A Boosting approach for the detection of faint compact sources in wide field aperture synthesis radio images

VICOROB
http://vicorob.udg.edu

Albert Torrent [1], Marta Peracaula[1], Xavier Lladó[1], Jordi Freixenet[1], Juan Ramón Sánchez-Sutil[2], Josep M. Paredes[3] and Josep Martí[2]
[1]Universitat de Girona, [2]Universidad de Jaen, [3]Universitat de Barcelona
{atorrent,martapb,llado,jordif}@eia.udg.edu, {jrssutil,jmarti}@ujaen.es, jmparedes@ub.edu

Universidad de Jaén

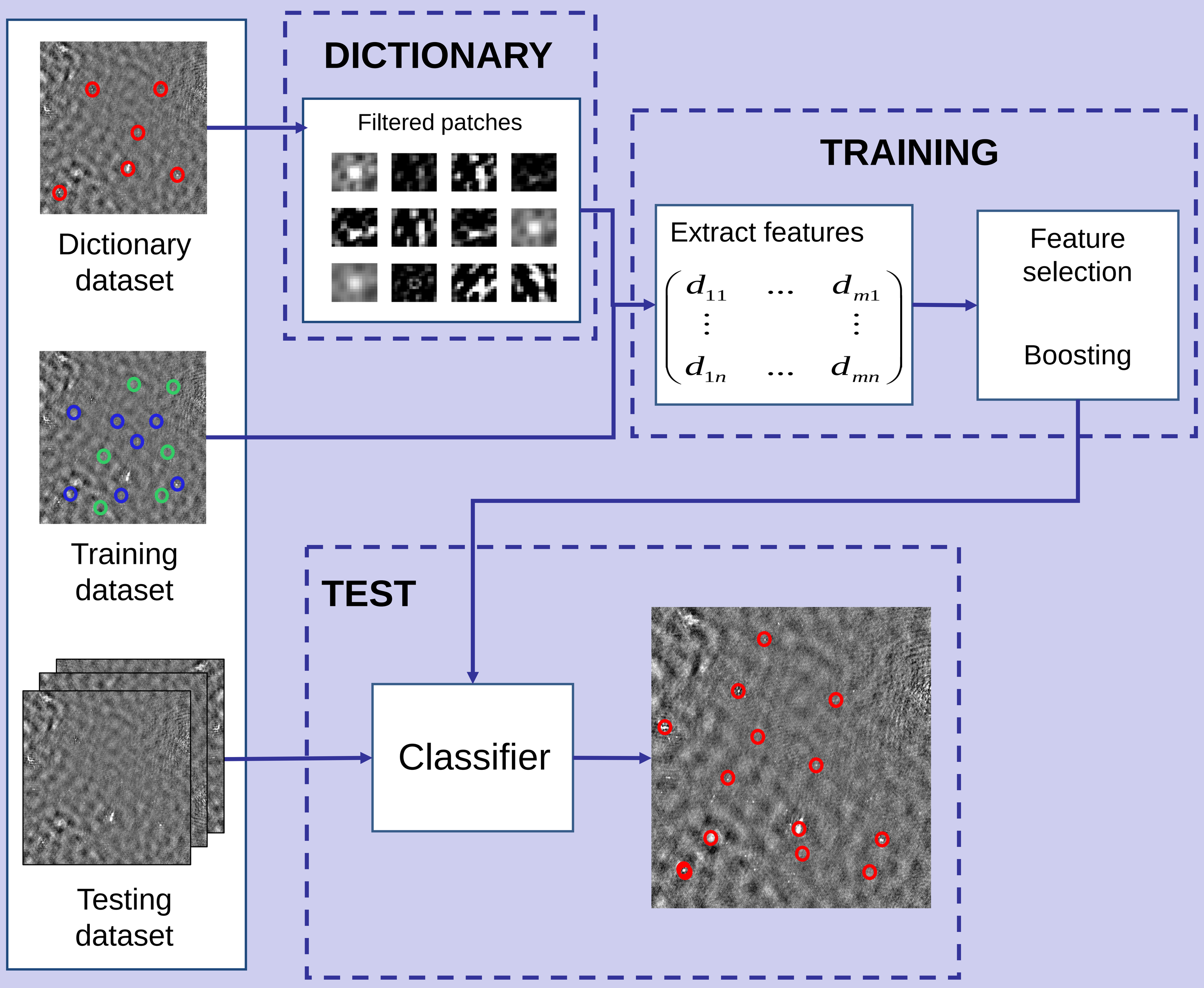
Abstract

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 deg x 2.5 deg 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.

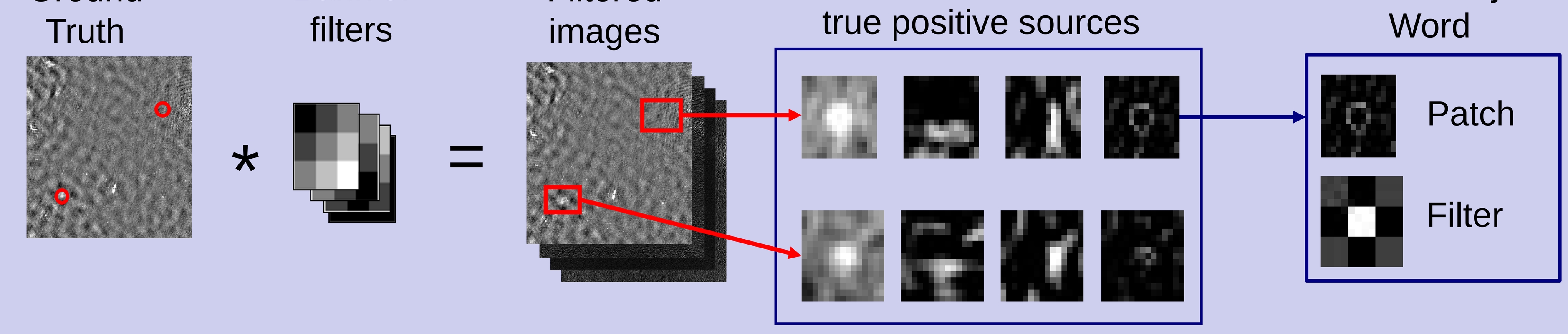
1. Motivation

- Recent wide field radiointerferometric surveys show a large amount of faint compact objects with intensities very near to noise levels.
- The images are usually very complex with the presence of diffused extended emission and interferometric patterns.
- Automated detection methods working at low signal-to-noise ratios are necessary in order to create reliable catalogues of the compact sources.

2. Boosting detection

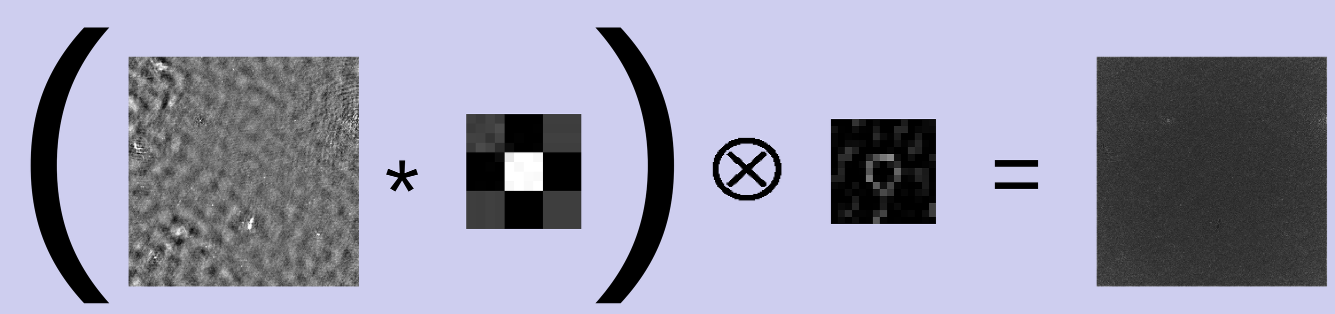


1. **Dictionary:** We generate a dictionary of words which defines what a source is. A word is composed by a filtered patch of a source and the filter used.

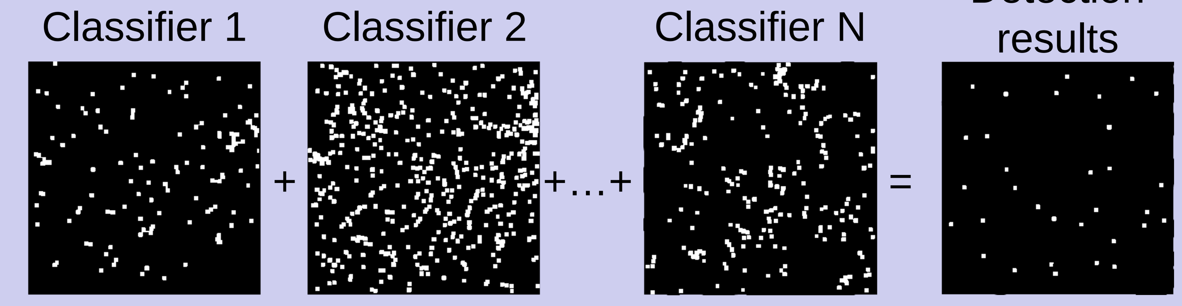


2. **Training:** We characterize a set of training images using the normalized cross correlation of the filtered images with the dictionary words. Some positive points (sources) and negative points (non sources) are selected to perform the training by using a Boosting classifier.

Feature = (image $*$ filter) $\otimes$ patch

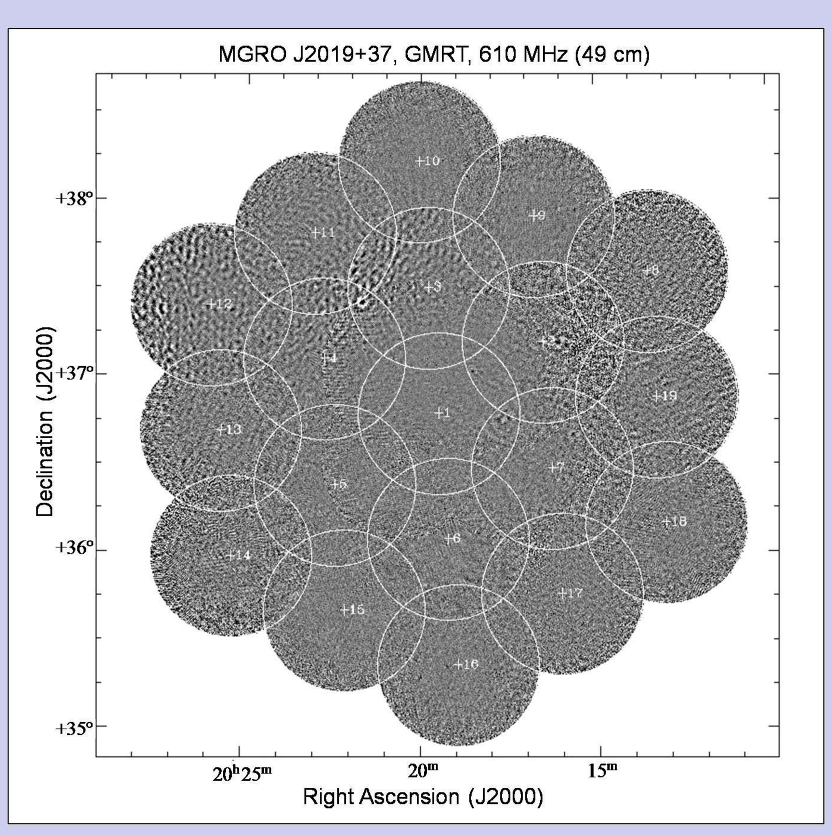


3. **Testing:** We combine all the weak classifiers obtained from the Boosting training to perform the source detection.



3. Application to a set of images

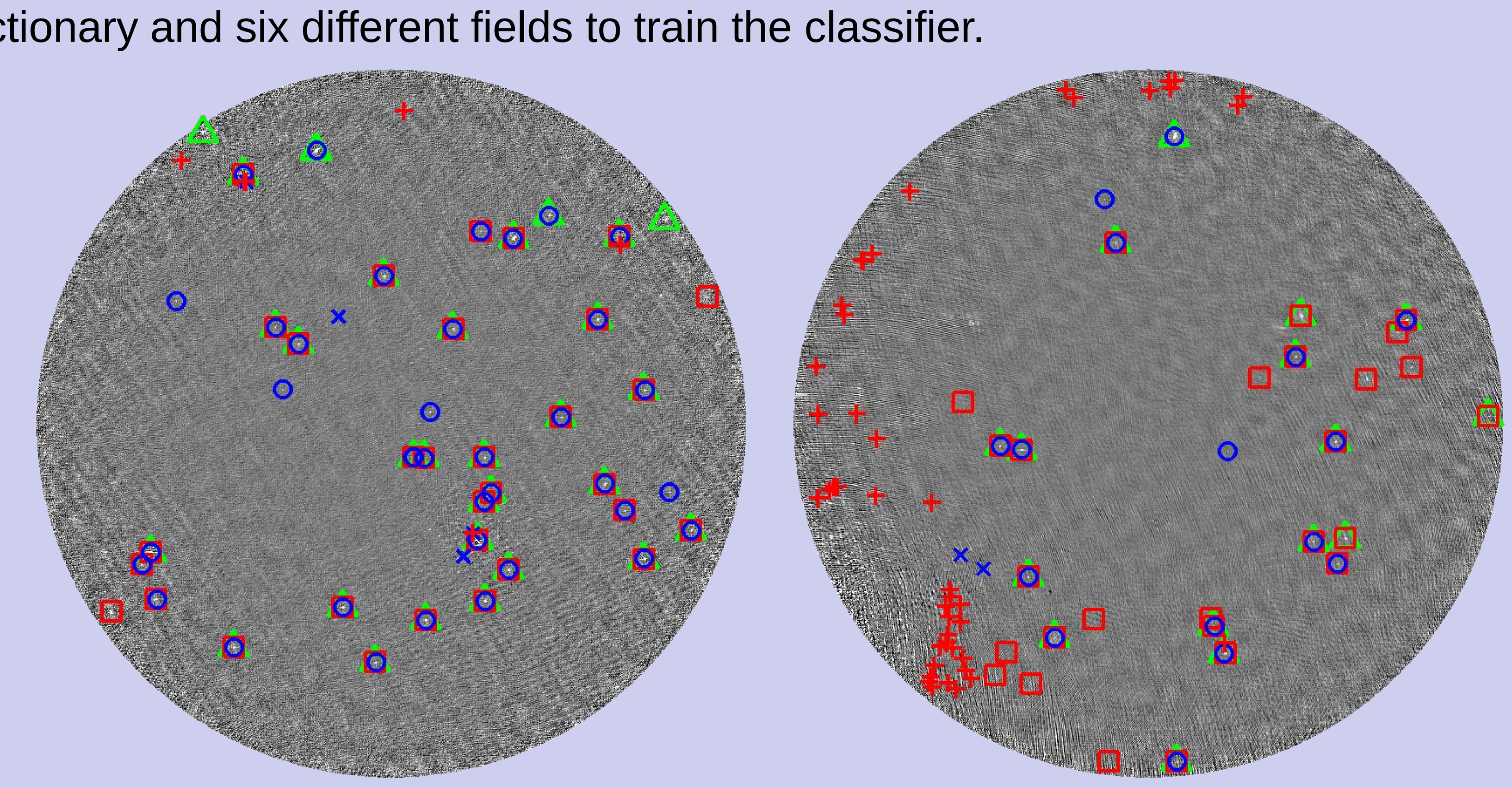
To illustrate our method we use the 19 deep radio images obtained by Paredes et al. (A&A, accepted July 2009) at 610 MHz (49 cm) to survey a 2.5°x2.5° region centered on the MGRO J2019+37 peak of high energy gamma-rays emission. Each of the images covers a 28' radius circular region and they are partly overlapped in a hexagonal pattern to compose the final mosaic.



These images are ideal for testing automated detection methods because:
- They show a significant amount of detail due to its high spatial dynamic range (over 1000).
- They have a remarkable population of compact sources (i.e. star-like objects) and show extended diffuse emission.
- They show unwanted interferometric pattern mainly caused by deconvolution artifacts and grating rings from strong sources both inside and outside the primary beam.

4. Experimental results

The results presented have been obtained using one field to generate the dictionary and six different fields to train the classifier.



Results from field 14 — Results from field 15

We compare our results with two catalogues: the first one was produced by Paredes et al. 2008 (private communication) using SAD task of AIPS. The second one was obtained using SExtractor published by Paredes et al. (A&A, accepted July 2009).
For a better comparison of our method with these two catalogues we used the same exclusion zones. These zones generally contain a great amount of artifacts near the bright sources.
The following table shows the symbol code used in the figures and the number of True Positive (TP) and False Positive (FP) detections found in each work:

	TP detections	FP detections
Boosting	○ 505	x 96
SAD	□ 455	+ 474
SExtractor	∆ 473	Not Available

Classification of true and false positive has been performed by close visual inspection by the author of each work.

5. Conclusions

Our Boosting approach succeeds in reducing the number of FP with respect to SAD (information not available for SExtractor), while providing a similar number of TP. However, some of these TP are not coincident in position:
- Our approach finds 368 TP which coincide with the results reported by SAD (80%) and 402 TP with respect to those results reported by SExtractor (85%).
- 321 TP are coincident in all three methods (Boosting, SAD and SExtractor).

In summary, the results reported in this work show the validity of our Boosting approach to perform the automated detection of faint objects.

Acknowledgements: The authors acknowledge support by DGI of the Spanish Ministerio de Educación y Ciencia (MEC) under grants AYA2007-68034-C03-01, AYA2007-68034-C03-02 and AYA2007-68034-C03-03, as well as partial support by the European Regional Development Fund (ERDF/ FEDER). GMRT is run by the National Centre for Radio Astrophysics of the Tata Institute of Fundamental Research.