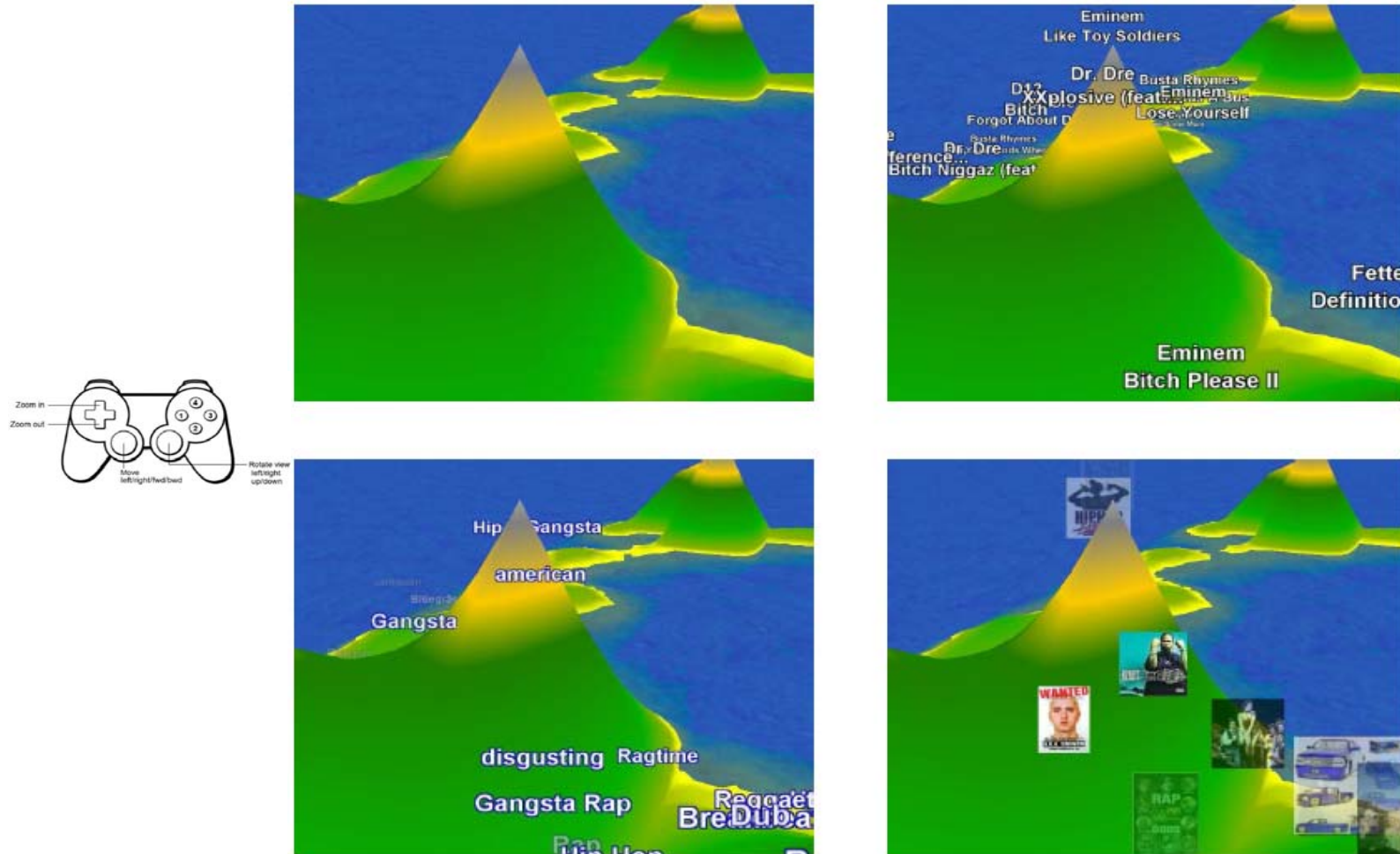


Video Summarization



© Carnegie Mellon University, 1995

Music Exploration [Knees2006]



Feature Extraction

- **Image**

- Color, texture, shape, spatial layout

- **Audio**

- MFCC, note frequency, rhythm, speech

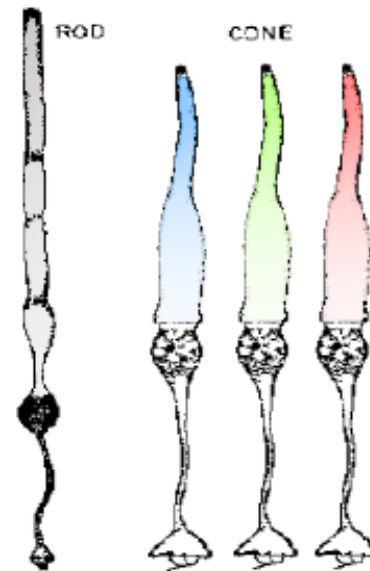
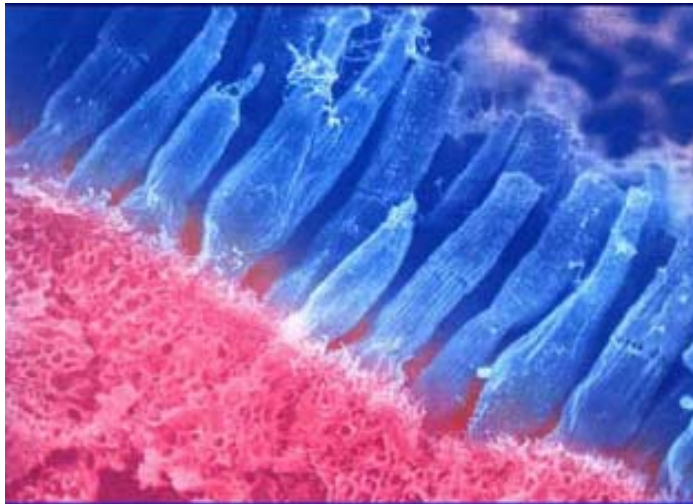
- **Video**

- 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



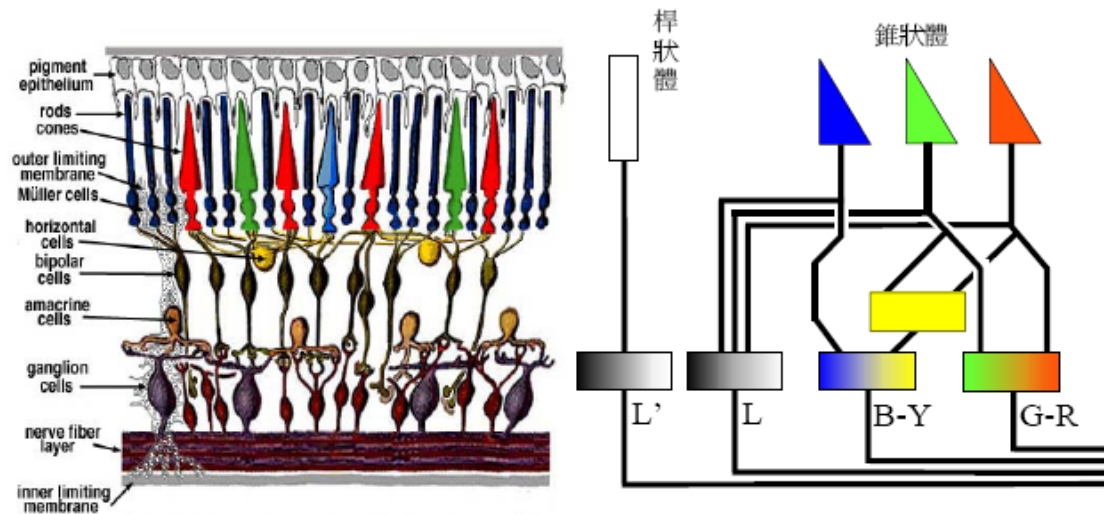
網膜下的桿狀細胞與椎狀細胞

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

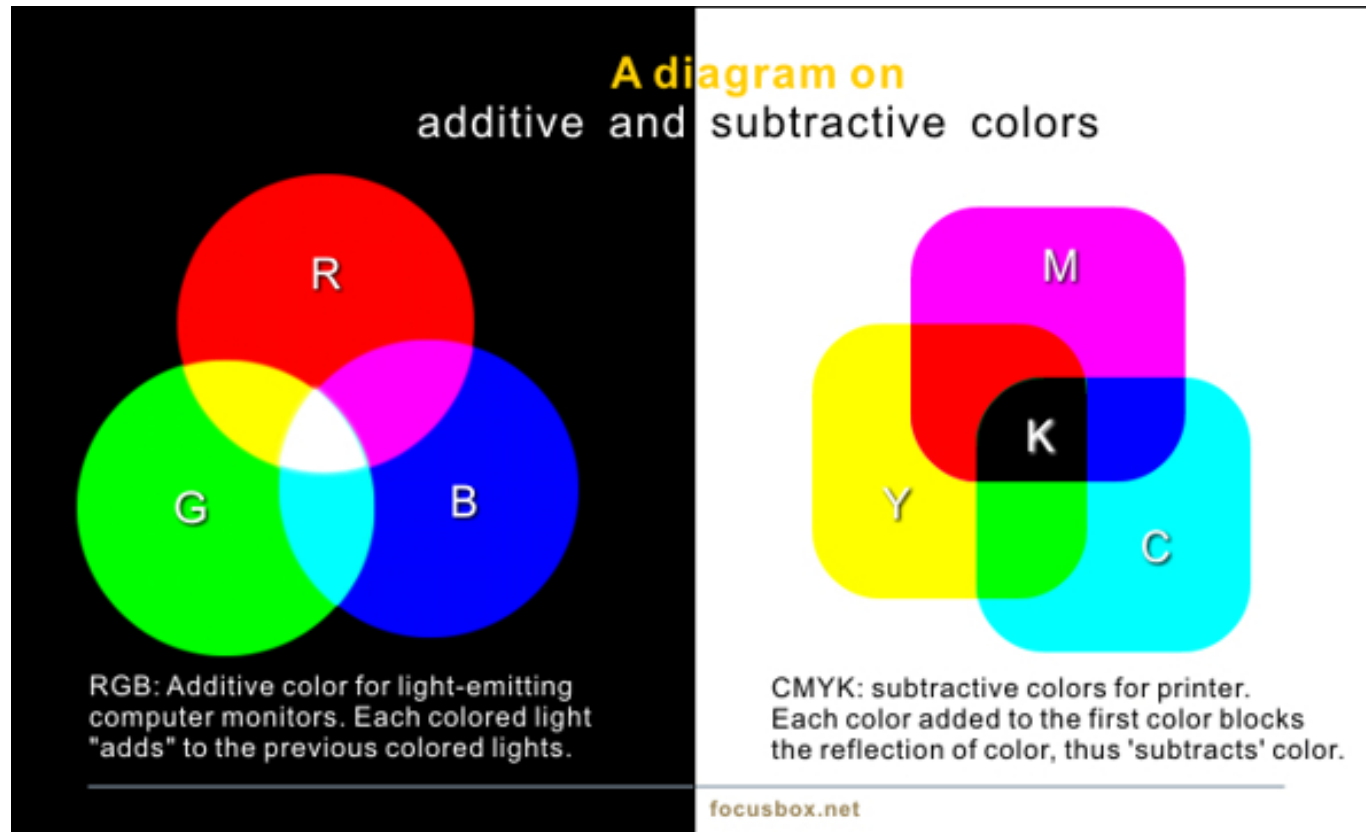
Visual Primitive

- 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



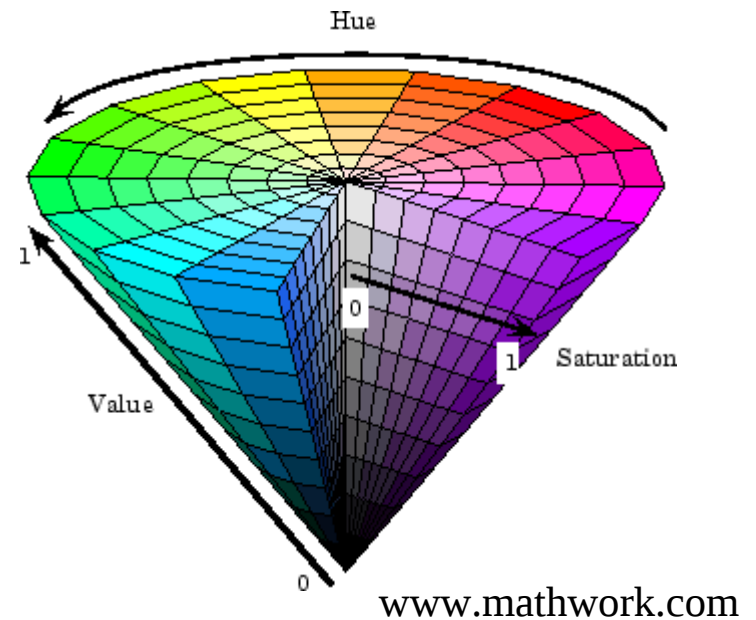
色光加法混色

色料減法混色

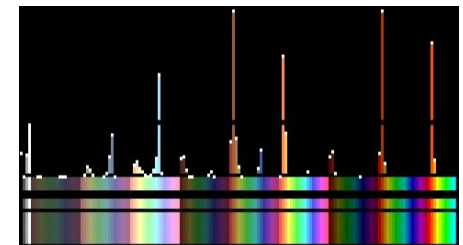
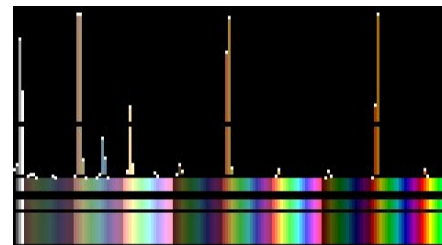
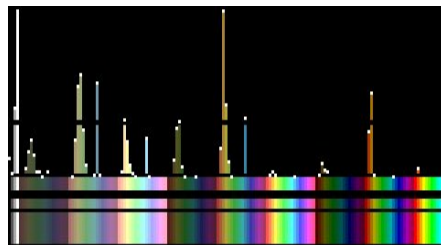
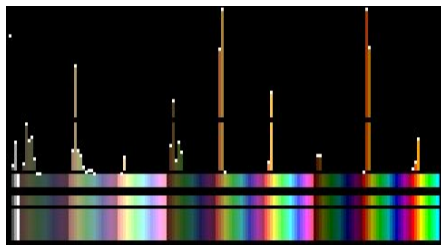
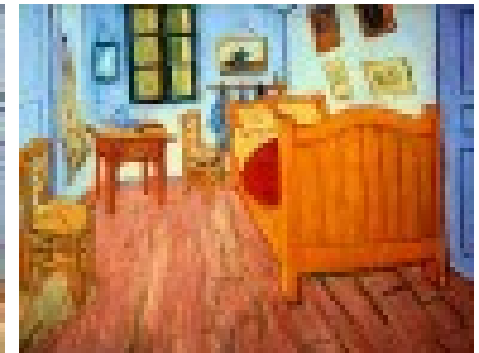


Color Space

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



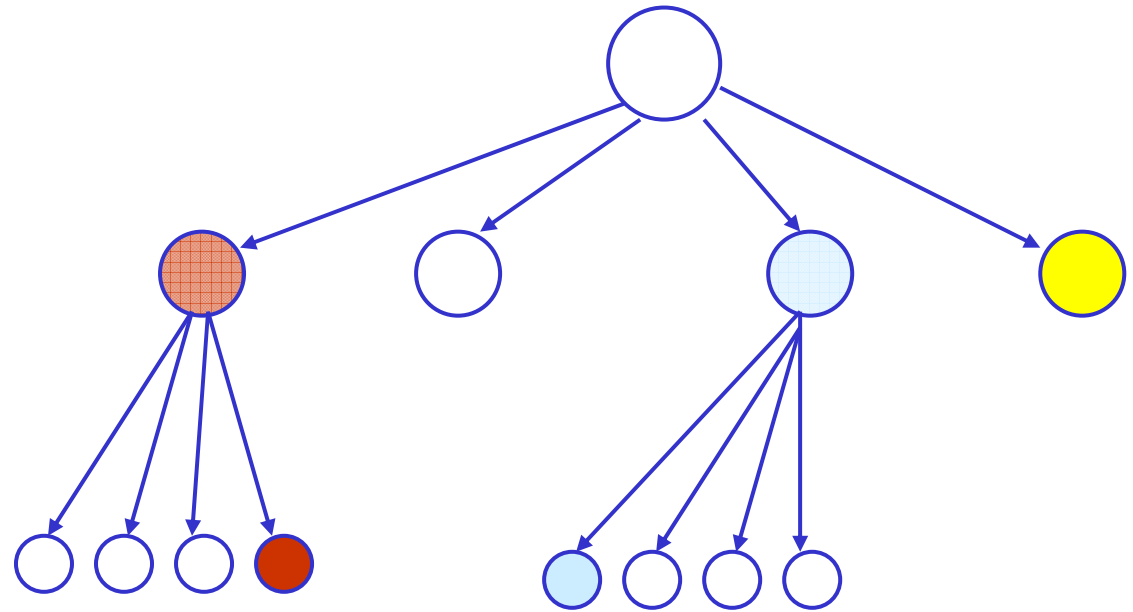
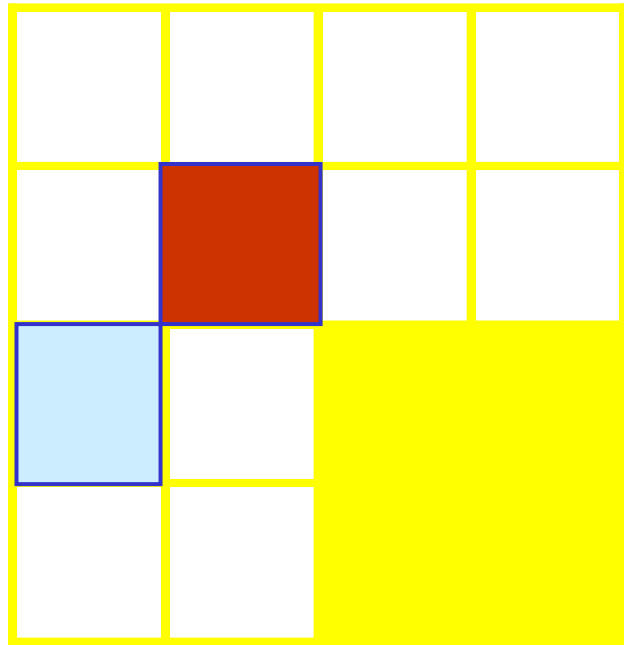
Color Histogram



Color Layout

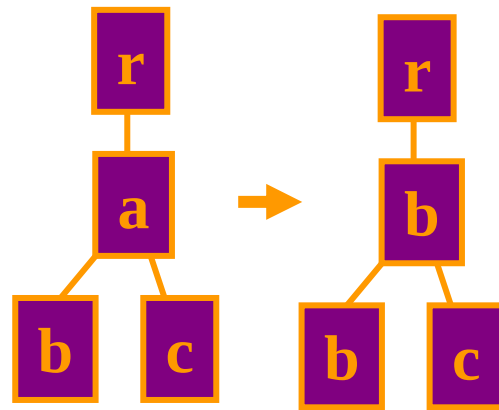
- More accurate color description
 - Histogram do not reflect the spatial information
- Quad-trees representations
 - Similarity computation between two quad-trees

Quad-tree Representation

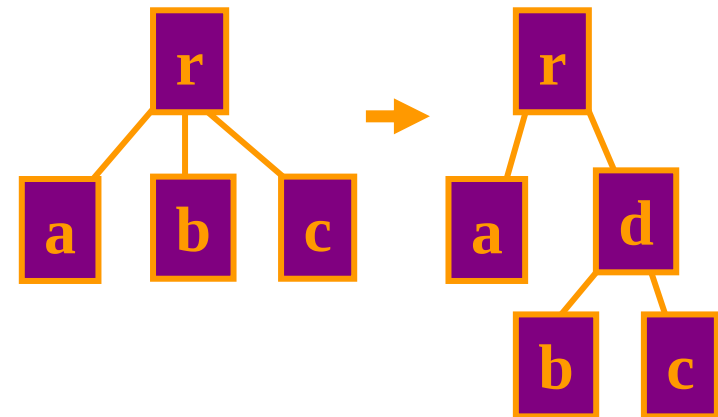


Quad-tree Similarity

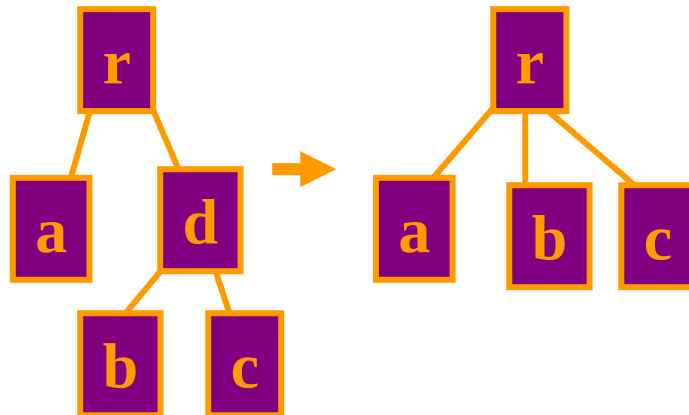
Replace



Insert

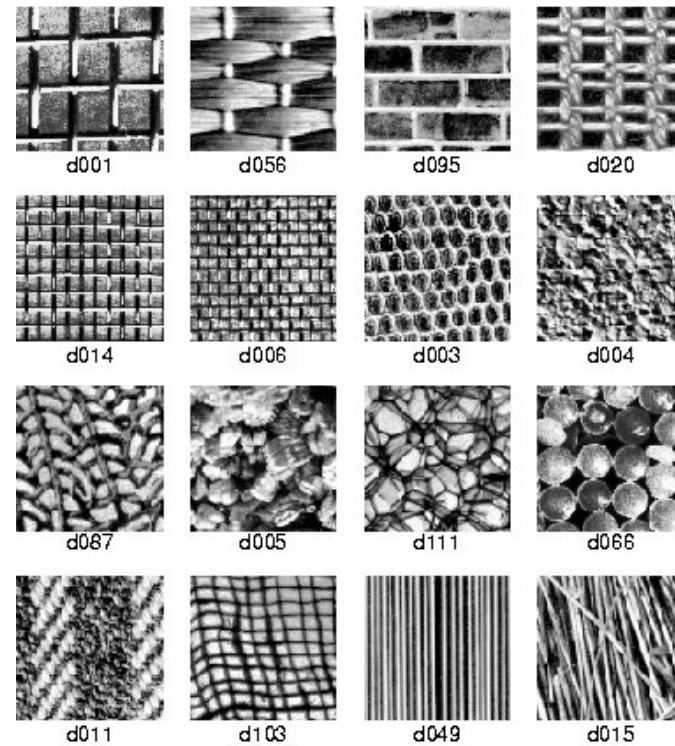
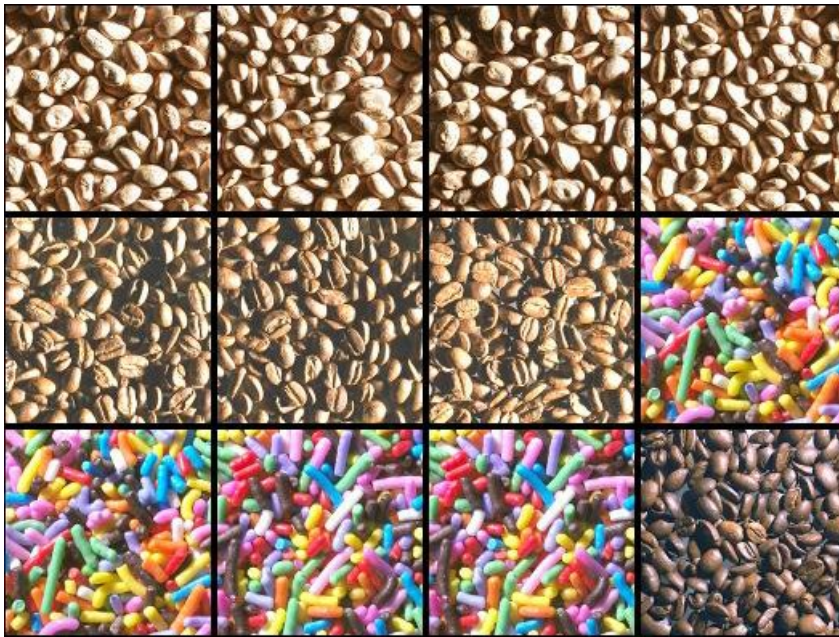


Delete



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?

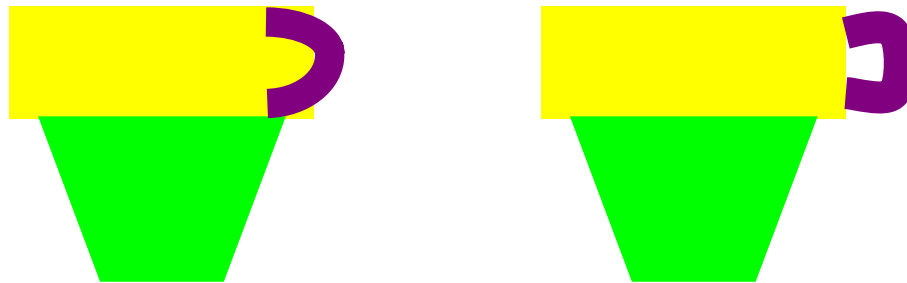
<i>Question</i>	<i>Yes</i>	<i>No</i>	NA
<i>"Did you need an image ..."</i>			
<i>"...with a particular object on it?"</i>	122	41	7
<i>"...with a particular color on it?"</i>	25	137	8
<i>"...with a particular texture on it?"</i>	23	137	10

(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** [Banz99]



Which is the canonical view?

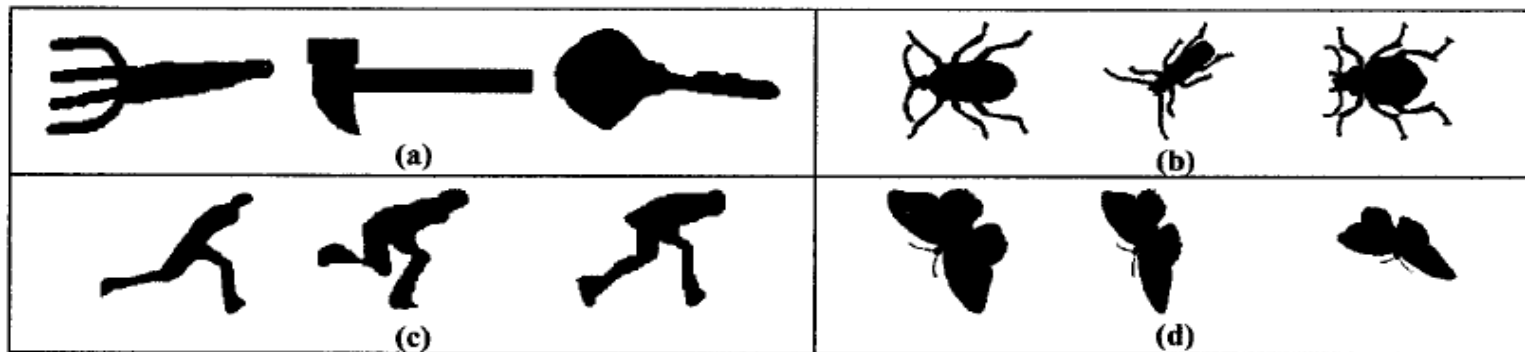
Shape Feature

- 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

	C			
A			G	F
		D		
	B			
			E	

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

D				
			E	
	A			F
C		G		
		B		

(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
 - $X=abcdgh, Y=aedfhr, \text{LCS}(X,Y)=adh$
 - 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

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

Directional and Topological Relation

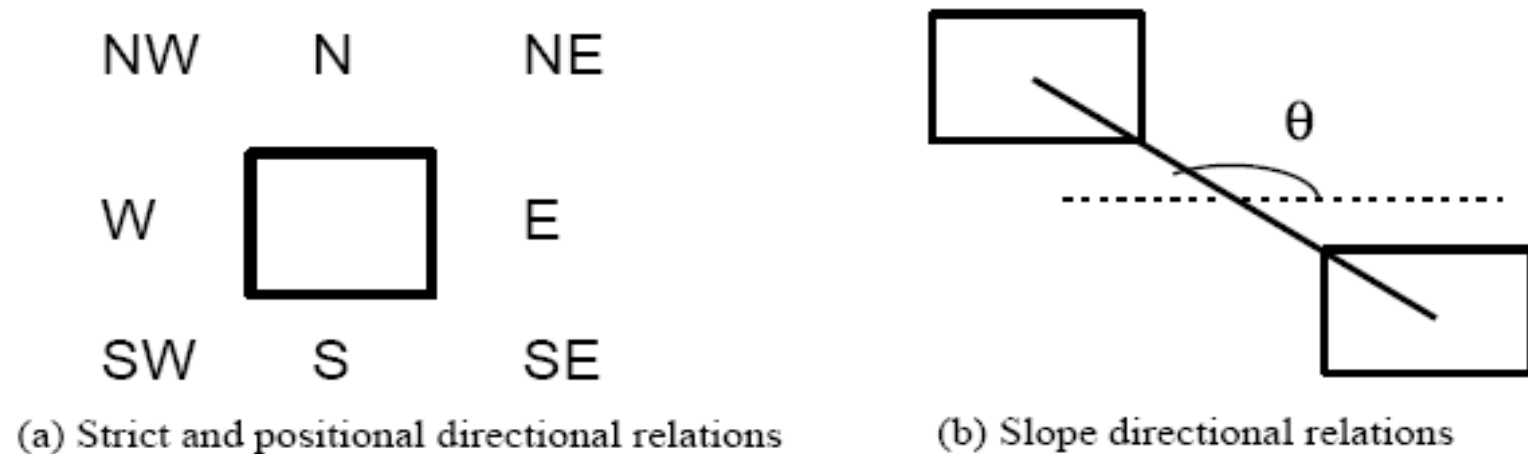


Fig. 2. Directional relations.

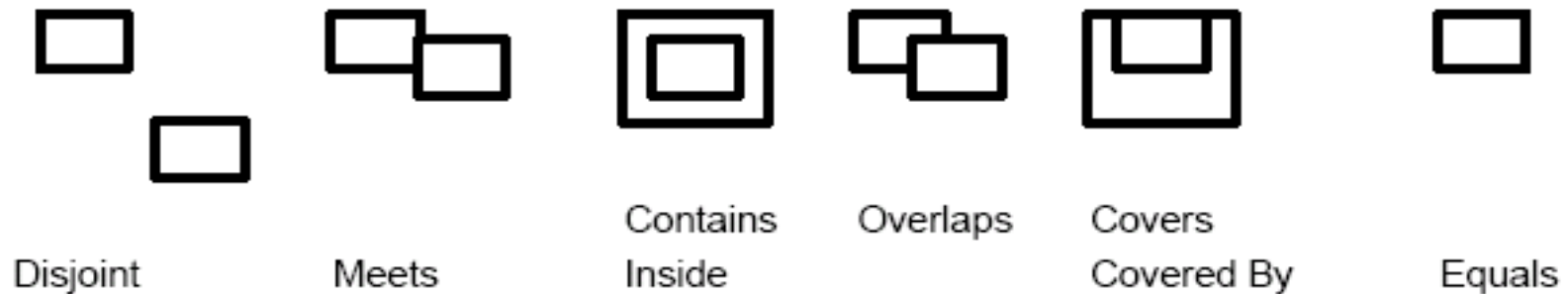


Fig. 3. Topological relations.

How People Organize Images in Computers

[Rodden99]

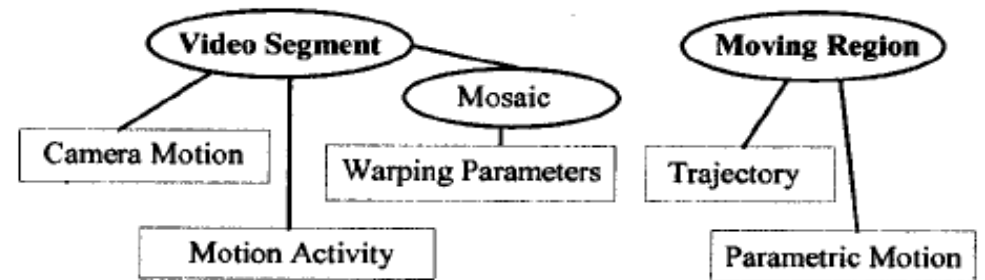
- Adding a title to an image 2.0
 - Typing notes to associate with an image or group of images 1.8
 - Speaking notes to associate with an image or group of images 2.5
 - Having spoken notes automatically recognized 2.2
 - Searching for images based on the text of your notes 1.5
 - Searching for images 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

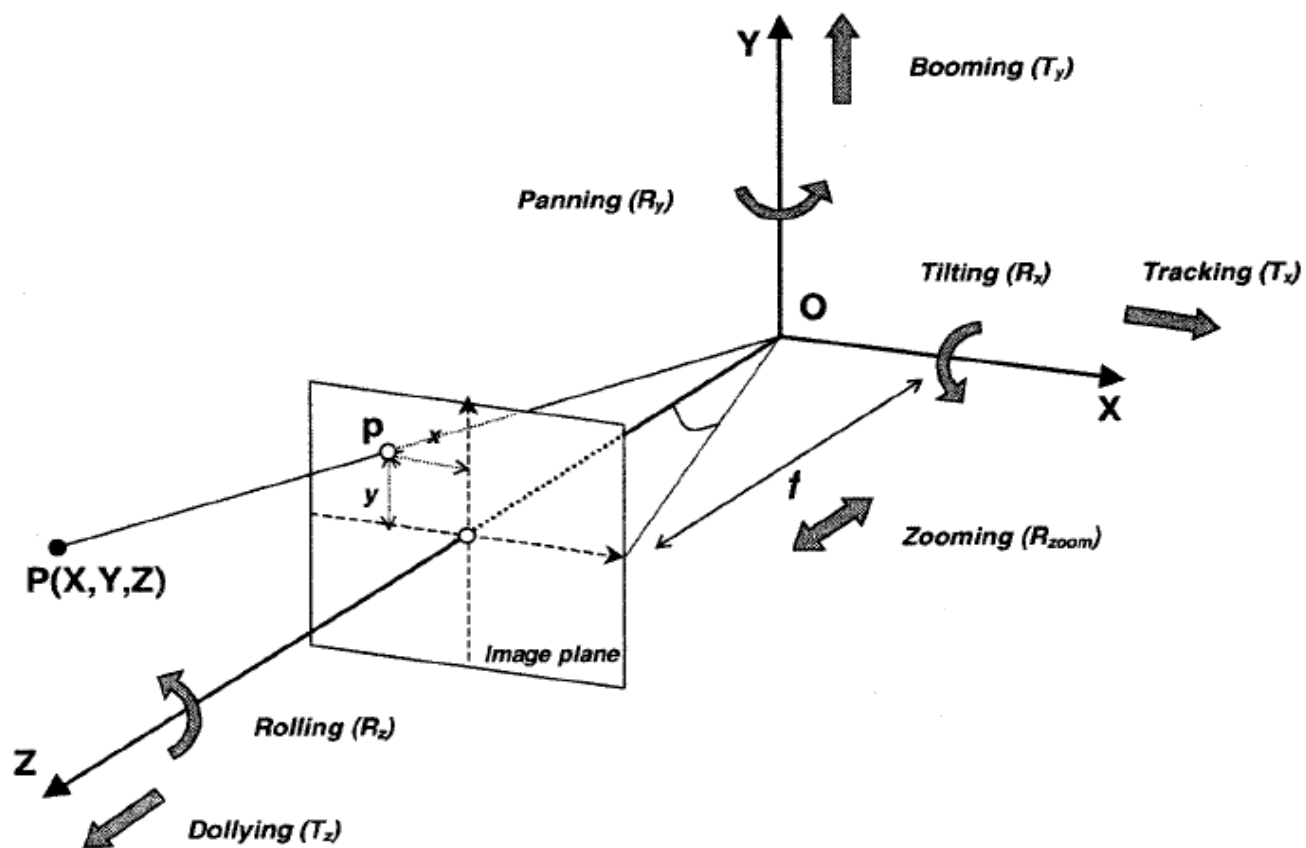
- The motion features of a video sequence provide the easiest access to its **temporal dimension**
- In combination with indexing based on still-image features, motion-based indexing is useful in contexts where motion has a **higher-level semantic meaning**

Motion Feature

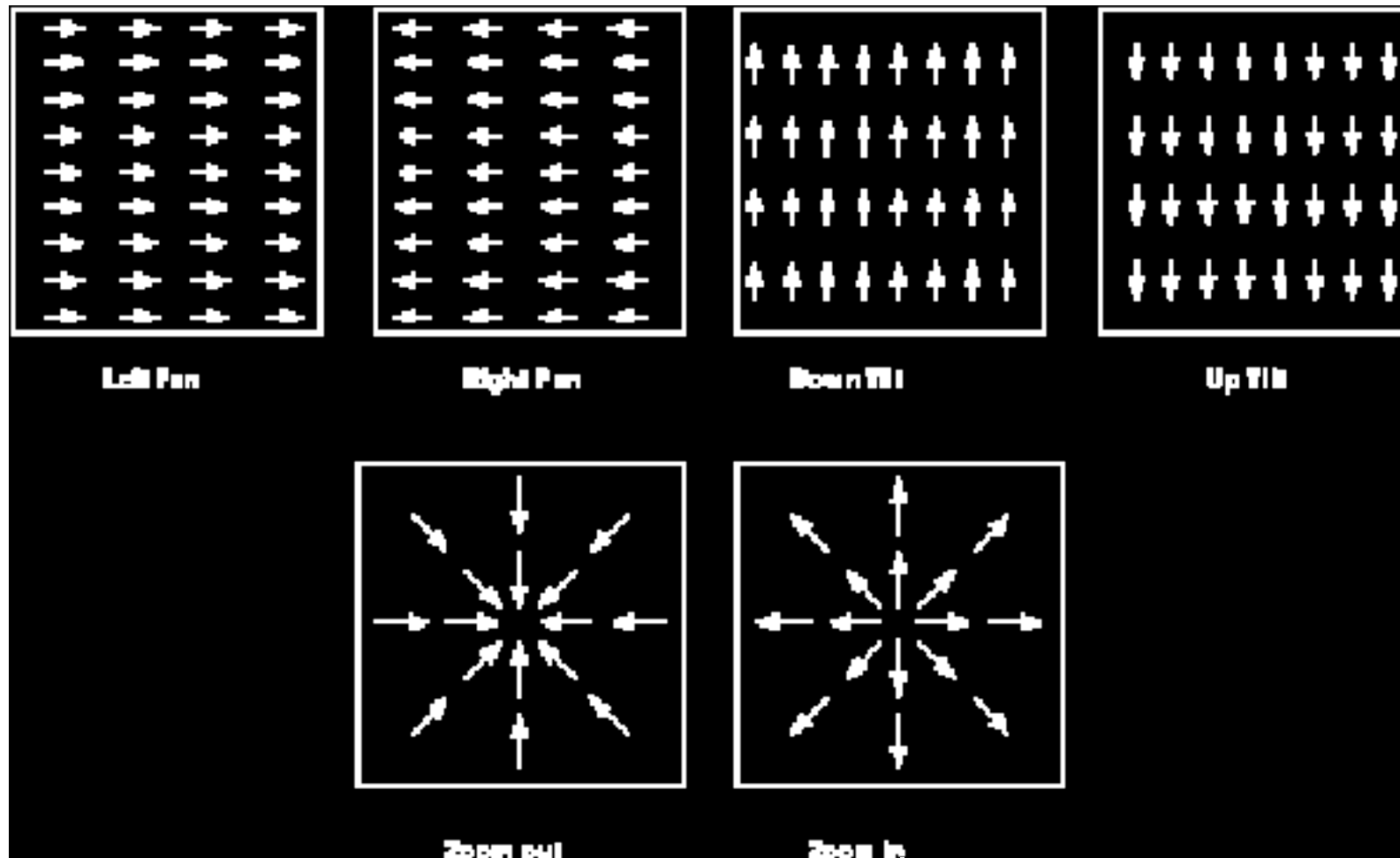
- Global motion
 - Motion activity
 - Camera motion
 - Warping parameters
- Local motion
 - Motion trajectory
 - Parametric motion



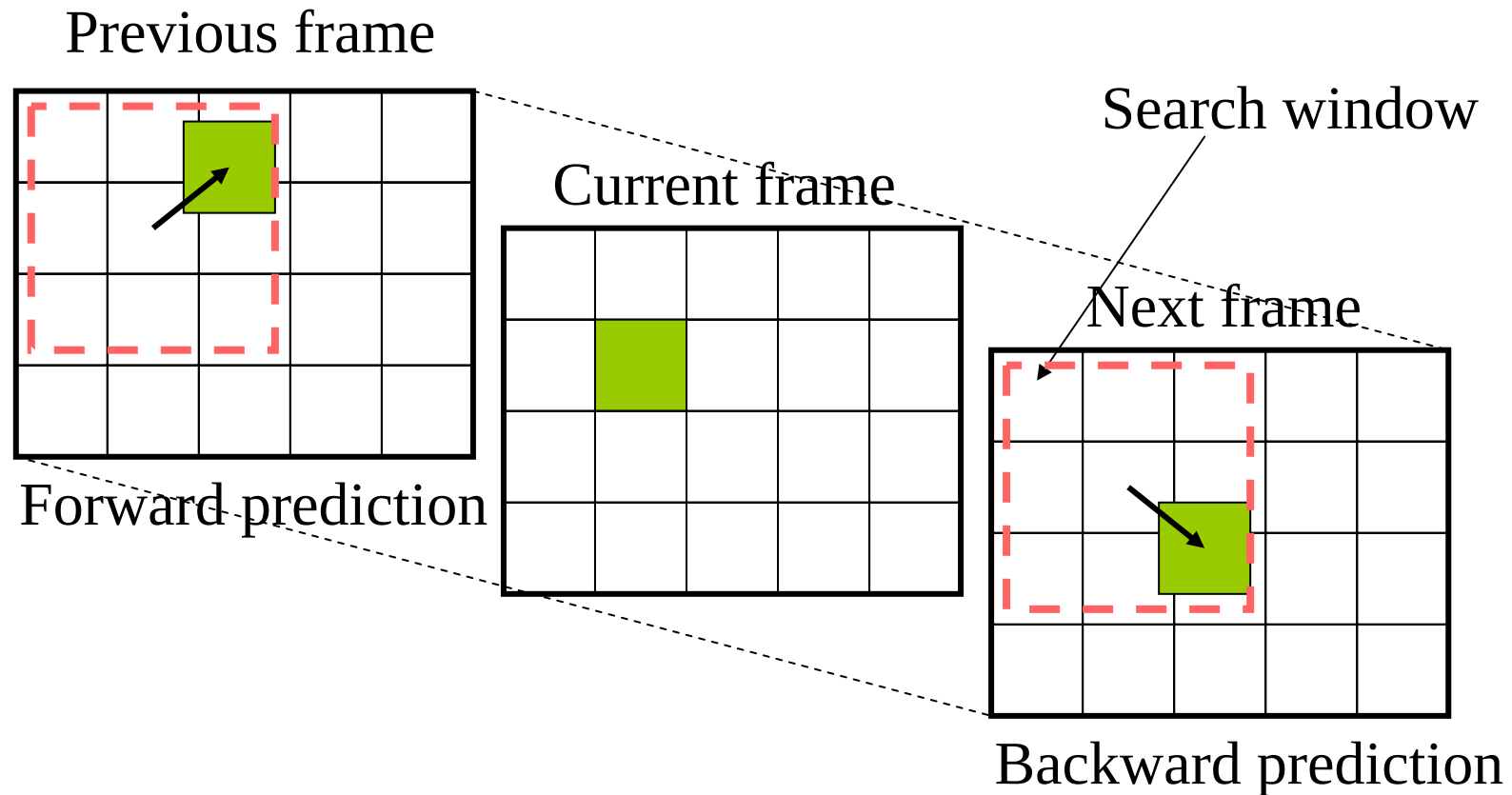
Camera Operation



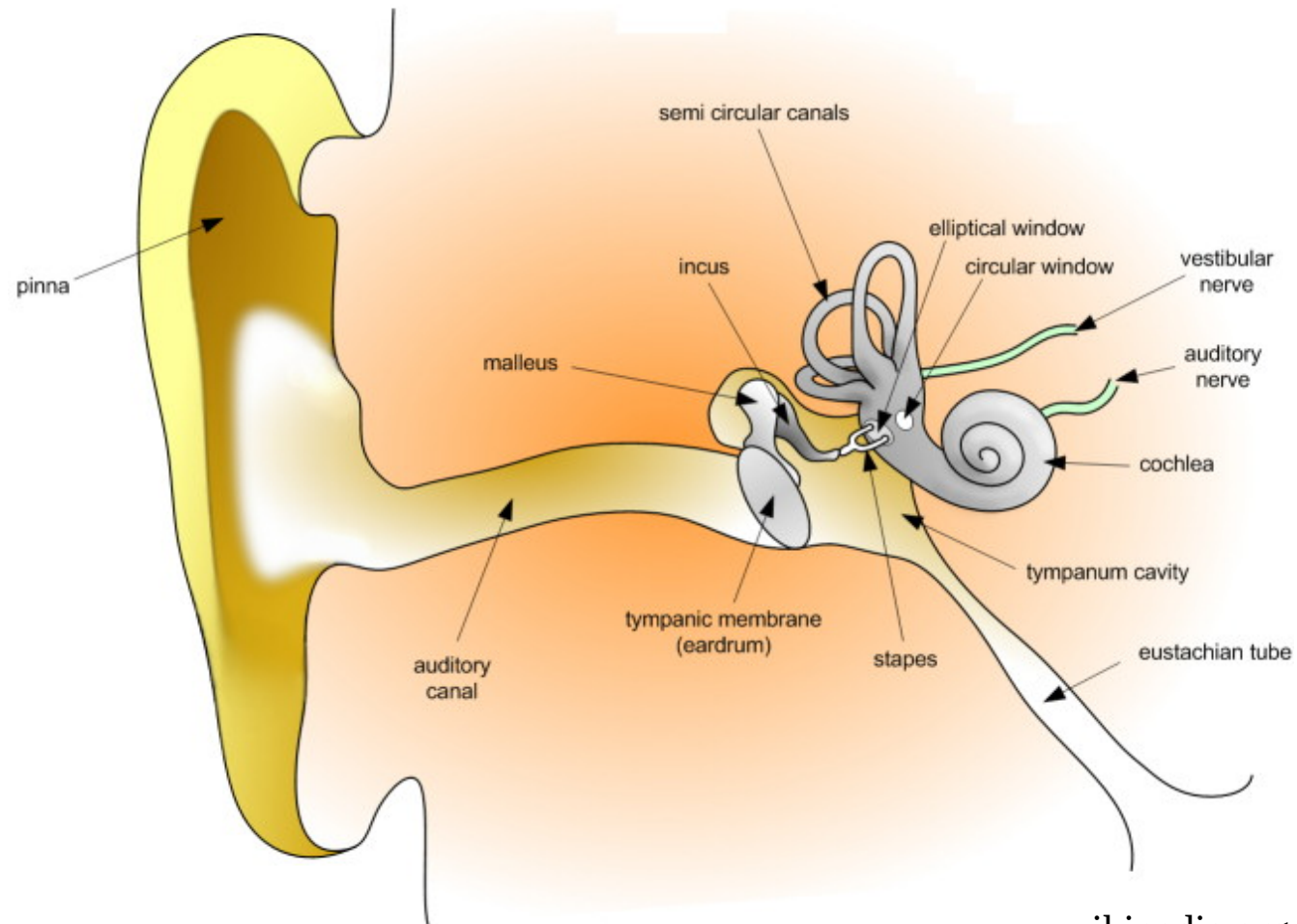
Optical Flow of Camera Motion



Motion Vector in Video Compression



Audio Primitive

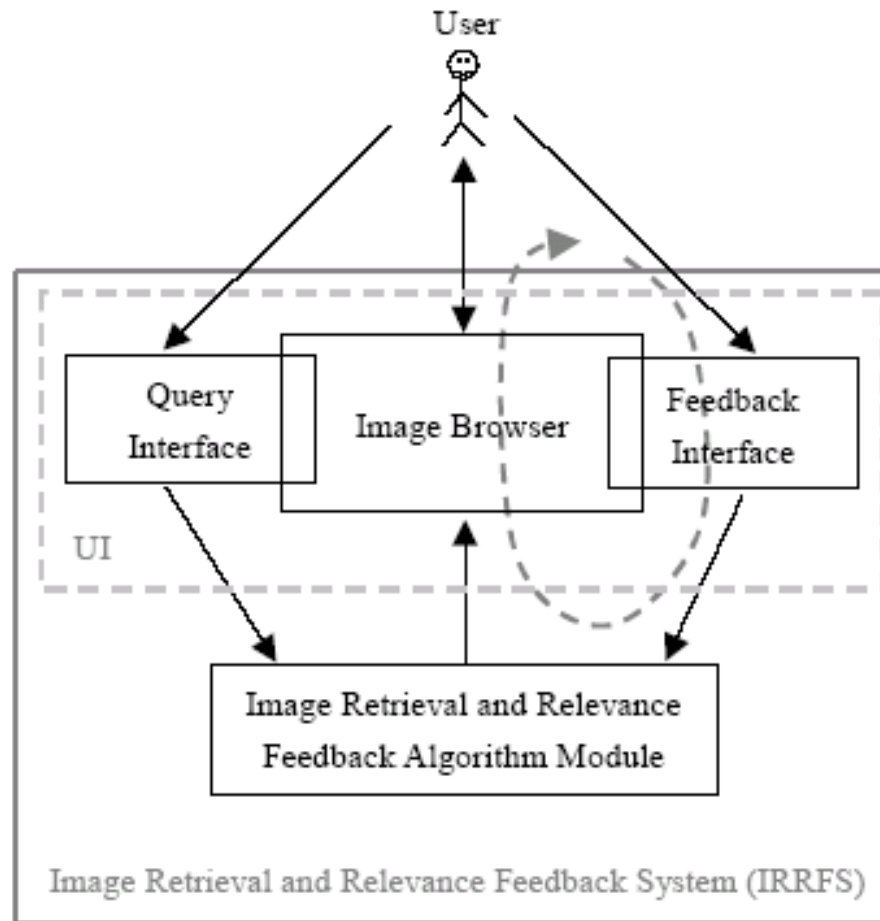


www.wikipedia.org

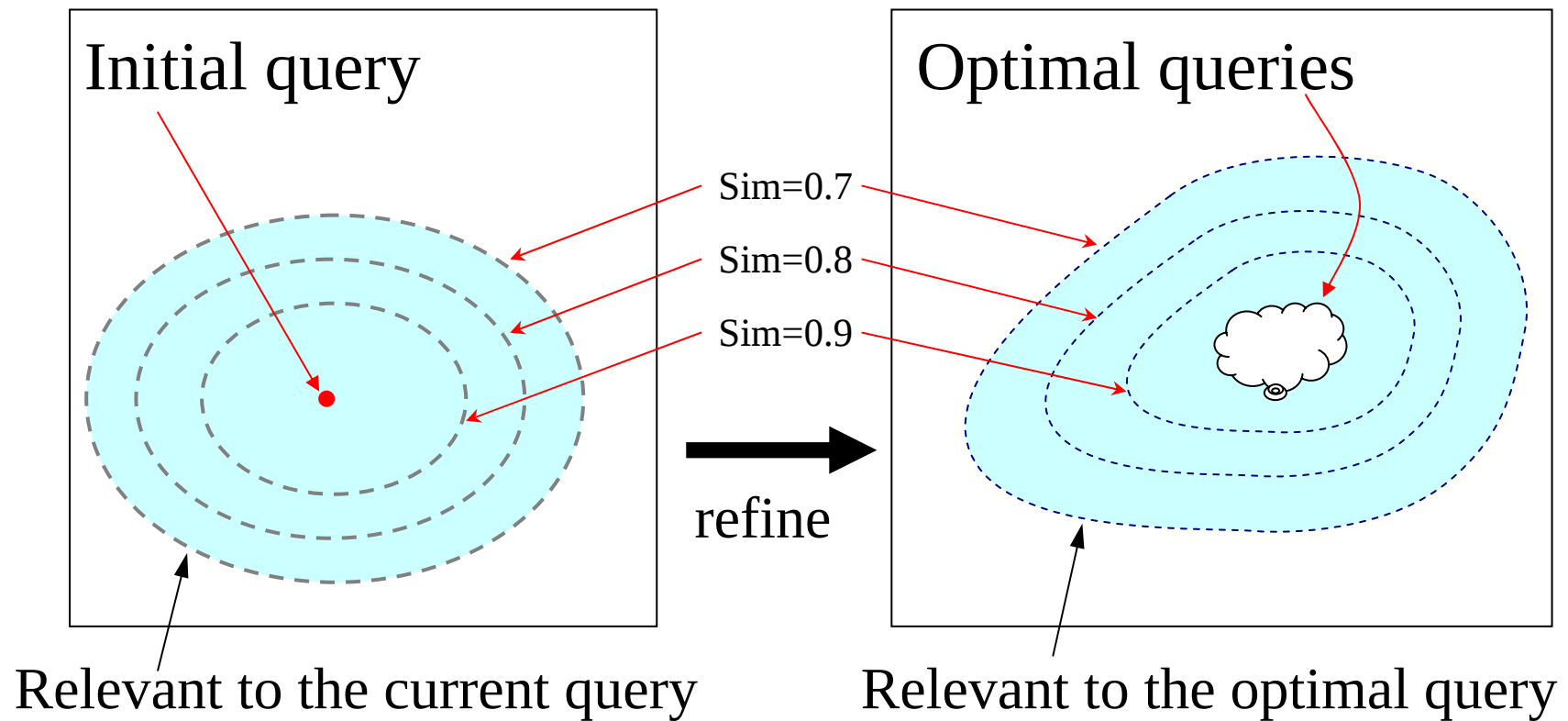
Audio Feature

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

Relevance Feedback Interface

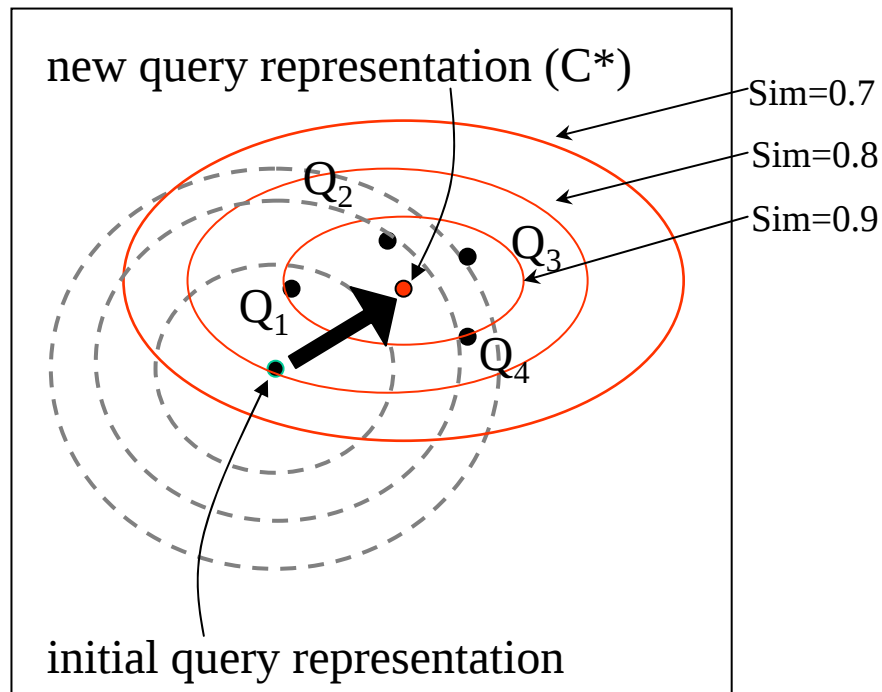


Relevance Feedback

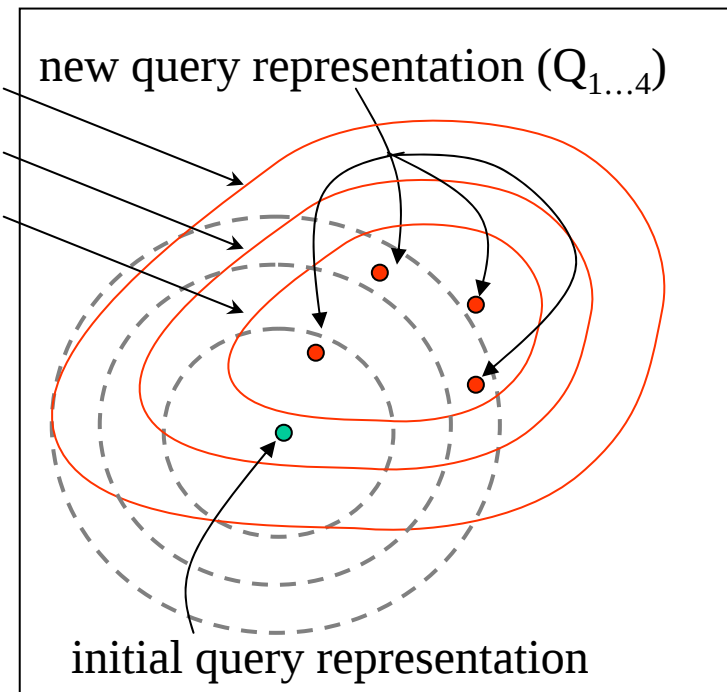


Relevance Feedback

Query Point Movement



Query Expansion



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