Astronomical Data Analysis Software and Systems XIX
ASP Conference Series, Vol. 434, © 2010
Y. Mizumoto, K.-I. Morita, and M. Ohishi, eds.

A Boosting Approach for the Detection of Faint Compact Sources in Wide Field Aperture Synthesis Radio Images

Albert Torrent, Marta Peracaula, Xavier Lladó, and Jordi Freixenet

Institut d'Informàtica i Aplicacions, Universitat de Girona, Spain

Juan Ramón Sánchez-Sutil

Departamento de Física, Universidad de Jaén, Spain

Josep M. Paredes

Departament d'Astronomia i Meteorologia, Facultat de Física, Universitat de Barcelona, Spain

Josep Martí

Departamento de Física, Universidad de Jaén, Spain

Abstract. Several thresholding techniques have been proposed so far in order to perform faint compact source detection in wide field interferometric radio images. Due to their low intensity/noise ratio, some objects can be easily missed by these automatic detection methods. In this paper we present a novel approach to overcome this problem. Our proposal is based on using local features extracted from a bank of filters. These features provide a description of different types of faint source structures. Our approach performs an initial training step in order to automatically learn and select the most salient features, which are then used in a Boosting classifier to perform the detection. The validity of our method is demonstrated using 19 images that compose a $2.5^\circ \times 2.5^\circ$ radio mosaic, obtained with the Giant Metrewave Radio Telescope, centered on the MGRO J2019+37 peak of gamma emission at the Cygnus region. A comparison with two previously published radio catalogues of this region (task SAD of AIPS and SExtractor) is also provided.

1 Introduction

Recent wide field radiointerferometric surveys show a large amount of faint compact objects with intensities very near to noise levels. On top of that, these images are usually very complex with the presence of diffused extended emission and interferometric patterns. Therefore, automatic detection methods working at low signal-to-noise ratios become necessary in order to create reliable catalogues of the compact sources. In previously presented works we have explored methods that use wavelet decomposition (Peracaula et al. 2008) and contrast radial function (Peracaula et al. 2009) to detect faint compact sources. Here we enlarge the set of possible methods presenting a boosting approach.

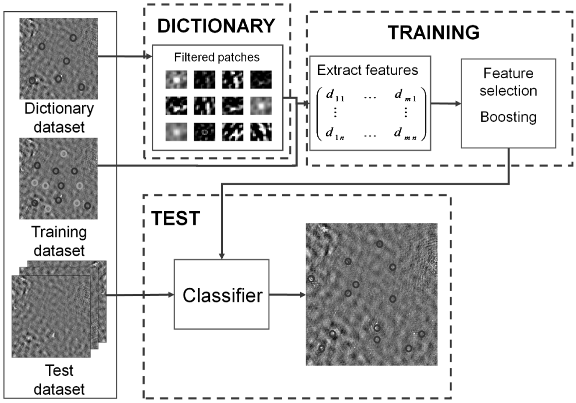

Figure 1. Graphical representation of our proposal: 1) building the dictionary, 2) training process, 3) testing process with new images.

2 Our Boosting Classification System

Our approach for the detection of faint compact sources is based on a work of Murphy et al. 2005 for object detection using local features and using a boosting classifier. The proposed approach is divided in three parts as is shown in Figure 1. First we build a feature dictionary which is composed of patches of faint sources. Afterwards, dictionary words are used to extract features and the system is trained using a boosting algorithm. For both dictionary and training image sets ground truth is available. Finally, the system is tested with new images. In the following sections we will describe in more detail the important aspects of these three parts of our approach.

2.1 Building the Dictionary

The first task of our system consists in building a feature dictionary. The definition of this dictionary contains the visual words (patches) that will be then used to extract features for training and testing. First of all, we randomly select a set of wide field aperture synthesis radio images to create the dictionary. Afterwards, these images are convolved with a bank of filters (see Murphy et al for more details). All the filtered images are then used to extract different patches centered on faint sources. Figure 2 shows the building vocabulary process. These patches become the words of our dictionary. Notice that as well as the patch, the filter used is also needed to extract the image features, since each patch is convolved with the corresponding filtered image.

When the dictionary has been built, the pixels of an image can be characterized using the following equation:

$$v = (I * f) \otimes p \tag{1}$$

where v are the characterized image pixels, I the original image, f the filter, and p the filtered patch. Therefore, the image is convolved ($*$) with the filter and then a cross correlation ($\otimes$) with the patch is performed. As a result we obtain a probability image with high values on the regions similar to the patch.

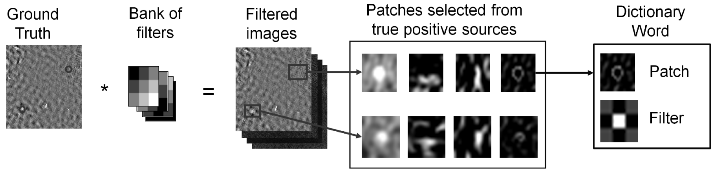

Figure 2. Vocabulary building process.

2.2 Training and Testing Processes

The goal of the training process is to learn which features are the best in order to detect compact faint sources. Our training data is created as follows. Instead of using all the pixels of the training images, which will be computational expensive, we select some points of each training image. In particular, we select the center of the faint sources (positive training examples) and some random locations of the background (negative training examples). Afterwards, we apply a boosting algorithm to perform the training. Boosting algorithms are based on the simple idea that the sum of weak classifiers can produce a strong classifier (Friedman et al. 2000). In our case the weak classifiers (h_t) are simple regression stumps with one of the features, so at each round the feature with less error is selected. The weak classifier function used is shown in Equation 2,

$$h(x) = a(x > th) + b \tag{2}$$

where th is a threshold that determines if a pixel belongs to the object class or not, and a and b are parameters selected to minimize the error of the classifier.

Our final faint sources classifier ($H(x)$) is the sign of the result of the sum of weak classifiers. Once the classifier is built, we can move to the testing process where the classifier is applied to new images in order to perform the detection of compact faint sources. Observe also that we build a pixel-based classifier and therefore we provide as a result a probability image where high values mean more confidence to be a faint source.

3 Experimental Results and Conclusions

In order to validate the performance of our approach we used the 19 deep radio images obtained by Paredes et al. 2009 at 610 MHz (49 cm) to survey a $2.5^\circ \times 2.5^\circ$ region centered on the MGRO J2019+37 peak of high energy gamma-rays emission. Each image covers a $28'$ radius circular region and all of them are partly overlapped in a hexagonal pattern to compose the final mosaic. These images are an ideal bench testing set for automated detection methods since: 1) they show a significant amount of detail due to its high spatial dynamic range (over 1000); 2) they have a remarkable population of compact sources (i.e. star-like objects) and show extended diffuse emission; and 3) they also show unwanted interferometric pattern mainly caused by deconvolution artifacts and grating rings from strong sources both inside and outside the primary beam.

The results presented here have been obtained using 1 field to generate de dictionary and 6 different fields to train the boosting classifier. Afterwards, we tested the 19 images with our approach, comparing our results with two different catalogues. The first one was produced by Paredes et al. 2008 (private communication) using the task SAD of AIPS. While the second one was obtained using the well-known SExtractor and published by Paredes et al. 2009. For a better comparison with these two previously existing catalogues we excluded exactly the same image zones than those of the respective papers. These zones contain a great amount of artifacts near to the bright sources.

Table 1 shows the quantitative analysis in terms of number of True Positive (TP) and False Positive (FP) detections obtained for the 19 images and using all the approaches. Observing these results one can see that our boosting approach succeeds in reducing the number of FP with respect to SAD, while the number of TP detections is very similar, and even better than the ones reported by the other methods. Notice also that FP were not available for the SExtractor approach. We want to stress that some of these TP were not coincident in position. Our approach found 368 TP coincident with the results reported by SAD (80%) and 402 with respect to the results reported by SExtractor (85%). While 321 TP were coincident in all the approaches.

Approach	TP Detections	FP Detections
Boosting	505	96
SAD	455	474
SExtractor	473	N/A

Table 1. Quantitative analysis, comparing our approach with the SAD and SExtractor methods.

In summary, we want to emphasize the simplicity of implementing our boosting approach as well as the satisfactory obtained results which demonstrate the validity of our proposal for the detection of compact faint sources.

Acknowledgments. This work has been supported by MEC grants AYA2007-68034-C03-(01/02/03) and DPI2007-66796-C03-02. Albert Torrent holds a FPU grant AP2007-01934.

References

Murphy, K., Torralba, A., Eaton, D., & Freeman, W.T. 2005, in Towards Category-Level Object Recognition, ed. J. Ponce, M. Hebert, C. Schmid & A. Zisserman, 393.

Friedman, J., Hastie, T. & Tibshirani, R. 2000, The Annals of Statistics, ed. H. Künsch & J. O. Berger, 28, 337

Paredes, J.M. et al. 2009 A&A, 507, 541

Peracaula, M., Freixenet, J., Martí, J., Martí, J., Paredes, J.M. 2008, in ASP Conf. Ser. 394, ADASS XVII, ed. Robert W. Argyle, Peter S. Bunclark, & James R. Lewis (San Francisco: ASP), 547

Peracaula, M., Freixenet, J., Lladó, X., Martí, J., Paredes, J.M. 2009, in ASP Conf. Ser. 411, ADASS XVIII, ed. D.A. Bohlender, D. Durand, & P. Dowler (San Francisco: ASP), 255