

MULTIWAVELENGTH IMAGE ANALYSIS OF THE SMALL MAGELLANIC CLOUD USING HIERARCHICAL MARKOVIAN SEGMENTATION

Mireille Louys^b, Anaïs Oberto[‡], Caroline Bot[‡], Christophe Collet^b

^bUniversité Strasbourg I, LSIT : UMR CNRS 7005

LSIT, Pole API, Bd S. Brant - 67400 Illkirch - France

[‡]Observatoire astronomique de Strasbourg, UMR CNRS 7550

<http://astro.u-strasbg.fr/Obs.html>, ^b<http://picabia.u-strasbg.fr/lsiit/> name@lsiit.u-strasbg.fr

Abstract

This paper addresses the segmentation of astronomical multiband images with missing data. We present some results obtained on Multiwavelength images of the Small Magellanic Cloud, by using the Marginal Posterior Mode (MPM) estimator on a quadtree structure under Markovian assumption : the estimation of the model parameters is then addressed with Expectation-Maximization (EM)-type algorithms, allowing unsupervised hyperparameter estimation. The main interest of this modeling effort lies in its generality : the algorithm handles multiwavelength data (possibly with missing data) in a single Causal-in-scale Markovian model. It is an interesting tool for astronomical image analysis, which exhibits very large dynamic range of intensities and missing data on the sampling grid in this case.

1. MULTIBAND IMAGES IN ASTRONOMY

The study of star formation mechanisms and their relationship with interstellar medium is one of the most dynamic fields of research in astronomy. The interstellar medium is composed of a mixture of gas (mainly hydrogen and helium) in different phases and dust grains (mainly carbon and silicium). Stars are forming from the gas in the heart of molecular Hydrogen complexes. The molecular clouds themselves result from the formation of molecules via chemical reactions in dense atomic gas. The ultra violet emission of massive newborn stars ionizes the surrounding gas, giving birth to the so-called HII regions. Observations at various wavelengths are necessary to study the different states of gas and dust and their mutual relationships as well as their links to star formation. The emission line at 21cm in the radio wavelength range gives the column density of atomic hydrogen (HI), which is the number of hydrogen atoms in an unitary section cylinder along the line of sight. Emissions in the far Infrared at 100 or 170 microns are directly due to the thermic emission of big dust grains. The H α emission, at 656.3nm, is the strongest Hydrogen recombination line in the HII regions and is generally superimposed on the optical continuum of stars[1]. Molecular Hydrogen is difficult to detect, but estimation of their density can be inferred from the intensity of the CO lines at millimeter wavelengths, assuming that this CO and molecular hydrogen are formed at quite the same time[2].

2. OBSERVATIONS

The region observed here is the Small Magellanic Cloud (SMC), a close satellite galaxy showing physical properties different of that of the Milky Way but close enough to allow high resolution maps to be

obtained. The images we studied here are centered at equatorial coordinates : 00h52m07s -72d49m ($J2000$) covering a large fraction of the SMC. These are the following :

- The "HI image", Figure 2, observed at 21.1cm is obtained from both the Australian Telescope Compact Array (ATCA) radio interferometer and from the Parkes telescope. These combined observations allow a good spatial frequency coverage ;
- The "PHOT image", Figure 3, coding the luminosity value recorded at 170 μ m by the ISOPHOT photometer on board of ISO, the Infrared Space Observatory operated by the European Space Agency (ESA). On this particular image, some missing pixels appear as a regular lattice of undefined pixels values. These gaps come from accidental undersampling of the observations ;
- The "H α image", Figure 4, observed at 656.3nm for the same region is used as explanatory data in order to examine correlation with the star forming regions ;
- The "IRAS100 image", Figure 5, observed at 100 μ m from the IRAS instrument (Infrared Astronomical Satellite) is used in conjunction with the PHOT image to estimate the temperature of regions of different components.

All these images have been registered and resampled with respect to the ISOPHOT image. The goal consists in using "HI image" and "PHOT image" to segment the SMC cloud into different classes, in order to be able to estimate physical parameters such as the temperature for each of them. In order to do that, we use a bi-wavelength Markovian-in scale segmentation.

3. IMAGE ANALYSIS WITH THE MARKOVIAN HIERARCHICAL METHOD

For segmentation tasks, with strongly noisy images, Markovian assumption allows the description of global behaviors by considering, on a predetermined spatial neighborhood around each pixel, statistical relationships between observation field and label field (*i.e.*, segmented image). Nevertheless, Markov Random Fields (MRFs) lead to robust but iterative procedures (*i.e.*, computing demanding due to slow convergence[3]). To avoid such difficulties, hierarchical modeling allows the definition of different coarse-to-fine strategies, under (spatially or/and in-scale) Markovian assumptions[4]. Since 1995, hierarchical framework has been extensively studied with success for a large variety of applications[5, 6, 7] requiring restoration, segmentation or classification steps. In this way, we consider a special class of Markov models, which helps to circumvent the latter drawbacks (*i.e.*, iterative often computational intensive estimation algorithms). Indeed, the Causal-in-scale Markov Random models,

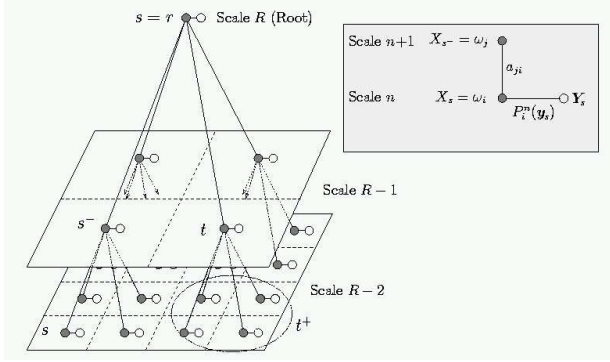


Fig. 1. Quadtree structure and its dependency graph. The upper-right gray box shows the Markovian link in scale (a_{ij} stands for the probability of down-transition from class ω_j to class ω_i) whereas $P_i^n(\mathbf{y}_s)$ corresponds to the Gaussian likelihood linking observation \mathbf{y}_s and class label $x_s = \omega_i$.

attached to the nodes of a quadtree, are defined in a general manner, allowing to analyse simultaneously multiband and/or multiresolution images [8, 9]. The great interest of this model results in an interesting causality property through scale, which allows the design of exact and non-iterative inference algorithms which are similar to those used in the context of Markov Chain Models (MCM).

In our example, on both HI and ISOPHOT images, we applied the markovian quad-tree segmentation [9]. We obtain the 8-classes segmentation map given in Figure 6(a) and the corresponding edge map overlaid on top of the H α image, in Figure 6(b). The noise statistics are supposed to be gaussian for each class. We drove the analysis in 2 steps :

1. First, we segmented the pixel population in 2 classes only, and derived a mask focusing on the high response regions in both bands ;
2. Then for each of these 2 classes, we applied a 4-class segmentation ;
3. We finally obtained a map with 8 classes, Figure 7(a), more efficient to discriminate the pixels in the region of interest than the straightforward 8-class segmentation on the original image shown on Figure 6(a).

One important point of the Markovian quad-tree segmentation, is its capability to overcome the missing zones. Indeed, it could be interesting having a map of segmentation without superimposed missing pixels : the quad-tree propagation of the labels and transitions laws in the method, allows to overcome the missing pixels problem and spacing artefacts, as shown on Figure 6(a).

All the regions identified here are homogeneous with respect to their spectrophotometrical information. Further analysis of physical properties in the classes can then take place, particularly the class by class correlated temperature assumption with the IRAS map.

4. RESULTS ON BISPECTRAL ASTRONOMICAL IMAGES

Thus, by combining the segmentation map and the missing pixels list, we derive a tool for selecting the valid pixels of the different classes. For each of them, statistical, geometrical or physical parameters can be evaluated. For example, the correlation between IRAS100 μm and PHOT170 μm intensities is computed and the regression line is drawn for each class on Figure 7(c). Each line, fitted

for each cloud of points corresponding to one segmented class, provides a reasonable assumption of the temperature of the corresponding region. Higher is the slope of the regression line, colder is the temperature in the related region.

We can also see on Figure 7(b) that the class 8 associated to excess of infrared emission, also presents the highest slope in the temperature graph. That implies that this region is cold, and then may contain molecular hydrogen. Moreover, our results are confirmed by recent CO observations of the SMC [2]. This confirms the validity of the segmentation obtained with the Markovian quad-tree.

As a conclusion, the Markovian hierarchical approach using a quad-tree offers here an effective way to identify the different compositions of regions in the SMC. In each class, the interpretation of physical properties is focused on the adequate pixels and is therefore more accurate, even if some observations are missing on "PHOT image" (i.e., spatial filtering methods are inadequate in that case).

Acknowledgments: The following figures can be viewed in color at <http://picabia.u-strasbg.fr/Isiit/perso/collet/Images/PSIP2003.htm>. The authors thank F. Boulanger for fruitful discussions and F. Bonnarel, (CDS - Strasbourg Astronomical Observatory) for astronomical image interpretation and management. This research is supported by the French government (ACI-Grid / IDHA project: Action Concertée Incitative-Globalisation des Ressources Informatiques et des Données / Images Distribuées Hétérogènes pour l'Astronomie, 2001-2003)

5. REFERENCES

- [1] C. Bot, F. Boulanger, K. Okumura, and B. Stepnik, "Multi-wavelength analysis of the dust emission in the small magellanic cloud," *Proc. of symp. "Exploiting the ISO data archive - Internet Astronomy in the Internet Age*, vol. C. Gry et al. eds., ESA SP 511, 24-27 June 2002.
- [2] N. Mizuno, M. Rubio, A. Mizuno, R. Yamaguchi, and al., "First results of a co survey of small magellanic cloud with nanten," *Publ. Astron. Soc. Japan*, vol. 53, pp. L45-L49, 25 December 2001.
- [3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721-741, November 1984.
- [4] C. Graffigne, F. Heitz, P. Pérez, F. Prêteux, M. Sigelle, and J. Zérubia, "Hierarchical Markov random field models applied to image analysis : a review," in *SPIE Neural Morphological and Stochastic Methods in Image and Signal Processing*, San Diego, 10-11 July 1995, vol. 2568, pp. 2-17.
- [5] C.A. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. on Image Processing*, vol. 3, no. 2, pp. 162-177, 1994.
- [6] Z. Kato, M. Berthod, and J. Zérubia, "A hierarchical Markov random field model and multitemperature annealing for parallel image classification," *Graphical Models and Image Processing*, vol. 58, no. 1, pp. 18-37, 1996.
- [7] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical mrf model," *IEEE Trans. on Image Processing*, vol. 9, no. 7, pp. 1-17, July 2000.
- [8] M.R. Luetzgen, W.C. Karl, A.S. Willsky, and R. Tenney, "Multiscale representation of markov random fields," *IEEE Trans. Image Process.*, vol. 41, no. 12, pp. 3377-3395, December 1993.
- [9] P. Rostaing, J.-N. Provost, and C. Collet, "Unsupervised multispectral image segmentation using generalized Gaussian noise model," in *Proc. International Workshop EMMCVPR'99 : Energy Minimisation Methods in Computer Vision and Pattern Recognition, Springer Verlag - Lecture Notes in Computer Science*, York, UK, July 1999, vol. 1654, pp. 141-156.

Acronyms: **ATCA** : Australian Telescope Compact Array.

Other references :

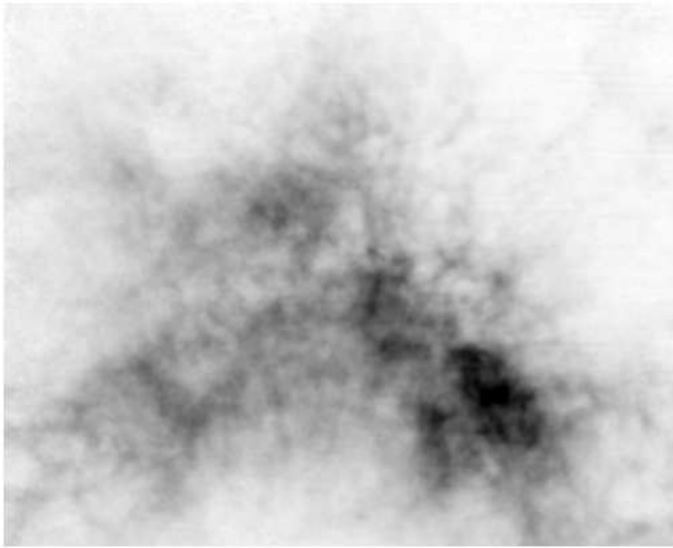
IRAS <http://space.gsfc.nasa.gov/astro/iras/intro.html>

ISO : <http://isowww.estec.esa.nl/>

Some links for more information about the SMC:

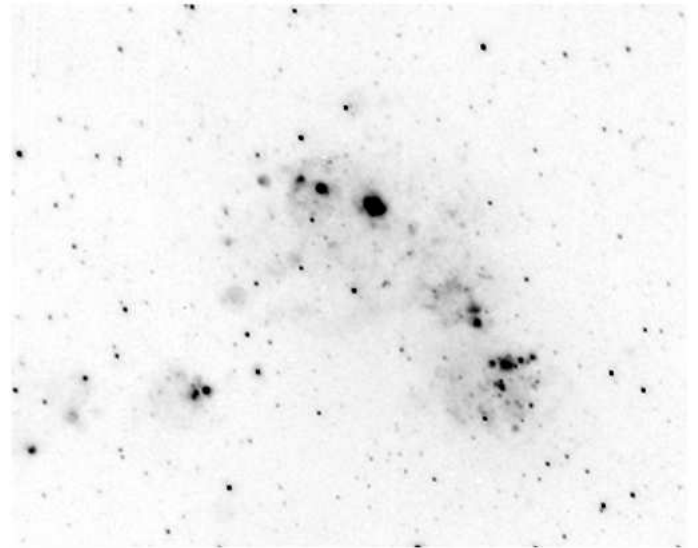
<http://www.obspm.fr/actual/communique/1998-07-24-Heydari.fi.shtml>

<http://heritage.stsci.edu/2000/30/table.html>



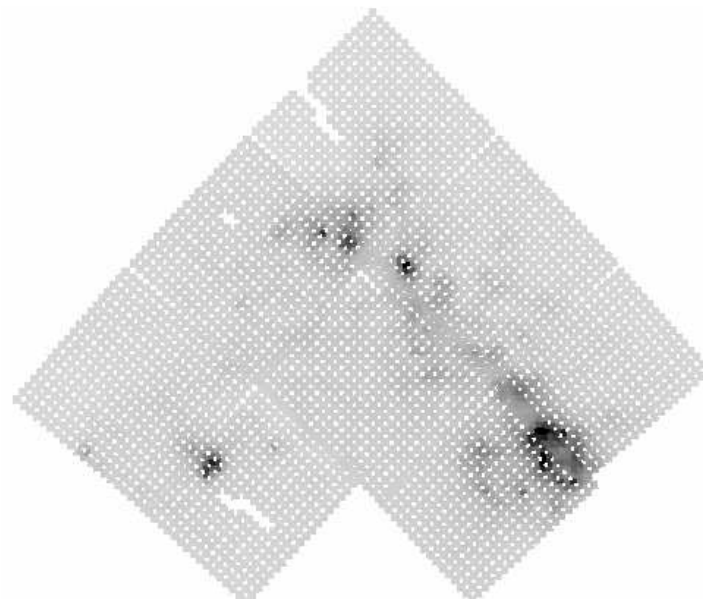
"HI image"

Fig. 2. The image have been resampled to provide a pixels grid of 30" spacing compliant to the PHOT image (Figure 3).



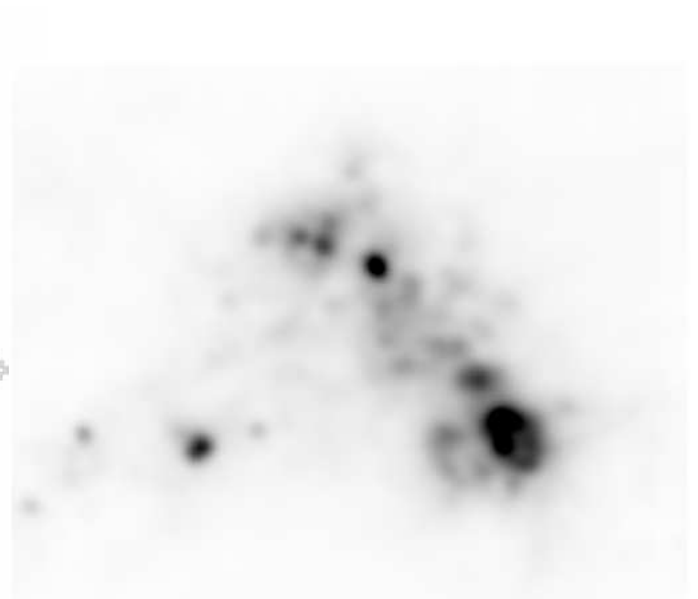
"H α image"

Fig. 4. The same region in the H α band resampled to 30" per pixel in order to match the PHOT image resolution (Figure 3).



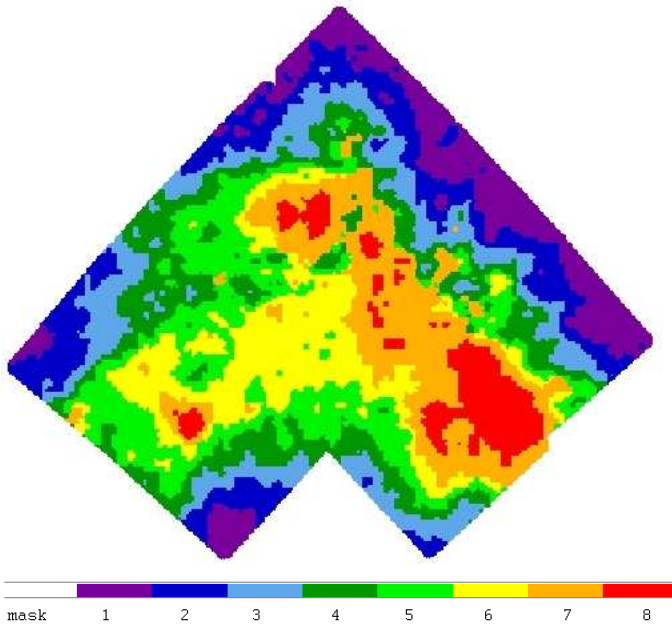
"PHOT image"

Fig. 3. On this picture, composed of a mosaic of 9 observations, the missing pixel data, due to sampling adjustment problems, appear as a regular lattice of white dots.



"IRAS100 image"

Fig. 5. The images have been resampled to provide a pixels grid of 30" spacing compliant to the PHOT image (Figure 3).



(color used for each class)

Fig 6(a)

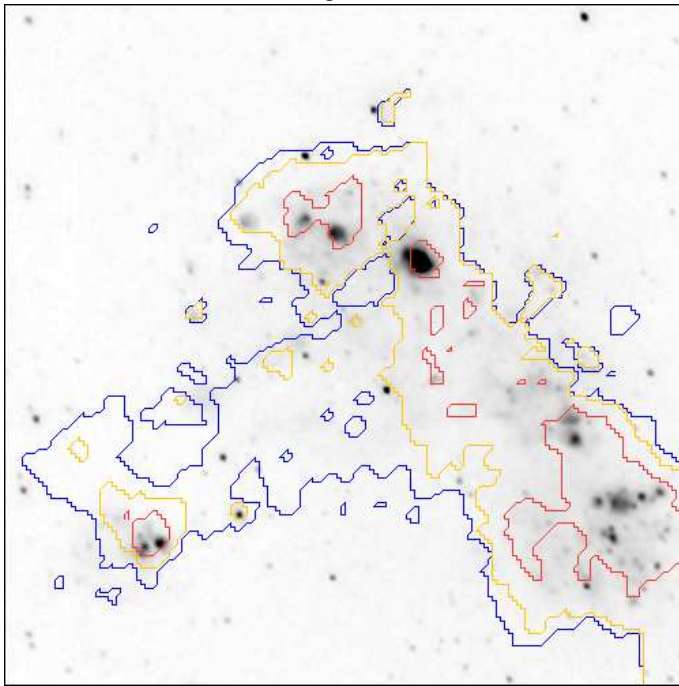


Fig 6(b)

Fig. 6. Figure (a) displays the 8-classes segmented picture using a Markovian-in-scale analysis taking into account simultaneously the HI and PHOT images. This analysis takes into account simultaneously the 2 spectral bands (HI and PHOT) for the segmentation task whereas the missing data on "PHOT image" (figure 3) are completed on the segmentation map (fig. (a)) by the label propagation thanks to the transition laws a_{ij} (Figure 1) up to the leaves of the tree. Figure (b) shows the edges of the segmentation map superimposed on the $H\alpha$ image.

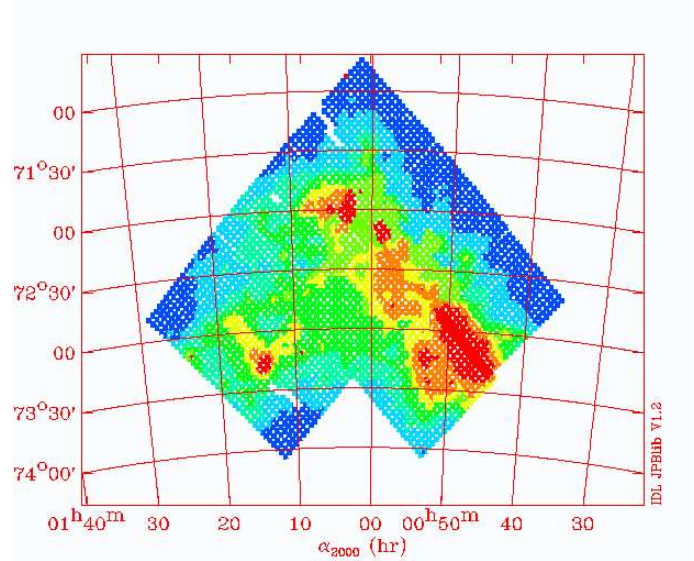


Fig 7(a) Segmented picture using a Markovian analysis taking into account simultaneously the HI image and PHOT image.

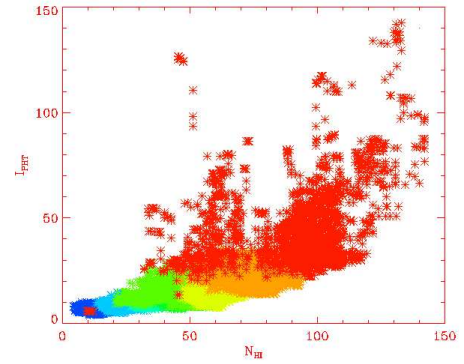


Fig 7(b) Coupled distribution by class on PHOT versus HI intensities.

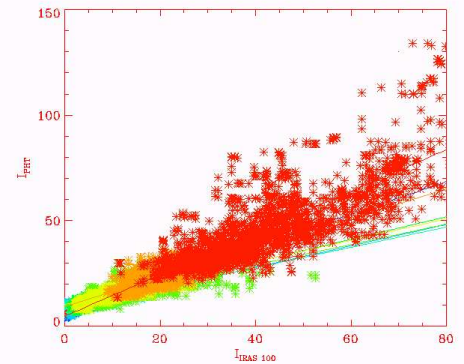


Fig 7(c) Coupled distribution by class on PHOT versus IRAS100 intensities

Fig. 7. Class 8 associated to excess of infrared emission in (b), is also associated to a higher slope in the temperature graph (c), suggesting that this region is cold, and probably contains molecular hydrogen.