1 What is multimedia information retrieval?
   1.1 Information retrieval
   1.2 Multimedia
   1.3 Semantic Gap?
   1.4 Challenges of automated multimedia indexing
2 Basic multimedia search technologies
   2.1 Meta-data driven retrieval
   2.2 Piggy-back text retrieval
   2.3 Automated annotation
   2.4 Fingerprinting
   2.5 Content-based retrieval
   2.6 Implementation Issues
3 Evaluation of MIR Systems
4 Added value
1 What is multimedia information retrieval?
   1.1 Information retrieval
   1.2 Multimedia
   1.3 Semantic Gap?
   1.4 Challenges of automated multimedia indexing

2 Basic multimedia search technologies
   2.1 Meta-data driven retrieval
   2.2 Piggy-back text retrieval
   2.3 Automated annotation
   2.4 Fingerprinting
   2.5 Content-based retrieval
   2.6 Implementation Issues

3 Evaluation of MIR Systems

4 Added value
Why content-based?

Actually, what is content-based search?

Is human thinking content-based?

Metadata annotation (text) is good but
Features and distances

Feature space

X
X
X
X
O
Architecture

- information need
- query [example]
- database index
- similarity ranking
- result presentation

relevance feedback
Features

Visual
  Colour, texture, shape, edge detection, SIFT/SURF

Audio

Temporal

How to describe the features?
  For people
  For computers
Histogram

1: 0 – 31
2: 32 – 63
3: 64 – 95
4: 96 – 127
5: 128 – 159
6: 160 – 191
7: 192 – 223
8: 224 – 255
phenomenon of human perception
three-dimensional (RGB/CMY/HSB)
spectral colour: pure light of one wavelength

- blue
- cyan
- green
- yellow
- red

spectral colours: wavelength (nm)
Colour histogram
Exercise

Sketch a 3D colour histogram for

- **black**: RGB (0, 0, 0)
- **red**: RGB (255, 0, 0)
- **green**: RGB (0, 255, 0)
- **blue**: RGB (0, 0, 255)
- **cyan**: RGB (0, 255, 255)
- **magenta**: RGB (255, 0, 255)
- **yellow**: RGB (255, 255, 0)
- **white**: RGB (255, 255, 255)
Solution
Other Colour Spaces

HSV, HSL, CIELAB/CIELUV
HSB colour model

hue (0°-360°) = spectral colour

saturation (0% - 100%) = spectral purity

brightness (0% - 100%) = energy or luminance

chromaticity = hue + saturation
HSB colour model
disadvantage: hue coordinate is not continuous

0 and 360 degrees have the same meaning
but there is a huge difference in terms of numeric distance

example:
red = (0,100%,50%) = (360,100%,50%)

advantage: it is more natural to describe colour changes “brighter blue”, “purer magenta”, etc
Texture

- coarseness
- contrast
- directionality
Texture histograms

Coarseness  contrast  Directionality

Gabor filter

Query

[with Howarth, CLEF 2004]
shape = class of geometric objects invariant under translation
scale (changes keeping the aspect ratio)
rotations
information preserving description (for compression)
non-information preserving (for retrieval)
boundary based (ignore interior)
region based (boundary + interior)
Shape Analysis

- boundary based
  - perimeter & area
  - corner points
  - circularity
  - chain codes
- region based (considering interior and holes, ...)
  - not covered here
Perimeter and area

Parameterised curve \( x(t), y(t) \)

Boundary pixel count vs count pixels in area
Circularity

\[ T = 4\pi \frac{A}{p^2} \]

A = area, P = perimeter
T is 1 for a circle
T is smaller than 1 for all other shapes
circularity is aka compactness
Convexity

ratio of perimeter of convex hull and the original curve
1 for convex shapes, less than 1 otherwise
Sound
Audio Features

• **Spectrogram**
  – graph of frequencies/energy/time
• **tempo, pitch, mode**
Histograms

Condensed
Content-based
Real-valued vector
Summarising
Sparseness
Statistical moments
Histograms

Feature vectors → histograms

145 173 201 253 245 245
153 151 213 251 247 247
181 159 225 255 255 255
165 149 173 141 93 97
167 185 157 79 109 97
121 187 161 97 117 115

1: 0 – 31 5: 128 – 159
2: 32 – 63 6: 160 – 191
3: 64 – 95 7: 192 – 223
4: 96 – 127 8: 224 – 255
Central moments

Simple statistics

\( \mu \)  Mean

\( p_2 \)  Variance (squared standard deviation)

\( p_3 \)  3\(^{rd}\) central moment (skewness)

\[
\overline{p}_n = \frac{1}{wh} \sum_{i=1}^{w} \sum_{j=1}^{h} (p(i, j) - \mu)^n
\]

where \( w \) is image width and \( h \) is image height
Moment features

\((\mu, \sqrt{p_2}, \text{sign}(p_3) \sqrt[3]{|p_3|})\)
Moment features

\[
\left( \mu_r, \sqrt{r_2}, \text{sign}(r_3)^3 \sqrt{|r_3|}, \\
\mu_g, \sqrt{g_2}, \text{sign}(g_3)^3 \sqrt{|g_3|}, \\
\mu_b, \sqrt{b_2}, \text{sign}(b_3)^3 \sqrt{|b_3|} \right)
\]
Global vs local

Global histogram also matches polar bears, marble floors, ...
Localisation

64% centre

36% border
Tiled Histograms
Segmentation

foreground

background
Points of interest

Many PoI, i.e., many feature vectors
Quantised feature vectors ≈ words
Bag of word model ≈ text retrieval
“Bag of Words”
Exercise

- http://192.168.1.5:8080/uBase
- Find an example query image that works well
- Find an example query image that doesn't work
- Try changing the features weights, can you improve the results?
• Anticipation Trailer
• Segmentation Equations
gradual transition detection (eg, fade)
accumulate distances
long-range comparison

audio cues
silence and/or speaker change

motion detection and analysis
camera motion, zoom, object motion
MPEG provides some motion vectors
Vlad Tanasescu: Anticipation, SCiFi trailer]
At time $t$ define distance $d_n(t)$

- compare frames $t-n+i$ and $t+i$ ($i=0,...,n-1$)

- average their respective distances over $i$

Peak in $d_n(t)$ detected if

$$d_n(t) > \text{threshold and } d_n(t) > d_n(s) \text{ for all neighbouring } s$$

Shot = near-coincident peaks of $d_{16}$ and $d_8$
Features and distances

Feature space
Distances and similarities

assumes coding of MM objects as data vectors

distance measures
   Euclidean, Manhattan

correlation measures
   Cosine similarity measure
   histogram intersection for normalised histograms

\[ \text{sim}(h, q) = \sum_{i} \min(h_i, q_i) \]
\[ d_p(v, w) = \sqrt[p]{\sum_i |v_i - w_i|^p}, \quad p \geq 1 \]

\[ d_2(v, w) = \sqrt{\sum_i |v_i - w_i|^2} \]

\[ d_1(v, w) = \sqrt[1]{\sum_i |v_i - w_i|} = \sum_i |v_i - w_i| \]

\[ d_{\infty}(v, w) = \]
\[ d_p(v, w) = \sqrt[p]{\sum_i |v_i - w_i|^p}, \quad p \geq 1 \]

\[ d_2(v, w) = 2\sqrt{\sum_i |v_i - w_i|^2} \]

\[ d_1(v, w) = \sqrt{\sum_i |v_i - w_i|^1} = \sum_i |v_i - w_i| \]

\[ d_\infty(v, w) = \max_i |v_i - w_i| \]
What happens at $p < 1$?
Other distance measures

- Squared chord
- Earth Mover's Distance
- Chi squared distance
- Kullback-Leibler divergence (not a true distance)
- Ordinal distances (for string values)
Implementation

speed vs flexibility vs precision

Process:
1. best abstracted representation of your media
2. best method for calculating difference/similarity
3. implement efficiently, considering responsiveness and scalability
Sketch a block diagram showing how you would implement a multimedia information retrieval system for one of these scenarios:

1. Browsing wallpaper patterns in a home decorator store
2. Finding “interesting” photos in a personal collection of holiday snaps
3. Managing industrial design pattern templates for a manufacturing company

Think about:
- what types of features you might use
- what would the query be
- the user interface
For example

“Where is the big pineapple?”

Specific (“known item”)

“Family group photo taken last Christmas”

“The song I heard at the restaurant yesterday”

General

“Family vacation pics at Surfers – like this one”

“Music to go with my vacation photo slide show”
1 What is multimedia information retrieval?
   1.1 Information retrieval
   1.2 Multimedia
   1.3 Semantic Gap?
   1.4 Challenges of automated multimedia indexing

2 Basic multimedia search technologies
   2.1 Meta-data driven retrieval
   2.2 Piggy-back text retrieval
   2.3 Automated annotation
   2.4 Fingerprinting
   2.5 Content-based retrieval
   2.6 Implementation Issues

3 Evaluation of MIR Systems

4 Added value