

Rapport de fin de parcours¹

EFIGI (Extraction des Formes Idéalisées de Galaxies en Imagerie)

1 Liste des équipes impliquées

- Equipe TERAPIX, Institut d’Astrophysique de Paris (IAP, UMR7095), Paris
- Equipe Origine et Evolution des Galaxies, Institut d’Astrophysique de Paris (IAP, UMR7095), Paris
- Groupe Traitement et Interprétation des Images, Département Traitement Signal-Images (TSI), LTCI (UMR5141), ENST, Paris
- Groupe Traitements Statistiques et Applications aux Communications (TSAC), Département TSI, LTCI (UMR5141), ENST, Paris
- LRDE, EPITA, Paris
- PSEG, Laboratoire d’Astrophysique de Marseille (LAM, UMR6110), Marseille
- Equipe Physique des Galaxies, Laboratoire d’Astrophysique de l’Observatoire Midi-Pyrénées (LAOMP, UMR5572), Toulouse
- Equipe Galaxies, Centre de Recherche Astronomique de Lyon (CRAL, UMR5574), Lyon
- Service d’Astrophysique (SAP), DSM-DAPNIA, Commissariat à l’Energie Atomique, Saclay

2 Liste des participants au 17/10/2007

- IAP
 - Bertin Emmanuel, équipe TERAPIX, Astronome-adjoint, 50%
 - Baillard Anthony, équipe TERAPIX, thésitif, 100%
 - de Lapparent Valérie, équipe Origine et Evolution des Galaxies, CR1 CNRS, 40%
 - Magnard Frédéric, équipe TERAPIX, IR2 CNRS, 10%
 - Mellier Yannick, équipe TERAPIX, Astronome, 20%
 - McCracken Henry Joy, équipe TERAPIX, Astronome-adjoint, 10%
- LTCI
 - Maître Henri, Professeur, TII, 10%
 - Campedel Marine, Ingénieur de Recherche, TII, 20%
 - Cardoso Jean-François, DR2 CNRS, TSAC, 5%
 - Dumoncel Jean, stagiaire du 01/04/2007 au 31/08/2007, TII, 100%
 - Groisne Mélanie, stagiaire du 01/04/2006 au 31/08/2006, TII, 100%
 - Lienou Marie, stagiaire du 01/04/2005 au 31/08/2005, TII, 100%
- LRDE
 - Géraud Thierry, Maître de conférence, LRDE, 20%
 - Dehak Réda, Maître de conférence, LRDE, 20%
 - Demaille Akim, Maître de conférence, LRDE, 10%
 - Ricou Olivier, Maître de conférence, LRDE, 20%
- LAM
 - Arnouts Stéphane, CR1 CNRS, équipe PSEG, 30%

¹La version PDF de ce rapport contient de nombreux liens illustratifs vers le site de la collaboration. L'accès à certains liens est protégé pour des raisons de confidentialité (C.R. de réunions, etc.) : le nom d'utilisateur est **efigi**, et le mot de passe **mastro**

- Tasca Lidia, post-doctorante, équipe PSEG, 10%
- LAOMP
 - Fouqué Pascal, Astronome, équipe Physique des Galaxies, 20%
 - Le Borgne Jean-François, Astronome, équipe Physique des Galaxies, 20%
 - Pelló Roser, Astronome, équipe Physique des Galaxies, 20%
- CRAL
 - Prugniel Philippe, Astronome, équipe Galaxies, 10%
 - Paturel Georges, Astronome, 10%
 - Makarov Dmitry, chercheur invité, 20%
 - Makarova Lidia, chercheur invité, 20%
- SAp
 - Refregier Alexandre, Ingénieur-chercheur CEA, SAp, 5%

3 Changements significatifs intervenus dans le projet

- Lidia Tasca, post-doctorante au LAM, et spécialiste en morphométrie des galaxies (mais non-financée par EFIGI), a rejoint la collaboration en mai 2006.
- Anthony Baillard, initialement ingénieur CDD EFIGI, a démarré une thèse de l'école doctorale EDITE en décembre 2005 sur la classification automatique des galaxies au sein de la collaboration EFIGI.
- Albert Bijaoui, Astronome à l'observatoire de Nice-Côte d'Azur, a rejoint la collaboration en 2006 ; il co-encadre avec Emmanuel Bertin la thèse d'Anthony Baillard.

4 Résumé des principales avancées

Construction de l'échantillon "morphologique" de référence (IAP + LAOMP + CRAL) Les critères de la classification morphologique des galaxies ont été établis dans les années 1970 par quelques spécialistes qui ont classé visuellement plusieurs milliers d'images d'origine photographique. Un effort préliminaire a été de comprendre les bases objectives de cette classification puis de définir de nouveaux critères, plus précis et adaptés aux connaissances actuelles de la physique des galaxies. L'échantillon de référence RC3 (1992), corrigé par Paturel et al. (2003) a été affiné dans le contexte de la base de données HyperLeda (Prugniel et al. 2008) afin d'enrichir la base de données étiquetée (voir ci-dessous) utilisée pour le raccordement entre le nouveau système et l'ancien.

Constitution d'une base de données étiquetée (IAP + LAOMP + LAM + CRAL) Une partie significative des efforts déployés durant le projet a consisté à produire une grande base étiquetée (20 attributs morphologiques) de 22000 images de galaxies, extraites à partir de 4 To de données du relevé SDSS. A ces jeux de données s'ajoutent 10000 images SDSS sur lesquelles l'équipe du LAM a appliqué des estimateurs morphométriques "classiques", des images issues de simulations, des relevés CFHTLS et GalEx produites ou sélectionnées au sein de la collaboration (Paris, Marseille). Une interface web autorisant 10 astronomes de la collaboration à étiqueter à distance les images stockées dans la base de données a été développée à l'IAP. Le fastidieux travail d'étiquetage s'est étalé sur plus d'un an ; un article présentant le catalogue est en cours d'écriture (Baillard et al. 2008).

Simulation d'images de galaxies (IAP) L'évaluation et la comparaison des techniques de mesure morphométrique mises en place dans EFIGI passe aussi par la comparaison entre les paramètres insérés dans des simulations réalistes et ceux mesurés. Deux outils développés initialement pour TERAPIX (STUFF et SKYMAKER) ont dû subir des améliorations importantes afin de répondre aux exigences d'EFIGI : ajout d'une loi d'évolution dans les brillances des disques, support d'une PSF externe, parallélisation et *repackaging* pour supporter le déploiement en environnement de calcul distribué.

Prétraitement et mise en forme des données (LRDE + IAP) Les galaxies se présentent assez rarement de manière isolée sur les images : un "nettoyage" (repérage des pixels affectés par les

sources voisines ou des pollutions lumineuses et “remplissage” par *inpainting*) et un prétraitement sont nécessaires avant d’effectuer l’analyse proprement-dite. Les équipes du LRDE et de l’IAP ont mis en commun leur expérience pour développer des techniques et des outils de segmentation des images à base de filtres connectés, optimisées pour les images astronomiques de grande dynamique (Baillard et al. 2007, Berger et al. 2007, Baillard et al. 2008). Ces outils sont en cours d’implémentation au sein de la bibliothèque logicielle Olena.

Mesures morphométriques : Réduction dimensionnelle par décomposition linéaire (LTCI + LRDE + IAP + SAp) La réduction dimensionnelle par décomposition linéaire initialement proposée pour ce projet (Shapelettes, ACP ou ACI après remise en forme), bien que donnant de bons résultats pour la classification en type de Hubble (Baillard et al. 2006), s’est avérée peu adaptée à la mesure de paramètres de forme spécifiques. Seule la classification de textures par filtres de Gabor appliquée à la mesure de floculence des galaxies (Dumoncel et al. 2007) a finalement bénéficié de cette approche.

Mesures morphométriques : Réduction dimensionnelle par ajustement non-linéaire (LTCI + IAP) Cette approche plus classique chez les astronomes (correspondant au “plan B” du projet initial) a finalement été adoptée suite aux difficultés rencontrées avec les décompositions linéaires non-paramétriques avec qualité d’image variable, et aux travaux exploratoires menés en 2006 par l’équipe du LTCI. Outre une totale optimisation et un avantage en vitesse d’un ordre de grandeur par rapport aux travaux concurrents, l’outil EFIGI ajoute la mesure des barres, bras spiraux et anneaux à celle des bulbes et des disques. Le logiciel sera rendu public en janvier 2008, et utilisé au sein des pipeline TERAPIX et DES dès 2008.

Mesure automatique de la réponse impulsionnelle (IAP) La mesure de la morphologie des galaxies distantes, mal-résolues, nécessite une connaissance précise de la réponse impulsionnelle locale. Bertin et al. (2008, en préparation), ont développé une nouvelle version du logiciel PSFEX destiné à mesurer automatiquement la réponse impulsionnelle des images, y compris dans les conditions de sous-échantillonnage des expériences spatiales. Les réponses impulsionnelles des images de l’échantillon étiqueté ont été calculées avec cet outil.

Calcul distribué et “web services” (LRDE + IAP) La mesure des paramètres morphologiques d’une galaxie avec l’outil EFIGI prend entre 0.2 et 100s par coeur de processeur, selon le nombre d’attributs laissés en paramètres libres et la résolution de l’image. Les relevés d’imagerie astronomique contenant de quelques centaines de milliers à plusieurs dizaines de millions de galaxies, le succès d’EFIGI repose sur la capacité d’effectuer les traitements de manière massivement parallèle. O.Ricou et al. (2007) ont développé et déployé sur le *cluster* TERAPIX à l’IAP un système basé sur Condor, permettant d’exécuter EFIGI en parallèle sur plusieurs dizaines de processeurs. Pour l’ensemble de la communauté, l’accès au prototype EFIGI se fait, au choix, par l’interface web, par un web-service respectant les recommandations du *Virtual Observatory*, ou par un service de grille de calcul basé sur Globus. Grâce à la généralité du système mis en place, il est prévu d’ étoffer courant 2008 l’offre des services ouverts à la communauté astronomique sur les serveurs de TERAPIX par des outils de simulation et de distribution d’images.

Décalages vers le rouge “photométriques” et application (LAM + LAOMP) L’apparence d’une galaxie dépend de la longueur d’onde à laquelle on l’observe. Pour que les mesures morphométriques des galaxies distantes soient réellement exploitables scientifiquement, il est important de connaître les décalages vers le rouge que l’expansion de l’Univers leur fait subir. En l’absence d’information spectroscopique, cela ne peut se faire qu’à partir des couleurs des galaxies. Les travaux du LAM sur l’optimisation des patrons de distributions d’énergie spectrales couplés à une approche bayésienne ont permis une percée considérable en termes de fiabilité et de précision des décalages spectraux (Ilbert et al. 2006, Arnouts et al. 2007, Dahlen et al. 2008) ; ces deux paramètres sont désormais de qualité suffisante pour notre application. Inversement, l’utilisation des paramètres morphologiques mesurés par EFIGI dans les *a priori* de la recherche du décalage vers le rouge est envisagée.

5 Réalisations obtenues dans le cadre du projet

- Version 3 du logiciel de simulation d’images astronomiques SKYMAKER (public).
- Logiciel de “nettoyage” automatique d’images de galaxies NFIGI (public).
- Interface web d’étiquetage des données ManClass et outil de gestion de la base de données associé.
- Base de données fusionnées d’images multi-bande de galaxies étiquetée et son interface web d’interrogation (distribution publique prévue pour Q1 2008).
- Logiciel PSFEX de mesure automatique de la réponse impulsionnelle d’une image (public).
- Logiciel d’ajustement paramétrique de modèles de galaxies EFIGI (inclus dans SEXTRACTOR, remplacement prévu de l’ancienne version publique par celle-ci en janvier 2008).
- web service, grid service, et interface web de mesure morphologique automatisée à partir d’une image (public dans sa version simple, une version améliorée doit être rendue publique courant 2008).

6 Réunions et Conférences organisées dans le cadre du projet

5 grandes réunions de collaboration : 29/11/2004, 05/10/2005, 13/06/2006, 11/12/2006, 26/06/2007.

7 Soutiens obtenus en liaison avec ce projet

7.1 Postes chercheurs

Thèse d’Anthony Baillard (de Q4 2005 à Q3 2008, financement actuel INSU+PNC)

7.2 Postes ingénieurs

CDD IE Anthony Baillard (de Q4 2004 à Q2 2006, financement EFIGI)

7.3 Contrats nationaux

7.4 Contrats européens

7.5 Contrats internationaux hors CEE

7.6 Contrats industriels

7.7 Contacts internationaux dans le cadre de ce projet

- Morphométrie : Contacts avec les consortia scientifiques Astro-WISE, COSMOS et Dark Energy Survey concernant l’adoption d’outils EFIGI dans l’analyse des images.
- Décalages vers le rouge “photométriques” : application à la mission spatiale SNAP

8 Publications obtenues dans le cadre du projet

- Journaux internationaux
 - “Accurate photometric redshifts for the CFHT Legacy Survey calibrated using the VIMOS VLT Deep Survey”
Ilbert O. et al., 2006, *Astronomy & Astrophysics* 457, 841
 - “The SWIRE-VVDS-CFHTLS surveys : stellar mass assembly over the last 10 Gyears. Evidence for a major build up of the red sequence between $z=2$ and $z=1$ ”
S. Arnouts et al. 2007, *Astronomy & Astrophysics* 476, 137

- “SuperNova Acceleration Probe (SNAP) : Investigating Photometric Redshift Optimization”
T. Dahlen, B. Mobasher, S. Jouvel, J.-P. Kneib, O. Ilbert, S. Arnouts, G. Bernstein, J. Rhodes, 2008,
soumis à *Astrophysical Journal*
- Conférences internationales :
 - “L’analyse des images dans les grands relevs d’astronomie visible/proche-infrarouge (revue invitée)”
E. Bertin, 2005, *Traitement du Signal et des Images*,
20ème colloque GRETSI, Presses Univ. de Louvain, 299
 - “the EFIGI project : Automatic classification of galaxies”
A. Baillard et al., 2006,
Astronomical Data Analysis Software and Systems XV, ASP Conf. Series, 351, 236
 - “VOTables in TERAPIX software”
E. Bertin & G. Tisserand, 2007,
Astronomical Data Analysis Software and Systems XVI, ASP Conf. Series, 376, 507
 - “Algorithme de Calcul de l’Arbre des Composantes avec Applications à la Reconnaissance des Formes
en Imagerie Satellitaire”
A. Baillard, C. Berger, E. Bertin, T. Géraud, R. Levillain, N. Widynski, 2007,
22ème colloque GRETSI sous presse
 - “Effective Component Tree Computation with Application to Pattern Recognition in Astronomical
Imaging”
C. Berger, T. Géraud, R. Levillain, N. Widynski, A. Baillard, E. Bertin, 2007,
IEEE International Conference on Image Processing, sous presse
 - “Web services at TERAPIX”
O. Ricou, A. Baillard, E. Bertin, F. Magnard, C. Marmo, Y. Mellier, 2007,
Astronomical Data Analysis Software and Systems XVII, ASP Conf. Series, sous presse
 - “An automatic method to determinate the degree of flocculence of a galaxy”
J. Dumoncel et al., 2007,
Astronomical Data Analysis Software and Systems XVII, ASP Conf. Series, sous presse
- Journaux nationaux
- Conférences nationales
 - “Le projet EFIGI : Extraction des Formes Idéalisées des Galaxies en Imagerie”
Baillard A. et al., 2005,
Journées Masses de Données PaRISTIC.
 - “Automatic quantitative morphology of galaxies”
Baillard A. et al., 2006, Semaine de l’Astrophysique Française.
 - “EFIGI : Extraction des Formes Idéalisées de Galaxies en Imagerie”
E. Bertin et al., 2006, Journées Masses de Données PaRISTIC