

Computer Vision for the *Solar Dynamics Observatory* (SDO)

P.C.H. Martens · G.D.R. Attrill · A.R. Davey · A. Engell · S. Farid · P.C. Grigis · J. Kasper · K. Korreck · S.H. Saar · A. Savcheva · Y. Su · P. Testa · M. Wills-Davey · P.N. Bernasconi · N.-E. Raouafi · V.A. Delouille · J.F. Hochedez · J.W. Cirtain · C.E. DeForest · R.A. Angryk · I. De Moortel · T. Wiegmann · M.K. Georgoulis · R.T.J. McAteer · R.P. Timmons

Received: 3 February 2010 / Accepted: 1 December 2010 / Published online: 7 January 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract In Fall 2008 NASA selected a large international consortium to produce a comprehensive automated feature-recognition system for the *Solar Dynamics Observatory* (SDO). The SDO data that we consider are all of the *Atmospheric Imaging Assembly* (AIA) images plus surface magnetic-field images from the *Helioseismic and Magnetic Imager* (HMI). We produce robust, very efficient, professionally coded software modules that can keep up with the SDO data stream and detect, trace, and analyze numerous phenomena, in-

The Solar Dynamics Observatory

Guest Editors: W. Dean Pesnell, Phillip C. Chamberlin, and Barbara J. Thompson.

P.C.H. Martens · G.D.R. Attrill · A.R. Davey · A. Engell · S. Farid · P.C. Grigis · J. Kasper · K. Korreck · S.H. Saar · A. Savcheva · Y. Su · P. Testa · M. Wills-Davey
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

P.C.H. Martens (✉)

Department of Physics, Montana State University, 247 EPS, Bozeman, MT 59717-3880, USA
e-mail: martens@physics.montana.edu

A. Savcheva

Astronomy Department, Boston University, 725 Commonwealth Avenue, Boston, MA 02215, USA

P.N. Bernasconi · N.-E. Raouafi

Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, MD 20723, USA

V.A. Delouille · J.F. Hochedez

SIDC-Royal Observatory of Belgium, Avenue Circulaire 3, 1180 Brussels, Belgium

J.W. Cirtain

Marshall Space Flight Center-NASA, Mail Code:VP 62, Marshall Space Flight Center, AL 35812, USA

C.E. DeForest

Southwest Research Institute, 1050 Walnut Street Suite 300, Boulder, CO 80302, USA

R.A. Angryk

Department of Computer Science, Montana State University, 362 EPS, Bozeman, MT 59717-3880, USA

cluding flares, sigmoids, filaments, coronal dimmings, polarity inversion lines, sunspots, X-ray bright points, active regions, coronal holes, EIT waves, coronal mass ejections (CMEs), coronal oscillations, and jets. We also track the emergence and evolution of magnetic elements down to the smallest detectable features and will provide at least four full-disk, non-linear, force-free magnetic field extrapolations per day. The detection of CMEs and filaments is accomplished with *Solar and Heliospheric Observatory* (SOHO)/*Large Angle and Spectrometric Coronagraph* (LASCO) and ground-based H α data, respectively. A completely new software element is a trainable feature-detection module based on a generalized image-classification algorithm. Such a trainable module can be used to find features that have not yet been discovered (as, for example, sigmoids were in the pre-*Yohkoh* era). Our codes will produce entries in the *Heliophysics Events Knowledgebase* (HEK) as well as produce complete catalogs for results that are too numerous for inclusion in the HEK, such as the X-ray bright-point metadata. This will permit users to locate data on individual events as well as carry out statistical studies on large numbers of events, using the interface provided by the Virtual Solar Observatory. The operations concept for our computer vision system is that the data will be analyzed in near real time as soon as they arrive at the SDO Joint Science Operations Center and have undergone basic processing. This will allow the system to produce timely space-weather alerts and to guide the selection and production of quicklook images and movies, in addition to its prime mission of enabling solar science. We briefly describe the complex and unique data-processing pipeline, consisting of the hardware and control software required to handle the SDO data stream and accommodate the computer-vision modules, which has been set up at the Lockheed-Martin Space Astrophysics Laboratory (LMSAL), with an identical copy at the Smithsonian Astrophysical Observatory (SAO).

Keywords Instrumentation and data management · Solar Dynamics Observatory

1. Introduction

The open and immediate dissemination of solar data from all NASA solar missions is one of the greatest assets in solar physics, and the *Solar Dynamics Observatory* (SDO) represents

I. De Moortel

School of Mathematics & Statistics, University of St Andrews, North Haugh, St Andrews, KY16 9SS, UK

T. Wiegmann

Max-Planck-Institut für Sonnensystemforschung, Max-Planck Strasse 2, 37191 Katlenburg-Lindau, Germany

M.K. Georgoulis

Research Center for Astronomy and Applied Mathematics, Academy of Athens, 4 Soranou Efessiou Street, Athens 11527, Greece

R.T.J. McAteer

School of Physics, Trinity College Dublin, Dublin 2, Ireland

R.T.J. McAteer

Department of Astronomy, New Mexico State University, P.O. Box 30001, MSC 4500, Las Cruces, NM 88003-8001, USA

R.P. Timmons

Lockheed Martin Advanced Technology Center, Bldg. 252, 3251 Hanover Street, Palo Alto, CA, 94304, USA

a new frontier in quantity and quality of solar data. At about 1.5 TB per day, the data will not be easily digestible by solar physicists using the same methods that have been employed for images from previous missions. The availability of this imagery in a form that is useful for scientists will be crucial to the success of the mission.

In order for solar scientists to use the SDO data effectively they need metadata that will allow them to identify and retrieve data sets that address their particular science questions. Providing this metadata, via the pipeline processes described below, is the core of our computer vision project.

We are building a comprehensive computer-vision post-processing pipeline for SDO, abstracting complete metadata on many of the features and events detectable on the Sun without human intervention. Our feature-finding team will debug, deploy, and support a modular, extensible pipeline framework to augment the existing data system, and comprises experts in each subfield of computer vision. Our project unites more than a dozen individual, existing codes into a systematic tool that can be used by the entire solar community. In addition to static, standards-based codes that detect well-defined and well-studied features, we are developing and deploying a trainable feature recognition system for post-processing. It will enable the systematic study of human-recognizable features that are yet to be identified. This unique capability allows flexible scientific exploration of the SDO data set as new types of features gain interest.

2. Operations Concept and Overview of Modules

The feature-finding codes described here are part of the SDO *Event Detection System* (EDS) at the *Joint Science Operations Center* (JSOC; joint between Stanford and the Lockheed Martin Solar and Astrophysics Laboratory (LMSAL)). The basic purpose of the EDS is to analyze the SDO imagery for events and features as soon as the data reach the JSOC, *i.e.* in near real time. This concept implies that the analysis must be automated. The setup allows for the generation of space-weather alerts, relevant quicklook images and movies, as well as timely alerts for the solar-physics community. The feature-finding modules that are described in the remainder of this article form the core part of the EDS, but there will also be room for other, community-produced tools.

The metadata produced by the EDS software modules are stored in the *Heliophysics Event Knowledgebase* (HEK), which is accessible on-line for the rest of the world directly or via the *Virtual Solar Observatory* (VSO). Solar scientists will be able to use the HEK to select event and feature data to download for science studies. Such a preselection is necessary because it is impossible to retrieve and store the full 1.5 TB per day of SDO imagery. SDO data is available from archives at JSOC, the Smithsonian Astrophysical Observatory (SAO), and European sites. In order for our feature-finding algorithms to survive the throughput of SDO data, we are producing software that is modular (there is an independent, separate program for each feature-finding task), robust, and efficient.

Some of the feature-finding modules described here have been designed to monitor the data stream and extract metadata continuously. Others (such as the flare module) are self-triggered: they carefully monitor part of the data stream for an alert and then spring into full action when such an alert is issued. A third group of modules does not operate continuously, but is triggered by one or more alerts from other modules (*e.g.*, the wave-detection algorithms).

In Tables 1 and 2 we present an overview of the software modules produced by our Feature Finding Team (FFT), specifying their names, modes of operation, scientific purpose,

Table 1 Elements of the SDO computer vision system related to space weather alerts.

Science Target S/W Module Name	Flares Flare Detective	Coronal Dimming Dimming Detector	Magnetic Feature Tracking SWAMIS	Filaments AAFDC
Triggered or Continuous? Trigger Source Data Trigger Source Module(s) Binned Image for Trigger Cadence for Trigger	Triggered AIA 193 Self 16 X 16 Full	Triggered AIA 193 Self 512 X 512 5-6 minutes	Continuous	Continuous
Source Data Source Data Binning Source Data Cadence	All AIA Channels None Full	AIA 512 X 512 5-6 minutes	HMI Time Averaged Magnetograms None 6 minutes	Global Hi-res Ha Network Full; some smoothing Full (~ 4/day eventually)
S/W Module Data Programming Language Module Location Computing Requirement Responsible Team Member Originating Organization Output Data Volume First Pipeline Installation	Flare Detective IDL LMSAL Paolo Grigis SAO June 2010	Dimming Detector IDL LMSAL 20 minutes CPU/day Meredith Wills-Davey SAO 10 kB/event, 5 MB/movie July 2010	SWAMIS Perl/PDL LMSAL 8-16 CPUs to operate at 4x realtime Craig DeForest SwRI 120 MB/day nominal November 2010	AAFDC IDL and C SAO Most < 1 min CPU/Image Pietro Bernasconi JHU-APL ~ 1.5 MB per Ha image March 2010
Heritage	EIT, TRACE, RHESSI	New	MDI	GBO H-alpha
Science Target S/W Module Name	Active Regions SPoCA	Sigmoids Sigmoid Sniffer	PIL Mapping PIL Finder	CMEs CME Detector/Tracker
Triggered or Continuous? Trigger Source Data Trigger Source Module(s) Binned Image for Trigger Cadence for Trigger	Continuous	Triggered AIA 211 Self Full Resolution 10 min	Continuous	Triggered LASCO C2 and C3 Flares, Dimmings, Sigmoids Full Resolution Full LASCO Cadence
Source Data Source Data Binning Source Data Cadence	AIA 171/195, 211/335, 94 None 15 minutes	AIA 94, 131 None 10-20 s	HMI Magnetograms Depends on science specs ~5 min	LASCO C2 and C3 None; front smoothing Full (~1 hr)
S/W Module Data Programming Language Module Location Computing Requirement Responsible Team Member Originating Organization Output Data Volume Scheduled Installation	SPoCA C++ LMSAL 30 sec/pair of full-res AIA Veronique Delouille ROB 6 kB/event October 2010	Sigmoid Sniffer IDL LMSAL 10 min CPU/image Nour-Eddine Raouafi JHU-APL 5 MB/day November 2010	PIL Finder IDL SAO Standard desktop Alexander Engell SAO ~ 2 Mb/day December 2010	CME Detector/Tracker IDL SAO 30 minutes/event Meredith Wills-Davey SAO 1-8 Mb/day Spring 2011
Heritage	EIT	SXT, SXI, XRT	Kitt Peak, SOLIS	LASCO

module authors, *etc.* The modules in Table 1 include those that spawn near real-time space weather alerts (flares, coronal dimmings, and emerging flux), and others that are closely related to space weather. The modules in Table 2 will be added during the mission and also run on the already archived data to complete the HEK entries for the entire mission. A detailed individual description of each module is given in the remainder of this paper. At the end, we also describe the hardware setup of the data pipeline in which the modules are operating during the mission. Identical copies of the data pipeline have been installed at LMSAL and SAO. The source codes of the software modules that we are developing will be made available to the scientific community via SolarSoft. With that in mind we list, where known, the CPU and output requirements for each module in the remainder of the article.

3. Flare Detection

The flare detection module of the computer-vision system serves a twofold purpose:

- Provide rapid flare alerts for the space-weather community in near real time (*trigger component*)
- Generate a statistical survey of flares and measure physical parameters relevant for flare science (*analysis component*)

This module is named *flare detective* (FD).

Table 2 Further elements of the SDO computer vision system.

Science Target S/W Module Name	Coronal Holes SPoCA	X-ray Bright Points BP Finder	Sunspots SWAMIS	Global NLFFs Optimization Code for Full Disk
Triggered or Continuous? Trigger Source Data Trigger Source Module(s) Binned Image for Trigger Cadence for Trigger	Continuous	Continuous	Continuous	Continuous
Source Data Source Data Binning Source Data Cadence	AIA 171/195, 211/335, 94 None 15 minutes	AIA 171, 195, 211 None (TBC) Full	HMI Magnetograms None TBD (5-60 min)	HMI+AIA for comparison Binned to 256x256 for ARs ~5 min
S/W Module Data Programming Language Module Location Computing Requirement Responsible Team Member Originating Organization Output Data Volume First Pipeline Installation	SPoCA C++ LMSAL Same as for Active Regions Veronique Delouille ROB 6 kB/event January 2011	BP Finder IDL SAO Steve Saar SAO 600 kB/10 minutes February 2011	SWAMIS PDL LMSAL < 1 CPU Craig DeForest SwRI < 1 MB/day Spring 2011	Optimization Code C MPS 2 hours/vector-magnetogram Thomas Wiegelmann Max Planck Sonnenforschung 256x256x256 datacube When HMI VMG's become routine
Heritage	EIT	EIT	MDI	SOT
Science Target S/W Module Name	Jets Jet Detector	Oscillations Oscillation Finder	"EIT Waves" EIT Wave Tracker	Trainable Feature Recognition
Triggered or Continuous? Trigger Source Data Trigger Source Module(s) Binned Image for Trigger Cadence for Trigger	Triggered TBD AIA Channels XRBP & CH 1024x1024 (TBC) 1 min	Triggered AIA 193 Flares, CMEs, Dimmings N/A N/A	Triggered AIA 193 Dimmings 1024x1024 Full	Continuous
Source Data Source Data Binning Source Data Cadence	AIA 193 None 1 min	Appropriate AIA Channels None None	AIA 193 None; smoothing applied Full	All AIA Channels, some HMI Data 128 X 128 subimages Cadence TBD
S/W Module Data Programming Language Module Location Computing Requirement Responsible Team Member Originating Organization Output Data Volume First Pipeline Installation	Jet Detector IDL SAO 15 seconds per event Antonia Savcheva SAO 90 MB/jet February 2011	Oscillation Finder IDL SAO 60 minutes CPU/event James McAtsee New Mexico State Univ. Spring 2011	EIT Wave Tracker IDL SAO Very small Meredith Wills-Davey SAO 2 Mb/event Spring 2011	Trainable Feature Recognition C/IDL SAO Rafal Angryk MSU 0.1% of image volume December 2010 for image vectors
Heritage	XRT	TRACE, EIT	TRACE	TRACE

The trigger component of the FD needs to be very fast and works on heavily binned *Atmospheric Imaging Assembly* (AIA) images of 16×16 macropixels, each covering approximately a $150'' \times 150''$ square (see Figure 1 for an example). A peak-detection algorithm is applied to the integrated signal in each macropixel, thus providing the start, peak, end time, and approximate location of the flare. This approach allows detection of simultaneous flares in different active regions (ARs). We use a detection procedure based on the flare-identification algorithm used on *Reuven Ramaty High Energy Solar Spectroscopic Imager* (RHESSI) data by Christe *et al.* (2008), which works well for noisy and background-affected lightcurves. Basically, this algorithm smooths the lightcurves and detects flares as intervals of positive derivative and negative derivative around a local maximum. We are running the trigger in the 193 Å band, because its behavior is well known from past observations. However, if during the mission it is found that some flares are better observed in the new, hotter bands 131 Å and 94 Å, we will use those as well.

The more computationally intensive analysis component runs only on the subset of data close in time and space to the flare (as detected by the trigger component). The key parameters determined for each flare are: flare timing, location, area, peak intensity, plasma temperature, and emission measure, lightcurves in each EUV channel, characteristic time scales, associated NOAA AR number, *Geostationary Operational Environmental Satellite* (GOES) class, peak emissivity in selected EUV and SXR emission lines from SDO's *Extreme ultraviolet Variability Experiment* (EVE), movies of the flare from rise to end, and

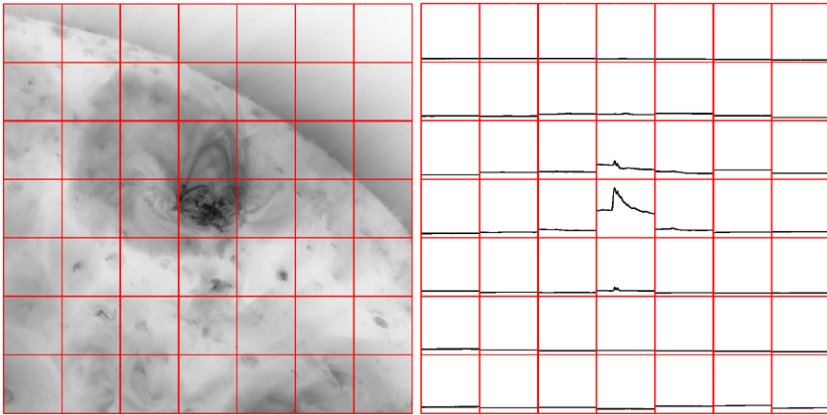


Figure 1 Left: AIA 193 Å image of active region (AR) 11069, taken on 5 May 2010 at 17:02 UT. Each macropixel consists of 256×256 AIA pixels, a square with about $150''$ sides. Right: Lightcurves for each macropixel over a time interval of one hour. The brightness scale is linear without background subtraction. An M1 flare is clearly visible in the lightcurve corresponding to the macropixel containing the core of the AR.

a classification of the flare in terms of flare type (*e.g.* two-ribbon flares). This classification will be provided by the generic feature-recognition system described in Section 16.

The analysis component determines the flare area and centroid from difference images at peak time *versus* flare onset. A set of background-subtracted lightcurves in all EUV channels is generated for the flare interval. These lightcurves are used to estimate the plasma temperature and to provide approximate emission-measure values. Summary quicklook movies will be produced for each event in the 94 Å, 131 Å, 171 Å, and 193 Å channels. Our approach provides complete coverage at least down to GOES C1-level flares, with possible extension to smaller events.

The set of statistical information about the flares provided by the FD over a substantial fraction of a solar cycle can be used to determine the distribution of flares in space, time, size, and thermal energy. This information is of critical importance to study the energy-release mechanisms in flares.

4. H α Filaments

Filaments in the solar chromosphere are well-known, large-scale structures of relatively dense and cool plasma suspended in the hot and thin corona. They are particularly well visible in H α filtergrams. Filaments and their sources, filament channels, are known to align with photospheric magnetic polarity inversion lines (PILs; Martin, 1998). All solar eruptions occur above PILs. In addition, filaments are known to involve helical magnetic fields, twisted beyond their minimum-energy, current-free, magnetic configuration (Martin, Bilimoria, and Tracadas, 1994; Rust and Martin, 1994; Pevtsov, Balasubramaniam, and Rogers, 2003). Non-potential (*i.e.*, helical) magnetic fields are invariably involved in solar eruptions and give rise to coronal mass ejections (CMEs). Filaments themselves often erupt fully or partially into CMEs, leading to a complete or partial filament disappearance from the solar disk (Gilbert *et al.*, 2000; Asai *et al.*, 2003; Jing *et al.*, 2004). If one knows the sense of twist (chirality) in a filament before its disappearance, then one has additional clues about the magnetic helicity of the CME that might

be useful in assessing the CME's possible geoeffectiveness (e.g., Yurchyshyn *et al.*, 2001; Rust *et al.*, 2005).

In recent years many algorithms have been developed for filament detection; see, e.g., Gao, Wang, and Zhou (2002), Shih and Kowalski (2003), and Qu *et al.* (2005). The code by Fuller, Abouadarham, and Bentley (2005) was recently implemented as part of the European Grid of Solar Observations (EGSO) project. Most of these codes do not go far beyond the mere detection of filaments. The *Advanced Automated Filament Detection and Characterization Code* (AAFDCC) that we have developed is a step beyond a typical filament identification code. Besides mere identification, *i*) it determines the filament's complete shape, spine, and orientation angle, *ii*) if a filament is broken up into two or more pieces, it correctly identifies them as a single entity, *iii*) it finds the filaments' magnetic chirality (sense of twist), and *iv*) it tracks them from image to image for as long as they are visible on the solar disk. The code was originally developed by Bernasconi, Rust, and Hakim (2005) and is fully functional, tested, and validated. It currently runs daily on an APL server and so far has fully automatically generated a database of daily solar-filament properties from July 2000 until the present with only occasional gaps due to lack of H α images for specific days. To our knowledge, our code is the only one that accomplishes all of these important tasks at once without any human intervention.

We have described the algorithm extensively in Bernasconi, Rust, and Hakim (2005) so here we only highlight its most important steps. At regular intervals (currently four times per day) the code checks if new (most recent) full-disk H α images have been posted online by the Global High-Resolution H α network (<http://www.bbso.njit.edu/Research/Halpha>). If they have, it downloads the most recent image, performs some preprocessing to clean and standardize it, and begins the analysis by identifying all dimmings in the image (candidate filaments). It then eliminates all very small, or round and very dark features that most likely are not filaments, but sunspots. For each positively identified filament it determines its chain code outlining the shape, and its core skeleton, or spine. The algorithm can also merge filament segments close to each other, thus uncovering the true shape. Once the spine is identified, the code checks for filament barbs whose orientation (left/right-handed) implies the filament magnetic chirality (dextral/sinistral, respectively: Martin, 1998). The helicity of a given filament is determined if a decisive majority of barbs have a common orientation and is left undetermined if the number of right-bearing and left-bearing barbs is evenly split. Finally, the code compares the location of each filament found on the most recent image with the location of filaments detected in an older image. This allows us *i*) to track the evolution of filaments in time; *ii*) to detect the appearance of a new filament; *iii*) to detect if a filament has disappeared from the visible disk, thus indicating a possible filament eruption (this requires a fast cadence which is not yet available). A typical result is shown in Figure 2. We have validated our code by comparing its results with the filament list of Pevtsov, Balasubramaniam, and Rogers (2003). The list was reproduced with an accuracy of 72%. The main difference between our results and those of Pevtsov *et al.* was that in our case a larger number of filaments had undetermined chirality. We were able to identify the orientation of all filament barbs in all cases, so we attribute this discrepancy to our unbiased automated chirality finder, as opposed to the biased chirality determination performed by a human operator (see Bernasconi, Rust, and Hakim, 2005).

For the SDO computer vision project, we have modified the 2005 code to improve its filament identification and characterization performance. We are now applying an adaptive-threshold method that changes the threshold for identification of a filament's outline. This allows the code to identify and characterize thin filaments in active regions that usually are difficult to detect with a single threshold for the entire solar disk, as was done in the 2005

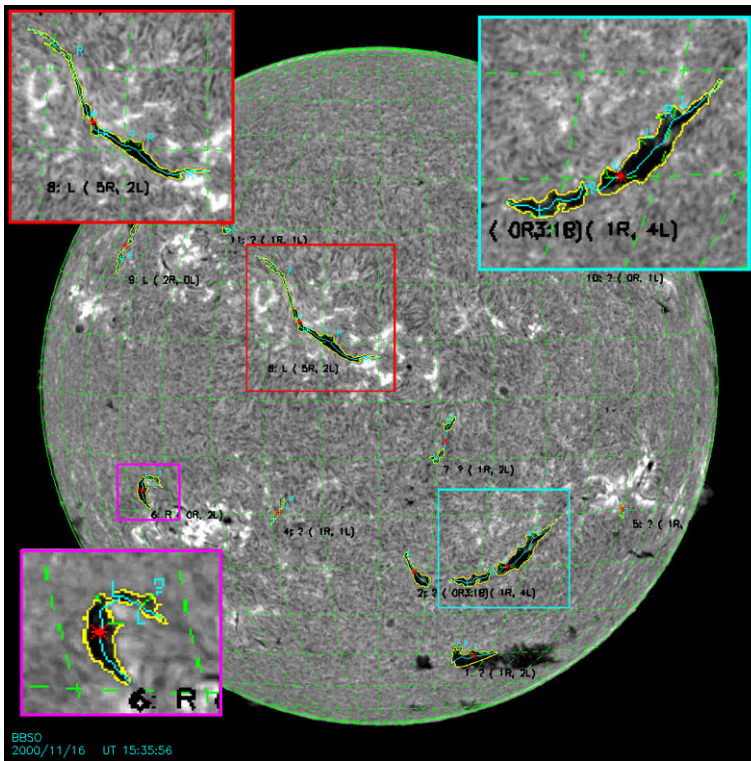


Figure 2 Typical results of the filament finder algorithm of Bernasconi, Rust, and Hakim (2005), applied to an H α image taken by the Big Bear Solar Observatory (BBSO) H α telescope at 15:36 UT on 16 November 2000. Insets at the sides show three well-formed filaments magnified: one sinistral (red box) and two dextral (magenta and cyan boxes). Yellow curves outline the filaments, cyan curves indicate their spines, and the red points indicate the intensity-weighted centroid of the filaments.

version. We have also made the detection of merging in a segmented filament more accurate. Another enhancement we have implemented with respect to the 2005 version is adding to the tracking algorithm the capability to determine whether a filament is the result of merging of two or more previously identified filaments, and if an older filament has split into two or more components in a subsequent image. For the SDO feature-finding project we have also produced a new format of module output. Like the current version, it still produces full-disk H α maps with contours of the identified filaments, location of filament center, filament spine, and barbs, and printed filament ID with chirality (if any is determined). In addition, for each identified filament the module provides a VOEvent entry which contains information about the filament location, outline (chain code), spine skeleton, area and length, number of barbs and in which direction they point, and finally the filament chirality based upon the barb-direction analysis. The module has been deployed since March 2010 and posts results at least twice a day to the HEK. The Big Bear Solar Observatory/New Jersey Institute of Technology and the Global High-Resolution H α Network provide and maintain the H α FTP data archive from which we derive our data. The filament code can be used either for statistical study of filament properties over long periods of time spanning a full solar cycle, or if run in near real-time it can provide useful information for space-weather forecasting. When the code detects the disappearance of a large filament, it can deliver a CME warning. Where the code

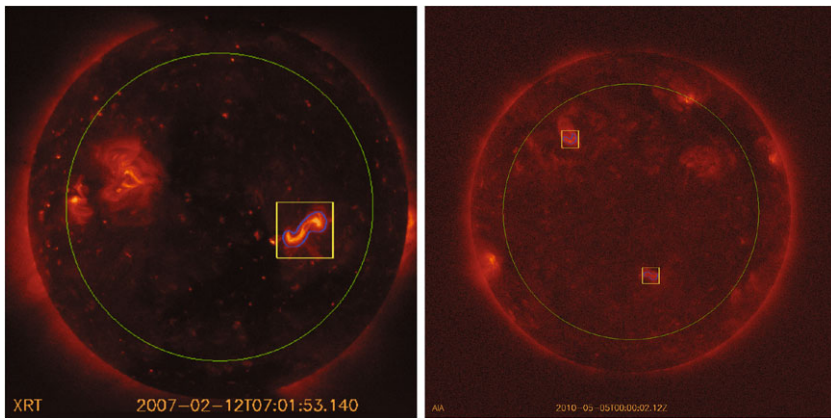


Figure 3 Examples of the results of the sigmoid sniffer module: (left) a right-handed sigmoid is detected in an *Hinode*/XRT full-disk image of 12 February 2007; (right) a left-handed sigmoid (bottom box) detected in a 94 Å image of 5 May 2010 from SDO/AIA. The feature in the top box is a misidentification by the code.

has determined the chirality of the disappeared filament, it provides information about the orientation of the erupting flux-rope and thus the geoeffectiveness of the associated CME.

5. Sigmoids

The solar X-ray corona frequently presents us with bright sigmoidal structures. These *sigmoids* (Rust and Kumar, 1996) are considered to be directly or indirectly related to magnetic structures twisted above their lower energy state (for reviews see Gibson *et al.*, 2006; Green *et al.*, 2007) and, as such, they are broadly recognized as telltale signatures of unstable coronal magnetic-flux systems. This view is supported by the realization that X-ray sigmoids are associated with solar eruptions (Rust and Kumar, 1996). Canfield, Hudson, and McKenzie (1999) found that active regions (ARs) with sigmoidal structures are about 70% more likely to erupt than non-sigmoidal ones, with most ($\approx 84\%$, according to Canfield *et al.*, 2007) triggering solar eruptions within one or two days. Eruptions typically result in the disappearance of an X-ray sigmoid and the launch of a CME that is thought to be associated with the twisted/stressed sigmoidal pre-eruption fields.

Not all CMEs are preceded by sigmoids. Further, not all sigmoids are the same. The most persistent and fainter of them may or may not trigger an eruption, while the brightest, more-transient sigmoids show a much more robust association with eruptions (Gibson *et al.*, 2006). From this viewpoint, although only a fraction of CMEs are associated with sigmoids, it has been accepted that an efficient, automatic, sigmoid recognition offers an unbiased way of identifying short-term progenitors for many CMEs. What is perhaps more important than CME prediction is the physical understanding of what an X-ray sigmoid is and why it occurs. Two model classes are proposed to explain this phenomenon: *i*) The classical view (Rust and Kumar, 1996) treats sigmoids as flux ropes undergoing the helical kink instability. *ii*) More recent views, assisted by numerical simulations, consider upward or downward kinking, inverse bald-patch magnetic geometries, or hyperbolic flux tubes to infer sigmoids with helicity similar or opposite to that of a Titov and Démoulin (1999) flux rope undergoing either the helical kink instability or the torus instability (*e.g.*, Fan and Gibson, 2003, 2004; Kliem, Titov, and Török, 2004; Török and Kliem, 2005, 2007; Gibson *et al.*, 2006;

Green *et al.*, 2007). Regardless of their complicated similarities and differences, all of these models tend to treat sigmoids as *single* flux ropes or current sheets surrounding these flux ropes. Competing with the flux-rope models are models where the flux rope forms before the eruption by a sheared arcade created along a photospheric magnetic PIL (Pneuman, 1983; van Ballegooijen and Martens, 1989, 1990), and models where the sheared arcade gives rise to a flux rope only after the eruption (Antiochos *et al.*, 1994, 1999). Evidence in support of all of these models also exists in recent observations. Of the 107 X-ray sigmoids that Canfield *et al.* (2007) studied using full-resolution observations by the *Soft X-ray Telescope* (SXT; Tsuneta *et al.*, 1992) onboard the *Yohkoh* mission (Ogawara *et al.*, 1991), none appeared to be a single S - or inverse S -shaped loop (see also McKenzie and Canfield, 2008). Instead, they appeared to consist of *multiple* loop patterns arranged such that they are discerned as single sigmoidal structures in limited-resolution images. If these results are confirmed, many sigmoid models may need substantial revision. For the purpose of understanding sigmoids and predicting CMEs, we will use the automatic pattern recognition “sigmoid sniffer” algorithm, first described by LaBonte, Rust, and Bernasconi (2003). A detailed description of the algorithm is given by Bernasconi, Raouafi, and Georgoulis (2010). Full-disk X-ray images are inserted as input to the algorithm, which first uses multiple brightness thresholds to discern persistent bright structures. If one or more candidate sigmoids are identified, the code infers the orientation-angle profile of successive points along the structures’ outline and compares it to the expected profile of the theoretical S -shaped curve. From the fit, one recovers the handedness (forward S [right-handed], or inverse S [left-handed]), the orientation, and the aspect ratio of the sigmoid. The code further issues a CME warning in case it identifies a sigmoid. This information, along with the location and the total size of the sigmoid, will be provided as output of the module. The sigmoid sniffer algorithm has been applied extensively to full-disk X-ray images from *Yohkoh*/SXT and the GOES/*Solar X-Ray Imager* (SXI) instrument, resulting in the automatic identification of numerous sigmoids. For this project, it will be applied to data from the following instruments.

- i) *X-Ray Telescope* (XRT; Golub *et al.*, 2007) on *Hinode* (Kosugi *et al.*, 2007): XRT images have significantly higher spatial resolution (1 arcsec per pixel) than *Yohkoh*/SXT images, so the sigmoid sniffer routinely identifies and delivers crisper, finer sigmoids in *Hinode*/XRT data. The sigmoid sniffer has been implemented to automatically process quicklook *Hinode*/XRT data at SAO. The reported XRT charge-coupled device (CCD) contamination (Narukage *et al.*, 2011, SolarNews, 4 September 2007) covering $< 4\%$ of the XRT field of view is not a major concern for sigmoid detection. The reported irregularities in the *Hinode* downlink have affected the cadence of the incoming XRT images, but even the present cadence of ≈ 1 to a few minutes for the quick-look XRT data is quite sufficient for the purposes of the sigmoid sniffer.
- ii) SDO/AIA: Here we will identify sigmoids primarily in the highest-energy channels (94 Å and 131 Å), which correspond to flaring active regions ($10^{6.8-7.2}$ K), but also in some of the lower energy channels (211 Å and 335 Å), which correspond to the (not necessarily flaring) active-region corona ($10^{6.3-6.4}$ K). SDO/AIA data have just become available through JSOC and/or via the Virtual Solar Observatory (VSO), so final preparations for pipeline deployment are being carried out.

An example of the output is depicted in Figure 3. Two sigmoids are detected in *Hinode*/XRT (left) and SDO/AIA (right) full-disk images. Among other characteristics of the detected features, the sigmoid sniffer determines the sigmoid’s handedness, which is right for the XRT sigmoid and left for the AIA one in the lower box. Noisy data may result in misidentification, such as the feature shown in the upper box of the AIA image. Reducing misidentifications to an absolute minimum is one of our current development goals.

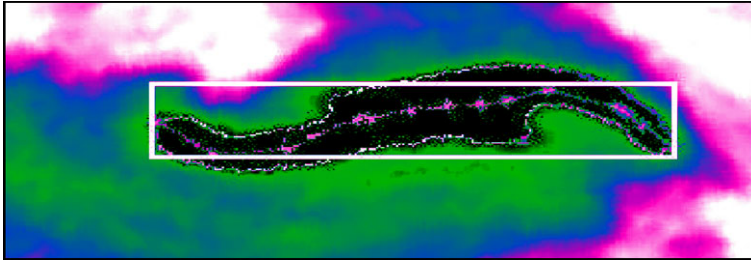


Figure 4 Example of the interaction between the sigmoid sniffer and the filament identification, and the characterization module. The background is a false-color *Yohkoh/SXT* image, with the sigmoid appearing dark. The image box is $51.5'' \times 9.7''$ in size. The white box is drawn by the sigmoid sniffer, while the curved sigmoid outline and the pink sigmoid “spine” stem from the filament-identification code (adapted from Rust and LaBonte, 2005).

The high-resolution *Hinode/XRT* and *SDO/AIA* observations may help resolve the controversy concerning the morphology of sigmoids as single, monolithic, or multiple structures. In addition, it will be very interesting to see whether the sigmoidal morphology persists in different AIA temperature channels. Serious morphological differences in different temperatures would again constrain the modeling of sigmoids, especially the spatial distribution of temperature enhancements along them. For example, is the transient sigmoid caused by upward kinking (peak temperature on top), giving rise to a classical flare cusp (Masuda *et al.*, 1994), or is it mostly due to a downward kinking toward a possible low-lying bald patch (peak temperature at bottom)? Moreover, what is the physical difference between a *complete* and a *partial* sigmoid eruption, where the sigmoid reappears shortly after its disappearance?

The sigmoid sniffer will be cross-correlated with the filament identification and characterization module of Section 4. In an illustrative example, Rust and LaBonte (2005) first applied the sigmoid sniffer to a *Yohkoh/Soft X-Ray Telescope* (SXT) image to identify a sigmoid and then applied the filament code to it to identify its “spine” and infer its aspect ratio more accurately. The example has been reproduced in Figure 4. Further developments of the sigmoid sniffer based on wavelet filtering aim to improve edge detection of solar features and are expected to enhance the overall efficiency of the package.

6. CME Recognition and Tracking

Over the years, coronal mass ejection (CME) recognition and tracking has become increasingly robust, evolving from observer-dependent methods (*e.g.* Gopalswamy *et al.*, 2009) to fully automated detection. A number of tracking algorithms have been developed, each implementing different methods and criteria; techniques include modified Hough transforms (Robbrecht and Berghmans, 2004), threshold segmentation (Olmedo *et al.*, 2008), multiscale filtering (Byrne *et al.*, 2009), adaptive filtering and segmentation (Boursier *et al.*, 2009), and forward-modeling approximations (Boursier, Lamy, and Llebaria, 2009).

Our algorithms are most comparable to those of Boursier *et al.* (2009) and Boursier, Lamy, and Llebaria (2009) in that CME fronts are tracked organically in two dimensions, using a polar coordinate system. This makes it possible to measure both radial and lateral dynamics, enabling us to follow all aspects of the CME expansion.

By monitoring local intensity changes, events are detected as they emerge from behind the *Large Angle and Spectrometric Coronagraph* (LASCO) C2 and C3 coronagraph disks.

After initial identification, our algorithm tracks the front edge of the CME, identified as the outermost strong positive gradient with respect to the background. Lateral expansion is incorporated by measuring intensity increases along adjacent radial coordinates. Each point along the leading front is then tracked across the field of view, recording the two-dimensional structure of the leading edge as a function of time.

To determine the most energetically favorable course of propagation, the measured fronts are used to create a smoothed “time surface.” To do this, the points along each front are splined and assigned a time-equivalent value (usually assuming units of seconds); they are then simultaneously mapped onto a null background and subjected to a recursive-smoothing kernel. The result is a two-dimensional array in which each pixel value corresponds to a point in time. This “time surface” is then subdivided into contours separated by constant time intervals (typically about five minutes) with the original measured fronts also included.

With the data processed in this way, it becomes possible to find two-dimensional velocity trajectories by using Huygens’s method to track points along the CME front smoothly forward in time. As a result, we can calculate plane-of-sky velocity and acceleration at any point along the CME front as it propagates.

In cases where the CME is isolated, meets a sufficient brightness threshold, and has an obvious near-side source region (such as a flare or a dimming region), one can use the central axis of the associated “ice cream cone” model (Fisher and Munro, 1984; Leblanc *et al.*, 2001; Michałek, Gopalswamy, and Yashiro, 2003; Xie, Ofman, and Lawrence, 2004; Xue, Wang, and Dou, 2005) to estimate its three-dimensional propagation direction. From this, we can calculate a three-dimensional velocity vector, a quantity far more valuable than the oft-used plane-of-sky CME vectors. This three-dimensional velocity, when combined with the CME mass, will allow for the calculation of the CME kinetic energy.

Determining the CME mass requires a knowledge of the volume and density of the event. By fitting the “ice cream cone” model to the angular width and leading CME front, one can calculate a rough volume for the CME. Since the plasma is optically thin and any LASCO intensity signal is linear in electron column density, the observed Thomson-scattered intensity of a CME is proportional to its column density brightness. Using a base–difference imaging technique, we can find the excess heliospheric density due to the CME to within a factor of two (DeForest, Plunkett, and Andrews, 2001). We note that Webb (2000) and Bemporad *et al.* (2007) have demonstrated that the mass of a CME is not constant, but increases over time. Therefore, mass and probably kinetic energy are calculated as a dynamic quantity.

The algorithm that we use for extracting these metadata is based on a well-established tracking and measurement code (Wills-Davey, 2006). This code is run on calibrated, background-subtracted *Solar and Heliospheric Observatory* SOHO/LASCO coronagraph data. The tracking module monitors the data stream continuously, in order to catch the appearance of a CME from behind the coronagraph disk. In cases of isolated, well-defined CMEs, metadata outputs are extracted and converted to VOEvent format (*e.g.* CME_AngularWidth and CME_Mass) for dissemination via the HEK.

7. Coronal Dimming Regions

The coronal-dimming detection and metadata-extraction algorithm, together with the flare-detection and flux-emergence algorithms, is designed to provide space-weather alerts in near real time. The relationship between coronal dimmings and CMEs, acknowledged since the 1970s (*e.g.*, Rust and Hildner, 1976), has recently been confirmed statistically by Bewsher,

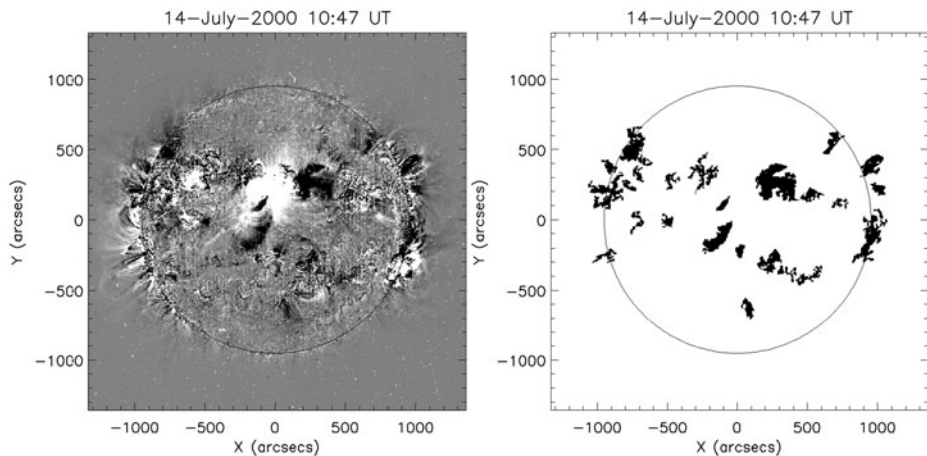


Figure 5 Left: base-difference image of EIT 195 Å data (10:47–09:24 UT), showing the coronal dimmings (black regions) during the 14 July 2000 Bastille Day event. White regions indicate increased intensity compared to the pre-event (base) image at 09:24 UT. Regions that do not show any significant change in intensity with respect to the pre-event image appear gray. Right: the result of the coronal dimming algorithm. Only the dimming regions are extracted.

Harrison, and Brown (2008). Coronal dimmings have been noted as reliable indicators of front-side (halo) CMEs, which can be difficult to detect in white-light coronagraph data. In the absence of a coronagraph (such as with SDO), coronal dimmings can act as an important indicator of the launch of a CME. The algorithm described in detail by Attrill and Wills-Davey (2010) is designed to detect and extract coronal-dimming signatures in the low corona, which are associated with eruptive events, such as CMEs.

This algorithm has two main components: *i*) detection and *ii*) metadata extraction. The detection relies on the analysis of statistical properties of EUV images in which coronal dimmings occur. This method was first introduced by Podladchikova and Berghmans (2005). Our implementation differs in that it is adapted to run on non-differenced images, thus reducing the required computer resources. The metadata extraction is implemented with the detection of a coronal dimming. To extract the coronal dimmings, we adopt the Reinard and Biesecker (2008) threshold of more than one σ below the mean pre-event difference image value. Figure 5 shows an example of our algorithm, in which the coronal dimmings associated with the 2000 Bastille Day event are extracted from the corresponding base-difference data. Once the coronal-dimming regions have been identified, our algorithm analyzes intensity changes over time using non-differenced data. Further metadata outputs of the algorithm include area, lightcurves, location coordinates, volume, and mass of the dimmings. Such outputs can make an important contribution to the study of coronal dimmings and their interplanetary CME counterparts. This algorithm is discussed in more detail by Attrill and Wills-Davey (2010).

8. Jets

An extensive study of polar coronal-hole (CH) X-ray jets was conducted with the *X-Ray Telescope* (XRT) onboard the *Hinode* satellite in early 2007. Coronal-hole jets provide grounds for testing models of reconnection and may prove to be one of the sources of the solar wind. During several days of observation, 44 jets were identified in the 171 Å TRACE

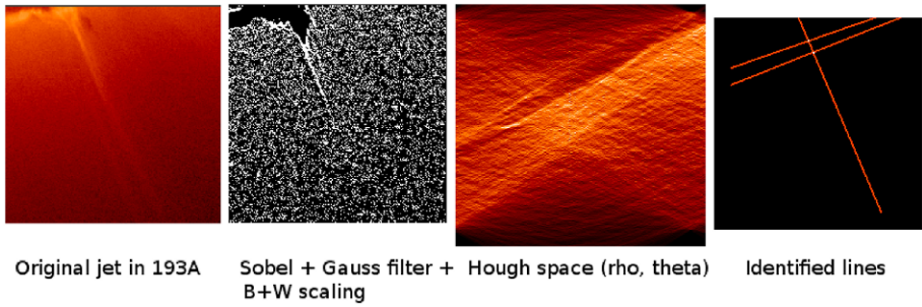


Figure 6 Sequence of images illustrating the steps taken by the jet-finder algorithm, as described in the text. The AIA 193 Å image on the left was taken on 16 June 2010, at 00:04:08 UT. The cutout of the full-disk image represents a location on the limb, inside the South pole coronal hole, just west of the pole.

images. AIA, similar to TRACE in resolution and wavelengths, will have full-Sun field of view and consistent cadence. Therefore, AIA will be able to observe a large number of jets. The coronal jets found in coronal holes are the easiest to identify. Here, the coronal jet (CJ) stands for a jet inside a coronal hole.

The CJ detection calculations will take place at SAO, once triggered by both a bright point (BP) and CH identification by our system (see Sections 11 and 12). The CH boundary is necessary to ensure that the BP is inside the CH. After identifying the boundary of the CH, the BP finder runs on every other image, and the pixels within BPs are identified. The CJ detection and parameter-determination algorithms work on data cubes covering a box enclosing the BP and extending forward in time. Our methods for determining the CJ parameters are described in detail by Savcheva *et al.* (2007). The detection algorithm has been tested on XRT images. The algorithm has been proven to work in the general case for low irregularity in the background, as is the case with XRT images. Further refinements handle the more dynamic background in corresponding EUV images that are taken with AIA. TRACE images have been used to implement this step. Figure 6 illustrates the sequence of steps in the algorithm behind the jet finder part of the jet module for AIA data. The first plot shows the original jet as seen in AIA 193 Å. The second one shows the same portion of the image after a Sobel and a Gaussian filter ($\sigma = 15$ pixels) and linear scaling have been applied – the resultant image consisting only of pixels with values 0 or 1. The third image shows the part of the $\rho - \theta$ space that the scaled image spans after the linear Hough transform has been applied to it. The brightest spot in this image represents the most prominent line in the image in terms of distance from the origin and inclination. The last image shows a reverse Hough transform of the three brightest spots in the Hough space, which is represented by three lines: one for the jet and two for the limb.

The final result from the algorithm is an entry into the HEK containing the following VO Parameters (see http://www.lmsal.com/helio-informatics/hpkb/VOEvent_Spec.html): a jet ID and the corresponding BP ID, the position of the origin of the jet in heliospheric coordinates, duration, length, width, line-of-sight velocity, and BP size. The inclination with respect to the N–S direction and transverse velocity are external parameters that are saved in a text file with the other parameters to create a metadata cube.

9. Oscillations

The study of large data sets (both in area and duration) is needed to improve the statistics of current results regarding coronal oscillations as well as localized, short-lived events. Certain

dynamical events such as CMEs, flares, and coronal dimmings are known to be triggers for wave processes, and hence programs for wave analysis will be run on data sets associated with such events, which comprise only a small fraction of the observations ($< 0.1\%$). Other potential wave triggers, such as flux emergence, will also be studied. The codes have been optimized for robustness and accuracy, even at the expense of speed. Wavelet analysis plays an important role in this process, as it is capable of identifying periodicity in a given data set and can also provide both spatial and time localization (De Moortel and Hood, 2000; De Moortel, Hood, and Ireland, 2002; Ireland and De Moortel, 2002). This ability to localize a periodicity in time and space within a given data set can be exploited to “train” the wavelet code to identify certain events.

The adopted code was originally developed by McAteer *et al.* (2004) and De Moortel and McAteer (2004). For a given data cube, it performs a wavelet analysis of the time series in each pixel and outputs the periodicities that are present, the duration (number of cycles) of the detected periodicities, and their start and end times. These data will be entered into the VOEvent catalog. This information allows the team to optimize the wavelet code to detect a variety of events and features found in the solar atmosphere. Using such an automated detection technique on specified observations (*i.e.* observations associated with CMEs, flares, and coronal dimming events) facilitates a thorough investigation into the link between dynamical (explosive) events and observed solar oscillations. Indeed, De Moortel, Munday, and Hood (2004) showed that various wavelet parameters can be altered to favor either the time or frequency resolution. Additionally, by making a detailed study of the wavelet transform of the observed oscillations, it is possible to infer some coronal-plasma properties, as demonstrated by De Moortel and Hood (2000) and De Moortel, Hood, and Ireland (2002).

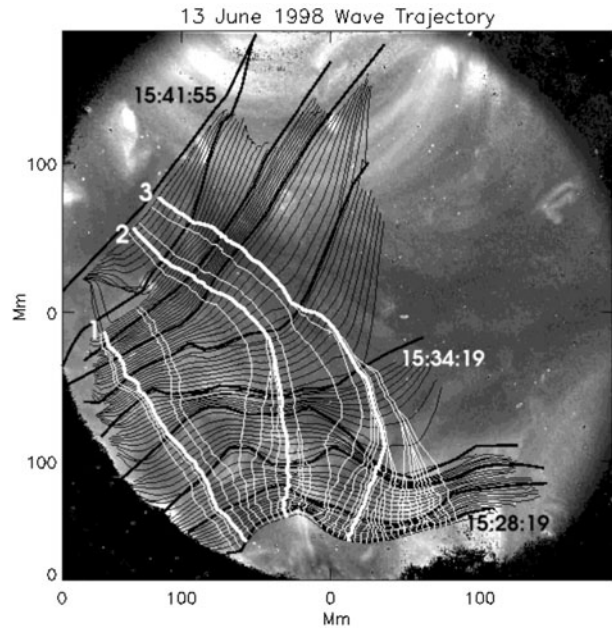
We are currently exploring a variety of options to improve performance of the code, for example, by binning (which does not seem to affect wave detection). The code will be made available to users for searches with user-specified optimization of search parameters.

10. EIT Wave Tracking

As part of the SDO computer vision project, EIT waves will be tracked and their metadata extracted using algorithms based on the work of Wills-Davey (2006) and Podladchikova and Berghmans (2005). Both sources rely on calibrated, derotated EUV data. Wills-Davey (2006) derives her output from TRACE 171 Å and 195 Å percentage base – difference images, while Podladchikova and Berghmans (2005) use data from the SOHO/*Extreme ultraviolet Imaging Telescope* (EIT) 195 Å “CME Watch” and *Solar Terrestrial Relations Observatory* (STEREO)/EUVI 171 Å and 195 Å data. The Wills-Davey (2006) method assumes that the EIT wave is a coherent front of intensity enhancement and measures intensity increases across the field of view. Podladchikova and Berghmans (2005) instead find EIT waves by measuring the first four moments of a given full-disk EUV image; in particular, the global kurtosis offers a way to find EIT waves that may displace structures while producing only dim fronts. This new algorithm takes aspects of both of these methods, enabling us to find a larger range of events.

Observations suggest that both EIT waves and coronal-dimming regions are highly correlated with CMEs, and many models assume that these events are a product of CME initiation, with EIT waves and dimming regions originating cospatially. Because EIT waves are tracked using a polar coordinate system, the origin of these coordinates is determined by cross-referencing with the coronal-dimming region tracking module, and finding the starting location of the dimming.

Figure 7 Automated tracking of an EIT wave observed on 13 June 1998.



Two methods are used for finding the EIT wave location. No assumptions are made about the wave's propagation characteristics, other than that it will propagate roughly away from the eruptive region. In cases where a bright-front is observed, the position of the intensity maximum is determined from one frame to the next in the polar coordinate frame. Figure 7 shows an example of this automated bright-front tracking. This analysis is combined with measurements of local kurtosis across overlapping 25×25 pixel sections. This use of local kurtosis allows us to track subtle motion at a resolution of order $10''$. Since EIT waves are global structures, such a resolution is more than sufficient for determining structural displacement due to coherent wave propagation. Velocity is calculated using the same Huygens plotting technique discussed for the CME tracker (Section 6).

When intensity measurements show the existence of a bright front, we record the intensity enhancements due to the wave as a function of time. Intensity cross sections can either be calculated as a three-dimensional contoured array, or can be considered as “slices” along the Huygens-plotted velocity trajectories. As EIT waves are defined as propagating intensity increases, only the positive values from the difference images are kept. These data will also be used to find the center-of-front amplitude as a function of time.

By combining intensity enhancement measurements from several overlapping EUV passbands, it is possible to measure the density increase above the pre-event background due to the wave front. Intensity measurements also allow us to calculate the entrained energy of a given wave event as follows: $E = \int d^3V (P_0 \Delta n_e^{2\gamma} n_e^{-2\gamma})$, where E is the entrained energy, V is the volume of interest, P_0 is the equilibrium pressure, n_e is the electron density, and $\gamma = 5/3$.

All but the strongest EIT waves, while easy to pick out by eye in running-difference movies, are nonetheless extremely difficult to find and track *a priori* because noise tends to dominate the difference images. Additionally, studies by Rachmeler and Wills-Davey (2005) show that false-positive detections may be more likely than previously expected. Because of this possibility of false positives, as well as the excessive resource requirements for wave

detection from the full data stream, the wave detection program will be conducted *post facto*, triggered by other events detected by our system, particularly flares and coronal dimming regions. Output parameters will be input into the HEK, using the appropriate VOEvent format. With these triggers we expect to observe 100–300 events year⁻¹.

11. Detection and Analysis of Active Regions and Coronal Holes

The *Spatial Possibilistic Clustering Algorithm* (SPoCA) we have developed produces a segmentation of EUV solar images into regions that we call “classes” corresponding to active regions (AR), coronal holes (CH), and the quiet Sun (QS); see Barra *et al.* (2005), Barra, Delouille, and Hochedez (2008), Barra *et al.* (2009). Other segmentation methods have been proposed; an overview is given in Barra *et al.* (2009). We have selected SPoCA because of the maturity and flexibility of the program. SPoCA uses a multichannel, fuzzy-logic, clustering procedure. It has been applied successfully to a series of EIT image pairs (171 and 195 Å) spanning almost a full solar cycle; see Barra *et al.* (2009). The classes are determined by minimization of intra-class variance. The method is generic and therefore portable to other instruments, and in particular to SDO/AIA. SPoCA involves a preprocessing by which the limb brightness discontinuity is attenuated. It can take transformed EUV images as input, such as differential emission measure (DEM) maps obtained from AIA images (using software supplied by AIA investigators). This might help address the problem of line-of-sight confusion.

The level-1 product of the procedure is a set of maps giving the membership value to each class. Several higher-level products can be generated from these maps of memberships:

- A segmentation map, attributing a class to the pixel according to its highest membership value;
- A probability-density function giving, for each pixel intensity value, the probability that this pixel belong to a particular class;
- Computation of quantities such as area, integrated intensity, and first statistical moments for each class.

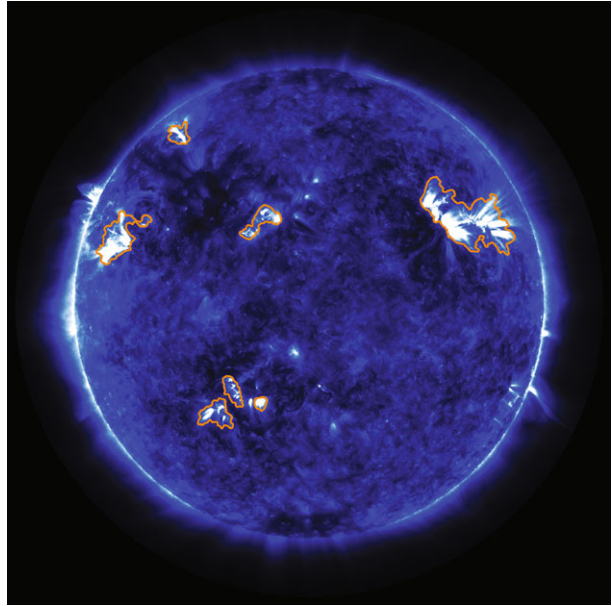
From the maps of, *e.g.* ARs, connected AR pixels are then gathered by means of a region-growing technique; see Figure 8 for an example of our results with AIA data. It provides the instantaneous location of the barycenter, the area, the coordinates of the bounding box, and a mask for each AR.

These elements are computed over time and handled as dynamical quantities. The current tracking method uses an optical-flow algorithm to locate the barycenter of the AR in the next fuzzy map, where the same region-growing technique updates the parameters of the AR being tracked. A starting date for an AR is defined when a new set of connected pixels is identified. An AR “end date” is recorded either when the tracking algorithm can no longer find a connected set, or when the AR disappears over the west limb. The algorithm also handles merging of ARs.

Fuzzy CH maps can be treated exactly as the AR maps, producing area, location of barycenter, and location of boundary for a connected element of the CH map. However, filament channels seen in coronal EUV passbands are often erroneously classified as “coronal holes” in the segmentation. Using H α images from ground-based observatories, the filament-detection module (see Section 4) separates filaments from the CHs.

On AIA images we will identify AR, QS, and CH using our multichannel segmentation algorithm on pairs of 171 and 193 Å images, 211 and 335 Å images, as well as on 94 Å

Figure 8 AIA 171 Å image from 29 April 2010, 00:00:42 UT, together with overlays of segmented ARs.



images. When similar morphological information is available in two bandpasses, a multi-channel approach provides a more robust segmentation than a mono-channel one.

12. X-ray Bright Points

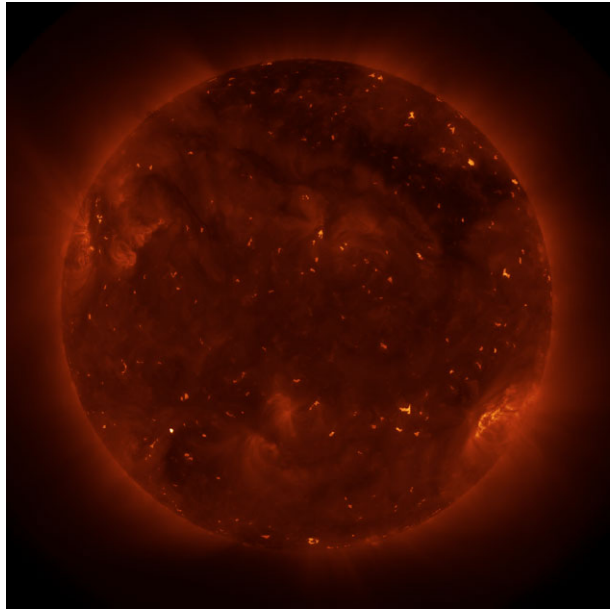
AIA delivers a set of full-frame images every ten seconds in each of its seven EUV passbands. The bright point (BP) detection algorithm will run on the three coolest coronal passbands at wavelengths 171 Å, 193 Å, and 211 Å.

We use a modified version of the BP finder developed by McIntosh and Gurman (2004) to successfully identify and track bright points in SOHO/EIT 171 Å, 193 Å, and 284 Å over an entire solar cycle. The algorithm, written in IDL, uses full-Sun synoptic images to define areas of interest based on levels of intensity above the local background, and number of contiguous pixels that define shape and size (see Figure 9). This process is described in detail in Davey and McIntosh (2007); see also McIntosh and Gurman (2005), McIntosh *et al.* (2009). The bright-point detection algorithm works by applying an $N \times N$ boxcar smooth function to determine the local background intensity. The noise level (σ_{noise}) is also estimated. The smoothed background is subtracted from the original image, and intensity enhancements (BP candidates) are classified by their separation from the background in units of σ_{noise} , with some restrictions on size and shape.

We are implementing several steps to improve the existing code:

- We have included an upper and lower limit on the size of the BPs, weighted by limb angle.
- The McIntosh code can pick up strings of “BPs” which are actually the brighter portions of longer ribbon/loop-like features. After initial passes of BP detection, we slightly reduce the noise threshold to see if spatially close BPs merge. If so, the underlying ribbon structure is then reassigned.

Figure 9 Full-disk 193 Å AIA image, scaled logarithmically. The pixels found by the BP finder module have been identified and highlighted in the original image. The image was taken on 29 July 2010 at 15:27:56 UT.



- The McIntosh code can pick up patches of locally enhanced quiet Sun (QS) pixels, often with complex boundaries, which are clearly not BPs. We have placed an additional constraint on the ratio of the two major axes of the BP and its perimeter/area ratio to prevent the detection of such QS conglomerations. This also eliminates long ribbon-like structures which are typically remnant loops of decaying active regions.

Once the bright points are defined by heliographic coordinates and time of observation, obtaining statistical information is straightforward. The total number, intensity (mean count rate), size, perimeter, major/minor axes, and area of bright points are determined for each of the regions in each bandpass. To determine lifetimes, the feature-finding program tracks the intensity-weighted center of mass and area in any of the AIA passbands, with the position compared with projected rotation rates as a function of latitude. The lifetime of BPs is found from their appearance and disappearance, or from their rotating on and off the disk. Figure 9 shows an early example of the application of the BP-finder algorithm to a full-disk 193 Å AIA image.

The output of the region finder, made publicly available through the HEK and a full catalog at SAO, includes time of observation, heliospheric coordinates, total area, integrated intensity for each wavelength, and environment associated with each bright point (QS, AR, CH). Also associated with a select number of BPs are 128×128 pixel images of the region that can be made into QuickTime movies, accessible, *e.g.* through the VSO interface.

The data-processing time for finding and tracking bright points in three full-disk EIT 171 Å images is four to five seconds using a 2×3 GHz Dual-Core Intel Xeon Macintosh HD. The actual processing time will depend on the AIA image-preparation software. Detection parameters (background, dark subtraction, sigma enhancement in each filter) will be adjusted after in-flight calibrations when the AIA image preparation software becomes available. Algorithms that correlate the feature-finding database with other databases will be developed within the first year after launch depending on when the level-2 algorithms are implemented.

12.1. Correlation with Other Databases

The heliographic coordinates determined from the feature-detection program will be correlated to other databases within the pipeline to create other level-3 data products. The magnetic-flux tracker SWAMIS identifies and tracks magnetic-flux concentrations above a particular threshold (see Section 13). Using coordinates of bright points identified in the AIA 171 Å channel, the magnetic-flux database is cross-referenced for footpoints within an area of the bright-point coordinates. The magnetic flux for each of the bright points can then be calculated by summing over the total area in each of the footpoints. The lifetime and unsigned magnetic flux of the region are then appended to the bright-point catalog.

12.2. Correlation with Jets

The algorithm for jet detection uses the location of coronal holes and bright points from the feature-finding program (see Section 8). This algorithm is run every few hours. After jets are detected, the bright-point catalog is updated with a keyword [BPJ] associating the bright-point number to the jet number. The algorithms that correlate the feature-finder database with other databases will be developed after launch, and can be executed retroactively independent of the near real-time pipeline.

Because the basic McIntosh code uses levels of σ_{noise} above background to identify features, with a few modifications the program can also be used to define and track not only bright points, but also active regions and coronal holes. These features will then be used to cross-check the features identified with the segmentation algorithm (see Section 11).

13. Magnetic Feature Tracking and Sunspots

We will provide comprehensive data on flux emergence, interaction, and cancellation over the whole solar disk. Two main types of magnetic-tracking data are available: feature-level data, which represent the motion of resolved line-of-sight magnetic fields over the entire surface of the Sun within 45° of the sub-Earth point; and large-scale emergence data, which identify and highlight emerging-flux regions associated with ephemeral active regions on the supergranular scale and above.

The feature-level data provide comprehensive motion and history information on small magnetic details. They enable correlative statistical studies of flux origin, interaction with coronal features, and diffusion over the solar surface. For example, it will be possible to extract and search on the simplified magnetic geometry, and hence number of coronal null points and field complexity, in the neighborhood of any feature or event in the entire SDO data set. Further, every emergence event observed by the SDO/*Helioseismic and Magnetic Imager* (HMI) over the entire mission life will be cataloged and sorted by position, time, and event size and strength, permitting statistical surveys of the interaction of emerging flux with existing magnetic structures in the chromosphere and corona.

Magnetic feature tracking is easy to prototype but difficult to perform reproducibly and reliably. The pipeline uses a variant of the SWAMIS code described by DeForest *et al.* (2007). For feature-level tracking, we use the “downhill” quasi-watershed method that they discuss; for large-scale emergence tracking we are prototyping both the multiresolution method pioneered by Hagenaar and Cheung (2008) and a post-track feature clustering method. Large-scale flux emergence is challenging to detect with a simple feature finder operating at high spatial resolution; this has led to studies of several methods, including post-track clustering

analysis of feature centers, unsigned-flux counting in large image “macropixels” (Hagenaar and Cheung, 2008), and multiple passes through the image data at different spatial resolutions. We identify large-scale flux emergence using the multiresolution approach, in which the data are tracked at multiple spatial resolutions. Larger-scale emergences are identified in degraded copies of the original data. The multiresolution technique is not expected to significantly impact computing resources, as the degraded copies of the data require far less memory and CPU power than the full-resolution stream.

SWAMIS has been used to study flux the emergence patterns with SOHO/MDI (*e.g.*, Lamb *et al.*, 2008) and the *Hinode/Solar Optical Telescope (SOT)–Narrowband Filter Imager (NFI)* (see, *e.g.*, DeForest *et al.*, 2008; Lamb *et al.*, 2007). Our basic tracking cadence is six minutes; lifetime studies using SOHO/MDI data show that this is fast enough to capture evolution at the ≈ 1 arcsec resolution of SDO/HMI. Simple feature location and evolution data are the most basic products of every feature recognition code. In addition, we provide origin and demise information on each feature detected. SWAMIS classifies birth events into five categories: Isolated Appearances (a large fraction of new flux features appear in isolation, far from any other feature), as described by Lamb *et al.* (2008), Fragmentations (in which a feature “calves” from a like-sign feature that is undergoing shredding), Emergences (in which a new feature appears in a manner that approximately conserves flux, either due to a new opposing feature appearing nearby or to growth of an existing nearby opposing feature), Errors (which do not appear to conserve flux), and Complex events, where more than two features appear to be interacting directly. Less than 1% of events greater than $10\times$ the detection threshold result in Error events. In each case, SWAMIS also identifies the associated opposing region. Cancellation is classified using the same scheme, reversed in time, to yield Disappearances, Mergers, Cancellations, Errors, and Complex cancellations.

Identification of every feature-interaction event is crucial for the new field of “event-selected ensemble imaging” (ESEI), which allows deep-field study well below the noise floor of an individual observation. ESEI, which uses a combination of magnetic and other feature detections, is a unique and powerful tool to distinguish models of small-scale activity that cannot be resolved any other way.

The exact SWAMIS code used in the SDO pipeline will be made available for download, inspection, and local modification as free software.

14. Polarity Inversion Line Mapping

Identifying the location of *Polarity Inversion Lines (PILs)*—often also called neutral lines—can be of great importance for phenomenological and theoretical studies. Historically, neutral lines in active regions have been useful tools for predicting the locations of flares and CMEs (Falconer, Moore, and Gary, 2002). They also can be used to map out coronal structures (McIntosh, 1994) and are associated with filaments and filament channels (*e.g.*, Martin, Bilimoria, and Tracadas, 1994; Chae *et al.*, 2001).

Until recently, little work appears to have been done to develop automated PIL mapping. Inversion lines have typically either been drawn by hand or determined by contouring algorithms within larger programming packages. The use of a PIL mapping algorithm not only allows for user-independent inclusion in a pipeline, but standardizes the mapping with prespecified, explicit resolution scales.

We have successfully developed a code that identifies PILs based on a well-established code previously developed for National Solar Observatory (NSO)/*Kitt Peak Vacuum Telescope (KPVT)* magnetogram data (Jones, 2004). This method invokes techniques such as

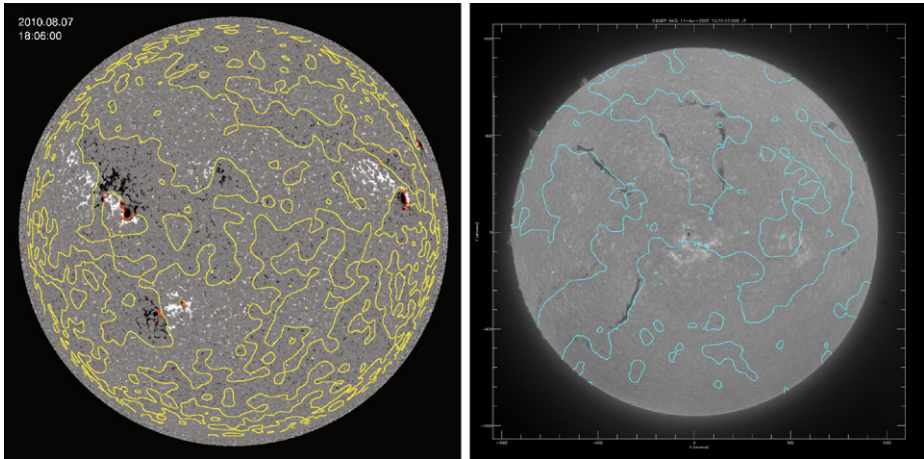


Figure 10 Left: Polarity-inversion contours (in yellow), determined by the PIL module, overlaid on the source SDO/HMI magnetogram. Red dots indicate flare-prone regions of high magnetic gradient. Right: Polarity-inversion lines in blue from a SOLIS magnetogram, calculated with the same code, overlaid on a nearly simultaneous $H\alpha$ image from BBSO.

edge detection, smoothing, and block averaging to determine the appropriate scale at which to measure the neutral lines. The main difference, and the one that makes this code more desirable and sophisticated, is a multiple-smoothing method. There are three ranges of flux corresponding with the values of the pixels, each of which receives a different amount of smoothing. If the flux is above the upper threshold, little smoothing is applied; if it is below the lower threshold, a lot of smoothing is applied. For all other fluxes between the upper and lower thresholds, a medium amount of smoothing is applied. See Figure 10 (left) for a recent example of PIL code output overlaid on an HMI magnetogram. Areas of high magnetic gradients are identified and indicated in the figure in red. Projection has been taken into account.

To verify our results early on, we overlaid a PIL map determined from a *Synoptic Optical Long-term Investigations of the Sun* (SOLIS) image on an $H\alpha$ image from BBSO (Figure 10, right). We tested on SOLIS magnetograms because their resolution is closest to that of HMI. Parameters in the PIL routine may be varied in accordance with the user requirements, allowing mapping to be done on a variety of length scales. The output from this code is converted to VOEvent format as a mask data type for the HEK.

15. Nonlinear Force-free Field Extrapolations

The solar magnetic field is key to understanding the physical processes in the solar atmosphere. Unfortunately, we can measure the magnetic field vector routinely with high accuracy only in the photosphere. These measurements are extrapolated into the corona under the assumption that the field is force free, because the magnetic pressure is several orders of magnitudes higher than the plasma pressure. We have to solve the equations $(\nabla \times \mathbf{B}) \times \mathbf{B} = \mathbf{0}$ and $\nabla \cdot \mathbf{B} = 0$.

We solve by minimizing the functional proposed in Wheatland, Sturrock, and Roumeliotis (2000) and effectively encoded in Cartesian and spherical geometry by Wiegelmann

(2004, 2007).

$$L = \int_V [B^{-2} |(\nabla \times \mathbf{B}) \times \mathbf{B}|^2 + |\nabla \cdot \mathbf{B}|^2] d^3x, \quad (1)$$

where an observed preprocessed vector magnetogram specifies the photospheric boundary conditions. Preprocessing of the measured photospheric vector magnetograms is necessary, because nonmagnetic forces such as plasma pressure and gravity are present in the photosphere, and consequently the measured magnetic-field data are not consistent with force-free consistency criteria as defined in Aly (1989). We developed a minimization procedure that uses the measured photospheric field vectors as input to approximate a more chromospheric-like field (Wiegelmann, Inhester, and Sakurai, 2006). The procedure includes force-free consistency integrals and spatial smoothing. Direct chromospheric observations can be taken into account for an improved match to the field direction, as inferred from H α fibrils by Wiegelmann *et al.* (2008). Recently, the preprocessing routine has also been implemented and tested in spherical geometry by Tadesse, Wiegelmann, and Inhester (2009).

A recent comparison of different nonlinear force-free extrapolation codes has revealed that minimizing the functional (1) is the fastest and most accurate currently available method (Schrijver *et al.*, 2006; Metcalf *et al.*, 2008).

The Cartesian version of the code has been applied to vector magnetographs from spaceborne data (*e.g.*, from *Hinode/SOT* by Jing *et al.* (2008)) and ground-based data (*e.g.*, from SOLIS in Thalmann, Wiegelmann, and Raouafi (2008)). The computed data cubes have been analyzed regarding the content of free magnetic energy and the magnetic topology. An ongoing effort is the application of the spherical code to SOLIS data. We use the large and high-resolution HMI field-of-view (FOV) vector-magnetograms to produce estimates several times per day for the free magnetic energy for the ARs on disk, and produce 3D magnetic-field maps outlining the general coronal magnetic-field topology. These data will be cross-referenced with the AR and CH data obtained through other methods (see Section 11).

Further complications in using current vector magnetograms (*e.g.*, from *Hinode/SOT*) are a limited FOV, missing data, and high noise in the transverse B -field measurements (DeRosa *et al.*, 2009). For meaningful extrapolations, it is important to use large and high-resolution FOV vector-magnetograms—as provided by HMI—and to deal with measurement errors and nonmagnetic forces by preprocessing. An updated version of the nonlinear force-free field (NLFFF) code takes the measurement errors into account by adding another term to Equation (1) with a Lagrangian multiplier; see Wiegelmann and Inhester (2010) for details.

For reliable estimates of the free magnetic energy and topology from the extrapolated coronal magnetic-field model, one should validate the model field by additional coronal observations, *e.g.* compare projected field lines with AIA images. This comparison can be done with a newly developed tool to extract coronal loops from EUV images (Inhester, Feng, and Wiegelmann, 2008). This tool has been applied so far to images taken from the STEREO/SECCHI/EUVI instruments, and an application to AIA is straightforward. Wiegelmann, Inhester, and Feng (2009) proposed an algorithm [an additional term in Equation (1)] to incorporate extracted coronal information into the computation of nonlinear force-free coronal magnetic fields, and we will study the use of this tool for our purpose here.

16. Trainable Feature Recognition and Retrieval

Humans have an amazing generic feature-recognition ability that has been hard to match in computers. For example, a solar scientist can instruct student in less than an hour to

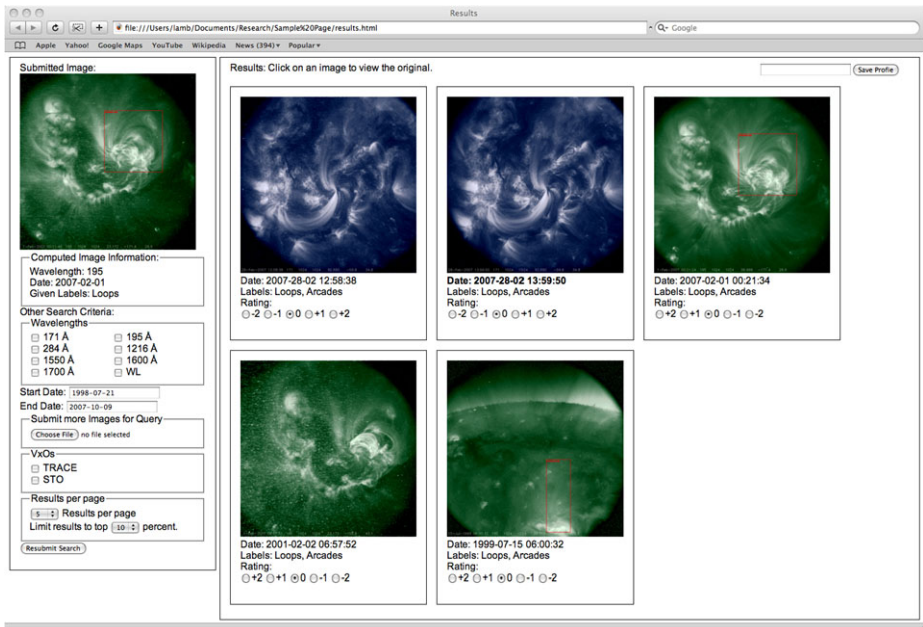


Figure 11 An example of the querying interface showing sample results with the option to rate the returned images and resubmit the image-based query.

recognize sunspots, filaments, loops, and arcades in solar imagery, and the student can then easily produce a catalog of these features from a given set of images. However, a computer feature-finding algorithm takes months to years to develop, and the development must be repeated almost from scratch for every new feature.

Motivated by the successful development and implementation of a generic, more human-like, feature detection method for mammography (*e.g.*, Yang *et al.*, 2007), we are in the process of creating an automated solar-feature-retrieval system that is generic in nature; *i.e.*, the software can detect features of any kind, and rank the returned images based on their similarity to an image provided by the user (see Figure 11 for a sample of the user interface). Rather than developing a new task-specific application to identify each separate feature, our generic feature-recognition software can detect and catalog a wide range of solar features, even involving serendipitous discovery. This work will benefit the solar community in three ways. It will deliver *i)* a benchmark data set that includes balanced representations of common types of phenomena and that can be used to perform comparative evaluation of different image recognition systems, which will be made freely available. (Our benchmark data set for TRACE images is available at <http://www.cs.montana.edu/angryk/SDO/data/>.) It will also deliver *ii)* catalogs for common types of phenomena with database indices, speeding up data search and retrieval, and *iii)* a content-based image retrieval system with a ranking mechanism that returns images based on their similarity to the image provided as a query, and that can learn from users' feedback.

16.1. Method

There are two steps to our method. The first is implemented at the end of the metadata pipeline. The second is totally separate from it and can be performed by any user at any

Table 3 Texture parameters used for analyzing TRACE images in Lamb (2008).

Name	Equation
Mean	$m = \frac{1}{L} \sum_{i=0}^{L-1} z_i$
Standard deviation	$\sigma = \sqrt{\frac{1}{L} \sum_{i=0}^{L-1} (z_i - m)^2}$
Third moment	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$
Fourth moment	$\mu_4 = \sum_{i=0}^{L-1} (z_i - m)^4 p(z_i)$
Relative smoothness	$R = 1 - \frac{1}{1 + \sigma^2(z)}$
Entropy	$E = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$
Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$

time, even on a laptop, without requiring SDO JSOC resources. Step one consists of calculating for each AIA image the texture parameters, which are then stored in the image texture catalog. Specifically, each AIA image is subdivided in 1024 128×128 sections, and for each of those a set of texture parameters such as entropy, average intensity, standard deviation, kurtosis, *etc.* is calculated. The parameters that we have used for analyzing TRACE images are given in Table 3. Assuming that we derive 16 parameters per image section, each 16 bits, the information compression from AIA image to catalog entry is roughly a factor of 1000. Therefore, the size of the AIA image-texture catalog will not be prohibitive. For the SDO mission we are investigating the option of even further compressing our catalog entries by discretization of our texture parameters (Banda and Angryk, 2009). We estimate that the calculation of statistics for a single image entry into the catalog requires one second of computing time for the whole grid-segmented image.

Suppose now that a user wants to build a catalog of loop arcades by teaching the algorithm how to detect them. The user then downloads a selection, say several dozens, of AIA full-disk images. In those images the user identifies arcades via the simple point-and-click interface shown in Figure 11; there is no need to identify all arcades on the disk. The program then calculates the texture parameters for each of the image segments containing the identified arcades. Using these parameters as a feature definition, the entire AIA image-texture catalog can be quickly searched to filter out irrelevant images and then fine-tune the search for image segments with similar parameters. These segments will contain arcades if the image-texture parameters are adequate for the type of image one is analyzing.

Our work on TRACE images (Lamb, 2008; Lamb, Angryk, and Martens, 2008; Banda and Angryk, 2009) has focused on the crucial step of determining the correct texture parameters: those that have good properties, first from the perspective of computational costs to keep up with our pipeline, and second from the perspective of distinguishing between different types of solar phenomena. Table 4 summarizes the results from Lamb (2008). Here the receiver operating characteristic (ROC) curve plots the true-positive rate on the y-axis and the false-positive rate on the x-axis. The area under the curve represents the overall accuracy of the classifier. The closer the area under the curve is to 1.0, the more accurate the classifier; at 0.5 the classifier would not be any better than randomly picking a class from an evenly distributed set of phenomena. Random under sampling (RUS) and random over sampling (ROS) are different sampling methods to treat data sets with unbalanced numbers of different phenomena (*e.g.*, relatively common images of quiet Sun and coronal loops, *versus* relatively rare occurrences of flares and filaments). The column headers represent different classification techniques. (The method C4.5 performs entropy-based classification, SVM

Table 4 Average area under an ROC curve using two different sampling techniques (RUS and ROS), and two different classification algorithms (C4.5 and SVM) for different types of solar phenomena (Lamb, Angryk, and Martens, 2008).

Phenomenon	C4.5		SVM		AdaBoost C4.5		AdaBoost SVM	
	RUS	ROS	RUS	ROS	RUS	ROS	RUS	ROS
Quiet Sun	0.795	0.872	0.940	0.923	0.920	0.912	0.939	0.920
Coronal loop	0.897	0.905	0.932	0.922	0.911	0.917	0.912	0.912
Sunspot	0.890	0.917	0.922	0.958	0.901	0.932	0.944	0.943
Filament	0.838	0.960	0.832	0.848	0.875	0.783	0.898	0.897
Flare	0.977	0.970	0.988	0.980	0.977	0.967	0.977	0.976

stands for support vector machines, and AdaBoost is a commonly used boosting technique that emphasizes misclassified samples during classifier training.)

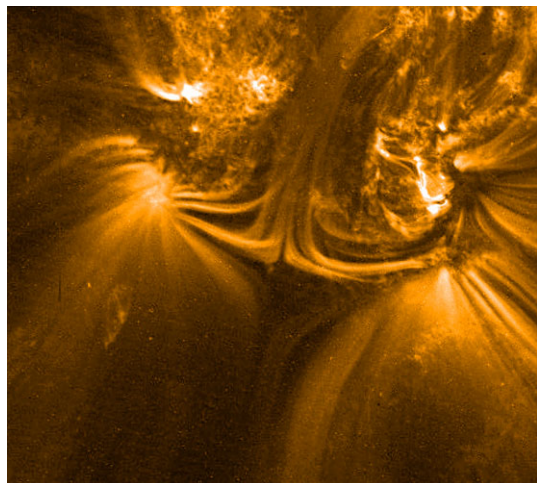
We continue to evaluate our texture parameters and improve our classifiers using TRACE images. The TRACE results will carry over easily to AIA because their imagery is similar. We emphasize again that step two is completely separate from the data pipeline, and that the AIA data archive need not be accessed in its entirety; only a small number of images must be downloaded to provide a training sample to our software. Since the algorithm is further trainable, a user can fine-tune the selection, *e.g.* to distinguish between left-skewed and right-skewed arcades.

Our current list of solar features for which we intend to generate VOEvent entries (time and location at a minimum), using this method, includes: cusps, arcades, null-geometries, flare ribbons (starting with the flare catalog), keyholes (previously detected by *Skylab* and *Yohkoh/SXT*), circular filaments (starting with the filament catalog), faculae, pores, surges, arch filaments, δ -spots (from magnetograms), plumes, and anemones (also called rosettes).

16.2. Discovery of New Features

A promising advantage of generic feature-recognition methods over task-specific ones is the potential for rapid discovery and cataloging, even serendipitously, of new features. An

Figure 12 TRACE 171 Å AR image showing a loop geometry that is consistent with the presence of a magnetic null. Field is 3.8×4.2 arcmins.



example is given in Figure 12, a TRACE image of an active region having a special type of loop geometry. The user can start with this image to try to find similar ones in the TRACE or AIA catalog. Not many such images have been observed with TRACE, probably because they are rare, and for an instrument such as AIA, searching the database to find more of such images is a daunting task. Our image-texture characterization makes such a search possible and even rather simple without the user needing to know how to specify the parameters of the search. The scientific interest of Figure 12 is clear: this is what one would expect to see in the neighborhood of a magnetic null, making it a good candidate for inclusion in a study of such topologies.

This then points to a more general application for the image-texture catalog. A user may see a number of images that look very similar and which pique interest without even being able to define exactly what is exceptional about these images (Figure 12 would be an example for someone with limited knowledge of MHD). Our feature-recognition method can then be used to find similar images without an excessive effort, even though the users do not even know exactly what they are looking for!

The image texture catalog, together with a manual and supporting software, will be made available online as a complement to the VOEvent catalog.

17. A Pipeline for the Generation of Feature and Event Metadata

17.1. Computing Facilities at LMSAL

LMSAL has provided a computer infrastructure to enable the running of appropriate feature and event modules. This includes a cluster of Linux servers consisting of either four- or eight-core 2.2 GHz processors with up to 16 GB of memory per server. LMSAL has available > 100 TB of storage which is Network File System (NFS) mounted to these servers. Data transfer between Stanford and LMSAL is achieved by means of a 10 Gbps fiber link, giving LMSAL almost instant access to the SDO data. As part of their commitment to implementing the *Heliophysics Events Knowledgebase* (HEK: Hurlburt *et al.*, 2010), LMSAL has provided for interacting with the *Data Record Management System/Storage Unit Management System* (DRMS/SUMS) and extracting data needed by the feature and event detection modules. LMSAL has created a control system for running and monitoring these modules and tools that enables the easy contribution of events to the HEK. In Section 2.1.1 of their paper, Hurlburt *et al.* (2010) describe the *Event Detection System* (EDS), which continually acquires incoming data and makes it available to a variety of feature and event modules. The same article provides flow charts for the various processes.

Because the available resources on which to run the various feature and event detection modules are finite, SAO provides additional computing power required to successfully run the modules. LMSAL has ported their event pipeline control software to SAO to help with this process.

17.2. SAO Computing Facilities

SAO has deployed an 11-node computer cluster consisting of two dual-core 3 GHz processors with between 16 and 32 GB of memory per node. This has been integrated with a 220 TB storage array from SUN, available to nodes via NFS, which has shown to be more than capable of delivering data at the required rates. SAO uses the storage array to construct a rolling archive of the most recent months of SDO/HMI and AIA level 1.0 data. SAO has implemented and tested a remote version of the DRMS/SUMS setup that is in use at Stanford.

Stanford uses SLONY-I (a master–slave replication system that replicates large databases to a number of slave systems, see <http://www.slony.info>) to replicate meta-information to a limited number of external sites, including SAO. SAO transfers and stores SDO data on the local system as it becomes available. This gives SAO rapid access to SDO data, and enables local processing and easy comparison between SDO data and data from other sources (such as H α data in the filament detection module).

17.3. Interfacing to DRMS

LMSAL has provided a dedicated Java/IDL-based pipeline framework to interface DRMS/SUMS with the IDL-based feature- and event-detection modules. SAO makes use of the same code locally for any IDL module processed there. For non-IDL modules, *e.g.* the magnetic feature-tracking module, written in *Perl Data Language* (PDL), the code still uses the pipeline software, but with an additional interface layer providing the communication between module and pipeline software.

SAO runs algorithms that either detect events not relevant to the SDO/HEK effort at LMSAL or report exhaustive findings (*e.g.* bright point locations), and so the SAO computer-vision center will port the EDS software to one of their clusters. SAO will also run detection algorithms that do not use SDO data, such as the filaments and CME modules, as well as modules that key off from detection events created by other detection algorithms (*e.g.* flare-detection events will spawn the process to look for oscillations). Initially, there remains a need for human interaction for some event types, *e.g.* EIT waves and oscillations, to ensure that the detections represent actual events.

17.4. Module Testing

The facilities available at SAO have enabled us to create a framework similar to that available for running the feature and event detection modules at LMSAL. We utilize this framework for end-to-end testing and evaluation of the modules before they are deployed at LMSAL, and for prototyping algorithmic or coding improvements to the modules. Part of the evaluation will include investigation of standardization opportunities, particularly in the case of the IDL codes, which will have to be modified to work within the pipeline framework and provide output in VOEvent format, using the environment provided by LMSAL.

17.5. Community Access to Metadata at SAO

The primary target for the feature and event detection modules' output is the SDO/JSOC feature metadata catalog (*i.e.* HEK at LMSAL). In addition to supplying the HEK with features and events, SAO will utilize the newly redesigned VSO catalog infrastructure to make features and events available, with search capabilities that complement those currently proposed for the HEK. This includes the ability to use the catalogs as datasets for science research. In addition, the VSO catalog infrastructure will allow users to comment on data and catalog entries. This information can be used to enhance the data being provided. Additionally, some of the feature- and event-detection modules run at SAO (*e.g.* for non-SDO data), requiring SAO to host any secondary data products generated as the module runs. By utilizing the VSO infrastructure, LMSAL and SAO can easily make data available to other virtual observatories of the heliosphere (such as the Virtual Heliospheric Observatory and the Virtual Wave Observatory) and provide enhanced data integration for all of heliophysics. Dissemination of the pipeline capabilities will be accomplished through the use of a guidebook, two annual workshops, and a web interface; see <http://helio.cfa.harvard.edu/sdosc/>.

18. Conclusions

The original motivation for NASA to submit an announcement of opportunity for SDO automated feature-finding projects was that the continuous SDO data stream would be simply overwhelming for the solar community: an AIA image every ten seconds in each of the eight channels, an HMI magnetogram every 45 seconds, and a vector magnetogram every 90 seconds, all with no interruption. This would be simply too much to inspect for any solar research group, and for the community, even with the best coordination of efforts. Automated event alerts, from near real-time space-weather warnings to automated cataloging of features and events, must provide at least part of the solution here. We emphasize that automated feature recognition has found wide application in other areas where the amount of imagery has become overwhelming for human observers, *e.g.* elementary particle physics with particle colliders, diagnosis of medical X-ray images, monitoring of urban traffic, and security surveillance in public areas such as airports and sport stadiums. Hence, application to solar physics seems a natural development.

Thus motivated, our group set out to develop a computer-vision system for SDO. We consulted the SDO instrument PIs to select the events and features that should have the highest priority, and began developing a suite of 16 modules, described here, that would meet these priorities best. For most of the modules we were able to find development teams with proven track records of their modules for pre-SDO data, thus minimizing the risk of unwelcome surprises in the little time that we had for development pre-launch. From the beginning we did not want to restrict our computer-vision project to SDO data; for example, it is obvious that for scientific and space-weather uses automated monitoring of filaments and CMEs is required, and hence we incorporated those modules.

As the computer-vision project is taking shape it has become increasingly evident that the production of very high accuracy catalogs of solar features and events is desirable for the field of solar physics in its own right. Due to the fragmented nature of solar observations – different observatories, different time coverage, limited image cadence – for too long solar physicists have limited themselves to the study of single or a few events, often finding that physical models that seem to apply very well in a number of cases turn out to be irrelevant for others.

Examples of the potential unleashed by the computer vision project are: *i)* Draw a butterfly diagram for Active Regions, *ii)* Find all filaments that coincide with sigmoids, and then correlate sigmoid handedness with filament chirality, *iii)* Correlate EUV jets with small-scale flux emergence in coronal holes, *iv)* Draw PIL maps with regions of high shear and large magnetic field gradients overlaid, to pinpoint potential flaring regions. Then correlate with actual flare occurrence. All of these tasks will be accomplished with great ease; the power of this method is limited merely by the imagination of the researcher. In addition, the space-weather alerts generated by our modules allow for the automation of the production and online publication of flare lightcurves and other basic characteristics, quicklook movies, and also automated email alerts to interested parties, from space-weather forecasters to instrument planners. Correlating phenomena will be merely a matter of a few line commands on cataloged metadata.

We are convinced that the presence of automated feature recognition will facilitate the paradigm shift of solar physics from a discipline mostly focused on the analysis of single events, or very limited sets of events, to a discipline capable of the analysis of very large representative sets of events. This, in turn, will enable the discovery of statistical patterns, leading to the recognition of the underlying physical mechanisms, as well as the prediction of the probability of space-weather events.

The multipurpose, trainable, feature-recognition module that we are developing has, as discussed in Section 16, the potential of dealing with solar phenomena for which no task-specific modules exist (*e.g.* cusps, arcades, faculae, pores, arch filaments, *etc.*), as well as quickly identifying newly discovered phenomena (such as sigmoids were in the *Yohkoh* era), and even phenomena that the user is not quite able to define yet (see Section 16 for the example of coronal nulls). We anticipate that the trainable module will find wide applications in heliophysics, not just limited to coronal imagery and magnetograms. Thus, the trainable module has the potential of saving considerable amounts of programming time and effort in comparison with developing task-specific modules for every phenomenon.

The trainable module will also be used to help quantify the accuracy of the task-specific modules. In this article we have not discussed the issue of accuracy in detail, since we are still in the initial phase of the computer-vision project. However, we have developed plans to quantify the accuracy and precision of all our modules in the second phase of our work, which we shall briefly review here. Each module will be tested against human observers. For that we will identify an appropriate time period (say, a week or a month) for which human volunteers will annotate all the AIA and relevant HMI images for the features identified by the task-specific modules. Comparison of the discrepancies between the human-produced data and the module data will allow us to correct in part for the misidentifications by the human observers and produce a dataset that we will consider “true”. Discrepancies between the module data and this set can then be used to pinpoint the accuracy and precision of the module-produced metadata. In addition, we will apply the trainable module to the same dataset, and cross-compare the results. Comparison with the output of other automated feature detection algorithms, such as the GOES flare catalog, and the output from modules developed by other groups will allow for further analysis. These quantitative results then define error bars for statistical analysis of events and correlations. Only very limited quantitative verification of feature-recognition algorithms has been carried out so far; one exception is the verification of an early version of our filament detection algorithm AAFDCC, using the filament list by Pevtsov, Balasubramaniam, and Rogers (2003); see Section 4 for more details.

We have some final comments on further feature-recognition modules produced by the solar-physics community. The Feature Finding Team has produced the modules that were deemed most urgent for the SDO mission, but there is absolutely no reason or intent to limit SDO feature finding to these modules. We are aware of multiple efforts by groups in Europe and the US to produce additional SDO feature-recognition algorithms. For the modules of this article we have developed general data input and metadata output subroutines that are available to any group intent on developing feature-finding routines. In addition, we point out that any modules that run on the continuous data stream, even if for only a single AIA passband, must be able to operate in a semi-autonomous manner in a pipeline setup, because of the data volume. This requires a different approach to module programming than is customary in the solar-physics community. A near real-time pipeline exists at LMSAL, but pipelines operating with a lag time of the order of a day can be set up by any institution that receives at least part of the data stream. As described above, SAO operates a pipeline identical to that at LMSAL that runs several of our modules and that is open, in principle, to community-produced algorithms. The metadata produced by additional modules can be delivered to the HEK in the appropriate format, but there is no reason to limit delivery to the HEK. Some of our modules (the bright-point module and the trainable module) do not deliver all their metadata to the HEK because the former produces a volume of metadata that is too large, while the latter produces metadata that are completely incompatible with the HEK.

In closing, we present our vision for the near and medium future of helioinformatics. We foresee a situation in which heliospheric virtual observatories, such as the HEK and the VSO, provide a simple and seamless interface between data, metadata, and computer-vision software systems. Heliospheric virtual observatories and computer-vision systems will work together to monitor the Sun constantly, provide space weather warnings, populate catalogs of metadata, analyze trends, and produce real-time online imagery of current events. Supported by this extensive apparatus, the field of solar physics will transition from an event-driven research mode to a system-oriented approach and will thus develop better predictive capabilities.

Acknowledgements This research and development project is supported by NASA Grant NNX09AB03G to the Smithsonian Astrophysical Observatory with subcontracts to the Southwest Research Institute, Montana State University, the Applied Physics Laboratory of Johns Hopkins University, and the Lockheed Martin Solar and Astrophysics Laboratory, and with the participation of the NASA Marshall Space Flight Center. This project would be impossible without the major contributions from our European collaborators. The institutions involved are the Royal Observatory of Belgium, the Academy of Athens, Trinity College Dublin, the University of St Andrews, and the Max-Planck-Institut für Sonnensystemforschung.

We thank the referee for helpful and constructive comments, and Trae Winter for carefully proofreading the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Aly, J.J.: 1989, On the reconstruction of the nonlinear force-free coronal magnetic field from boundary data. *Solar Phys.* **120**, 19–48.
- Antiochos, S.K., Dahlburg, R.B., Klimchuk, J.A.: 1994, The magnetic field of solar prominences. *Astrophys. J.* **420**, 41–44. doi:[10.1086/187158](https://doi.org/10.1086/187158).
- Antiochos, S.K., DeVore, C.R., Klimchuk, J.A.: 1999, A model for solar coronal mass ejections. *Astrophys. J.* **510**, 485–493. doi:[10.1086/306563](https://doi.org/10.1086/306563).
- Asai, A., Ishii, T.T., Kurokawa, H., Yokoyama, T., Shimojo, M.: 2003, Evolution of conjugate footpoints inside flare ribbons during a great two-ribbon flare on 2001 April 10. *Astrophys. J.* **586**, 624–629. doi:[10.1086/367694](https://doi.org/10.1086/367694).
- Attrill, G.D.R., Wills-Davey, M.J.: 2010, Automatic detection and extraction of coronal dimmings from SDO/AIA data. *Solar Phys.* **262**, 461–480. doi:[10.1007/s11207-009-9444-4](https://doi.org/10.1007/s11207-009-9444-4).
- Banda, J.M., Angryk, R.: 2009, On the effectiveness of fuzzy clustering as a data discretization technique for large-scale classification of solar images. In: Feng, G.G. (ed.) *Proceedings of the 18th IEEE International Conference on Fuzzy Systems, FUZZ-IEEE '09*, IEEE, New York, 2019–2024. doi:[10.1109/FUZZY.2009.5277273](https://doi.org/10.1109/FUZZY.2009.5277273).
- Barra, V., Delouille, V., Hochedez, J.F.: 2008, Segmentation of extreme ultraviolet solar images via multi-channel fuzzy clustering. *Adv. Space Res.* **42**, 917–925. doi:[10.1016/j.asr.2007.10.021](https://doi.org/10.1016/j.asr.2007.10.021).
- Barra, V., Delouille, V., Hochedez, J.F., Chainais, P.: 2005, Segmentation of EIT images using fuzzy clustering: a preliminary study. In: Danesy, D., Poedts, S., De Groof, A., Andries, J. (eds.) *The Dynamic Sun: Challenges for Theory and Observations*, SP-600, ESA, Noordwijk, 71–80.
- Barra, V., Delouille, V., Kretschmar, M., Hochedez, J.: 2009, Fast and robust segmentation of solar EUV images: algorithm and results for solar cycle 23. *Astron. Astrophys.* **505**, 361–371. doi:[10.1051/0004-6361/200811416](https://doi.org/10.1051/0004-6361/200811416).
- Bemporad, A., Raymond, J., Poletto, G., Romoli, M.: 2007, A comprehensive study of the initiation and early evolution of a coronal mass ejection from ultraviolet and white-light data. *Astrophys. J.* **655**, 576–590. doi:[10.1086/509569](https://doi.org/10.1086/509569).
- Bernasconi, P.N., Rust, D.M., Hakim, D.: 2005, Advanced automated solar filament detection and characterization code: description, performance, and results. *Solar Phys.* **228**, 97–117. doi:[10.1007/s11207-005-2766-y](https://doi.org/10.1007/s11207-005-2766-y).
- Bernasconi, P.N., Raouafi, N.E., Georgoulis, M.K.: 2011, The sigmoid sniffer. *Solar Phys.*, submitted.

- Bewsher, D., Harrison, R.A., Brown, D.S.: 2008, The relationship between EUV dimming and coronal mass ejections. I. Statistical study and probability model. *Astron. Astrophys.* **478**, 897–906. doi:[10.1051/0004-6361:20078615](https://doi.org/10.1051/0004-6361:20078615).
- Boursier, Y., Lamy, P., Llebaria, A.: 2009, Three-dimensional kinematics of coronal mass ejections from STEREO/SECCHI-COR2 observations in 2007–2008. *Solar Phys.* **256**, 131–147. doi:[10.1007/s11207-009-9358-1](https://doi.org/10.1007/s11207-009-9358-1).
- Boursier, Y., Lamy, P., Llebaria, A., Goudail, F., Robelus, S.: 2009, The ARTEMIS catalog of LASCO coronal mass ejections. Automatic recognition of transient events and Marseille inventory from synoptic maps. *Solar Phys.* **257**, 125–147. doi:[10.1007/s11207-009-9370-5](https://doi.org/10.1007/s11207-009-9370-5).
- Byrne, J.P., Gallagher, P.T., McAteer, R.T.J., Young, C.A.: 2009, The kinematics of coronal mass ejections using multiscale methods. *Astron. Astrophys.* **495**, 325–334. doi:[10.1051/0004-6361:200809811](https://doi.org/10.1051/0004-6361:200809811).
- Canfield, R.C., Hudson, H.S., McKenzie, D.E.: 1999, Sigmoidal morphology and eruptive solar activity. *Geophys. Res. Lett.* **26**, 627–630. doi:[10.1029/1999GL900105](https://doi.org/10.1029/1999GL900105).
- Canfield, R.C., Kazachenko, M.D., Acton, L.W., Mackay, D.H., Son, J., Freeman, T.L.: 2007, *Yohkoh* SXT full-resolution observations of sigmoids: structure, formation, and eruption. *Astrophys. J. Lett.* **671**, 81–84. doi:[10.1086/524729](https://doi.org/10.1086/524729).
- Chae, J., Martin, S.F., Yun, H.S., Kim, J., Lee, S., Goode, P.R., Spirock, T., Wang, H.: 2001, Small magnetic bipoles emerging in a filament channel. *Astrophys. J.* **548**, 497–507. doi:[10.1086/318661](https://doi.org/10.1086/318661).
- Christe, S., Hannah, I.G., Krucker, S., McTiernan, J., Lin, R.P.: 2008, RHESSI microflare statistics. I. Flare-finding and frequency distributions. *Astrophys. J.* **677**, 1385–1394. doi:[10.1086/529011](https://doi.org/10.1086/529011).
- Davey, A.R., McIntosh, S.: 2007, The SoHO/EIT brightpoint database: mining the database for science. *Bull. Am. Astron. Soc.* **38**, 327.
- De Moortel, I., Hood, A.W.: 2000, Wavelet analysis and the determination of coronal plasma properties. *Astron. Astrophys.* **363**, 269–278.
- De Moortel, I., McAteer, R.T.J.: 2004, Waves and wavelets: an automated detection technique for solar oscillations. *Solar Phys.* **223**, 1–11. doi:[10.1007/s11207-004-0806-7](https://doi.org/10.1007/s11207-004-0806-7).
- De Moortel, I., Hood, A.W., Ireland, J.: 2002, Coronal seismology through wavelet analysis. *Astron. Astrophys.* **381**, 311–323. doi:[10.1051/0004-6361:20011659](https://doi.org/10.1051/0004-6361:20011659).
- De Moortel, I., Munday, S.A., Hood, A.W.: 2004, Wavelet analysis: the effect of varying basic wavelet parameters. *Solar Phys.* **222**, 203–228. doi:[10.1023/B:SOLA.0000043578.01201.2d](https://doi.org/10.1023/B:SOLA.0000043578.01201.2d).
- DeForest, C.E., Plunkett, S.P., Andrews, M.D.: 2001, Observation of polar plumes at high solar altitudes. *Astrophys. J.* **546**, 569–575. doi:[10.1086/318221](https://doi.org/10.1086/318221).
- DeForest, C.E., Hagenaar, H.J., Lamb, D.A., Parnell, C.E., Welsch, B.T.: 2007, Solar magnetic tracking. I. Software comparison and recommended practices. *Astrophys. J.* **666**, 576–587. doi:[10.1086/518994](https://doi.org/10.1086/518994).
- DeForest, C.E., Lamb, D.A., Berger, T., Hagenaar, H., Parnell, C., Welsch, B.: 2008, The small-scale field measured with *Hinode*/SOT and feature tracking: where is the mixed-polarity flux? *AGU Spring Meeting Abstracts*, 1.
- DeRosa, M.L., Schrijver, C.J., Barnes, G., Leka, K.D., Lites, B.W., Aschwanden, M.J., Amari, T., Canou, A., McTiernan, J.M., Régnier, S., Thalmann, J.K., Valori, G., Wheatland, M.S., Wiegelmann, T., Cheung, M.C.M., Conlon, P.A., Fuhrmann, M., Inhester, B., Tadesse, T.: 2009, A critical assessment of nonlinear force-free field modeling of the solar corona for active region 10953. *Astrophys. J.* **696**, 1780–1791. doi:[10.1088/0004-637X/696/2/1780](https://doi.org/10.1088/0004-637X/696/2/1780).
- Falconer, D.A., Moore, R.L., Gary, G.A.: 2002, Correlation of the coronal mass ejection productivity of solar active regions with measures of their global nonpotentiality from vector magnetograms: baseline results. *Astrophys. J.* **569**, 1016–1025. doi:[10.1086/339161](https://doi.org/10.1086/339161).
- Fan, Y., Gibson, S.E.: 2003, The emergence of a twisted magnetic flux tube into a preexisting coronal arcade. *Astrophys. J. Lett.* **589**, 105–108. doi:[10.1086/375834](https://doi.org/10.1086/375834).
- Fan, Y., Gibson, S.E.: 2004, Numerical simulations of three-dimensional coronal magnetic fields resulting from the emergence of twisted magnetic flux tubes. *Astrophys. J.* **609**, 1123–1133. doi:[10.1086/421238](https://doi.org/10.1086/421238).
- Fisher, R.R., Munro, R.H.: 1984, Coronal transient geometry. I – The flare-associated event of 1981 March 25. *Astrophys. J.* **280**, 428–439. doi:[10.1086/162009](https://doi.org/10.1086/162009).
- Fuller, N., Aboudarham, J., Bentley, R.D.: 2005, Filament recognition and image cleaning on Meudon H α spectroheliograms. *Solar Phys.* **227**, 61–73. doi:[10.1007/s11207-005-8364-1](https://doi.org/10.1007/s11207-005-8364-1).
- Gao, J., Wang, H., Zhou, M.: 2002, Development of an automatic filament disappearance detection system. *Solar Phys.* **205**, 93–103.
- Gibson, S.E., Fan, Y., Török, T., Kliem, B.: 2006, The evolving sigmoid: evidence for magnetic flux ropes in the corona before, during, and after CMES. *Space Sci. Rev.* **124**, 131–144. doi:[10.1007/s11214-006-9101-2](https://doi.org/10.1007/s11214-006-9101-2).
- Gilbert, H.R., Holzer, T.E., Burkepile, J.T., Hundhausen, A.J.: 2000, Active and eruptive prominences and their relationship to coronal mass ejections. *Astrophys. J.* **537**, 503–515. doi:[10.1086/309030](https://doi.org/10.1086/309030).

- Golub, L., Deluca, E., Austin, G., Bookbinder, J., Caldwell, D., Cheimets, P., Cirtain, J., Cosmo, M., Reid, P., Sette, A., Weber, M., Sakao, T., Kano, R., Shibasaki, K., Hara, H., Tsuneta, S., Kumagai, K., Tamura, T., Shimojo, M., McCracken, J., Carpenter, J., Haight, H., Siler, R., Wright, E., Tucker, J., Rutledge, H., Barbera, M., Peres, G., Varisco, S.: 2007, The X-ray telescope (XRT) for the *Hinode* mission. *Solar Phys.* **243**, 63–86. doi:[10.1007/s11207-007-0182-1](https://doi.org/10.1007/s11207-007-0182-1).
- Gopalswamy, N., Yashiro, S., Michalek, G., Stenborg, G., Vourlidas, A., Freeland, S., Howard, R.: 2009, The SOHO/LASCO CME catalog. *Earth Moon Planets* **104**, 295–313. doi:[10.1007/s11038-008-9282-7](https://doi.org/10.1007/s11038-008-9282-7).
- Green, L.M., Kliem, B., Török, T., van Driel-Gesztelyi, L., Attrill, G.D.R.: 2007, Transient coronal sigmoids and rotating erupting flux ropes. *Solar Phys.* **246**, 365–391. doi:[10.1007/s11207-007-9061-z](https://doi.org/10.1007/s11207-007-9061-z).
- Hagenaar, H., Cheung, M.: 2008, Magnetic flux emergence on different scales. In: *12th European Solar Physics Meeting 2*, 2.53. <http://espm.kis.uni-freiburg.de>.
- Hurlburt, N., Cheung, M., Schrijver, C., Chang, L., Freeland, S., Green, S., Heck, C., Jaffey, A., Kobashi, A., Schiff, D., Serafin, J., Seguin, R., Slater, G., Somani, A., Timmons, R.: 2010, Heliophysics event knowledgebase for the solar dynamics observatory and beyond. *Solar Phys.* doi:[10.1007/s11207-010-9624-2](https://doi.org/10.1007/s11207-010-9624-2).
- Inhester, B., Feng, L., Wiegelmann, T.: 2008, Segmentation of loops from coronal EUV images. *Solar Phys.* **248**, 379–393. doi:[10.1007/s11207-007-9027-1](https://doi.org/10.1007/s11207-007-9027-1).
- Ireland, J., De Moortel, I.: 2002, Application of wavelet analysis to transversal coronal loop oscillations. *Astron. Astrophys.* **391**, 339–351. doi:[10.1051/0004-6361/20020643](https://doi.org/10.1051/0004-6361/20020643).
- Jing, J., Yurchyshyn, V.B., Yang, G., Xu, Y., Wang, H.: 2004, On the relation between filament eruptions, flares, and coronal mass ejections. *Astrophys. J.* **614**, 1054–1062. doi:[10.1086/423781](https://doi.org/10.1086/423781).
- Jing, J., Wiegelmann, T., Suematsu, Y., Kubo, M., Wang, H.: 2008, Changes of magnetic structure in three dimensions associated with the X3.4 flare of 2006 December 13. *Astrophys. J. Lett.* **676**, 81–84. doi:[10.1086/587058](https://doi.org/10.1086/587058).
- Jones, H.P.: 2004, Counting magnetic bipoles on the Sun by polarity inversion. In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) *Knowledge-Based Intelligent Information and Engineering Systems: 8th International Conference, KES 2004, Lecture Notes in Computer Science* **3215**, 433–439.
- Kliem, B., Titov, V.S., Török, T.: 2004, Formation of current sheets and sigmoidal structure by the kink instability of a magnetic loop. *Astron. Astrophys.* **413**, 23–26. doi:[10.1051/0004-6361/20031690](https://doi.org/10.1051/0004-6361/20031690).
- Kosugi, T., Matsuzaki, K., Sakao, T., Shimizu, T., Sone, Y., Tachikawa, S., Hashimoto, T., Minesugi, K., Ohnishi, A.A., Yamada, T., Tsuneta, S., Hara, H., Ichimoto, K., Suematsu, Y., Shimojo, M., Watanabe, T., Shimada, S., Davis, J.M., Hill, L.D., Owens, J.K., Title, A.M., Culhane, J.L., Harra, L.K., Doschek, G.A., Golub, L.: 2007, The *Hinode* (Solar-B) mission: an overview. *Solar Phys.* **243**, 3–17. doi:[10.1007/s11207-007-9014-6](https://doi.org/10.1007/s11207-007-9014-6).
- LaBonte, B.J., Rust, D.M., Bernasconi, P.N.: 2003, An automated system for detecting sigmoids in solar X-ray images. *Bull. Am. Astron. Soc.* **35**, 814.
- Lamb, D.A., DeForest, C.E., Hagenaar, H.J., Parnell, C.E., Welsch, B.T.: 2008, Solar magnetic tracking. II. The apparent unipolar origin of quiet-sun flux. *Astrophys. J.* **674**, 520–529. doi:[10.1086/524372](https://doi.org/10.1086/524372).
- Lamb, D., DeForest, C.E., Hagenaar, H.J., Parnell, C.E., Welsch, B.T.: 2007, Feature tracking of *Hinode* magnetograms. *AGU Fall Meeting Abstracts*, 1066.
- Lamb, R.: 2008, An information retrieval system for images from the TRACE satellite. Master's thesis, Montana State University, Bozeman, MT, USA. <http://www.cs.montana.edu/techreports/2008/Lamb.pdf>.
- Lamb, R., Angryk, R., Martens, P.: 2008, An example-based image retrieval system for the TRACE repository. In: Ejiri, M., Kasturi, R., Sanniti di Baja, G. (eds.) *Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08)*, IEEE, New York, 1–4. doi:[10.1109/ICPR.2008.4761078](https://doi.org/10.1109/ICPR.2008.4761078).
- Leblanc, Y., Dulk, G.A., Vourlidas, A., Bougeret, J.L.: 2001, Tracing shock waves from the corona to 1 AU: Type II radio emission and relationship with CMEs. *J. Geophys. Res.* **106**, 25301–25312. doi:[10.1029/2000JA000260](https://doi.org/10.1029/2000JA000260).
- Martin, S.F.: 1998, Conditions for the formation and maintenance of filaments (invited review). *Solar Phys.* **182**, 107–137. doi:[10.1023/A:1005026814076](https://doi.org/10.1023/A:1005026814076).
- Martin, S.F., Bilimoria, R., Tracadas, P.W.: 1994, Magnetic field configurations basic to filament channels and filaments. In: Rutten, R.J., Schrijver, C.J. (eds.) *Solar Surface Magnetism*, Kluwer, Dordrecht, 303–338.
- Masuda, S., Kosugi, T., Hara, H., Tsuneta, S., Ogawara, Y.: 1994, A loop-top hard X-ray source in a compact solar flare as evidence for magnetic reconnection. *Nature* **371**, 495–497. doi:[10.1038/371495a0](https://doi.org/10.1038/371495a0).
- McAteer, R.T.J., Gallagher, P.T., Bloomfield, D.S., Williams, D.R., Mathioudakis, M., Keenan, F.P.: 2004, Ultraviolet oscillations in the chromosphere of the quiet Sun. *Astrophys. J.* **602**, 436–445. doi:[10.1086/380835](https://doi.org/10.1086/380835).
- McIntosh, P.S.: 1994, *YOHKOH* X-ray image interpretation with overlays of H α neutral lines. In: Uchida, Y., Watanabe, T., Shibata, K., Hudson, H.S. (eds.) *X-ray Solar Physics from Yohkoh*, Universal Academy Press, Tokyo, 271–272.

- McIntosh, S.W., Gurman, J.B.: 2004, EIT EUV brightpoints over the SOHO mission so far. In: Walsh, R.W., Ireland, J., Danesy, D., Fleck, B. (eds.) *SOHO 15 Coronal Heating, ESA Special Publication 575*, 235.
- McIntosh, S.W., Gurman, J.B.: 2005, Nine years of EUV bright points. *Solar Phys.* **228**, 285–299. doi:[10.1007/s11207-005-4725-z](https://doi.org/10.1007/s11207-005-4725-z).
- McIntosh, S.W., Sitongia, L., Markel, R., Judge, P.G., Davey, A.R.: 2009, Mining a massive brightpoint database for science. *Bull. Am. Astron. Soc.* **41**, 839.
- McKenzie, D.E., Canfield, R.C.: 2008, *Hinode* XRT observations of a long-lasting coronal sigmoid. *Astron. Astrophys.* **481**, 65–68. doi:[10.1051/0004-6361:20079035](https://doi.org/10.1051/0004-6361:20079035).
- Metcalf, T.R., Derosa, M.L., Schrijver, C.J., Barnes, G., van Ballegoijen, A.A., Wiegelmann, T., Wheatland, M.S., Valori, G., McTiernan, J.M.: 2008, Nonlinear force-free modeling of coronal magnetic fields. II. Modeling a filament arcade and simulated chromospheric and photospheric vector fields. *Solar Phys.* **247**, 269–299. doi:[10.1007/s11207-007-9110-7](https://doi.org/10.1007/s11207-007-9110-7).
- Michalek, G., Gopalswamy, N., Yashiro, S.: 2003, A new method for estimating widths, velocities, and source location of Halo coronal mass ejections. *Astrophys. J.* **584**, 472–478. doi:[10.1086/345526](https://doi.org/10.1086/345526).
- Narukage, N., Sakao, T., Kano, R., Hara, H., Shimojo, M., Bando, T., Urayama, F., DeLuca, E., Golub, L., Weber, M., Grigis, P., Cirtain, J., Tsuneta, S.: 2011, Coronal-temperature-diagnostic capability of the *Hinode*/X-Ray Telescope based on self-consistent calibration. *Solar Phys.* doi:[10.1007/s11207-010-9685-2](https://doi.org/10.1007/s11207-010-9685-2).
- Ogawara, Y., Takano, T., Kato, T., Kosugi, T., Tsuneta, S., Watanabe, T., Kondo, I., Uchida, Y.: 1991, The solar-A mission – an overview. *Solar Phys.* **136**, 1–16. doi:[10.1007/BF00151692](https://doi.org/10.1007/BF00151692).
- Olmedo, O., Zhang, J., Wechsler, H., Poland, A., Borne, K.: 2008, Automatic detection and tracking of coronal mass ejections in coronagraph time series. *Solar Phys.* **248**, 485–499. doi:[10.1007/s11207-007-9104-5](https://doi.org/10.1007/s11207-007-9104-5).
- Pevtsov, A.A., Balasubramaniam, K.S., Rogers, J.W.: 2003, Chirality of chromospheric filaments. *Astrophys. J.* **595**, 500–505. doi:[10.1086/377339](https://doi.org/10.1086/377339).
- Pneuman, G.W.: 1983, The formation of solar prominences by magnetic reconnection and condensation. *Solar Phys.* **88**, 219–239. doi:[10.1007/BF00196189](https://doi.org/10.1007/BF00196189).
- Podladchikova, O., Berghmans, D.: 2005, Automated detection of EIT waves and dimmings. *Solar Phys.* **228**, 265–284. doi:[10.1007/s11207-005-5373-z](https://doi.org/10.1007/s11207-005-5373-z).
- Qu, M., Shih, F.Y., Jing, J., Wang, H.: 2005, Automatic solar filament detection using image processing techniques. *Solar Phys.* **228**, 119–135. doi:[10.1007/s11207-005-5780-1](https://doi.org/10.1007/s11207-005-5780-1).
- Rachmeler, L.A., Wills-Davey, M.J.: 2005, Observations of unusual “EIT Wave” dynamics. *AGU Spring Meeting Abstracts*, 9.
- Reinard, A.A., Biesecker, D.A.: 2008, Coronal mass ejection-associated coronal dimmings. *Astrophys. J.* **674**, 576–585. doi:[10.1086/525269](https://doi.org/10.1086/525269).
- Robbrecht, E., Berghmans, D.: 2004, Automated recognition of coronal mass ejections (CMEs) in near-real-time data. *Astron. Astrophys.* **425**, 1097–1106. doi:[10.1051/0004-6361:20041302](https://doi.org/10.1051/0004-6361:20041302).
- Rust, D.M., Hildner, E.: 1976, Expansion of an X-ray coronal arch into the outer corona. *Solar Phys.* **48**, 381–387.
- Rust, D.M., Kumar, A.: 1996, Evidence for helically kinked magnetic flux ropes in solar eruptions. *Astrophys. J. Lett.* **464**, 199–202. doi:[10.1086/310118](https://doi.org/10.1086/310118).
- Rust, D.M., LaBonte, B.J.: 2005, Observational evidence of the kink instability in solar filament eruptions and sigmoids. *Astrophys. J. Lett.* **622**, 69–72. doi:[10.1086/429379](https://doi.org/10.1086/429379).
- Rust, D.M., Martin, S.F.: 1994, A correlation between sunspot whirls and filament type. In: Balasubramaniam, K.S., Simon, G.W. (eds.) *Solar Active Region Evolution: Comparing Models with Observations CS-68*, Astron. Soc. Pac., San Francisco, 337.
- Rust, D.M., Anderson, B.J., Andrews, M.D., Acuña, M.H., Russell, C.T., Schuck, P.W., Mulligan, T.: 2005, Comparison of interplanetary disturbances at the NEAR spacecraft with coronal mass ejections at the Sun. *Astrophys. J.* **621**, L524–L536. doi:[10.1086/427401](https://doi.org/10.1086/427401).
- Savcheva, A., Cirtain, J., Deluca, E.E., Lundquist, L.L., Golub, L., Weber, M., Shimojo, M., Shibasaki, K., Sakao, T., Narukage, N., Tsuneta, S., Kano, R.: 2007, A study of polar jet parameters based on *Hinode* XRT observations. *Publ. Astron. Soc. Japan* **59**, 771–778.
- Schrijver, C.J., Derosa, M.L., Metcalf, T.R., Liu, Y., McTiernan, J., Régnier, S., Valori, G., Wheatland, M.S., Wiegelmann, T.: 2006, Nonlinear force-free modeling of coronal magnetic fields, Part I: A quantitative comparison of methods. *Solar Phys.* **235**, 161–190. doi:[10.1007/s11207-006-0068-7](https://doi.org/10.1007/s11207-006-0068-7).
- Shih, F.Y., Kowalski, A.J.: 2003, Automatic extraction of filaments in H α solar images. *Solar Phys.* **218**, 99–122. doi:[10.1023/B:SOLA.0000013052.34180.58](https://doi.org/10.1023/B:SOLA.0000013052.34180.58).
- Tadesse, T., Wiegelmann, T., Inhester, B.: 2009, Nonlinear force-free coronal magnetic field modeling and preprocessing of vector magnetograms in spherical geometry. *Astron. Astrophys.* **508**, 421–432.
- Thalmann, J.K., Wiegelmann, T., Raouafi, N.E.: 2008, First nonlinear force-free field extrapolations of SO-LIS/VSM data. *Astron. Astrophys.* **488**, 71–74. doi:[10.1051/0004-6361:200810235](https://doi.org/10.1051/0004-6361:200810235).

- Titov, V.S., Démoulin, P.: 1999, Basic topology of twisted magnetic configurations in solar flares. *Astron. Astrophys.* **351**, 707–720.
- Török, T., Kliem, B.: 2005, Confined and ejective eruptions of kink-unstable flux ropes. *Astrophys. J. Lett.* **630**, L97–L100. doi:[10.1086/462412](https://doi.org/10.1086/462412).
- Török, T., Kliem, B.: 2007, Numerical simulations of fast and slow coronal mass ejections. *Astron. Nachr.* **328**, 743–746. doi:[10.1002/asna.200710795](https://doi.org/10.1002/asna.200710795).
- Tsuneta, S., Hara, H., Shimizu, T., Acton, L.W., Strong, K.T., Hudson, H.S., Ogawara, Y.: 1992, Observation of a solar flare at the limb with the *YOHKOH* soft X-ray telescope. *Publ. Astron. Soc. Japan* **44**, 63–69.
- van Ballegoijen, A.A., Martens, P.C.H.: 1989, Formation and eruption of solar prominences. *Astrophys. J.* **343**, 971–984. doi:[10.1086/167766](https://doi.org/10.1086/167766).
- van Ballegoijen, A.A., Martens, P.C.H.: 1990, Magnetic fields in quiescent prominences. *Astrophys. J.* **361**, 283–289. doi:[10.1086/169193](https://doi.org/10.1086/169193).
- Webb, D.F.: 2000, Understanding CMEs and their source regions. *J. Atmos. Solar-Terr. Phys.* **62**, 1415–1426.
- Wheatland, M.S., Sturrock, P.A., Roumeliotis, G.: 2000, An optimization approach to reconstructing force-free fields. *Astrophys. J.* **540**, 1150–1155. doi:[10.1086/309355](https://doi.org/10.1086/309355).
- Wiegelmann, T.: 2004, Optimization code with weighting function for the reconstruction of coronal magnetic fields. *Solar Phys.* **219**, 87–108. doi:[10.1023/B:SOLA.0000021799.39465.36](https://doi.org/10.1023/B:SOLA.0000021799.39465.36).
- Wiegelmann, T.: 2007, Computing nonlinear force-free coronal magnetic fields in spherical geometry. *Solar Phys.* **240**, 227–239. doi:[10.1007/s11207-006-0266-3](https://doi.org/10.1007/s11207-006-0266-3).
- Wiegelmann, T., Inhester, B.: 2010, How to deal with measurement errors and lacking data in nonlinear force-free coronal magnetic field modelling? *Astron. Astrophys.* **516**, 107. doi:[10.1051/0004-6361/201014391](https://doi.org/10.1051/0004-6361/201014391).
- Wiegelmann, T., Inhester, B., Sakurai, T.: 2006, Preprocessing of vector magnetograph data for a nonlinear force-free magnetic field reconstruction. *Solar Phys.* **233**, 215–232. doi:[10.1007/s11207-006-2092-z](https://doi.org/10.1007/s11207-006-2092-z).
- Wiegelmann, T., Inhester, B., Feng, L.: 2009, Solar stereoscopy – where are we and what developments do we require to progress? *Ann. Geophys.* **27**, 2925–2936.
- Wiegelmann, T., Thalmann, J.K., Schrijver, C.J., Derosa, M.L., Metcalf, T.R.: 2008, Can we improve the preprocessing of photospheric vector magnetograms by the inclusion of chromospheric observations? *Solar Phys.* **247**, 249–267. doi:[10.1007/s11207-008-9130-y](https://doi.org/10.1007/s11207-008-9130-y).
- Wills-Davey, M.J.: 2006, Tracking large-scale propagating coronal wave fronts (EIT waves) using automated methods. *Astrophys. J.* **645**, 757–765. doi:[10.1086/504144](https://doi.org/10.1086/504144).
- Xie, H., Ofman, L., Lawrence, G.: 2004, Cone model for halo CMEs: application to space weather forecasting. *J. Geophys. Res. (Space Phys.)* **109**(A18), 3109–3121. doi:[10.1029/2003JA010226](https://doi.org/10.1029/2003JA010226).
- Xue, X.H., Wang, C.B., Dou, X.K.: 2005, An ice-cream cone model for coronal mass ejections. *J. Geophys. Res. (Space Phys.)* **110**(A9), 8103–8114. doi:[10.1029/2004JA010698](https://doi.org/10.1029/2004JA010698).
- Yang, L., Jin, R., Sukthankar, R., Zheng, B., Mummert, L., Satyanarayanan, M., Chen, M., Jukic, D.: 2007, Learning distance metrics for interactive search-assisted diagnosis of mammograms. In: *SPIE CS-6514*. doi:[10.1117/12.710076](https://doi.org/10.1117/12.710076).
- Yurchyshyn, V.B., Wang, H., Goode, P.R., Deng, Y.: 2001, Orientation of the magnetic fields in interplanetary flux ropes and solar filaments. *Astrophys. J.* **563**, 381–388. doi:[10.1086/323778](https://doi.org/10.1086/323778).