



# EFIGI



http://www.efigi.org

## Projet EFIGI : Extraction des Formes Idéalisées de Galaxies en Imagerie

A. BAILLARD<sup>1</sup>, E. BERTIN<sup>1</sup>, Y. MELLIER<sup>1</sup>, R. DEHAK<sup>2</sup>, T. GERAUD<sup>2</sup>, O. RICOU<sup>2</sup>, H. MAITRE<sup>3</sup>, M. CAMPEDEL<sup>3</sup>, M. LIENOU<sup>3</sup>, P. FOUQUE<sup>4</sup>, R. PELLO<sup>4</sup>, J-F. LEBORGNE<sup>4</sup>, S. ARNOUITS<sup>5</sup>, G. PATUREL<sup>6</sup>, A. REFREGIER<sup>7</sup>

<sup>1</sup>Institut d'Astrophysique de Paris, Paris

<sup>2</sup>Laboratoire Traitement et Communication de l'Information, Paris

<sup>3</sup>Laboratoire d'Astronomie de Marseille, Marseille

<sup>7</sup>Service d'Astrophysique du Commissariat à l'Energie Atomique, Saclay

<sup>2</sup>Laboratoire de Recherche et Développement d'EPITA, Paris

<sup>4</sup>Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées, Toulouse

<sup>6</sup>Centre de Recherche Astronomique de Lyon, Lyon

Nous proposons un système de classification automatique d'images de galaxies de résolution variable. Attribuer un type morphologique à une galaxie est une tâche difficile, particulièrement pour les galaxies distantes convoluées par une réponse impulsionnelle et souffrant d'un rapport signal sur bruit faible.

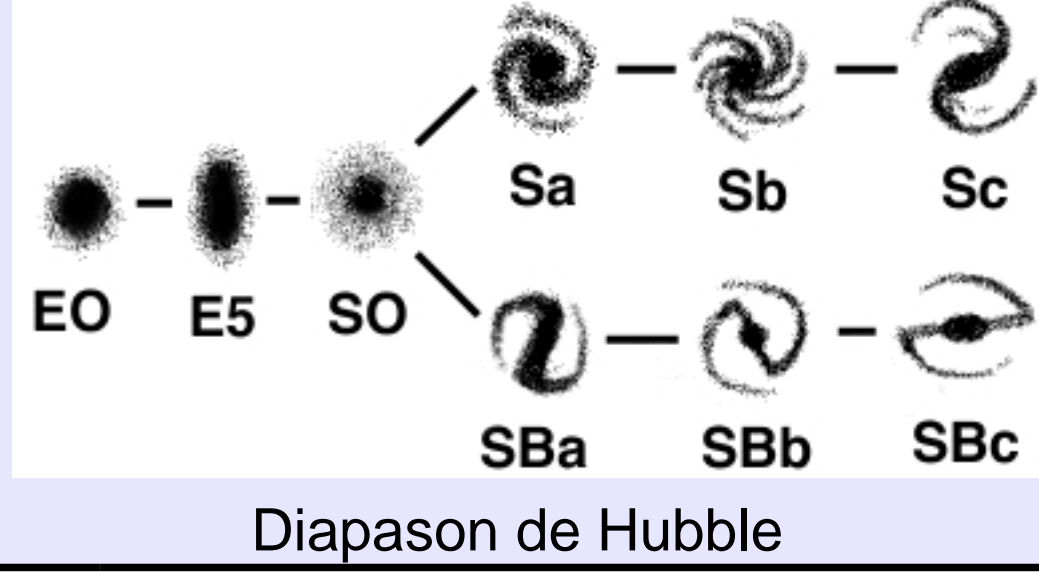
Nous expliquons les difficultés et les enjeux du projet EFIGI ainsi que la nature spécifique des données à traiter. Nous présentons le fonctionnement de notre logiciel automatique, l'interface interactive de classification manuelle et des résultats préliminaires obtenus après la première étape de développement du projet. Finalement, nous discutons des travaux que nous souhaitons mener avant de délivrer le logiciel à la communauté sous forme de "web service".

### Introduction

#### Motivation

La description morphologique et la classification des galaxies constituent des informations cruciales sur la façon dont les galaxies se forment, évoluent et sont altérées avec le temps cosmique. Une étude approfondie des types morphologiques et de leur sensibilité à tous les paramètres locaux et globaux nécessite la classification de plusieurs milliers de galaxies qui ne peut être faite manuellement. Des outils de classification automatisés robustes pouvant fournir des résultats fiables sont donc nécessaires. Le premier schéma de classification est le diapason de Hubble (*tuning fork*, 1936), mis à jour par de Vaucouleurs pour obtenir le *Revised Hubble System* (RHS, 1959). D'autres schémas existent, comme celui de Morgan (1985) ou le *DDO* (Van Den Bergh 1960).

Une classification globale a été proposée par la NASA sous le nom de *revised morphological types*. Elle unifie les schémas existants.



Diapason de Hubble

### Données

#### Galaxies

Le nombre de galaxies observées et classifiables depuis la Terre s'élève actuellement à environ un million et continue de croître avec l'amélioration des technologies et des programmes d'observation.

#### Images

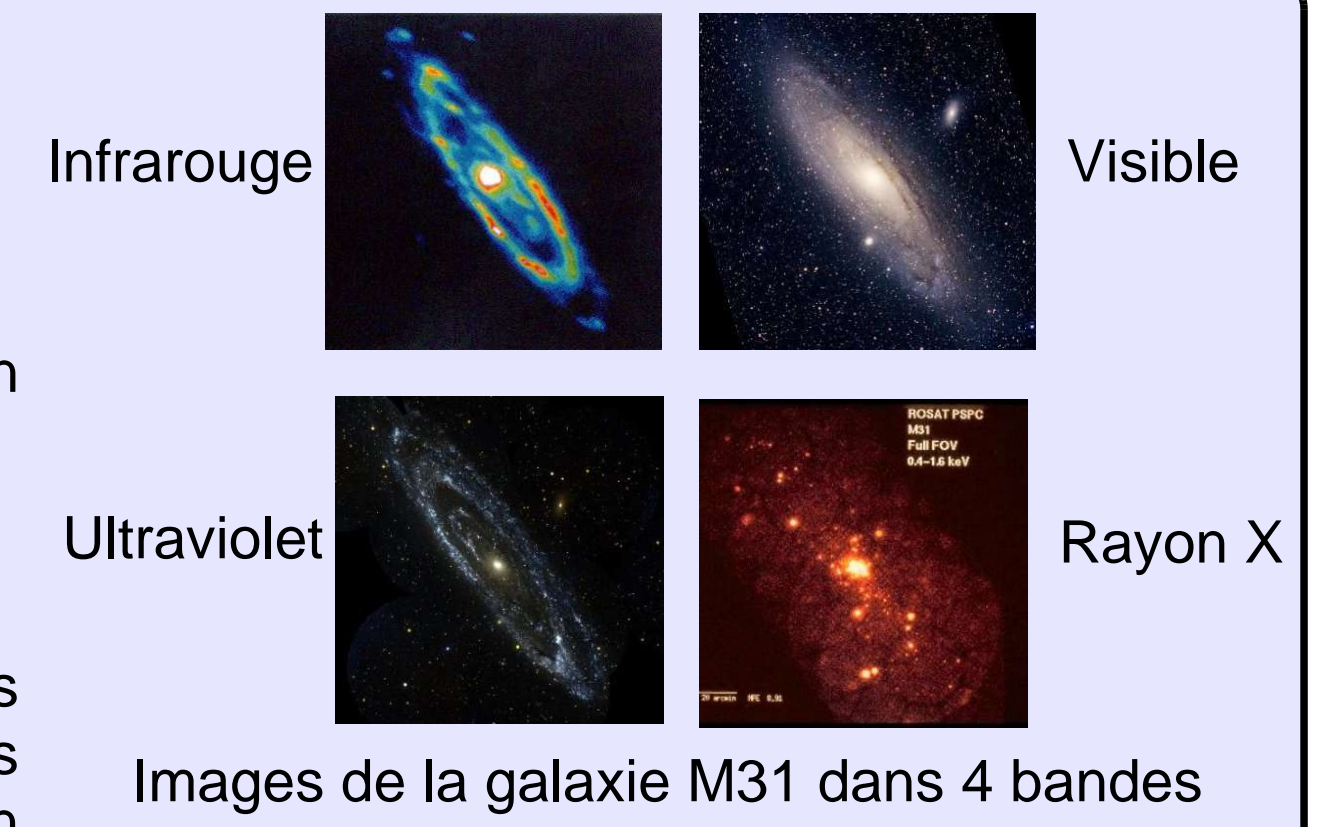
Les images de galaxies sont variées, elles peuvent provenir d'instruments opérant dans plusieurs longueurs d'onde du visible, dans l'ultra-violet ou encore dans l'infrarouge. L'aspect d'une galaxie est très différent en fonction de la bande d'observation. La résolution des images est également variable en fonction de la qualité des instruments, des observations et de la distance de l'objet étudié.

#### Catalogues

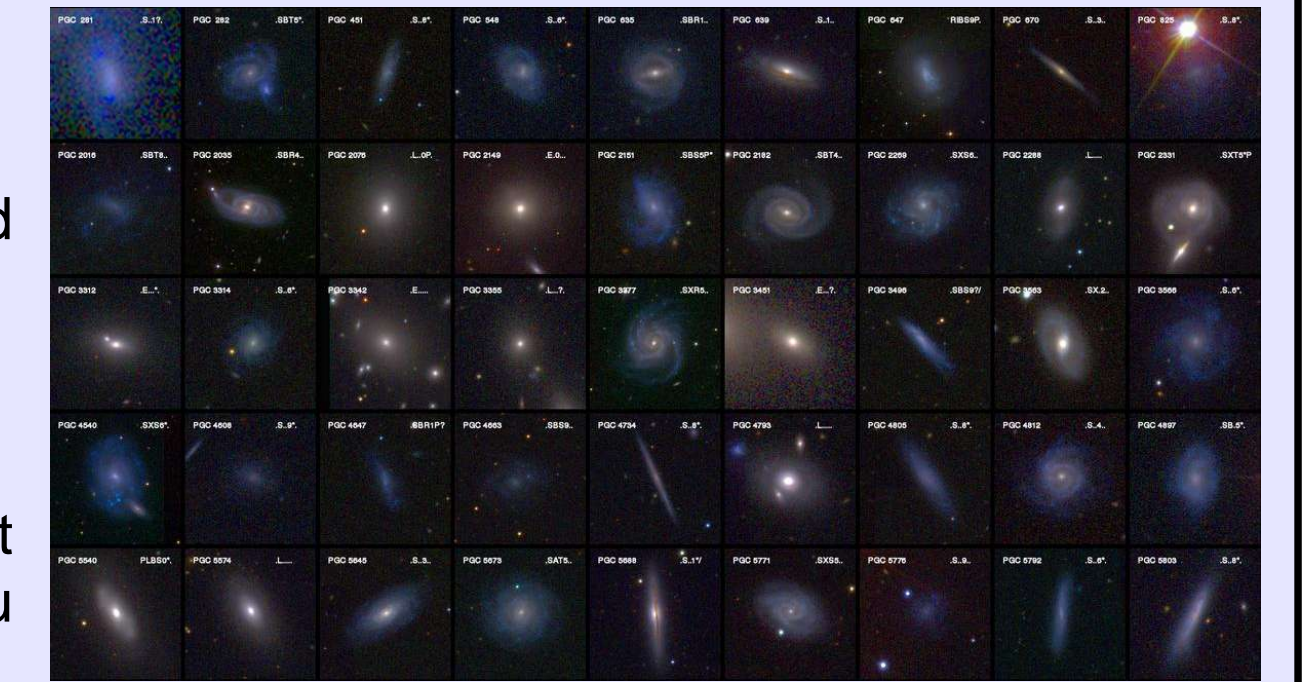
Actuellement, les catalogues de galaxies recensent les objets les plus connus mais il existe un grand nombre d'objets ne disposant d'aucune information morphologique.

#### Collecte des images

Une des tâches essentielles du projet est de collecter la grande quantité d'images de galaxies provenant de différents relevés astronomiques. Actuellement, nous disposons de 11 000 images en 5 bandes du *Sloan Digital Sky Survey* (soit 55 000 images pour un total de 2 To brut).



Images de la galaxie M31 dans 4 bandes



Extrait de l'ensemble de galaxies collecté

### Système automatique

#### Nettoyage

Les sources parasites comme les étoiles sont nettoyées sur les images grâce à l'utilisation d'opérateurs simples et efficaces issus de la morphologie mathématique.

#### Calcul d'invariants

Des opérations linéaires sont appliquées aux images afin de représenter les galaxies selon leur morphologie et non selon leur présentation (direction de l'axe principal, échelle, sens de rotation des bras).

#### Analyse en Composantes Principales

Nous obtenons une base de Karhunen-Loève directement sur les pixels et des coefficients de projection pour chaque image.

#### Apprentissage supervisé

L'utilisation d'un nombre réduit de coefficients permet d'entraîner des systèmes d'apprentissage supervisé : un perceptron multi-couches (PMC) et une machine à vecteurs de support (SVM).

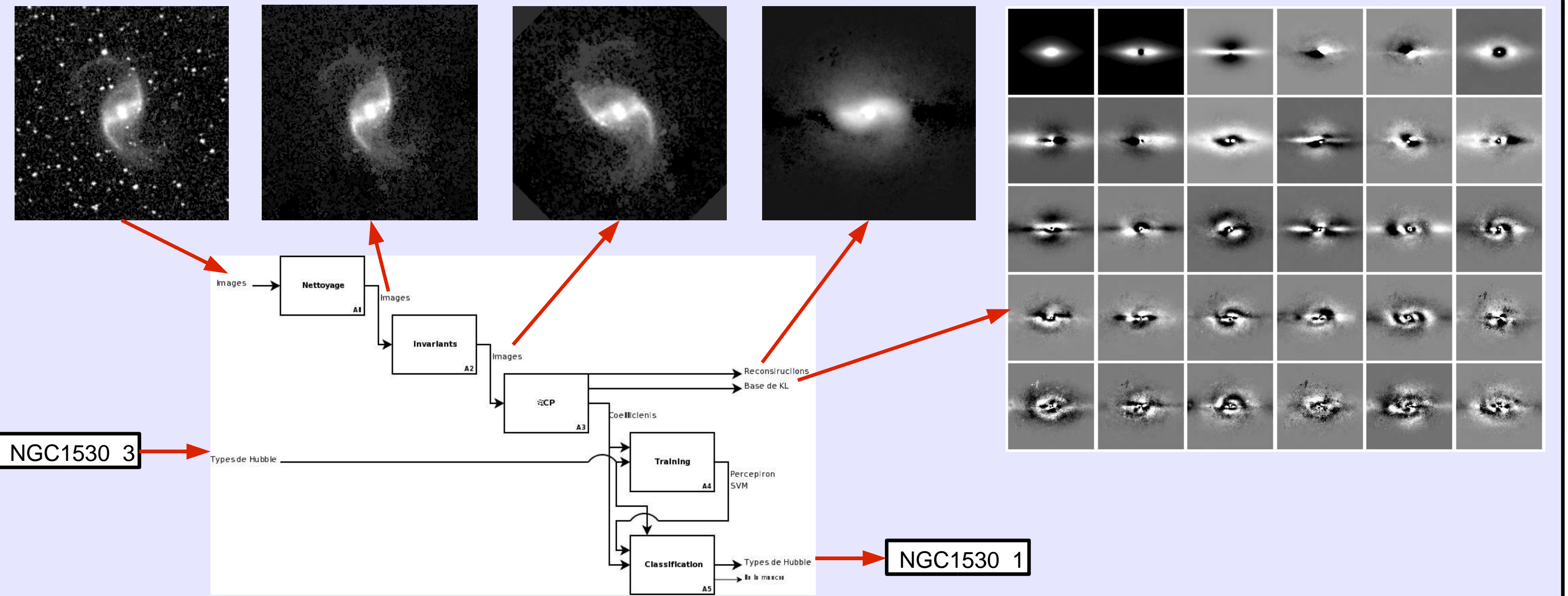
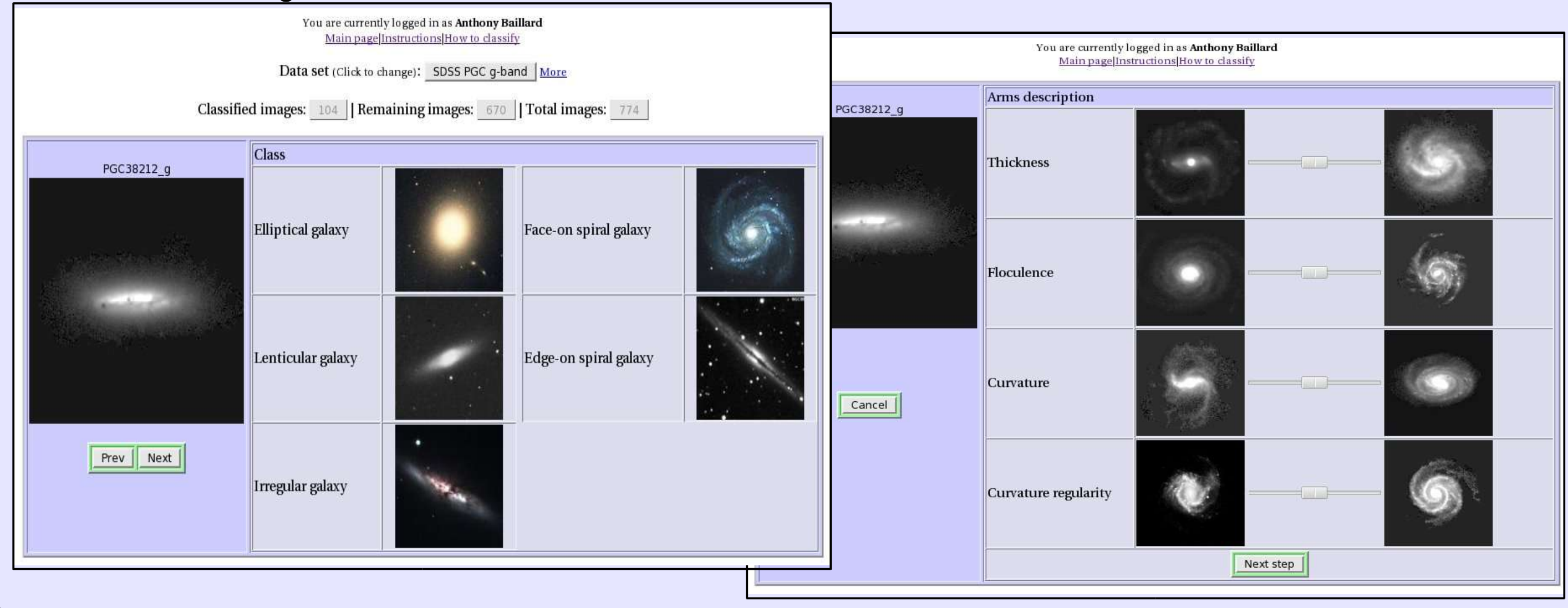


Schéma de fonctionnement global et exemple de la galaxie NGC 1530 en bande g

### Classification manuelle

#### Interface web

Une interface interactive liée à une base de données est proposée aux astronomes pour classer manuellement les galaxies.



### Résultats

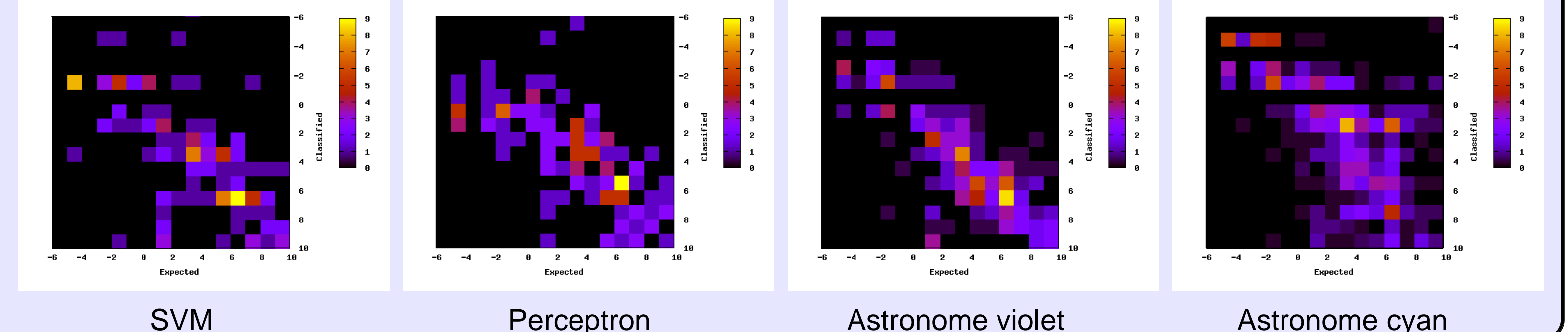
#### Performance de classification

Perceptron (à 2 types de Hubble près) : 68%  
SVM (à 2 types de Hubble près) : 75%

Astronome violet (à 2 types de Hubble près) : 74%  
Astronome cyan (à 2 types de Hubble près) : 66%

#### Matrices de confusion

On estime la qualité de la classification en comparant le type de Hubble attendu ( $T_{PGC}$  en abscisse) au type de Hubble donné par le classifieur ( $T_C$  en ordonnées).



### Comparaisons

#### Résidus ( $T_C - T_{PGC}$ )

$\sigma_1$  et  $\sigma_2$  (Odewahn 2004)

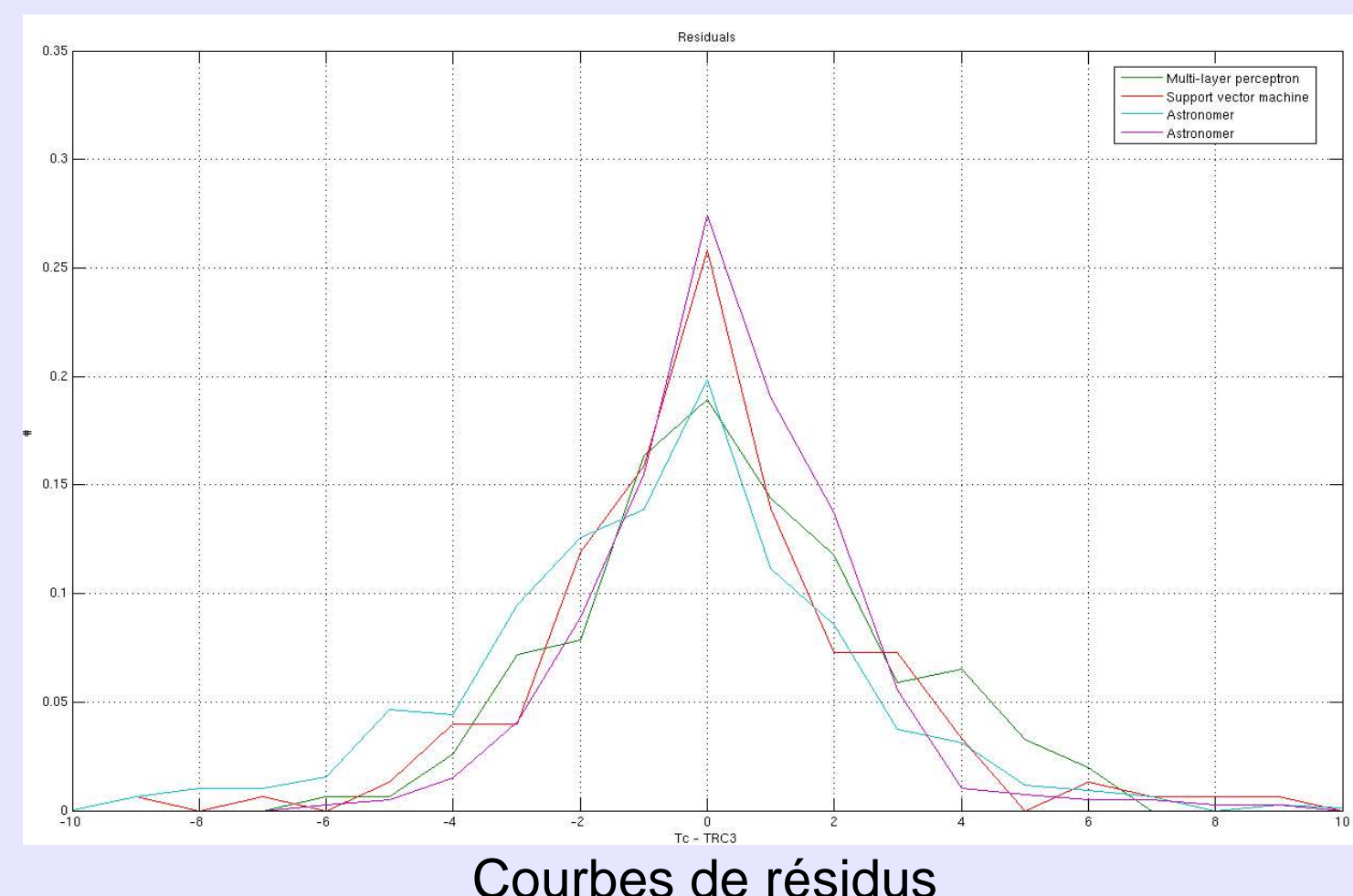
- On calcule d'abord une régression linéaire sur la matrice de confusion afin d'obtenir une pente  $\alpha$  indiquant l'adéquation entre le classifieur et le catalogue (une pente de 1 indiquant une adéquation parfaite). On calcule ensuite une déviation standard  $\sigma_1$  entre les données de la matrice de confusion et la régression linéaire obtenue.

- On calcule ensuite une déviation standard  $\sigma_2$  représentant la dispersion intrinsèque aux données :

$$\sigma_2 = \sqrt{\frac{\sum (T_C - T_{PGC})^2}{(N-1)}}$$

| Classifieur        | $\alpha$ | $\sigma_1$ | $\sigma_2$ |
|--------------------|----------|------------|------------|
| Astronome (violet) | 0.87     | 0.57       | 2.03       |
| Astronome (cyan)   | 0.65     | 1.49       | 3.18       |
| Perceptron         | 0.81     | 0.83       | 2.97       |
| SVM                | 0.83     | 0.70       | 2.66       |

Tableau récapitulatif



Courbes de résidus

### Conclusion

#### Critique

- + Le classificateur obtenu a montré des performances équivalentes à celles d'un spécialiste tout en étant beaucoup plus rapide (moins de 2 minutes au lieu de 10 à 20 heures).
- + Les formes des galaxies peuvent être décomposées en un nombre relativement faible de composantes et rester reconnaissables.
- Le classificateur actuel n'a été testé que sur des images de bonne résolution.
- La classification est pour le moment limitée au type 7 de Hubble.

#### Perspectives

Des données supplémentaires de l'UV au proche infrarouge doivent être récupérées et manuellement reclassifiées afin d'étendre le jeu d'entraînement.

Le système devra gérer les images de résolution variable, en opérant soit au niveau de la base de projection, soit en associant les informations de résolution aux descripteurs actuels. Le système devra aussi intégrer le contexte multi-spectral en gérant le décalage vers le rouge des galaxies faibles.

Au final, le web service permettra à un utilisateur de fournir une image quelconque de galaxie (accompagnée éventuellement d'une image de PSF et d'un masque) et d'obtenir en retour une liste d'attributs morphologiques détectés.