Large Scale Multimedia Data:

Applications, processing and search techniques

Laurent Amsaleg (CNRS)

15 janvier 2014

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTEMES ALÉATOIRES

UMR IRISA

Multimedia data

- your pictures, videos, CD, podcasts
- pics from the whole family, from friends
- online contents of your local library
- TV, radio, INA archive, BBC archive …
- UGC “User Generated Contents”

Large scale?

- you have:
  - 10 000 – 100 000 images
  - few 100s hours of music, of videos
- INA has:
  - 60 years of TV + 70 years of radio
  - 5 000 000 hours to play with
  - recording 24*365: 120 channelsTV+radio
  - size+= 600 000 h/y

Large scale?

- 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

What for?

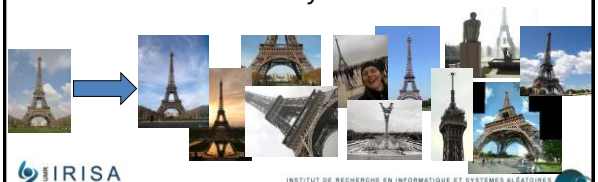
- search, navigate, compare, visualize, summarize, recognize, detect, predict, annotate, classify, find correlations, “understand”
- This is search engines! Google does it all.
  - yes and no

Search engines

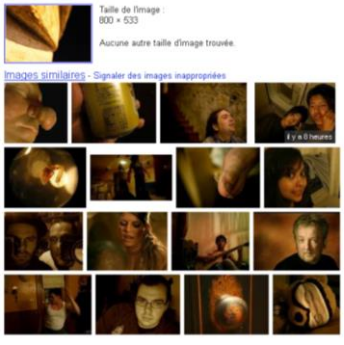
- the king is: keyword
- multimedia context:
  - manual annotations
  - ambiguity, painful task
  - good semantics
- Narrows the very *nature* of searches
  - poorly copes with some descriptions (faces!)
  - poorly supports similarity searches

Content based retrieval

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



what about Google?



Content Analysis

- How to
  - compare 2 documents
  - compare 1 document and 10^9 others
- If we can compare then we can:
  - create categories
  - summarize a photo collection
  - decompose a video into elementary units
  - do non-linear navigation within a video
  - …

Some applications

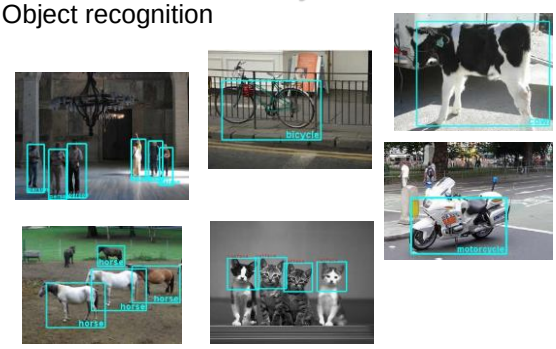
- For everyone
  - boost your images, your music
  - futur of digital TV: a smarty decoder
- Professional world
  - audience measurements
  - recommendation
  - visual diversity
  - security oriented: biometry, video surveillance, copyright enforcement, dismantling networks…

For everyone

- Enhance/Complete your holidays pics
- a posteriori geo-localization
- sort your family pictures
- Shazam
- “Query by humming”
- Audio mood

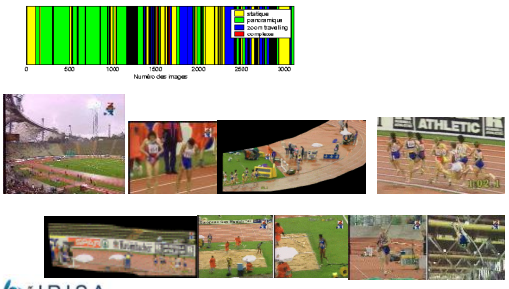
For everyone

- Object recognition



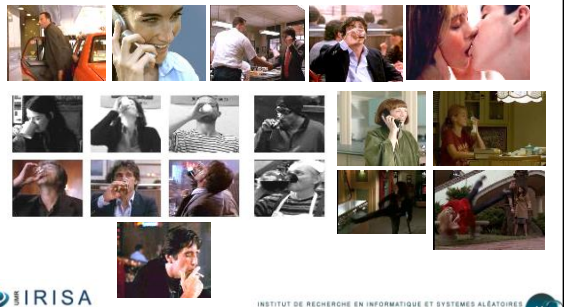
For everyone

- Smart TV
  - summary of sport events



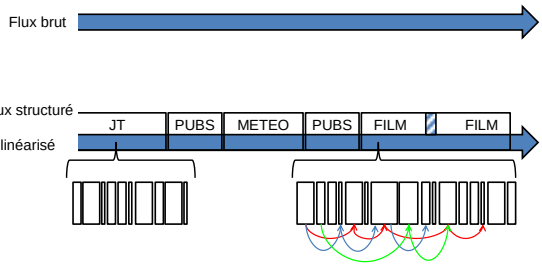
For everyone

- Smart TV
  - action recognition



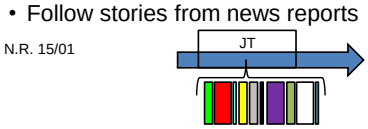
For everyone

- Smart TV
  - structuring and de-linearizing TV streams

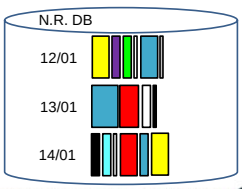


For everyone

- Smart TV
  - Follow stories from news reports



- new subjects
- evolution
- fusions
- navigation

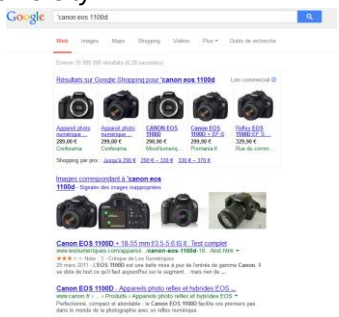


Professional

- audience measurements
  - speech time
  - ads, photos
- recommendations
  - collaborative filtering → content based recomm.
  - VOD, Amazon, …
- augmented reality
  - museums, glasses

Professional

- visual diversity



Professional

- visual diversity

Professional

- Tag propagation
- Duplicate removal
- Facebook has 1000 billion pics
  - how many duplicates?
  - how much does their management costs?

Professional

- Security
  - contents filtering; contents detection
  - video surveillance
  - biometry

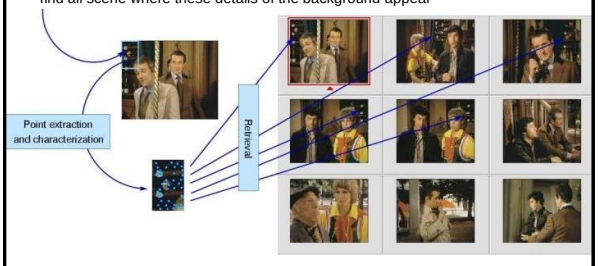
Professional

- Detecting / filtering contents

Professional

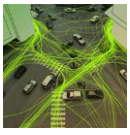
- Detail identification

find all scene where these details of the background appear



Professional

- Video surveillance

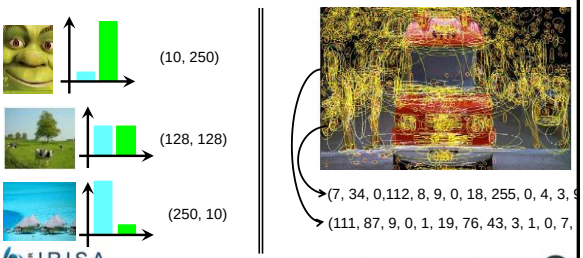


How does this work?

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

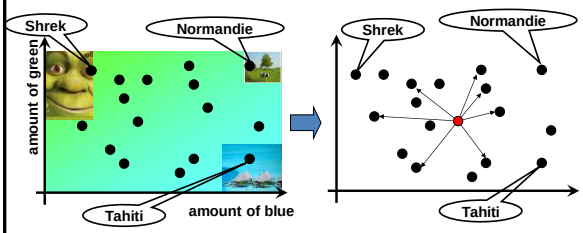
Analyzing contents

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



Analyzing contents

- Similarity
  - distance calculations, nearest neighbors, L_2

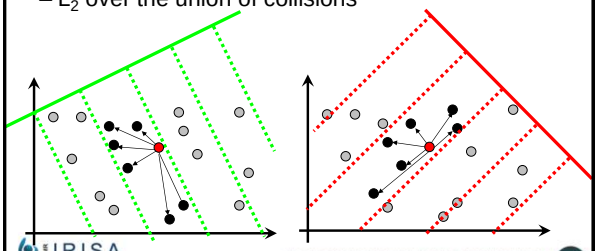


Efficient search techniques

- high dimensional indexing
  - approximate searches
  - close enough points groups into cells
  - quickly identify the best cells
  - compute distances
  - raffinements

Efficient search techniques

- LSH
  - random projections + segmentation = hashing
  - multiples hashing lines
  - L_2 over the union of collisions



Efficient search techniques

- Bag of Visual Words
  - learn a visual vocabulary: k-means (k goes to millions)

  - image turn into a word occurrence vector
  - use of inverted files, stop-words, tf-idf

What we can do today

- 10M images on a laptop
- 100M on a desktop
  - 30.10^9 descriptors, dimension 128
  - 4 To on disks
- performance
  - response time = few ms if RAM
  - response time = few seconds if disks
  - few ms if goal is throughput and batch-querying

Research directions

- Better descriptions
- Better indexing techniques
- Bridge the semantic gap
  - mono- vs. multi-modality
- Open problems
  - privacy
  - multi-modality
  - multimedia sequences at large scale