Gestion de grandes bases d'images

équipe LinkMedia

Laurent Amsaleg (CNRS)

15 mars 2016

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTEMES ALÉATOIRES

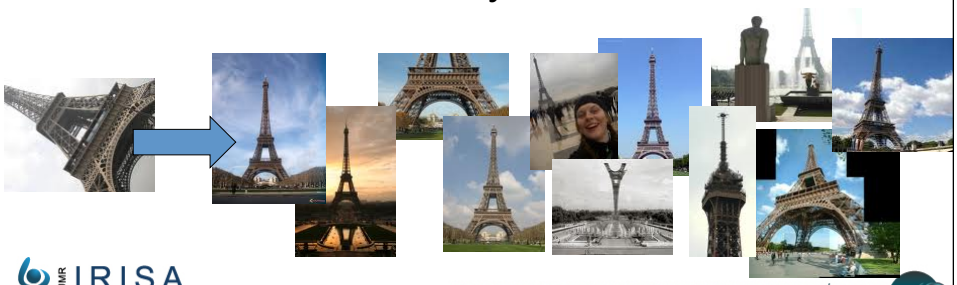
UMR IRISA

Search engines

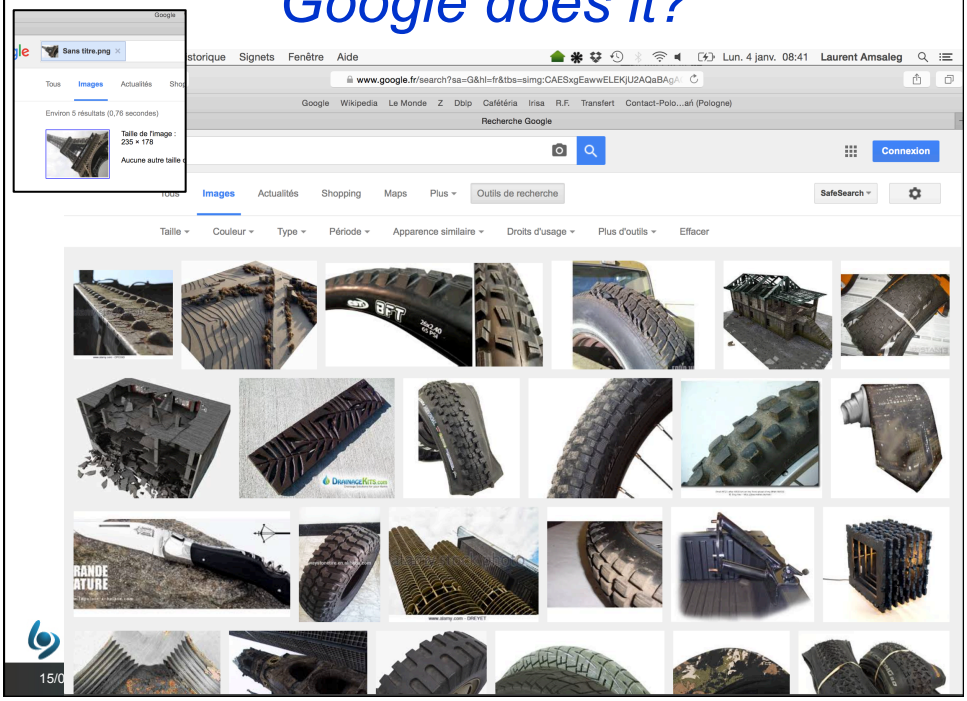
- the king is: keyword
- multimedia context:
  - manual annotations
  - ambiguity, painful task
  - good semantics
- Narrows the very *nature* of searches
  - poorly copes with some descriptions (faces!)
  - poorly supports similarity searches

Content based retrieval

- Image search
  - query= photo
  - result= similar photos
- no words here
- builds on visual similarity



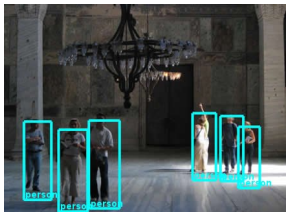
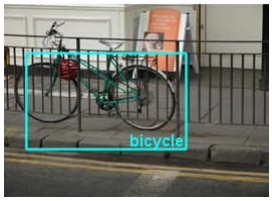
Google does it?

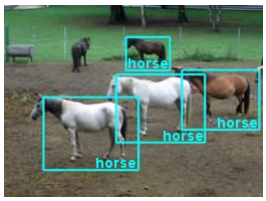


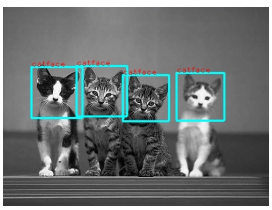
Content Analysis

- How to
  - compare 2 documents
  - compare 1 document and 10^9 others
- If we can compare then we can:
  - search, navigate, compare, visualize, summarize, recognize, detect, predict, annotate, classify, find correlations, “understand”
  - detect quasi-copies, alterations, re-use

- Object recognition

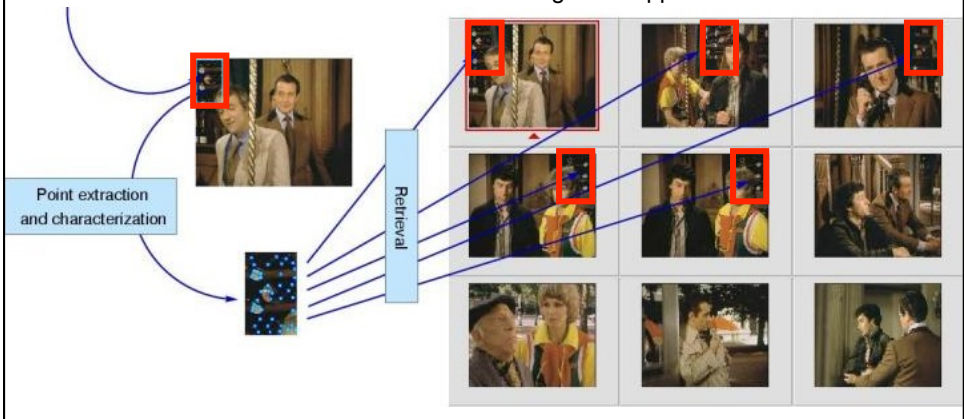




- Smart TV
  - action recognition

- Detail identification

find all scenes where these details of the background appear



Requirements: Scale, Dynamicity, Durability

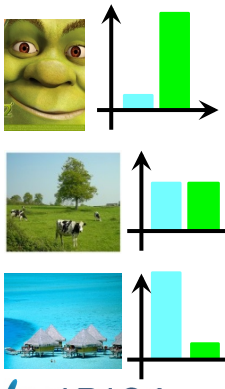
- you have:
  - 10 000 to 100 000 images
  - few 100s hours of music, of videos
- Flickr:
  - > 6.10^9 high resolution images
  - size+= $1,5.10^6$ pics per day
- Facebook:
  - > 1 000 billion pictures
  - size+= 200.10^6 pics per day

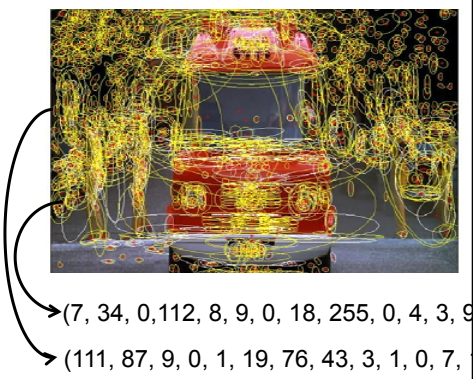
How does this work?

- Intensive CPU; Intensive Storage
- Focus on images
- two families of techniques
  - signal processing for describing contents
  - efficient search strategies

Analyzing contents

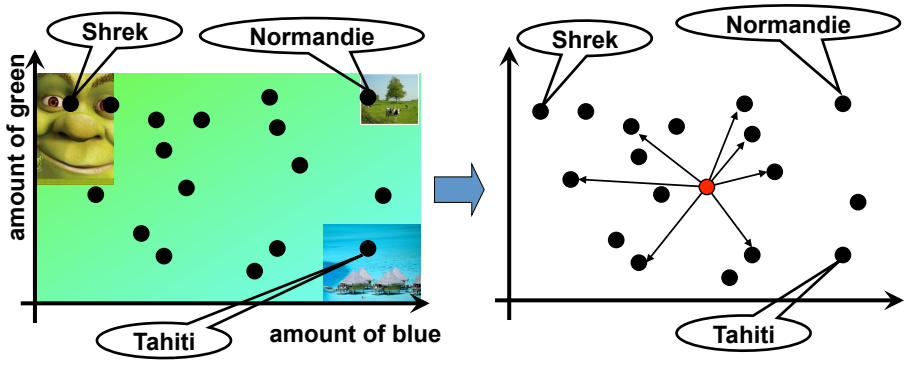
- feature extraction
  - high dimensional vectors
  - color histograms, local textures, …





Analyzing contents

- Similarity
  - distance calculations, nearest neighbors, L_2

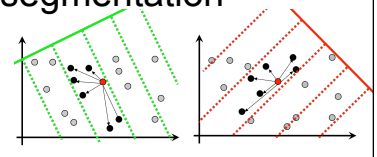


Efficient search techniques

- high dimensional indexing
  - approximate searches
  - close enough points grouped into cells
  - quickly identify the best cells
  - compute distances
  - refinements (geometrical consistency)

Efficient search techniques
state of the art

- LSH (Locality Sensitive Hashing)
  - multiple random projections + segmentation
  - L_2 over the union of collisions



- Bag of Visual Words
  - learn a visual vocabulary: k-means (k goes to millions)
  - images turned into a word occurrence vector
  - use of inverted files, stop-words, tf-idf

Efficient search techniques
local contributions

- NV-Tree
  - multiple random projections + segmentations
  - ranking, no distance calculations
  - unlimited scalability, dynamicity, durability
- DeCP
  - visual vocabulary
  - distributed architectures: Hadoop+MR; Spark
  - batch-oriented search

What we can do today

- 10M images on a laptop
- 100M on a desktop with disks
  - 43.10^9 descriptors, dimension 128, 4 TB on disks
- performance single server
  - indexing time: few days
  - response time: few ms (RAM), few seconds (disks)
- performance distributed settings
  - indexing time: 10h (Hadoop), 5.5h (Spark)
  - throughput / image: 30ms (Hadoop), 18ms (Spark)

Research directions

- Descriptions: very fine grain recognition
- Indexing: scalability, dynamicity, durability
- Accuracy: compensate approx; dim. curse
- Sequences: unsupervised motif discovery
- Multimedia: mono- vs. multi-modality
- Privacy & Security: spoofing, poisoning
- Tasks: classif, reco, labelling, diversity
- Tasks: analytics