Classification and Machine Learning techniques for CBIR: introduction to the RETIN system

Matthieu Cord

ETIS CNRS UMR 8051

Content-Based Image Retrieval

- Retrieve large categories of pictures in generalist image database
- Vector-based description of images
- User interaction
- Statistical learning approach

  $\longrightarrow$ Multimodality (category retrieval)

  $\longrightarrow$ Efficient strategies in text retrieval

  $\longrightarrow$ Interactive strategies (active learning)

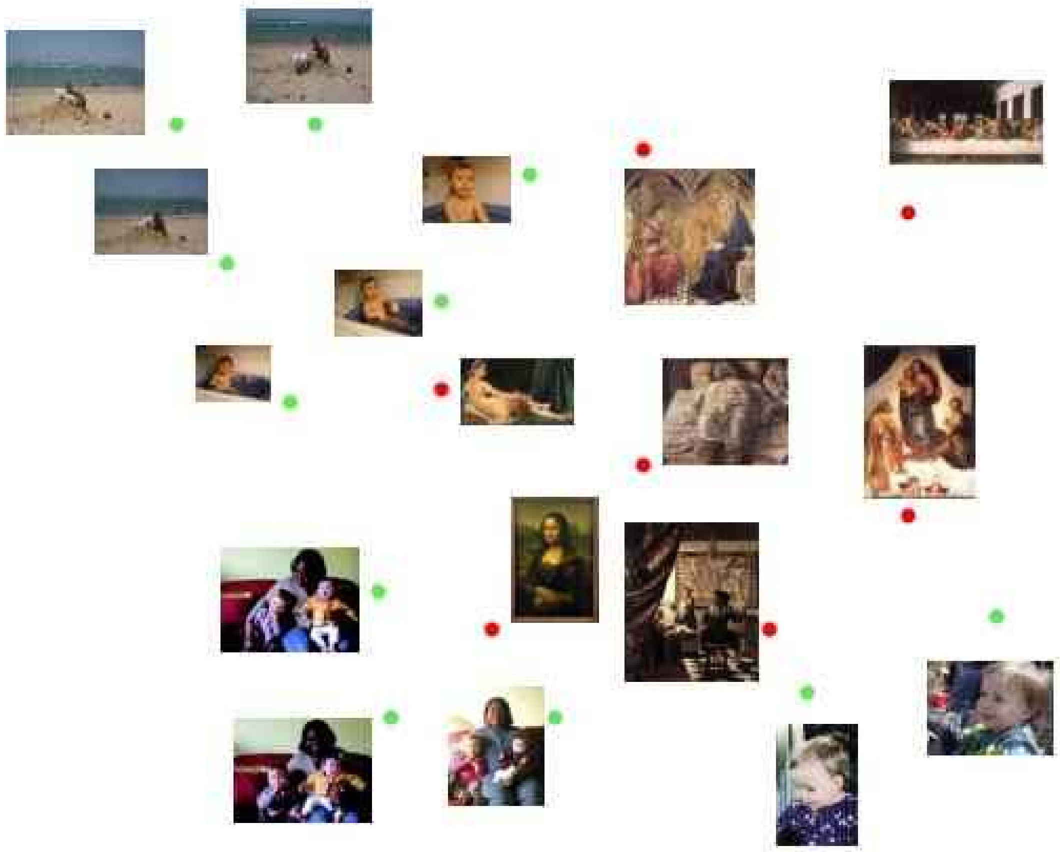
Outline

1. Binary Classification for CBIR
2. Active learning:
   (a) Error Reduction and Uncertainly-Based strategies
   (b) RETIN scheme: Boundary Correction and diversity
3. Semi-supervised classification
4. Long Term Learning

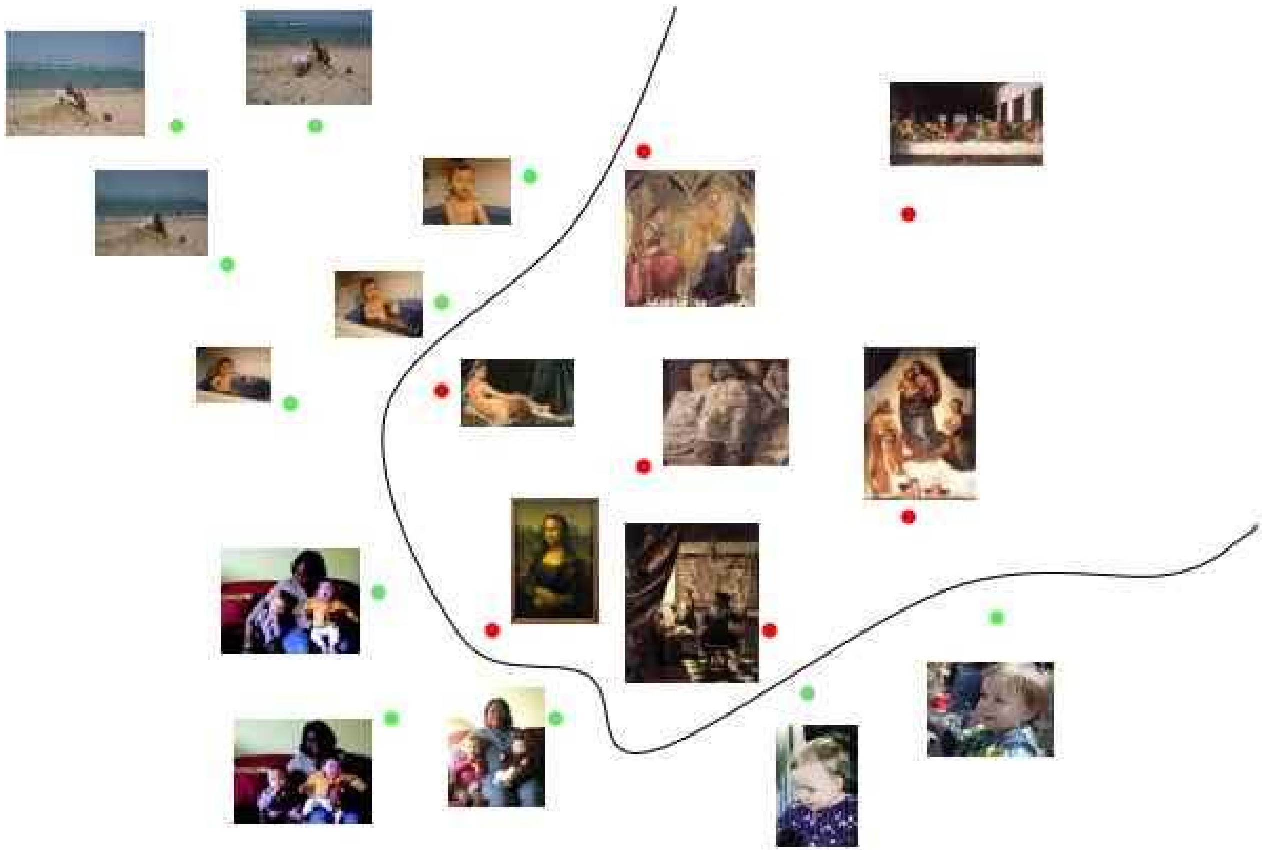
Supervised Classification for CBIR

Introduction

- Vector-based description of images;



- binary classification



Supervised Classification

Three representative methods for CBIR:

- Bayes Classifiers (Vasconcelos)
- k-Nearest Neighbors
- Support Vector Machines (Chapelle)

Specific characteristics [Chang ICIP’03]:

(c1) High dimension and non-linearity of input space

(c2) Few training data

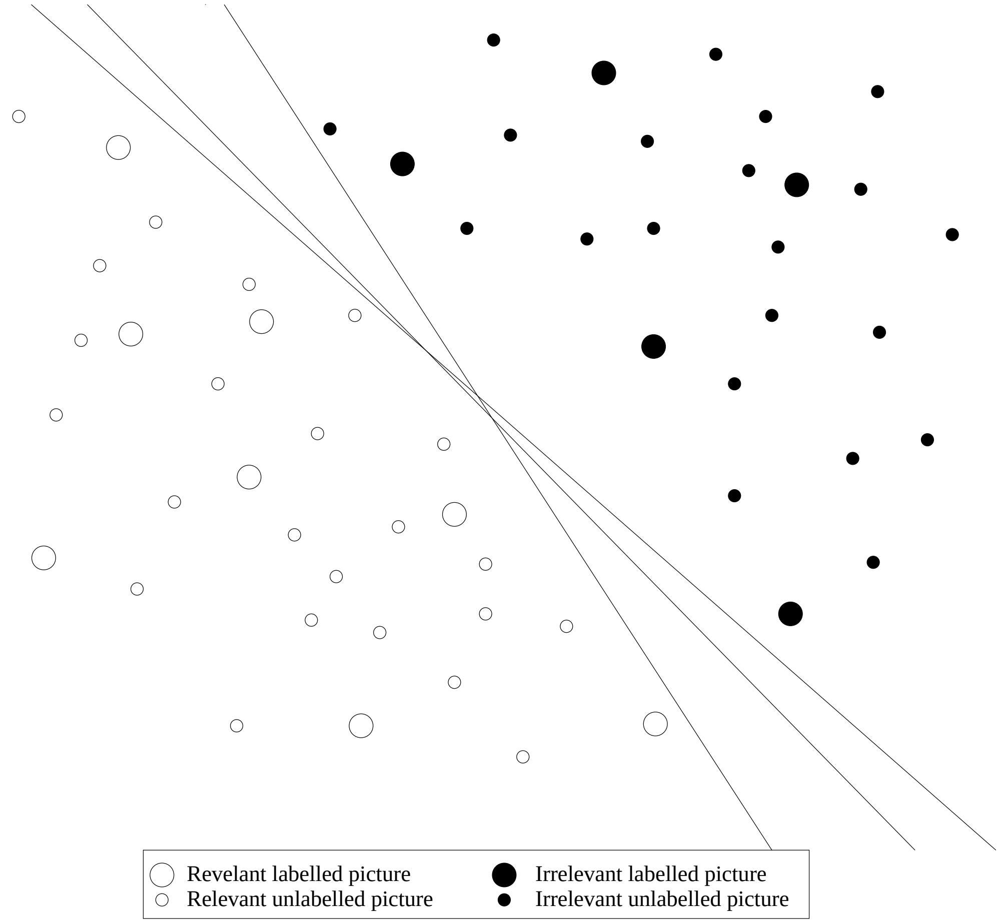
(c3) Many unlabelled data

(c4) Interactive learning (Relevance feedback)

(c5) Unbalanced training data

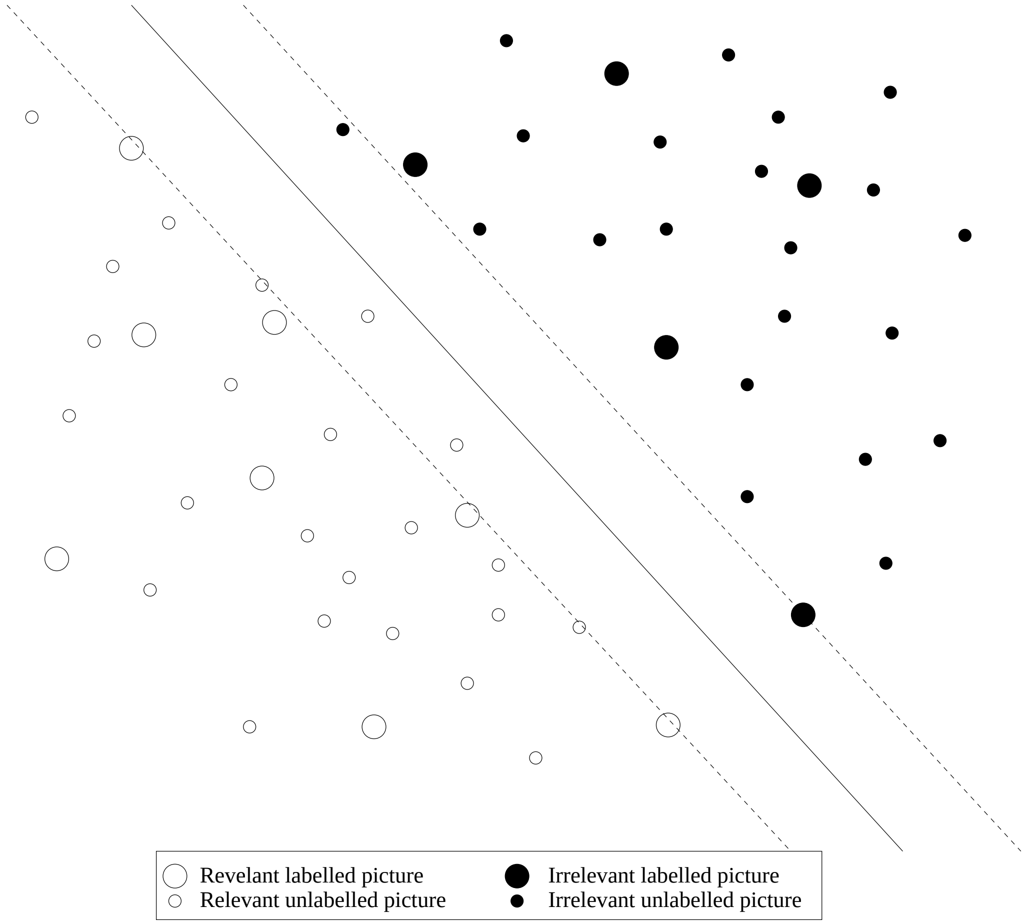
Support Vector Machines (1/4)

Classification by an hyperplan:



Support Vector Machines (2/4)

Choose the hyperplan which maximizes the margin:



Support Vector Machines (3/4)

Quadratic problem:

$$\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\text{argmax}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j < \mathbf{x}_i, \mathbf{x}_j >$$

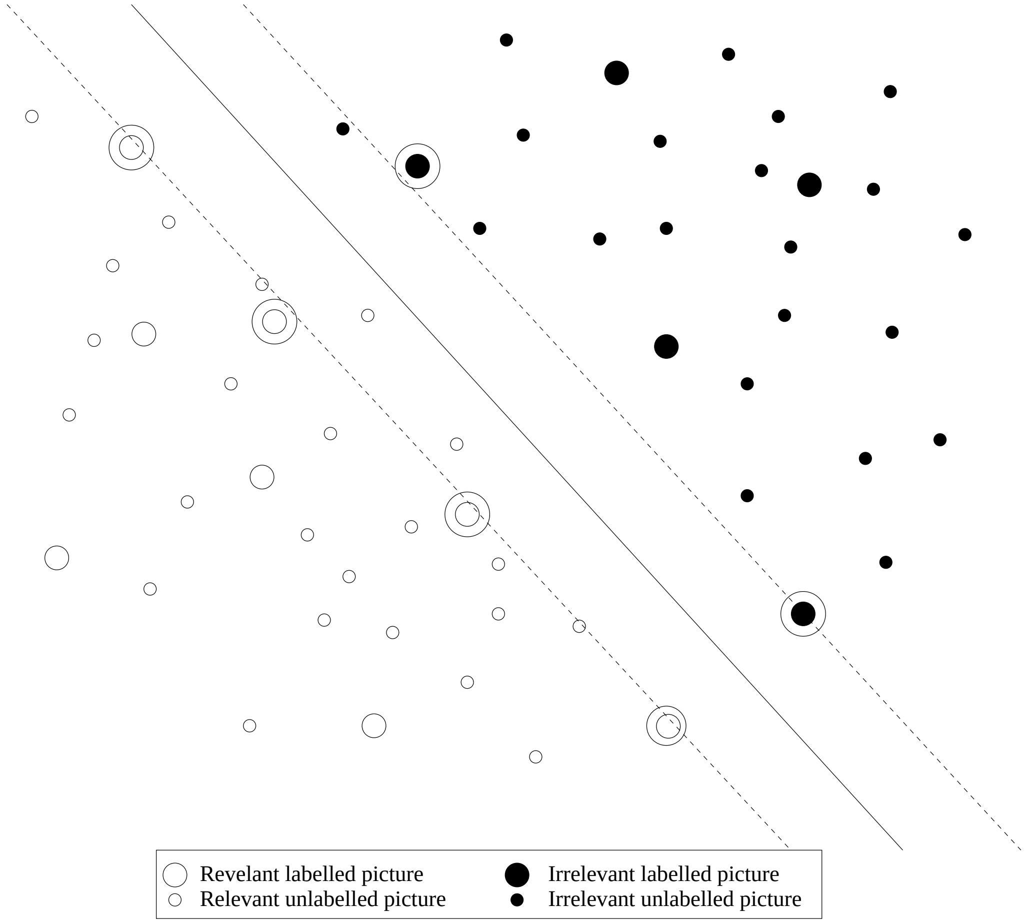
$$\text{with} \begin{cases} \displaystyle\sum_{i=1}^{n} \alpha_i y_i = 0 \\ \forall i \in [1, n] \quad 0 \leq \alpha_i \leq C \end{cases}$$

Decision function:

$$f(\mathbf{x}) = \sum_{i=1}^{n} y_i \alpha_i^\star < \mathbf{x}, \mathbf{x}_i > + b$$

Support Vector Machines (3/4)

Support Vectors:



"Kernelization"

Kernelization of SVM:

- SVM decision function:

$$f(\mathbf{x}) = \sum_{i=1}^{N} y_i \alpha_i^\star < \mathbf{x}, \mathbf{x}_i > +b \qquad (1)$$

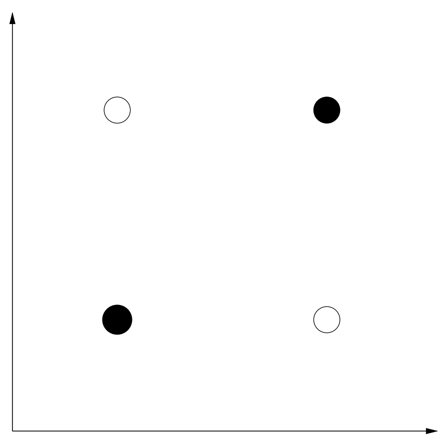
- "Kernelized" version:

$$f(\mathbf{x}) = \sum_{i=1}^{N} y_i \alpha_i^\star k(\mathbf{x}, \mathbf{x}_i) + b \qquad (2)$$

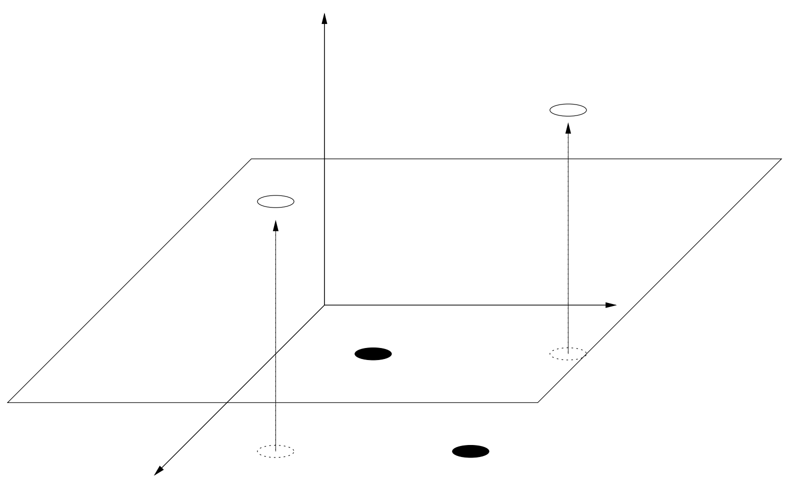
Dealing with the class of kernels k corresponding to dot product in an induced space $\mathcal{H}$ via a map Φ:

$$\begin{aligned}\Phi \quad : \quad & \mathbb{R}^p \longrightarrow \mathcal{H} \\ & x \longmapsto \Phi(x)\end{aligned}$$

that is $k(\mathbf{x}, \mathbf{x}') =< \Phi(\mathbf{x}), \Phi(\mathbf{x}') >$



Initial: X



Induit: $\Phi(X)$

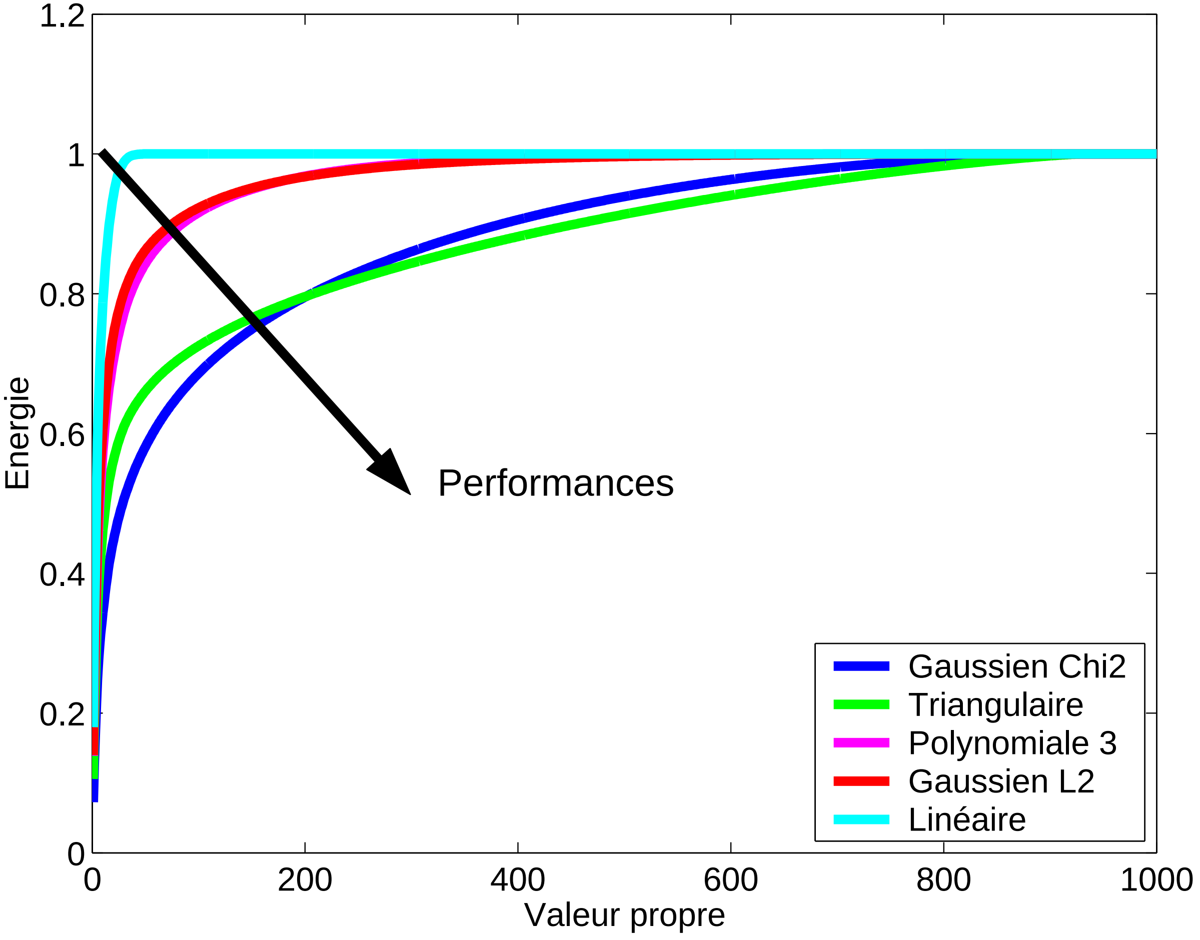
Kernels

- Usual kernels: Lin., Polynomial, Sigmoid, RBF ...
- Choice of a kernel depends on the database and its usage:

$\rightarrow$ Different levels of performances for two different kernels;

- In our experiments: Gaussian kernels give the best results
  - $\rightarrow$ The most adapted to CBIR;
  - $\rightarrow$ In the following experiments: Gaussian kernels with χ^2 distance, because feature vector are distributions.

Spectral analysis of kernel matrices



Large distribution $\Rightarrow$ high performances;

SVM and Kernels

Deal with (c1) high dimension and non-linear input space:

$\rightarrow$ Use of a kernel function to induce a feature space

$\rightarrow$ Relevance function f using Kernel in SVM:

$$f(\mathbf{x}) = \sum_{i=1}^{N} y_i \alpha_i^{\star} \mathbf{k}(\mathbf{x}, \mathbf{x}_i) + b$$

When a method cannot be directly "kernelized": KPCA.

Experiments

Protocol:

- COREL Photo database (6,000 images);
- 50 categories, 100-300 size;
- Training set of 200 points (unbalanced).
- Statistical measure: Mean Average Precision MAP

Methods	MAP(%)	Time
No learning	8	-
Bayes/Parzen	18	0.09s
k-NN	16	0.20s
SVM	20	0.13s

- SVM selected [Gosselin CVDB04]

Experiments

Training with 10 examples => poor top-similarity ranking results

$\rightarrow$ User interaction (c4) to enhance the retrieval

2 components: the parameter tuning of f and the optimization of the set of examples

Active learning for CBIR

Active Learning

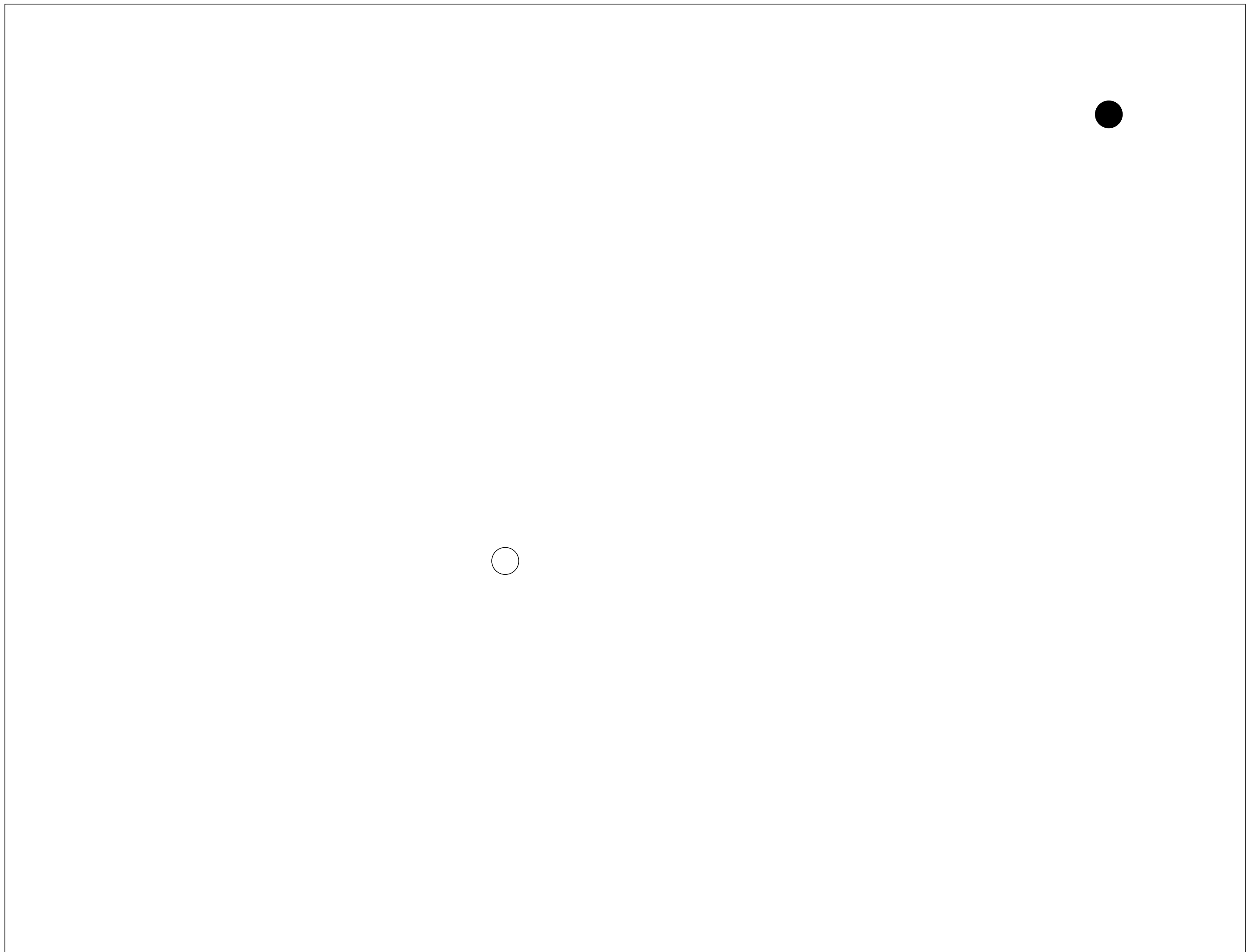
Deal with the few training data (c2) and interactive learning (c4) characteristics

$\rightarrow$ optimize training data to get the best classification with as few as possible user labeling

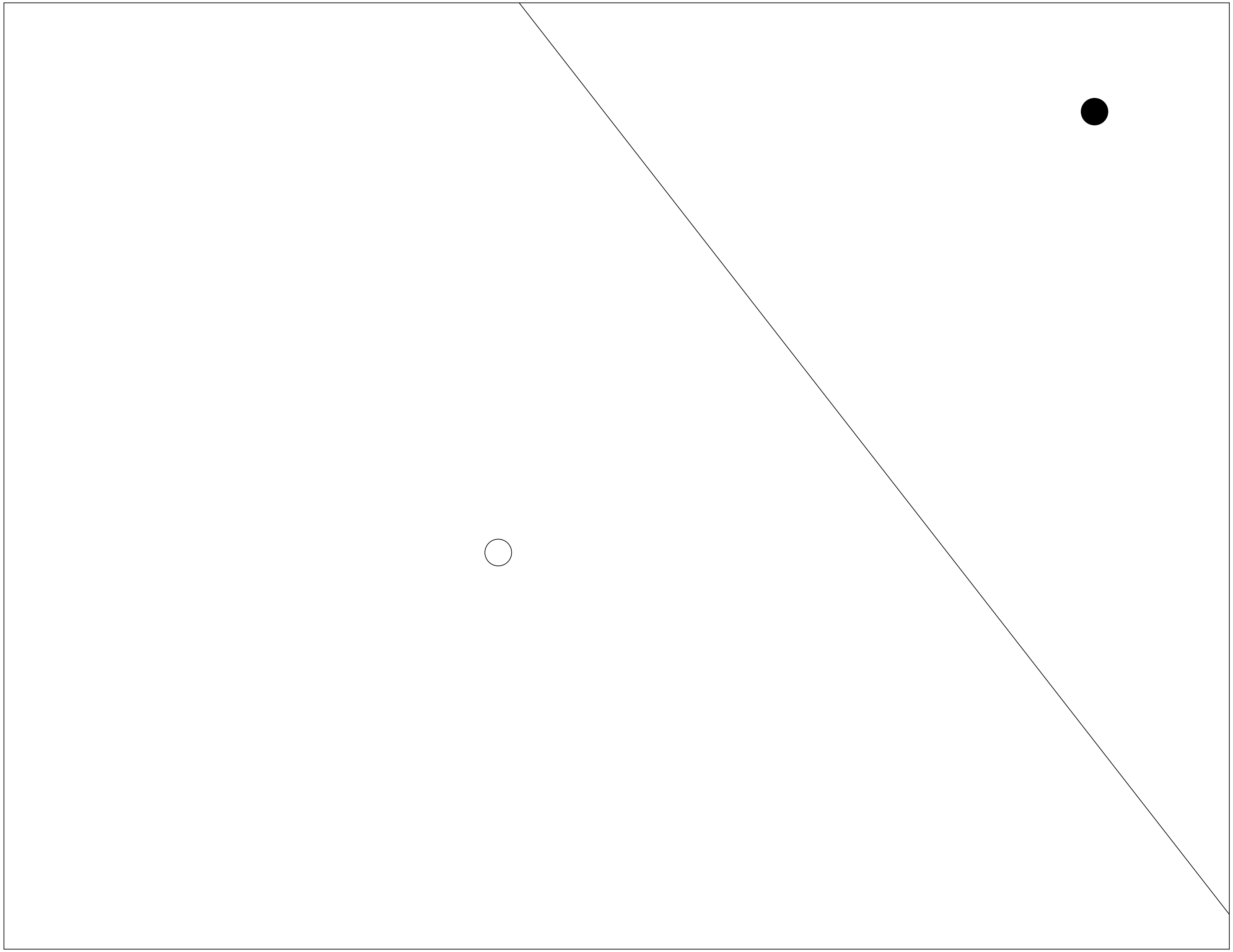
Strategies of **selective sampling**:

- Relevance-Based (RB):
  $\rightarrow$ Select the most relevant image
- Uncertainly-Based (UB)
- Error Reduction (ER)
  $\rightarrow$ Priority to the classification error minimization

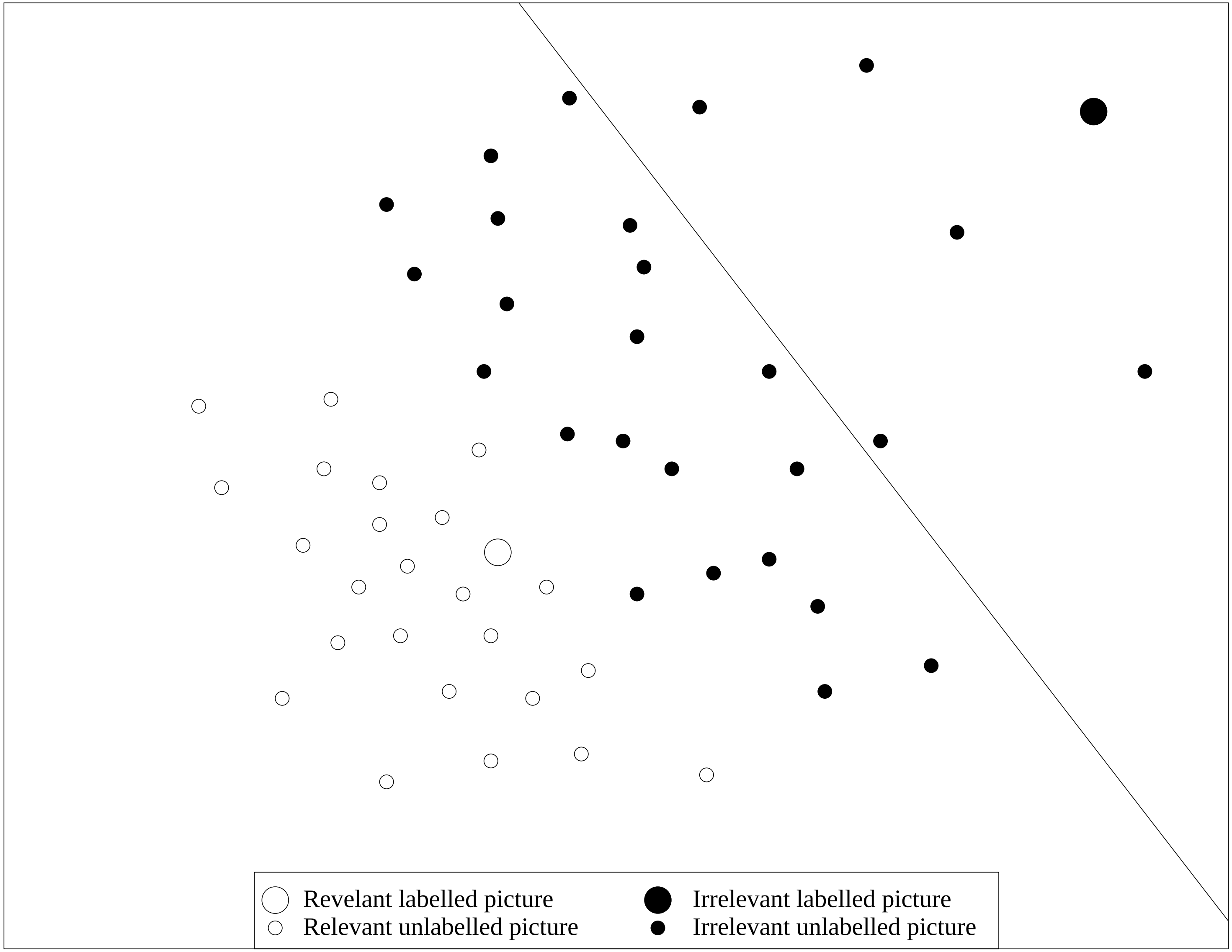
Labelling the most relevant (RB)



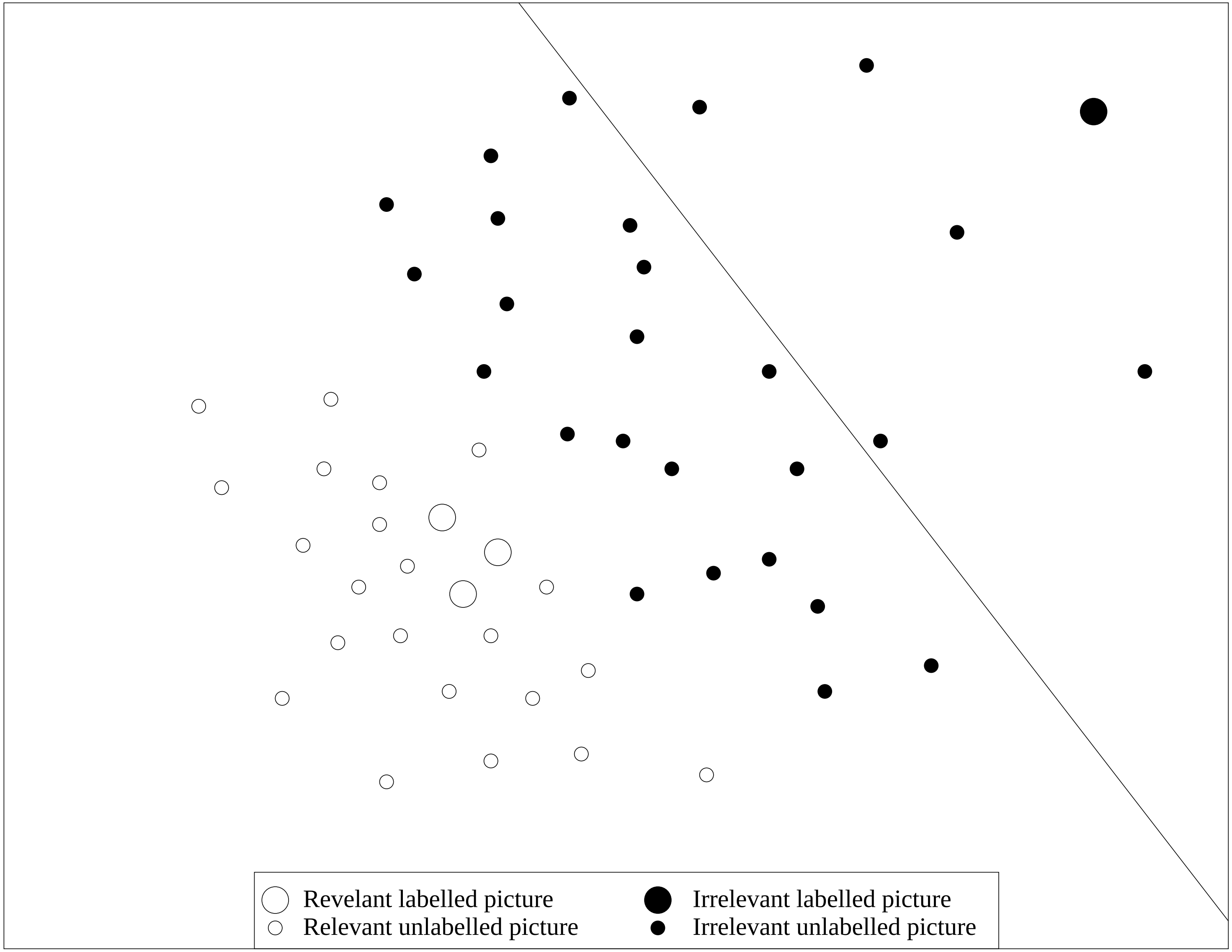
Labelling the most relevant (RB)



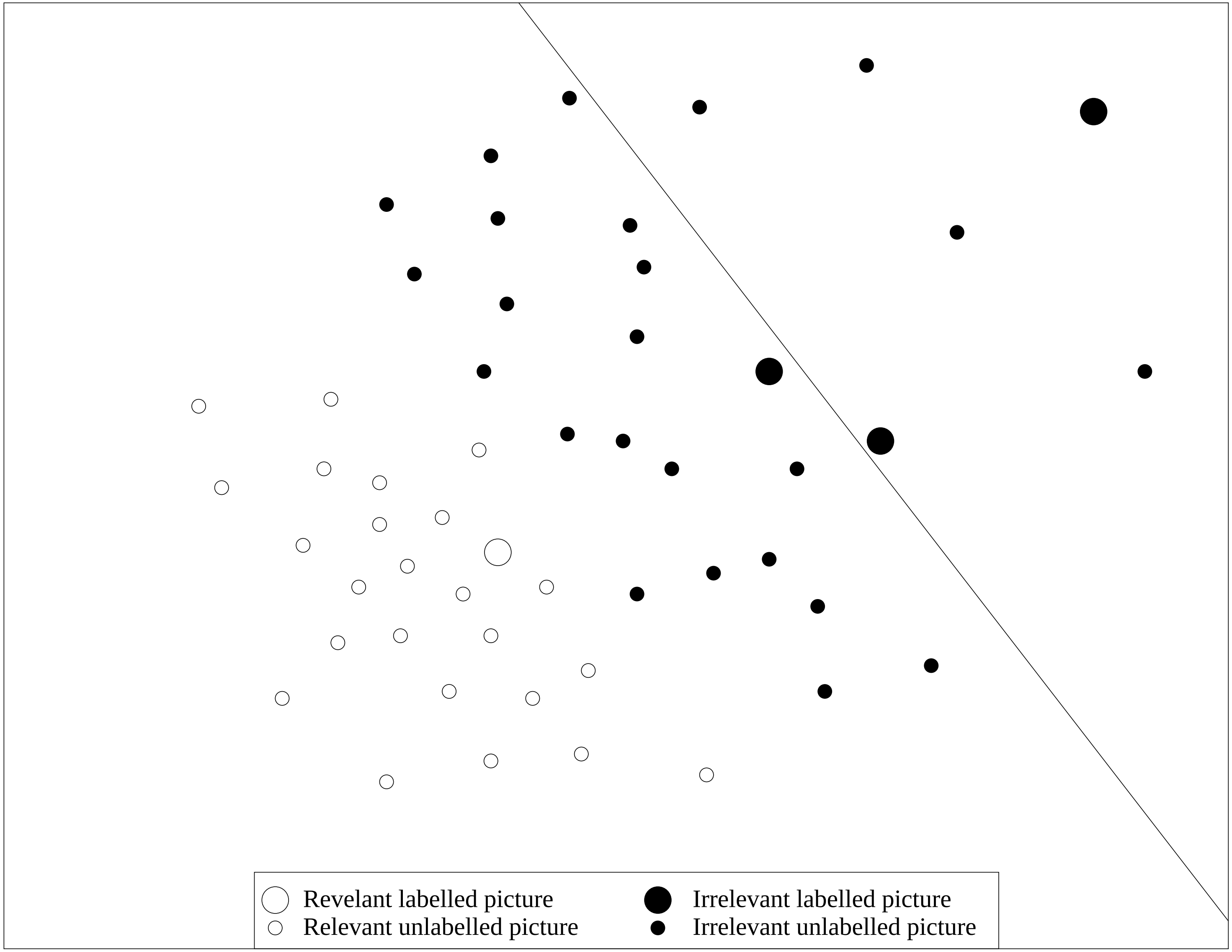
Labelling the most relevant (RB)



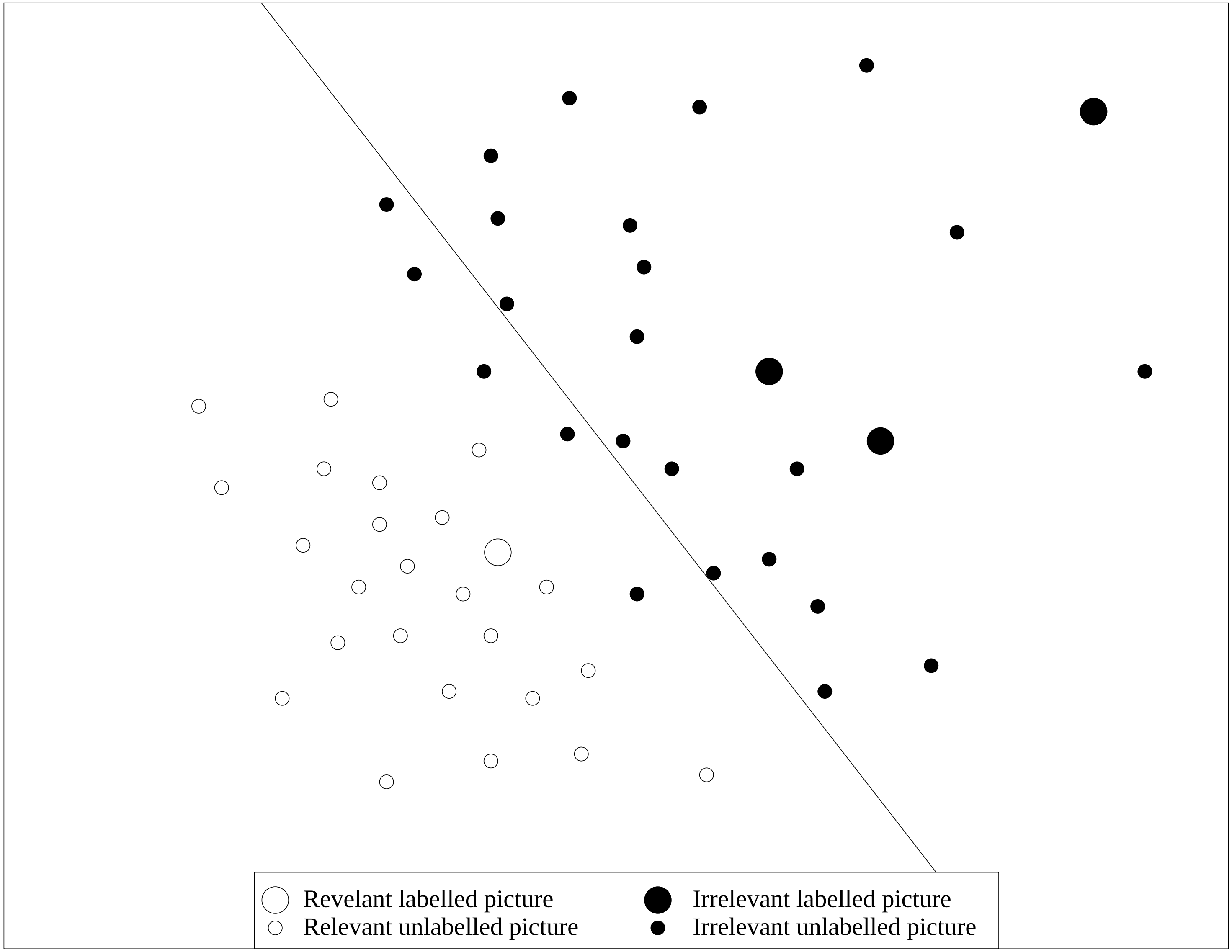
Labelling the most relevant (RB)



Labelling the most difficult to classify (UB)



Labelling the most difficult to classify (UB)



Active Learning

The aim of an active learner is to select the most interesting picture $\mathbf{x}^\star$

$\rightarrow$ We propose to express the following methods as the minimization of a cost function $g(\mathbf{x})$:

$$\mathbf{x}^\star = \operatorname*{argmin}_{\mathbf{x}} g(\mathbf{x})$$

For Relevance-Based active learning: $g(\mathbf{x}) = -f(\mathbf{x})$ where $f(\mathbf{x})$ is the relevance function

Uncertainly-based (UB)

Active learner:

$$\mathbf{x}^\star = \underset{\mathbf{x}}{\operatorname{argmin}}\ g(\mathbf{x})$$

- UB strategy selects the picture which is the most difficult to classify:

$$g(\mathbf{x}) = |f(\mathbf{x})|$$

- Method:
  - SVM$_{active}$ (Tong):
    $\longrightarrow$ Works in the version space
    $\longrightarrow$ Needs an accurate estimation of the boundary

Error Reduction (ER)

- ER strategy (Roy and McCallum): select the picture which will minimize the new expected test error:

$$g(\mathbf{x}) = \sum_{c \in \{-1,1\}} E_{\hat{P}_{\mathcal{A}+(\mathbf{x},c)}} \hat{P}_{\mathcal{A}}(c|\mathbf{x})$$

  with:
  - $\hat{P}_{\mathcal{A}}(c|\mathbf{x})$ the ***estimation*** of the probability of class c given $\mathbf{x}$, with the training set $\mathcal{A}$
  - $E_{\hat{P}_{\mathcal{A}+(\mathbf{x},c)}}$ the ***estimation*** of the expectation of the test error, with training set $\mathcal{A}+(\mathbf{x},c)$
- Require an accurate estimation of $\hat{P}_{\mathcal{A}}(c|\mathbf{x})$

RETIN

- Active learners select only one example
- In image retrieval, several images labeled during each feedback step $=$ *batch processing*
  How to select other ones ?
  - Iteration of the active selection:
    $\longrightarrow$ Problem: close images may be selected.
  - Diversity: select different images:
    $\longrightarrow$ Clustering or using angle Diversity AD (Brinker):
    $I^\star$ set of selected images

$$g_{I^\star}(\mathbf{x}_i) = \lambda \underbrace{g(\mathbf{x}_i)}_{\text{active criteria}} + (1-\lambda) \underbrace{\max_{j \in I^\star} \frac{|k(\mathbf{x}_i, \mathbf{x}_j)|}{\sqrt{k(\mathbf{x}_i, \mathbf{x}_i) k(\mathbf{x}_j, \mathbf{x}_j)}}}_{\text{angle criteria}}$$

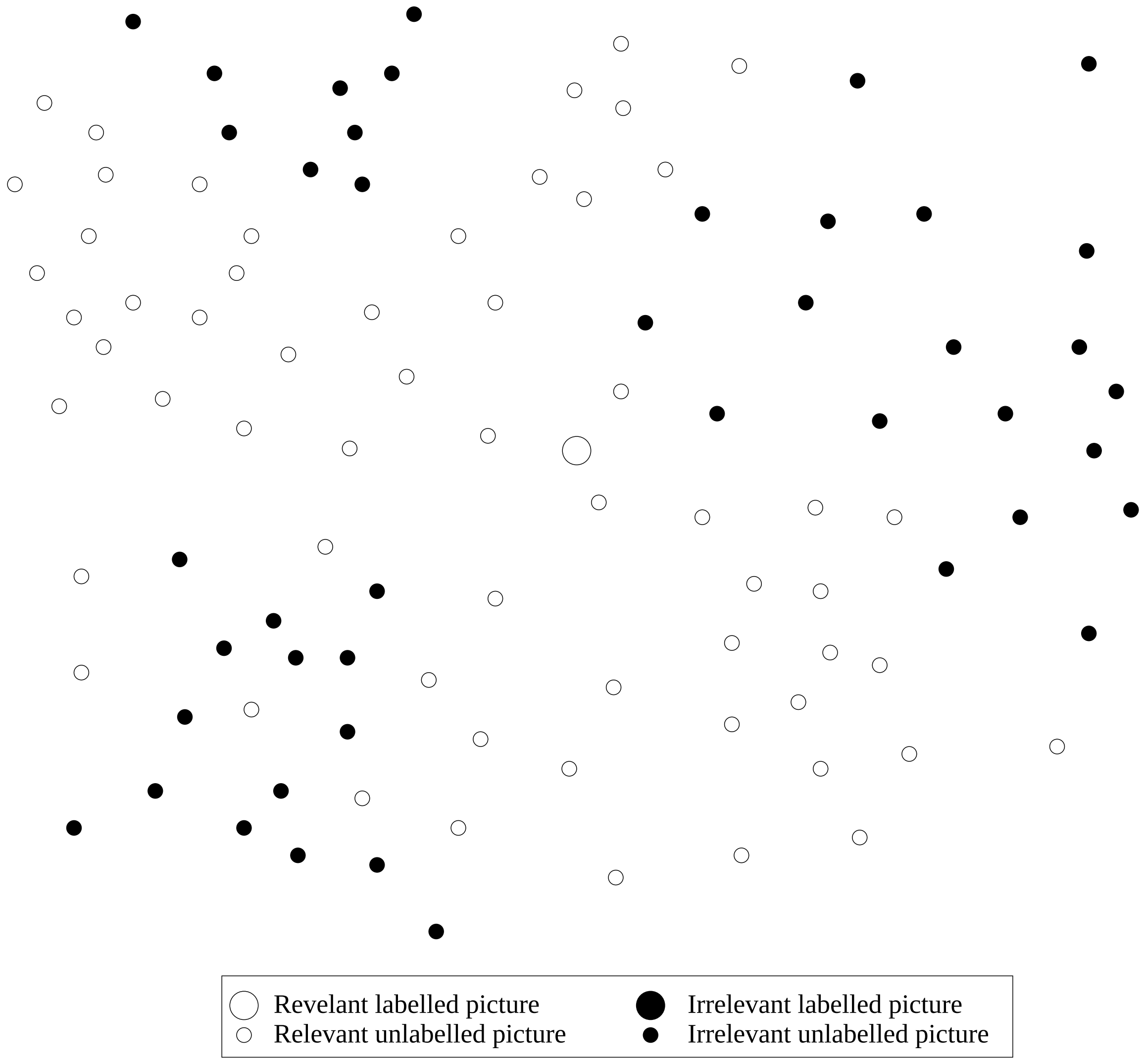
RETIN

- Second: having a good estimation of f near the boundary
  - Exploit again the *batch processing*:
    user labels $\neq$ expected labels (balanced set)
  - too many positive labels => go further (and *vice versa*)

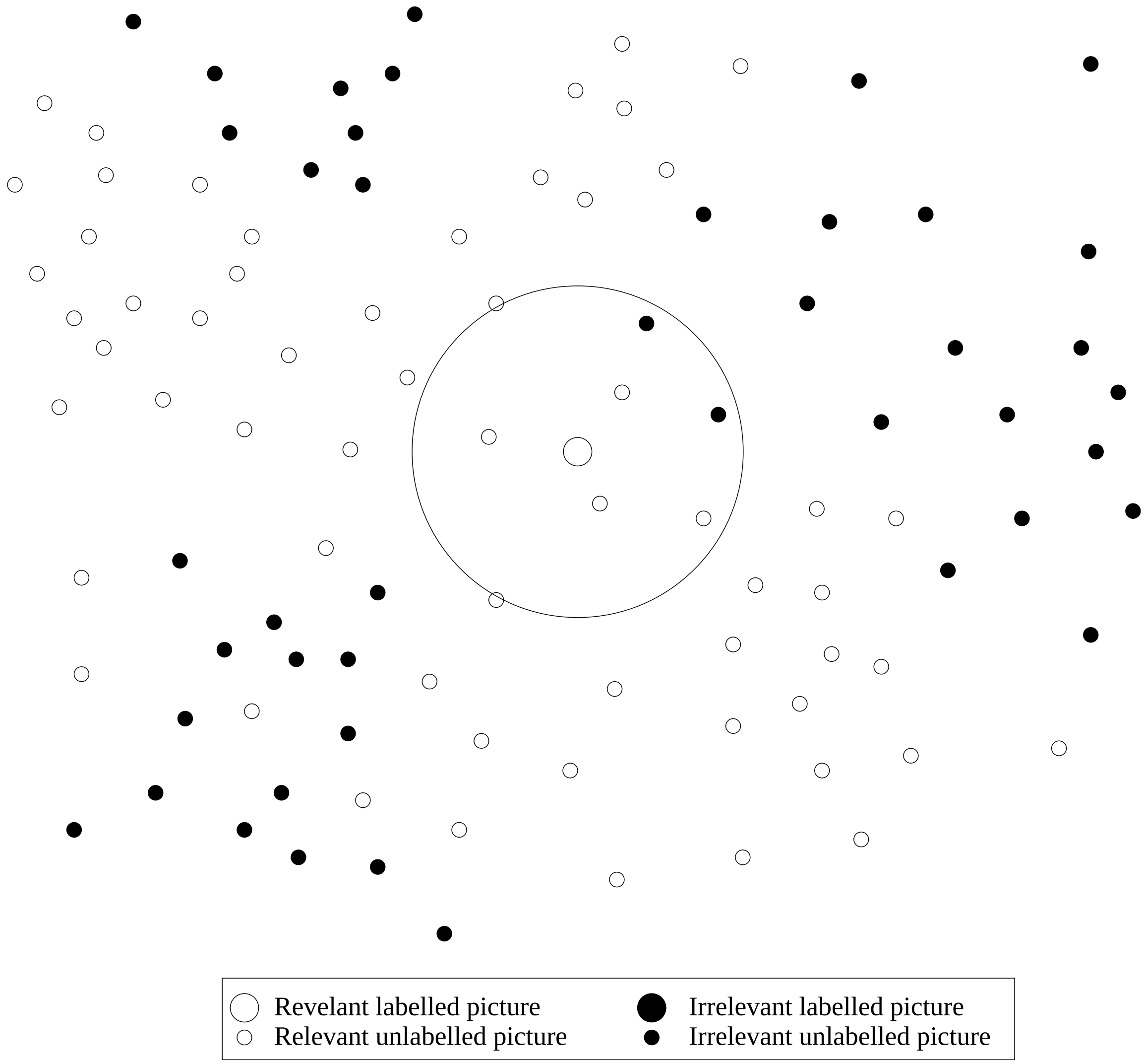
- Boundary Correction (BC):
  - $O = \mathrm{argsort}\ f$, and s the index of the current threshold: $f^\star(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}_{O_s})$

  $\rightarrow$ Update: $s(t+1) = s(t) + 2(\mathrm{pos}(t) - \mathrm{neg}(t))$ with:
    - $\mathrm{pos}(t)$ (resp. $\mathrm{neg}(t)$) $=$ number of relevant (resp. irrelevant) labels
    - t = feedback iteration number
  - Efficient during the first feedback steps

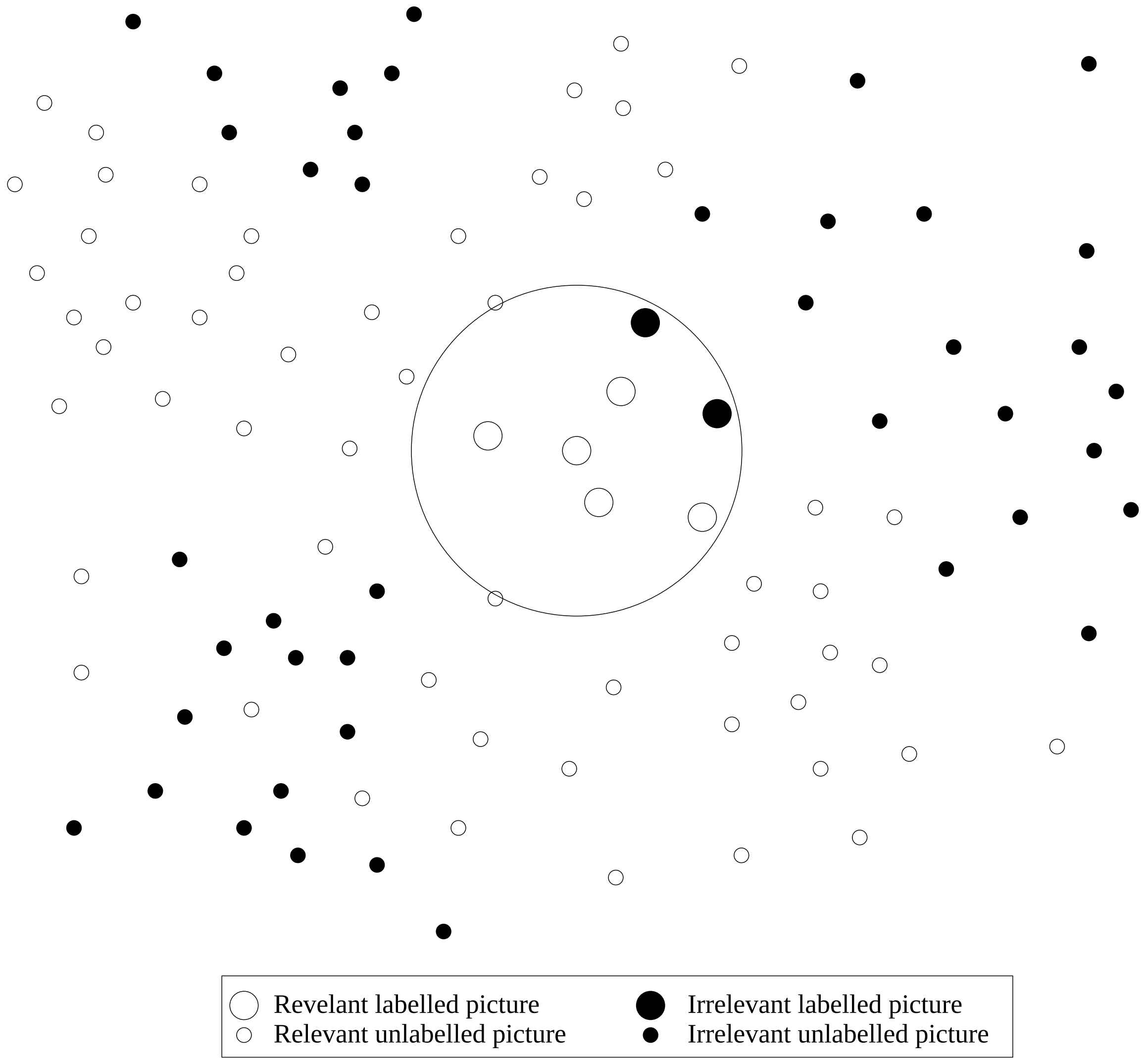
An example of retrieval session (1/8)



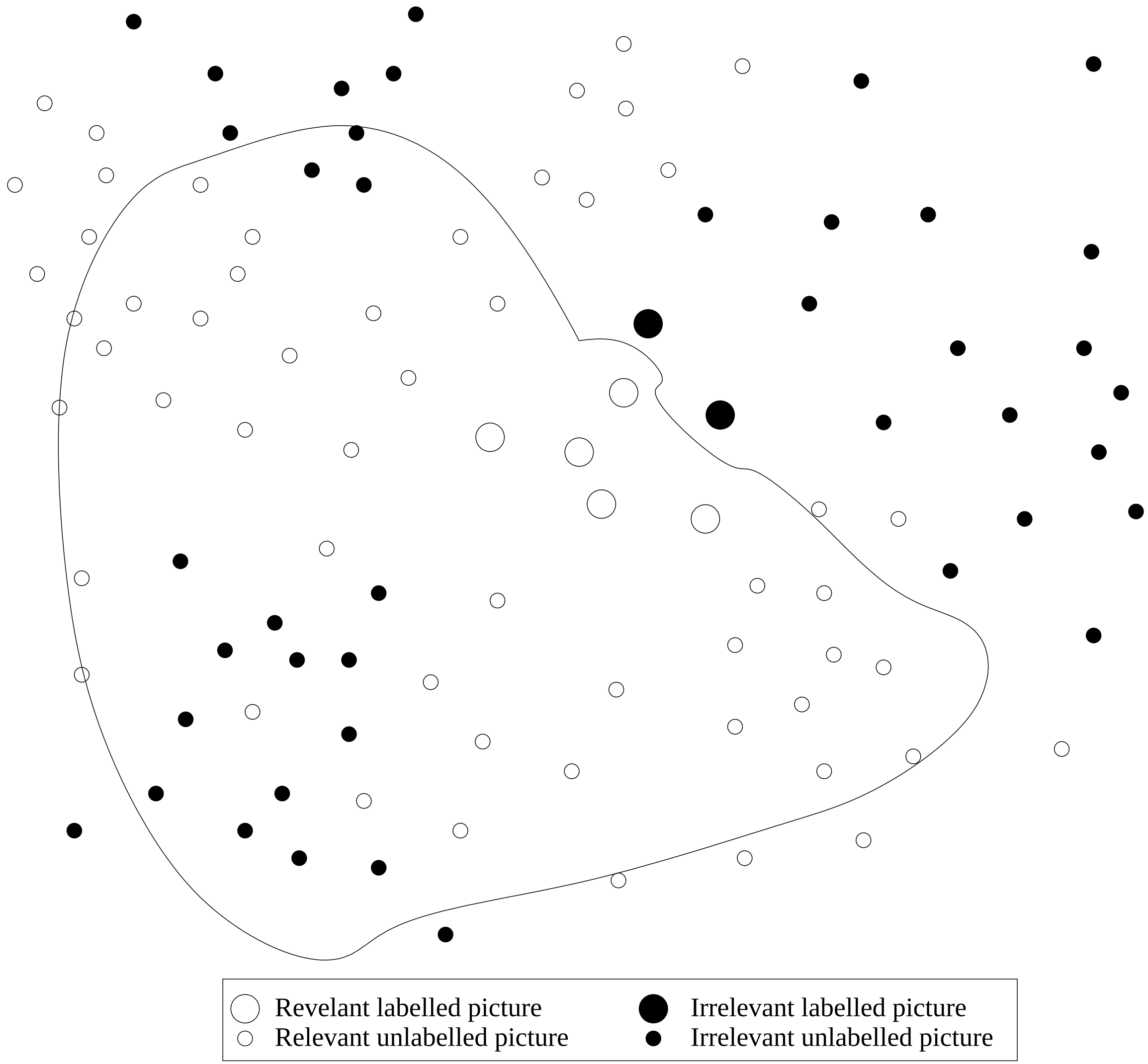
An example of retrieval session (2/8)



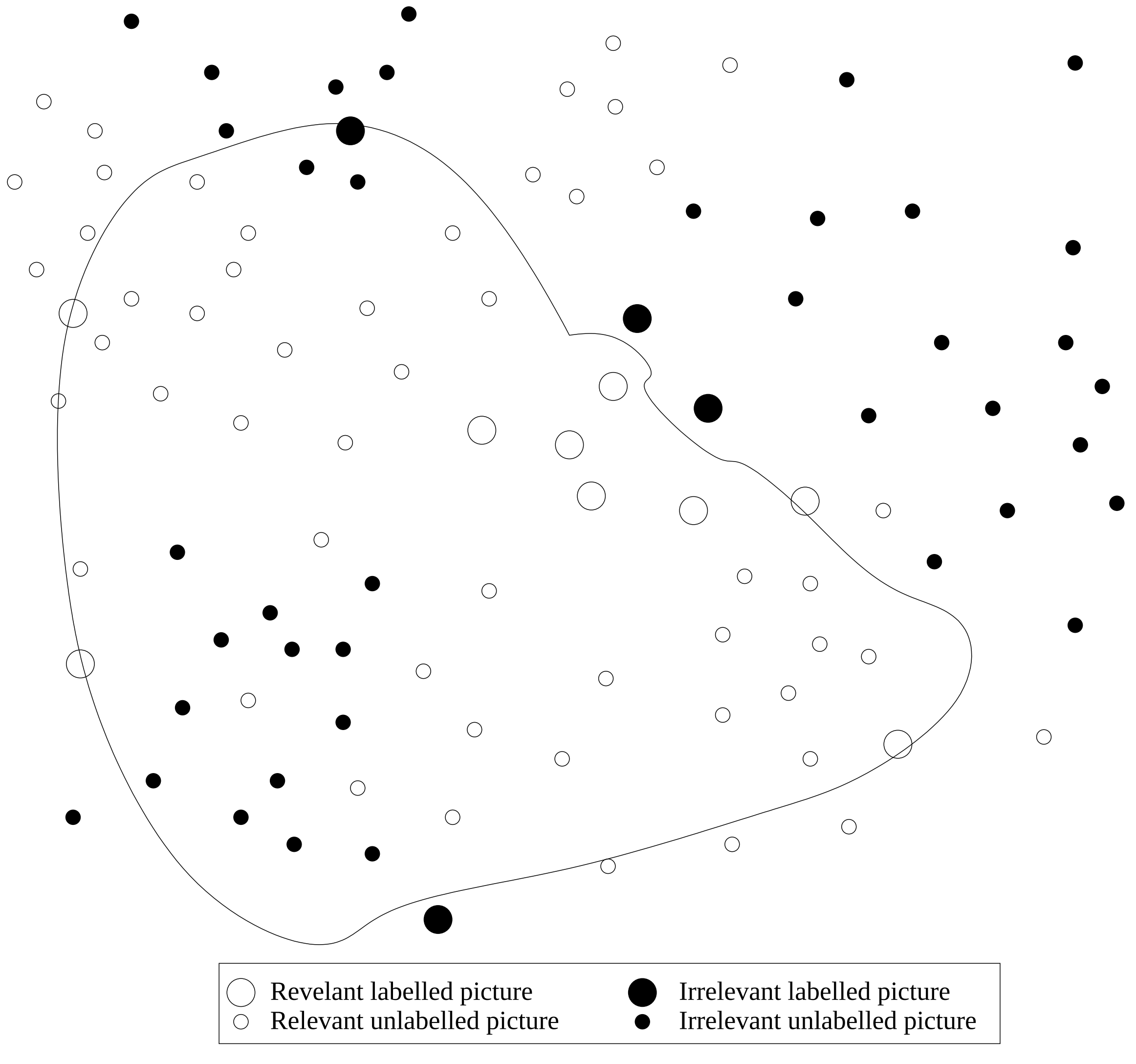
An example of retrieval session (3/8)



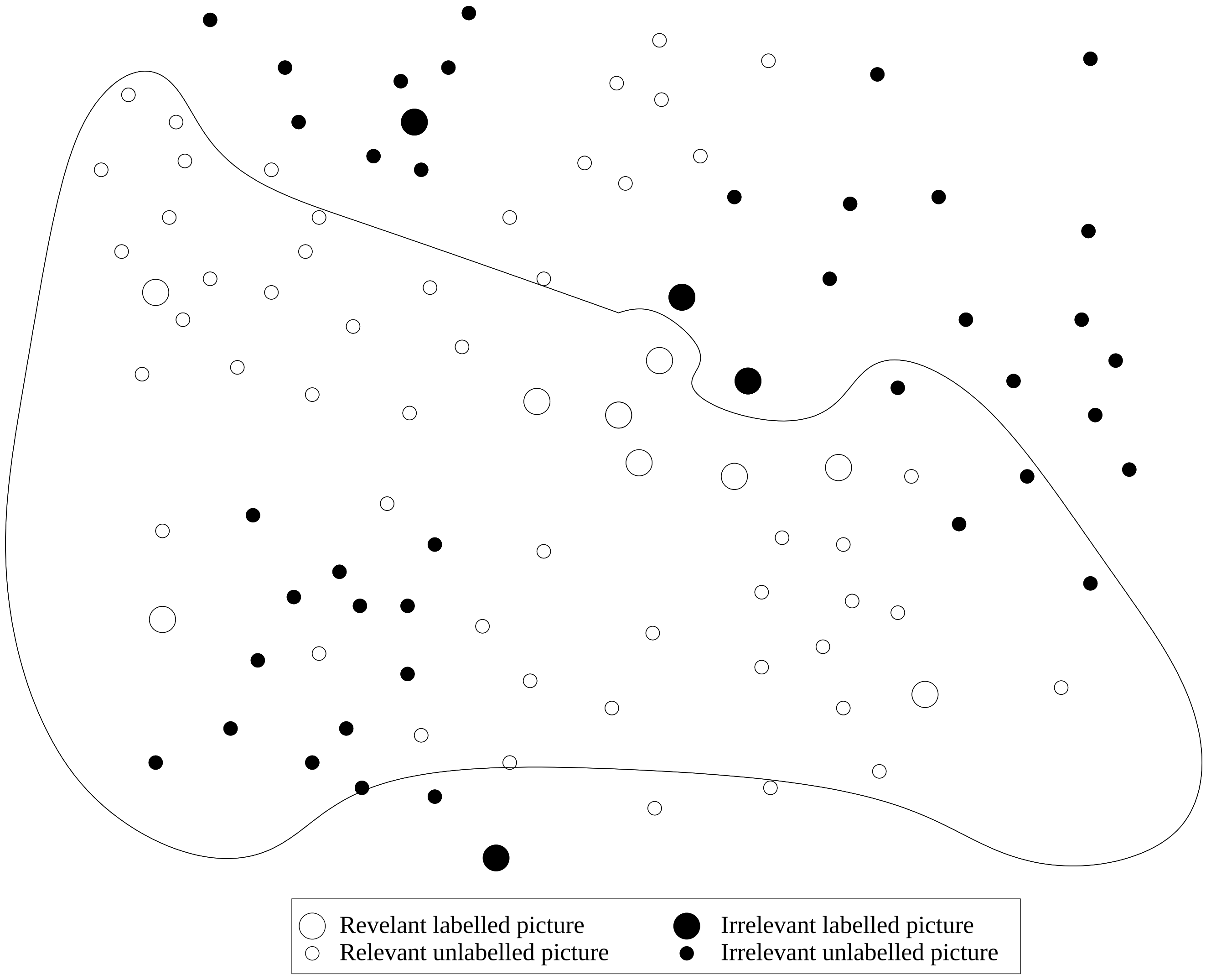
An example of retrieval session (4/8)



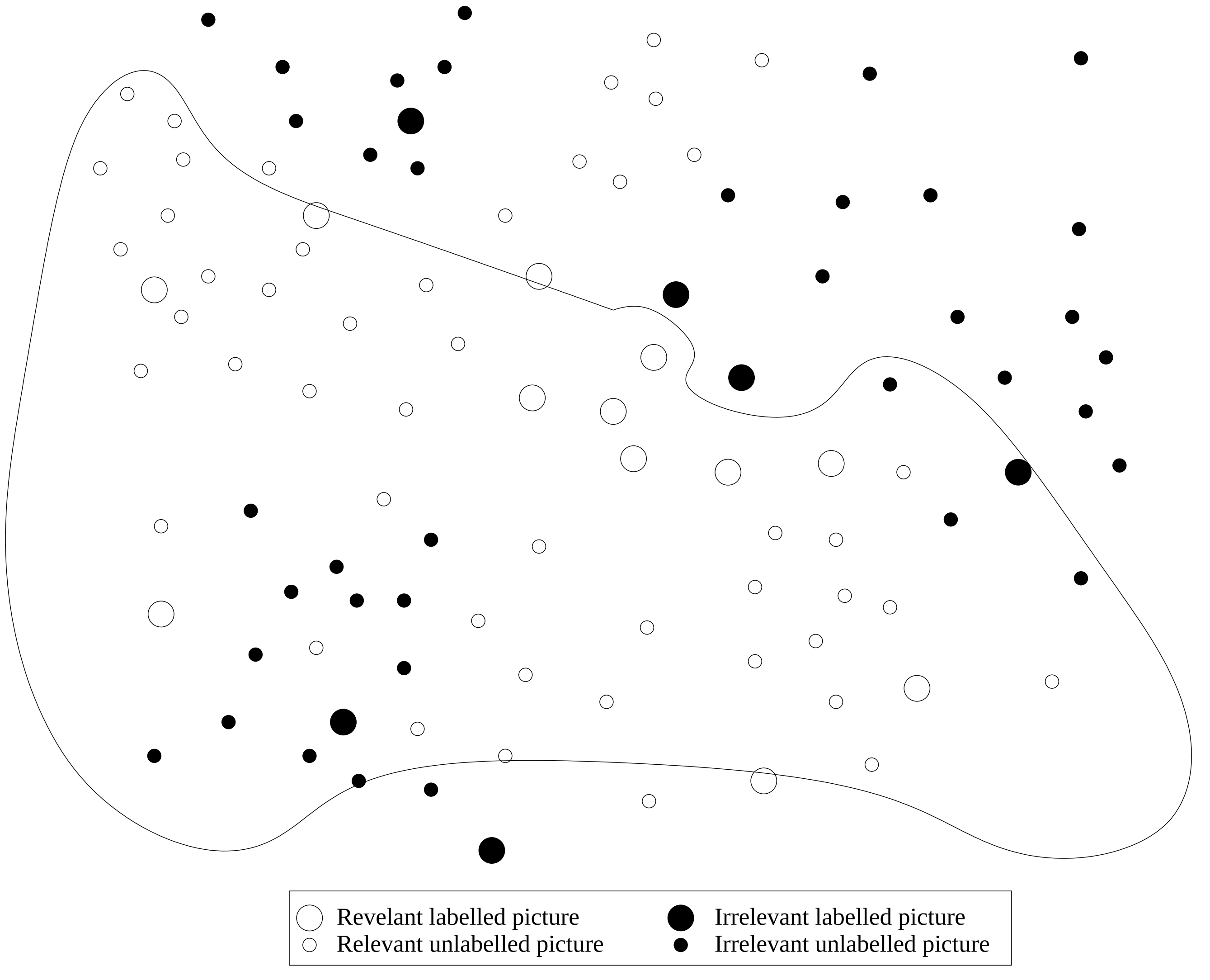
An example of retrieval session (5/8)



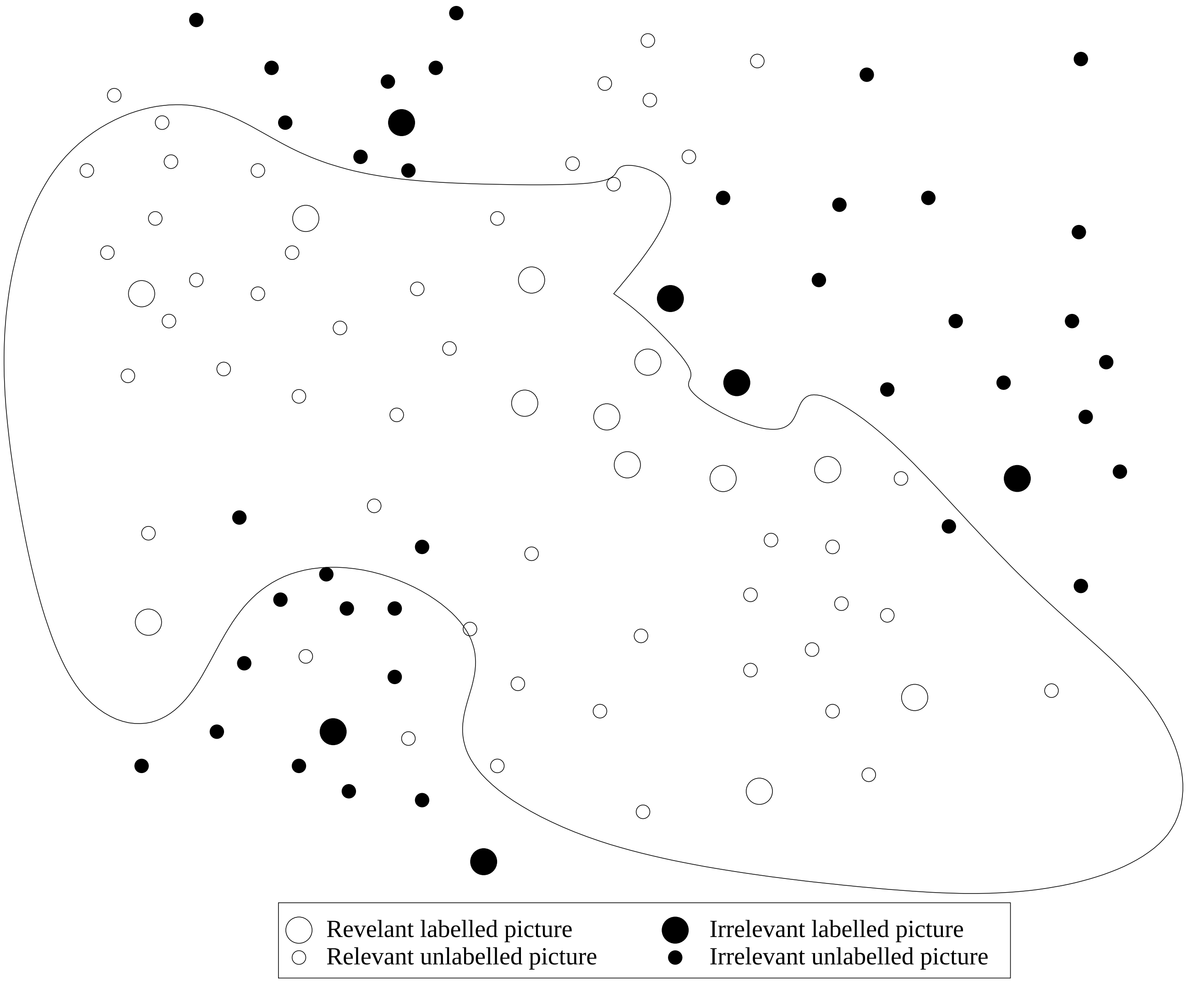
An example of retrieval session (6/8)



An example of retrieval session (7/8)



An example of retrieval session (8/8)



Experiments

Methods	Top-100(%)	Time
no AL	16	0.07s
UB	28	0.41s
ER	30	600s
UB+AD	31	60s
ER+AD	34	700s
BC+UB+AD	36	60s
BC+ER+AD	35	700s

Protocol:

- COREL Photo database (6,000 images);
- 50 categories, 100-300 size;
- Training set of 50 points (10 feedback steps, 5 labels per step).

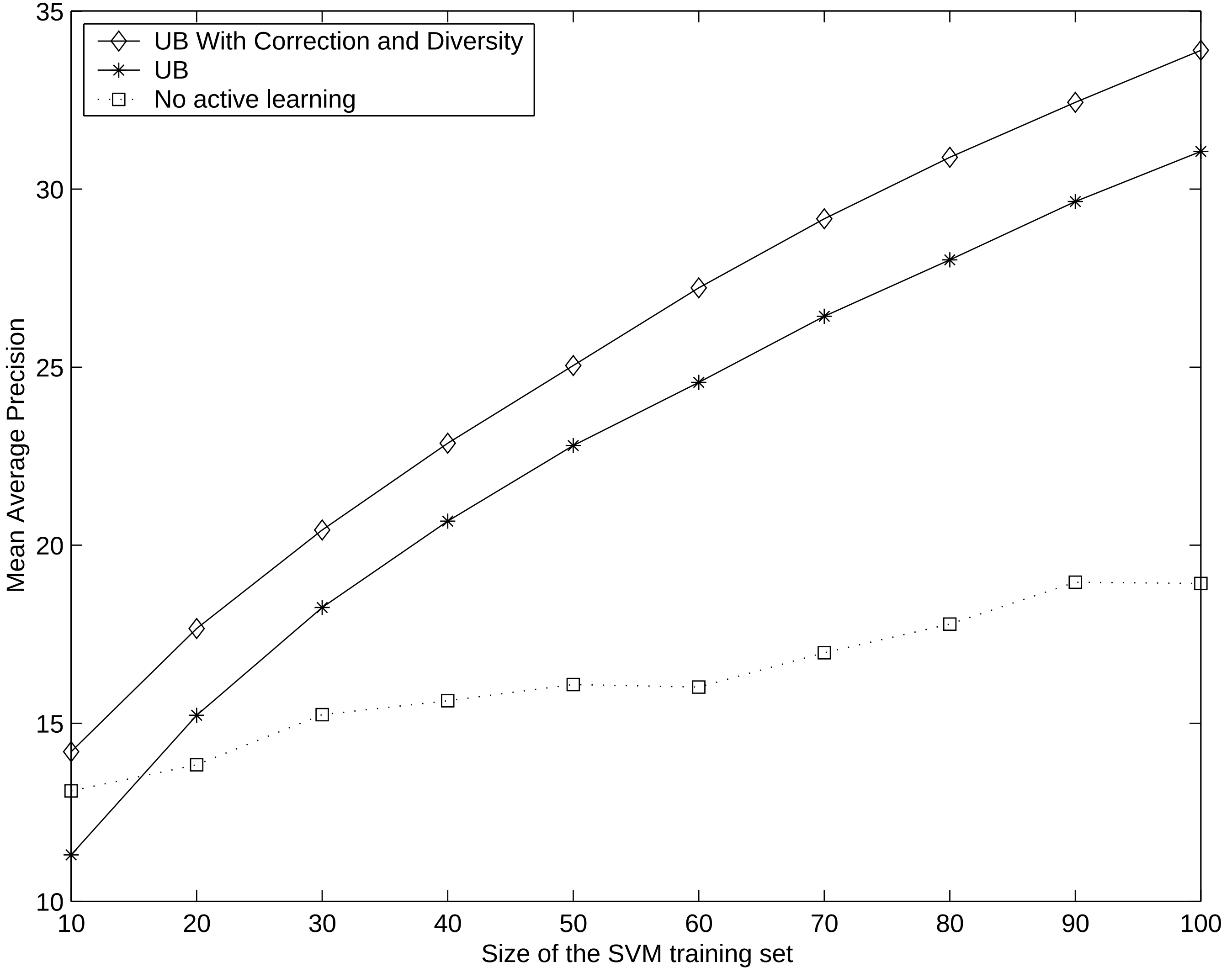
Experiments

Methods	MAP(%)
no AL	20
UB	31
ER	32
UB+AD	37
ER+AD	37
BC+UB+AD	39
BC+ER+AD	38

Protocol:

- COREL Photo database (6,000 images);
- 50 categories, 100-300 size;
- Training set of 200 points (20 feedback steps, 10 labels per step).

Experiments



Experiments

Experiments

Conclusion for Active Learning

- Active learning gives a theoretical framework for relevance feedback
- Boundary correction efficient for uncertainly-based techniques
- Adding diversity improves the performances

- Active learning framework can be used for other optimization problems in CBIR:

  $\longrightarrow$ Kernel parameters

  $\longrightarrow$ Long-term, similarity matrix learning

Semi-supervised classification for CBIR

Semi-supervised classification

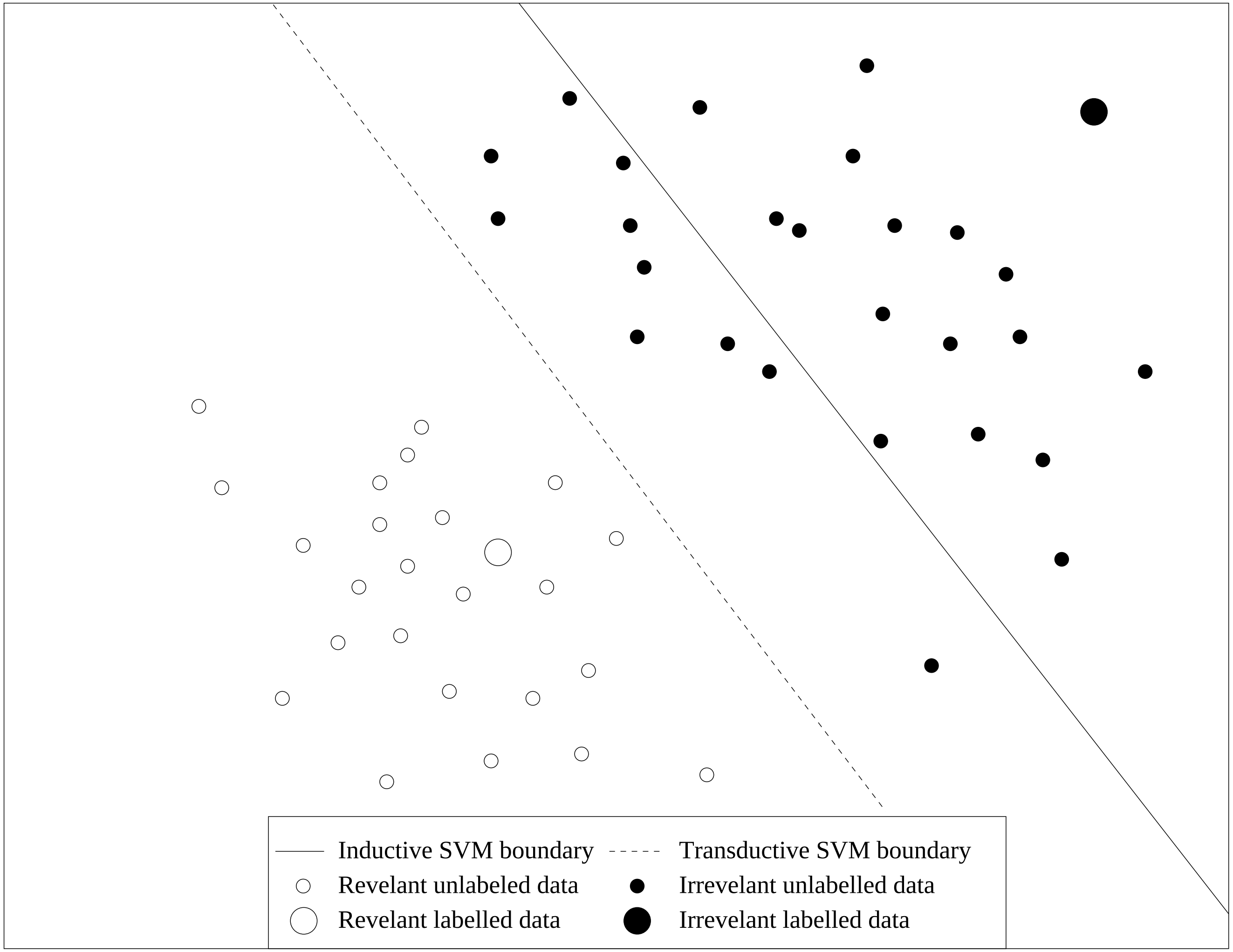
Deal with the few training data (c2) and many unlabelled data (c3) characteristics

$\rightarrow$ use unlabelled data to compensate scarcity of training data.

Three representative methods:

- Transductive SVM (Joachims):
  $\rightarrow$ Maximize the margin considering all data.
- Gaussian Mixture (Najjar):
  $\rightarrow$ Estimate the gaussians using all data.
- Gaussian Fields (Zhu):
  $\rightarrow$ Estimate the densities using harmonic functions.

Transductive SVM (Joachims)



Experiments

Methods	Error(%)	MAP(%)	Time
SVM	2.29	20	0.13s
TSVM	2.29	20	10.7s
GM	20.2	9	12.1s
GF	?	?	>10min

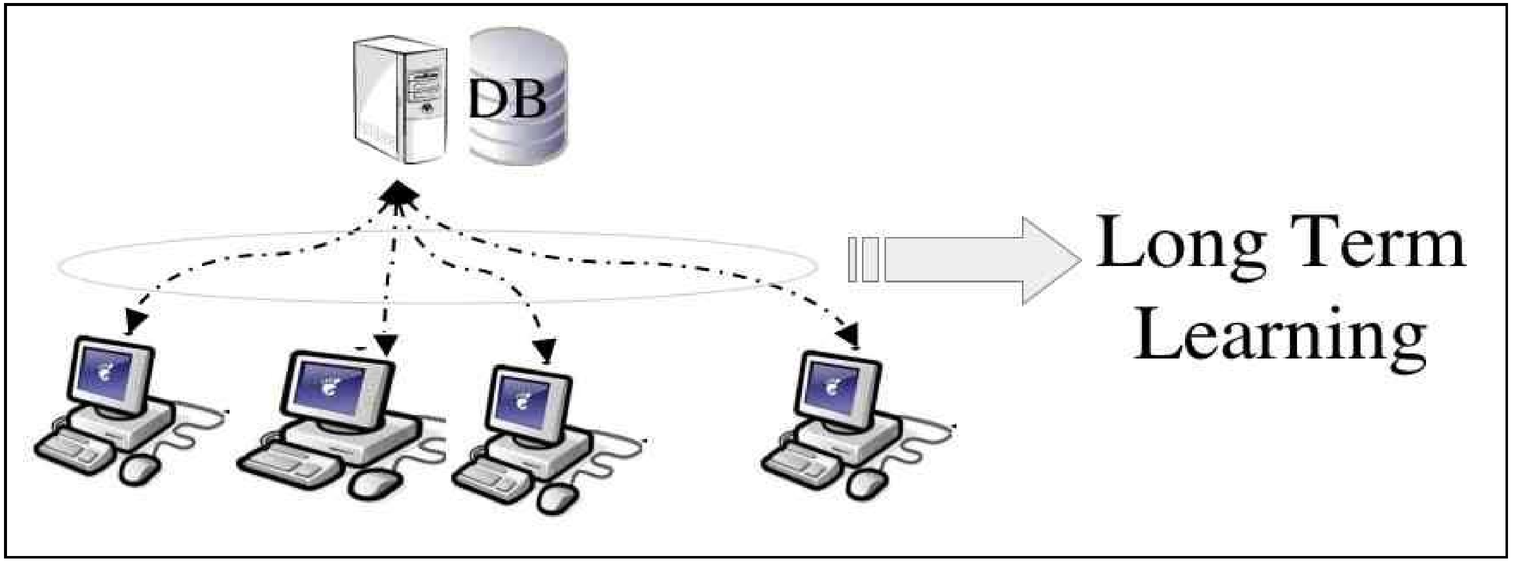
Protocol:

- COREL Photo database (6,000 images);
- 50 categories, 100-300 size;
- Training set of 200 point (unbalanced).

Long Term or Semantic learning for interactive image retrieval

Challenge

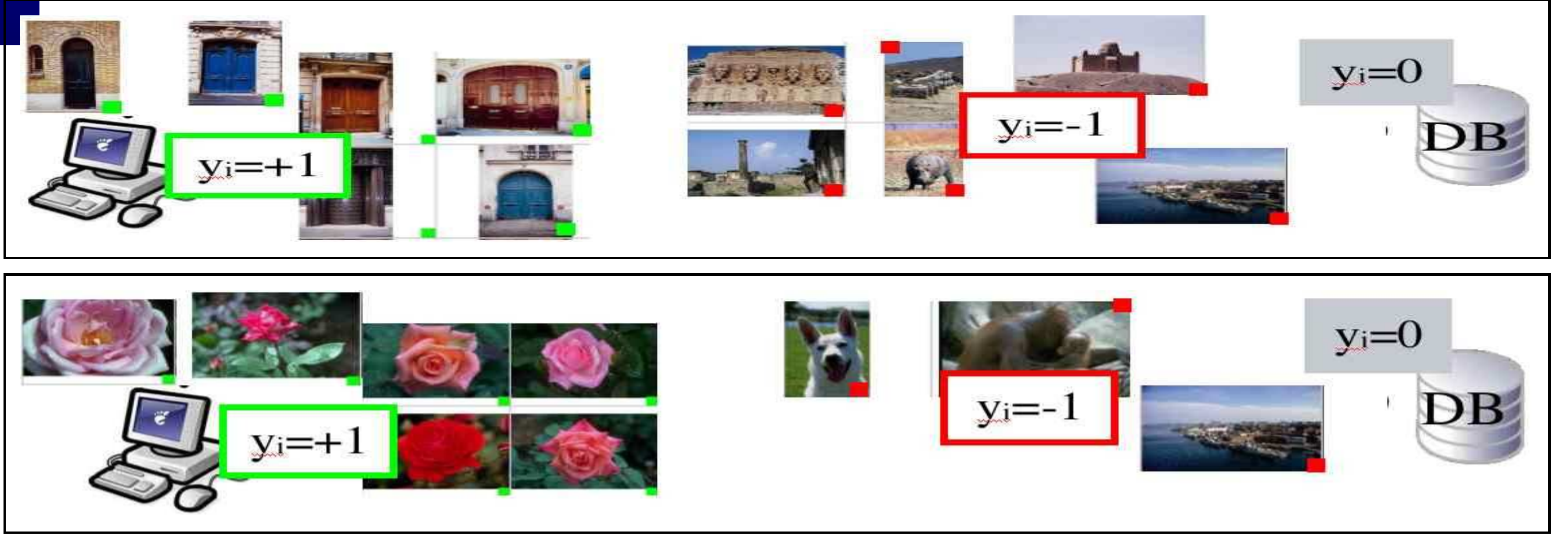
- Active research: Interactive learning techniques of image categories using relevance feedback;



- Limitation: Knowledge lost at the end of each retrieval session;
- Proposition: Learn from past retrieval sessions

Retrieval sessions

- Interactive learning (relevance feedback, SVM, active learning) to retrieve an image category , a subset of the database
- In our framework, binary labels:



- <u>At the end of a retrieval session:</u> Collect all these labels in a vector $\mathbf{y}$

Training set

- All vectors $\mathbf{y}$ in a matrix $\mathbf{Y}$:

$$
\mathbf{Y}: \begin{array}{c|ccccccccc}
 & \mathbf{y}(1) & \mathbf{y}(2) & \mathbf{y}(3) & \mathbf{y}(4) & \mathbf{y}(5) & \mathbf{y}(6) & \mathbf{y}(7) & \ldots & \mathbf{y}(M) \\
\hline
\mathbf{x}_1 & 1 & 1 & 0 & 0 & -1 & 1 & 0 & \ldots & 0 \\
\mathbf{x}_2 & 1 & 1 & 1 & 1 & -1 & 0 & 1 & \ldots & 0 \\
\mathbf{x}_3 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & \ldots & 1 \\
\mathbf{x}_4 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & \ldots & 1 \\
\mathbf{x}_5 & -1 & 0 & 0 & 1 & 1 & -1 & 0 & \ldots & 0 \\
\mathbf{x}_6 & 0 & 0 & -1 & 0 & 1 & 0 & -1 & \ldots & -1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
\mathbf{x}_N & 0 & 1 & 0 & 0 & 0 & -1 & 0 & \ldots & 0
\end{array}
$$

- This matrix $\mathbf{Y}$ is the training set

Exotic learning problem

- Partial knowledge

  $\longrightarrow$ Not possible to identify 1 category from 1 retrieval session
- Unknown category for each retrieval session

  $\longrightarrow$ Not possible to easily rebuild categories
- Mixed categories

  <u>Proposition:</u>
- Do not work on learning some explicit categories but on the learning of the database similarities
- Long-term learning or semantic (from $\mathbf{Y}$) learning
- Assumption: search for a finite nb of categories

A Gram matrix approach

- Aim: Optimize the matrix $\mathbf{K}$ of similarities $k(\mathbf{x}_i, \mathbf{x}_j)$ between the N images of the database
- Naïve approach: increase (resp. decrease) the similarity between two images in the same (resp. different) category using an heuristic function
- Problems:
  - No more guaranty on metric properties
  - Relevance feedback techniques can not be used anymore
- Proposition: consider a **definite positive matrix** of similarities $\mathbf{K}$ (a kernel matrix)

  $\rightarrow$ Update $\mathbf{K}$ under constraints to keep dp properties

Adaptive Method RETIN SL

- Reinforce kernel values corresponding to positive values in $\mathbf{y}$, statistical accumulation

  $\rightarrow$ Gram matrix updating:

$$\begin{aligned}\mathbf{K}(t+1) &= \mathrm{update}(\mathbf{K}(t), \mathbf{y}(t), \rho(t)) \\ &= (1-\rho(t))\mathbf{K}(t) + \rho(t) \times \mathrm{merge}(\mathbf{K}(t), \mathbf{y}(t))\end{aligned}$$

  with $\mathbf{K}(t)$ the Gram matrix at iteration t, and $\mathbf{y}(t)$ a randomly selected vector of $\mathbf{Y}$, and $\mathrm{merge}(\mathbf{K}(t), \mathbf{y}(t))$ an operator to merge knowledge in $\mathbf{K}(t)$ and $\mathbf{y}(t)$;

- Merging strategy: the resulting form with two components:

$$\mathrm{merge}(\mathbf{K}, \mathbf{y}) = a \times (\mathbf{TKT}^t + b\mathbf{K_u})$$

First condition: unbalance updating

- Two class problem:

$$\forall(\mathbf{x}_i, \mathbf{x}_j), \qquad \mathbf{y}_i\mathbf{y}_j > 0 => k(\mathbf{x}_i, \mathbf{x}_j) \nearrow$$

- Unbalance updating:

$$\text{XClass handling} \quad \begin{cases} \text{If } \mathbf{y}_i > 0 & \text{then} \quad \Delta k(\mathbf{x}_i, \mathbf{x}_j) \quad \text{high} \\ \text{ow} & \qquad\quad \Delta k(\mathbf{x}_i, \mathbf{x}_j) \quad \text{small} \end{cases}$$

- $\forall(\mathbf{x}_i, \mathbf{x}_j), \qquad \mathbf{y}_i\mathbf{y}_j < 0 => k(\mathbf{x}_i, \mathbf{x}_j) \searrow$
- Merging strategy: $\mathbf{K_u} = \mathbf{u}\mathbf{u}^t$
  where $\mathbf{u}_k = 1$ if $\mathbf{y}_k > 0$, $\mathbf{u}_k = -\gamma$ if $\mathbf{y}_k < 0$, otherwise 0
- dp properties clearly preserved

Second condition

- Homogenize similarities in one shot $\mathbf{y}$:

(1) Cst values inside $\left\{ \begin{array}{rcll} \forall(\mathbf{y}_i, \mathbf{y}_j) & = & +1 & in\ \mathbf{y} \\ k(\mathbf{x}_i, \mathbf{x}_j) & \rightarrow & cst & (+1) \end{array} \right.$

(2) Cst V outside $\forall \mathbf{y}_i = +1\ \forall \mathbf{x}_q \in db,\ k(\mathbf{x}_q, \mathbf{x}_i) \rightarrow cst_q$

- Algebraic trick:
- $\mathbf{K} \leftarrow \mathbf{TKT}^t$, $\mathbf{T} = \left[\begin{array}{cc|ccc} 1 & 1 & & & \\ 1 & 1 & & & \\ \hline & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{array} \right]$

  - Averaging of similarities and (1) and (2) performed
  - dp also preserved

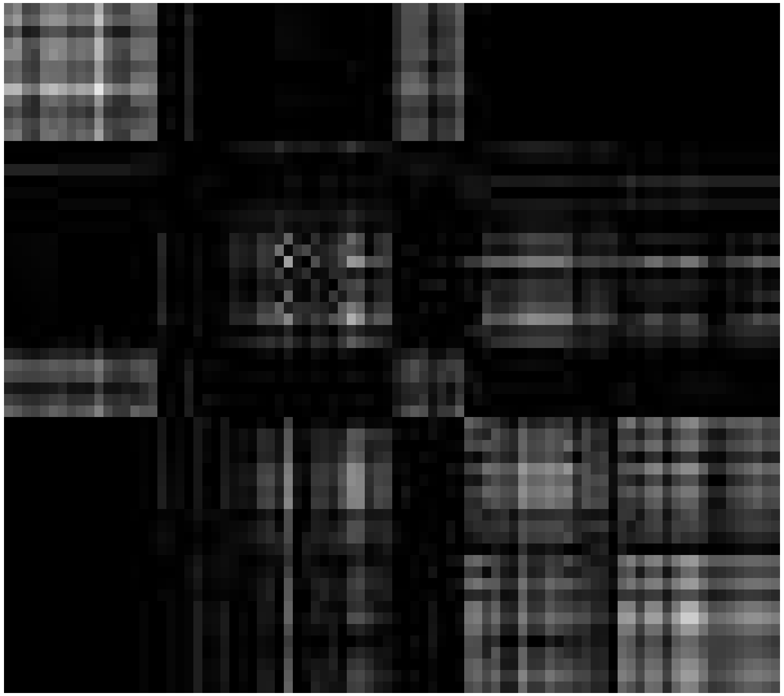
Comments

- Good news:
  - Fast evolution of the $\mathbf{K}$ similarity matrix because all the similarities between database images and labeled images are updated
  - Nice control of the matrix rank

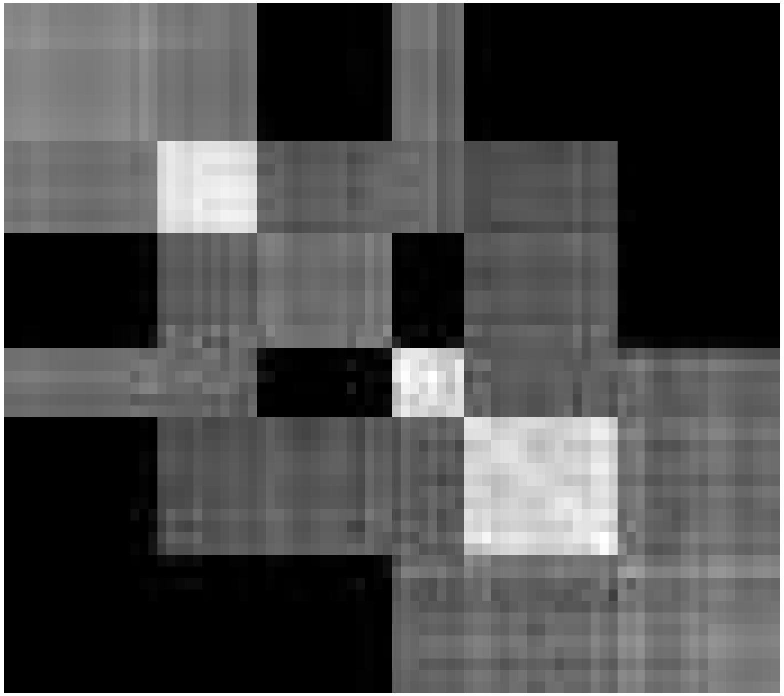
Similarity matrix

- Similarity/kernel matrix:

Before optimization



After optimization



Vector-based method

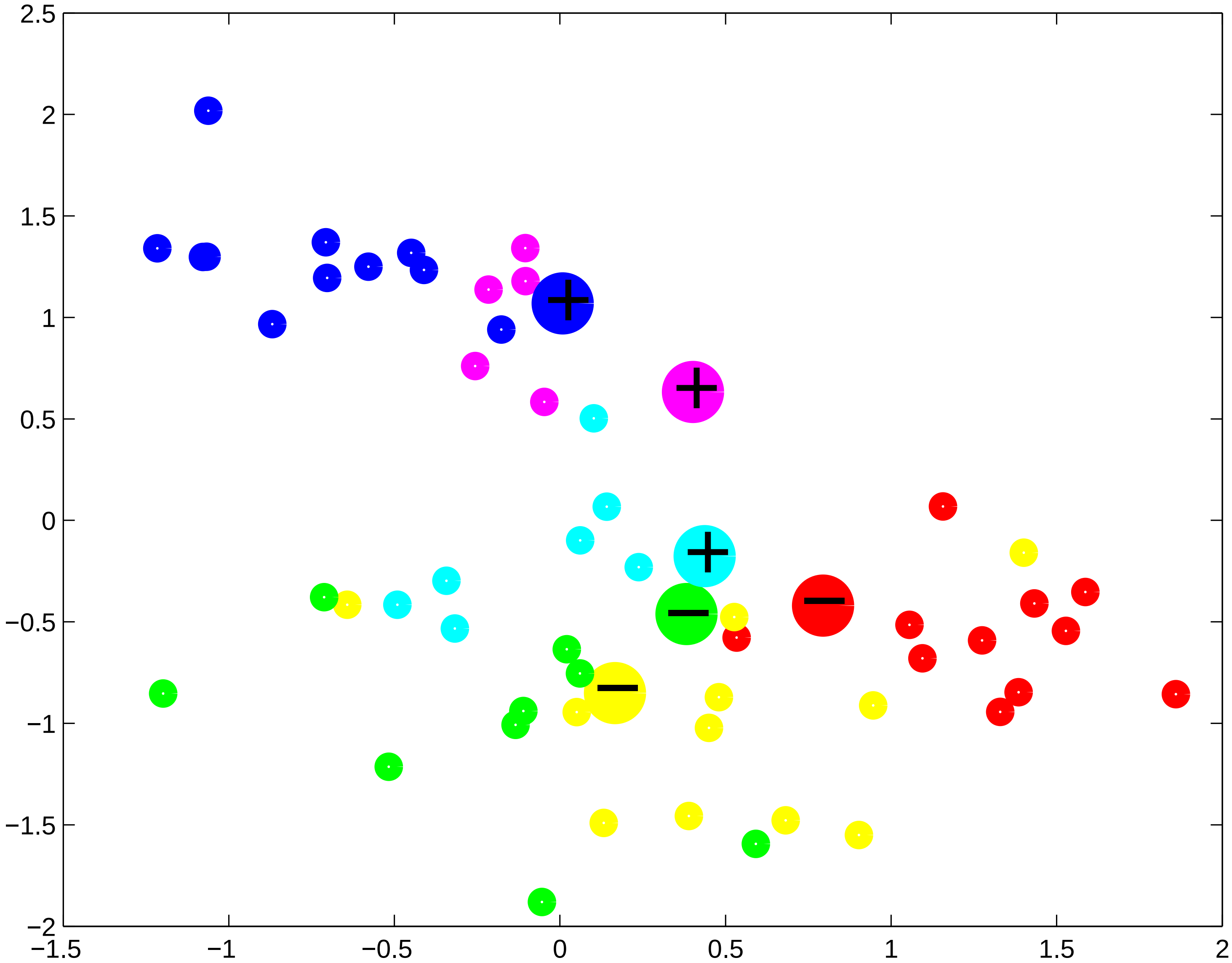
- Operator $\mathbf{TKT}^\top$ equivalent to:

$$\forall i \in I_1 \quad \mathbf{x}_i \leftarrow \frac{1}{n_1} \sum_{j \in I_1} \mathbf{x}_j$$

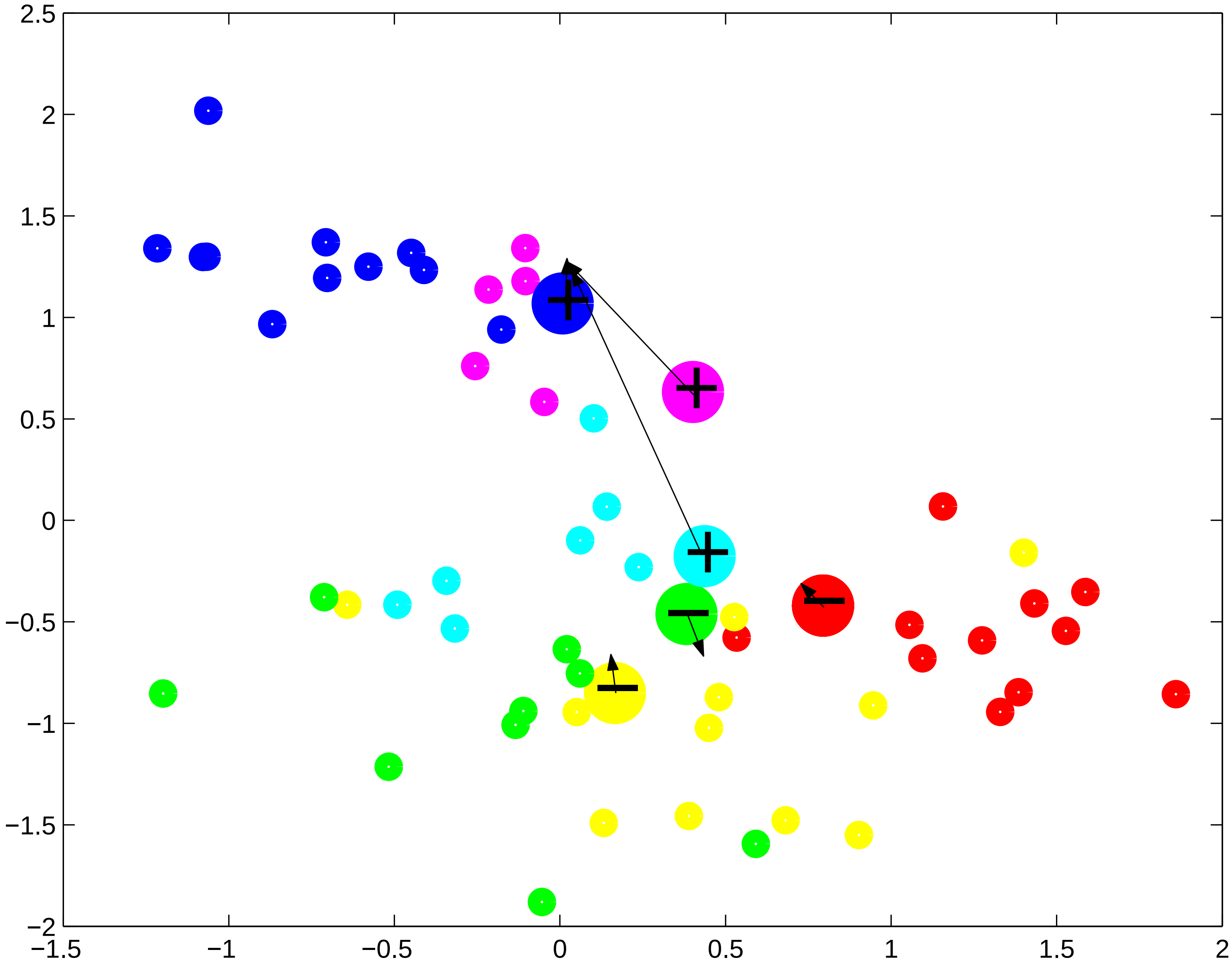
with $I_1 = \{$*relevant labeled images*$\}$

- Idea: move feature vectors;
- General scheme:
  - Group together images in clusters

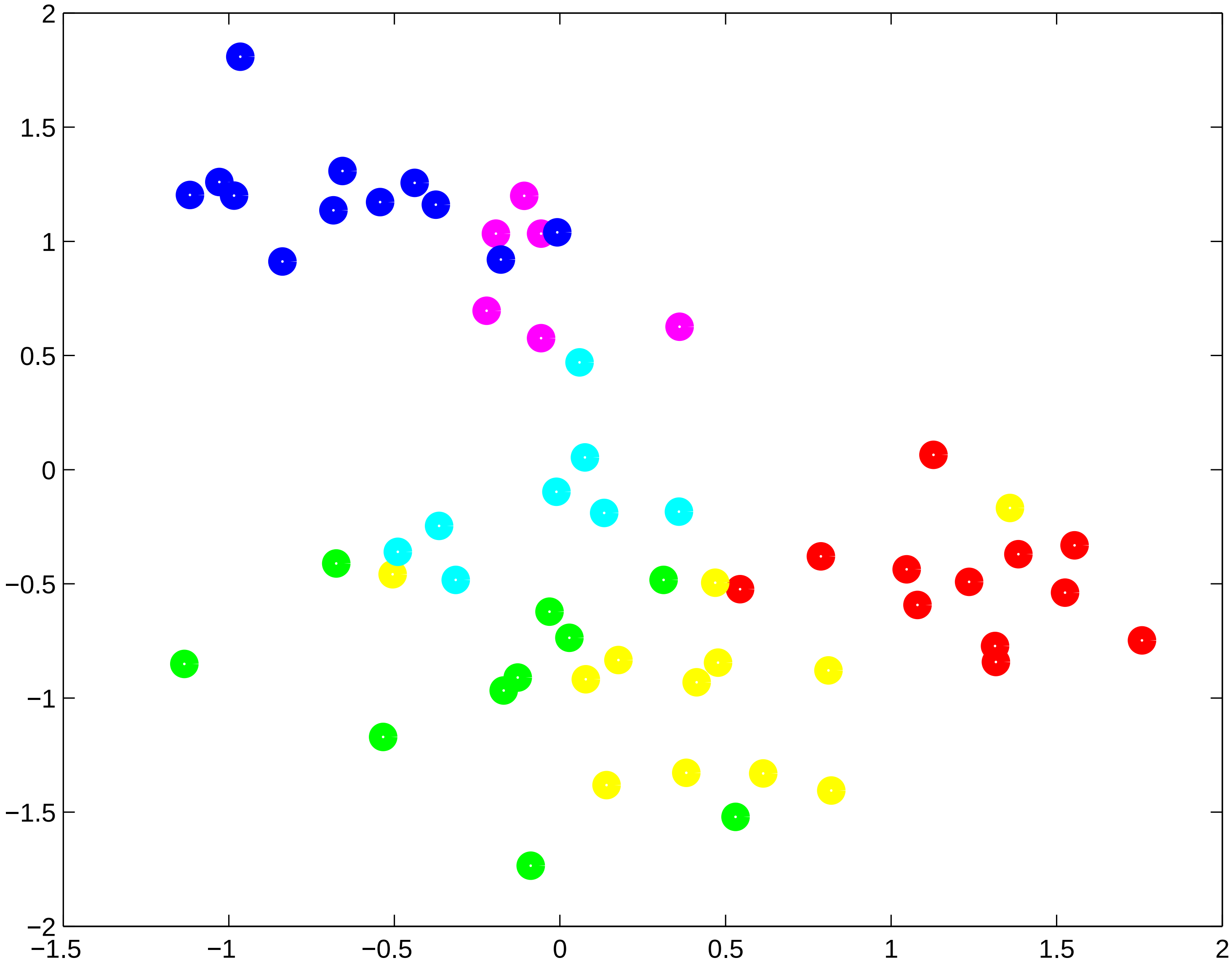
second method: Vector-based method



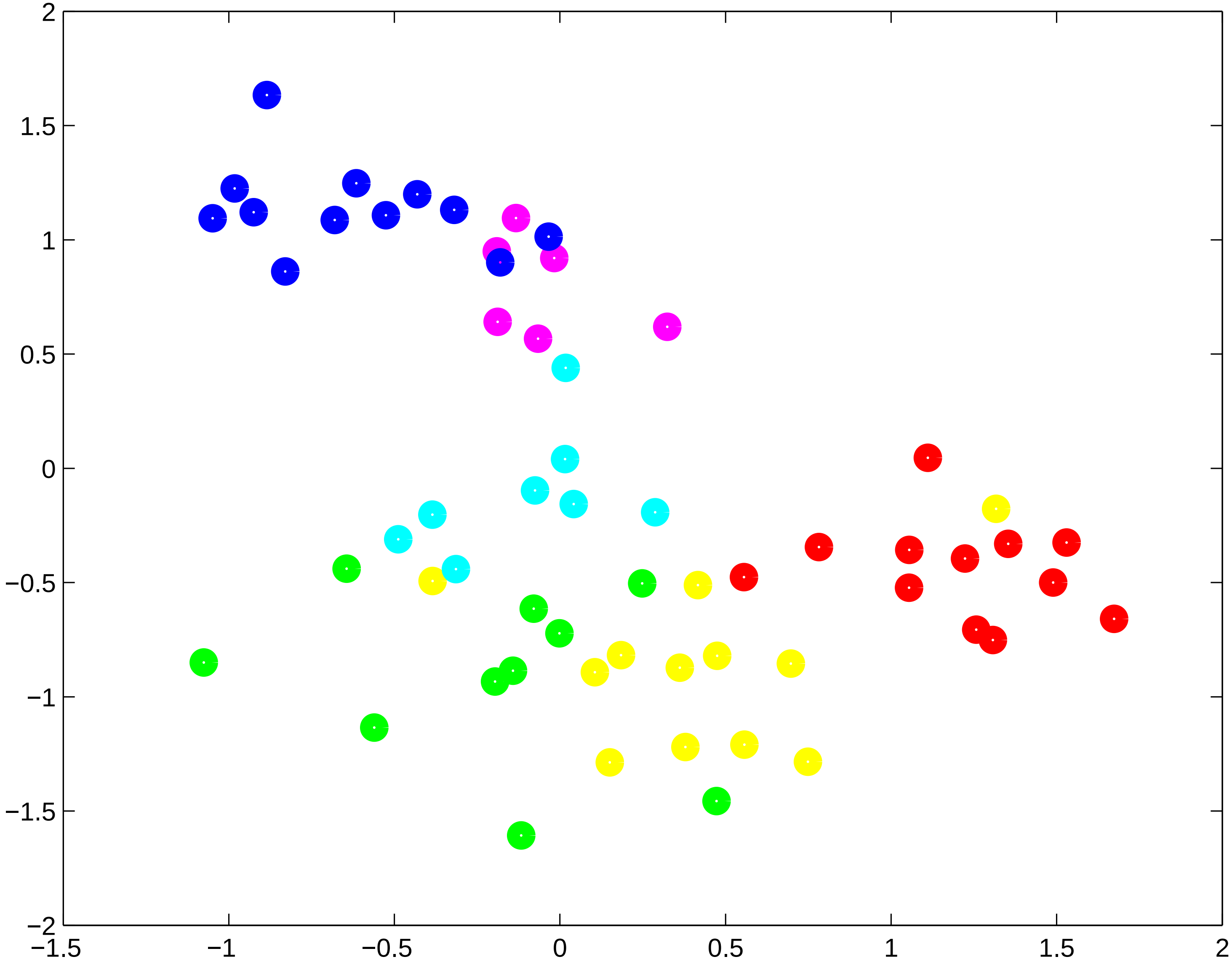
Vector-based method



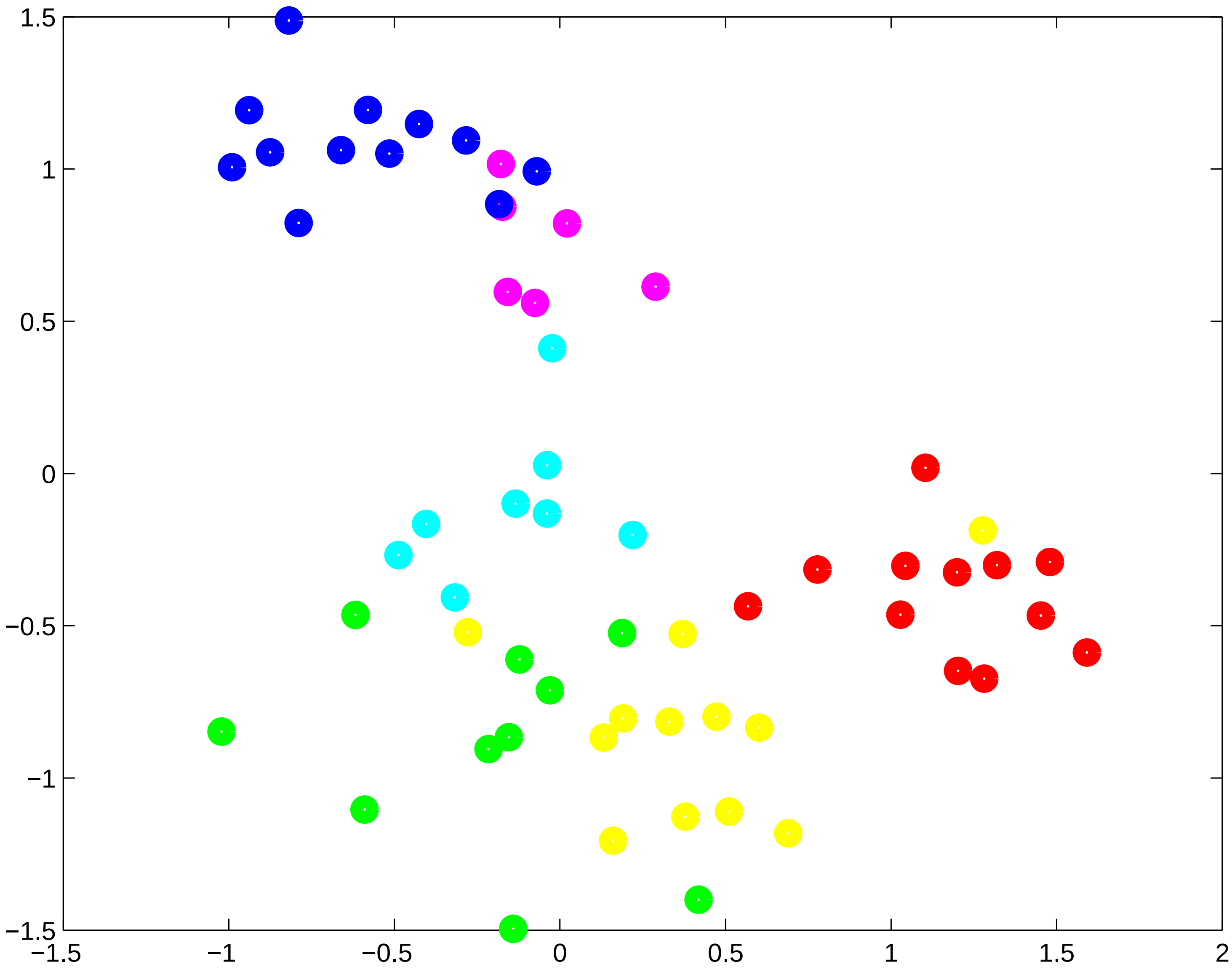
Vector-based method



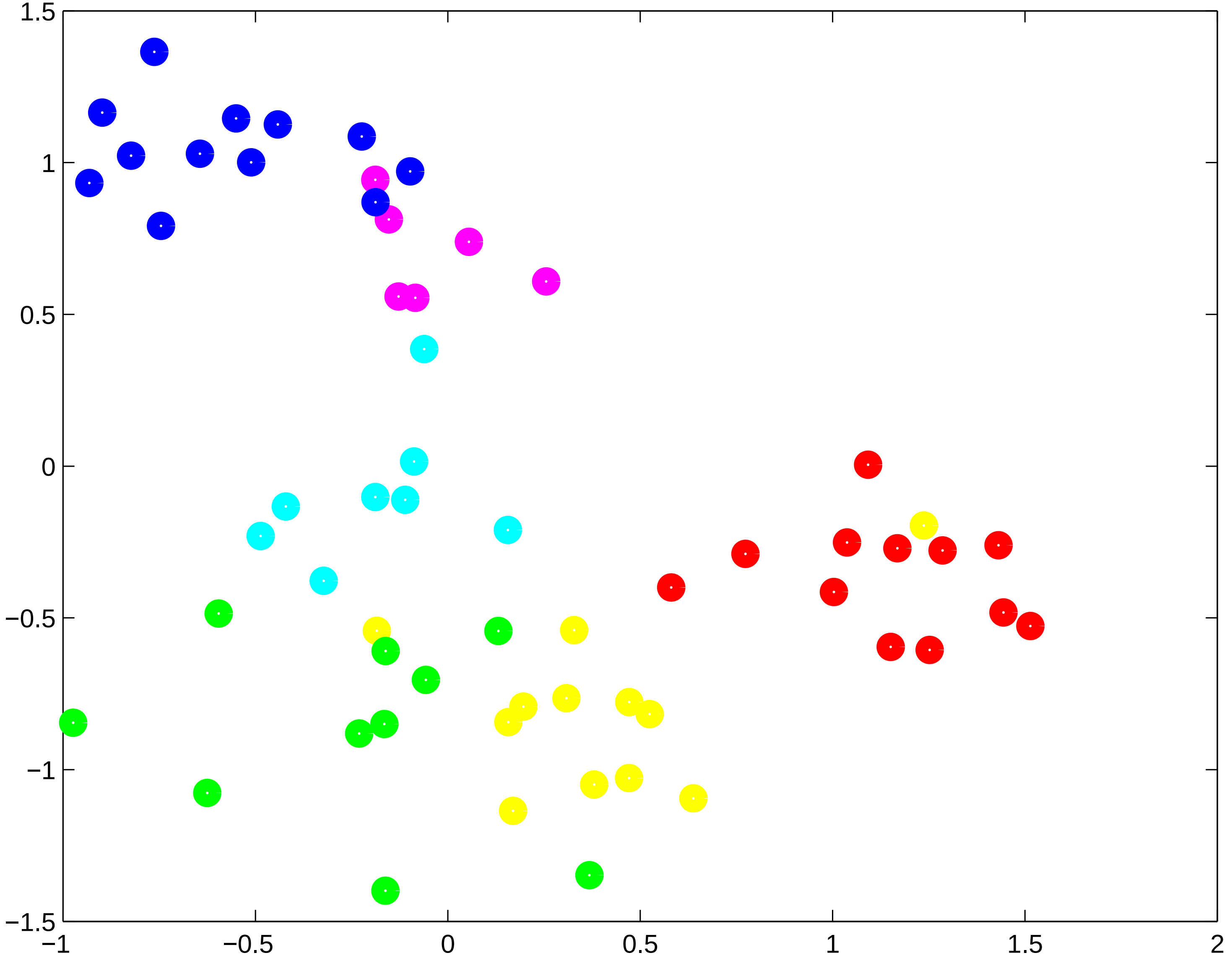
Vector-based method



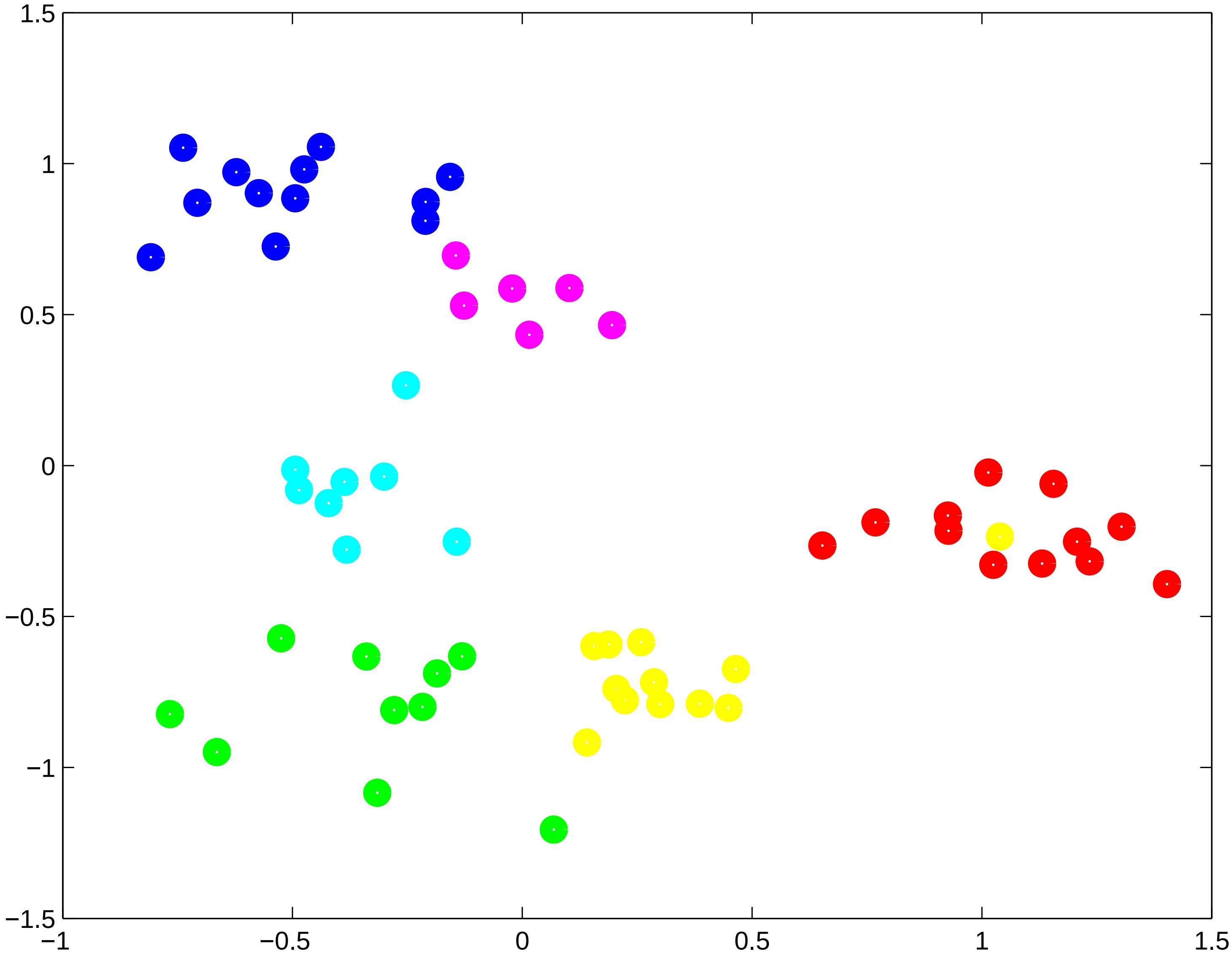
Vector-based method



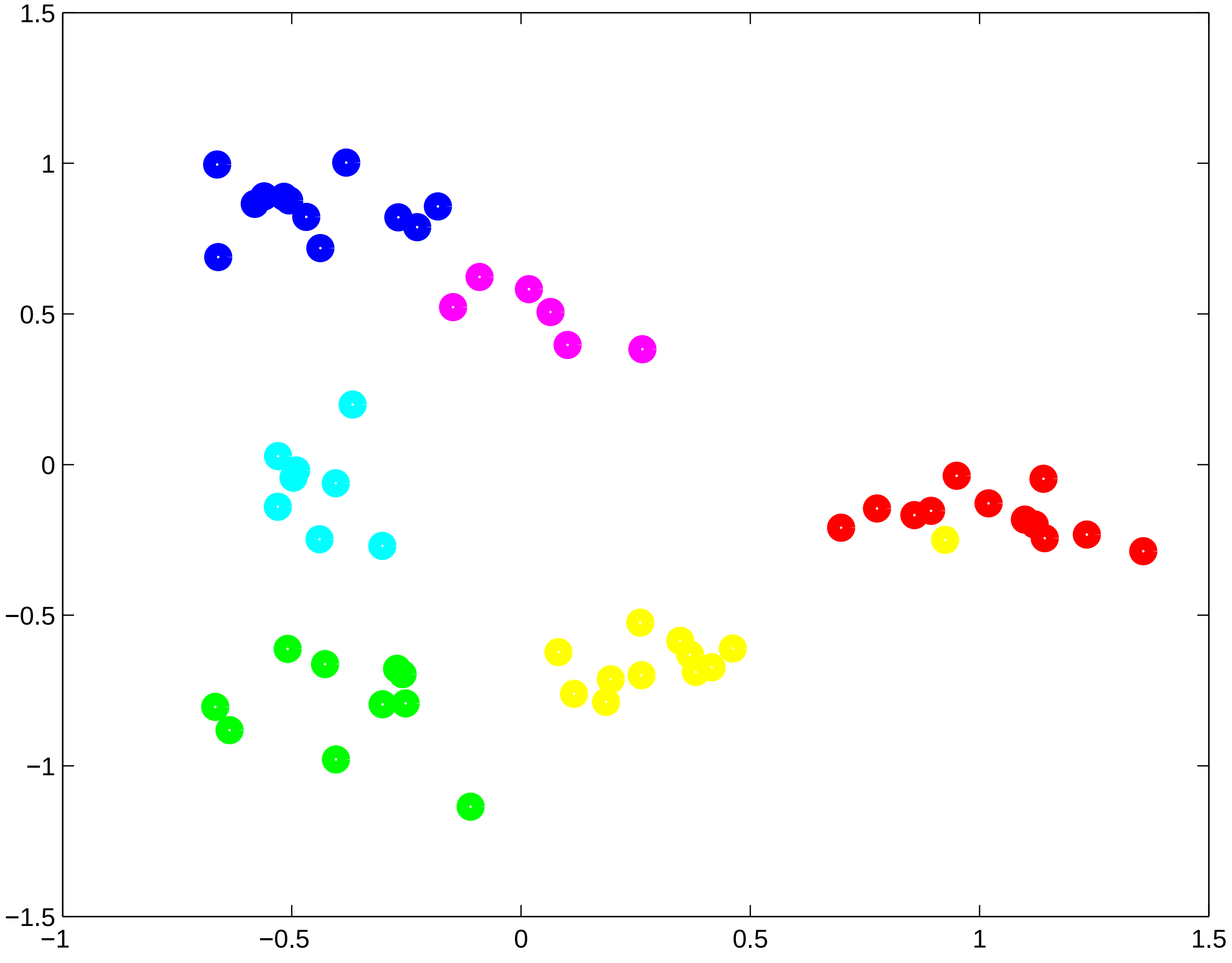
Vector-based method



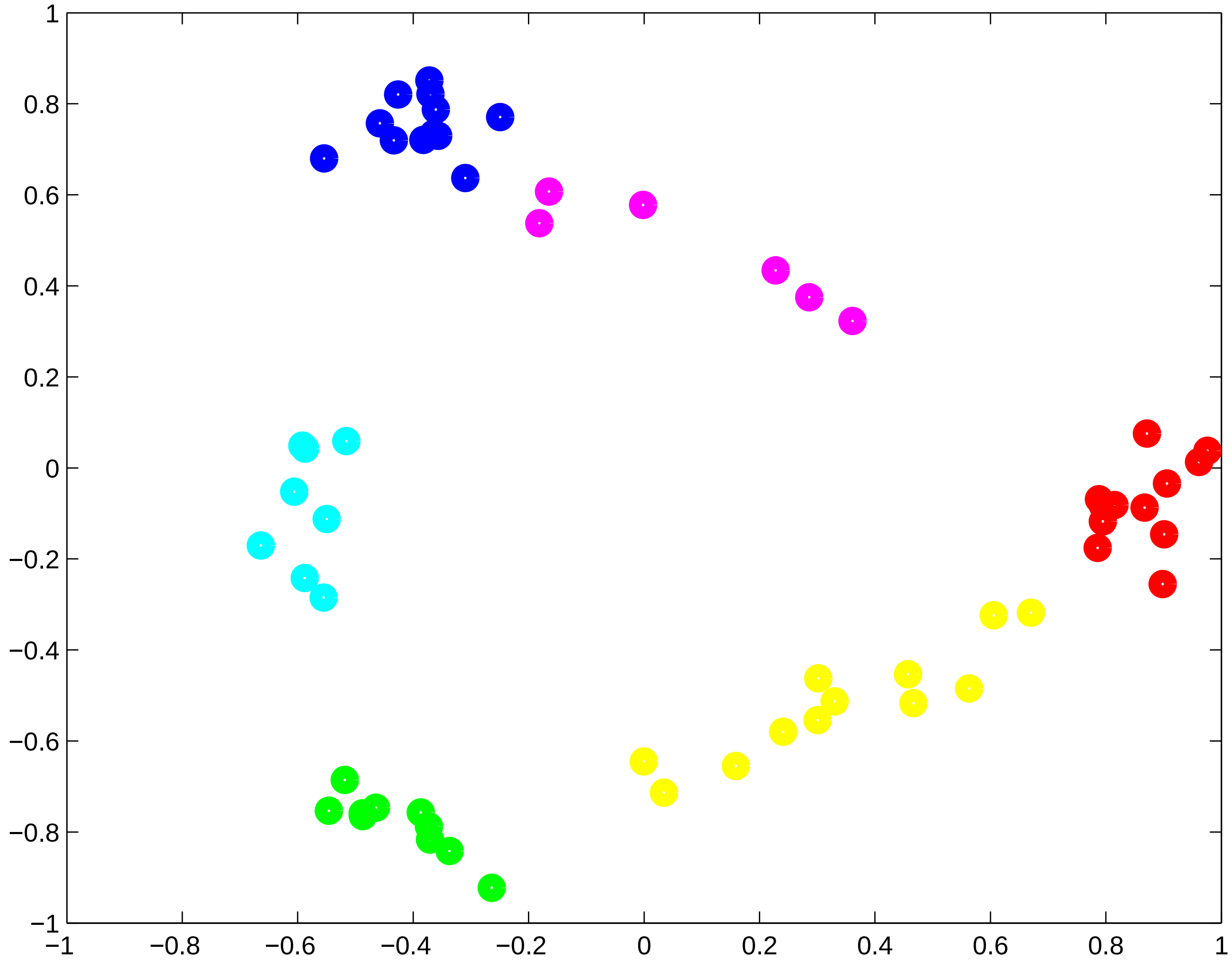
Vector-based method



Vector-based method



Vector-based method



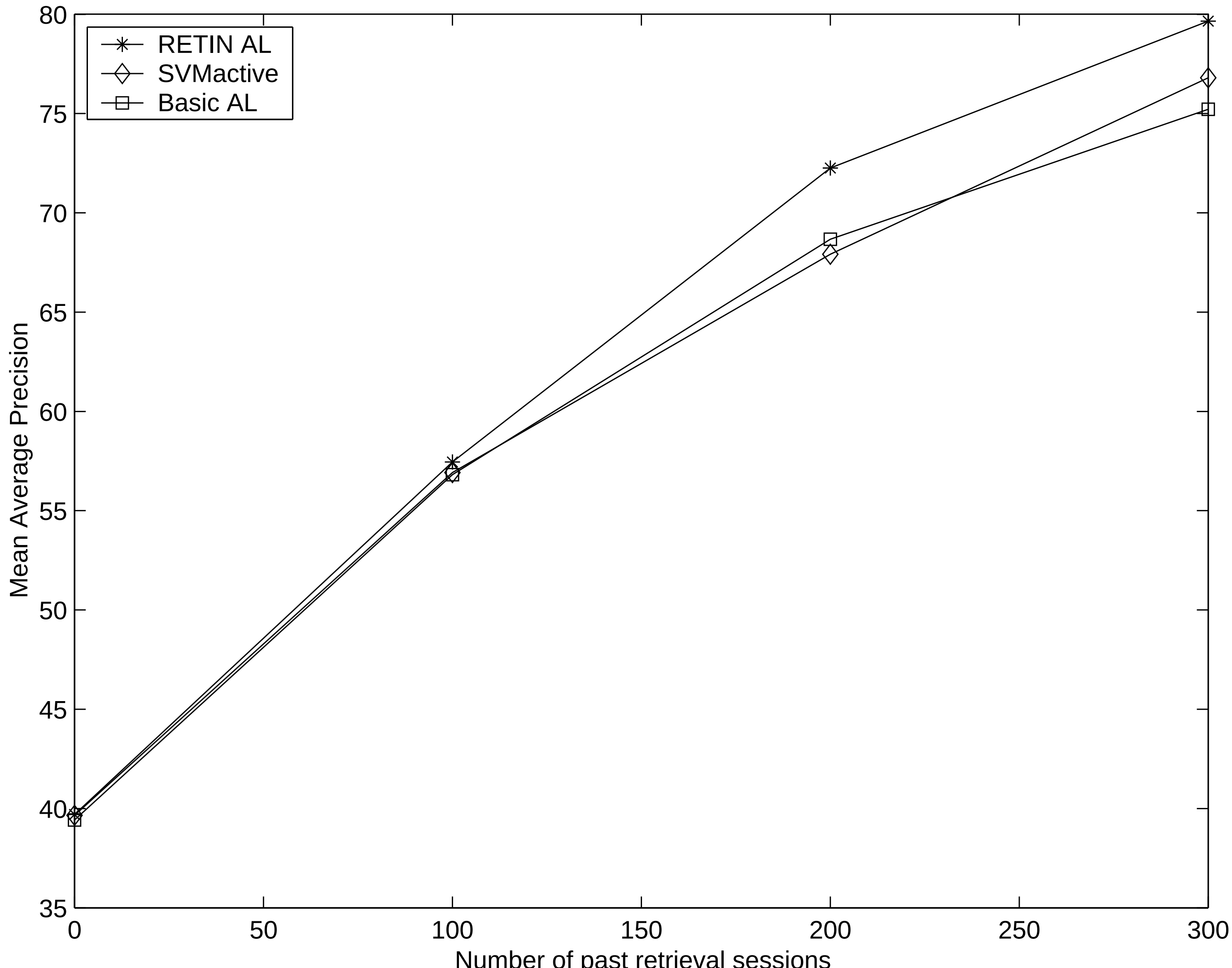
Experiments

Generalist image database:

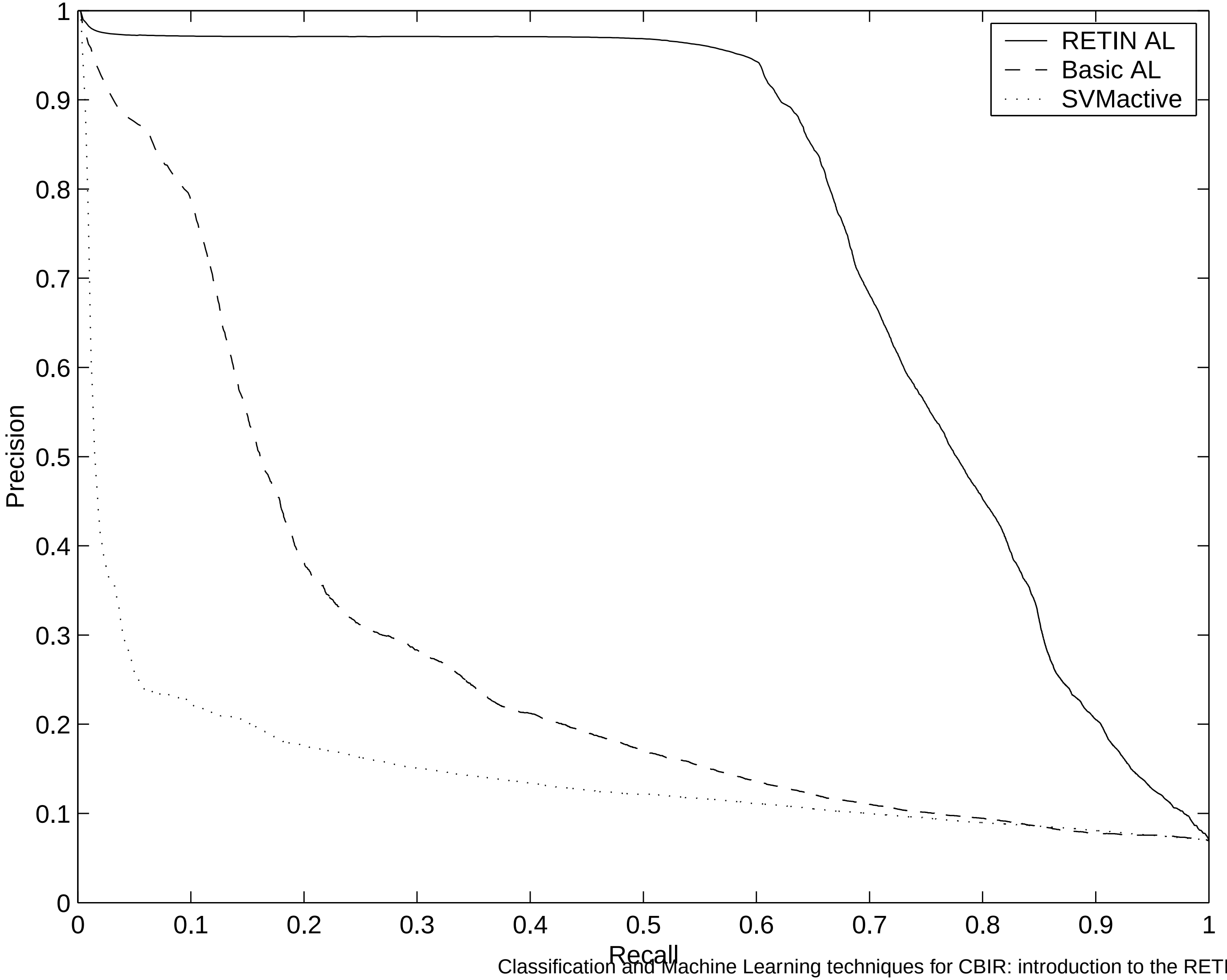
- 6,000 from COREL photo database;
- Features: $L^\star a^\star b^\star$ and Gabor filters;
- 50 mixed categories, with size from 50 to 300 images, from simple (monomodal) to complex (multimodal);
  Learning set:
- Vectors $\mathbf{y}$ with 100 non-zero values;
- From 0 to 300 vectors $\mathbf{y}$.

Experiments

Mean performance for each active learner:

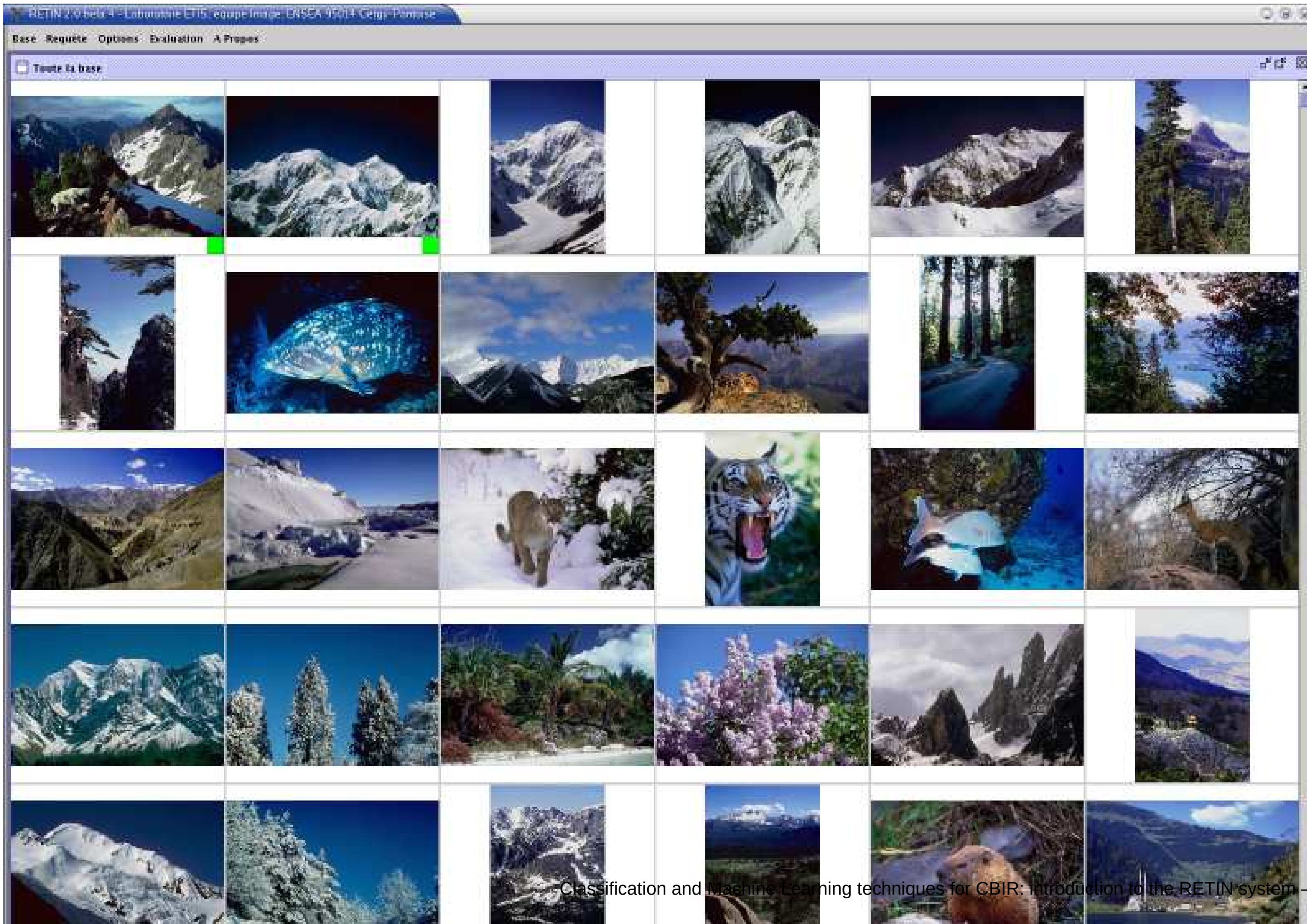


Precision/Recall curve for the 'savana' category:



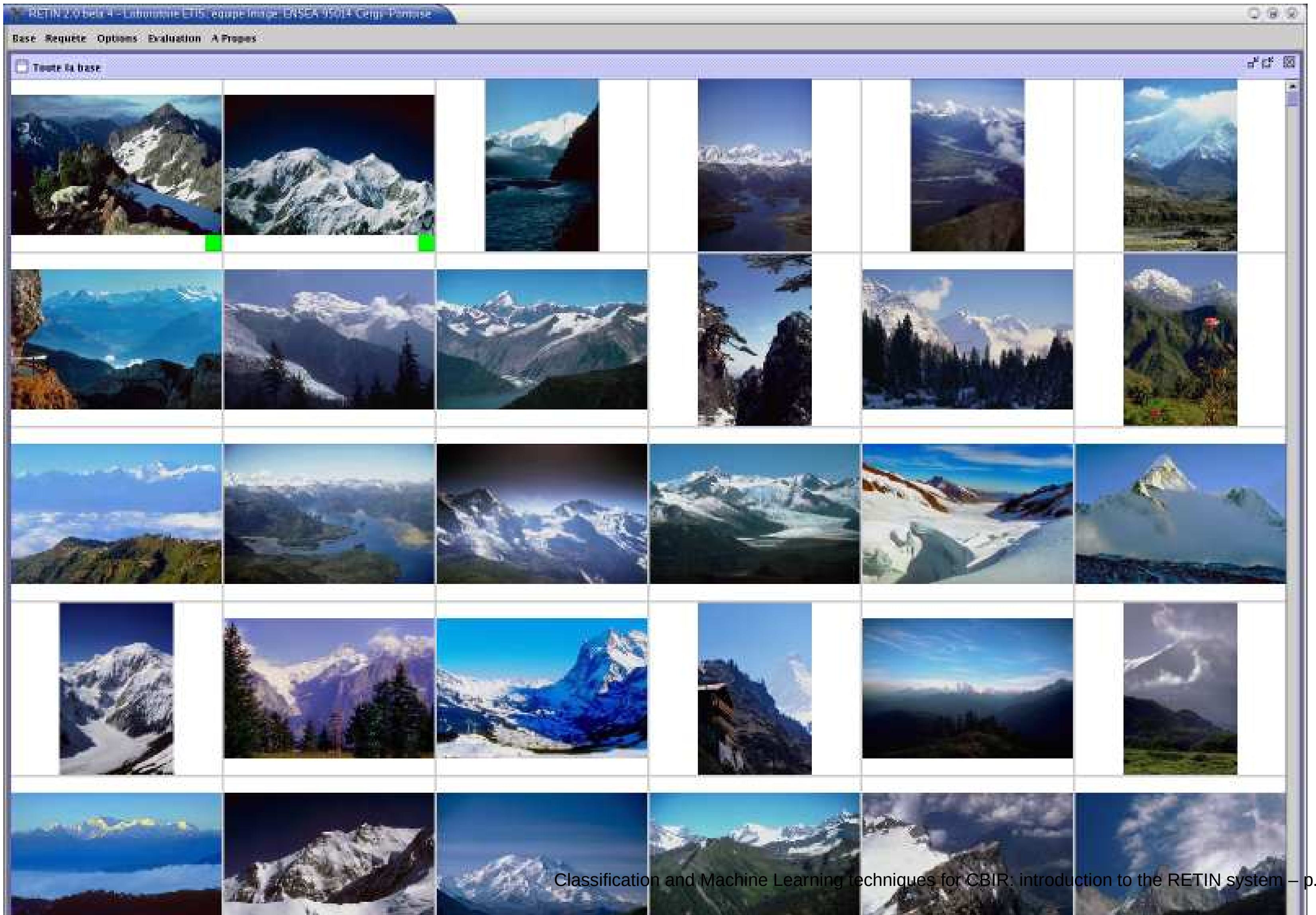
Experiments

Example : before optimisation:



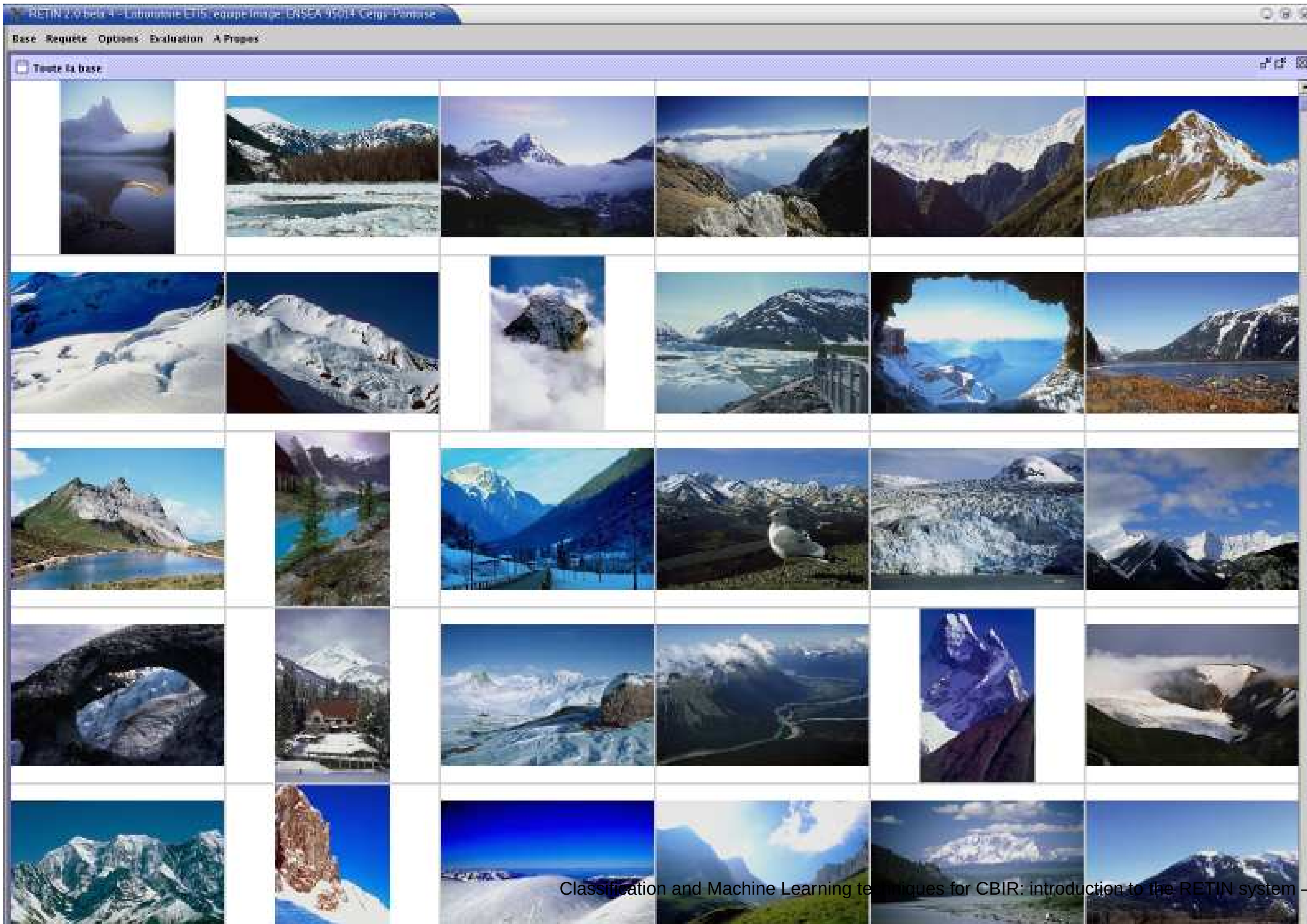
Experiments

After optimisation:

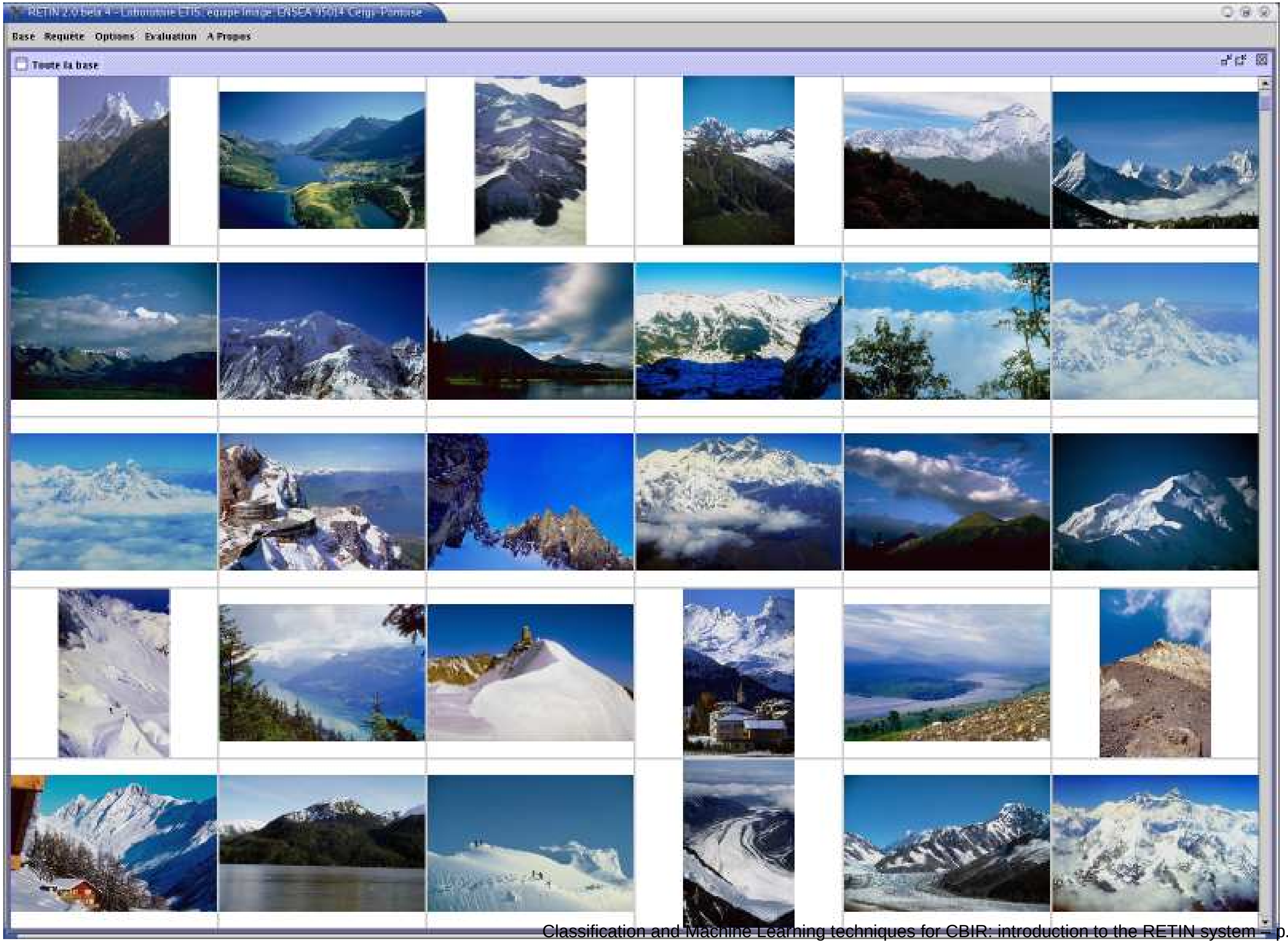


Experiments

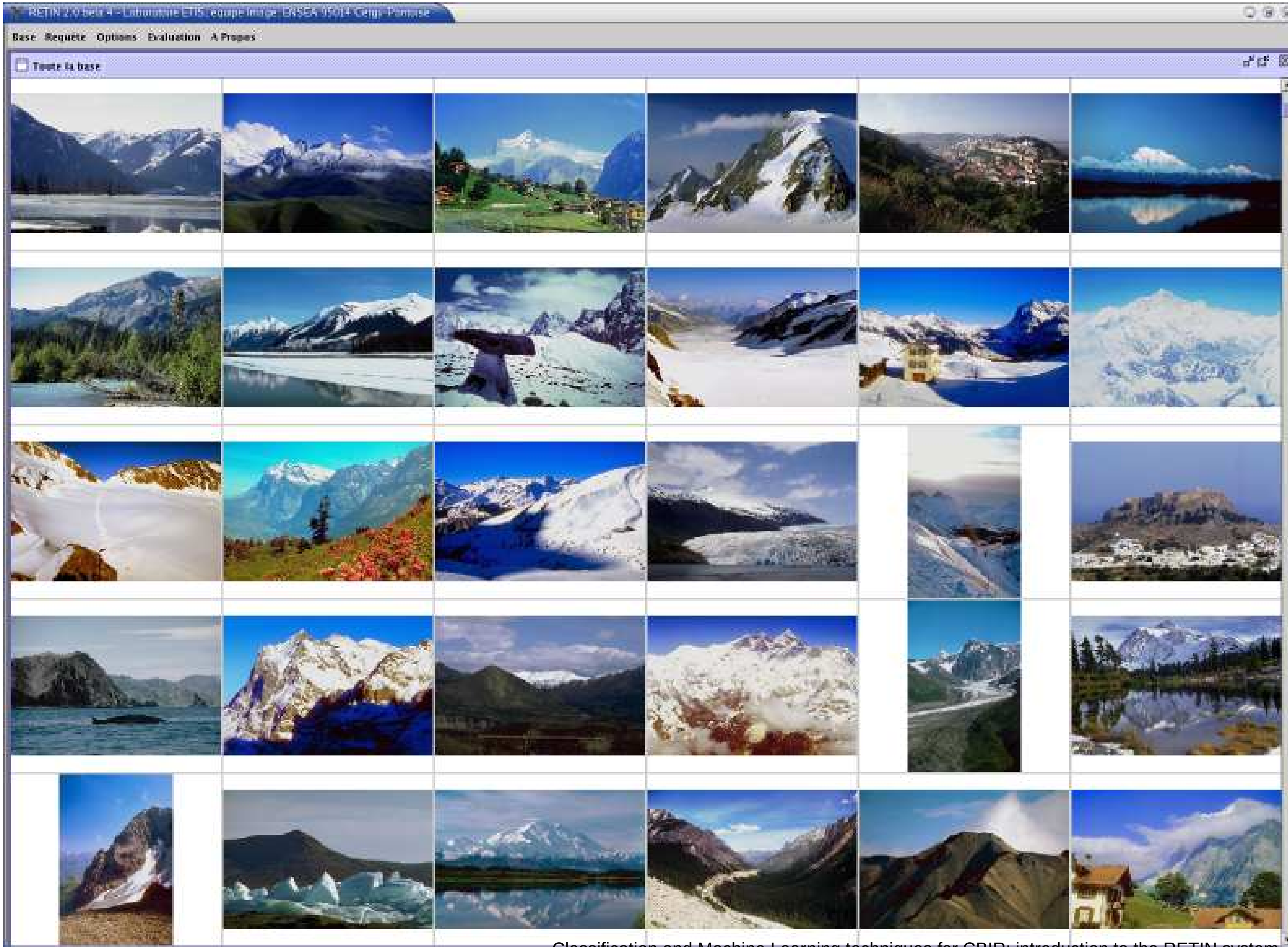
After optimisation (second screen):



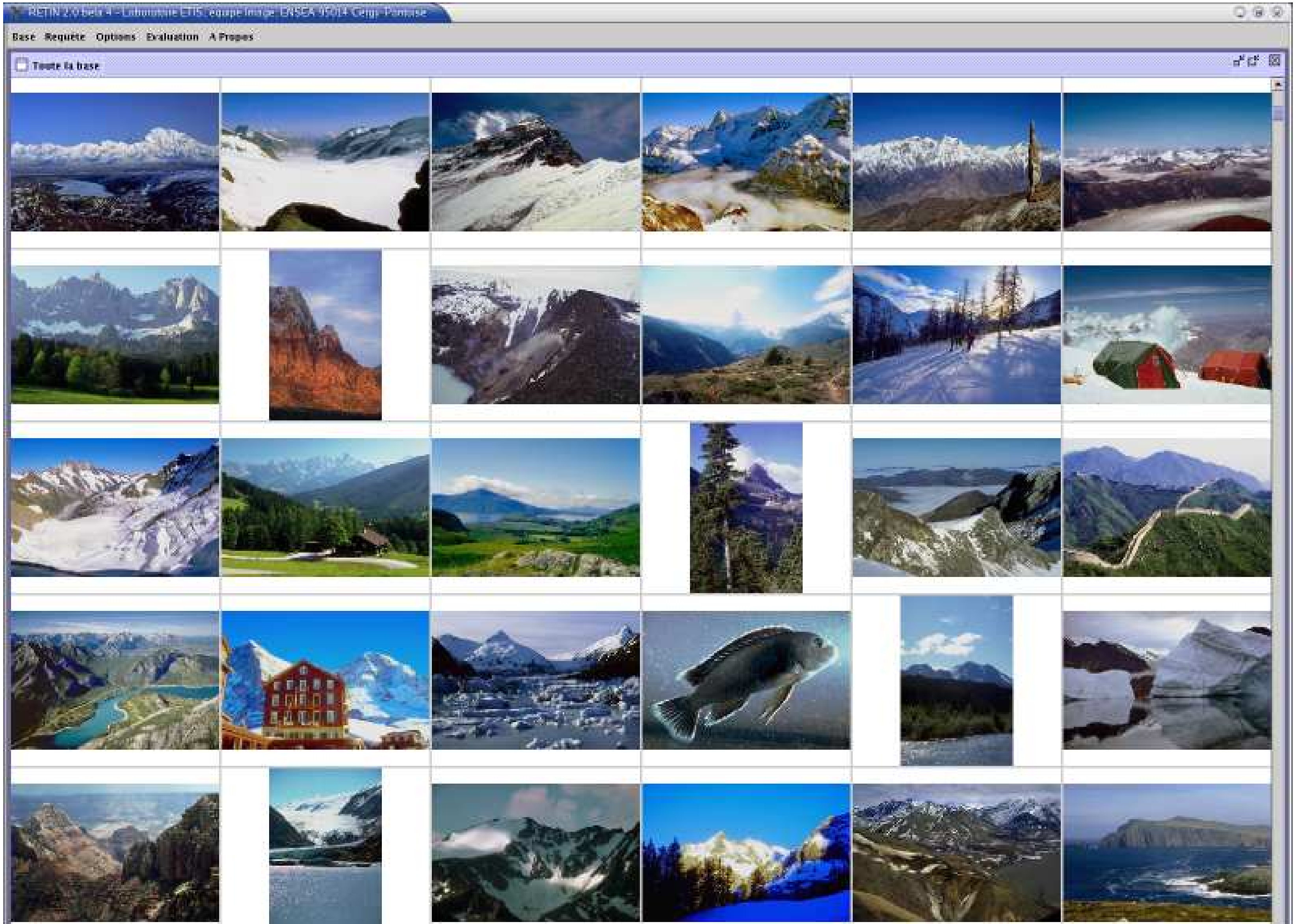
Experiments



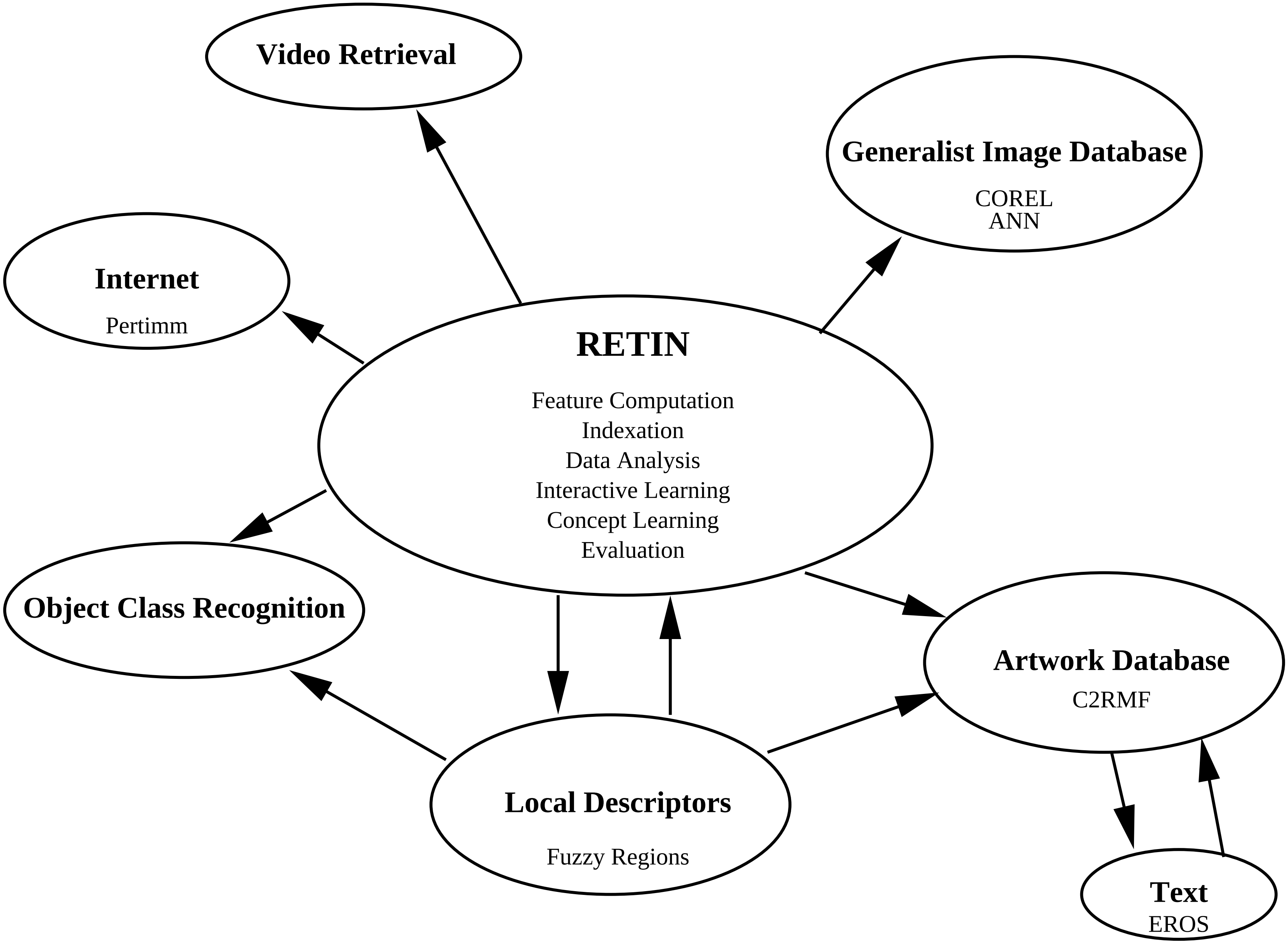
Experiments



Experiments



ETIS Lab.


Video Retrieval
Generalist Image Database
COREL
ANN
Internet
Pertimm
RETIN
Feature Computation
Indexation
Data Analysis
Interactive Learning
Concept Learning
Evaluation
Object Class Recognition
Artwork Database
C2RMF
Local Descriptors
Fuzzy Regions
Text
EROS

ETIS
CNRS UMR 8051 (ENSEA / Université de Cergy)
http://www-etis.ensea.fr/

Matthieu Cord (cord@ensea.fr), Philippe-Henri Gosselin, Sylvie Philipp Foliguet
http://perso-etis.ensea.fr/˜cord/

RETIN demo : http://dupont.ensea.fr/ ruven/start.php