

Apprentissage interactif et collaboratif pour la recherche dans les bases multimédia

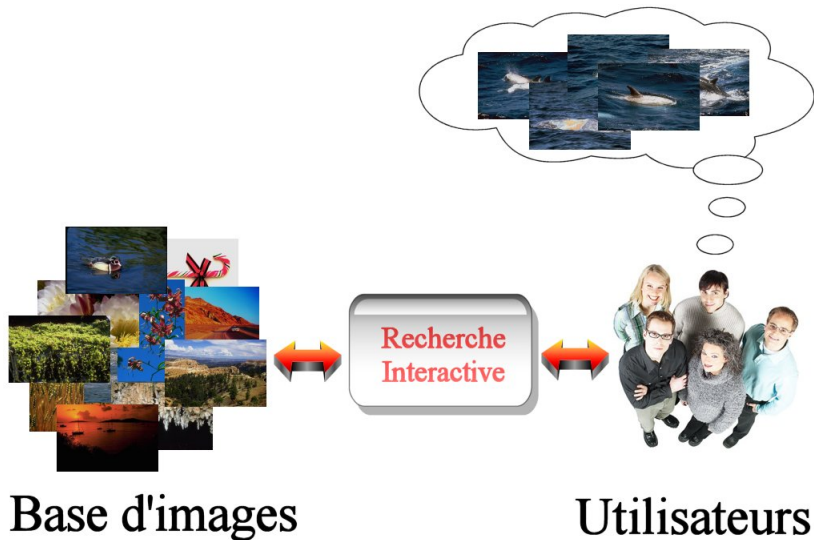
Philippe-Henri Gosselin

Habilitation à Diriger des Recherches
Université de Cergy-Pontoise

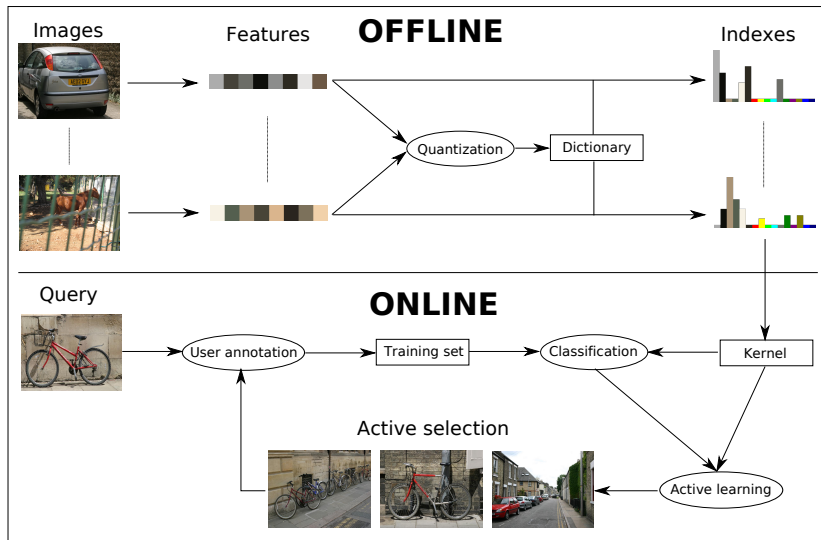
10 novembre 2011



La recherche interactive multimédia



Architecture classique





Plan

- 1 Noyaux sur Sacs et Graphes
- 2 Noyaux Dynamiques
- 3 Boosting Interactif
- 4 Conclusion



Plan

- 1 Noyaux sur Sacs et Graphes
- 2 Noyaux Dynamiques
- 3 Boosting Interactif
- 4 Conclusion



Noyaux sur Sacs et Graphes

Indexation pour la recherche d'objets

Étapes *hors ligne* :

- 1 Extraction descripteurs visuels
⇒ Une image = un ensemble (ou sac) de descripteurs
- 2 Définition métrique entre sacs de descripteurs
- 3 (Éventuellement) Structures index rapides

Sacs de descripteurs

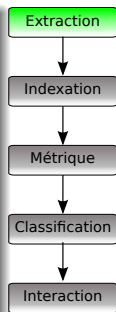
Sacs de descripteurs

- 1 Extraction de primitives
 - ⇒ Regions,
 - ⇒ Points d'intérêt,
 - ⇒ ...
- 2 Description des primitives
 - ⇒ Couleurs, textures,
 - ⇒ HoG, SIFT,
 - ⇒ ...
- 3 Sacs de descripteurs $B_i = \{\mathbf{b}_{ri}\}_r$
 - ⇒ \mathbf{b}_{ri} : descripteur r de l'image i



\mathbf{b}_1	\mathbf{b}_2	\mathbf{b}_3	\mathbf{b}_4	\mathbf{b}_5	\mathbf{b}_6	\mathbf{b}_7	\mathbf{b}_8
37	37	78	90	19	38	23	77
83	09	83	12	78	29	36	82
83	61	38	72	88	99	01	32
92	43	82	72	66	72	28	39

$$\text{Bag } B = \{\mathbf{b}_r\} \in \mathcal{B}$$



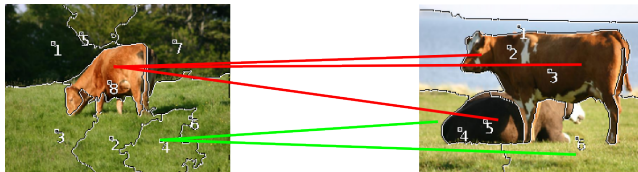


Métrique entre sacs de descripteurs

Méthodes pour la recherche d'objets (2005)

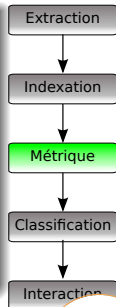
- Systèmes basé vote
 - + Très performants
 - Coûteux calculs (sans index rapide)
 - Non compatible classification
- Dictionnaires visuels
 - + Performants
 - + Rapides
 - Difficile à régler
- Fonction noyaux (+MIL)
 - + Très performants (idem vote)
 - + Facile à régler
 - Coûteux calculs (sans index rapide)

Comparaison des Sacs



Mise en correspondance

- $B_i = \{\mathbf{b}_{ri}\}_r$ sac descripteurs image i
- $B_j = \{\mathbf{b}_{rj}\}_r$ sac descripteurs image j
- Système vote $S(B_i, B_j) =$ nombre descripteurs communs entre i et j
- Fonction noyau $K(B_i, B_j) = \text{idem}$ et $\exists \Phi : \mathcal{X} \rightarrow \mathcal{H}$ t.q.
 $K(B_i, B_j) = \langle \Phi(B_i), \Phi(B_j) \rangle_{\mathcal{H}}$



Noyaux sur Sacs

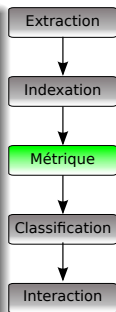
“Soft maximum” (Shawe-Taylor)

$$K_{softmax}(B_i, B_j) = \sum_{\mathbf{b}_{ri} \in B_i} \sum_{\mathbf{b}_{sj} \in B_j} k(\mathbf{b}_{ri}, \mathbf{b}_{sj})$$

- k fonction noyau $\Rightarrow K_{softmax}$ fonction noyau
- Compare le barycentre des éléments de chaque sac :

$$\Phi(B_i) = \sum_{\mathbf{b}_{ri} \in B_i} \phi(\mathbf{b}_{ri})$$

$$\begin{aligned} K_{softmax}(B_i, B_j) &= \langle \Phi(B_i), \Phi(B_j) \rangle \\ &= \sum_{\mathbf{b}_{ri} \in B_i} \sum_{\mathbf{b}_{sj} \in B_j} \langle \phi(\mathbf{b}_{ri}), \phi(\mathbf{b}_{sj}) \rangle \end{aligned}$$



Noyaux sur Sacs

Noyaux définis

- “Enhanced Soft Maximum” (Lyu) :

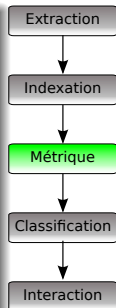
$$K_{lyu}(B_i, B_j) = \frac{1}{|B_i|} \frac{1}{|B_j|} \sum_{\mathbf{b}_{ri} \in B_i} \sum_{\mathbf{b}_{sj} \in B_j} k(\mathbf{b}_{ri}, \mathbf{b}_{sj})^q$$

- “Power” (Gosselin) :

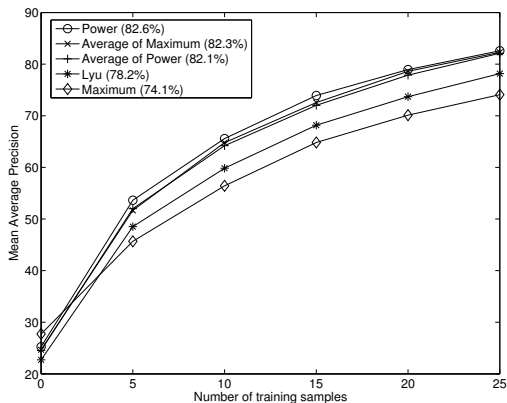
$$K_{power}(B_i, B_j) = \left(\sum_{\mathbf{b}_{ri} \in B_i} \sum_{\mathbf{b}_{sj} \in B_j} k(\mathbf{b}_{ri}, \mathbf{b}_{sj})^q \right)^{\frac{1}{q}}$$

1. Gosselin, Cord and Philipp. *Kernel on Bags of Fuzzy Regions for fast object retrieval*. ICIP 2007.

2. Gosselin, Cord and Philipp. *Kernel on Bags for multi-object database retrieval*. CIVR 2007.



Noyaux sur Sacs

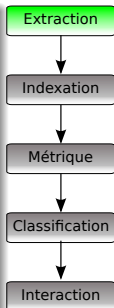
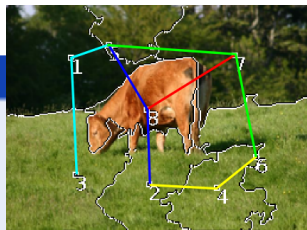


Précision Moyenne(%) sur une base d'objets (Columbia) sur fond aléatoire (ANN) avec différents noyaux.

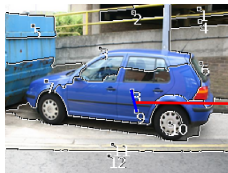
Représentation par Graphes

Représentation par Graphes

- Intégrer les relations spatiales
- Graphes relationnels valués
- Sommet d'un graphe :
 - Région
 - Point d'intérêt
- Arête d'un graphe : Relation spatiale
 - Gauche/droite
 - Haut/base



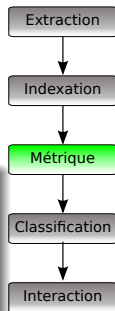
Noyaux sur Graphes



Approche par Sacs de chemins

- Énumération des chemins $h \in H(G)$ issus du graphe G
- Mise en correspondance entre chemins :

$$K_{\text{graphe}}(G, G') = K_{\text{sac}}(H(G), H(G'))$$



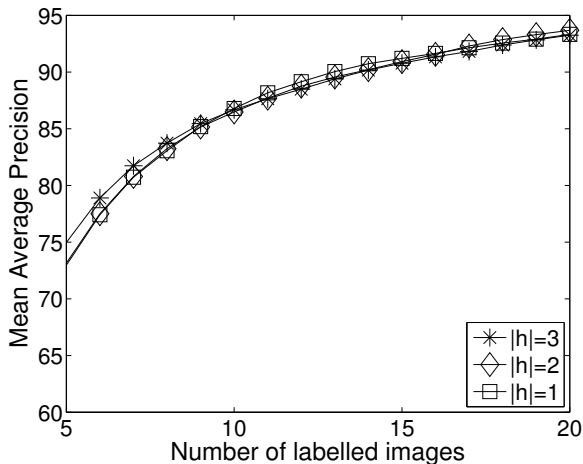


Noyaux sur Graphes

Thèses

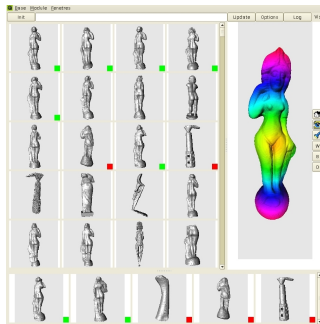
- Justine Lebrun (2006–2011)
 - Allocation ministérielle
 - « *Fonctions noyaux sur Graphes pour la recherche d'objets dans les bases multimédia* »
- Jean-Emmanuel Haugeard (2007–2010)
 - Projet ANR (iTowns)
 - « *Extraction et reconnaissance d'objets dans les façades de Paris à l'aide d'appariement de graphes* »

Recherche d'objets



1. Lebrun, Gosselin, Philipp. *Inexact Graph Matching Based on Kernels for Object Retrieval in Image Databases*. IVC 2011.
2. Lebrun, Gosselin, Philipp. *Image retrieval with graph kernel on regions*. ICPR 2008.

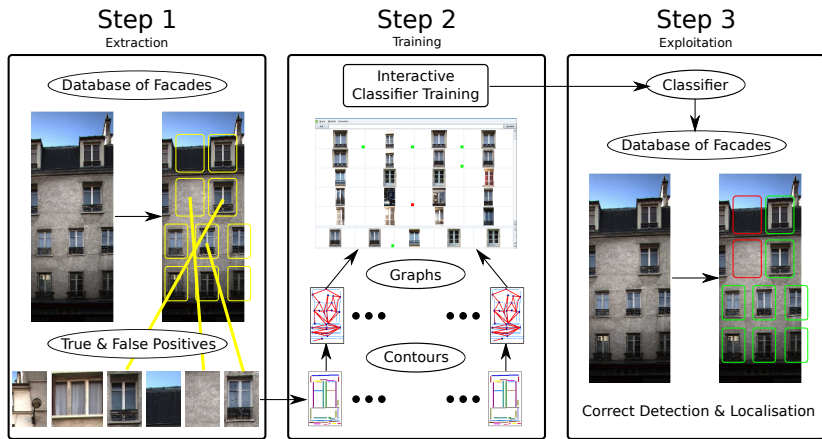
Recherche d'objets 3D



- Projet ANR (EROS 3D)
- 750 objets 3D (maillages sans textures)

1. Philipp, Jordan, Fuzier, Gosselin. *Indexing of 3D Models Based on Graph of Surfacic Regions*. ACM Multimedia 2010.

Reconnaissance dans les façades de Paris



Projet ANR (iTowns)

1. Haugeard, Philipp, Gosselin. **Kernel on Graphs based on Dictionary of Paths for Image Retrieval**. ICPR 2010.
2. Haugeard, Gosselin, Philipp. **Kernel on Graphs of Contours for Image Retrieval**. Soumis à Pattern Recognition Letters



Noyaux sur Sacs et Graphes

Avantages

- Reproduisent bien les systèmes de vote
- Compatibles avec les machines à noyaux
- Peu de paramètres

Inconvénient

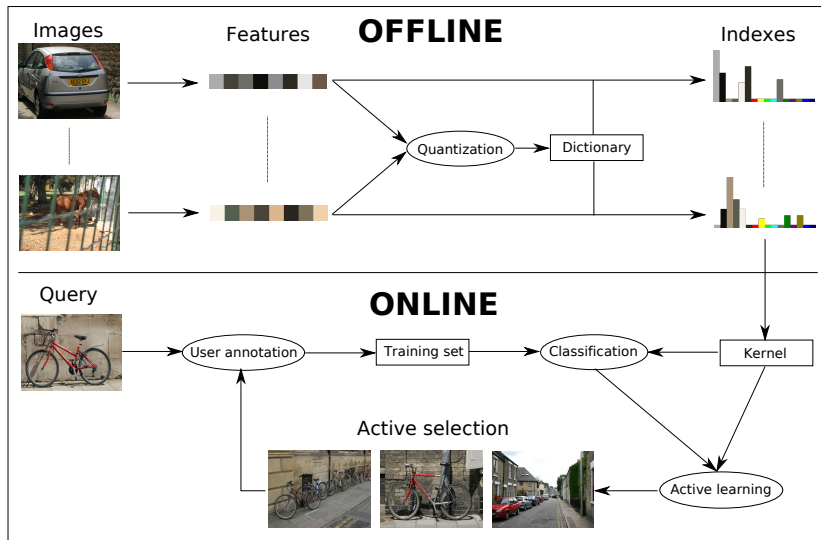
- Temps de calculs !



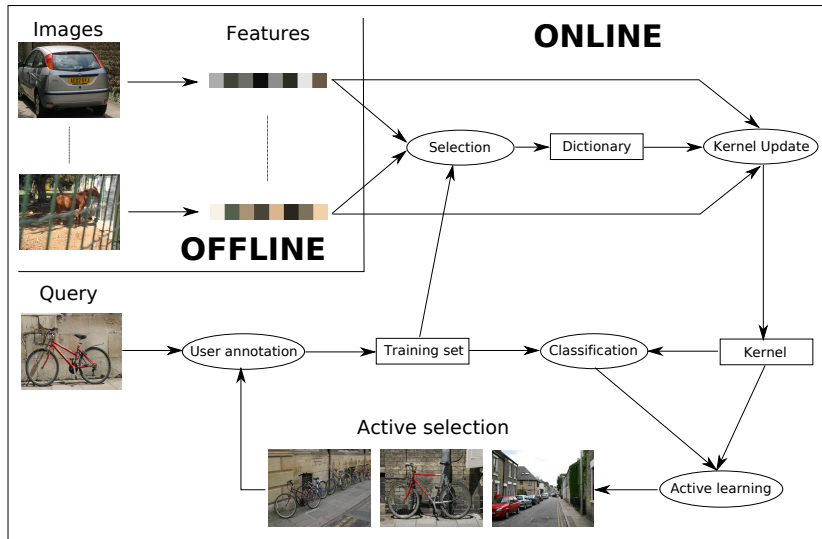
Plan

- 1 Noyaux sur Sacs et Graphes
- 2 Noyaux Dynamiques
- 3 Boosting Interactif
- 4 Conclusion

Architecture classique



Architecture dynamique



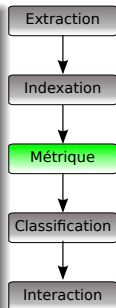
Noyaux Dynamiques

Objectif

- Apprendre combinaison noyaux (MKL)

$$K = \sum_{t=1} \beta_t k_t$$

- Contraintes :
 - Adapter la combinaison à la requête utilisateur courante
 - Gérer le faible nombre d'annotations
 - Apprendre rapidement



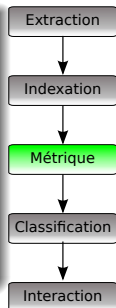
Noyaux Dynamiques

Approche itérative

$$K_t^* = K_{t-1} + \beta_t^* k_t^*$$

avec

- $\beta_t^*, k_t^* = \operatorname{argmax}_{\beta > 0, k \in \mathcal{K}} \mathcal{A}_L^H(K_{t-1} + \beta k)$
- $\mathcal{A}_L(K)$ Alignement du Noyau K sur les annotations L
- \mathcal{K} pool de noyaux mineurs



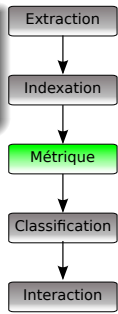
1. Gosselin, Precioso, Philipp. *Incremental Kernel Learning for Active Image Retrieval without Global Dictionaries*. PR 2010.



Noyaux Dynamiques

Noyaux mineurs

- Un noyau mineur \Leftrightarrow Un mot de dictionnaire
- Une combinaison de noyaux \Leftrightarrow Un dictionnaire visuel





Noyaux Dynamiques

Categorie	Taille Dictionnaire			Dynamique
	64/64	256/256	64/160	
bicycle	43	45	44	48
bus	48	57	56	64
car	65	69	70	79
cat	28	28	30	31
cow	47	35	40	41
dog	22	23	23	25
horse	22	24	23	25
motorbike	46	55	50	61
person	33	35	34	37
sheep	48	38	48	36
all	40	41	42	45

TABLE: Precision Moyenne(%) sur l'ensemble test de VOC2006. Initialisation avec 1 image, 10 bouclages, 5 labels par bouclage.

Reconnaissance d'objets dans les façades de Paris

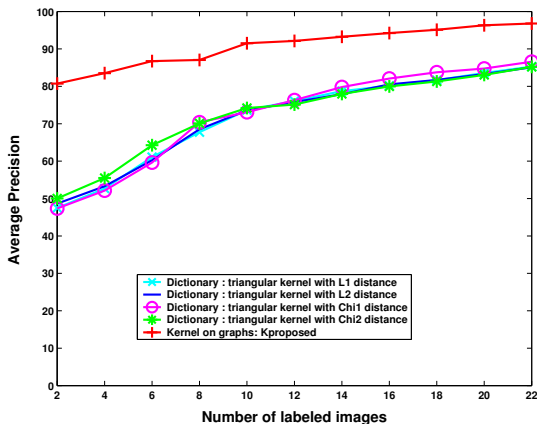


FIGURE: Précision Moyenne(%) avec un noyau sur sac (en rouge) et un noyau dynamique (autres).



Reconnaissance d'objets dans les façades de Paris

Nombre de sommets	26	30	40	124
Noyau sur Graphe	40ms	96ms	184ms	511ms
Noyau dynamique	5ms	6ms	9ms	48ms

TABLE: Temps d'exécution pour l'évaluation d'une valeur de fonction noyau, en fonction de la taille des graphes.



Noyaux Dynamiques

Propriétés

- Conception à la volée d'un dictionnaire/noyau adapté à la requête courante
 - ⇒ Pas de dictionnaire à cross valider
- Entraînement rapide avec peu d'exemples
- Classification plus rapide (dictionnaires plus petits)

Batch learning

- Fusion très rapide de sorties de classifieurs



Plan

- 1 Noyaux sur Sacs et Graphes
- 2 Noyaux Dynamiques
- 3 Boosting Interactif
- 4 Conclusion



Boosting

Une philosophie très différente

- Postulat
 - Il est facile de trouver des règles généralement justes...
 - ...mais il est difficile de trouver des règles toujours justes
- Principe
 - Construire un ensemble de règles simples
 - ⇒ Pool de classifieurs "faibles"
 - Combiner ces règles
 - ⇒ Formation d'un classifieur "fort"



Boosting

Thèses

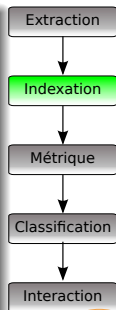
- Alexis Lechervy (2009–2012)
 - Projet DGA
 - « *Apprentissage interactif et collaboratif pour la détection d'évènement dans les vidéos* »
- Jefersson Alex dos Santos (2009–2012)
 - Cotutelle avec Université Brésilienne de Campinas
 - « *Multi-Scale Classification of Remote Sensing Images* »

Boosting pour les bases généralistes



Classifieurs faibles

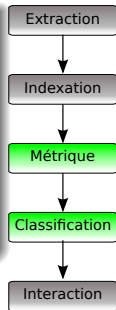
- Description Visuelle
 - Découpage en 9 blocs
 - 1 Région = 1 combinaison bloc (511 possibilités)
 - Calcul descripteur dans chaque région
- Classifieur faible :
 - Généré à partir d'un exemple positif
 - Se focalise sur 1 type de région
 - Similarité au descripteur dans la région



Boosting pour les bases généralistes

Pool dynamique de classifieurs faibles

- Peu d'exemples en recherche interactive
- Solution : pool dynamique de classifieurs faibles
 - Pour chaque image annotée positive
 - Pour chaque type de région
 - Pour chaque type de descripteur
- ⇒ Fabriquer un classifieur faible



1. Lechery, Gosselin, Precioso. *Active Boosting for interactive object retrieval*. ICPR 2010.

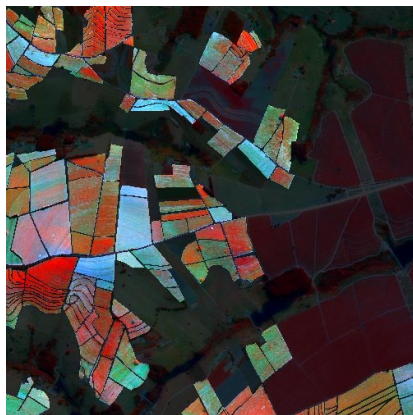


Boosting pour les bases généralistes

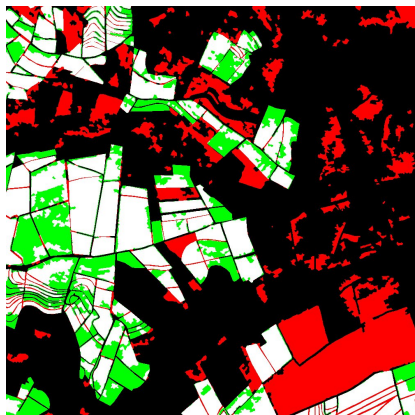
Dictionnaire Méthode	32 mots		64 mots	
	SVM	Boosting	SVM	Boosting
bicycle	34	44	43	38
bus	40	45	48	46
car	57	68	65	68
cat	26	30	28	28
cow	32	40	47	37
dog	20	25	22	24
horse	16	19	22	19
motorbike	35	42	46	41
person	34	38	33	38
sheep	46	53	48	44
all	35	40	40	38

TABLE: Precision Moyenne(%) sur l'ensemble test de VOC2006. Initialisation avec 1 image, 10 bouclages, 5 labels par bouclage.

Boosting pour les images satellites

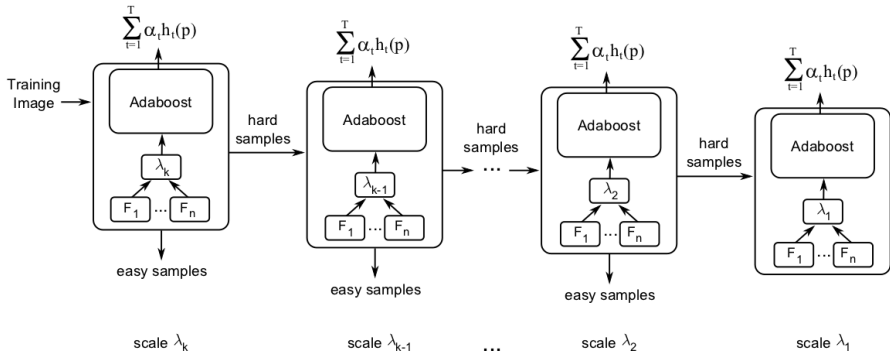


Vérité terrain



Classification

Boosting pour les images satellites



1. Santos, Gosselin, Philipp, Torres and Falcao. **Multi-Scale Classification of Remote Sensing Images**. GRS (Révision).



Boosting pour les images satellites

Méthode	Précision(%)	Kappa
Baseline SVM	77.47	0.7054
Boosting	80.57	0.7502
Boosting hiérarchique	82.09	0.7772

TABLE: Résultats de classification d'images satellites (catégorie café).



Boosting interactif

Propriétés

- Conception plus libre que machines à noyaux
- Entraînement plus lent
- Classification plus rapide
- Fonctionne avec peu d'exemples

En cours

- Boosting collaboratif images/vidéos (A. Lechervy)
- Boosting interactif images satellites (J. Alex)



Plan

- 1 Noyaux sur Sacs et Graphes
- 2 Noyaux Dynamiques
- 3 Boosting Interactif
- 4 Conclusion



Noyaux sur ensembles

Bilan Noyaux sur Sacs et Graphes

- + Performants, compatible SVM
- Coûteux en calculs

En cours

- Réduction temps calculs noyaux sur sacs
- Approche linéarisation avec le formalisme des tenseurs¹
- Thèse démarrée 2011 : Romain Négrel « *Représentation optimales pour la recherche dans les bases patrimoniales* », Labex Patrima

Perspectives à moyen terme

- Extension résultats noyaux sur graphes
- Passage aux très grandes bases (1 million +)



Représentations dynamiques

Bilan

- Noyau dynamique performant mais contraignant
- Boosting moins contraignant mais moins performant

En cours

- Cadre formel intermédiaire
- Apprentissage collaboratif
- Application vidéo

Perspectives à moyen terme

- Classification temps réelle
- Mise à jour on-line du classifieur



Perspectives à long terme

Apprentissage dans le flux

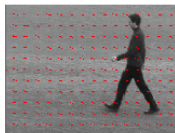
- Bases vivantes / flux vidéos
 - ⇒ Ajout régulier de documents
 - ⇒ Suppression/archivage régulier de documents
- Objectifs classification instables
 - ⇒ Évolution rapide des centres d'intérêts
 - ⇒ Communautés d'utilisateurs
- Données d'apprentissage de mauvaise qualité
 - ⇒ Micro ensembles d'apprentissage
 - ⇒ Beaucoup d'erreurs
- Passage à l'échelle



Publications

	Revues	Année	8
	Elsevier Image and Vision Computing	2011	1
	Elsevier Pattern Recognition	2010	1
	Elsevier Computer Vision and Image Understanding	2009	1
	IEEE Transactions on Image Processing,	2008	2
	Elsevier Computer Vision and Image Understanding		
	Elsevier Image and Vision Computing	2007	1
	Elsevier Pattern Recognition,	2006	2
	EURASIP Journal on Applied Signal Processing		
	Conférences internationales	Période	17
	ICIP, ICPR, ACM	2008–2011	5
	ICIP, CIVR, ...	2004–2007	12
	Conférences nationales	Période	4
	RFIA, EGC	2008–2011	2
	RFIA, CAP	2004–2007	2

Tenseurs pour la description vidéo



Polynômes générés

$$V^1 = -0.467823 - 0.366877x_1 - 0.061723x_2$$

$$V^2 = 0.279198 + 0.224533x_1 - 0.06561x_2$$



Tenseur généré

$$T = \begin{matrix} 0.218858 & 0.171634 & 0.028875 & -0.130615 & -0.105042 & 0.030697 \\ 0.171634 & 0.134599 & 0.022645 & -0.102431 & -0.082376 & 0.024073 \\ 0.028875 & 0.022645 & 0.003810 & -0.017233 & -0.013859 & 0.004050 \\ -0.130615 & -0.102431 & -0.017233 & 0.077951 & 0.062689 & -0.018320 \\ -0.105042 & -0.082376 & -0.013859 & 0.062689 & 0.050415 & -0.014733 \\ 0.030697 & 0.024073 & 0.004050 & -0.018320 & -0.014733 & 0.004306 \end{matrix}$$



Tenseurs pour la description vidéo

Stages/Thèse

- Romain Negrel
 - Stage ETIS printemps/été 2011
- Virginia Fernandes Mota
 - Collaboration avec Université Brésilienne de Juiz de Fora
 - Stage ETIS printemps/été 2010
 - Thèse démarrée en 2011 « *Tensor based representations for action retrieval in videos* »