

FONDS NATIONAL DE LA SCIENCE

ACI Masses de Données

Rapport de fin de contrat

(à transmettre en 3 exemplaires)

I - FICHE D'IDENTITE DU PROJET

Numéro de référence du projet : 502190

Titre du Projet : EFIGI : Extraction des Formes Idéalisées de Galaxies en Imagerie

Coordinateur du projet : Emmanuel BERTIN

Tél du coordinateur du projet : 01 44 32 81 51

Mél du coordinateur du projet : bertin@iap.fr

Laboratoire de rattachement du coordinateur (*en toutes lettres, plus sigle et numéro d'unité le cas échéant*) : Institut d'Astrophysique de Paris (IAP, UMR7095)

Adresse postale du laboratoire : 98bis, bd Arago, F-75014 Paris

Numéro d'unité (*s'il existe*) : UMR7095

Montant global (en KEurosTTC) : 330

Informations de cadrage du projet :

Durée : ● 1 an ● 2 ans 3 ans

Répartition budgétaire initialement prévue (en KEuros TTC) :

Equipement : 50

Fonctionnement : 280 (dont vacations 55,8)

CDD : 18 mois

Nom des personnels recrutés, niveau de recrutement et durée des contrats : Anthony Baillard, CDD IE2, 18 mois

Partenaires du Projet	Laboratoire (pas de sigle)
Bertin Emmanuel (50%) Baillard Anthony (100%) Dantel-Fort Mireille (5%) de Lapparent Valérie (40%) Magnard Frédéric (10%) Mellier Yannick (20%) McCracken Henry (10%) <i>départs :</i> van Waerbeke Ludovic (départ à UBC, Canada) Domisse Laurent (change- ment d'équipe) Missonnier Gilles (change- ment d'équipe)	Institut d'Astrophysique de Paris
Maître Henri (10%) Campedel Marine (20%) Cardoso Jean-François (5%) <i>départs :</i> Cappé Olivier (désintéret) Gousseau Yann (désintéret) Moulines Eric (désintéret)	Laboration Traitement et Communication de l'Information
Géraud Thierry (20%) Dehak Réda (20%) Demaille Akim (10%) Ricou Olivier (20%)	Laboratoire de Recherche & Développement de l'EPITA
Arnouts Stéphane (30%) <i>arrivées :</i> Tasca Lidia (post-doc spécialiste du sujet, 10%)	Laboratoire d'Astrophysique de Marseille
Fouqué Pascal (20%) Le Borgne Jean-François (20%) Pelló Roser (20%)	Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées
Prugniel Philippe (10%) <i>arrivées :</i> Makarov Dmitry (invité, 20%) Makarova Lidia (invitée, 20%) <i>départs :</i> Paturel Georges (retraite)	Centre de Recherche Astrophysique de Lyon
Refregier Alexandre (5%)	Service d'Astrophysique du Commissariat à l'Energie Ato- mique

N.B : S'il y a eu modifications des partenaires initialement prévus, en expliquer les raisons.

II – BILAN DETAILLE DU PROJET

1 – Rappel des objectifs initiaux du projet :

L’objectif global du projet était de fournir à la communauté astronomique un outil logiciel générique permettant d’évaluer automatiquement la forme des galaxies à partir d’images. Pour atteindre ce but plusieurs étapes étaient prévues :

1. Concertation entres spécialistes afin de définir un jeu de descriptifs morphologiques génériques pertinents pour la mesure automatique, et leur relation vis-à-vis de descripteurs existants (moins génériques et/ou dont la mesure est difficilement automatisable).
2. Evaluation des techniques proposées par les membres de la communauté STIC (Sciences du Traitement de l’Information et de la Communication) : réduction de dimensionalité (estimation paramétrique, non-paramétrique ; linéaire, non-linéaire), prise en compte des effets de la réponse impulsionnelle et du bruit de fond, relation au paramètres physiques des galaxies, techniques d’apprentissage
3. Compilation de données existantes et réalisation d’échantillons d’images de galaxies, dont certains étiquetés par les astronomes de la collaboration, pour l’apprentissage, l’évaluation et la validation des algorithmes.
4. Ecriture des logiciels, mise en place d’un système de calcul distribué et d’un point d’accès internet au service de mesure morphométrique pour la communauté astronomique, en conformité avec les protocoles de l’*Astronomical Virtual Observatory*¹.

2 – Rapport final

2.1 Programme des actions engagées

1. en cas de modification du programme des travaux préciser les raisons et les réorientations
2. préciser les modalités concrètes de travail interdisciplinaire

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. Une tâche 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. Le défaut essentiel est la confusion entre les galaxies elliptiques (normales, E) et les galaxies elliptiques diffuses (dE). Lorsque le système de classification morphologique a été crée, les dEs, qui sont des objets intrinsèquement faibles, étaient mal connues et n’ont pas été différenciées des E. Or on sait aujourd’hui que cette classe de galaxies est numériquement dominante et possède des caractéristiques physiques très différentes des E. Des critères de distinction (brillance de surface) ont été introduits pour améliorer la classification de référence, et ont servi à la définition de la description morphologique EFIGI. L’échantillon de référence RC3 (1992), corrigé par Paturel et al. (2003) a été affiné dans le contexte de la base de données HyperLeda² (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.

¹<http://www.ivoa.net>

²<http://leda.univ-lyon1.fr>

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, reconstruites (extraction, reprojection, soustraction du fond de ciel, mosaïcage et combinaison) à partir de 4 To de données du relevé SDSS³. 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 CFHTLS⁴ et GaLEX⁵ produites ou sélectionnées au sein de la collaboration (Paris, Marseille). Une interface web⁶ autorisant 10 astronomes de la collaboration à étiqueter à distance les images stockées dans la base de données a été développée à l'IAP (Fig. 1). Le fastidieux travail d'étiquetage s'est étalé sur plus d'un an ; l'équipe a notamment fait appel à des spécialistes des galaxies proches, D. Makarov et L. Makarova, pour certains objets. Une attention particulière a été apportée à tenter de distinguer les éléments du jugement visuel qui étaient objectivement fondés sur des caractéristiques de l'image de ceux affectés par l'expertise astrophysique (comme la prise en compte implicite des modifications morphologiques en fonction de l'orientation ; l'astronome veut une classification indépendante de l'angle de vue de l'objet). Le contenu de la base sera mis à disposition du public début 2008 à travers une interface⁷ en voie de finalisation à l'IAP. L'article présentant le catalogue est toujours en cours d'écriture (Baillard et al. 2008), en raison des délais importants suscités par l'ampleur des tâches d'étiquetage.

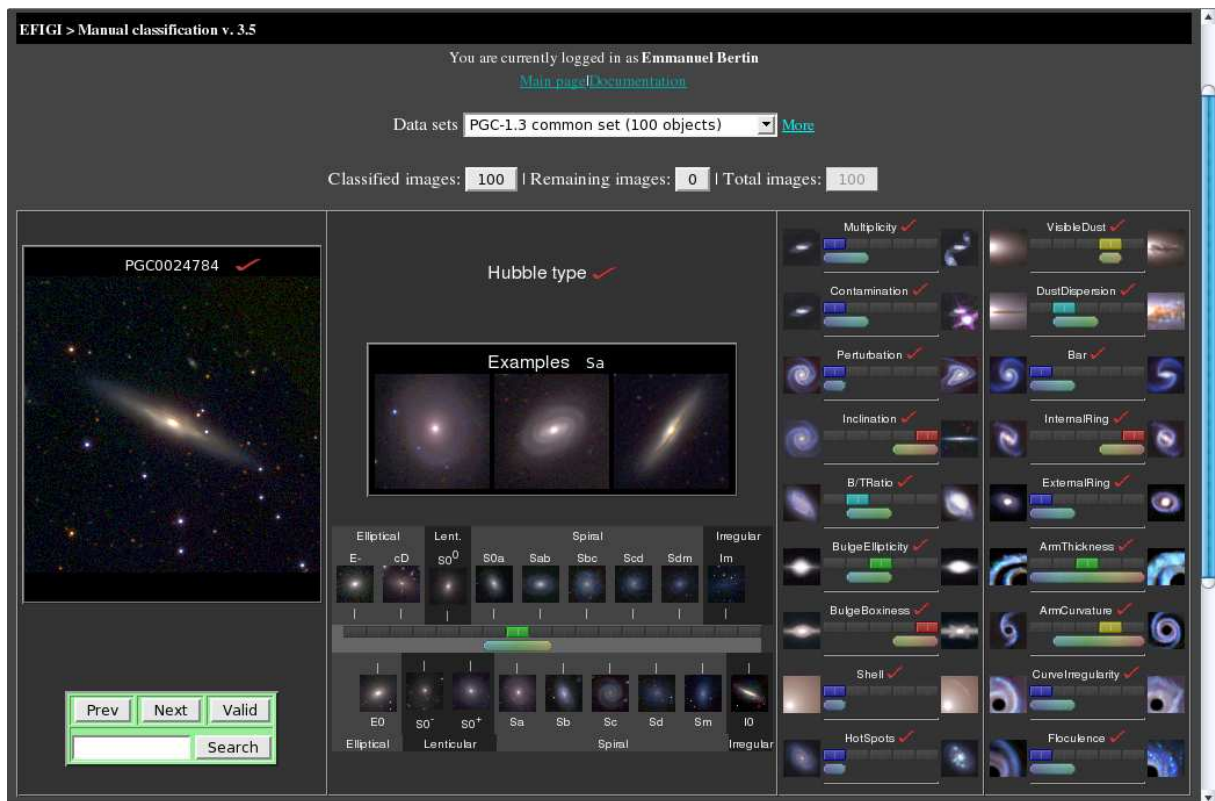


FIG. 1 – L'interface "ManClass" d'étiquetage manuel des galaxies de référence.

Simulation d'images de galaxies (IAP) L'évaluation et la comparaison des techniques de mesure morphométrique mises en place dans EFIGI passent aussi par la com-

³<http://www.sdss.org>

⁴<http://www.cfht.hawaii.edu/Science/CFHLS/>

⁵<http://www.oamp.fr/lam/projets/fich-galex.html>

⁶<http://efigix.iap.fr/manclass/interface.html>

⁷<http://efigix.iap.fr/cfigi/>

paraison entre les paramètres insérés dans des simulations réalistes et ceux mesurés. Deux outils développés initialement pour TERAPIX, STUFF⁸ et SKYMAKER⁹, 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. Plusieurs millions d’images d’étoiles et de galaxies ont été simulées afin de tester les différentes composantes d’EFIGI.

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, optimisés 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 OLENA¹⁰ (stages de Guillaume Duhamel, Matthieu Garrigues et Simon Nivault, en cours au LRDE) ; le logiciel de nettoyage d’images de galaxies NFIGI¹¹ est quant-à-lui déjà disponible pour la communauté.

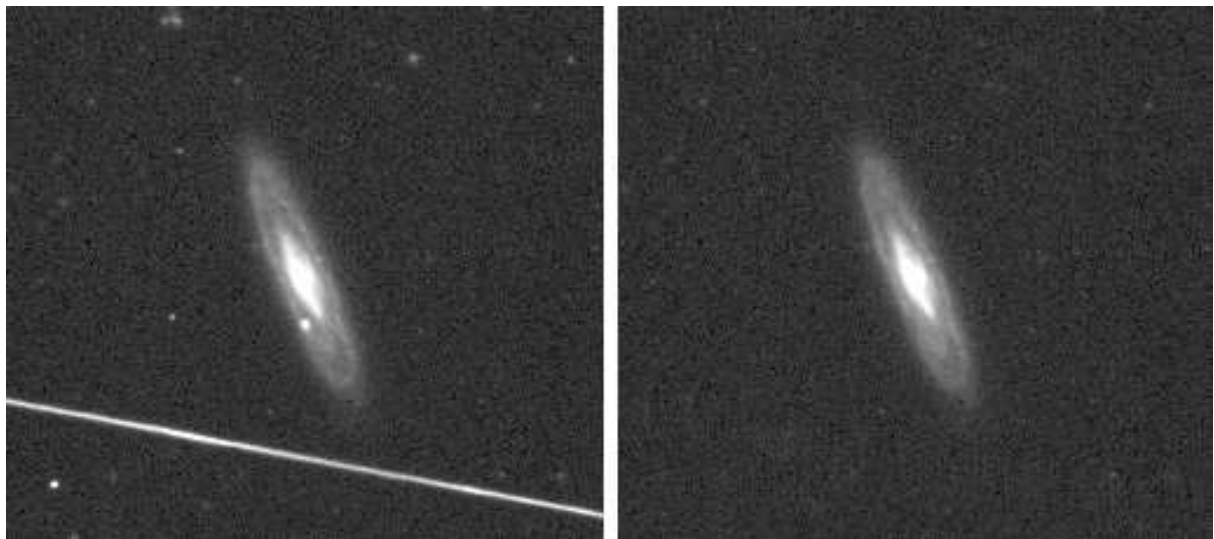


FIG. 2 – *A gauche* : image du relevé SDSS, contaminée par une étoile d’avant plan et une traînée de satellite. *A droite* : version “nettoyée” par l’outil NFIGI.

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, Fig. 3), s’est avérée peu adaptée à la mesure de paramètres de forme spécifiques, malgré le temps et les efforts importants investis dans cette direction. 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.

⁸<http://terapix.iap.fr/soft/stuff>

⁹<http://terapix.iap.fr/soft/skymaker>

¹⁰<http://olena.lrde.epita.fr/>

¹¹<http://terapix.iap.fr/soft/nfigi>

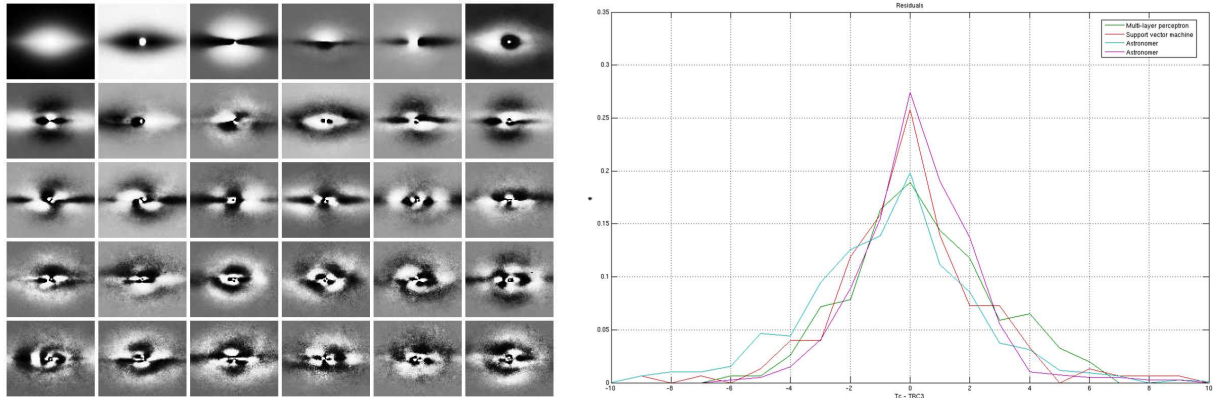


FIG. 3 – *A gauche* : trente premiers vecteurs de la base de Karhunen-Loève obtenu sur un sous-échantillon d’images de galaxies de référence après remise en forme et nettoyage. *A droite* : Ecart entre le type morphologique révisé du catalogue RC3 et les estimations de deux classifieurs automatiques (Perceptron multi-couche et Séparateur à Vaste Marge), ainsi que celle de deux astronomes de la collaboration.

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 gain en vitesse d’un ordre de grandeur par rapport aux travaux concurrents, l’outil EFIGI ajoute la mesure des barres, bras spiraux, bandes de poussière et anneaux à celle des bulbes et des disques (Fig. 4). 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 PSFEX¹² destinée 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 (Fig. 5).

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 cœur 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 web¹³, par un web-service¹⁴ respectant les recommandations du *Virtual Observatory*¹⁵ (VO), ou par un service de grille de calcul¹⁶ basé sur Globus. Les données renvoyées en sortie sont au

¹²<http://terapix.iap.fr/soft/psfex>

¹³http://efigix.iap.fr/ws/efigi_wi.html

¹⁴<http://efigix.iap.fr/ws/ws.html>

¹⁵<http://www.ivoa.net/>

¹⁶<http://efigix.iap.fr/ws/globus.html>

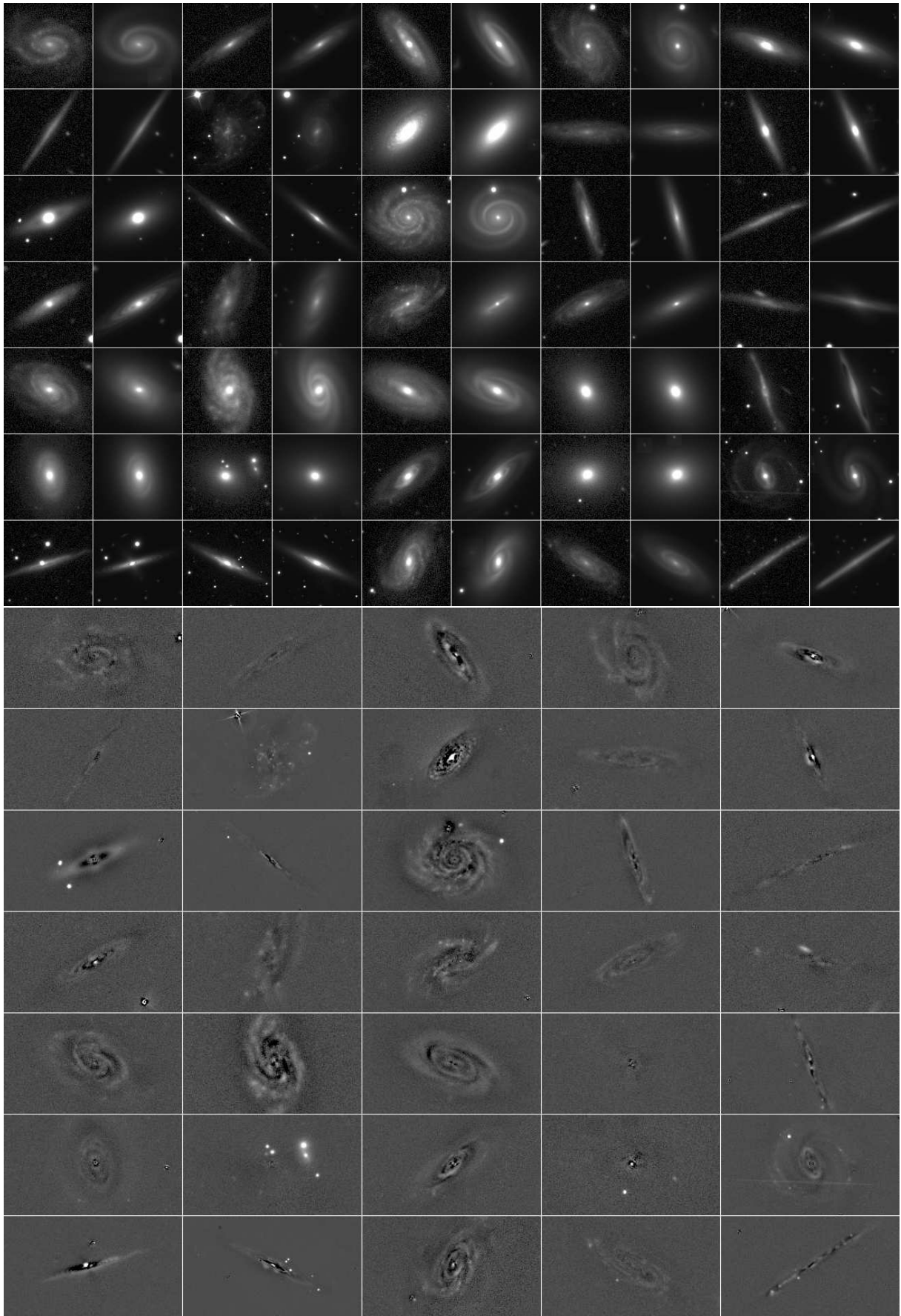


FIG. 4 – *En haut* : Exemples d’ajustements d’images de l’échantillon de référence avec un modèle bulbe+disque+barre+bras spiraux obtenus dans le canal g (pour chaque galaxie : image d’origine à gauche, modèle ajusté à droite). *En bas* : Images des résidus d’ajustement correspondantes ; .

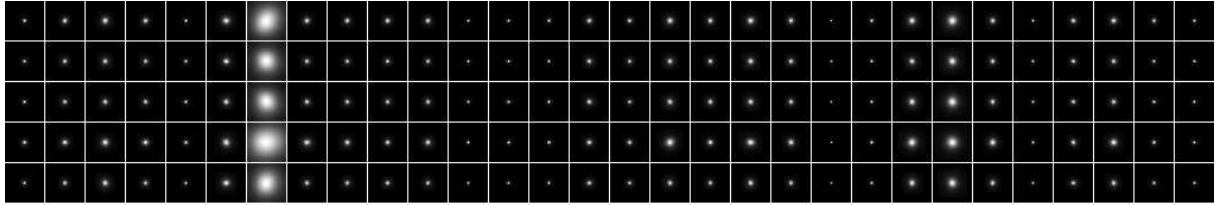


FIG. 5 – Extraits de modèles de réponse impulsionnelle locale obtenus sur les images SDSS dont est issu l'échantillon étiqueté. De haut en bas : canaux u, g, r, i et z .

format XML-VOTable (Bertin & Tisserand 2007) adopté par le VO.

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 indicateurs ont désormais atteint un niveau suffisant 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.

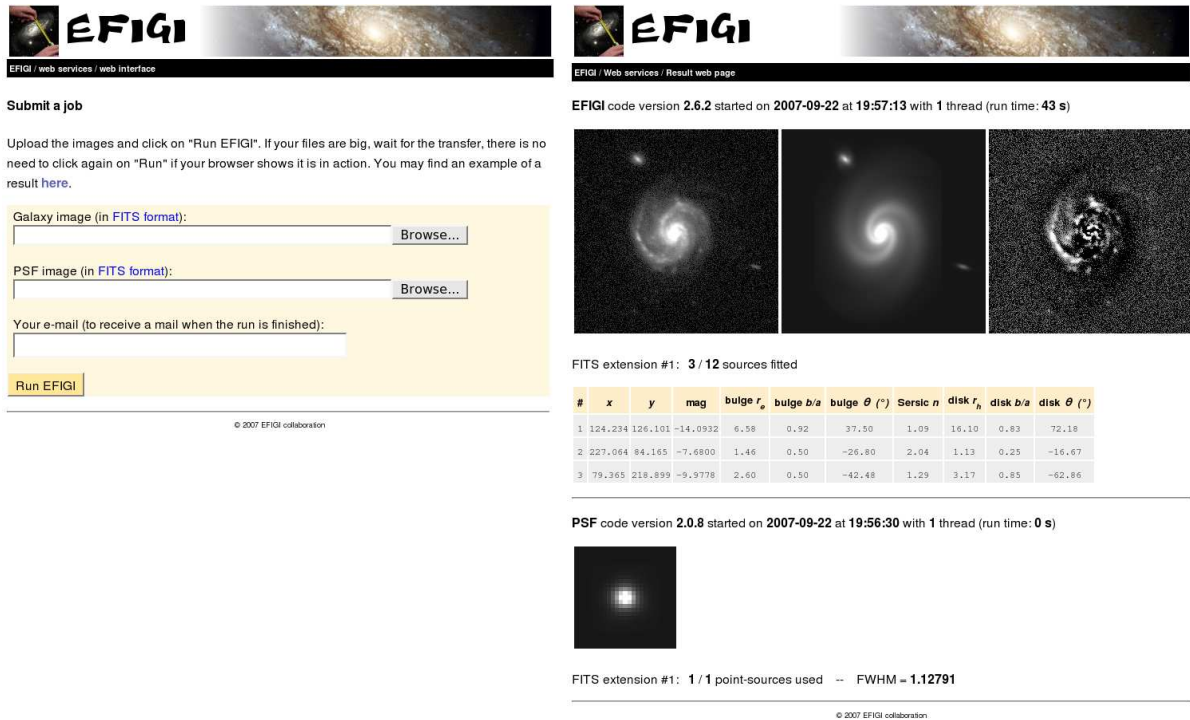


FIG. 6 – *A gauche* : Prototype de l'interface web au service de morphométrie EFIGI (http://efigix.iap.fr/ws/efigi_wi.html). *A droite* : Page de résultats de mesure morphométriques sur un exemple (il s'agit en fait des données de sortie au format XML-VOTable traduites en HTML au moyen d'un filtre XSLT).

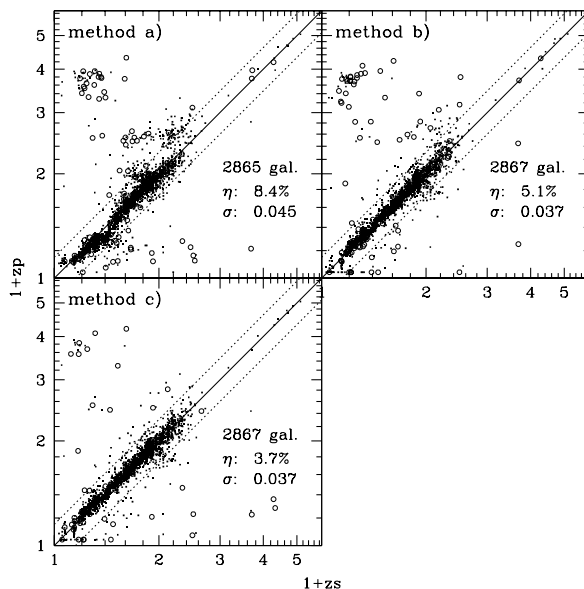


FIG. 7 – Optimisation des estimations du décalage vers le rouge z à partir des couleurs des galaxies (Ilbert et al. 2006). *a*) : décalages retrouvés en fonction du décalage réel par minimisation du χ^2 . *b*) : idem avec optimisation des patrons de distributions d’énergie spectrale. *c*) : comme *b*), avec prise en compte des a priori sur la distribution en z . η est la fraction d’objets avec des erreurs sur le z estimé “catastrophiques” (> 0.15).

2.2 Apports et résultats scientifiques, y compris les avancées permises par la collaboration interdisciplinaire

EFIGI est avant-tout une collaboration orientée “applications” :

- La collaboration a permis aux équipes des STIC d’investir un domaine d’application particulièrement exigeant : opérations sur des images géantes, de très large dynamique, aux rapports Signal-sur-Bruit par pixel souvent inférieurs à 1, avec des contraintes scientifiques fortes sur la maîtrise des biais statistiques. Ces conditions ont amené l’équipe du LRDE à proposer un algorithme peu gourmand en mémoire et efficace avec des fortes quantifications, voire en virgule flottante pour les constructions d’arbres de composantes (Berger et al. 2007, Baillard et al. 2007).
- Après soustraction d’un modèle paramétrique ajusté, une fraction des images de galaxies présente une texture dont les caractéristiques (“floculence”, présence de bandes de poussière, ponctuation par des “points chauds”) renseignent sur les conditions physiques qui y règnent. L’équipe du LTCI a dû repenser l’approche utilisée traditionnellement pour la classification de textures sur images satellite afin de prendre en compte notamment la géométrie particulière des galaxies et les variations de réponse impulsionnelle (Dumoncel et al. 2007).
- En attendant une analyse étendue à l’ensemble du relevé SDSS (voir plus bas), l’échantillon d’images étiqueté de galaxies (Baillard et al. 2008) représente déjà, au delà de sa finalité première de jeu d’apprentissage et de validation, un ensemble inédit de référence morphologique de l’Univers local. Les 20 attributs mesurés sur une échelle de 0 à 4 par une dizaine d’astronomes professionnels et accompagnés d’un intervalle de confiance sont à comparer aux trois ou quatre indicateurs le plus souvent binaires des catalogues précédents ou des efforts communautaires actuels comme GALAXYZOO¹⁷.
- Grâce aux logiciels de nettoyage d’image (nFIGI), d’extraction automatique de la réponse impulsionnelle (PSFEx), d’ajustement non linéaire de modèles de galaxies

¹⁷<http://www.galaxyzoo.org>

(EFIGI), et du développement des couches web/grid, tous effectués au sein de la collaboration, il a été possible de mettre en ligne un véritable service de “morphométrie” librement accessible à la communauté¹⁸. La version publique actuelle est limitée à une décomposition bulbe+disque (11 paramètres libres), testée sur un grand nombre de simulations et d’images de référence (Fig. 8). La version complète, permettant l’ajustement de modèles de barres, bras spiraux, anneaux et bandes de poussière (Fig. 4) nécessite une validation scientifique plus poussée en raison d’une plus grande sensibilité aux conditions initiales, et un risque supérieur de dégérescence sur les images mal-résolues ; actuellement en phase de test, sa disponibilité à l’ensemble de la communauté n’est prévue qu’au printemps 2008.

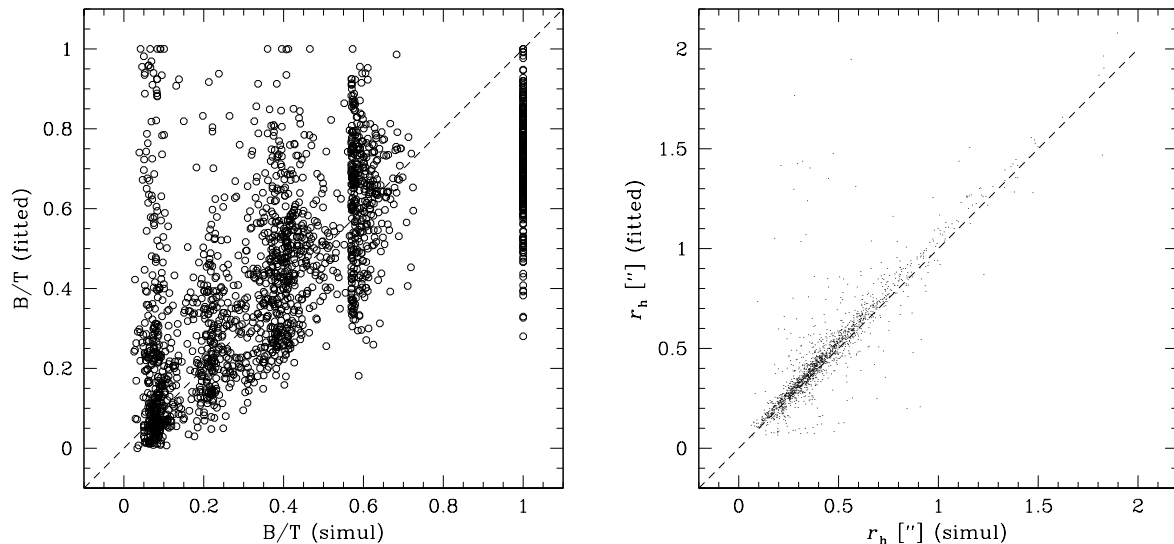


FIG. 8 – Comparaison entre valeurs entrées dans des simulations d’images de galaxies faibles et mesures morphométriques effectuées par EFIGI, dans des conditions typiques d’observation au télescope Canada-France-Hawaii. *A gauche* : fraction de luminosité contenu dans le bulbe. *A droite* : Longueur d’échelle du disque.

2.3 Collaborations internationales

- University of Illinois at Urbana-Champaign¹⁹, USA : installé sur les machines du National Center for Supercomputing Applications²⁰ (NCSA), le pipeline du Dark Energy Survey²¹ est avec celui de TERAPIX à Paris l’un des deux pipelines de traitement de données où est déployé et testé EFIGI de manière massivement parallèle.
- University of Hawaii²² / Canada-France-Hawaii-Telescope corp.²³, USA : où sont développés les codes de mesure des décalages spectraux à partir des couleurs des galaxies.
- Special Astronomical Observatory²⁴, Russie : caractérisation des galaxies de l’Univers proche et implication dans HyperLeda.

¹⁸<http://efigix.iap.fr/ws>

¹⁹<http://www.uiuc.edu>

²⁰<http://www.ncsa.edu/>

²¹<http://www.darkenergysurvey.org/>

²²<http://www.ifa.hawaii.edu/>

²³<http://www.cfht.hawaii.edu>

²⁴<http://www.sao.ru>

2.4 Publications par chercheurs impliqués (joindre en annexe : les résumés - 1/2 à 1 page - des publications liées au projet (en précisant si elles sont soumises, acceptées ou publiées), le sommaire des ouvrages publiés. . .)

- “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
- “Accurate photometric redshifts for the CFHT Legacy Survey calibrated using the VIMOS VLT Deep Survey”
Ilbert O. et al., 2006, *Astronomy & Astrophysics* 457, 841
- “VOTables in TERAPIX software”
E. Bertin & G. Tisser, 2007,
Astronomical Data Analysis Software and Systems XVI, ASP Conf. Series, 376, 507
- “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
- “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
- “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*

3 articles décrivant respectivement la technique de nettoyage des images, d’extraction de la réponse impulsionnelle et le catalogue étiqueté sont en outre en préparation pour la revue astronomique *Astronomy & Astrophysics*. Les versions en cours peuvent être consultées sur le site d’EFIGI²⁵.

2.6 Autres résultats : structuration, création d’équipe, nouvelles collaborations, thèses. . .

- Le travail d’Anthony Baillard, initialement recruté en CDD d’ingénieur EFIGI, a débouché sur une thèse en IAD (Intelligence Artificielle et Aide à la Décision) dont le sujet concerne la classification morphologique automatique et dont la soutenance est prévue pour le troisième trimestre 2008.

²⁵http://terapix.iap.fr/rubrique.php?id_rubrique=249, utilisateur `efigi`, mot-de-passe `mastro`.

- La mise en fonction du service de “morphométrie”, offre aux astronomes un potentiel scientifique considérable ; ainsi la prochaine distribution du relevé CFHTLS, dont le processing démarre dès janvier 2008 à TERAPIX, et qui devrait contenir plusieurs millions de galaxies faibles, inclura dans son catalogue les mesures effectuées par EFIGI en version bulbe+disque dans plusieurs canaux. Les galaxies du CFHTLS étant distantes, il est impératif de corriger leurs mesures morphométriques des effets du décalage spectral ; ce qui sera fait automatiquement grâce à la technique mise en place par Ilbert et al. (2006).
- Le travail effectué sur l’échantillon de référence issu des images du relevé SDSS a convaincu l’équipe de l’IAP réunie autour d’EFIGI de la nécessité d’étendre les analyses photométriques et morphométriques automatiques à l’ensemble des galaxies bien résolues de ce relevé, soit environ 100.000 objets. Cela représente 8To de données d’images à archiver et analyser. Un tiers de ce volume a déjà été téléchargé depuis les USA, et l’ensemble des données sera prêt à traiter début 2008. Les informations récoltées à l’issue de l’analyse devraient être rapidement rendues disponibles sur la base HYPERLEDA.
- Grâce à la généricité du système de web service mis en place par l’équipe du LRDE à l’IAP, il est prévu d’offrir courant 2008 l’offre des services ouverts à la communauté astronomique sur les serveurs EFIGI par des outils de simulation et de distribution d’images. Cette nouvelle collaboration LRDE-IAP se fera sous les auspices de l’action spécifique VO-France²⁶.

2.7 Valorisation : colloques, formation, expositions...

- 8 stages rémunérés d’école d’ingénieur ou de master ont été effectués dans le cadre d’EFIGI : Marta Arroyo, Guillaume Duhamel, Jean Dumoncel, Matthieu Garrigues, Mélanie Groisne, Thibaut Levieuge, Marie Lienou et Simon Nivault.
- Les logiciels développés dans le cadre d’EFIGI sont distribués sous licence GPL (libre) à l’ensemble de la communauté sur les sites du LRDE²⁷ et de l’IAP²⁸, et bénéficient pour la plupart d’un support technique appuyé par un réceptacle de développement SubVersion²⁹ et d’un forum de discussion³⁰ pour les utilisateurs.

²⁶<http://www.france-vo.org/>

²⁷<http://olena.lrde.epita.fr/>

²⁸<http://terapix.iap.fr/soft>

²⁹<http://terapix.iap.fr/wsvn>

³⁰<http://terapix.iap.fr/forum>

3 – Rapport financier

Rappel des dépenses figurant sur le relevé de dépenses d'étape ou de fin d'opération : fonctionnement, équipement, CDD (ventilation par partenaire dans le cas des projets en concertation).

Institut d'Astrophysique de Paris

Type	Montant (€ HT)
Nov. 2004 - Nov. 2005	
Personnel (CDD)	32.840,09
Equipement	10.404,40
Fonctionnement	8.339,90
Nov. 2005 - Nov. 2006	
Personnel (CDD)	12.636,50
Equipement	7.848,18
Fonctionnement	16.625,84
Nov. 2006 - Nov. 2007	
Personnel (CDD)	0,00
Equipement	51.393,49
Fonctionnement	37.726,67
Total sur 3 ans	
Personnel (CDD)	45.476,59
Equipement	69.646,07
Fonctionnement	62.692,41
Grand total :	177.815,07

note : la somme totale allouée à l'IAP étant supérieure à celle demandée, et la durée du contrat CDD limitée à 18 mois (notre demande de le porter à 24 mois comme requis initialement ayant été refusée), les crédits excédentaires ont été affectés à l'achat d'équipement informatique (nœuds de calcul pour le service de morphométrie).

Laboratoire Traitement et Communication de l'Information

Type	Montant (€ HT)
Nov. 2004 - Nov. 2005	
Equipement	0,00
Fonctionnement	4.590,00
<i>dont vacations</i>	<i>4.590,00</i>
Nov. 2005 - Nov. 2006	
Equipement	0,00
Fonctionnement	7.793,10
<i>dont vacations</i>	<i>7.613,10</i>
Nov. 2006 - Nov. 2007	
Equipement	0,00
Fonctionnement	7.766,12
<i>dont vacations</i>	<i>2.091,50</i>
Grand total :	20.149,12

Laboratoire de Recherche & Développement de l'EPITA

Type	Montant (€ TTC)
Nov. 2004 - Nov. 2005	
Equipement	8.147,51
Fonctionnement	1.794,00
Nov. 2005 - Nov. 2006	
Equipement	4.452,21
Fonctionnement	1.722,24
Nov. 2006 - Nov. 2007	
Equipement	4.781,61
Fonctionnement	11.657,41
<i>dont vacations</i>	<i>6.261,06</i>
Total sur 3 ans	
Equipement	17.381,33
Fonctionnement	15.173,65
<i>dont vacations</i>	<i>6.261,06</i>
Grand total :	38.816,04

Laboratoire d'Astrophysique de Marseille

Type	Montant (€ HT)
Nov. 2004 - Nov. 2005	
Equipement	0,00
Fonctionnement	544,00
Nov. 2005 - Nov. 2006	
Equipement	0,00
Fonctionnement	4.803,00
Nov. 2006 - Nov. 2007	
Equipement	0,00
Fonctionnement	16.465,00
Total sur 3 ans	
Equipement	0,00
Fonctionnement	21.812,00
Grand total :	21.812,00

Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées

Type	Montant (€ HT)
Nov. 2004 - Nov. 2005	
Equipement	8.890,00
Fonctionnement	2.083,02
Nov. 2005 - Nov. 2006	
Equipement	0,00
Fonctionnement	1.537,86
Nov. 2006 - Nov. 2007	
Equipement	0,00
Fonctionnement	21.796,22
<i>dont vacations</i>	7.748,00
Total sur 3 ans	
Equipement	8.890,00
Fonctionnement	25.417,71
Grand total :	34.307,71

Centre de Recherche Astrophysique de Lyon

Type	Montant (€ HT)
Nov. 2004 - Nov. 2005	
Equipement	0,00
Fonctionnement	1.112,84
Nov. 2005 - Nov. 2006	
Equipement	5.475,00
Fonctionnement	2.200,00
Nov. 2006 - Nov. 2007	
Equipement	4.744,92
Fonctionnement	8.154,38
Total sur 3 ans	
Equipement	10.219,92
Fonctionnement	11.467,22
Grand total :	21.687,14

III - Annexe : résumés des articles

L’analyse des images dans les grands relevs d’astronomie visible /proche-infrarouge (Bertin 2005)

Le volume et la nature des données issues des grands relevés d’imagerie astronomique posent des problèmes spécifiques en matière de traitement et d’analyse. Je présente les réponses actuelles, mais aussi les questions en suspens concernant la détection et l’analyse morphologique des sources dans le domaine visible/proche-infrarouge.

The EFIGI project : Automatic classification of galaxies (Baillard et al. 2006)

We propose an automatic system to classify images of galaxies with varying resolution. Morphologically typing galaxies is a difficult task and this is particularly true for distant galaxies convolved by a point- spread function and suffering from a poor signal-to-noise ratio. In the context of the first phase of the project EFIGI (extraction of the idealized shapes of galaxies in imagery), we present the three steps of our software : cleaning, dimensionality reduction and supervised learning. We present preliminary results derived from a subset of 774 galaxies from the Principal Galaxies Catalog (observed in g-band by the Sloan Digital Sky Survey) and compare them to human classifications made by astronomers. Finally, we discuss future improvements which we intend to implement before releasing our tool to the community.

Accurate photometric redshifts for the CFHT Legacy Survey calibrated using the VIMOS VLT Deep Survey (Ilbert et al. 2006)

We present photometric redshifts for an uniquely large and deep sample of 522286 objects with $i'_{AB} < 25$ in the Canada-France Legacy Survey “Deep Survey” fields, which cover a total effective area of 3.2 deg². We use 3241 spectroscopic redshifts with $0 < z < 5$ from the VIMOS VLT Deep Survey as a calibration to derive these photometric redshifts. We devise a robust calibration method which removes systematic trends in the photometric redshifts and significantly reduces the fraction of catastrophic errors. We use our unique spectroscopic sample to present a detailed assessment of the robustness of the photometric redshift sample. For a sample selected at $i'_{AB} < 24$, we reach a redshift accuracy of $\sigma_{\Delta z/(1+z)} = 0.037$ with $\eta = 3.7\%$ of catastrophic error. The reliability of our photometric redshifts is lower for fainter objects : we find $\sigma_{\Delta z/(1+z)} = 0.029, 0.043$ and $\eta = 1.7\%, 5.4\%$ for samples selected at $i'_{AB} = 17.5-22.5$ and $22.5-24$ respectively. We find that the photometric redshifts of starburst galaxies in our sample are less reliable : although these galaxies represent only 18% of the spectroscopic sample they are responsible for 54% of the catastrophic errors. We find an excellent agreement between the photometric and the VVDS spectroscopic redshift distributions at $i'_{AB} < 24$. Finally, we compare the redshift distributions of i' selected galaxies on the four CFHTLS deep fields, showing that cosmic variance is already present on fields of 0.8 deg².

VOTables in TERAPIX software (Bertin & Tisserier 2007)

The management of metadata is one of the critical tasks a data processing pipeline must also perform. VOTables represent a convenient and powerful standard to exchange metadata between the data- processing modules and the pipeline control software. We show how one can benefit from the VOTable output offered by the latest versions of

SExtractor, SCAMP and SWarp (three packages of the TERAPIX software suite) in a pipeline environment.

The SWIRE-VVDS-CFHTLS surveys : stellar mass assembly over the last 10 Gyrs. Evidence for a major build up of the red sequence between $z = 2$ and $z = 1$ (Arnouts et al. 2007)

We present an analysis of the stellar mass growth over the last 10 Gyrs using a large 3.6μ selected sample. We split our sample into active (blue) and quiescent (red) galaxies. Our measurements of the K-LFs and LD evolution support the idea that a large fraction of galaxies is already assembled at $z \sim 1.2$. Based on the analysis of the evolution of the stellar mass-to-light ratio (in K-band) for the spectroscopic sub-sample, we derive the stellar mass density for the entire sample. We find that the global evolution of the stellar mass density is well reproduced by the star formation rate derived from UV dust corrected measurements. Over the last 8Gyrs, we observe that the stellar mass density of the active population remains approximately constant while it gradually increases for the quiescent population over the same timescale. As a consequence, the growth of the stellar mass in the quiescent population must be due to the shutoff of star formation in active galaxies that migrate into the quiescent population. From $z = 2$ to $z = 1.2$, we observe a major build-up of the quiescent population with an increase by a factor of 10 in stellar mass, suggesting that we are observing the epoch when an increasing fraction of galaxies are ending their star formation activity and start to build up the red sequence.

Algorithme de calcul de l'arbre des composantes avec application à la reconnaissance des formes en imagerie satellitaire (Baillard et al. 2007)

Cet article présente un nouvel algorithme de calcul de l'arbre des composantes d'une image. Vis-à-vis de l'état de l'art, cet algorithme ne fait pas un usage excessif de la mémoire et travaille efficacement sur les images dont les valeurs ont une forte quantification et sur les images à valeurs à virgule flottante. Nous décrivons également une application de cet algorithme à l'identification d'objets pertinents dans des images d'astronomie.

Effective Component Tree Computation with Application to Pattern Recognition in Astronomical Imaging (Berger et al. 2007)

In this paper a new algorithm to compute the component tree is presented. As compared to the state of the art, this algorithm does not use excessive memory and is able to work efficiently on images whose values are highly quantized or even with images having floating values. We also describe how it can be applied to astronomical data to identify relevant objects.

Web services at TERAPIX (Ricou et al. 2007)

We present an implementation of V.O.-compliant web services built around software tools developed at the TERAPIX centre. These services allow to operate from a remote site several pipeline tasks dedicated to astronomical data processing on the TERAPIX cluster, including the forthcoming EFIGI morphological analysis tool.

An automatic method to determinate the degree of flocculence of a galaxy (Dumoncel et al. 2007)

We propose a new method to determine the flocculence of a galaxy image. Flocculence is characterized by a texture feature computed using a bank of Gabor filters. These filters, inspired by the human visual system, uniformly cover the spatial-frequency domain. Texture features are obtained by extracting statistics from sub-windows in the filtered images. Flocculent regions are then detected using a machine learning approach. First results are presented on the EFIGI dataset.

SuperNova Acceleration Probe (SNAP) : Investigating Photometric Redshift Optimization (Dahlen et al. 2008)

The aim of this paper is to investigate ways to optimize the accuracy of photometric redshifts for a SNAP like mission. We focus on how the accuracy of the photometric redshifts depends on the magnitude limit and signal-to-noise ratio, wave-length coverage, number of filters and their shapes and observed galaxy type. We use simulated galaxy catalogs constructed to reproduce observed galaxy luminosity functions from GOODS, and derive photometric redshifts using a template fitting method. By using a catalog that resembles real data, we can estimate the expected number density of galaxies for which photometric redshifts can be derived. We find that the accuracy of photometric redshifts is strongly dependent on the signal-to-noise (S/N) (i.e., $S/N > 10$ is needed for accurate photometric redshifts). The accuracy of the photometric redshifts is also dependent on galaxy type, with smaller scatter for earlier type galaxies. Comparing results using different filter sets, we find that including the U-band is important for decreasing the fraction of outliers, i.e., “catastrophic failures”. Using broad overlapping filters with resolution 4 gives better photometric redshifts compared to narrower filters (resolution $> \approx 5$) with the same integration time. We find that filters with square response curves result in a slightly higher scatter, mainly due to a higher fraction of outliers at faint magnitudes. We also compare a 9-filter set to a 17-filter set, where we assume that the available exposure time per filter in the latter set is half that of the first set. We find that the 9-filter set gives more accurate redshifts for a larger number of objects and reaches higher redshift, while the 17-filter set is gives better results at bright magnitudes.