Heliophysics Integrated Observatory

Project No.: 238969
Call: FP7-INFRA-2008-2

Solar Image Processing

Title:	**Document Title**
Document No.:	HELIO-xxx
Date:	31/01/10
Editor:	**David Pérez-Suárez**, TCD
Contributors:	Paul A. Higgins (TCD), Nicolas Fuller (OP)
Distribution:	Project

Revision History

Version	Date	Released by	Detail
Draft	2009-11-31	D. Pérez-Suárez	Draft – overview of the codes available
Draft	2010-01-31	D. Pérez-Suárez	Draft – extended description of the codes, some figures added, and radio and filaments contributions from N. Fuller

Note: This document will continue to undergo revisions during the implementation phase of HELIO to incorporate changes and improvements.

This page intentionally left blank.

Introduction

The volume of heliospheric data obtained today is much larger than what researchers have time to study. The automatic recognition of features and events is needed for reduction to concise products, which researchers can then focus on. This is HELIO's main objective: to catalogue events and associate them when they have the same origin. HELIO will also allow the user to run the algorithms which feed the catalogues on their own set of images, allowing them to change the parameters of the detection.

This feature detection report includes the description of some candidate algorithms to be included in the HELIO work-flow. However, the nature of this document is to be a living description of the techniques and algorithms used in HELIO, so new methods, algorithms and features will be added in revised versions. As of now, only solar feature detection is covered. The addition of heliospheric features is very complex and requires a process of selection that it will be done during the next months.

The document is structured in 3 sections; the first one covers image processing (future editions will be refer to this more generally as data processing) in which a brief description of the techniques used by the automated detection algorithms is presented. The next section describes the algorithms that run on solar images to detect features, such as coronal holes, filaments, active regions, coronal mass ejections and radio bursts. Future versions of this document will include the last section which will talk about the detection of features in other kinds of heliospheric data.

Image Processing: Basics

The amount of data expected from coming missions (SDO) is overwhelming, so the use of automated processing techniques is a must. In the following subsections, image pre-processing and some of the most common techniques for feature recognition are outlined. However, for a more insightful view of the methods used, there are many books which specialize on this subject.

Image Pre-processing

The procedures presented here normally use calibrated data unless near real-time data is required and calibration is not possible, or standard pre-processing the image is modified in such a way that it is not appropriate for the detection of the desired feature. The calibration is instrument dependant, but in general all digital images follow some similar pre-processing steps (such as flat-field correction). Most of the images used now a days are taken with CCDs (Charged Coupled Devices, Tompsett et al., 1970) embedded on a optical system.

Other pre-processing methods are performed to facilitate the extraction or characterization of the feature. An example of processing for feature extraction is "histogram equalization" which consists of increasing the global contrast, spreading out the most frequent intensity values along the histogram. An example facilitating characterization is an image transformation which goes from Cartesian to polar or to Lambert projection (equal area per pixel).

Image segmentation

Normally, every image obtained from a CCD contains different regions or objects (real or produced by the instrument or the detector, e.g. electronic noise). Solar images present many different features which are interesting for solar physicists and may lead to a better understanding of how the Sun works. Different digital image processing techniques allow to recognize and characterize such features in an image. Image segmentation is one of the most popular yet difficult tasks in image processing. Segmentation subdivides an image into its constituent regions or objects.

Segmentation is mainly based on image intensity: where it is discontinuous and where it is similar. To detect discontinuities, we differentiate those regions where there is a abrupt change in intensity, such as edges. To detect regions of similar intensity we must define a criterion. Some techniques used to evaluate the criterion are thresholding, region growing, region splitting and region merging. Below we discuss the techniques used in the algorithms under review.

Thresholding

Thresholding is the primary technique for segmentation and is used by all of the other segmentation techniques. It first requires the selection of an intensity value, *T*. The image is then segmented in two groups, one below and another above the value *T*. In other words, for any point (x,y) in the image (f(x,y)) we obtain a segmented image g(x,y) which is given by:

$$g(x,y) = \begin{matrix} 1 \, if \; f(x,y) > T \\ 0 \, if \; f(x,y) \leqslant T \end{matrix}$$

This process is called global thresholding when *T* is a constant applicable over the entire image. In practice, this method is not very useful if there has not been some pre-processing done before to remove irregularities on the image as noise, variation on background intensity, etc. If instead of just one value *T* we are discerning of two or more that is known as multiple thresholding. In case *T* changes over an image, we use the term variable thresholding. When *T* depends on the properties of the pixel in the neighbourhood it is known as local or regional, but if it instead depends on the spatial coordinates it is called dynamic or adaptive.

The difficulty of this method is in the selection of *T*, which depends on the problem to be solved. In the techniques reviewed, some use a static threshold for all of the images §\ref{AR.TCD}, others use a global threshold, such as 10% of the median of each image §\ref{CH.HH} or a global threshold based on local minima of an image partitioned in different sizes §\ref{CH.TCD}.

Region-based segmentation

Region-based segmentation relies on the spatial characteristics of different features. Different techniques fit in this category of segmentation; some used here are region growing and region splitting and merging.

Region growing is a procedure that groups pixels or sub-regions into larger regions based on predefined criteria for growth. Starting with a set of "seed" points, regions are grown

by appending neighbouring pixels using the criteria.
Criteria such as intensity values are local in nature and do not take into account the "history" of region growth. Additional criteria that increase the power of a region-growing algorithm utilize the concept of size, likeness between candidate pixels and the ones grouped already, and the shape of the region being grown. The use of these criteria is based on the assumption that a model of expected results is at least partially available. Figure 1 shows with a 8-connected and absolute difference of the intensities between the seed and pixel < *T*.

Figure 1. Active Region NOAA 10486 seen by MDI and segmented using region growing technique.

Region splitting and merging is a partitioning process where the image is successively subdivide into smaller and smaller quadrant regions so that, for any new region a set of properties are searched. If the properties are not found the quadrant is subdivide into subquadrants, and so on. It is common that the final partition contains adjacent regions with identical properties. This is solved by allowing merging when the properties of the union of both have the requirements setted. Figure shows an EIT image segmented with this technique which the common property of values subquadrants larger than 2 sigma.

Figure 2. EIT image segmented with region splitting an merging technique where the standard deviation is greater than 2 sigma.

Edge-based segmentation

Edge-based segmentation is based on finding those locations were there is an abrupt change in intensity. Local changes can be detected using derivatives, which are easily calculated with local differences. Similarly, second derivatives (also called Laplacian) can be calculated. Local differences are calculated using a 3x3 running kernel. There are numerous kernels, each one with its advantages and disadvantages, some offering greater sensibility to certain directions than others. Some examples are: Roberts, Sobel, and Prewit. More advanced techniques, such as Canny or taking the Laplacian of a Gaussian, are techniques for extracting regions which take into account factors such as image noise and the nature of edges themselves.

Segmentation using Morphological Watershed

The concept of watersheds is based on visualizing an image in three dimensions. Any local minimum can be thought of as the point to which water falling on the surrounding region drains. The boundaries of watersheds lie on the tops of ridges. The watershed operator labels each watershed region with an unique index and sets the boundaries to zero.

Fuzzy clustering segmentation

Fuzzy logic and fuzzy set theory \citep{fuzzy 65} is able to simultaneously represent and manage imprecise and uncertain knowledge. In fuzzy set theory, a fuzzy measure is a representation of uncertainty, giving for each event to be analysed a coefficient in [0,1] assessing the degree of certitude for this event. So, strictly speaking, fuzzy clustering is not a segmentation theory. But once we have the coefficient for each pixel in each event,

the segmentation is just a simple task of choosing the pixels with the largest coefficients for each category.

Application to Solar images

HELIO will provide two frameworks, one will run a set of features detection algorithms to feed a knowledge database and the other will make the algorithms available to the community for incorporation in work-flows, allowing parameters within the algorithms to be altered, such as thresholds, features sizes, etc.

In the following section, the algorithms selected for HELIO are described. In future versions of this report, new methods will be added together with modifications that will be made on the original codes in order to adapt them to HELIO. There will also be methods related to the detection of features in the heliosphere.

Coronal holes

Coronal holes (CHs) are low density regions in the solar corona that have less ultraviolet and X-ray emission than quiet-sun and ARs. The magnetic field within the regions characterized are mostly unipolar (open magnetic field lines) that extend beyond the corona into the interplanetary medium and give rise to high-speed solar wind streams (Altschuler et al., 1972). Due to their properties, they are an important factor in producing recurring magnetic disturbances at Earth on time scales of days to months.

CHs can be observed in EUV and X-Rays from rocket telescopes or spacecraft as well as in He I 10830 chormospheric line from the ground-base telescopes. Their appearance is dark on the corona whereas, these regions, appear bright on the chromosphere. Their uni-polarity nature usually have helped to confirm their detection through the use of magnetograms.

The possibility of ground-base observations has made possible to manually catalogued the CHs since 1974 from Kitt Peak Vacuum Telescope. Trying to automatize this task various teams have become with different approaches depending of the instruments used. Most of them are based on threshold segmentation, however it is not clear which criteria is used by some of them for the threshold selection. Four algorithms are shown below, the first uses 195 A Extreme ultraviolet images from SoHO, whereas the second detect CHs from ground based observations. The last two uses more than one dataset for the segmentation.

Krista and Gallagher (2009) method provides a CH identification which it also compares the CH properties with in-situ measurements of the solar wind at ~1AU. Their algorithm can be used as a space weather forecast tool to predict the fast solar wind streams with different propagation models (Parker, 1958).

The CH identification aspect of their code has been tested with 195 A images from the *Extreme Utraviolet Imager Telescope* (EIT/SoHO), the *Extreme Ultraviolet Imager* on the *Solar Terrestrial Relations Observatory* (EUVI-SECCHI/STEREO) and on full-disk Ti-poly images from the *X-Ray telescope* (XRT/Hinode), with a good agreement with the high-speed solar wind in-situ measurements from the *Advanced Composition Explorer* (ACE) and *Plasma and Suprathermal ion Composition instrument* (PLASTIC/STEREO).

The threshold detection is the key point of this method. Intensity histograms of a solar EUV image gives a multimodal distribution where each frequency distribution

corresponds to a feature in the Sun, i.e. low intensity regions (LIRs), quiet sun (QS) and ARs. The CH boundary corresponds to the location of a local minimum between the LIRs and QS distribution. As Figure 3 shows, that minimum can be enhanced using a partitioning operation, and by doing so the solar image and local histograms obtained have a more defined minima which aids in the determination of the threshold desired. The sub-images used are 1/2, 1/3, 1/4 and 1/5 of the original Lambert equal-area projected image. Each sub-image size steps across the image and obtains local intensity histograms from overlapping sub-images. The threshold minima are defined as local minima that are located within 30-70% of the QS intensity and are wider than 6 digital numbers.

Since this method depends on the QS intensity, it is reliable at any point in the solar cycle, regardless of the change in the overall intensity. However, the CH boundaries acquired during solar maxima are less accurate due to bright coronal loops intercepting the line-of-sight and obscuring parts of the CHs boundaries.

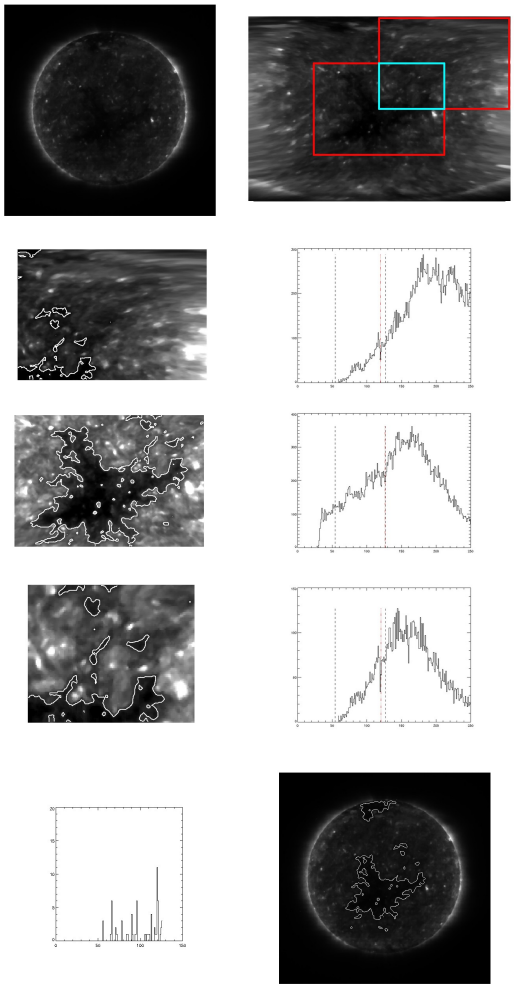


Figure 3. Visualization of the partitioned to extract the proper threshold used by Krista and Gallagher's code. Original 195 EIT/SoHO image (top-left) is transformed to an Lambert equal area projection (top-right). The image is divided in sub-images and a optimal value is obtained from the histograms of each partition. The three middle rows show the sub-images and their histograms of the frames outlined on the Lambert projection. The number of occurrence of the optimal thresholds found is shown on the bottom left, and the last image shows the contour with the most frequent threshold.

Once the EUV -or X-ray- image has been segmented into LIR and non-LIR, the next step is to determine whether they are CHs or filaments. Filaments are thin and elongated structures that have a bi-polar distribution of polarities, whereas CHs show a dominant polarity. The polarity properties of the LIRs are easily extracted from a magnetogram using the segmented mask. The skewness of the field strength distributions of the different regions provide a numeric value of how bi-polar the regions are (i.e. a filament will have a value close to 0 and a CH will be skewed towards the positive or negative side). Other methods also use polarity information of the regions detected to distinguish CHs from other LIRs, as for example Henney & Harvey algorithm.

Henney & Harvey (2005) developed an algorithm motivated by the conclusion of operations of the KPVT in 2003 and the start synoptic observations by the SOLIS *Vector SpectroMagnetograph* (VSM) helium spectroheliograms and photospheric magnetograms.

Their detection uses two-day average he I 10830 spectroheliogram and two-day average photospheric magnetogram weighted by a expression involving their time difference. A mask is created with those values above 1/10 the median of positive values. A morphological closing operation using a square kernel function as the shape operator is applied to fill the gaps and connects nearby regions. For physical reasons (Harvey and Recely, 2002) all those areas smaller than certain size (two supergranules) are removed.consecutive observations in He I and two magnetograms. The mask is then multiplied by a very large value and fill the correspondent pixels on the first segmented image, the smoothing of the resultant fills small gaps and holes. This is follow by a morphological opening operation which removes small features while preserves the size and shape of small regions obtaining the final mask of CHs candidates.

The magnetic field of each candidate region is extracted from the magnetograms and their porcentage of unipolarity examinated. Those with a value below a variable threshold are deprecated. The variable threshold relative to latitude and central meridian distance was obtained from a study of 11241 hand-drawn CHs done in KPVT from data observed between 1992 and 2003.

There has not been tested the two algorithms shown here together. However, both obtain satisfied result when compared with other sources. Krista and Gallagher have compared their fast solar wind predictions with the observed values obtaining a positive correlation between the observation speed and the CHs area. On the other hand Henney and Harvey compared their results with the hand-drawn ones obtaining just an area 3% smaller.

Scholl & Habbal (2008) developed a technique that detect and classify CHs and filaments from EUV and Magnetogram observations. They compared the results from three different EUV wavelengths (171,195 and 304 A) with EGSO filament catalogue and He I 10830 A CH database from Kitt Peak. For the purpose of comparison they run the code following the availability of the data in this order between 1997 and 2003: Kitt Peak, SFC-EGSO (filaments database), EIT and MDI.

The data preparation and processing works as explained below:

- Co-rotate MDI, Kitt Peak maps and SFC contours when needed.
- EIT: Normalization, noise cleaning, contrast enhancement (histogram equalization), image-segmentation (combination of region-based and edge-based

methods), region extraction, cleanup (remove small features). These steps differ between 304 and 171 or 195 A.

- Comparison with magnetograms: skewed distribution and mean magnetic flux.
- Classification.

They found, in general, better agreement during the declining phase of solar activity. But they find a great agreement between the different wavelenghts for features lying close to Sun center, when line-of-sight effects are minimal.

SPoCA (Barra et al., 2005, 2008, 2009), the Spatial Possibilistic Clustering Algorithm is a a multi-channel unsupervised spatially-constrained fuzzy clustering method that automatically segments solar EUV images into regions of interest, which in this case are, at least, CHs, active regions and quiet sun. However, in the way that this code works new features can be detected, as coronal bright points, and new wavelengths can be added.

The algorithm try to find through an iterative minimization equation the cluster centers of the features being detected. Then each pixel will obtain an amount of belonging to one or other feature group, as a probabilistic value, depending on itself, its closer neighbours and the whole image. The improvements done in the last version include a radial line-of-sight equalization, the inclusion of an automatic evaluation of the segmentation with a sursegmentation method (i.e. segmenting the image into a number of classes strictly superior to the intuitively expected number of classes in the image, and then finding and aggregation criterion of the resulting partition that shows the relevant classes), the smoothing of the edges using a morphological opening with a circular isotropic element of size one and finally the CHs and filaments are differentiated using an H-alpha image from a ground-based telescope. This last improvement is not good for real-time data analysis.

Comparison

	Photosphere (magnetic)	Chromosphere	Corona	Heliosphere
Krista & Gallagher	MDI/SoHO		EIT/SoHO,SECCHI/STEREO, XRT/Hinode	ACE
Henney & Harvey	SOLIS/VSM	SOLIS/VSM HE I 10830		
Scholl & Habbal	MDI/SoHO		EIT/SoHO	
SPoCa		Halpha(Kanzelhoehe)	EIT/SoHO	

Filaments

Filaments are large regions of very dense and cool plasma held in place by magnetic fields. They usually appear long, dark, and thin when seen against the solar disk, and at the limb appear as bright fuzzy arches, called prominences. As previously stated, filaments appear bipolar in magnetograms, which allows them to be differentiated from CHs. Even though filaments are observed in the corona, H-alpha (chromospheric) images provide the best outline of the feature.

H-alpha observations are made from ground-base telescopes (Hinode contained an H-alpha filter, but it suffered problems during launch and its use was not recommended). The images thus require pre-processing to correct for the constantly varying observing conditions (mostly caused by atmospheric seeing). Some of the corrections done are the same for all ground-based observations but others are instrument dependant. We refer the reader to the documentation of each algorithm for further details.

It is well established that the sudden disappearance (or eruption) of filaments from an image is usually associated with CMEs. The characterization of filaments can provide information to predict the orientation of the magnetic field associated with CMEs and the probability of a CME hitting the Earth.

Bernasconi et al. (2005) produced a very complete automated filament detection and characterization algorithm which is based on a code written by Shih and Kowalski (2003). They used full-disk H-alpha images observed from Big Bear Solar Observatory (BBSO), as the one shown on the left panel of figure 4. The filament detection is performed by creating a mask using a threshold segmentation and an advanced morphological filtering operation. The first step is to remove the sunspots (their cores are usually darker than filaments). All pixels below a certain threshold (determined manually) are marked as sunspot seeds. Using a region-growing morphological operation, the seeds are grown by adding adjacent pixels until the number of pixels exceeds a value higher than the threshold level for filaments. All the regions detected, with an area smaller than 2000 pixels, are set to a larger value than the threshold level for filaments.

A binary mask is created by setting every pixel to unity, below the filament detection threshold and the rest to zero. It requires a filtering operation to extract only those regions with elongated shapes and to remove unwanted regions. These shapes can be isolated from other by separately applying to the filament mask eight opening morphological operations with the eight linear structuring elements shown in left-top panel of figure 4 (Soille and Talbot, 2001). Pixels that survive at least two of such opening operations are used as seeds for a region-growing morphological filter to expand them until they surpasses a second filament threshold. Regions smaller than 300 pixels are deprecated.

Once the mask have been created each separate cluster is numbered and the characterization of the detection proceed. In this process the: position, length, area, average tilt of axis with respect to the Sun's equator and chirality of the magnetic flux rope are extracted. The determination of the Filament's spine is done using a multi-step iterative technique as it is shown on the bottom left panel of Figure 4. The first iteration starts by determining the location of the two spine ends points. Then by adding the middle point and through an optimization process it gets another vertex. The iteration of those steps until each segment is smaller than certain length gives an array with the coordinates of the filament's spine.

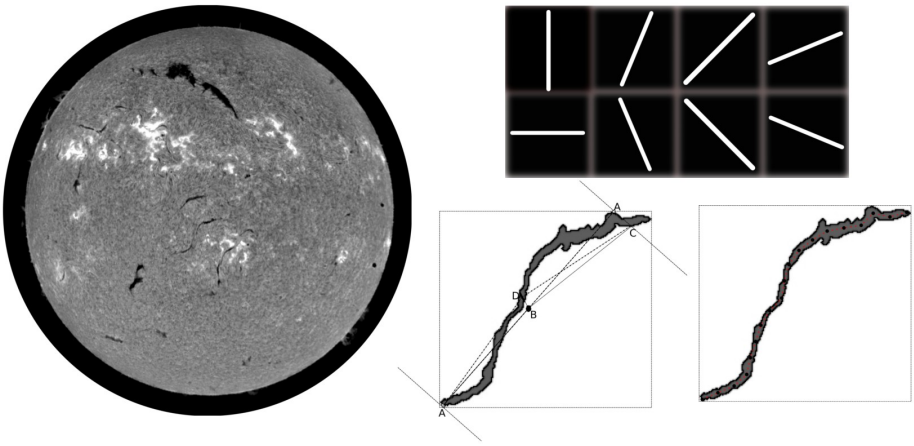


Figure 4. Full disk H-alpha image where the filaments can be seen as long dark and thin structures. On the right side; top: the eight directional linear structuring elements used by advanced morphological filter. On the bottom the first step and final result of the algorithm that determine the filament's spine. The labels refer to the order in which the points are found.

The filament's barbs is a important characteristic to take into account because it gives information of the chirality of the flux rope in which the filament is embedded. The barbs are detected as the farthest boundary points to the spine, which are longer than 3 pixels and that the ratio between the barb's height and the width in its base is greater than 0.3. Depending of the angle of each barb respective to the closest spine segment determine them as a bear-right or a bear-left direction. The difference between the number of left and right bears establish the chirality of the filament as left or right-handed.

The technique used for others automatic filament detection differ mainly on how the threshold is selected. On the algorithm used in EGSO (Fuller et al., 2005) the lower and upper threshold used to find the seeds to create a mask with the segmentation of the image is calculated according to local statistics after divide the image into smaller areas. To obtain the skeleton of the filaments this code use the thinning process based on the HitOrMiss transorm (Sonka et al., 1999), removing the branches of the skeleton after computing iteratively the end points.

N. Fuller et al. (2005) developed a code to automatically detect filaments on Halpha observations in the frame of the EGSO Solar Feature catalogue. This method is based on region growing.

Starting from a standardized solar image the following steps are applied:

1. Optional cleaning
   - Remove possible dark lines on image: the method combines global thresholding and morphological thinning to detect dust lines.
   - Flattening: remove large variation in background intensity (center to limb attenuation already removed with standardization process):

a- Apply median filtering on a smaller scale image
b- filtered image is thresholded and highest/lowest values are replaced by the median filtered image values
c- second median filtering of the image calculated above (i.e. reducing the effect of bright regions and dark filaments)
d- Filtered image is scaled back to the original size and subtracted from original image, mean intensity is restored
- Sharpen the image using High-boost filtering (subtract the second derivative of image: laplacian)$g(x,y) = f(x,y) - \tilde{N}^2 f(x,y)$

2. Find the region seeds
- A windowed threshold is applied. Statistics of pixels values are calculated within the window, discarding the highest and lowest ones to get an estimation of the quiet sun level. Then the threshold which is applied to get the seed pixels is calculated.

3. Region growing process
- A bounding rectangle is computed for each seed.
- Within the rectangle, statistics of pixels are determined
- Upon intensity criteria the pixels are appended to the region or not.
- Some limits are set like region maximum and minimum size, seed minimum size, seed distance to the limb etc.

The final result is a segmented image with pixel corresponding to filaments set to 1 and others to 0. Some more morphological operations
are then applied to group close regions and to describe the detected features (contour, skeleton, area, length etc.)

Comparison of codes

The technique used for others automatic filament detection differ mainly on how the threshold is selected. On the algorithm used in EGSO the lower and upper threshold used to find the seeds to create a mask with the segmentation of the image is calculated according to local statistics after divide the image in smaller square windows. To obtain the skeleton of the filaments this code use the thinning process based on the HitOrMiss transform [*ref* *Sonka,Hlavac, Boyle 1999*], removing the branches of the skeleton after that computing iteratively the end points.

Active regions

Active regions (ARs), or "sunspots" as they are called when they are seen in the photosphere, are high concentrations of magnetic flux emerging over the solar surface, which are a direct proxy of solar activity. They are clearly identified in different wavelengths: in the corona they appear like high concentrations of EUV emission where magnetic loops can be sometimes be distinguished, whereas they appear as sunspots in the lower atmospheric layers (normally with a dark inner umbra and surrounding penumbra). Figure 5 shows how different an AR (NOAA 10708) appears when seen at different temperatures compared to the magnetic field in the photosphere. Due to differences between how this feature appears in different wavelengths, there is some controversy over the definition of AR. Throughout this document, an AR will be defined as a concentration

of magnetic flux seen in the photosphere. However, it should be noted that NOAA only designates numbers to those which have a white light signature (ie. Sunspot).[i]

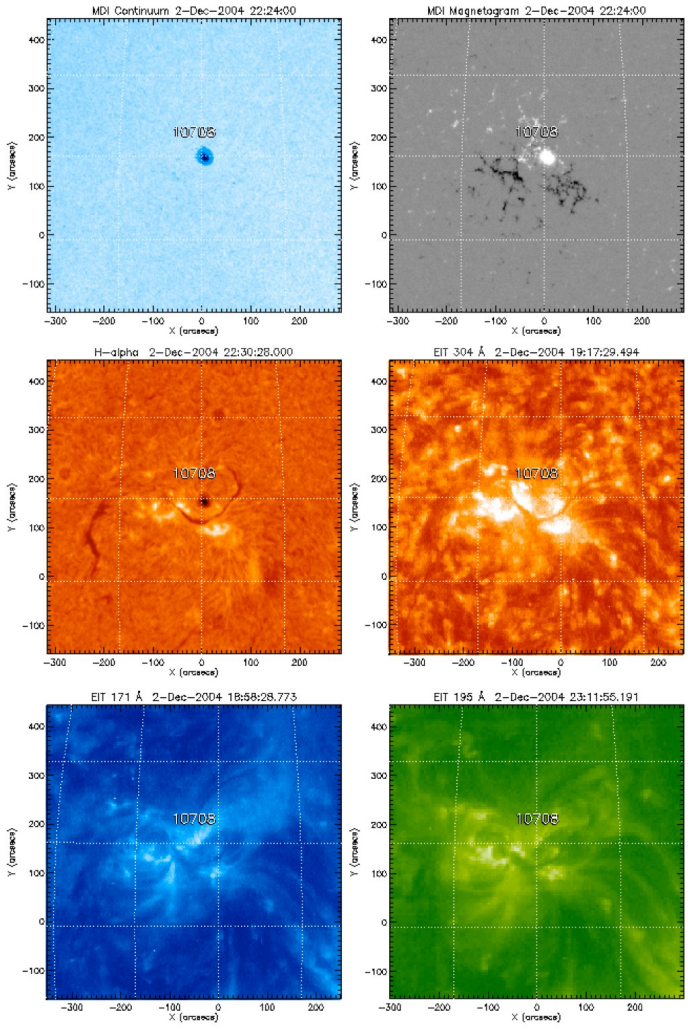


Figure 5. These six images show NOAA AR 10708 as is seen on, from left to right and top to bottom: continuum (photosphere) and magnetogram from MDI, Halpha (chromosphere) from Big Bear Solar Observatory and EIT 304 (higher chromosphere), 171 (corona) and 195 (corona) from SoHO. Extracted from http://solarmonitor.org

As can be seen from Figure 5, the same technique cannot be used for every image. However, for cataloguing purposes there is no need to detect the feature in all of them. There are different algorithms to detect features in each layer of the atmosphere. The one explained here uses magnetogram images, which is the one from which the most information can be extracted. From the space weather point of view, the detection of ARs are very useful because they are the source of other events - flares and coronal mass ejections (CMEs) - that can produce effects on Earth and other planets (and the instruments and astronauts orbiting them). Currently, region-based flare (and CME) forecasting requires determining the magnetic properties of the AR in question and its surroundings (Conlon et al., 2009; Zhang et al., 2009 and therein)

The SolarMonitor Active Region Tracker (**SMART**; Higgins et al., 2010) is a newly developed algorithm used for detecting, tracking and cataloguing emerging active regions throughout their evolution and decay. It uses the SoHO Michelson Doppler Interferometer (MDI) full-disk line-of-sight magnetograms to extract magnetic properties such as the region size, total flux, flux imbalance, flux emergence rate, Scrhijver's R-value, R* and Falconer's measurement of non-potentiality.

SMART can be separated in four phases, first the magnetograms are segmented into individual feature masks, a characterization algorithm is then run on each extracted region to determine its properties. The properties are classified and finally catalogued (with tracking information included). Here we are interested only in the segmentation technique, so we refer the reader to the SMART documentation for further information on the other aspects of the algorithm.

Magnetograms, after calibration, contain values in gauss, so there is negative (usually represented as black) and positive (as white) values. SMART uses two consecutive magnetograms (denoted as B_t and B_{t-dt} to remove transient features and to extract time-dependent properties later.

The first steps, shown on Figure 6, are: smooth the images with a 10x10 pixels^2 2D Gaussian with a full-width at half maximum of 5 pixels to remove ephemeral regions which have scales on the order of 10 Mm (7 MDI pixels); remove the background using a static threshold of ~70 G[ii]; line-of-sight component of the field is corrected using a cosine correction factor. At this stage, B_{t-dt} is differentially rotated to time, t.

Binary masks, M_t and M_{t-dt} , are created from the corrected magnetograms, setting all pixes above the threshold to one. Features of less than 50 pixels and those which are not present in both masks are removed after dilating each mask by 10 pixels and subtracting them, which results in a transient features mask. This mask is removed from the un-grown M_t, which is then dilated by 10 pixels and results in a mask of the desired features. Individual contiguous features are indexed by assigning ascending integer values in order of decreasing feature size. The segmentation output is an indexed grown mask (IGM_t).

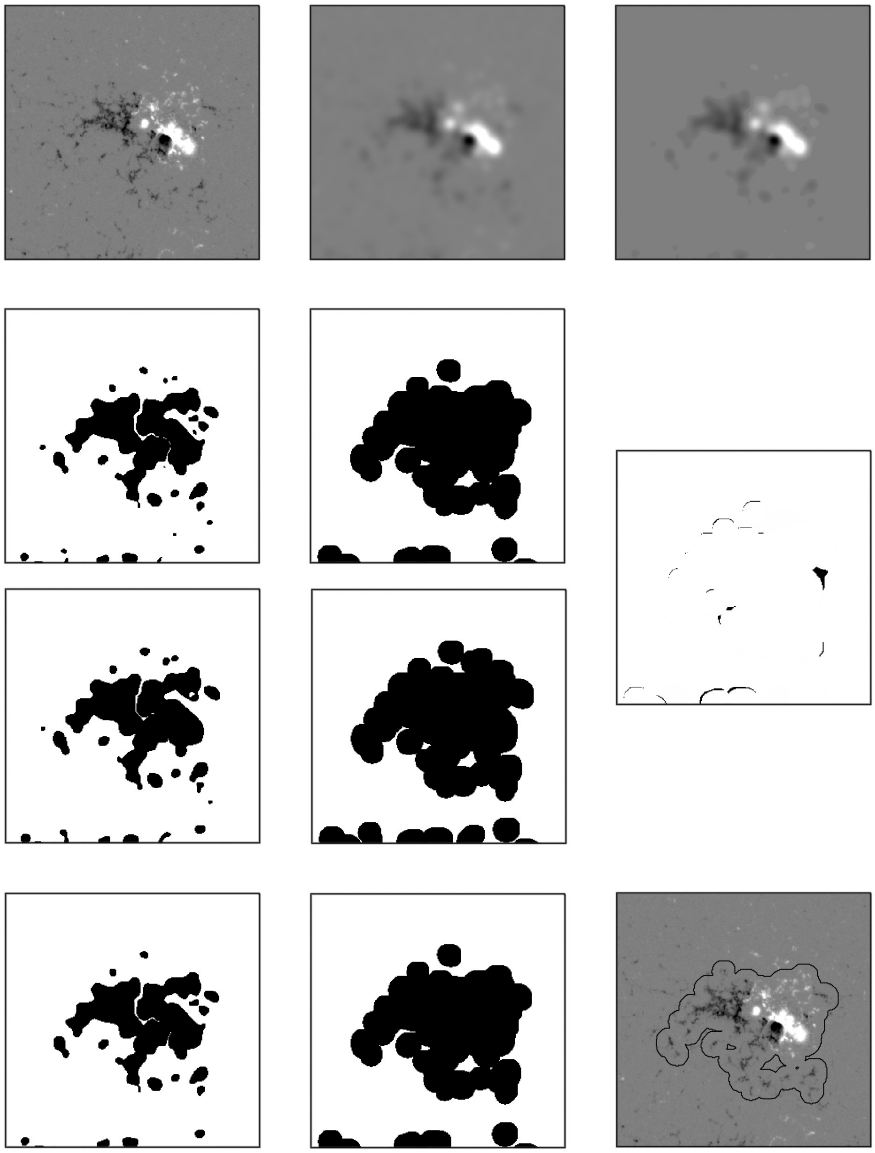


Figure 6. SMART's steps. First image, top-left, show the magnetogram of an AR. Next images on the left show the smoothing and threshold cut of 70 G respectively. Next two rows show the masks from that image and the one obtained 96 minutes before, the growing step and their difference. The last row shows the first mask after subtracting the difference and after dilation, which is the final segmentation mask used to plot the contours on the last image.

Each feature is then characterized individually by their physical values obtained from the image resulting from the multiplication of the IGM_t by B_t .

SMART algorithm is unique among automated AR extraction algorithms in that it allows the temporal analysis of magnetic properties from birth ant through multiple solar rotations. It is planned for further versions of SMART to incorporate a flare event probability. This will be done in part by adapting the McAteer et al. (2009) and Conlon et

al. (2009) multifractal technique to the SMART feature characterization. Conlon et al. (2009) propose a 2D wavelet transform modulus maxima method (WTMM) to study the multifractal properties of AR magnetic fields. The segmentation of the AR is obtained by using WTMM, which applying it to the AR provides an adaptive space-scale partition of the fractal distribution. It is shown in their work that the study of the fractal nature of an AR can be used as a flare forecasting tool.

Other algorithms are interested in the classification of sunspots as it has been done manually by human observers. That is the case of the code written by (Colak and Qahwaji, 2008) where, using white light images and magnetograms from MDI/SoHO and neural network techniques, the sunspots are detected and classified automatically according to the McIntosh classification system. This system has the advantage that it compares well the NOAA AR identification scheme.

Also, two codes presented in **Zharkova et al. (2005)** were developed for EGSO at Meudon observatory to study ARs by comparing Ca II k1 emission line ground-based images and MDI white-light images with MDI magnetograms, obtaining information about the sunspots properties, i.e. size of the umbra and penumbra and the intrinsic magnetic properties.

The code which works with white-light images from MDI/SoHO and Ca II K1 line images from a ground-based telescope at Meudon Observatory using edge detection techniques.

Once the data is preprocessed (limb fitting, removal of geometrical distortion, centre position and size standardisation) the detection starts. The segmentation is obtained using the Sobel edge detection technique on the photospheric images after a Gaussian smoothing was applied. The use of a global threshold segments the edges obtained. The existing gaps are filled with the close and watershed morphological operators and they are preceded by new segmentation based on dynamic thresholding (constant for the MDI data, variable for ground-based images due to the Earth's atmosphere unstable conditions) to extract the umbra and penumbra parts of the sunspots.

These two codes are complementary, whereas one extracts the magnetic field properties of the ARs, the other extracts the properties of the sunspots. Not all the ARs, as defined by SMART, produce the photospheric signature seen on the continuum as sunspots, therefore studying the output of both codes will help to the understanding of the solar magnetic field and its production.

Comparison of their results

Other algorithm are interested on the classification of sunspots as it has been done manually by human observers. That is the case of the code done by (ref Colak and Qahwaji) where using white light images and magnetograms from MDI/SoHO and neural network techniques the sunspots are detected and classified automatically according to the McIntosh classification system. This system has the advantage that it compares well the NOAA AR identification scheme. Also, two codes were developed for EGSO at Meudon observatory to study the ARs comparing respectively Ca II k1 line ground-base images and MDI white-light with MDI magnetograms. Those obtained information of the characteristics of the sunspots, i.e. size of the umbra and penumbra and the intrinsic magnetic properties. (Falconer's???)

CME's

Coronal Mass Ejections are large-scale eruptions of plasma and magnetic field with high energies. The particles expelled in a CME travel from the Sun at velocities of hundreds up to several thousand kilometres per second and they have consequences on space-borne instruments and on the planets, auroras are an example seen on Earth and other magnetic-field-protected planets as Saturn (Prangé et al., 2004). Their diffuse and transient nature makes them a difficult object to automatically identify. CMEs are observed with the help of coronagraphs. A coronograph is a telescope attachment designed to block out the solar disk allowing to see the surrounded corona (which is fainter than the disk). A natural example of a coronagraph happen during the solar eclipses, when the moon cover the sun. On space, SoHO and STEREO are equipped with them. *Large Angle and Spectrometric Coronagraphs* (LASCO) C2 and C3 are onboard SoHO and STEREO is equipped with Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI) which contains two coronographs namely COR1 and COR2. SECHHI also counts with a wide-angle visible-light imaging system called H1.

There is available a few catalogues obtained from those instruments mantained for their main teams: CDAW Catalog and NRL LASCO CME List from SoHO and COR1 CME Catalog and HI1 Event List from STEREO. Those catalogues provide information of when a CME happen and its properties, including the position angle, angular width, height, velocity and acceleration. However, since they are human-based process they are time consuming and the parameters measured are subject to human bias. Therefore automation of their detection is a must. To date several automated CME detection algorithms have been proposed to used with LASCO C2 data and extensible to COR1. The Computer Aided CME Tracking (**CACTus**) was the first one of its kind developed at the Royal Observatory of Belgium (Berghmans et al., 2002), its development was followed by the Solar Eruptive Event Detection System (**SEEDS**; Olmedo et al., 2008) from the Geroge Mason University and Automatic Recognition of Transient Events and Marseille Inventory from Synoptic maps (**ARTEMIS**; Boursier et al., 2009) from the LASCO team at Laboratoire d'Astrophysique de Marseille. All those automated relies on the uses of more than one frame to detect a CME; **Byrne et al., (2009)** from Trinity College Dublin propose a method to overcome this problem for real-time detection.

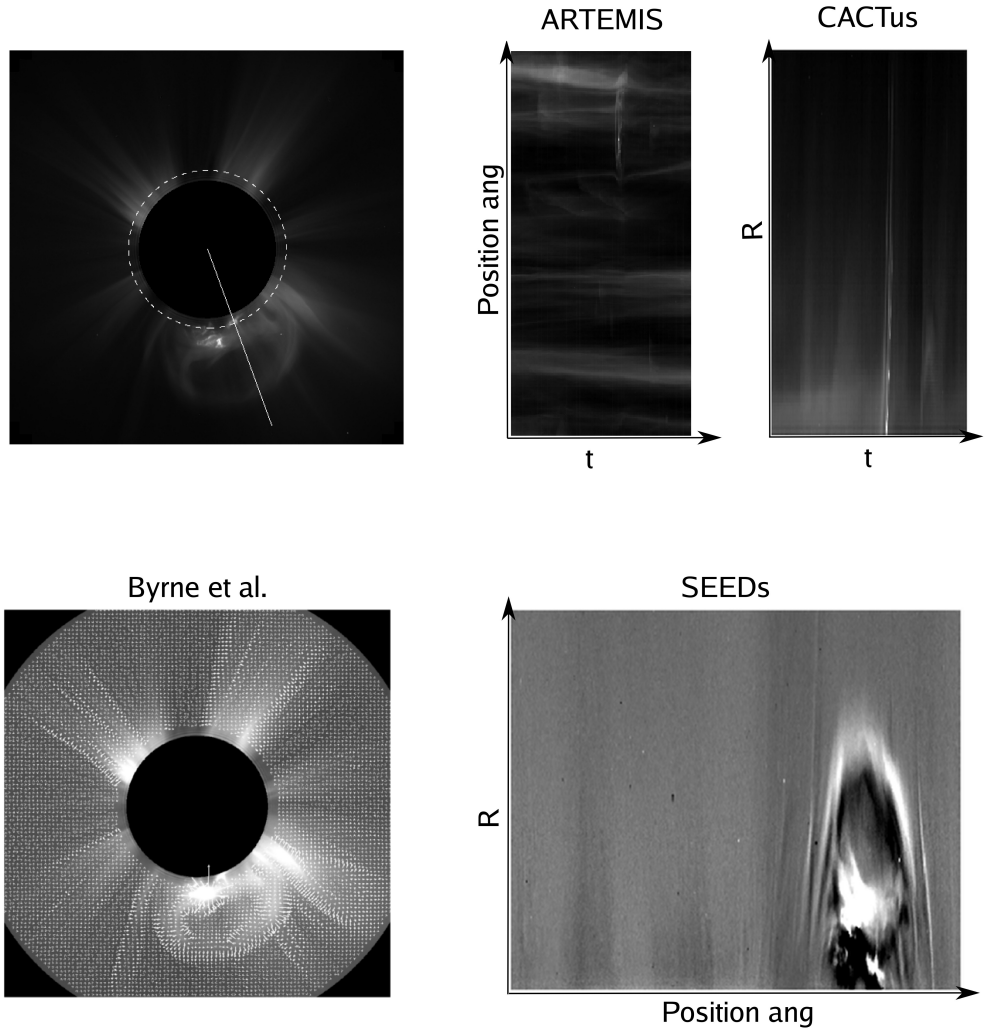


Figure 7. Example of the same CME seen by the different algorithms presented here. Top-left image show a single frame from LASCO C2/SoHO with a dashed line showing the radio shown in Artemis figure and with a solid line the angle at where CACTus method is shown. ARTEMIS and CACTus visualization at the radius and position angle told before agains time. Second row shows Byrne et al. detection with vector representation of the multiscale detection and SEEDs representation of the running difference images at the same time.

The preprocessed of the images followed for these groups differ from the standard methods proposed by the instrument teams. This is due because the standard reduction is not optimized to detect CMEs (background stars, planets and comets appearance are usual on those images. It is worth to mention that LASCO is the most successful comet discoverer in history, having found over *one thousand six-hundred* comets in over thirteen years of operation[iii]. The images are exposure time normalized, corrected from cosmic rays, stars or planets by different methods and transformed to their preferred coordinate system. Figure 7 shows an example of the different transformation used by the methods exposed. CACTus transform each coronographic image from its native Cartesian coordinate system to a polar coordinate system [r,position angle], where r is the radial distance from the centre of the Sun and the position angle is the angle from certain

reference point (in here the position angle is measured anticlockwise from the ecliptic). The successive transformed images are stacked to the first one constructing a [r,position angle,t] datacube, which is iteratively processed to estimate the background and to remove the dust corona and rotating streamers. [r,t] slices are extracted from the cleaned datacube to proceed with the CMEs detection. ARTEMIS creates synoptic maps which consists on the generation of [position angle,t] images, being complementary to CACTus approach. Finally SEEDS works in polar coordinates [r,position angle] after the construction of running-difference image of two consecutive images. Byrne et al. method no produce any coordinate transformation previously to the CME detection.

It is easily seen on the Figure 6 that the same CME present different signatures on each of the transformations, therefore the techniques for the detection are different. CMEs appear as inclined lines on CACTus, which relies on the Hough transform for it detection, whereas on ARTEMMIS the CMEs appear with different morphologies which they have classified in 4 different types: undistorted vertical streaks without temporal dispersion, quasi-symmetric arc shapes, arc shapes followed by a second structure with a dark zone in between and a last type which groups events of unclear signature. ARTEMIS detection involves three main steps: filtering, segmentation and merging with high-level knowledge. The filtering is done line by line removing the background with a median filter with a 7 pixels window. The segmentation process is done by a simple thresholding process with a value selected by *experience*, followed by the application of the Line Adjacency Graph (LAG; Pavlidis, 1986) which: removes small artificial "holes" by performing a morphological closure operation, identify regions of interest (ROIs), compute their geometrical and their statistical parameters and remove those smaller than certain size. Finally each ROIs is associated to the same CME if satisfy simultaneously three empirically-determined conditions which form the High-Level knowledge; they are: the difference between the x-coordinates and y-coordinates of their center differs by less than certain amount, and the ration of the radiances at the center of both ROIs on the original synoptic map ranges between given values.

A CME on SEEDs appear as a bright leading-edge enhancement (positive values) followed by a dark area deficient in brightness (negative values), the background appear as grey indicating zero-change. The quasi-static features such as coronal streamers have been removed by the running-difference process. The extraction of the CME is done using a threshold-segmentation technique and region-growing algorithm. The positives values of the image are projected to one dimension along the angular axis, obtaining the angular intensity profile. The threshold value is obtained from such profile as a number of times the standard deviation over the mean of the profile. The number of times of the standard deviation is chosen by experimental methods and its value is often between two and four. This gives the angle of the core, which it is made to grow to cover the whole CME, the region growing connects those values between the maxima of the core-angle and a second threshold calculated as the previous time but just over the values outside the core-angle.

Byrne et al. method, an application of Young and Gallagher (2008), is not automatized yet but it presents a multiscale method for CME detection using a single image. The algorithm create a mesh of vector arrows from the magnitude and angle information obtained from the Canny edge detection. A threshold is specified to the angle information in order to chain pixels along maxima, and the CME front is fitted with an ellipse.

The next step is the characterization of the CMEs' kinematics. It is not much connected with image analysis as it is with the application of a model. However, the nature of Hough transform used by CACTus constrains the CMEs to have constant velocity. Boursier et al. (2009) showed a comparison of the catalogues produced by CACTus, SEEDs and ARTEMIS together with the man-made CDAW showing that the automated catalogues tend to report more than twice as many CMEs as are identified by visual detection, but the major interest on the development of those automated algorithms is to produce not human biased results and be able to produce robust statistical analysis.

Radio Bursts

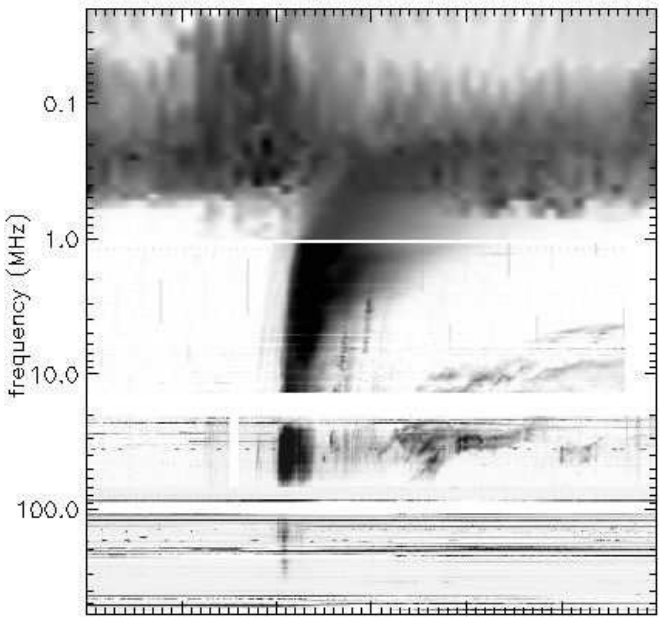
Automatic detection of radio type events or features is not an easy task. The goal of this short document is to give an overview of what has been done so far, of possible data sources and of the main difficulties that we might face. This first draft is not a complete review of the knowledge in this field and should be improve in the future.

Reminders

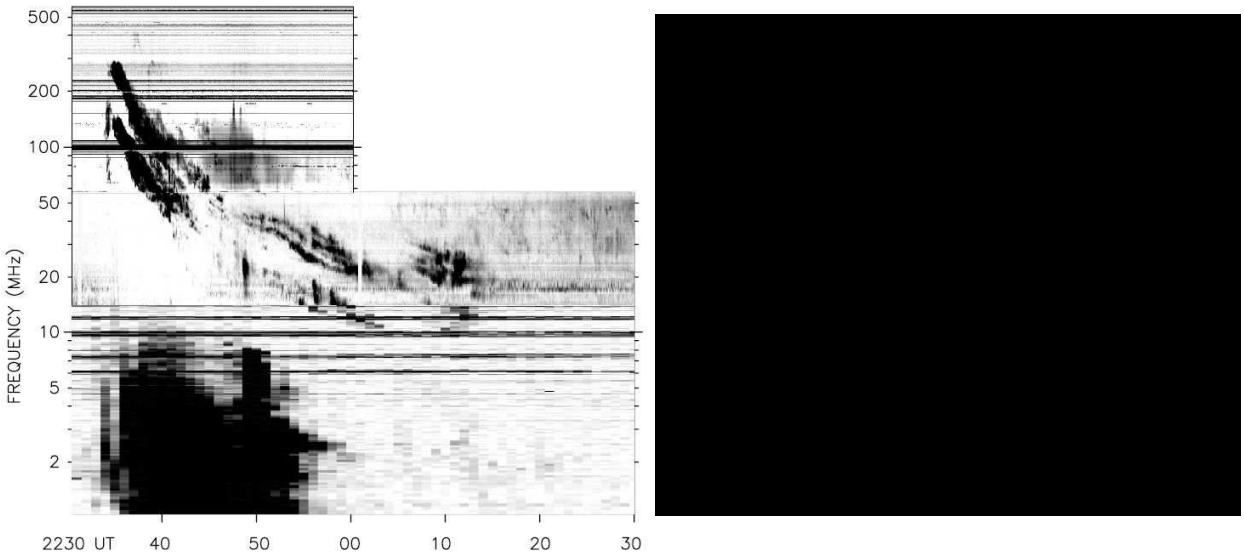
Let's remind here that ground base radio observations are limited by the ionospheric absorption. Below a limit of 10-15 MHz radio observations are performed from space. Waves instruments observe below 14MHz, also called the decametric-hectometric (D-H) domain. This means that observations from space are the main source of information concerning interplanetary radio events. The higher is the frequency, the closer to the Sun you can observe. The BIRS observatory which covers the range 5MHz-62.5 Mhz is the only ground base radio observatory at such low frequency (only when the ionospheric absorption is low). It's also important to note that ground observatories allow a higher observation cadence (<1s) but are limited to about 7 or 8 hours a day.

Radio type events

We only investigate here type II and III radio events. Type III bursts are usually associated with solar flares and are produced by electron beams travelling along the coronal and interplanetary magnetic field lines. Several typeIII can occur in a very short time before the flare.
Here is an example taken from secchiRH tool (combined observations from Waves, DAM and Artemis):



Interplanetary type II bursts (or kilometric type II) are usually the signature of the bow shock of fast CME's. The origin of metric type II burst, which might not be connected to the interplanetary counterpart, remains unclear but maybe associated with a fast CME or a blast wave. Both type II show a slow drift on time-frequency plots. Below is an example of a metric type II (combined observation of Culgoora, BIRS and Waves data) which disappear before reaching the interplanetary space. On the left: an interplanetary type II driven by an ICME from Waves observations (see the different time scales)

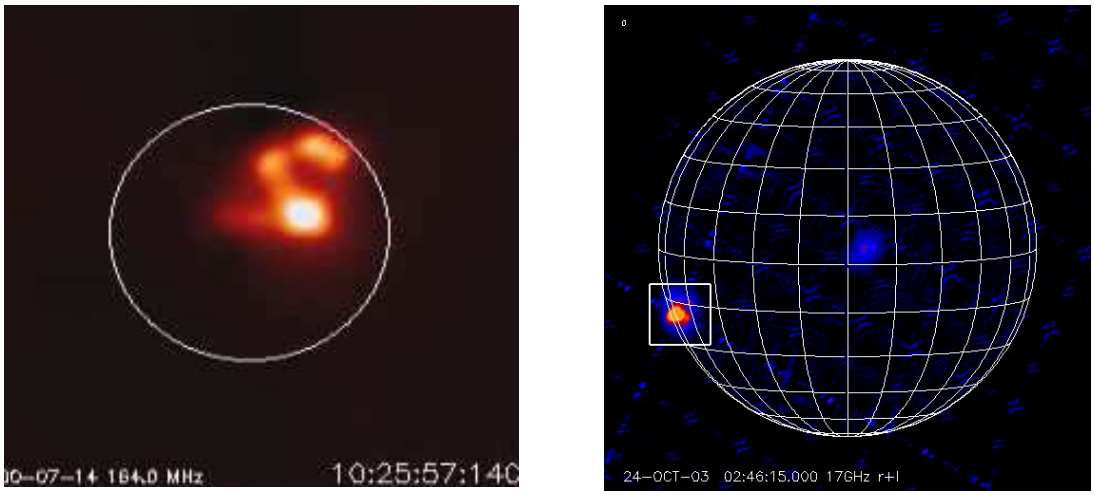


Side informations:

- There is not always a type III before a type II
- A type II can precede a type III (the type II leading to a non flaring type III)
- A type III can originate from Jupiter (Volcanic activity on Io)

"Radio active" regions

Nancay RH (164 MHz) and Nobeyama RH (17GHz)



A software has been developed by C. Renie from Paris observatory to extract informations from the NRH synthetized images. An elliptic gaussian is fitted on the brightest source in

order to get a location and, by subtracting this source, extract informations about the secondary sources. Analysis of these images is not straightforward, on need to distinguish noise storms from flare contribution. See below comments from Pr Shibasaki for the NoRH images, which observes at higher frequencies (17GHz, 34 GHz, 10 images/sec). The Nancay Radioheliograph provides 2-D images of the sun at 5 frequencies (between 150 and 450 Mhz) at a cadence of 8 images/second.

Automatic detection of type II and III

Automated detection of Type II Solar Radio Bursts in Radiospectrograms
T. Dudok de Wit, M. Tamaazousti

They developed a method based on the Hough transform to detect interplanetary Type II automatically from wind/waves data.The method still needs some optimizations and has not been applied to a large set of data to determine its efficiency. No publications, but TDdW agrees to provide informations.

Automated detection of solar radio burst and plasma structure observed with Wind/Waves
N.C. Parikh, M.L. Rilee

In the summary:
"*Multi-spacecraft space science missions such as CLUSTER (launch 2000) and Magnetospheric Multi Scale (launch 2008) carry radio receivers that operate below the ionospheric limit of 10-15 MHZ down to a few KHz and produce large amounts of in situ and remote observations of a great vaiety of Solar System phenomena.(...) In this paper we analyze dynamic spectra from the thermal noise receiver of the Waves experiment on board the Wind spacecraft to determine contours of the local plasma emission and typeII and III solar radio bursts*".
No access to full article yet.

Automatic recognition of type III solar radio bursts:automated radio burst identification system method and first observations
C.V. Lobzin, I. H. Cairns, P.A. Robinson, G. Steward, G. Patterson

They developed a method, which seems quite interesting, based on the binarization of the dynamic spectra and the radon transform (and eventually smoothing). They claim a 84% performance, by comparing their results to the NGDC list of events for 13 days in 2002 (ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_RADIO/SPECTRAL/SPEC_NEW.02). They use data from the Radio Solar Telescope Network (RSTN) and from Culgoora. This is the first step of a more complex system called ARBIS (Automated Burst Identification System, see also below) which aims is to provide space weather forecasting based on radio spectrograph data.

Automatic Recognition Of Coronal Type II Radio Bursts: The Automated Radio Burst Identification System Method And First Observations. (*published 2010 January 22*)
Vasili V. Lobzin, Iver H. Cairns, Peter A. Robinson, Graham Steward and Garth Patterson

From the same set of observations, ARBIS2 also detects coronal type II events. The method is a bit more complex, with a preprocessing step involving filtering and morphological operations. But the main idea is the same: using the Hough transform (similar to the radon transform) to find straight segments on a binary image. The authors seem to mix up coronal and interplanetary type II in the introduction, whereas there is no clear evidence that both are linked.

Automatic recognition of low-frequency radio planetary signals. *Planetary Radio Emissions IV*, Austrian Academic Science Press, Vienna, pp. 359–368.
H. de Lassus, Ph. Daigremont, F. Badran, S. Thiria and A. Lecacheux

In the summary :
We address the problem of autonomous decision making in classification of radioastronomy spectrograms from spacecraft. It is known that the assessment of the decision process can be divided into acceptation of the classification, instant rejection of the current signal classification, or rejection of the entire classifier model. We propose to combine prediction and classification with a double architecture of Time Delay Neural Network (TDNN) to optimize a decision minimizing the false alarm risk. Results on real data from URAP experiment aboard Ulysses spacecraft show that this scheme is tractable and effective.
Full article might be retrieved from A. Lecacheux

Japanese experience from Pr. Shibasaki (Nobeyama)

"*We have been using an automatic burst detection method developed during the initial operation of NoRH in 1992. Listed events in "http://solar.nro.nao.ac.jp/norh/html/event/" were detected automatically using this method. We use correlation plots " http://solar.nro.nao.ac.jp/norh/html/cor_plot/" for burst detection. These time series data are generated every day after the end of daily observation. They are averaged visibility amplitudes of longer baseline combinations every second. Visibilities of longer baselines are very sensitive to brightenings of compact structures such as radio bursts. They are not much influenced by weather conditions. By differentiation of these time series, we can easily identify the suddend increase of the visibility values. At this moment, no one is working on this software but simply using routinely. Shimojo san (shimojo@nro.nao.ac.jp) has been working on automatic detection of prominence eruptions using synthesized images. The results are listed in "http://solar.nro.nao.ac.jp/norh/html/prominence/". Automatic detection of prominences are not robust enough, so it is necessary to delete false events by hands.*"

Application to Heliospheric data

It is much more difficult to define some generic method for remote radio data and in-situ measurements as techniques used are closely related to the instrument. Each instrument has its specific noise and its own set of interferences and then information extraction has to deal with that. As a matter of fact, many such instrument teams have already developed their own data processing code. In a forthcoming update of this document, we shall provide more information concerning this kind of data.

REFERENCES

Altschuler, M. D., Trotter, D. E., and Orrall, F. Q. (1972). Coronal Holes. *Solar Physics*, 26:354–365.

Attrill, G. D. R. and Wills-Davey, M. J. (2009). Automatic Detection and Extraction of Coronal Dimmings from SDO/AIA Data. *Solar Physics*, pages 143–+.

Barra, V., Delouille, V., and Hochedez, J. (2008). Segmentation of extreme ultraviolet solar images via multichannel fuzzy clustering. *Advances in Space Research*, 42:917–925.

Barra, V., Delouille, V., Hochedez, J., and Chainais, P. (2005). Segmentation of EIT Images Using Fuzzy Clustering: a Preliminary Study. In *The Dynamic Sun: Challenges for Theory and Observations*, volume 600 of ESA Special Publication.

Barra, V., Delouille, V., Kretzschmar, M., and Hochedez, J. (2009). Fast and robust segmentation of solar EUV images: algorithm and results for solar cycle 23. *Astronomy & Astrophysics*, 505:361–371.

Berghmans, D., Foing, B. H., and Fleck, B. (2002). Automated detection of CMEs in LASCO data. In A. Wilson, editor, *From Solar Min to Max: Half a Solar Cycle with SOHO*, volume 508 of ESA Special Publication, pages 437–440.

Bernasconi, P. N., Rust, D. M., and Hakim, D. (2005). Advanced Automated Solar Filament Detection And Characterization Code: Description, Performance, And Results. *Solar Physics,* 228:97–117.

Bewsher, D., Harrison, R. A., and Brown, D. S. (2008). The relationship between EUV dimming and coronal mass ejections. I. Statistical study and probability model. *Astronomy & Astrophysics*, 478:897–906.

Boerner, W., Cole, J. B., Goddard, W. R., Tarnawecky, M. Z., Shafai, L., and Hall, D. H. (1983). Impacts of solar and auroral storms on power line systems. *Space Science Reviews*, 35:195–205.

Boursier, Y., Lamy, P., Llebaria, A., Goudail, F., and Robelus, S. (2009). The ARTEMIS Catalog of LASCO Coronal Mass Ejections. Automatic Recognition of Transient Events and Marseille Inventory from Synoptic maps. *Solar Physics*, 257:125–147.

Byrne, J. P., Gallagher, P. T., McAteer, R. T. J., and Young, C. A. (2009). The kinematics of coronal mass ejections using multiscale methods. *Astronomy & Astrophysics*, 495:325–334.

Colak, T. and Qahwaji, R. (2008). Automated McIntosh-Based Classification of Sunspot Groups Using MDI Images. *Solar Physics*, 248:277–296.

Conlon, P. A., Kestener, P., McAteer, R., and Gallagher, P. (2009). Magnetic Fields, Flares & Forecasts. In *AAS/Solar Physics Division Meeting*, volume 40 of AAS/Solar Physics Division Meeting, pages 16.02–+.

Eddy, J. A. (1976). The Maunder Minimum. *Science*, 192:1189–1202.

Fuller, N., Aboudarham, J., and Bentley, R. D. (2005). Filament Recognition and Image Cleaning on Meudon H-alpha Spectroheliograms. *Solar Physics*, 227:61–73.

Harvey, K. L. and Recely, F. (2002). Polar Coronal Holes During Cycles 22 and 23. *Solar Physics*, 211:31–52.

Henney, C. J. and Harvey, J. W. (2005). Automated Coronal Hole Detection using He 1083 nm Spectroheliograms and Photospheric Magnetograms. In K. Sankarasubramanian, M. Penn, & A.

Pevtsov, editor, *Large-scale Structures and their Role in Solar Activity*, volume 346 of Astronomical Society of the Pacific Conference Series, pages 261–+.

Hudson, H. S., Acton, L.W., and Freeland, S. L. (1996). A Long-Duration Solar Flare with Mass Ejection and Global Consequences. *Astrophysics Journal*, 470:629–+.

Krista, L. D. and Gallagher, P. T. (2009). Automated Coronal Hole Detection Using Local Intensity Thresholding Techniques. *Solar Physics*, 256:87–100.

McAteer, R. T. J., Gallagher, P. T., and Conlon, P. A. (2009). Turbulence, Complexity, and Solar Flares. *ArXiv e-prints.*

McIntosh, S. W. and Gurman, J. B. (2005). Nine Years Of Euv Bright Points. *Solar Physics*, 228:285–299.

Olmedo, O., Zhang, J., Wechsler, H., Poland, A., and Borne, K. (2008). Automatic Detection and Tracking of Coronal Mass Ejections in Coronagraph Time Series. *Solar Physics*, 248:485–499.

Parker, E. N. (1958). Dynamics of the Interplanetary Gas and Magnetic Fields. *Astrophysics Journal,* 128:664–+.

Pavlidis, T. (1986). A vectorizer and feature extractor for document recognition. *Computer Vision, Graphics, and Image Processing*, 35(1):111–127.

Podladchikova, O. and Berghmans, D. (2005). Automated Detection Of Eit Waves And Dimmings. *Solar Physics*, 228:265–284.

Prang«e, R., Pallier, L., Hansen, K. C., Howard, R., Vourlidas, A., Courtin, R., and Parkinson, C. (2004). An interplanetary shock traced by planetary auroral storms from the Sun to Saturn. *Nature*, 432:78–81.

Sanchez, F., Collados, M., and Vazquez, M. (1992). *Solar observations: Techniques and interpretation.*

Shih, F. Y. and Kowalski, A. J. (2003). Automatic Extraction of Filaments in H Solar Images. *Solar Physics*, 218:99–122.

Soille, P. and Talbot, H. (2001). Directional morphological filtering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1313–1329.

Sonka, M., Hlavac, V., and Boyle, R. (1999). *Image Processing Analysis and Machine Vision.*

Stix, M. (2004). *The sun : an introduction.*

Tompsett, M. F., Amelio, G. F., Smith, G. E., (1970). "Charge Coupled 8-bit Shift Register". *Applied Physics Letters* **17**: 111–115.

Thompson, B. J., Plunkett, S. P., Gurman, J. B., Newmark, J. S., St. Cyr, O. C., and Michels, D. J. (1998). SOHO/EIT observations of an Earth-directed coronal mass ejection on May 12, 1997. *Geophys. Res. Lett.*, 25:2465–2468.

Young, C. A. and Gallagher, P. T. (2008). Multiscale Edge Detection in the Corona. *Solar Physics*, 248:457–469.

Zhang, J., Wang, Y., and Robinette, S. (2009). Toward Creating A Comprehensive Digital Active Region Catalog. In *AAS/Solar Physics Division Meeting*, volume 40 of AAS/Solar Physics Division Meeting, pages 09.21–+.

Zharkova, V. V., Aboudarham, J., Zharkov, S., Ipson, S. S., Benkhalil, A. K., and Fuller, N. (2005).

Solar Feature Catalogues In Egso. *Solar Physics*, 228:361–375

[i] http://www.ngdc.noaa.gov/stp/SOLAR/ftpsunspotnumber.html

[ii] The selection of a static threshold is shown as acceptable even though the sun changes its activity over the years. In the paper cited, it is demonstrated why

[III] http://sungrazer.nrl.navy.mil/