KNOWLEDGE MEDIA INSTITUTE

KNOWLEDGE MEDIA INSTITUTE

KNOWLEDGE MEDIA INSTITUTE

Why content-based?

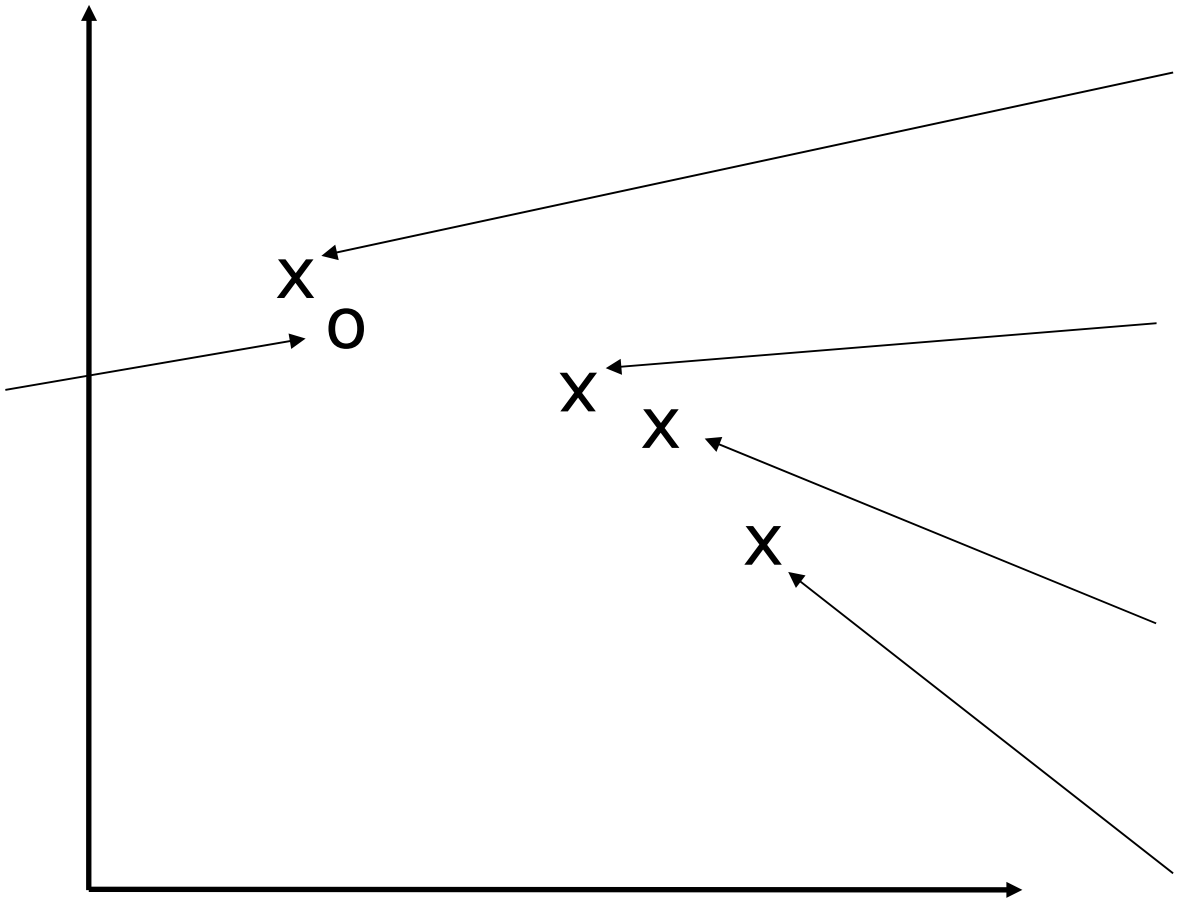
Actually, what is content-based search?

Is human thinking content-based?

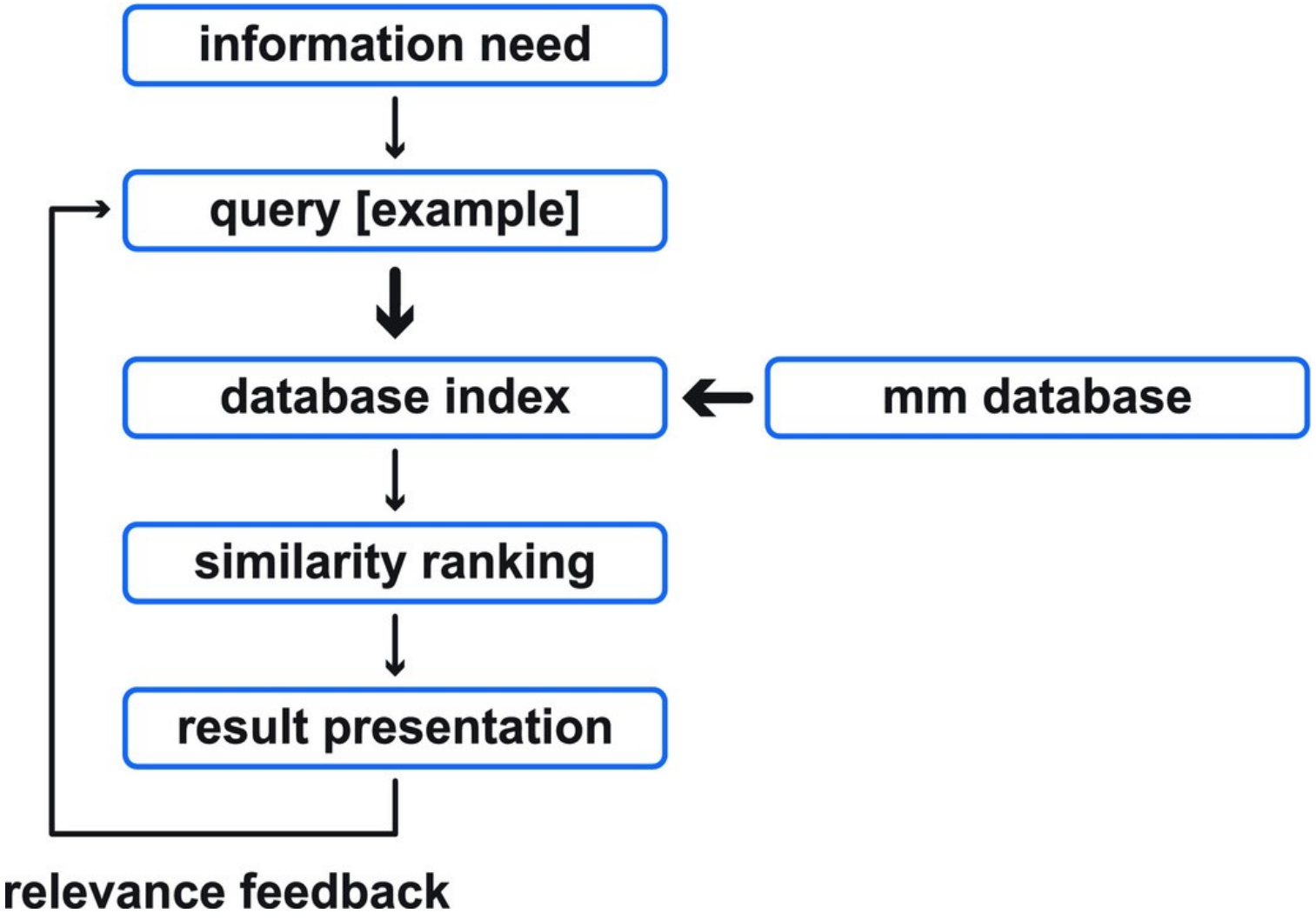
Metadata annotation (text) is good but

-
-
-
-

Features and distances



Architecture



KNOWLEDGE MEDIA INSTITUTE

Features

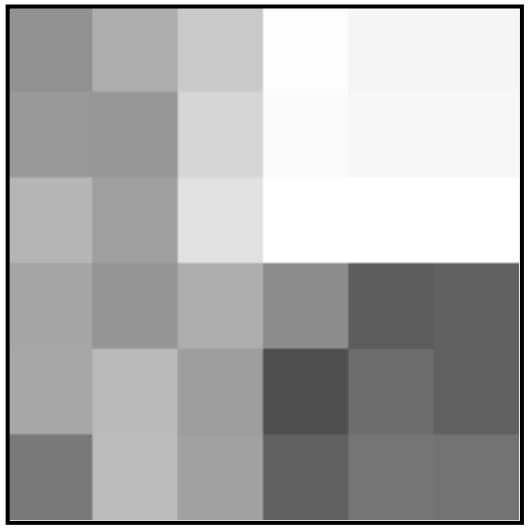
- Visual
  - Colour, texture, shape, edge detection, SIFT/SURF
- Audio
- Temporal
- How to describe the features?
  - For people
  - For computers

KNOWLEDGE MEDIA
KMi
INSTITUTE

Digital Images

KNOWLEDGE MEDIA INSTITUTE

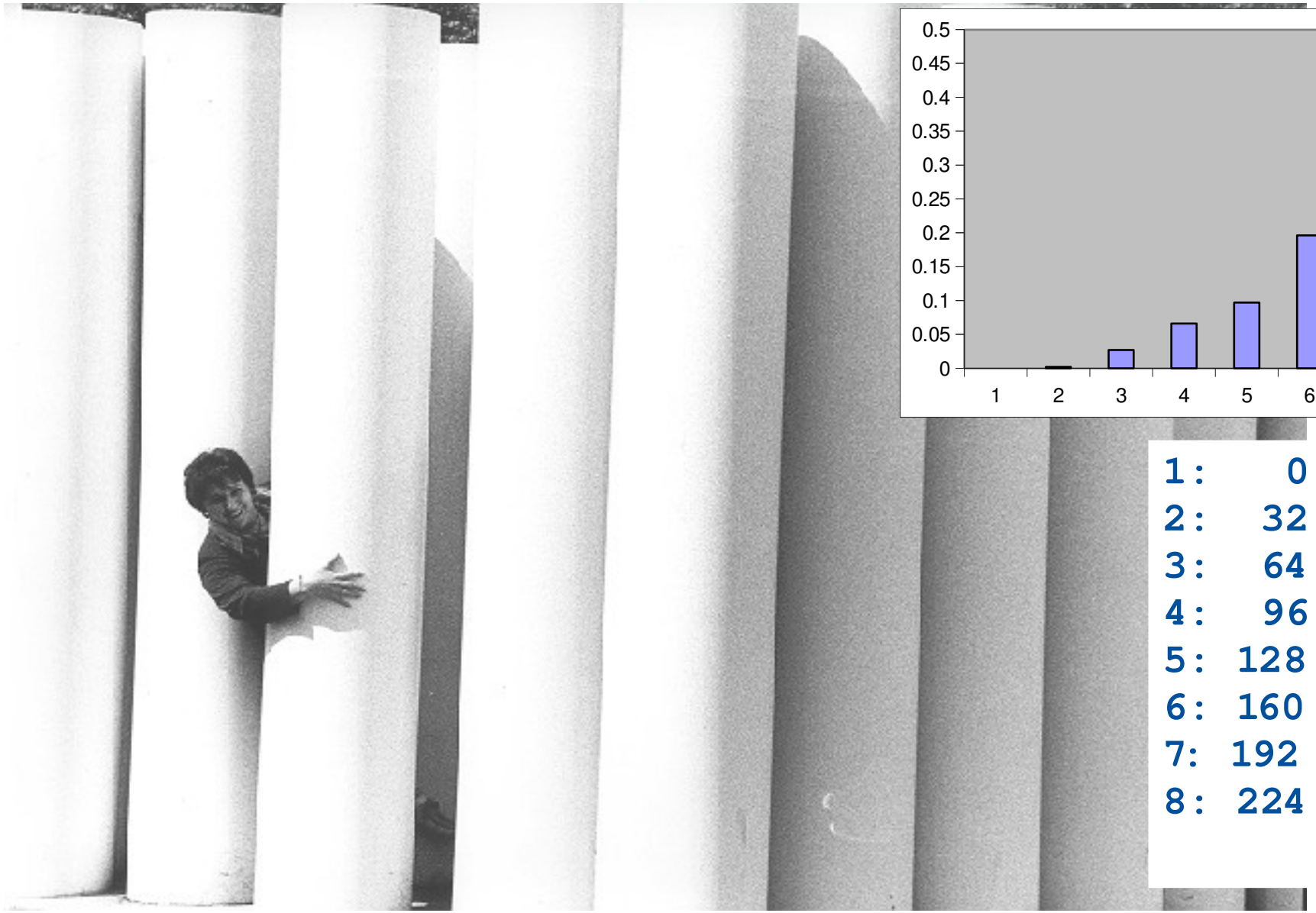
Content of an image



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

KNOWLEDGE MEDIA KMi INSTITUTE

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
```

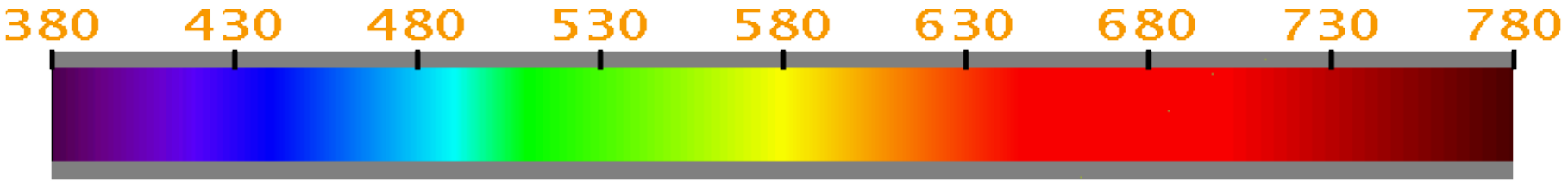
Colour

phenomenon of human perception

three-dimensional (RGB/CMY/HSB)

spectral colour: pure light of one wavelength

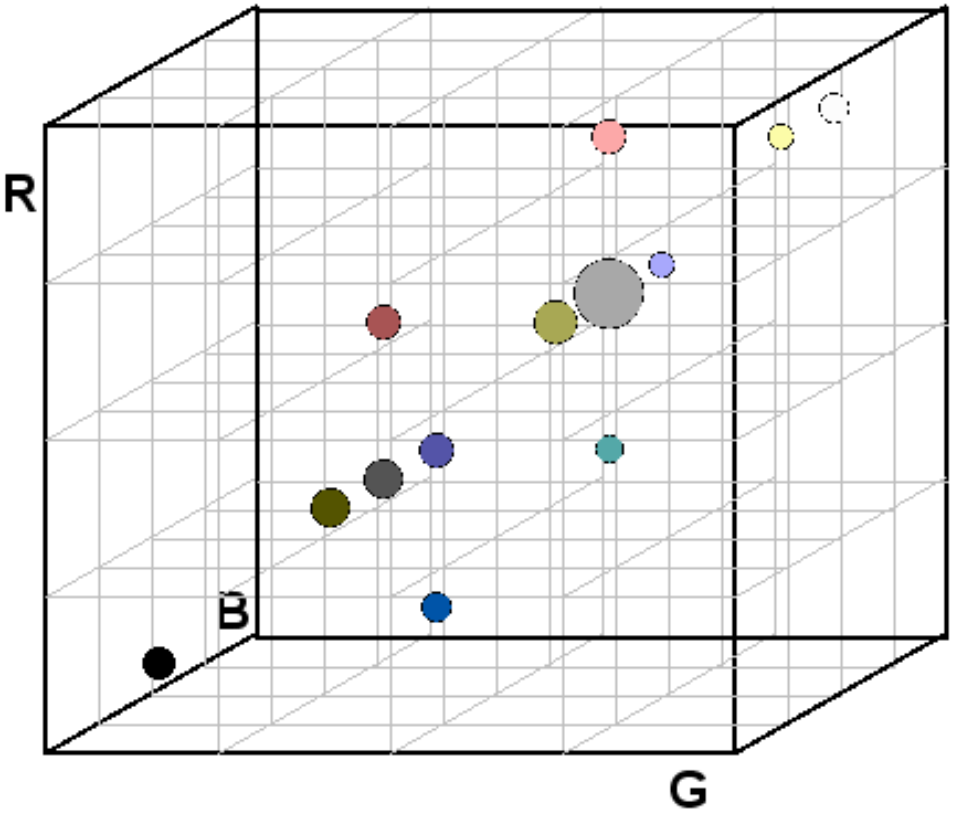
blue cyan green yellow red



spectral colours: wavelength (nm)

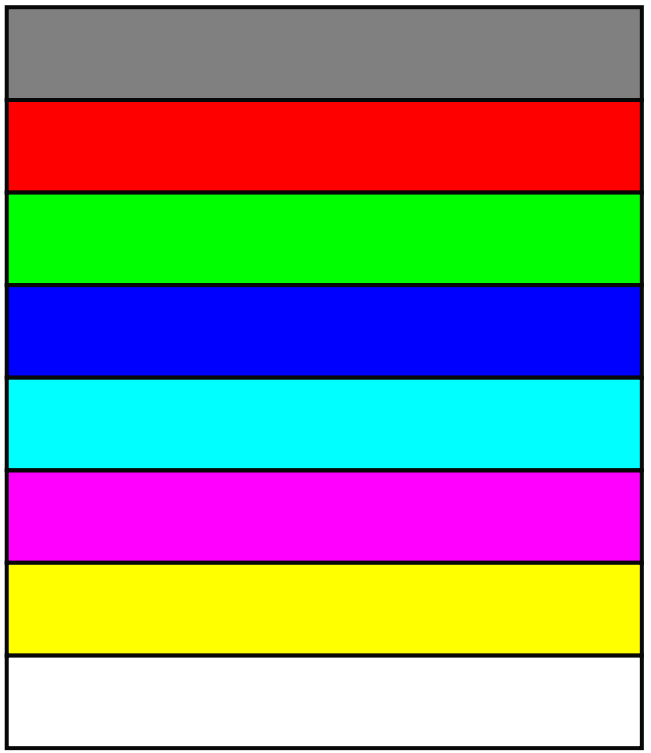
KNOWLEDGE MEDIA INSTITUTE

Colour histogram



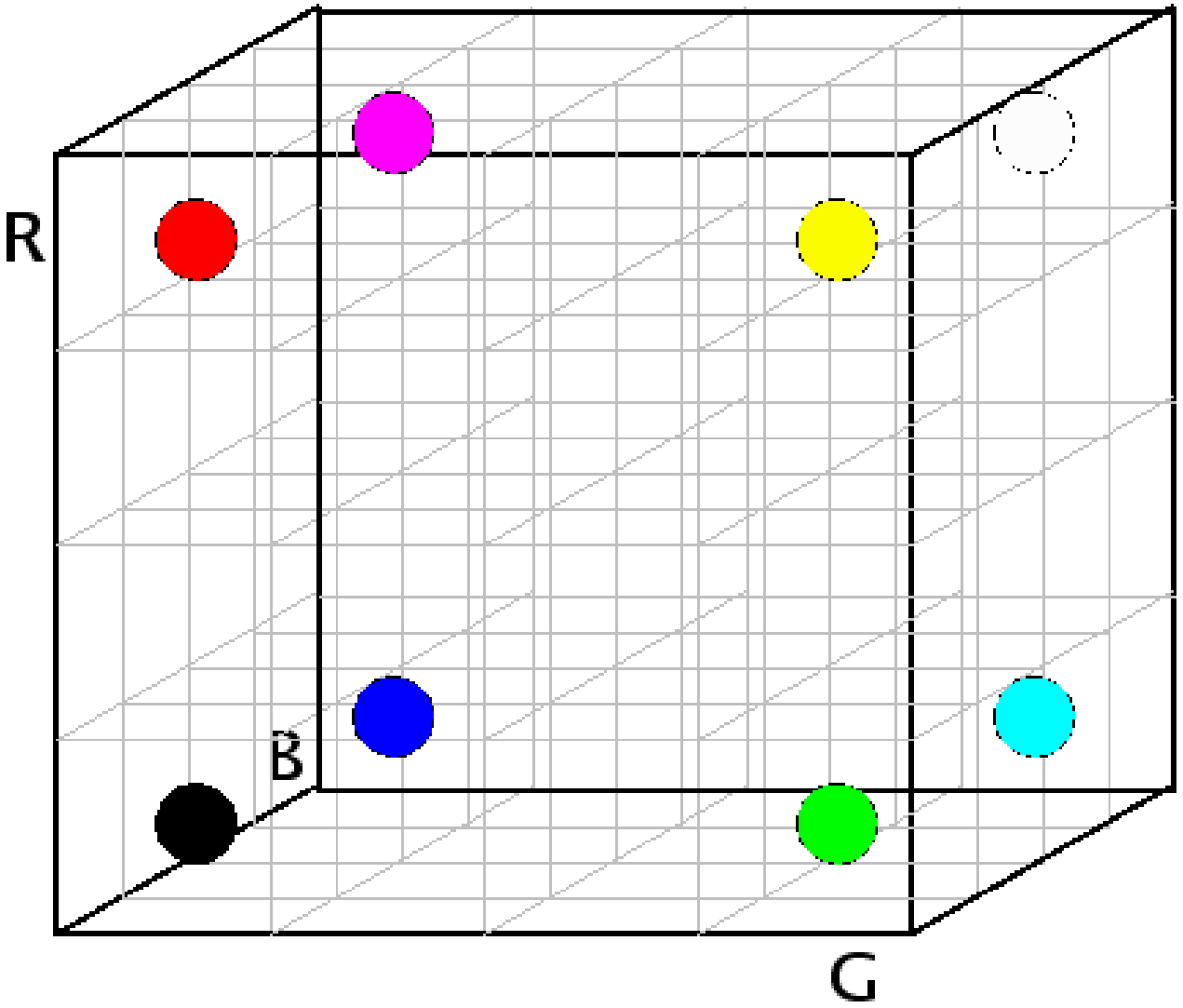
Exercise

Sketch a 3D colour histogram for



R	G	B	
0	0	0	black
255	0	0	red
0	255	0	green
0	0	255	blue
0	255	255	cyan
255	0	255	magenta
255	255	0	yellow
255	255	255	white

Solution

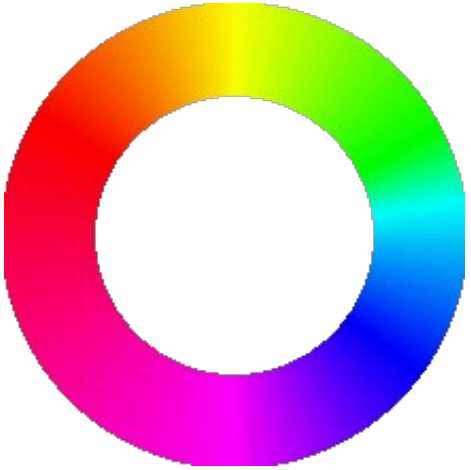


Other Colour Spaces

HSV, HSL, CIELAB/CIELUV

KNOWLEDGE MEDIA KMi INSTITUTE

HSB colour model



saturation (0% - 100%)
= spectral purity

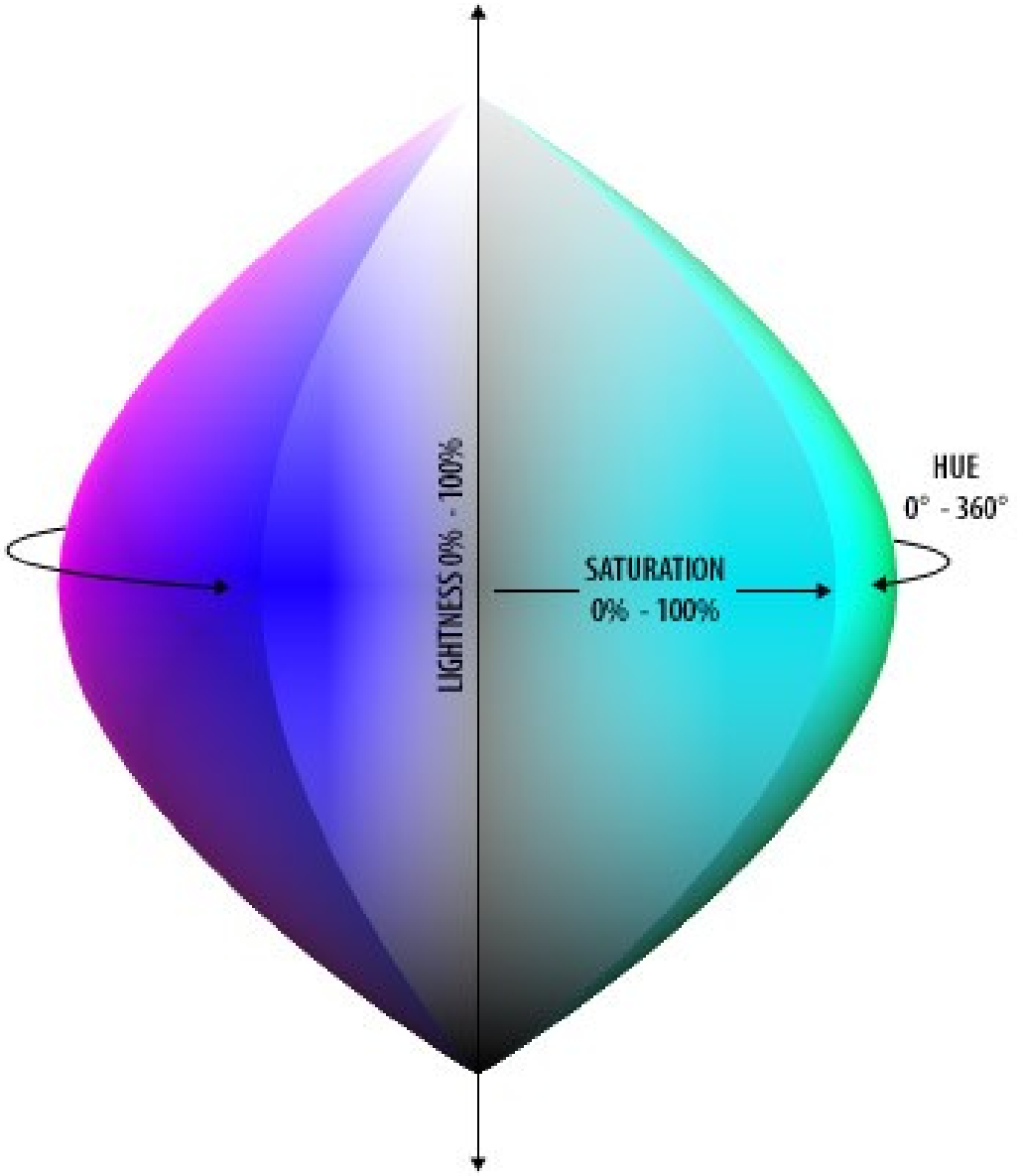
hue (0°-360°)
spectral colour

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

chromaticity = hue+saturation

KNOWLEDGE MEDIA INSTITUTE

HSB colour model



The Open University

HSB model

disadvantage: hue coordinate is not continuous

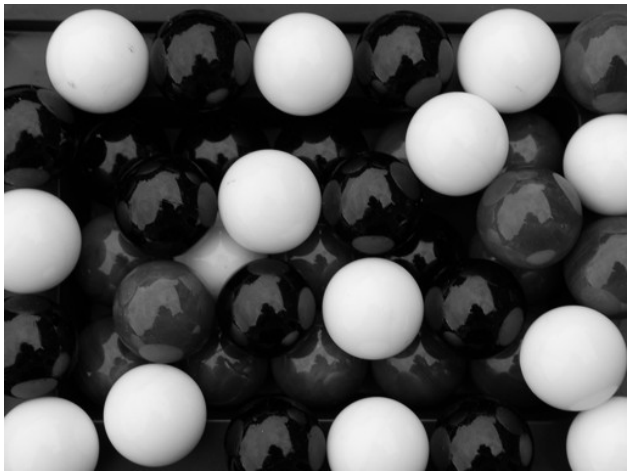
- 0 and 360 degrees have the same meaning
- but there is a huge difference in terms of numeric distance
- example:
  red = (0,100%,50%) = (360,100%,50%)

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

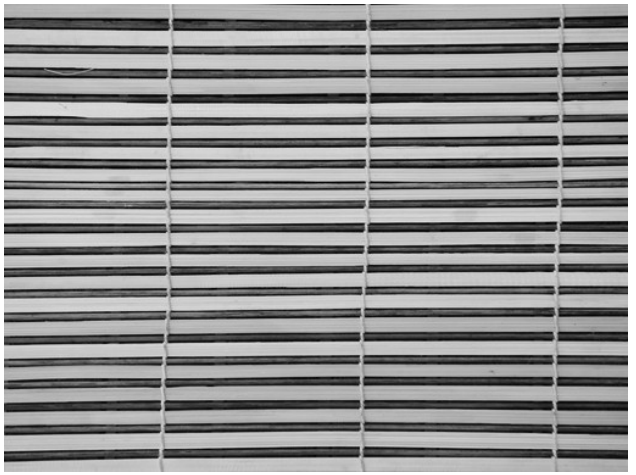
KNOWLEDGE MEDIA INSTITUTE

Texture

coarseness



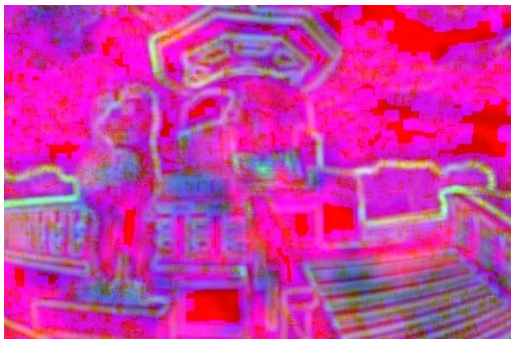
contrast

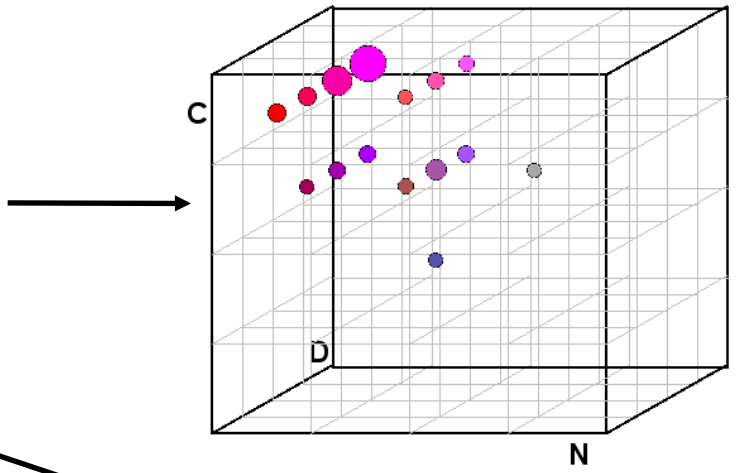


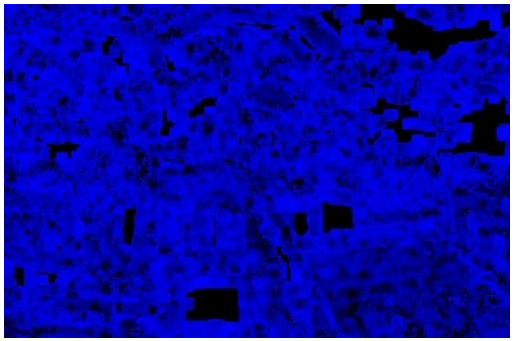
directionality

KNOWLEDGE MEDIA KMi INSTITUTE

Texture histograms







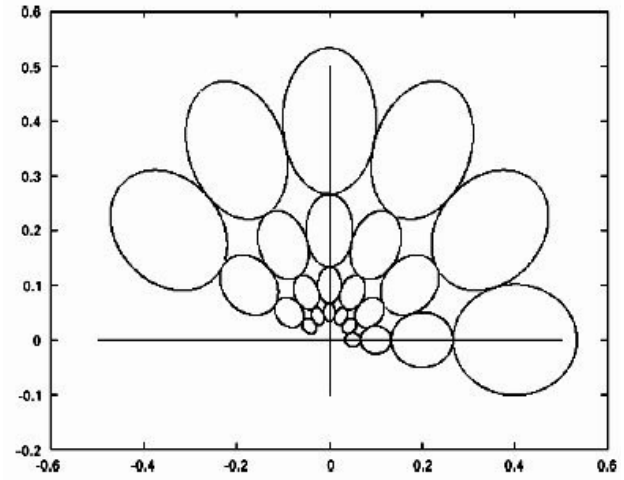
Coarseness coNtrast Directionality

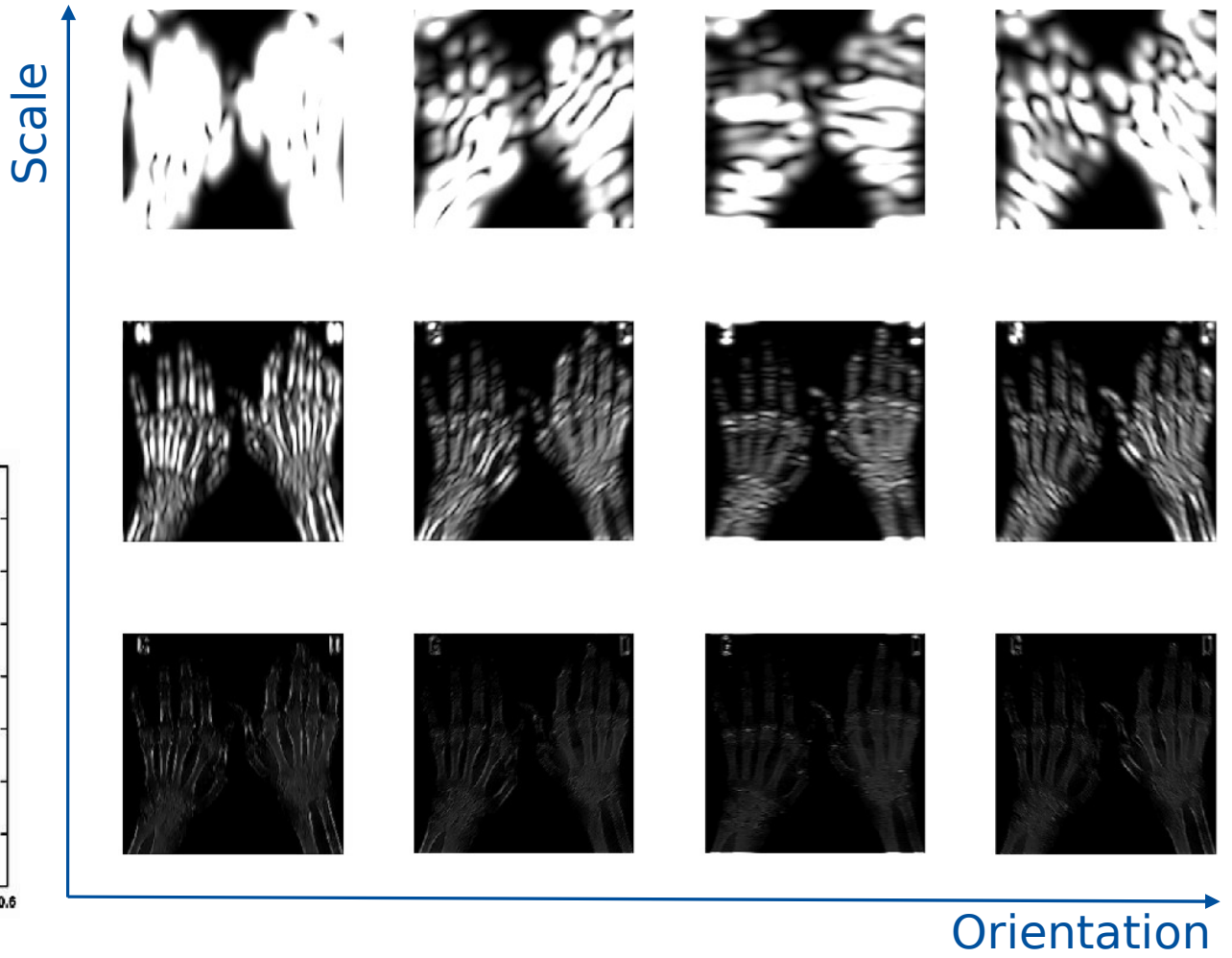
[with Howarth, *IEE Vision, Image & Signal Proc* 15(6) 2004; Howarth PhD thesis]

KNOWLEDGE MEDIA KMi INSTITUTE

Gabor filter

Query





[with Howarth, CLEF 2004]

KNOWLEDGE MEDIA INSTITUTE

Shape Analysis

shape = class of geometric objects invariant under
- translation
- scale (changes keeping the aspect ratio)
- rotations

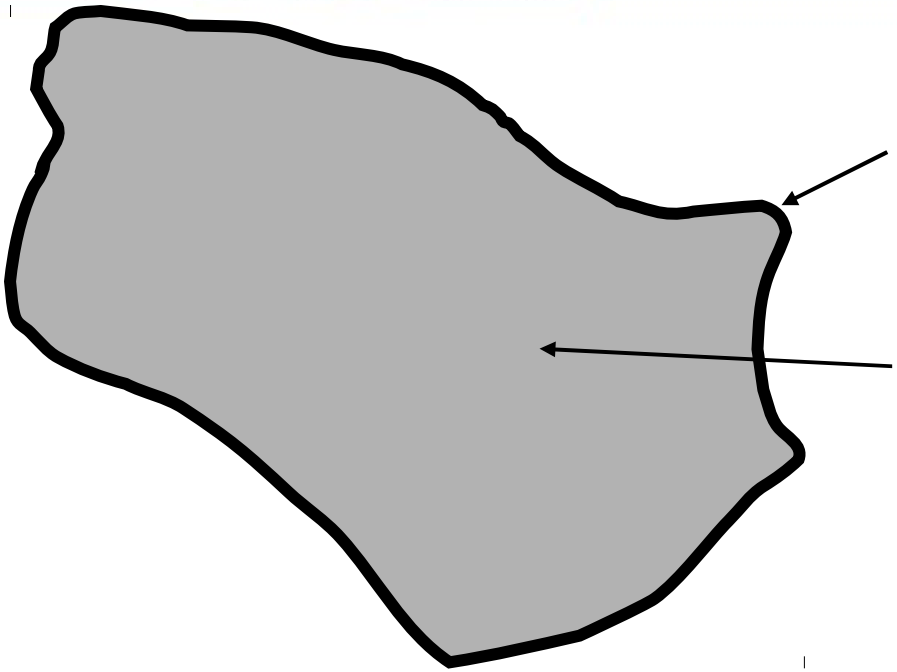
information preserving description (for compression)

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



$$P = \int \sqrt{x'^2(t) + y'^2(t)}dt$$

$$A = \iint_R dxdy$$

~~boundary pixel count~~

count pixels in area

Circularity

$$T = 4\pi \frac{A}{P^2}$$

A=area, P=perimeter

T is 1 for a circle

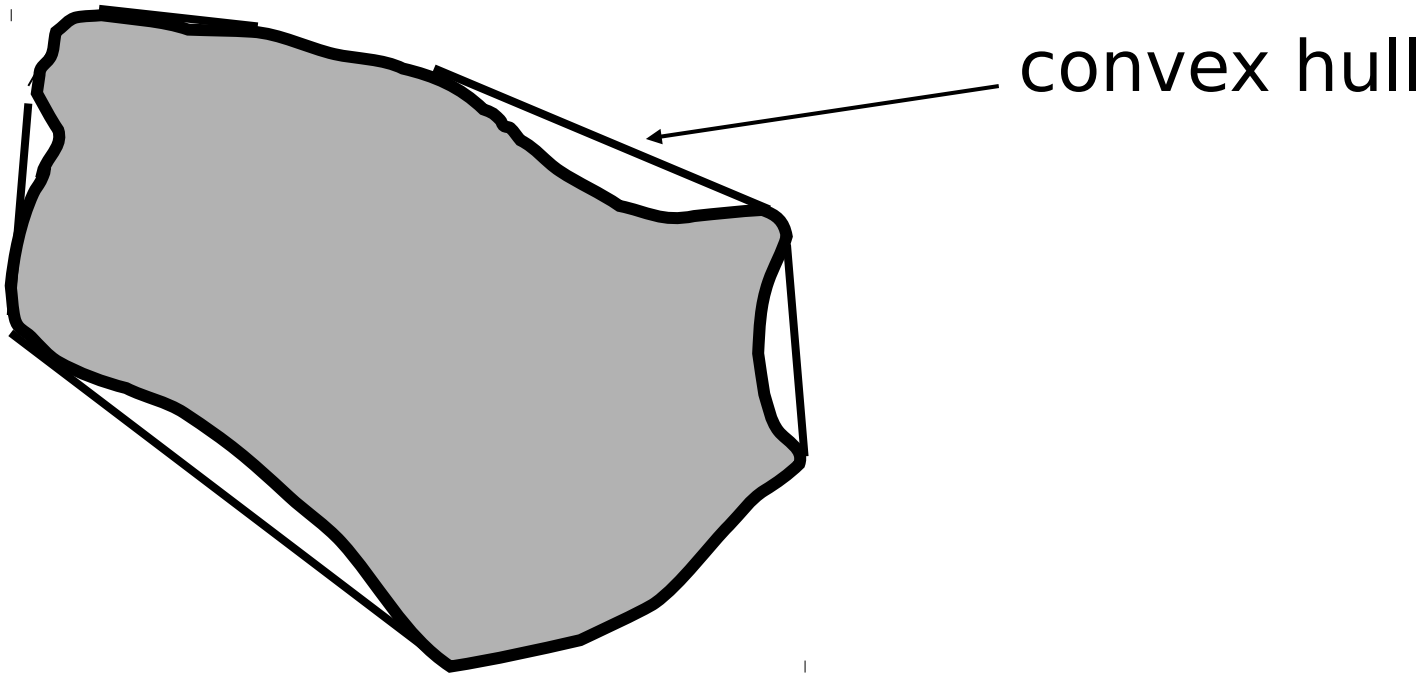
T is smaller than 1 for all other shapes

circularity is aka compactness

Convexity

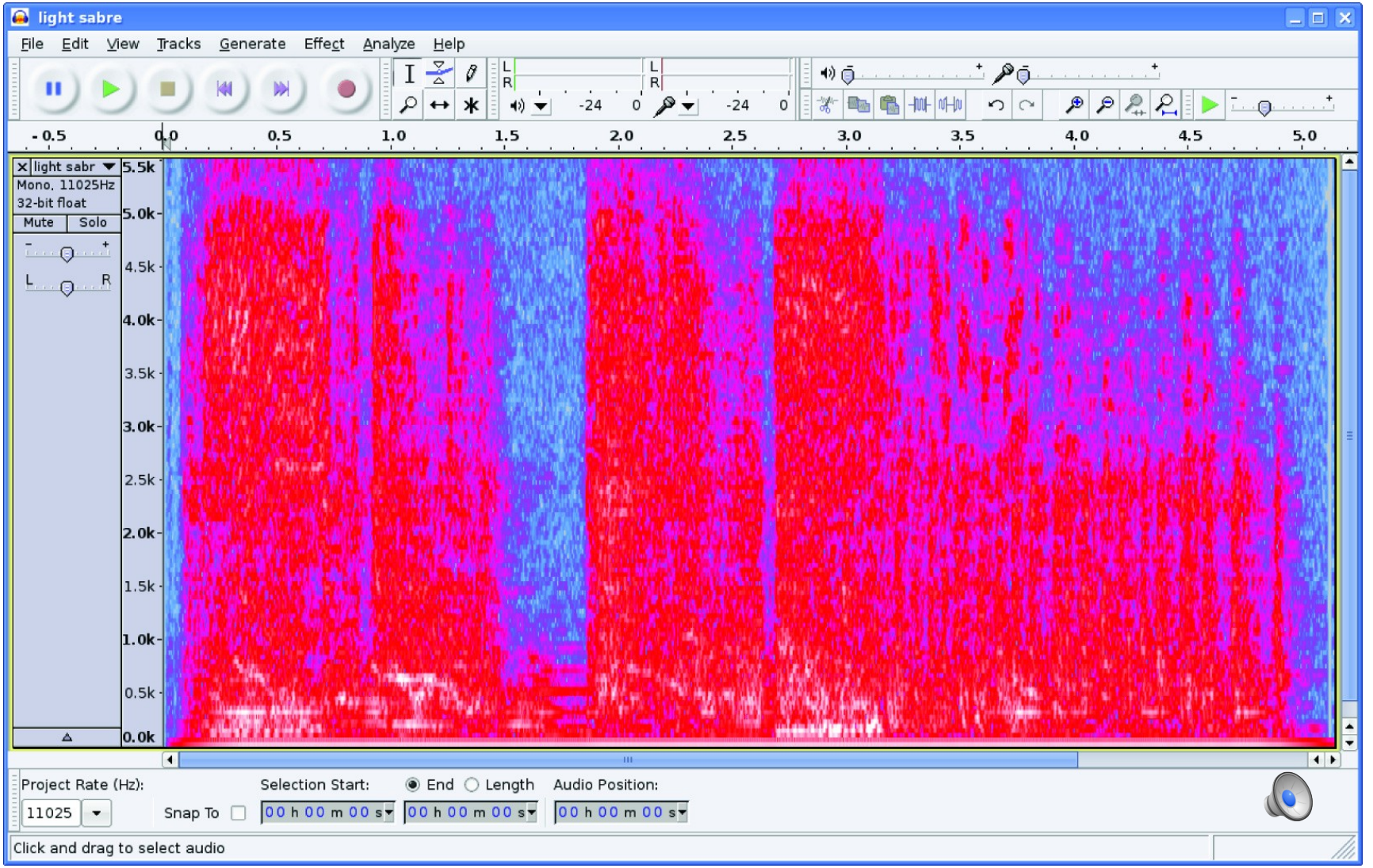
ratio of perimeter of convex hull and the original curve

1 for convex shapes, less than 1 otherwise



KNOWLEDGE MEDIA
KMi
INSTITUTE

Sound



The Open University

KNOWLEDGE MEDIA KMi INSTITUTE

Audio Features

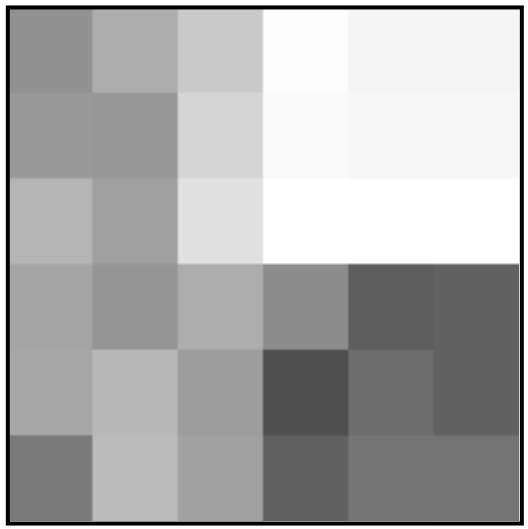
- Spectrogram
  - graph of frequencies/energy/time
- tempo, pitch, mode
- See Z Liu, Y Wang and T Chen (1998). Audio feature extraction and analysis for scene segmentation and classification. *VLSI Signal Processing 20*(1-2), 69-79.

Histograms

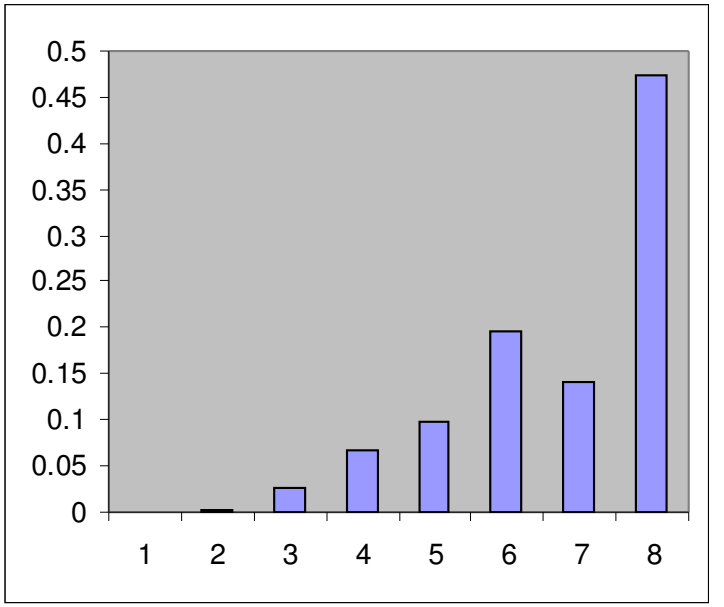
- Condensed
- Content-based
- Real-valued vector
- Summarising
- Sparseness
- Statistical moments

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

- μ Mean
- $\bar{p}_2$ Variance (squared standard deviation)
- $\bar{p}_3$ 3rd central moment (skewness)

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

where w is image width and h is image height

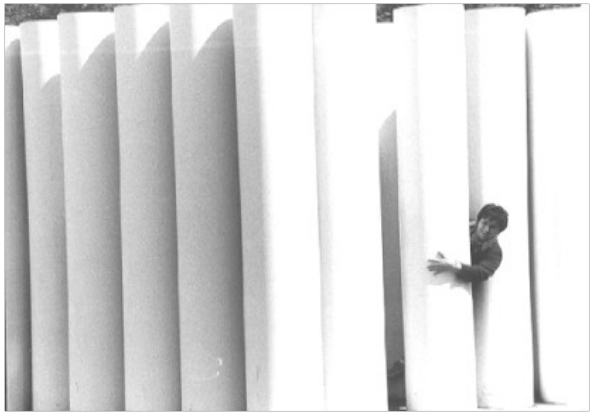
Moment features

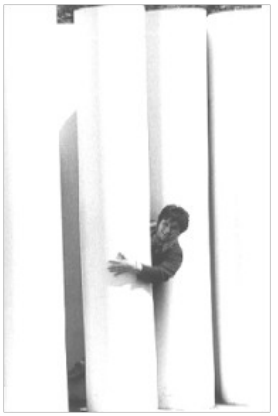
$$(\mu,\ \sqrt{\bar{p}_2},\ \text{sign}(\bar{p}_3)\sqrt[3]{|\bar{p}_3|})$$

Moment features

$$\Big(\mu_r,\ \sqrt{\bar{r}_2},\ \mathrm{sign}(\bar{r}_3)\sqrt[3]{|\bar{r}_3|},$$
$$\mu_g,\ \sqrt{\bar{g}_2},\ \mathrm{sign}(\bar{g}_3)\sqrt[3]{|\bar{g}_3|},$$
$$\mu_b,\ \sqrt{\bar{b}_2},\ \mathrm{sign}(\bar{b}_3)\sqrt[3]{|\bar{b}_3|}\Big)$$

Global vs local

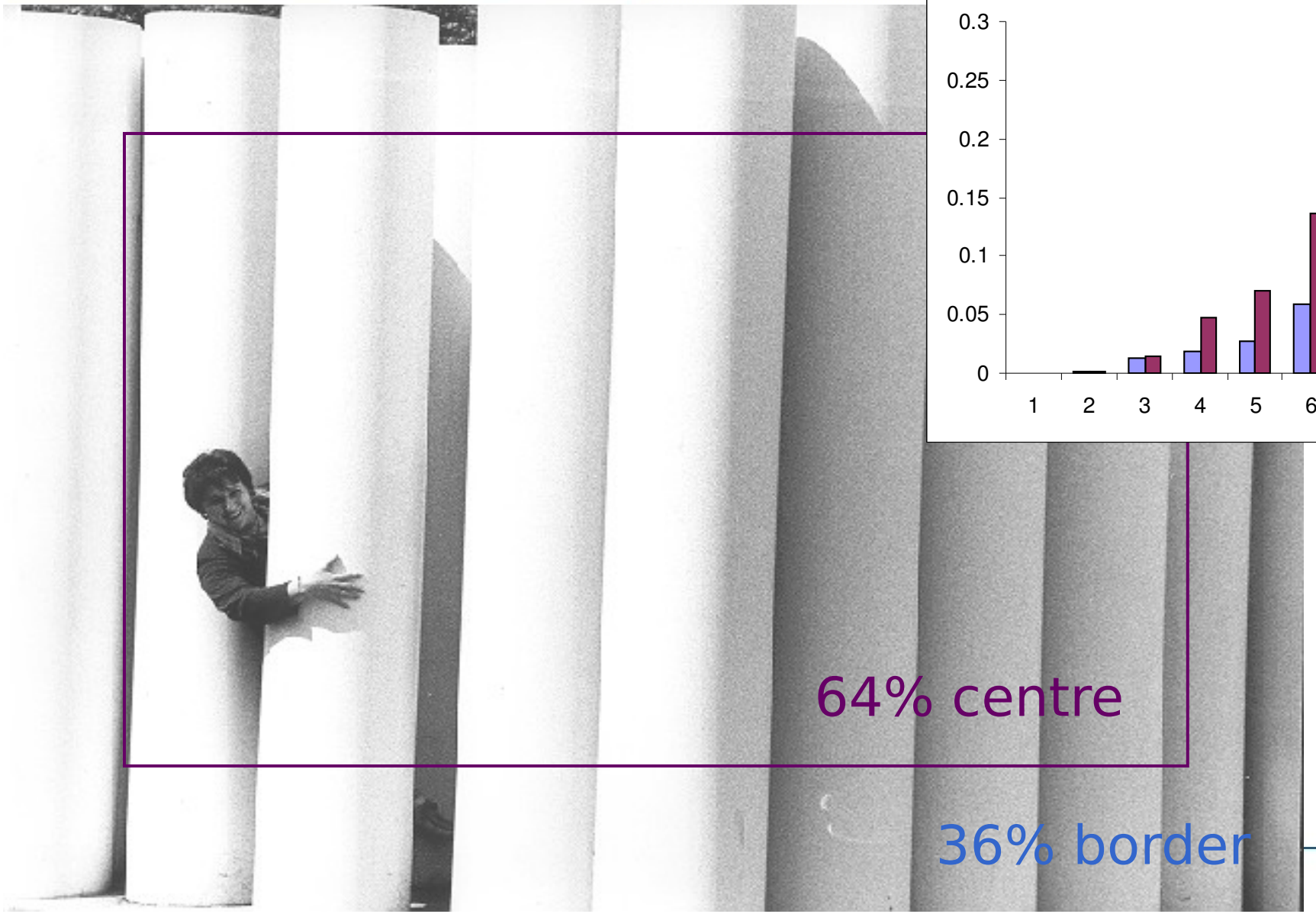




Global histogram also matches polar bears, marble floors, ...

KNOWLEDGE MEDIA KMi INSTITUTE

Localisation



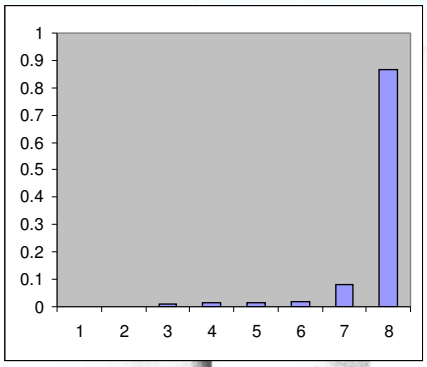
64% centre

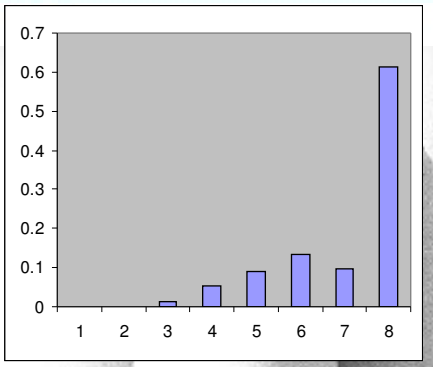
36% border

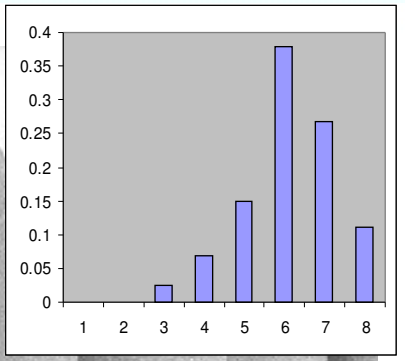
The Open University

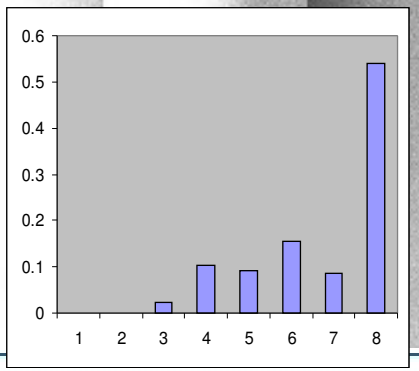
KNOWLEDGE MEDIA
KMi
INSTITUTE

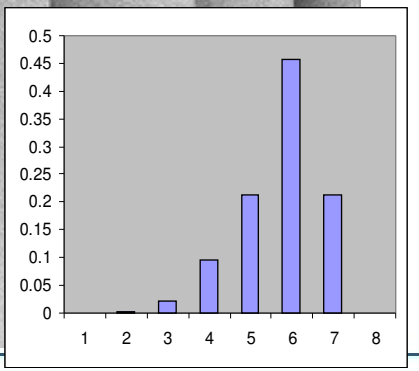
Tiled Histograms







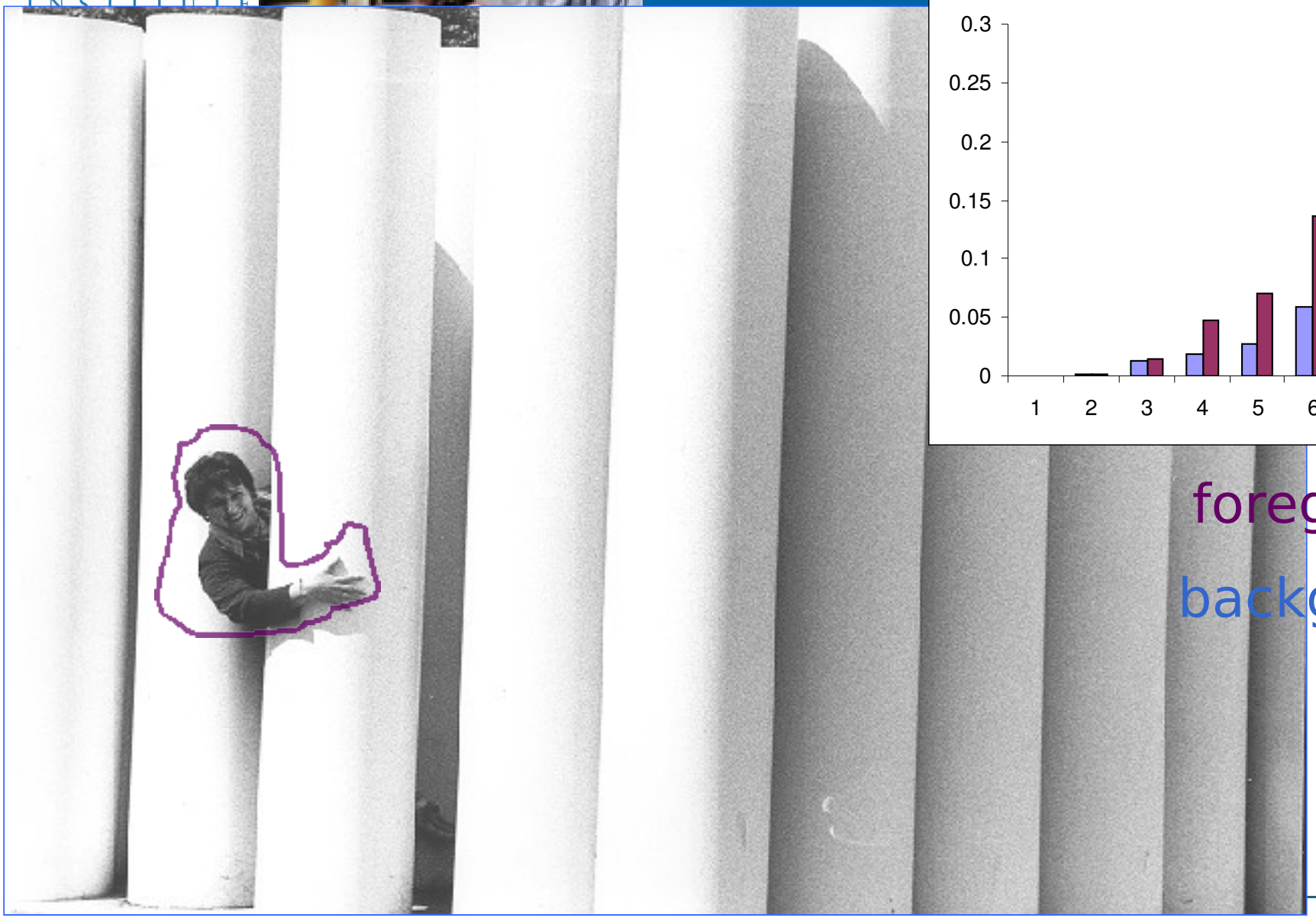




The Open University

KNOWLEDGE MEDIA INSTITUTE

Segmentation

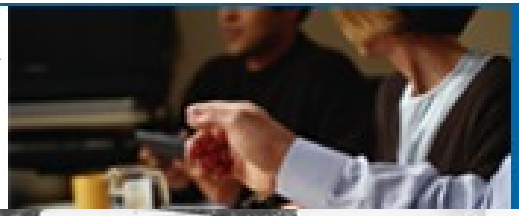


foreground

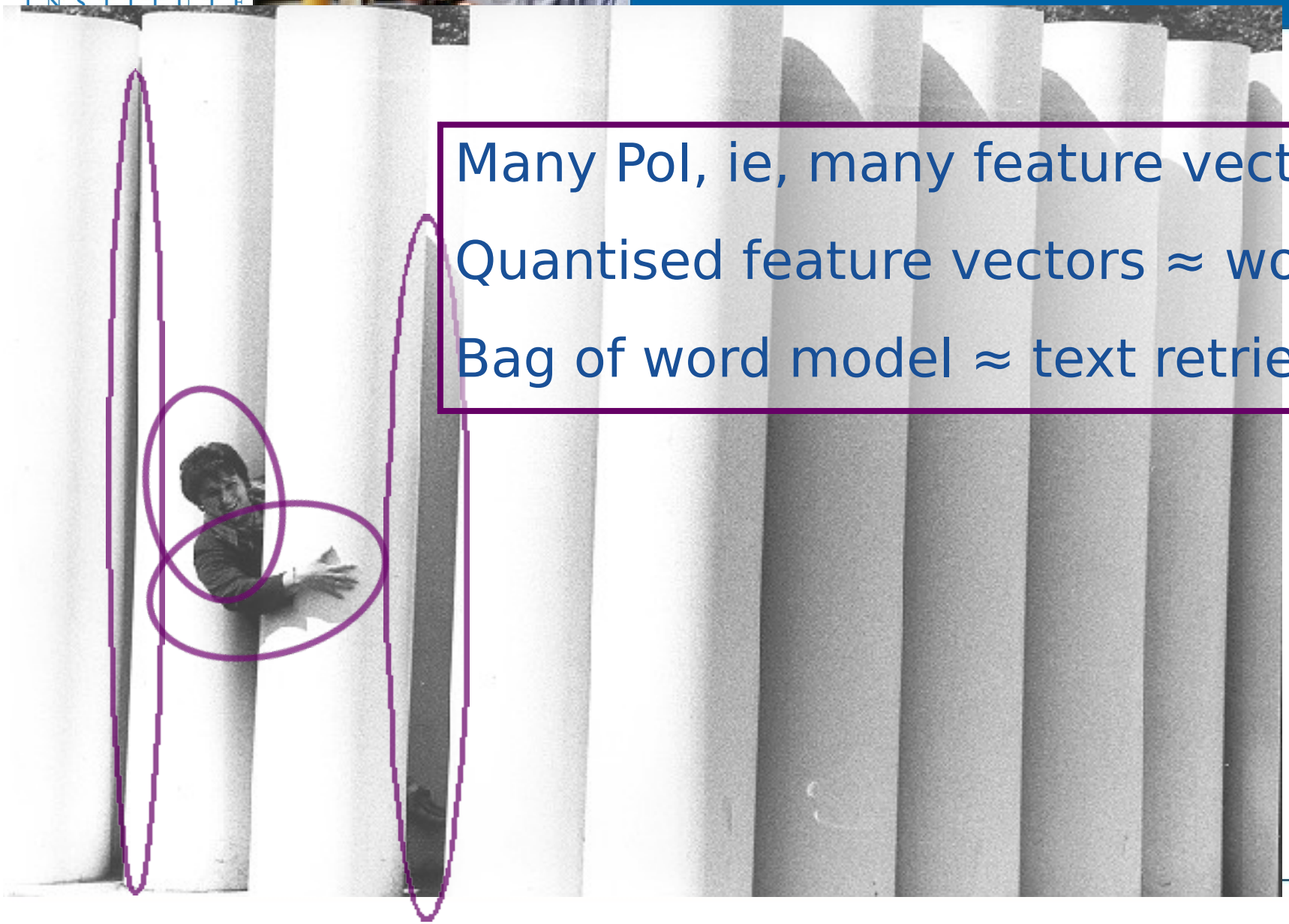
background

The Open University

KNOWLEDGE MEDIA KMi INSTITUTE



Points of interest



Many PoI, ie, many feature vectors

Quantised feature vectors ≈ words

Bag of word model ≈ text retrieval

The Open University

KNOWLEDGE MEDIA INSTITUTE

“Bag of Words”

KNOWLEDGE MEDIA INSTITUTE

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?

KNOWLEDGE MEDIA
KMi
INSTITUTE

Video Segmentation

- Anticipation Trailer
- Segmentation Equations

KNOWLEDGE MEDIA INSTITUTE

Video Segmentation

gradual transition detection (eg, fade)
- accumulate distances
- long-range comparison

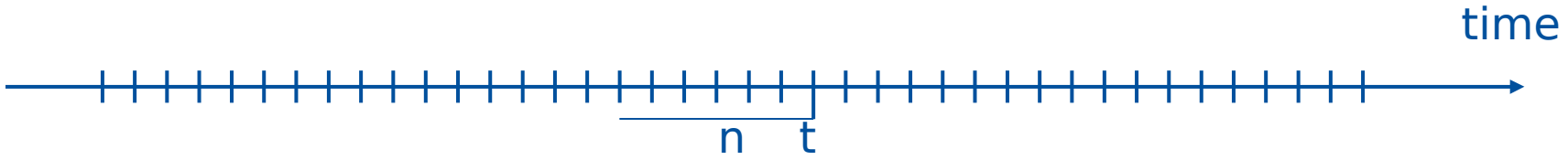
audio cues
- silence and/or speaker change

motion detection and analysis
- camera motion, zoom, object motion
- MPEG provides some motion vectors

[Vlad Tanasescu: Anticipation, SCiFi trailer]]

Long range comparison

At time t define distance $d_n(t)$



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

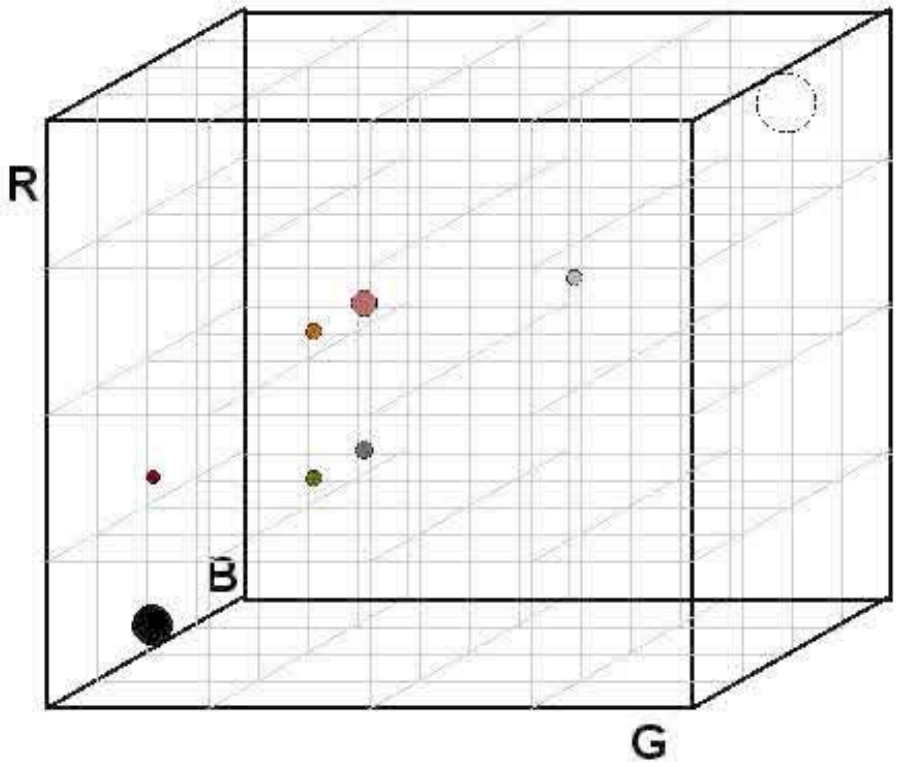
- average their respective distances over i

Peak in $d_n(t)$ detected if

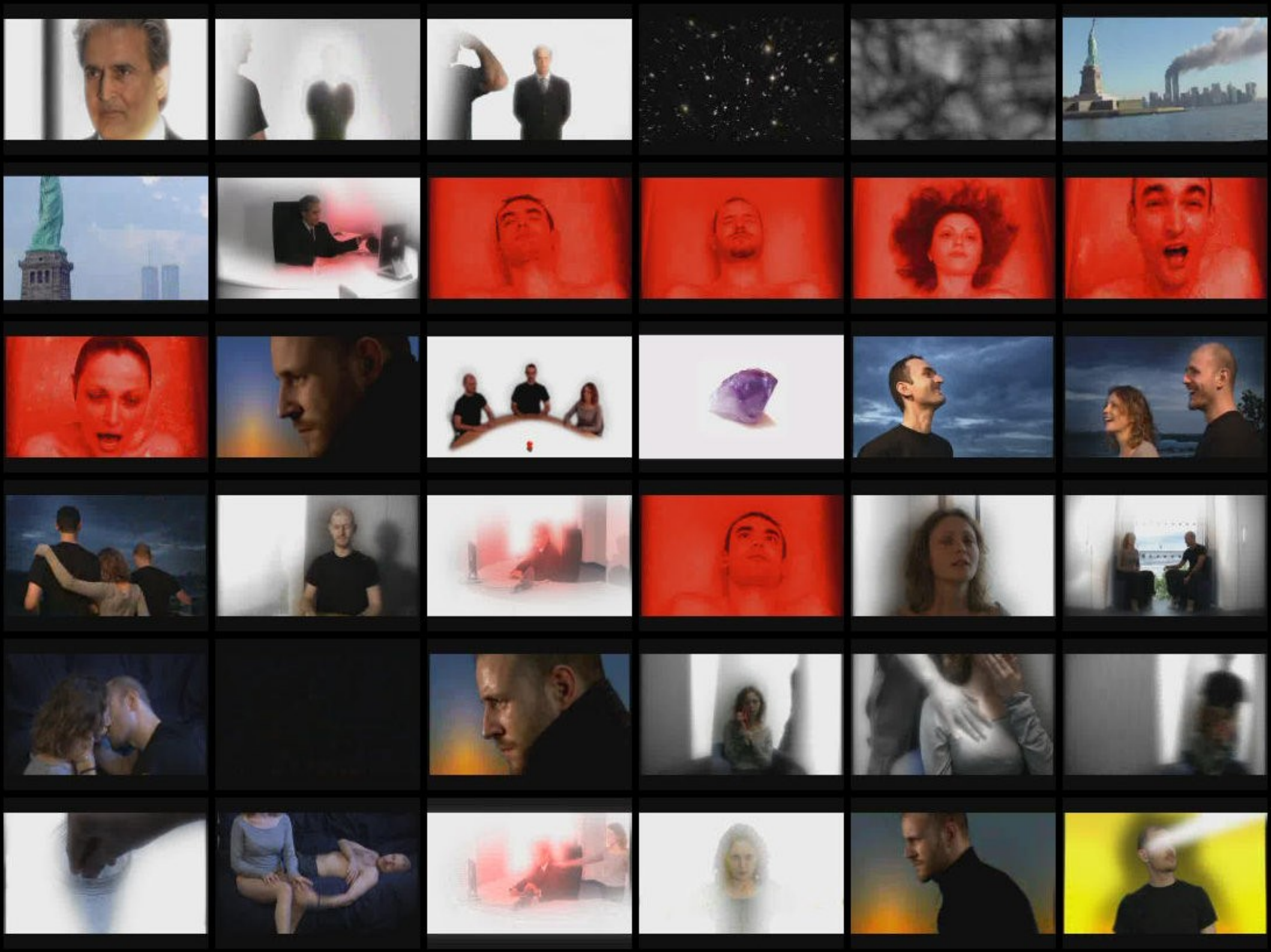
$d_n(t)$>threshold and

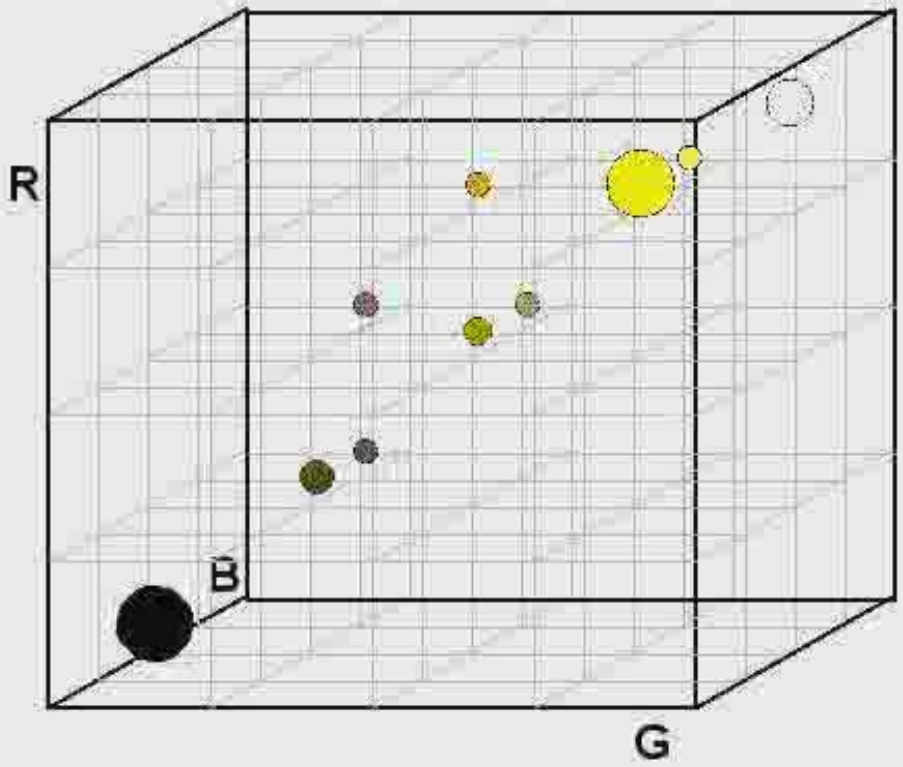
$d_n(t)>d_n(s)$ for all neighbouring s

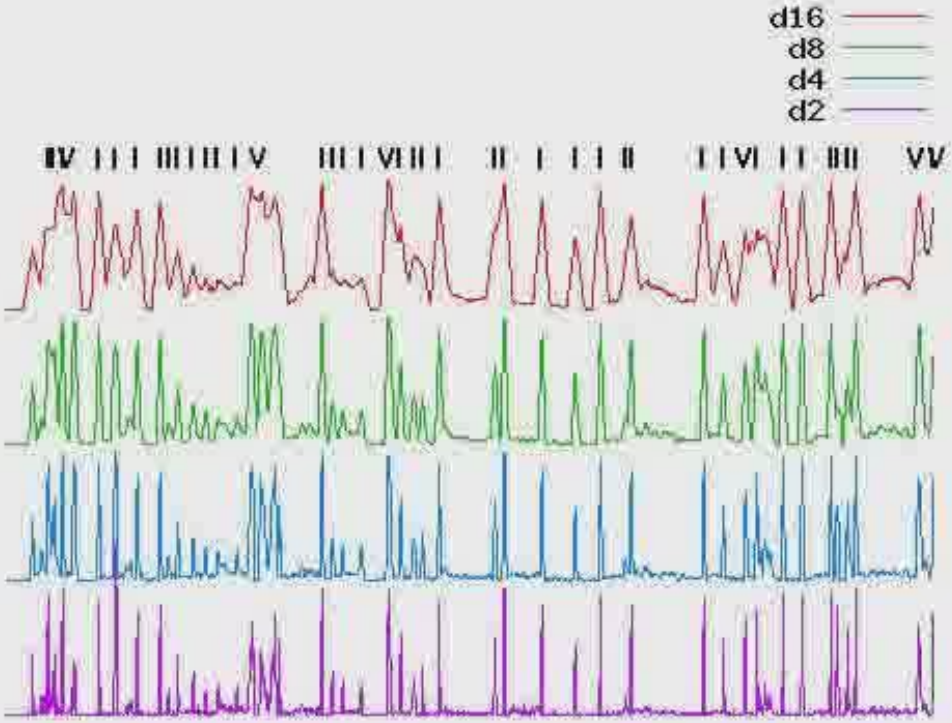
Shot = near-coincident peaks of d_{16} and d_8



d16
d8
d4
d2

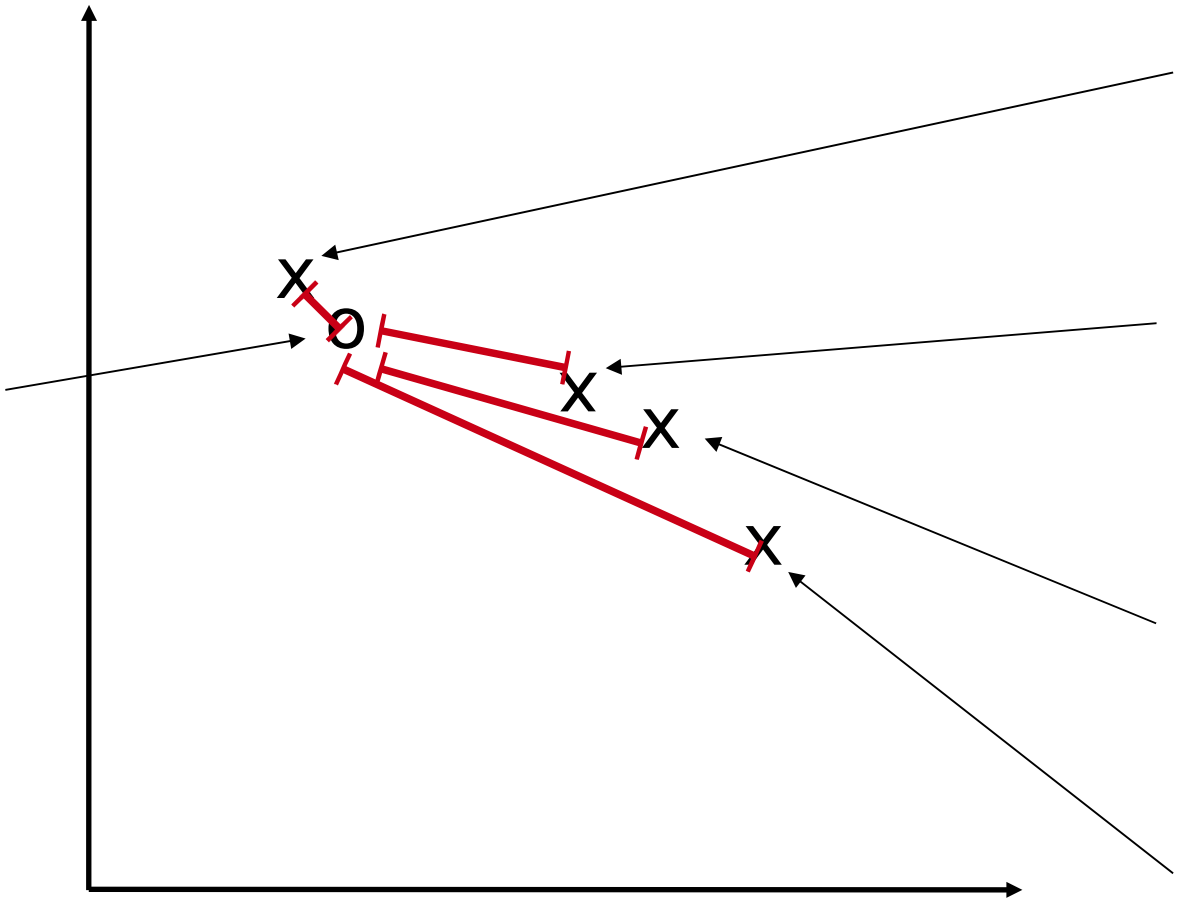



R
B
G


d16
d8
d4
d2

KNOWLEDGE MEDIA
KMi
INSTITUTE

Features and distances



The Open University

KNOWLEDGE MEDIA KMi INSTITUTE

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

$$sim(h, q) = \sum_i min(h_i, q_i)$$

L_2 vs L_1

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

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

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

$$d_\infty(v, w) =$$

L_2 vs L_1

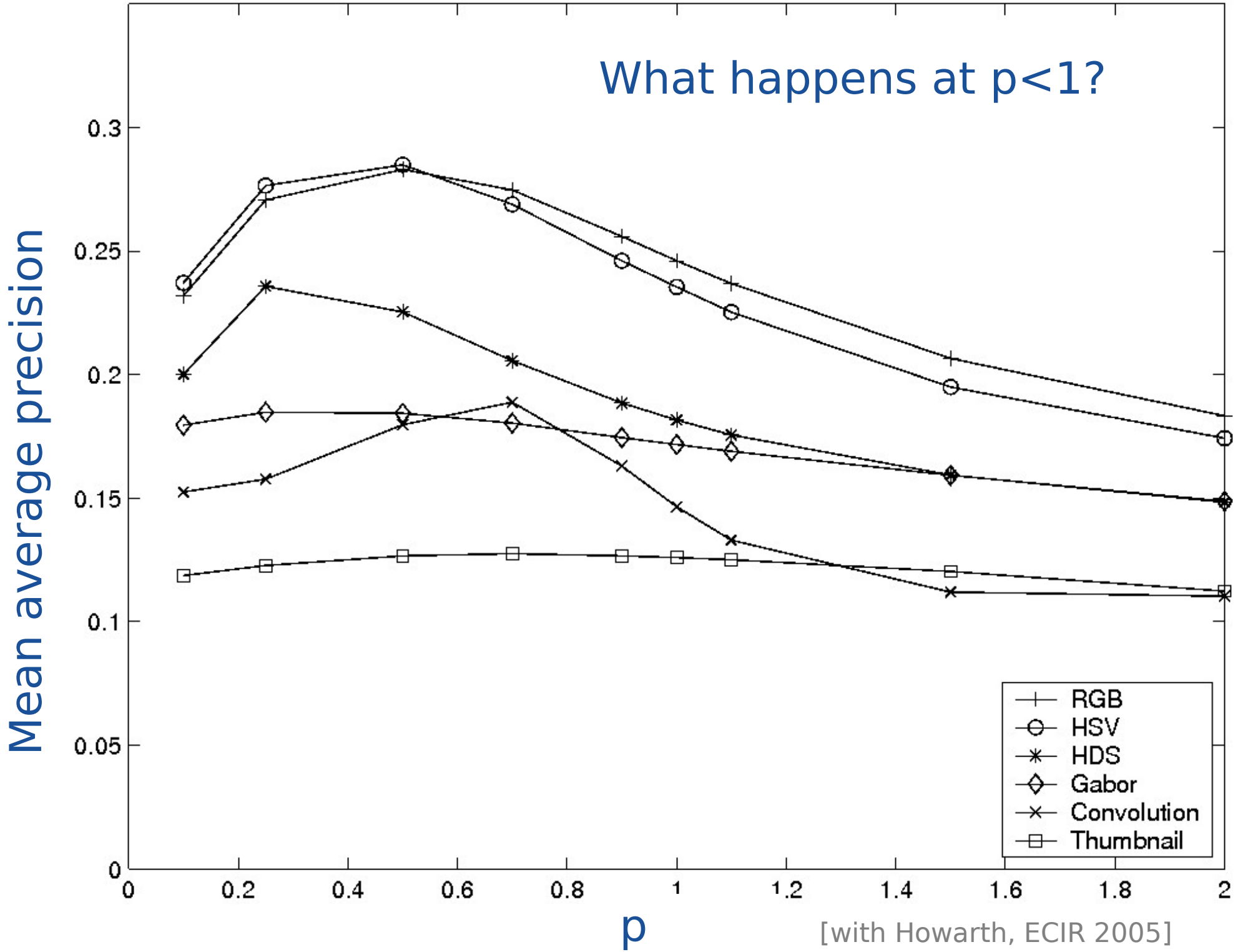
$$d_p(v, w) = \sqrt[p]{\sum_i |v_i - w_i|^p}, \quad p \geq 1$$

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

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

$$d_\infty(v, w) = \max_i |v_i - w_i|$$

What happens at p<1?



[with Howarth, ECIR 2005]

KNOWLEDGE MEDIA INSTITUTE

Other distance measures

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

KNOWLEDGE MEDIA INSTITUTE

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

KNOWLEDGE MEDIA INSTITUTE

Exercise

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

KNOWLEDGE MEDIA INSTITUTE

Information Retrieval

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"

KNOWLEDGE MEDIA INSTITUTE