ECS736 Information Retrieval

Lecture: Visual Information Retrieval

Tuomas Ketola, Ammar Yasir

Roadmap of this lecture

- What is Visual Information Retrieval?
- Digital Media
  - Some Facts
  - Basic Definitions
  - Digital Media Content Management
  - Digital Media Databases
    - Characteristics
    - Main tasks
    - Key challenges

Roadmap of this lecture

- Information Retrieval
  - Annotation
  - Text Based Retrieval
  - Content Based Retrieval
    - Distance Metrics
    - Colour Representation
    - Dominant Colour
    - Colour Histogram
    - Colour Layout
    - Texture
- Multimedia Applications

What is Visual Information Retrieval

"Visual information retrieval is the activity of obtaining (i.e. search and retrieval) visual information resources relevant to an information need (i.e. a query) from a collection of visual information resources (e.g. image and video database)"

Digital Media

- Over the past few decades – significant proliferation of digital media
- Huge amounts of data and information
- People take pictures and record videos regularly – stored and/or published on the web
- **Challenges**: efficient searching, browsing, accessing, using and storing

Challenges

- Growing deluge requires more effective solutions for organizing, managing and searching media content
- Manual indexing is costly, time-consuming and inadequate
- New technologies are needed to automate processing and unlock value of large repositories
- Metadata standards are needed to support interoperable search

Some facts*

1,987,262,613,861,770,000,000+

- Bytes of information created since January 1st, 2011
- In 2011, the cost of dealing with information is down to 1/6 of what it was in 2005
- However, since 2005, money spent on managing data has increased 50% to €4 trillion

*http://www.emc.com/leadership/programs/digital-universe.htm

Some facts cont'd.

- For each of the following countries, the amount of data created, replicated and consumed each year from 2012-2020 is:
  - China x 24
  - India x 23
  - US x 2 every three years
  - Western Europe x 2 every two and a half years

Some facts cont'd.

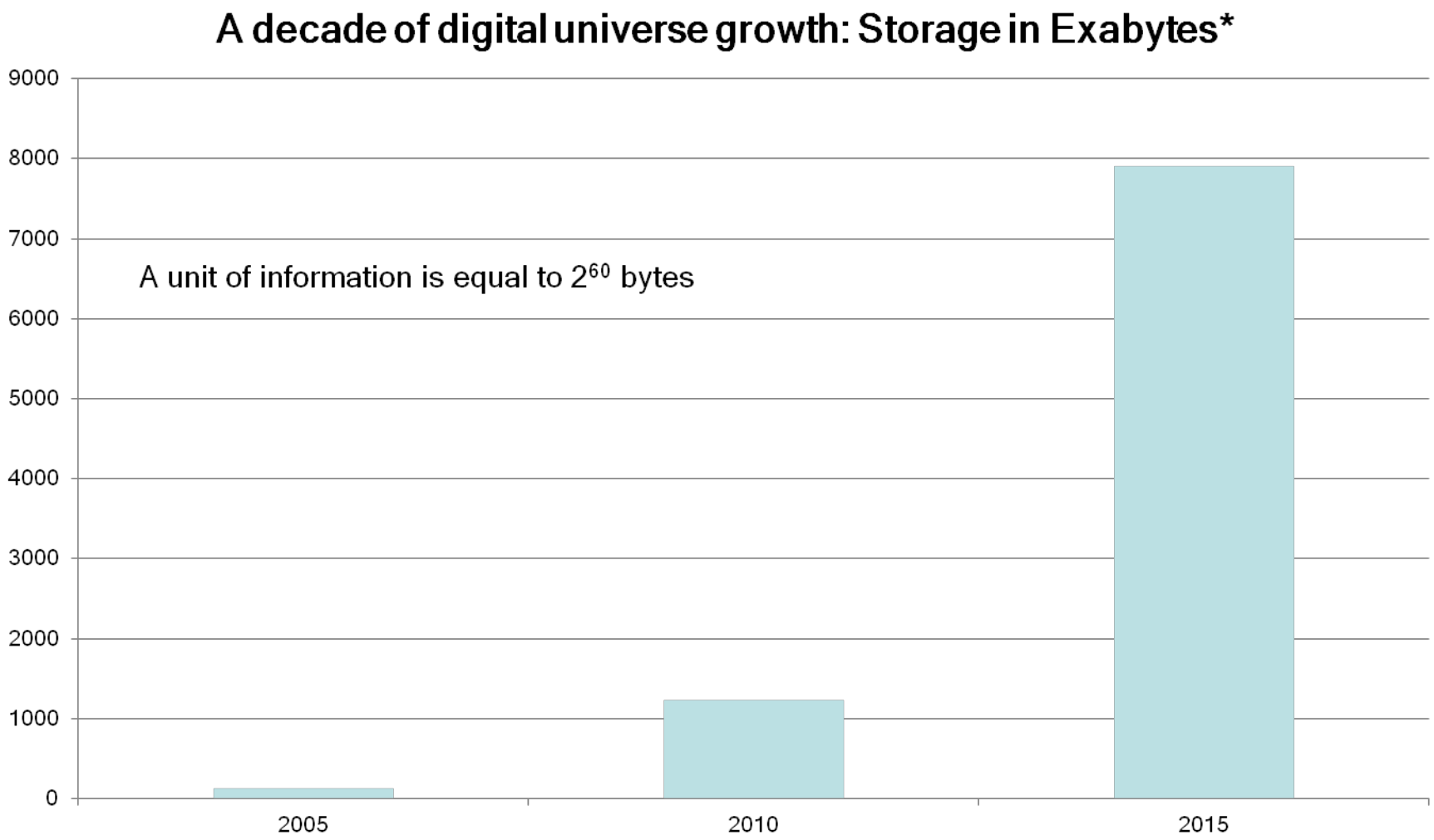
- 1 billion Facebook users in 2012
- 350 million photos uploaded to Facebook daily
- 400 million tweets a day in 2012
- 20 trillion texts in 2012
- 1 billion unique users to YouTube monthly
- 6+ billion hours of YouTube videos are watched each month
- 100 hours of video are uploaded to YouTube every minute

Some facts cont'd.

- Migration from analogue to digital:
  - 1970s, 1980s – computers and word processors replaced typewriters and filing cabinets
  - 1990s, 2000s – digital cameras overtake film
  - 2000s, 2010s – TV and movies start to go digital
  - 2010s, 2020s – RFID* tags and sensors surpass 1 trillion

Radio frequency identification – wireless non-contact use of radio-frequency electromagnetic fields to transfer data, e.g. Oyster cards

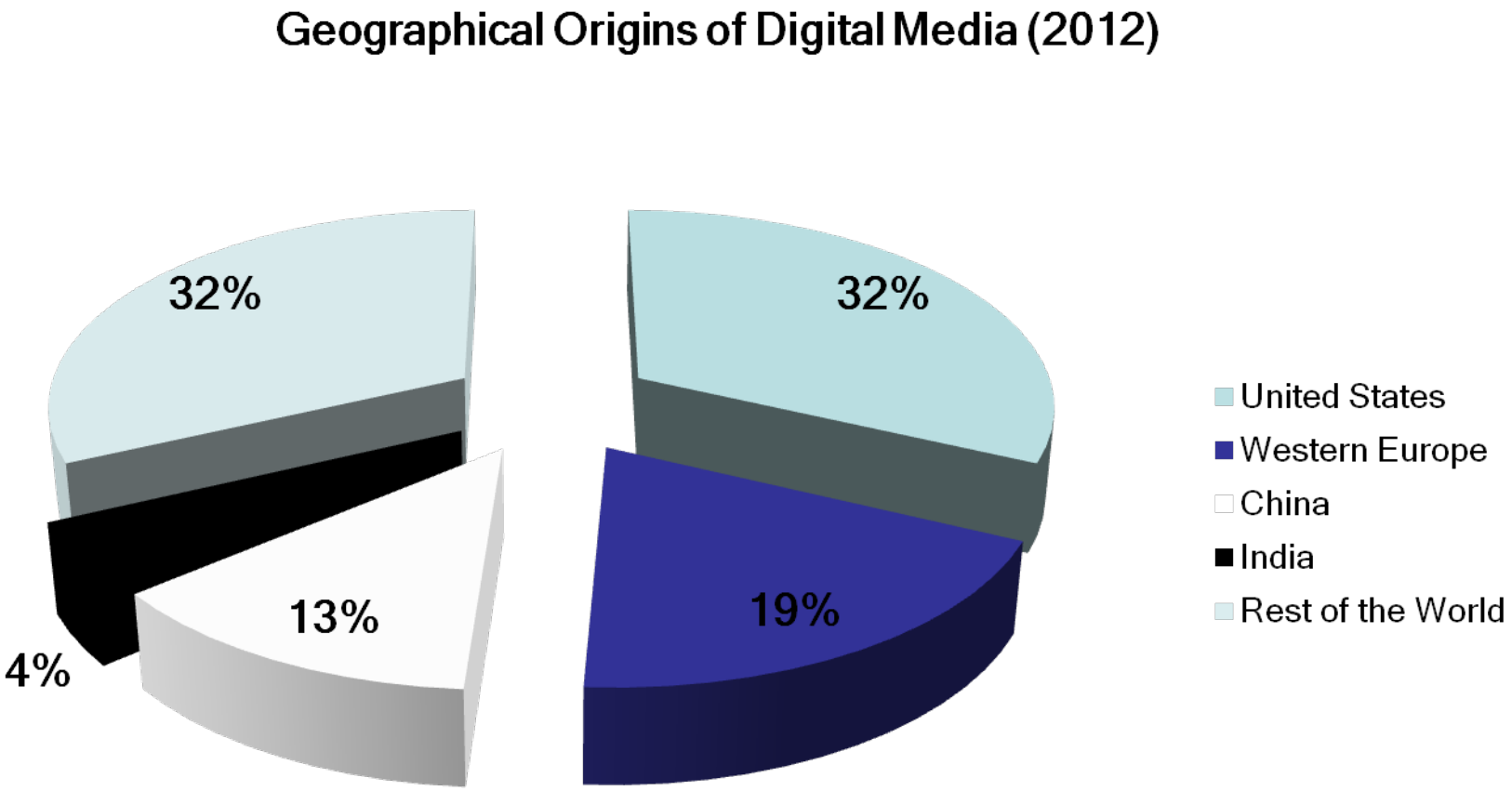
Some facts cont'd.



Some facts cont'd.

- In 2005, 48% of the digital information came from just the US and Western Europe
  - Emerging markets accounted for < 20%
- In 2012, emerging markets accounted for 36%
- In 2020, emerging markets will account for 62%
  - China alone will generate 21% of all data

Some facts cont'd.



Basic Definitions

- **Digital media**: combination of different media used to present multimodal information in conjunction with computer technology

- **Multimedia Database**: collection of related digital media data

- **Image Classification**: analyses the numerical properties of various image features and organizes data into categories in order to allow reuse

Basic Definitions

- **Image Clustering**: grouping images into meaningful categories to reveal useful information

- **Image Search and Retrieval**: browsing, searching and retrieving images from a large database of digital images

Digital Media

- Common multimedia data types are:
  - Text
  - Graphics – drawings, sketches and illustrations
  - Images – colour, black and white, photographs, maps, paintings
  - Animation sequences – animated images or graphic objects
  - Video – sequence of images (frames)
  - Audio
  - Composite multimedia – combination of two or more of the above data types

Digital Media

- Can be divided into 2 major classes
  - Continuous: media that changes over time, such as audio and video
  - Discrete: time-independent media such as text, still images and graphics
- Overwhelming number of file representations
  - TIFF, BMP, PPT, IVUE, FPX, JPEG, MPEG, AVI, MID, WAV, DOC, GIF, EPS, PNG, etc.

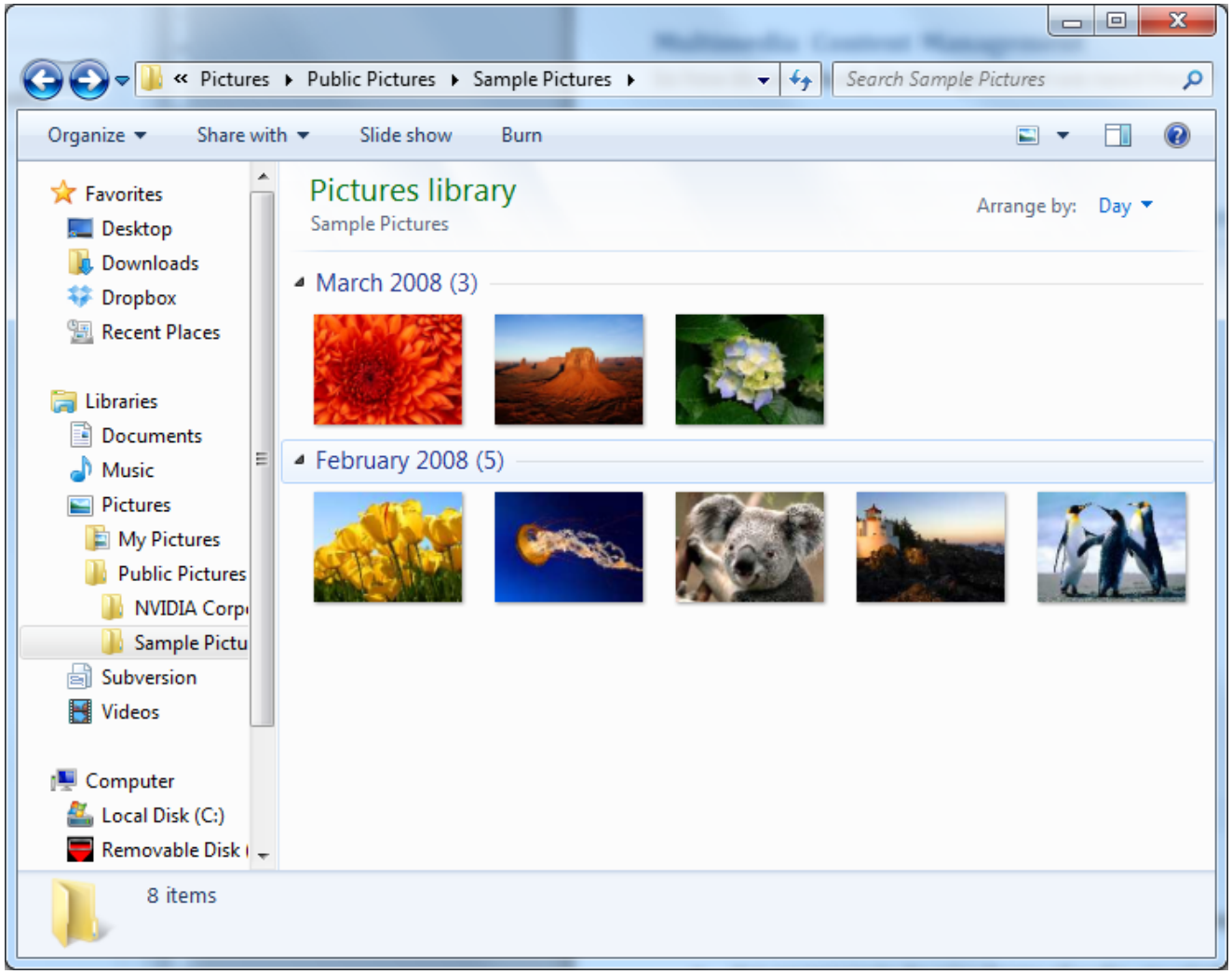
Digital Media

- Digital Media data
  - Diverse with different characteristics
    - Lack of structure
    - Temporality
    - Massive volume
    - Logistics
  - Multidimensional and hierarchically structured
    - Can have some relations among them, e.g. video has both temporal and spatial behaviour. The image sequences of the video should be displayed in order and in some dedicated time

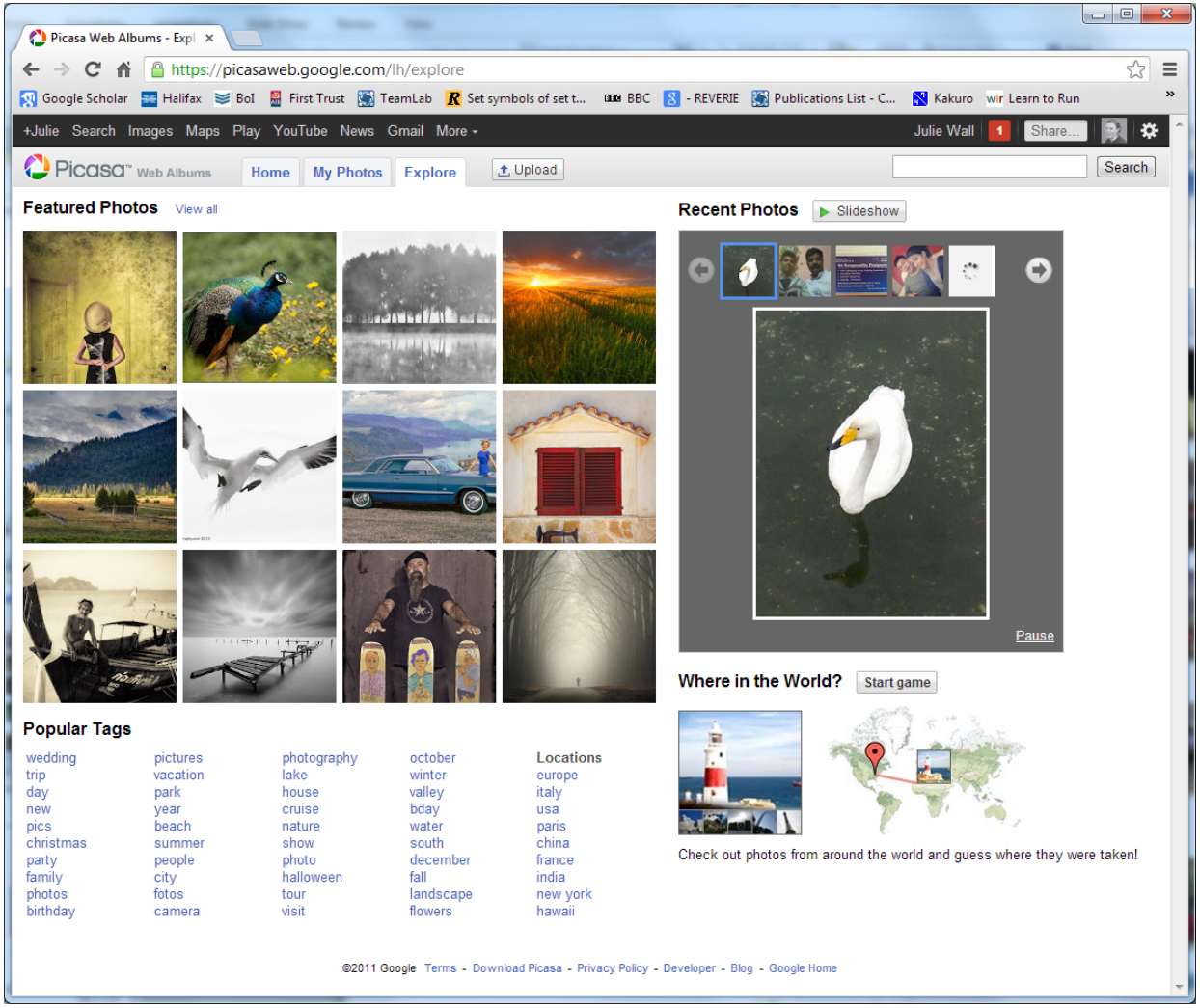
Digital Media Content Management

- So, how do we actually find images when we need them?
  - Using a clever directory structure?
  - Using ‘sophisticated’ applications?

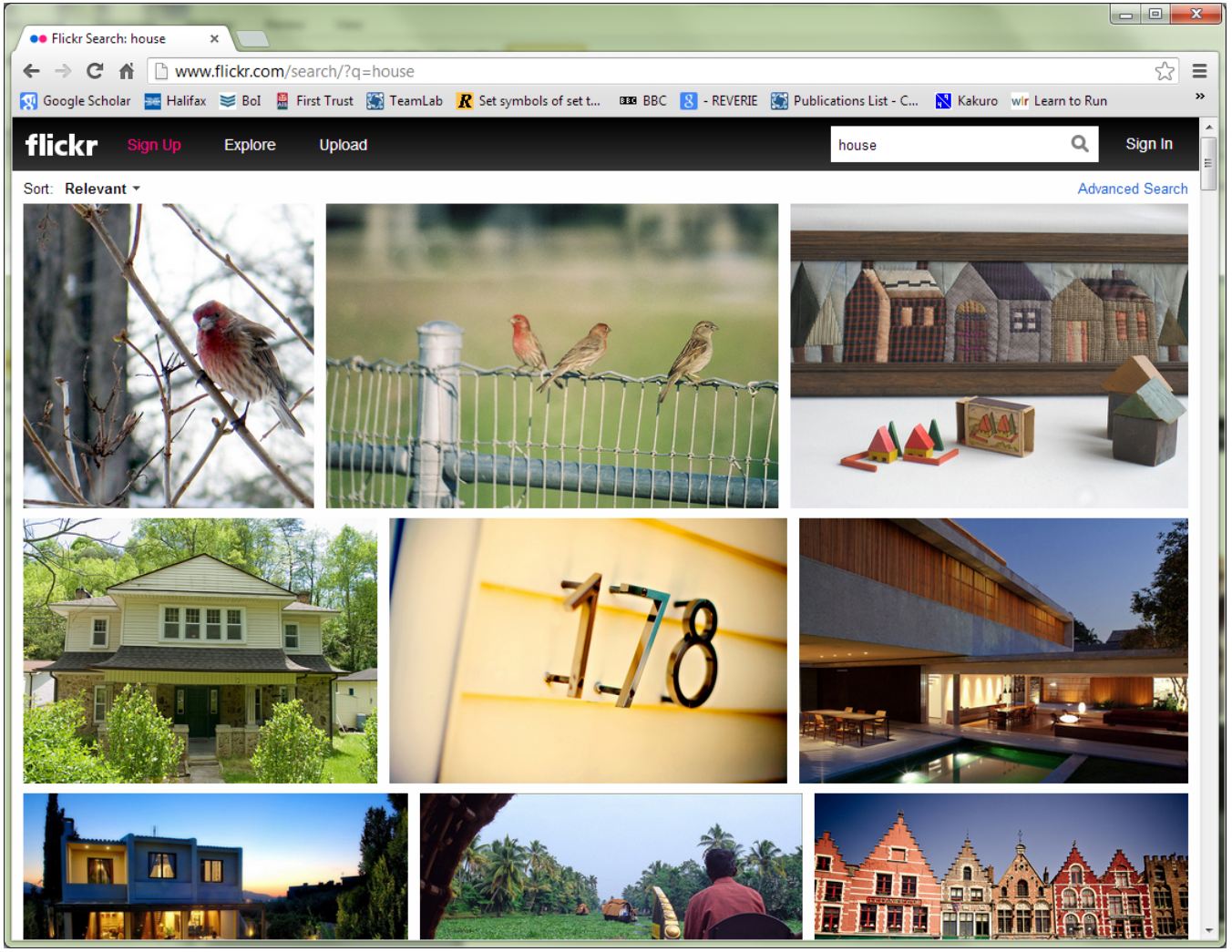
Windows folder structure with search facility



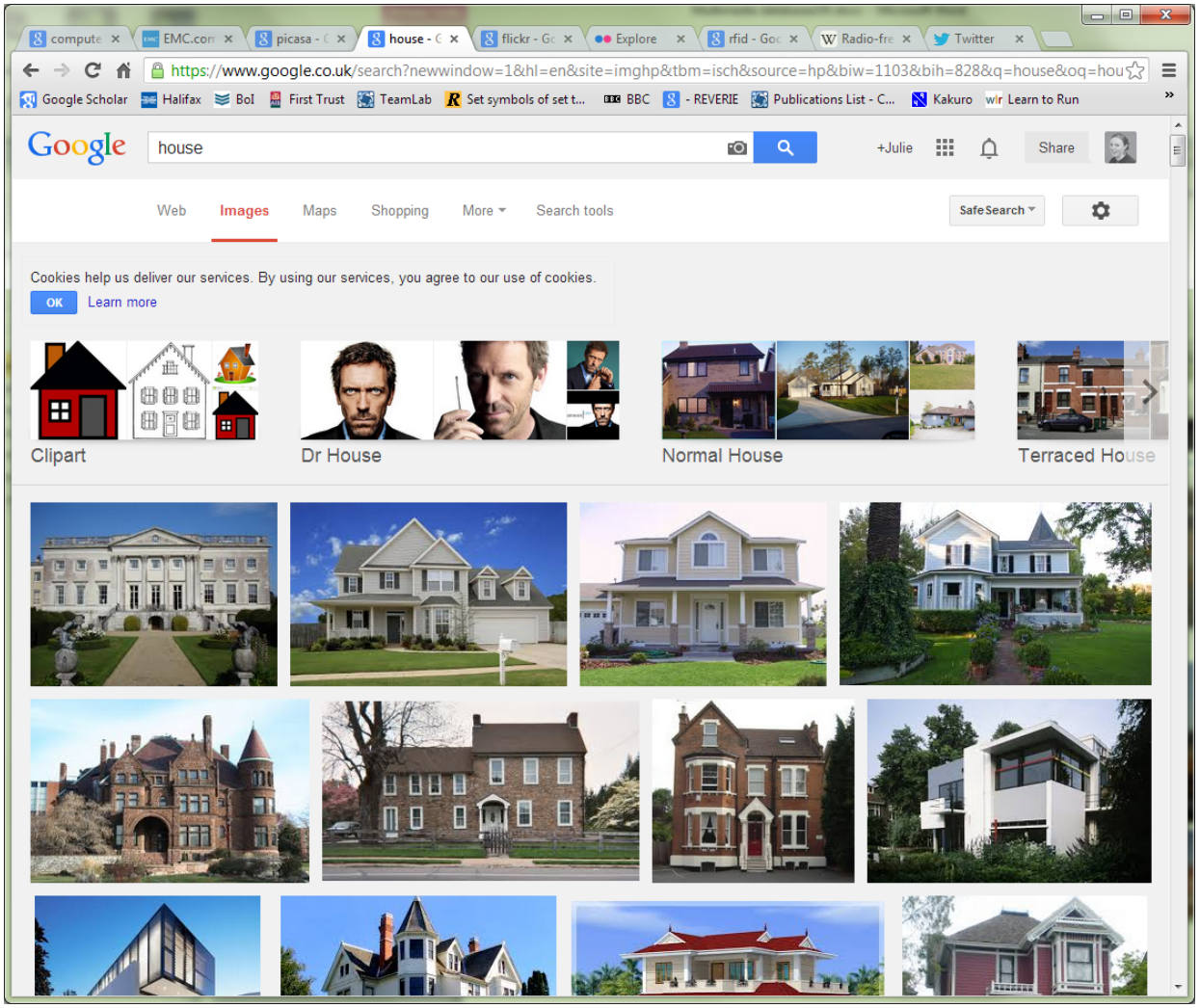
Picasa web albums



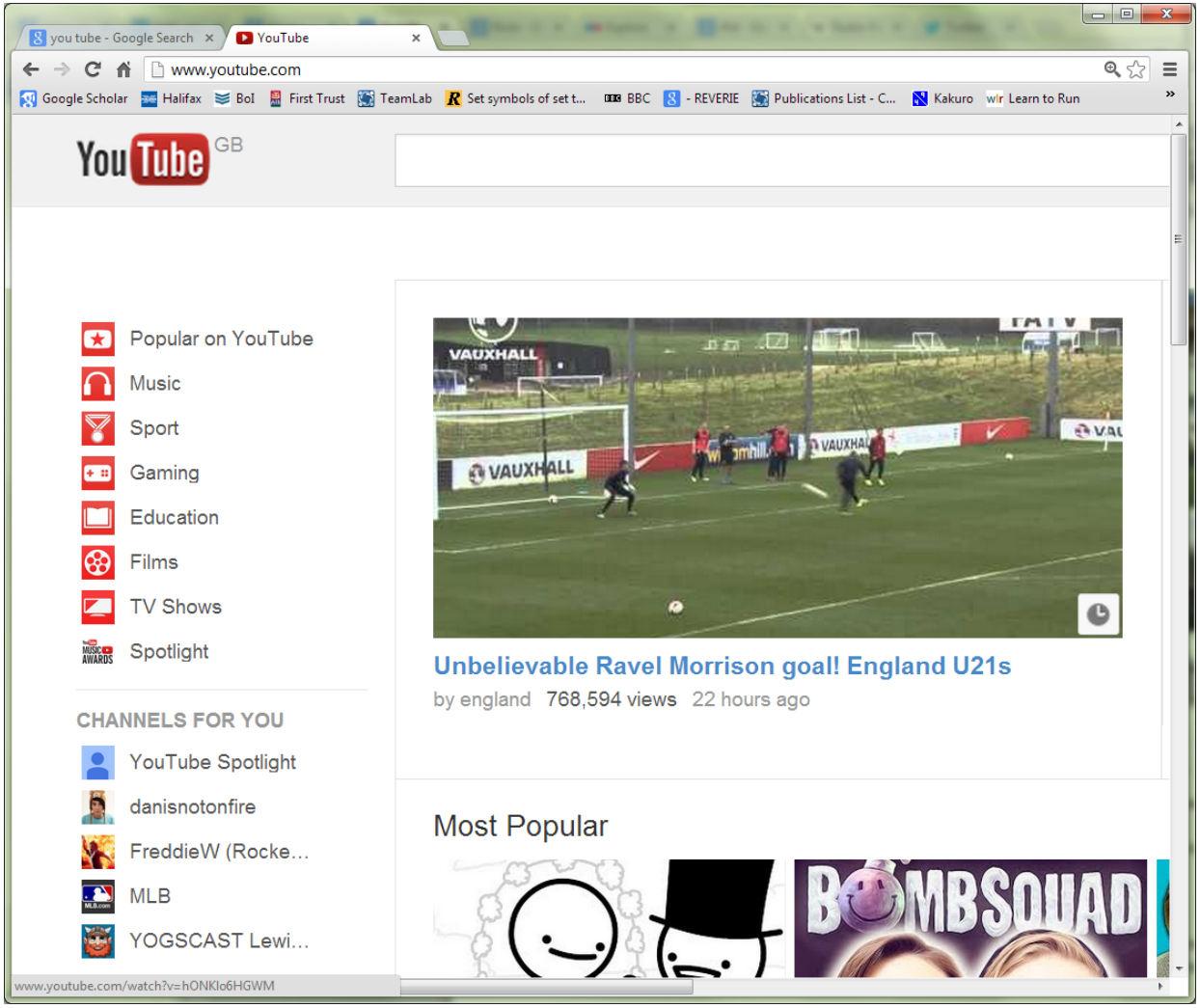
Flickr


Flickr Search: house
www.flickr.com/search/?q=house
Google Scholar
Halifax
BoI
First Trust
TeamLab
Set symbols of set t...
BBC
- REVERIE
Publications List - C...
Kakuro
Learn to Run
flickr
Sign Up
Explore
Upload
house
Sign In
Sort: Relevant
Advanced Search
178

Google image search



YouTube video search



Digital Media Databases

- Characteristics
  - Large object size
  - Synchronous delivery of multimedia objects
  - Digital media objects may have embedded timing constraints
  - Digital media object composed of multiple components
  - Queries may not be text or numeric based, but content-based
  - Most digital media transactions are long and require long processing and retrieval times
  - Digital media object presentation is very important

Digital Media Databases

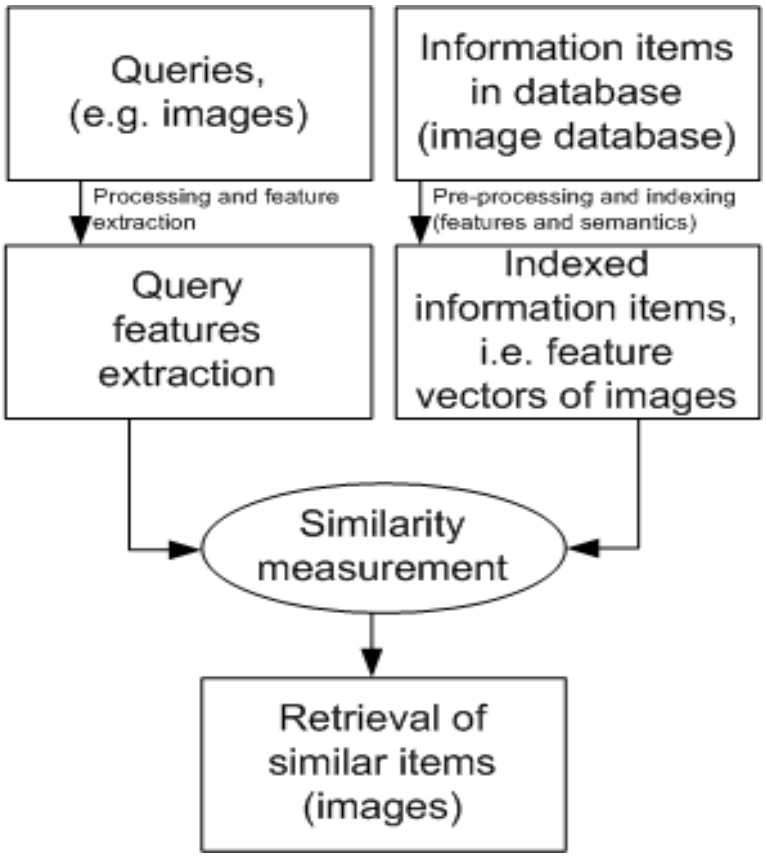
- Main tasks:
  - Provide cost effective storage management scheme
  - Provide basic operations as supported by traditional databases
    - insert, delete, search and update
  - Composition and decomposition of multimedia objects
  - Security and intellectual property protection

Digital Media Databases

- Main tasks cont'd:
  - Concurrency control and locking mechanism
  - Recovery
  - Support asynchronous operations
  - Organize, classify and query multimedia
  - Support abstract data analysis for search engine
  - Indexing and clustering
  - Search and retrieval of multimedia data should be accurate, quick and easy

Visual Information Retrieval

- Vast amount of multimedia data stored
- It has special characteristics and requirements
  - Very different to alphanumeric data
  - Text document information retrieval, e.g. Google search, has a limited capacity to handle effectively multimedia data



Visual Information Retrieval

- There are several existing techniques for retrieval of multimedia data
  - Text based retrieval
  - Content based retrieval
  - Region based retrieval
  - Context based retrieval
- We will discuss the first two in this lecture.

Annotation

- Images are indexed with a set of relevant text phrases which describe the content
- Subjective – different people may use the same phrases to describe the same or very similar content
- Laborious – takes a lot of man-hours to label large multimedia databases
  - Never complete
  - Goal is required

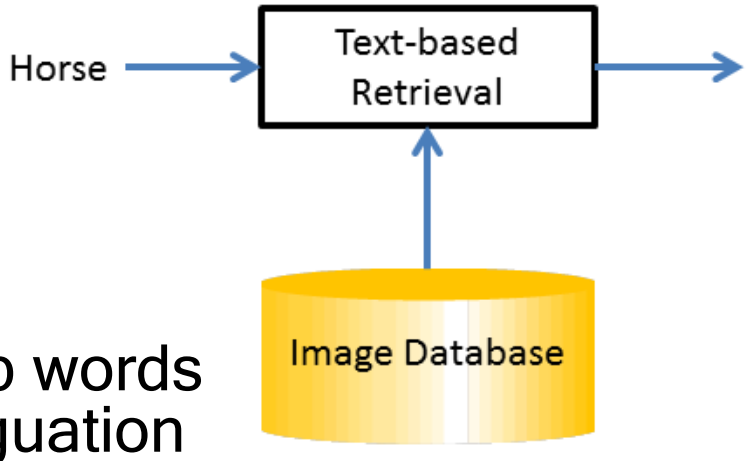
Annotation

- Many things are hard to express, feelings, situations, what is scary, etc.
- Mistakes such as spelling errors, spelling differences (US vs. UK), or weird abbreviations (medical data) cause problems

Annotation phrases for this image could include: Mother, child, vegetable, yellow, green, purple, etc.

Text Based Retrieval

- Textual data is associated with the object
  - Caption,
  - tags,
  - description,
  - comments,
  - etc.
- The text itself is processed for stop words removal, disambiguation



Text Based Retrieval

- Advantages:
  - Easy to construct queries
  - Image retrieval is fast
  - String matching is a relatively resource friendly task

Text Based Retrieval

- Disadvantages:
  - Availability of textual information is time consuming and search has no automation
  - Manually generated information is subjective
  - Relevance of textual information is unknown, single words can mean radically different things
  - Relevant images might be left out due to:
    - Lack of specific keyword in the query
    - Mistakes or misspellings in the query string
  - Retrieval depends on the annotator and retriever sharing some common vocabulary or language
  - Use of synonyms would result in missed results

Content Based Retrieval

- Makes direct use of the content of the multimedia data
  - Does not rely on human annotation
  - Computer vision techniques
  - How difficult is it to encode, perceive, convey and measure similarity, e.g. between two images?

Content Based Retrieval

Are these images
similar?

Content Based Retrieval

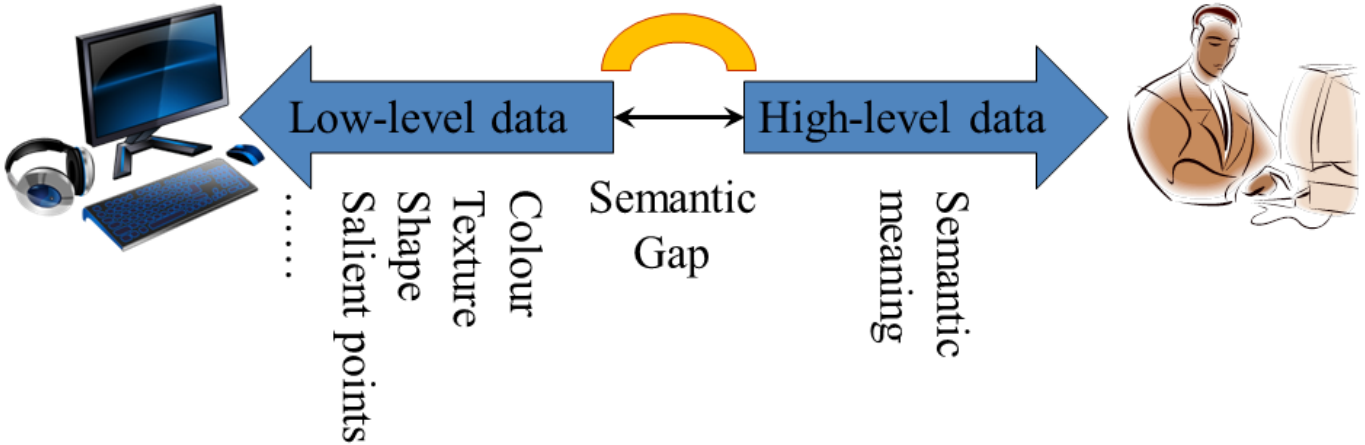




Which of these small images are most similar to the big one?

Content Based Retrieval

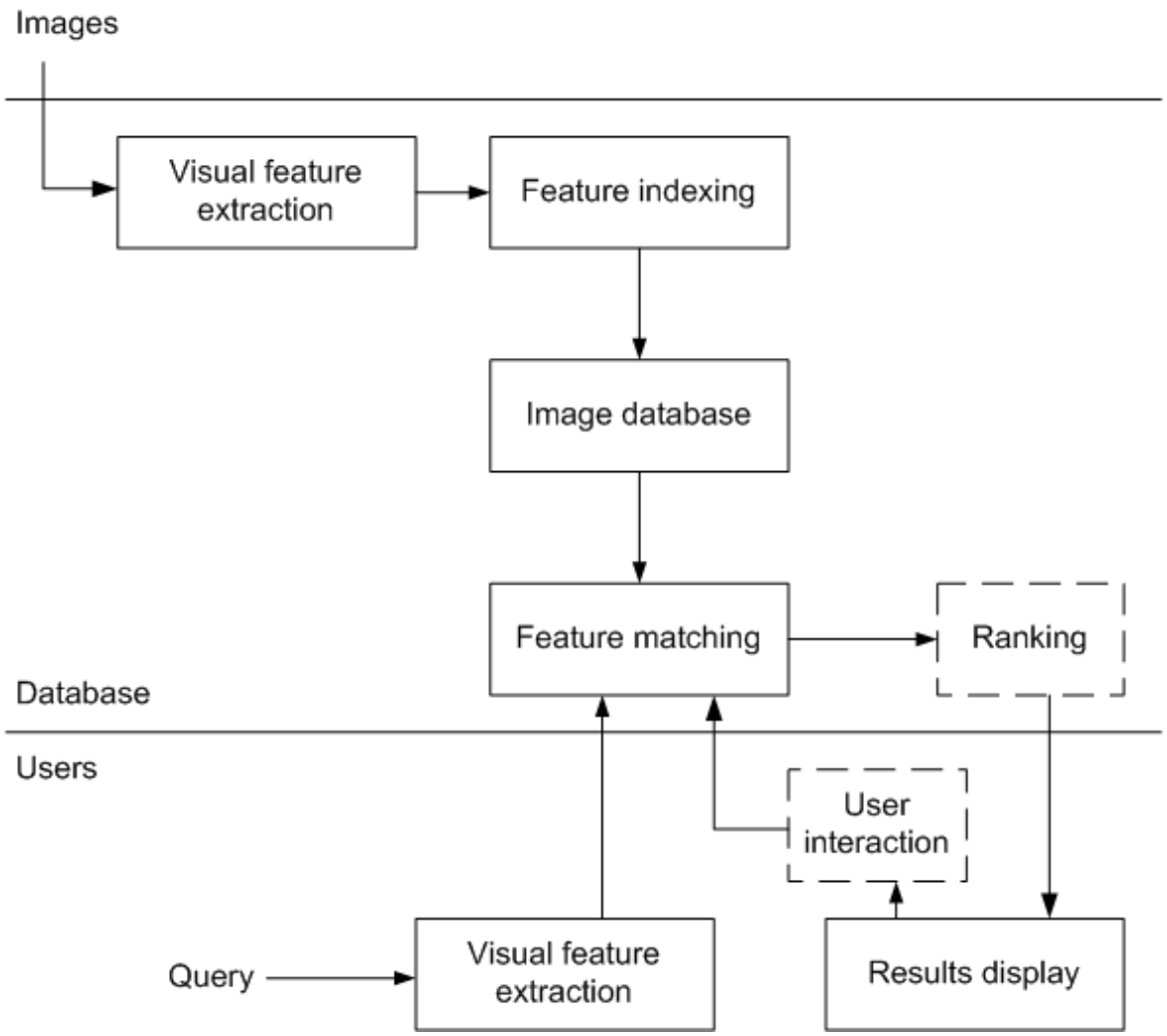
- The semantic gap characterises the differences between the two descriptors
  - low-level data (metadata)
  - high-level data (semantic meaning)
- Leads to an inability of automatic understanding.



Content Based Retrieval

- Low-level images features include:
  - Colour features:
    - Colour histograms
    - Dominant colours
    - Colour distribution
    - Colour correlogram
  - Texture features:
    - Tamura features
    - Edge histogram
    - Local features
  - Shape features
  - ……

Content Based Retrieval



Colour Representation

- Colour is very powerful in description
- Provides easy extraction from natural images
  - Considerable changes in variance, e.g. illumination
  - Orientation of the surface
  - Viewing geometry of the camera
- Light of different wavelengths produce different colour sensations
  - E.g. in broad regions (violet, blue, green, yellow, orange and red).

Colour Representation

What soft drink?

Colour Representation

What soft drink?

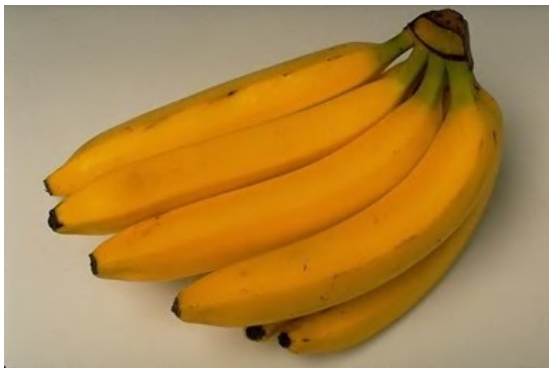
Colour Representation

What soft drink?

Which fruit?

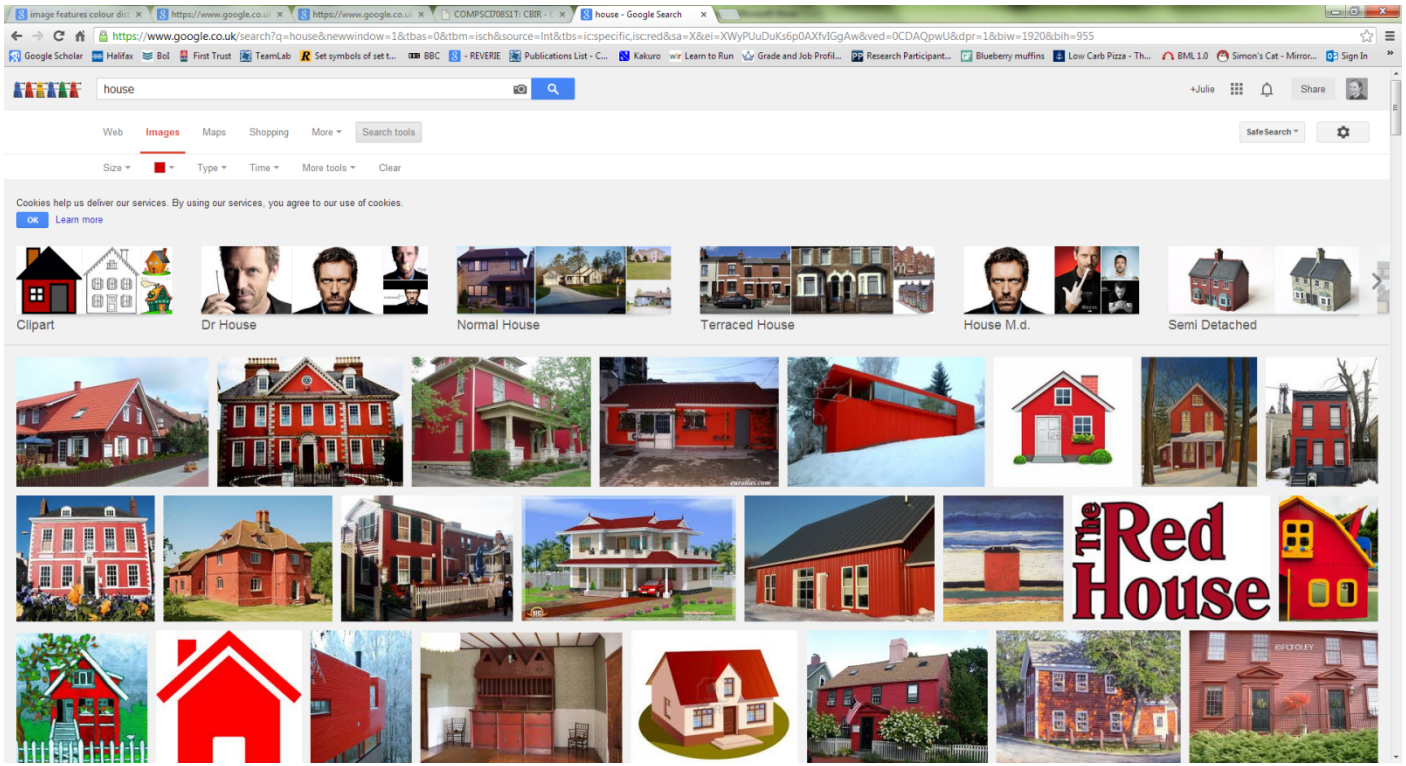
Colour Representation

What soft drink?



Which fruit?

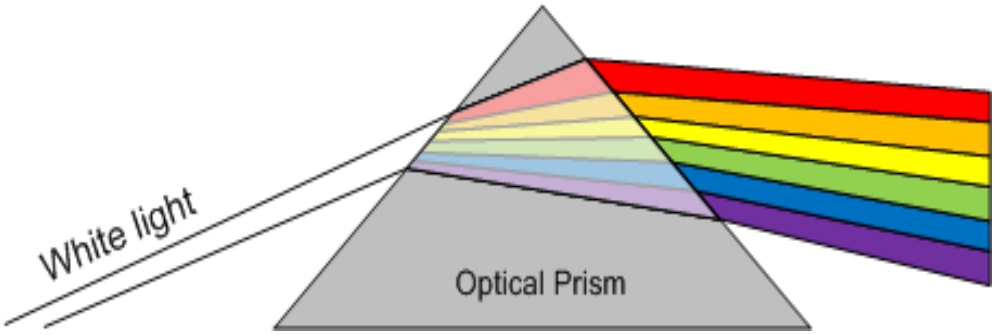
Colour Representation



Google image search filtered by colour

Colour Representation

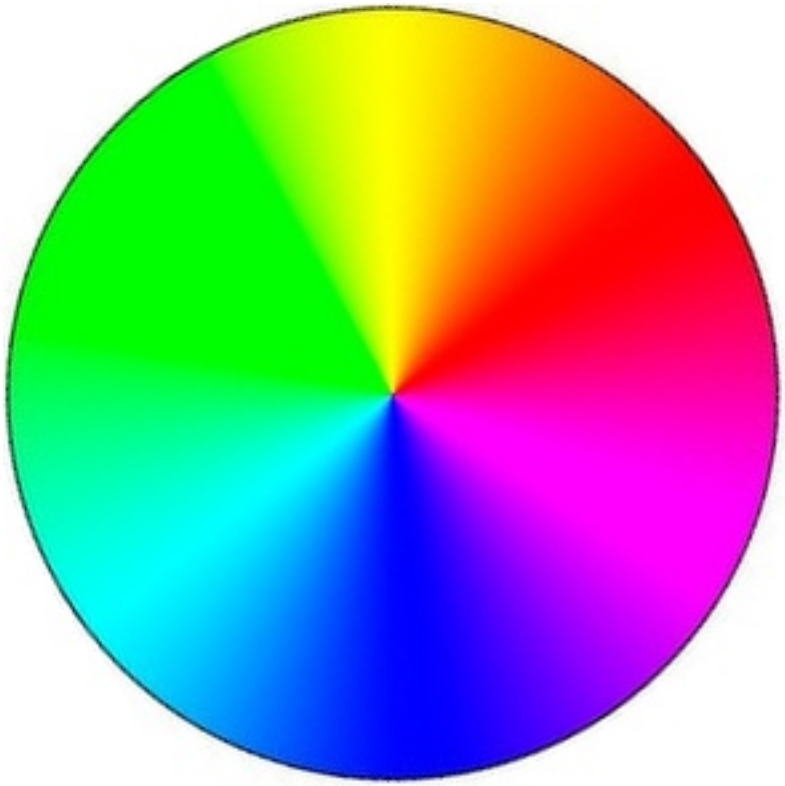
- Colours that humans perceive in an object
  - Determined by the nature of the light reflected from the object



  - Visible light: electromagnetic radiation
  - Spectrum wavelength range - 400 to 780 nm.
  - Red, green and blue - additive primary colours
    - Any colour can be specified giving weights of these

Colour Representation

- RGB colour space
  - Most important representation
  - (r-value, g-value, b-value)
    - Value: percentage of pure light
    - (100%, 0%, 0%) - pure saturated primary red
    - (50%, 0%, 0%) - darker red
    - (0%, 0%, 0%) - black
    - (100%, 100%, 100%) - white
  - Cartesian coordinate system defined to measure each colour with a vector



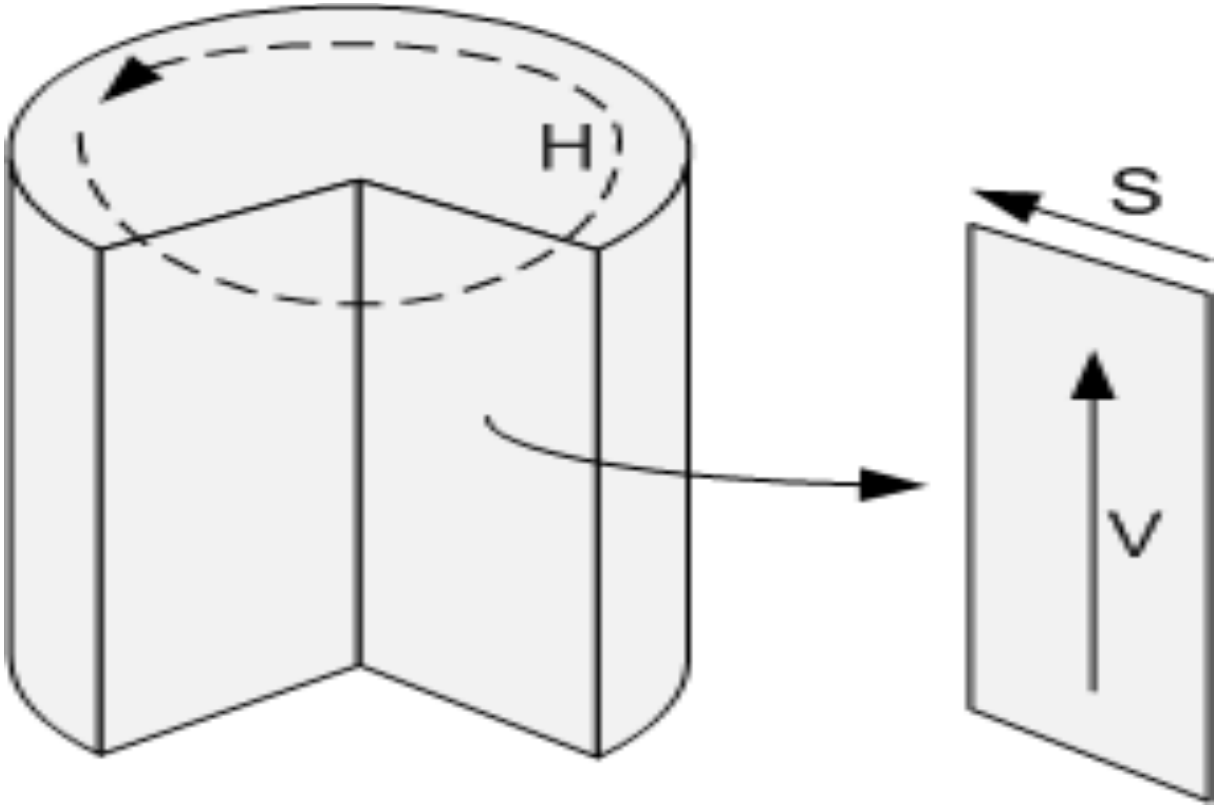
Colour Representation

- RGB colour space cont'd.
  - Value range for each primitive: 0 to 255 (8-bit byte)
    - RGB colour can be represented by 24 bits / 3 bytes
  - For practical systems, RGB colour can hold different bit depths
    - 24-bit full RGB colour space
    - 15-bit 5-bit for R, 5-bit for G and 5-bit for B
    - 12-bit 4 bit for R, 4-bit for G and 4-bit for B

Colour Representation

- HSV colour space
  - Physical properties of colour radiation
    - Hue
      - Attribute of a visual sensation according to which area appears similar to one of the perceived colours, such as red, yellow, green and blue.
      - Range: 0-360˚
    - Saturation
      - Colourfulness of an area judged in proportion to its brightness
      - Pure colour has saturation 100%
      - White colour has saturation 0%
    - Value: either Intensity of Brightness of pixel

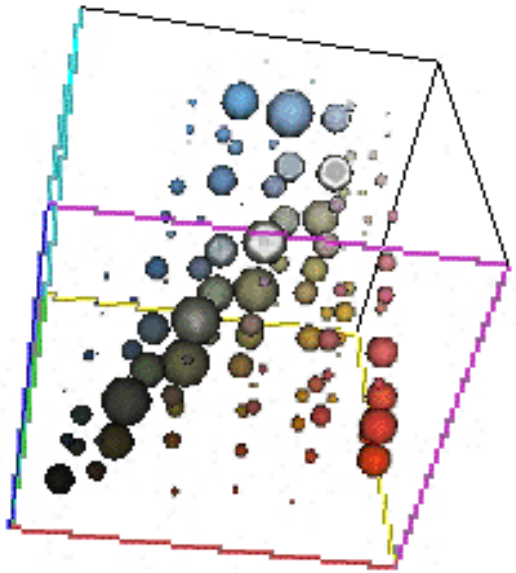
Colour Representation



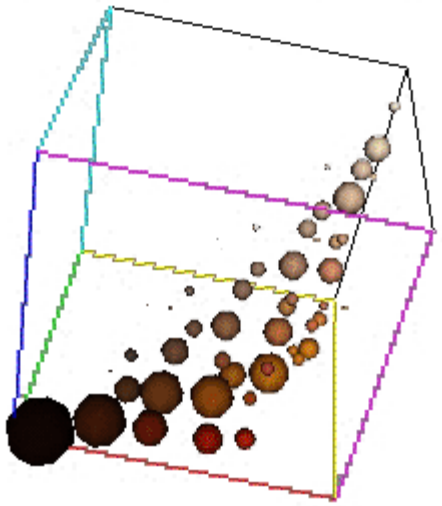
Content Based Retrieval - Dominant Colour

- Representative colours are selected from the image
- Colours present in a given image/region obtained after segmentation are clustered to retain only a small number of colours and the percentages of those colours
- Dissimilarity function:
  - Difference in amount (percentage)
  - Difference between colours

Content Based Retrieval - Dominant Colour



Content Based Retrieval - Dominant Colour



Content Based Retrieval - Dominant Colour

- Advantages:
  - Small feature vectors
  - Easily understandable and intuitive
  - Invariant to rotation, translation and reflection
- Disadvantages:
  - Similarity of colour pairs is no trivial problem
  - Colours might not represent semantics

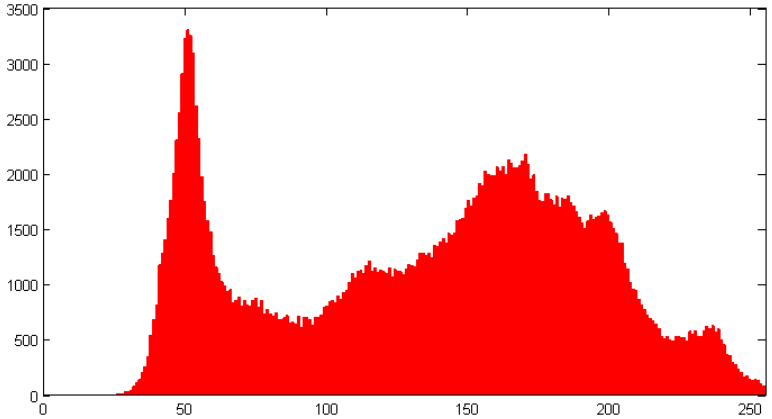
Content Based Retrieval - Colour Histogram

- Characterises the distributions of colours in an image both globally and locally
  - Count how often each colour is used in the image – partition the colour space
- Each pixel described by three colour components
  - Histogram for each components describes the distribution of pixels for that component colour in a quantitative level
    - A quantized colour bin
    - The levels can be 265, 64, 32, 16, 8, 4, 1 (8-bit byte).

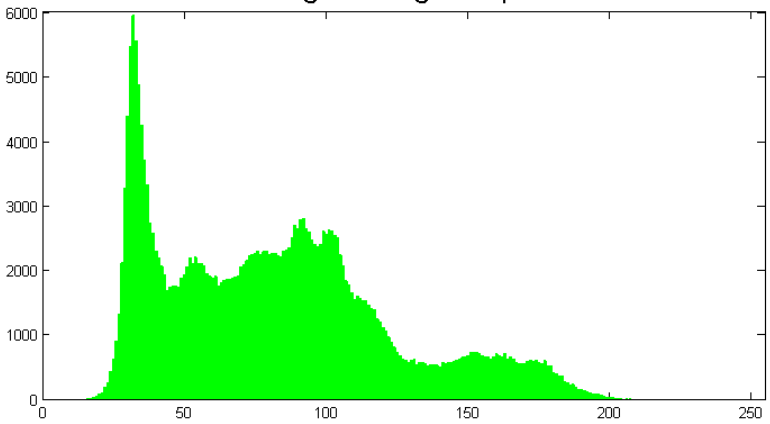
Content Based Retrieval - Colour Histogram

Original Color Image

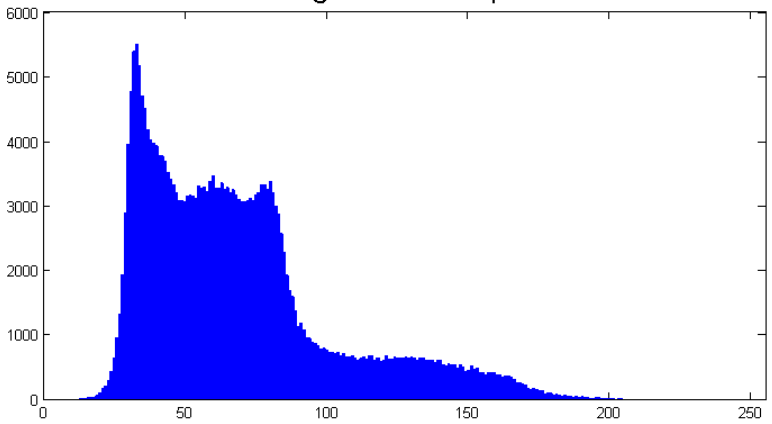
Histogram of red plane



Histogram of green plane



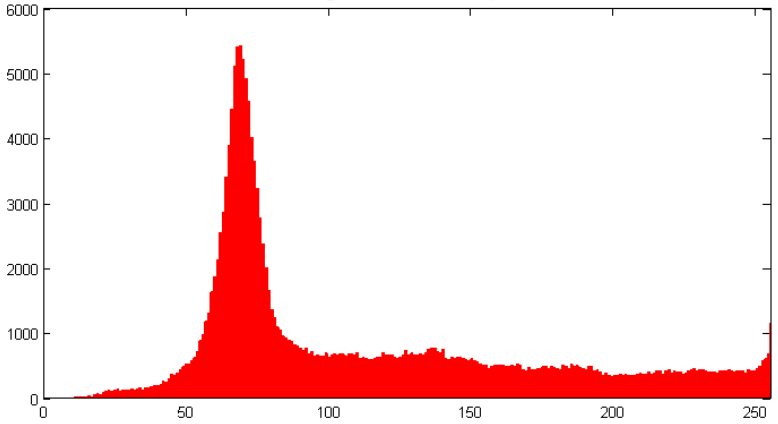
Histogram of blue plane



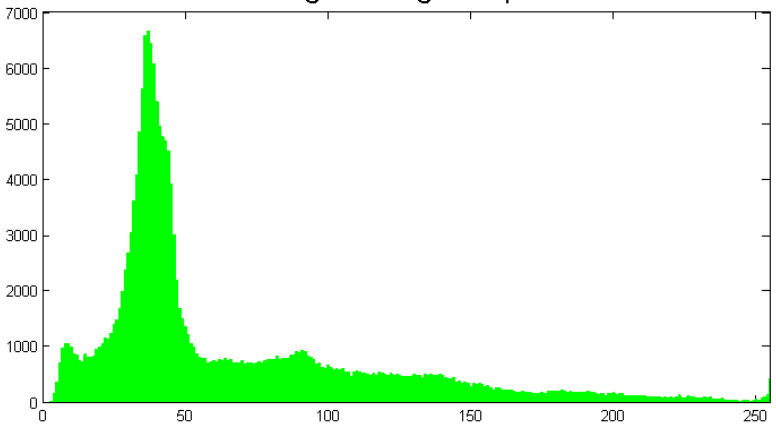
Content Based Retrieval - Colour Histogram

Original Color Image

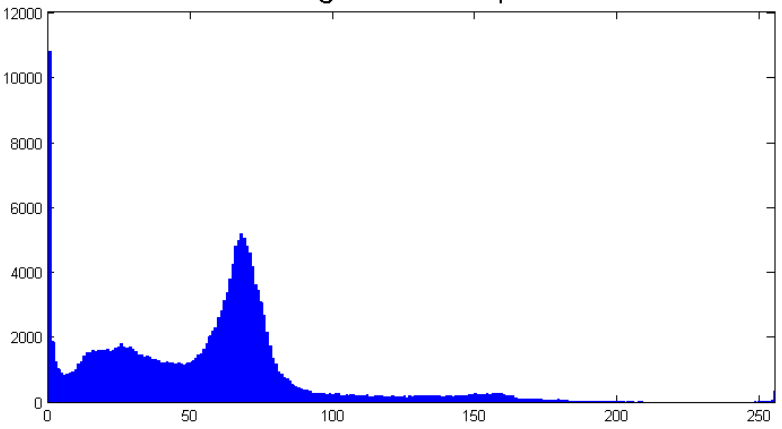
Histogram of red plane



Histogram of green plane

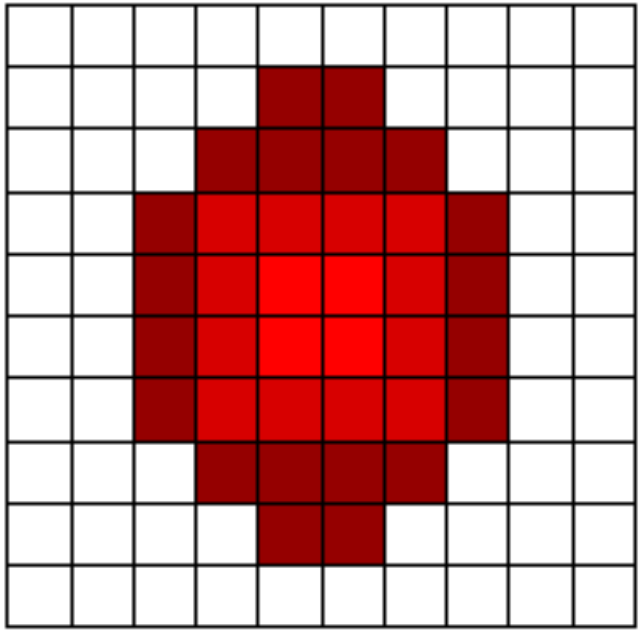


Histogram of blue plane



Content Based Retrieval - Colour Histogram

- Algorithm:
  1. Allocate int array h with dim = # of colours
  2. Visit next pixel -> it has a colour with index i
  3. Increment h[i]
  4. IF pixels left THEN goto line 2



4 colours, 10×10 pixels
Histogram: [4, 12, 20, 64]

Content Based Retrieval - Colour Histogram

- Strategies:
  - Quantize (compress) if too many colours
  - Normalize histogram (different image sizes)
  - Weight colours according to use case
  - Use colour space according to domain
- Distance/similarity
  - Assume all images have the same colours
  - Use the suitable distance metrics

Content Based Retrieval - Colour Histogram

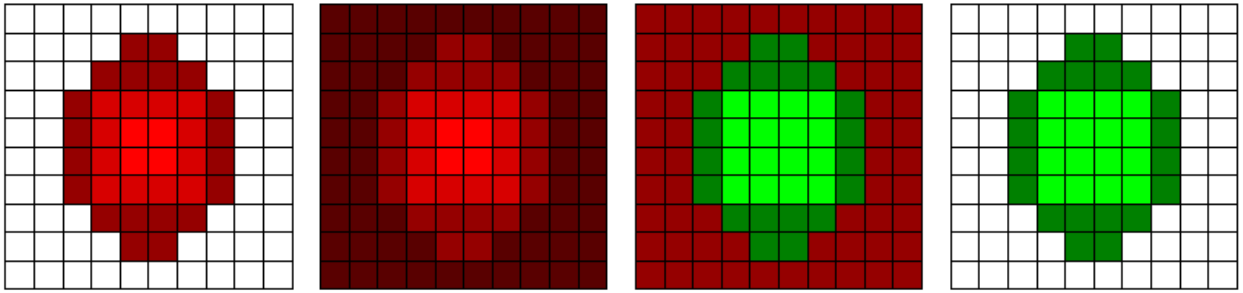
- Advantages:
  - Easy to compute, not dependent on pixel order
  - Matches human perception quite well
  - Quantization allows for scaling size of histogram
  - Invariant to rotation, translation and reflection

Content Based Retrieval - Colour Histogram

- Disadvantages:
  - Distribution of colours not taken into account
  - Colours might not represent semantics
  - Find quantization fitting to domain or perception
  - Image scaling might be a problem

Content Based Retrieval - Colour Histogram

- Dissimilarity example using Manhattan distance:



- 4 images, 7 colours
- 1: [4, 12, 20, 0, 64, 0, 0]
- 2: [4, 12, 20, 64, 0, 0, 0]
- 3: [0, 0, 64, 0, 0, 16, 20]
- 4: [0, 0, 0, 0, 64, 16, 20]

$$d_1(p,q) = \|p-q\| = \sum_{i=1}^{n} |p_i - q_i|$$

Content Based Retrieval - Colour Histogram

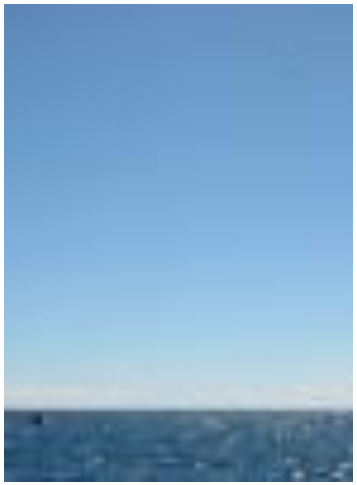
- d(1,2) = 128
- d(1,3) = 160
- d(1,4) = 72
- Why is the distance between images 1 and 4 the least dissimilar (most similar)?

Content Based Retrieval - Colour Histogram

- If more bins are defined in histogram calculation
  - represents more discrimination power
- Increases computation cost if combined colour bin histogram systems are used
  - R×G×B = 256×256×256 = 16777216 bins!
  - May generate colour indexes for image database inappropriately
  - Might not help image retrieval performance

Content based Retrieval – other techniques

- Colour Layout
  - Divides image into grid
  - Searches for similar colours in equivalent grid position
  - e.g blue sky



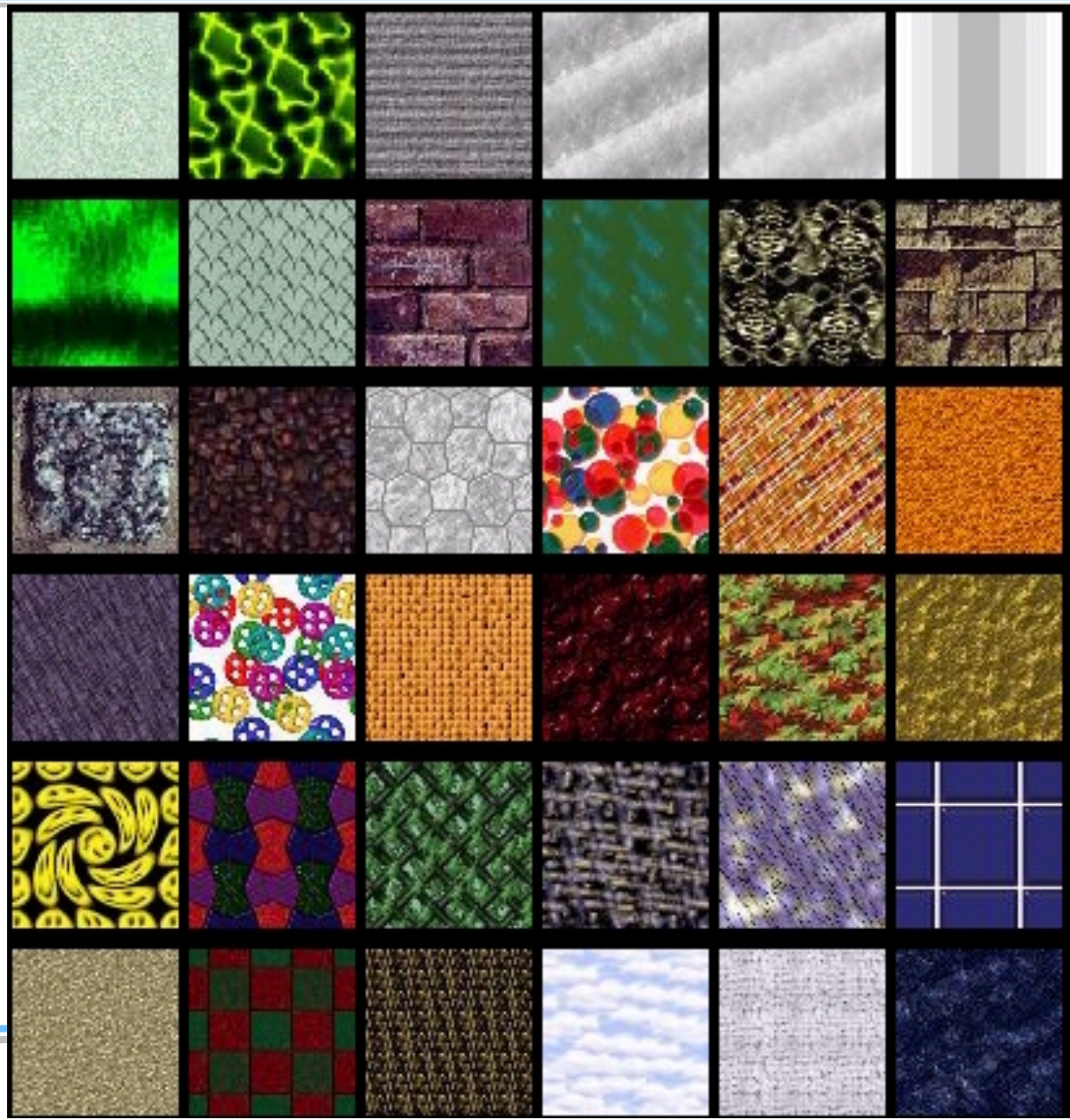
Content based Retrieval – Texture

- Uses orientation and spacing of edges
- Histogram of edge features can be used
- Also texture layout (as colour layout)

Content based Retrieval – Texture

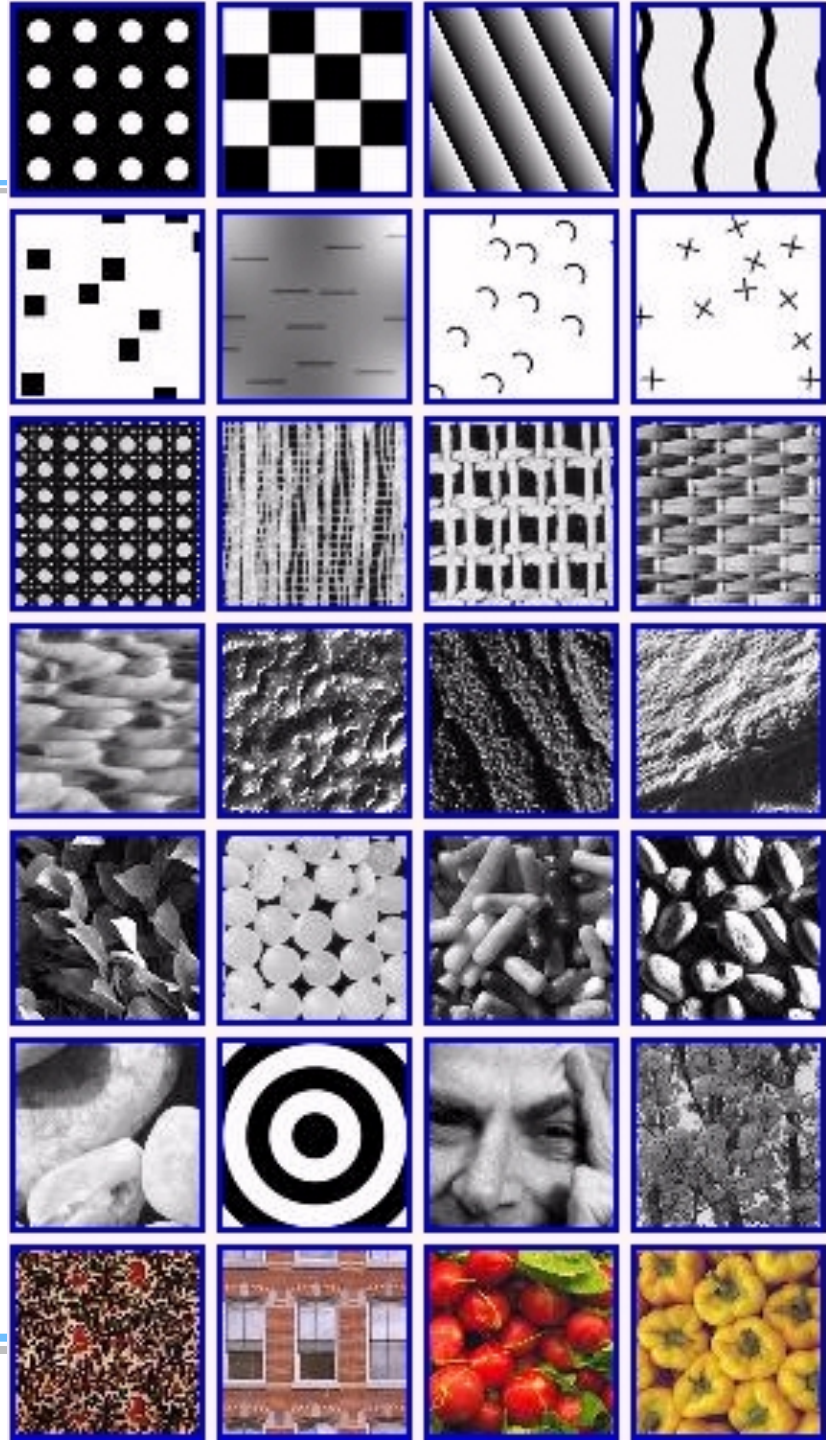
A Technical Definition of texture:

Texture is a broad term used in pattern recognition to identify image patches (of any size) that are characterized by differences in brightness.



Content based Retrieval – Texture

- Techniques to extract meaningful texture descriptors from image are many, based on different models and assumptions.
- An effective representation of textures can be based on statistical and structural properties of brightness patterns.



Content based Retrieval – Texture

- Texture
  - Textures may be described according to their spatial, frequency or perceptual properties.
    - Periodicity,
    - coarseness,
    - preferred direction,
    - degree of complexity
    - ...
  - Feature spaces based on these attributes are particularly interesting for image retrieval by texture similarity.

Content based Retrieval – Texture

- Space based models
  - **Auto-correlation function** A texture can be represented taking into account the spatial size of grey-level primitives.
    - Fine textures have a small size of their grey-level primitives.
    - Coarse textures have a large size.
  - **Co – occurrence matrix** A different way of measuring textures is by taking into account the spatial arrangement of grey-level primitives.

Content based Retrieval – Texture

- Calculate texture

$$\sum_{i,j} P(i,j)^2$$ energy

$$\sum_{i,j} P(i,j) \log P(i,j)$$ entropy

$$\sum_{i,j} (i-j)^2 P(i,j)$$ contrast

Segmentation issues

- Considered as a difficult problem
- Not reliable
- Segments regions, but not objects
- Different requirements from segmentation:
  - Shape extraction: High Accuracy required
  - Layout features: Coarse segmentation may be enough

Content based Retrieval – Shape

- Shape
  - Color and texture are both global attributes of an image.
  - Shape refers to a specific region of an image.
  - Shape goes one step further than color and texture in that it requires some kind of region identification process to precede the shape similarity measure.
  - Segmentation is still a crucial problem to be solved.
  - Shape matching will be discussed here.

Content based Retrieval – Shape

- 2-D shape recognition is an important aspect of image analysis.
- Comparing shapes can be accomplished in several ways – structuring elements, region adjacency graphs etc.
- They tend to expensive in terms of time.
- In CBIR we need the shape matching to be fast.
- The matching should also be size, rotational and translation invariant.

Content based Retrieval – Shape

- Shape histogram
  - Histogram distance simply an extension from color and texture.
  - The biggest challenge is to define the variable on which the histogram is defined.
  - Histogram matching types:
    - **Projection matching** - using horizontal and vertical projections of the shape in a binary image.
    - **Boundary matching** - the extraction and representation of the boundaries of the query shape and image shape.

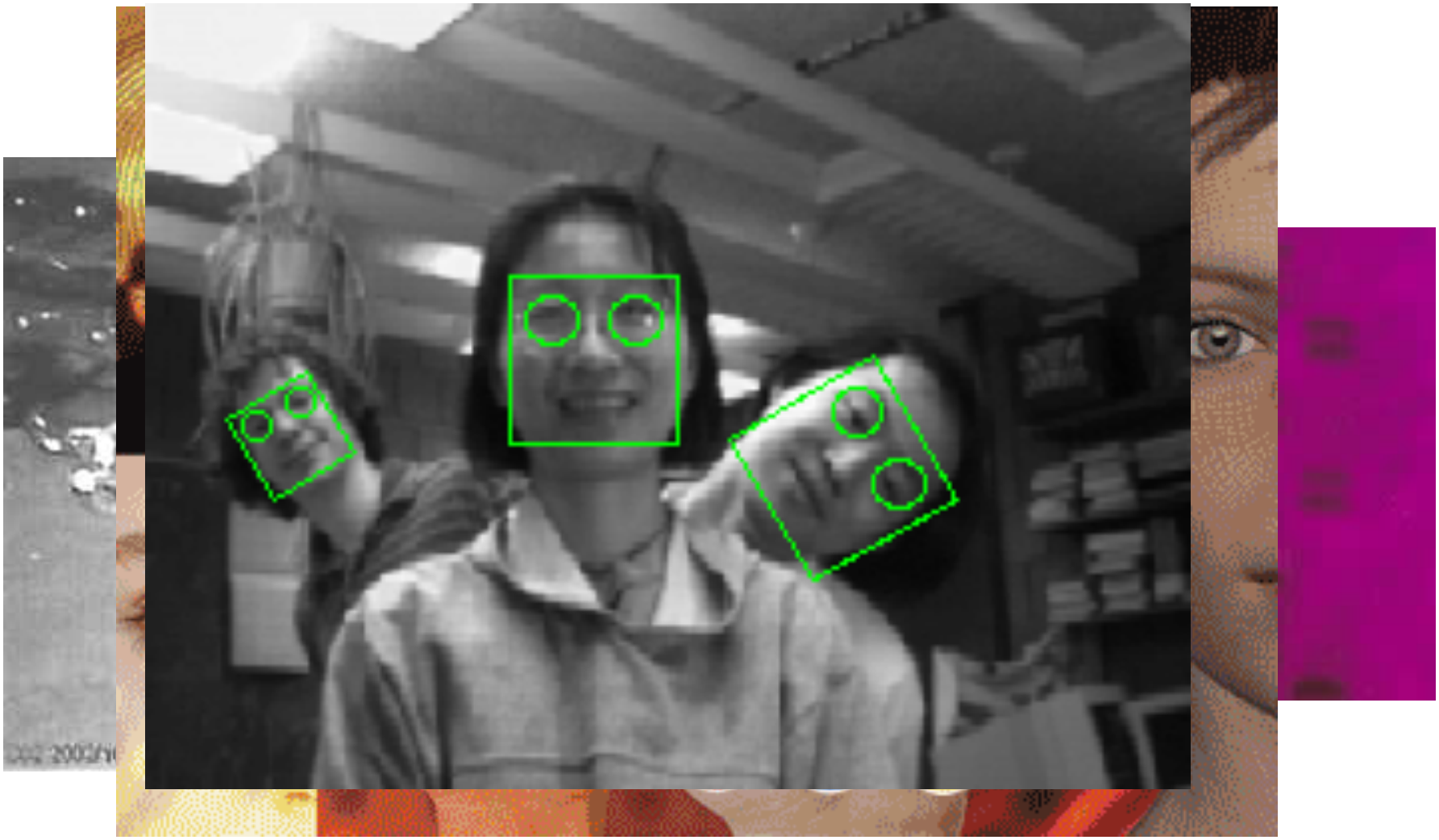
Content based Retrieval – Sketch

- Sketch matching systems allow the user to input a rough sketch of the major edges in an image and look for matching images.
  - In the ART MUSEUM system, the DB consists of color images of famous paintings. The following preprocessing step are performed to get an *abstract image* of all the images in the DB.
- When the user enters a rough sketch, it is also converted to the normalized size, binarized, thinned and shrunk, resulting in a *linear sketch.*
- Now the linear sketch must be matched to the abstract image.
- The matching algorithm can (gridded) correlation-based.

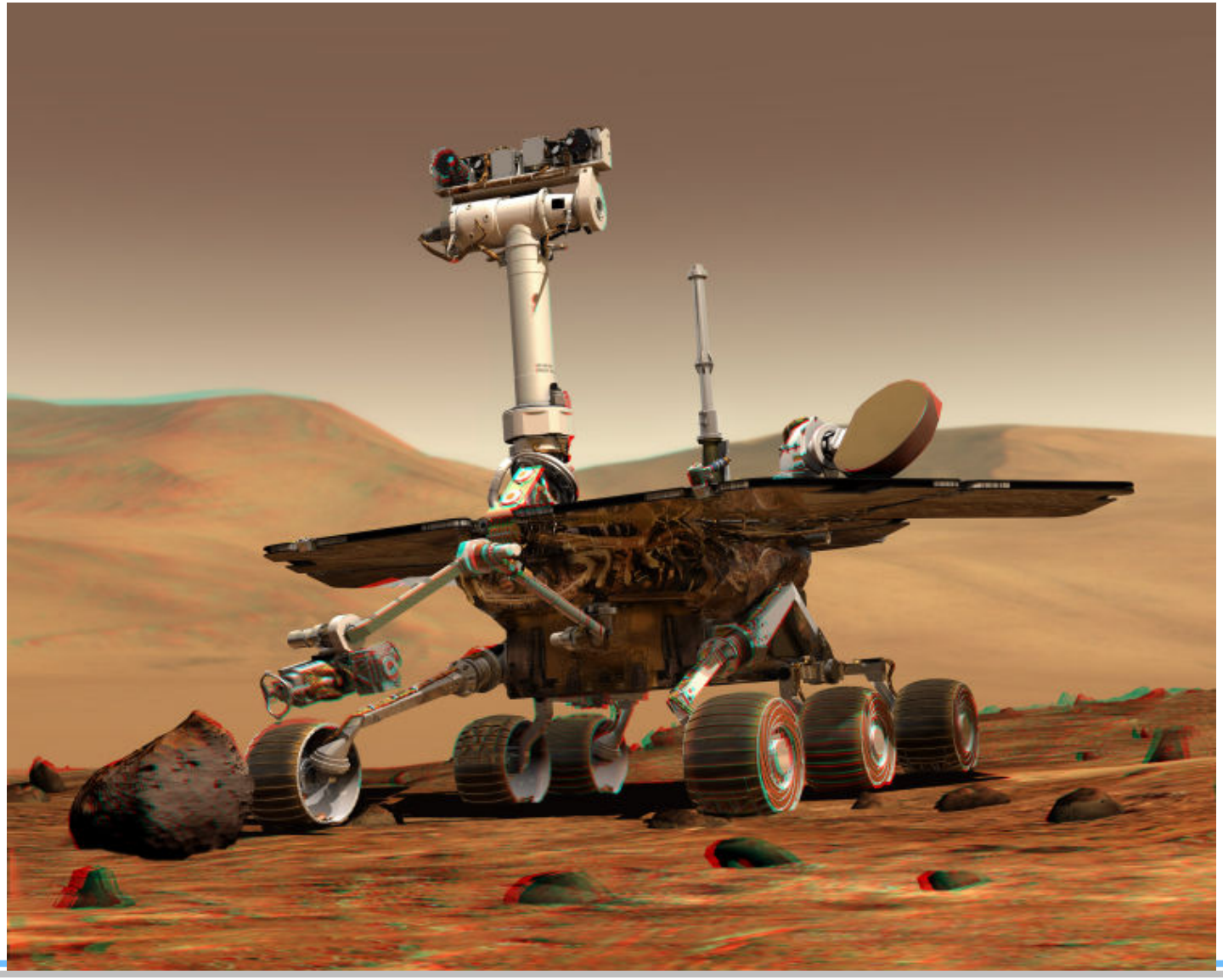
Video Retrieval

- Three major processes to prepare a video for retrieval:
  - **video segmentation** - divides the video into a number of segments by detecting the camera breaks.
  - **index extraction** - manual indexing, image analysis and computer vision and object recognition
  - **keyframe extraction** - select representative image frames from each video segment to represent the segment. These keyframes may be used for browsing and for presentation.
- From another perspective, video retrieval could be considered simpler than image retrieval
  - since video reveals its objects more easily as the points corresponding to one object move together.
- In addition, video has a linear timeline, as important to the narrative structure of video as it is in text.

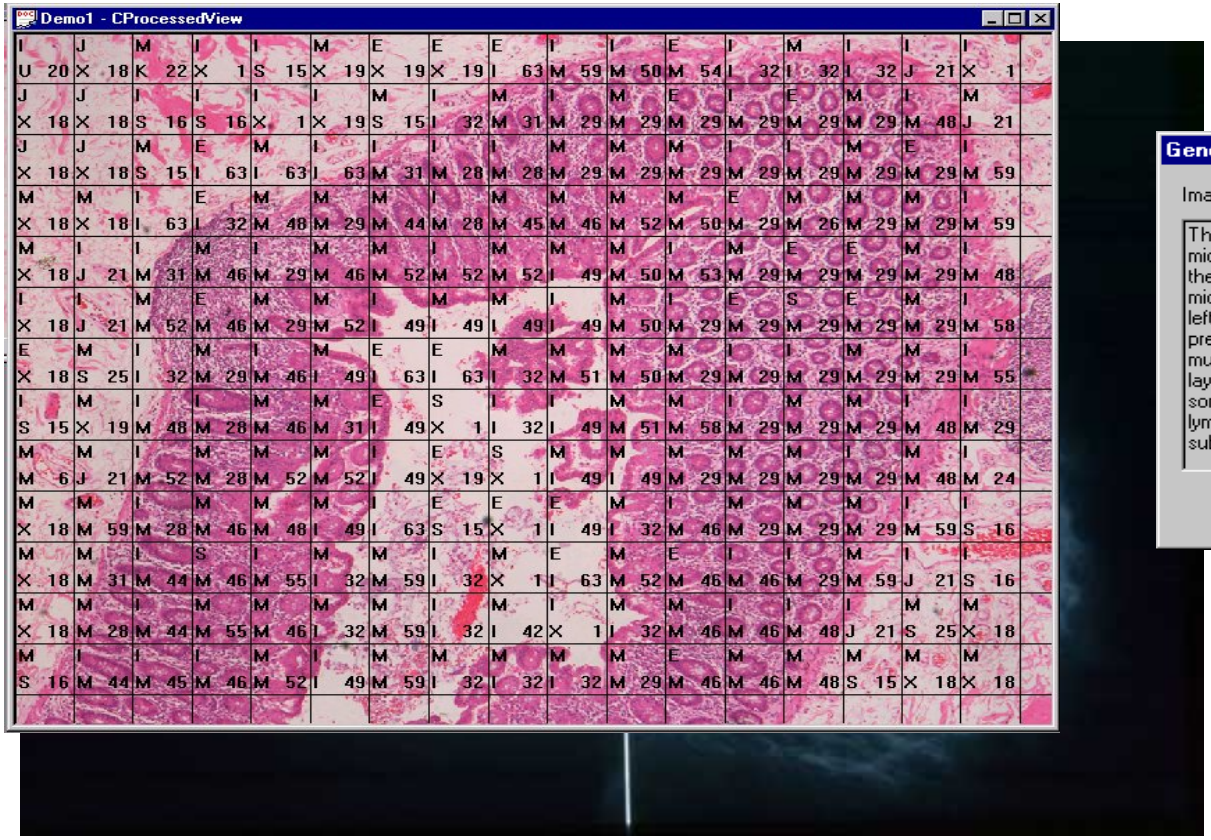
Applications: Police surveillance, genome research, biometrics, security

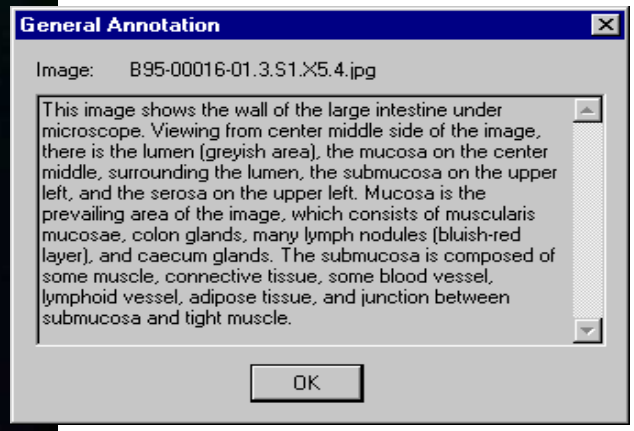


Applications: Remote sensing, astronomy, GIS, Earth/Planetary observation, monitoring, exploration



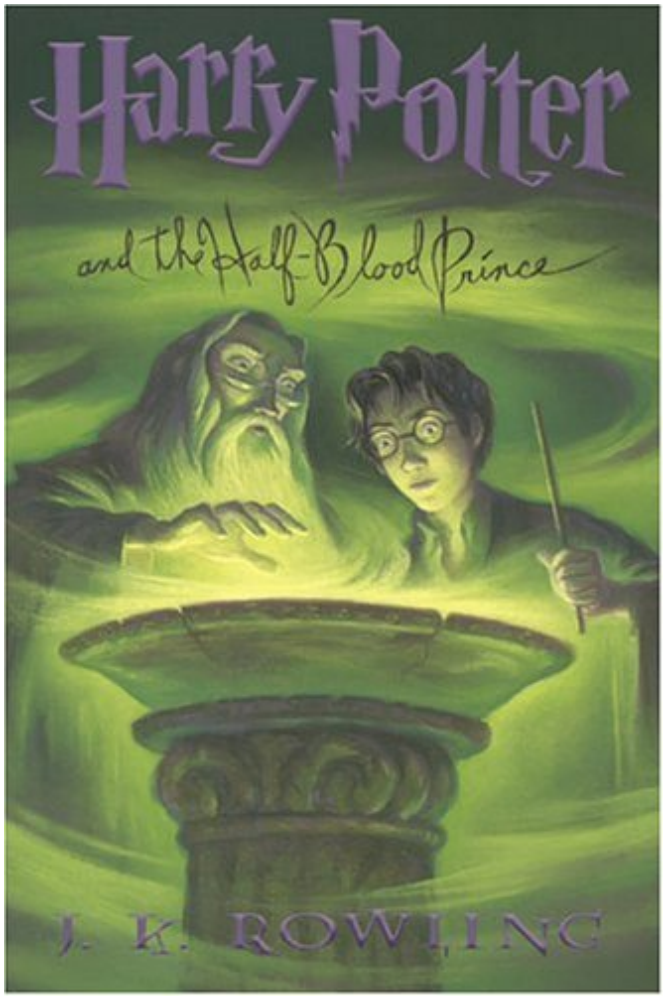
Applications: Medical imaging





Applications: entertainment

film and TV, DTV, news and sport, creative media, art, museums



Related areas

- Big data science
- Artificial intelligence (deep learning)
- Social media (user generated content)
- Security
- Healthcare and wellbeing
- News
- Broadcasting
- Finance and law
- …. Many others