Learning Visual Similarity Measures for Comparing *Never Seen* Objects

Eric Nowak - Frédéric Jurie
LEAR Group
http://lear.inrialpes.fr/people/{nowak,jurie}

CVPR 2007 – Minneapolis

Motivation

This is a car
you have
never seen before...

... can you find it
in these images?

Motivation

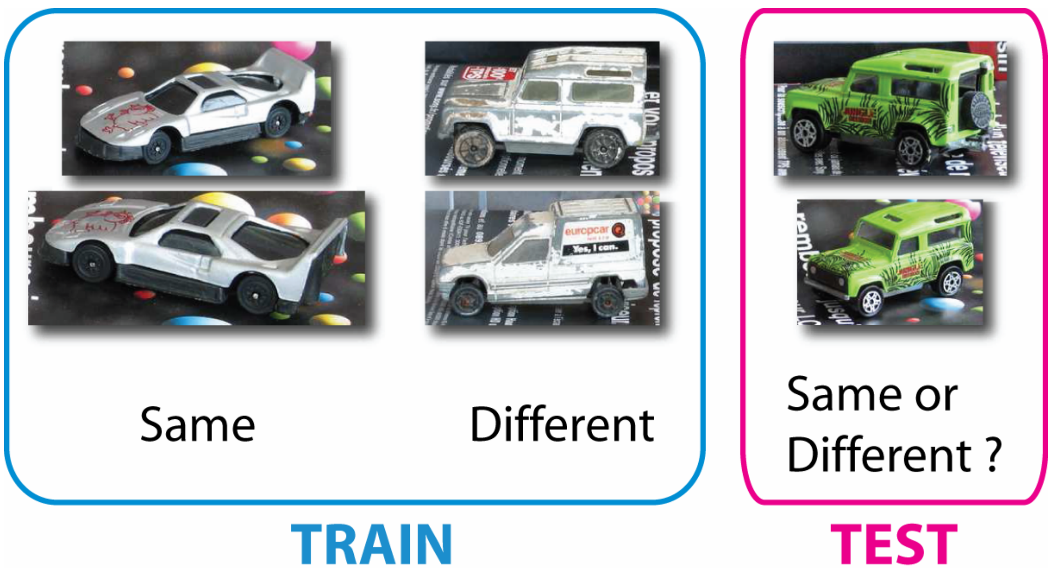
- Humans: specific knowledge (cars, faces, etc.)

⇒Recognize a car seen only once

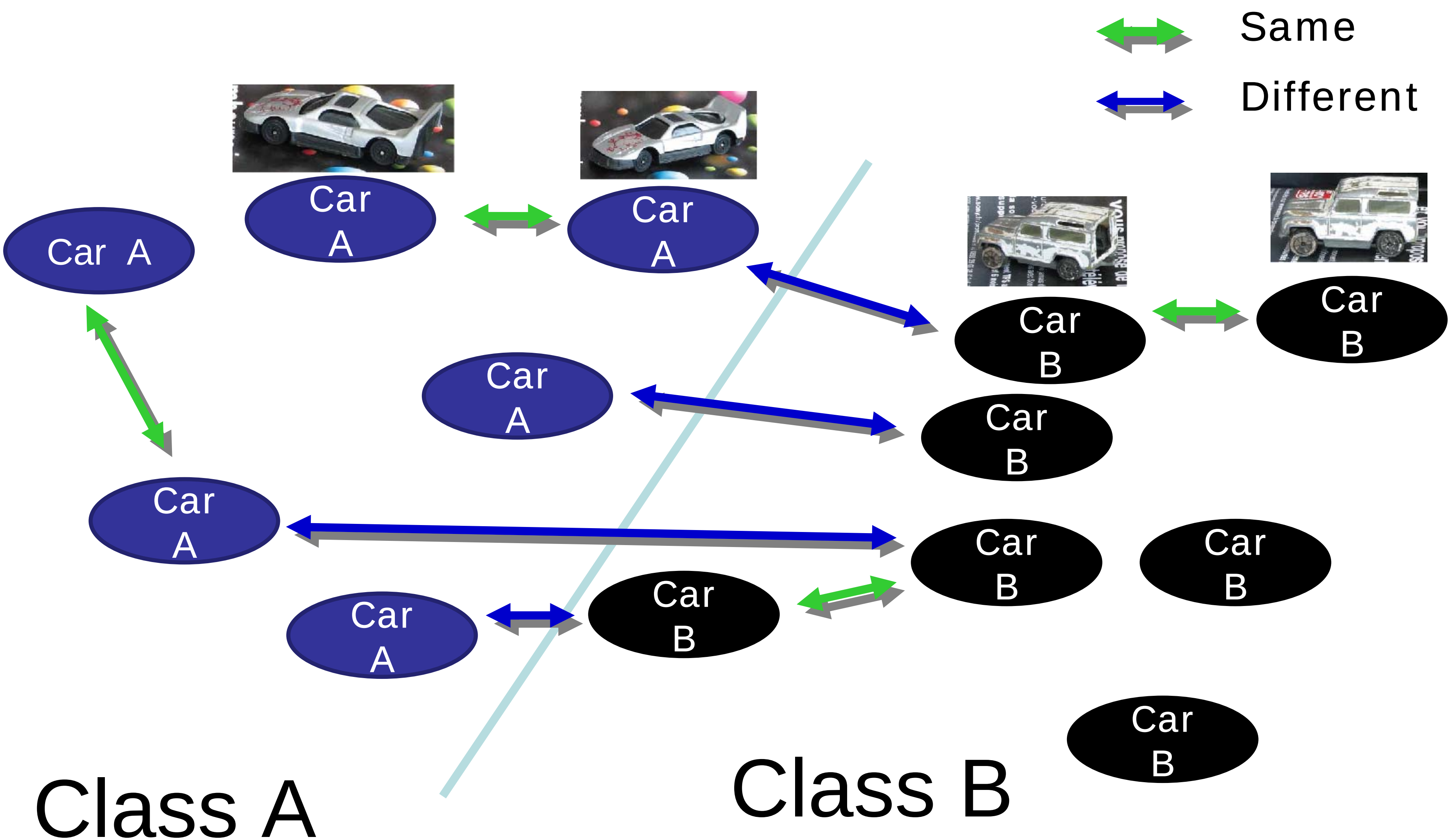
- Algorithm:
  also has to **integrate specific knowledge**

Our Goal

- Computing the visual similarity of two **never seen** objects
- Based on training pairs labeled “Same” or “Different” (equivalence constraints)
- Despite occlusions, changes in pose, light, background



Equivalence Constraints ?



Equivalence Constraints

- ☹ Less informative than Class Labels
  - “car model X and car model Y”
  - “same/different car model”
- ☺ Cheaper to obtain
  - e.g. space of class labels too large
- ☺ Deal with **new objects**.
  - Which model? CANNOT answer
  - Same or Different? CAN answer

How to compare images ?

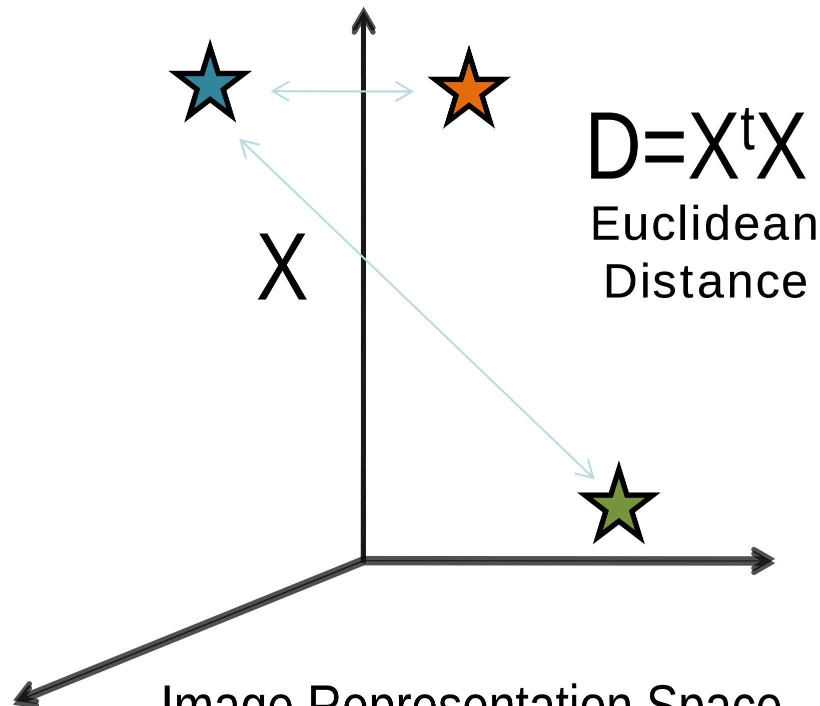
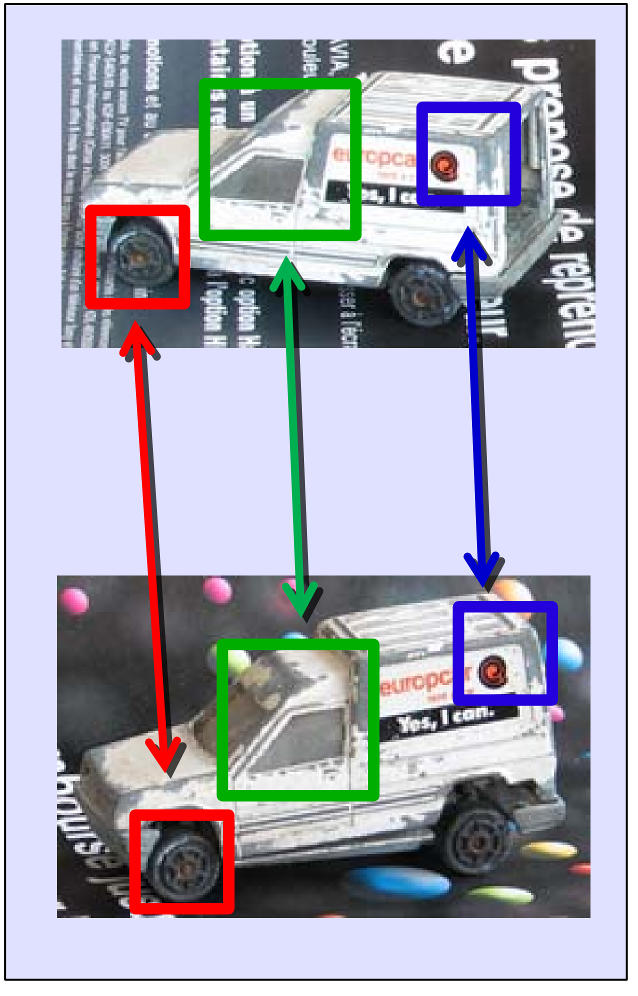


Image Representation Space
(Histograms, pixels, etc.)

Occlusions, view point changes, …

Global descriptors not adapted

How to be robust to occlusions, view point changes ?



Consider **local** representations

Get corresponding patch pairs

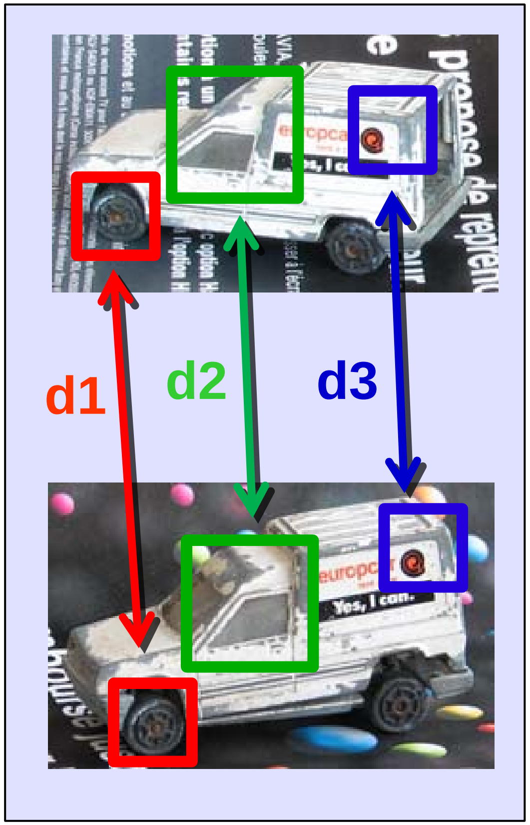
Vocabulary for Local Representations

- Text → vocabulary of ***words***
  “car”, “wheel”, “glass”, “motor”, …
- Image → vocabulary of ***visual words***

- Image pair → vocabulary of ***visual differences***

HOW *do the patches differ?*
=> Characterize local differences

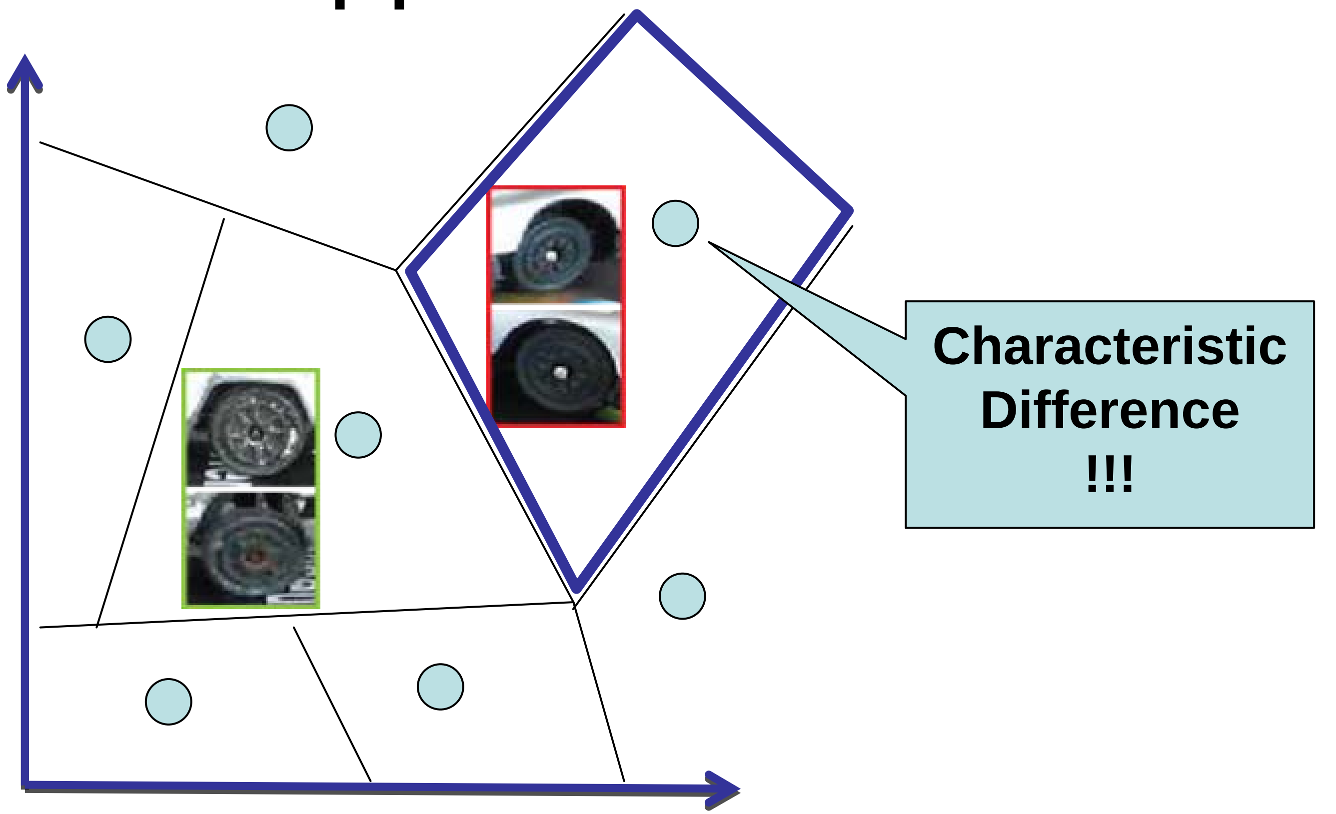
Characterizing local differences (Ferencz et al, ICCV 05)



D(I1,I2) = f(d1,d2,d3)

☹ d1, d2, d3: weak characterization of the differences

Characterizing local differences: our approach

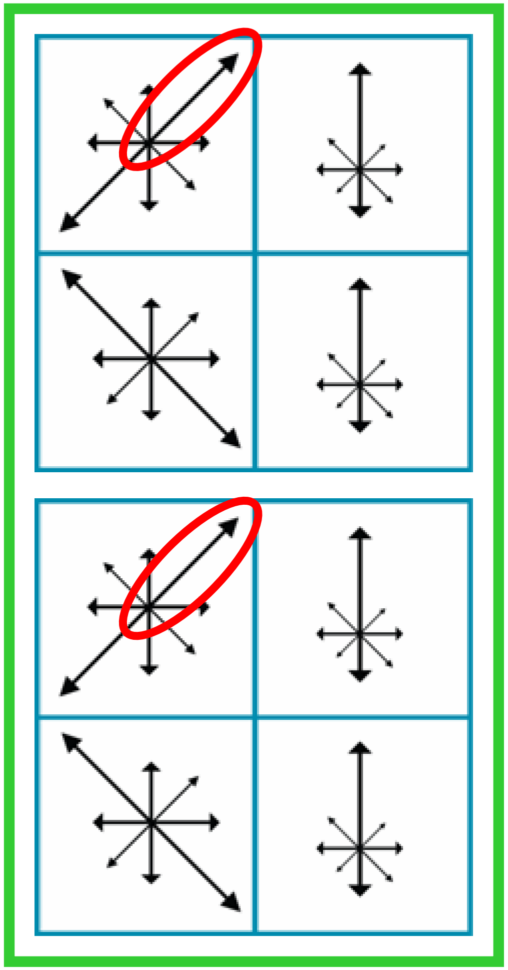


Patch Pair Space (ND)

☺ Much more information than a simple distance

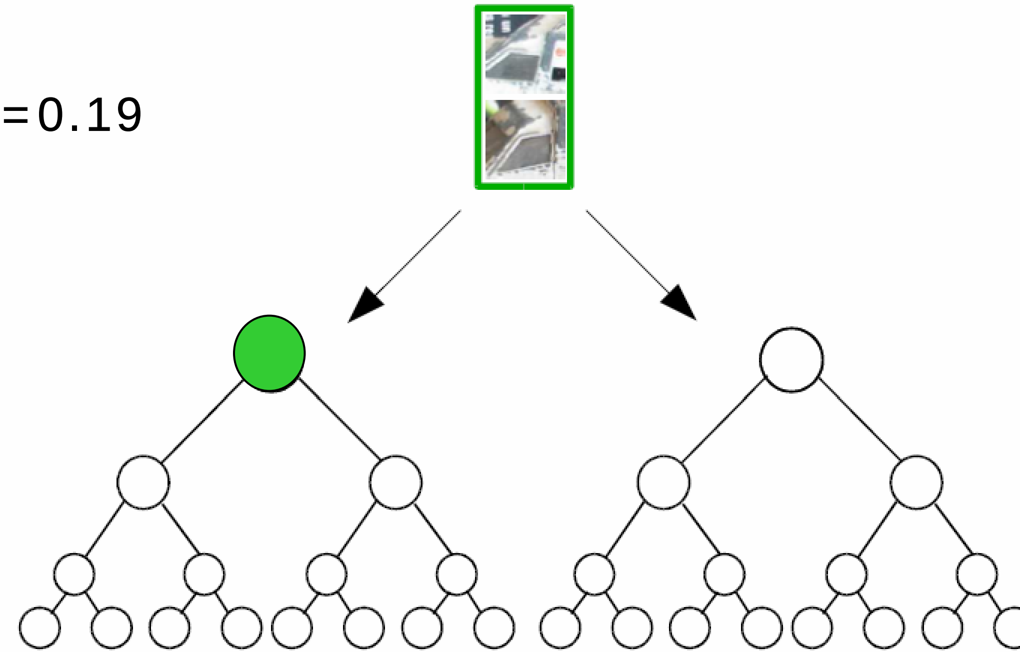
HOW TO COMPUTE THIS QUANTIZATION?

Patch pair quantization algorithm



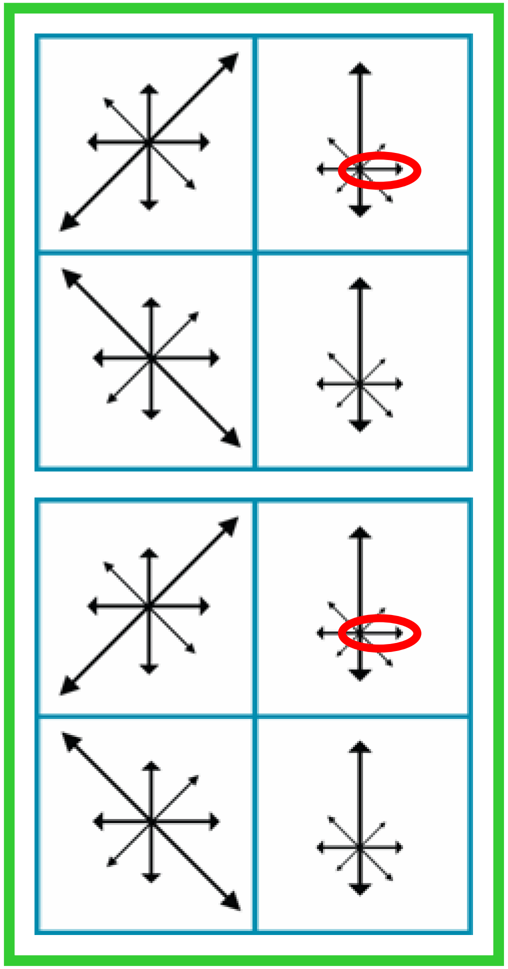
2 SIFT descrip.

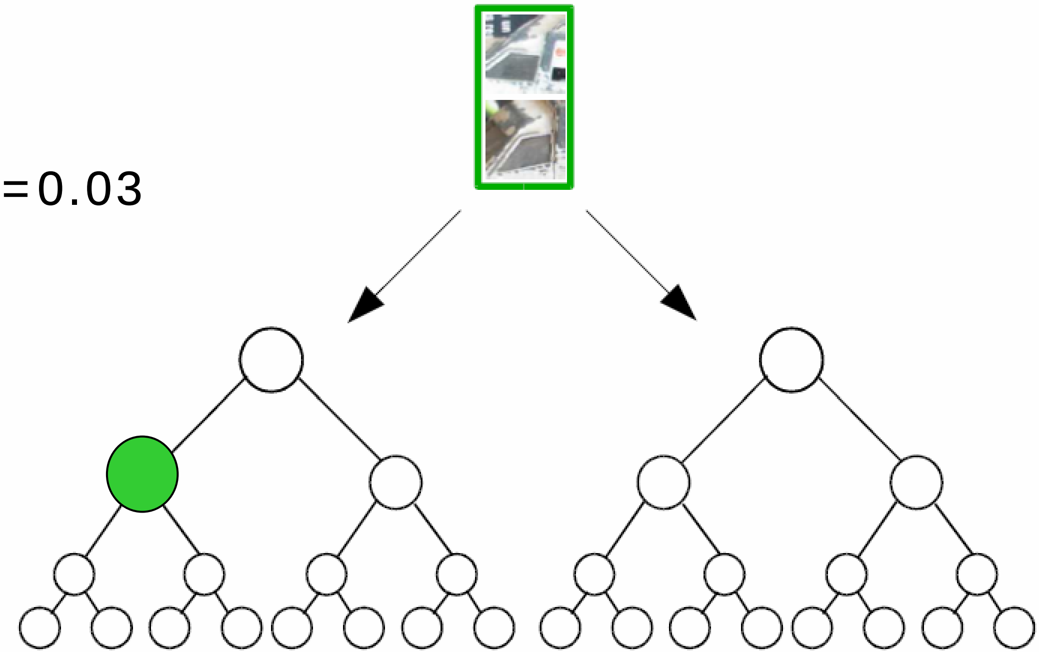
Thr=0.19



Both larger than 0.19 ?
False → left child
True → right child

Patch pair quantization algorithm

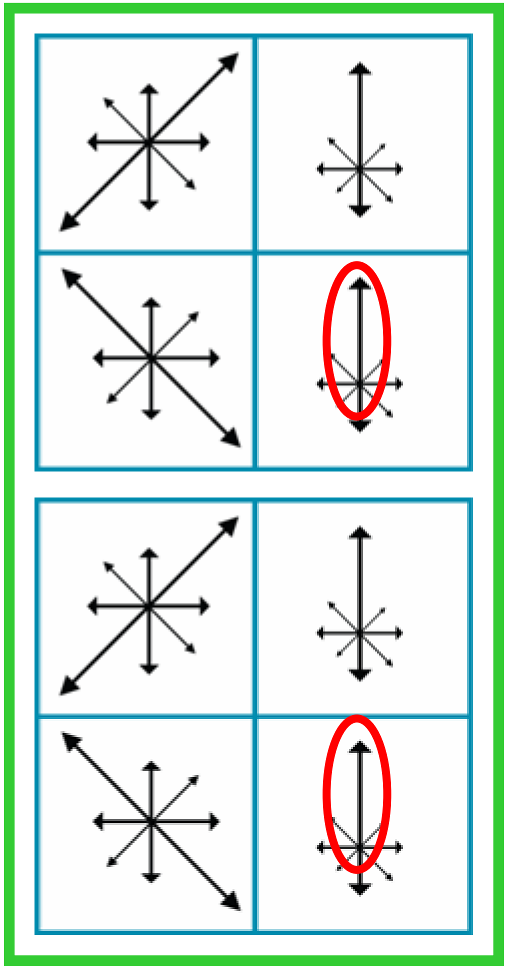


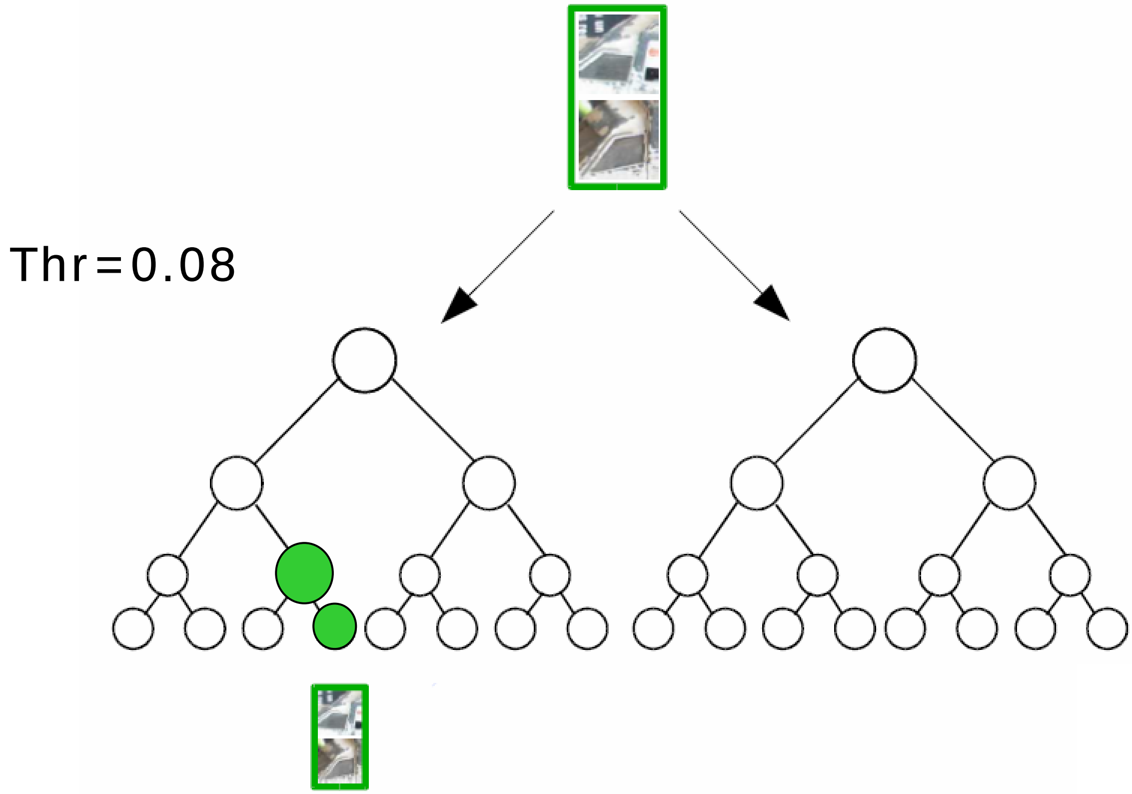


Thr=0.03

2 SIFT descrip.

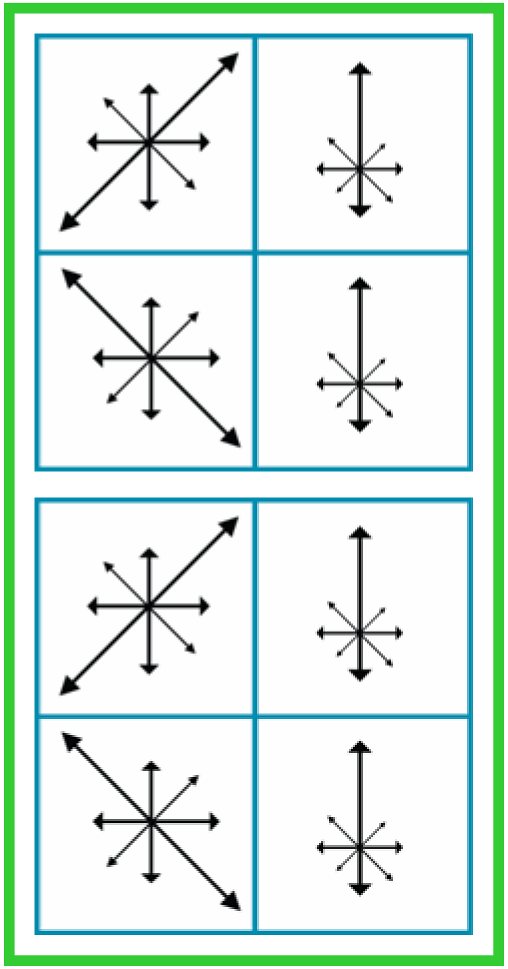
Patch pair quantization algorithm



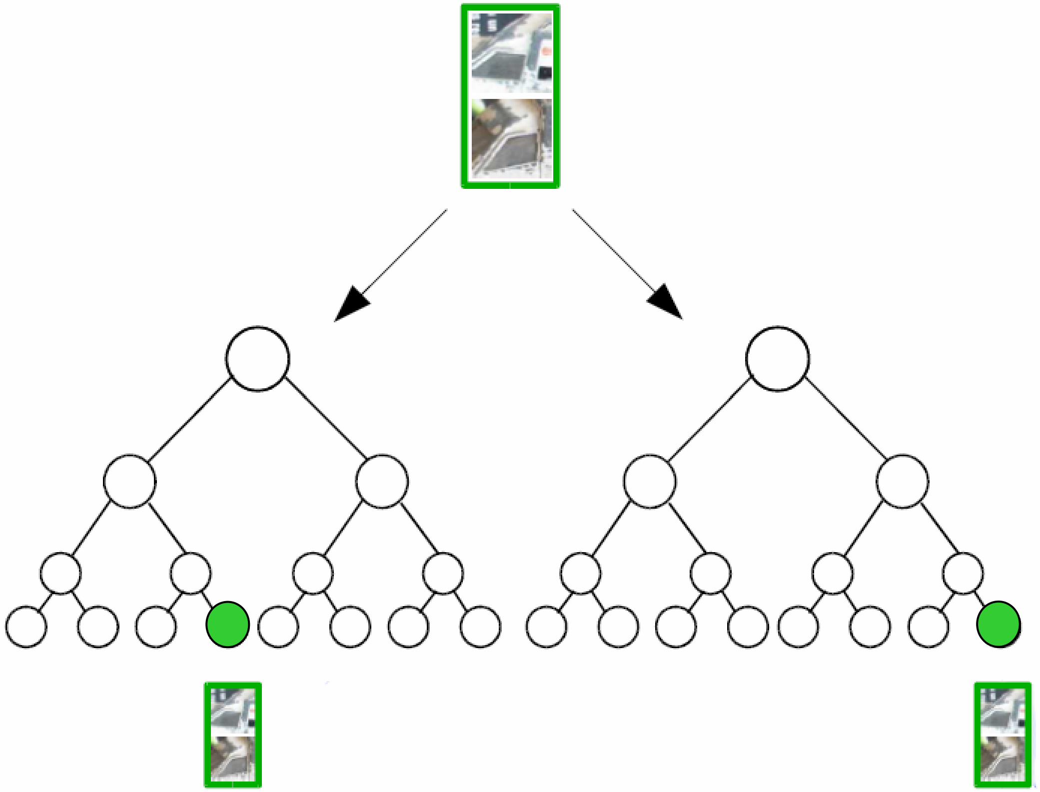


2 SIFT descrip.

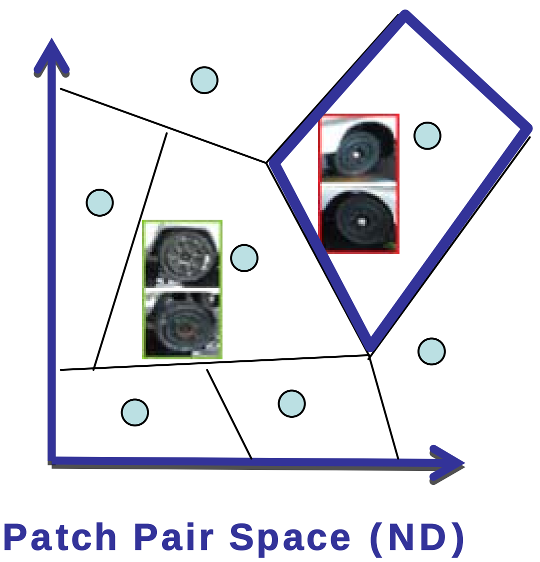
Patch pair quantization algorithm

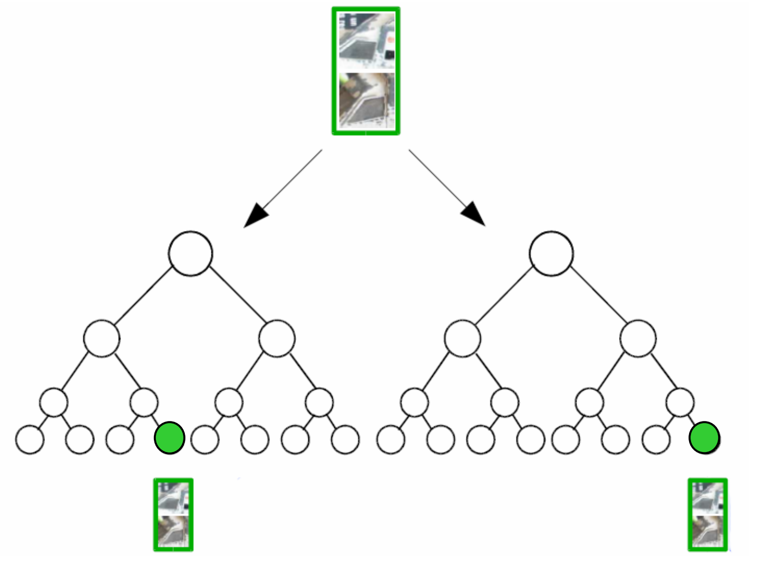


2 SIFT descrip.



Patch pair quantization algorithm



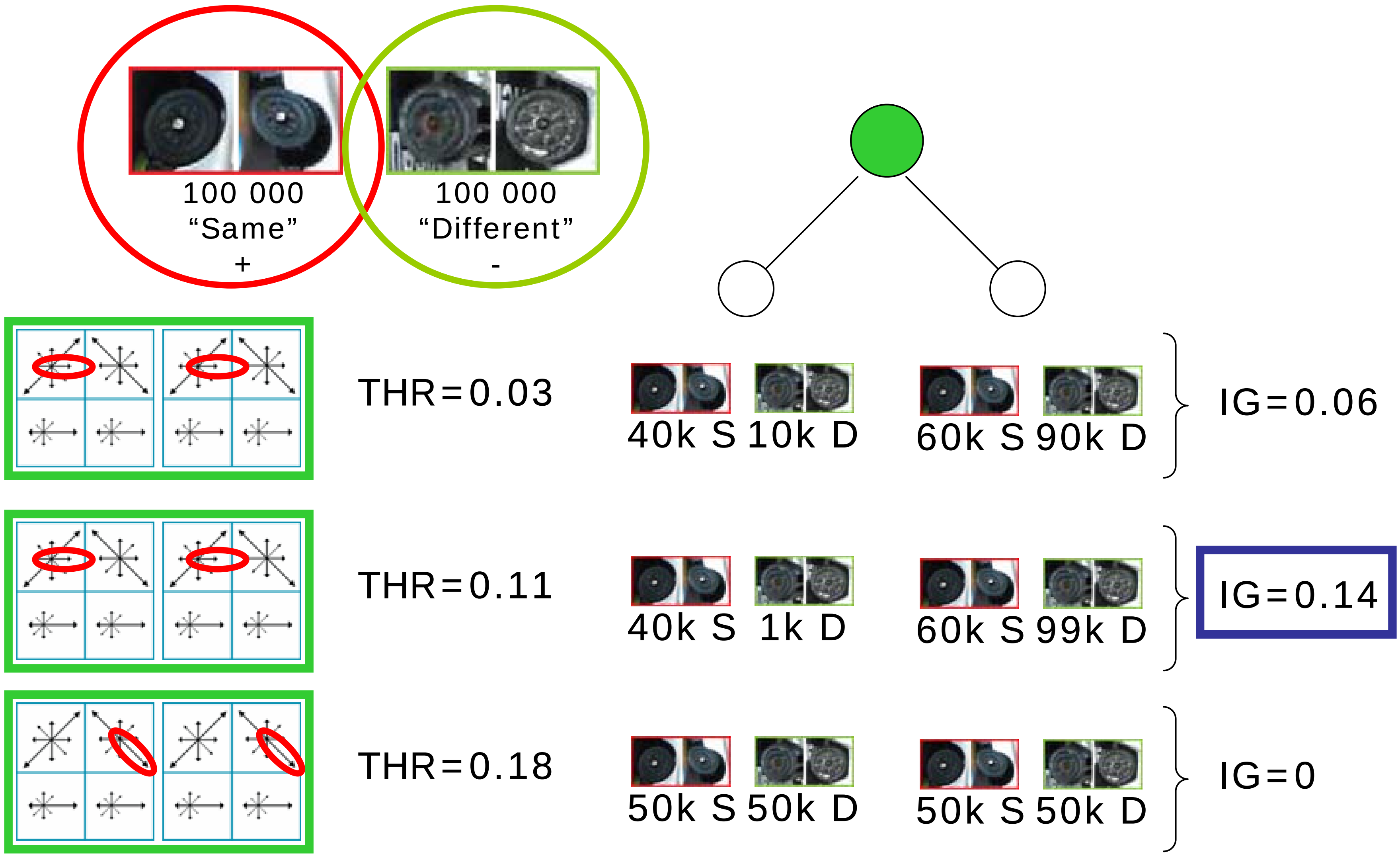


- Quantizer / Clusterer
  - Defined by the trees
- Cluster centers (characteristic differences)
  - defined by the leaves

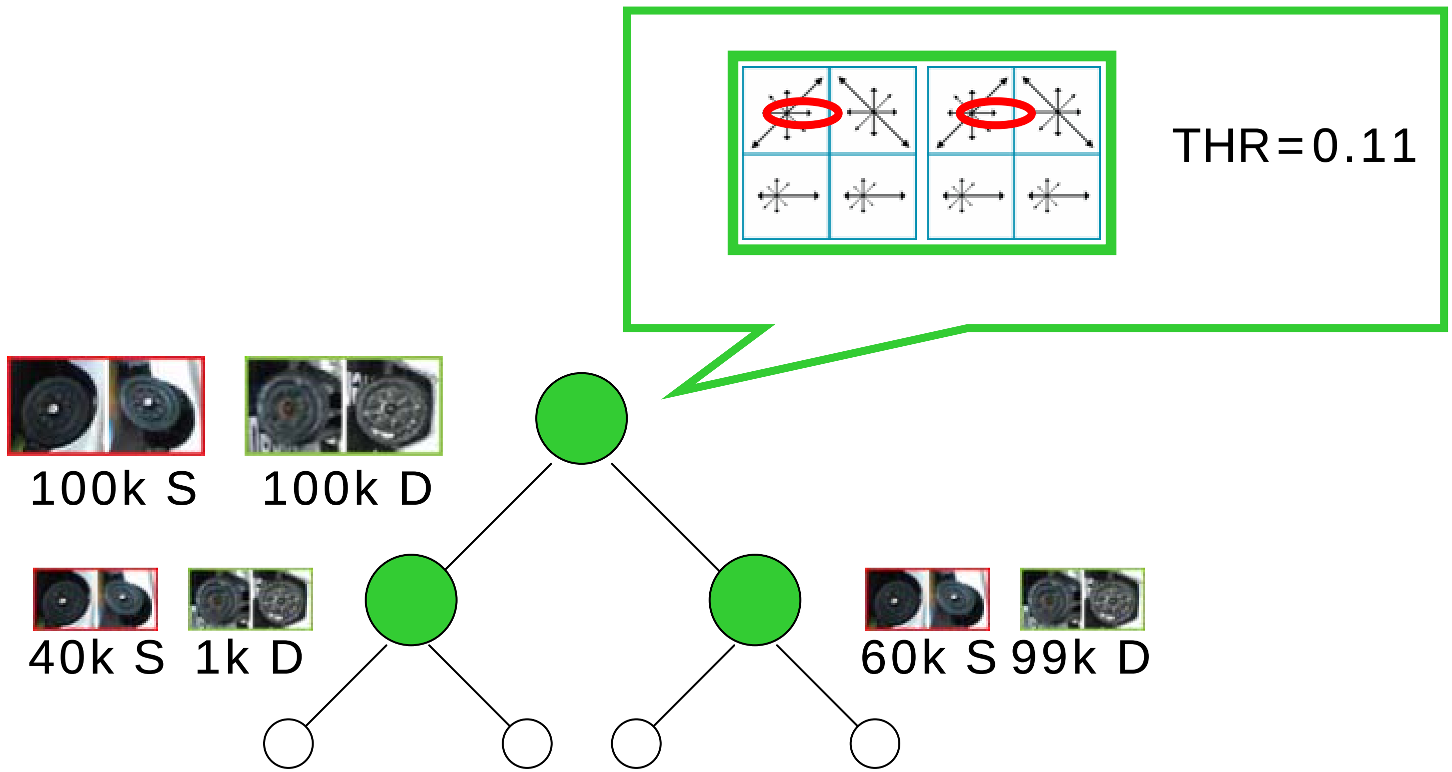
How to learn the trees?

- Classical decision trees
  – For each node select the best feature [which SIFT dimension] and the best threshold
- Extremely Randomized Decision Trees (Geurts 06)
  – Ensemble of decision trees + combination rule
  – Each node is suboptimal
  ☺Variance is small
  ☺Fast to learn
  ☺Good for clustering (Moosman, Triggs and Jurie 06)

How to learn a tree?



How to learn a tree?

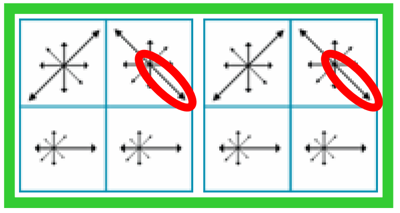


...

Until leaves contain only positive or negative elements

=> **discriminative clustering**

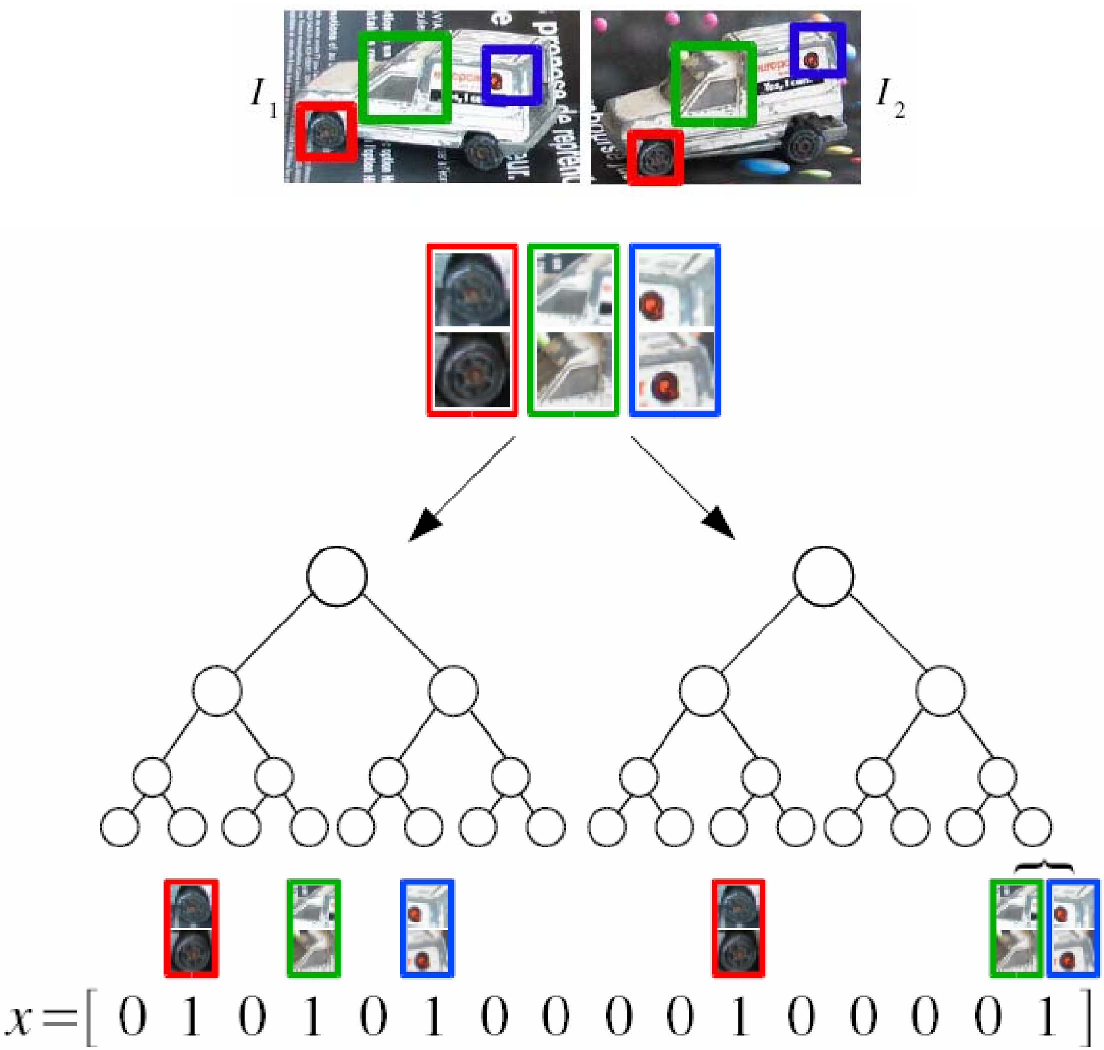
How to learn a tree?



THR=0.18

- Two kinds of Split Condition
- Type 1: **SIFT based**
  - Consider a SIFT dimension and a threshold
  - Feature value above (or below) threshold for the two patches?
- Type 2: **Geometry based**
  - Patch P0 from the first image sampled from a given region (position & scale) ?

An image pair descriptor

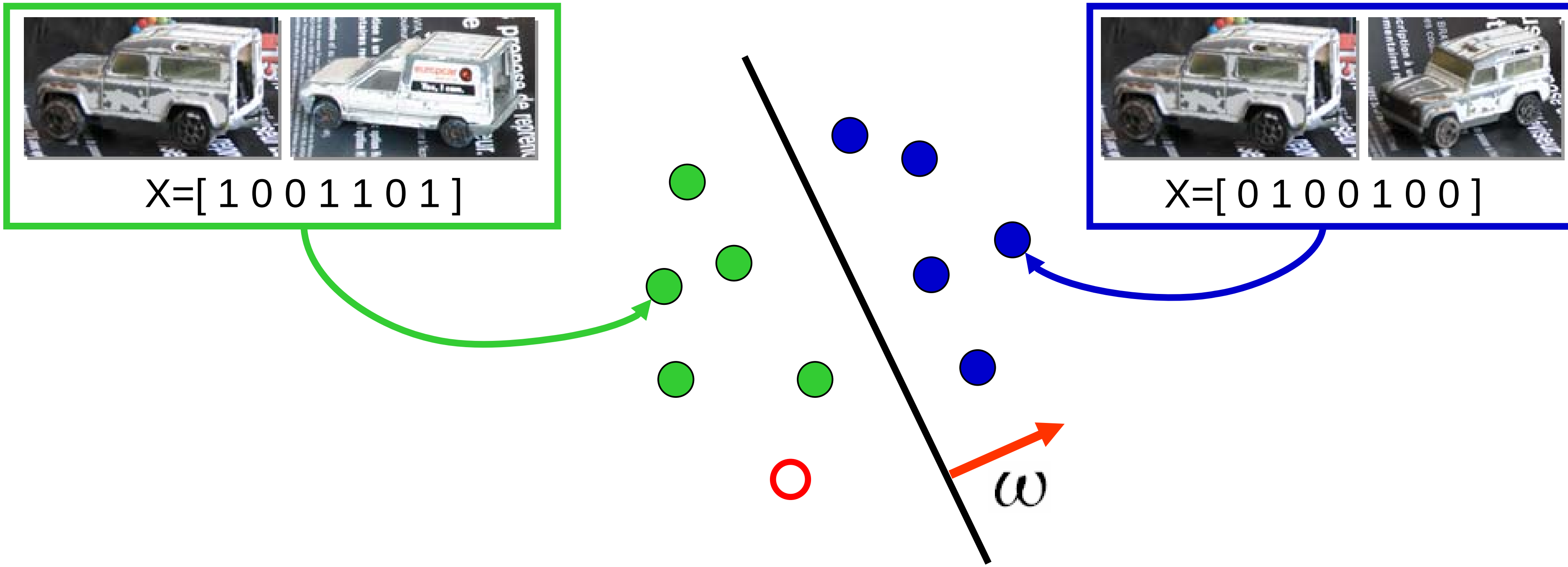


a) sample corresponding patch pairs

b) cluster them with the forest

c) Update a global image pair descriptor

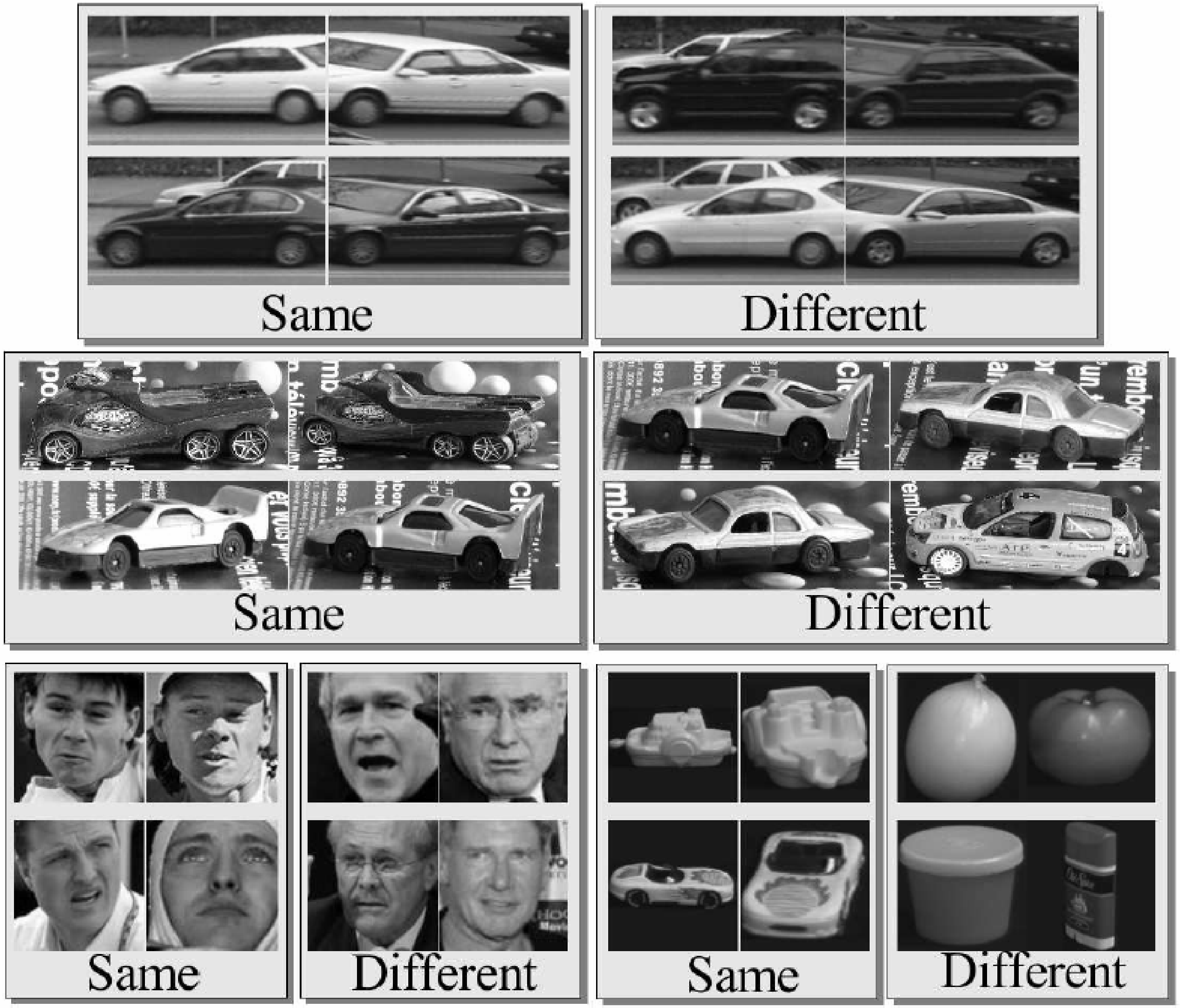
Similarity Measure Computation



Our Goal!

$$S(I1, I2) = S(x) = \omega^t x$$

Datasets



Ferencz et al:
cars
distortions, tiny details, crop

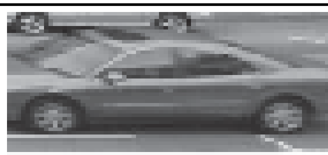
Our dataset:
toycars
view point, light,
background

Jain et al:
“faces in the news”
light, expression, pose,
quality, annotation errors

Fleuret et al:
COIL 100
full rotation, heterogeneous

Generic vs. Specific Knowledge

- The algorithm learns **trees** and **weights**:
  Two kinds of **KNOWLEDGE …**
- Knowledge:
  - generic information for similarity computation?
  - or information specific to a dataset?

Trees	Weights	Test	EER-PR
			91.0%
			86.5% → **-4.5**
			63.0% → **-28.0**

CCL: we ARE embedding specific knowledge

Comparison with State of the Art: Equal Error Rate of Precision

Method	Toy cars	Ferencz	Faces	Coil 100
Others	-	84.9 [4]	70.0 [12]	88.6±4 [7]
Ours	85.9±0.4	91.0±0.6	84.2±3.1	93.0±1.9
Gain	-	6.1	14.2	4.4

Ferencz

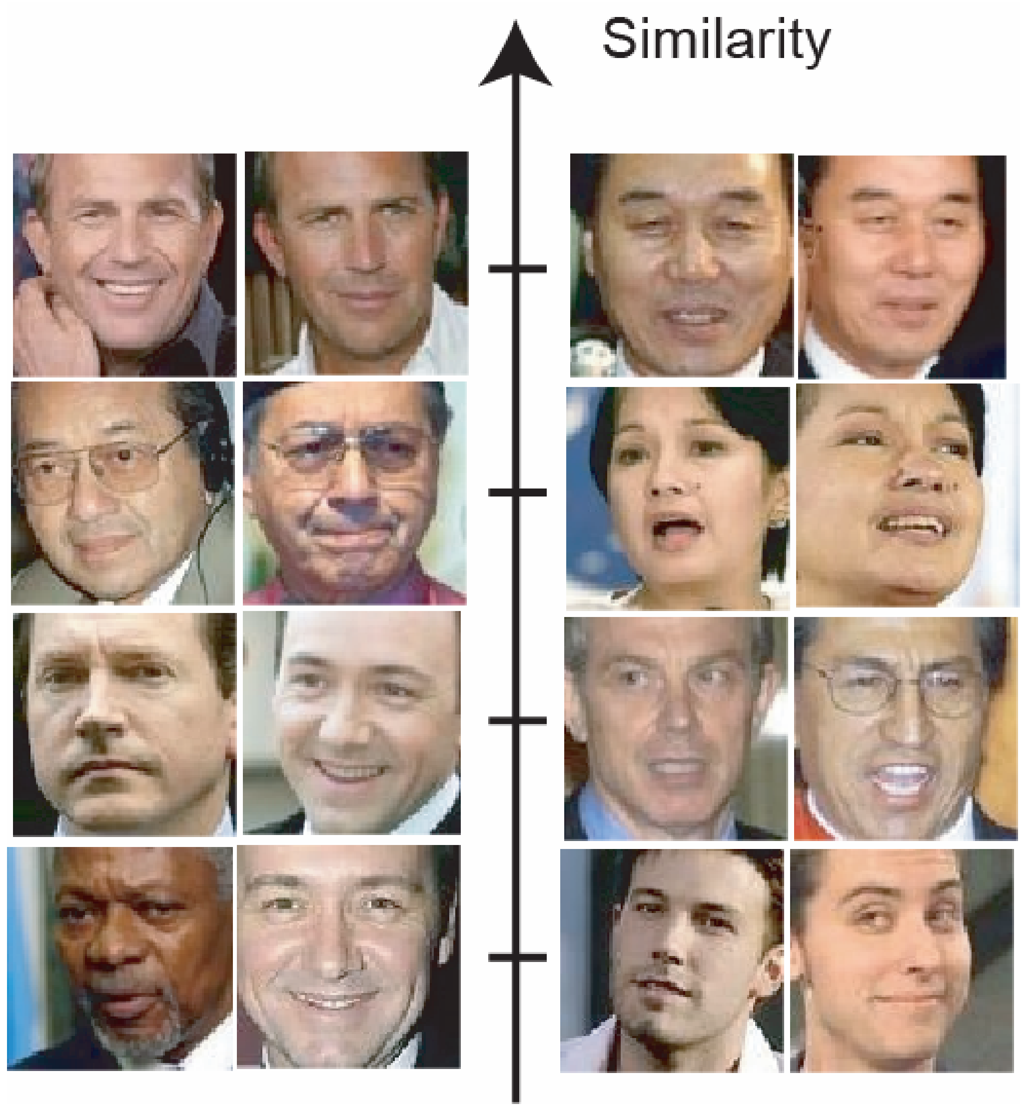
Toycars

Left: Faces

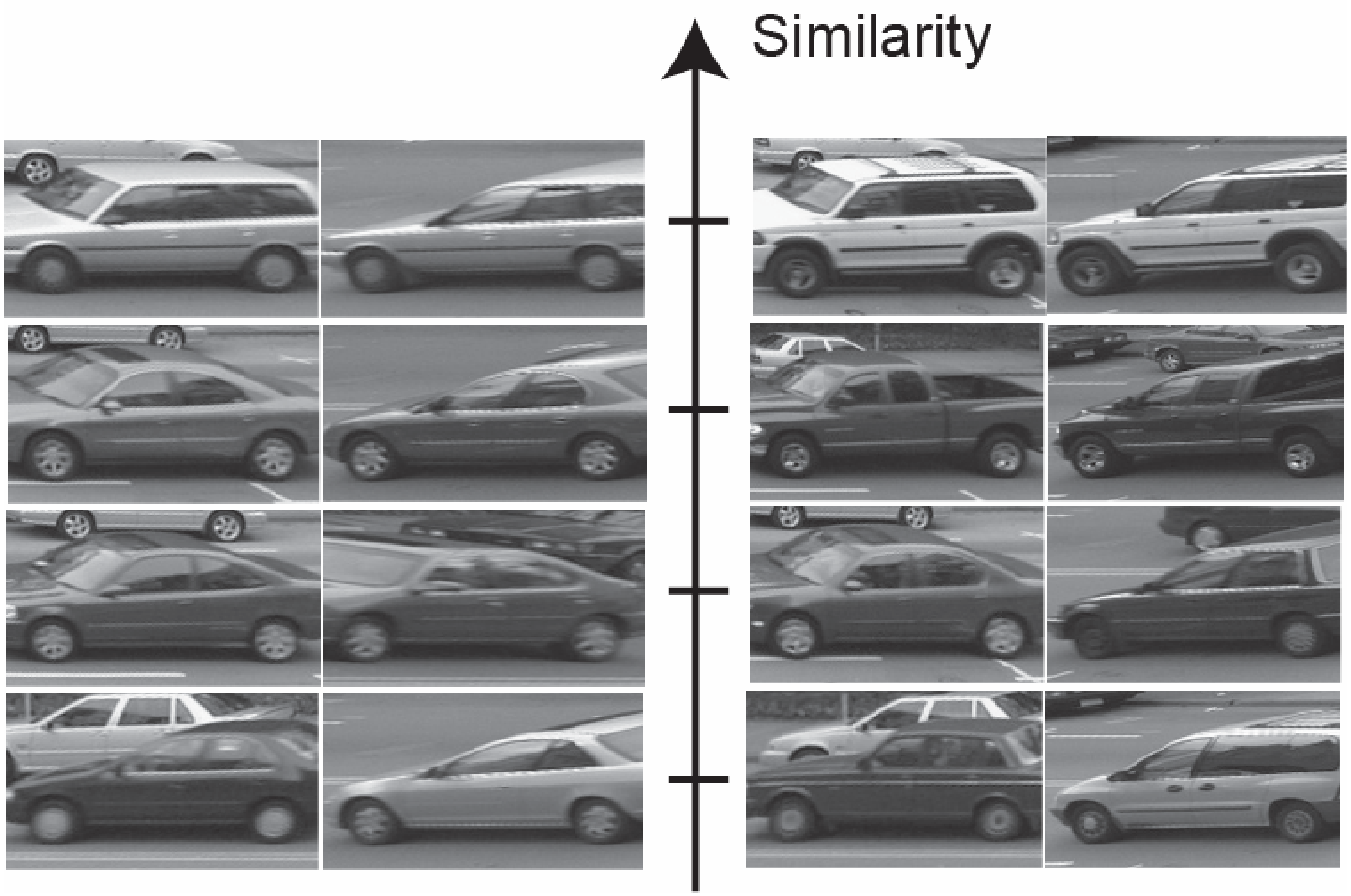
Right: COIL 100

Nowak – Jurie – CVPR07 – Learning Visual Similarity Measures for Comparing *Never* Seen Objects

Visualizations

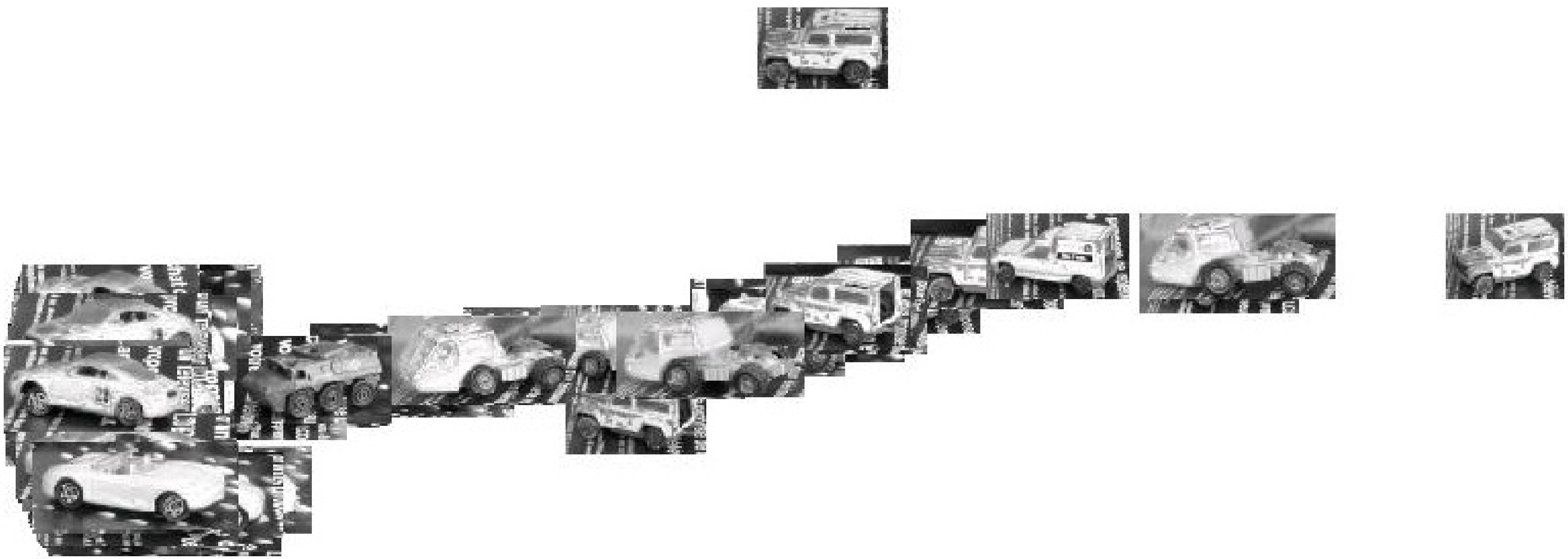


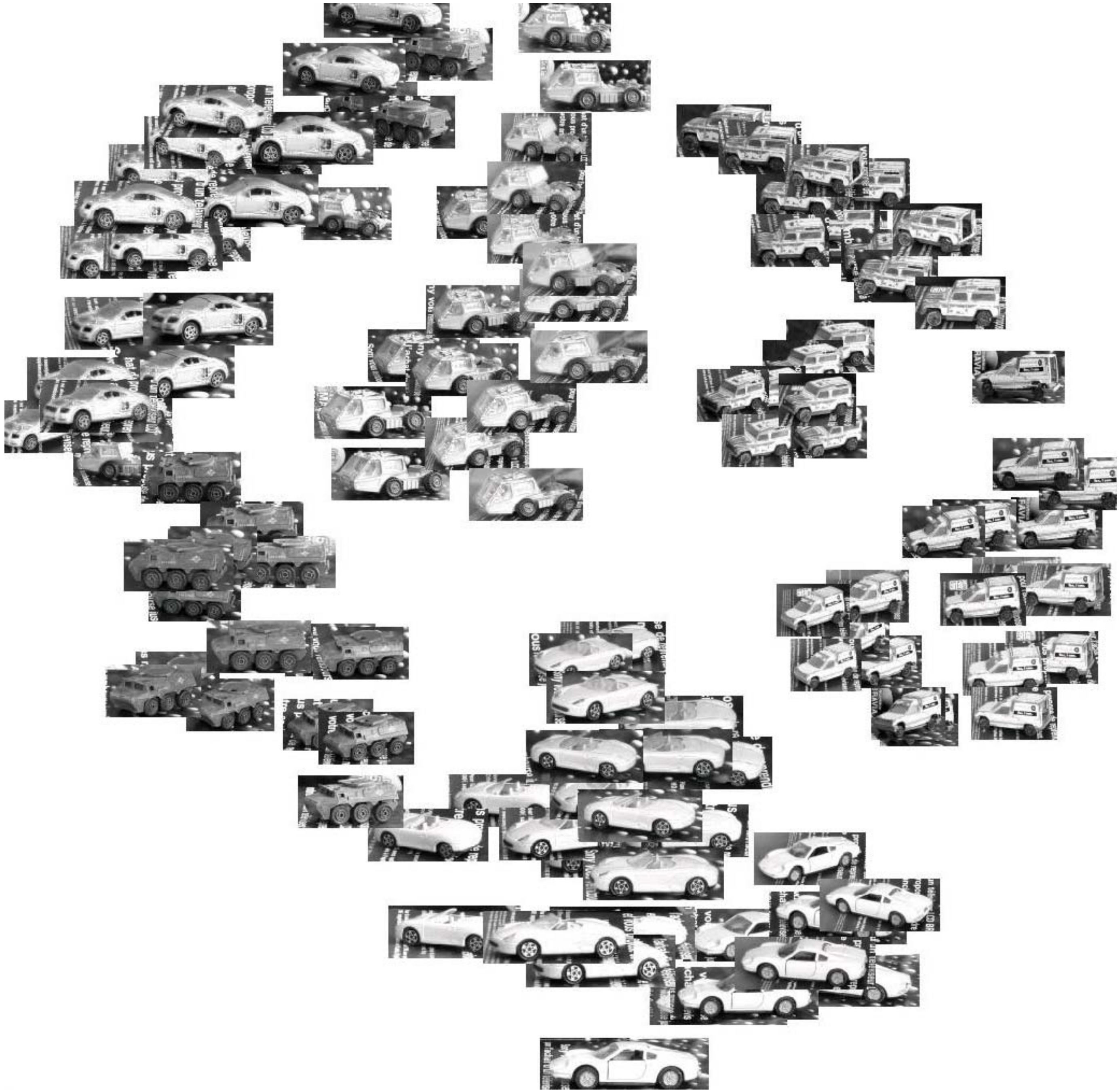
Visualizations



Visualizations

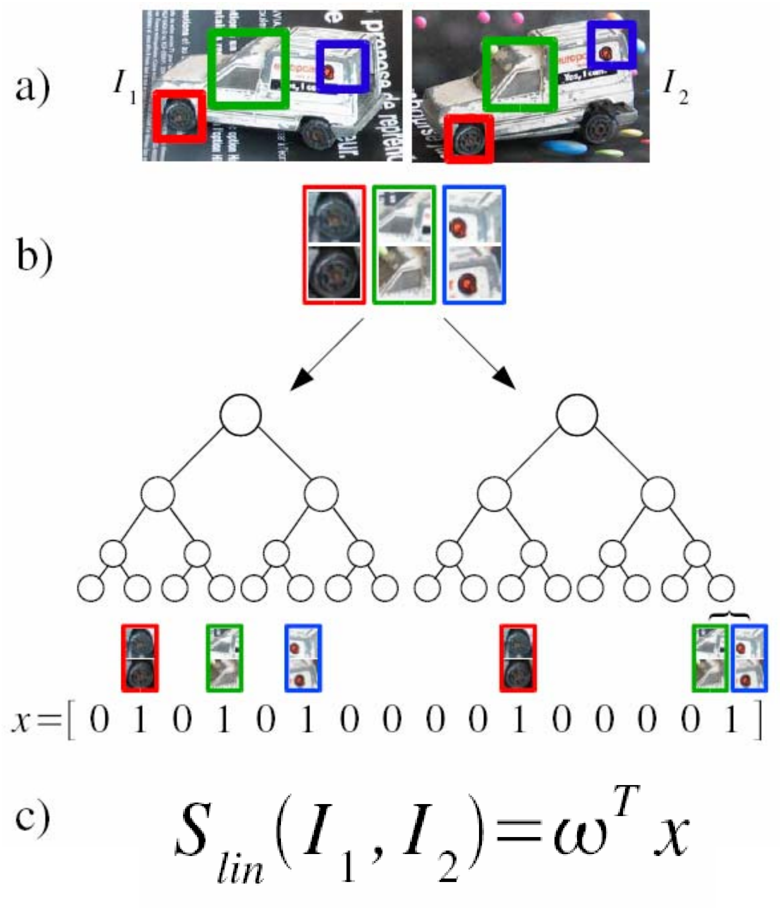
- Multi dimensional scaling (2D):
  L2 distance in 2D as close as possible to the pairwise similarity matrix
- Below: simple bag of words representation
- Next page: our similarity measure





Method Summary

- Consider corresp. local regions
- Quantize patch pair differences
  Extremely Randomized Clustering Forest
- Get global image pair descriptor
- Similarity measure is a weighted sum



Future Works

- Deal with object categories instead of object instances
- Use and combine more features
  - e.g. color
- Applications
  - Photo collection browsing
  - Face identification
  - …

Binaries, Dataset, …

http://lear.inrialpes.fr/people/nowak

Thank you for
your attention!