

Work Package 5

Solar Feature Catalogue; Automatic Image Classification Tools

WP5.1 Survey of Image Processing Techniques

This activity is aimed surveying all the technologies already in use for automatic image classification, and those potentially useful in this project. Internal report WP1-IR1 will provide input regarding the type of features that could be recognized.

The algorithms used to create the feature catalogue should work on various features, seen at different wavelengths. The appearance of features varies with wavelength - those seen in optical observations are often more distinct, though more complex in shape than things seen in X-rays. Thus, some wavelengths will be easier to accommodate than others. In some cases, information from one wavelength can be used as the basis for the search in a different wavelength - this should increase the overall reliability of the algorithms.

WP5.2 Experimentation and Preliminary Design

This activity will evaluate technologies relevant for the project, and perform a preliminary selection on the basis of the preliminary science requirements document (WP1-IR1). When appropriate, a limited experimentation of technologies of interest can be performed to gain more experience. This activity will produce the deliverable WP5-D1, which will summarize the work of this and the previous activity ("Survey of Image Processing Techniques")

WP5.3 Detailed Design and Definition of Evaluation Criteria

Based on WP5-D1 and WP1-D2, this activity will include the detailed design and implementation of the automatic image classification algorithms and their integration in the content-based image retrieval mechanisms that will be available in the final system. This feature recognition will be greatly aided by having multiple datasets combined in the Grid. By comparing images from multiple datasets, additional, multi-wavelength characteristics can be used in the classification algorithms that will increase their accuracy. The activity should also take account of the overall system architectural design included in WP1-IR3.

The final selection of technologies to be deployed or to be developed in the project will be made at the start of the activity. The work will also finalize the definition of the features that will be considered for content-based image retrieval mechanisms, and the techniques to be used to implement them. This activity will produce internal report WP5-IR2.

Definition of performance evaluation criteria

Automatic image classification cannot achieve 100% accuracy, because of the problems intrinsic to automatic recognition of complex patterns. This activity will define a procedure to evaluate the algorithms implemented, in terms of classification accuracy, and it will include:

- Exact definition of what accuracy means in this context.
- Selection of the data sets suitable for content-based search (for example, spectra could not be in this category).
- Selection of features that the system will be able to recognize.
- Selection of a set of images, stored in the archives managed by partners of the consortium, which will be used for benchmarking the automatic classification methods.
- Definition of the procedures for performance evaluation and to compute figures of merit for the
- automatic classification methods.

The results of this activity will be included in internal report WP5-IR2.

WP5.4 Implement Image Processing Techniques

In this activity the code required for the recognition of features at different wavelengths will be written. At optical wavelengths, this will draw on experience from the Observatory of Paris-Meudon (partner no. 5), but will provide an implementation that is much more rigorous than has previously been achieved. Since the features seen at one wavelength often have counterparts at other wavelengths, the ability to use features found at one wavelength as a starting location at the other wavelength will be

developed.

This activity runs in parallel to the evaluation package. Because of the difficulty in recognizing some features, an iterative approach will be required, with the implementation and evaluation being undertaken by different groups.

When finalized, the tools will be used to produce the feature catalogue (WP5-D3), but could also be used on a stand alone basis on other images. For example, the tools might be able to identify small X-ray bright points in an image, but these would not be included in the catalogue because of their number. However, their location could be of interest to researchers in certain areas of solar physics who would be able to use the tools on an image by image basis when comparing data at several wavelengths.

WP5.5 Evaluate Performance of Image Processing Techniques

Based on WP5-IR2, this activity will apply the performance evaluation procedure to image classification mechanisms developed in a parallel activity in WP5. The evaluation runs in parallel with implementation, because the results of the evaluation may be used to improve the algorithms or to tune the parameters of the implementation. The evaluation of the final version of the content-based search mechanisms will be reported in WP5-IR4.

WP5.6 Version 0 of the Feature Catalogue Format

The Solar Feature Catalogue (SFC) is a new entry point into the solar data. Tools to interrogate the SFC will be developed within WP3, and it is essential to provide a preliminary version of the SFC to this workpackage so that the tools can be developed – this is the purpose of this task. The design of the feature catalogue is deliverable WP5-IR3.

WP5.7 Create prototype SFC using Published Lists

A number of lists of solar features have been published in the literature. Although some lists are exhaustive using data that are continuous (e.g. the list of CMEs observed by SOHO-LASCO in Gopalswamy et al., 2000), others are less rigorous. The purpose of WP5 is to redress this problem, but a prototype of the Solar Feature Catalogue will be created from such lists (WP5-D2) to allow more complete testing of the search tools of WP3, in advance of the catalogue created using the image recognition software (WP5.8).

WP5.8 Create SFC from Identified Images

Once confidence in the functionality of the image classification routines has been established for some types of images, the task of creating the Solar Feature Catalogue will commence. Using the format defined in document WP5-IR3, the different types of images will be treated in turn, initially producing preliminary catalogues covering a limited time interval. These catalogues will be available for users (both inside and outside the consortium) to use and comment on. The full catalogue will be created following the inclusion of feedback from users, and will result in deliverable WP5-D3. This will be integrated with other parts of the project within WP1.