

## Image Retrieval: from Perception to Emotion

Allan Hanbury

## Outline

- Image retrieval
- Content of images
- Image search at the perceptual level
- Image search at the semantic level
- Image search at the impressions, emotions and meaning level
- Summary

2

## Image retrieval

- The most obvious step beyond text retrieval is usually image retrieval, due to the ubiquity of images
- But general images can be similar in so many different ways...

3

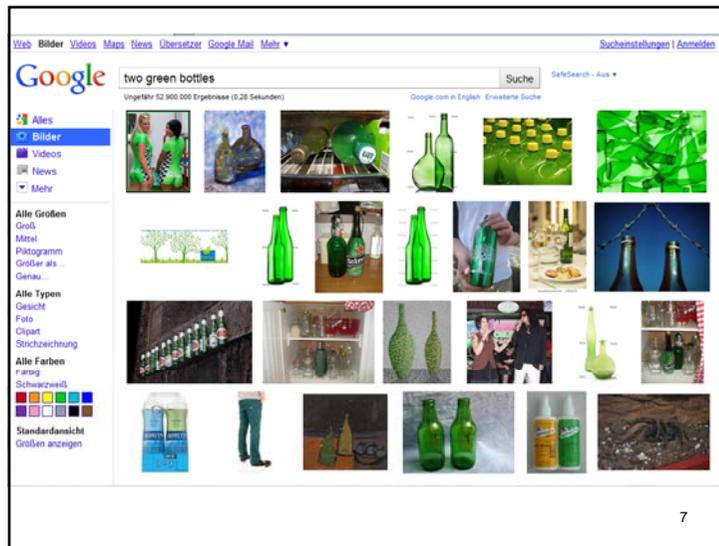
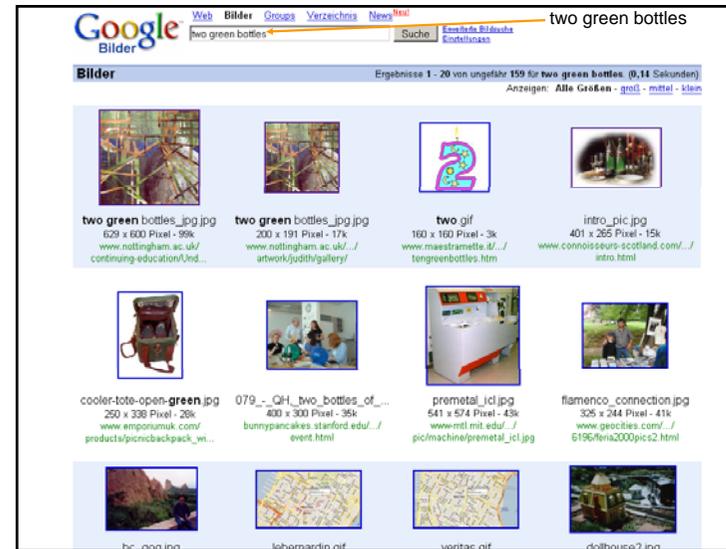
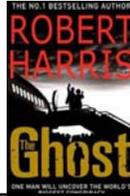
 <p>FIND SIMILAR IMAGE S</p>	 <p>FIND SIMILAR IMAGE S</p>	 <p>FIND SIMILAR IMAGE S</p>
 <p>FIND SIMILAR IMAGE S</p>	 <p>FIND SIMILAR IMAGE S</p>	 <p>FIND SIMILAR IMAGE S</p>
 <p>FIND SIMILAR IMAGE S</p>	 <p>FIND SIMILAR IMAGE S</p>	 <p>FIND SIMILAR IMAGE S</p>

Search results from LTU technologies: <http://corbis.lutec.com/>

## Image search in spy books...

“I switched to Google Images and was offered a mosaic of bizarre illustrations, a cat, a couple of acrobats in leotards ... This is the trouble with internet research, in my experience. The proportion of what’s useful to what’s dross dwindles very quickly ... What’s important is to ask the right question, and somehow I sensed I was getting it wrong.”

Robert Harris, The Ghost



7

## Image Search Applications

- Photo search
  - Personal
  - Photo archives
  - The internet
- Medical image search
- Technical document or patent image search
- ...

8

## Content of Images

- Three levels of content:
  - Perception level
  - Semantic level
  - Impression, emotion and meaning level

9

10

## Content of images

- **Perceptual properties**, like colour, texture, shape and spatial relationships.

Concentric circles

Yellow, red, black

Yellow is surrounded by red

11

- **Semantic primitives** corresponding to abstractions, like objects, roles and scenes.

A Scene from the film "2001 - Space Odyssey"

A HAL 9000 Computer

12



- Impressions, emotions and meaning associated with the combination of perceptual features.

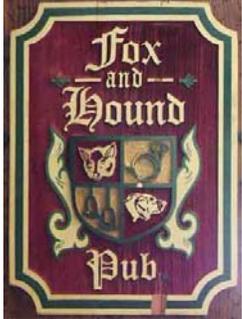
Advanced technology

Fear of advanced technology

13

### Another example: image content

- 3 levels:
  - Perceptual:
    - Mostly Mauve, with some green and yellow
    - Green and yellow frame
  - Semantic:
    - a pub sign
  - Impressions, Emotions and Meaning:
    - One can get a beer there!



14

### Understanding of “content”

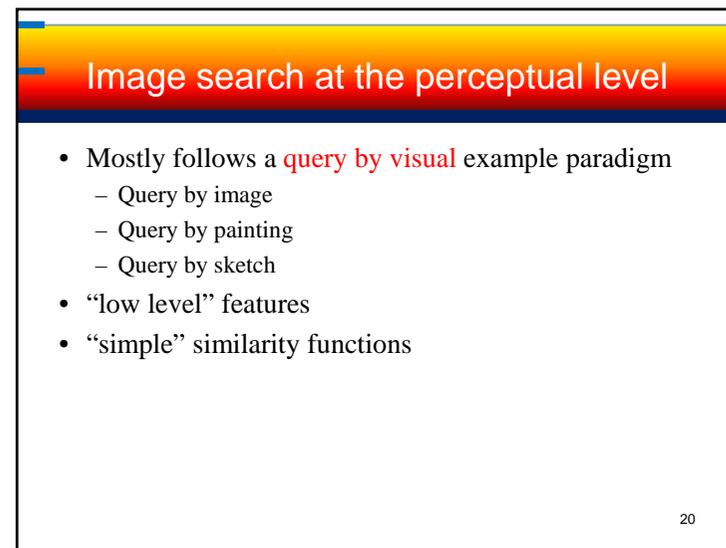
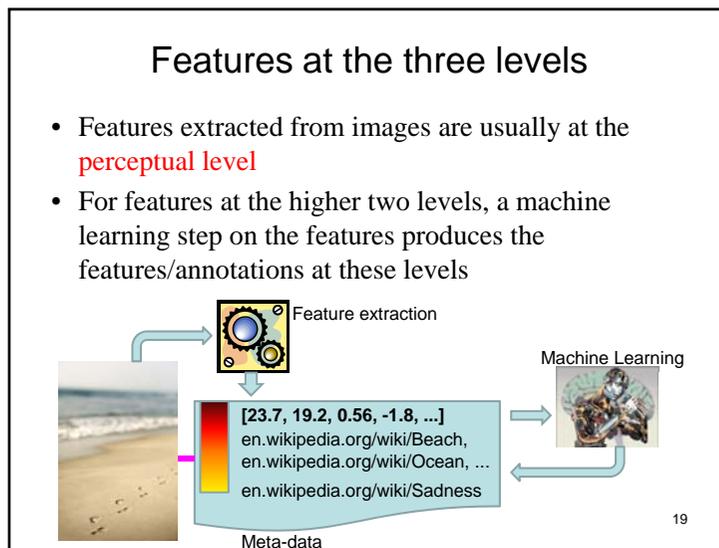
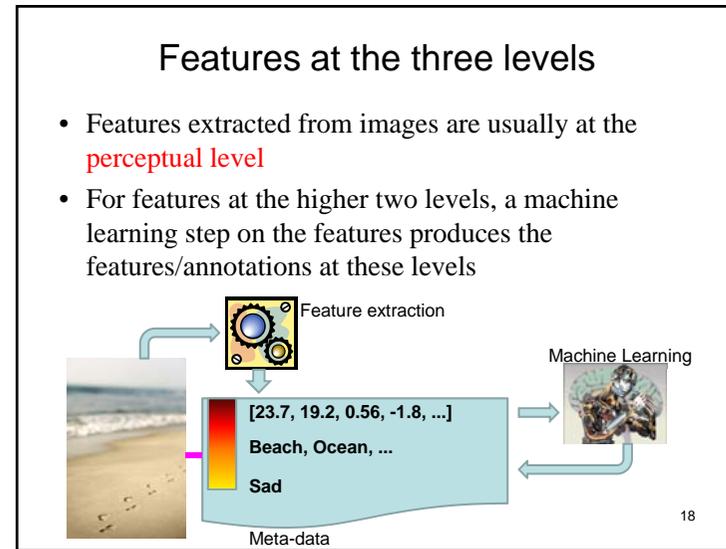
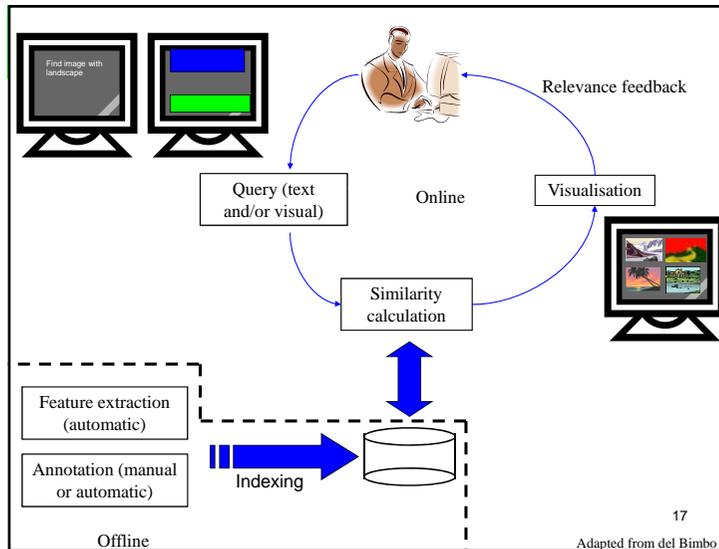
- Generally, the word “content” often has had different meanings in the computer science and humanities disciplines:
  - Computer scientists generally use “content” to mean perceptual properties
  - In the humanities, “content” normally refers to meaning
- Result – some disappointment with *content-based* image retrieval systems!

15

### Image search systems

- Offline part:
  - Image feature calculation
  - Image annotation
  - Indexing
- Online part:
  - Query specification
  - Similarity calculation
  - Relevance feedback

16



- Query by image

- The user gives an example image and the system presents the images that are closest to this image using the chosen features and similarity function
- Sometimes the user has the possibility to set the importance of colour, texture, etc.



IKONA System, INRIA

- Query by painting

- The user draws the colours to be retrieved and their spatial arrangement.

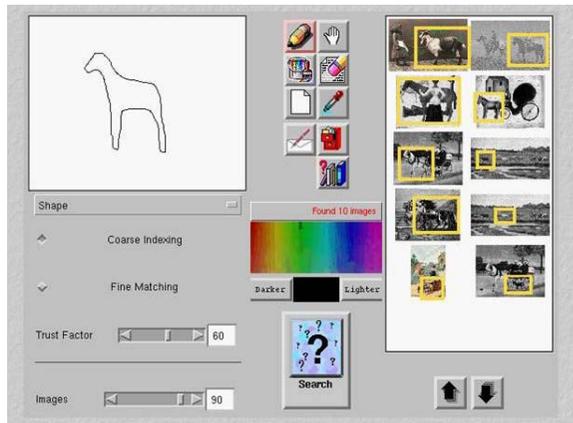


Picasso System, University of Florence

[Link to QBIC](#)

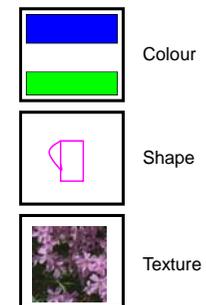
- Query by sketch

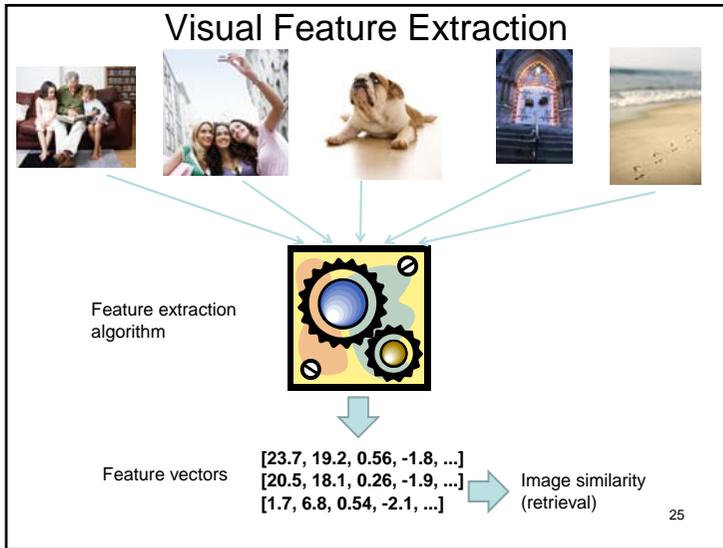
- The user sketches the shape of the object to be found.



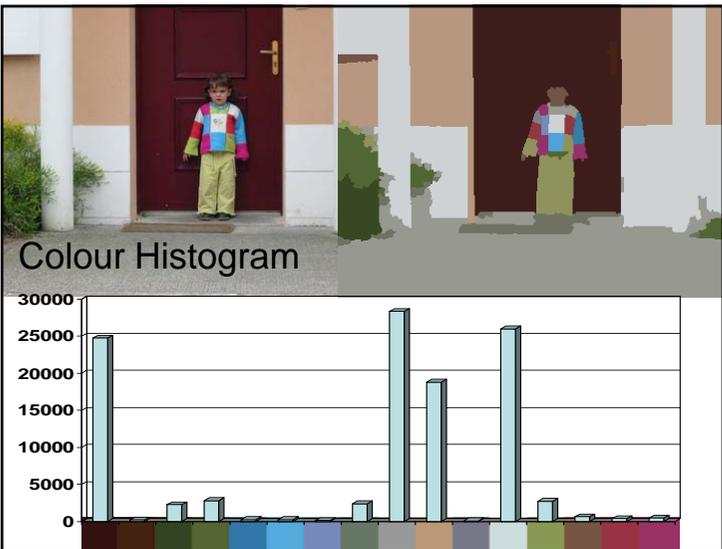
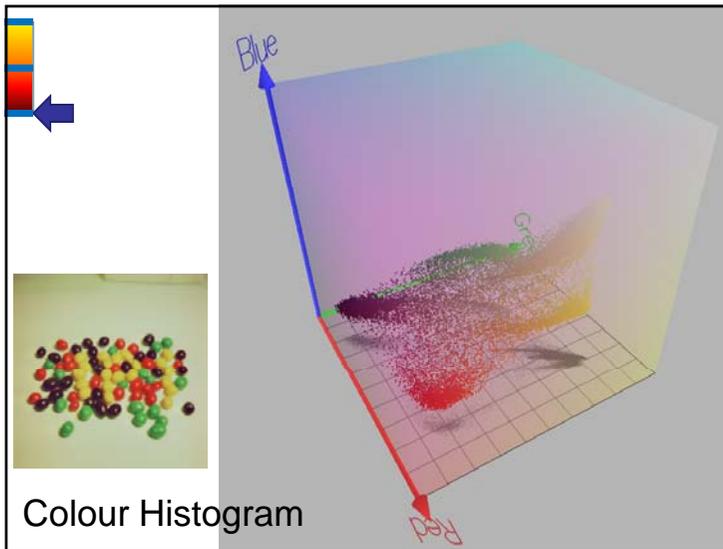
Picasso System, University of Florence

- The following perceptual properties are often used in query by example:





- ### Perceptual Features
- Colour features
  - Visual words
  - Perceptual descriptions
- 26



## Similarity by colour histogram



29

<https://www.rocq.inria.fr/cgi-bin/media/circario.cgi/demos>

## Problem with colour histograms

- Similar colour distribution, but not so similar content



## Colour option already in Google

Google Bilder

Suche:  Bilder-Suche [Erweiterte Bildsuche](#)

Ergebnisse 1 - 20 von etwa 332.000.000 (0,22 Sekunden)

• Beliebige Größe  
 Klein  
 Groß  
 Symbol  
 Größe als...  
 Genau...

• Beliebiger Typ  
 Gesicht  
 Foto  
 Clipart  
 Lineart

• Beliebige Farben  
 Alle Farben  
 Schwarzweiß  
 Bestimmte Farbe

• Beliebiger Typ  
 Gesicht  
 Foto  
 Clipart  
 Lineart

• Beliebige Farben  
 Alle Farben  
 Schwarzweiß  
 Bestimmte Farbe

• Optionen zurücksetzen

31

Web Bilder Maps News Groups Blogger Google Mail More

Google bilder

Suche:  Bilder-Suche [Erweiterte Bildsuche](#)

Ergebnisse 1 - 20 von etwa 10.300.000 (0,28 Sekunden)

• Beliebige Größe  
 Klein  
 Groß  
 Symbol  
 Größe als...  
 Genau...

• Beliebiger Typ  
 Gesicht  
 Foto  
 Clipart  
 Lineart

• Beliebige Farben  
 Alle Farben  
 Schwarzweiß  
 Bestimmte Farbe

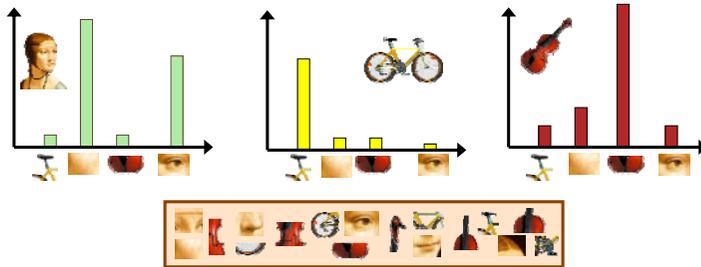
• Optionen zurücksetzen

32



## Bag of Visual Words

- Independent features
- Histogram representation



37  
Slide from Li Fei-Fei,  
CVPR 2007 Tutorial

## Video Google

- Approach to matching duplicates
- A system which aims to retrieve objects or scenes in a film
- <http://www.robots.ox.ac.uk/~vgg/research/vgoogle/>

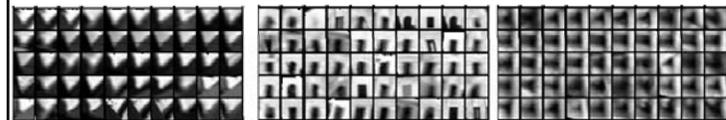
- Interest points and their surrounding regions are detected



- The SIFT features for each region are calculated (128 dimensional feature vector)

39  
From <http://www.robots.ox.ac.uk/~vgg/research/vgoogle/>

- The feature vectors are clustered to form **visual words** (with K-means clustering)



- These visual words can be used to index the frames in which they occur
- Text retrieval methods can then be used to find the frames which are being asked for

40  
From <http://www.robots.ox.ac.uk/~vgg/research/vgoogle/>

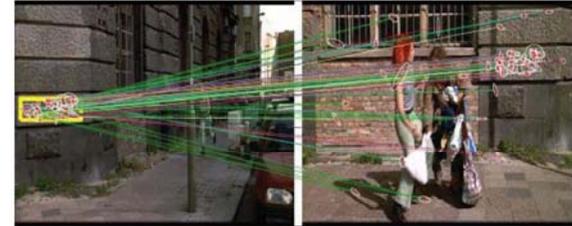
- Example of a query:
  - The user selects a query region



41

From <http://www.robots.ox.ac.uk/~vgg/research/vgoogle/>

- Images containing the corresponding visual “words” are found:

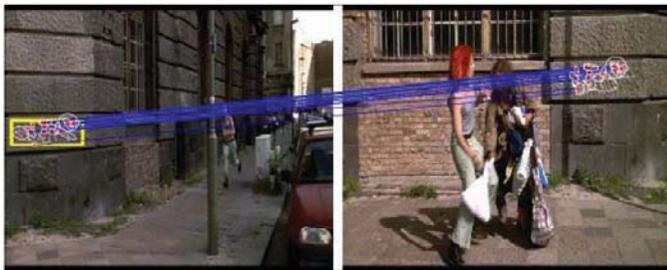


- Words which occur very often in the whole dataset are removed (stop words):



42

- Spatial consistency is considered to further reduce false matches:



- Another example is shown on the next slide

43

From <http://www.robots.ox.ac.uk/~vgg/research/vgoogle/>



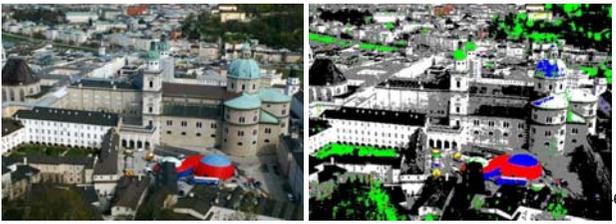
## ← Words describing perceptual features

- Colour naming
  - Map colours to colour names
- Texture naming
- Shape naming

45

## Colour naming

- 11 basic colour terms have been defined through English linguistic analysis:
  - black, blue, brown, green, grey, orange, pink, purple, red, white, yellow
- Every RGB value is mapped to one of these colours



- Colours learnt from images retrieved using Google image search.
- Retrieved using queries “red+color”, “green+color”, etc.

Figure 1. Google-retrieved examples for color names. The red bounding boxes indicate false positive images. The same images can be retrieved with various color names, such as the flower image which appears in the red and the yellow set.

- Results in a Look-Up-Table mapping every RGB triplet to one of the 11 colours.

For details see (and image from): van de Weijer, Schmid, Verbeek, “Learning Color Names from Real-World Images”, CVPR 2007



## Perceptual features summary

- Can be good for some applications, such as choosing images for a page layout in a magazine
- In general, image similarity by perceptual features cannot satisfy many information needs
- Usually perceptual features are useful in combination with higher level features

49

## Image search at the semantic level

- Need to know what things (**concepts**) are in the image
- Can get these in two ways
  - Manually
  - Automatically (object recognition)

50

## Text Annotation

- Is more difficult than it appears!
- Can be done as:
  - Free-text annotation
    - E.g. "Fishing vessels in Northern Ireland"
  - Keyword annotation
    - Free choice of keywords
    - Controlled Vocabulary
  - Hierarchy, Ontology
    - In its simplest form, a set of keywords with a hierarchical structure, but there exist more complicated relations.

Example from the Iconclass hierarchy  
(28 000 definitions in total)

4	Society, Civilization, Culture
47	crafts and industries
47D	machines; parts of machines; tools and appliances
47D3	machine driven by wind
47D31	windmill

## Vocabularies

- It is more useful, but also more work, to get images annotated by concepts from a controlled vocabulary
  - Avoids typos and spelling mistakes
  - Can help resolve word ambiguity
- A structure in the vocabulary (e.g. hierarchy) can
  - Help in relating concepts to each other (e.g. a dog is an animal)
  - Resolve synonyms, etc.
  - Map words in multiple languages to the same concept
- Examples of vocabularies
  - WordNet
  - Wikipedia/dbpedia
  - ...

54

Search on Yotophoto using the keyword "leopard"

55

### Example: GettyImages keywords

- The GettyImages Archive (<http://www.gettyone.com>) has a list of keywords used for annotating their image archive.
- Goes beyond objects, e.g.:
  - Life Events:
    - Childbirth, Recruitment, Working Overtime, Office Romance, Wedding, Retirement, Lifestyles, Social Issues, Age Contrast, ...
  - Concepts:
    - Friendship, Funky, Futility, Gambling, Gratitude, Greed, Growth, ...
  - Individual People:
    - One Person, One Man Only, One Woman Only, One Baby Boy Only, One Baby Girl Only, One Boy Only, One Teenage Boy Only, ...

See this document for many more keywords:  
[http://corporate.gettyimages.com/marketing/m01/PDF/Keyword\\_UK\\_1\\_Jan\\_05.pdf](http://corporate.gettyimages.com/marketing/m01/PDF/Keyword_UK_1_Jan_05.pdf)

57

### Location-specific annotation

- Most often, a list of concepts is assigned to the image
- Location-specific annotation means assigning the concepts to image regions

58

### Manual annotation

- Manually annotated images useful for:
  - Search
  - Training object recognition algorithms
  - Evaluating search algorithms
- Always too few annotated images available!
- Who can annotate the images?

59

## Students

Image (images/00/25.jpg)

**Freelast Annotation**

**Title:** Plaza de Armas

**Description:** Plaza de Armas; yellow house with white columns in background; two palm trees in front of house; cars parked in front of house; woman and child walking over the square.

**Notes:** The Plaza de Armas is one of the most visited places in Cochabamba. The locals are very proud of the colourful buildings.

**Titel:** Plaza de Armas

**Beschreibung:** Plaza de Armas, gelbes Haus mit weißen Säulen im Hintergrund; zwei Palmen vor dem Haus; geparkte Autos vor dem Haus; Frau und Kind spazieren über den Platz.

**Anmerkungen:** Der Plaza de Armas ist einer der populärsten Plätze Cochabambas. Die Einheimischen sind sehr stolz auf die bunten Gebäude.

**Titulo:** Plaza de Armas

**Descripcion:** Plaza de Armas; casa amarilla con dos columnas blancas al fondo; dos palmeras delante de la casa; coches aparcados delante de la casa; mujer con hijo caminando por la plaza.

**Observaciones:** La Plaza de Armas es una de las plazas más visitadas en Cochabamba. La gente es muy orgullosa de las casas multicolores.

taken by André Kwitz, 1 February 2003, Cochabamba (Bolivia)

IAPR-TC12 dataset

- Photo sharing sites like Flickr also provide many annotations (tags), but these are generally noisy

**A 2 Headed Zebra?**

A pair of zebras enjoying the fresh, spring grass at the African Lion Safari. I watched them as they passed each other going in opposite directions, trying to get a "2 headed Zebra" picture LOL.

Enjoy!  
Here's a link to the African Lion Safari. It's an awesome place, where the animals roam free and the humans are caged up in their vehicles...the way it should be.

**Tags:** Zebra • zebra poop • LOL :)) • African Lion Safari • grass • stripes • impress@Beauty

## Game players

- Games with a purpose (GWAP)
- ESP Game

score: 400

time: 1:54

What do you see?

task words: insect, tree, moth

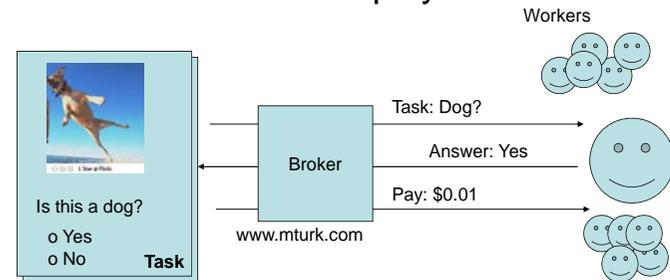
guesses: brown

Author prediction: 5000 people continuously playing the game could assign a label to all images indexed by Google in 31 days.

<http://www.gwap.com>

66

## Virtual employees



- Amazon Mechanical Turk
- E.g. 100 000 images × \$0.01 = \$1000
- <http://visionpc.cs.uiuc.edu/~largescale/index.html>
- Potential problem: Variance in judgements, Bad judgements

From Sorokin and Forsyth

## Summary: who annotates images?

- People who you have influence over
  - Students
  - Annotation parties!
- People who want to sell photos or become famous
  - Professional and amateur photographers
- People with spare time
  - GWAP players
- People who want to earn money
  - Virtual employees (e.g. Mechanical Turk)
- Others?

68

## If nobody annotates your images...

- Take the words where you can find them...
- The image search techniques used by Google, Yahoo, etc. are basically text searches
- The algorithms look for the search text in
  - Names of the image files
  - Text around images in web pages
  - *alt* attribute in the *image* tag
  - ...

69

Simple search term

The screenshot shows a Google image search for the term "sunset". The search bar contains "sunset" and the search button is labeled "Google Suche". Below the search bar, there are navigation links for "Web", "Bilder", "Glossar", and "Verzeichnisse". The search results are displayed in a grid format, showing various sunset images. Each image is accompanied by its filename and a brief description of the source website. For example, the first result is "sunset.jpg" from "www.crh.noaa.gov/arc/photos.shtml".

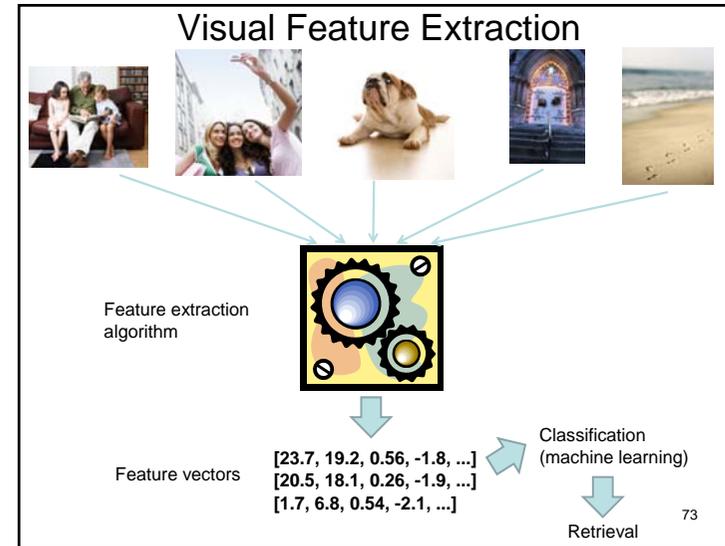
The screenshot shows a Google image search for the term "sunset". The search bar contains "sunset" and the search button is labeled "Suche". Below the search bar, there are navigation links for "Web", "Bilder", "Videos", "Maps", "News", "Übersetzer", "Google.Mail", and "Mehr". The search results are displayed in a grid format, showing various sunset images. On the left side of the page, there is a sidebar with navigation options: "Bilder", "Videos", "News", "Mehr", "Alle Größen", "Alle Typen", "Alle Farben", "Standardsicht", and "Größen anzeigen".

71

## Automatic annotation

- How well can an algorithm determine what is in an image?
- Usually involves some sort of machine learning on the features extracted from the images
  - The classifier output is indexed as an image feature to be used in retrieval
- Detecting if there is a face or faces in images works well, even recognising the faces in some cases
- Detecting a large number of general objects in images is an unsolved problem

72



## Searching for portraits

A screenshot of a Google search for 'otto wagner'. The search results page shows a grid of images. An orange arrow points from the search bar to the first portrait in the grid. The left sidebar shows various filters like 'Alle Typen' and 'Alle Farben'. The top navigation bar includes 'Bilder', 'Videos', 'Maps', 'News', 'Übersetzer', 'Google Mail', and 'Mehr'.

## Face detection in image search

A screenshot of a Google search for 'otto wagner' showing face detection. The search results page is similar to the previous one, but the grid of images is overlaid with a blue 'Gesicht' (Face) detection box. The detected faces are highlighted with small blue boxes, showing that the search engine has identified human faces in the image results.

## Challenges in general object recognition

- Objects can look different when seen from different directions and from different distances
- Objects of the same **category** (e.g. bicycle) can still show variability in appearance
- Objects can be occluded

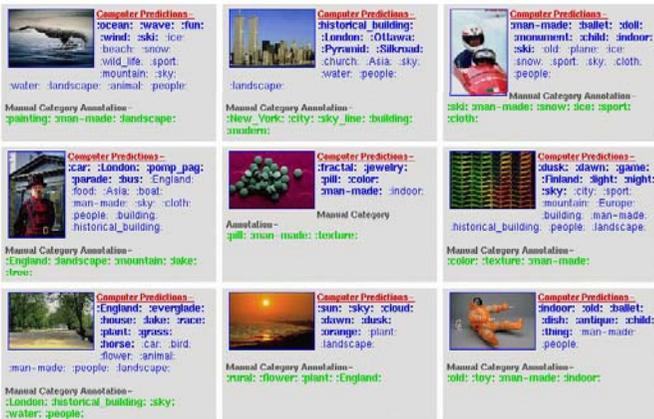


## Some systems for automatic image annotation

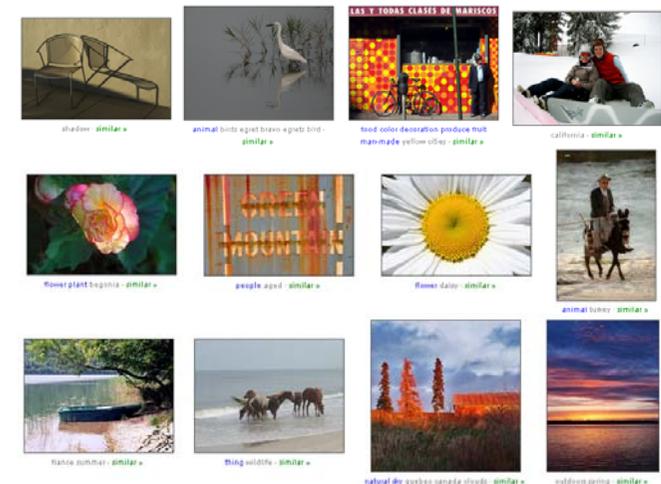
- The *a*-LIP system (Li and Wang, 2003):
  - Training phase:
    - Uses a training set of images labelled with keywords
    - Extracts colour and texture features
    - A pattern recognition algorithm learns which keywords are associated with which feature values
  - Usage:
    - The same features are extracted from each image to be annotated
    - The trained pattern recognition algorithm associates keywords with these features
- The ALIPR system (Li and Wang, 2006) is a real-time version of the above system. <http://alipr.com>

77

## Some a-LIP system results



## Some Alipr results

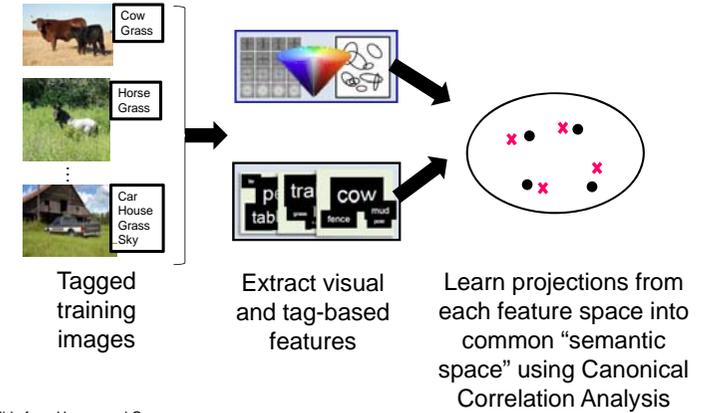


## Using words and image features

- The latest approaches often attempt to combine retrieval by both text and image features
- Example from S. J. Hwang and K. Grauman, *Accounting for the Relative Importance of Objects in Image Retrieval*, In Proc. British Machine Vision Conference (BMVC), Aberystwyth, UK, September 2010

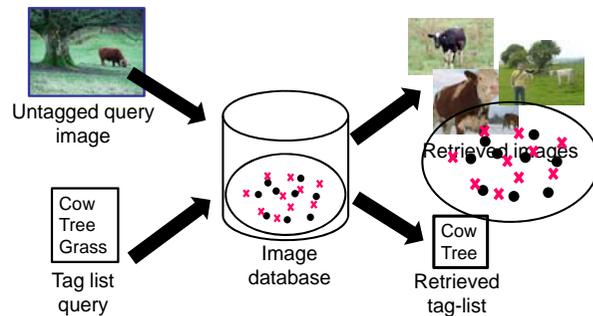
80

## Approach overview: Building the image database



Slide from Hwang and Grauman

## Approach overview: Retrieval from the database



- Image-to-image retrieval
- Tag-to-image retrieval
- Image-to-tag auto annotation

Slide from Hwang and Grauman

## Image-to-image retrieval results



Slide from Hwang and Grauman

### Image-to-image retrieval results

Query Image 

Visual only 

Hwang and Grauman method 

Slide from Hwang and Grauman

### Image-to-tag auto annotation results

	Boat Person Water Sky Rock		Tree Boat Grass Water Person
	Bottle Knife Napkin Light fork		Person Tree Car Chair Window

Slide from Hwang and Grauman

## Image search at the impressions, emotions and meaning level

- Image Affect
- Also involves machine learning, but here the features are chosen to measure aspects related to a higher level of image content

86

### Image Affect

- Is it possible to determine the emotional effect of on image on a person using computational methods?

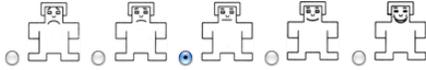
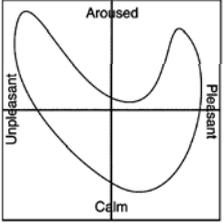
		
		

87



### How to measure affect?

- Individual vs. common
- Psychological model
  - **Pleasure (Valence)**

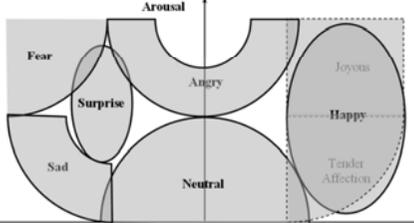
- **Arousal**



- (Dominance)

### Emotional Categories

- Emotional categories by Mikels et al.:
  - Amusement
  - Awe
  - Excitement
  - Contentment
  - Anger
  - Disgust
  - Fear
  - Sad



### Feature extraction

- Color
- Texture
- Composition
- Content





## Psychological Colour

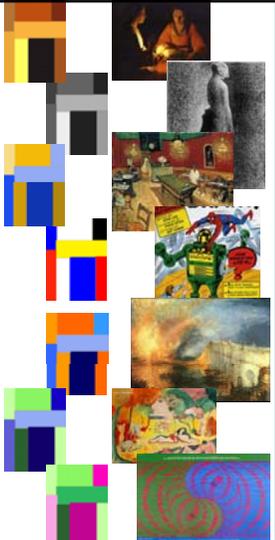
- Psychological Experiments done by Valdez and Mehrabian, 1994
- 250 students to rated their emotional response to single colours on the Pleasure – Arousal – Dominance scale.
- Regression produced surprisingly high correlations:
 
$$\begin{aligned} \text{Pleasure} &= 0.69 \text{ Brightness} + 0.22 \text{ Saturation} \\ \text{Arousal} &= -0.31 \text{ Brightness} + 0.60 \text{ Saturation} \\ \text{Dominance} &= 0.76 \text{ Brightness} + 0.32 \text{ Saturation} \end{aligned}$$
- Average over the whole image

## Examples

	-0.72	0.30	0.66
Pleasure			
Arousal	-0.25	-0.02	0.32

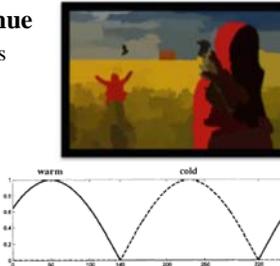
## Itten Contrasts

- Johannes Itten defined 7 contrasts used by artists to produce optical and psychological sensations:
  - contrast of saturation**
  - contrast of light and dark**
  - contrast of complements** - formed by visually complementary colors - colors that are opposite each other on the color wheel,
  - contrast of hue** - the maximal contrast of complements is the contrast of hues,
  - contrast of warm and cold** - formed by the combinations of colors that are considered “warm” and “cold”,
  - contrast of extension** - proportion of the sizes of the color patches in relation to the visual “weight” of their color,
  - simultaneous contrast** - when contrasting colors are missing, boundaries of the color patches “vibrate”, creates color illusions or invokes a sense of tension.



## Features inspired by Itten contrasts

- Contrast of **saturation** and of **light and dark**
  - Standard deviation of the region saturations and lightnesses
- Contrast of **complements** and of **hue**
  - Standard deviation of the region hues
  - Max and average difference of hues between all region pairs
- Contrast of **warm and cold**
  - Contrast from membership values
  - Area of warm and cold colours
- Simultaneous contrast**
  - Basically means that complementary contrast is low, but difficult to find beyond this





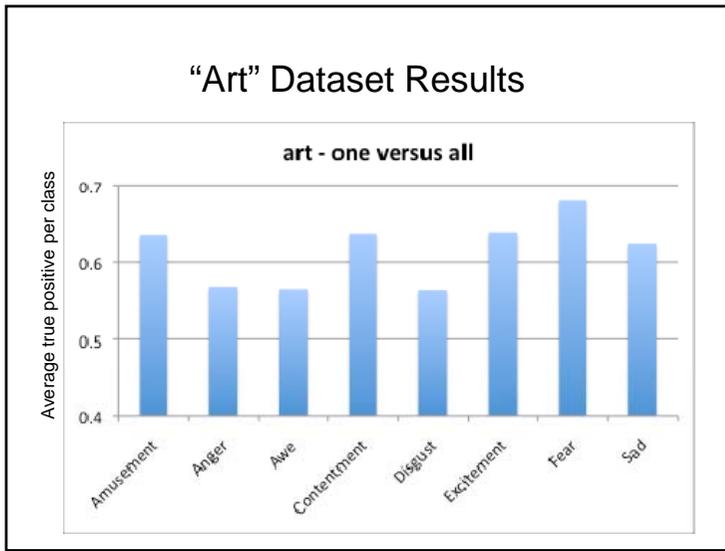
## Content Features

- Human Faces
  - Viola-Jones frontal face detection
- Skin
  - 



## “Art” dataset

- “Art” photos from an art-sharing web-site
  - “art” = images with intentional expression & conscious use of design
  - Artists use tricks (or follow guidelines) to create the proper atmosphere of their images
  - Data set assembled by searching for images with emotion words in image title or keywords/tags
  - Images are from the art-sharing web community deviantArt.com
  - 807 images

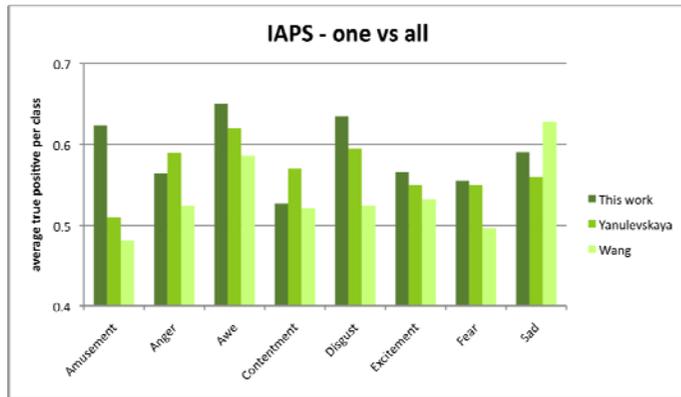



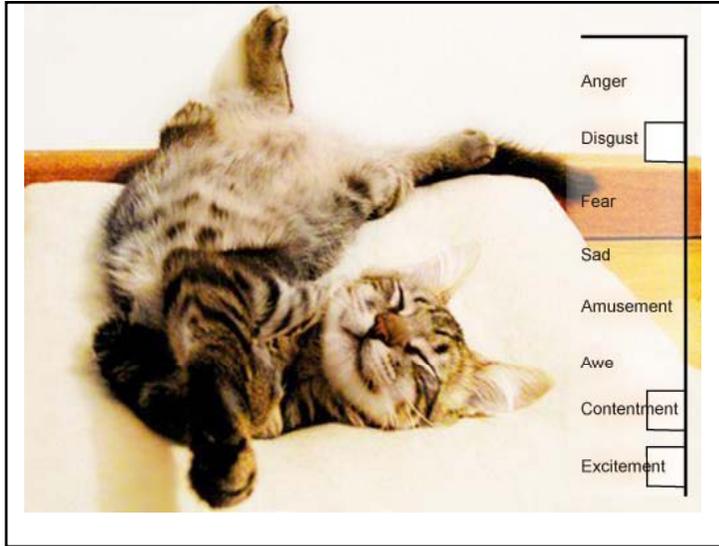
## IAPS Dataset

- IAPS – International Affective Picture System
  - 369 general, “documentary style” photos, covering various scenes
    - e.g. insects, puppies, children, poverty, diseases, portraits, etc.
  - Rated with affective words in psychological study with 60 participants



## IAPS Dataset Results





## Summary

- Three levels of image content:
  - Perceptual
  - Semantic
  - Impressions, emotions and meaning
- Retrieval by perceptual features is easy, but not always useful
- Algorithms for detecting objects are improving, but crowd-sourcing is a valid alternative
- Research is starting on impressions and emotions. Automatically understanding image meaning is a challenge

106