Content Based Image Retrieval
Thanks to John Tait

Outline

- Introduction
  - Why image retrieval is hard
  - How images are represented
  - Current approaches
- Indexing and Retrieving Images
  - Navigational approaches
  - Relevance Feedback
  - Automatic Keywording
- Advanced Topics, Futures and Conclusion
  - Video and music retrieval
  - Towards practical systems
  - Conclusions and Feedback

Scope

- General Digital Still Photographic Image Retrieval
  - Generally colour
- Some different issues arise
  - Narrower domains
    - E.g.Medical images especially where part of body and/or specific disorder is suspected
  - Video
  - Image Understanding - object recognition

Thanks to

- John Tait
- Chih-Fong Tsai
- Sharon McDonald
- Ken McGarry
- Simon Farrand
- And members of the University of Sunderland Information Retrieval Group

Introduction

Why is Image Retrieval Hard ?

- What is the topic of this image
- What are right keywords to index this image
- What words would you use to retrieve this image ?

→ **The Semantic Gap**

Problems with Image Retrieval

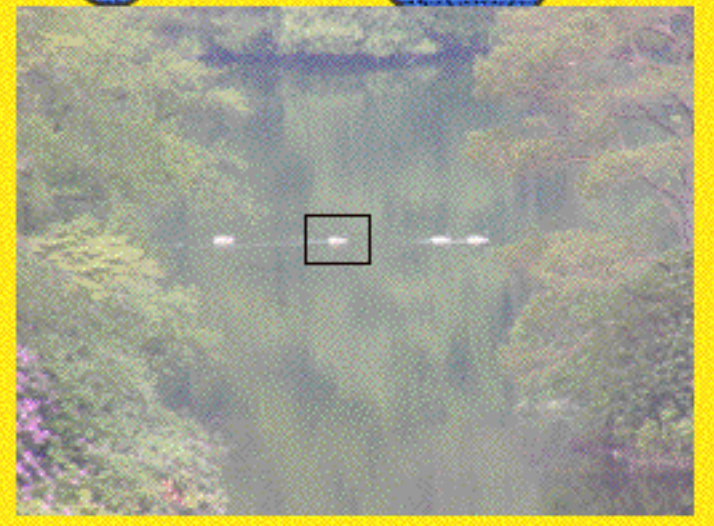
- A picture is worth a thousand words
- The meaning of an image is highly individual and subjective

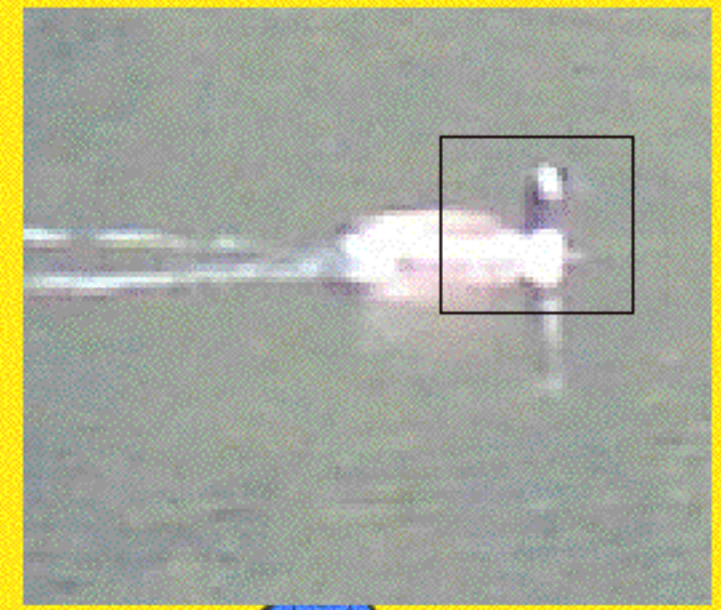
How similar are these two images

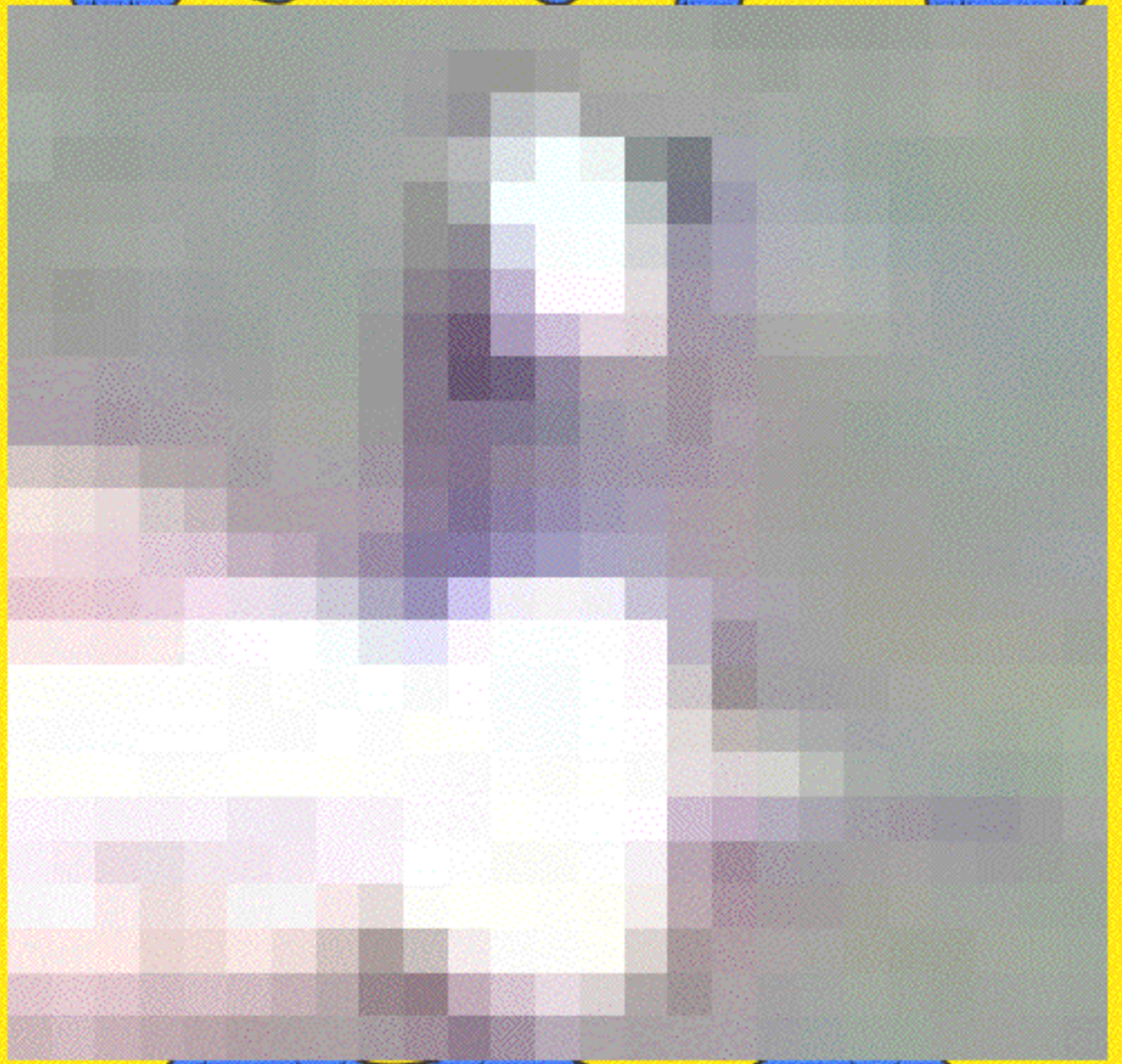
How Images are represented

Digital Images

- Represented as Pixel's
  - Lot's of little coloured dots on a regular grid
- Pixilation
- Also called Raster







Compression

- In practice images are stored as compressed raster
  - Jpeg
  - Mpeg
- Cf Vector …
- Not Relevant to retrieval

Image Processing for Retrieval

- Representing the Images
  - Segmentation
  - Low Level Features
    - Colour
    - Texture
    - Shape

Image Features

- Information about colour or texture or shape which are extracted from an image are known as ***image features***
  - Also a low-level features
    - Red, sandy
  - As opposed to high level features or concepts
    - Beaches, mountains, happy, serene, George Bush

Image Segmentation

- Do we consider the whole image or just part ?
  - Whole image - global features
  - Parts of image - local features

Global features

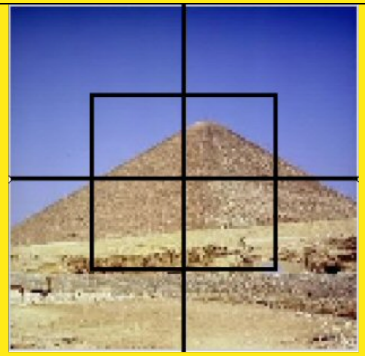
- Averages across whole image
- ✗ Tends to loose distinction between foreground and background
- ✗ Poorly reflects human understanding of images
- ✓ Computationally simple
- ✓ A number of successful systems have been built using global image features including Sunderland's CHROMA

Local Features

- ***Segment*** images into parts
- Two sorts:
  - *Tile Based*
  - *Region based*

Regioning and Tiling Schemes

Tiles



(a) 5 tiles

(b) 9 tiles

Regions

(c) 5 regions

(d) 9 regions

Tiling

- ❑ Break image down into simple geometric shapes
- ✗ Similar Problems to Global
  - ✗ Plus dangers of breaking up significant objects
- ✓ Computational Simple
- ✓ Some Schemes seem to work well in practice

Regioning

- Break Image down into visually coherent areas

✓ Can identify meaningful areas and objects

✗ Computationally intensive

✗ Unreliable

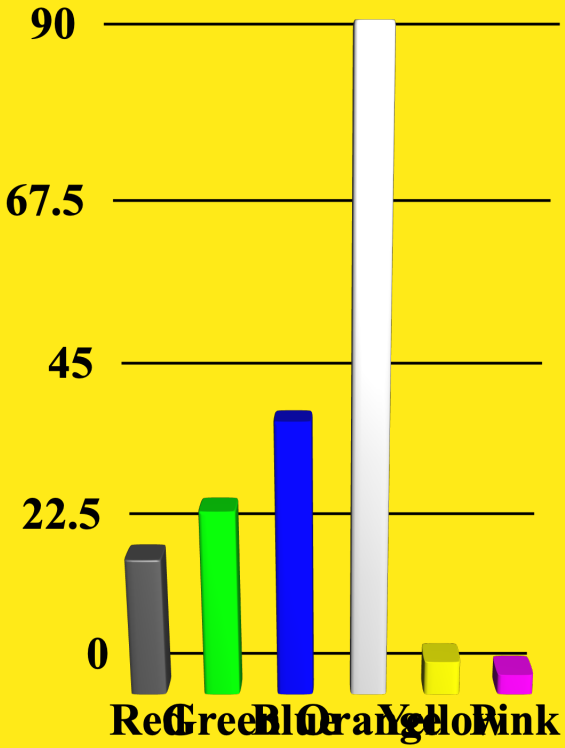
Colour

- Produce a colour ***signature*** for region/whole image
- Typically done using colour correllograms or colour histograms

Colour Histograms

Identify a number of buckets in which to sort the available colours (e.g. red green and blue, or up to ten or so colours)
Allocate each pixel in an image to a bucket and count the number of pixels in each bucket.
Use the figure produced (bucket id plus count, normalised for image size and resolution) as the index key (signature) for each image.

Global Colour Histogram



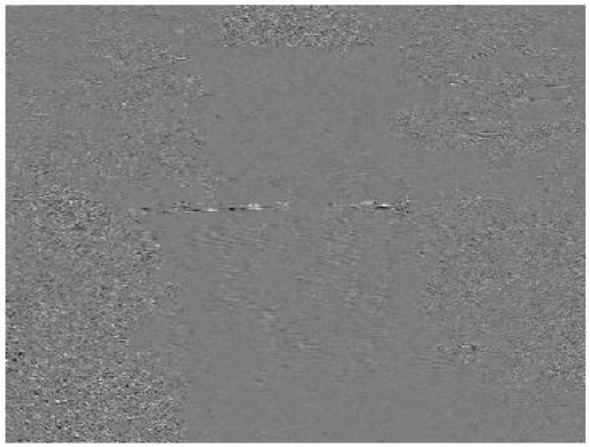
Other Colour Issues

- Many Colour Models
  - RGB (red green blue)
  - HSV (Hue Saturation Value)
  - Lab, etc. etc.
- Problem is getting something like human vision
  - Individual differences

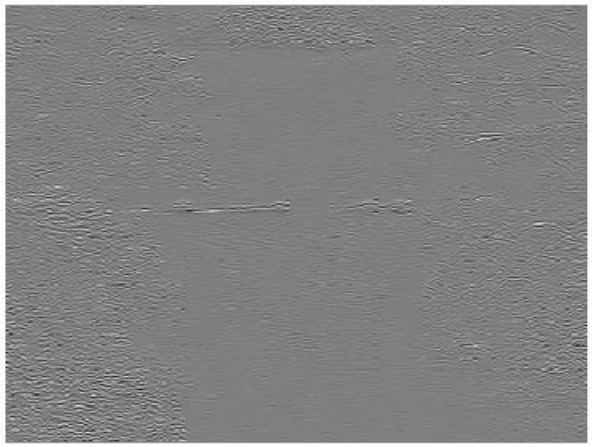
Texture

- Produce a mathematical characterisation of a repeating pattern in the image
  - Smooth
  - Sandy
  - Grainy
  - Stripey

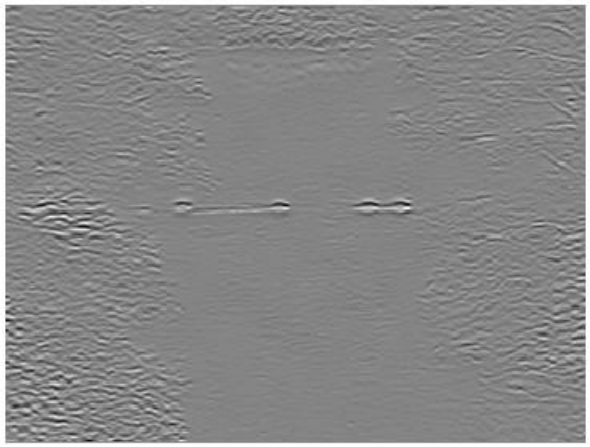
Horizontal Bands At Four Scale Levels



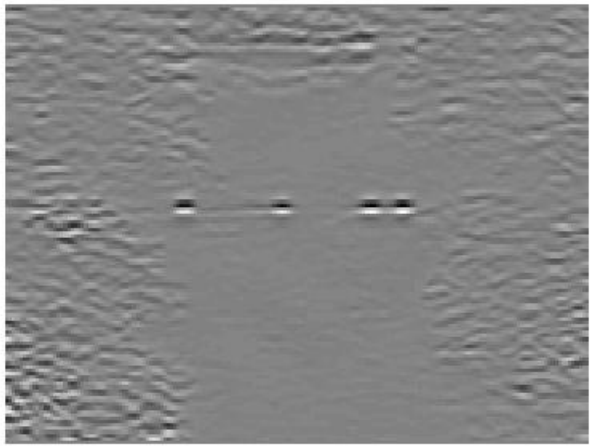
Range: [-0.162, 0.167]
Dims: [1024, 1360] / 8



Range: [-0.266, 0.265]
Dims: [512, 680] / 4

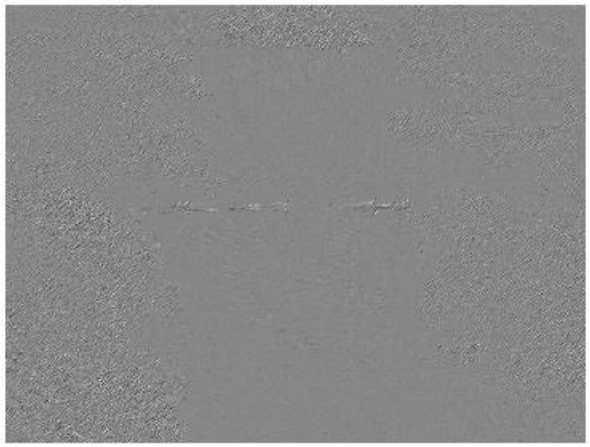


Range: [-0.528, 0.487]
Dims: [256, 340] / 2

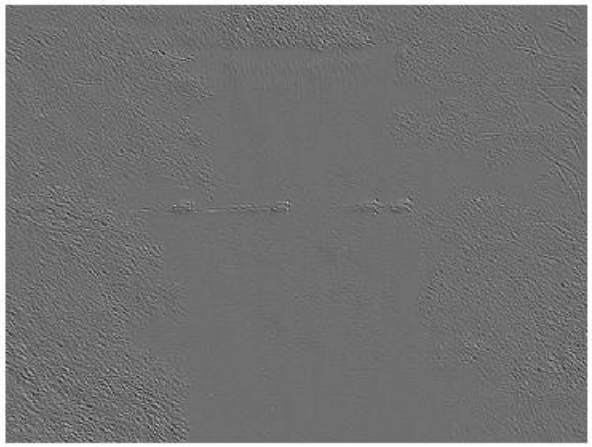


Range: [-1, 0.94]
Dims: [128, 170] / 1

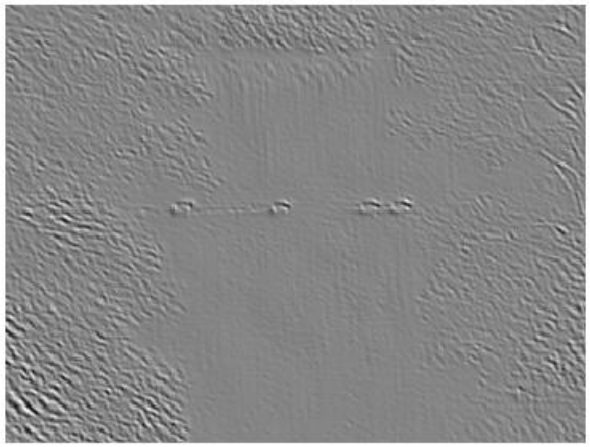
Diagonal Bands At Four Scales



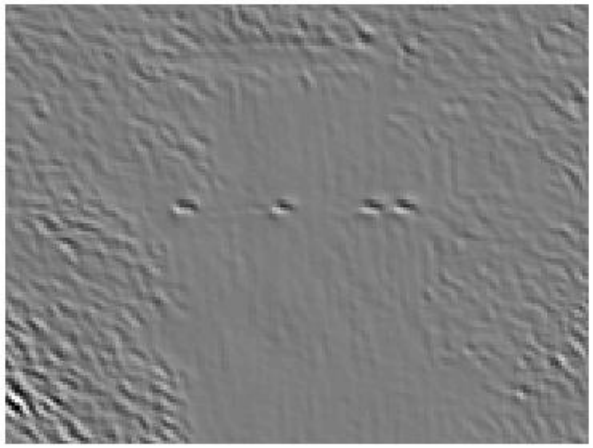
Range: [-0.146, 0.13]
Dims: [1024, 1360] / 8



Range: [-0.233, 0.3]
Dims: [512, 680] / 4



Range: [-0.445, 0.406]
Dims: [256, 340] / 2



Range: [-0.93, 0.945]
Dims: [128, 170] / 1

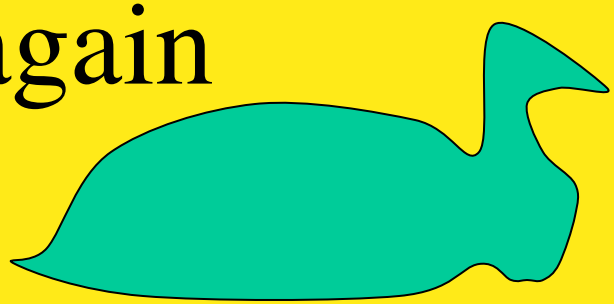
Texture

- Reduces an area/region to a (small - 15 ?) set of numbers which can be used a signature for that region.
- Proven to work well in practice
- Hard for people to understand

Shape

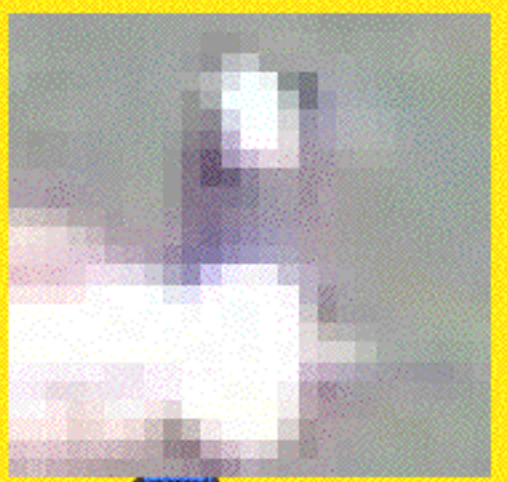
- Straying into the realms of object recognition
- Difficult and Less Commonly used

Ducks again



- All objects have closed boundaries
- Shape interacts in a rather vicious way with segmentation
- Find the duck shapes

Match the Duck



Summary of Image Representation

- Pixels and Raster
- Image Segmentation
  - Tiles
  - Regions
- Low-level Image Features
  - Colour
  - Texture
  - Shape

Indexing and Retrieving Images

Overview of Section 2

- Quick Reprise on IR
- Navigational Approaches
- Relevance Feedback
- Automatic Keyword Annotation

Reprise on Key Interactive IR ideas

- Index Time vs Query Time Processing
  - Query Time
    - Must be fast enough to be interactive
  - Index (Crawl) Time
    - Can be slow(ish)
    - There to support retrieval

An Index

- A data structure which stores data in a suitably abstracted and compressed form in order to faciliate rapid processing by an application

Indexing Process

- Index all the images in the database
- Store the indexes in a suitable form
- Search will be done on indexes

Navigational Approaches to Image Retrieval

Essential Idea

- Layout images in a virtual space in an arrangement which will make some sense to the user
- Project this onto the screen in a comprehensible form
- Allow them to navigate around this projected space (scrolling, zooming in and out)

Notes

- Typically colour is used
  - Texture has proved difficult for people to understand
  - Shape possibly the same, and also user interface - most people can't draw !
- Alternatives include time (Canon's Time Tunnel) and recently location (GPS Cameras)
- Need some means of knowing where you are

Observation

- It appears people can take in and will inspect many more images than texts when searching

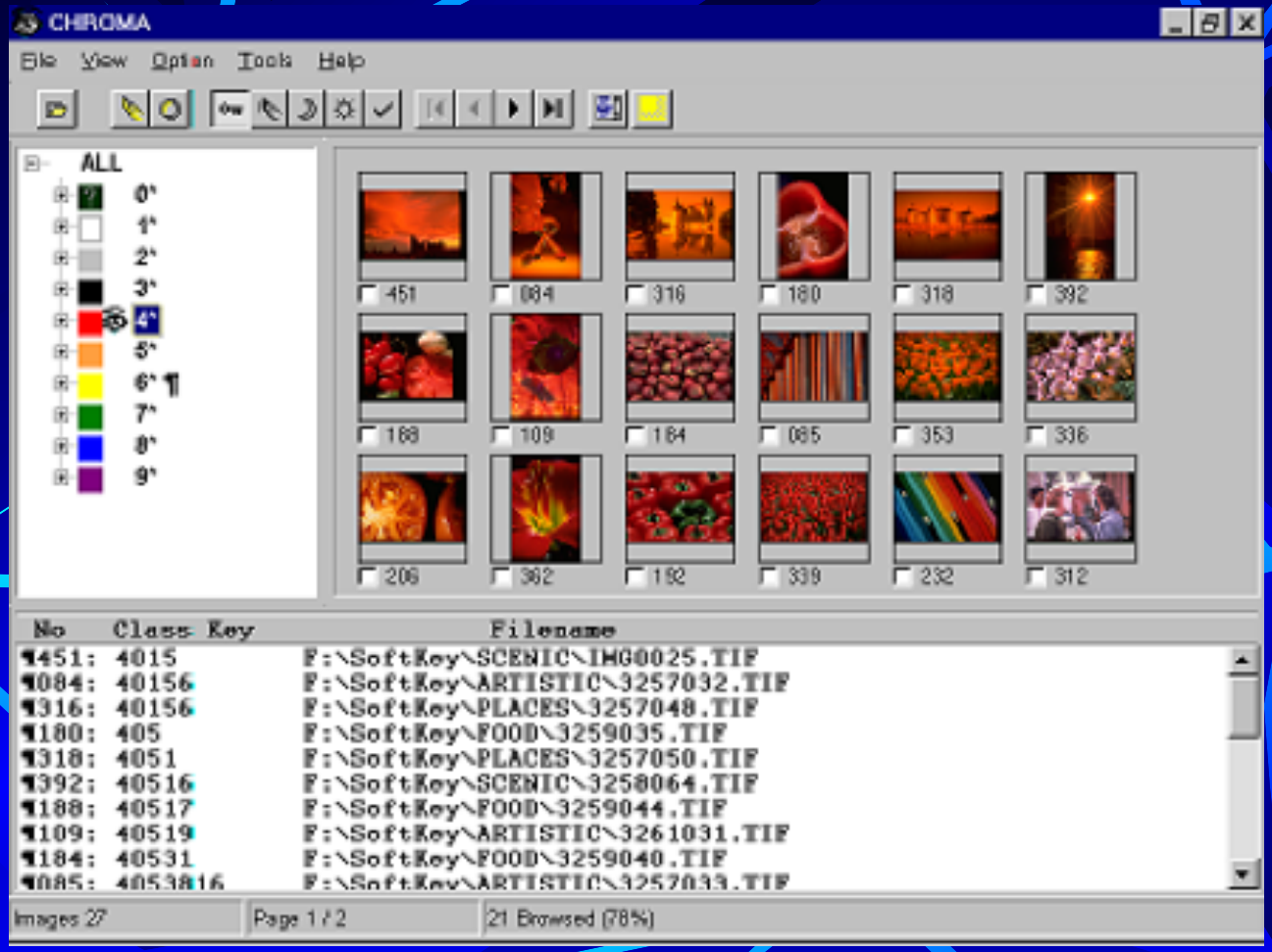
CHROMA

- Development in Sunderland:
  - mainly by Ting Sheng Lai now of National Palace Museum, Taipei, Taiwan
- **Structure Navigation System**
- **Thumbnail Viewer**
- **Similarity Searching**
- **Sketch Tool**

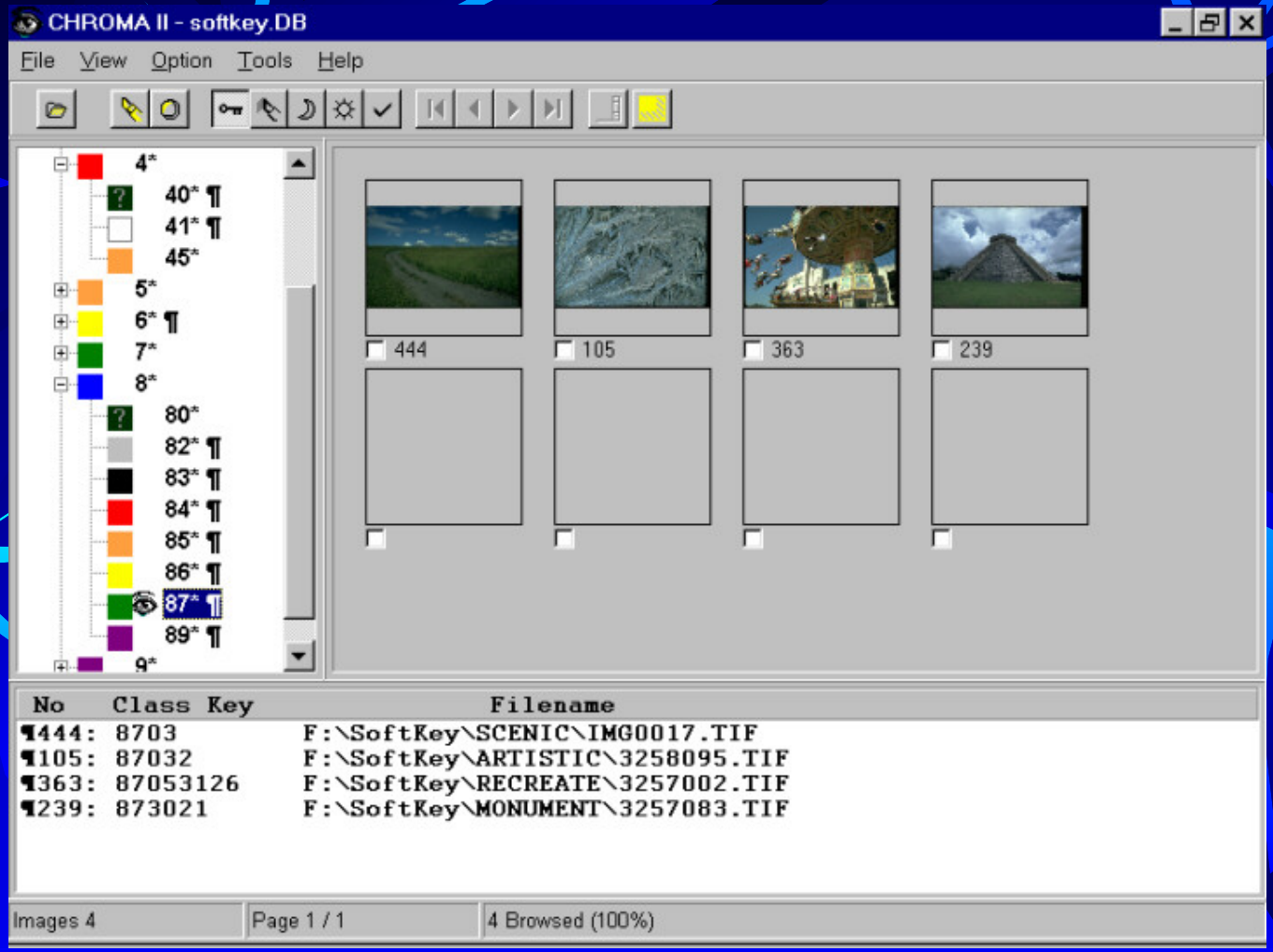
The CHROMA System

- General Photographic Images
- Global Colour is the Primary Indexing Key
- Images organised in a hierarchical classification using 10 colour descriptors and colour histograms

Access System



The Navigation Tool



Technical Issues

- Fairly Easy to arrange image signatures so they support rapid browsing in this space

Relevance Feedback

More Like this

Relevance Feedback

- Well established technique in text retrieval
  - Experimental results have always shown it to work well in practice
- Unfortunately experience with search engines has shown that it is difficult to get real searchers to adopt it - too much interaction

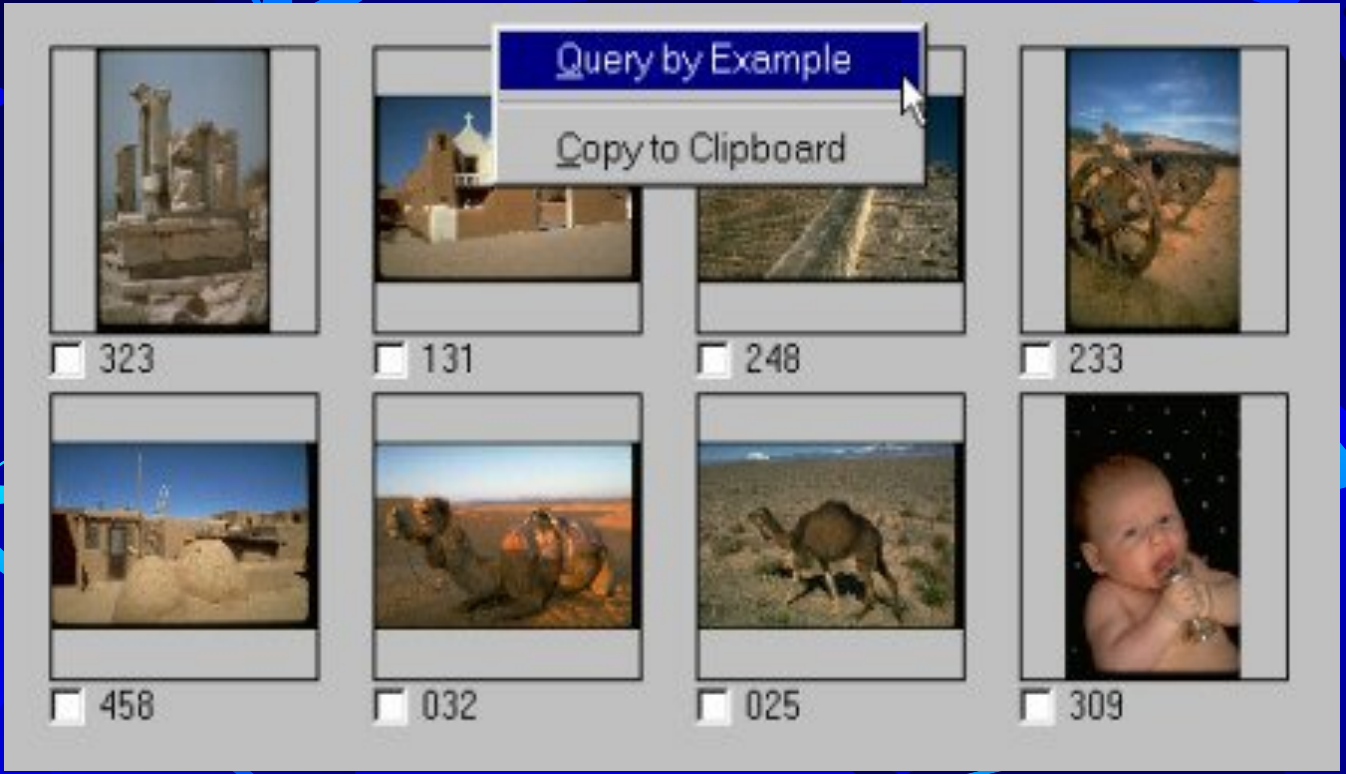
Essential Idea

- User performs an initial query
- Selects some relevant results
- System then extracts terms from these to augment the initial query
- Requeries

Many Variants

- Pseudo
  - Just assume high ranked documents are relevant
- Ask users about terms to use
- Include negative evidence
- Etc. etc.

Query-by-Image-Example



Why useful in Image Retrieval?

1. Provides a bridge between the users understanding of images and the low level features (colour, texture etc.) with which the systems is actually operating
2. Is relatively easy to interface to

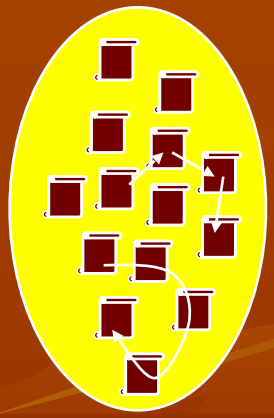
Image Retrieval Process

Green

Water Texture

Leaf Texture

Ducks



Observations

- Most image searchers prefer to use key words to formulate initial queries
  - Eakins et al, Enser et al
- First generation systems all operated using low level features only
  - Colour, texture, shape etc.
  - Smeulders et al

Ideal Image Retrieval Process

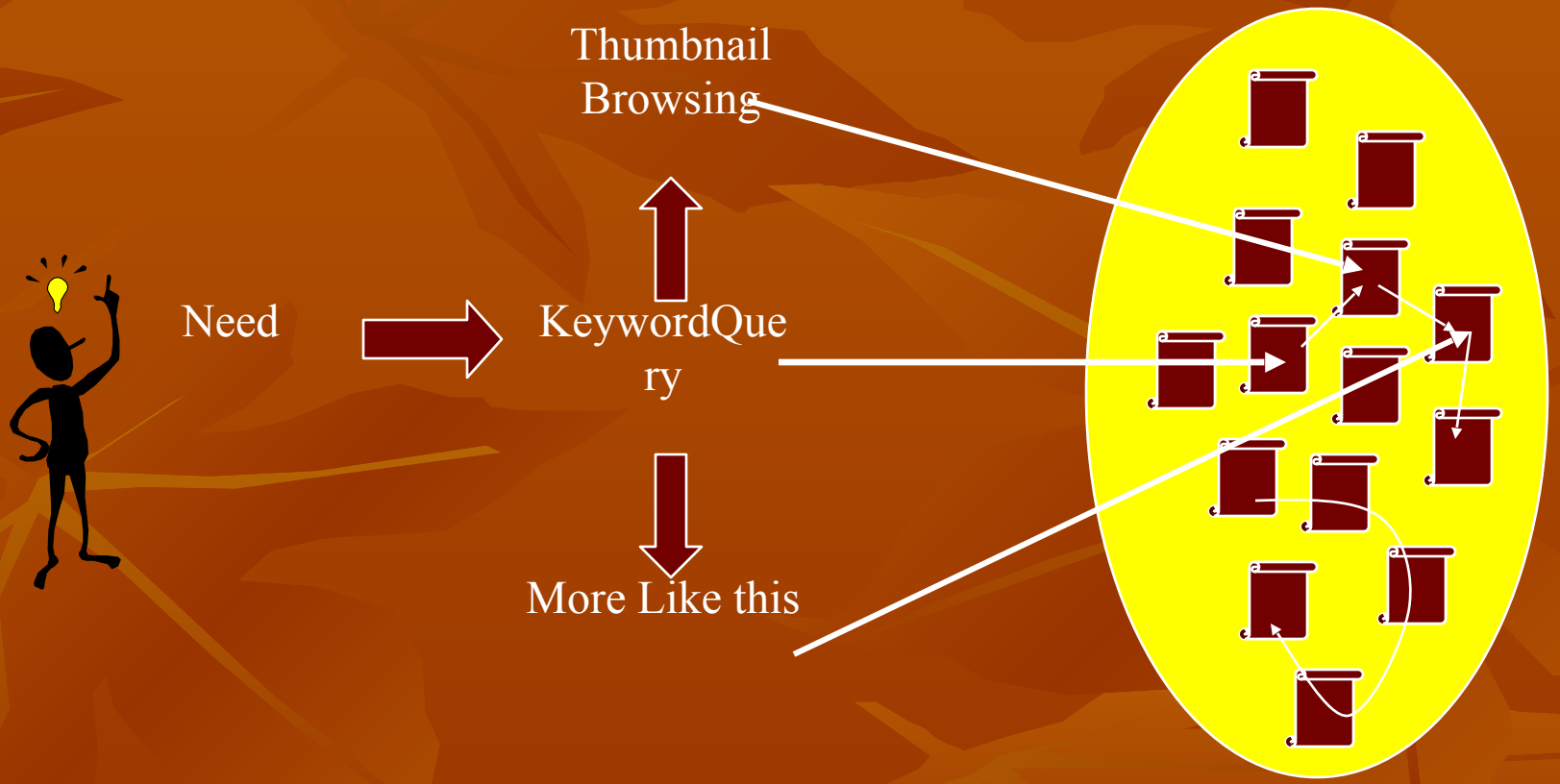


Image Retrieval as Text Retrieval

What we really want to do is make the image retrieval problem as a text retrieval problem

Three Ways to go

- Manually Assign Keywords to each image
- Use text associated with the images (captions, web pages)
- Analyse the image content to automatically assign keywords

Manual Keywording

- Expensive
  - Can only really be justified for high value collections – advertising
- Unreliable
  - Do the indexers and searchers see the images in the same way
- Feasible

Associated Text

- Cheap
- Powerful
  - Famous names/incidents
- Tends to be “one dimensional”
  - Does not reflect the content rich nature of images
- Currently Operational - Google

Possible Sources of Associated text

- Filenames
- Anchor Text
- Web Page Text around the anchor/where the image is embedded

Automatic Keyword Assignment

A form of Content Based Image Retrieval

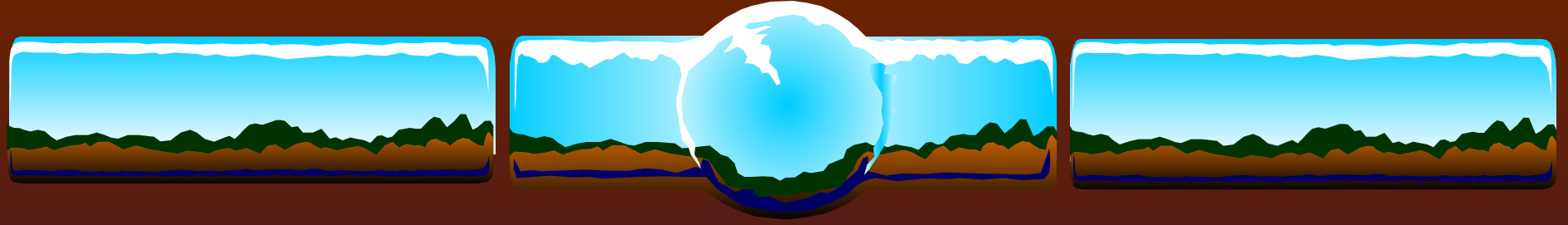
- Cheap (ish)
- Predictable (if not always “right”)
- No operational System Demonstrated
  - Although considerable progress has been made recently

Basic Approach

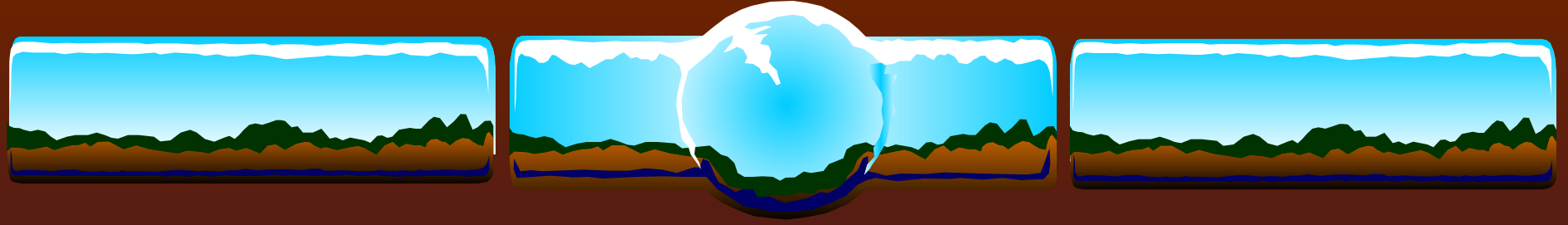
- Learn a mapping from the low level image features to the words or concepts

Two Routes

1. Translate the image into piece of text
   - Forsyth and other s
   - Manmatha and others
2. Find that category of images to which a keyword applies
   - Tsai and Tait
   - (SIGIR 2005)

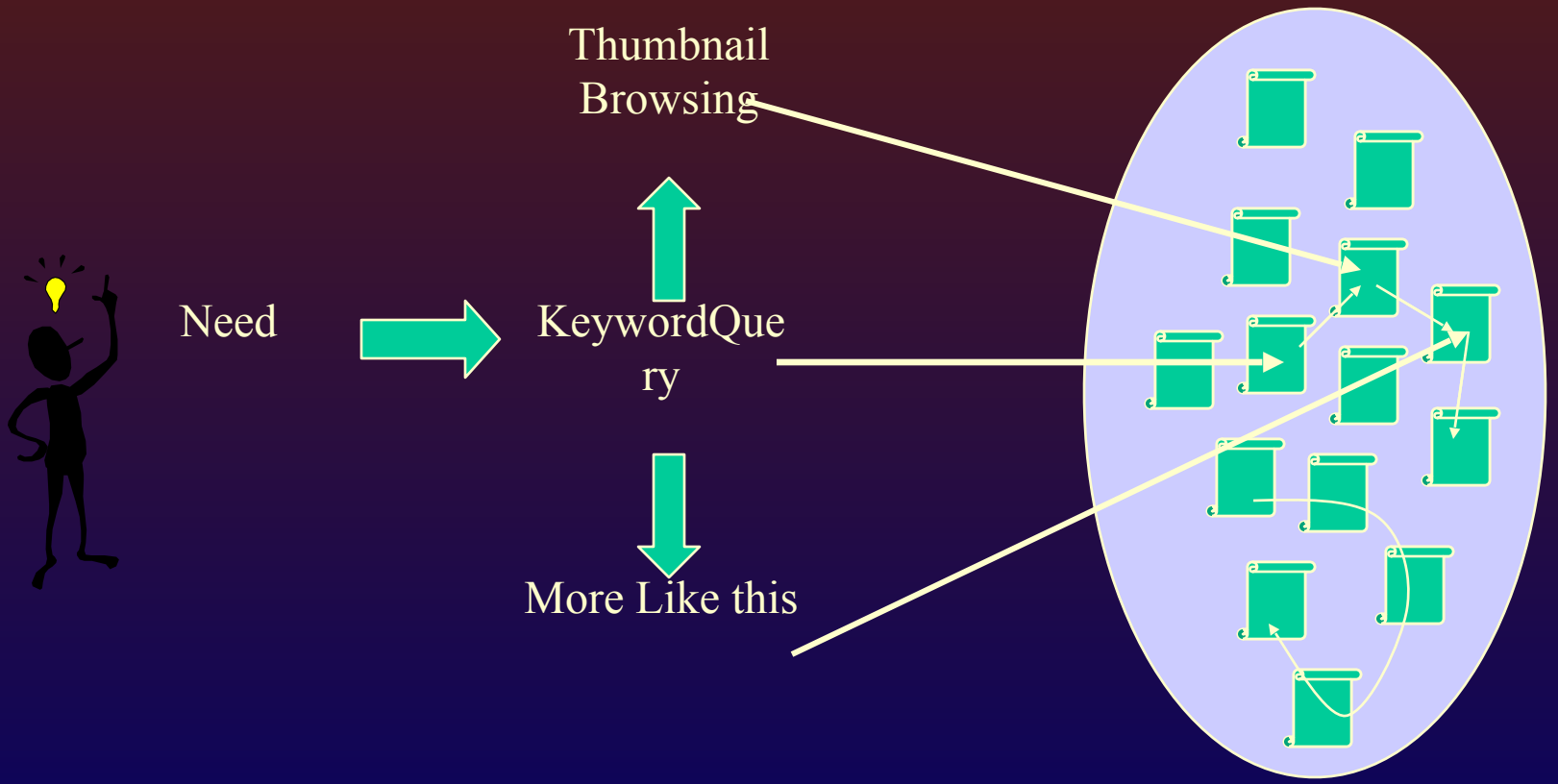


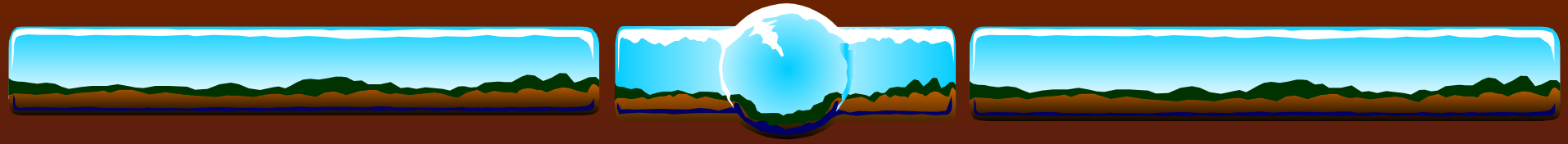
Advanced Topics, Futures and Conclusions



Towards Practical Systems

Ideal Image Retrieval Process

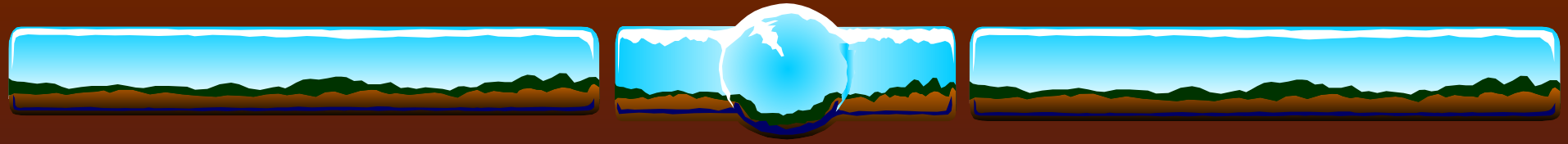




Requirements

- > 5000 Key word vocabulary
- > 5% accuracy of keyword assignment for all keywords
  - > 5% precision in response to single key word queries

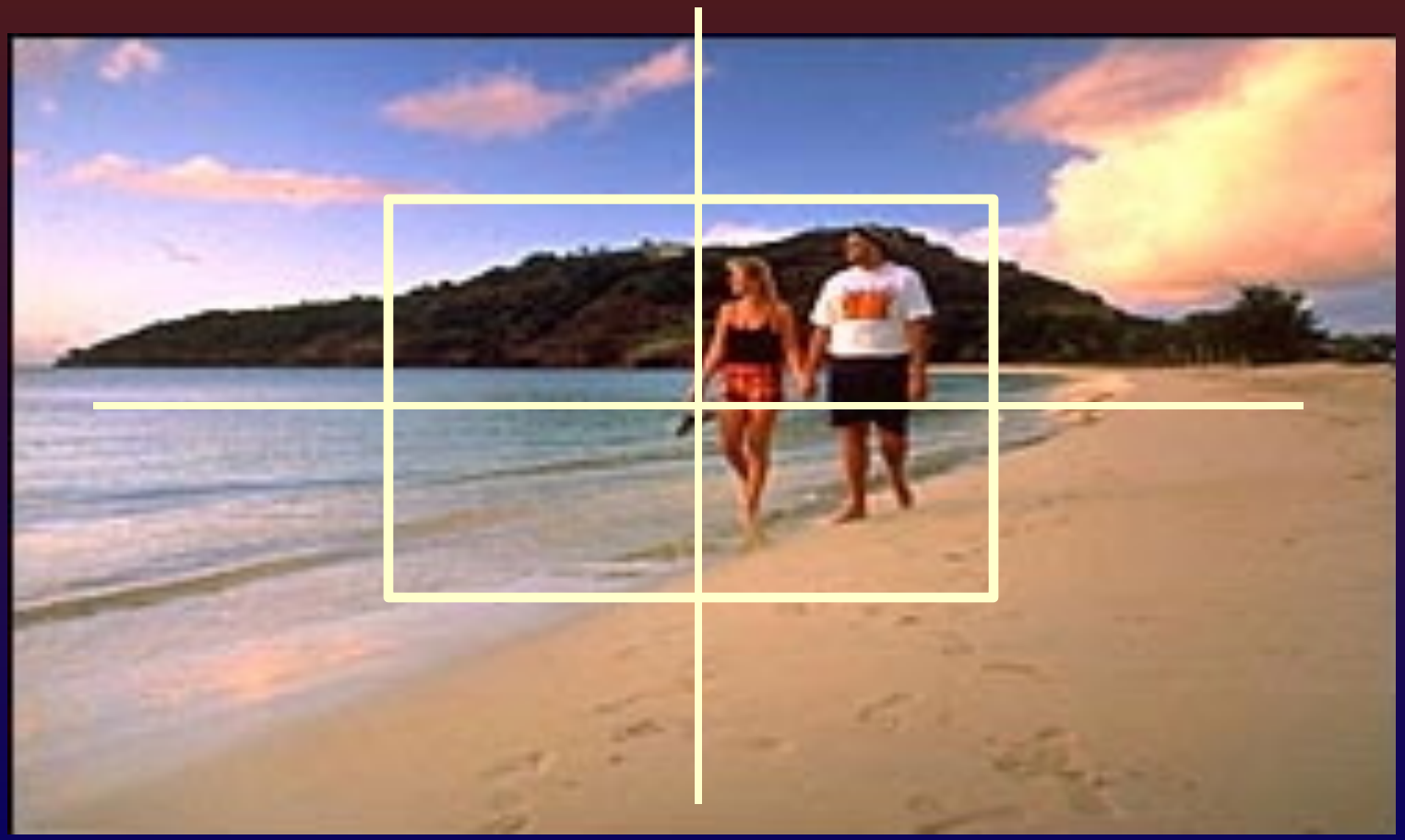
The Semantic Gap Bridged!



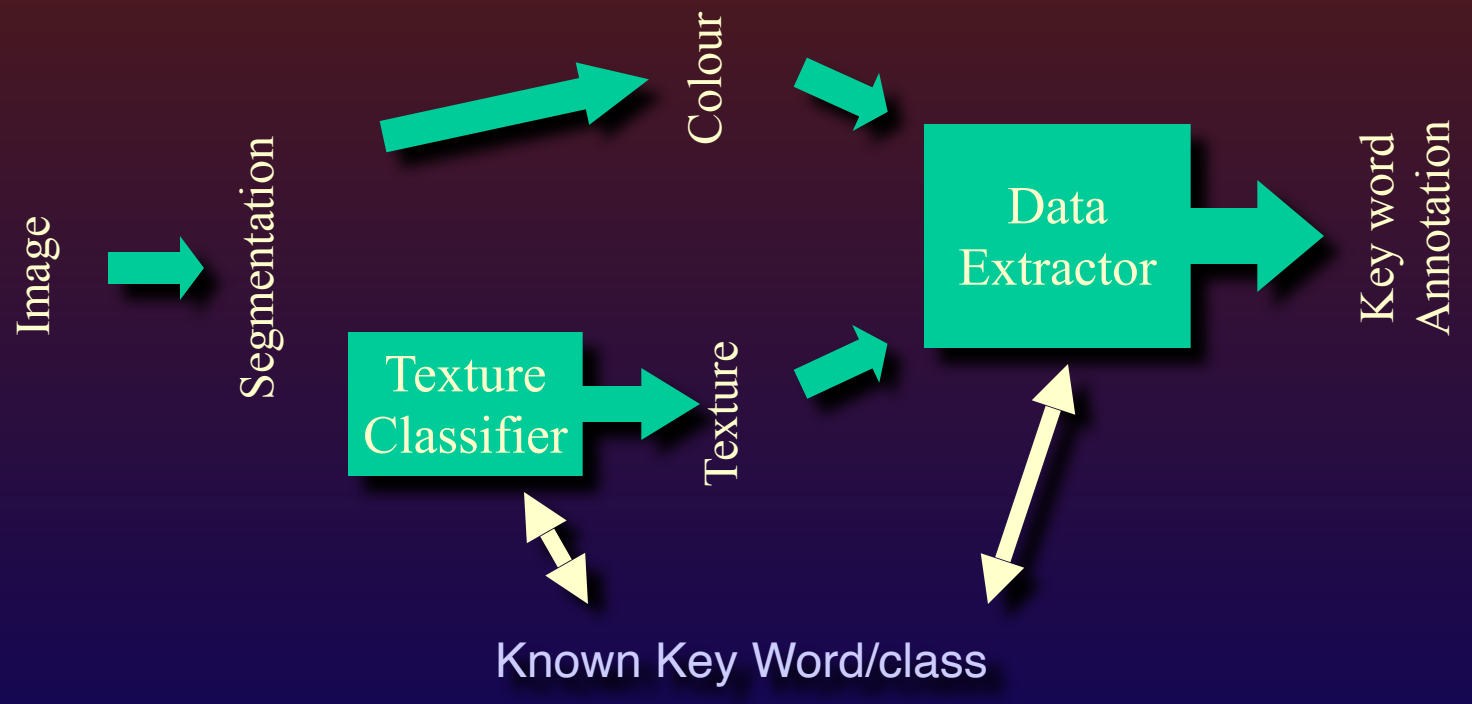
CLAIRE

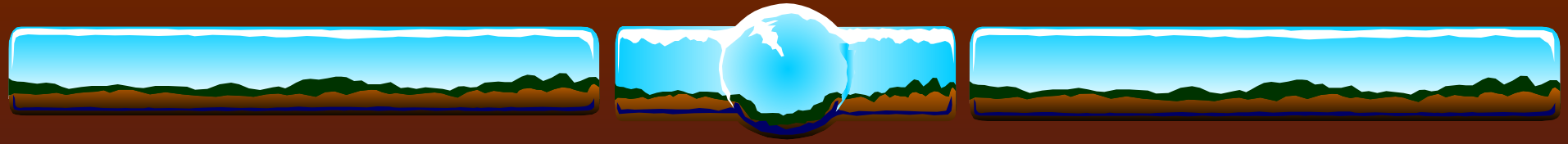
- Example State of the Art Semantic CBIR System
- Colour and Texture Features
- Simple Tiling Scheme
- Two Stage Learning Machine
  - SVM/SVM and SVM/k-NN
  - Colour to 10 basic colours
  - Texture to one texture term per category

Tiling Scheme



Architecture of Claire

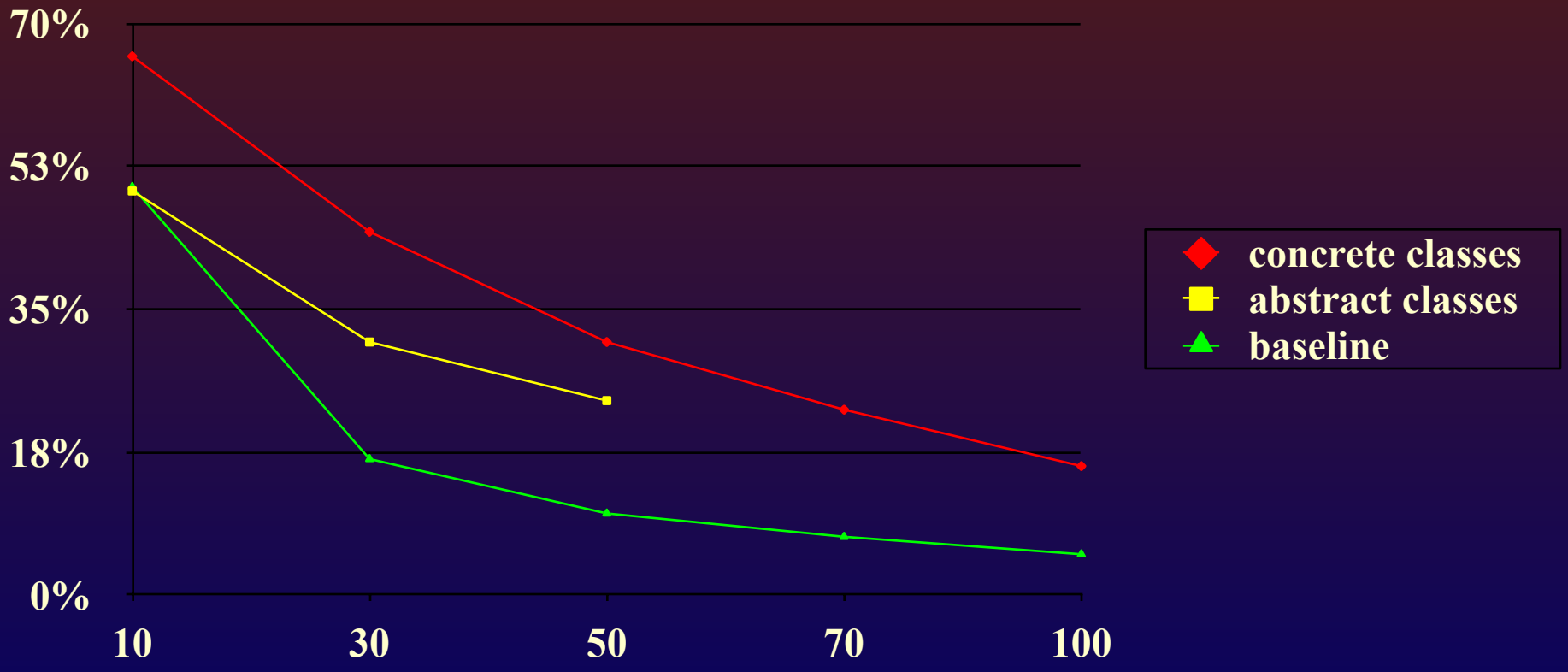


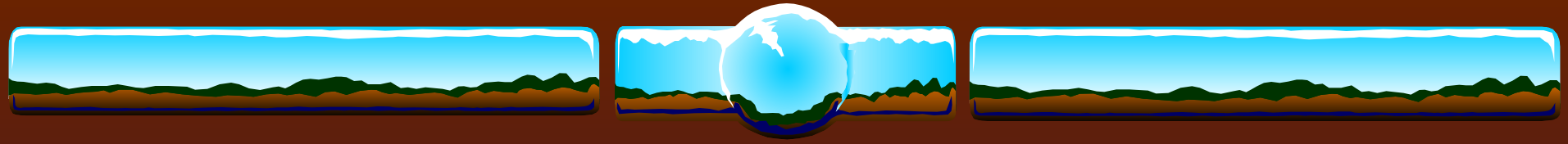


Training/Test Collection

- Randomly Selected from Corel
- Training Set
  - 30 images per category
- Test Collection
  - 20 images per category

SVM/SVM Keywording with 100+50 Categories

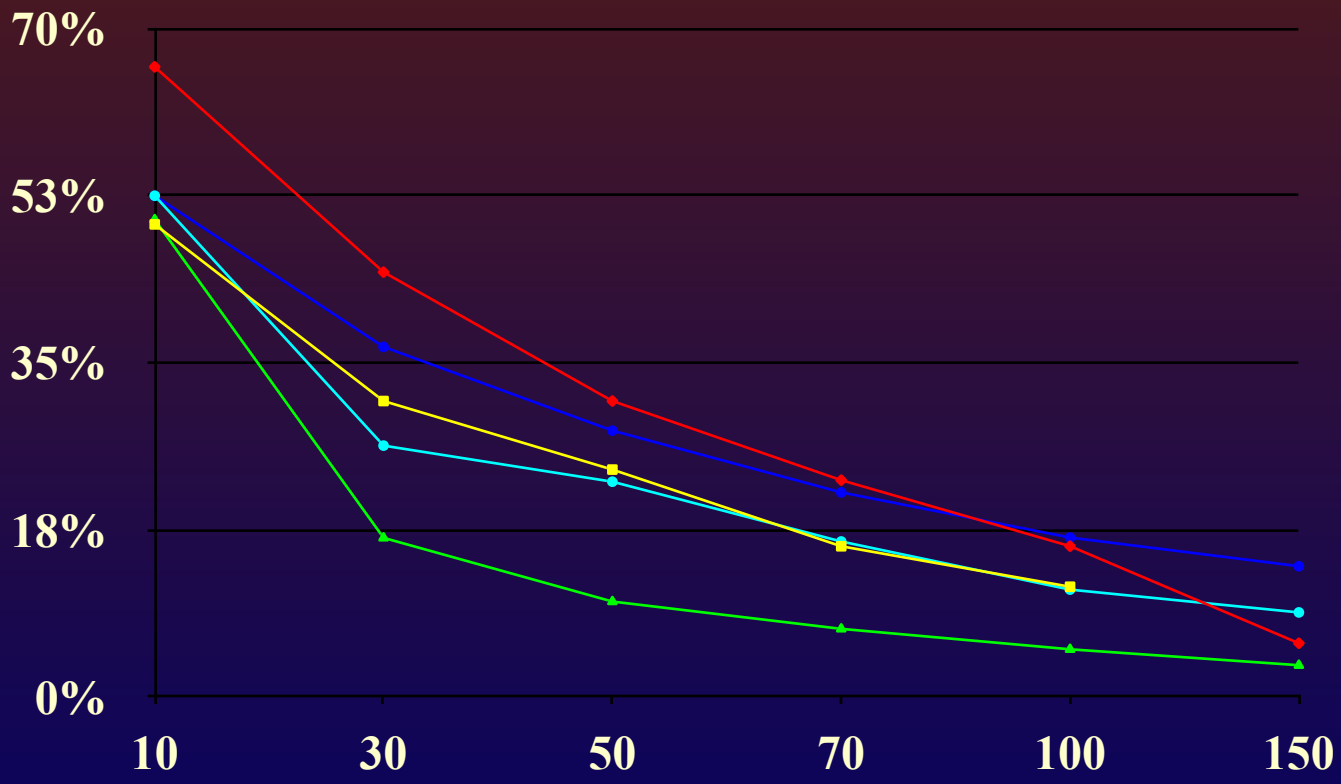




Examples Keywords

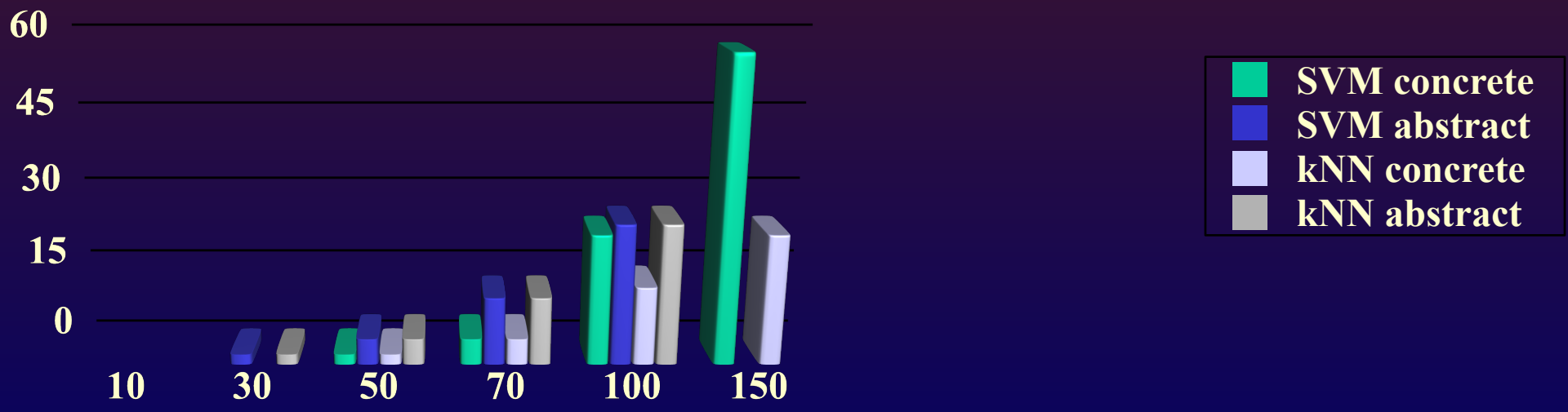
- Concrete
  - Beaches
  - Dogs
  - Mountain
  - Orchids
  - Owls
  - Rodeo
  - Tulips
  - Women
- Abstract
  - Architecture
  - City
  - Christmas
  - Industry
  - Sacred
  - Sunsets
  - Tropical
  - Yuletide

SVM vs kNN

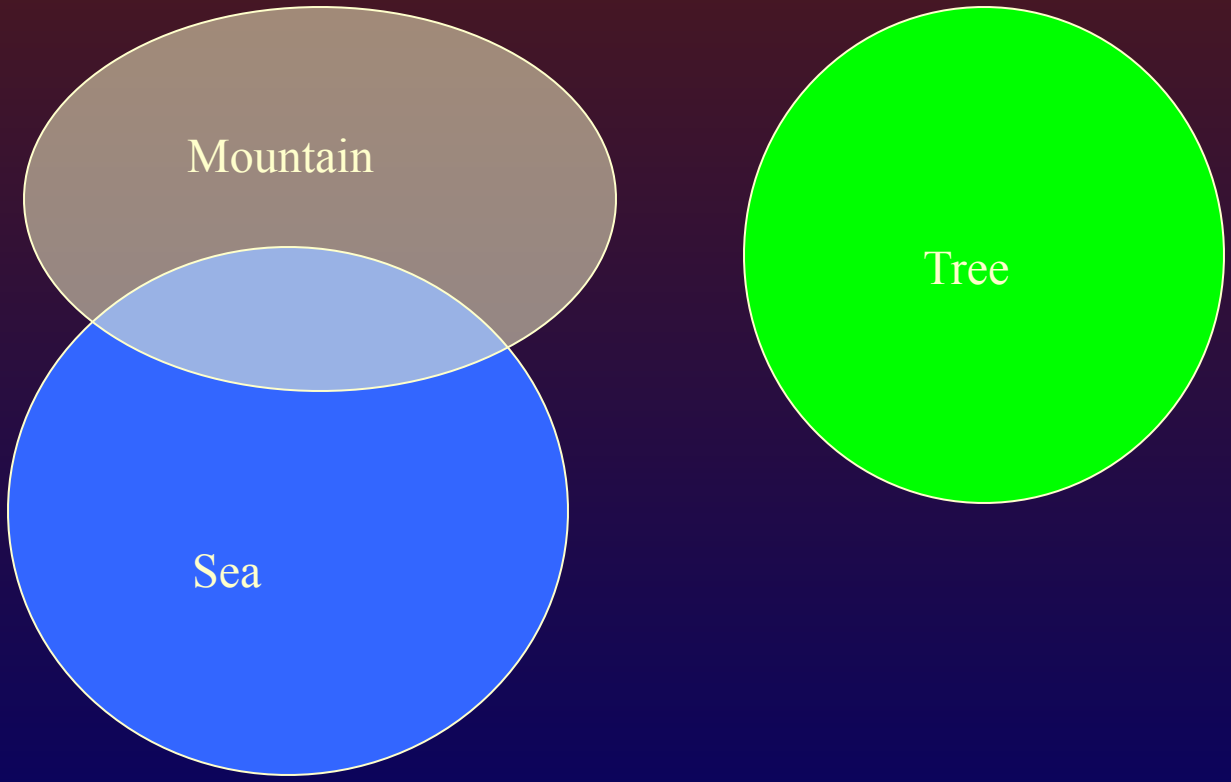

70%
53%
35%
18%
0%
10
30
50
70
100
150
SVM concrete
SVM abstract
baseline
kNN abstract
kNN concrete

Reduction in Unreachable Classes

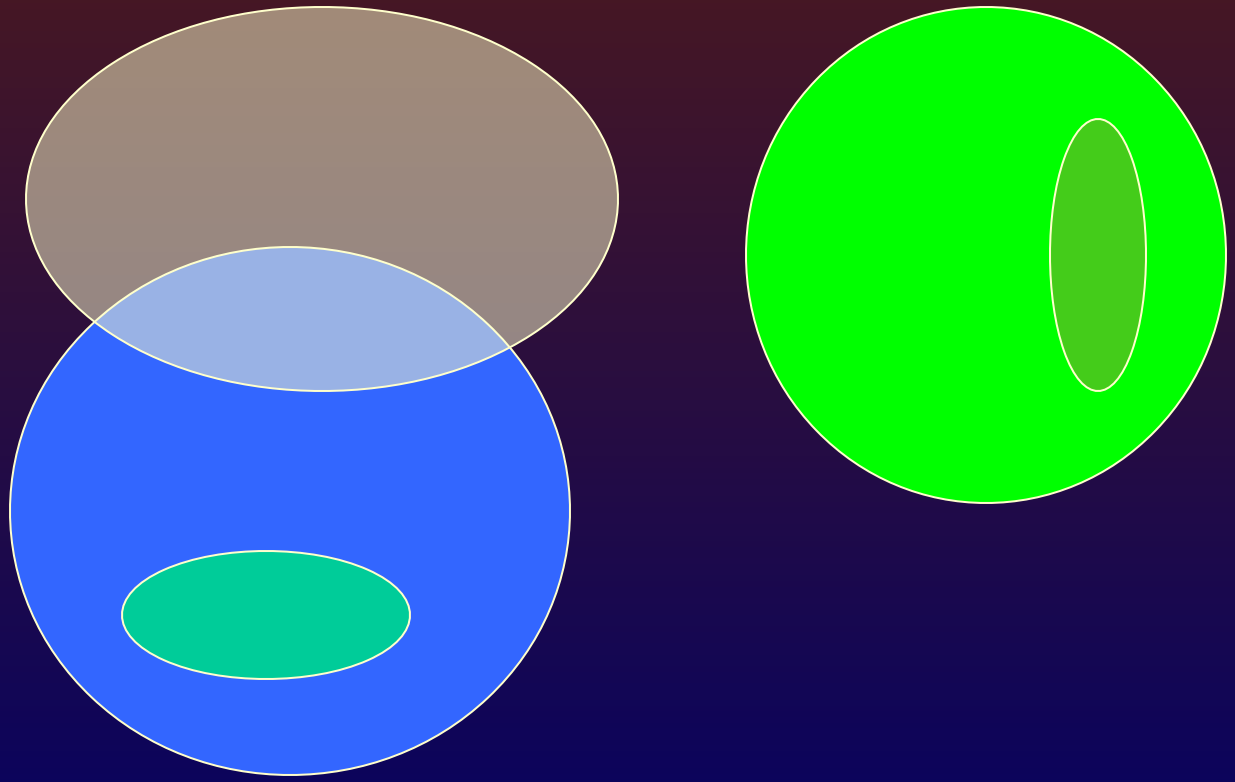
Missing Category Numbers



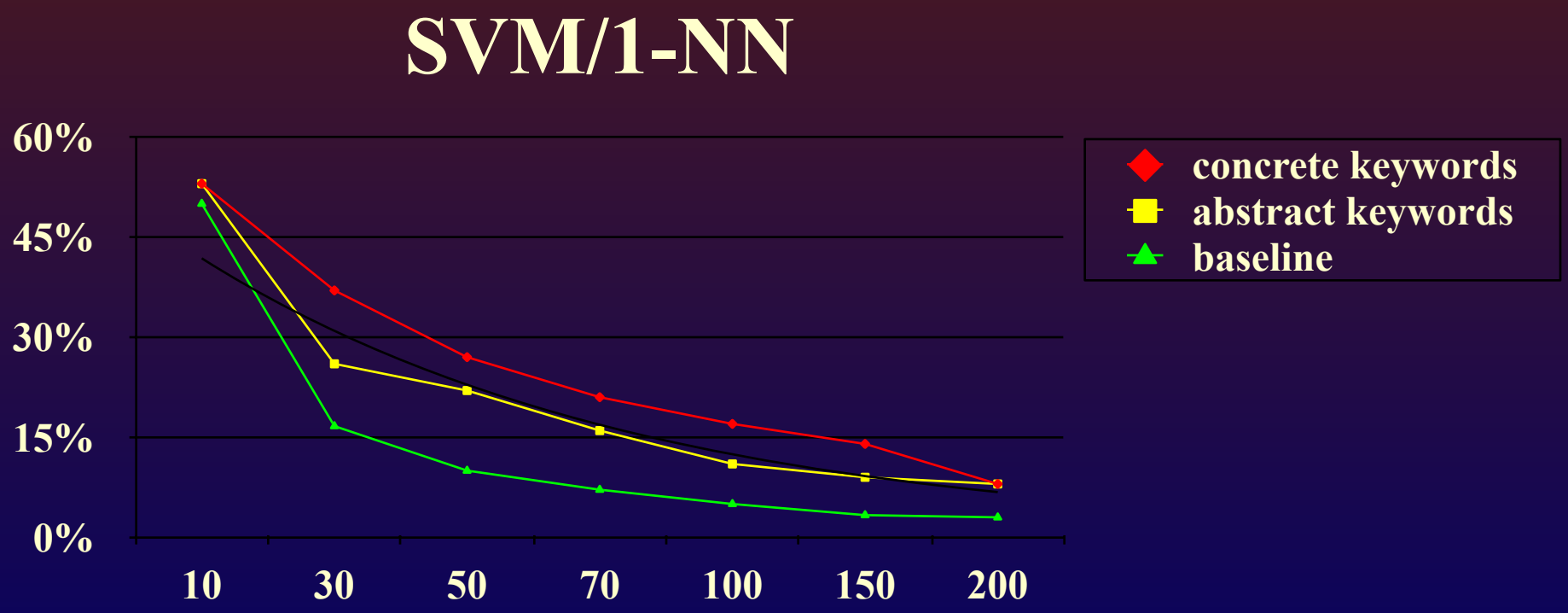
Labelling Areas of Feature Space

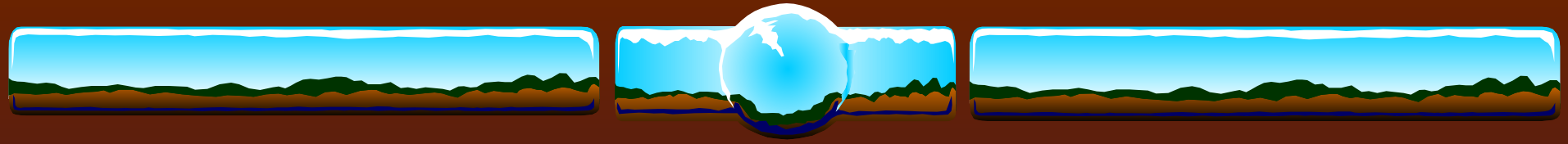


Overlap in Feature Space



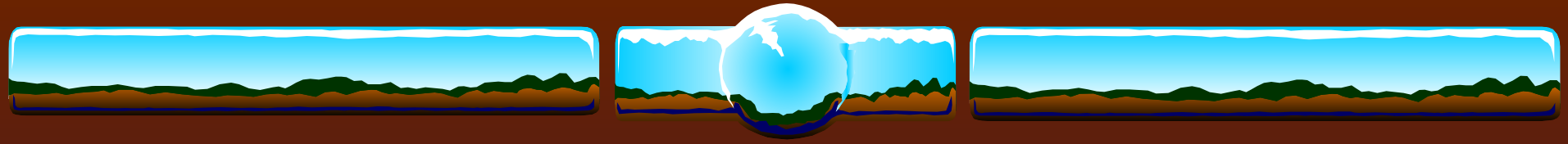
Keywording 200+200 Categories





Discussion

- Results still promising 5.6% of images have at least one relevant keyword assigned
- Still useful - but only for a vocabulary of 400 words !
  - See demo at http://osiris.sunderland.ac.uk/~da2wli/system/silk1/
- High proportion of categories which are never assigned

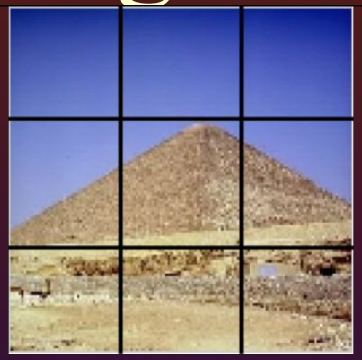


Segmentation

Are the results dependent on the specific tiling/ regioning scheme used ?

Regioning

(a) 5 tiles

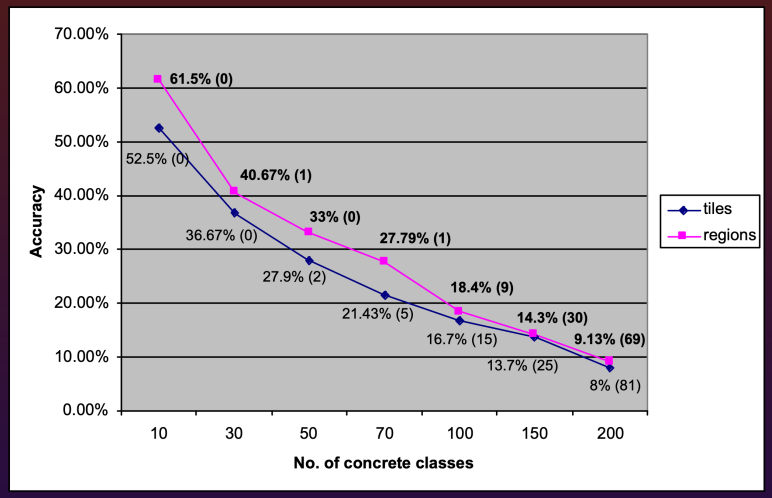


(b) 9 tiles

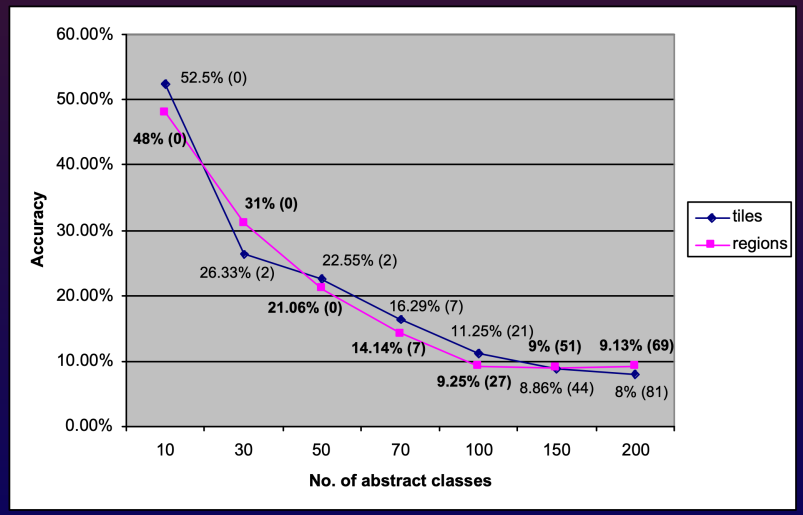
(c) 5 regions

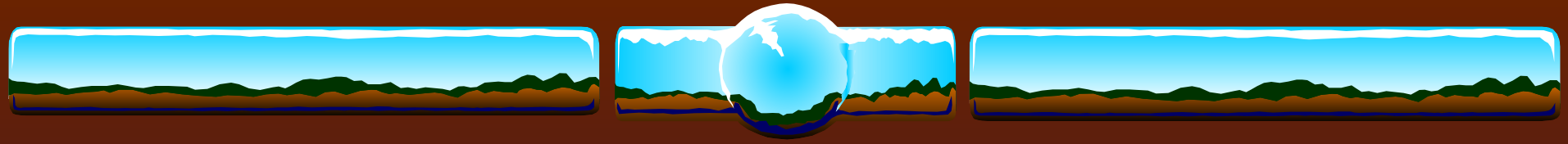
(d) 9 regions

Effectiveness Comparison



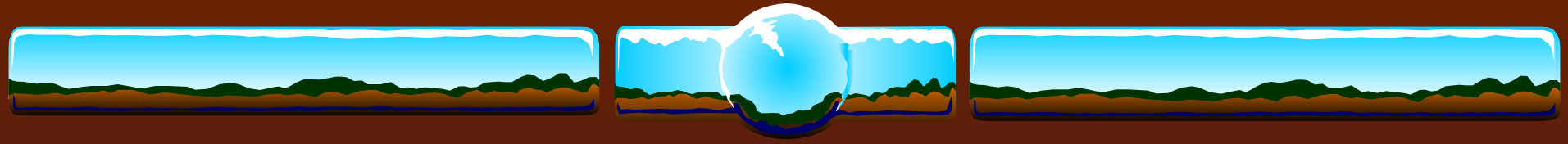
Five Tiles vs Five Regions
1-NN Data Extractor





Next Steps

- More categories
- Integration into complete systems
- Systematic Comparison with Generative approach pioneered by Forsyth and others



Conclusions

- Image Indexing and Retrieval is Hard
- Effective Image Retrieval needs a cheap and predictable way of relating words and images
- Adaptive and Machine Learning approaches offer one way forward with much promise

Selected Bibliography

- Early Systems

  The following leads into all the major trends in systems based on colour, texture and shape

  - A. Smeaulder, M. Worring, S. Santini, A. Gupta and R. Jain "Content-based Image Retrieval: the end of the early years" IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12):1349-1380, 2000.

- CHROMA
  - Sharon McDonald and John Tait "Search Strategies in Content-Based Image Retrieval" Proceedings of the 26th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003), Toronto, July, 2003. pp 80-87. ISBN 1-58113-646-3
  - Sharon McDonald, Ting-Sheng Lai and **John Tait,** "Evaluating a Content Based Image Retrieval System" Proceedings of the 24th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001), New Orleans, September 2001. W.B. Croft, D.J. Harper, D.H. Kraft, and J. Zobel (Eds). ISBN 1-58113-331-6 pp 232-240.

- Translation Based Approaches
  - P. Duygulu, K. Barnard, N. de Freitas and D. Forsyth "Learning a Lexicon for a Fixed Image Vocabulary" European Conference on Computer Vision, 2002.
  - K. Barnard, P. Duygulu, N. de Freitas and D. Forsyth "Matching Words and Pictures" Journal of machine Learning Research 3: 1107-1135, 2003.

  Very recent new paper on this is:

  - P. Virga, P. Duygulu "Systematic Evaluation of Machine Translation Methods for Image and Video Annotation" Images and Video Retrieval, Proceedings of CIVR 2005, Singapore, Springer, 2005.

- Cross-media Relevance Models etc
  - J. Jeon, V. Lavrenko, R. Manmatha "Automatic Image Annotation and Retrieval using Cross-Media Relevance Models" Proceedings of the 26th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003), Toronto, July, 2003. Pp 119-126

  See also recent unpublished papers on

  http://ciir.cs.umass.edu/~manmatha/mmpapers.html
- More recent stuff
  - G Carneiro and N. Vasconcelos "A Database Centric View of Sentic Image Annotation and Retrieval" Proceedings of the 28th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005), Salvador, Brazil, August, 2005
  - M. Srikanth, J. Varner, M. Bowden, D. Moldovan "Exploiting Ontologies for Automatic Image Annotation" Proceedings of the 28th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005), Salvador, Brazil, August, 2005

  See also the SIGIR workshop proceedings

  http://mmir.doc.ic.ac.uk/mmir2005