Astronomical Data Analysis Software and Systems XVII **P5.2**
ASP Conference Series, Vol. XXX, 2008
J. Lewis, R. Argyle, P. Bunclarck, D. Evans, and E. Gonzales-Solares, eds.

An automatic method to determinate the degree of flocculence of a galaxy

Jean Dumoncel[1], Marine Campedel[1], Henri Maître[1], Anthony Baillard[2], Emmanuel Bertin[2], Valerie de Lapparent[2], Yannick Mellier[2], Pascal Fouqué[3], Jean-Francois Le Borgne[3], Roser Pelló[3], Dmitry Makarov[4], Lidia Makarova[4], Phillipe Prugniel[4] and Stephane Arnouts[5]

[1]*LTCI, Ecole Nationale Supérieure des Télécommunications, Paris, France*

[2]*Institut d'Astrophysique, Paris, France*

[3]*Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées, Toulouse, France*

[4]*Centre de Recherche Astronomique, Lyon, France*

[5]*Laboratoire d'Astronomie, Marseille, France*

Abstract. 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.

1. Introduction

The EFIGI[1] project proposes to address both the computational and algorithmic aspects of the extraction of useful morphological information from galaxy images in a fully automated way. Flocculence is a prominent feature in some galaxies, characterised by a patchy, fragmented ("flocculent") texture of the disk, in opposition to a smooth and regular spiral structure. Although the degree of flocculence in a galaxy can be related to "concentration", "coarseness", "asymmetry" or the Gini coefficient (e.g. Abraham et al. 2003, Conselice et al. 2003, Lotz et al. 2004, Yamauchi et al. 2005), these simple estimators merely estimate the degree of irregularity of the images, but are not meant to isolate a particular feature such as flocculence. In the present work, the problem is studied from the point of view of texture classification.

[1]`http://www.efigi.org`

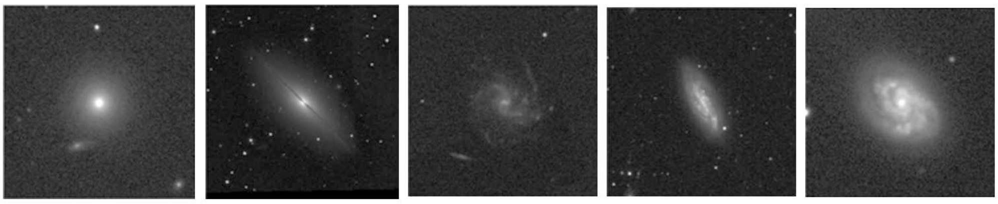

Figure 1. Some galaxies with flocculence indices ranging from 0 (left) to 4 (right).

2. The EFIGI dataset

Identifying flocculent features requires well-resolved galaxy images. The dataset used in this study consists of 4462 SDSS g-band images centered on galaxies with typical diameters (D25) of 1 arcmin, extracted from the RC3 catalogue (de Vaucouleurs et al. 1991). A "flocculence (integer) index" ranging from 0 (no flocculence) to 4 (strong flocculence, see Fig. 1) has been assigned by eye to each galaxy by a pool of 9 astronomers, and its homogeneity checked by one of us (E.B.).

3. Feature extraction

All the galaxy images of our catalogue are background-subtracted, resampled, and rescaled to 255x255 pixels. Sub-windows are then extracted within the isophotal limits of each galaxy; the centres of sub-windows are defined by their polar coordinates with respect to the barycentre or the parent galaxy. From now on a galaxy image is treated as a mosaic of textures.

In order to characterise the textures, the content of all sub-windows is filtered with a bank of Gabor filters. This filter bank works at multiple scales and orientations and shares some similarities with the receptive fields of the mammalian visual pathway. A Gabor function is the product of a Gaussian function and a sine function (e.g. Manjunatah & Ma 1996):

$$h(x,y,u,\theta) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}(\frac{x_\theta^2}{\sigma_x^2}+\frac{y_\theta^2}{\sigma_y^2})} cos(2\pi u x_\theta), \tag{1}$$

where $x_\theta = x\cos(\theta) + y\sin(\theta)$, $y_\theta = -x\sin(\theta) + y\cos(\theta)$, and u and θ are the frequency and phase along the x-axis, respectively. The content of sub-windows is convolved with the set of Gabor filters for 4 orientations (0, 45, 90 and 135°) and 4 scales (Fig. 2). The mean and the standard deviation of the resulting sub-images are used to characterise the region, yielding a feature vector with 32 components per sub-window:

$$F = (\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01}, ..., \mu_{44}, \sigma_{44}), \tag{2}$$

Although the orientation of the Gabor filters in the original vector is given relative to the image axis, flocculent features are expected to be stablest when measured relative to the tangential (or radial) direction. Therefore feature vector

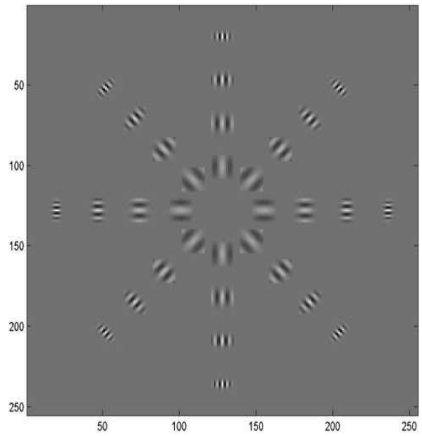

Figure 2. Set of two-dimensional Gabor filter functions in a 24x24 pixels window for 4 scales and 4 orientations.

components must be permuted to compensate for the position angle of the current sub-window. For example if the position angle is in the interval [22.5°..67.5°] then F in (2) becomes:

$$F = (\mu_{10}, \sigma_{10}, \mu_{11}, \sigma_{11}, ..., \mu_{04}, \sigma_{04}). \tag{3}$$

4. Unsupervised classification and histogramming

In principle, the Gabor characteristics could be used directly as inputs to a classifier. But the relatively large number of samples and the variety of presentation of flocculence makes it worth to carry out an intermediate classification in a larger number of classes. To estimate the optimal number of classes in our catalogue of images, we adopt a criterion based on Minimum Description Length (MDL) (Kyrgyzov et al. 2007), providing 20 classes C_j together with their centre coordinates in feature space. Each of the n sub-windows X_k (with $k \in [1..n]$) of a galaxy is assigned a class C_i. For each galaxy, a histogram H is built, with components

$$H_{C_i} = \sum_{k=1}^{n} (X_k \in C_i). \tag{4}$$

H can now be used as a new feature vector.

5. Supervised classification

Our supervised classifier is a Support Vector Machine (SVM, Cortes & Vapnik 1995) with Gaussian kernel, trained to separate each of the 5 classes (flocculence indices) against all the others from the set of H vectors. Not unexpectedly, the best classification success rates measured through cross-validation are obtained

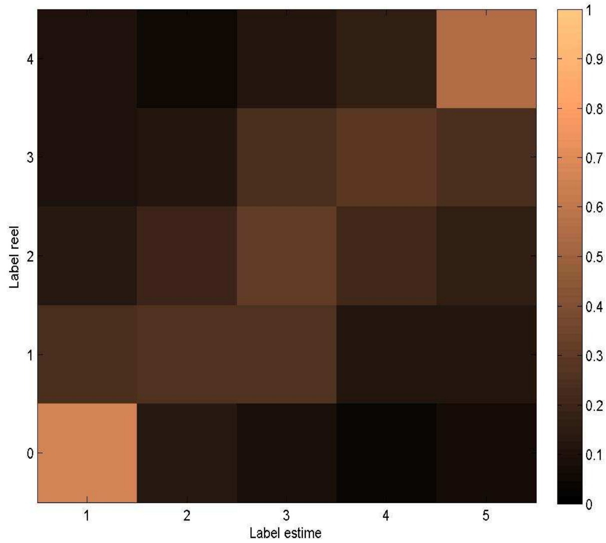

Figure 3. Confusion matrix for the flocculence index using the SVM classifier

for indices 0 and 4 ($> 80\%$). Intermediary flocculence indices seem more difficult to discriminate (40% success rate for exact matching), but these numbers increase to a respectable 70% if a tolerance of ± 1 index is allowed (Fig. 3), comparable to what is expected from interpersonal variability amongst astronomers.

6. Conclusion

We have introduced a supervised texture classification algorithm of well-resolved galaxy images using a fixed set of Gabor filters. The classifier is able to discriminate between flocculent and non-flocculent galaxies with a $\geq 70\%$ success rate. Future work will focus on dealing with images more strongly affected by seeing.

References

Abraham R.G., Van Den Bergh S., Nair P., 2003, ApJ 588, 218

Conselice C.J., 2003, ApJS 147, 1

Cortes C., Vapnik V., 1995, Machine Learning 20, 273

Kyrgyzov I.O., Kyrgyzov O.O., Maître H., Campedel M., 2007, in LNCS 4571, 203

Lotz J.M., Primack J., Madau P., 2004, ApJ128, 163

Manjunatah B.S., Ma W.Y., 1996, IEEE Transactions on Pattern Analysis and Machine Intelligence 18, 837

de Vaucouleurs G., de Vaucouleurs A., Corwin H.G., Buta R.J., Paturel G., Fouqué P., 1991, Third Reference Catalogue of Bright Galaxies (RC3), Springer-Verlag, New York

Yamauchi C., Ichikawa S.-I., Doi M., Yasuda N., Yagi M., Fukugita M., Okamura S., Nakamura O., Sekiguchi M., Goto T. 2005, ApJ 130, 1545