

Classification morphologique des galaxies à l'aide des shapelets

Marie LIENOU

Responsable: Marine CAMPEDEL

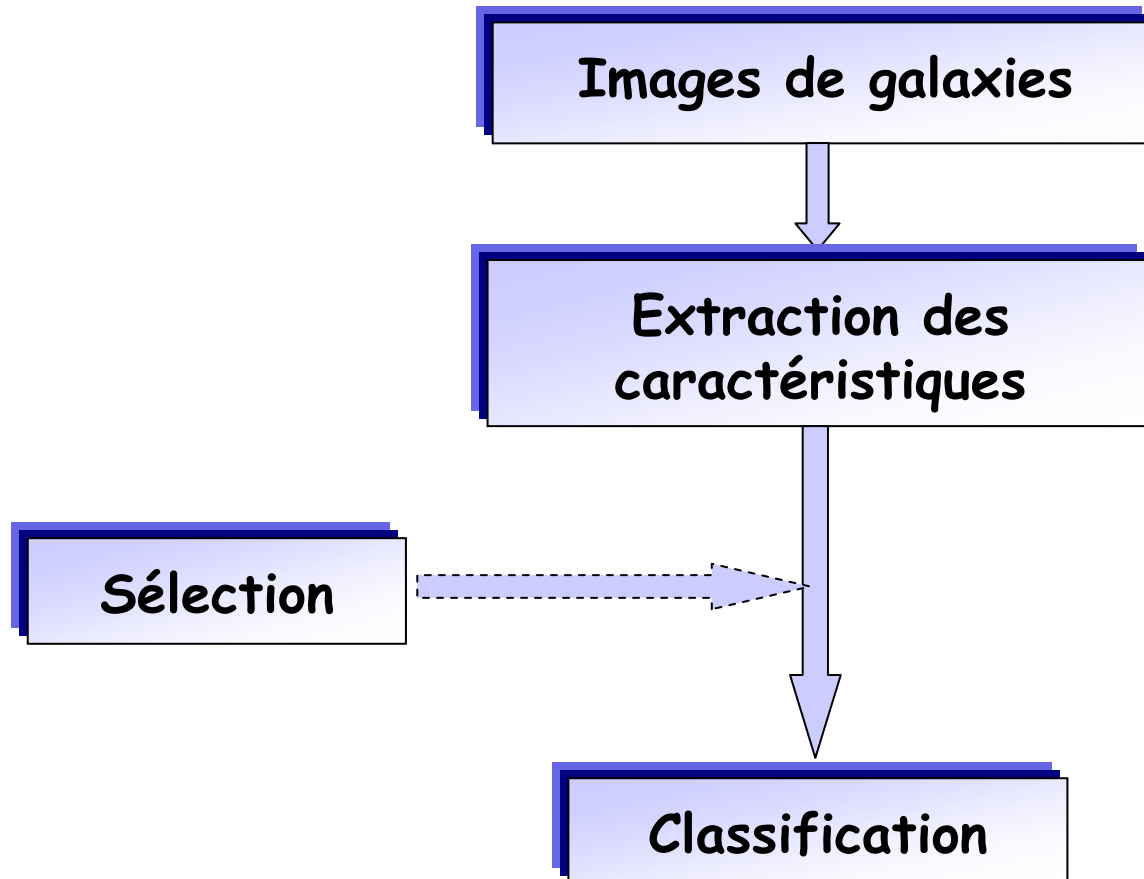
Meeting EFIGI - Octobre 2005

Plan

- ❑ Schéma de la classification
- ❑ La base d'images
- ❑ Extraction des caractéristiques
- ❑ Sélection et classification
- ❑ Tests et exploitation des résultats
- ❑ Conclusions

1. Schéma de la classification

La classification automatique



2. La base d'images

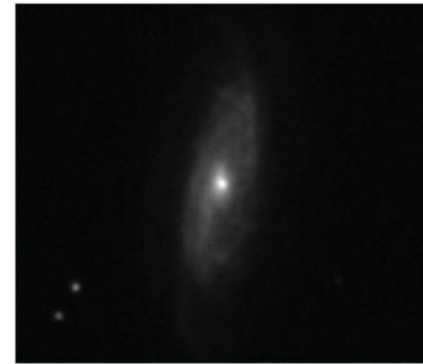
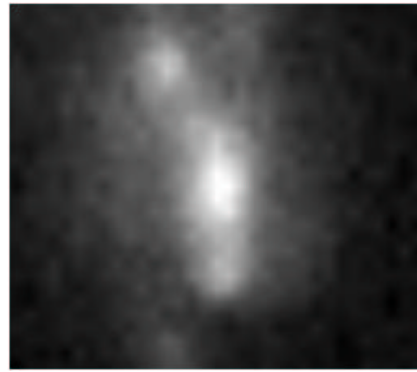
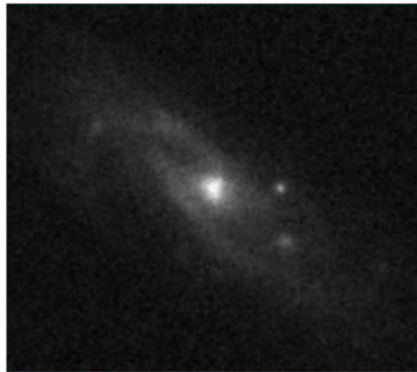
- ❑ 368 images extraites du catalogue PGC
- ❑ Inégale répartition des images entre les 4 grandes classes
- ❑ Présence d'images très bruitées → tri

201 retenues:

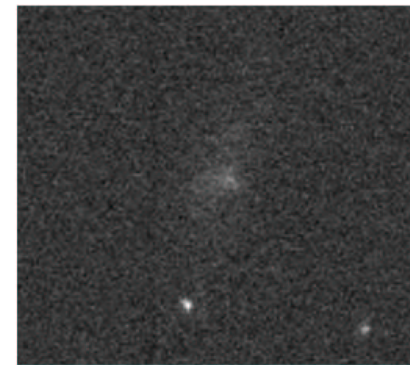
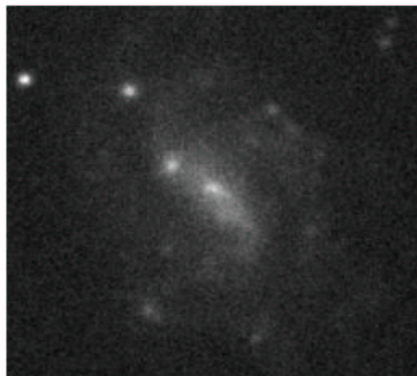
- 9 elliptiques
- 6 irrégulières
- 37 lenticulaires
- 149 spirales

2. La base d'images

Exemples d'images retenues



Exemples d'images rejetées



3. Extraction des caractéristiques

Les shapelets

- fonctions de base localisées, de formes différentes
- solutions de l'oscillateur harmonique quantique (QHO):

$$\begin{aligned}\phi_n(x) &= (2^n \pi^{\frac{1}{2}} n!)^{-\frac{1}{2}} H_n(x) e^{-\frac{x^2}{2}} \\ B_n(x; \beta) &= \beta^{-\frac{1}{2}} \phi_n(\beta^{-1} x)\end{aligned}$$

où $H_n(x)$: polynôme d'Hermite d'ordre n

β : paramètre d'échelle, proche de la taille de l'objet

- orthonormales et complètes:

$$f(\mathbf{x}) = \sum_{\mathbf{n}} f_{\mathbf{n}} B_{\mathbf{n}}(\mathbf{x}; \beta) \quad \text{et} \quad f_{\mathbf{n}} = \int_{-\infty}^{\infty} d^2 x f(\mathbf{x}) B_{\mathbf{n}}(\mathbf{x}; \beta)$$

3. Extraction des caractéristiques

Les shapelets cartésiennes 2-D

Obtenues à partir du produit tensoriel de 2 shapelets 1-D:

$$\phi_{\mathbf{n}}(\mathbf{x}) = \phi_{n_1}(x_1) \phi_{n_2}(x_2) \quad \text{et} \quad B_{\mathbf{n}}(\mathbf{x}; \beta) = \frac{1}{\beta} \phi_{\mathbf{n}}(\beta^{-1}\mathbf{x})$$

où $\mathbf{n} = (n_1, n_2)$ et $\mathbf{x} = (x_1, x_2)$

Les shapelets polaires 2-D

Interprétation physique plus directe et plus intuitive

$$\chi_{n,m}(r, \theta; \beta) = \frac{(-1)^{\frac{n-|m|}{2}}}{\beta^{|m|+1}} \left[\frac{\left(\frac{n-|m|}{2}\right)!}{\pi \left(\frac{n+|m|}{2}\right)!} \right]^{\frac{1}{2}} \times r^{|m|} L_{\frac{n-|m|}{2}}^{|m|} \left(\frac{r^2}{\beta^2} \right) e^{\frac{-r^2}{2\beta^2}} e^{im\theta}$$

où $L_p^q(x)$: polynômes de Laguerre

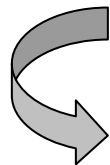
3. Extraction des caractéristiques

□ Propriétés:

- Simplification du calcul du **flux** et du **centroïde**
- Expressions simplifiées pour les **transformations** : translation, dilatation, rotation, cisaillement
- Convolution

Invariance par la transformation de Fourier, à un coefficient et un changement d'échelle près

$$\tilde{B}_n(k; \beta) = i^n B_n(k; \beta^{-1})$$



Convolution \cong simple multiplication de matrices!!!

3. Extraction des caractéristiques

- Symétrie rotationnelle

$$\chi_{n,-m}(r, \theta; \beta) = \chi_{n,m}^*(r, \theta; \beta) = \chi_{n,m}(r, -\theta; \beta)$$

Pour une fonction réelle $f(x)$, $f_{n,-m} = f_{n,m}^*$

Les spiralets

- Shapelets avec une dépendance radiale
- Améliorent la représentation des structures spirales

Décomposition en shapelets:

image de galaxie 2-D \rightarrow vecteur de coefficients dans un espace multidimensionnel

3. Extraction des caractéristiques

Nos caractéristiques

- 4 types de caractéristiques extraites:
 - Coefficients shapelets polaires
 - Coefficients spiralets
 - Flux
 - L'erreur moyenne de reconstruction
- Extraction de 300 coefficients par image
- Choix du paramètre d'échelle

4. Sélection et classification

Classification automatique

- Analyse du pouvoir discriminant des différents ensembles de caractéristiques
- Méthode supervisée: les classes sont connues à priori
- 3 classificateurs:

- k-plus proches voisins (k=3)

L'objet considéré est affecté à la classe majoritaire parmi ses k voisins les plus proches

- classificateur linéaire de Fisher

Basé sur la discrimination linéaire:

4. Sélection et classification

Problème à 2 classes: trouver une direction w perpendiculaire à un hyperplan, telle que la projection des données sur cet hyperplan les sépare au mieux

Plus de 2 classes: ensemble de problèmes à 2 classes du type 1 contre 1

- **SVM linéaire:**

Maximise la marge de séparation lorsque les données sont linéairement séparables

- Evaluation des performances par validation croisée

4. Sélection et classification

Sélection

- ❑ Identification des paramètres pertinents et non redondants
- ❑ Extraction d'un sous-ensemble permettant de classifier au mieux les données
- ❑ Algorithme de Fisher

Exploitation des poids de l'hyperplan estimé par l'analyse discriminante de Fisher utilisée en 1 contre 1

5. Tests et exploitation des résultats

Procédure

- Base petite et inhomogène → classification selon 2 classes: spirales et non spirales

- 2 types de tests:
 - Classification sans sélection
 - Classification avec la sélection Fisher

- Utilisation des caractéristiques seules, puis concaténations

5. Tests et exploitation des résultats

Classification sans sélection: caractéristiques seules

Coefficients shapelets 201 exemples - 2 classes, F = 3			
β	3-ppv	Fisher	svm
4	21.4 \pm 0.8	19.4 \pm 4.4	26.4 \pm 3.7
5	24.9 \pm 0.8	20.9 \pm 5.3	28.4 \pm 8.3
6	22.4 \pm 6.8	20.9 \pm 5.3	28.9 \pm 8.4
7	23.9 \pm 7.4	21.4 \pm 6.2	31.8 \pm 3.1
8	25.4 \pm 2.5	20.9 \pm 5.9	30.8 \pm 5.6
9	25.4 \pm 2.9	19.9 \pm 4.5	32.3 \pm 5.2
10	26.9 \pm 3.9	18.4 \pm 3.1	34.3 \pm 2.5

5. Tests et exploitation des résultats

- ❑ Résultats similaires pour les coefficients spiralets
- ❑ Dans les 2 cas, erreurs minimales pour Fisher

Caractéristiques	Erreur minimale	Matrice de confusion	
Coefficients shapelets	18.4 ± 3.1	0.7144	0.1584
		0.2856	0.8416
Coefficients spiralets	17.9 ± 2.5	0.7217	0.1528
		0.2783	0.8472

5. Tests et exploitation des résultats

Classification sans sélection: concaténations

□ Plusieurs combinaisons:

- Coefficients **shapelets/flux**
- Coefficients **shapelets/spiralets**
- Coefficients **spiralets/flux/mse**
- Coefficients **shapelets/spiralets/flux**
- etc...

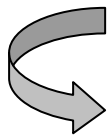
□ Résultats similaires au cas des caractéristiques seules

- Erreur du même ordre de grandeur, plus stable
- Erreurs minimales : Fisher et $\beta = 5$ et 10 (18.4 %)

5. Tests et exploitation des résultats

Conclusions

- ❑ La concaténation n'apporte pas d'amélioration notable de la capacité de reconnaissance
- ❑ Erreur de reconstruction: caractéristique quasi transparente
- ❑ Cause possible: nombre important de caractéristiques utilisées



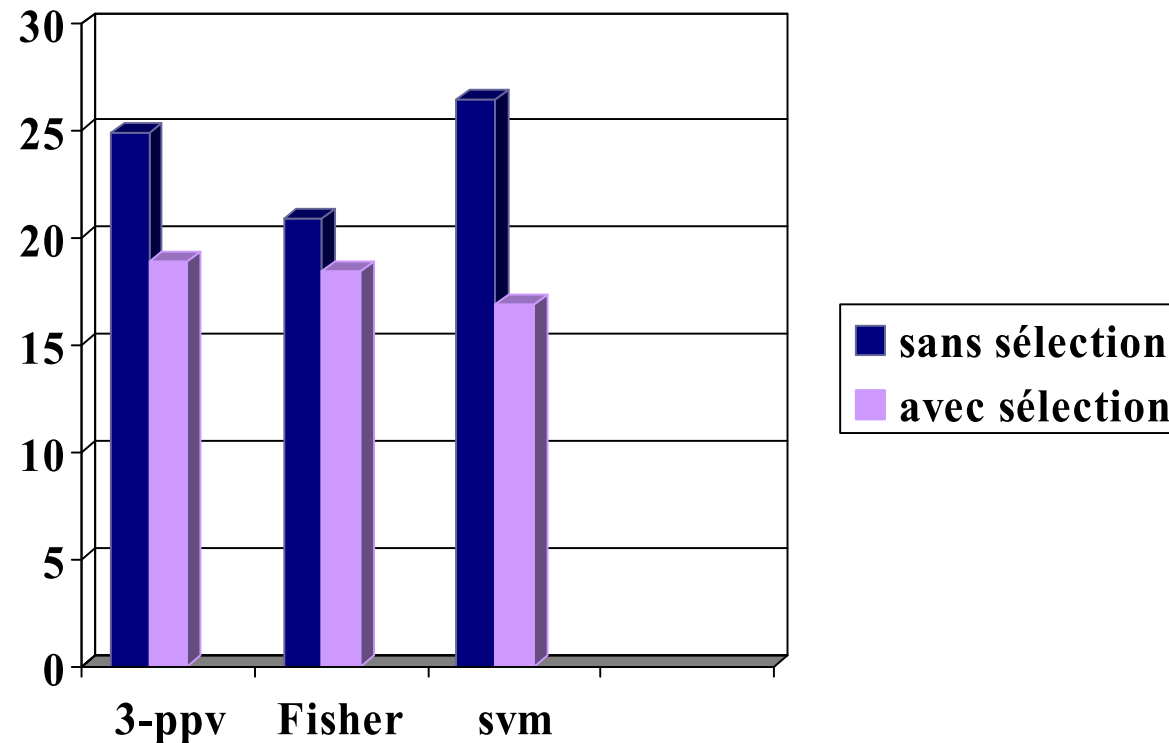
Augmentation du nombre de paramètres non pertinents et redondants

→ Intérêt de la sélection!!!

5. Tests et exploitation des résultats

Classification avec sélection Fisher: améliorations

Comparaison des performances de classification avec et sans sélection pour les coefficients spiralets et $\beta = 5$



5. Tests et exploitation des résultats

Classification avec sélection Fisher: améliorations

Sélection Fisher, 201 exemples, $\beta = 5$			
	Fisher +3-ppv	Fisher + Fisher	Fisher + svm
shapelets	17.4 ± 2.2 172	19.9 ± 4.5 20	17.9 ± 4.4 21
spiralets	18.9 ± 2.2 66	18.4 ± 2.1 152	16.9 ± 7.5 23
shap.+flux	17.9 ± 1.4 157	18.4 ± 4.5 35	17.9 ± 3.9 26
spir.+flux	16.9 ± 1.7 74	18.4 ± 3.1 146	16.9 ± 4.5 9
shap.+spir.	18.4 ± 0.8 143	18.9 ± 3.4 293	19.4 ± 2.9 12
shap.+spir.+flux	17.4 ± 0.8 147	18.9 ± 3.4 330	17.9 ± 2.9 12

5. Tests et exploitation des résultats

Classification avec sélection: performances

Caractéristiques	Erreur minimale	Classificateur	Rapport de sélection	β
shapelets	17.4 \pm 2.2	3-ppv/svm	172/300	5
spiralets	17.9 \pm 2.5	Fisher/svm	134/300	10
shap.+flux	17.9 \pm 1.4	3-ppv/svm	157/301	5
spir.+flux	16.9 \pm 1.7	3-ppv/svm	74/301	5
shap. + spir.	18.4 \pm 0.8	3-ppv/svm	143/600	5
shap. + spir.+ flux	17.4 \pm 0.8	3-ppv/svm	147/601	5

5. Tests et exploitation des résultats

Classification avec sélection Fisher

- ❑ Caractéristiques retenues: les plus grands coefficients
- ❑ Flux: caractéristique dominante
- ❑ Sélection efficace sur l'ensemble spiralets/flux
- ❑ Efficacité du classificateur svm qui utilise peu de paramètres
- ❑ Paramètre d'échelle $\beta = 5$

5. Tests et exploitation des résultats

Conclusions

- Intérêt de la sélection:
 - Réduction du nombre de caractéristiques sans perdre en efficacité de reconnaissance
 - performances meilleures

- Mauvaise classification pour les images de galaxies contenant plusieurs objets et bruitées

- Taille et inhomogénéité de la base

5. Tests et exploitation des résultats

Nouvelle base d'images

- ❑ 3773 images extraites du catalogue PGC-1.1
- ❑ Classification selon les 4 grandes classes
- ❑ Classificateurs utilisés:
 - k-plus proches voisins (k=10)
 - classificateur linéaire de Fisher
 - SVM gaussien
- ❑ Evaluation des performances par validation croisée (5 boucles)

5. Tests et exploitation des résultats

Caractéristiques:

- ❑ Coefficients shapelets: 210
- ❑ Composantes principales (ACP): 20

Classification sans sélection: $\beta = 5$

3773 exemples - 4 classes, F=5			
Caractéristiques	10-ppv	Fisher	svm gaussien
Coefficients shapelets	22.7 ± 4.9	48.1 ± 17.5	12.5 ± 6.5
Composantes ACP	24.8 ± 3.1	39.8 ± 8.5	25.3 ± 0.8

5. Tests et exploitation des résultats

Comparaison

Caractéristiques	Erreur minimale	Matrice de confusion			
Coefficients shapelets	12.5 ± 6.5 Svm gaussien	0.8350	0	0.0433	0.0250
		0.0114	0.8100	0.0041	0.0353
		0.0400	0.0364	0.8206	0.0614
		0.1136	0.1536	0.1321	0.8783
Composantes ACP	24.8 ± 3.1 10-ppv	0.3527	0	0.1468	0.0335
		0.0256	0.5685	0.0224	0.0491
		0.2767	0.0667	0.3979	0.0810
		0.3450	0.3648	0.4329	0.8365

Conclusions

- ❑ Coefficients shapelets, plus discriminants que les composantes ACP
- ❑ Résultats meilleurs avec la procédure de sélection de Fisher
- ❑ Utilisation des coefficients spiralets et le flux

Merci de votre attention!