Fast learning methods adapted to the user specificities: application to earth observation image information mining

Pierre Blanchart

PhD defense - September 26, 2011

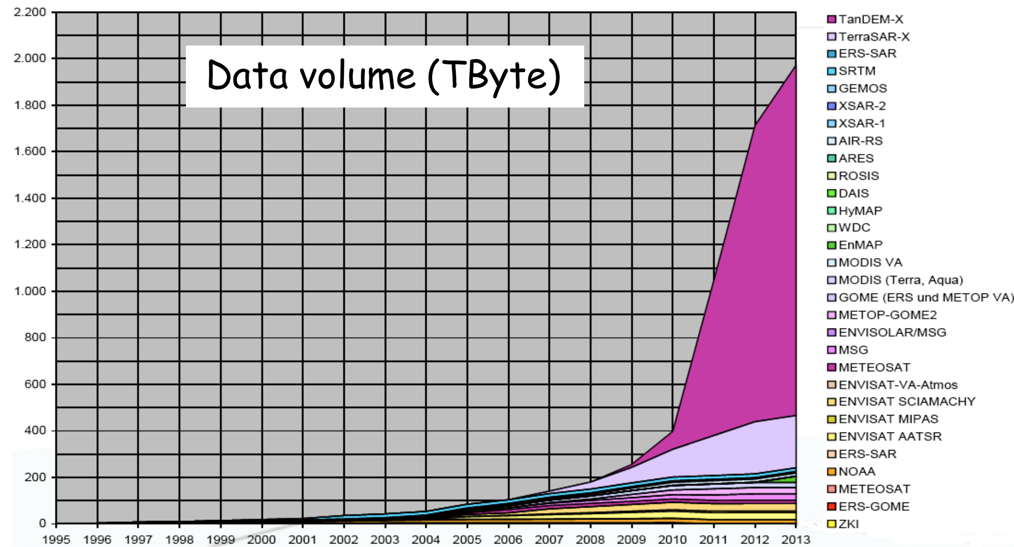
Supervisor: Mihai Datcu

Co-supervisor: Marin Ferecatu

Context and related problems

→ Exponential growth of EO image databases

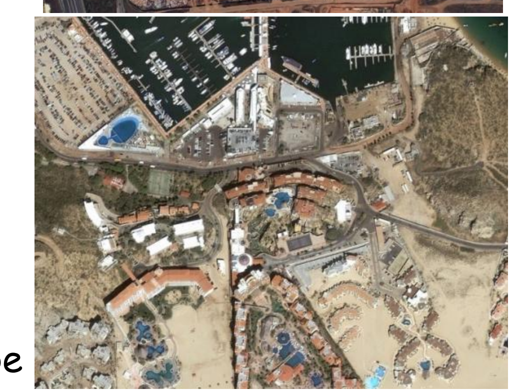
→ Need for efficient methods to index these databases

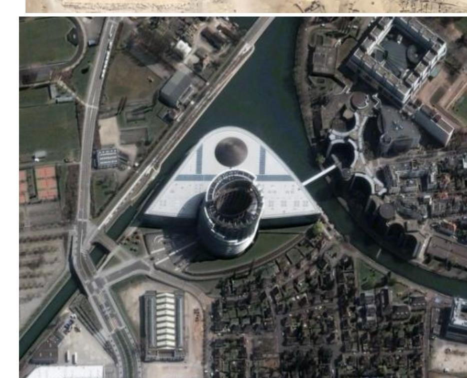


Main Challenges:

→ Exhaustive training databases do not exist

→ Efficient indexing methods to extract image semantics are computationally expensive and often require a high level of supervision





©DigitalGlobe

Target objectives

→ Learning from small and non-exhaustive training datasets

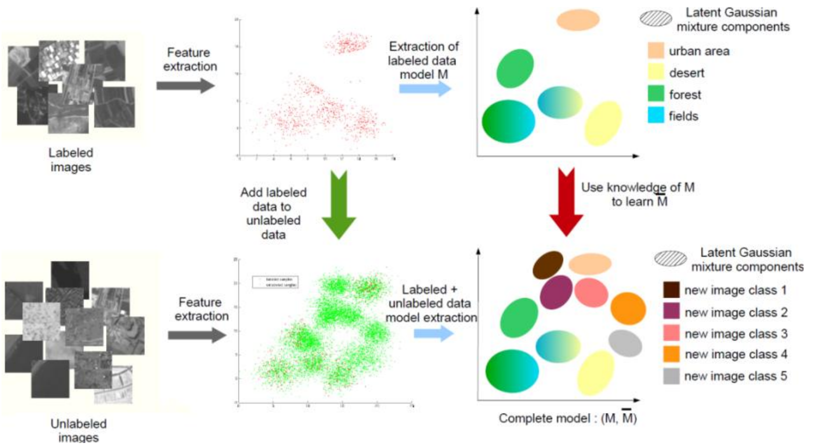
→ Designing learning algorithms scaling up to large data volumes

→ Adapting to a human user (need for fast learning methods allowing fluid user-system interactions)

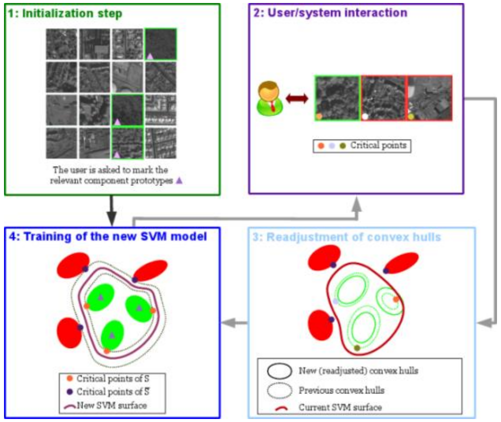
→ Envisaging these problems successively from the point of view of auto-annotation and interactive image search engines

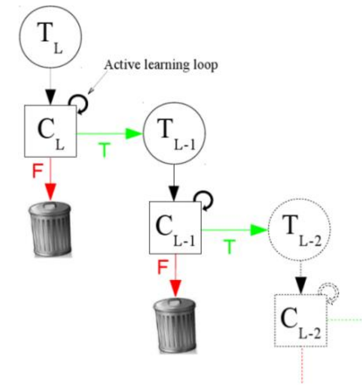
Contributions of the thesis

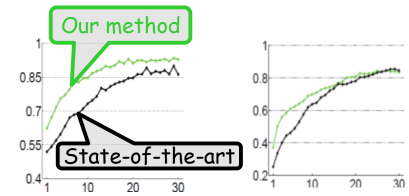
- **auto-annotation systems**
  - identify unknown semantic structures to handle the non-exhaustiveness of training databases

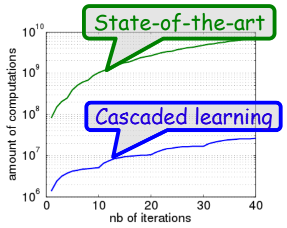


- **interactive category search engines**
  - accelerate the learning of the targeted image category by using the statistical distribution of the data
  - use of coarse-to-fine strategies to considerably reduce the number of evaluation of the decision function to perform object retrieval inside an active learning scheme









Outline of the presentation

I. Semi-supervised auto-annotation in the context of non-exhaustive training datasets

II. Accelerated semi-supervised active learning in the framework of interactive image search engines

III. Interactive object detection in large satellite image repositories using a cascaded active learning scheme

IV. Conclusion and perspectives

Outline

I. Semi-supervised auto-annotation in the context of non-exhaustive training datasets

II. Accelerated semi-supervised active learning in the framework of interactive image search engines

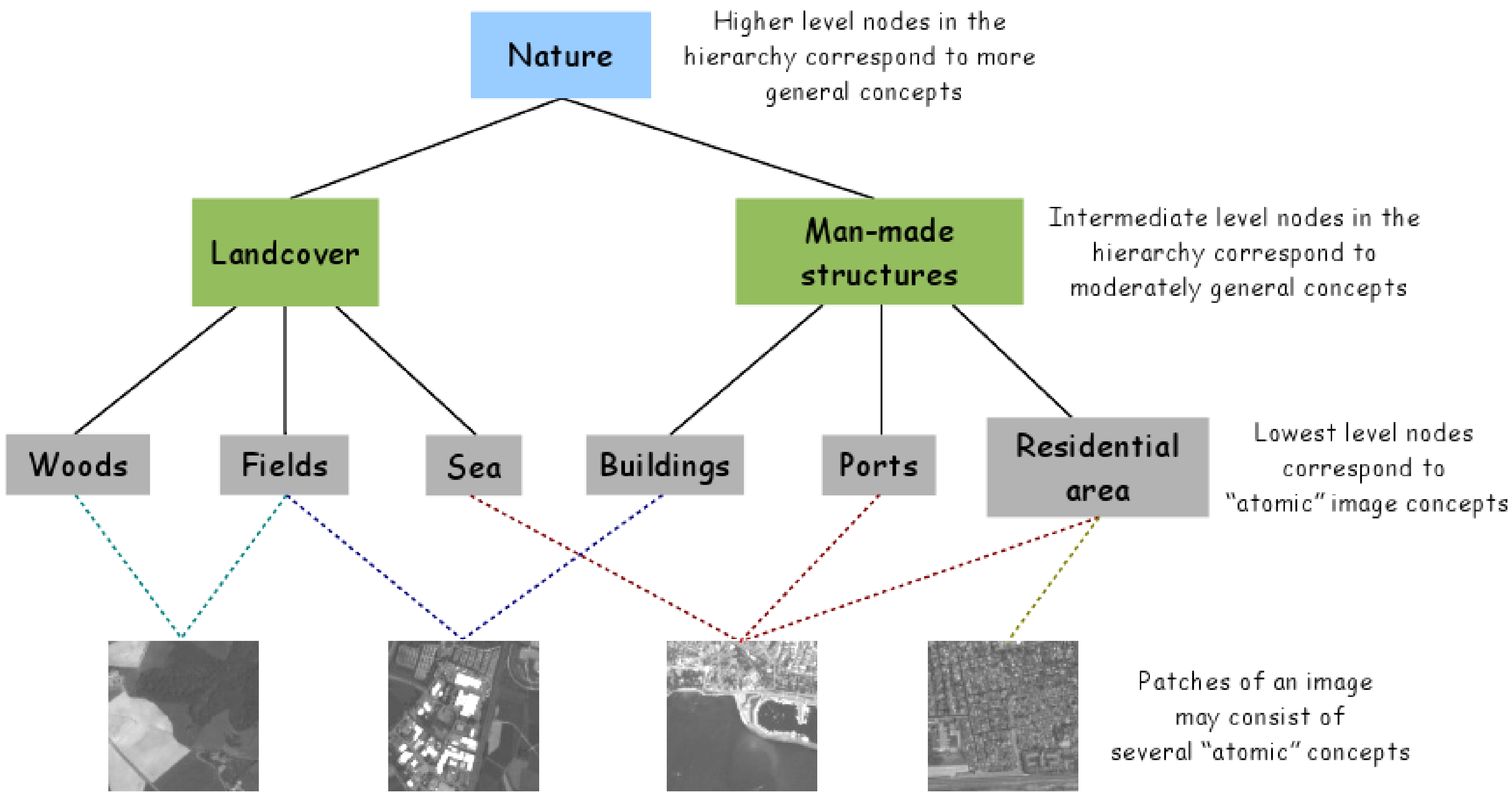
III. Interactive object detection in large satellite image repositories using a cascaded active learning scheme

IV. Conclusion and perspectives

Auto-annotation systems and unknown semantic structures discovery

Goal: associate words belonging to a predefined vocabulary with images

→ Building of an annotation model linking low-level image descriptors to high-level semantic concepts



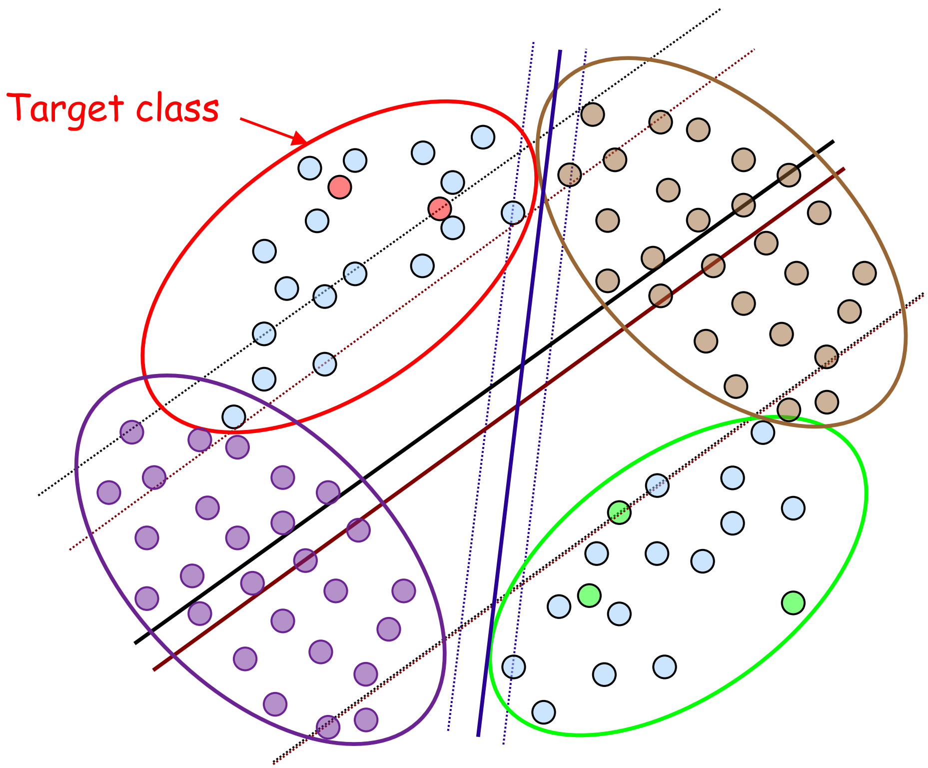
Positioning of the problem

Objective: exploiting the information from both the labeled and the unlabeled part of the database

→ Semi-supervised methods fit naturally inside this framework

→ **Problem:** avoiding the common assumption of semi-supervised methods i.e. considering that the distribution of unlabeled data fits that of labeled data: $p(x) \rightarrow p(y|x)$

→ not verified in the case of non-exhaustive training datasets



Proposed concept

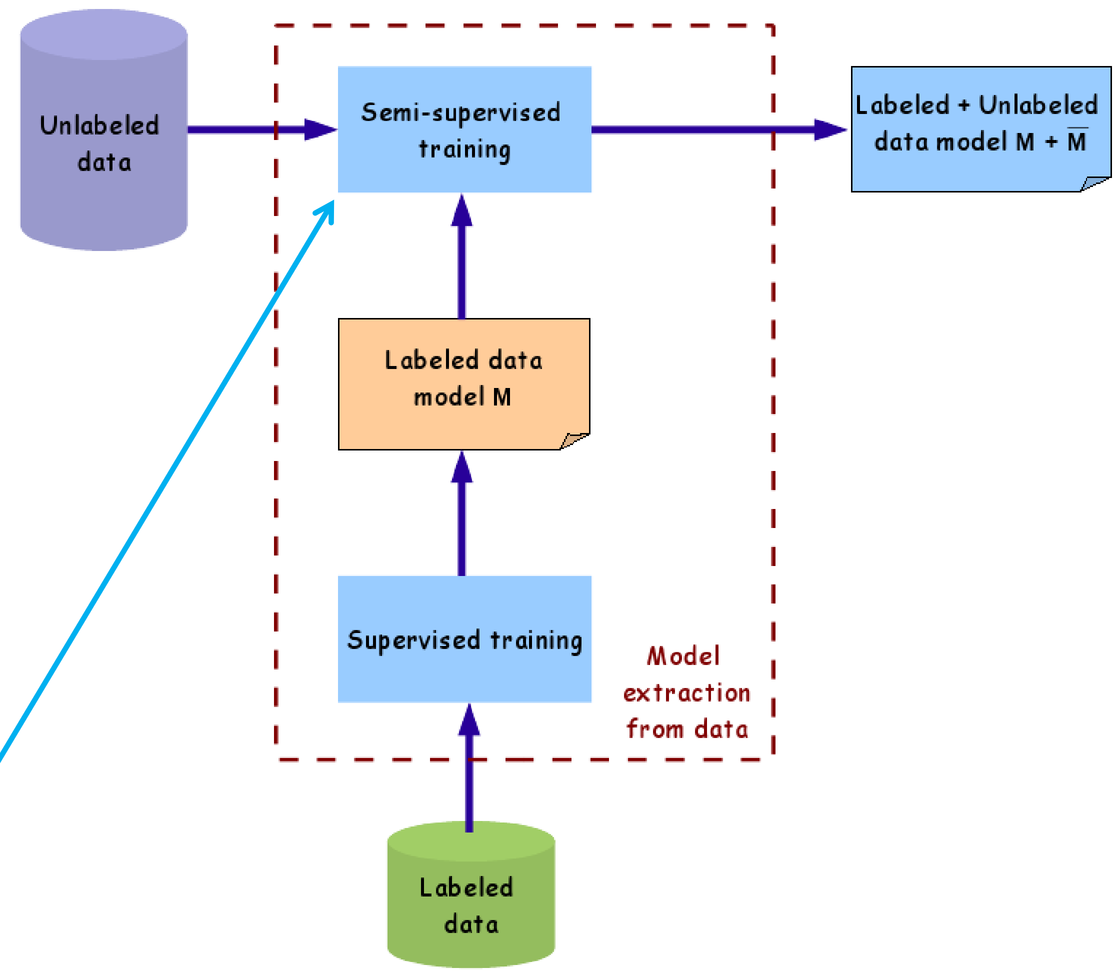
Objective:

→ Computing in a joint way the model of labeled data and the model of unlabeled data

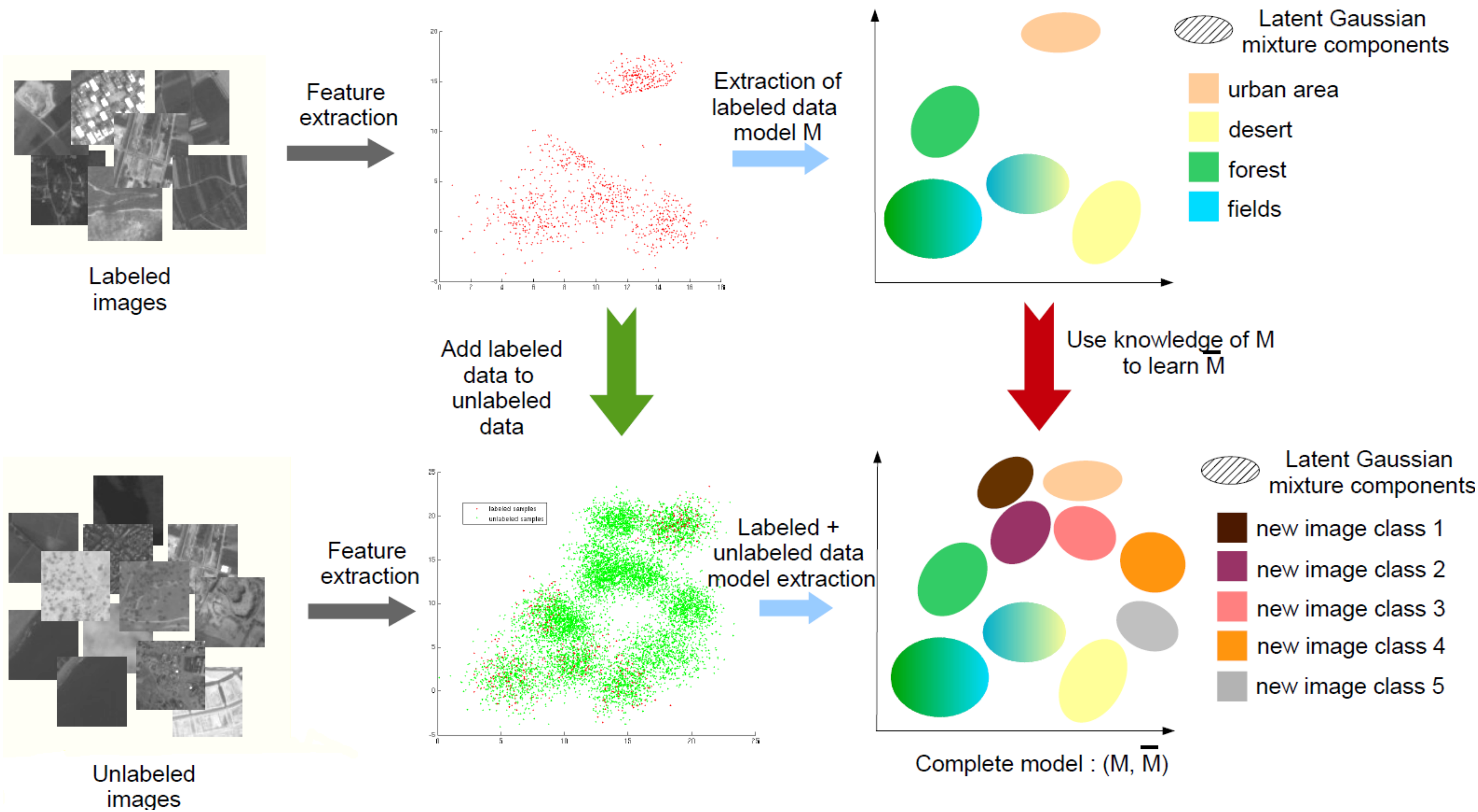
✓ Labeled data contain known image classes

✓ Unlabeled data contain both known and unknown image classes

✓ Unlabeled data are used to learn the model $\overline{M}$ and to refine the model M



Synopsis of the system



Auto-annotation model (1/3)

- Multi-labeled and multi-valued training dataset

$$v_1^{d^1}, v_2^{d^1}, \ldots, v_{N_{d^1}}^{d^1}$$
$$a_1^{d^1}, a_2^{d^1}, \ldots, a_{N_{d^1}}^{d^1}$$

document 1

$$v_1^{d^K}, v_2^{d^K}, \ldots, v_{N_{d^K}}^{d^K}$$
$$a_1^{d^K}, a_2^{d^K}, \ldots, a_{N_{d^K}}^{d^K}$$

document K

$v_1^{d^1}$ $v_2^{d^1}$ $v_3^{d^1}$ ...

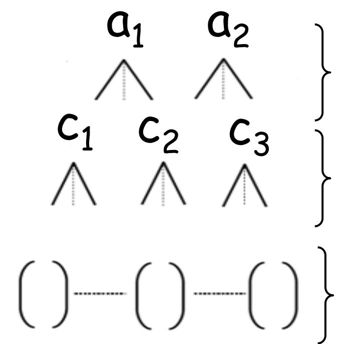
Urban area, road, landing strip, sea ...

document 1

Auto-annotation model (2/3)

Purpose: Predicting unigram models for each image

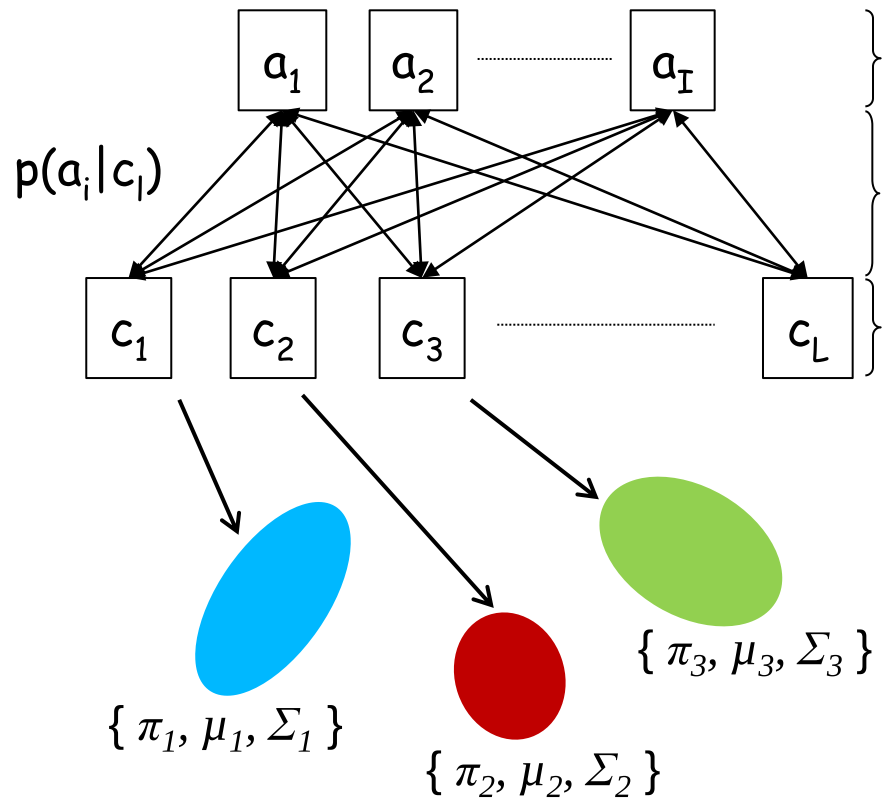
$$p(a_i|Img) = \sum_{l=1}^{L}\left[\pi_l \cdot p(a_i|c_l) \cdot \prod_{v_j \in Img} p(v_j|c_l)\right]$$



Supervised part of the model:

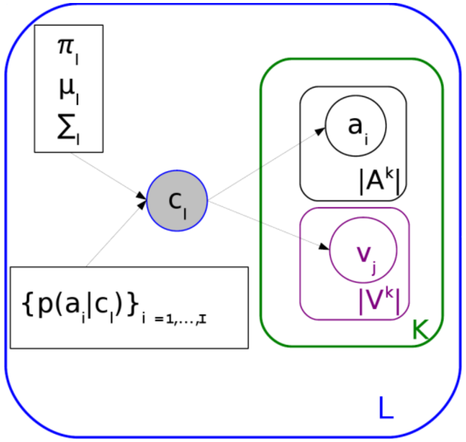
- Learning the parameters of hierarchical Bayesian model to perform bottom-up inference

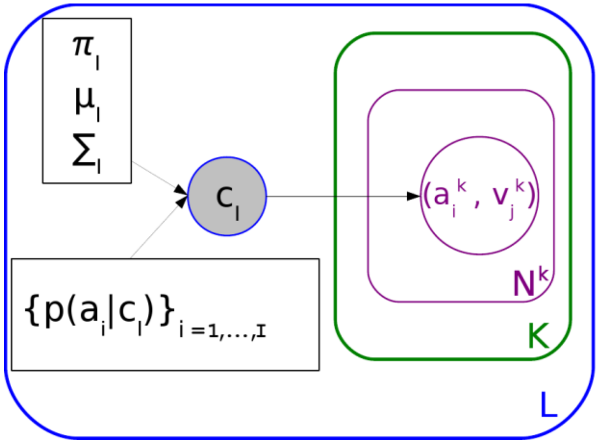
$$\theta = \left\{\pi_l, \mu_l, \Sigma_l, \{p(a_i|c_l)\}_{i=1,\ldots,I}\right\}_{l=1,\ldots,L}$$



Auto-annotation model (3/3)

- Two cases:
  - No-explicit associations between feature vectors and image patches
  - Explicit associations





- Two-fold bag-of-words assumption

Inside a document d_k:

- ✓ Conditional independence of atomic concepts given a latent mixture component ➡ $p(A^k|c_l) = p(a_1^k, \dots, a_{|A^k|}^k|c_l) = \prod_{i=1}^{|A^k|} p(a_i^k|c_l)$
- ✓ Conditional independence of feature vectors given a latent mixture component ➡ $p(V^k|c_l) = p(v_1^k, \dots, v_{|V^k|}^k|c_l) = \prod_{j=1}^{|V^k|} p(v_j^k|c_l)$

$$p(d_k|c_l) = \left[\prod_{j=1}^{|V^k|} p(v_j^k|c_l)\right] \cdot \left[\prod_{i=1}^{|A^k|} p(a_i^k|c_l)\right] \Longrightarrow p(d_k|\theta) = \sum_{l=1}^{L} \pi_l \cdot p(d_k|c_l, \theta)$$

Incorporating the unlabeled data

Purpose: inferring the existence of unknown structures given the known part of the model and obtain more reliable estimators for the low-level statistics

$$D_{tot} = D_a \cup D_u^{known} \cup D_u^{unknown}$$

$$L(D_{tot},\theta) = \sum_{d_k \in D_a} E_{H_l^k}^{l \in C_d}\left[\log p_a(d_k, H_l^k|\theta)|d_k\right] \quad D_a$$

$$+ \sum_{d_k \in D_u^{known}} E_{H_l^k}^{l \in C_d}\left[\log p_u(d_k, H_l^k|\phi^d)|d_k\right] \quad D_u^{known}$$

$$+ \sum_{d_k \in D_u^{unknown}} E_{H_l^k}^{l \in C_{nd}}\left[\log p_u(d_k, H_l^k|\phi^{nd})|d_k\right] \quad D_u^{unknown}$$

- Use of a three-part log-likelihood objective function optimized with a modified EM algorithm

- Addition of a decision step between the expectation and maximization steps in order to decide to which part („known" or „unknown") of the model an „unlabeled training sample" will contribute

→ Set $D_u^{known} = \varnothing$ and $D_u^{unknown} = D_u$

→ Iterate

- Step 1: E-step
- Step2: decision step: Set

$$D_u^{known} = \{d_k \in D_u \,|\, \mathcal{L}_{\phi^d}(d_k) > \beta\}$$

and $D_u^{unknown} = D_u \backslash D_u^{known}$

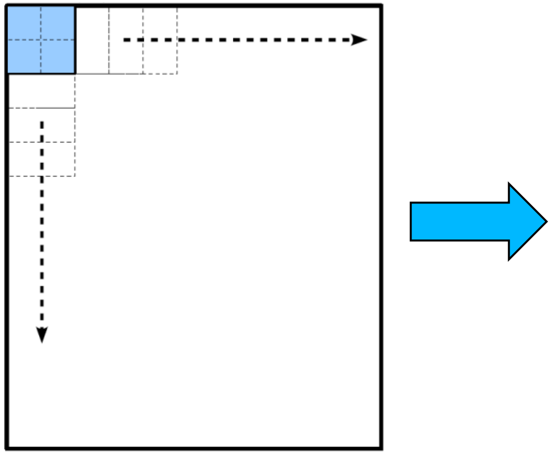
- Step 3: M step

Experimental results (1/2)

→ Database of 64 SPOT 5 panchromatic scenes size 3000 x 3000

→ Resolution of 2.5m

→ Classification is performed on small patches of size 64 x 64 extracted through the help a sliding window using overlapping

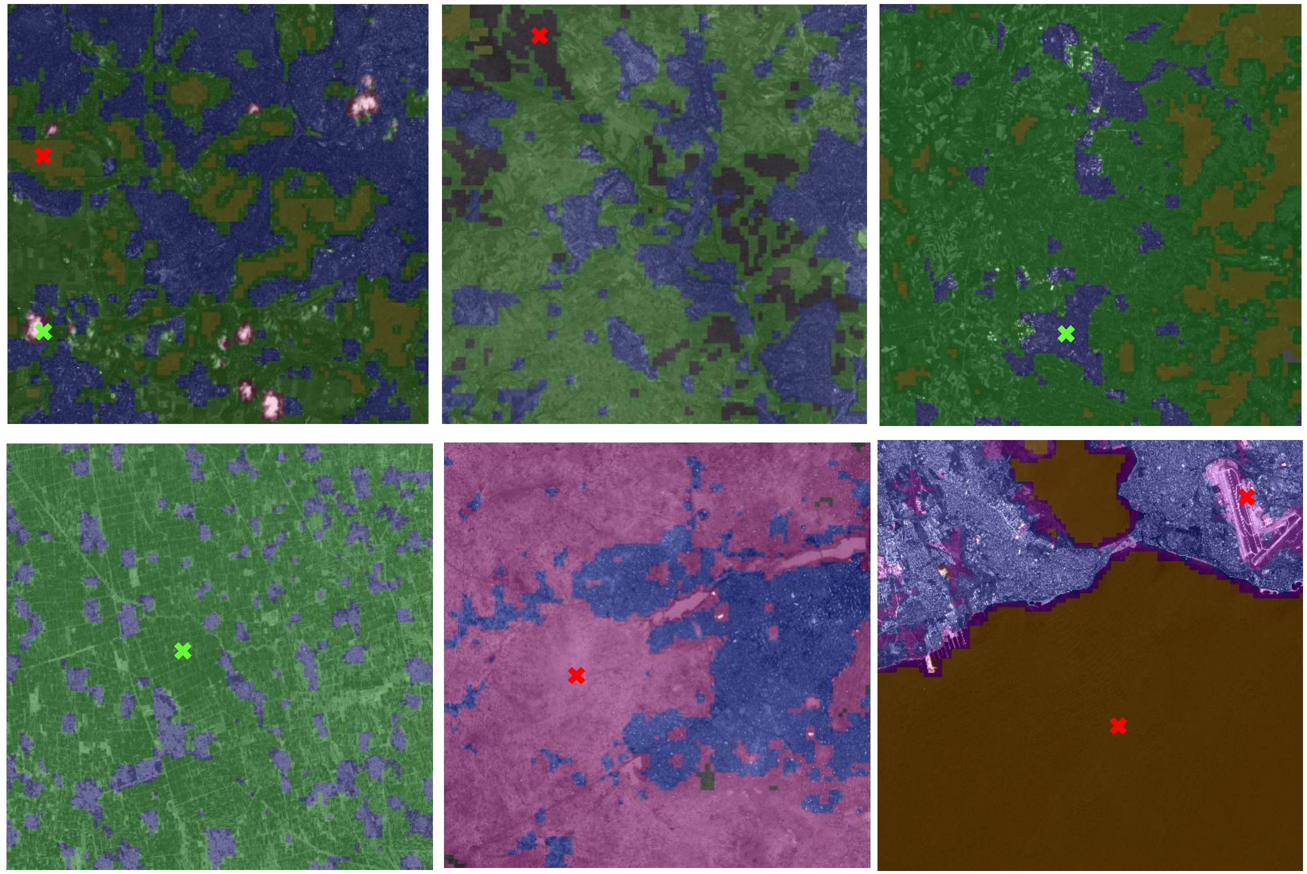


Over 600 000 patches are extracted over the whole database

Experimental results (2/2)

Unknown semantic structures discovered

Known image classes (training data)



	annotated		
truth	urban	fields	clouds
urban	**0.74**	0.17	0.09
fields	0.15	**0.77**	0.08
clouds	0.03	0.11	**0.86**

Semi-supervised SVM

	annotated				
truth	urban	fields	clouds	desert	sea
urban	**0.79**	0.07	0.05	0.06	0.03
fields	0.05	**0.81**	0.02	0.09	0.03
clouds	0.01	0.12	**0.8**	0.07	0
desert	0.06	0.13	0	**0.78**	0.03
sea	0.02	0.02	0.03	0.01	**0.92**

Proposed approach with unknown structures discovery

Outline

I. Semi-supervised auto-annotation in the context of non-exhaustive training datasets

II. Accelerated semi-supervised active learning in the framework of interactive image search engines

III. Interactive object detection in large satellite image repositories using a cascaded active learning scheme

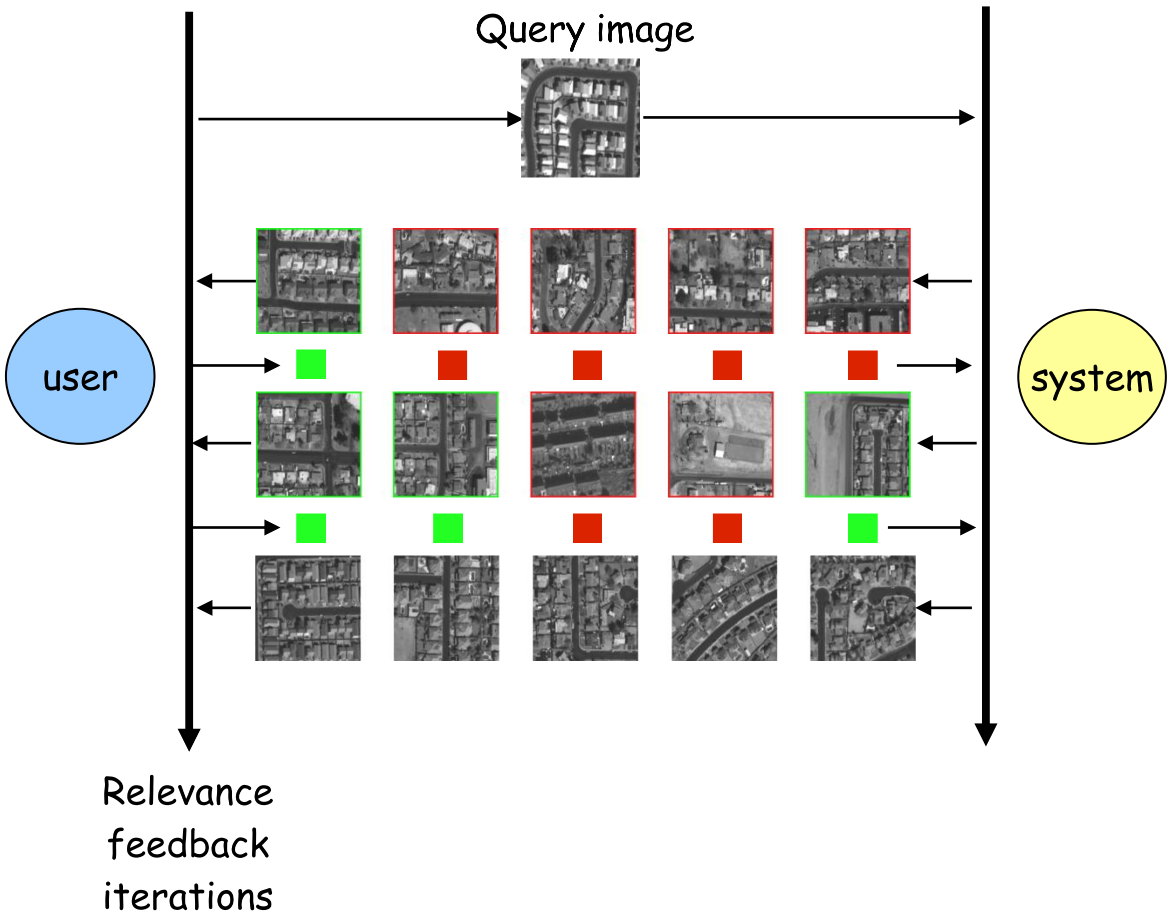
IV. Conclusion and perspectives

Accelerated active learning using the data distribution

Interactive image search engines
try to achieve two goals:

→ learn the targeted image category as accurately and as exhaustively as possible

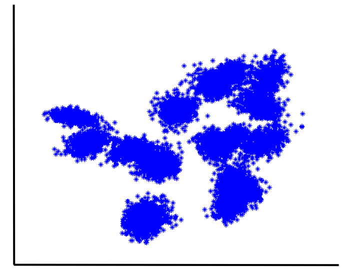
→ and also, as fast as possible → we focus on this second goal in the following



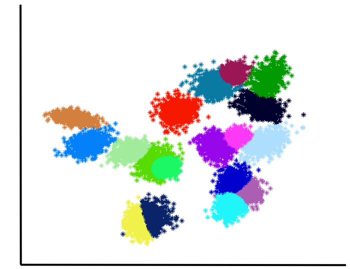
Proposed concept (1/4)

Concept:

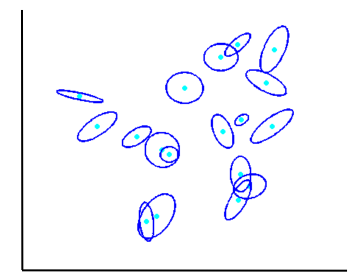
→ **Changing the granularity of the description space** and exploiting the inner structure of the data → instead of working at point level, we work on bigger entities such as Gaussian mixture components



Descriptors extracted from images



clustering/GMM parameters estimation



Change of granularity

→ Structuring of the descriptor space under the form of a mixture of Gaussian components

$$p\left(\nu; \{\pi_l, \mu_l, \Sigma_l\}_{l=1,\ldots,L}\right) = \sum_{l=1}^{L} \pi_l \cdot \mathcal{N}(\nu; \mu_l, \Sigma_l)$$

Structuring of the input space / estimation of the parameters of a Gaussian mixture model

→ Gaussian mixture modeling, restrictive assumption ?

Proposed concept (2/4)

Enabling hypothesis:

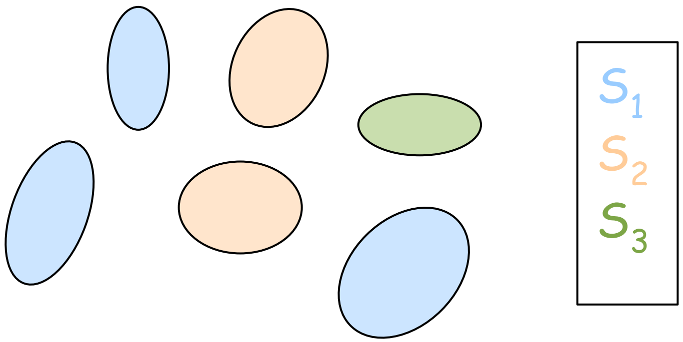
→ Classes can be roughly approximated by a union of Gaussian „clusters“

→ Simple associative model between mixture components and semantic concepts

→ Use in a first approach of binary associations between the mixture components c_l and the targeted concept $S \rightarrow p(S|c_l) = \delta_l^S$

→ Problems of the binary associative model:

- ✓ It is not discriminative enough
- ✓ Gaussian assumption never completely true



$$p(S|\nu;\theta) = \sum_{l=1}^{L} p(S, c_l|\nu;\theta) = \frac{1}{p(\nu;\theta)} \sum_{l=1}^{L} p(S|c_l;\nu;\theta) \cdot p(c_l;\theta) \cdot p(\nu|c_l;\theta)$$
$$= \frac{1}{p(\nu;\theta)} \sum_{l=1}^{L} p(S|c_l) \cdot \pi_l \cdot \mathcal{N}(\nu;\mu_l,\Sigma_l)$$

S : targeted category
ν : descriptor associated with an image
θ : mixture model parameters

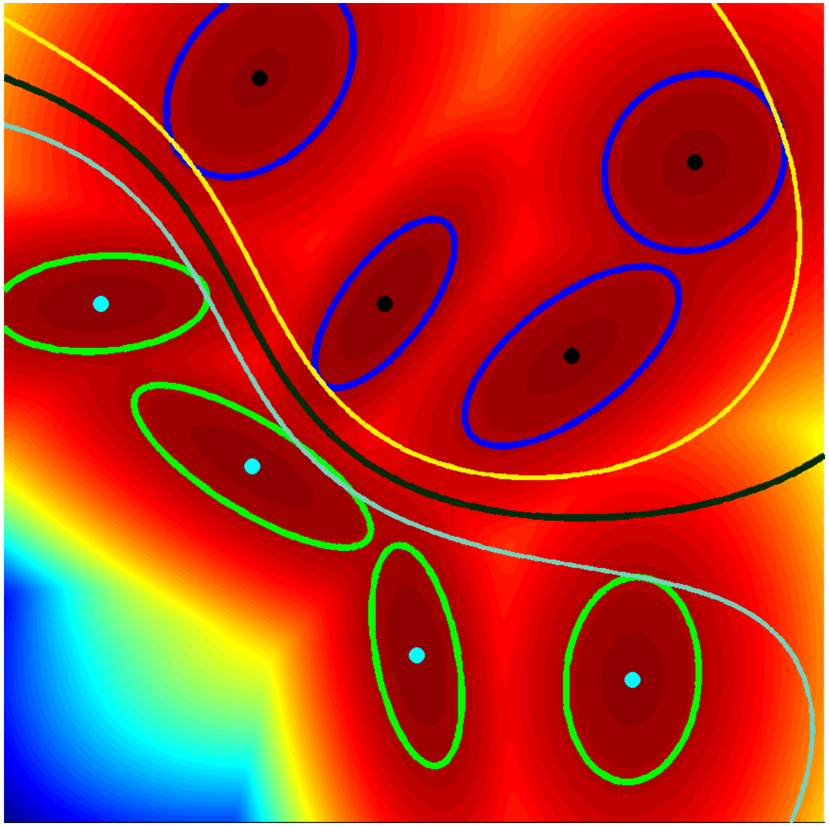
Proposed concept (3/4)

Concept:

→ Exploiting the data intrinsic distribution to speed up the learning of a discriminative classifier → semi-supervised paradigm

→ Objective: design of a method which respects the two fundamental assumptions of semi-supervised learning:

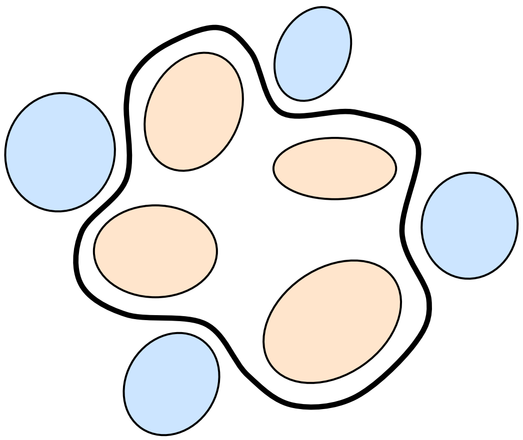
- ✓ **Low-density separation**
- ✓ **Intra-cluster coherence (cluster assumption)**



Proposed concept (4/4)

→ Component-based SVM working directly on the Gaussian mixture components

Idea: using the simple binary associative model as a first approximation of the SVM separation surface



➡ SVM surface approximating the „equi-likelihood" surface:

$$\{ v \,;\, p(S|v, \theta) = p(\bar{S}|v, \theta) \}$$

Solving the component-based SVM problem (1/3)

→ Naïve formulation:

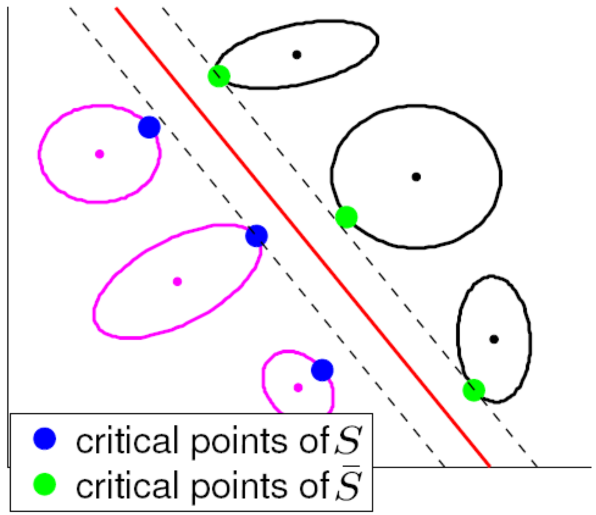
$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \begin{cases} \text{sign}(\mathscr{L}_p(v_i) - \mathscr{L}_{np}(v_i)) \cdot (w \cdot \phi(v_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0, \forall i = 1, \ldots, N \end{cases}$$

→ Problem of this formulation:
all the database points are used for the training

→ Solution: identifying points belonging to the mixture components such as the SVM trained with these points has the smallest possible margin

→ notion of critical points

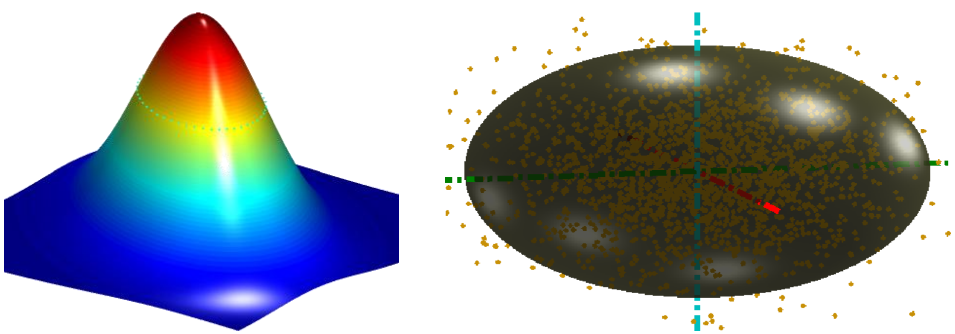


Solving the component-based SVM problem (2/3)

Step 1: defining hard delineations for the mixture components

The equiprobable envelope of a Gaussian mixture component c_l is an ellipsoid of equation:
$(v_l - \mu_l)^T \ \Sigma_l^{-1} \ (v_l - \mu_l) = \rho_l$

→ A mixture component is a probabilistic notion → need to define a „hard delineation" for each component



→ the constants ρ_l control the size of the envelopes → they are adjusted during the active learning process

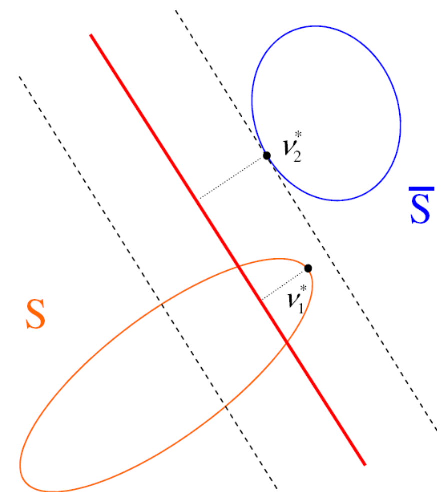
→ component-based SVM formulation:

$$\max_{v_1^*,\dots,v_L^*} \min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum_{l=1}^{L}\xi_l$$

$$\text{s.t.} \begin{cases} (2\delta_l^S - 1)\cdot(w\cdot\phi(v_l^*) + b) \geq 1 - \xi_l \\ \xi_l \geq 0, \forall l = 1,\dots,L \\ (v_l^* - \mu_l)^T \Sigma_l^{-1}(v_l^* - \mu_l) \leq \rho_2^l, \forall l = 1,\dots,L \end{cases}$$

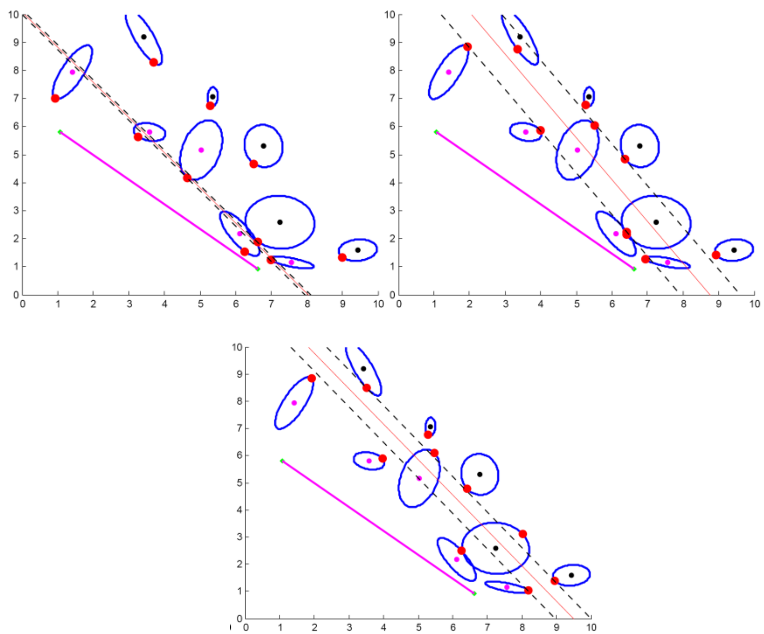
Solving the component-based SVM problem (3/3)

Step 2: solving the „max min" optimization problem



→ the critical points are defined as the points of the SVM surface belonging to the convex enveloppes of mixture components which are the closest of the SVM separating surface (separable case) or the farthest on the „wrong" side of the SVM surface (non separable case)

→ iterative alternating maximization minimization scheme:



→ iterating between the following two steps until convergence

- Computing the SVM surface with the current set of critical points
- Recomputing the critical points as the points closest to the SVM surface

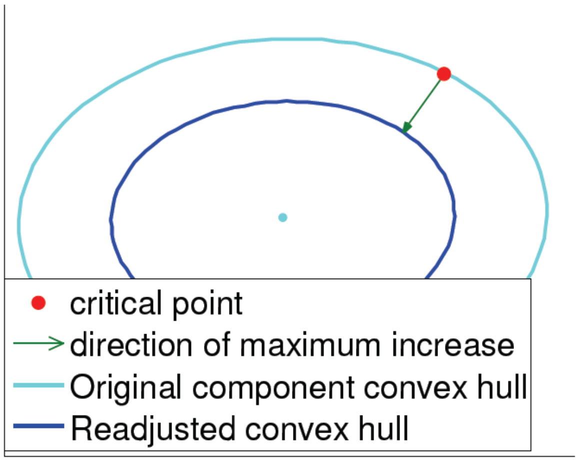
Integration into an active learning scheme (1/3)

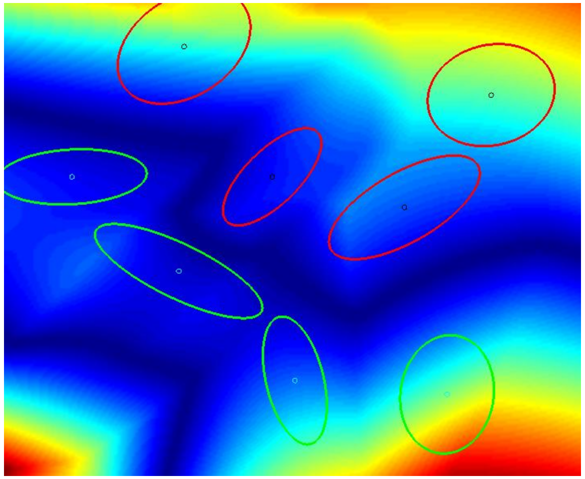
Idea:

→ Interactive readjustment of the mixture components envelopes by exploiting the user feedback in each critical point:

→ negative feedback => we reduce the size of the corresponding envelope
→ positive feedback => we increase the size of the corresponding envelope

→ Use of the binary associative model to perform the envelopes readjustement
→ the ratio $p(S|v, \theta)/p(\bar{S}|v, \theta) = L_p(v)/L_{np}(v)$ is used to define the direction in which to perform the readjustment
→ the readjustment step is fixed





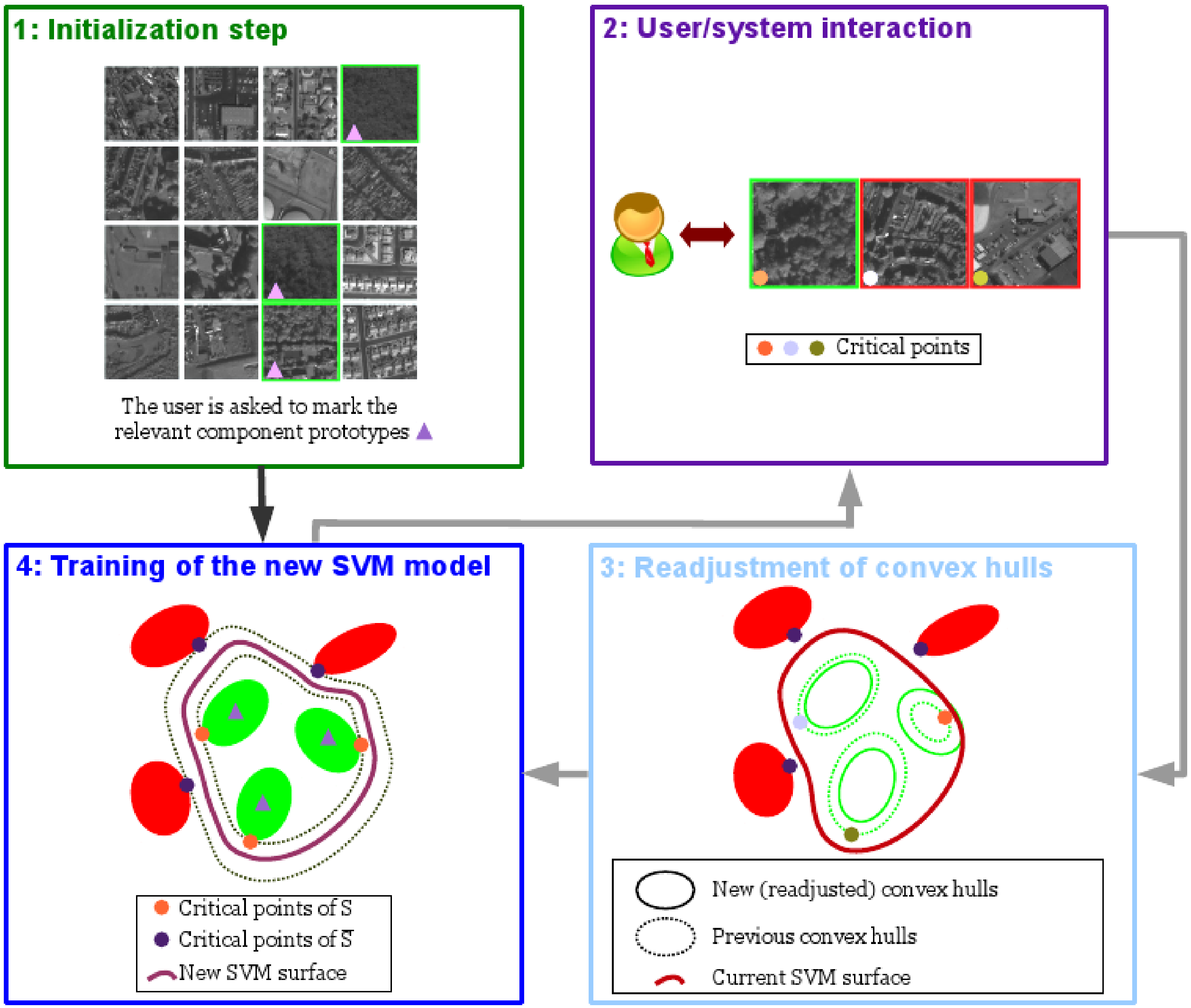
$$S_R = \frac{L_p}{L_{np}} + \frac{L_{np}}{L_p}$$

Integration into an active learning scheme (2/3)

Batch learning approach

→ Relevant mixture components are tagged all at once at the beginning of the active learning process

→ Not a realistic scenario → we want to be able to add relevant mixture components progressively during the interactive process

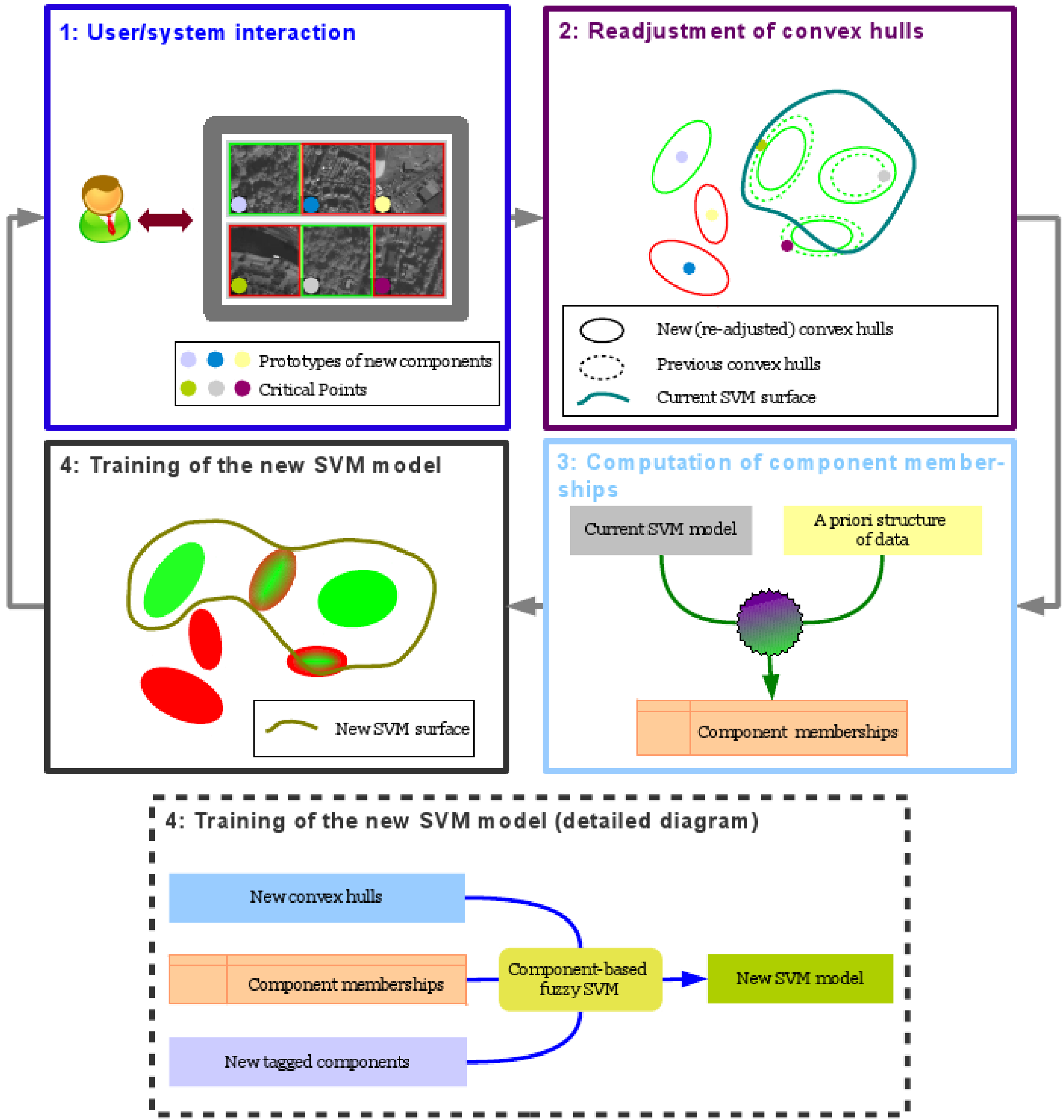


Integration into an active learning scheme (3/3)

Online learning approach

→ Relevant mixture components can be introduced progressively during the interactive learning process

→ Presence of an unlearning feature which allows to forget mixture components which have erroneously tagged as relevant by the user during previous learning iterations



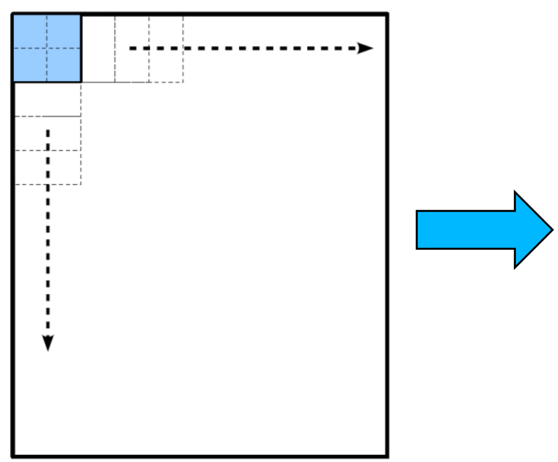
Results on QuickBird images (1/2)

- Database of 10 QuickBird panchromatic scenes of approximate size 30 000 x 30 000

- Resolution of 60cm

- Classification is performed on small patches of size 200 x 200 extracted through the help a sliding window using overlapping

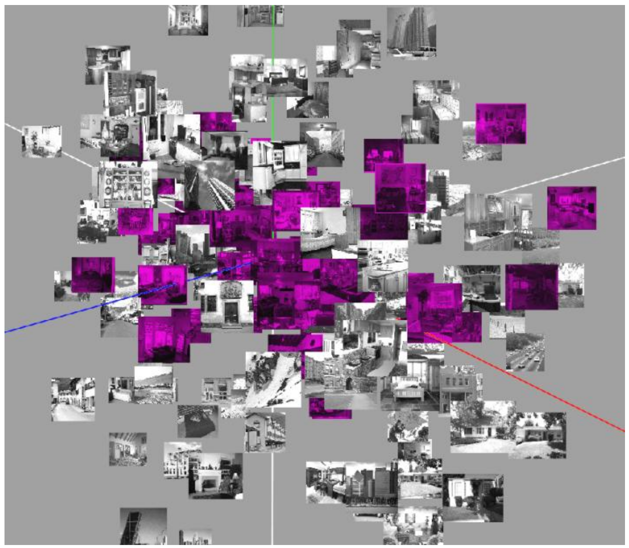
- What is the optimal patch size ?

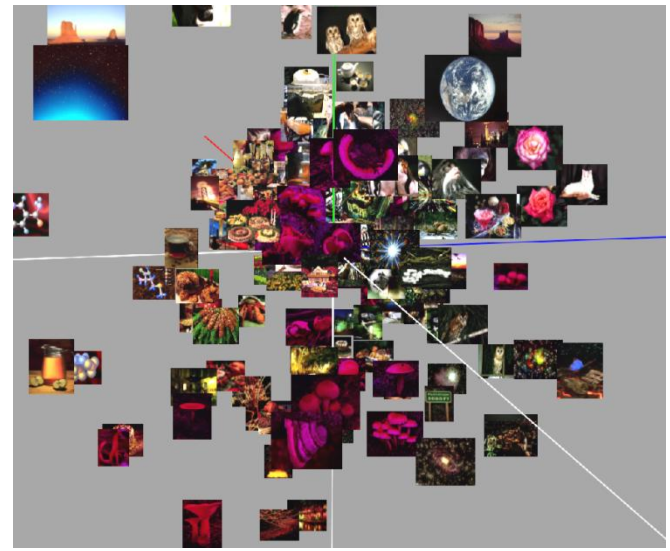


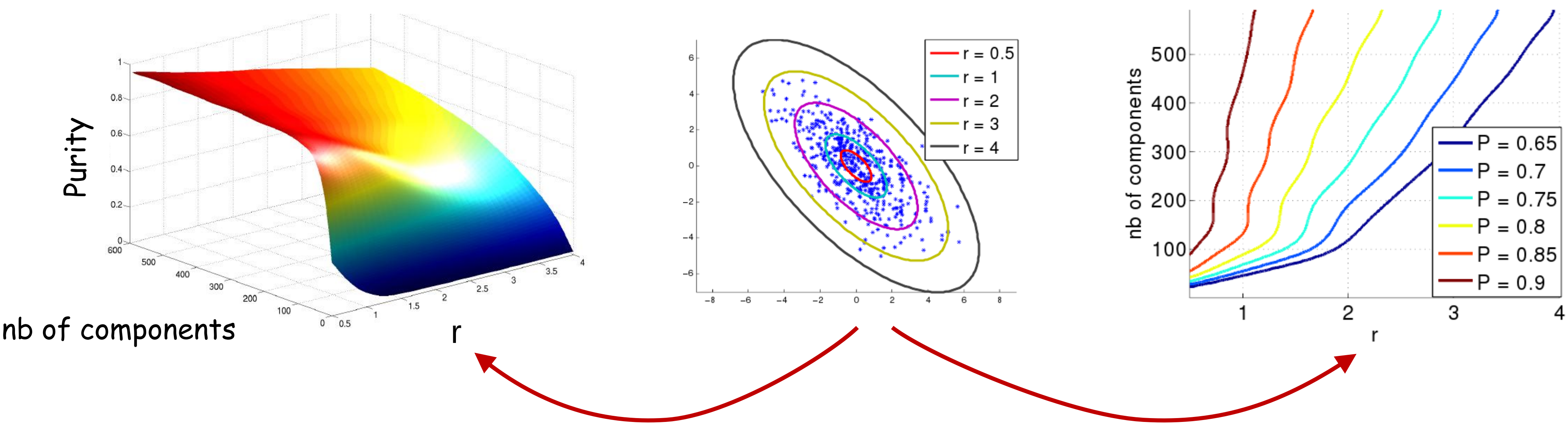
Over 1E6 patches are extracted over the whole database

Verifying the enabling assumption

Assumption: Gaussian clusters retain some semantic consistency near their center

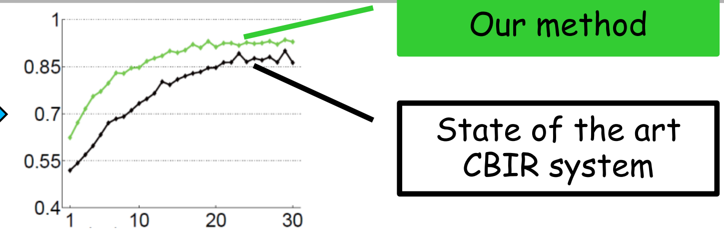




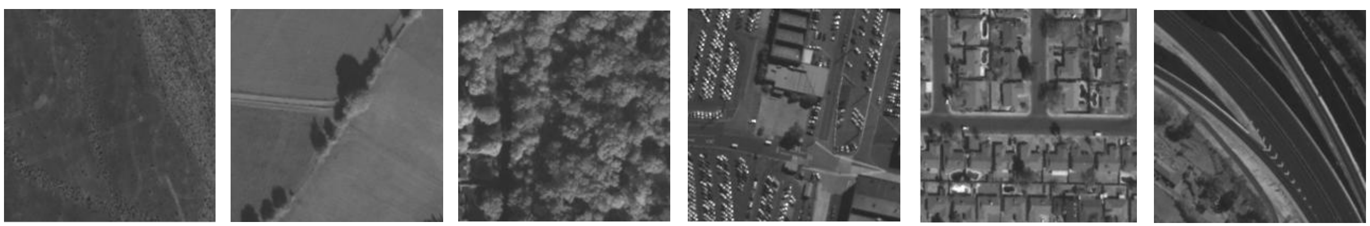


Results on QuickBird images (2/2)

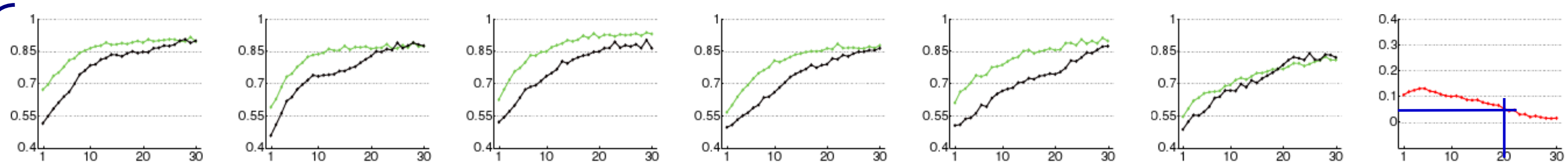
- precision
- recall



Nb of active learning iterations

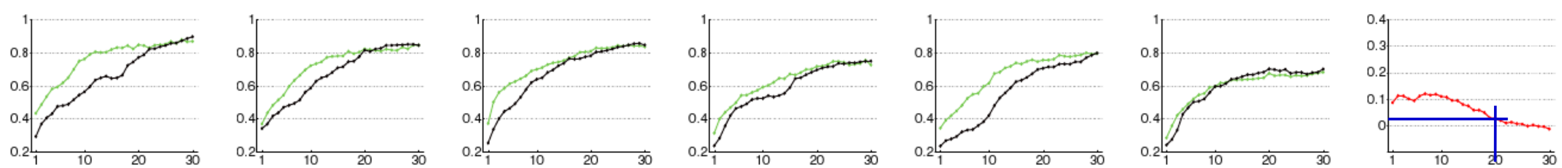


precision



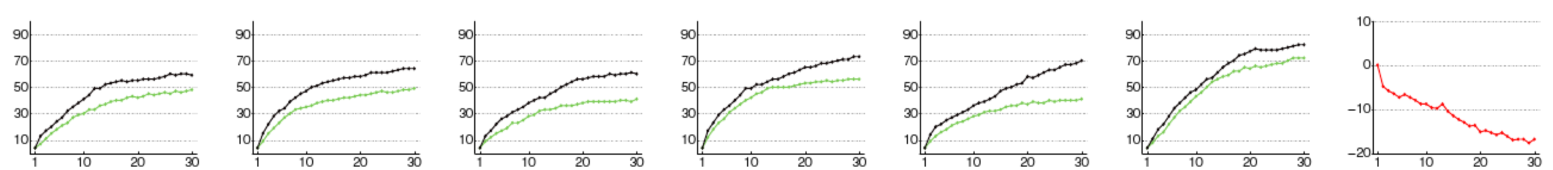
(a) desert (b) fields (c) forest (d) park. lots (e) res. area (f) roads (g) avg diff.

recall



(h) desert (i) fields (j) forest (k) park. lots (l) res. area (m) roads (n) avg diff.

Nb of SVs



(o) desert (p) fields (q) forest (r) park. lots (s) res. area (t) roads (u) avg diff.

We reduce by 30% the number of iterations needed to reach the same level of precision and recall

Outline

I. Semi-supervised auto-annotation in the context of non-exhaustive training datasets

II. Accelerated semi-supervised active learning in the framework of interactive image search engines

III. Interactive object detection in large satellite image repositories using a cascaded active learning scheme

IV. Conclusion and perspectives

Coarse-to-fine strategy and cascaded learning (1/3)

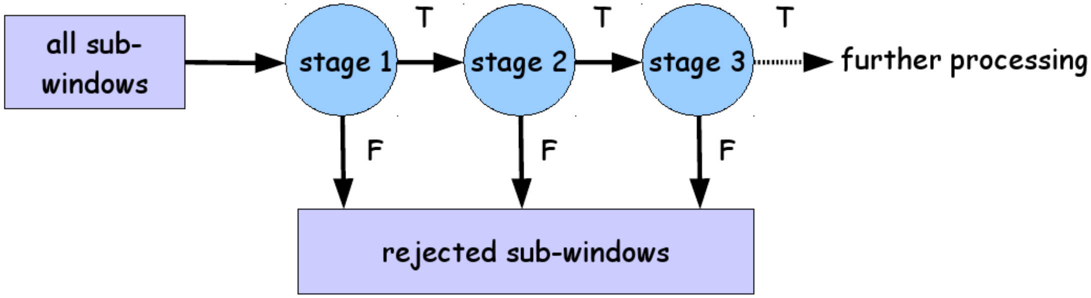
- Objects are very rare events in the images

- Use of a coarse-to-fine strategy to perform object retrieval

Goals:

- Eliminating as many subwindows as possible in the highest levels of the cascade to focus on a reduced number of subwindows in the lowest levels

- Applying more computationally expensive processing on the remaining subwindows



A subwindow is propagated to the next level of the cascade until it gets rejected

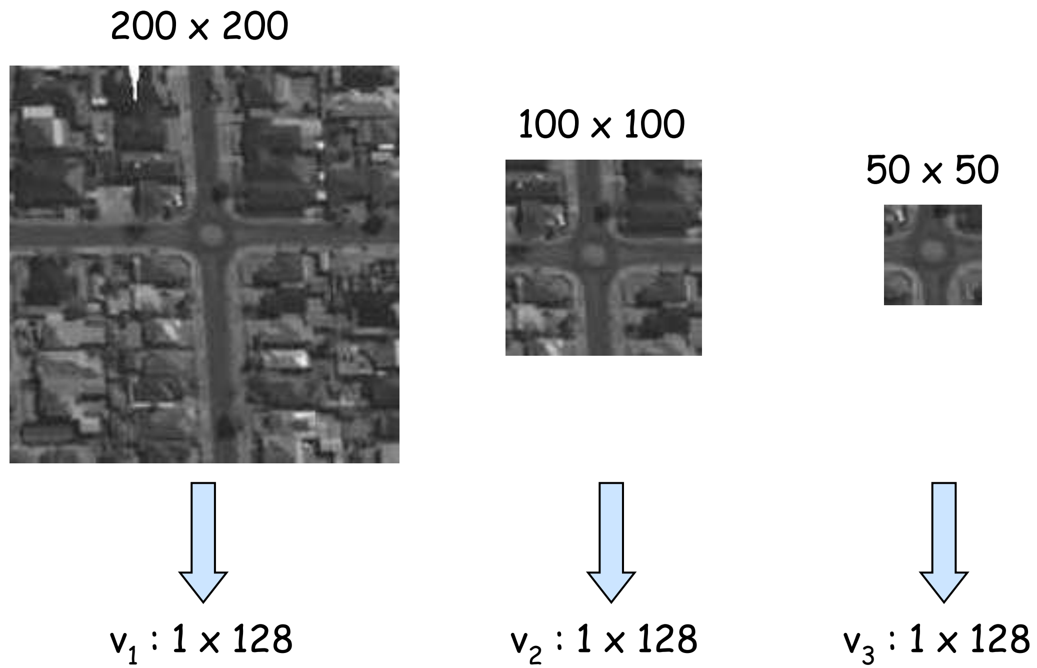
Coarse-to-fine strategy and cascaded learning (2/2)

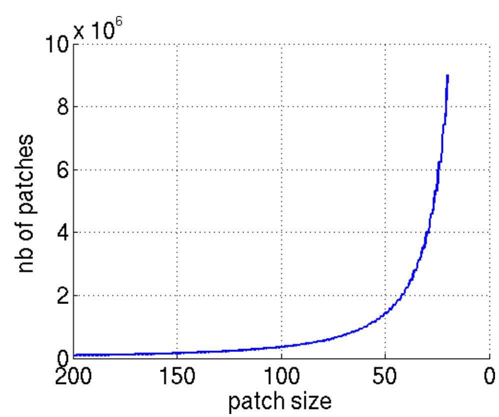
- At each stage of the cascade, we diminish the size of the patch in the patch-based representation of images

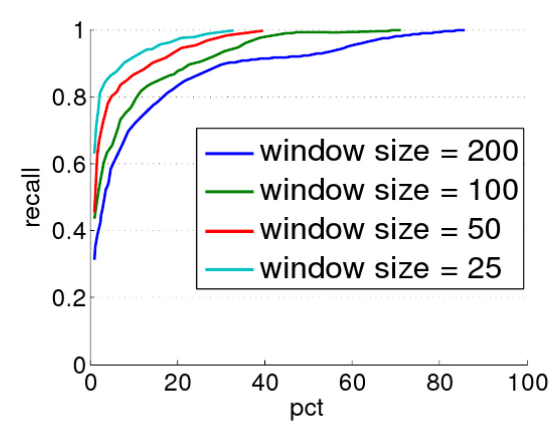
- Purpose: better capturing the properties of the object

- But at the expense of an exponential growth in the number of patches to process

- Interesting observation: at level 100, we can safely discard around 70% of the database, still ensuring a recall of 90%







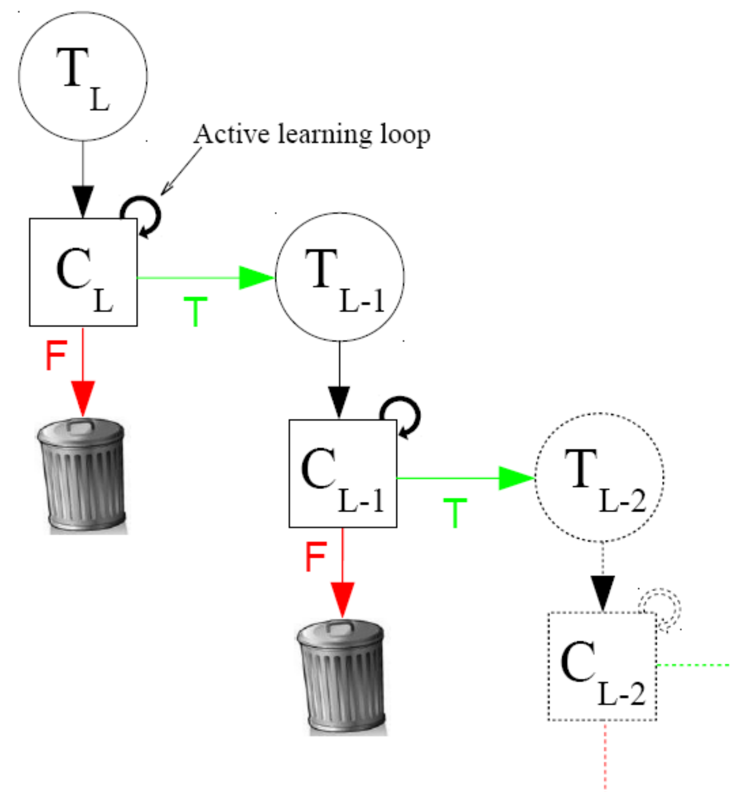
Object recognition using active learning (2/3)

- At each level of the cascade, we use a soft SVM classifier with a strategy to ensure sparsity of feedback samples

$$\arg \max_{i_1,\ldots,i_D \in S} \min_{\substack{(j_1,j_2)\in\{i_1,\ldots,i_D\} \\ \text{with } j_1<j_2}} d(v_{j_1}, v_{j_2})$$

→ we look for the D elements from the pool of unlabeled data whose minimum pairwise distance is maximum

- If at any stage in the cascade a classifier rejects the patch under inspection, no further processing is performed on this patch



$$\text{Detection rate} = \prod_{l=1}^{L} d_l$$

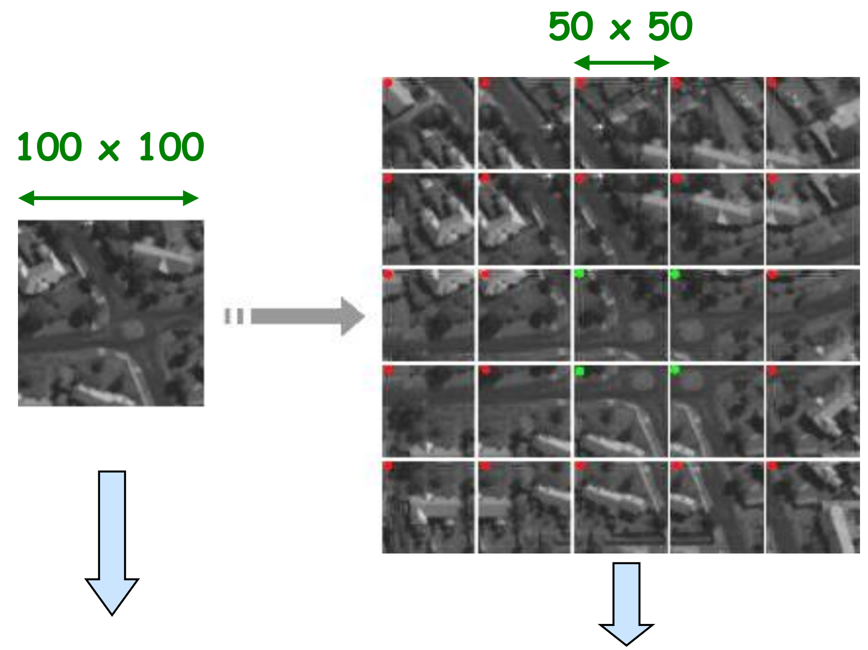
$$\text{False positive rate} = \prod_{l=1}^{L} f_l \longrightarrow 0$$

Object recognition using active learning (3/3)

Main difficulty:

- Propagating feedback examples from one level of the hierarchy to the other

- The problem can be formulated as a Multiple Instance Learning problem in which the purpose is to retrieve the positive examples from mixed bags of positive and negative examples



Positively tagged patch at level „100"

Intersecting patches at the level below (50 x 50)

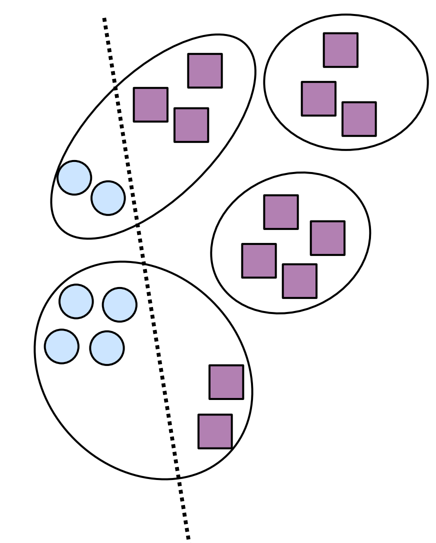
Multiple Instance Learning (MIL)

- There is an uncertainty on the labels of training instances

- Training data is available under the form of bags of instances with labels for the bags

- Purpose: retrieving the positive elements inside the positive bags

Positive bags Negative bags



- The maximum margin formulation of MIL leads to a combinatorial problem → we need to find a suboptimal way to solve the problem in a polynomial time

$$\min_{\{y_i\}_{i=1,\dots,N}} \min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \begin{cases} y_i(\langle w, \phi(x_i)\rangle + b) \geq 1 - \xi_i,\ \forall i = 1,\dots,N \\ \xi_i \geq 0,\ \forall i = 1,\dots,N \\ \sum_{j\in I_m} \frac{y_j+1}{2} \geq 1 \ \ \forall m \text{ s.t. } Y_m = 1 \\ y_j = -1 \ \ \forall j \in I_m \ \ \forall m \text{ s.t. } Y_m = -1 \end{cases}$$

MIL-SVM problem

Cascaded active learning and interlevel propagation of feedback examples (1/3)

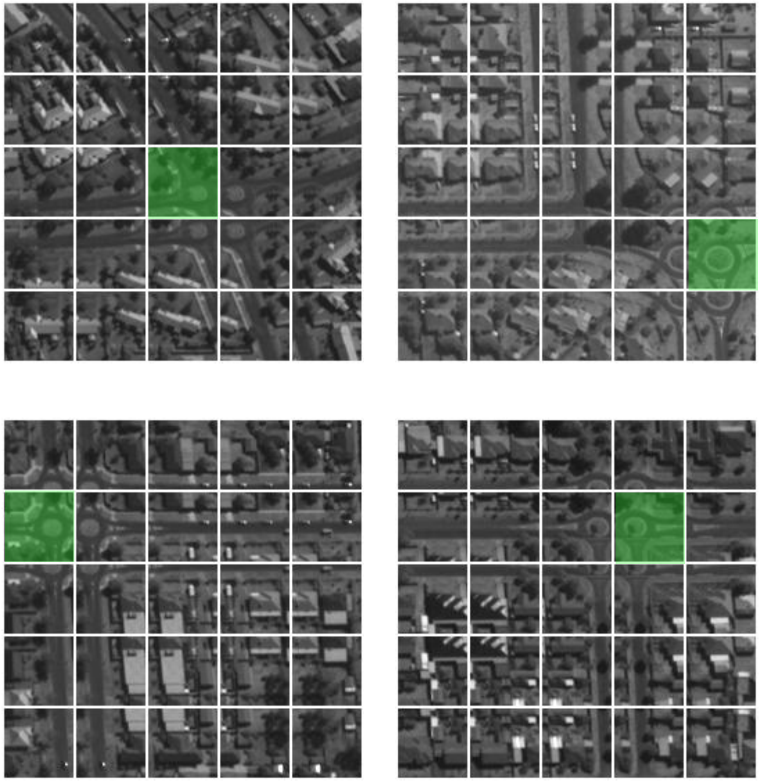
Positioning of the problem:

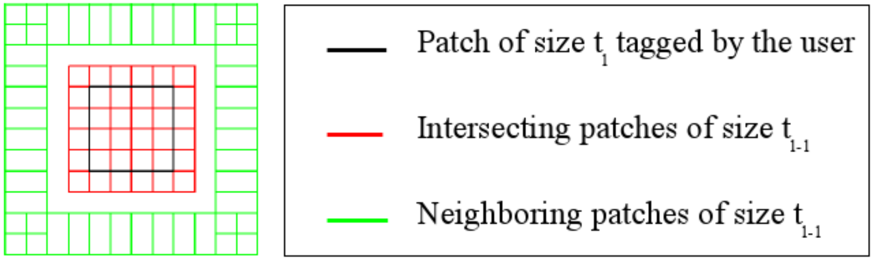
- We want to retrieve one element per positive bag

- Discarding all the negative bags

- Intuitively, we would like the negative elements of the negative bags to belong to the same class as the negative elements of the positive bags (principle of *noise cancellation*)

- Use of the context of each positive bag to do so



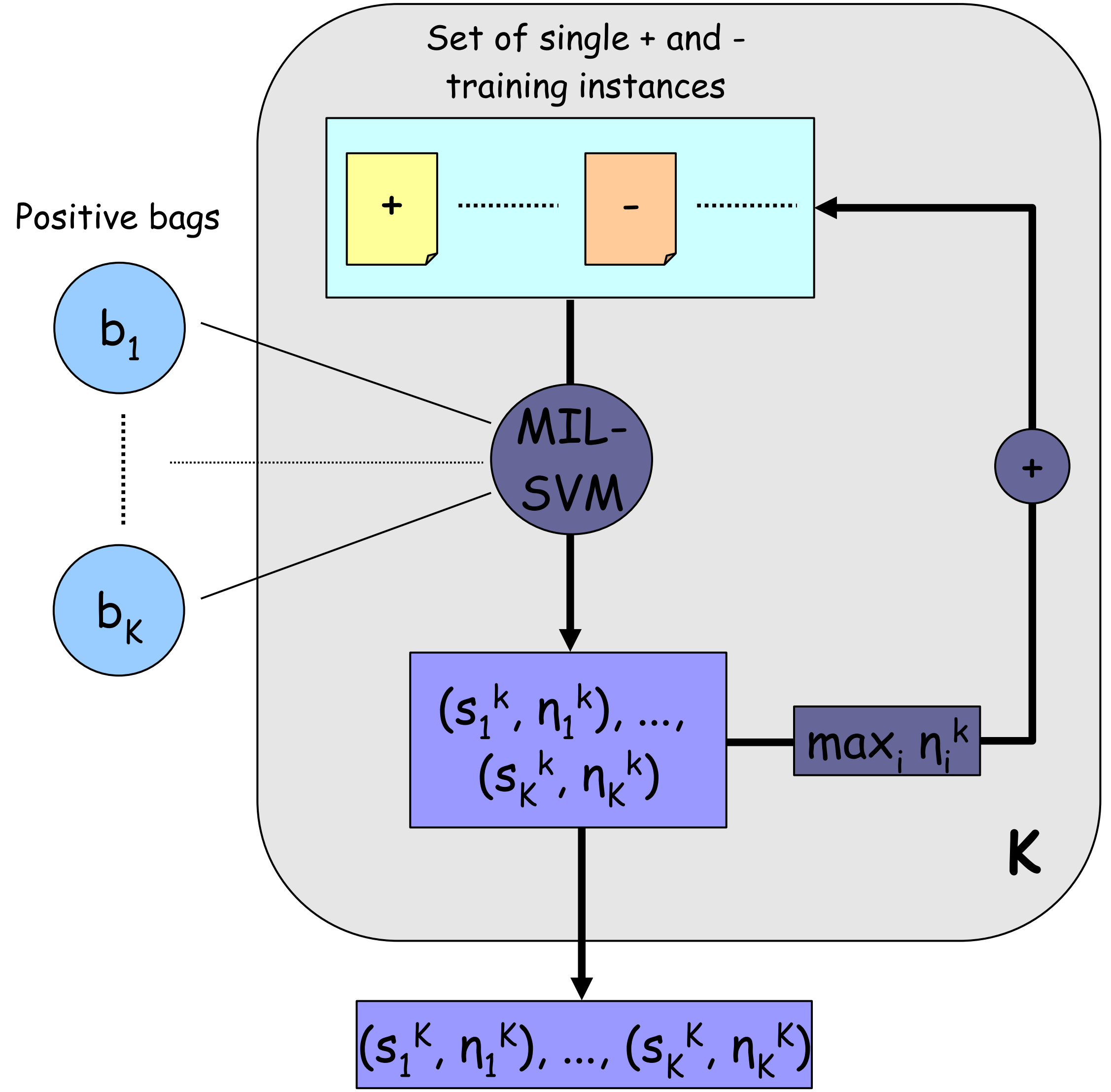


Cascaded active learning and interlevel propagation of feedback examples (2/3)

Solving the MIL problem with K-positive bags:

- Procedure consisting in solving K separate MIL-SVM problems and setting in common the obtained solutions to increase the learning performance

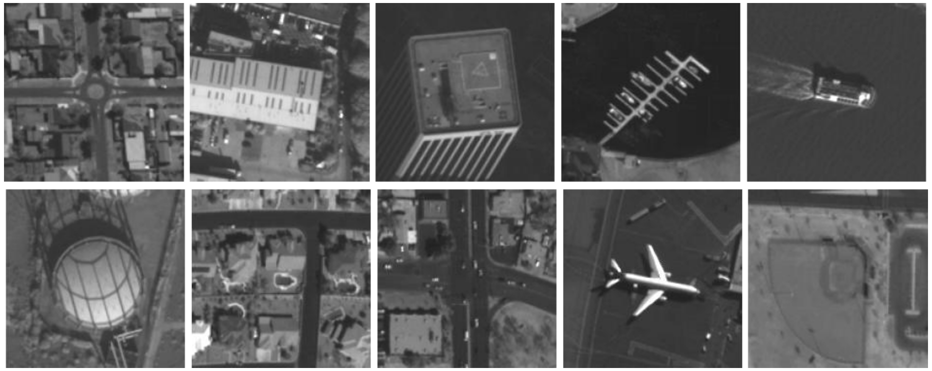
- Backtracking-like principle



Synopsis of the proposed MIL-SVM algorithm

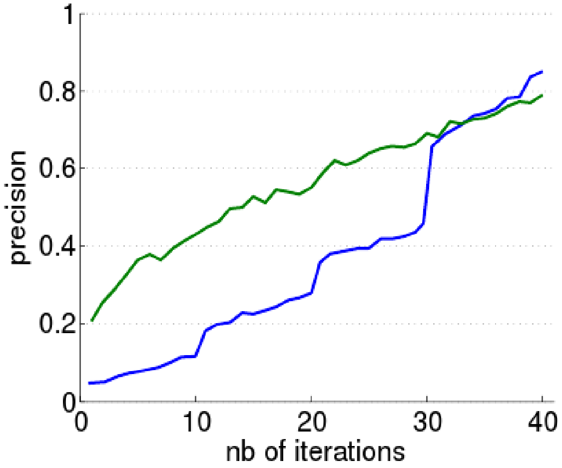
Object recognition using active learning: results (1/2)

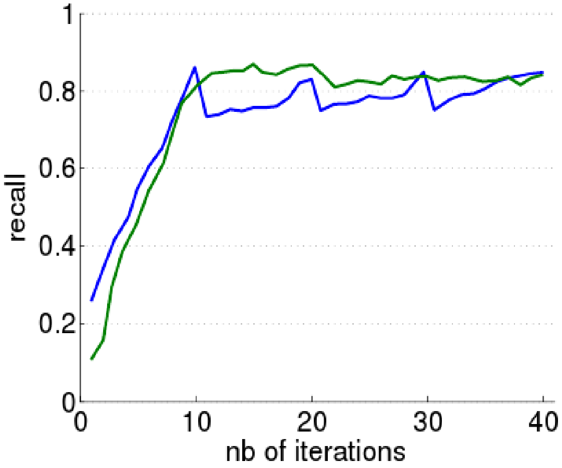
Results on ten classes of objects:



- Roundabouts, storehouses, buildings, marina, boats, gas holders, swimming pools, crossroads, planes, baseball grounds

- The precision and the recall are approximately the same at the end of the active learning process



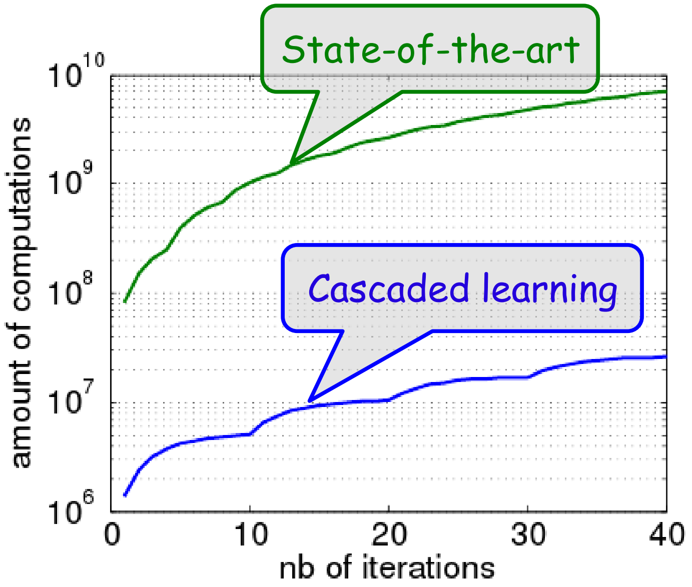


Object recognition using active learning: results (2/2)

• But: decrease in the number of evaluation of the classifier decision function by two orders of magnitude in average

→ Very important result in a context where the „fluidity" of the interactions between the user and the system is a crucial issue

The number of evaluation of the decision function is reduced by two orders of magnitude !!



Outline

I. Semi-supervised auto-annotation in the context of non-exhaustive training datasets

II. Accelerated semi-supervised active learning in the framework of interactive image search engines

III. Interactive object detection in large satellite image repositories using a cascaded active learning scheme

IV. Conclusion and perspectives

Conclusion, discussion and perspectives (1/4)

Summary of our main contributions:

(1) Semi-supervised algorithm to perform auto-annotation and „unknown semantic structures discovery" in satellite image databases

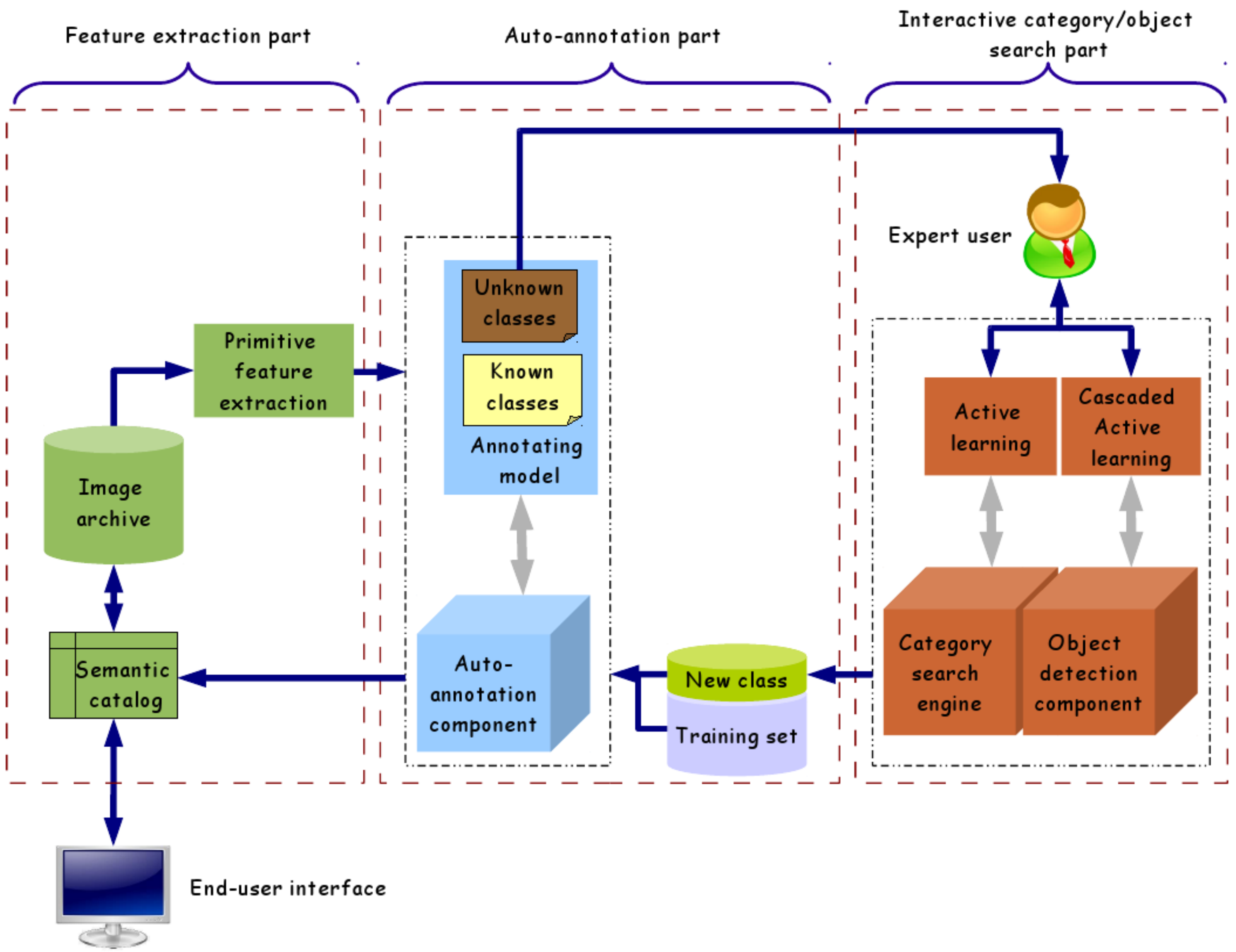
(2) Semi-supervised active learning algorithm to speed up the learning of the target image category in the framework of interactive image searc engines

(3) Object retrieval scheme using a cascaded active learning strategy.

(4) Our contributions were mainly tested on optical high-resolution satellite images. The results demonstrate the usefulness of our methods in the context of high-volume satellite image databases and very small/non-existent training sets

Concept for an image mining system exploiting the complementarity of auto-annotation and category search paradigms

- the category/object search engine is used to build the training database of the auto-annotation system
- the auto-annotation system is used to suggest prototypes of new categories to be searched for. It also provides an annotation model using natural keywords



Conclusion, discussion and perspectives (2/4)

Genericity of the proposed methods:

- Our first contribution dealing with the unknown structures discovery is well suited for low-resolution satellite images but not for high-resolution images and complex classes of objects

- In our second contribution, the proposed method is generic regarding the type of data → results demonstrated on a subset of the Corel dataset

- The third contribution could be generalized as well to other types of image data and other kind of problems such as face detection

- The descripors we use differ for each dataset but the methods we propose are independent of the type of signature used (the aim being to confer to our solutions the maximum generalizability over the type of data)

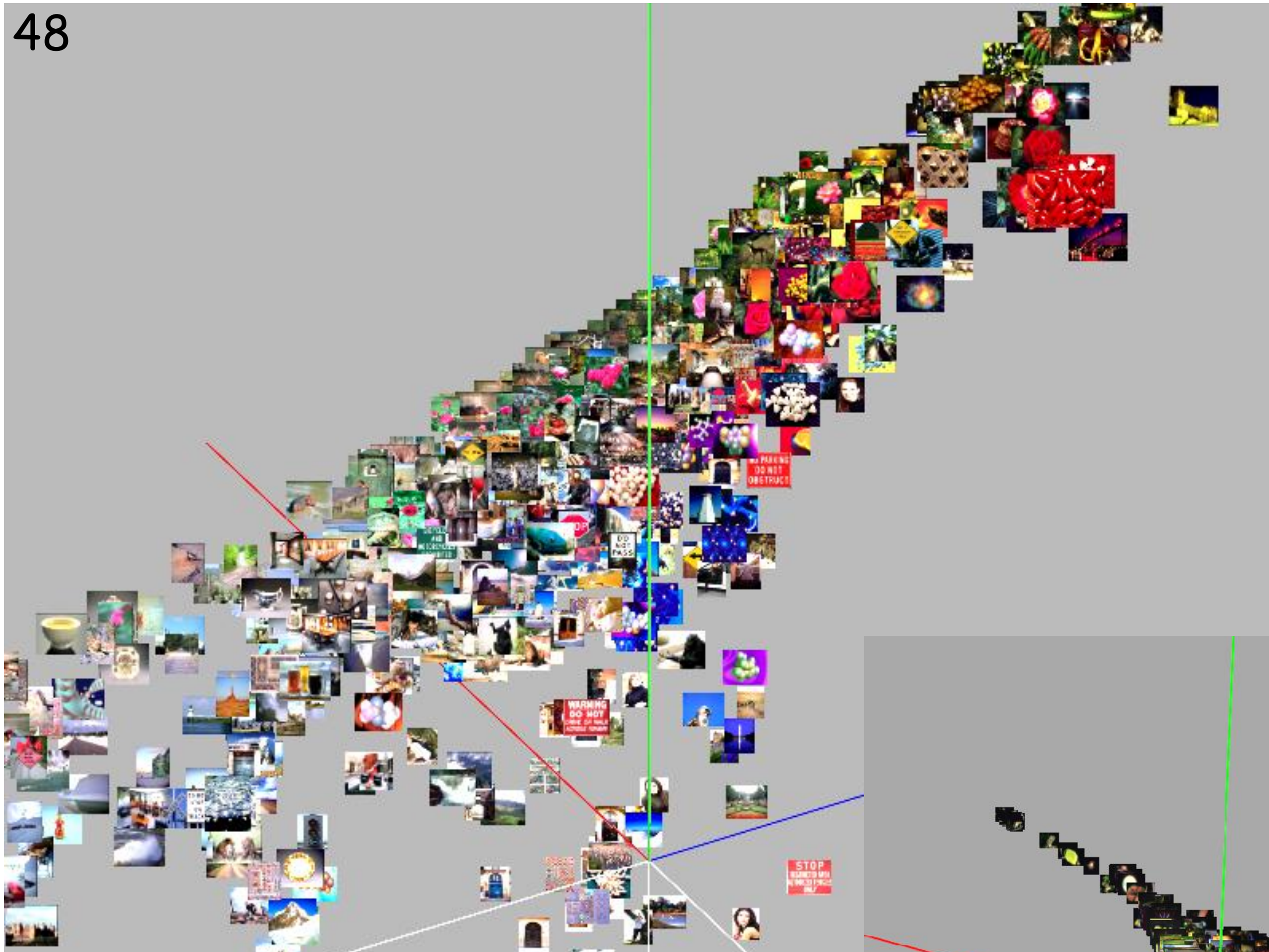
- Reusable concepts for other problems (MIL → error-prone training datasets …)

Conclusion, discussion and perspectives (3/4)

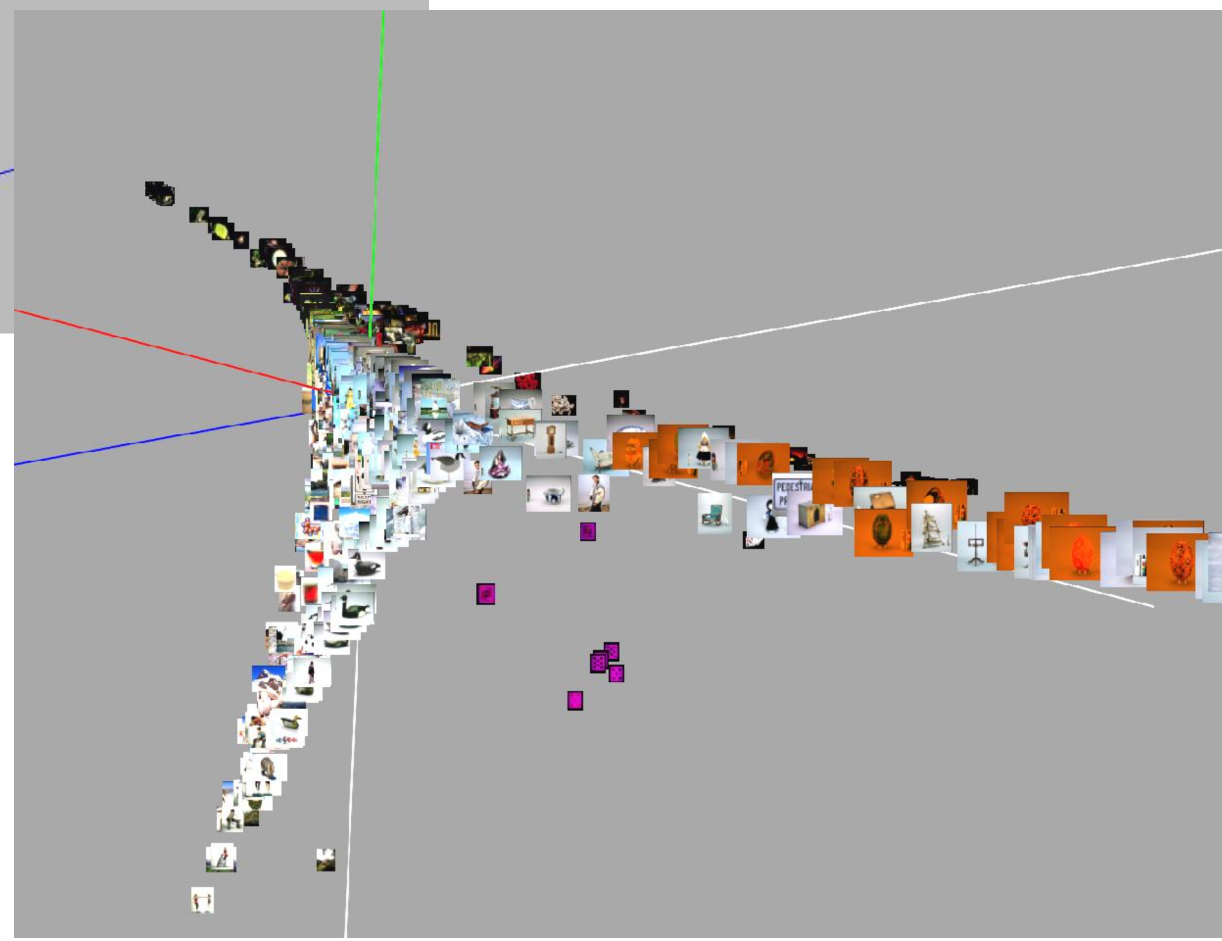
- Use of database technologies and of space-partitioning data structures to make the proposed algorithms scalable to large image databases:
  - In our second contribution: implementing the structuring of the input space under the form of ellipsoidal convex hulls using a space-partitioning data structure
  - In our third contribution: implementing the idea of hierarchy based on the patch size using quadtrees
  - The final goal being to implement the proposed procedures inside a geographical information sytem operating on thousands of terabytes of data

Conclusion, discussion and perspectives (4/4)

- Use of visual data mining techniques to speed up the learning of the target class in the case of interactive image search engines:
  - Use of dimensionality reduction techniques such as manifold learning algorithms to provide graphical representations of image databases in the purpose of performing active learning
  - The general idea is to replace standard user interfaces which have shown some limitations regarding the representation of the semantic content of image databases



Representation of the data in the 3D space using a Laplacian eigenmap



Semantically consistent groupings appear inside the data

Publications

- P. Blanchart and M. Datcu. *A Semi-Supervised Algorithm for Auto-Annotation and Unknown Structures Discovery in Satellite Image Databases*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3(4):698–717, 2010.

- P. Blanchart, M. Ferecatu, and M. Datcu. *Mining Image Databases with Adaptive Convex Hulls*. Submitted in Journal of Data Mining and Knowledge Discovery (DMKD), Springer, 2011.

- P. Blanchart, M. Ferecatu, and M. Datcu. *Object retrieval in large image databases using multiscale coarse-to-fine cascaded active learning*. Submitted in IEEE Transactions on Image Processing (TIP), 2011.

- P. Blanchart and M. Datcu. *Semi-supervised learning and discovery of unkown structures among data: Application to satellite image annotation*. In the proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2009.

- P. Blanchart, M. Ferecatu, and M. Datcu. *Apprentissage actif et utilisation de la structure a priori des données: application à une base d'images satellites haute-résolution*. In the proceedings of RFIA (Reconnaissance de Formes et Intelligence Artificielle), 2010.

- P. Blanchart, M. Ferecatu, and M. Datcu. *Active Learning Using the Data Distribution for Interactive Image Classification and Retrieval*. In the proceedings of the IEEE Symposium on Computer Intelligence and Data Mining, CIDM 2011.

- P. Blanchart, M. Ferecatu, and M. Datcu. *Indexation of Large Satellite Image Repositories Using Small Training Sets*. In the proceedings of ESA-EUSC-JRC 7-th Conference on Image Information Mining: Geospatial Intelligence from Earth Observation, JRC 2011.

- P. Blanchart, M. Ferecatu, and M. Datcu. *Mining large satellite image repositories using semi-supervised methods.* In the proceedings of IGARSS 2011.

- P. Blanchart,M. Ferecatu, and M. Datcu. *Cascaded Active Learning for Object Retrieval using Multiscale Coarse-to-fine Analysis*. In the proceedings of the IEEE Conference on Image Processing. ICIP 2011.

Cascaded active learning and interlevel propagation of feedback examples (3/3)

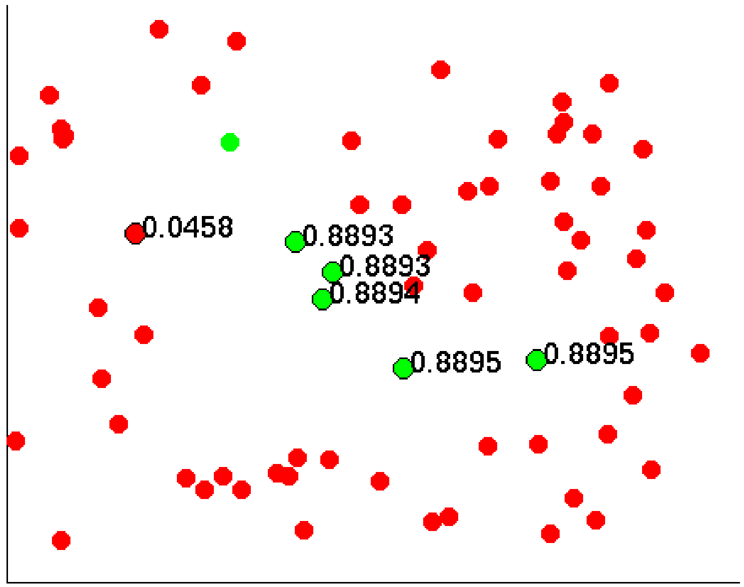
Computation of the confidence levels η_i^k:

- Must be performed in a completely unsupervised way

- We train a probabilistic SVM on the current training set at the iteration k of the MIL-procedure

- Use of a modified form of Platt's algorithm to compute the confidence levels associated with the current solutions of the K MIL-SVM problems

- Confidence levels obtained at the preceding iteration of the algorithm are taken as an out-of-sample model



$$\mathcal{L}_t(a,b) = -\sum_{i \mid v_i \in \{s_k^t\}_{k=1,\dots,K} \cup P} \left[t_i \log(p_i) + (1-t_i)\log(1-p_i) \right]$$

with

$$p_i = \frac{1}{1+\exp(a g_t(v_i)+b)}$$ and $t_i = \eta_i^k$

Platt's algorithm consists in finding the optimal (a, b) i.e. the couple (a, b) such as:

$$(a_t, b_t) = \mathrm{argmin}_{a,b} \mathcal{L}_t(a,b)$$

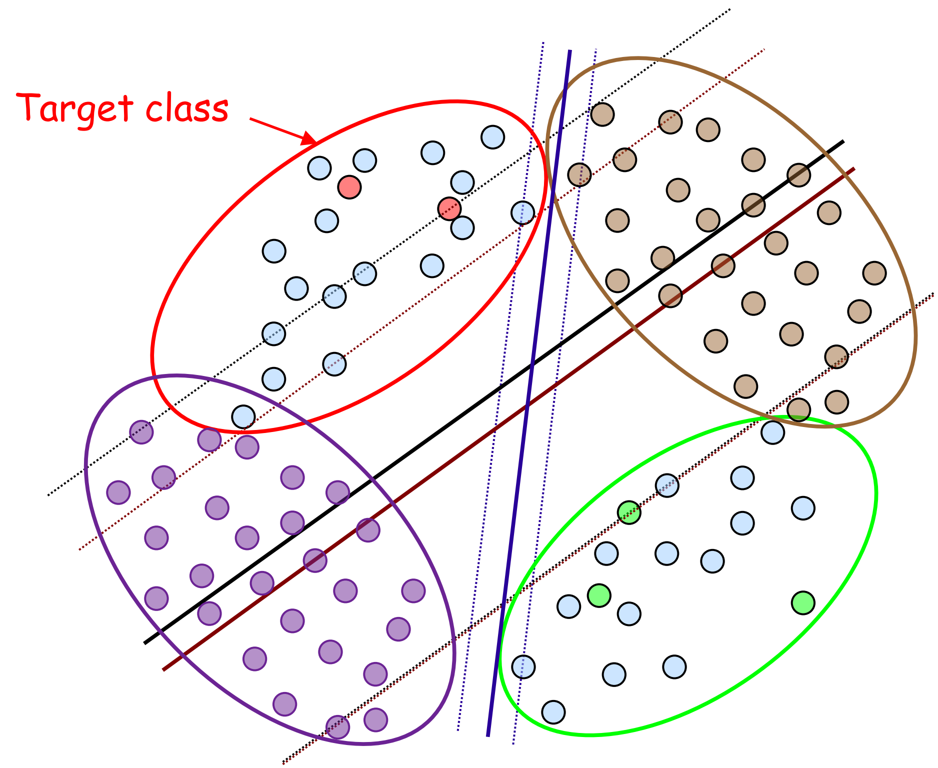
Positioning of the problem

Objective: exploiting the information from both the labeled and the unlabeled part of the database

→ Semi-supervised methods fit naturally inside this framework

→ Problem: avoiding the common assumption of semi-supervised methods i.e. considering that the distribution of unlabeled data fits that of labeled data: $p(x) \rightarrow p(y|x)$

→ not verified in the case of non-exhaustive training datasets



Positioning of the problem

Objective: designing an algorithm able to:

→ Exploit the information in the unlabeled data since the labeled data are only representative of a very small part of the semantic diversity inside the images

→ Adapt to non exhaustive training datasets which do not grant the distribution of unlabeled data to be the same as that of labeled data (it is almost never the case in the problems we consider)