KNOWLEDGE MEDIA
KMi
INSTITUTE

Multimedia Information Retrieval

Prof Stefan Rüger

Multimedia and Information Systems
Knowledge Media Institute
The Open University

http://kmi.open.ac.uk/mmis

MMIS
Multimedia and Information Systems

KNOWLEDGE MEDIA INSTITUTE

Why content-based?

Actually, what is content-based search?

Is human thinking content-based?

Metadata annotation (text) is good but

-
-
-
-

KNOWLEDGE MEDIA KMi INSTITUTE

Multimedia Information Retrieval

1. What are multimedia queries?
2. Fingerprinting
3. Metadata & piggy-back retrieval
4. Automated image annotation
5 Visual content-based retrieval I (challenges + features)
6 Visual content-based retrieval II (distances)
7. Evaluation
8. Browsing, search and geography

KNOWLEDGE MEDIA INSTITUTE

Where are the challenges?

Image content analysis (diversity, semantic gap, polysemy)

Mapping to higher level (semantic representation)

Time taken and resources needed ("scalability")

Automation, scale and coping with errors

KNOWLEDGE MEDIA KMi INSTITUTE

The semantic gap

1m pixels with a spatial colour distribution

faces & vase-like object

victory, triumph, ...

disappointment, ..

.

The Open University

KNOWLEDGE MEDIA INSTITUTE

Polysemy

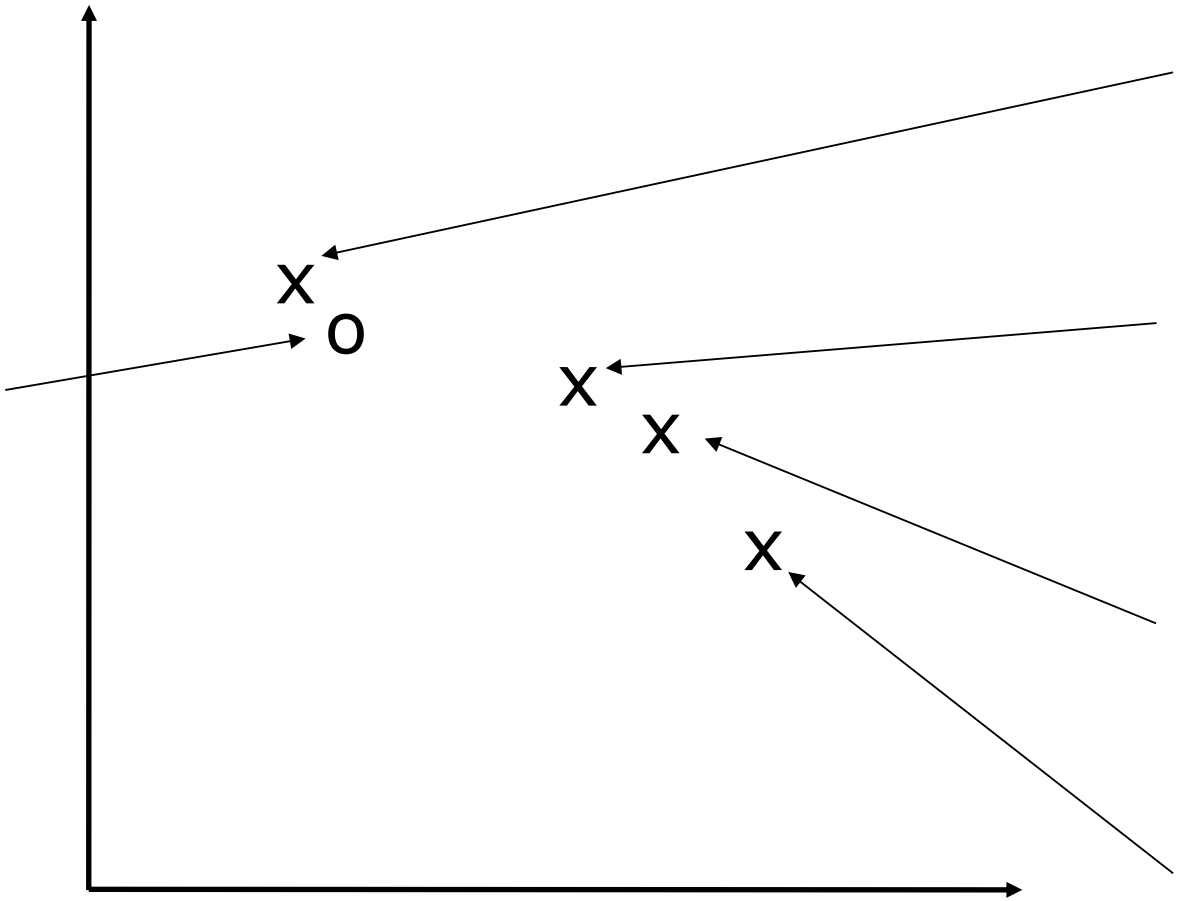
The Open University

KNOWLEDGE MEDIA INSTITUTE

Scalability

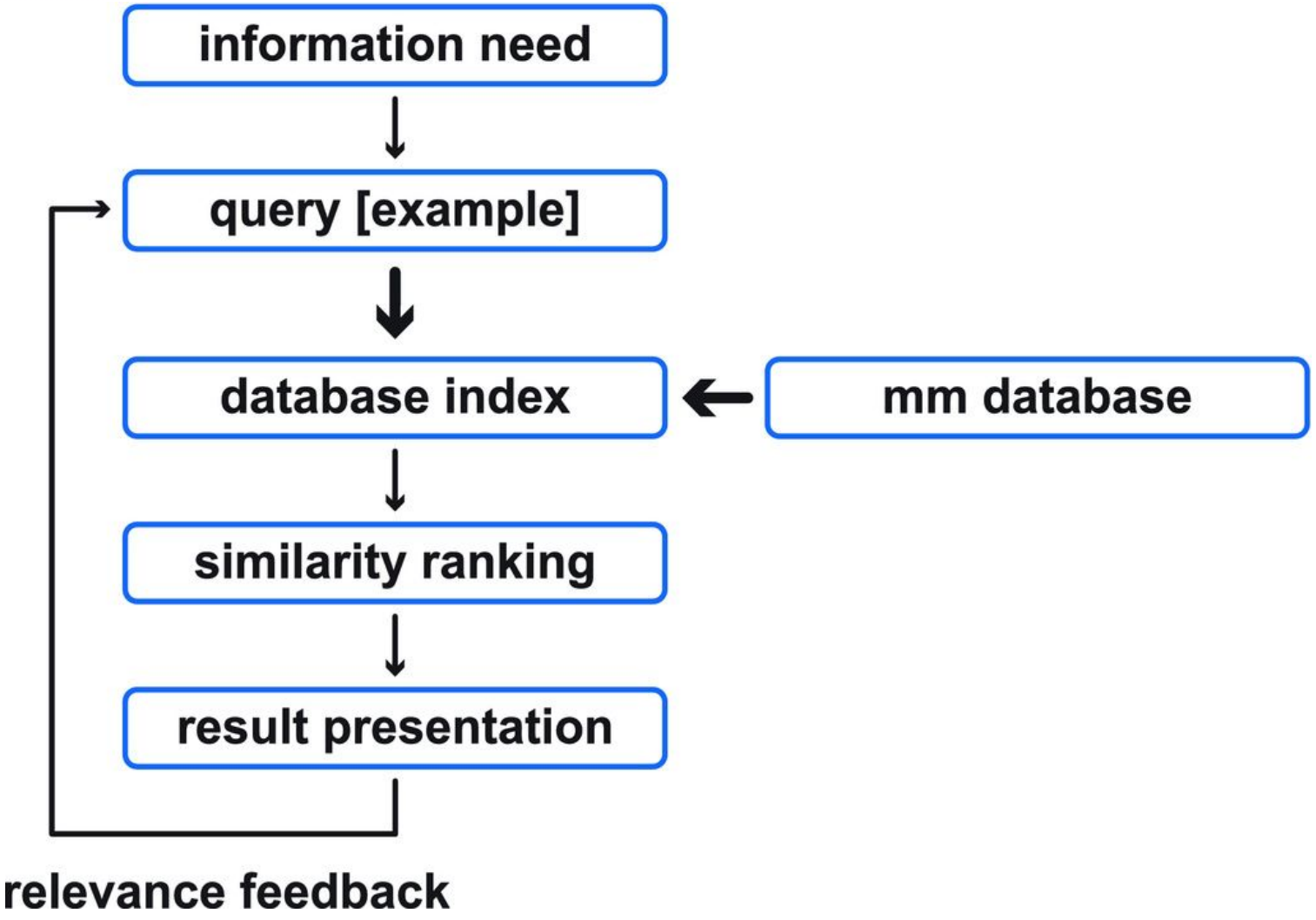
KNOWLEDGE MEDIA KMi INSTITUTE

Features and distances



The Open University

Architecture



Features

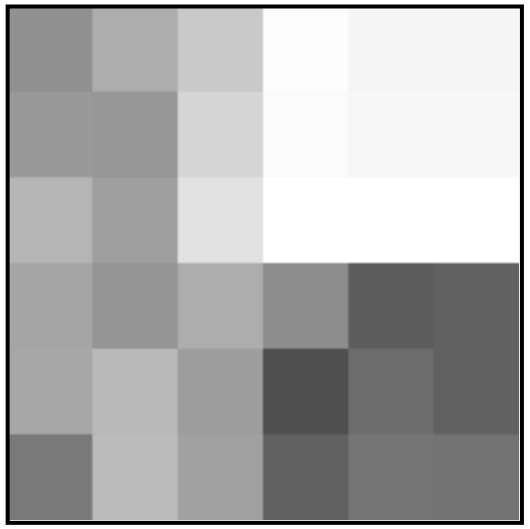
- Visual
  - Colour, texture, shape, edge detection, SIFT/SURF
- Audio
- Temporal
- How to describe the features?
  - For people
  - For computers

KNOWLEDGE MEDIA
KMi
INSTITUTE

Digital Images

KNOWLEDGE MEDIA INSTITUTE

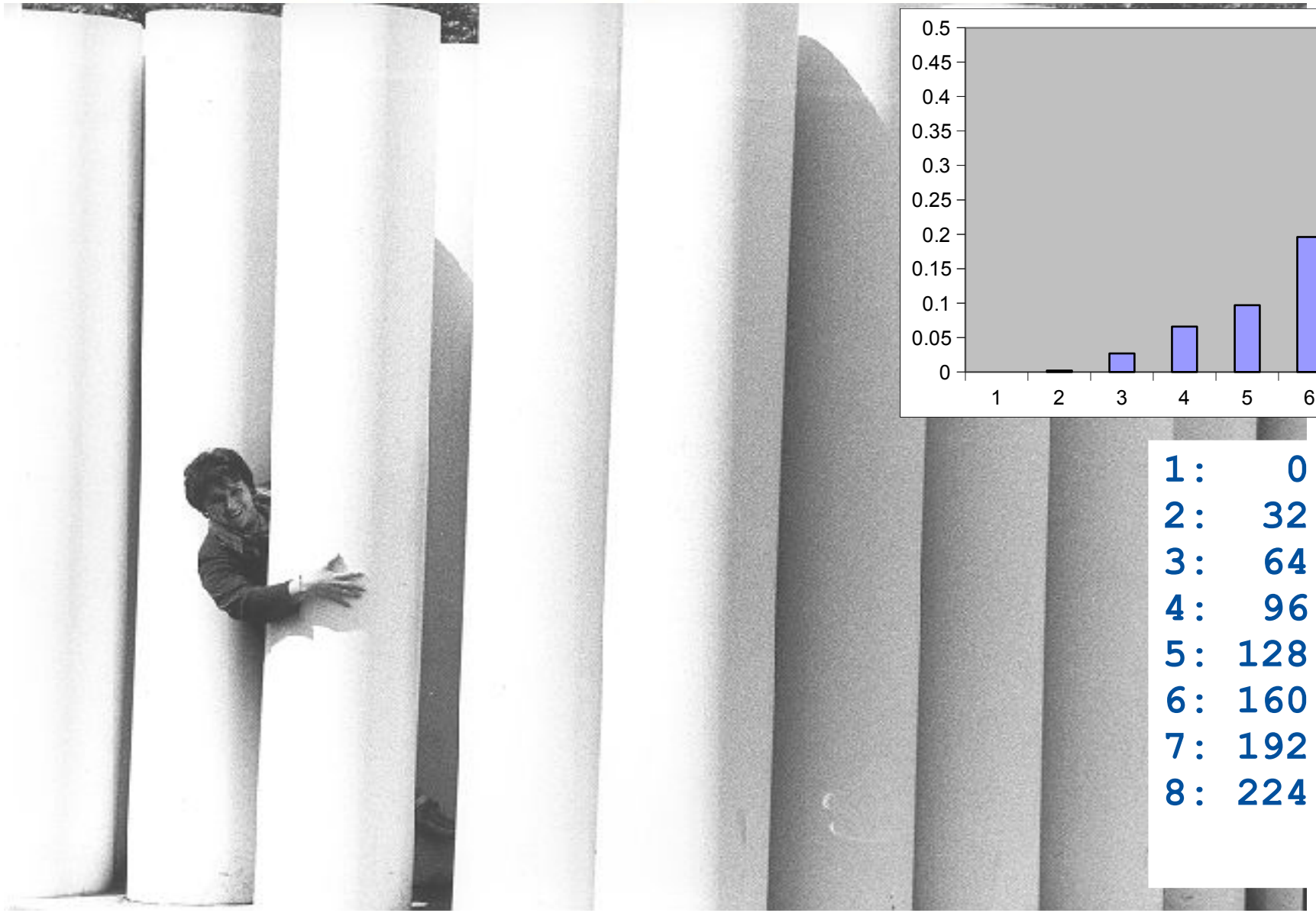
Content of an image



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



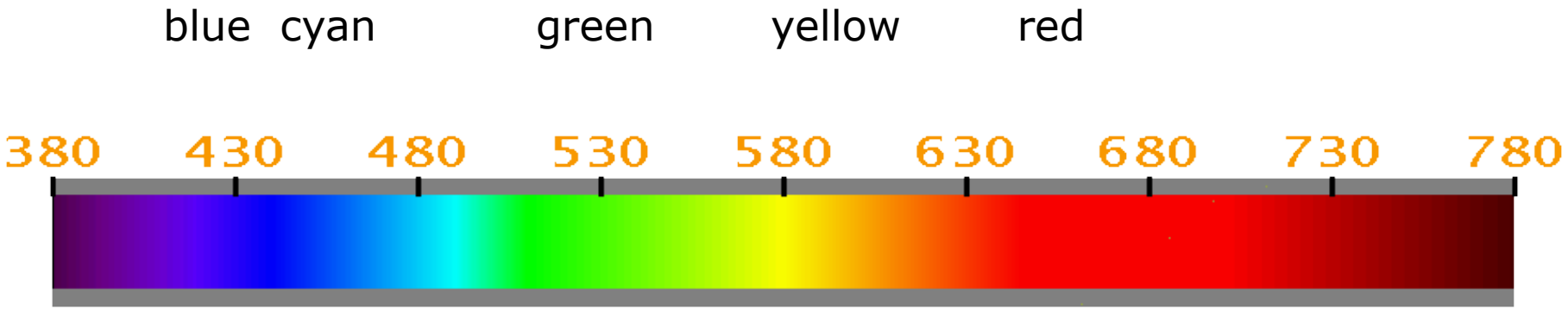

```
1:   0 -  31
2:  32 -  63
3:  64 -  95
4:  96 - 127
5: 128 - 159
6: 160 - 191
7: 192 - 223
8: 224 - 255
```

Colour

phenomenon of human perception

three-dimensional (RGB/CMY/HSB)

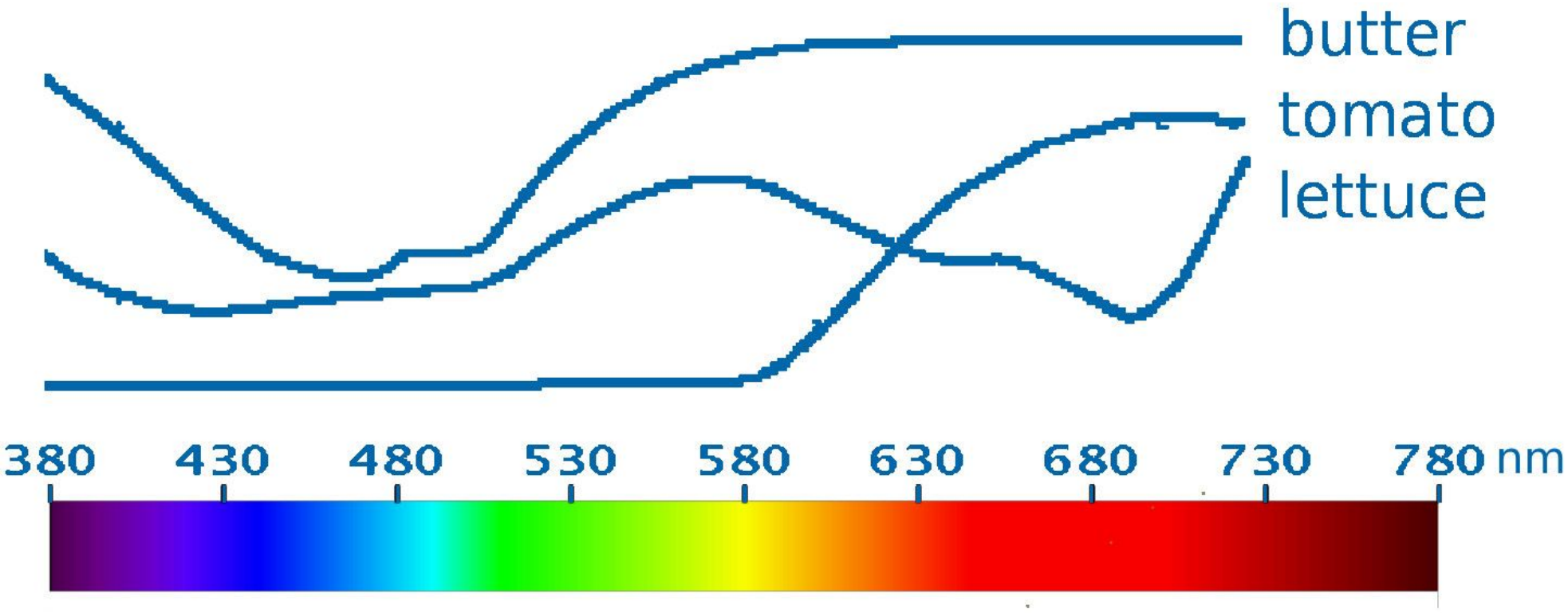
spectral colour: pure light of one wavelength

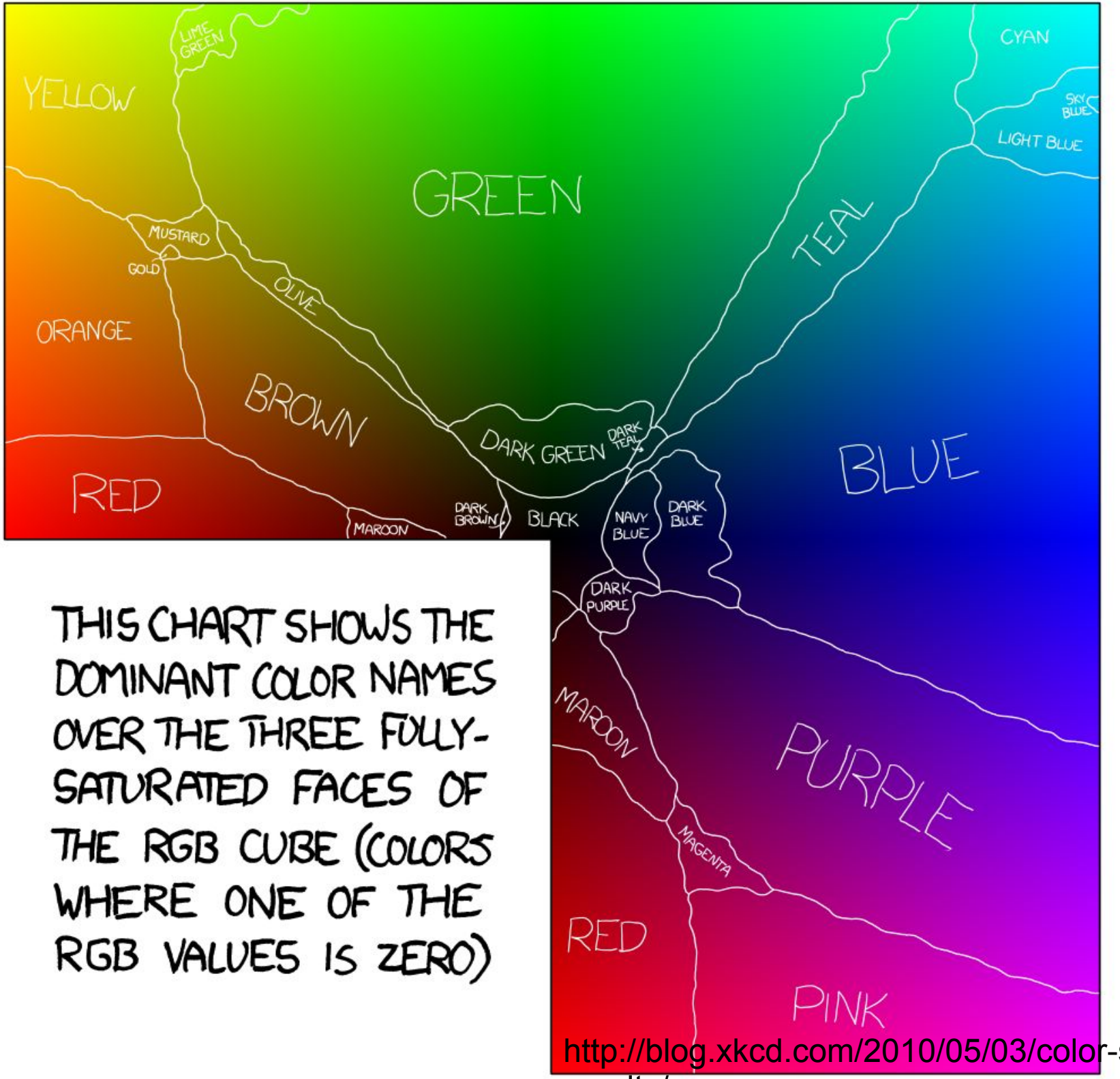


spectral colours: wavelength (nm)

KNOWLEDGE MEDIA INSTITUTE

Colour





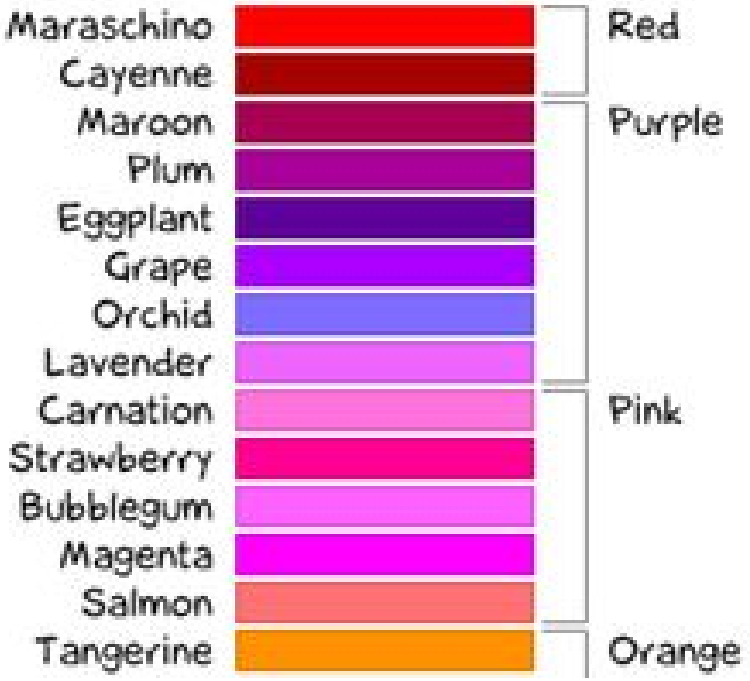
THIS CHART SHOWS THE DOMINANT COLOR NAMES OVER THE THREE FULLY-SATURATED FACES OF THE RGB CUBE (COLORS WHERE ONE OF THE RGB VALUES IS ZERO)

http://blog.xkcd.com/2010/05/03/color-survey-

KNOWLEDGE MEDIA INSTITUTE

Colour: subjective

Color names if you're a girl...

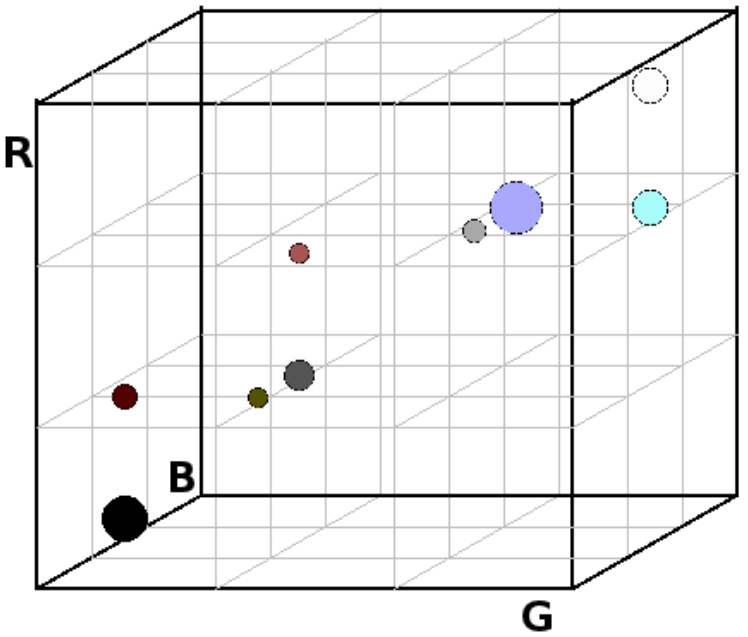


Color names if you're a guy...

http://eagereyes.org/blog/2011/you-only-see-colors-you-can-name

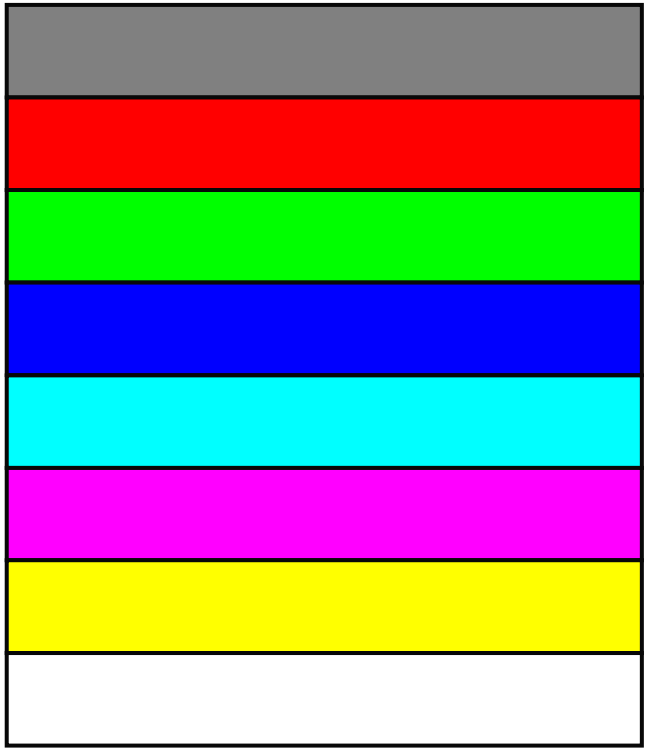
KNOWLEDGE MEDIA KMi INSTITUTE

Colour histogram



Exercise

a) 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

b) Sketch 2D colour histograms per channel

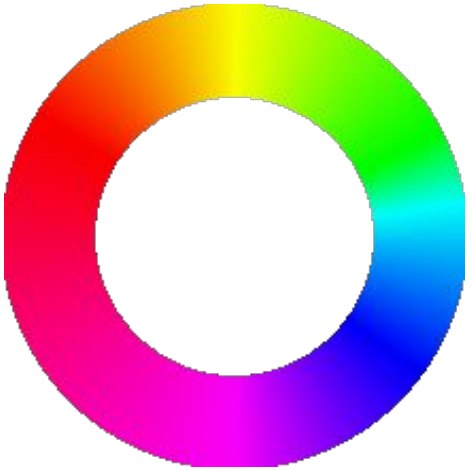
c) Which one is more informative? Why?

Other Colour Spaces

HSV, HSL, CIELAB/CIELUV

KNOWLEDGE MEDIA KMi INSTITUTE

HSB colour model



saturation (0% - 100%)
= spectral purity

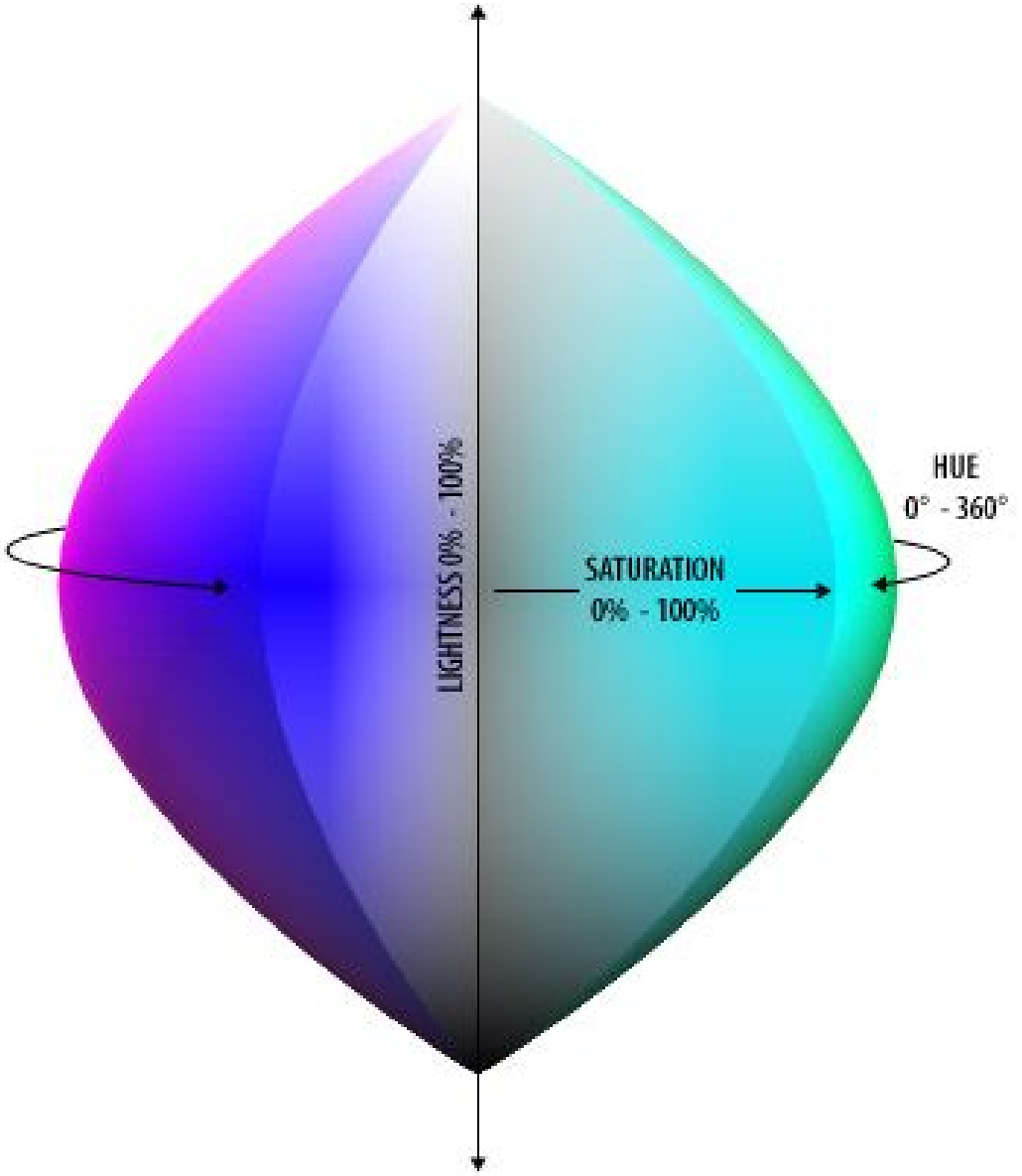
hue (0°-360°)
spectral colour

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

chromaticity = hue+saturation

KNOWLEDGE MEDIA INSTITUTE

HSB colour model



The Open University

HSB model

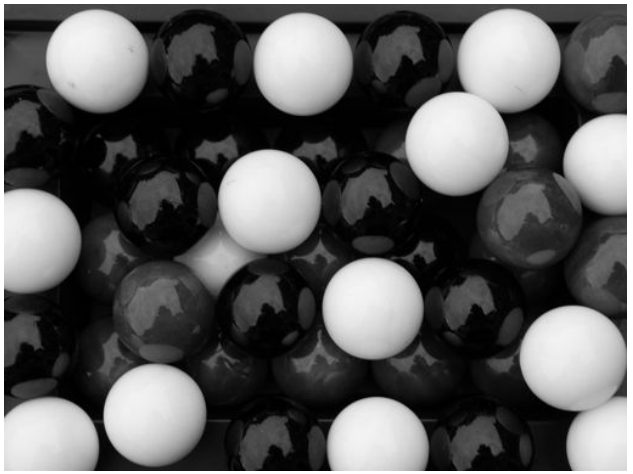
disadvantage: hue coordinate is not continuous

- 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%)

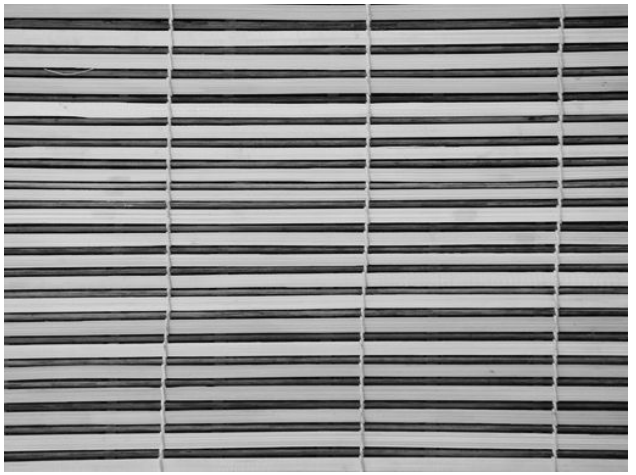
advantage: it is more natural to describe colour changes “brighter blue”, “purer magenta”, etc

Texture

coarseness



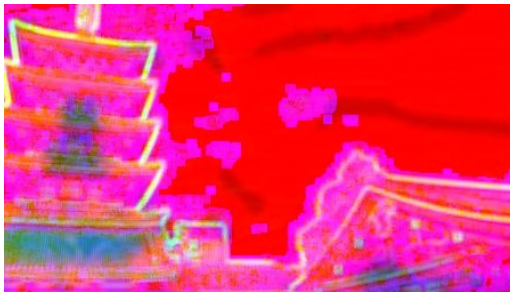
contrast

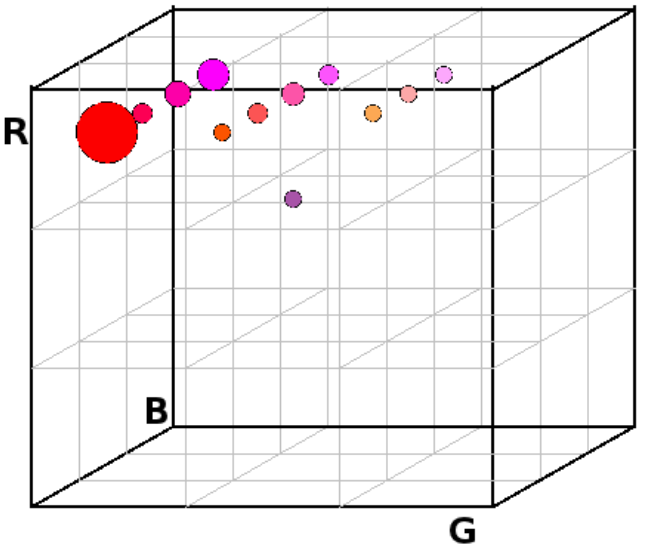


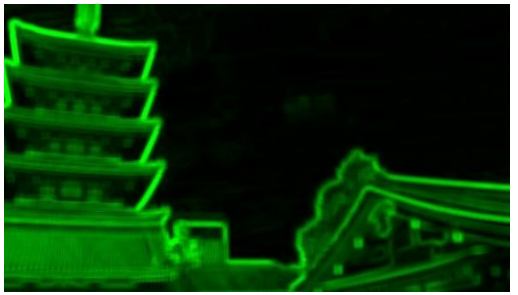
directionality

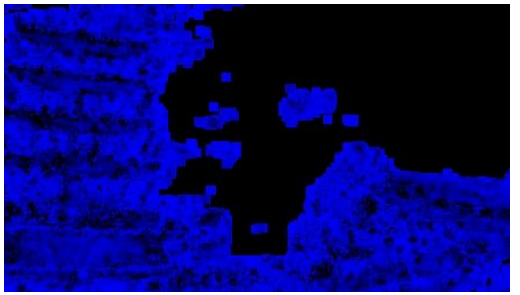
[Tamura et al, 1978]

Texture histograms









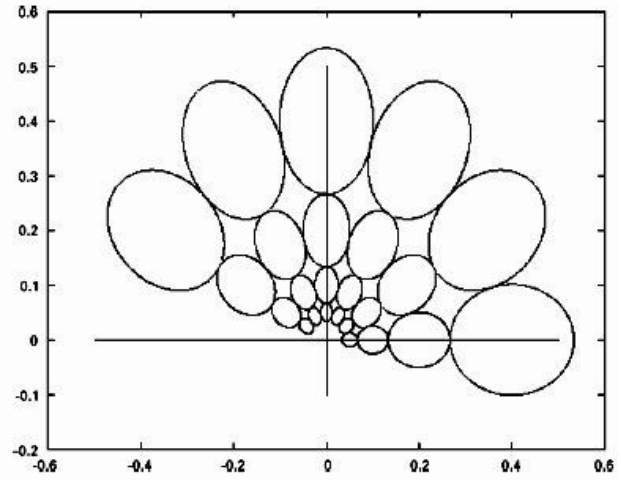
Coarseness coNtrast Directionality

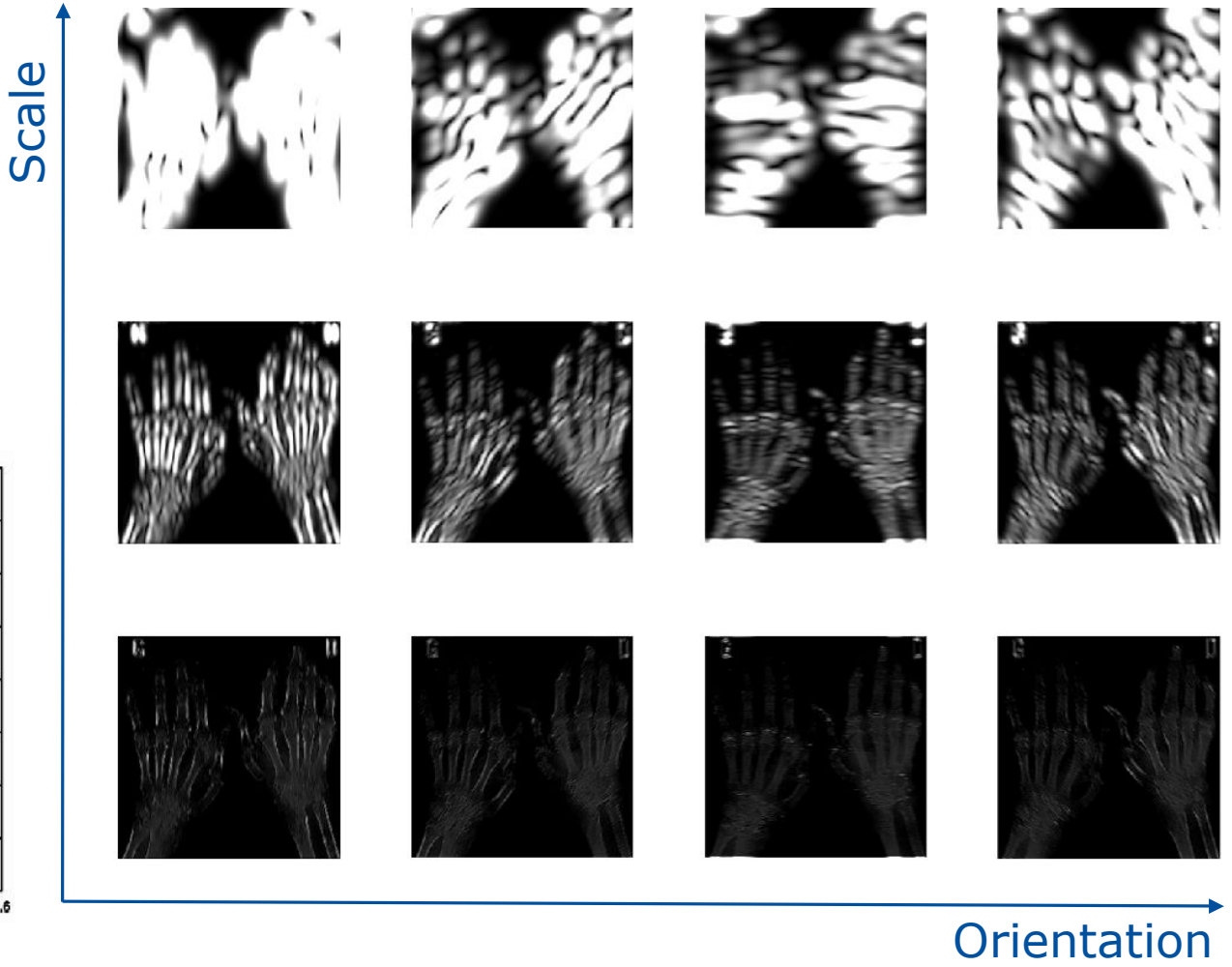
[with Howarth, *IEE Vision, Image & Signal Proc* 15(6) 2004; Howarth PhD thesis]

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)

Particularly important for drwings, eg, patent drawings

KNOWLEDGE MEDIA KMi INSTITUTE

MAREC research collection and CLEF-IP

Reserach corpus of 19m **patents**: XML docs + tif images

CLEF-IP track 2011

- Image-based prior art search
- Image classification

KNOWLEDGE MEDIA KMi INSTITUTE

MAREC subset

212,867 patents (34k EP; 167k US; 12k WO)
2,586,767 tifs as attachments (min 1 per patent)

A43B characteristic features of footwear;
parts of footwear

A61B diagnosis; surgery; identification

H01L semiconductor devices;
electric solid state devices not otherwise provided for

The Open University

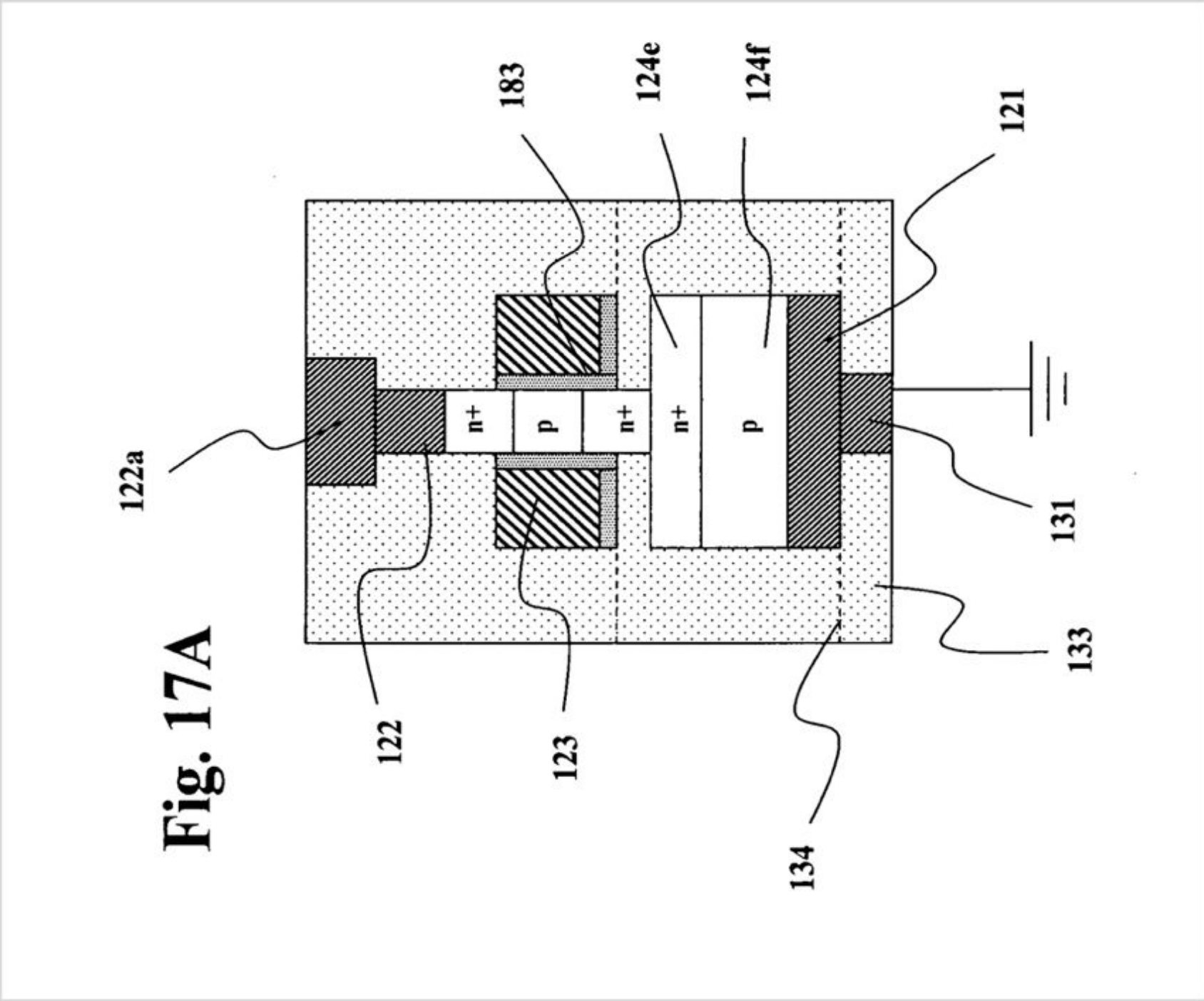


Fig. 17A

Randomly picked images (zooming out)

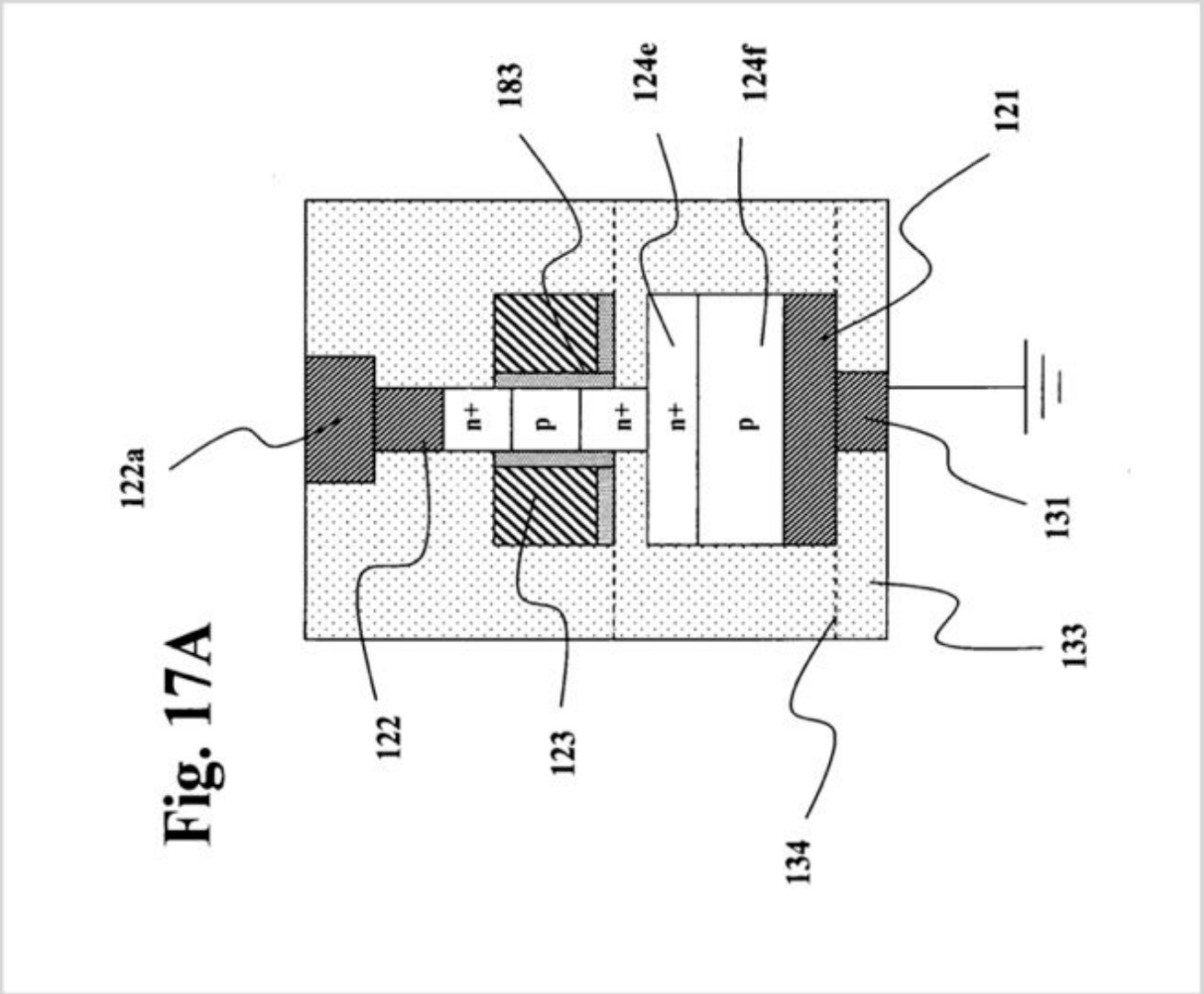
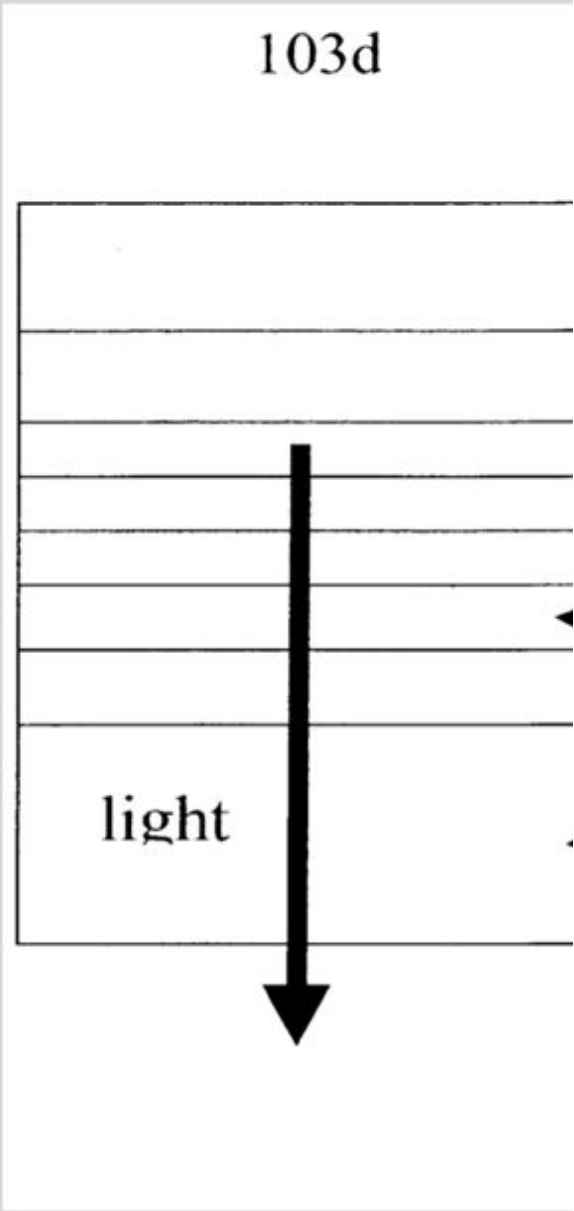


Fig. 17A



Randomly picked images (zooming out)

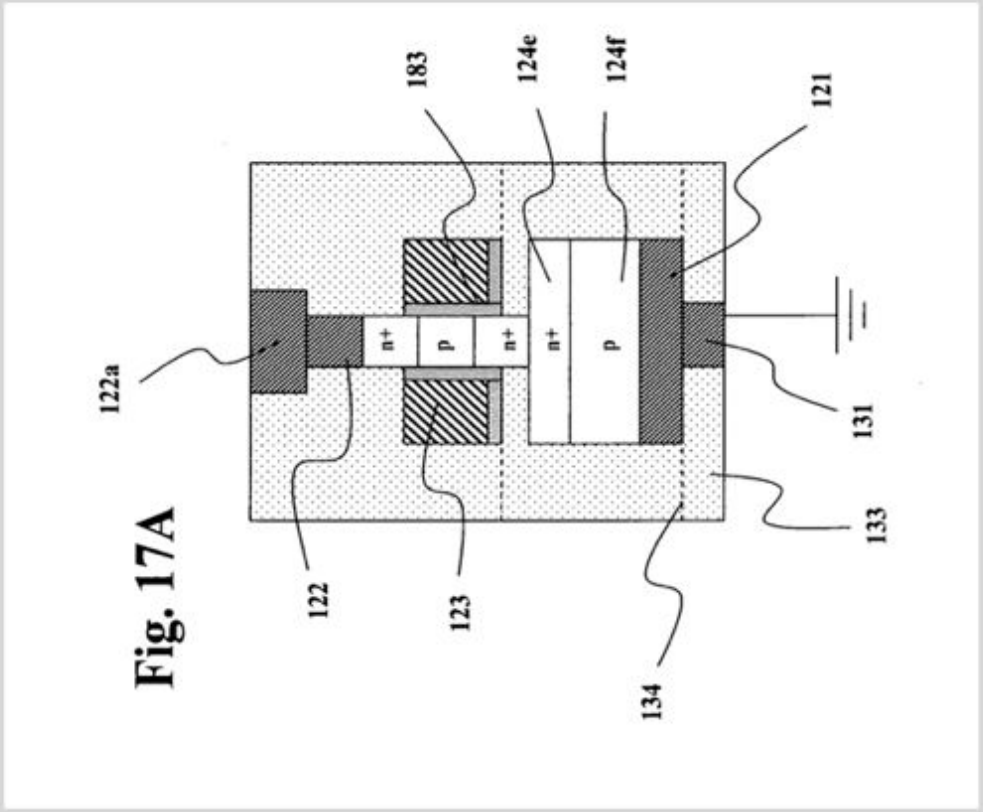
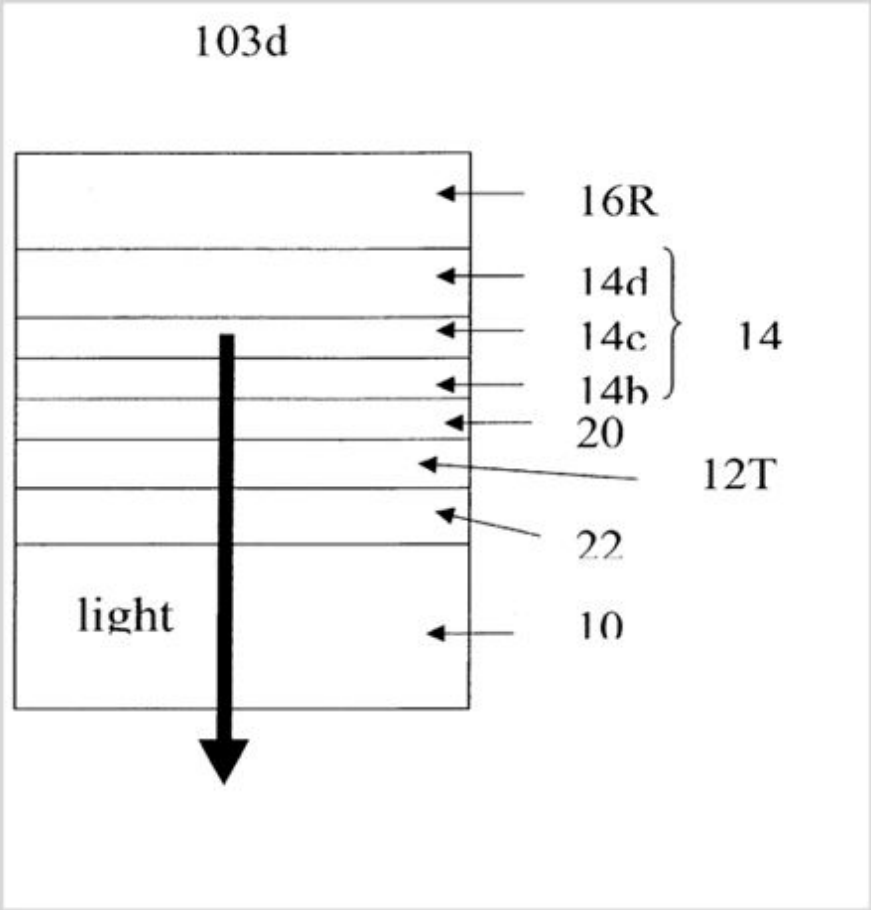
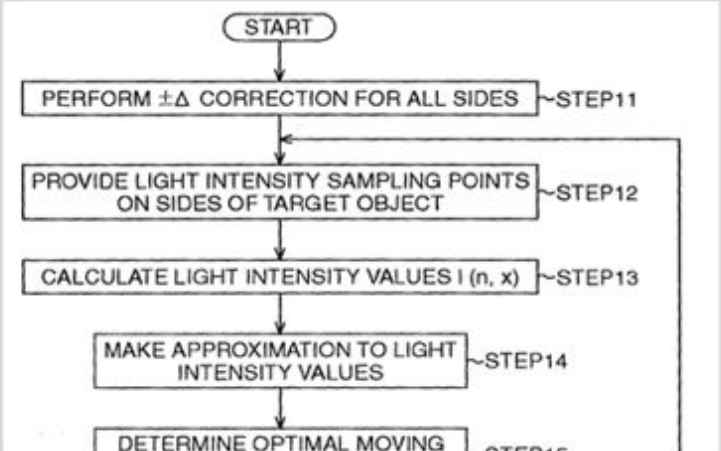
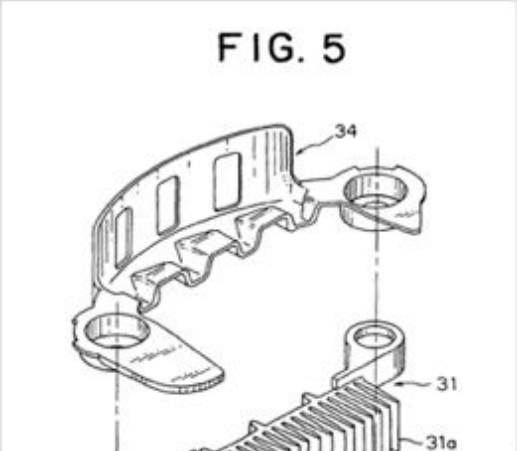


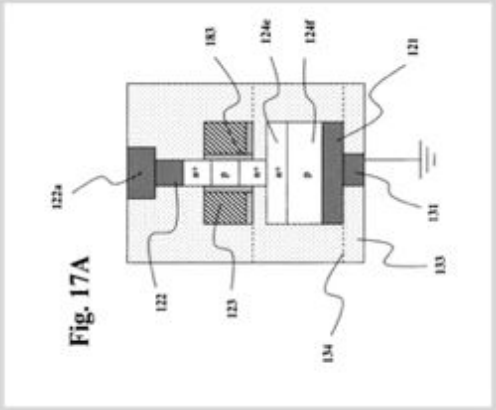
Fig. 17A

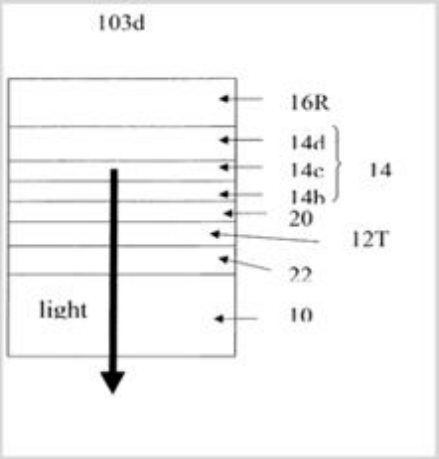




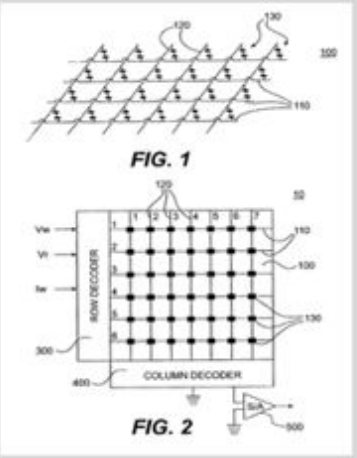


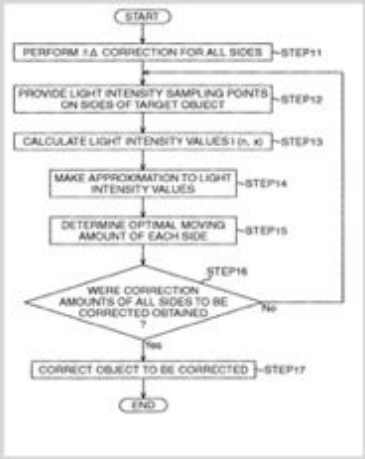
Randomly picked images (zooming out)

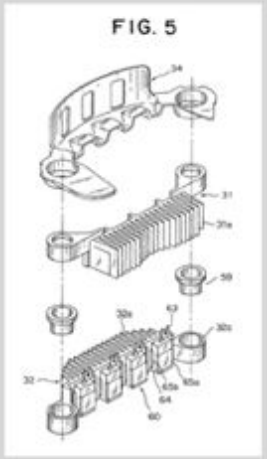


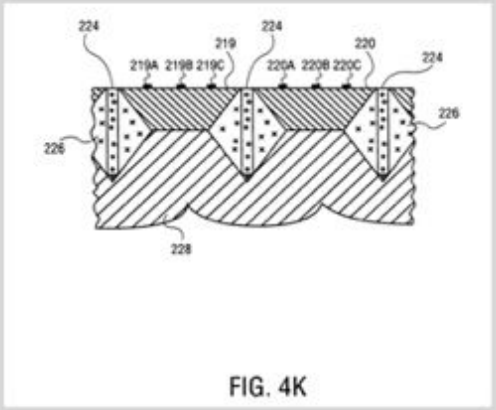


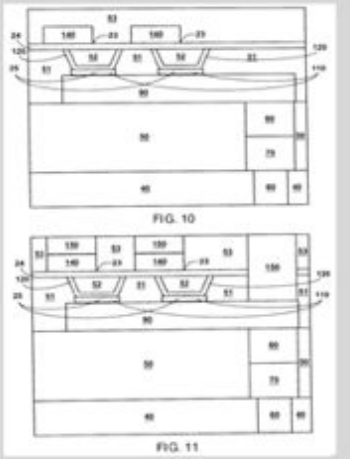
$$f_i[n] = \begin{cases} f[n], & n_{i-1} < n < n_i \\ 0, & \text{otherwise} \end{cases}$$

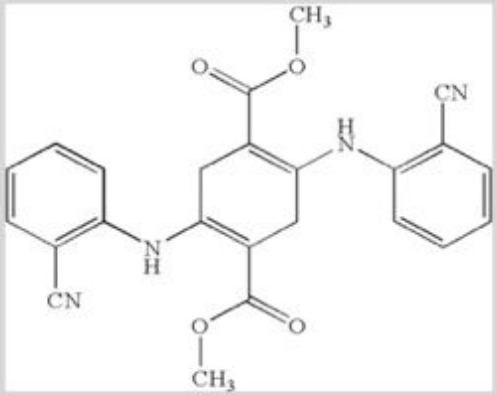


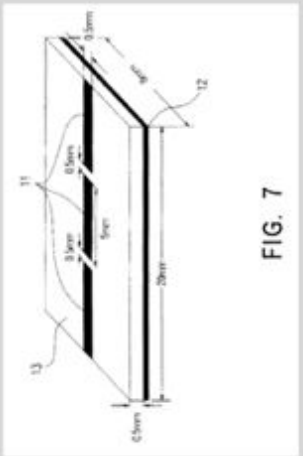




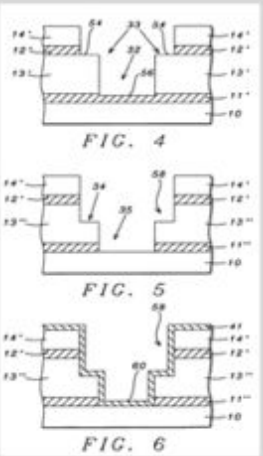




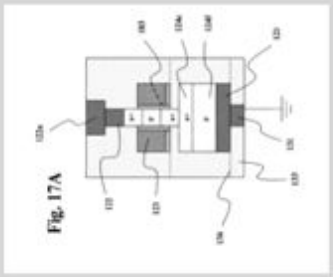


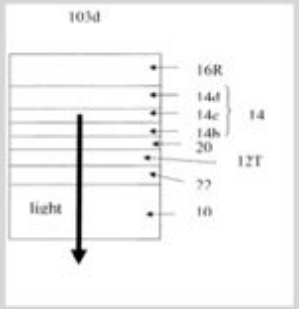


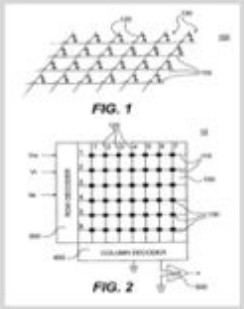
$$D_c \approx \frac{4\mu_r}{\mu_0 M_s^{\,2}} \left(\frac{2kT_c|K|}{a} \right)^{1/2}$$

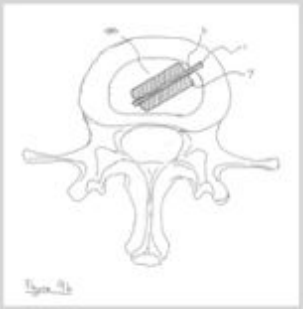


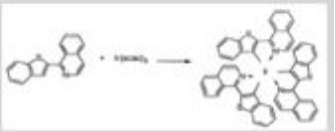
Randomly picked images (zooming out)

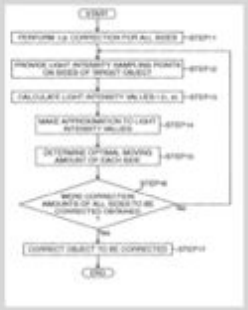


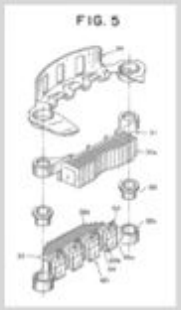


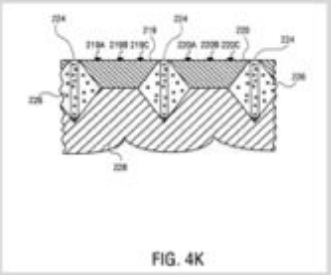


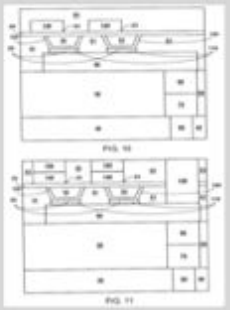


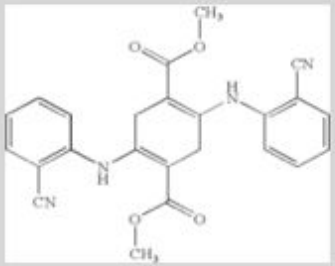


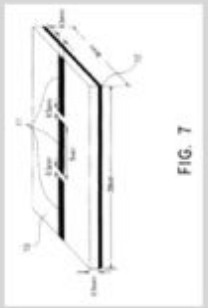




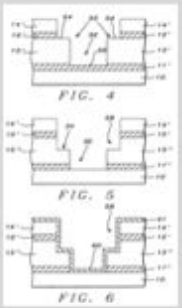


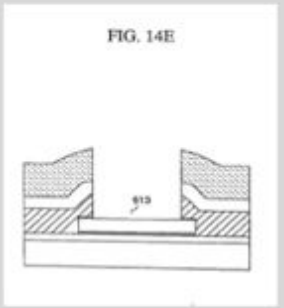


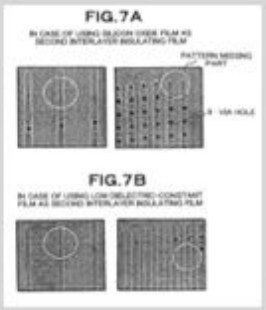


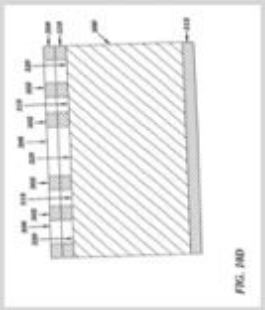


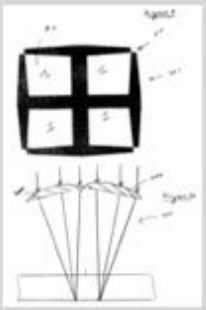
$$D_c \approx \frac{4\mu_r}{\mu_0 M_s^{\,2}}\left(\frac{2kT_c|K|}{a}\right)^{1/2}$$

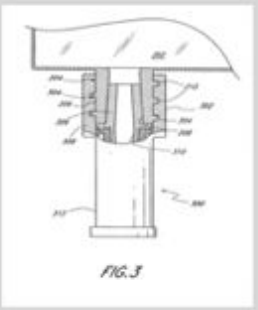


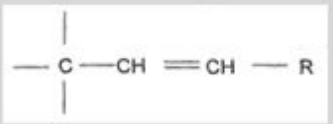


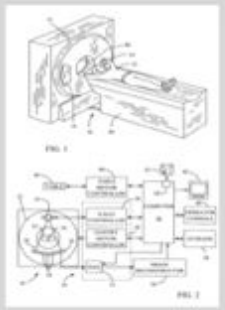


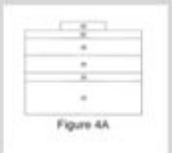


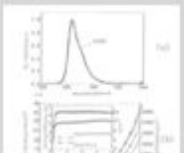


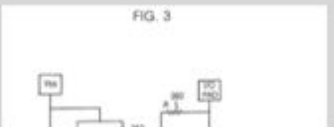




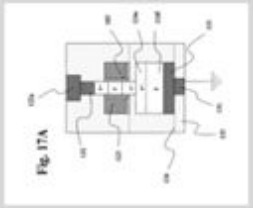


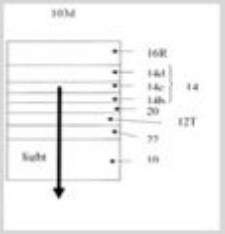


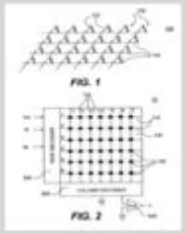


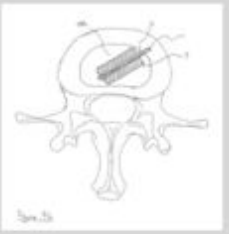


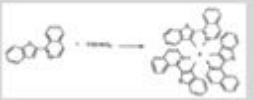
Randomly picked images (zooming out)

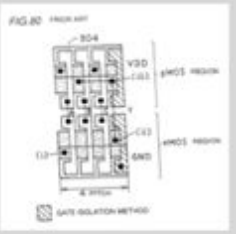


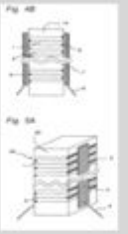


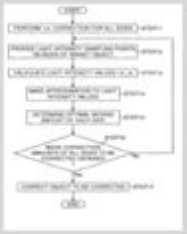


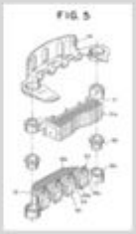


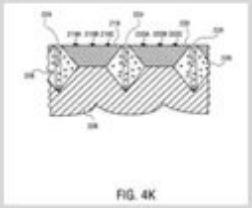


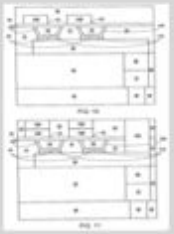


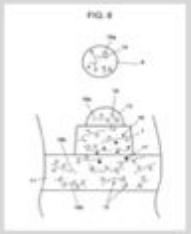


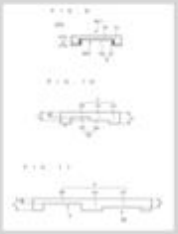


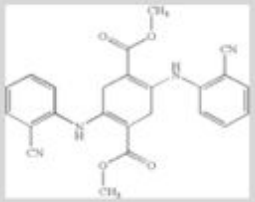


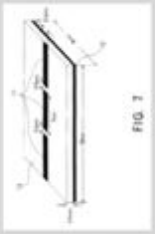




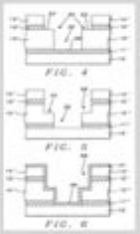


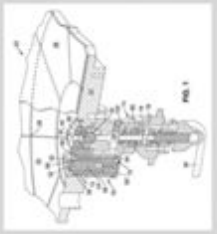


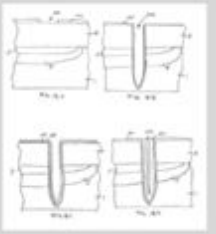


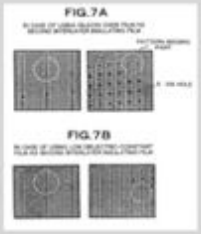


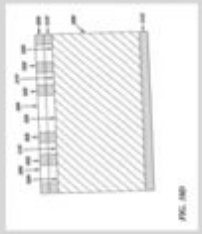
$$D_c \approx \frac{4\mu_c}{\mu_0 M_s^2}\left(\frac{2kT_c|K|}{a}\right)^{1/2}$$

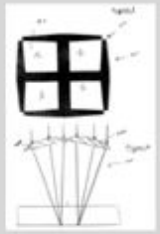


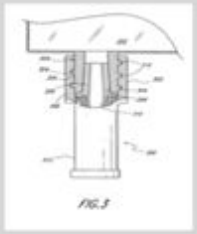


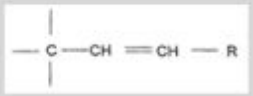


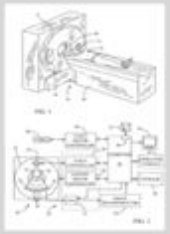


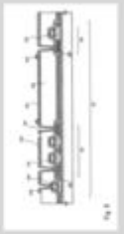


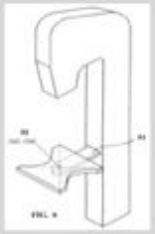


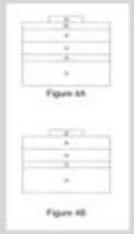


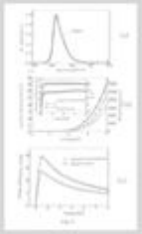


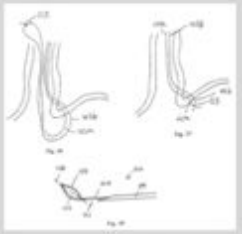


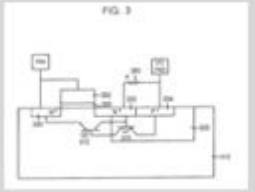


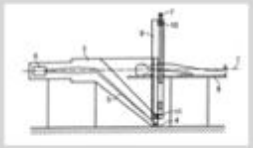




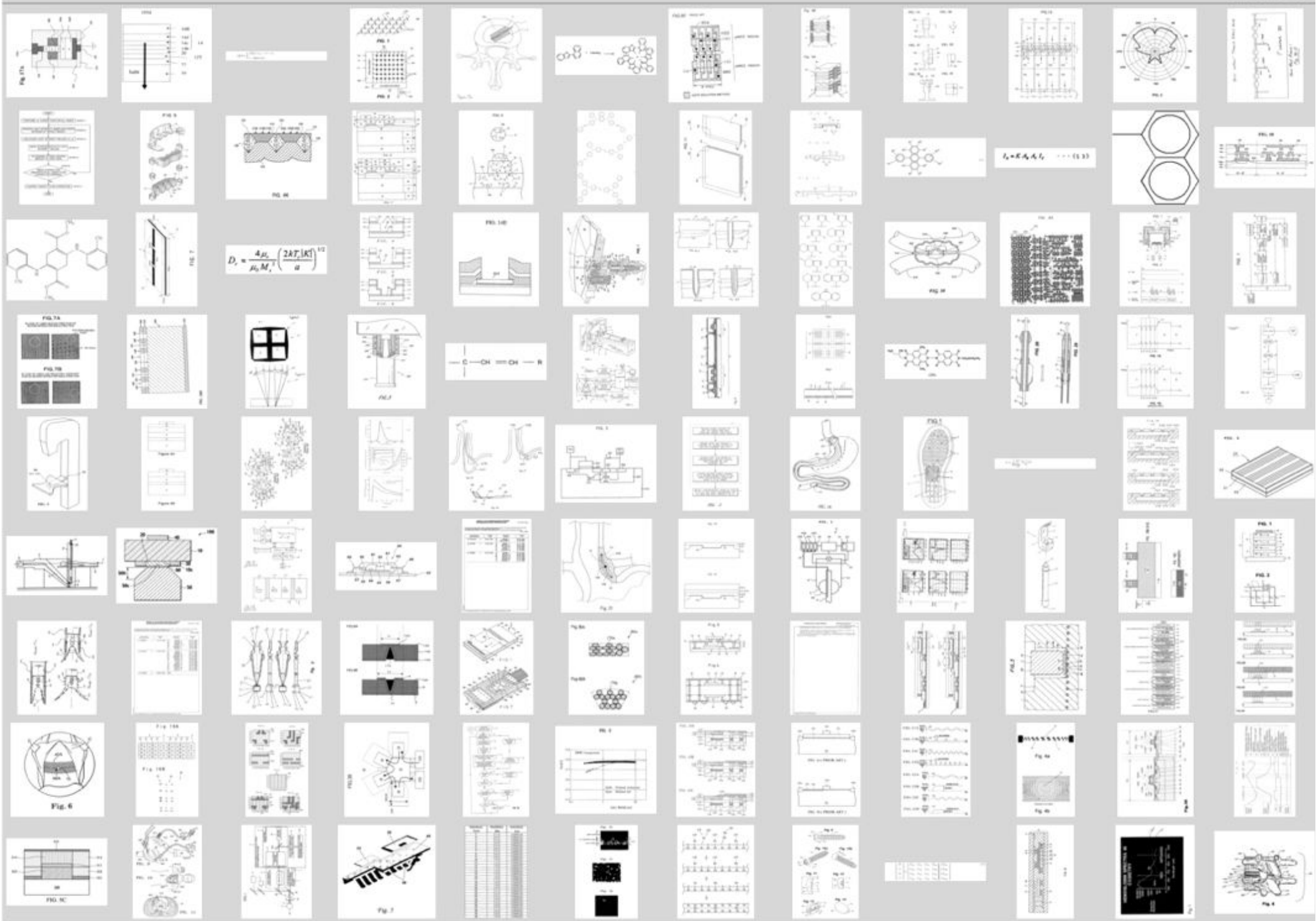






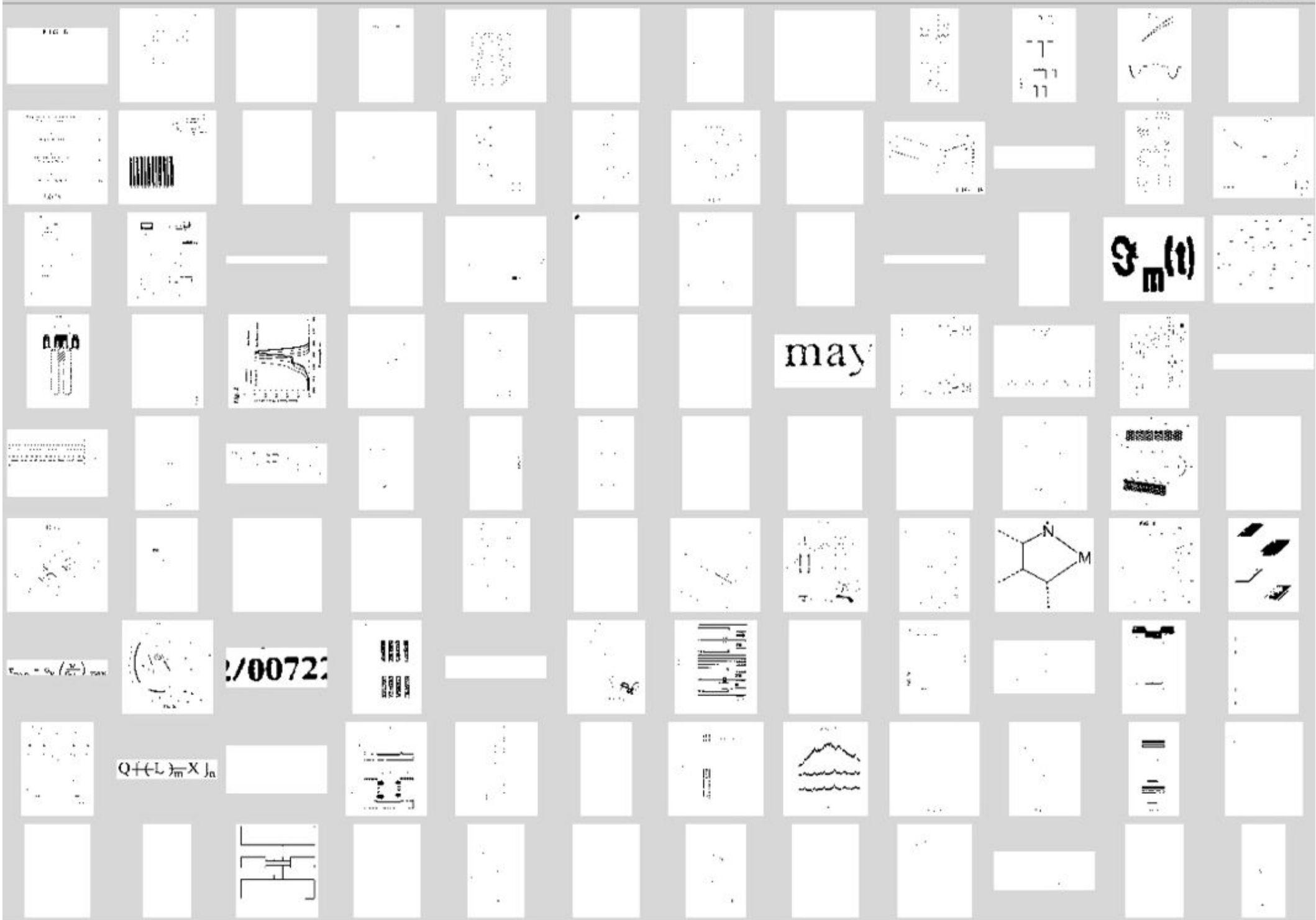


Randomly picked 108 images

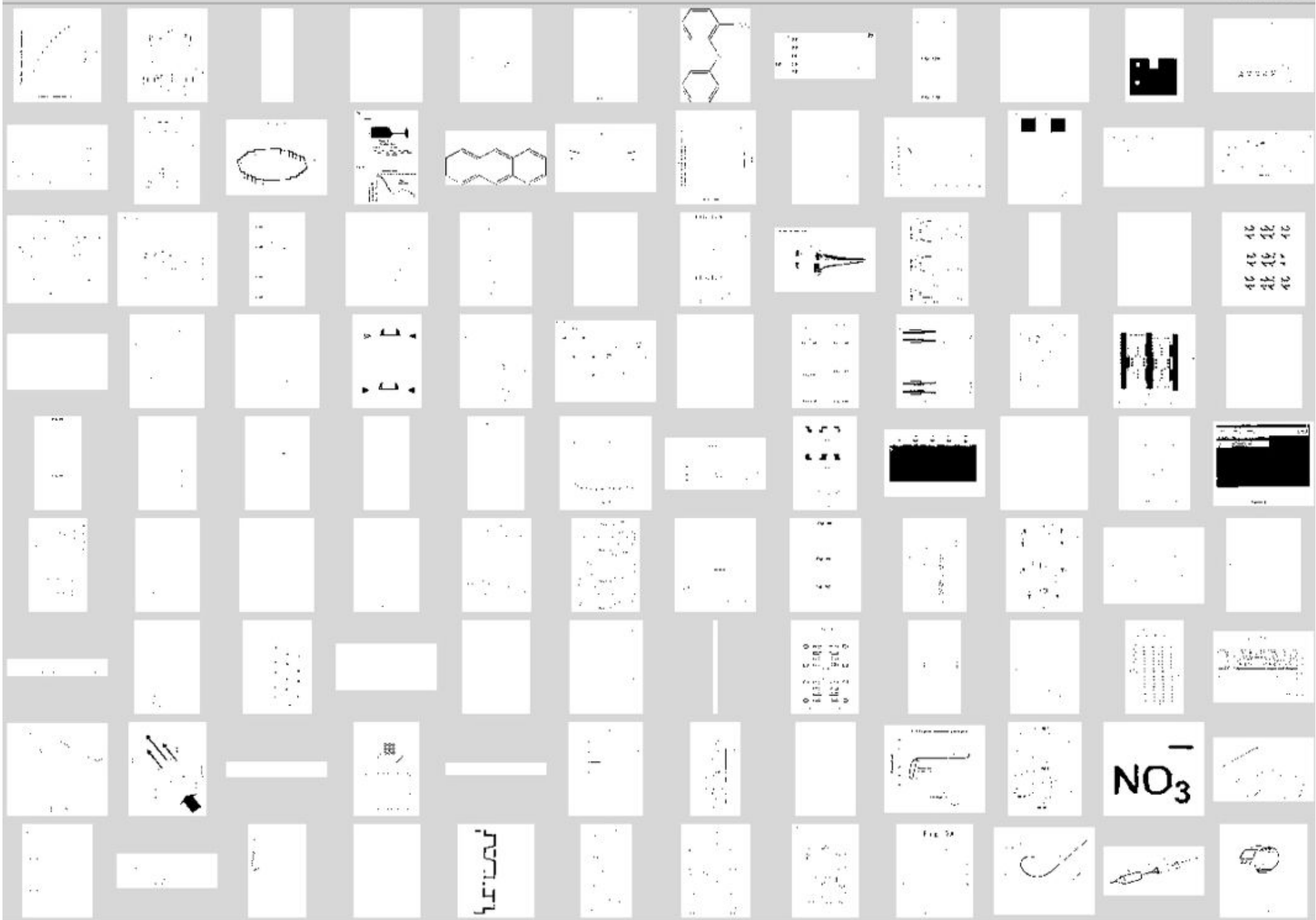


Randomly picked images (another set of 108)

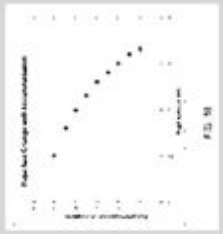
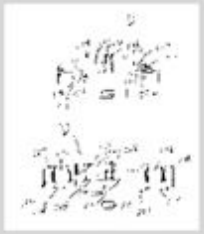
Randomly picked images (3rd set of 108)



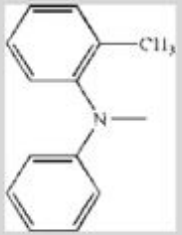
Randomly picked images (last set of 108)

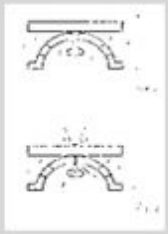
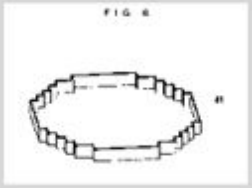


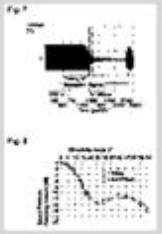
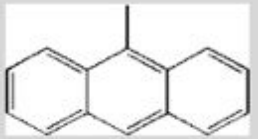
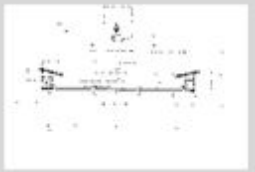
Randomly picked images (zooming in)

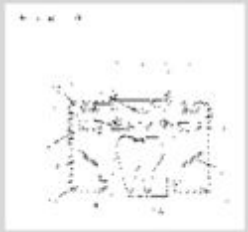
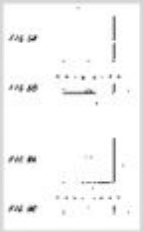


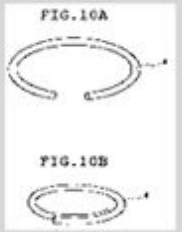




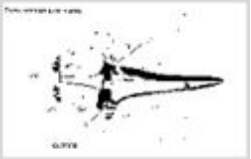




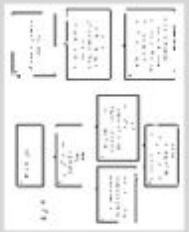


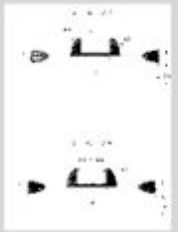
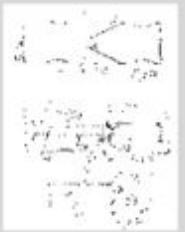
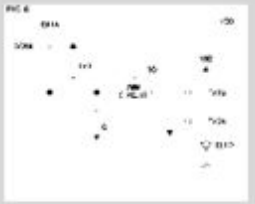



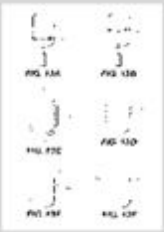






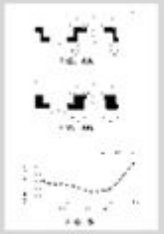




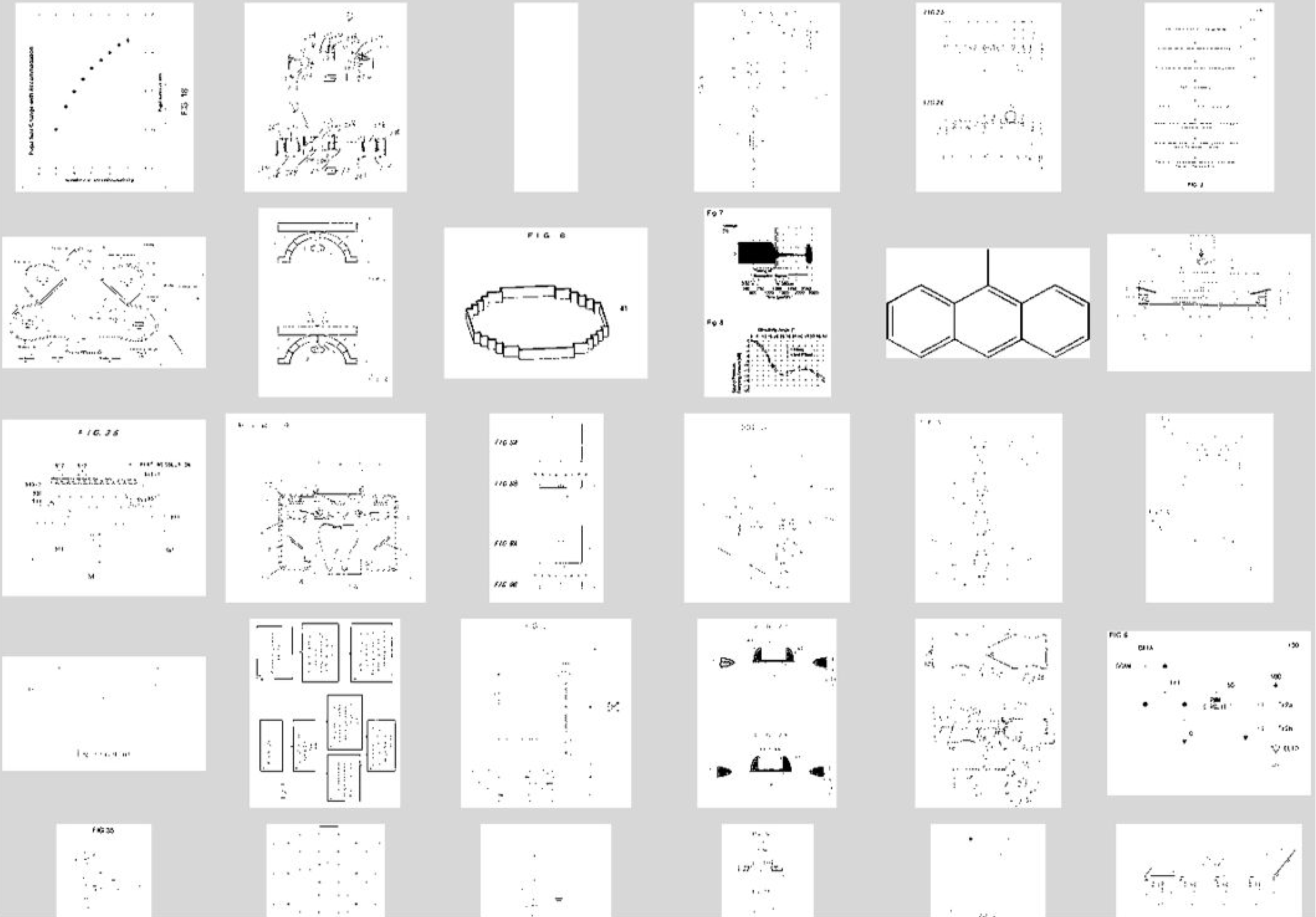




Randomly picked images (zooming in)

random-d-08-a



Randomly picked images (zooming in)

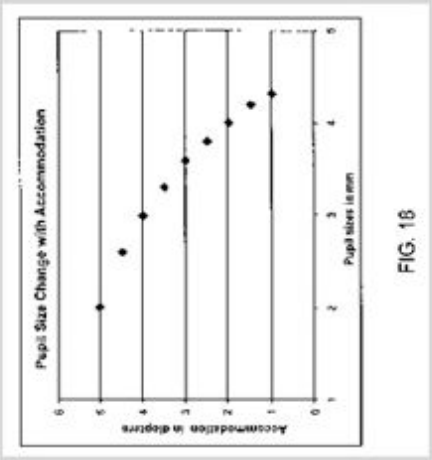
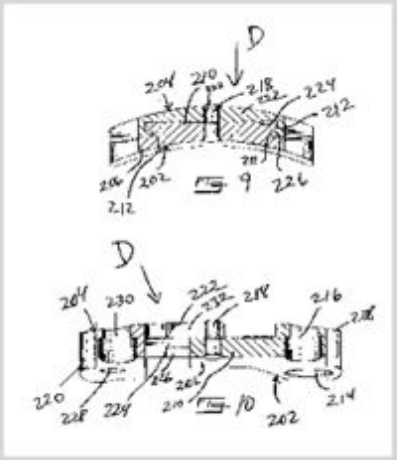


FIG. 18



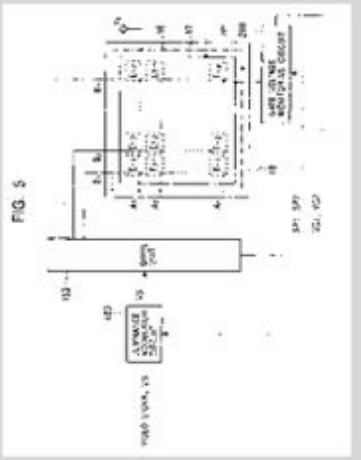
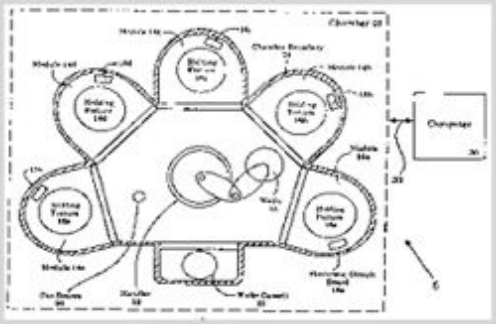


FIG. 5



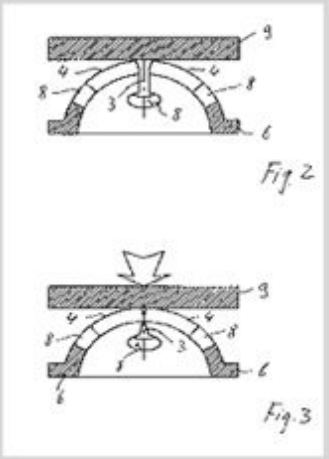
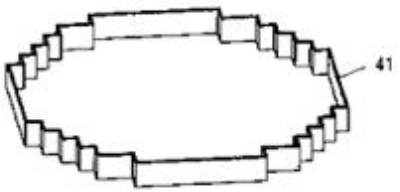
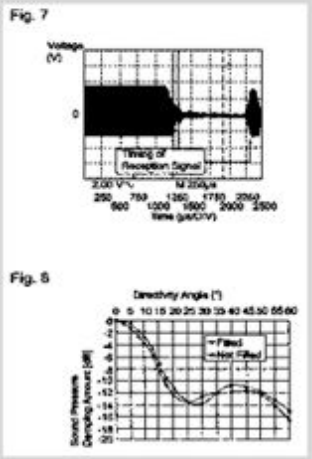


Fig. 2

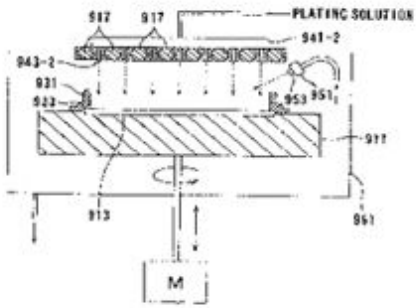
Fig. 3

F I G. 8

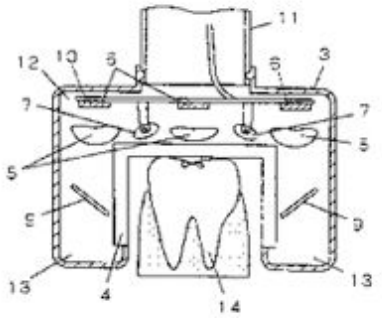




F I G. 3 5



F i g. 4



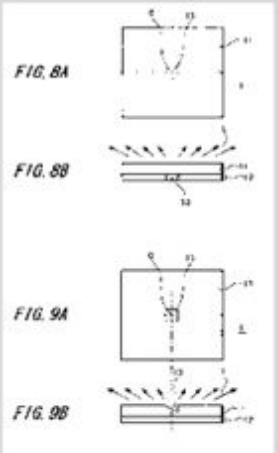
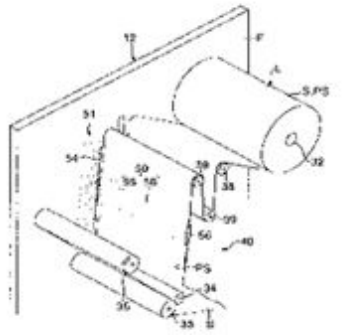


FIG. 5



Randomly picked images (zooming in)

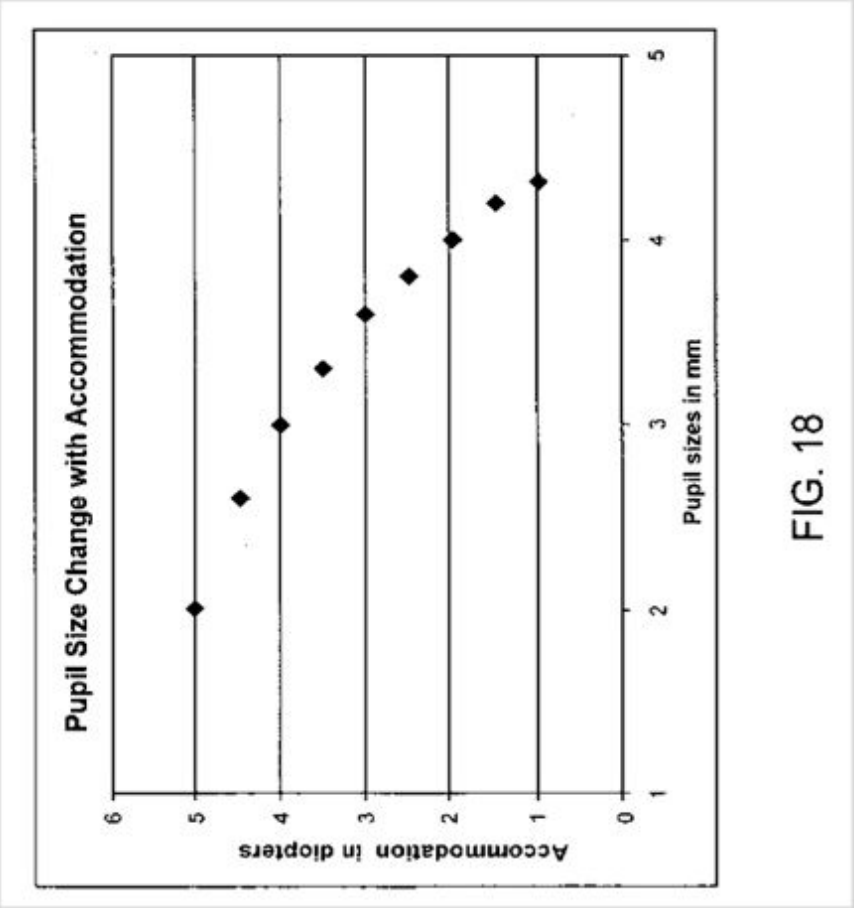
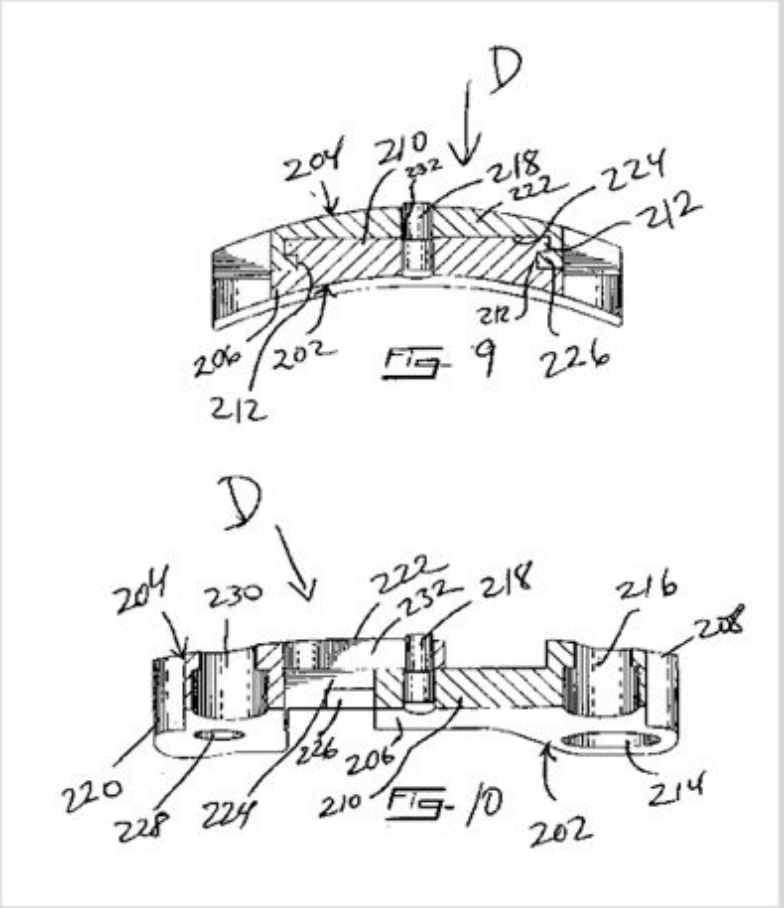
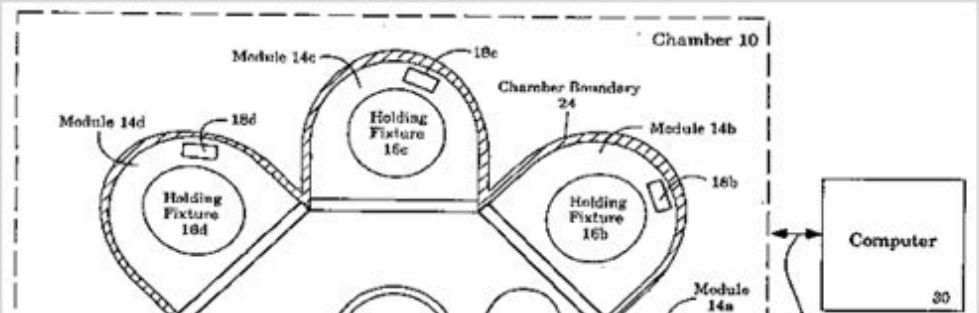


FIG. 18





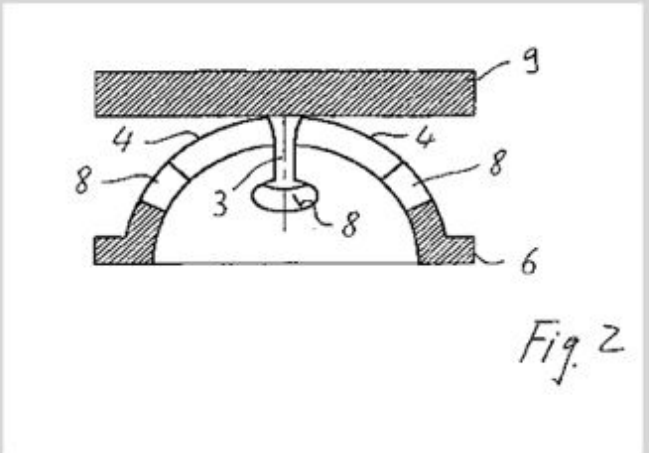


Fig. 2

Randomly picked images (zooming in)

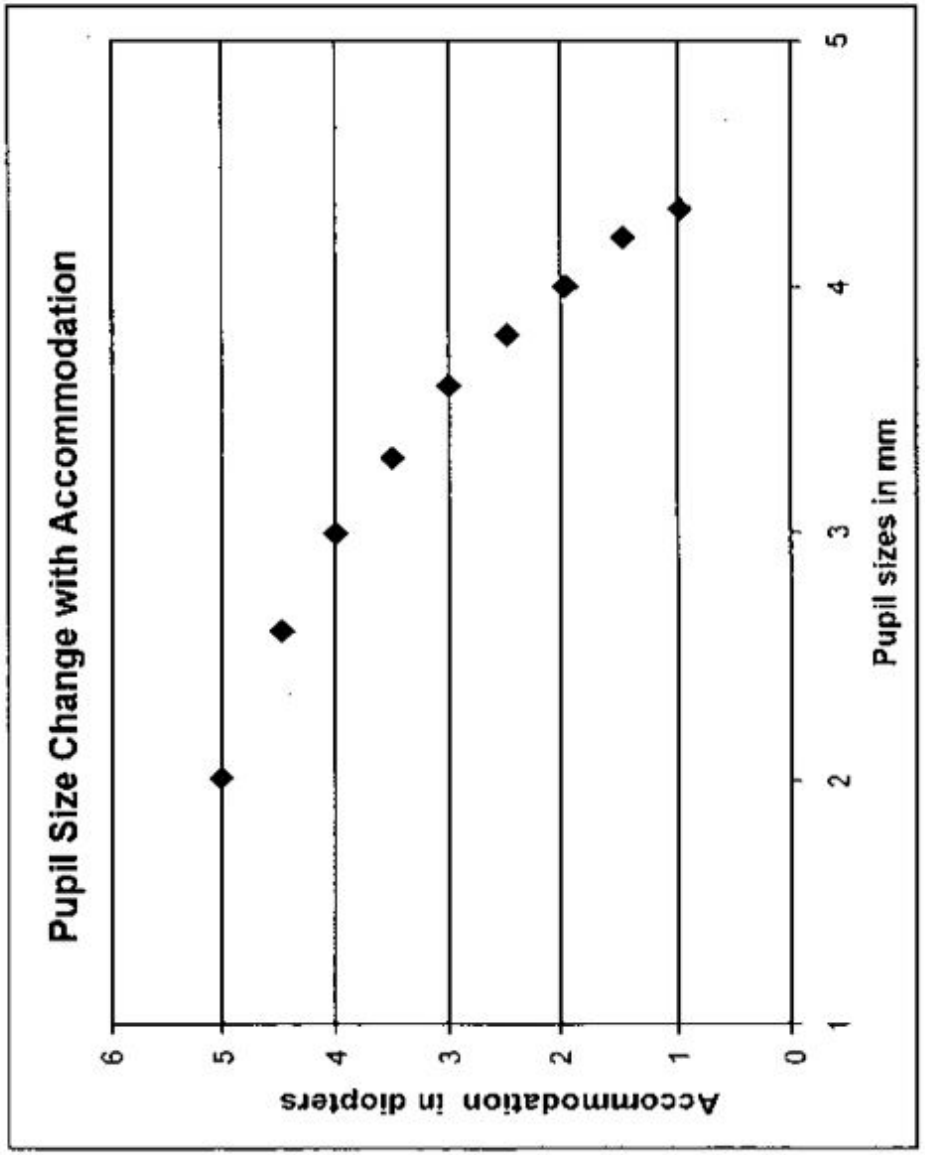
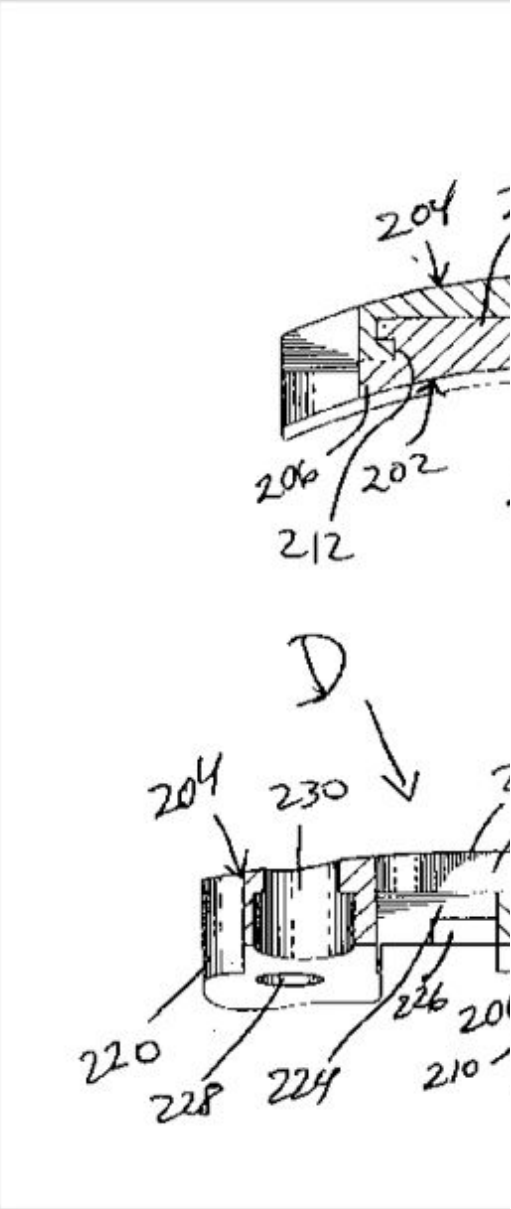


FIG. 18



Another random image

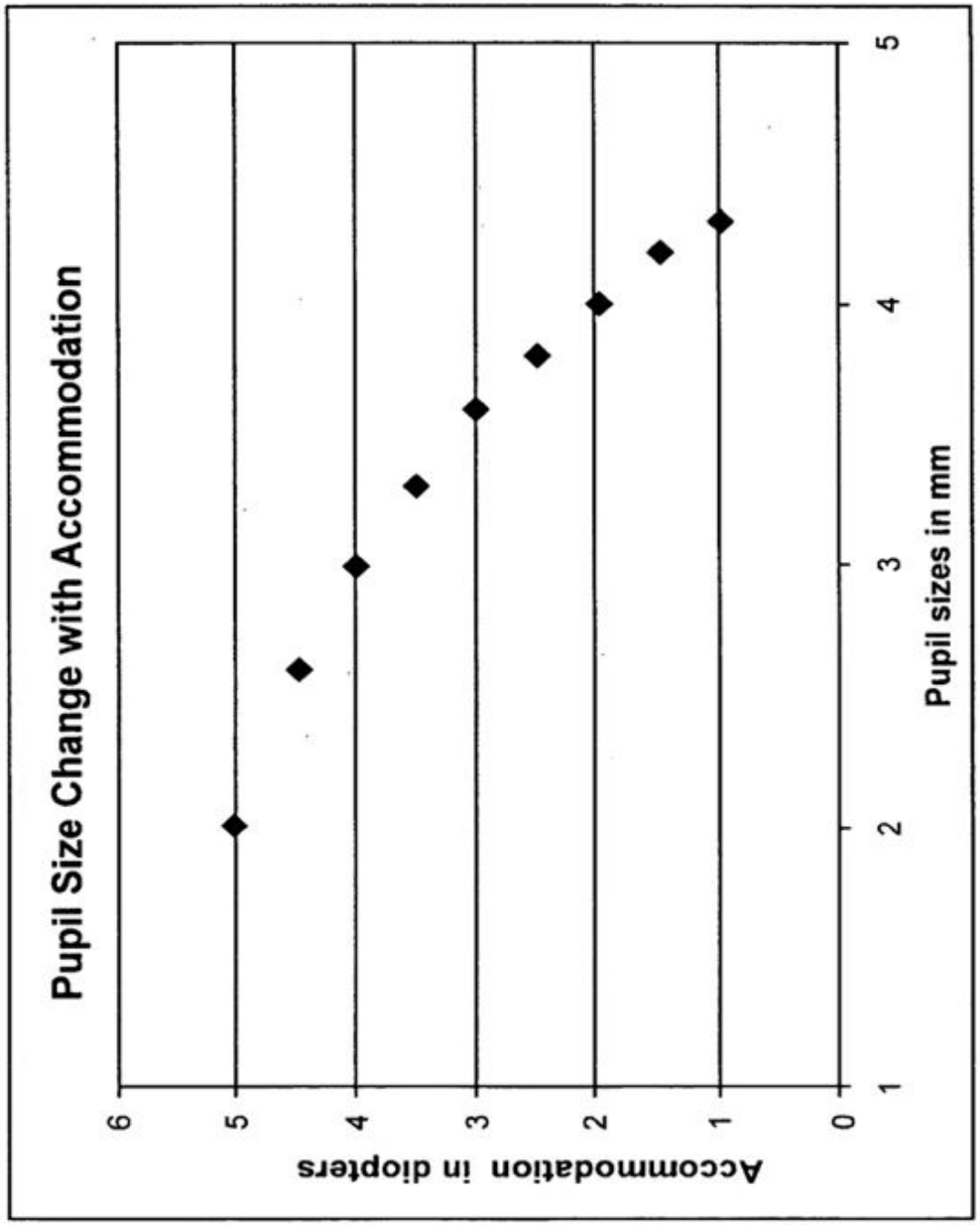


FIG. 18

KNOWLEDGE MEDIA INSTITUTE

What have we seen?

All black/white 1-bit bitonal
Different orientations, aspect ratio, sizes
Range of complexity (a few letters to hand-drawn sketches)
Range of different types – predominantly

- Abstract drawings
- Graphs
- Flow charts
- Gene sequences
- Program listings
- Symbols
- Chemical structures
- Tables
- Mathematics

KNOWLEDGE MEDIA KMi INSTITUTE

Ideal world?

- Semantically marked up entities (text & drawings obsolete)
  - Chemical Markup Language (CML)
  - Scalable Vector Graphics (SVG)
  - Flowchart markup language?
  - Tables and graphs?

KNOWLEDGE MEDIA INSTITUTE

Automated processing

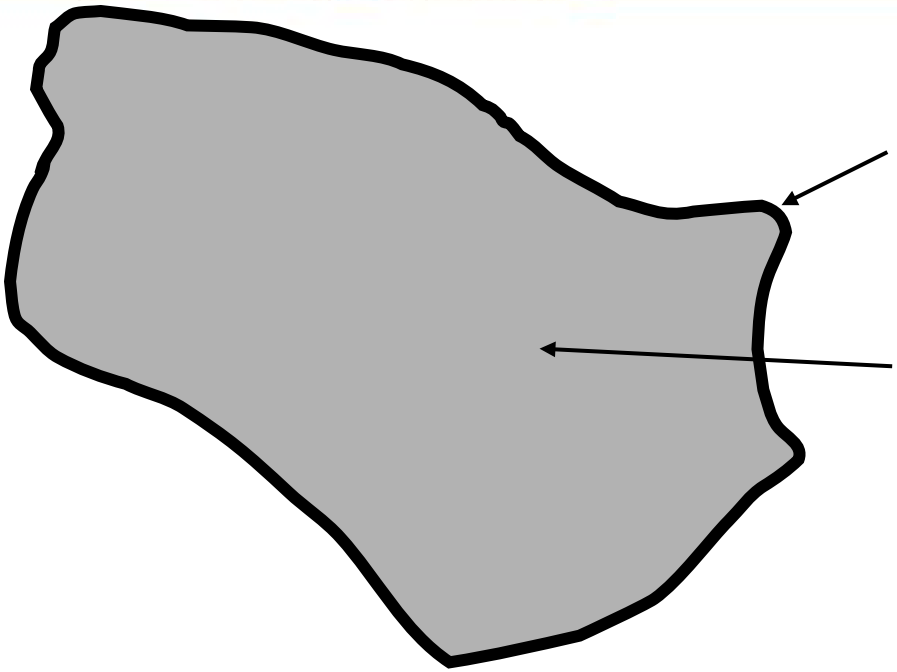
Need dedicated features for bitonal images!
What is the analogon for SIFT-type of special features?
Predict near duplicate detection is possible (and useful!)

Automated extraction of structure from scans?
  Word spotting

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

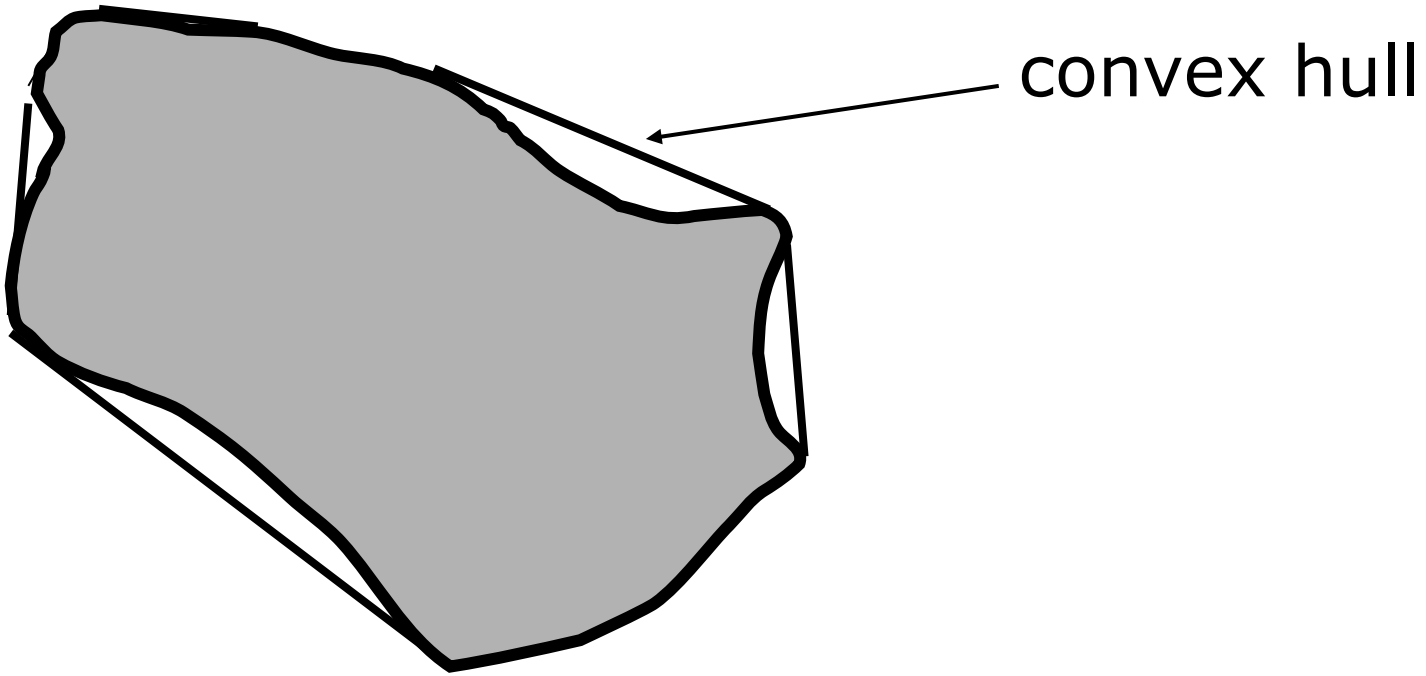
T is smaller than 1 for all other shapes

circularity is aka compactness

Convexity

ratio of perimeter of convex hull and the original curve

1 for convex shapes, less than 1 otherwise



Chain codes

contour description by a string of directions

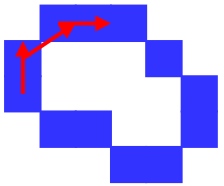
Freeman chain code

$f_1 f_2 f_3 \ldots$

eight possible directions for each pixel

translation invariant, not rotation invariant

3 2 1
4 ■ 0
5 6 7



= 210077654343

= 100776543432

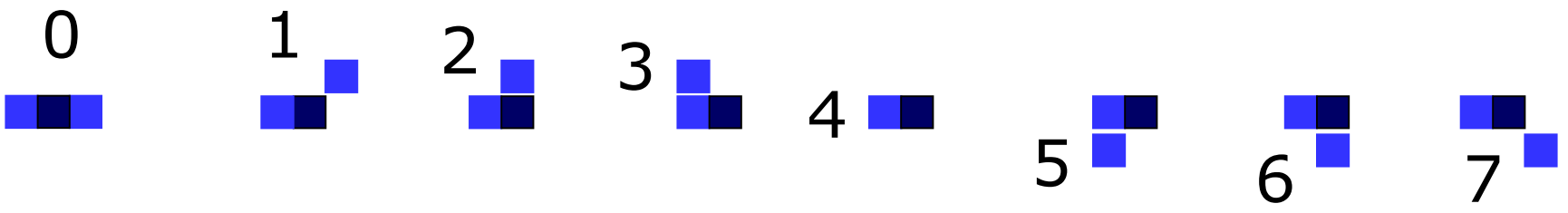
= 00776543432**1**

choose lowest number
for closed contours

KNOWLEDGE MEDIA KMi INSTITUTE

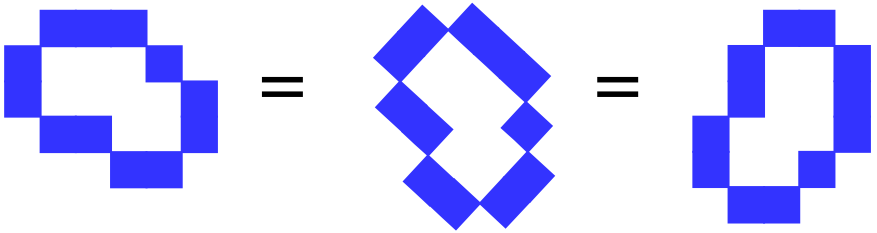
Difference chain codes

sequence of angles: $a_i = (f_{i+1} - f_i) \bmod 8$



0° 45° 90° 135° 180° 225° 270° 315°

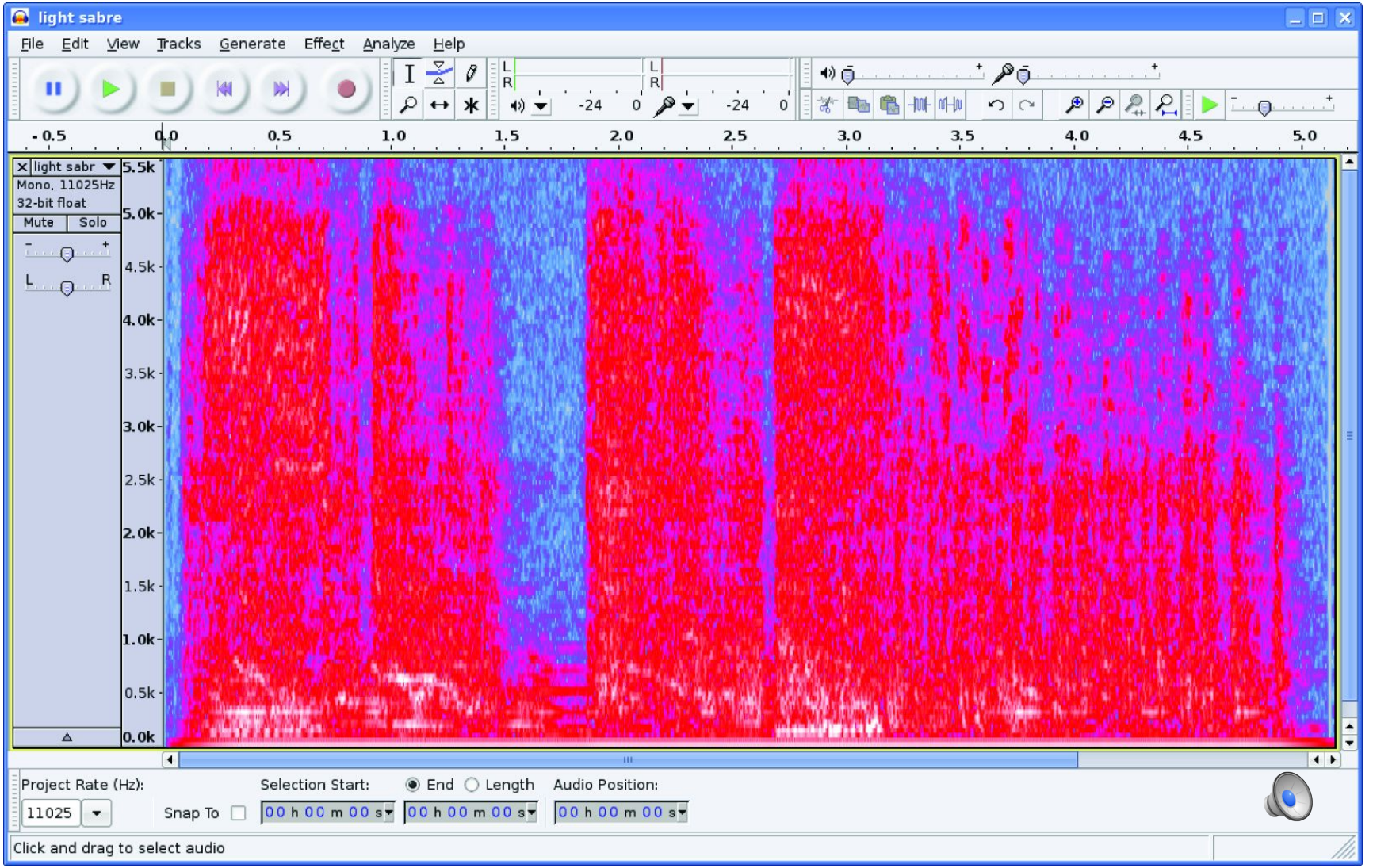
rotation invariant



= ... = 070777717777

KNOWLEDGE MEDIA
KMi
INSTITUTE

Sound



The Open University

KNOWLEDGE MEDIA INSTITUTE

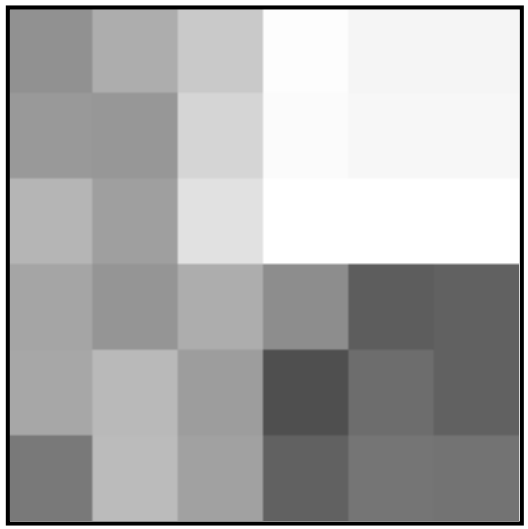
Audio Features

- Spectrogram
  - graph of frequencies/energy/time
- tempo, pitch, mode
- See

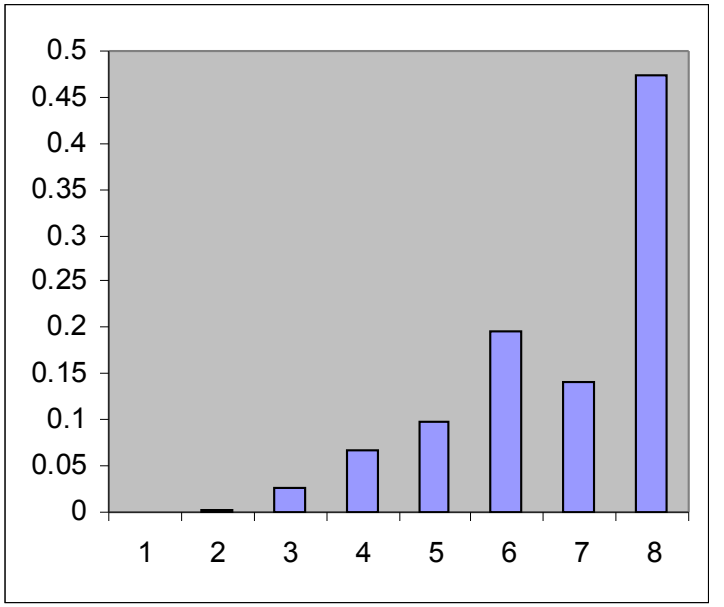
  Z Liu, Y Wang and T Chen (1998). Audio feature extraction and analysis for scene segmentation and classification. *VLSI Signal Processing 20*(1-2), 69-79.

Histograms

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

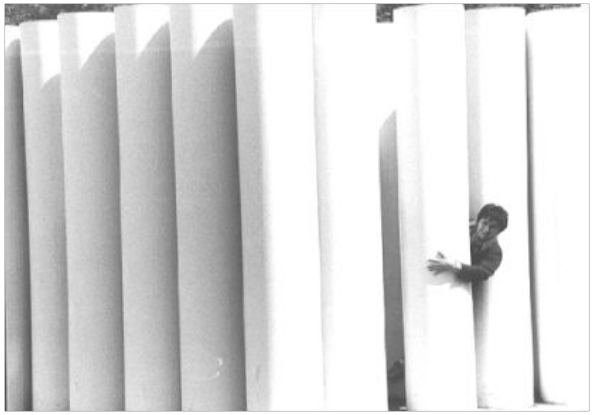
Moment features

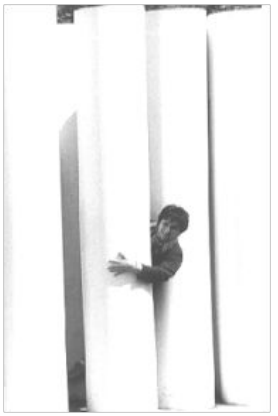
$$(\mu,\ \sqrt{\bar{p}_2},\ \text{sign}(\bar{p}_3)\sqrt[3]{|\bar{p}_3|})$$

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

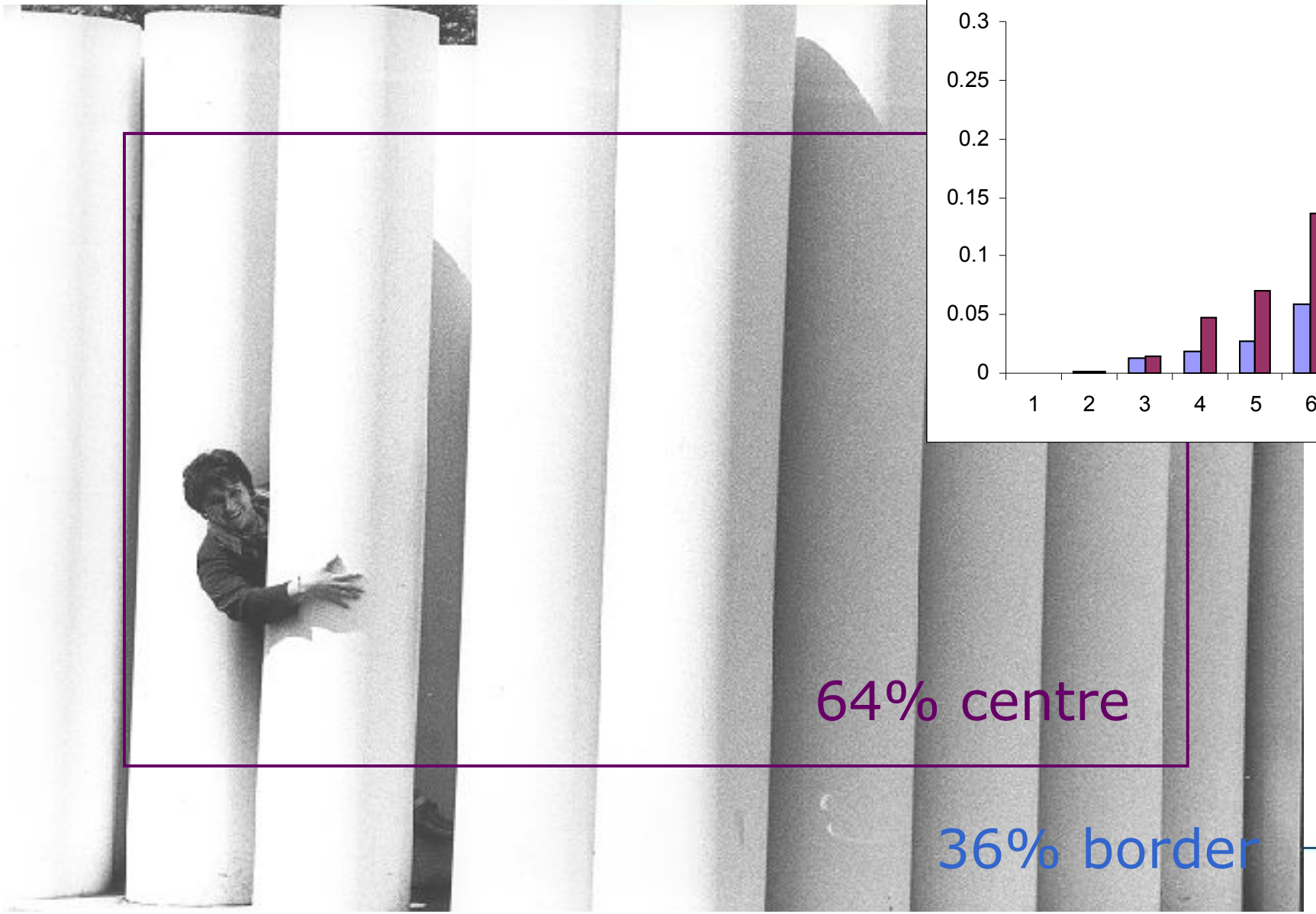




Global histogram also matches polar bears, marble floors, …

KNOWLEDGE MEDIA KMi INSTITUTE

Localisation



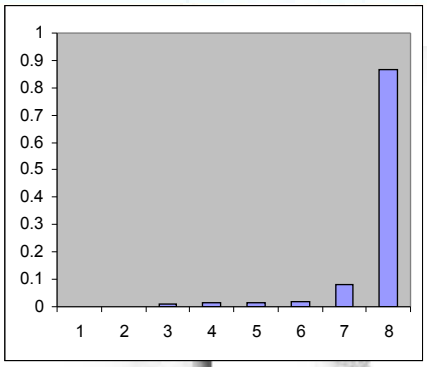
64% centre

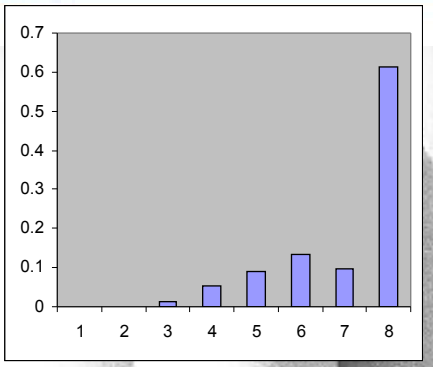
36% border

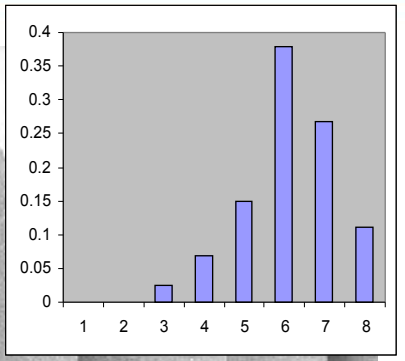
The Open University

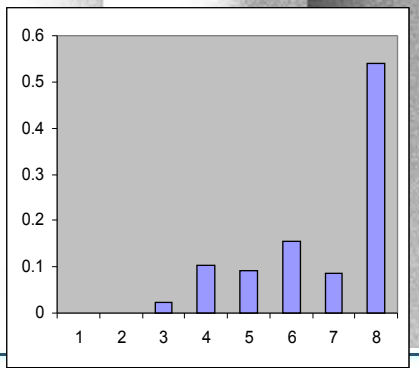
KNOWLEDGE MEDIA
KMi
INSTITUTE

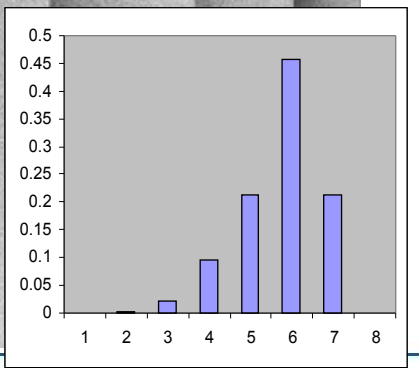
Tiled Histograms







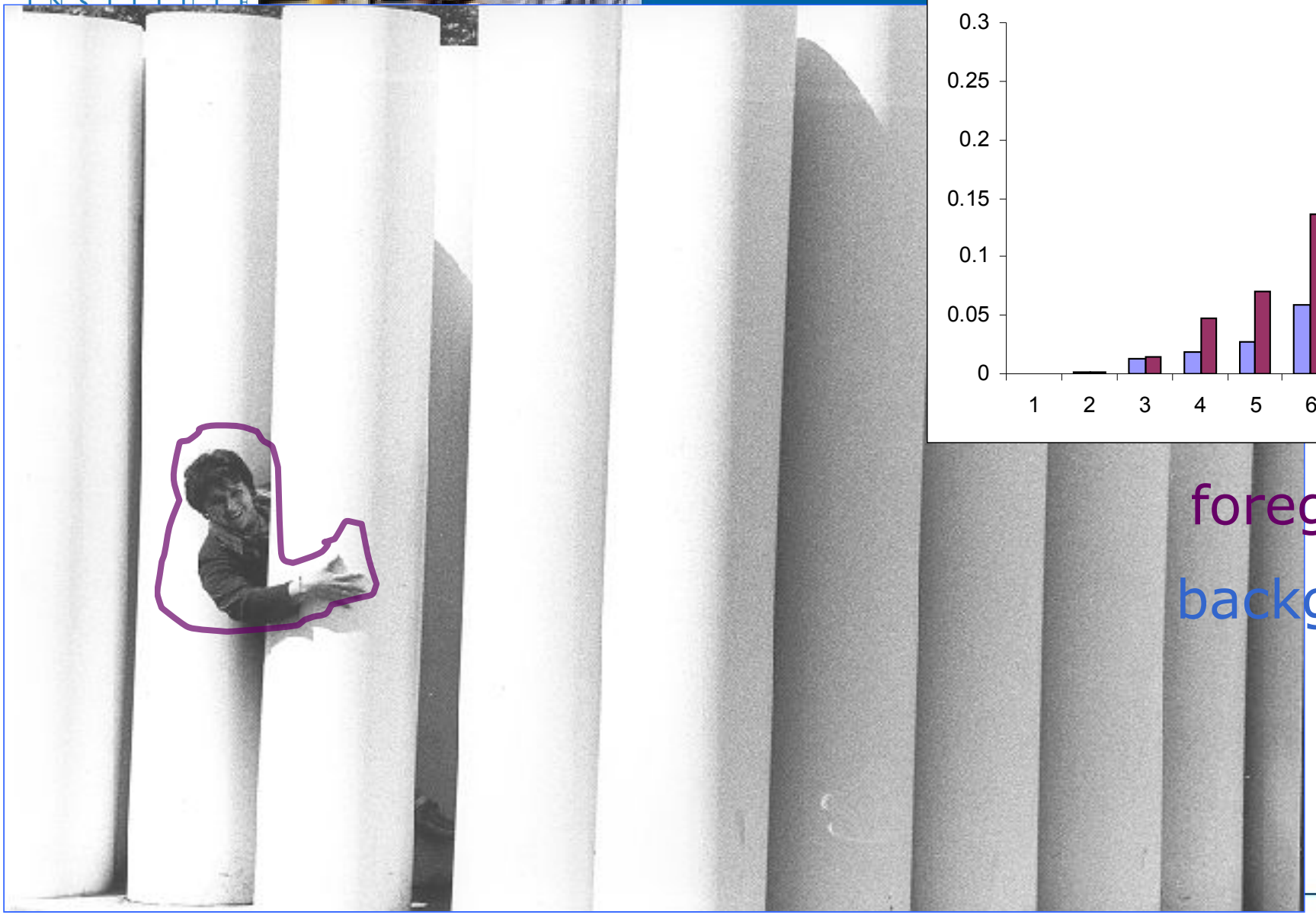




The Open University

KNOWLEDGE MEDIA
KMi
INSTITUTE

Segmentation



foreground

background

The Open University

KNOWLEDGE MEDIA 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 KMi 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

Content-Based Image Retrieval

Which features are best for searching? Depends on the information need:

Looking for sunset holiday pictures in your digital shoebox? Use colour histograms.

Want to build a wallpaper customer database? Use colour + texture.

Want to build a b/w sketch database for technical industrial designs? Use shape descriptors.

Not sure which features are best for a query (eg, if you also have abstract features such as Fourier coefficients)? Deploy **relevance feedback** and **let the system learn** the relevant features for this query...

KNOWLEDGE MEDIA KMi 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