

Quelques contributions pour la collaboration et l'interaction en analyse et interprétation d'images et de vidéos

Alain Boucher
Professeur IFI (AUF)
Hanoi, Vietnam

Présentation

■ Parcours personnel

→ *Canada*

- Ingénieur informatique, **Ecole Polytechnique de Montréal** [1994]

→ *France*

- Thèse, **Univ. Joseph Fourier de Grenoble** [1999]
- Ingénieur expert, **INRIA Sophia Antipolis** [1999 – 2002]

→ *Vietnam*

- Professeur à l'**Institut de la Francophonie pour l'Informatique (IFI - AUF)** [2002+]

■ Collaboration entre l'IFI et l'Université de La Rochelle

- Accord de double diplôme depuis 2006
 - option Systèmes Intelligents & Multimédia
- Etudiants de l'IFI venus au L3i en stage Master 2 et/ou en thèse
- Etudiants de l'ULR venus faire leur Master 2 à l'IFI et/ou stage M2

Contexte personnel de recherche

- **Domaine de recherche**
 - **Analyse / interprétation d'images** en utilisant des **techniques d'intelligence artificielle** avec un intérêt pour les **méthodes interactives / collaboratives**
- **Thèses en co-direction Vietnam-France (9 depuis 2002)**
 - *ULR (3), INRIA Sophia Antipolis (2), UPMC (2), LORIA (1), IRD (1)*
 - Etudiants VN recherchant formation + expérience à l'étranger
 - tout en faisant une partie de leur recherche au Vietnam
 - Conciliation des buts + projets de recherche des partenaires
- **Différentes équipes de recherche**
 - IFI, équipe MSI, IRD, UMI 209 UMMISCO
 - Centre de recherche MICA (2002-2006)

Introduction

Le traitement parfait n'existe pas !

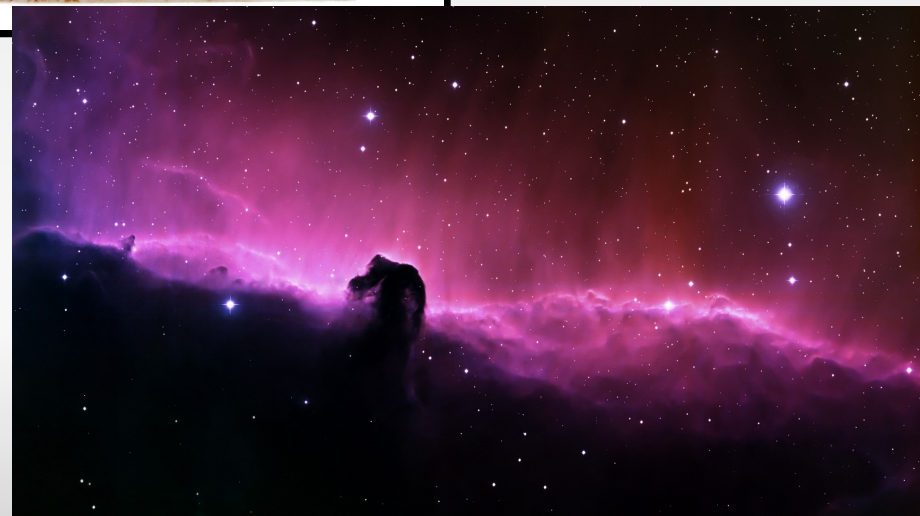
- Difficile / impossible d'avoir un traitement / processus effectuant parfaitement une tâche demandée
 - Besoin de souplesse et d'adaptation dans le traitement
 - Besoin de collaboration dans la résolution d'un problème
 - Addition des forces et pallier les faiblesses
- Exemples
 - **Avant** : segmentation d'images pré-requis à tout traitement
 - **Après** : points d'intérêt (*faible % utile*) puis sélection et traitement
 - **Avant** : reconnaissance des formes donnant solution unique
 - **Après** : recherche d'informations avec analyse de X solutions

Interaction comme processus de traitement

- **Systemes en vase clos de plus en plus difficiles à concevoir**
 - Sauf application spécialisée avec environnement contrôlé
 - Applications d'aujourd'hui de plus en plus complexes
- **Interaction devient une part essentielle du traitement**
 - **Avant** : X h de calculs intensifs aux résultats incertains
 - **Après** : 2 clics souris en 2 secondes par l'utilisateur
- **Connaissances ont besoin d'évoluer dans un système**
 - Adaptation au contexte
 - Evolution des conditions d'exécution

Sémantique à degré variable

- Sémantique donnée une image dépend de deux facteurs
 - Du niveau de connaissance et de la perception qu'à l'utilisateur de cette image (*description de l'image*)
 - De l'objectif visé par l'utilisateur (*contexte*)



Exemples avec différents travaux

- Différents travaux illustrant différentes possibilités d'intégrer l'humain ou la collaboration
- **Interaction avec l'humain**
 - Acquisition de connaissances avec un expert
 - Interaction en recherche d'images par le contenu
 - Interaction pour l'indexation de grandes bases d'images
- **Collaboration entre processus**
 - Analyse vidéo + recherche d'informations
 - Multimodalité (texte + parole)

Interaction avec un expert/utilisateur humain

Interaction avec un expert/utilisateur humain

Acquisition de connaissances avec un expert

- Un système semi-automatique de reconnaissance de pollens
 - *Post-doctorat (INRIA Sophia Antipolis) [1999- 2002]*

Interaction en recherche d'images par le contenu

- Représentations visuelles de concepts textuels pour la recherche et l'annotation interactives d'images
 - *Thèse de Nguyen Nhu Van (L3i / IFI-MSI) [2007-2011]*

Interaction pour l'indexation de grandes bases d'images

- Système interactif de structuration des index pour une recherche par le contenu dans des grandes bases d'images
 - *Thèse de Lai Hien Phuong (L3i / IFI-MSI) [2009-2012]*

Interaction avec un expert/utilisateur humain

Acquisition de connaissances avec un expert

- Un système semi-automatique de reconnaissance de pollens
 - *Post-doctorat (INRIA Sophia Antipolis) [1999- 2002]*

Interaction en recherche d'images par le contenu

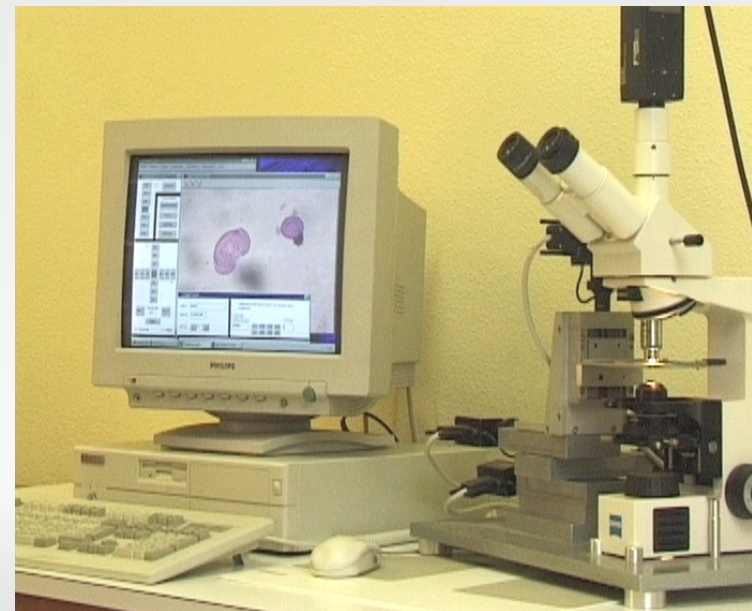
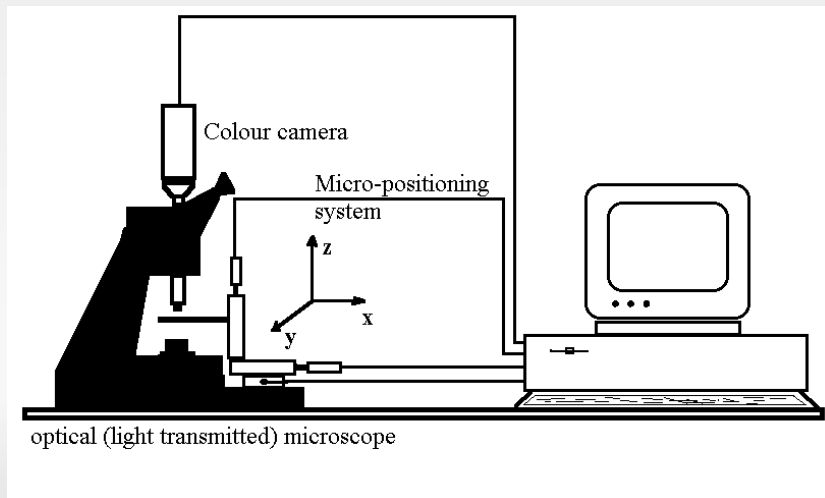
- Représentations visuelles de concepts textuels pour la recherche et l'annotation interactives d'images
 - *Thèse de Nguyen Nhu Van (L3i / IFI-MSI) [2007-2011]*

Interaction pour l'indexation de grandes bases d'images

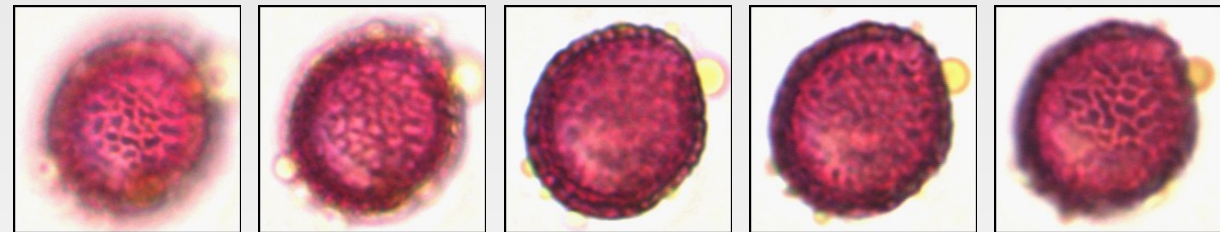
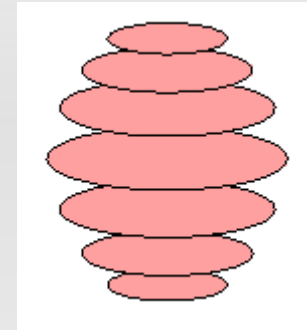
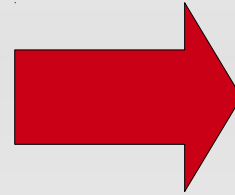
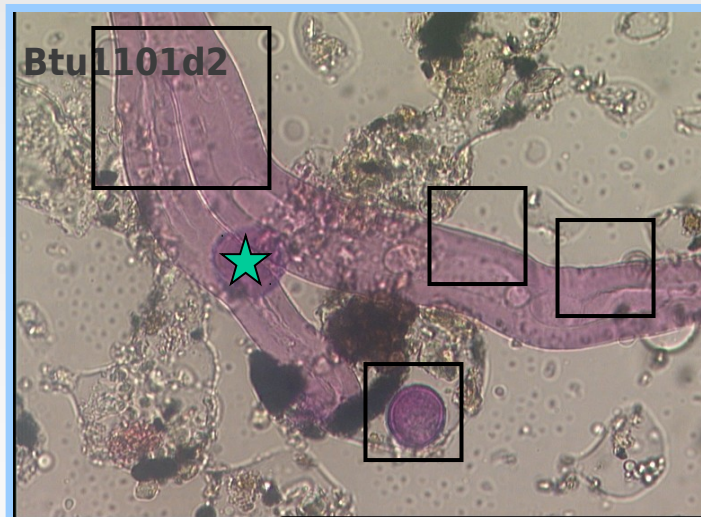
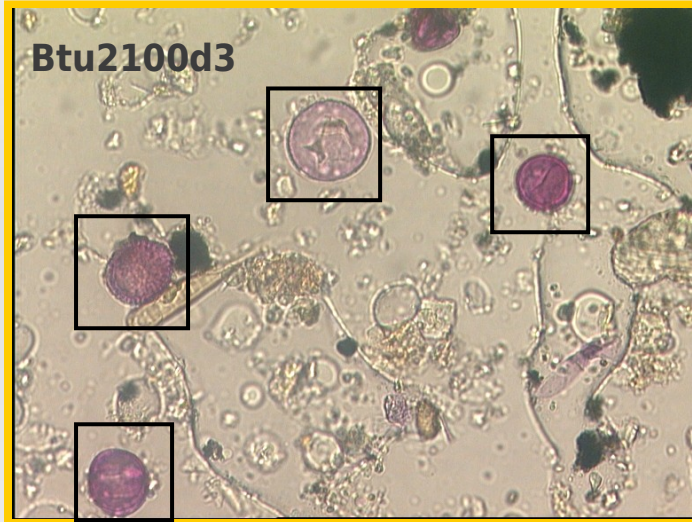
- Système interactif de structuration des index pour une recherche par le contenu dans des grandes bases d'images
 - *Thèse de Lai Hien Phuong (L3i / IFI-MSI) [2009-2012]*

Projet européen ASTHMA

- **Reconnaissance semi-automatique de grain de pollen**
 - identifier les espèces allergènes présentes dans l'air
 - émettre des bulletins de prévision de pollen
- **Processus manuel déjà existant**
 - Exécuter par des spécialistes (botanistes palynologues)
 - Processus long et répétitif



Détection et extraction des grains de pollen

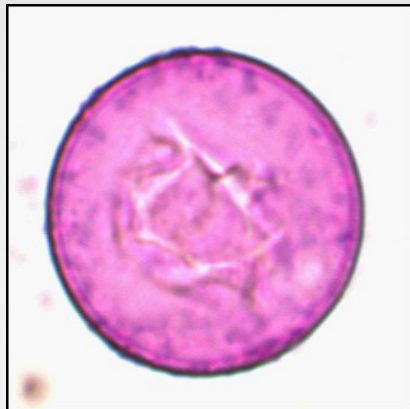


Numérisation des grains en 3D

Détection des grains en 2D

Reconnaissance des grains 3D de pollen

- Détecter les (4) espèces les plus allergènes en évitant la confusion avec d'autres espèces très semblables visuellement
 - Etape 1 : Classification grossière en 2D (*descripteurs images*)
 - Etape 2 : Classification fine en 3D (*structures spécialisées*)



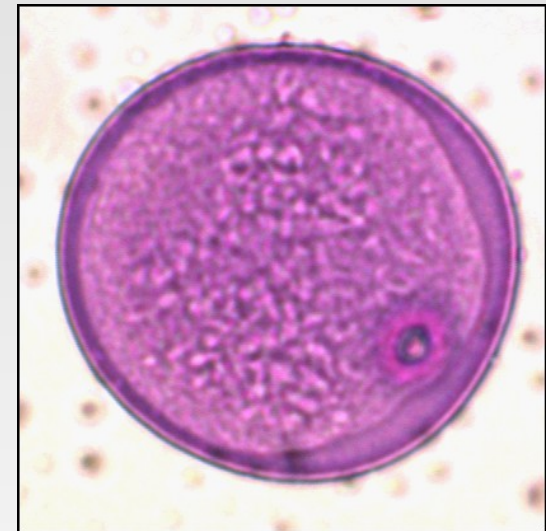
Cupressaceae



Olea



Parietaria



Poaceae

Populus

Brassicaceae

Broussonetia

Celtis

Fraxinus

Morus

Coriaria

Ligustrum

Urtica membranacea

Phillyrea

Salix

Acquisition de connaissances en 3 étapes

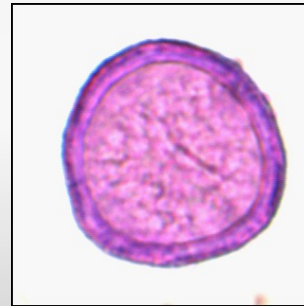
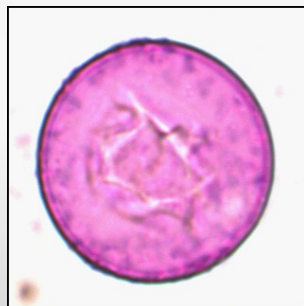
- Plusieurs étapes nécessaires pour acquérir les connaissances
 - Expérience d'un expert difficile à formaliser et à extraire

1^{ère} étape avec l'expert : Connaissance *descriptive*

Pollen type	Nr Apertures <small>ecuatorial=zono all=panto proximal=cata distal=ana</small>	Size <small>very small <10 µm small 10-25 µm med. 25-50 µm big. 50-100 µm very big 100-200 µm</small>	Polarity Apolar Isopolar Heteropolar	Symmetry Radial Bilateral Asymmetric	Shape equatorial view (e) polar view (p)	Surface <small>Psilate, Scabrate, Rugulate, Striate, Reticulate, Verrucate, Perforate, Foveolate, Echinata</small>	Exine thickness	Other characteristics <small>Onci/Opercule/Aerolate/Annulus/Mesocolpi/Apocolpi/Intine...</small>	Flowering period (more frequent)
Cupressaceae	Inaperturate	m	Apolar o heteropolar	Radial	e: circular p: circular	Microgemmate; gemmae irregularly distributed	1 µm	Star shape cytoplasm. Thick intine: 6,5 µm. Grains often split and then folded in a characteristic manner.	All the year (September-June)
<i>Populus</i>	Inaperturate	s-m	Apolar	Radial	approximately isodiametric	Scabrate, microgranules irregularly distributed	1,5 µm	Intine thick. <i>Anastomosis</i> in superficial gemmae	February-April

2^{ème} étape avec l'expert : Connaissance *discriminante*

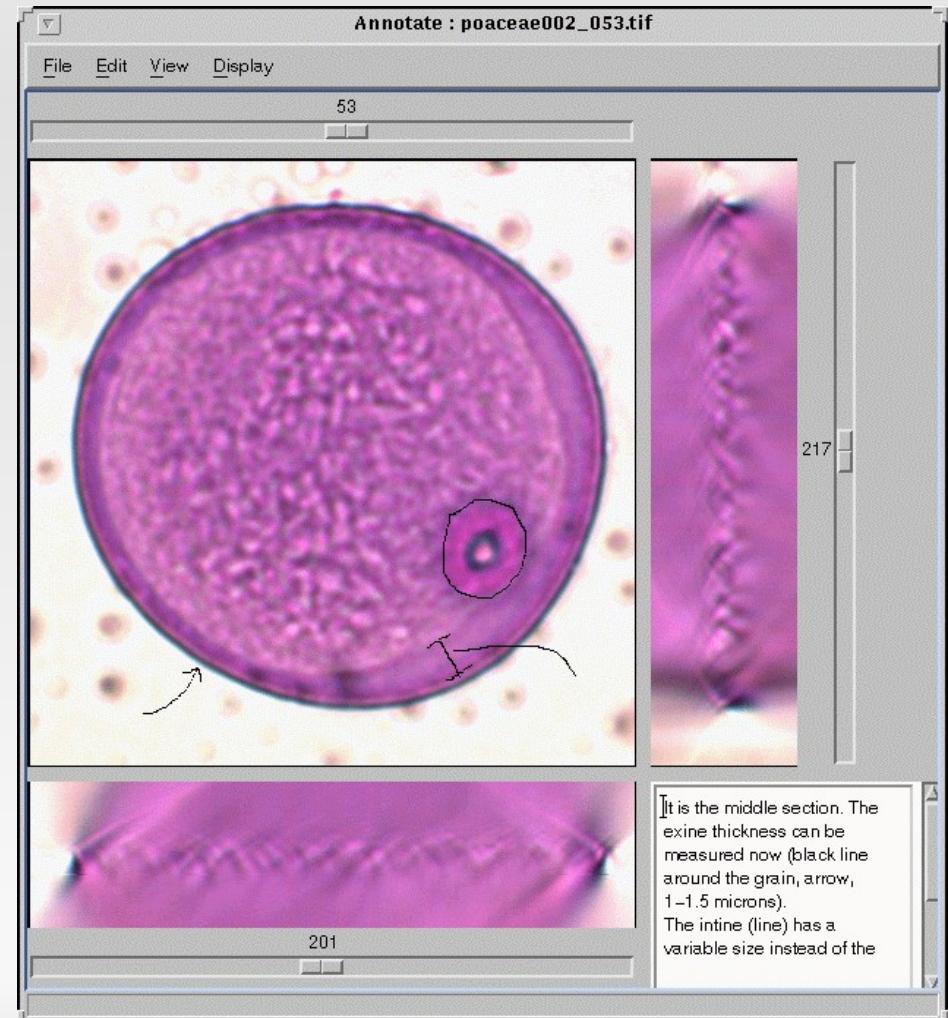
Cupressaceae	<i>Populus</i>
<p><u>Cytoplasm</u>: star-like structure (straight borders)</p> <p><u>Shape</u>: spheroidal, but often broken and open</p> <p><u>Intine</u>: thick, 50% total cell volume</p> <p><u>Surface</u>: few granules irregularly distributed</p>	<p><u>Shape</u>: outline irregular</p> <p><u>Intine</u>: 10% total cell volume</p> <p><u>Surface</u>: “sandy” appearance</p>



Acquisition de connaissances en 3 étapes

3^{ème} étape avec l'expert : Connaissance *pratique*

- Logiciel imitant le travail du palynologue avec un microscope
- Objectif : décrire sur des exemples précis le processus de reconnaissance d'un pollen



Reconnaissance de structure en 3D

Pore du Poaceae

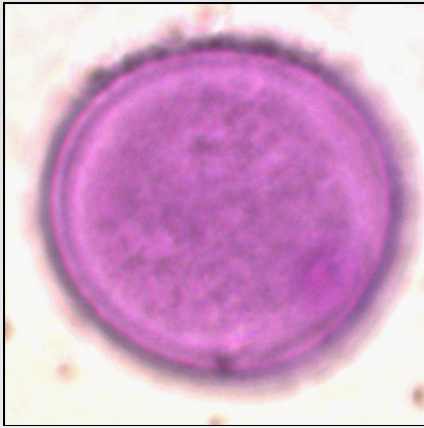


Image 35 / 100

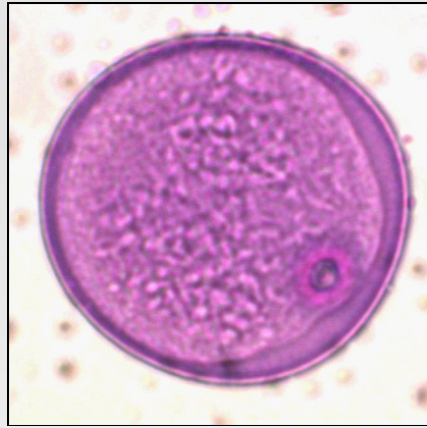


Image 50 / 100

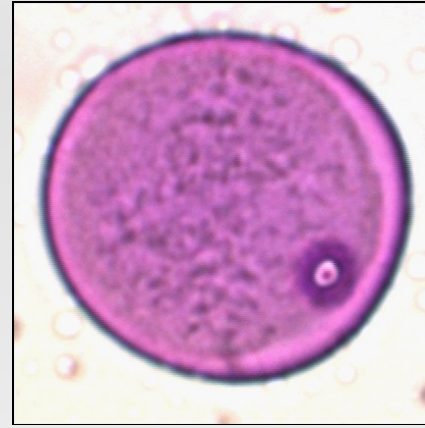
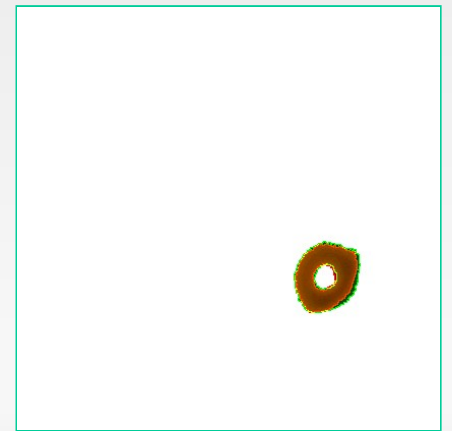
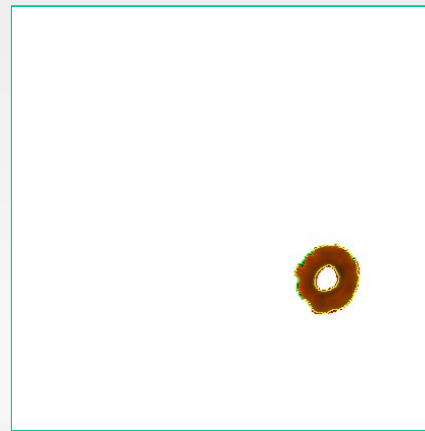
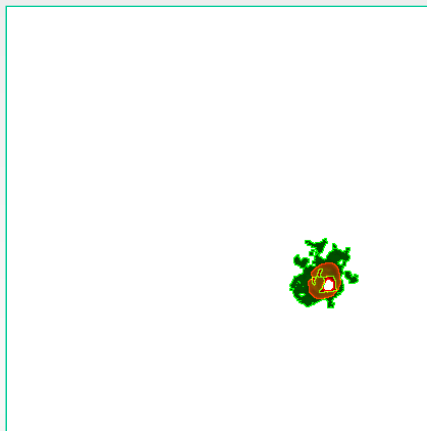


Image 65 / 100



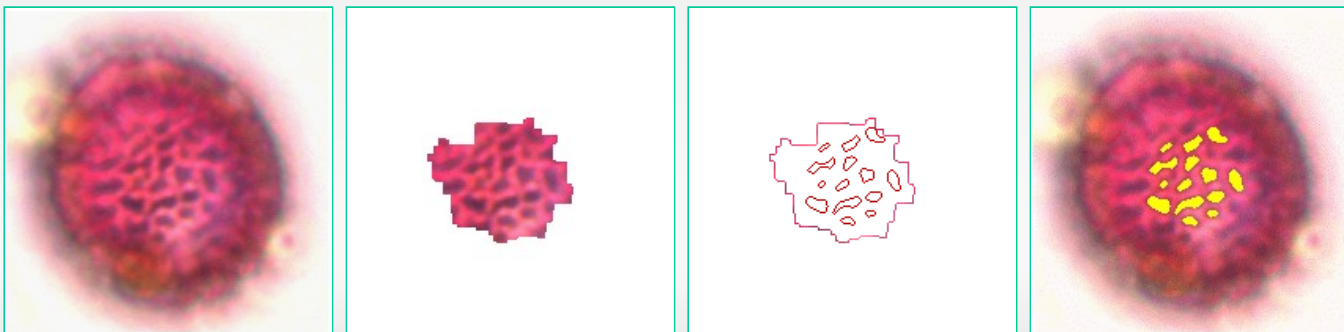
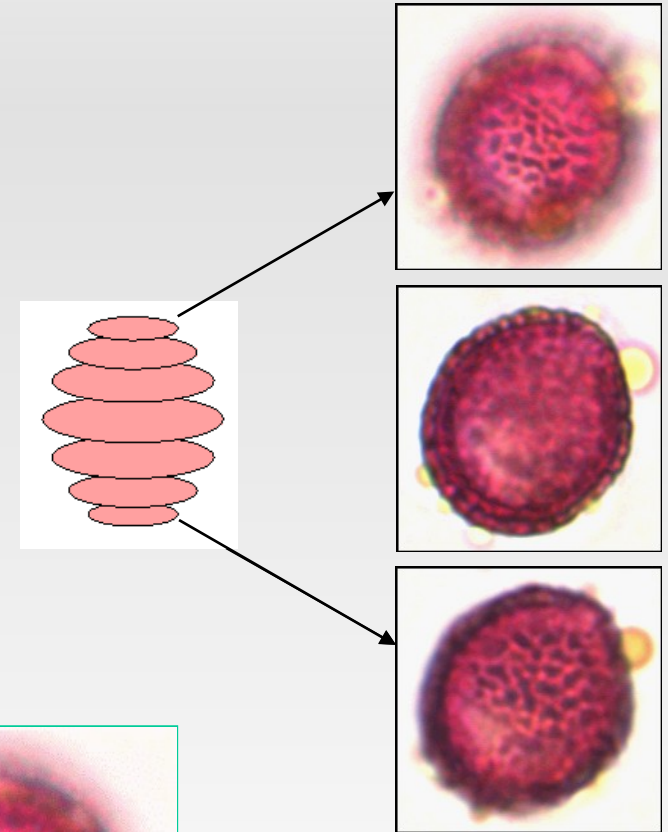
Image 80 / 100



Reconnaissance de structure en 3D

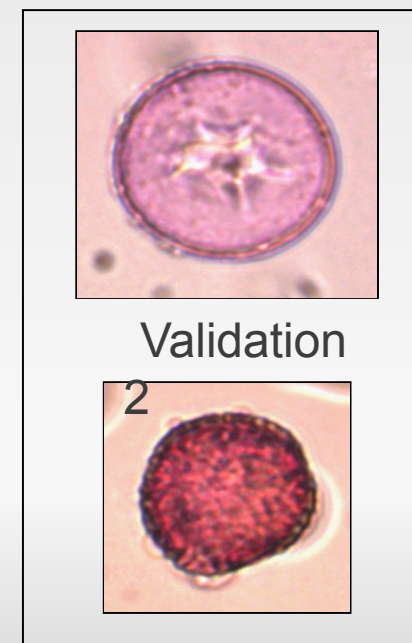
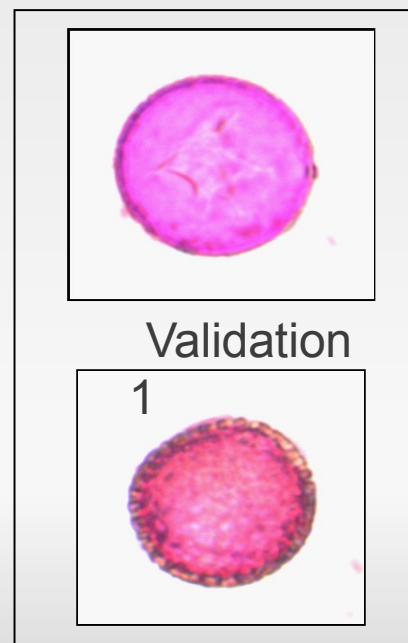
Réticule de l'Olea

- Le réticule (exine) est situé sur la surface du grain (supérieure et inférieure)
- Etapes à suivre:
 - Vérifier si le grain est réticulé
 - Localiser le réticule
 - Analyser le réticule



Résultats

- Deux validations :
 - Méthode *leave-one-out* sur ensemble de départ
 - Reconnaître les 4 types allergéniques (5 classes) : 99,7%
 - Reconnaître parmi 31 types (32 classes) : 77,7%
 - Nouvel ensemble d'images (*acquises 3 ans après les premières*)
 - Reconnaître les 4 types allergéniques (5 classes) : 45 % !!
 - Problèmes de calibration et de robustesse sur la couleur



Interaction avec un expert/utilisateur humain

Acquisition de connaissances avec un expert

- Un système semi-automatique de reconnaissance de pollens
 - *Post-doctorat (INRIA Sophia Antipolis) [1999- 2002]*

Interaction en recherche d'images par le contenu

- Représentations visuelles de concepts textuels pour la recherche et l'annotation interactives d'images
 - *Thèse de Nguyen Nhu Van (L3i / IFI-MSI) [2007-2011]*
Soutenance de thèse demain 13h00 à l'ULR

Interaction pour l'indexation de grandes bases d'images

- Système interactif de structuration des index pour une recherche par le contenu dans des grandes bases d'images
 - *Thèse de Lai Hien Phuong (L3i / IFI-MSI) [2009-2012]*

Contexte

- **Images of natural Disasters from robot Exploration in urban Areas** (*projet STIC-Asie MAE/CNRS/INRIA*)
 - Aide à la décision basée sur l'analyse d'images
 - Aider l'organisation des secours après une catastrophe naturelle



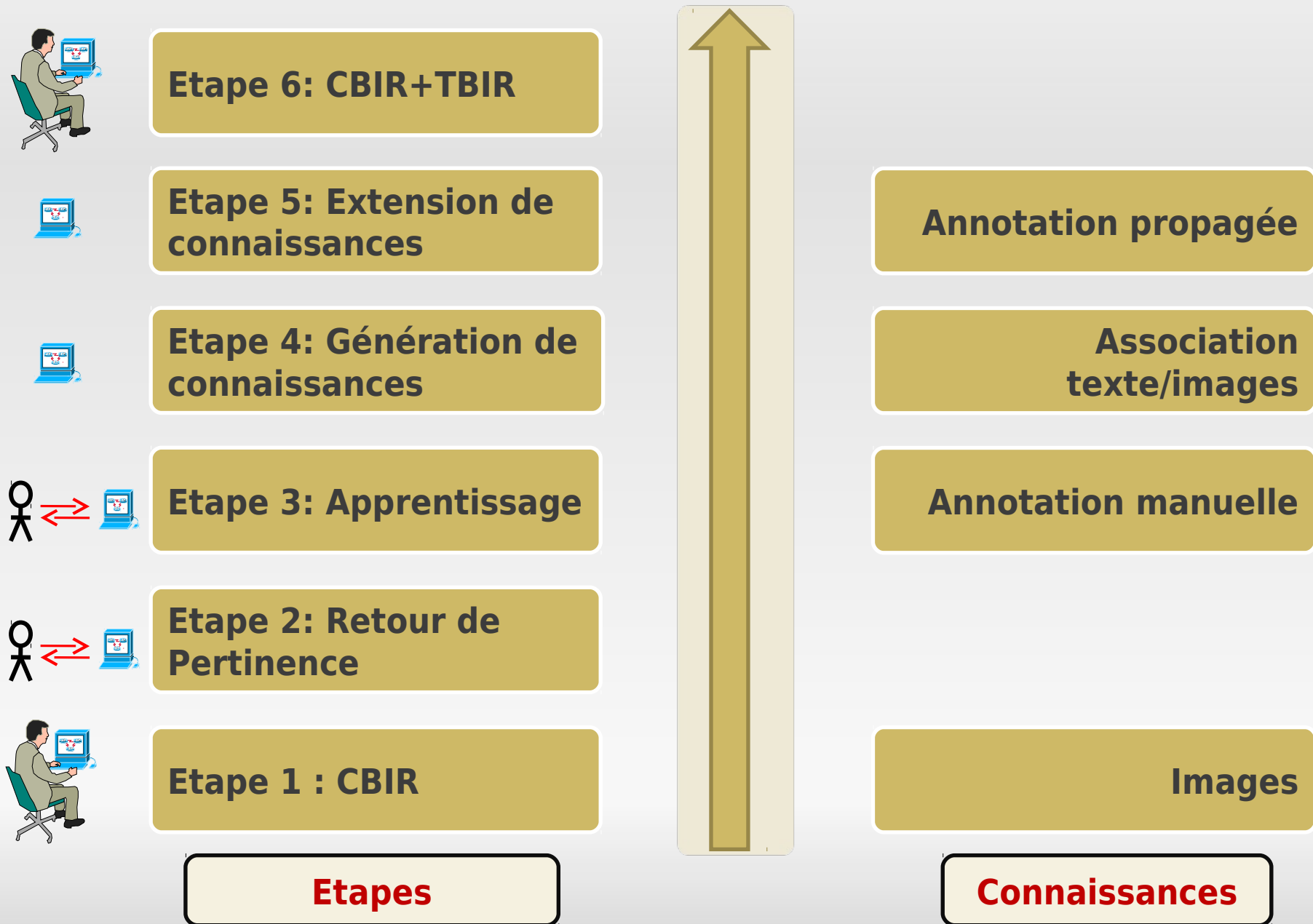
- **Propriétés de l'application**

- Application semi-spécialisée
- Pas de connaissance a priori
- Demande une grande adaptation au contexte d'exécution

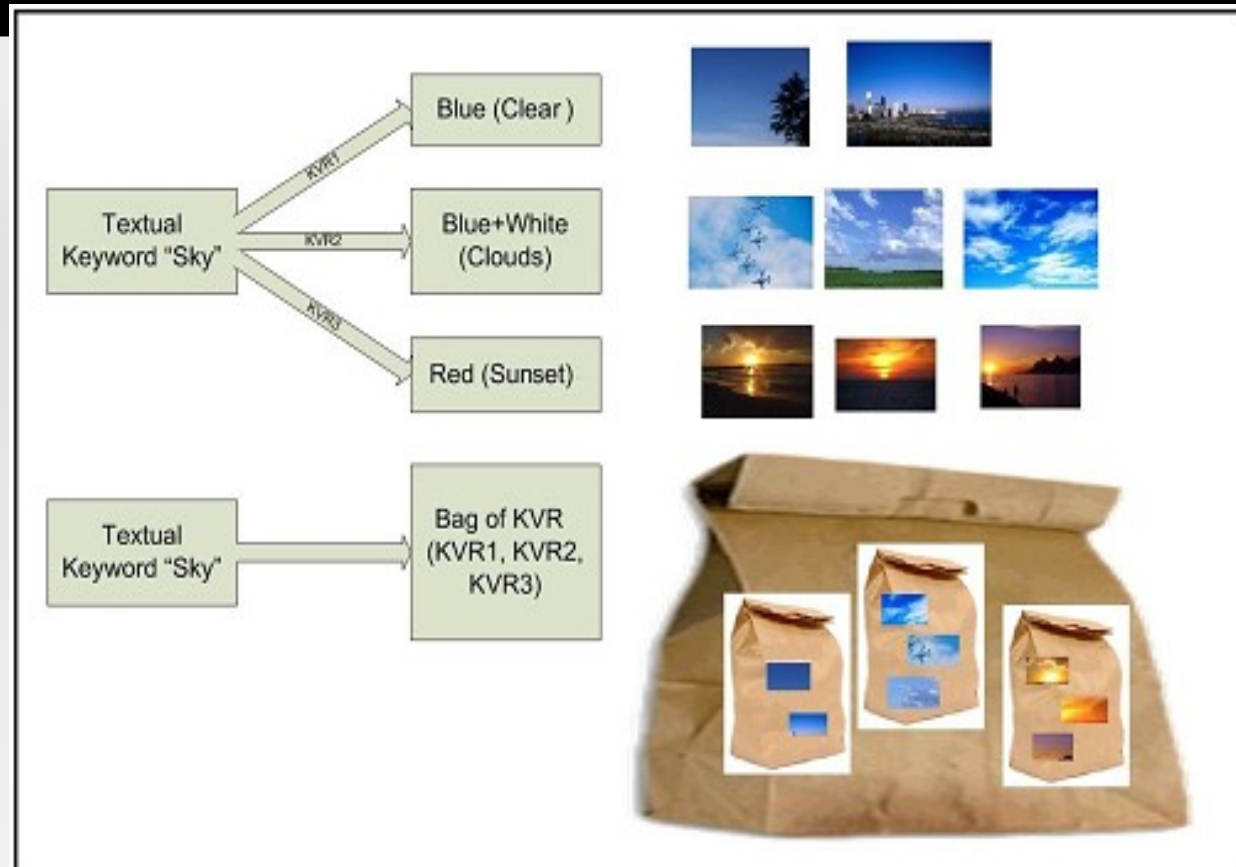
Idée : système s'appuyant sur l'interactivité

- **Au départ, aucune connaissance dans le système**
 - Le système utilise les images seulement
 - *CBIR classique*
- **Intégrer différents types d'interaction**
 - Retour de pertinence
 - *Permettre une recherche efficace dès le début*
 - Apprentissage de connaissances (par renforcement)
 - *Connaissances sur l'application dans le contexte précis*
- **Modèle d'association texte / images**
 - Construit interactivement d'après les retours de l'utilisateur
 - Modèle bidirectionnel : **image** → **texte** & **texte** → **image**

Modèle proposé



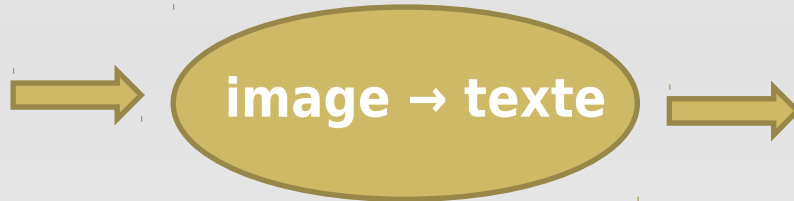
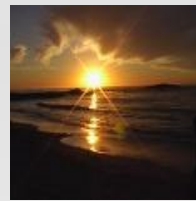
Modèle d'association texte / images



- Construction incrémentale à partir d'annotations de l'utilisateur
 - *Clustering des ensemble d'images annotées*
 - *Représentations des images par des sacs de mots visuels*
- Plusieurs représentations possibles pour un mot-clé
 - *Multiple ensembles d'exemples par mot-clé*
- Utilisable dès le début et s'améliorant avec le temps

Utilisation du modèle

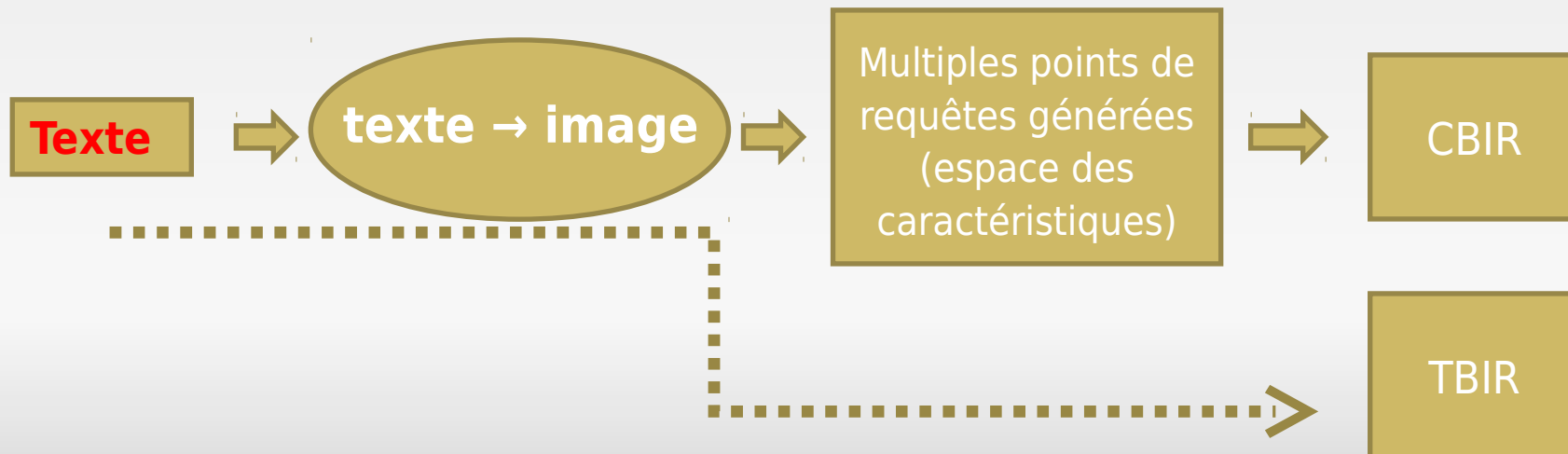
1 – Propagation d'annotations



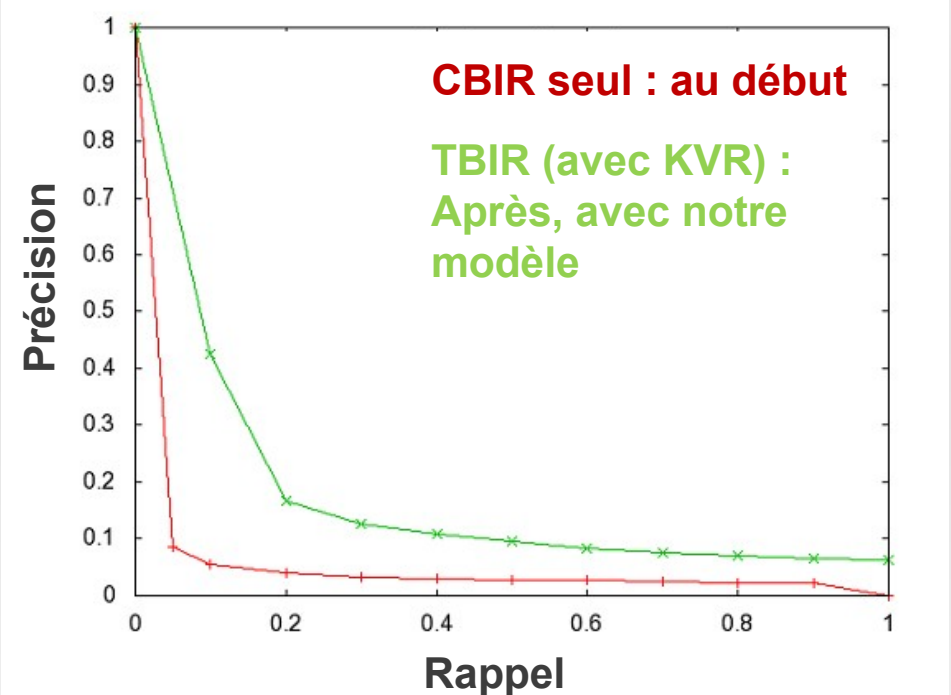
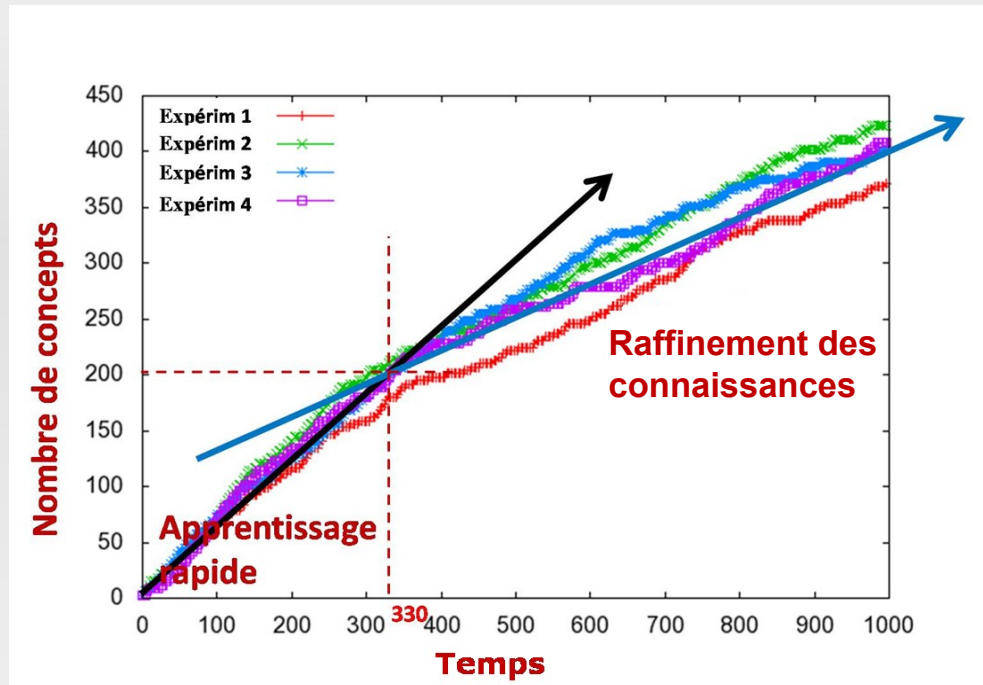
-Mot textuel 1 (d_1)
-Mot textuel 2 (d_2)
-Mot textuel 3 (d_3)
-Mot textuel 4 (d_4)
-Mot textuel 5 (d_5)

$d_1 < d_2 < \dots < d_n$

2 – Recherche d'images par le contenu avec requêtes textuelles



Résultats (base Corel 30K)



- **Modèle stable face à différentes conditions d'expérimentations**
 - Taux d'arrivée des images vs taux d'utilisation du système
- **Modèle intégrant différents types d'interaction**
 - Exploration & retour de pertinence & apprentissage interactif
- **Pas de connaissance a priori, mais adaptation au contexte**
 - Modèle générique spécialisable pour plusieurs applications

Interaction avec un expert/utilisateur humain

Acquisition de connaissances avec un expert

- Un système semi-automatique de reconnaissance de pollens
 - *Post-doctorat (INRIA Sophia Antipolis) [1999- 2002]*

Interaction en recherche d'images par le contenu

- Représentations visuelles de concepts textuels pour la recherche et l'annotation interactives d'images
 - *Thèse de Nguyen Nhu Van (L3i / IFI-MSI) [2007-2011]*

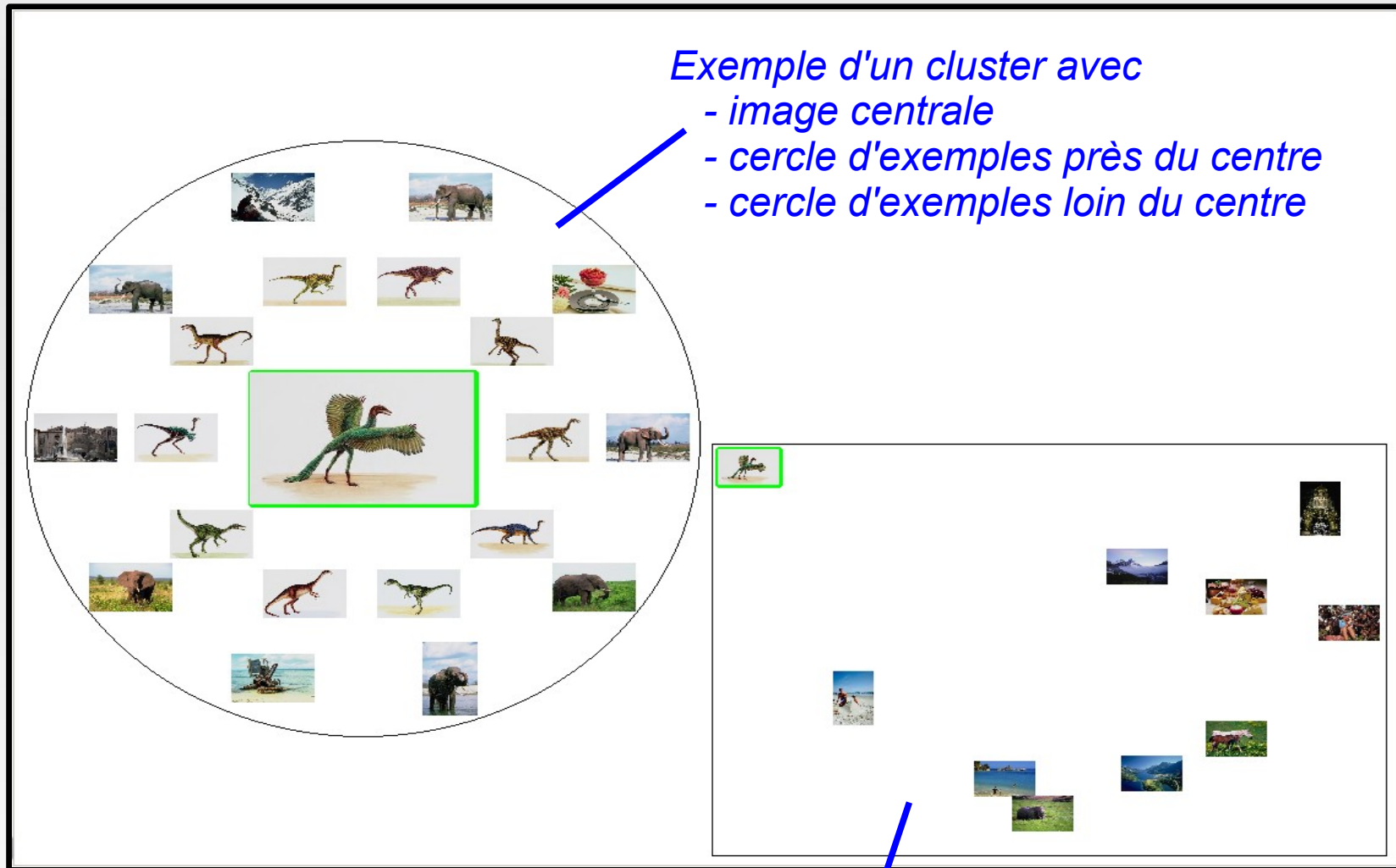
Interaction pour l'indexation de grandes bases d'images

- Système interactif de structuration des index pour une recherche par le contenu dans des grandes bases d'images
 - *Thèse de Lai Hien Phuong (L3i / IFI-MSI) [2009-2012]*

Contexte

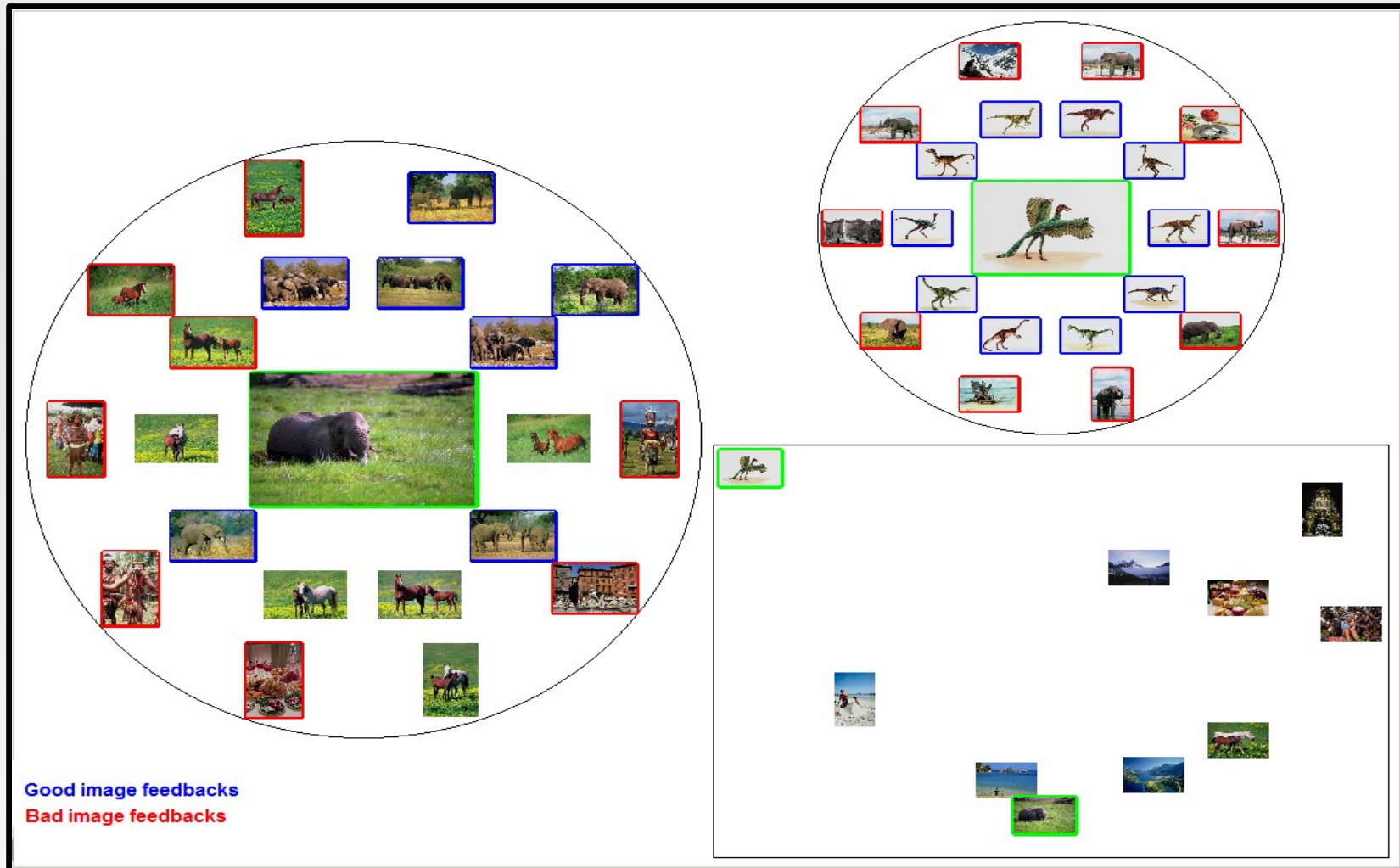
- **Indexation de bases d'images de plus en plus grandes**
 - mais besoin d'une assistance humaine pour l'indexation
 - tout en minimisant le nombre d'interactions nécessaires
- **Idée : bouclage entre**
 - le clustering effectué par le système
 - les retours de l'utilisateur
 - *Clustering interactif*
- **Différentes variantes du clustering interactif**
 - Retour positif : exemple dans le bon cluster
 - Retour négatif : avec indication du cluster pour déplacer
 - *création de paires « validées » (exemple, cluster)*

Clustering initial (automatique)



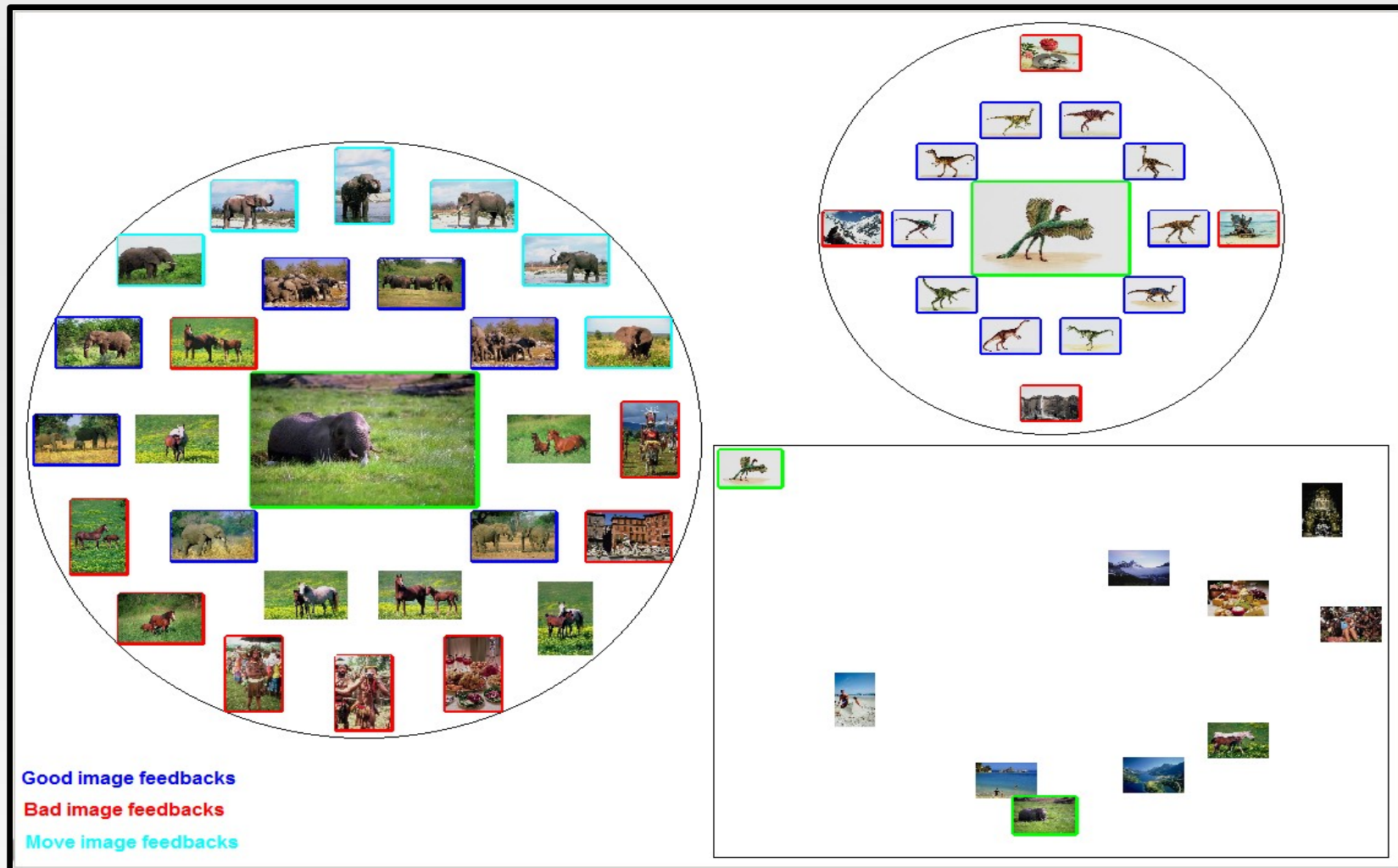
Pour un cluster, quelques exemples (21) seulement sont montrés sur les dizaines que contient le cluster

Interaction avec l'utilisateur



- L'utilisateur indique ses retours pour quelques clusters
 - *bons exemples pour renforcer la classe du cluster*
 - *mauvais exemples à déplacer vers d'autres clusters*

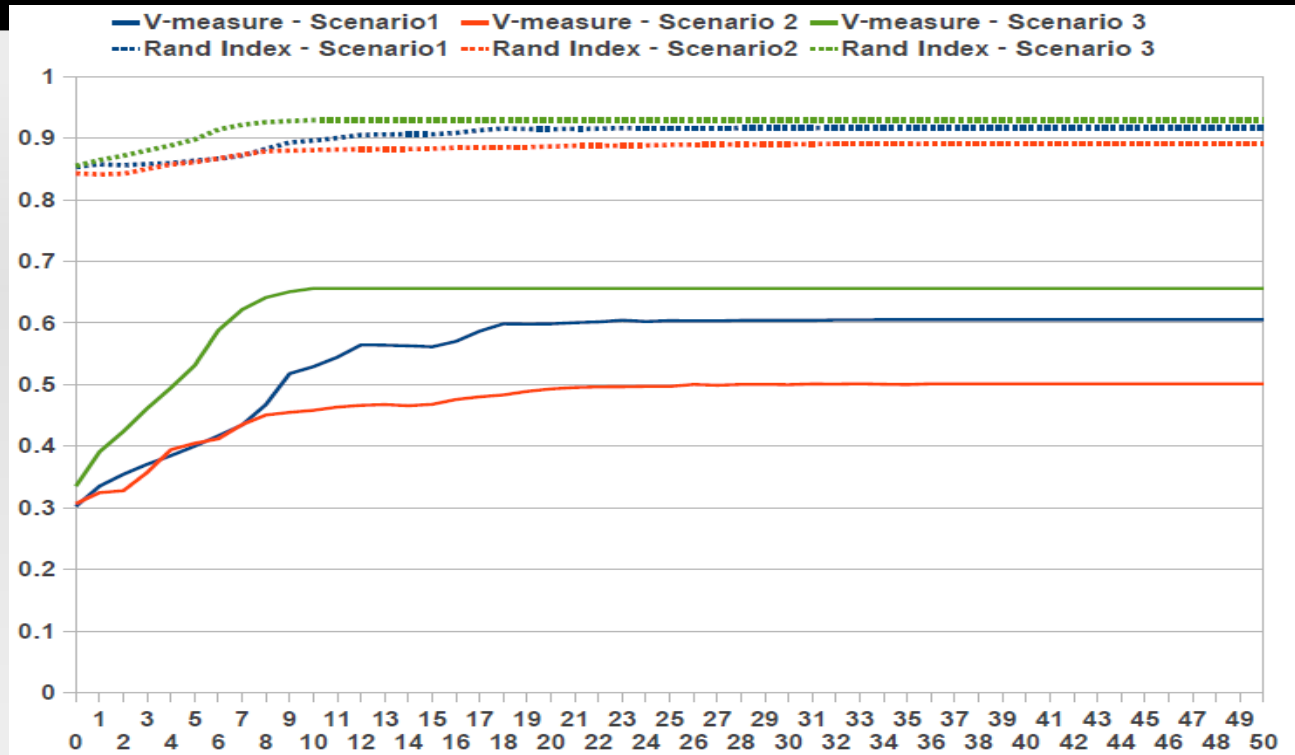
Re-clustering tenant compte des retours



- Clustering basé sur une minimisation de fonction
 - *tenant compte des distances entre points et de contraintes (retours)*
 - *défi : corriger plusieurs erreurs à chaque interaction*

Résultats

V-measure
Rand Index



Nb interactions

- **Trois scénarii d'interactions en clustering automatique**
 - Nb de retours utilisateurs, retours positifs / négatifs, ...
 - Base d'images de Wang (10 classes x 100 images/classe)
- **Deux mesures « externes » (tenant compte de la vérité terrain)**
 - V-measure et Rand Index
- **Amélioration sensible et rapide de la qualité de l'indexation**
- **Perspective** : remise en cause de la vérité terrain (limite du système)

Collaboration entre processus

Collaboration entre processus

- *Exemples avec applications spécialisées*

Analyse vidéo + Recherche d'informations

- Indexation et recherche de vidéos pour la vidéosurveillance
 - *Thèse de Le Thi Lan (INRIA Sophia Antipolis / MICA) [2005-2009]*

Multimodalité (texte + parole)

- Vers une reconnaissance multimodale du texte et de la parole pour l'analyse des documents vidéos pédagogiques
 - *Thèse de Thomas Martin (L3i / MICA) [2004-2009]*

Collaboration entre processus

- *Exemples avec applications spécialisées*

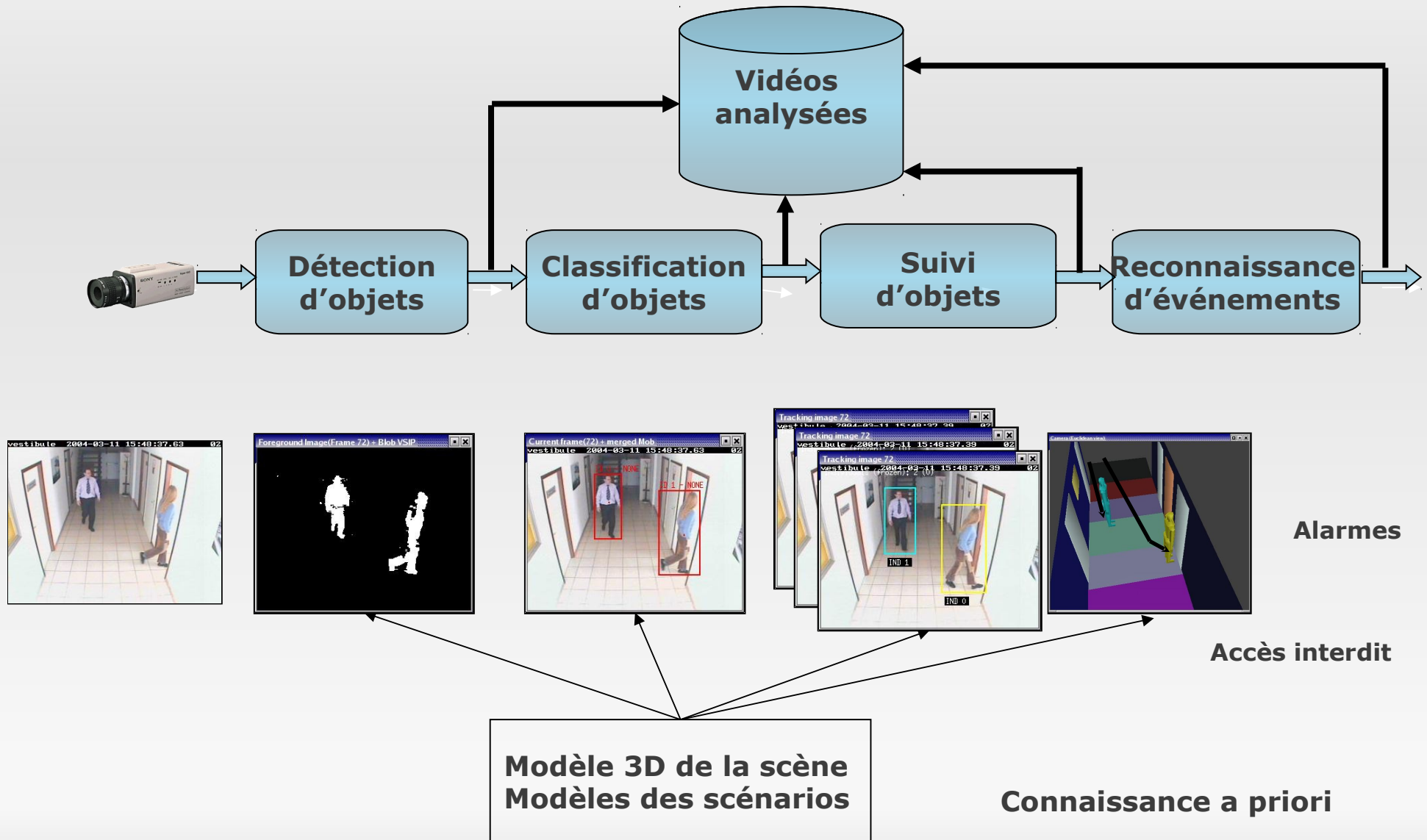
Analyse vidéo + Recherche d'informations

- Indexation et recherche de vidéos pour la vidéosurveillance
 - *Thèse de Le Thi Lan (INRIA Sophia Antipolis / MICA) [2005-2009]*

Multimodalité (texte + parole)

- Vers une reconnaissance multimodale du texte et de la parole pour l'analyse des documents vidéos pédagogiques
 - *Thèse de Thomas Martin (L3i / MICA) [2004-2009]*

Analyse vidéo pour la vidéosurveillance

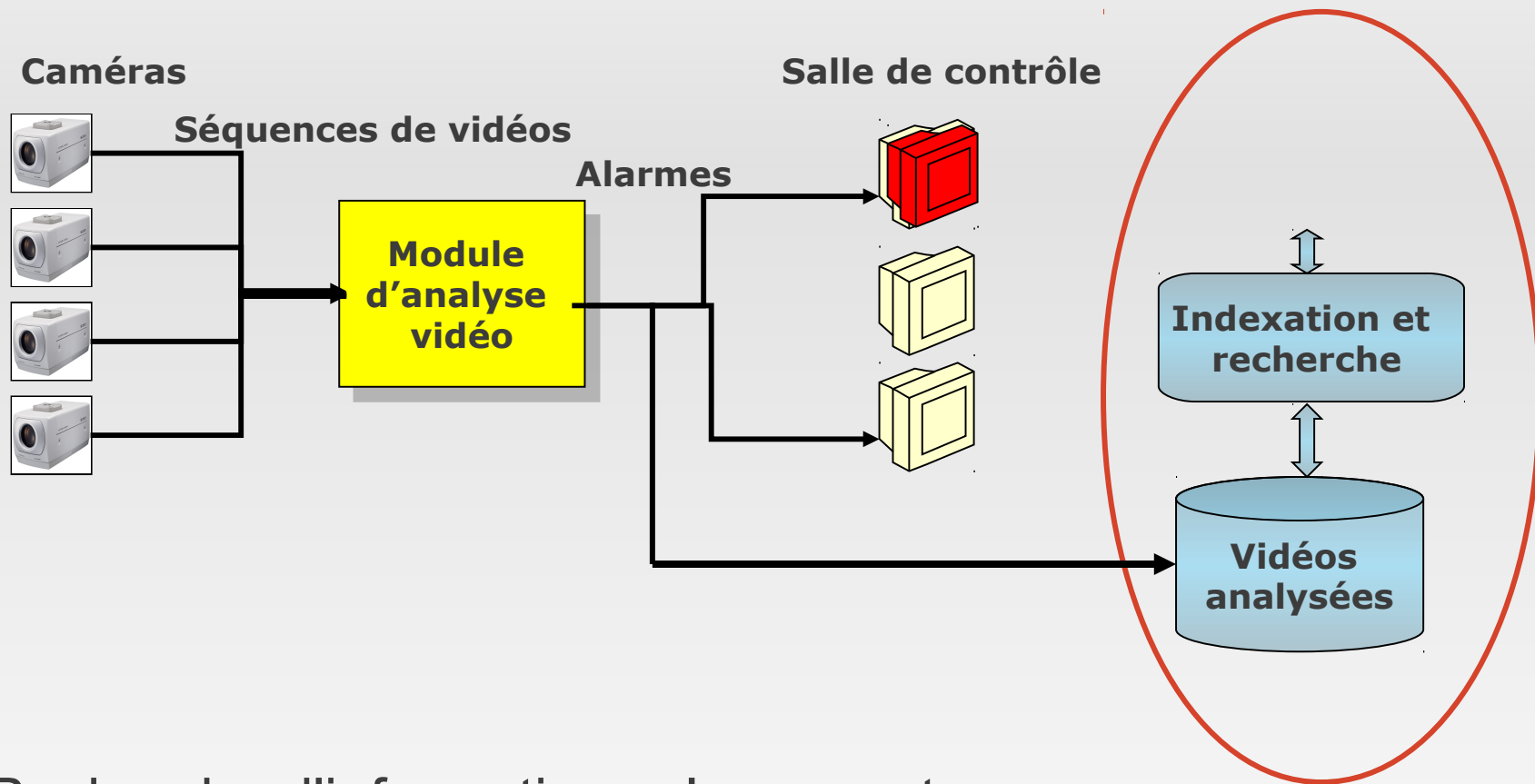


Source: [Thonnat 07]

Indexation et recherche en vidéosurveillance

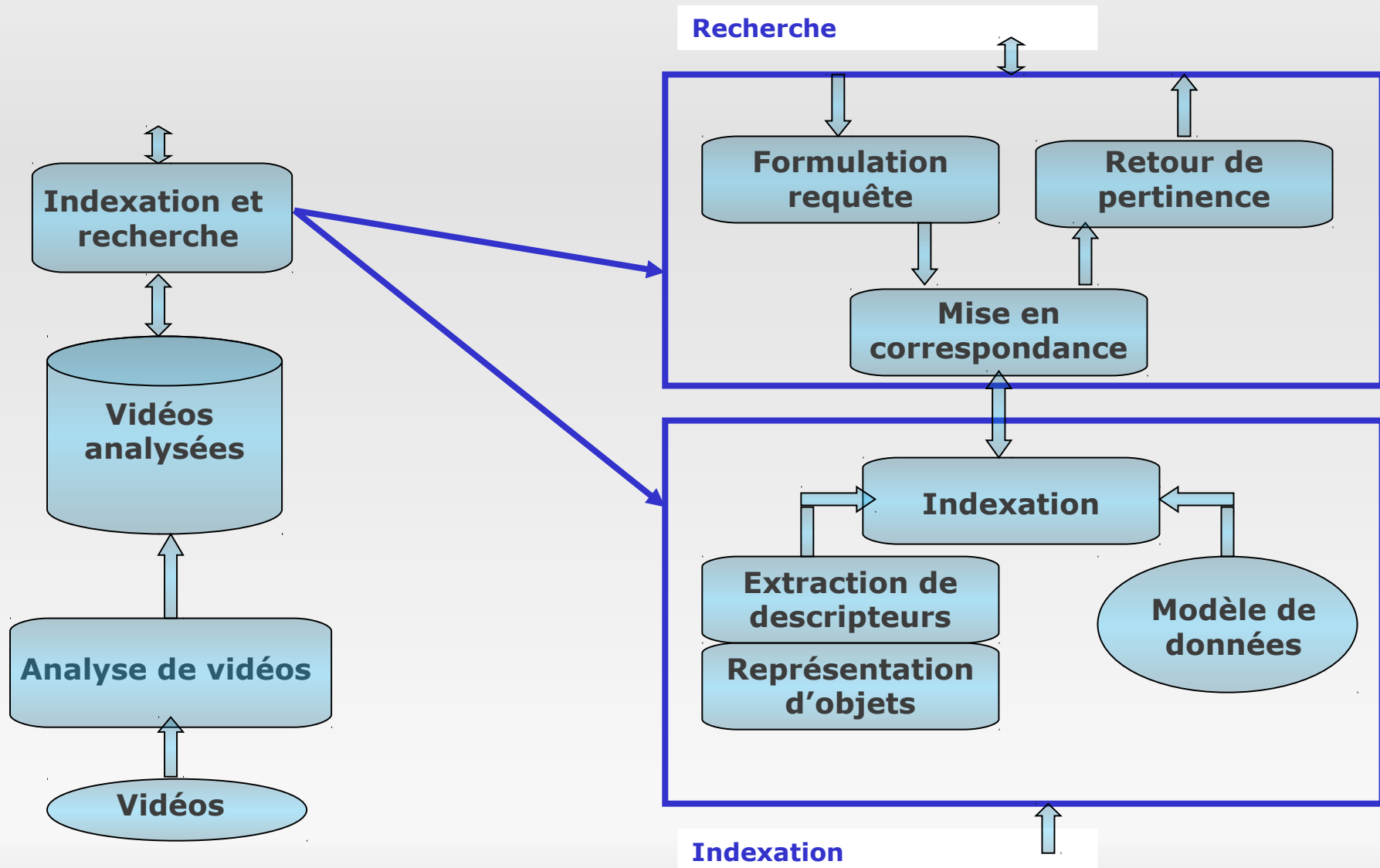
- **Modèle existant d'analyse vidéo**
 - Détection et suivi d'objets
 - Détection d'événements sur l'évolution et l'interaction des objets
 - Connaissances a priori
 - *Contexte : modélisation de la scène vide*
 - *Domaine : objets et événements d'intérêt pour l'application*
 - Objectif d'assister un opérateur humain (alarmes)
- **Application spécialisée**
 - Besoin de connaissances définies a priori
 - Pas/peu d'évolution de ces connaissances pendant l'exécution
 - Difficile de rechercher de nouvelles informations a posteriori
- **Proposition** : ajout d'un module de recherche d'informations

Extension pour la recherche d'informations



- Recherche d'informations s'appuyant sur
 - les séquences vidéos d'origine
 - les informations extraites par le module d'analyse vidéo
- Permettre l'exploitation des données archivées (vidéos + analyse)₃₈

Modèle proposé



Indexation en vidéosurveillance

Représentation des blobs



2 méthodes pour la détection des blobs représentatifs (objets en mouvement)

A - basée sur le changement d'apparence

- comparaison de distances
- efficace si bons résultats du module d'analyse vidéo

Distance

frame

B - basée sur le regroupement des blobs

- classification SVM + CAH
- plus lourd, hors ligne et supervisée
- compense les erreurs d'analyse

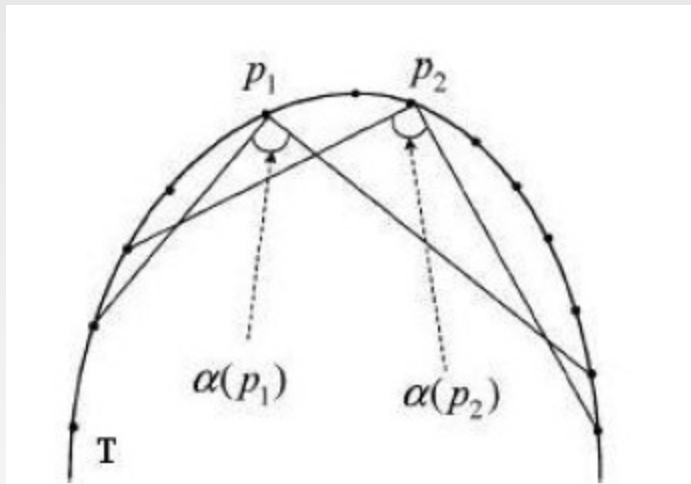


Indexation en vidéosurveillance

Représentation des trajectoires

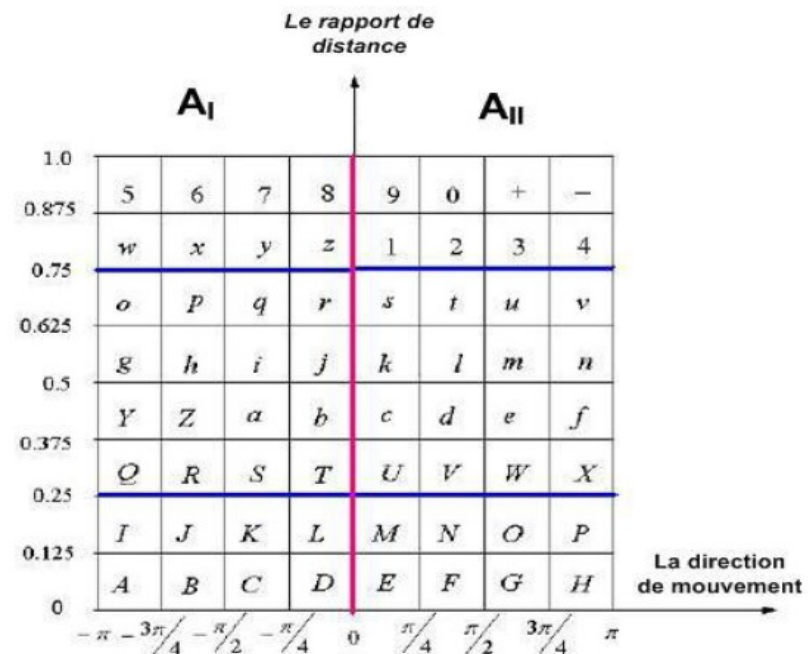
A – représentation numérique

- découpage en sections
- incréments de position relative



B – représentation symbolique

- discrétisation dans le plan direction/distance
- recherche similaire au texte



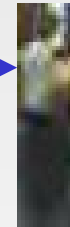
Recherche en vidéosurveillance

Langage de requête spécialisé

- Langage de requête spécialisée
 - Requêtes à 3 niveaux : images, objets et événements
 - Langage de type SQL pour formaliser différents types de requêtes

Requête : Retrouver une personne similaire à l'imagette et proche du distributeur de billets

```
SELECT * FROM CARE_1
WHERE :
ENTITIES ↑
((s : SubImage), (p : PhysicalObjects), (e: Events))
CONDITIONS
((p's Class ="Person") (e's Name =« close_to_VendingMachine1")
(s edge_histogram p))
```

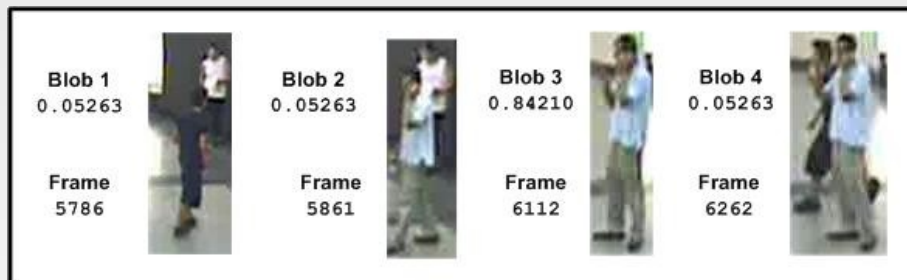


Imagette

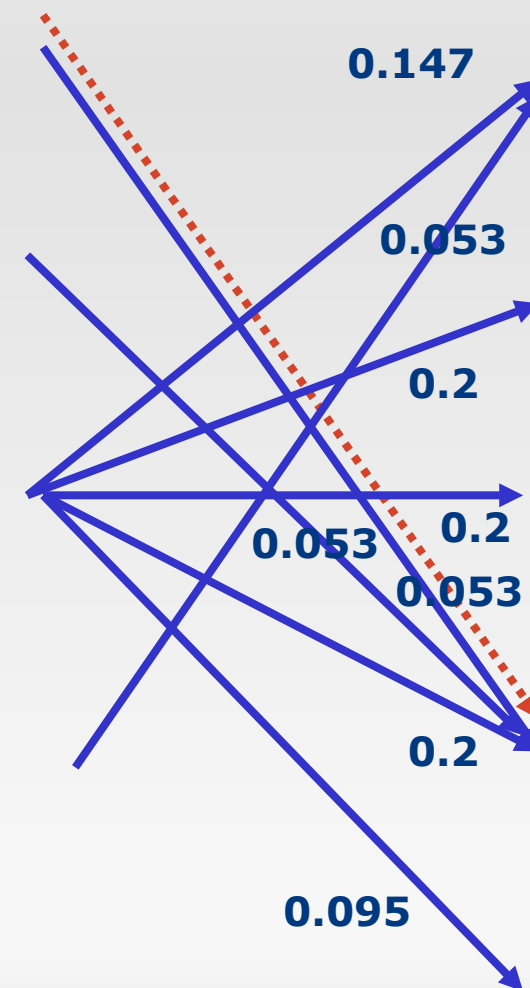
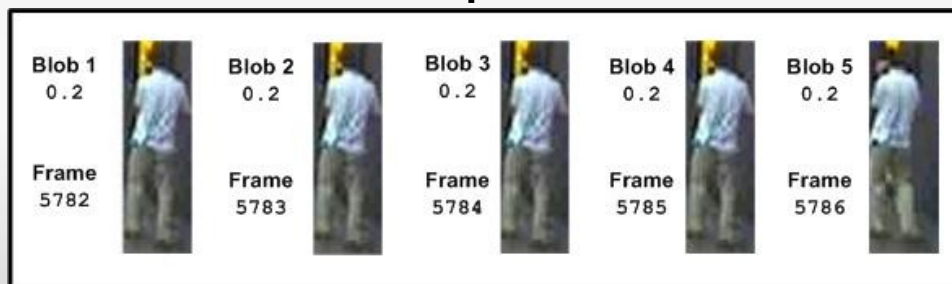
Recherche en vidéosurveillance

Mise en correspondance d'objets

- Mise en correspondance d'objet
 - Distance basée sur EMD
 - Mise en correspondance partielle
 - Prise en compte des poids des blobs
 - Suivi d'objets ayant une grande valeur de confusion d'étiquettes

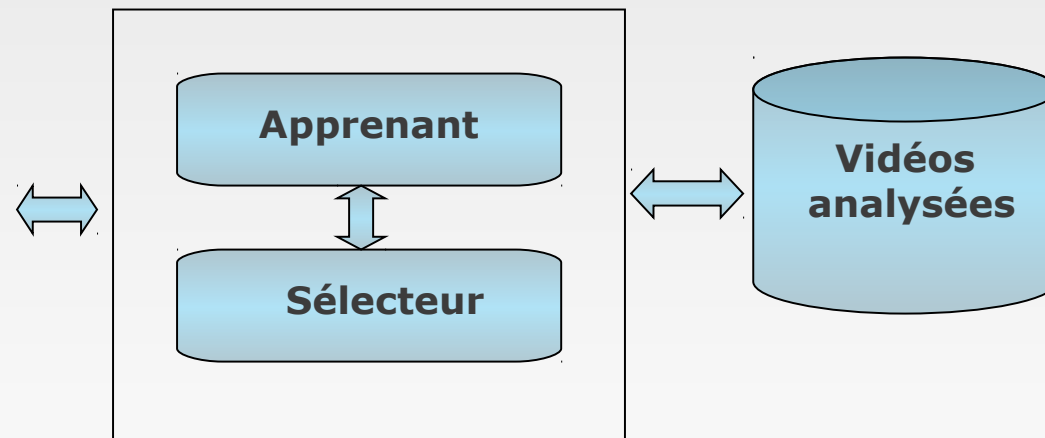


↓
mise en correspondance
↑



Recherche interactive (retour de pertinence)

- Basée sur la sélection par l'utilisateur de blobs exemples positifs
- Deux méthodes possibles
 - 1 - Modification de l'objet requête avec mise en correspondance EMD
 - 2 - Classification utilisant un SVM à une classe
- Peut exploiter (ou non) les informations de classes du module d'analyse vidéo



Résultats

- **Recherche d'informations spécialisée en vidéosurveillance**
 - Contraintes du domaines
 - Exploitation des informations générées par l'analyse
 - Complète efficacement les résultats de l'analyse vidéo
- **Modèle générique dans le cadre de ce domaine**
 - Peut exploiter (ou non) différents types de données des vidéos
 - Peut pallier les erreurs de l'analyse vidéo antérieure
- **Complémentarité**
 - Analyse vidéo utilisable en temps réel
 - Recherche d'informations plus adaptée a posteriori

Collaboration entre processus

- *Exemples avec applications spécialisées*

Analyse vidéo + Recherche d'informations

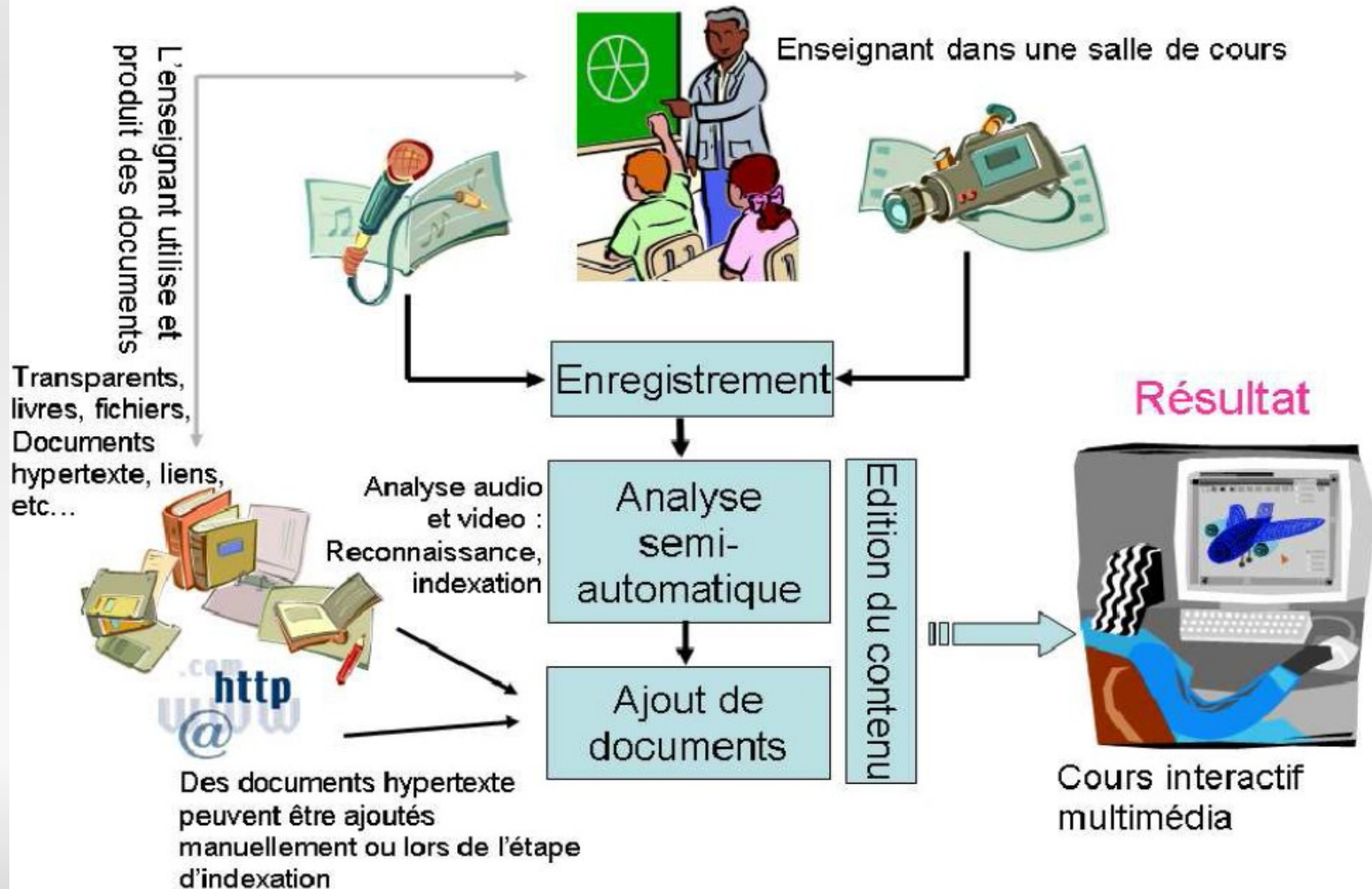
- Indexation et recherche de vidéos pour la vidéosurveillance
 - *Thèse de Le Thi Lan (INRIA Sophia Antipolis / MICA) [2005-2009]*

Multimodalité (texte + parole)

- Vers une reconnaissance multimodale du texte et de la parole pour l'analyse des documents vidéos pédagogiques
 - *Thèse de Thomas Martin (L3i / MICA) [2004-2009]*

Analyse de vidéos pédagogiques

Multimodal Analysis of Recorded Video for E-Learning



Modalités existantes

■ Audio

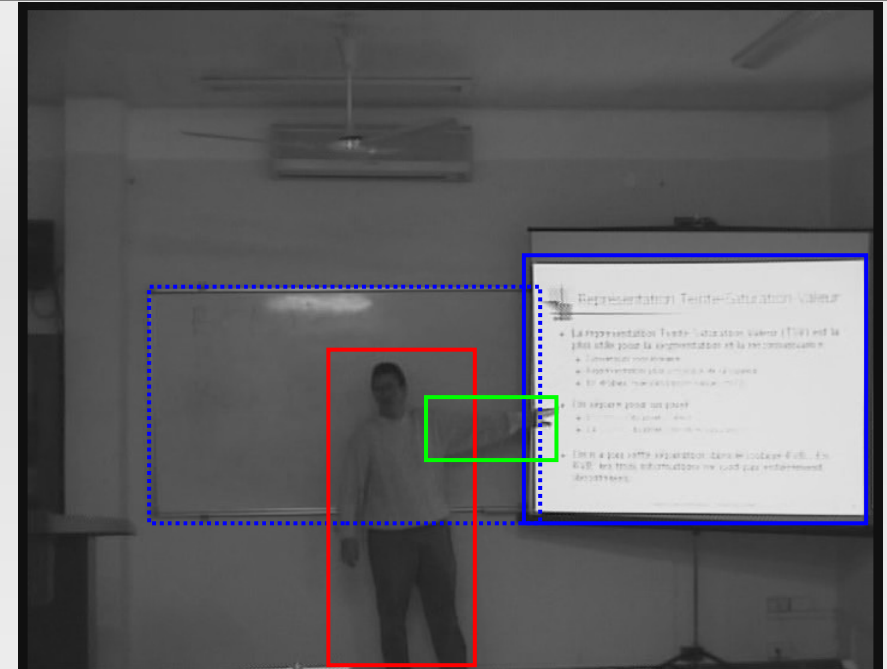
- Parole
- Bruit, silence

■ Vidéo

- Transparents
 - structure, graphiques
- Tableau
 - écriture
- Enseignant
 - gestuelle

■ Texte

- Texte imprimé (transparents)
- Texte manuscrit (tableau)
- Documents extérieurs



Représentation Teinte-Saturation-Valeur

- La représentation **Teinte-Saturation-Valeur** (TSV) est la plus utile pour la segmentation et la reconnaissance.
 - Conversion non-linéaire
 - Représentation plus *physique* de la couleur
 - En anglais *Hue-Saturation-Value* (HSV).
- On sépare pour un pixel
 - L'*intensité* du pixel (*valeur*)
 - La *couleur* du pixel (*teinte + saturation*).
- On n'a pas cette séparation dans le codage RVB. En RVB, les trois informations ne sont pas entièrement décorréliées.

→

13

Vision par ordinateur - Alain Boucher

Interactions entre les modalités

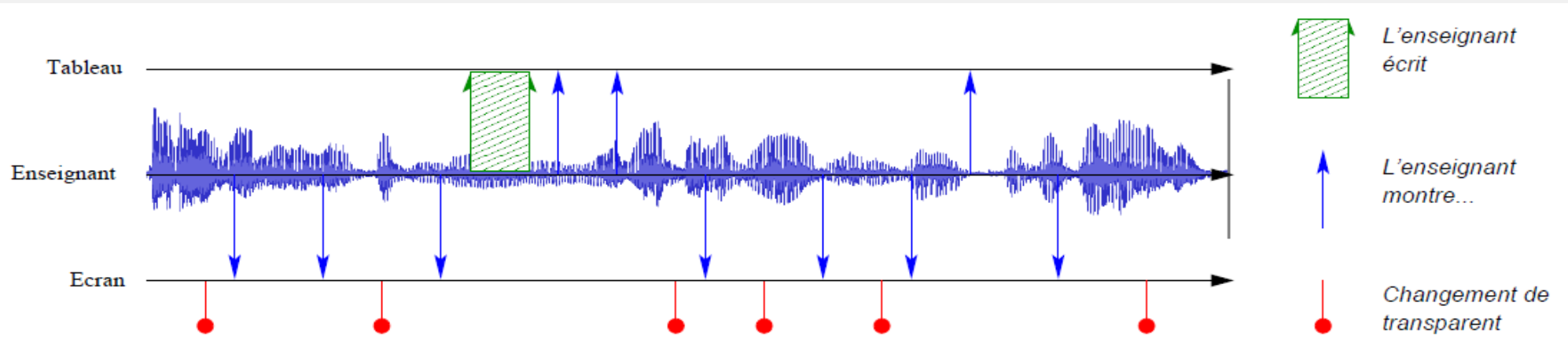
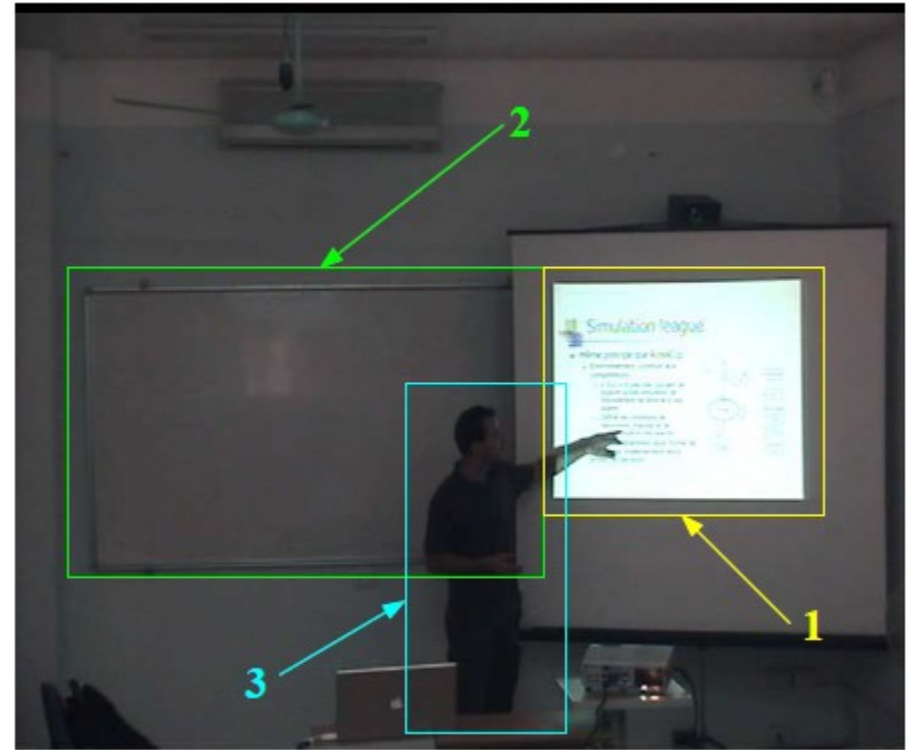
Deux types d'informations :

1 – Scénario du cours

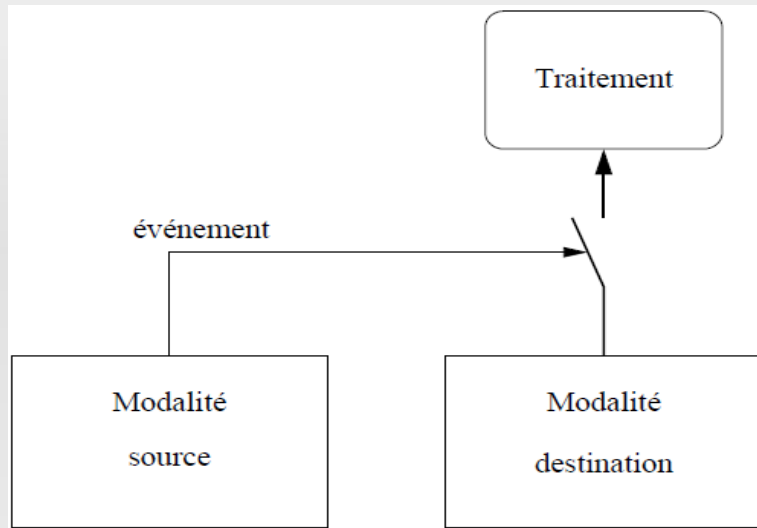
- séquences d'événements

2 – Contenu du cours

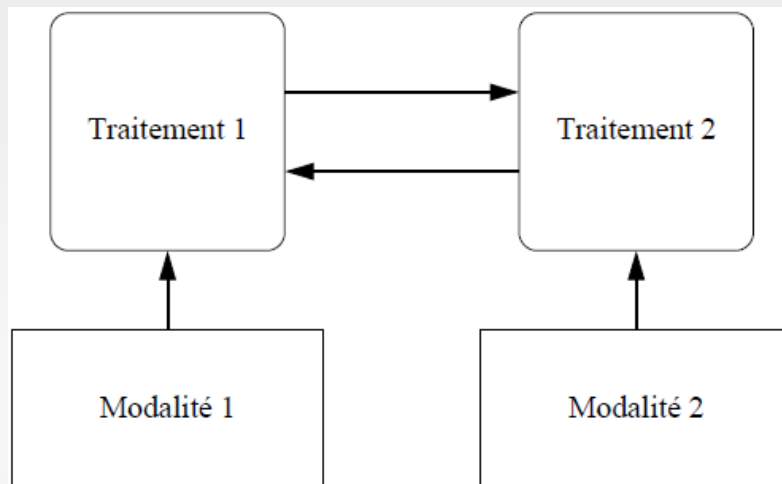
- Informations parole + texte



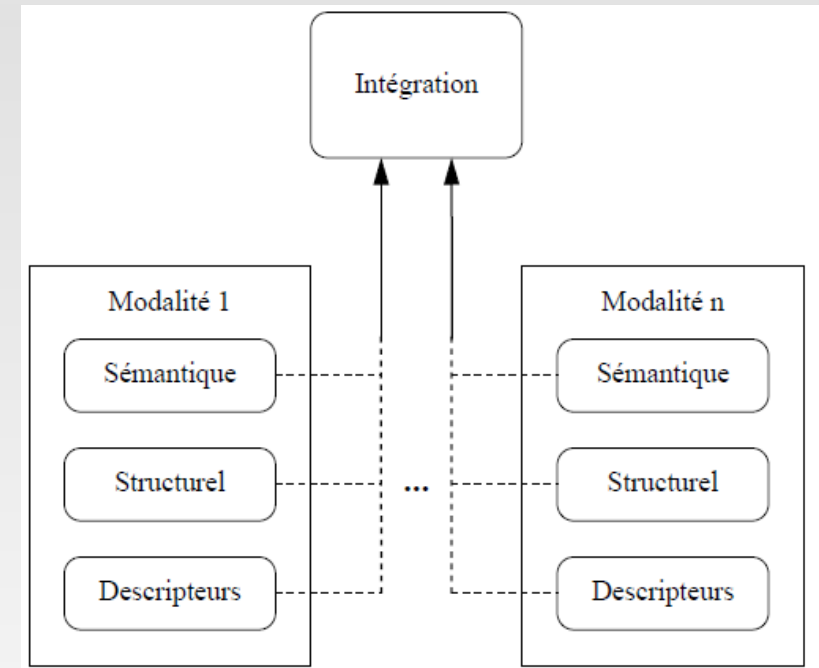
Trois types de relations entre les modalités



1 - Déclenchement



3 - Collaboration

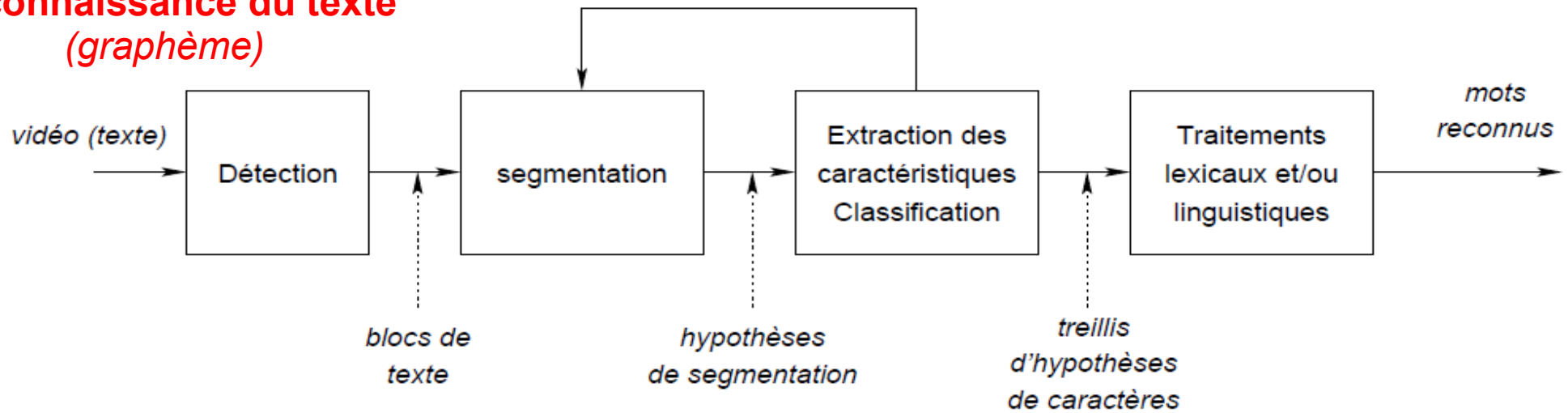


2 - Intégration

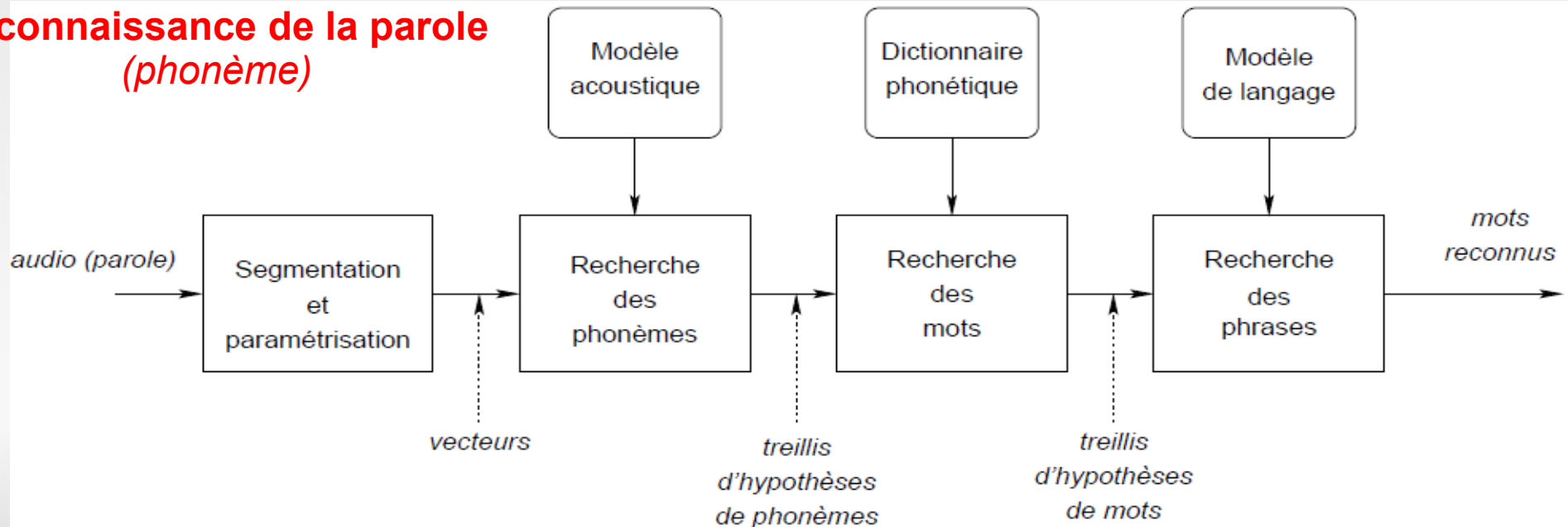
Collaboration texte + parole

Reconnaissance du contenu du cours

Reconnaissance du texte (graphème)



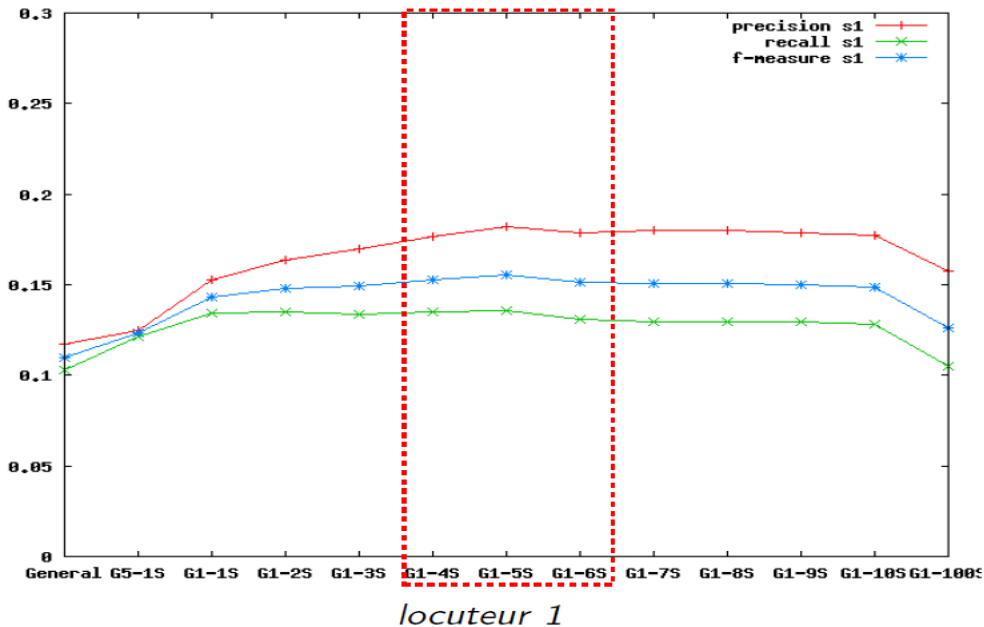
Reconnaissance de la parole (phonème)



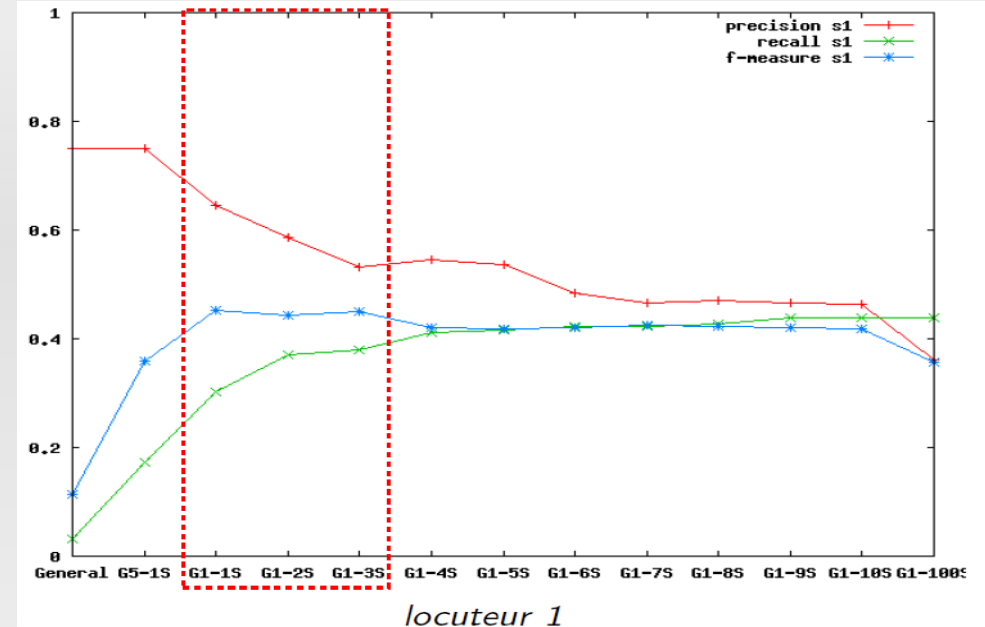
Collaboration texte + parole

- Relation unidirectionnelle
 - **Processus fort (texte) → processus faible (parole)**
- Idée : Thématiser un modèle générique du langage parlé avec un modèle spécialisé de langage écrit
 - *Moteur de reconnaissance de la parole générique*
 - *Extraction du vocabulaire spécialisé dans le texte*
- Construction d'un corpus spécialisé avec les transparents
 - *Corpus de faible taille*
- Reconnaissance du texte → reconnaissance de la parole
 - *Phonétisation des mots reconnus (texte)*
 - *Recherche dans le treillis d'hypothèses de phonèmes (parole)*

Résultats



Résultats sur tous les mots



Résultats sur les mots-clés

- Amélioration sensible du taux de reconnaissance de la parole
- Cependant insuffisant pour la transcription automatique
- **Perspective** : recherche d'informations dans une base de cours

Conclusions & perspectives

Conclusions

- **Différents modèles** intégrant interaction et collaboration
 - *Différents niveaux de spécialisation d'application*
- Humain dans la boucle permet d'**intégrer le contexte**
 - Apprentissage plus souple des connaissances
- Collaboration entre processus permet de **pallier les faiblesses**
 - *Différents modèles de collaboration selon les cas*
- Interactivité demande des **outils adaptés de validation**
 - *Simulation de l'interaction par des agents logiciels*
 - *Construction de scénarii plausibles d'exécution*

Perspectives : visualisation interactive

- **Modèles de visualisation interactives**

- *Besoin de modèles d'interfaces adaptés pour interagir avec de multiples sources d'informations (images, texte, ...)*



- Exemple de modèle SMA basé sur des forces d'attraction et de répulsion pour la recherche d'images

Stage Master 2 de Guillaume Chiron

Perspectives : niveaux de sémantique

- Capacités du système reposant sur le caractère visuel des images (« visualness »)
 - Sémantique reposant autant sur des aspects visuels qu'abstraits
 - Identifier une sémantique à plusieurs niveaux

Recherche d'image par la requête textuelle: « Ours+Neige »



Recherche d'image par la requête textuelle: « Peuple+Rue »



Perspectives : Internet

- Internet comme source de connaissances
 - *Internet source de connaissances du domaine*
 - *Complété par l'humain pour le contexte*

*Projet Master 1 IFI (TPE)
de Tran Thi Cam Giang*

