

Masses de Données en Astronomie (MDA)

Le Centre de Données astronomiques de Strasbourg

- Créé en 1972
- Développe des bases de données et des outils largement utilisés par la communauté astronomique internationale (30.000+ requêtes/jour)
- Liens CDS < > archives d'observation, journaux
- Astronomie: Données complexes, massives, hétérogènes, distribuées, essentiellement libres de droit (images, spectres, séries temporelles, cubes, textes, tables, ...) >>> Observatoire Virtuel

Relations avec des laboratoires STIC

- Depuis de nombreuses années:
 - LSIIT: images
 - IRIT: données textuelles
 - Comment labelliser et développer les interactions avec les laboratoires STIC (cf OV à l'étranger)?
 - Philosophie gagnant/gagnant:
Science STIC/nouvelles méthodes pour astro
- ACI GRID 2001: Images Distribuées Hétérogènes en Astronomie (essentiellement avec LSIIT)
 - Modèle de données
 - Images multispectrales

Le projet MDA

- Contexte Observatoire Virtuel Astronomique
- Trois aspects principaux:
 - Traitement d'images hyperspectrales
CDS, LSIIT, OCA, IAP
 - Ontologies, Web sémantique
CDS, IRIT, LORIA, Obs. Paris, CESR
 - Calcul et données distribuées
CDS (A. Schaaff), LSIIT (S. Genaud), Terapix (E. Bertin) *démarre*
- Plusieurs études intégrées progressivement

CDS

F. Genova
F. Bonnarel
A. Lançon
R. Ibata
Ch Pichon
C. Bot
B. Vollmer

OCA

A. Bijaoui
C. Benoist
E. Slézak
2 stagiaires

B. Guiderdoni, E. Bertin, Y. Mellier, A. Baillard (IAP)
J.-L. Starck (CEA, Orsay)
W. Pieczynski, A. Chardin (INT Paris)
G. Lagache (IAS Orsay)
F. Murtagh (QUB, Belfast)

Traitement d'images hyperspectrales



J.-J. Claudon (IE 2004-2005)



LSIIT

Ch. Collet (PU 2001)
A. Jalobeanu (CR CNRS 2004)
M. Louys (MdC)
F. Salzenstein (MdC 2002)
Ch Wolf (MdC 2004)

F. Flitti (Doct 2002)
M. Petremand (Doct 2003)

M. Hatt (DEA Info 2004)
O. Marchal (DEA Astro 2004)
M. Petremand (DEA Info 2003)
K.-P. Maalej (DEA Astro 2003)
A. Mokhtari (DEA PIC 2002)

Publications
Communications
Logiciel

Traitement de données images hyperspectrales (composante spectrale/spatiale)

Cubes de données multi/hyperspectraux

Visualisation de masses de données

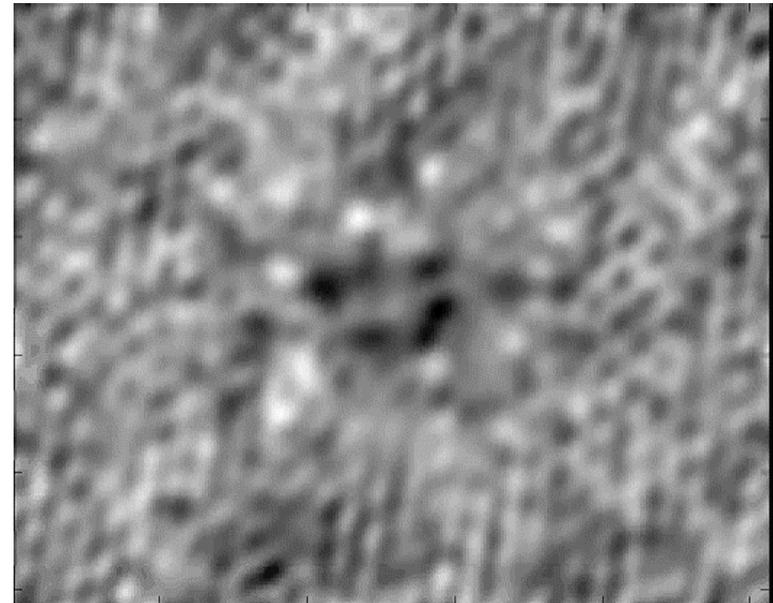
Segmentation markovienne avec interaction
permettant une convergence rapide

Techniques de réduction de données

Modélisation de l'information a priori:
champs de Markov flous

Algorithmes génétiques

Décomposition en ondelettes + quadarbre
markovien



48 bands around CO ray of the GG
Tauri system from the IRAM
interferometer

Image processing team

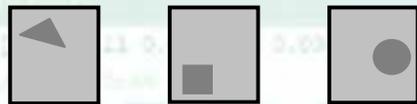
Multispectral Image analysis :



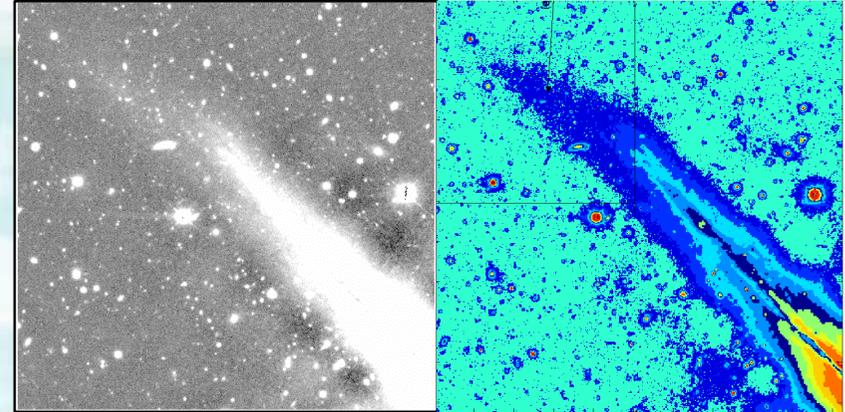
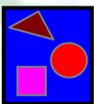
Bottom to up clustering
algorithm using
multiscale similarity
measure based on
normalized histograms
and barycenters



Reduction step :
PCA /ICA/Projection Pursuit
Gaussian Mixture model



Markovian
Segmentation



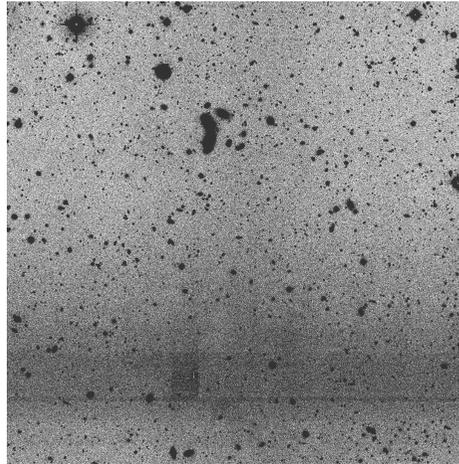
Example of application :

(F. Flitti, C. Collet, B. Vollmer, F. Bonnarel)
21cm data cube analysis of a Virgo
cluster galaxy based on Kinematic
segmentation using parameter maps

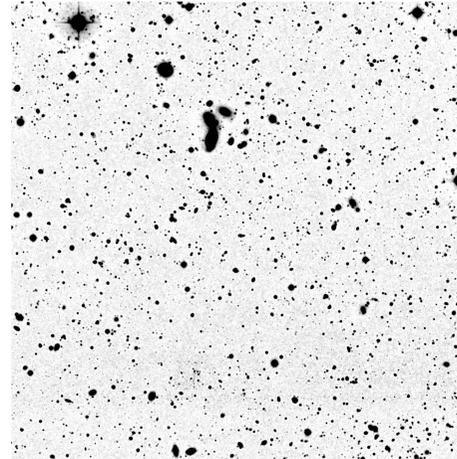
A l'OCA: aide à l'utilisation de nombreuses images d'un même champ

- Améliorer la détection
- Mesurer les objets
- Identifier des classes de pixels ou d'objets
- Problème des jeux d'images

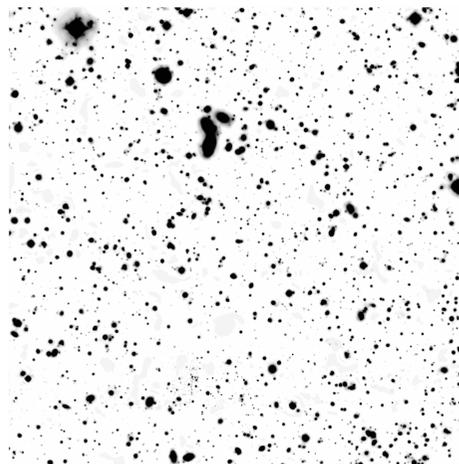
Détection (OCA): images de fusions



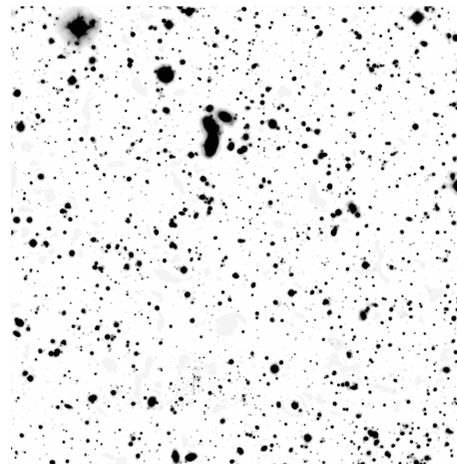
Brut



Débruitage simple



Lissage PSF

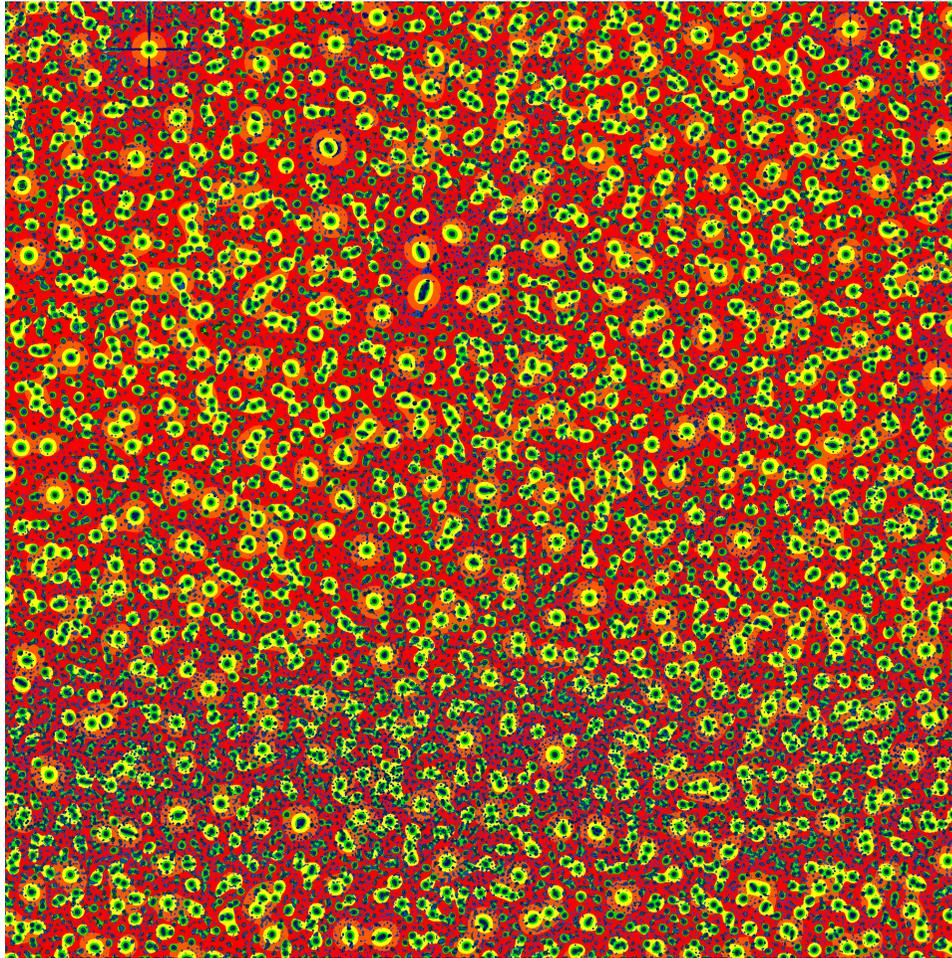
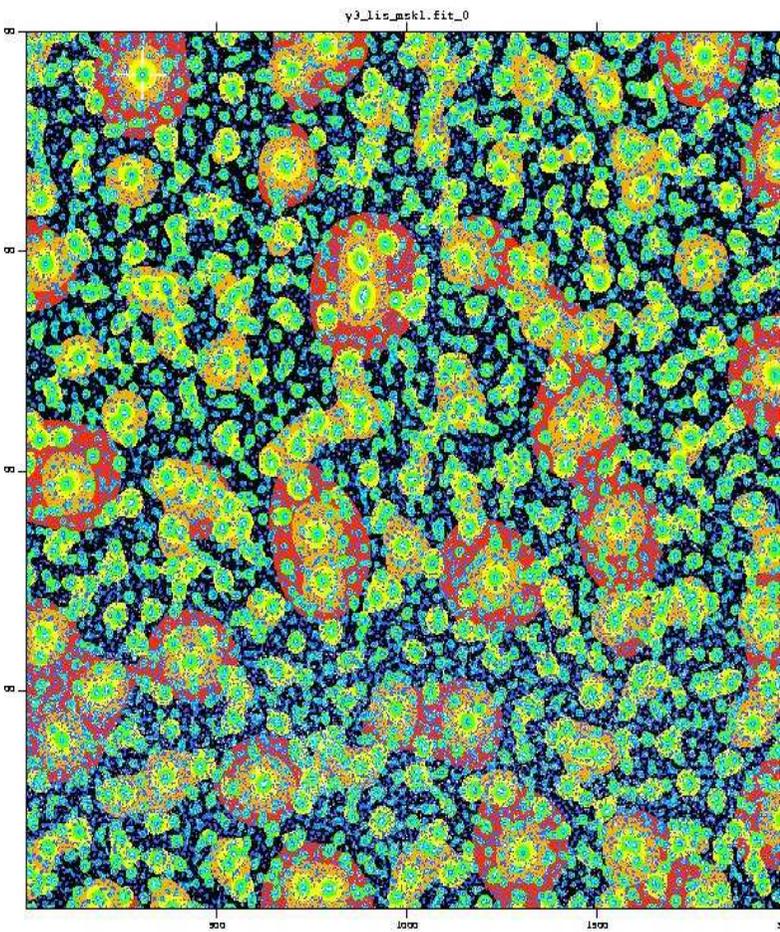


Déconvolution
Puis lissage PSF

Extraction: Superposition des plans des coefficients en ondelettes seuillés

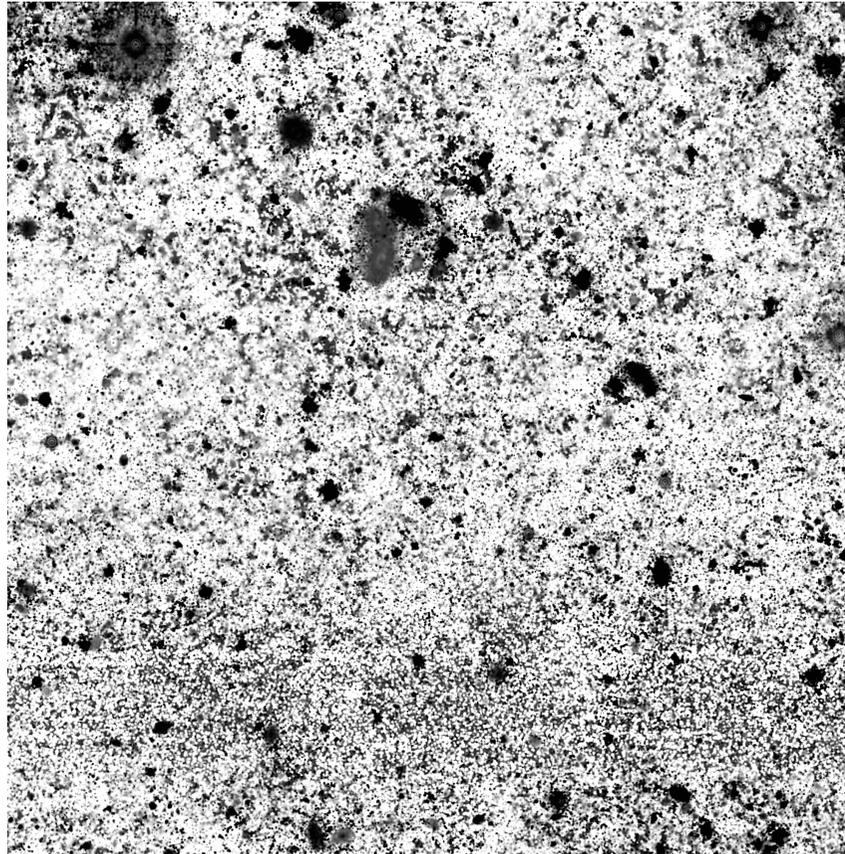
Masque de l'ensemble des échelles

Masque avec les petites échelles

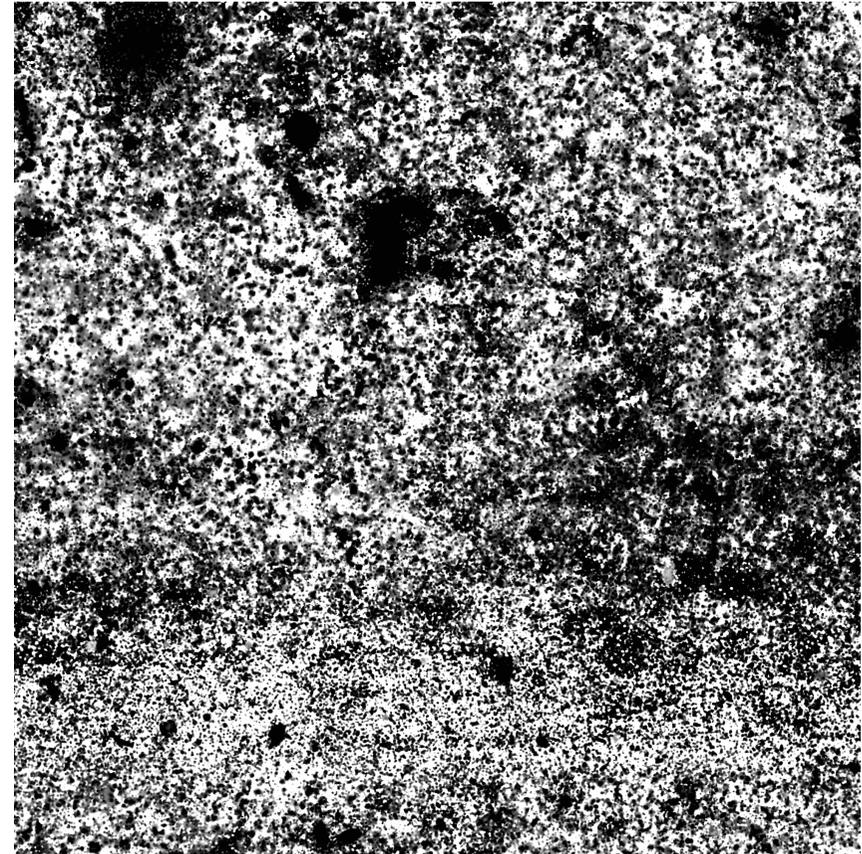


Analyse multibande: Résultat de l'ACP Pondérée

Composante 1

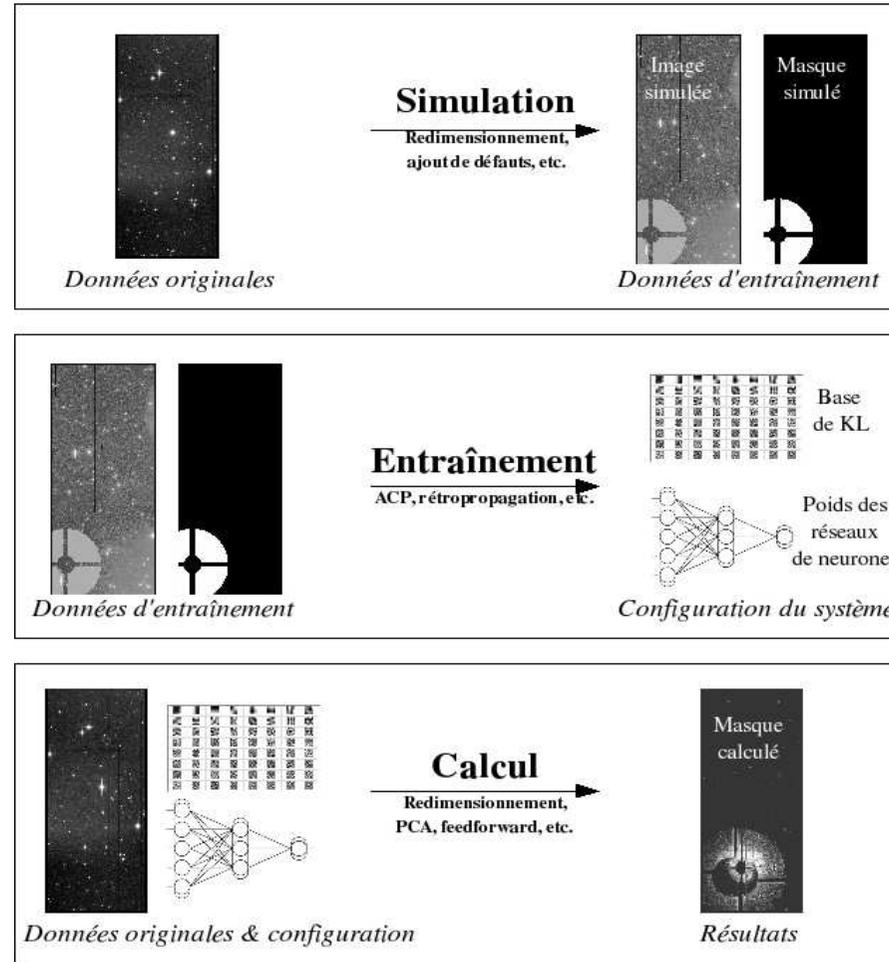


Composante 2



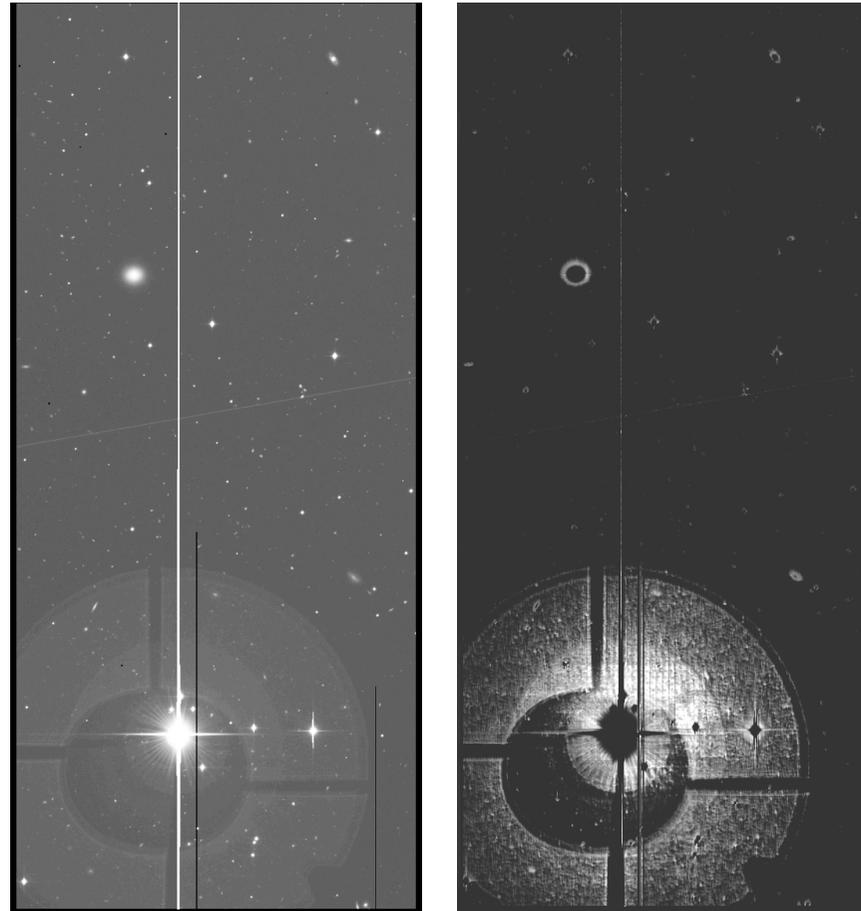
On voit bien sur la composante 1 la différence de couleur des 2 galaxies.
Le bruit est important et il augmente avec l'ordre de l'ACP

Terapix: Global process

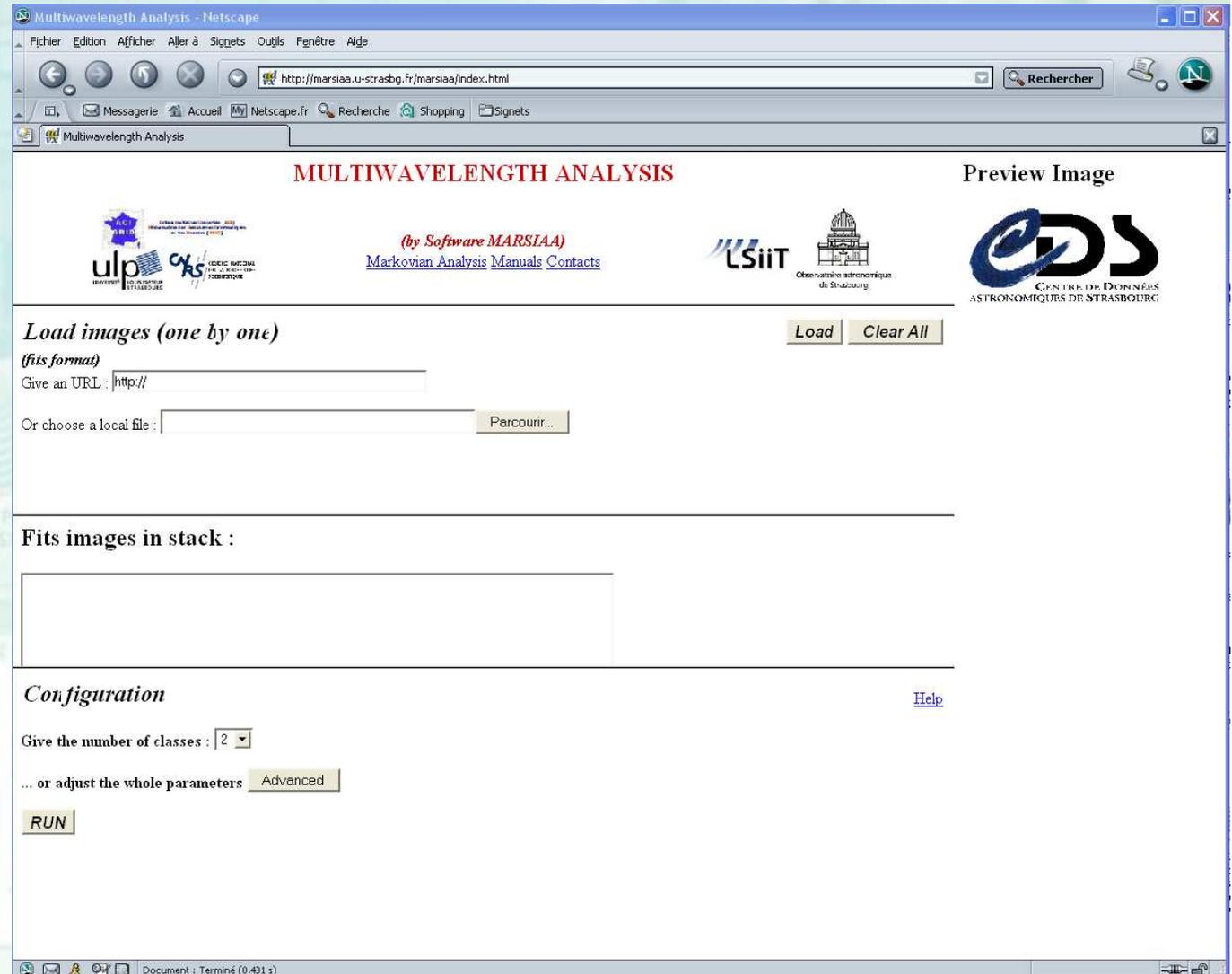


Masks

- Halo example
 - Rebin 4x4
 - 16 PC on 64
 - Network
 - 16 input nodes
 - 8 hidden nodes
 - 1 output node



WWW interface : MARSIAA



The screenshot shows a Netscape browser window titled "Multiwavelength Analysis - Netscape". The address bar contains the URL "http://marsiaa.u-strasbg.fr/marsiaa/index.html". The page content includes the following elements:

- Page Title:** MULTIWAVELENGTH ANALYSIS
- Navigation:** "Preview Image" link in the top right.
- Logos:**
 - ULP (Université Louis Pasteur de Strasbourg)
 - LSiIT (Laboratoire de Spectroscopie Interférométrique en Géométrie Inverse)
 - Observatoire astronomique de Strasbourg
 - CDS (Centre de Données Astronomiques de Strasbourg)
- Text:** "(by Software MARSIAA) Markovian Analysis Manuals Contacts"
- Form Section:**
 - Section: "Load images (one by one) (fits format)"
 - Buttons: "Load", "Clear All"
 - Input: "Give an URL : http://" with a text box and "Parcourir..." button.
 - Input: "Or choose a local file : " with a text box and "Parcourir..." button.
- Form Section:**
 - Section: "Fits images in stack :"
 - Input: A large empty text box.
- Form Section:**
 - Section: "Configuration" with a "Help" link.
 - Input: "Give the number of classes : 2" with a dropdown menu.
 - Text: "... or adjust the whole parameters" with an "Advanced" button.
 - Button: "RUN"

Ontologies

LORIA

A. Napoli (CNRS)
Y. Toussaint (INRIA)

CDS

F. Genova
P. Dubois
S. Derriere
S. Lesteven
A. Schaaff
S. Derivaux



H. Kou (post doc INRIA, 2004)



IRIT

J. Mothe (P.)
N. Hernandez
L. Cardonner

Physique atomique et moléculaire d'intérêt astrophysique

M.-L. Dubernet et al. (Observatoire de Paris)
J. Ph Bernard, A. Walters (CESR)

Bases de connaissance en astronomie

- Différents points de vue sur le domaine astronomie
- Thesaurus (bibliothécaires) – avec IRIT
- Unified Content Descriptors (UCD): description normalisée du contenu des tables publiées, catalogues, ... - avec LORIA – UCD phys. at. et mol.: OP
- Une ontologie ‘d’équipe’
- ...

Les UCD (1)

- Articles scientifiques :
 - tables de propriétés des objets célestes (et autres tables)
 - « readme » donnant en langue ‘naturelle’ la description des différentes propriétés

<i>Full</i>	<u>RAJ2000</u>	<u>DEJ2000</u>	<u>PKS</u>	<u>PKSdata</u>	<u>RAB1950</u>	<u>DEB1950</u>	<u>n</u>
	<u>"h.m.s"</u>	<u>"d.m.s"</u>	Bytes	Format	Units	Label	Explanations
<u>1</u>	00 03 27.3	-15 47 0					
<u>2</u>	00 03 57.6	-11 49 5					
<u>3</u>	00 05 17.0	-16 48 0	1- 3	A3	---	---	PKS
			5- 13	A9	---	PKS	PKS (BHHMM-DDD) designation
<u>4</u>	00 04 35.7	-47 36 2	15- 16	I2	<u>h</u>	RAh	PKSCAT90 right ascension (B1950)
			18- 19	I2	<u>min</u>	RAm	PKSCAT90 right ascension (B1950)
<u>5</u>	00 06 13.9	-06 23 3	21- 24	F4.1	<u>s</u>	RA s	PKSCAT90 right ascension (B1950)
<u>6</u>	00 08 00.4	-23 29 1	26	A1	---	DE-	PKSCAT90 declination sign (B1950)
			27- 28	I2	<u>deg</u>	DEd	PKSCAT90 declination (B1950)
			30- 31	I2	<u>arcmin</u>	DEm	PKSCAT90 declination (B1950)
			33- 34	I2	<u>arcsec</u>	DEs	PKSCAT90 declination (B1950)
			36- 37	I2	<u>h</u>	RArh	? New radio right ascension (B1950)
			39- 40	I2	<u>min</u>	RArm	? New radio right ascension (B1950)
			42- 46	F5.2	<u>s</u>	RArs	? New radio right ascension (B1950)

Les UCD (2)

- Ensemble de propriétés de référence que les astronomes peuvent utiliser pour décrire leurs objets
 - Unifier la description des propriétés
 - Normaliser par les Unified Content Descriptor
- Problème posé : assigner à une propriété d'une table l'UCD lui correspondant
- Considérer un UCD comme une classe caractérisée par :
 - Le texte de définition de UCD
 - Les textes de description de cette propriété (dans les différents readme) qui ont été assignés à cet UCD et validés manuellement

Assignment des UCD

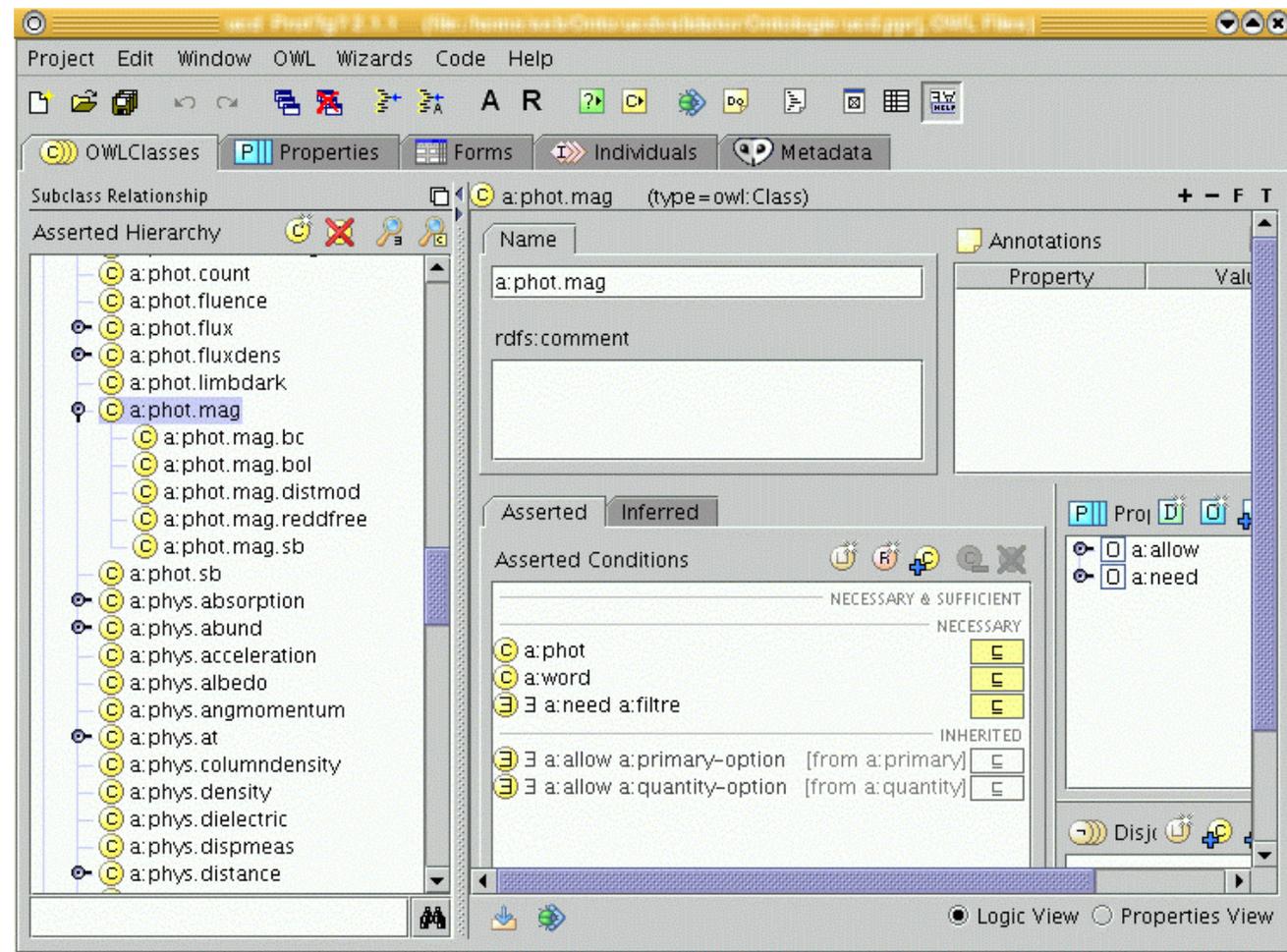
- Pour une description donnée, trouver les 3 UCD les plus proches
- Application extrême: textes très courts!
- 2 algorithmes : bayésien naïf, Rocchio
- Expérimentation :
 - 95% d'assignments correctes (93 UCD, 1500 descr.)
 - Détection d'erreurs dans les propriétés déjà validées
 - Tester sur les 1000 UCD et les 10^5 descriptions

Une nouvelle génération d'UCD

- Définition d'une ontologie des UCD
 - Considérer les UCD comme des éléments de connaissances (munis d'une syntaxe et d'une sémantique)
 - Construction d'une hiérarchie conceptuelle des UCD
 - Utilisation de l'éditeur de connaissances Protégé-2000 et du langage de représentation des connaissances OWL
 - Expérience et prototype en cours de développement (travail original dans le domaine): prise en compte des relations numériques

Ontologie prototype UCD

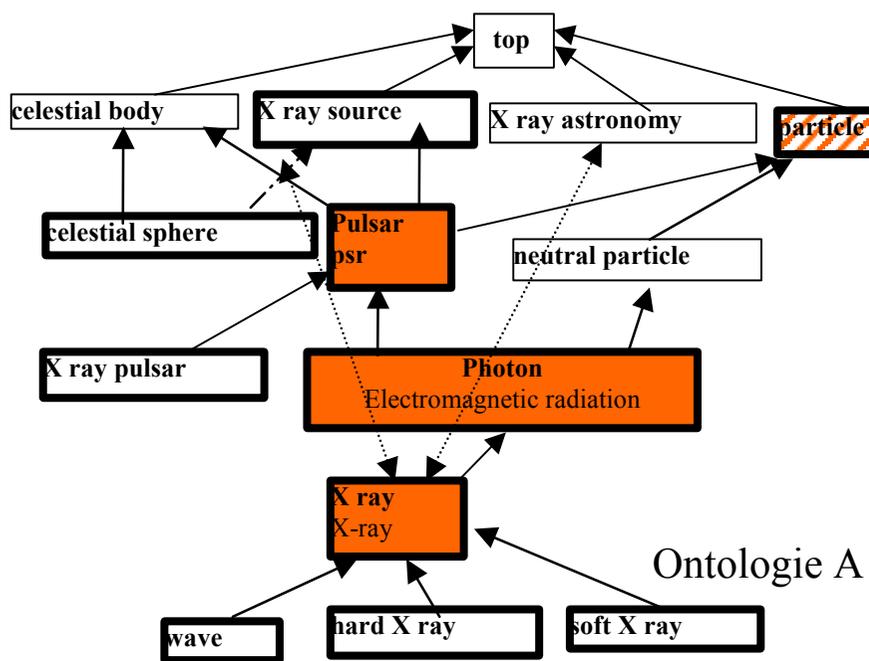
Prototype
pour
la validation
des UCD1+
avec une
ontologie



Thesaurus

- Transformation en première Ontologie (OWL)
 - Analyse : 30% des termes présents dans 4 corpus (A&A, APJ de 1995 à 2002)
 - Transformation en unités conceptuelles (vs mots)
- Connaissance de l'ontologie / connaissance des contenus
 - Paramètres de comparaison, définition de fonction d'adéquation
 - Analyse lexicale : % de couverture
 - Analyse conceptuelle: organisation des concepts, pouvoir représentatif des termes

•Corpus : 134 résumés de A&A 1995 sur “X-ray”



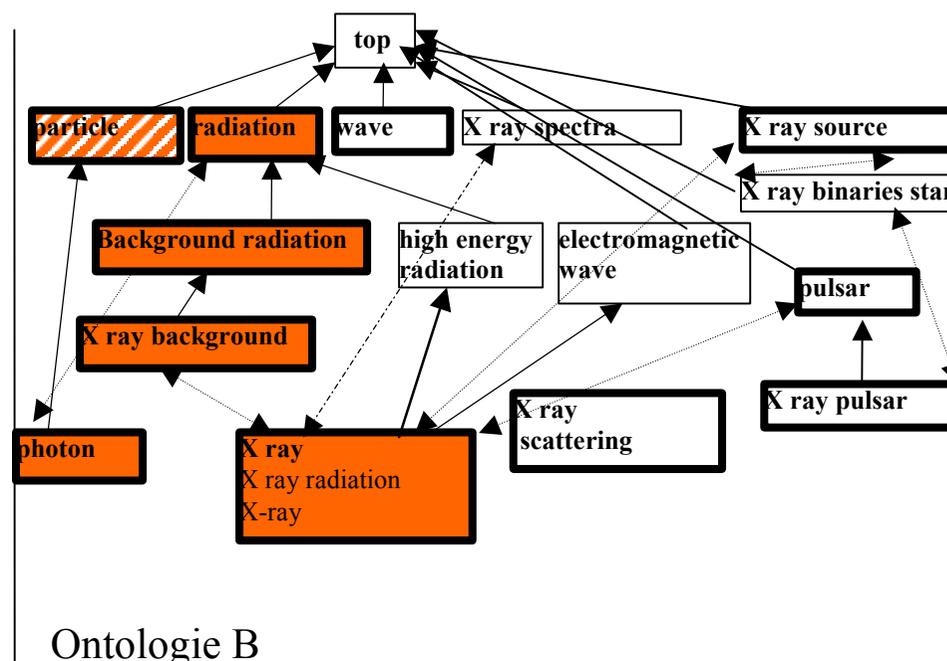
Ontologie A

-----> Partie de --> Est-un <-----> Est relié à

$$\text{Power}(\text{particle})=1,248$$

$$\text{Power}(\text{OntologyA})=12,4$$

F. Genova et équipe MDA, 16/09/2004



Ontologie B

$$\text{Power}(\text{particle})=3,5$$

$$\text{Power}(\text{OntologyB})=25,3$$

Réalisations en cours

- Spécification des relations entre concepts
 - Relation “est lié à”
 - Analyse: analyse distributionnelle [Syntex et Uppery]
 - Aide à la mise à jour automatique

L'Observatoire Virtuel astronomique

Des données d'observation aux résultats publiés, en intégrant les très grands relevés

« an enabling and coordinating entity to foster the development of tools, protocols, and collaborations necessary to realize the full scientific potential of astronomical databases in the coming decade »

NVO White Paper, juin 2000

Coordination internationale

- International Virtual Observatory Alliance

<http://www.ivoa.net>



15 membres, derniers arrivés: Hongrie, Espagne

- Chaque projet a ses propres objectifs (qui dépendent des financements obtenus ou espérés)
- Nombreux projets financés par des crédits e-science (USA/NSF-IT, GB/27 personnes!)
- En France: Action Spécifique INSU+CNES ‘incitative’
- Comment gérer l’implication des STIC autrement qu’au coup par coup??

