

Décomposition d'images multispectrales de galaxies au moyen d'algorithmes de Monte Carlo par chaînes de Markov

Benjamin PERRET¹, Vincent MAZET¹, Christophe COLLET¹, Éric SLEZAK²

¹Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection
UMR 7005 CNRS-Université de Strasbourg, France
Pôle API, Bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex

²Laboratoire Cassiopée
UMR 6202 CNRS-Observatoire de la Côte d'Azur
Boulevard de l'Observatoire, BP 4229, 06304 Nice, Cedex 4, France
{perret,mazet,collet}@lsiit.u-strasbg.fr
eric.slezak@oca.eu

Résumé – Nous proposons une méthode d'analyse d'observations multispectrales de galaxies dans le but d'utiliser l'information spectrale pour raffiner les classifications morphologiques existantes. Nous avons opté pour une approche globale qui consiste à décomposer la galaxie en utilisant un modèle paramétrique basé sur une interprétation physique. Nous espérons ainsi permettre une interprétation directe des résultats. La méthode développée est entièrement automatique et s'appuie sur des algorithmes de Monte Carlo par chaînes de Markov. Les évaluations réalisées sur des images simulées et des images réelles sur 5 bandes montrent que cette méthode est robuste et précise.

Abstract – In this paper we propose a way to analyse multispectral observations of galaxies aiming at refining existing classifications with spectral information. We propose a global approach which consists of decomposing the galaxy into a parametric model using physically meaningful structures. Physical interpretation of the results will be straightforward. The proposed approach is fully automatic and performed using Markov Chain Monte Carlo algorithms. Evaluation on simulated and real 5-band images shows that this new method is robust and accurate.

1 Introduction

La classification des galaxies est une étape importante pour comprendre les schémas d'évolution et d'interaction des galaxies. Les classifications actuelles reposent sur le schéma proposé par Hubble en 1936 et étendu en 1959 par de Vaucouleurs [1]. Elles sont basées sur la morphologie en lumière visible des galaxies projetées sur le plan du ciel : les cinq classes majeures (elliptique, lenticulaire, spirale, spirale barrée et irrégulière) sont caractérisées par la présence plus ou moins importante de certaines structures.

De nos jours, de nombreuses techniques d'acquisition ont été développées et des images couvrant un large spectre de fréquences sont disponibles en grande quantité. Bien que ces données multispectrales donnent accès à des traceurs importants de l'évolution galactique tels que l'histoire du taux de formation d'étoiles ou la quantité de poussières, les systèmes de classification actuels ne permettent pas de prendre en compte simultanément les informations morphologiques et spectrales disponibles. Nous cherchons donc à développer une méthode d'analyse automatique d'images multispectrales de galaxies afin de pouvoir proposer à terme une nouvelle classification combinant la totalité de ces informations.

Au fil des années, deux types d'approches ont été développées pour l'analyse et la classification des galaxies. La première repose sur la mesure directe de caractéristiques déterminées empiriquement (indice de symétrie, rayon de Pérosian, indice de concentration, *clumpiness*, etc.) [3]. La deuxième approche utilise des techniques de décomposition sur différentes bases :

shapelets, analyse en composantes principales, pseudo bases modélisant des structures fondamentales (bulbe et disque) [2]. Les paramètres mesurés au moyen d'une de ces méthodes sont ensuite injectés dans un classifieur (machine à vecteurs de support, perceptron multicouches ou modèle de mélanges gaussiens).

La méthode de décomposition en shapelets [2] ou celle basée sur la mesure de caractéristiques [3] ont été utilisées dans le cas multispectral en travaillant bande par bande. Le classifieur est alors utilisé pour réaliser la fusion des informations spectrales. Cependant, les paramètres mesurés par ces méthodes sont très souvent difficilement interprétables en termes physiques, il est alors difficile de justifier les nouvelles classes proposées par le classifieur. Pour pallier cette difficulté, nous proposons d'utiliser une décomposition en structures physiques telles que bulbe et disque dont chaque paramètre a également une signification physique.

Le modèle paramétrique de décomposition des galaxies en structures physiques est bien connu des astronomes. Plusieurs logiciels permettent de réaliser cette décomposition : Gim2D [4], Budda [5] et Galfit [6]. Galfit et Budda permettent tous les deux une décomposition en disque, bulbe et barre stellaire (au moins) et utilisent des algorithmes déterministes avec les avantages de rapidité d'exécution et de reproductibilité des résultats mais le désavantage de la sensibilité aux paramètres initiaux. Gim2D, le plus ancien, permet une décomposition en bulbe et disque au moyen d'algorithmes stochastiques permettant une automatisation aisée au détriment d'un temps de calcul im-

portant. Néanmoins aucun de ces logiciels ne permet une estimation automatique à partir de données multispectrales. Nous avons donc développé une nouvelle méthode d'estimation avec un modèle paramétrique modélisant une observation multispectrale d'une galaxie composée, pour le moment, d'un disque, d'un bulbe et d'une barre stellaire.

2 Modélisation

Les galaxies sont constituées d'un ensemble possible de structures qui se caractérisent par leurs paramètres de forme, leur population stellaire, la quantité de poussière et de gaz qu'elles hébergent et leur dynamique. Ainsi, les galaxies elliptiques et le bulbe central sont des concentrations sphéroïdale d'étoiles relativement anciennes et donc dominées par des étoiles naines ou des géantes plus rouges qu'une étoile comme le Soleil. Le disque est une structure planaire s'étendant au delà du bulbe, sa concentration stellaire est plus faible, il contient des étoiles de tous âges et notamment des étoiles jeunes avec une quantité plus ou moins grande de gaz et de poussières. Une barre stellaire dont la population stellaire est voisine de celle du bulbe traverse parfois la galaxie. Les bras spiraux sont la matérialisation d'ondes de densités qui parcourent le disque : ils sont caractéristiques des galaxies spirales. Ils se manifestent par une concentration de gaz et de poussières et la présence de régions de formation stellaire (régions HII) voyant en particulier la naissance d'étoiles massives très lumineuses de courte durée de vie. Les galaxies irrégulières ne rentrent par définition pas dans ce schéma de décomposition.

Dans cet article, nous tiendrons compte d'un bulbe (\mathfrak{B}), d'un disque (\mathfrak{D}) et d'une barre stellaire ($\mathfrak{B}a$). Les bras spiraux ne sont pas modélisés car il n'existe actuellement aucun modèle prenant en compte aussi bien leur forme que leur profil de brillance. Des travaux futurs porteront sur la recherche d'un tel modèle. La modélisation de chaque structure est divisée en deux parties : d'une part un descripteur de forme permet de définir la morphologie des isophotes (contour d'égale brillance), d'autre part un descripteur de brillance permet de décrire la variation de la brillance des isophotes en fonction de leur distance au centre (c'est-à-dire un profil).

Le descripteur de forme utilisé est commun aux trois structures et consiste en une ellipse généralisée dont les paramètres sont : abscisse c_x et ordonnée c_y du centre, élongation e (rapport du grand axe sur le petit axe), angle de position α et déformation c . La longueur de l'axe majeur d'une ellipse passant par le point (x, y) est alors donnée par :

$$r(x, y) = \left(|\cos \alpha (x - c_x) - \sin \alpha (y - c_y)|^{c+2} + \left| \frac{\cos \alpha (y - c_y) + \sin \alpha (x - c_x)}{e} \right|^{c+2} \right)^{\frac{1}{c+2}} \quad (1)$$

Grâce au paramètre de déformation, l'ellipse peut varier continuellement d'une forme quasi rectangulaire à une forme étoilée, ce qui permet de modéliser la barre stellaire, ainsi que l'aspect particulier de certains bulbes.

La fonction de Sérsic [7] est utilisée comme descripteur de brillance pour les trois structures : elle modélise le profil de brillance le long de l'axe majeur de l'ellipse. Elle comporte trois paramètres : le rayon effectif R_e (rayon contenant la moitié

du flux total de l'objet), la brillance au rayon effectif I_e et l'indice de concentration n :

$$I(x, y) = I_e e^{-k_n \left(\left(\frac{r(x, y)}{R_e} \right)^{\frac{1}{n}} - 1 \right)} \quad (2)$$

où k_n est une fonction permettant d'assurer que le rayon caractéristique correspond à la moitié du flux total.

Chaque structure est décrite par une ellipse généralisée qui définit sa forme dans toutes les bandes et autant de fonctions de Sérsic qu'il y a de bandes. Les paramètres de déformation (c) et de concentration (n) du disque sont constants. La barre possède une variable supplémentaire R_{max} appelée rayon de coupure au delà duquel sa luminosité est nulle. De plus on suppose que les structures ont un centre commun. Nous obtenons un total de $11 + 8B$ paramètres libres avec B le nombre de bandes d'observation.

Enfin, nous supposons que la PSF (*Point Spread Fonction*) due à l'atmosphère est stationnaire sur l'image et son effet sur l'observation est modélisé par une convolution. Les bruits qui sont d'origines variées (bruit de comptage, électronique, de quantification, etc.) sont approximés par un bruit additif gaussien de moyenne nulle. Le modèle général s'écrit donc :

$$Y = Hm + e \quad (3)$$

où Y est l'observation, H la PSF, $m = \sum_k I_k(x, y)$ la somme des brillances des différentes structures k et e le bruit.

TABLE 1 – Les paramètres et leur loi a priori. Toutes les lois candidates sont des gaussiennes dont la matrice de corrélation (ou l'écart type pour les scalaires) est donnée dans la dernière colonne. Dans le cas des algorithmes à direction adaptative, la loi candidate permet de générer les premiers échantillons avant que le système adaptatif ne prenne le relai. Dans le cas des variables de brillance, l'algorithme de simulation ne requière pas de loi candidate.

Structure	Variable	Support	Corrélation-Écart type
$\mathfrak{B}, \mathfrak{B}a, \mathfrak{D}$	(c_x, c_y)	Image	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
\mathfrak{B}	e	[1 ; 10]	1
	α	[0 ; 2 π]	0,5
	c	[-0,5 ; 1]	0,1
	I	\mathbb{R}^+	
	R	[0 ; 200]	$\begin{pmatrix} 0,16 & -0,02 \\ -0,02 & 0,01 \end{pmatrix}$
	n	[1 ; 10]	
\mathfrak{D}	e	[1 ; 10]	0,2
	α	[0 ; 2 π]	0,5
	I	\mathbb{R}^+	
	R	[0 ; 200]	1
$\mathfrak{B}a$	e	[4 ; 10]	1
	α	[0 ; 2 π]	0,5
	c	[0,6 ; 2]	0,1
	I	\mathbb{R}^+	
	R	[0 ; 200]	$\begin{pmatrix} 0,16 & -0,02 \\ -0,02 & 0,01 \end{pmatrix}$
	n	[0,5 ; 10]	
	R_{max}	[10 ; 100]	1

3 Estimation

L'estimation des paramètres est un problème délicat pour différentes raisons : la dimension de l'espace des solutions (51 pour une image sur 5 bandes), les nombreux minima locaux, la possible multiplicité des solutions acceptables. Ainsi, nous avons opté pour une formulation bayésienne du problème et choisi l'estimateur du maximum a posteriori (MAP). Les lois a priori utilisées sont pour le moment des lois uniformes (Tab. 1), mais nous cherchons actuellement à définir des a priori plus informatifs avec les experts. Dans un premier temps, pour valider le modèle, nous avons privilégié la robustesse de l'algorithme au détriment de la vitesse de calcul en utilisant des algorithmes de Monte Carlo par chaînes de Markov (MCMC) [8].

L'algorithme utilisé est un échantillonneur de Gibbs [9] qui permet d'obtenir la distribution a posteriori des paramètres en simulant itérativement chaque inconnue selon sa loi a posteriori conditionnelle. La plupart des variables sont échantillonnées séparément, néanmoins nous avons choisi d'échantillonner certaines variables conjointement en raison de leur forte corrélation (par exemple les paramètres de la fonction de Sérsic). En outre, excepté les variables de brillance qui sont distribuées suivant des gaussiennes à support positif et pour lesquelles il existe des algorithmes efficaces de simulation [10], l'échantillonnage des variables est réalisé au moyen de variations de l'algorithme de Hastings-Metropolis (échelle et direction adaptatives [11]). Les lois candidates ont été déterminées empiriquement par l'étude des lois a posteriori. En outre, la technique du recuit simulé est utilisé afin de faciliter l'exploration de l'espace des solutions.

4 Résultats

La méthode a d'abord été testée sur des images simulées. Pour chaque paramètre nous mesurons deux valeurs : le MAP et la variance de la chaîne. Cette dernière nous fournit une estimation de l'incertitude sur ce paramètre. Nous avons pu observer de bons résultats pour lesquels le résidu (différence entre l'observation et l'estimation) ne contient plus que du bruit et les structures non modélisées, tandis que la vérité terrain se trouve dans le domaine d'incertitude (voir Fig. 1). Certains paramètres comme l'élongation, l'angle de position ou le centre sont estimés avec une très grande précision (erreur relative inférieure à 0,1%). Par contre, les paramètres de la fonction de Sérsic sont plus délicats à estimer, en particulier dans le cas du bulbe où l'information est portée par un faible nombre de pixels (d'extension équivalente à la largeur à mi-hauteur de la PSF). Les erreurs relatives d'estimation vont alors de 5% pour le disque et peuvent monter jusqu'à 20% pour le bulbe. Néanmoins, ces erreurs semblent se compenser mutuellement car leur impact sur le flux total mesuré et la vraisemblance est faible.

Les tests sur images réelles ont également confirmé la pertinence du modèle et de son extension aux données multibandes. La figure 3 montre l'estimation de la galaxie PGC1282. On constate que le résidu ne contient plus que du bruit et les structures non modélisées : les bras spiraux ainsi que des étoiles se trouvant en avant plan. Des artefacts dus à une modélisation imparfaite de la PSF sont également présents au centre de la galaxie.

Par ailleurs, l'approche multi-spectrale permet d'accroître la

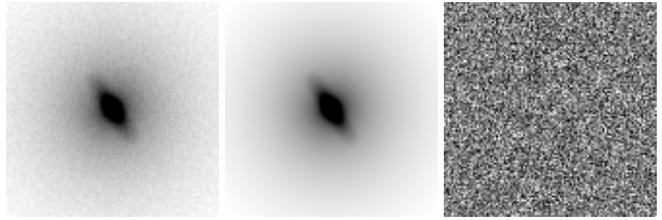


FIGURE 1 – Exemple d'estimation sur une image simulée (une seule des 5 bandes est montrée). À gauche : image simulée composée d'un bulbe, d'un disque et d'une barre stellaire. Au centre : résultat du processus d'estimation. À droite : résidu.

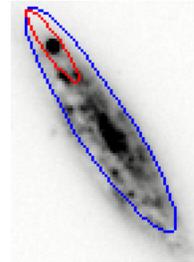


FIGURE 2 – Galaxie PGC10065 : un traitement monobande (g) de cette image donne un résultat incorrect (ellipse rouge), l'ajout d'une seconde bande (i) permet à l'algorithme d'obtenir une bonne estimation (ellipse bleue).

robustesse du processus et assure la cohérence des mesures entre les bandes (ce qui aurait été délicat avec un traitement bande par bande). Par exemple, dans le cas de la galaxie spirale PGC10065 (Fig. 2), lorsque la bande g est traitée seule, la détection est perturbée par une étoile très brillante au premier plan qui se trouve en haut de la galaxie (effet de projection), et l'algorithme se focalise sur cet objet très brillant. Lorsque l'on introduit une deuxième bande dans le processus (bande g et i), l'algorithme estime correctement la galaxie.

Enfin, nous avons comparé les résultats obtenus avec notre méthode avec ceux fournis par le logiciel Galfit. Nous avons utilisé 5 images monobandes (contrainte imposée par Galfit) extraites du catalogue PGC. Galfit demande une initialisation assez précise des paramètres initiaux pour fournir des résultats corrects : nous avons donc choisi d'utiliser les résultats de notre méthode légèrement modifiés par un bruit gaussien pour l'initialiser. Différents jeux de paramètres pouvant produire des solutions acceptables, une comparaison paramètre par paramètre n'est pas significative sans être accompagnée d'une évaluation qualitative, nous nous sommes donc restreint à comparer l'erreur quadratique moyenne et les flux intégrés de chaque structure (le flux intégré étant une caractéristique fondamentale d'une structure, on souhaite avoir des résultats aussi proche que possible entre les méthodes). On peut voir la bonne correspondance obtenue dans le tableau 2.

5 Conclusion

Nous avons proposé une extension du modèle de décomposition en bulbe, disque et barre stellaire des galaxies aux images multispectrales et un processus d'estimation entièrement automatique basé sur l'inférence bayésienne et des algorithmes

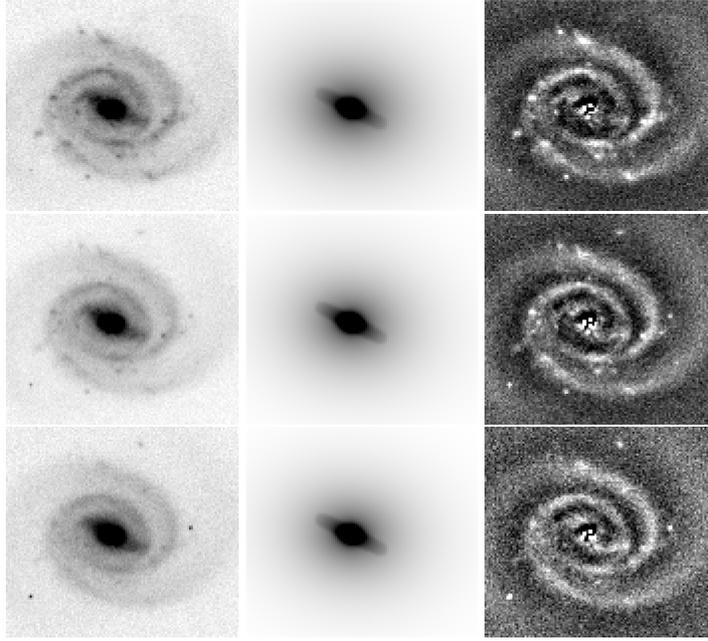


FIGURE 3 – Colonne de gauche : la galaxie PGC2182 (bandes g (4770 Å), r (6231 Å), et i (7625 Å)) est une galaxie spirale barrée. Colonne du centre : estimation. Colonne de droite : résidu. Les images sont données en niveaux de gris inversés avec une amélioration de contraste.

TABLE 2 – Comparaison entre notre méthode (N) et Galfit (G) pour la décomposition bulbe-disque sur 5 objets du catalogue PGC. Pour chaque ligne, la 1^{re} colonne donne le numéro PGC de l’objet, puis sont indiqués : l’erreur quadratique moyenne normalisée par pixel (EQM), la magnitude du disque (MD) et la magnitude du bulbe (MB) pour chaque méthode. (La magnitude est égale à $-2,5 \log_{10} F + F_0$ avec F le flux intégré de l’objet, le flux de référence F_0 est ici arbitrairement fixé à 26.)

Obj	(N)-EQM-(G)		(N)-MD-(G)		(N)-MB-(G)	
2182	2,53	2,69	12,6	12,55	15,65	15,35
2600	122,64	122,82	11,69	11,60	15,17	15,20
3563	8,48	9,65	13,13	13,13	12,24	12,24
3584	7,25	7,27	12,18	12,18	14,73	14,74
5055	1,56	1,59	14,92	14,95	13,24	13,24

MCMC. Notre approche décompose l’observation multispectrale de manière globale dans une base de fonctions correspondant à des structures physiques. En accord avec les experts, le lien entre les bandes est réalisé en liant certains paramètres. La méthode a été validée sur des images simulées et réelles et a été comparée avec les résultats d’un autre algorithme de décomposition.

Nous travaillons actuellement à l’enrichissement du modèle avec l’ajout d’une fonction fournissant une description multispectrale des bras spiraux. Nous souhaitons également définir des a priori plus informatifs avec les experts et effectuer des jeux d’évaluation plus exhaustifs.

Références

- [1] G. de Vaucouleurs. Classification and Morphology of External Galaxies. *Handbuch der Physik*, 53 :275, 1959.
- [2] B. C. Kelly et T. A. McKay. Morphological Classification of Galaxies by Shapelet Decomposition in the Sloan Digital Sky Survey. II. Multiwavelength Classification. *Astrophys. J.*, 129 :1287–1310, mars 2005.
- [3] S. Lauger, D. Burgarella et V. Buat. Spectro-morphology of galaxies : A multi-wavelength (UV-R) classification method. *Astron. & Astrophys.*, 434 :77–87, avril 2005.
- [4] L. Simard, C. N. A. Willmer, N. P. Vogt, V. L. Sarajedini, A. C. Phillips, B. J. Weiner, D. C. Koo, M. Im, G. D. Illingworth et S. M. Faber. The DEEP Groth Strip Survey. II. Hubble Space Telescope Structural Parameters of Galaxies in the Groth Strip. *Astrophys. J. Suppl.*, 142 :1–33, septembre 2002.
- [5] R. E. de Souza, D. A. Gadotti et S. dos Anjos. BUDDA : A New Two-dimensional Bulge/Disk Decomposition Code for Detailed Structural Analysis of Galaxies. *Astrophys. J. Suppl.*, 153 :411–427, aout 2004.
- [6] C. Y. Peng, L. C. Ho, C. D. Impey et H.-W. Rix. Detailed Structural Decomposition of Galaxy Images. *Astrophys. J.*, 124 :266–293, juillet 2002.
- [7] J. L. Sérsic. *Atlas de galaxies australes*. Cordoba, Argentina : Observatorio Astronomico, 1968.
- [8] R. W. Gilks, S. Richardson et D. J. Spiegelhalter. *Markov Chain Monte Carlo In Practice*. Chapman & Hall/CRC, Washington, D.C., 1996.
- [9] Ch. Robert. *Méthodes de Monte Carlo par chaînes de Markov*. Economica, Paris, 1996.
- [10] V. Mazet, D. Brie et J. Idier. Simuler une distribution normale à support positif à partir de plusieurs lois candidates. *GRETSI*, 6-9 septembre 2005, Louvain-la-Neuve, Belgique.
- [11] R. W. Gilks, O. G. Roberts et S. K. Sahu. Adaptive markov chain monte carlo through regeneration. *J. of the American Stat. Assoc.*, 93 :1045–1054, 1998.