IRIM@TRECVID2012
Hierarchical Late Fusion for Concept Detection in Videos

IRIM Group, GDR ISIS, FRANCE
http://mrim.imag.fr/irim

Alexandre Benoit,
LISTIC - Université de Savoie, Annecy, France,

TRECVID 2012 Workshop
November 25, 2012, Gaithersburg MD, USA

IRIM partners, from descriptors sharing to fusion methods

16 laboratories, 37 researchers

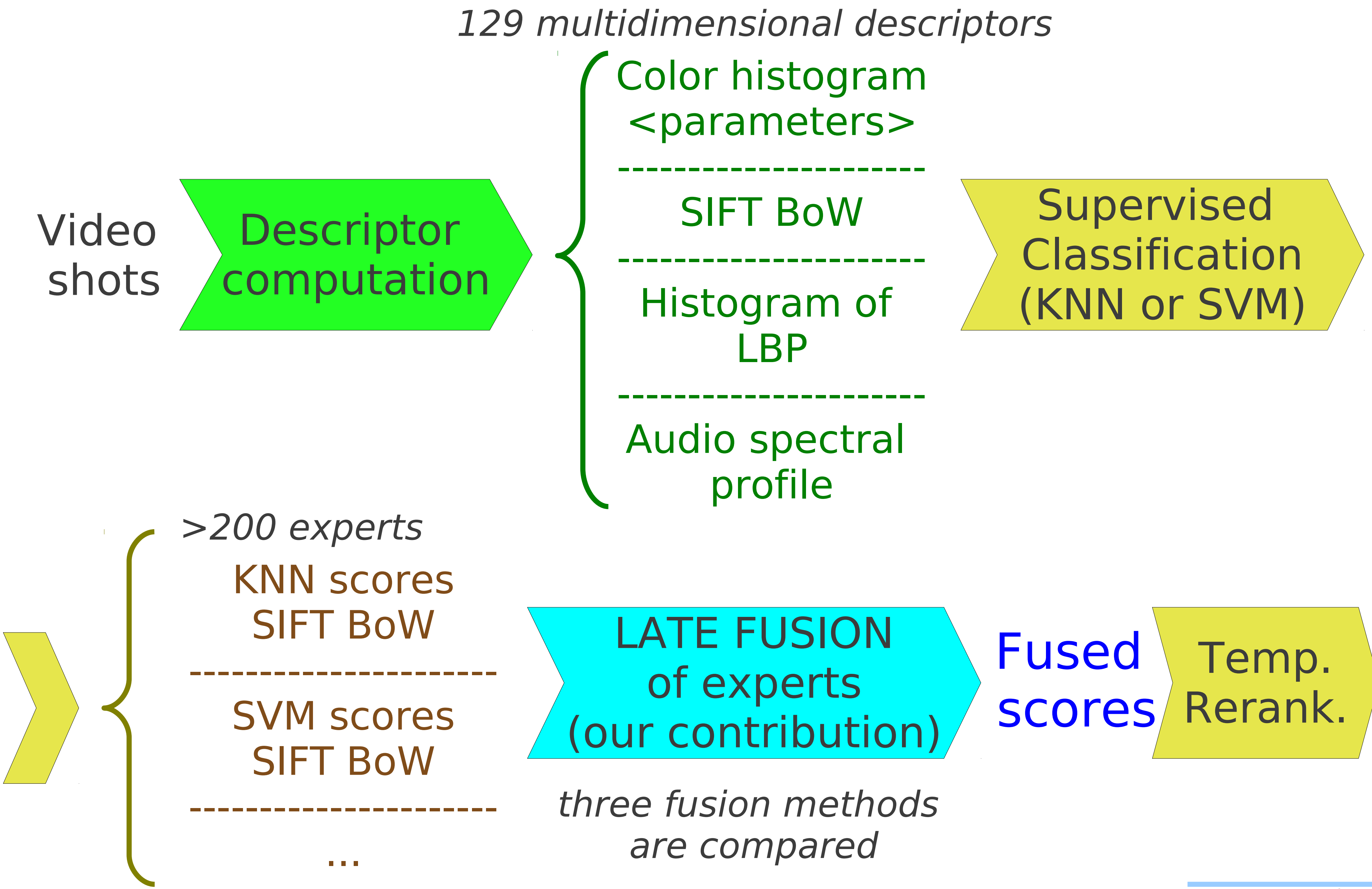
Nicolas Ballas (CEA, LIST)
Benjamin Labbé (CEA, LIST)
Aymen Shabou (CEA, LIST)
Hervé Le Borgne (CEA, LIST)
Philippe Gosselin (ETIS, ENSEA)
Miriam Redi (EURECOM)
Bernard Mérialdo (EURECOM)
Hervé Jégou (INRIA Rennes)
Jonathan Delhumeau (INRIA Rennes)
Rémi Vieux (LABRI, CNRS)
Boris Mansencal (LABRI, CNRS)
Jenny Benois-Pineau (LABRI, CNRS)
Stéphane Ayache (LIF, CNRS)
Abdelkader Hamadi (LIG, CNRS)
Bahjat Safadi (LIG, CNRS)
Franck Thollard (LIG, CNRS)
Nadia Derbas (LIG, CNRS)
Georges Quénot (LIG, CNRS)
Hervé Bredin (LIMSI, CNRS)

Matthieu Cord (LIP6, CNRS)
Boyang Gao (LIRIS, CNRS)
Chao Zhu (LIRIS, CNRS)
Yuxing tang (LIRIS, CNRS)
Emmanuel Dellandrea (LIRIS, CNRS)
Charles Edmond-Bichot (LIRIS, CNRS)
Liming Chen (LIRIS, CNRS)
Alexandre Benoit (LISTIC)
Patrick Lambert (LISTIC)
Sabin Tiberius Strat (LISTIC, LAPI Bucharest)
Joseph Razik (LSIS, CNRS)
Sébastion Paris (LSIS, CNRS)
Hervé Glotin (LSIS, CNRS)
Tran Ngoc Trung (MTPT)
Dijana Petrovska (MTPT)
Gérard Chollet (Telecom ParisTech)
Andrei Stoian (CEDRIC)
Michel Crucianu (CEDRIC)

Outline

- Processing chain : late fusion context
- IRIM descriptors
- Fusion principles
- Proposed fusion methods
- Results
- Conclusions

Processing chain : late fusion context



IRIM group *shared* descriptors

CEA LIST,
SIFT BoV
Local edge patterns

ETIS/LIP6,
VLAT
Color histograms

EURECOM,
Saliency moments

INRIA Rennes,
Dense SIFT, VLAD

LABRI,
face detection

LIF,
percept

LIG,
OppSIFT, STIP,
Concepts

LIRIS,
OCLBP BoW
MFCC BoW

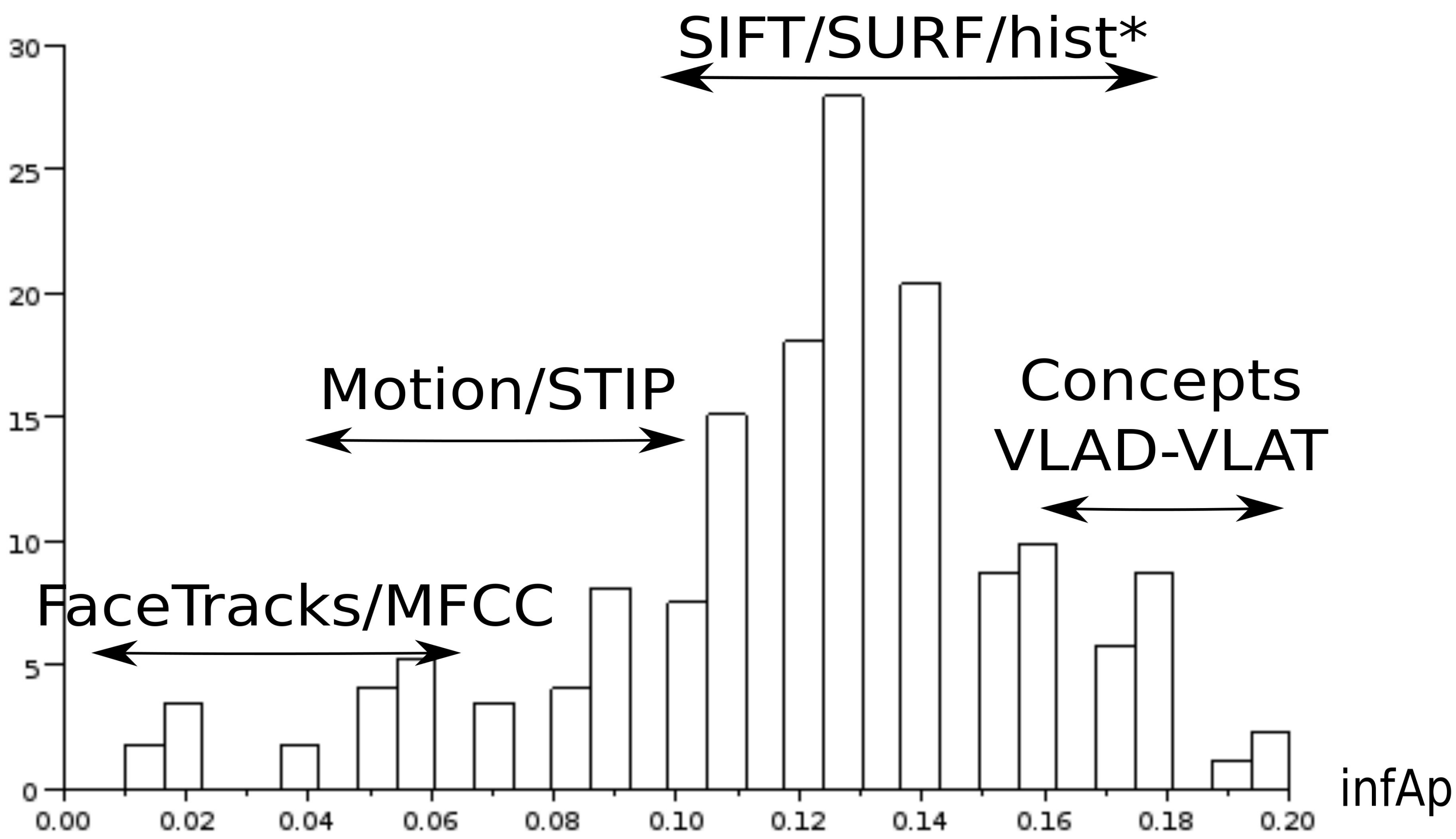
LISTIC,
SIFT retina BoW

LSIS,
MLHMS

MTPT,
superpixel color sift

IRIM descriptors

- Single descriptors initial infAp disribution
- Heterogeneous behaviors, each one can contribute more for specific concepts

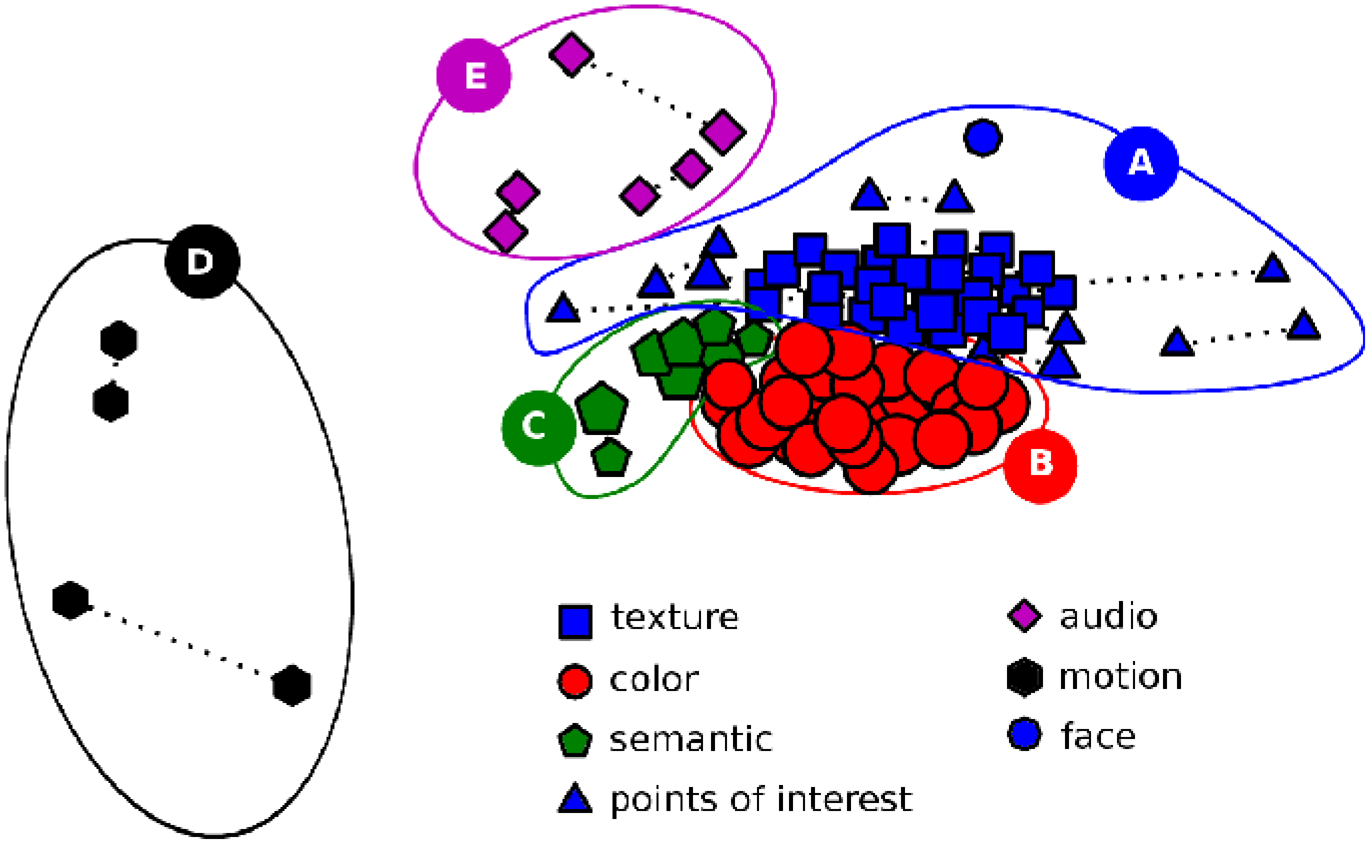


Late fusion principles

- Elementary expert =

  video descriptor + optimisation + machine learning algorithm
- *"schemes* (experts) *with dissimilar outputs but comparable performance are more likely to give rise to effective naive data fusion"* [Ng and Kantor]
- Experts of similar types tend to give similar shot rankings, but they are usually complementary with experts of different types
- Then fuse elementary experts to create higer level experts
  - First group similar elementary experts (clustering stage)
  - Fuse elementary experts in each group/family to balance the families (intra-group fusion)
  - Fuse the different groups together (inter-group fusion), which gives the main performance increase

Late fusion principles (II)



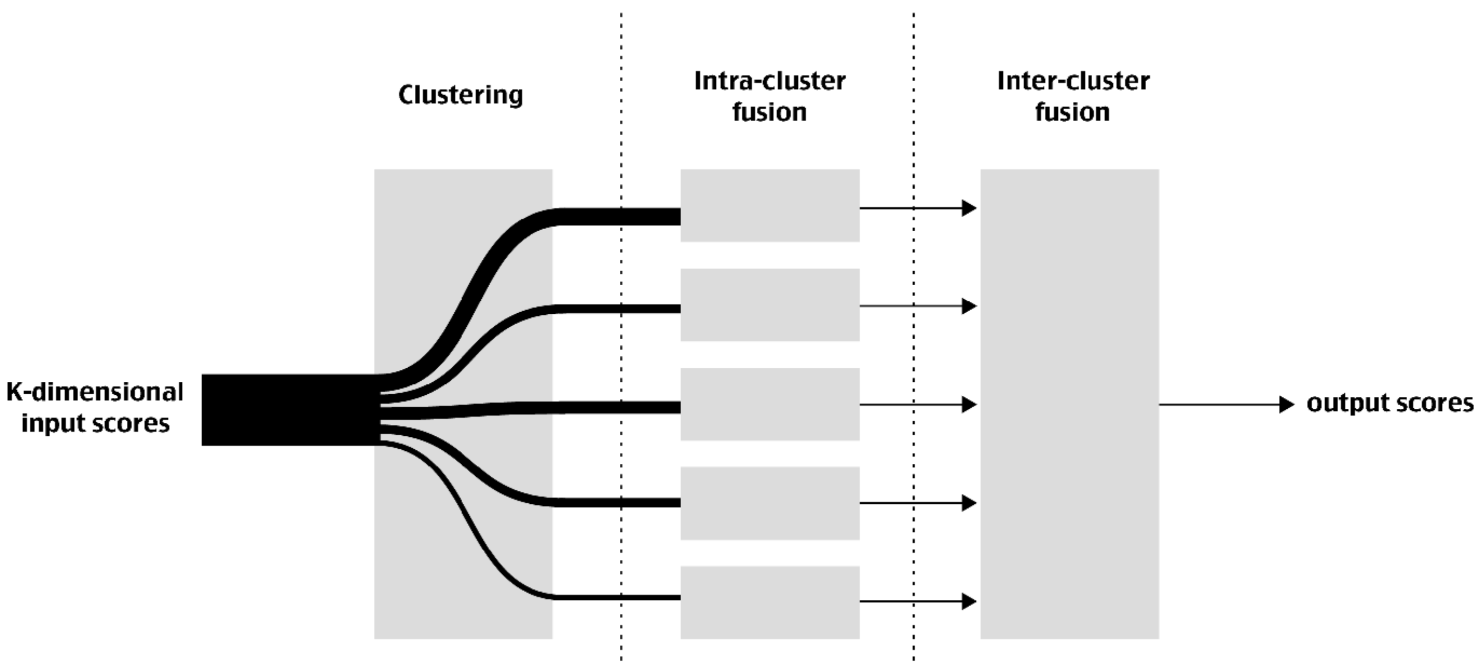
Grouping experts in families based on the similarity of outputs, for concept "Computers"

- Example of an automatic grouping (through automatic community detection)
- Experts of similar types tend to give similar rankings and achieve similar performances
- They are therefore automatically grouped in the same family

Proposed fusion methods

- Three fusion approaches are compared :
  - Manual hierarchical grouping
  - Agglomerative clustering
  - Community detection

- Common principles :
  - clustering stage (manual or automatic)
  - intra-cluster fusion
  - inter-cluster fusion



Manual hierarchical grouping

KNN scores
SIFT BoW 1024

SVM scores
SIFT BoW 1024

...

Fuse KNN-SVM pairs

arithmetic mean of normalized scores

ALLC scores
SIFT BoW 1024

ALLC scores
SIFT BoW 2048

ALLC scores
Color hist. 1X1

ALLC scores
Color hist 2x2

Fuse versions

ALLC scores
SIFT BoW all

ALLC scores
Color hist. all

ALLC scores
Audio spectral profile all

...

Fuse same modality

ALLC scores
visual all

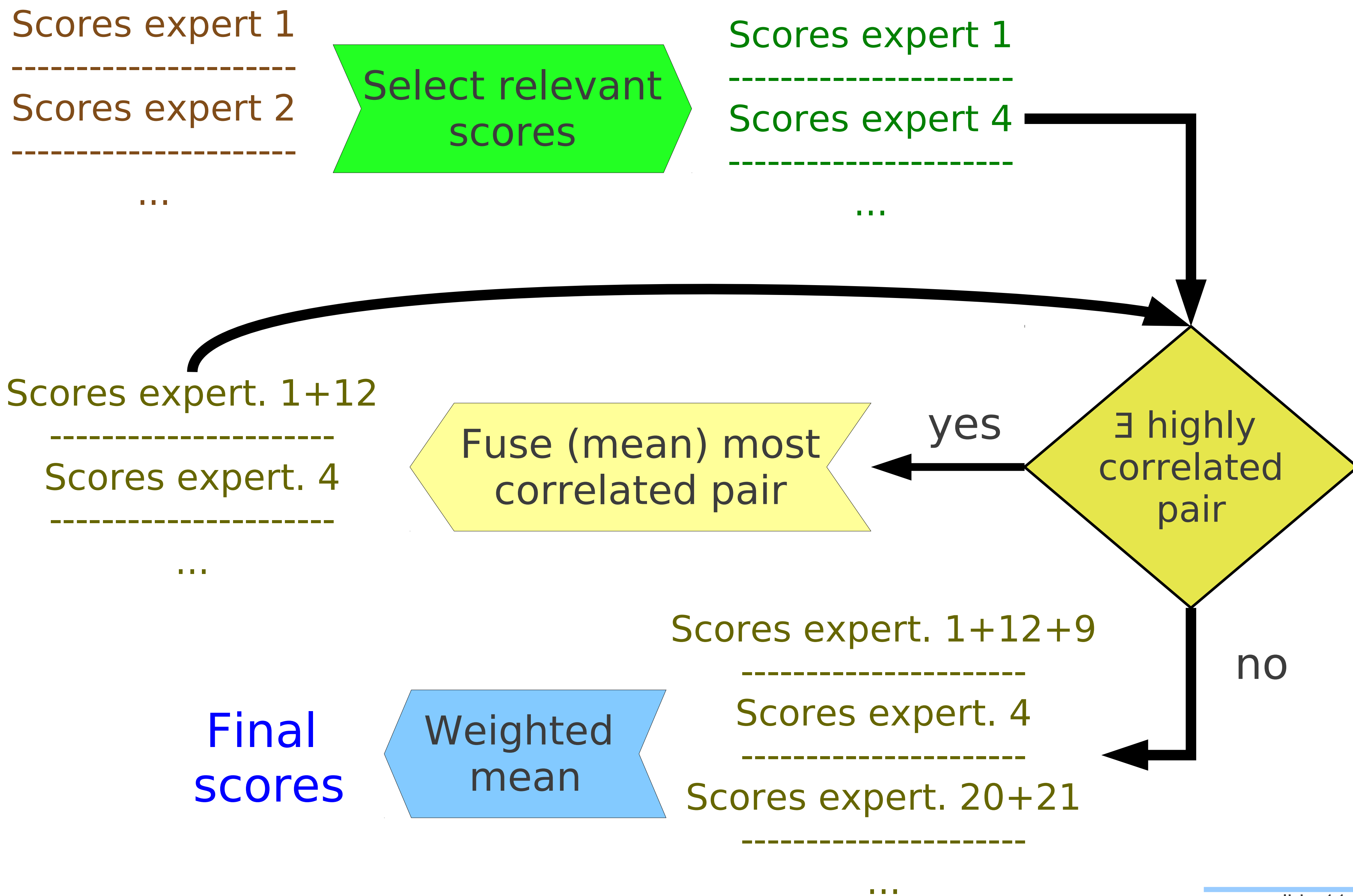
ALLC scores
audio all

Fuse different modalities

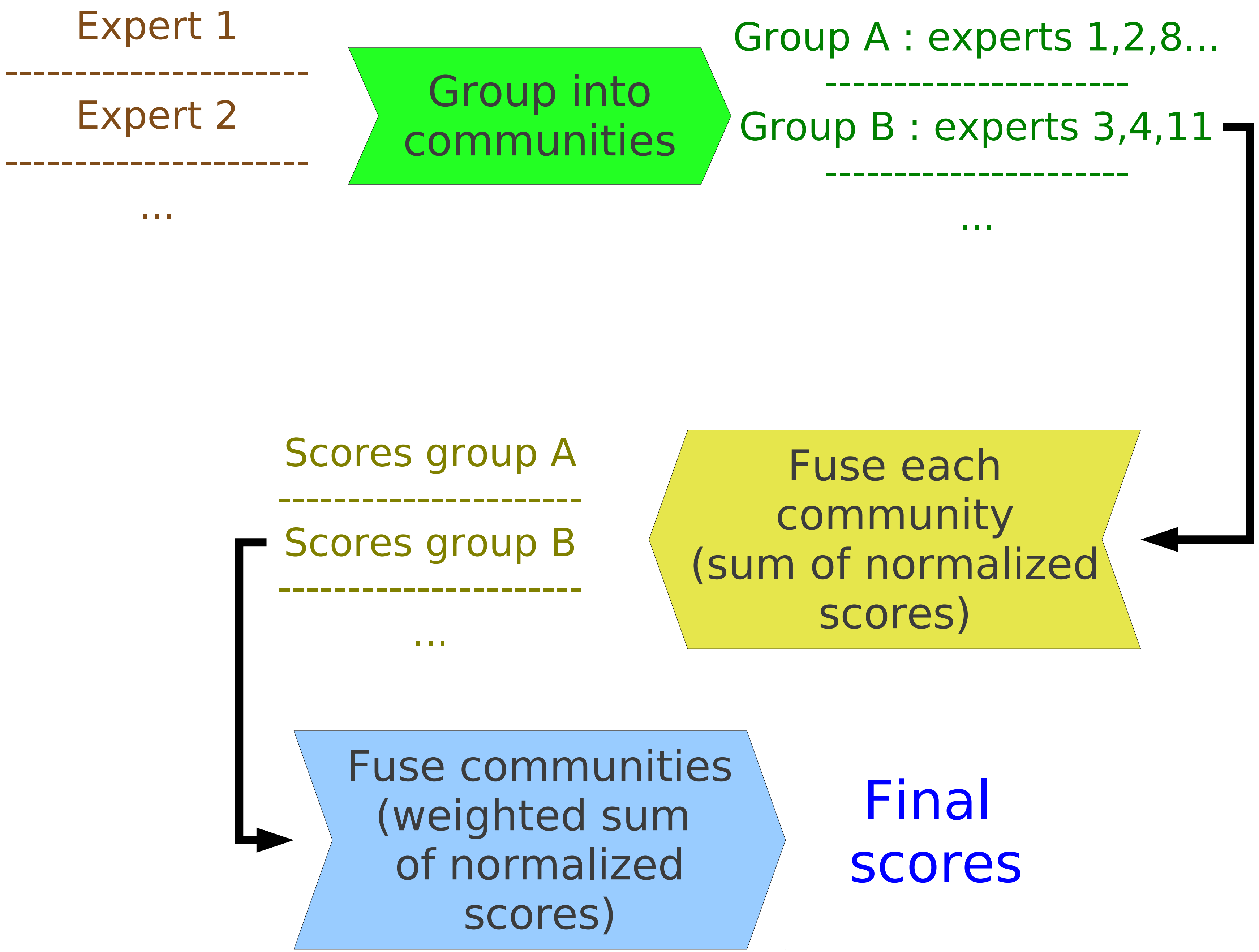
Final scores

weighted mean of normalized scores, optimized weights

Agglomerative clustering



Community detection



Community detection : details

Group into communities

Rank correlation coefficient

$$\rho_{ij} = \frac{\sum_{n=1}^{n=\mathrm{N}} \left(r_{in} - \overline{r_i}\right)\left(r_{jn} - \overline{r_j}\right)}{\sqrt{\sum_{n=1}^{n=\mathrm{N}} \left(r_{in} - \overline{r_i}\right)^2 \sum_{n=1}^{n=\mathrm{N}} \left(r_{jn} - \overline{r_j}\right)^2}}$$

Maximisation of modularity *[Blondel et al.]*

$$\mathcal{Q} = \frac{1}{\sum_{i,j} A_{ij}} \sum_{i,j} \left[A_{ij} - \frac{\sum_k A_{ik} \sum_k A_{kj}}{\sum_{i,j} A_{ij}} \right] \delta_{ij}$$

$$A_{ij} = \max\left(0, \rho_{ij}\right)$$

δ_{ij} = 1 if *i* and *j* in the same group

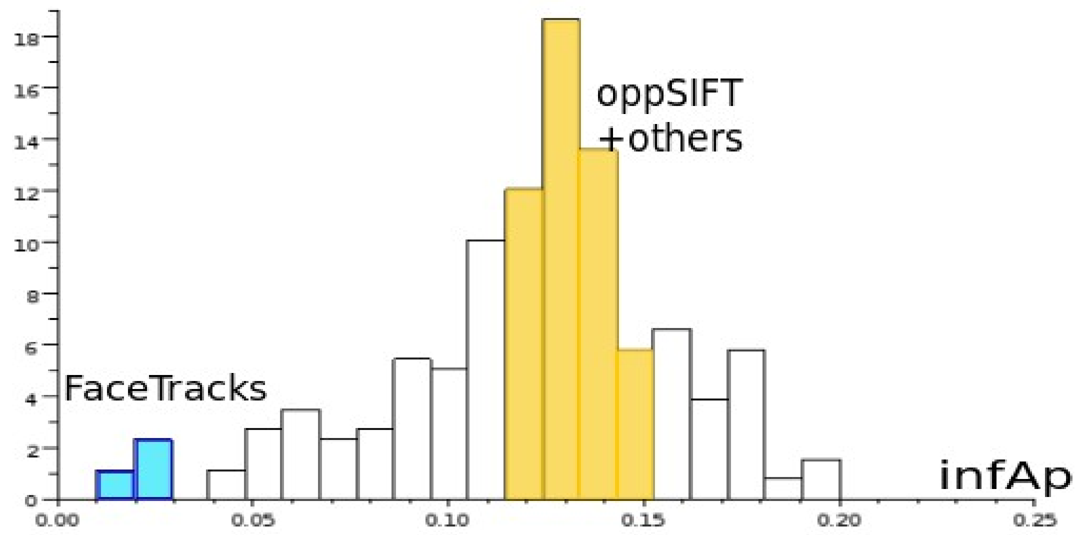
Score normalisation strategy

$$\widehat{x_{kn}} = \frac{1}{2}\left\{\tanh\left[0.01\left(\frac{x_{kn} - \mu_k}{\sigma_k}\right)\right] + 1\right\}$$

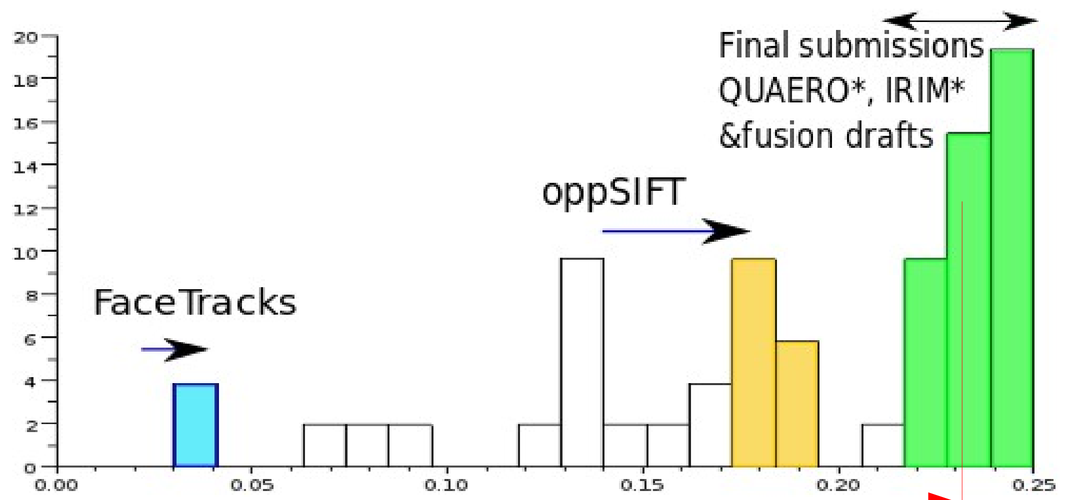
Descriptors fusion... and performance increase

Intra fusion + **inter fusion** improve performances !

Single experts
performance distribution



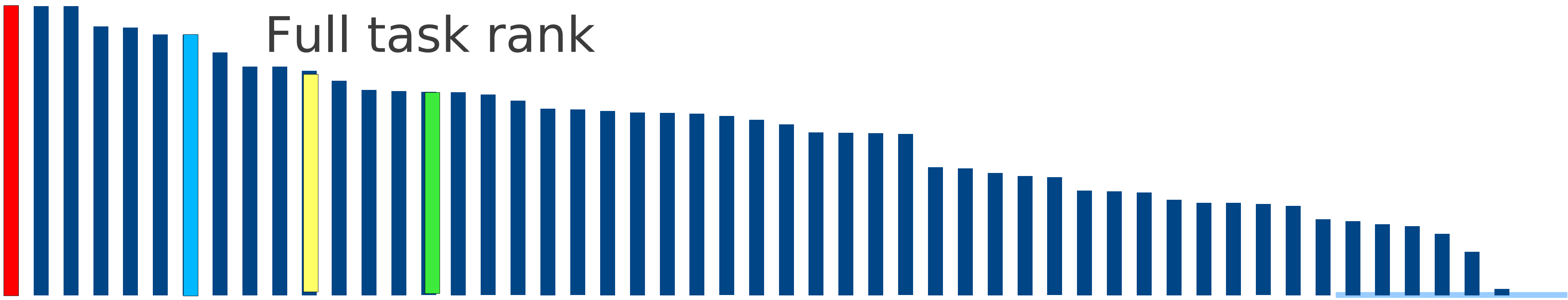
High level experts
performance distribution.
From intra fusion
to final inter fusion



Performances on TRECVID 2012 SIN

- Results when fusing available ALLC scores (KNN + SVM)
- Some slight differences between methods inputs

Type of fusion	infAP Full task	infAP Light task
Manual hierarchical fusion (*Quaero1_1*)	**0.2691**	**0.2851**
Agglomerative clustering (*IRIM1_1*)	0.2378	0.2549
Community detection (*IRIM2_2*)	0.2248	0.2535
Best performer (TokyoTechCanon2_brn_2)	***0.3210***	***0.3535***



Performances on TRECVID 2012 SIN (re-rank)

- Temporal re-ranking: video shots in the vicinity of a detected positive also have a chance of being positives *[Safadi and Quénot 2011]*

Type of fusion	**infAP no re-rank**	**infAP with re-rank**	**% increase**
Manual hierarchical fusion	**0.2487**	**0.2691**	**8.2**
Agglomerative clustering	0.2277	0.2378	4.4
Community detection	0.2154	0.2248	4.4

- **Temporal re-ranking increases average precisions**

Performances on TRECVID 2012 SIN

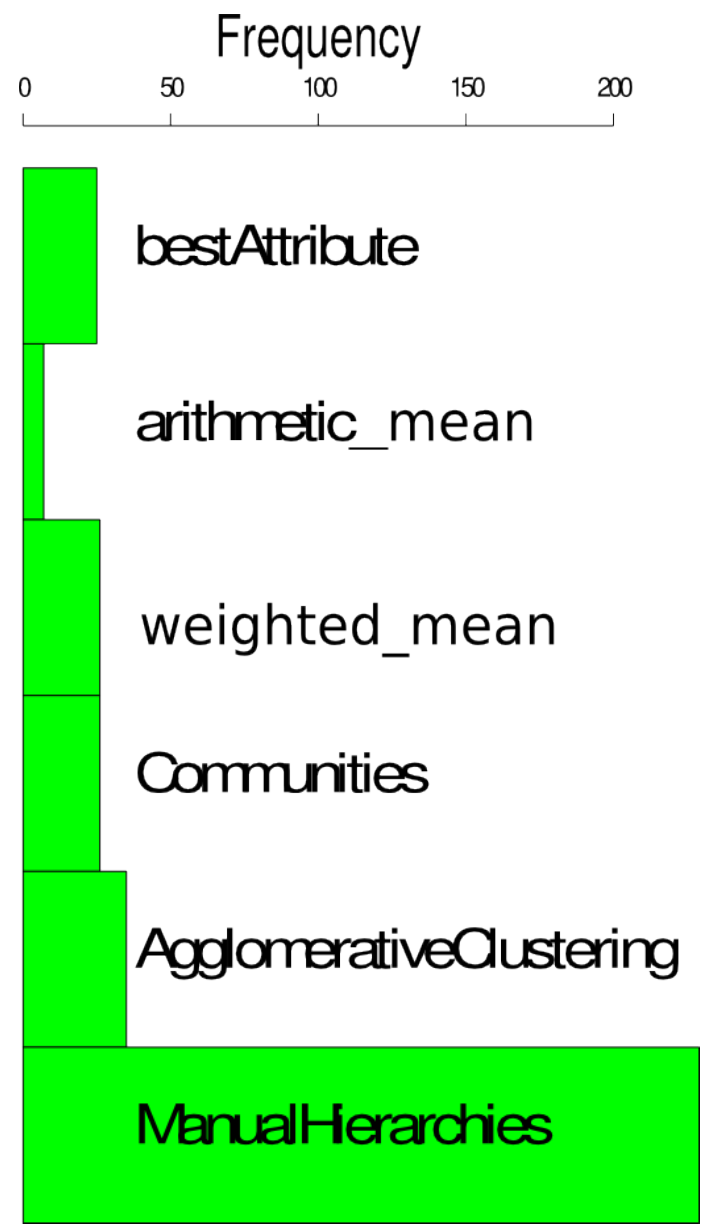
- 2012d (x=>y) subcollections analysis details
- Even the arithmetic mean greatly improves average precision.
- Manual and automatic fusion methods enhance results more

	infAP	**Performance evolution**	
Type of fusion	**Full task**	**over Best (%)**	**over arithm (%)**
Manual hierarchical fusion	**0.2469**	30.4	17.7
Agglomerative clustering	0.2247	18.6	7.2
Community detection	0.2206	16.5	5.2
Arithmetic mean	0.2097	10.7	0.0
Weighted mean	0.2183	15.3	4.1
Best expert per concept	0.1894	0.0	-9.7

Performances on TRECVID 2012 SIN

For how many concepts was a fusion algorithm the best ?

- 2012d subcollections ranking details
- The more complex fusion methods are more often better than the arithmetic (or weighted) mean
- Manual hierarchy definitely best performer



Performances : Method and Cost

- Manual hierarchical grouping:
    - best performer
    - low cost computational
    - requires human expertise

- Automatic fusion methods:
    - No human expertise needed (faster to apply)
    - Automatic update when adding new inputs

    - Agglomerative clustering: reduces input dataset
    - Community detection: keeps all input dataset

    - **… on the need of a fusion of the proposed fusion approaches ?**

Conclusions

- More experts lead to better results
- Even weak experts, especially if complementary, increase performance (resembles AdaBoost)
- All methods are better than Best expert for each concept
- Complex methods better than arithmetic mean (but not by much)
- Possible improvements: combine different fusion strategies, various normalization strategies at different levels

Acknowledgements

- This work was supported by the GDR 720 ISIS (Information, Signal, Images et ViSion) from CNRS. Experiments presented in this paper were carried out using the Grid'5000 experimental testbed, being developed under the INRIA ALADDIN development action with support from CNRS, RENATER and several Universities as well as other funding bodies. This work was also partly supported by the Quaero Program funded by OSEO (French State agency for innovation) and the VideoSense and QCompere projects, funded by ANR (French national research agency).

- Descriptors sharing:
  - The authors would also like to thank all the members of the IRIM consortium for the classifier scores used throughout the experiments described in this paper.

  - Share more, enhance more ! Let's extend the approach !

  TRECVid data sharing: http://mrim.imag.fr/trecvid (login with TRECVid active participants' identifier and password).