# Indexation sémantique des images et des vidéos par apprentissage actif

**Bahjat Safadi**

Soutenance de Thèse (Universités de Grenoble)

**Jury**:

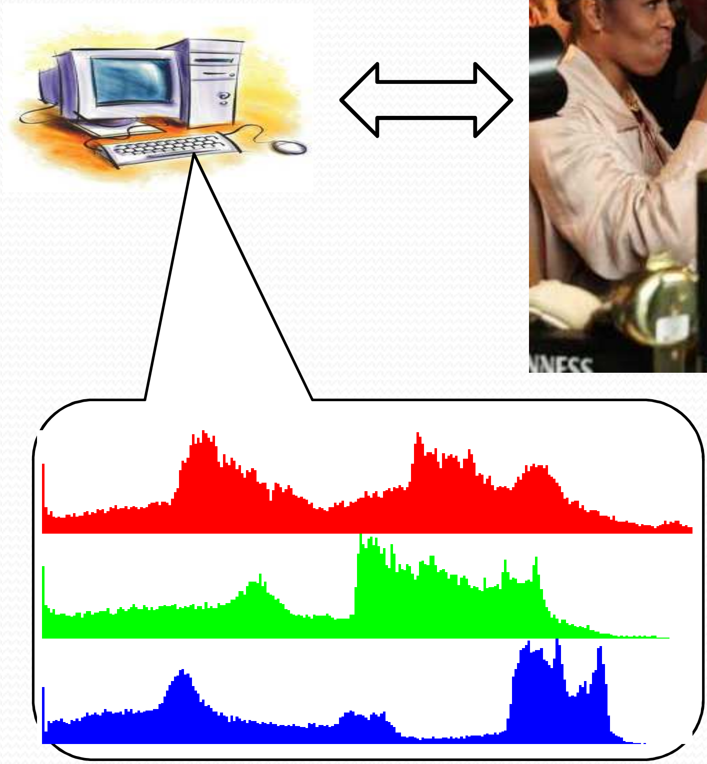
M. Stéphane Ayache, Université de la Méditerranée , Examinateur
M. Matthieu Cord, UPMC Sorbonne Universités, Rapporteur
M. Hervé Jégou, INRIA- Rennes, Examinateur
M. Denis Pellerin, Université Joseph Fourier, Président
M. Georges Quénot, CNRS, Directeur de thèse
M. Alan Smeaton, Dublin City University, Rapporteur

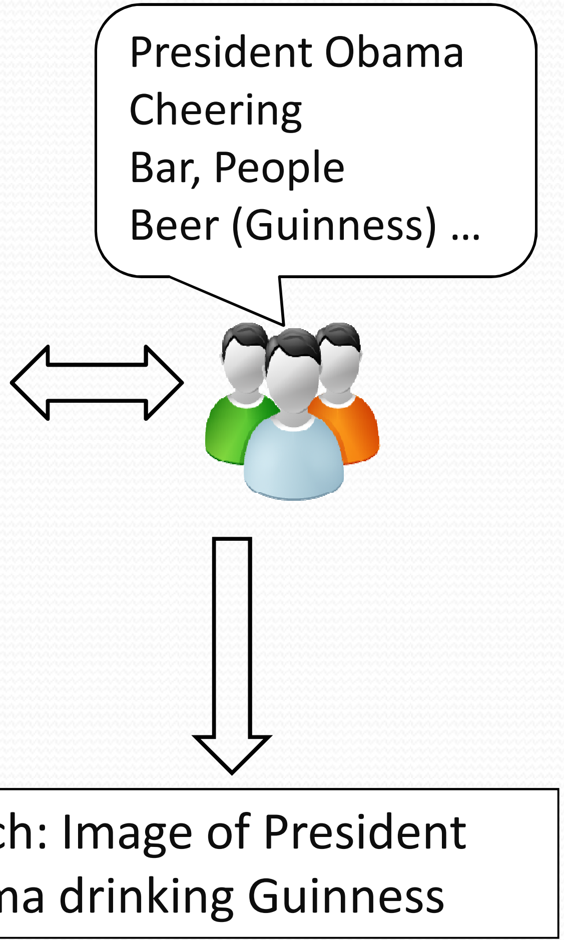
# Context

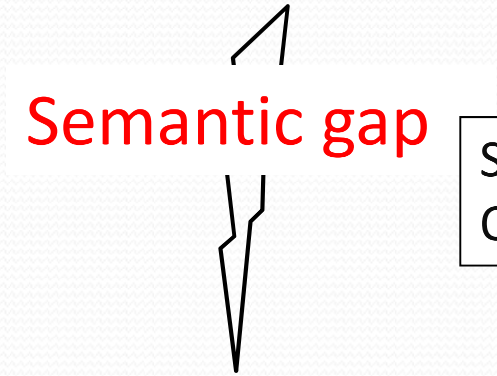
- Content-Base Information Retrieval (CBIR) in large multimedia collections

# Semantic gap

- (Smeulders et al [2000])



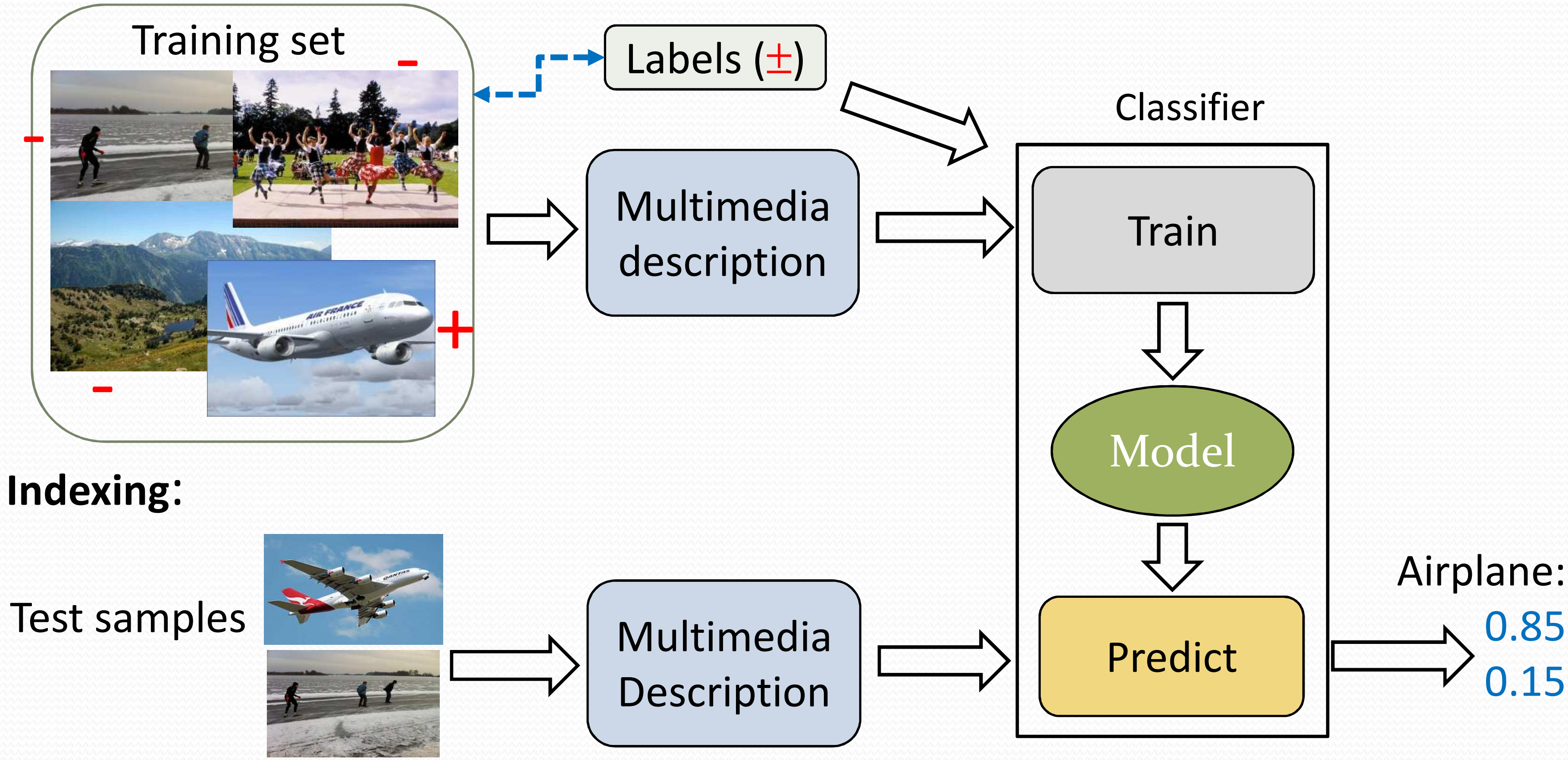


# Content-based Multimedia Indexing

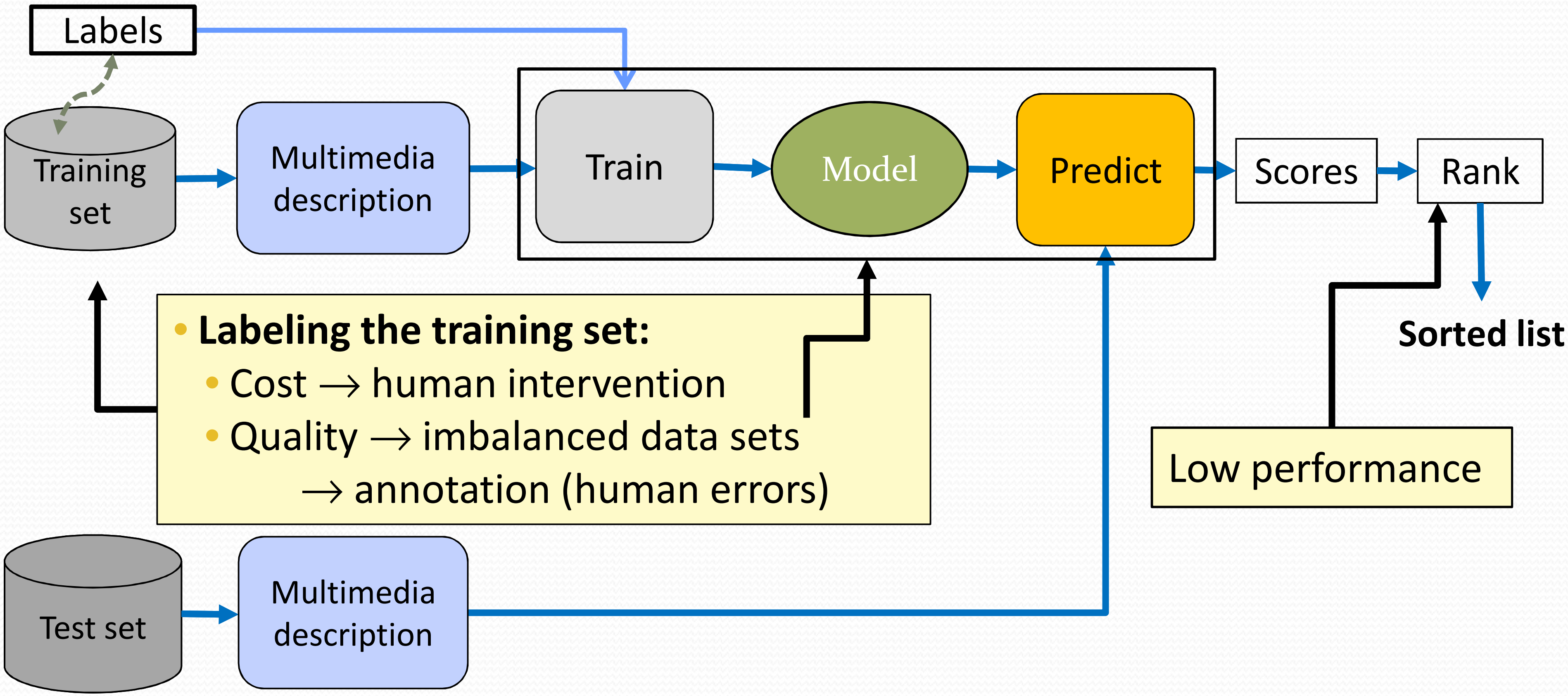
- For each concept (e.g. Airplane)

**Modeling**:

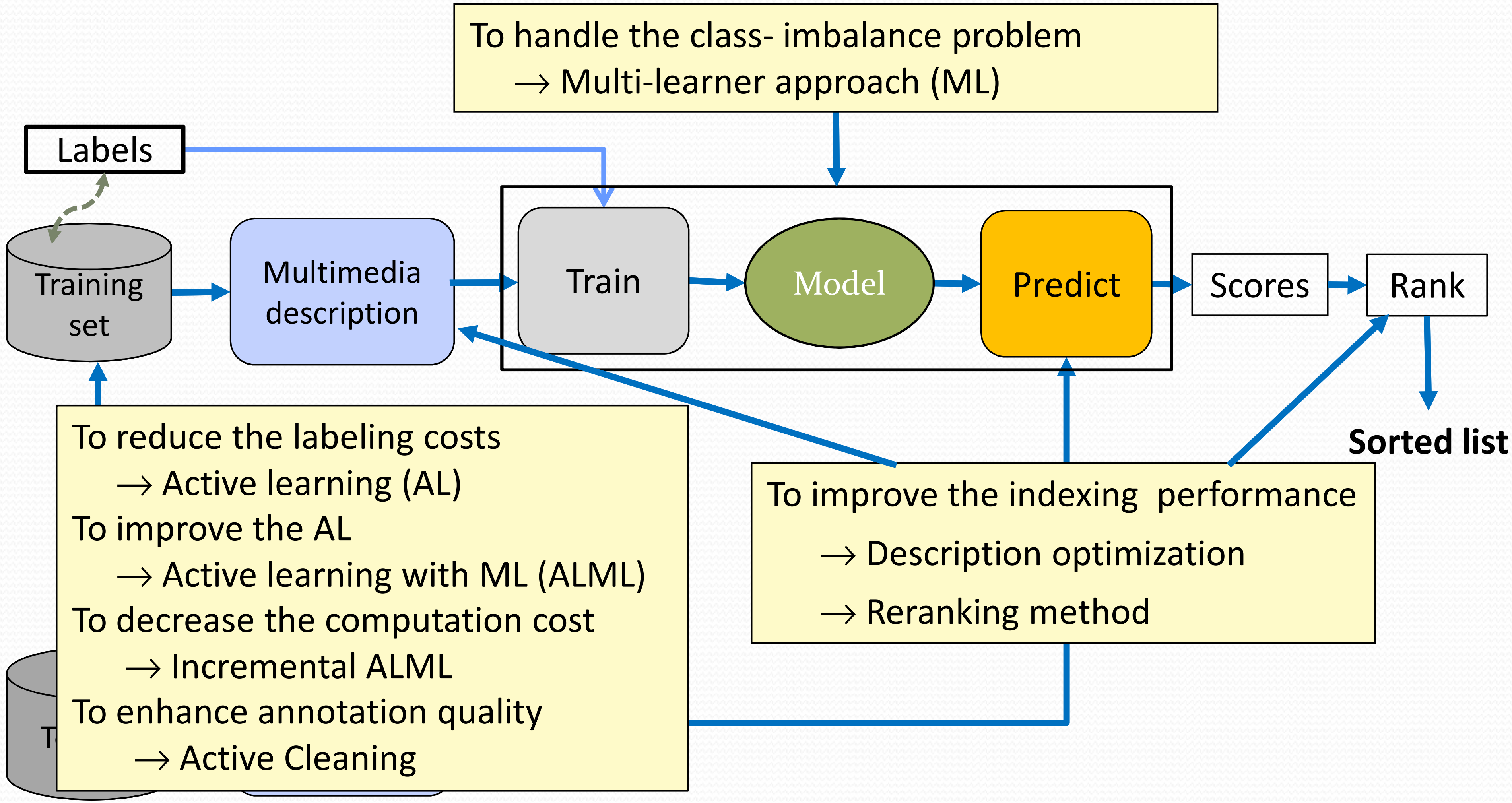


# Content-based Multimedia Indexing

- Major problems



# Proposals



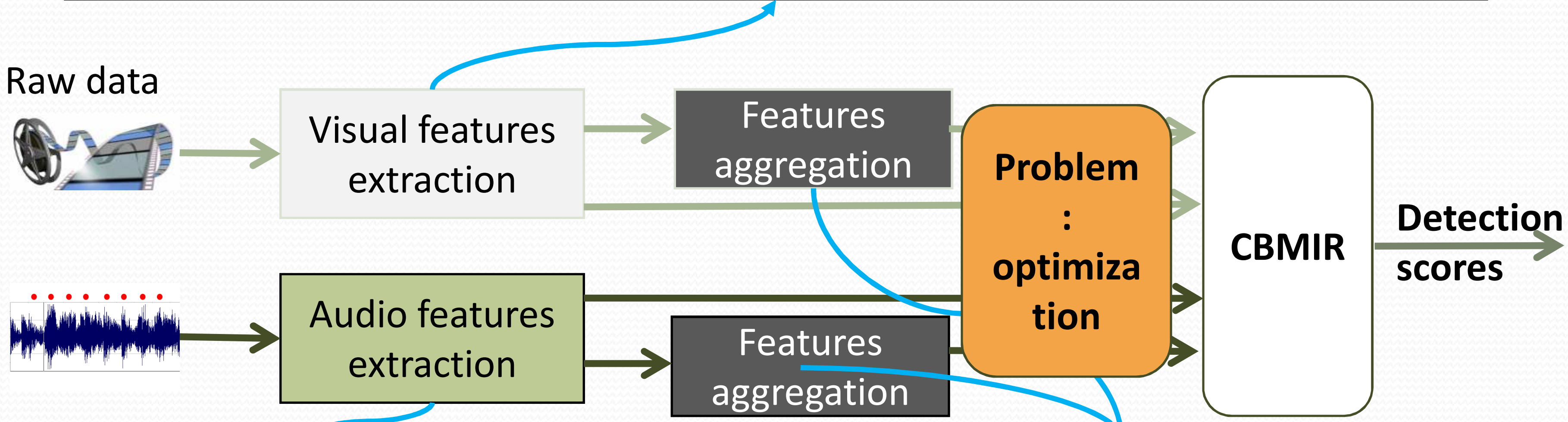
To handle the class- imbalance problem
→ Multi-learner approach (ML)
Labels
Training set
Multimedia description
Train
Model
Predict
Scores
Rank
Sorted list
To reduce the labeling costs
→ Active learning (AL)
To improve the AL
→ Active learning with ML (ALML)
To decrease the computation cost
→ Incremental ALML
To enhance annotation quality
→ Active Cleaning
To improve the indexing performance
→ Description optimization
→ Reranking method

# Outline

1. Introduction
2. State of the art
3. Proposals
4. Experiments
5. Conclusions and perspectives

# 2. Multimedia Content Description

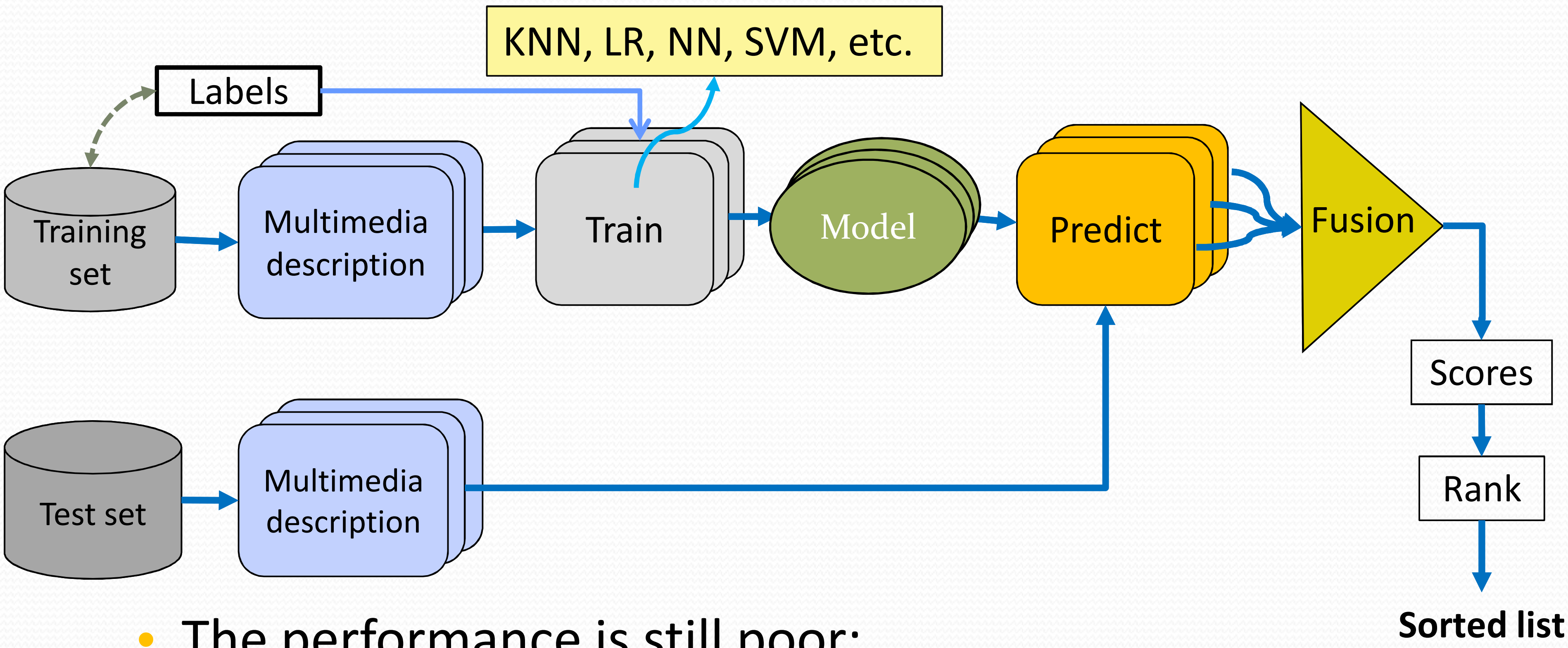
- **Global Features: Color**: Histogram, Dominant Color, Moments, **Texture**: Gabor transforms, **Shape**: Contour Shape, Region Shape, 2D/3D shape, **Motion**: Camera motion, Global motion, Parametric
- **Local Features**: SIFT (Lowe[1999]), STIP(Laptev[2003]), SURF (Bay et al.[2006])




- Spectral coefficients: MFCCs
- Temporal coefficients: Volume

- Bag of visual words (Sivic & Zisserman [2003], Csurka et al [2004])
- Fisher Kernel (Perronnin & Dance [2007])

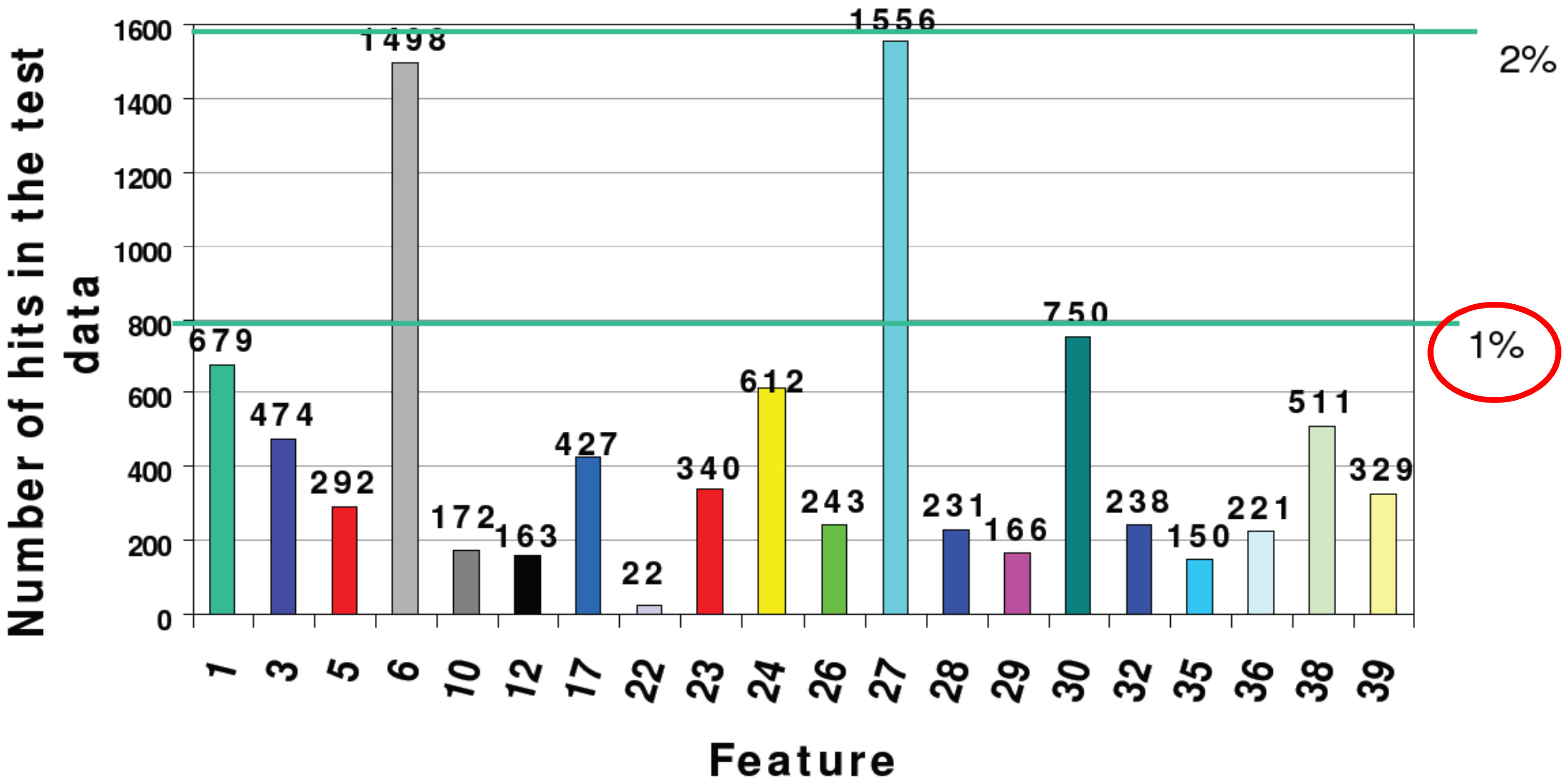
# 2. Generic Content-Based Indexing (CBI) Systems




- The performance is still poor:
  Mean Average Precision < 0.2 (Over et al. [2011])

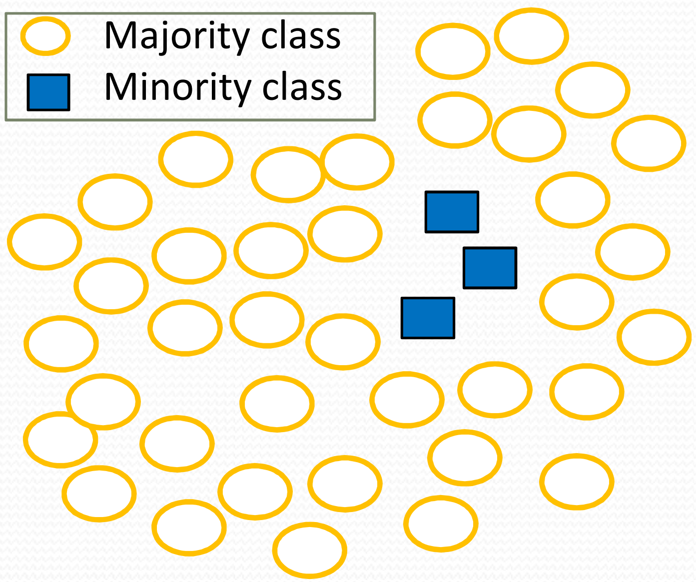
# 2. Class Imbalance

Concept frequency (Smeaton et al. [2006])



# 2. Class Imbalance

- **At the data level:** (Re-sampling)
  - Over- and Under-sampling
  - Active sampling
- **At the algorithmic level:**
  - Adjusting the costs of the classes

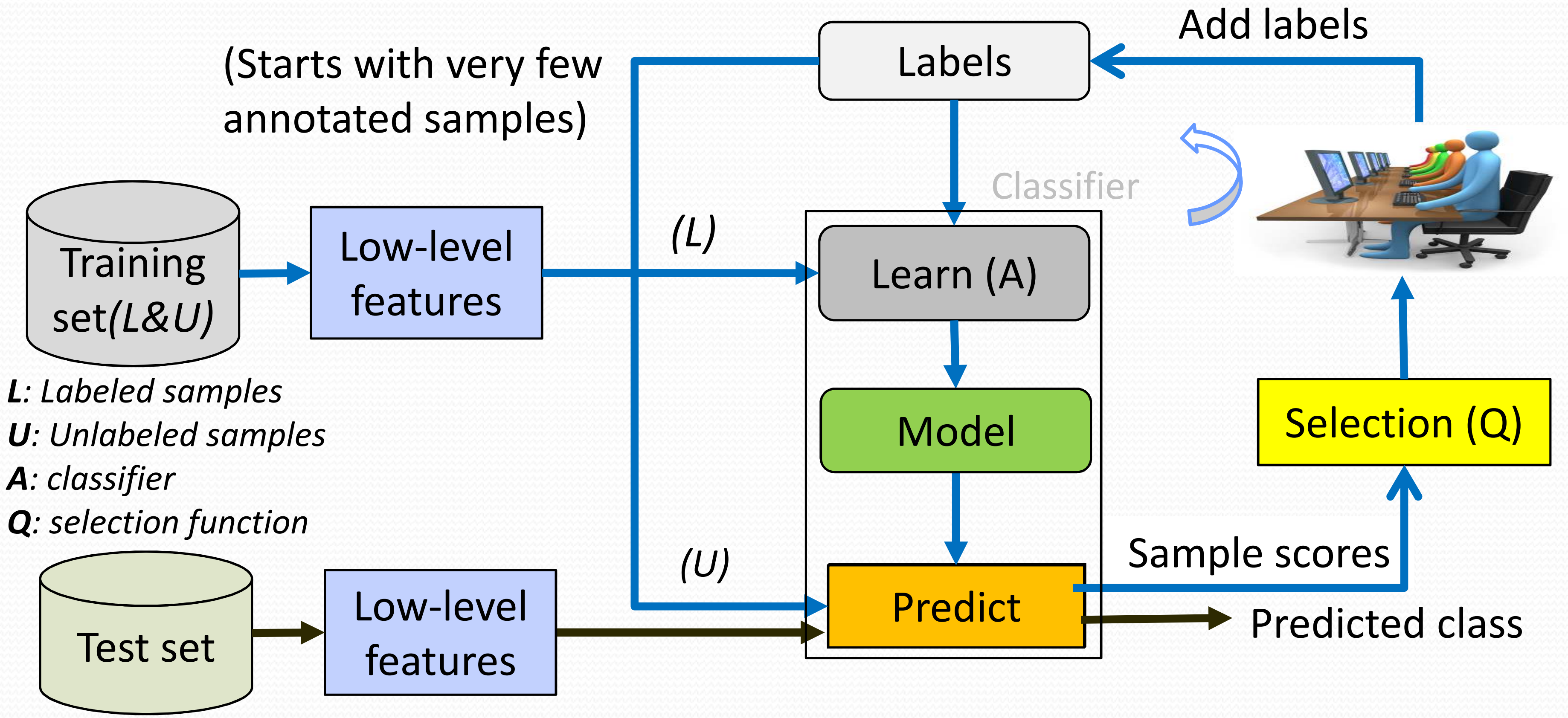



- **Methods:**
  - Random Under Sampling (RUS) (Bishop [2007])
  - Ensemble learning **(**Breiman [1996]**)**
  - Inverse Random Under Sampling (IRUS) (Tahir et al.[2009])

# 2. Active Learning

- Active learning is an approach in which an existing system is used to predict the usefulness of new samples *(Ghahramani and Cohn [1994])*
- **Objective**: select as few samples as possible to be manually labeled while getting a maximum increase of the classification performance
- Several strategies can be considered to predict samples' usefulness
  - Relevance sampling (Tong and Koller [2000])
  - Uncertainty sampling (Lewis and Catlett [1994])
  - Partition sampling (Souvanavong [2004])

# 2. Automatic Indexing System Based Active Learning




**AL = < L, U, A, Q >** *(Ghahramani & Cohn [1994])*

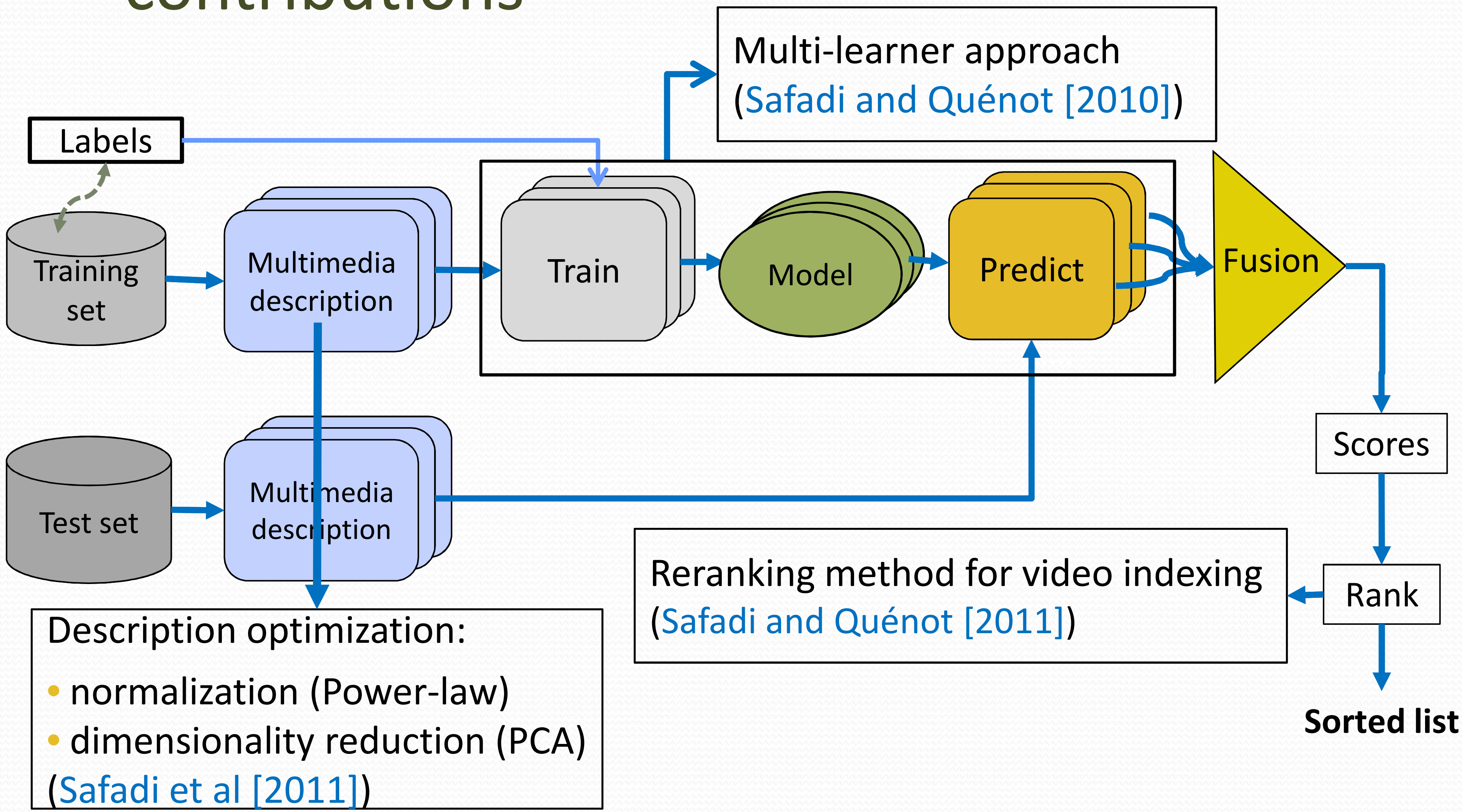
# 2. Summary

| | |
|---|---|
| **CBMI** | bridges the semantic gap<br>audio and visual descriptors + classification phase<br>combine several descriptors and classifiers<br>$\longrightarrow$ can be optimized |
| **Class-imbalance** | several methods<br>$\longrightarrow$ they are still not optimal |
| **Active learning** | has an impact on the imbalance problem<br>minimizes the labeling costs (human intervention)<br>may produces low quality annotations (human errors)<br>$\longrightarrow$ needs to be improved |

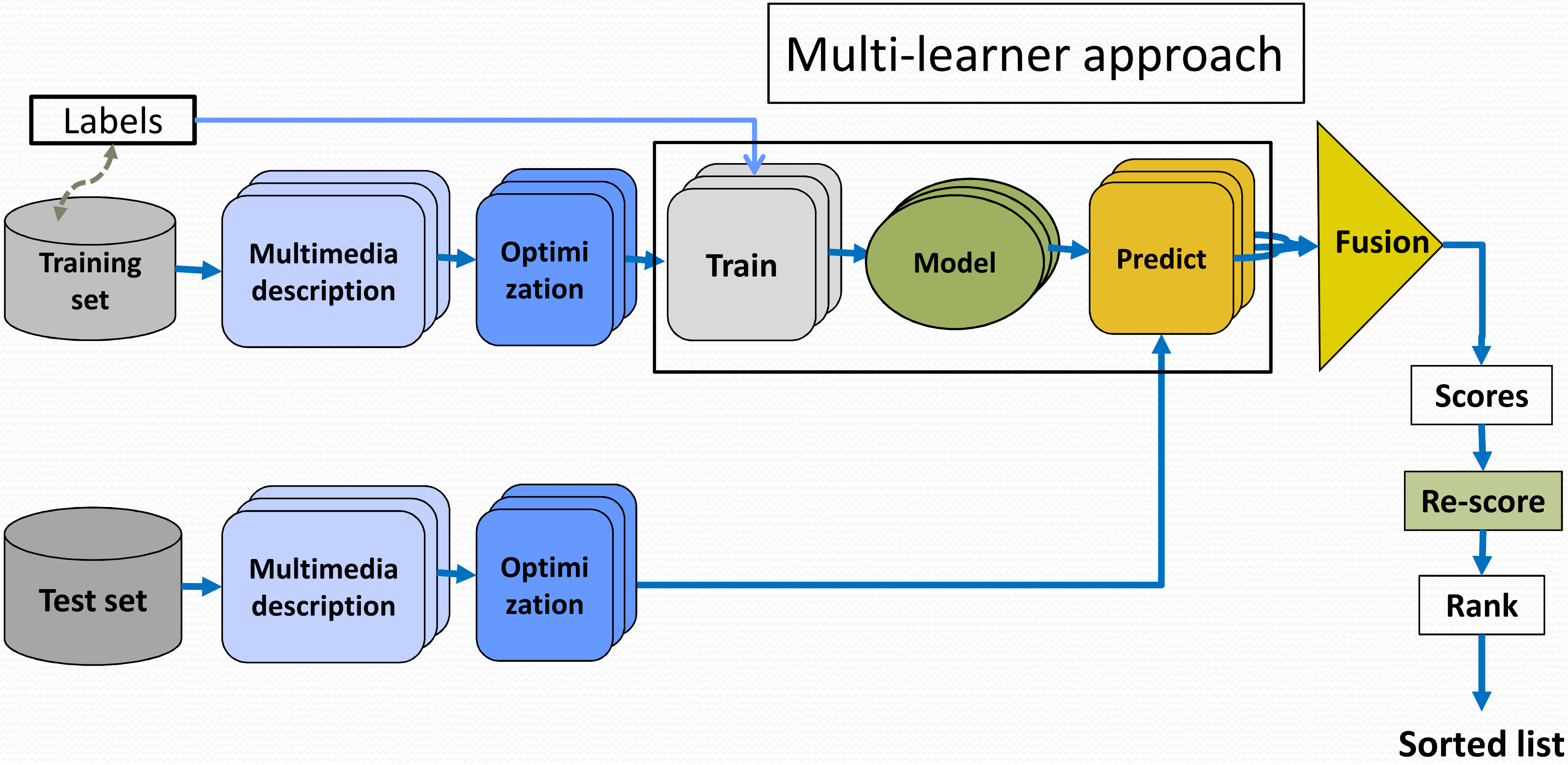
# Outline

1. Introduction
2. State of the art
3. Proposals
    1. Contribution to CBI
        1. The overall contributions
        2. The multi-learner approach
    2. Active learning methods for multimedia indexing
    3. Annotation quality
4. Experiments
5. Conclusions and perspectives

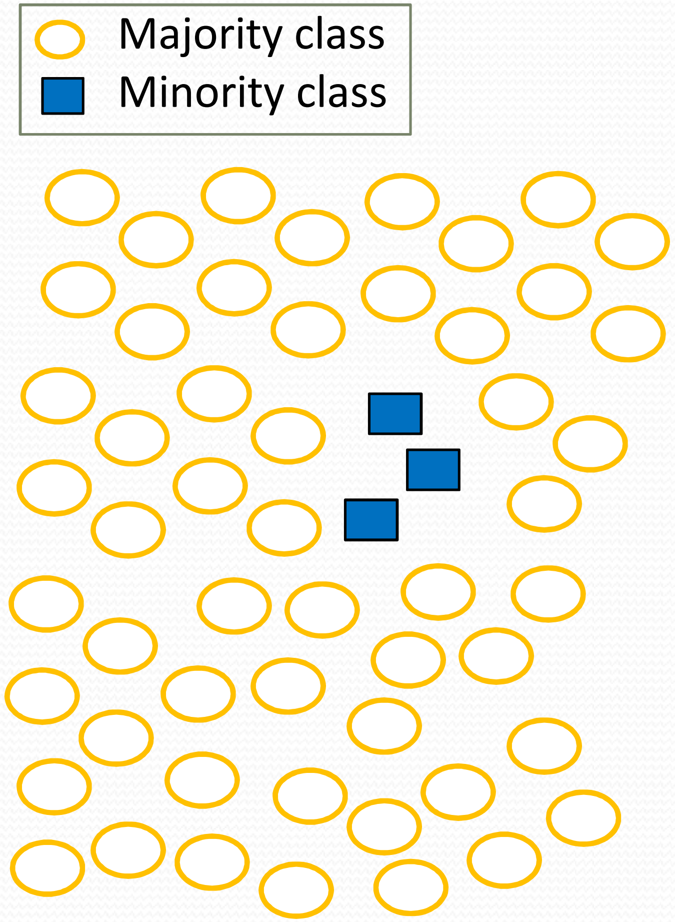
# 3.1 Contributions to CBI: The overall contributions



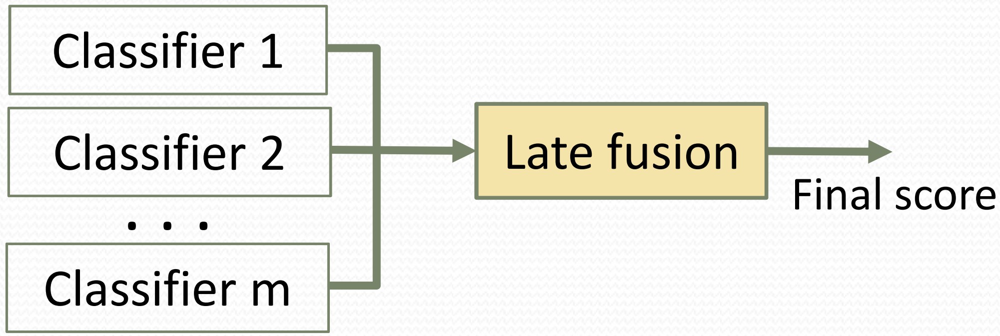
# 3.1 Contributions to CBI



# 3.1 Multi-Learner Approach for Class-Imbalance Problem




- **Multi-learner**:
  - combine the Random Under Sampling with Ensemble learning




- How do we re-sampling the majority class samples?
- How many learners do we propose to be trained? (m ????)

# 3.1 Multi-Learner Approach for Class-Imbalance Problem

Collection+ annotations

Descriptor

Defines wether to use a single- or multi- learner

The classifer + its hyper-parameters

$$\boldsymbol{ML}=\langle X, Ann, Desc, Mono, f_{maj}, f_{min}, App, App_{param}, Fu, Eval\rangle$$

The probability of each sample in the majority class to be seletcted

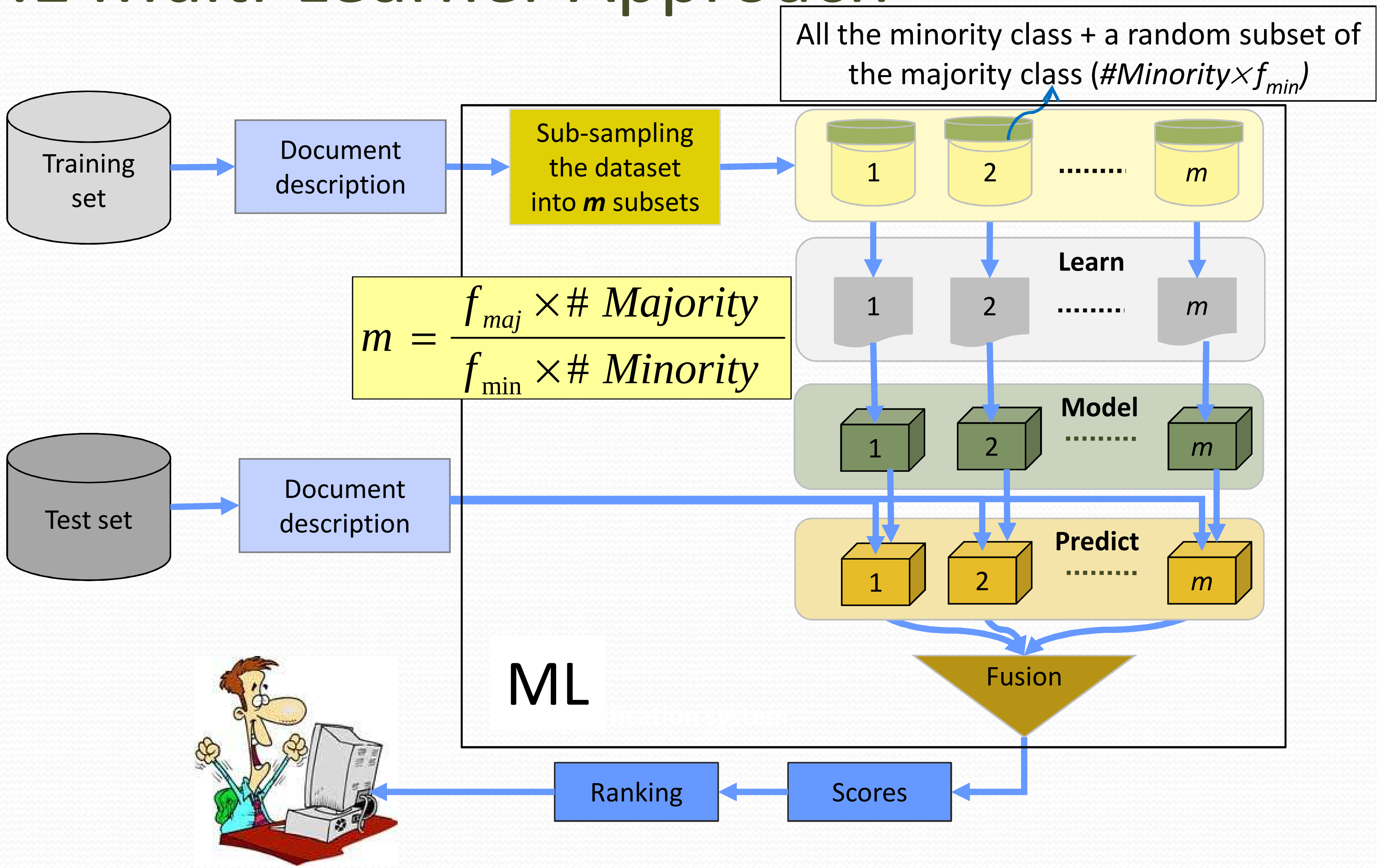
The desired ratio between the majority and the minority class

Fusion function

Evaluation metric

- $f_{maj}$ and $f_{min}$ need to be tuned

# 3.1 Multi-Learner Approach




$$m = \frac{f_{maj} \times \# \, Majority}{f_{\min} \times \# \, Minority}$$

# Outline

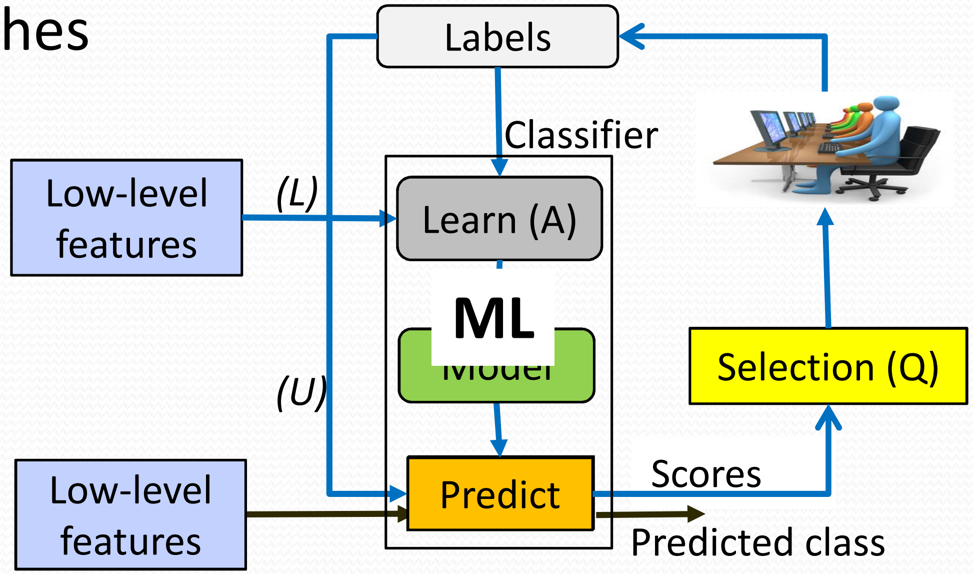
1. Introduction
2. State of the art
3. Proposals
   1. Contribution to CBI
   2. Active learning methods for multimedia indexing
   3. Annotation quality
4. Experiments
5. Conclusions and perspectives

# 3.2 Active Learning with Multiple Classifiers for Multimedia Annotation (ALML)

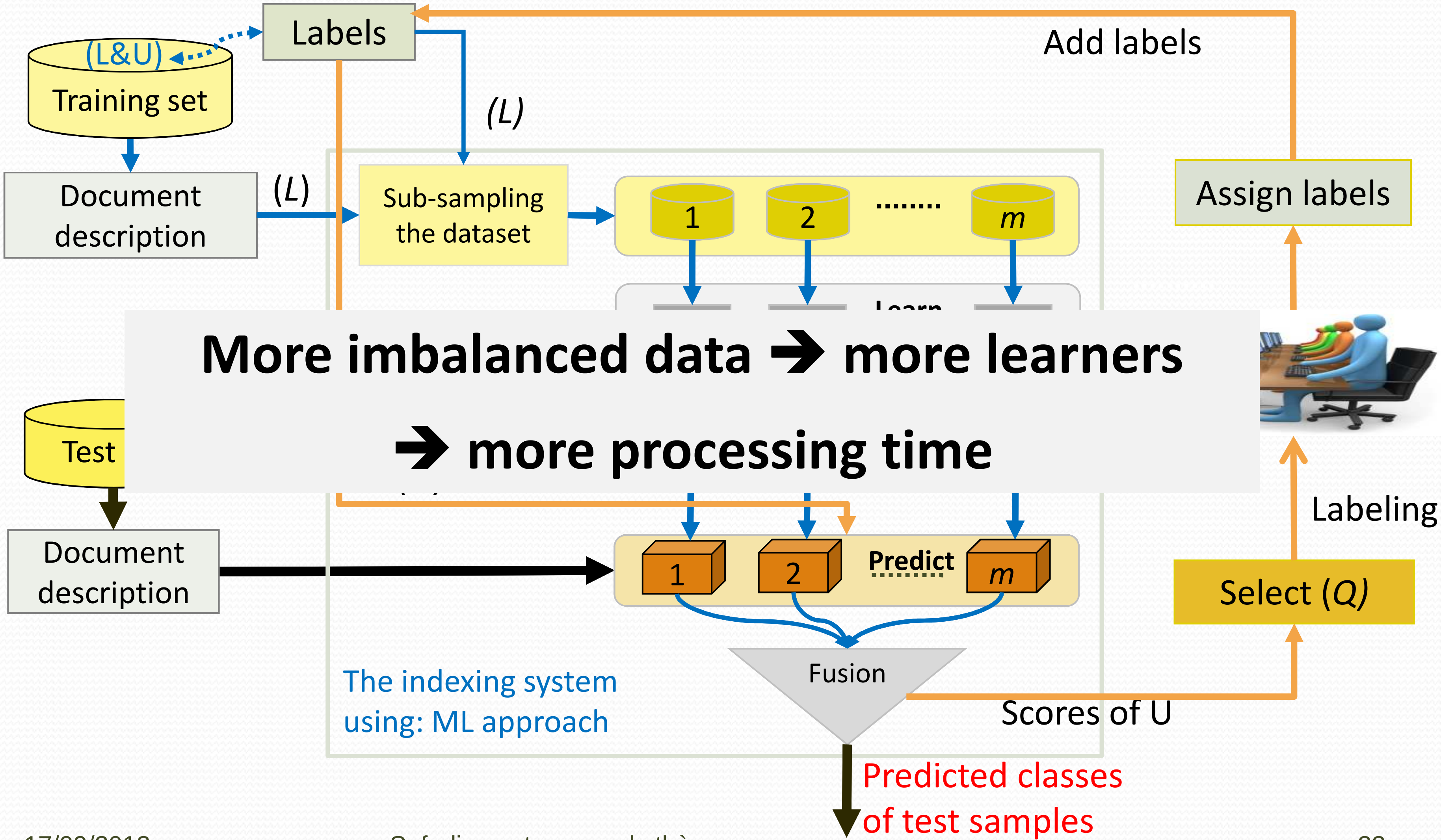
- Active learning and multi-learner approaches are two different ways of dealing with the imbalanced dataset problem
- Proposal: combine both approaches

**➔ ALML = < ML, Q >**

(Safadi and Quénot [2010])



# 3.2 ALML approach



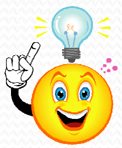

**More imbalanced data ➔ more learners**

**➔ more processing time**

# 3.2.1 Reducing the processing time

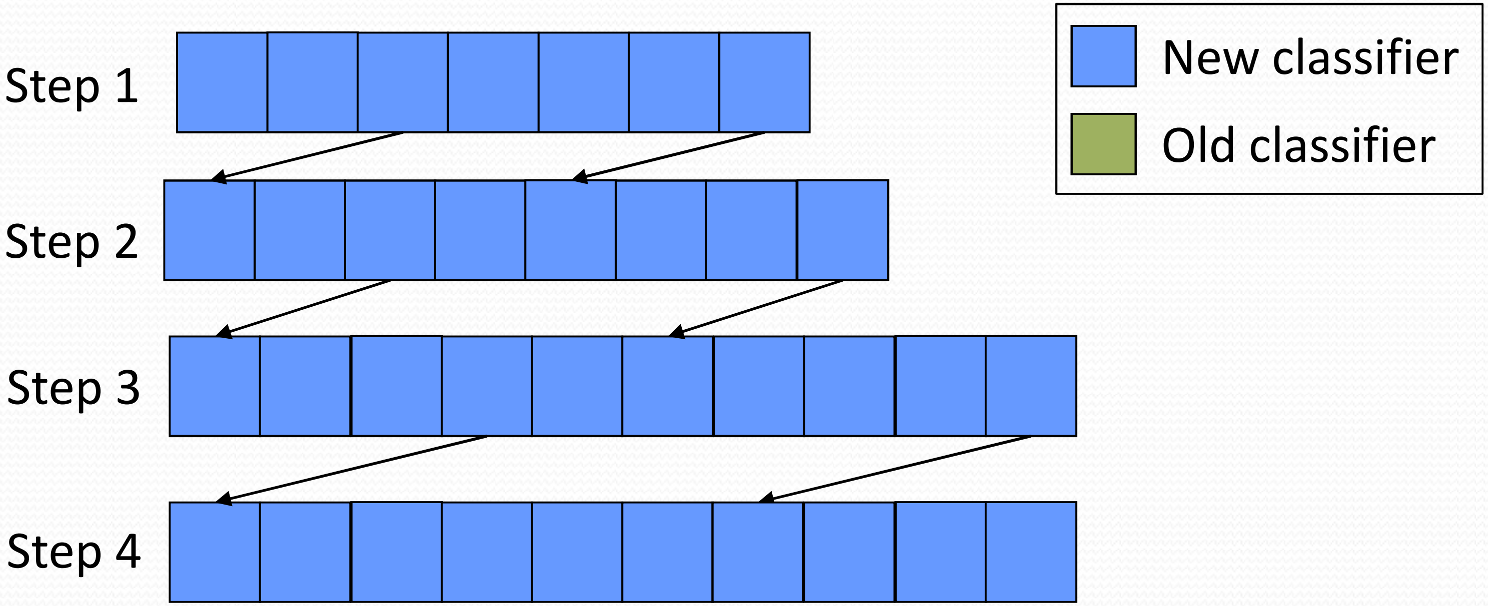
- We need to speed up the indexing system based on active learning with multi-learner (especially with MSVM)

- We may not need to train classifiers on all the subsets in the ML approach



➔ Incremental approach (Inc-ALML)
(Safadi et al [2011])

# 3.2.1 Proposed (Inc-ALML)




- By decreasing the number of learners to be learned, the processing time will be reduced
- But, how will this affect the performance of the system?
- →tuning of the refreshment rate

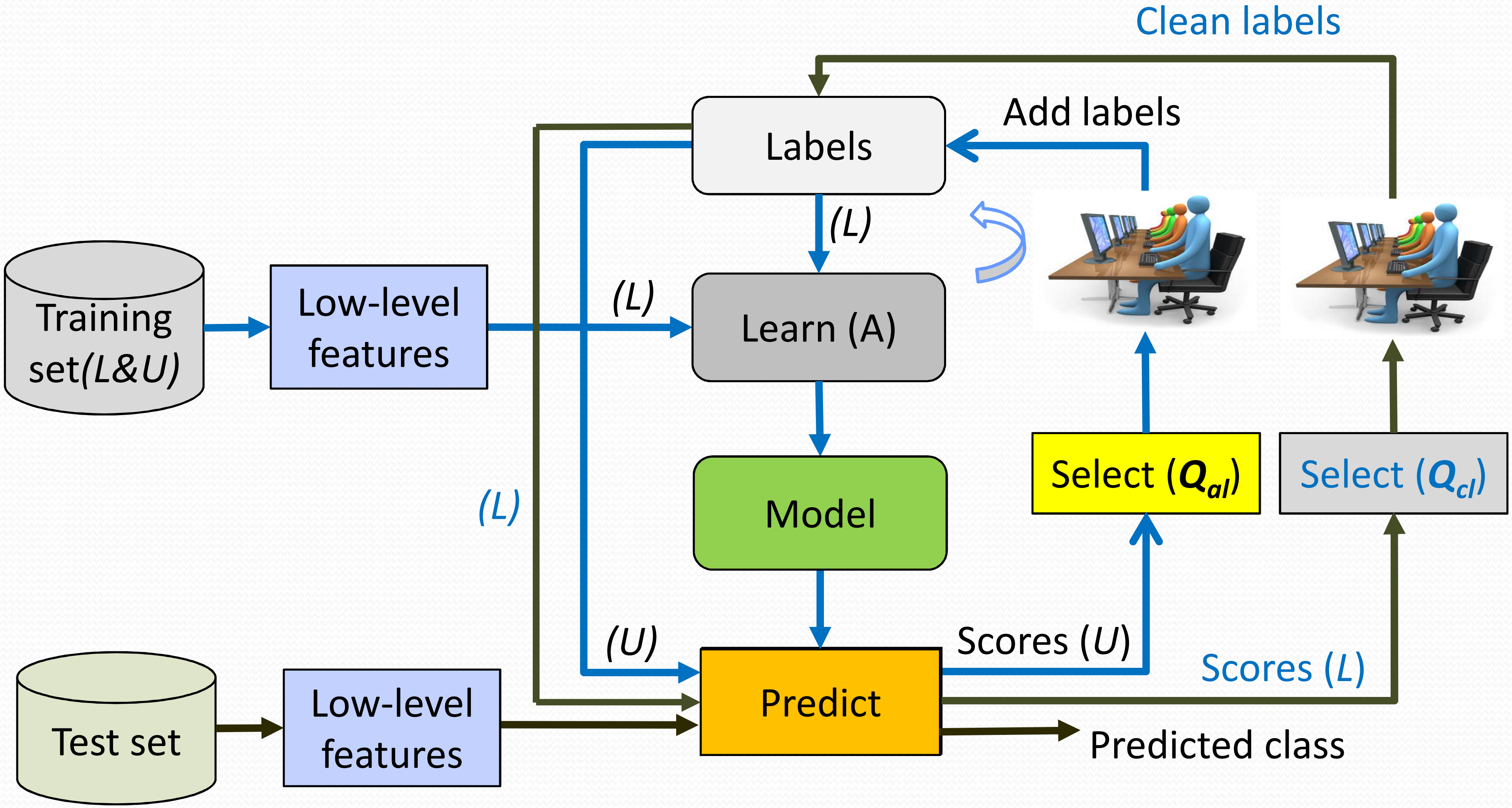
# Outline

1. Introduction
2. State of the art
3. Proposals
    1. Contribution to CBI
    2. Active learning methods for multimedia indexing
    3. Annotation quality
4. Experiments
5. Conclusions and perspectives

# 3.4 Annotation quality problem

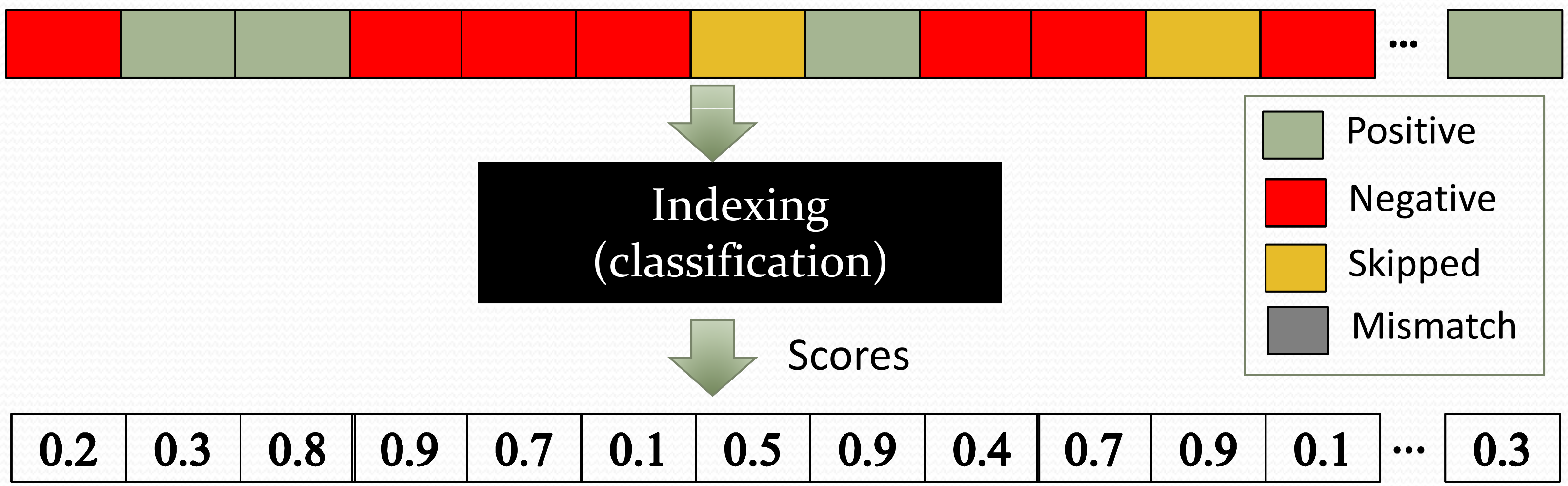
- Active learning: select the most informative sample for first annotation
- Active cleaning: select the most probably wrongly labeled samples for re-annotation (Safadi et al [2012])
- Same idea: select as few samples as possible to be manually labeled while getting a maximum increase of the classification performance

# 3.4 Active Cleaning



# 3.4 Active Cleaning Strategy $(Q_{cl})$:

- ***Cross-Val***: based on re-annotating the most probably wrongly labeled samples
  - Detecting the most probably wrongly labeled samples




  - Select fractions of these samples for a second annotation
  - Third annotation when conflicts annotations are detected

# Outline


1. Introduction
2. State of the art
3. Proposals
4. Experiments
    1. ALML and Inc-ALML
    2. Active cleaning
    3. Application to TRECVid Collaborative Annotation
    4. Participation to the TRECVid Semantic Indexing task
5. Conclusions and perspectives

# 4. Evaluation benchmark/campaign

- **Given:**
  - a set of $X$ data samples
    - Training ($X_{train}$): many hours of (*partly*) annotated videos
    - Testing ($X_{test}$): many hours of unseen videos
  - a set of $C$ semantic concepts:


Airplane

Mountain

Dancing

Skating

Single Female Person

- **Task:** (for each concept $c \in C$ )
  - detect a given concept $c$ in the $X_{test}$ samples?
  - rank $X_{test}$ based on the presence of a concept $c$.

# 4.1 Experiments (ALML)

- Evaluation on TRECVid 2008 HLF detection task (200 hrs of videos), 20 semantic concepts

| Collection | Hours | Shots |
|---|---|---|
| Development | 100 | 43616 |
| Test | 100 | 42461 |

- **Descriptors**: four descriptors from IRIM (GDR-ISIS) partners: *Color histogram &Gabor transform, Global-Tlep (color-texture combination), Global-Quaternion Wavelets, and Bow-SIFT*
- **Classifiers**: Logistic Regression and SVM with RBF kernel
- **AL Strategies**: Relevance and Uncertainty sampling
- **Baseline strategies**: Random and Linear sampling

# 4.1 Experiments (ALML)

- Cold-start: 10 positive and 20 negative samples
- Optimization on the development set, evaluation on the test set
- Goals:

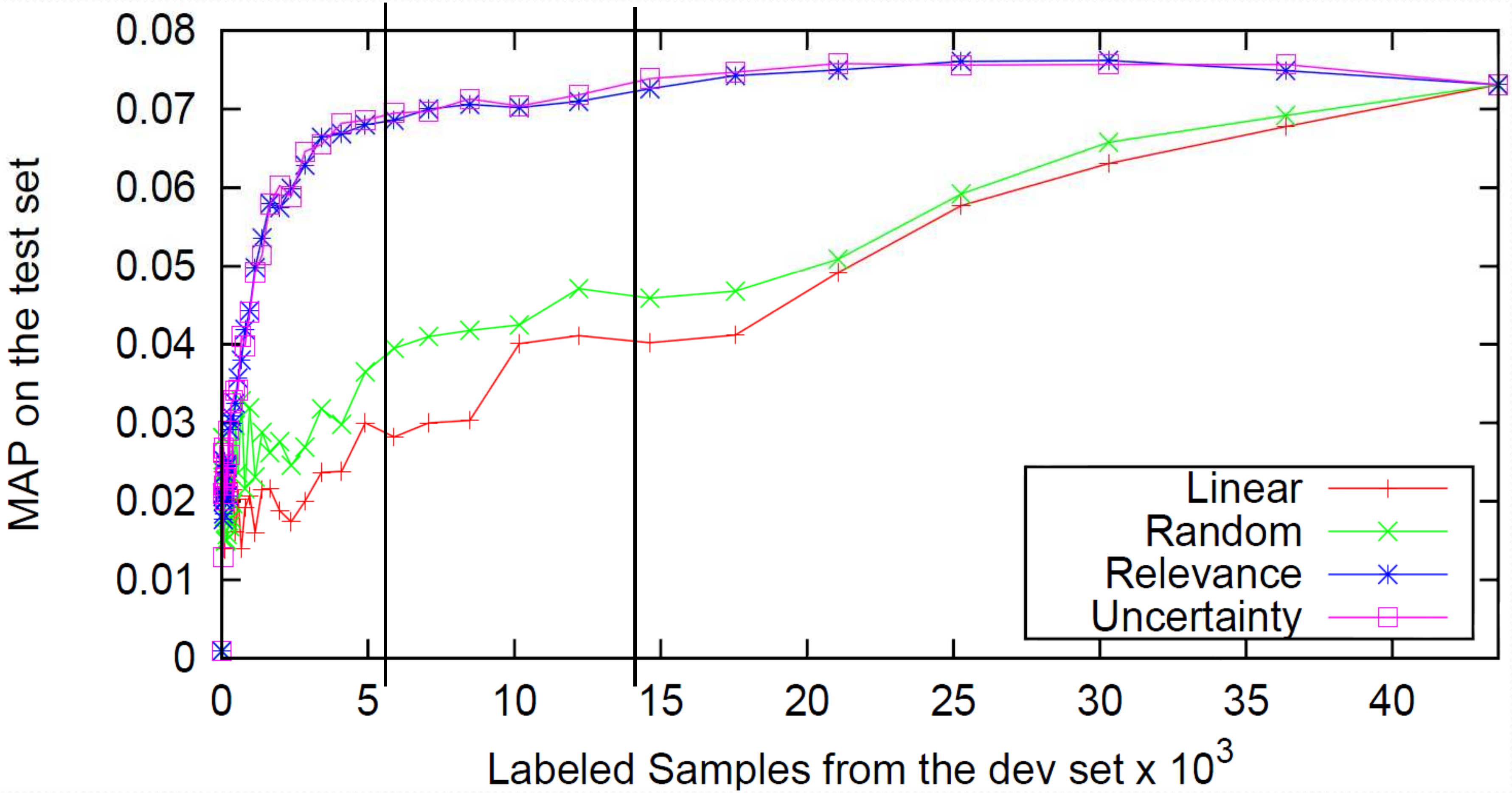
1. Comparing the performance of our method ALML with SVM-RBF (**MSVM**) with different AL strategies
2. Comparison our method **MSVM** with different learners:

| **Learner** | ***Mono*** | $f_{maj}$ | $f_{min}$ |
|---|---|---|---|
| **SSVM** | True | 1 | $\geq 1$ |
| **MLR** | False | 1 | $< 1$ |
| **MSVM** | False | 1 | $\geq 1$ |

3. Evaluation of the effectiveness of the **Inc-MSVM** approach

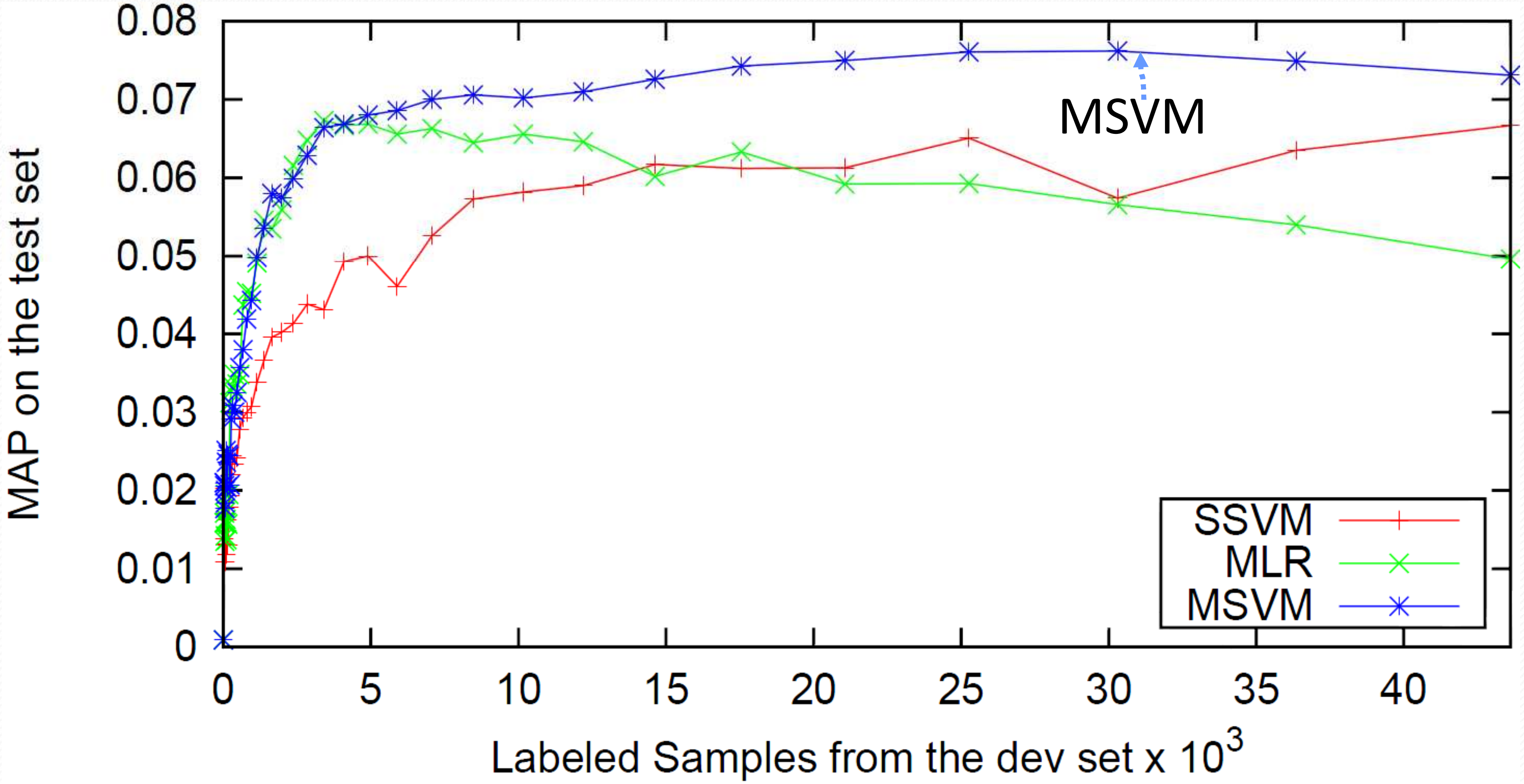
# 4.1.1 Comparison of active learning strategies

**Descriptor:** Global-Tlep; **Classifier**: MSVM



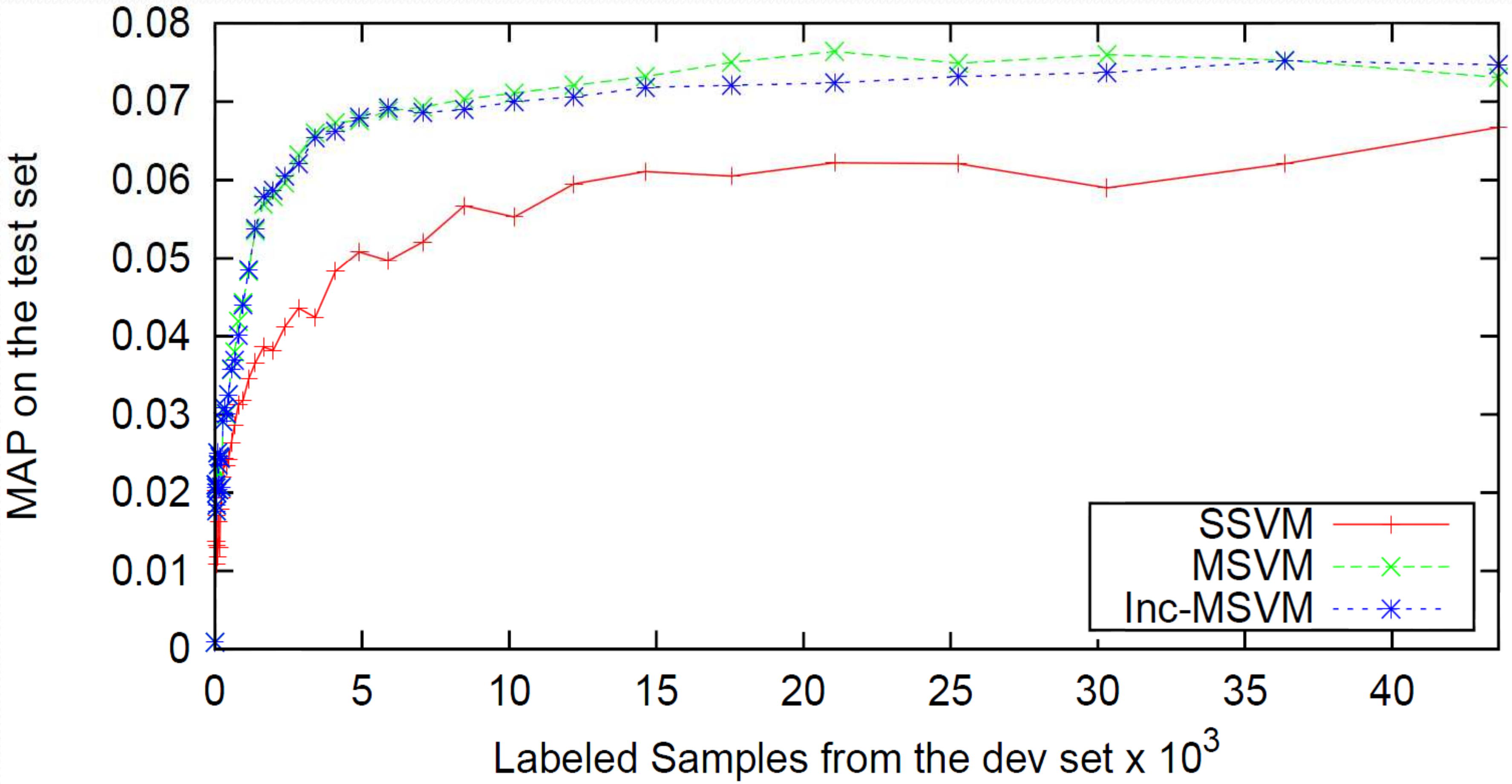
# 4.1.2 Comparison of Learners

**Descriptor:** Global-Tlep; **Classifier**: MSVM; **AL strategy**: Relevance sampling



# 4.1.3 Inc-ALML

**Descriptor:** Global-Tlep; **Classifier**: MSVM; **AL strategy**: Relevance sampling



# 4.1.3 Inc-ALML

The processing times (in hours) on TRECVid 2008

| Descriptor | dims | SSVM | MSVM | Inc-MSVM | Gain |
|---|---|---|---|---|---|
| LIG/Hg104 | 104 | 4.80 | 45.45 | 23.23 | 60% |
| CEALIST/global_tlep | 756 | 96.56 | 395.45 | 204.9 | 48% |
| ETIS/global_qwm | 768 | 46.17 | 460.57 | 212.3 | 54% |
| LEAR/bow_sift_1000 | 1000 | 181.0 | 592.10 | 300.6 | 49% |

Inc-ALML algorithm can achieve almost the same performance as ALML, with 50-60% of the calculation time saved

# Outline

1. Introduction
2. State of the art
3. Proposals
4. Experiments
    1. ALML and Inc-ALML
    2. Active cleaning
    3. Application to TRECVid Collaborative Annotation
    4. Participation to the TRECVid Semantic Indexing task
5. Conclusions and perspectives

# 4.2 Experiments Active Cleaning

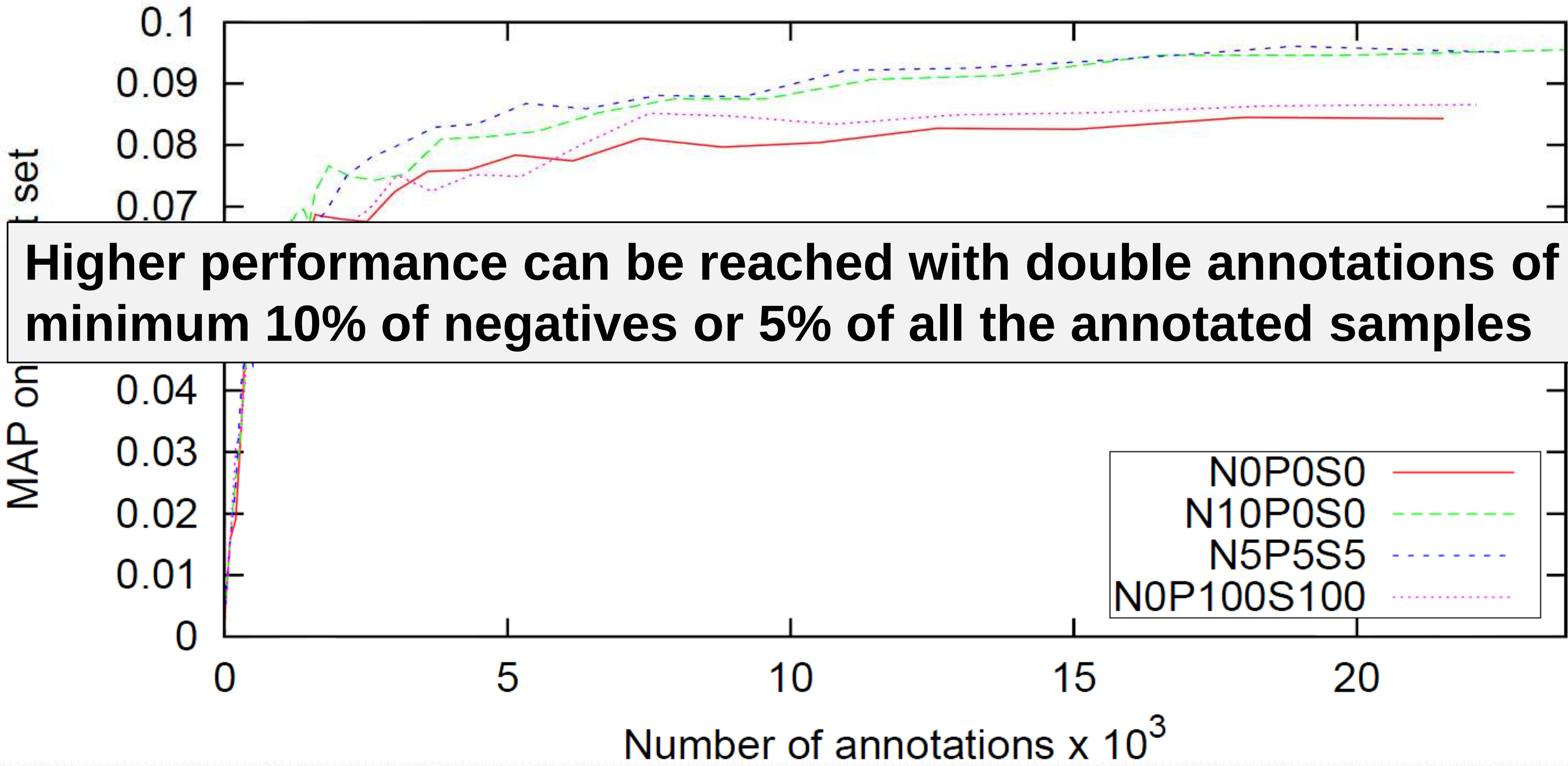
- **Collection:** TRECVid 2007 (100 hrs of videos), 20 concepts
- Three annotations for each (sample × concept):
  - Two from the Collaborative annotations of TRECVid **(CA)**
  - One from **MCG-ICT-CAS**
- Fusion with four descriptors
- MSVM-RBF and relevance sampling

| $Q_{cl}$ | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 |
|---|---|---|---|---|---|---|---|---|
| **Pos %** | 0 | 10 | 0 | 0 | 5 | 10 | 20 | 100 |
| **Neg %** | 0 | 0 | 0 | 10 | 5 | 10 | 20 | 0 |
| **Skip %** | 0 | 0 | 10 | 0 | 5 | 10 | 20 | 100 |

*The (P%,N%,S%) fraction values used in our evaluations*

# 4.2 Experiments Active Cleaning

Fusion of four descriptors; **Classifier**: MSVM; **AL strategy**: Relevance sampling



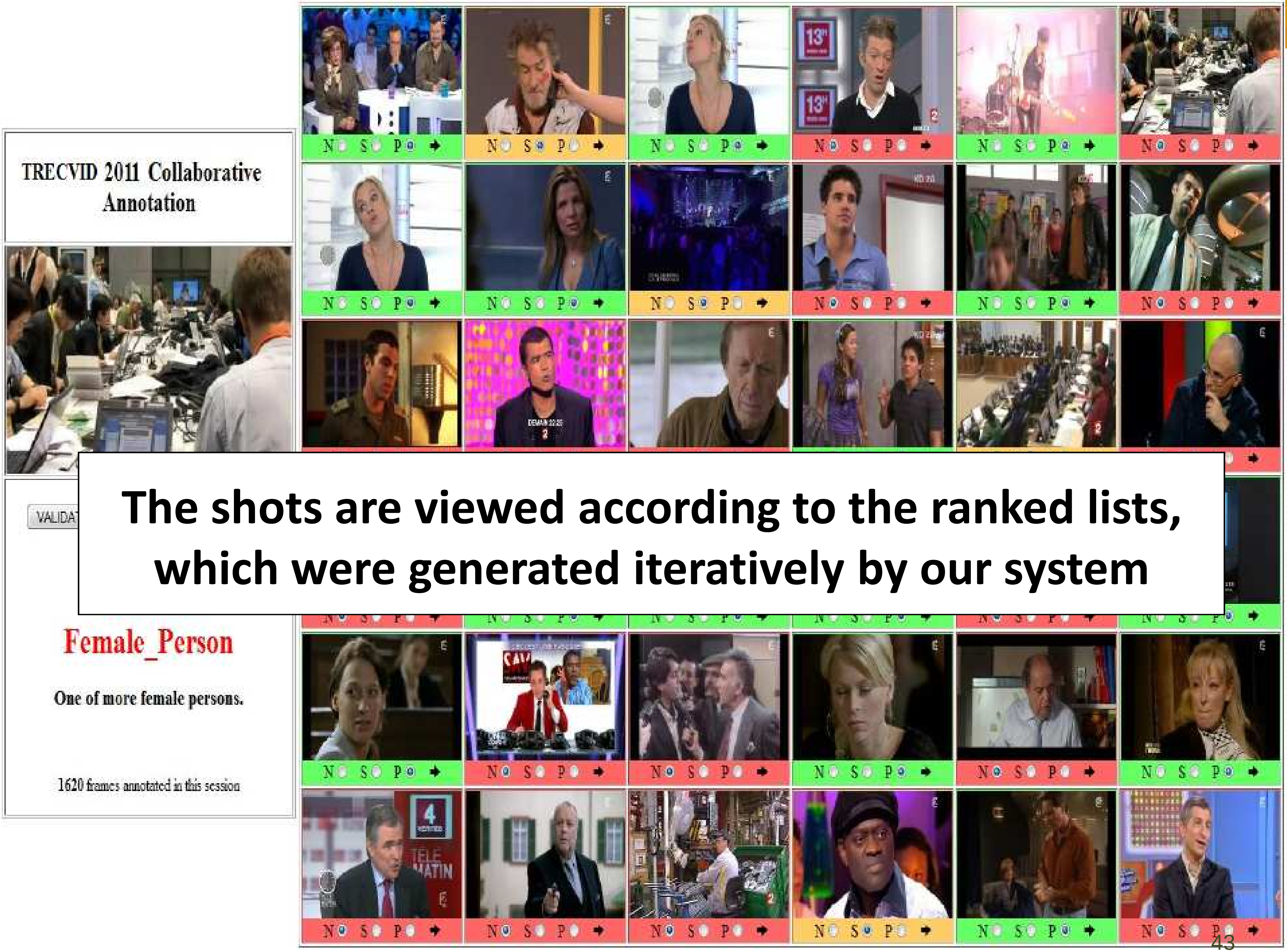

**Higher performance can be reached with double annotations of minimum 10% of negatives or 5% of all the annotated samples**

# Outline

1. Introduction
2. State of the art
3. Proposals
4. Experiments
    1. ALML and Inc-ALML
    2. Active cleaning
    3. Application to TRECVid Collaborative Annotation
    4. Participation to the TRECVid Semantic Indexing task
5. Conclusions and perspectives

# 4.3 Application to TRECVid collaborative annotation

- The collaborative annotations of the TRECVid 2010 and 2011 development sets.
- We have included the proposed methods in the collaborative annotation tool of TRECVid (Ayache and Quénot [2007])
- **Collection:** TRECVid 2011 dev set (400 hrs, from IACC)
  - 266473 shots and 500 target concepts
- **Goal**: produce as many coherent annotations as possible for the development set, with a cheapest cost and within a short time.




**The shots are viewed according to the ranked lists, which were generated iteratively by our system**

# 4.3 Application to TRECVid collaborative annotation

- 346 concepts (40 groups worldwide)
- 4.2 M single concept × shots annotations:
  - ≈ 88% were done once,
  - ≈ 9% were done twice,
  - ≈ 3% were done three or more times

Active learning (Inc-ALML)

Active cleaning

- The 4.2M were amplified to 18M usable annotations using relations between concepts (e.g. *cat* implies *animal*)
- Sparse annotation with AL: about 13% of all the possible annotations

# Outline

1. Introduction
2. State of the art
3. Proposals
4. Experiments
    1. ALML and Inc-ALML
    2. Active cleaning
    3. Application to TRECVid Collaborative Annotation
    4. Participation to the TRECVid Semantic Indexing task
5. Conclusions and perspectives

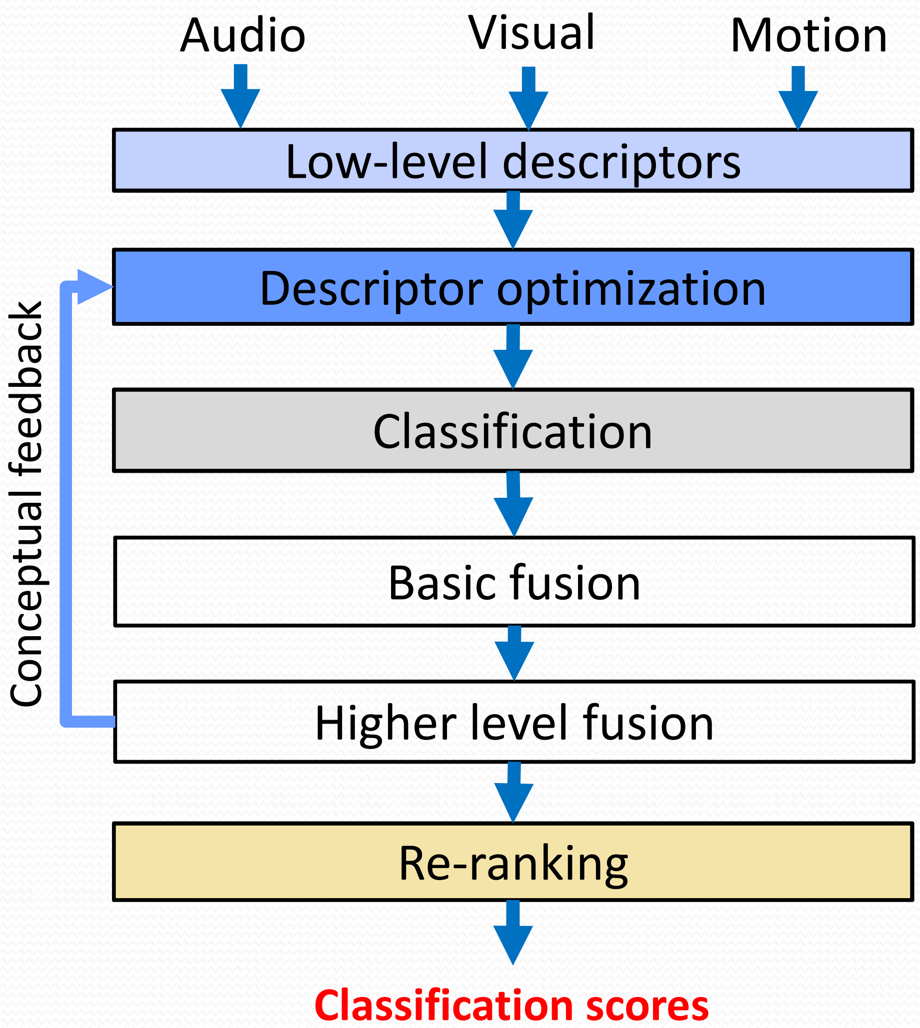
# 4.4 Participation to TRECVid (SIN)

- **Collection:** TRECVid 2011 (600 hrs), 346 semantic concepts

| Collection | Hours | Shots |
|---|---|---|
| Development | 400 | 266473 |
| Test | 400 | 137327 |

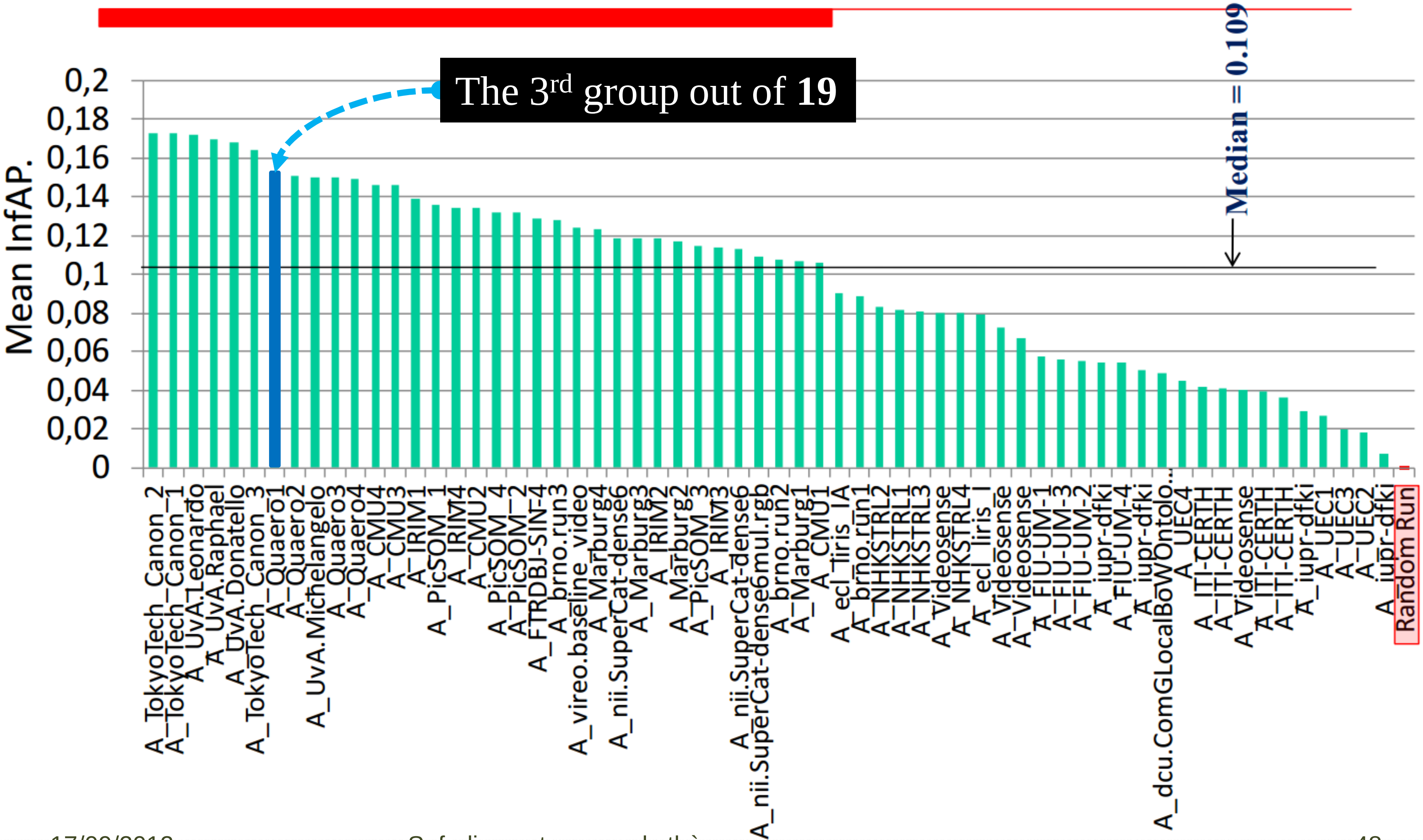
- **Evaluation:** MAP on 50 concepts
- **Participation:** 4 runs (Quaero)

# 4.4 Participation to TRECVid (SIN)




- 48 descriptors (IRIM & Quaero)
- Power-law followed by PCA
- KNN + MSVM (RBF kernel)
- Descriptor and classifer variants
- Higher level hirarchical fusion
- Generate a conceptual descriptor
- Our re-ranking method

# 4.4 Results on TRECVid (SIN)



# Outline

1. Introduction
2. State of the art
3. Proposals
4. Experiments
5. Conclusions and perspectives

# 5.1 Conclusions

- General framework: semantic indexing for retrieval of multimedia documents
- Two main difficulties: semantic-gap and the class-imbalance problems
- Focus on concept indexing of images and videos segments using active learning approaches
- Main objective: to increase the system performance while using as few labeled samples as possible, thereby minimizing the annotation cost

# 5.1 Contributions

**CBMI:**

Three contributions:

- Generalization of the *multi-learner* approach (ML)
- Improvement of a *re-ranking* method
- Generalization and evaluation of the power-law normalization and combination with PCA

# 5.1 Contributions

- **Active learning:**
  - ALML which enhanced the indexing performance
  - Inc-ALML which speeded up ALML (gain about 50-60%)
  - Active cleaning approach (re-annotating small fractions)
- Validations through several experiments on large-scale TRECVid collections
- Application to the TRECVid 2010-2012 collaborative annotations

# 5.2 Perspectives

- **Cold-Start**
  - How to best bootstrap AL?
- **Active Learning on Very Large-Scale Datasets**
  - At each iteration, predicting only on part of the unlabelled data
- **Crowd-sourcing: Annotations Quality and Annotators Surveillance**
  - How can we handle differences among workers in terms of the quality of annotations they provide?
  - How can we find and control noisier annotators?
  - Is it possible to identify ambiguous examples via annotator disagreements?

# Thank you for your attention!

# Publications

1. Bahjat Safadi and Georges Quénot. *Active learning with multiple classifiers for multimedia indexing*. **Multimedia Tools and Applications**, 1-15, 2010.

---

1. Bahjat Safadi, Stéphane Ayache and Georges Quénot. *Active Cleaning for Video Corpus Annotation*. **MMM 2012**, pages:518-528, Klagenfurt, Austria, Jan 2012.
2. Bahjat Safadi and Georges Quénot. *Re-ranking by Local Re-scoring for Video Indexing and Retrieval*. **CIKM** 2011, pages:2081-2084, Glasgow, Scotland, Oct 2011.
3. Bahjat Safadi and Georges Quénot. *Re-ranking for Multimedia Indexing and Retrieval*. **ECIR 2011**, pages:708-711, Dublin,Ireland, Apr 2011.
4. Bahjat Safadi, Yubing Tong and Georges Quénot. *Incremental Multiple Classifier Active Learning for Concept Indexing in Images and Videos*. **MMM 2011**, pages:240-250, Taipei, Taiwan, Jan 2011.
5. Bahjat Safadi and Georges Quénot. *Evaluations of multi-learners approaches for concepts indexing in video documents*. **RIAO 2010**, pages:88-91, Paris, France, Apr 2010.
6. Bahjat Safadi and Georges Quénot. *Active Learning with Multiple Classifiers for Multimedia Indexing*. The 8th IEEE Int. **CBMI 2010**, Grenoble, France, Jun 2010.
7. Bahjat Safadi, Yubing Tong and Georges Quénot. *Incremental Multi-Classifier Learning Algorithm on Grid'5000 for Large Scale Image Annotation*. **ACM Workshop on Very-Large-Scale Multimedia Corpus, Mining and Retrieval**, pages:1-6, Firenze, Italy, Oct 2010.

---

1. Bahjat Safadi and Georges Quénot. *Apprentissage Actif avec une Méthode de Rordonnancement Pour l'Indexation et la Recherche de Vidéos*. **CORIA 2011**, pages::231-245, Avignon, France, Mar 2011.
2. Bahjat Safadi and Georges Quénot. *Evaluation des approches multi-apprenants pour l'indexation des concepts dans les documents vidéo*. **CORIA 2010**, Sousse, Tunisie, Mar 2010.

# Experiments (ML)

Optimal ratios ($f_{min}$ )

| *Descriptor* | **Dim** | ***SRBF*** | ***MRBF*** | ***SLIN*** | ***MLIN*** | ***SLR*** | ***MLR*** |
|---|---|---|---|---|---|---|---|
| CEALIST/global_tlep | 756 | 8 | 4 | 2 | 0.5 | 2 | 0.2 |
| LEAR/bow_sift_1000 | 1000 | 8 | 4 | 4 | 1 | 2 | 0.2 |
| ETIS/global_qwm1x3 | 96 | 4 | 3 | 4 | 2 | 2 | 0.05 |
| LIG/hg104 | 104 | 4 | 2 | 2 | 0.05 | 2 | 0.05 |
| LIG/opp_sift_har | 4000 | 3 | 3 | 3 | 3 | 2 | 0.2 |

- Optimal ratios are lower for multiple learners
- Optimal ratios for LIN and LR $<$ than for RBF
- Results are quite stable against the descriptor types

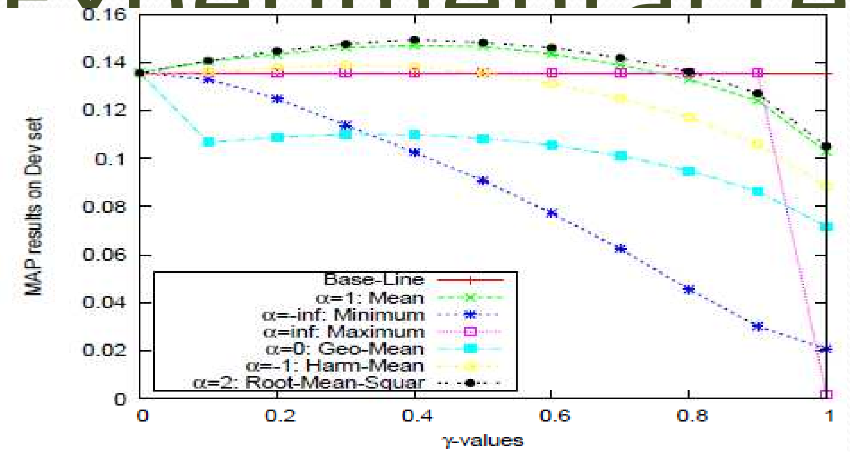
# Experiments (ML)

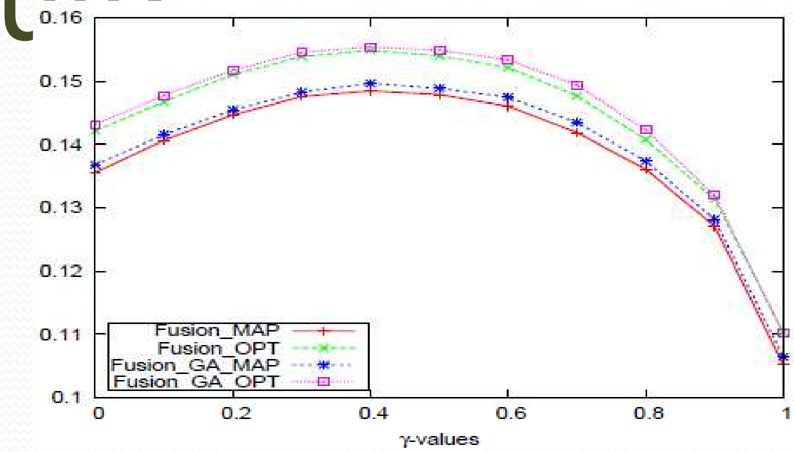
TRECVid 2008:

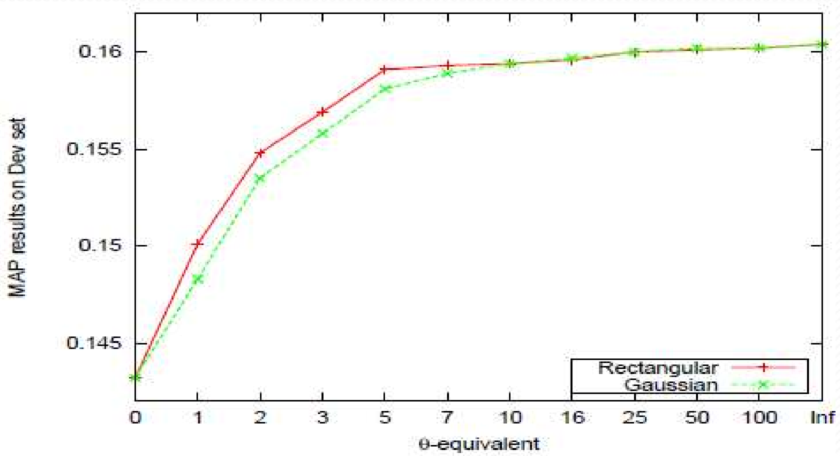
| *Descriptor* | *SRBF* | *MRBF* | *SLIN* | *MLIN* | *SLR* | *MLR* | *SKNN* |
|---|---|---|---|---|---|---|---|
| CEALIST/global_tlep | 0.0667 | **0.0751** | 0.0319 | 0.0405 | 0.0368 | 0.0598 | 0.0678 |
| LEAR/bow_sift_1000 | 0.0489 | **0.0561** | 0.0237 | 0.0345 | 0.0231 | 0.0469 | 0.0467 |
| ETIS/global_qwm1x3 | 0.0561 | 0.0566 | 0.0348 | 0.0465 | 0.0369 | 0.0469 | **0.0608** |
| LIG/hg104 | 0.0541 | **0.0596** | 0.0223 | 0.0310 | 0.0240 | 0.0481 | 0.0580 |
| LIG/opp_sift_har | 0.0651 | **0.0747** | 0.0485 | 0.0652 | 0.0486 | 0.0644 | 0.0621 |

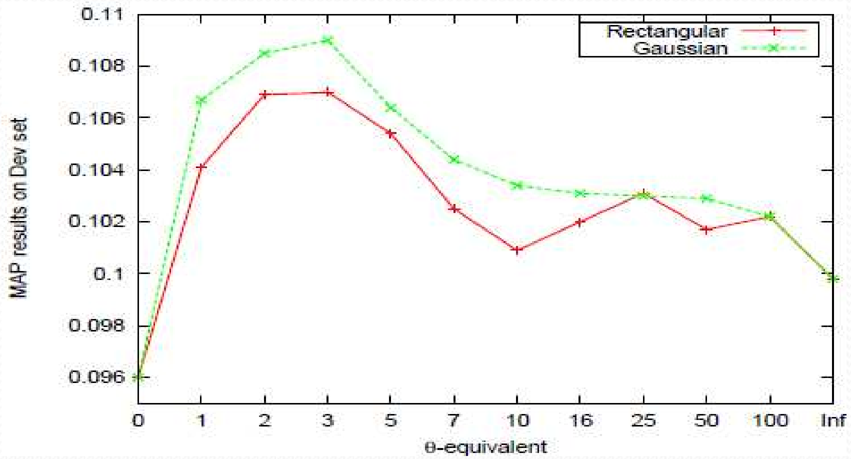
- Multiple learner is significantly better than single learner
- LR better than LIN but RBF better than LR

# Experimental results



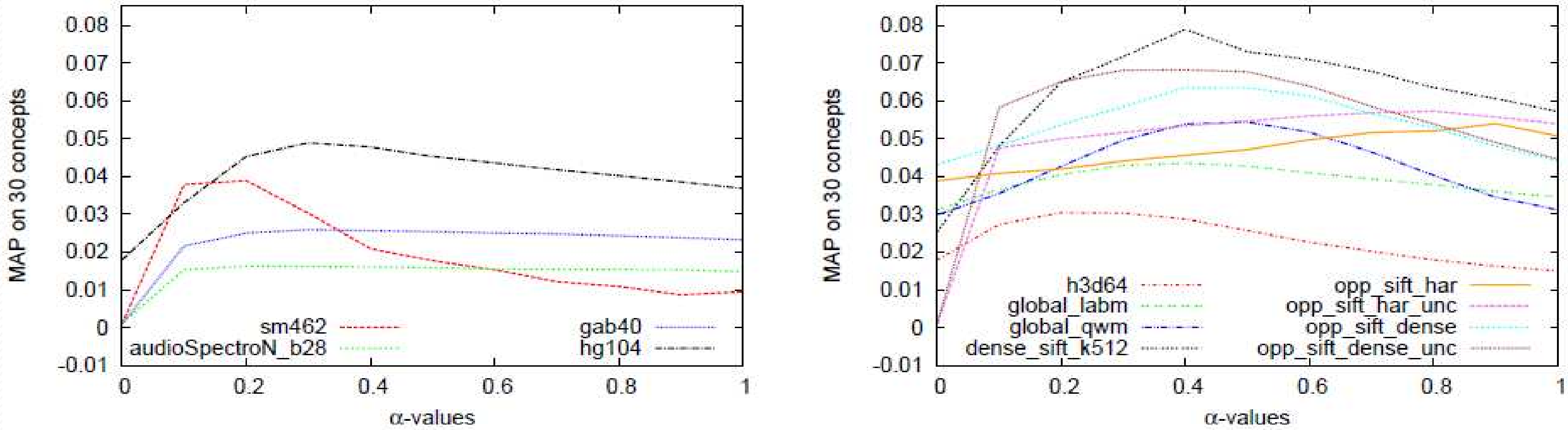



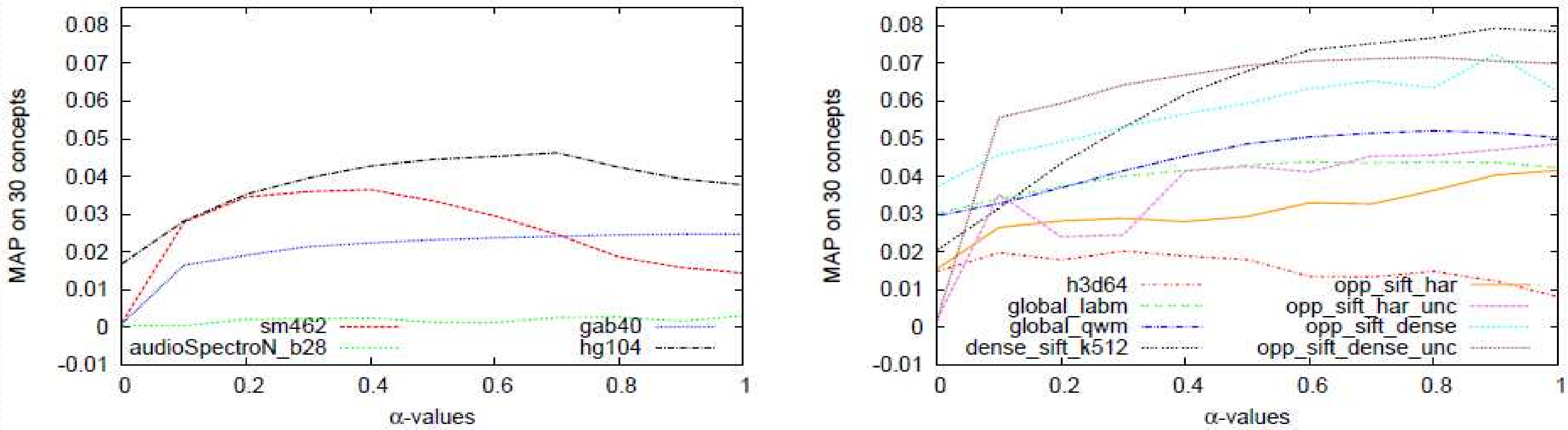
# Experimental results

Results of the re-ranking method on the test sets of TRECVid 2010 and 2008 with ($\gamma = 0.4$ *and* $\alpha = 2$).

| | **TV10** | | **TV08** | |
|---|---|---|---|---|
| | **θ /σ** | **MAP** | **θ /σ** | **MAP** |
| **Baseline** | 0 | 0.0480 | 0 | 0.099 |
| **ALL** | ∞ | 0.0568 **(+18%)** | ∞ | 0.101 **(+2%)** |
| **Rectangular** | θ = ∞ | 0.0568 **(+18%)** | θ = 3 | 0.112 **(+13%)** |
| **Gaussian** | σ = ∞ | 0.0568 **(+18%)** | σ = 2 | 0.109 **(+11%)** |

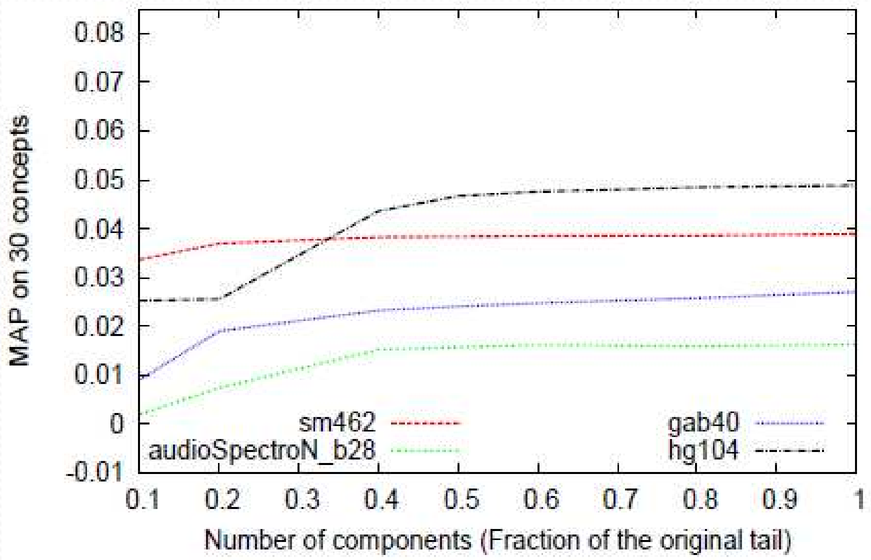


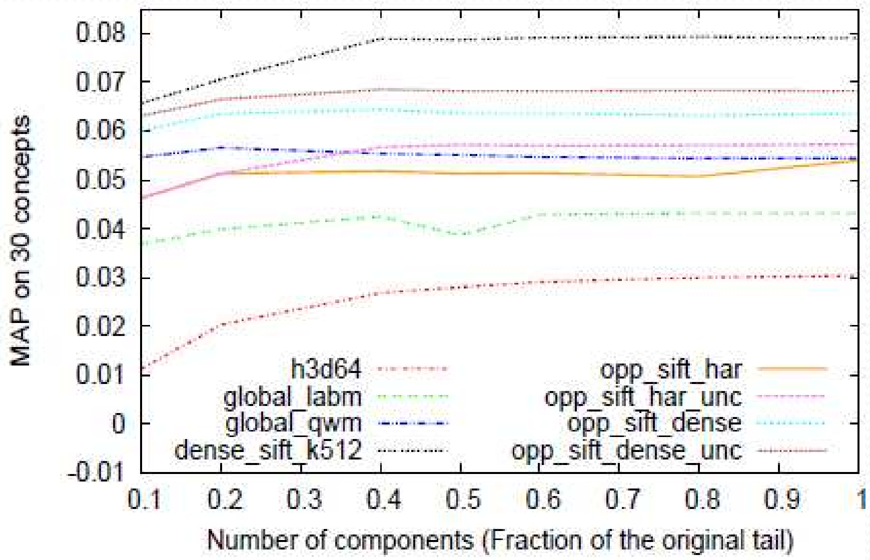

(a) Euclidean Distance




(b) Chi-square Distance

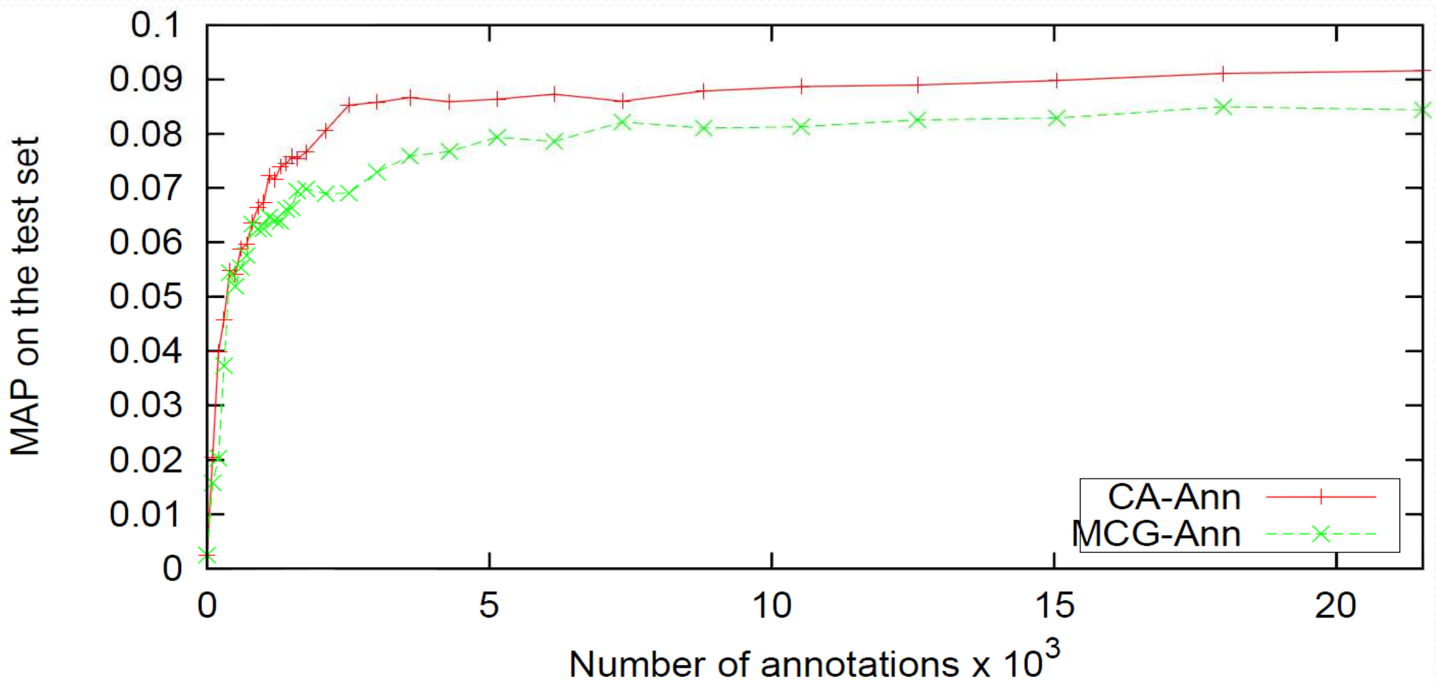
# Experiments and results



# (3/4) Experiments (AC)

The MAP calculated on 20 concepts with two different annotation sources.



# Experiments results (AC)

| | | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 |
|---|---|---|---|---|---|---|---|---|---|
| **MCG -CA** | MAP | 0.084 | 0.084 +0% | 0.086 +2% | **0.095** +14% | **0.096** +14% | 0.097 +15% | 0.097 +15% | 0.086 +2% |
| | #Ann | 21532 | +65 | +50 | +2100 | +1100 | +2200 | +4400 | +1150 |
| **CA-MCG** | MAP | 0.091 | 0.091 +0% | 0.092 +1% | **0.096** +5% | **0.095** +4% | 0.090 -1% | 0.095 +4% | 0.093 +2% |
| | #Ann | 21532 | +46 | +11 | +2150 | +1100 | +2215 | +4420 | +580 |

The result of the posteriori cleaning

| | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | Full |
|---|---|---|---|---|---|---|---|---|---|
| **MCG-CA** | 0.084 | 0.083 | 0.084 | 0.085 | 0.084 | 0.085 | 0.087 | 0.086 | **0.096** |
| **CA-MCG** | 0.091 | 0.091 | 0.092 | 0.091 | 0.092 | 0.091 | 0.092 | 0.093 | |