SYNCHRONIC CORONAL HOLE MAPPING USING MULTI-INSTRUMENT EUV IMAGES: DATA PREPARATION AND DETECTION METHOD

R. M. CAPLAN, C. DOWNS, AND J. A. LINKER
Predictive Science Inc., 9990 Mesa Rim Road, Suite 170 San Diego, CA 92121, USA; caplan@predsci.com
Received 2015 October 9; accepted 2016 March 20; published 2016 May 23

ABSTRACT

We describe a method for the automatic mapping of coronal holes (CHs) using simultaneous multi-instrument EUV imaging data. Synchronized EUV images from STEREO/EUVI A and B 195 Å and SDO/AIA 193 Å are preprocessed, including point-spread function deconvolution and the application of data-derived intensity corrections which account for center-to-limb variations (limb brightening) and inter-instrument intensity normalization. We systematically derive a robust limb-brightening correction that takes advantage of unbiased long-term averages of data and respects the physical nature of the problem. The new preprocessing greatly assists in CH detection, allowing for the use of a simplified variable-connectivity, two-threshold region-growing image segmentation algorithm to obtain consistent detection results. We generate synchronous EUV and CH maps, and show a preliminary analysis of CH evolution. Several data and code products are made available to the community (www.predsci.com/chd): for the period of this study (2010 June 10 to 2014 August 18), we provide synchronous EUV and CH map data at 6 hr cadence, data-derived limb-brightening corrections for STEREO/EUVI A and B 195 Å and SDO/AIA 193 Å, and inter-instrument correction factors to equate their intensities. We also provide the CH image segmentation code module (ezseg) implemented in both FORTRAN-OpenMP and GPU-accelerated C-CUDA. A complete implementation of our CH detection pipeline in the form of a ready-to-use MATLAB driver script euv2chm utilizing ezseg is also made available.

Key words: methods: data analysis – Sun: corona – Sun: evolution – techniques: image processing

1. INTRODUCTION

Coronal holes (CHs) manifest themselves as regions of relatively low intensity in EUV and X-Ray images (or correspondingly bright regions in He I 10830 absorption; Bohlin 1977). They have long been recognized as being associated with open magnetic fields (Altschuler et al. 1972), sources of fast solar wind streams (Krieger et al. 1973), and recurrent geomagnetic activity (Neupert & Pizzo 1974). The boundaries of CHs are recognized to play an important role in the origin of the slow solar wind. In theories that invoke a quasi-static origin, the slow wind arises from regions of large expansion factor near the boundaries (Wang & Sheeley 1990; Cranmer et al. 2007). Recently, it was shown that distance to CH boundaries is in fact a better predictor of solar wind speed than expansion factor (Riley et al. 2015). Theories that assert a dynamic origin for the slow wind require reconnection between the open and closed field regions (Fisk et al. 1998; Antiochos et al. 2011); the imprints of such reconnection should be seen near CH boundaries. Explanations for the quasi-rigid rotation of some extended CHs also require reconnection there (Wang et al. 1996; Fisk et al. 1998; Lionello et al. 2005). Understanding the evolution of CH boundaries is thus important not only for investigating coronal dynamics, but also for deducing the genesis of the solar wind.

The recent profile of the STEREO and SDO missions provides us with the opportunity to study CH evolution over medium time periods (days to weeks). However, a meaningful assessment, as well as comparison with models, requires an objective method for identifying CH boundaries in a consistent manner over many days.

It is difficult to systematically identify CH boundaries within EUV images because they are not always clear-cut, and the choice of boundary may be somewhat subjective (indeed, they can even differ based on arbitrary factors such as the chosen color table of the image). Even with a strict definition for a CH and a means to reject false positives (such as darker quiet-Sun regions and filaments), factors such as systematic instrumental effects, line of sight changes, and temporal evolution all may cause variations in CH intensity and/or appearance in EUV imaging data. Together, these factors can make it difficult to automatically detect CH regions in a consistent manner as they evolve in space and time.

All of these issues have therefore yielded many algorithms and methods for identifying CHs in the literature. Historically, CH detection was accomplished by manually tracing the perceived boundaries of the CHs by hand, frame by frame (Harvey & Recely 2002). Modern CH detection methods are typically built around adaptive image segmentation algorithms that use the properties of individual (or several) images to separate the image into distinct regions such as CHs or active regions (ARs). Image segmentation has been a subject of research for over 30 years. We refer the reader to the surveys of Zucker (1976) and Freixenet et al. (2002) for an overview.

Many of the algorithms that have been developed for image segmentation are difficult to implement or generally unsuitable for use in CH detection due to the noisy nature of EUV images, the ill-defined boundaries of CHs, as well as other unique features of EUV images, such as the similar intensities of dark quiet-Sun regions. For this reason, many applications of image segmentation to solar EUV images focus on the simplest techniques that yield acceptable results for the specific goal of each study; see Aschwanden (2010) for a review.

The most basic segmentation method, that of intensity thresholding, is the basis for many algorithms. Histograms can
be used to show different intensity value ranges in various structures with thresholds chosen to best separate these structures; for example, Gallagher et al. (1998) used this method to identify the structure of the quiet-Sun network. Methods using global and local histograms to select and apply thresholding segmentations have been used for CH detection (Scholl & Habbal 2008; Krista & Galleger 2009; Scholl & Habbal 2009) and utilized in studying short-term local CH evolution (Krista et al. 2011; Krista & Reinard 2013). A previous study that used AIA and EUVI data for synchronic CHs visible by SDO/AIA. The SPoCA algorithm uses a slightly modified version of the fuzzy C-means algorithm applied to the image histograms to clump similar areas together. Other authors use different methods of classifiers that utilize EUV images from multiple wavelengths to detect structures including CHs (De Wit 2006; Suresh 2013).

A core element in all of these techniques is the concept of identifying CHs using adaptive thresholds or adaptive segmentation logic that is based on the structural features presented in an image at a single time or in groups of images within a relatively short period (an exception is Henney & Harvey 2005, where the authors used 10-year averages of hand-drawn CHs from He 1083 nm images to form their threshold values). Thus, by design, the definition of a CH can shift or adapt depending on the relative distribution of structures in the current image(s), which may vary with time as the Sun rotates and evolves morphologically. In this work, we take a different philosophical approach to the problem. Here, we attempt to construct an image processing pipeline and detection algorithm that will allow us to employ an inherently fixed definition of a CH, one that ideally remains consistent as CHs evolve in time and space.

This idea is motivated by the goal of constructing high-quality full-Sun CH maps for a given time and to track the temporal evolution of such features in the simplest and most consistent manner as possible. Specifically, we aim to develop a technique that can: (a) account for systematic changes in brightness and contrast as CHs rotate throughout the field of view of one instrument (e.g., center-to-limb changes), (b) remain relatively insensitive to variations in the appearance and structure of the corona during rotation and the solar cycle (histogram variation), and (c) consistently detect the same features from heterogeneous, multi-spacecraft imaging data as features rotate from one viewpoint to another. Satisfying these goals opens the door to studying not only the long-term properties of CHs on the Sun (e.g., Lowder et al. 2014) or the short-term evolution of transient holes (e.g., Krista & Reinard 2013), but also the medium-term evolution of a CH as it rotates around the Sun. This is particularly relevant during the 2010–2014 time period that we cover in this study because the favorable orientation of the STEREO and SDO spacecraft allows for the instantaneous mapping of nearly the entire EUV corona (also known as a synchro map), meaning that the evolution of a single CH can be studied continuously for weeks at a time.

Our approach to achieve these goals is divided into two main elements. The first is the development of a data-driven, physically motivated, preprocessing pipeline to prepare EUV images for CH detection. In addition to using point-spread function (PSF) deconvolution to naturally enhance the CH contrast, we present a technique to derive intensity-dependent limb-brightening corrections from tailored year-long averages of imaging data, which enables the flattening of EUV intensity variation from the disk center to the limb in a manner that respects the inherent geometry and physics of the problem. We also use a similar technique to derive inter-instrument transformations (IITs) to relate intensity values across instruments, allowing the use of a single set of CH detection parameters for all three spacecraft. This preprocessing pipeline is very general and has a number of additional applications where the time-dependent mapping of EUV data is of interest.

The second element is the detection algorithm itself. Our code module, named ezseg, utilizes a two-threshold simplified implementation of a region-growing segmentation algorithm. It relies on seeding points that have a high probability of belonging to CHs, and subsequently growing the seeded region to intensities that are deemed to generally correspond to the boundary of CHs. A variation of this algorithm has been employed previously in other contexts, such as the popular SWAMIS code (DeForest et al. 2007) for magnetic feature tracking. Our implementation differs from SWAMIS in its algorithmic computation method, using manually selected threshold values, and an added variable-connectivity requirement (see Section 4). The ezseg algorithm is implemented as a modular CUDA-C routine (for running on Nvidia GPUs), as well as an OpenMP-enabled FORTRAN routine, both of which can be called from any custom-made driver code, including those written in scripting/data analysis languages (e.g., MATLAB, IDL, Octave, SciLab, R, etc.). For convenience and to provide a ready-to-use code, we make available a sample driver implementation in MATLAB in addition to the ezseg codes themselves.

This paper, which focuses on how we build and apply the above preprocessing and CH detection techniques to EUV imaging data from 2010 to 2014, is outlined as follows. Section 2 describes the process of data acquisition and the issues involved. Section 3 describes the preprocessing steps of PSF deconvolution, limb-brightening correction, and IITs. Validation of the preprocessing is also discussed. Section 4 describes our detection algorithm and mapping procedures. Sample results of the generated CH maps spanning the four-year period and a preliminary analysis of CH evolution are given in Section 5. We conclude in Section 6 where we provide the locations of the provided data products and codes, and comment on future applications of this work.

To aid the reader through the described data and processing pipeline, we provide an outline of each of the steps in Figure 1, noting the section number where each step is described in detail, and (where applicable) the name of the provided software examples that perform each step.

---

2 http://www.lmsal.com/hek/index.html
1) **Data Acquisition** (Secs. 2,3) [euv2chm_data_prep]
   - Obtain level-1 preprocessed EUV disk images using SSW in IDL.
   - Sample data formatting script made available.

2) **PSF Deconvolution** (Sec. 3.1)
   - Use publicly available PSFs and deconvolution algorithms to remove scattered light.

3) **Limb-brightening Correction** (Sec. 3.2) [euv2chm]
   - Flatten natural center-to-limb variation using a data-driven approach.
   - Correction data made available.

4) **Inter-instrument Transformation** (Sec. 3.3) [euv2chm]
   - Normalize intensities between the 3 instruments using a data-driven approach.
   - Transformation data made available.

5) **Coronal Hole Detection** (Sec. 4.1) [ezseg,euv2chm]
   - Detect the CHs using a region growing algorithm.
   - Software (EZSEG) made available.

6) **Mapping** (Sec. 4.2) [euv2chm]
   - Merge the 3 results to generate synchronic EUV and CH maps.
   - Final maps for 4 years at 6-hour cadence are made available.

Figure 1. Workflow to generate synchronic EUV and CH maps (examples are from 2011 February 03). The section describing the step and the name of the provided code example are indicated.
2. OBSERVATIONAL DATA

We use images from the Fe xiii 195 Å and 193 Å channels from the STEREO-EUVI and SDO AIA instruments, respectively, for our study. Fe xiii imaging observations have historically been used to detect CHs because the combination of high contrast and line strength (high count rates) make them ideal for CH detection. For example, standard differential emission measure (DEM) models employed by O’dwyer et al. (2010) predict around a ten-fold difference in AIA 193 Å photon count between typical CH and quiet-Sun regions. Although there are subtle differences in the temperature response functions of the AIA 193 Å and EUVI 195 Å channels, they are similar enough to be used together to generate synchrotron maps with proper preprocessing (see Section 3). To simplify notation, throughout this paper, we will use STA, STB, and AIA to refer to STEREO-A EUVI 195 Å, STEREO-B EUVI 195 Å, and SDO AIA 193 Å, respectively.

We obtain the EUV images from AIA and EUVI using SolarSoftWare (SSW; Freeland & Handy 1998) in IDL. For a given date range and cadence, the Virtual Solar Observatory and Joint Science Operation Center are queried to find sets of STA, STB, and AIA science images whose observation times are within a maximum time interval of 30 minutes (typically within 30 s) of each other. For this study, we obtain image sets at a 6 hr cadence from 2010 June to 2014 October, which yields a total of 6051 frames for each instrument.

Where possible, bad data frames are detected and replaced by nearby frames by checking header data. However, some bad EUVI data frames get through this process, so that they must be manually replaced, and/or properly handled by the processing codes. Most bad images can be detected automatically, but some (such as the rare out-of-focus images) are missed, requiring a visual inspection to identify them. These frames are then skipped in our processing and detection steps.

3. PREPROCESSING

Since CHs are identified as low-intensity regions in EUV images, it is very important to minimize intensity variations across the disk, both instrumental (noise, scattered light, scaling, etc.) and physical (limb brightening, line-of-sight obscuration, etc.).

We start by calling the standard image reduction and processing routines in SSW. EUVI images are reduced using seccchi_prep.pro, which includes exposure time correction and image registration, flat field correction, unit conversion, and bias. The AIA images, which are only available in a calibrated Level 1 format, are registered and exposure normalized using aia_prep.pro and binned to match the EUVI resolution (2048 × 2048). We refer to the results of this preliminary processing as the “original” line-of-sight disk images. The default units in SSW for EUVI data are photons/s, while those for AIA are DN/s. These units are similar enough to display the images using the same color scale (especially after applying our IITs (see Section 3.3)). Therefore, throughout this paper, we display the images (in log10) using the standard EIT green color table spanning image values between 0 and 3.

For images that undergo deconvolution (see below), the deconvolution step must be applied after calibration but before rotation to align the image with solar north. The images then receive the limb-brightening correction, and are scaled to each other as described in the remainder of this section. We note that the application of noise-reduction techniques to the images, such as the ongoing work by Kirk & Young (2014) for AIA, would also be a potentially beneficial preprocessing step. However, applications of sophisticated noise-reduction techniques to solar EUV imaging data are not currently available, and developing them is beyond the scope of this study. We also note that the image data from STA and STB are known to suffer from ICER compression artifacts, especially in dark regions off the limb. Although we have observed some small artifacts in the darker regions of CHs, their presence does not appear to have any noticeable effect on our data pipeline or CH detection.

The end goal of this section is to transform/flatten the EUV images such that every disk pixel appears (as much as possible) as if it were observed on the Sun at the disk center with a single idealized lens. This flattening of the images allows us to create synchrotron EUV maps and greatly assists our CH detection algorithm in Section 4.1.

3.1. PSF Deconvolution

Although many instrumental effects are removed from the images through the standard Level 1 preprocessing, optical effects such as scattered light, diffraction, and others remain. These can cause many CH detection algorithms to be less effective. For example, unwanted stray light from brighter regions can decrease the relative intensity contrast between CHs and nearby regions of the quiet-Sun, and cause CHs to be less uniformly dark. Some of these optical effects can be removed by deconvolving the images with the instrument’s PSF. Such PSFs can only be approximated and the deconvolution process is an ill-defined inverse problem. However, deconvolving the images with an approximate PSF can still greatly improve the image in the ways we desire for CH detection.

A full discussion for in-flight PSF estimation and correction using lunar transients for the EUVI and AIA instruments can be found in Shearer et al. (2012) & Poduval et al. (2013), respectively. For our purposes, the most relevant correction comes from the estimation of the stray light profile of the PSF. For EUVI, we obtain estimated PSFs tailored to each spacecraft (A and B) derived by Shearer (2013) from within the euvi_deconvolve.pro package accessible in SSW. For AIA, we use the freely available PSFs derived by Poduval et al. (2013). All of these PSFs contain an extended profile away from the PSF core that helps to correct stray light. Research on the formulation of PSFs for EUVI and AIA is ongoing (for example, see Gonzalez et al. 2016) and these efforts have the potential to improve further the correction of the images.

Once the estimated PSFs are obtained, the EUV images must be deconvolved with them. Due to the large number of images to be processed, we require a fast and efficient algorithm to reduce the overall processing time. To this end, we use the SGP deconvolution algorithm from Bonnetti et al. (2009) 3, which offers a speed improvement with similar results when compared to the popular Richardson–Lucy deconvolution algorithm (Lucy 1974). A stand-alone C and CUDA-C version of the SGP implementation is under development by the

---

3 See Prato et al. (2012) for the description of a freely available IDL implementation of the SGP algorithm (with support for the third-party IDL package GPULib: http://www.txcorp.com/home/gpulib).
authors of Prato et al. (2012), and binaries of the new codes were graciously provided to us, allowing us to utilize Nvidia GPUs for our processing (independent of IDL or third-party software). For example, deconvolving a single 4096 × 4096 image takes ≈7.5 minutes using the standard Richardson–Lucy routine in IDL SSW, ≈3 minutes using the SGP IDL scripts, and only ≈4 s using the stand-alone GPU SGP code. Using single precision drops the GPU compute time down further to only ≈2 s per image (without noticeable loss in image quality). We also note that a new GPU code for computing the Richardson–Lucy deconvolution for AIA is in development (Cheung 2015) that can process an AIA image in less than a second and will be incorporated into the standard AIA data pipeline when completed. This will eliminate the need to separately deconvolve the AIA images as done here.

The PSF deconvolution is applied to the calibrated images before rotation. The advantage of the deconvolution step as a contrast enhancement for CH detection is illustrated in Figure 2. Here, we show context images and slices through selected CHs for STA and AIA. While the nominal emission values for the quiet Sun and nearby brighter regions remain similar, the stray light correction removes contaminating emission from the darkest portion of CHs. This increases the relative contrast ratio between CHs and neighboring features. Note that because the base optical properties between the AIA and EUVI telescopes differ considerably, the relative contrast enhancement gained from PSF deconvolution is not expected to be the same.

3.2. Limb-brightening Correction

Because the extended EUV emitting portion of the corona is (mostly) optically thin, solar EUV images are affected by limb brightening due to the increasing line-of-sight path length toward the limb. Thus, the same solar feature observed near the disk center of the image will tend to be brighter in intensity near the limb. Since many CH detection algorithms make use of intensity thresholds (including ours), such variation in intensity can be problematic, particularly since we intend to track the evolution of CHs as they rotate across the solar disk and from one instrument viewpoint to the other.

In order to equalize (or flatten) the intensities across the disk, we seek limb-brightening correction curves (LBCCs) that will convert intensities observed at any position on the disk to their disk center equivalent. We choose to define the LBCCs as a function of μ, where

\[ \mu = \cos \theta, \quad \theta \in [0, \pi/2], \]

where \( \theta \) is the angle from the disk center (\( \mu = 1 \)) to the limb (\( \mu = 0 \)). We note that the “limb” here is defined as the disk edge in the coronal EUV images that, due to the height of the chromosphere and transition region, differs slightly from the photospheric limb (see Section 4.2 where this difference affects how we map the images). Throughout this paper, we choose a value for the base coronal radius of \( R_0 = 1.01 R_\odot \). We also choose to describe the LBCCs in terms of the \( \log_{10} \) of image intensity because EUV images are typically viewed in log-space.

Assuming that the rate of the increase in intensity from the disk center to the limb is independent of the disk center intensity (\( I_0 \)) itself (for example, in the case of a spherically symmetric solar atmosphere with constant temperature) leads to LBCCs of the form

\[ L(\mu) = \log_{10} \left( \frac{F(\mu)}{F_0} \right) = I(\mu) - I_0, \tag{1} \]

where \( F_0 = F(\mu = 1) \) is the image intensity at the disk center, and \( I = \log_{10} F \). Given an image of \( \log_{10} \) intensities \( I \), applying the LBCC is simply described as

\[ I'(\mu) = I(\mu) - L(\mu), \]

where \( I'(\mu) \) is the corrected image. The \( I_0 \)-dependent assumption of the LBCC of Equation (1) can, in some cases, be an adequate approximate limb-brightening correction, but we use it here to motivate and illustrate our data-derived approach for obtaining an \( I_0 \)-dependent correction for limb brightening.

In the following sections, we examine analytic approximations to LBCCs in the case of hydrostatic equilibrium, and then move on to describing the formulation of data-derived LBCCs, starting with our data selection process. Our final preferred method for limb-brightening correction—that of an \( I_0 \)-dependent limb-brightening correction image transformation—is then

---

*The CPU used for the timing was a Xeon W3680 3.3 GHz, while the GPU used was a GTX970.*
described in detail, which results in a more accurately flattened image and enhanced contrast.

3.2.1. Theoretical Limb-brightening Correction Curves

If one assumes an idealized single-temperature hydrostatic atmosphere and a weak dependence of the instrument’s temperature response function on density, a theoretical LBCC can be formulated (see Appendix) as

\[ L(\mu; T_e) = I(\mu; T_e) - I_0(T_e), \]  

(2)

where

\[ I(\mu; T_e) = \log_{10} \left( \int_0^\infty \exp \left( -2 \frac{R_0}{r} \frac{r - R_0}{\lambda(T_e)} \right) \, dx \right), \]  

(3)

and

\[ \lambda(T_e) = \frac{k_B T_e}{g_0 \mu m_p} \]

is the gravitational scale height, \( r = \sqrt{x^2 + 2 \mu R_0} \) is the distance to solar center, \( R_0 \) is the approximate radial distance of the base of the corona, \( x \) is the distance from the observer to \( R_0 \), \( \mu \) is the mean molecular weight (we use the coronal value of \( \mu = 0.6 \); Priest 2014), \( k_B \) is the Boltzmann constant, \( m_p \) is the proton mass, and \( g_0 = GM_s/R_0^2 \). It is important to note that under these assumptions, this LBCC over \( \mu \) is independent of the base density and of the temperature response function, and is only dependent on the temperature itself.

In Figure 3, we show the theoretical LBCCs computed numerically from Equation (2) for temperatures between 0.5 and 5 MK visualized as a function of \( \mu \) (left) and \( T_e \) (right). We see that the effect of limb brightening is much larger toward the limb than near the disk center and decreases overall with increasing temperature. It is also noticeable in the right plot that the temperature dependence of the curves can be well approximated by a linear function over small ranges of temperature (see Section 3.2.4 where we make use of this approximation).

Using theoretical LBCCs on the EUVI and AIA images is problematic however, because for any given EUV image of the Sun, distinct coronal structures will have different temperatures and densities, both of which are unknown. Furthermore, the total emission for a given pixel is a convolution of the differential emission measure distribution (DEM) of the plasma along that line of sight with the temperature response function of the given filter. Therefore, a limb-brightening curve of a single temperature will not be representative for the center-to-limb variation of all visible structures. For purposes of CH detection, one could choose a curve corresponding to the estimated average temperature of a CH (say 0.5–0.8 MK), but this would not be optimal either because it would over-correct the center-to-limb variation of hotter structures in the image, thus reducing their contrast with CHs. Additionally, even for a single temperature, additional effects such as opacity near the limb due to resonant scattering (Schrijver et al. 1994; Schrijver & McMullen 2000) are not taken into account in the optically thin model for limb brightening.

For these reasons, we opt for a more data-driven approach to model the center-to-limb intensity variance guided by the properties of the theoretical limb-brightening curves described above. The data-driven nature of the model is advantageous as it will inherently include effects not captured in the analytical model and can be used for the full range of structures within the EUV image.

3.2.2. Data Selection

To obtain data-derived limb-brightening corrections, we start by extracting intensity values as a function of \( \mu \) from the image. Due to solar rotation and the inherent inhomogeneity of
the corona, one must take care when determining the temporal and spatial selection of the data to be used.

To mitigate selection bias and to better sample variations of coronal structures, we average the image data over time. We have found that a one-year average is optimal for the time period we studied. This averaging window was chosen by observing the variations in the computed LBCC values (using the methods described in this section) as the averaging window was increased. The window size was fixed when the variations in the correction parameters were observed to be converging to a constant amplitude.

The discrete nature of the images require the selection of bins of image data to represent the intensity values \( I \) over \( \mu \). Using bins of data also helps to avoid the effects of noise and small transient structures exhibiting large intensity variation. A seemingly good choice for the \( \mu \)-bin locations/sizes is such that an equal area of data is contained in each bin. This could be done for the full disk for \( n \)-bins by using \( \mu \)-bin locations defined by \( \mu_i = \sqrt{1 - (i - 1)/(n - 1)}, i = 1 \rightarrow n \). However, such a choice is not optimal because solar rotation implies that structures visible at high latitudes will not rotate through all of the possible center-to-limb angles, causing them to contribute unequally to the \( \mu \)-bins. Instead, we require that all \( \mu \)-bins span the same range in heliographic latitude. In practice, this is done by determining the projected latitudes based on the current \( B_0 \) angle of the spacecraft and restricting bins to a strip that intersects the disk center such that all \( \mu \)-bins cover the same range of heliographic latitude during the data averaging window. We have found that a latitude limit of \( \pm \pi/64 \approx 2.8' \) from the center and bin sizes of \( \Delta \mu \approx 0.06 \) outside the center bin yield acceptable results for the time period discussed in this paper. It is also important to note that because the emission properties of the image understandably change rapidly past the photospheric edge, we only collect data up to \( r = R_\odot \), which, for our choice of \( R_0 = 1.01 R_\odot \), is represented by the \( \mu \) value of

\[
\mu(r = R_\odot) = \sqrt{1 - R_\odot^2/R_0^2} = 0.14037.
\]

Figure 4 shows examples of the latitude-limited \( \mu \)-bins used.

3.2.3. Data-derived LBCCs

The simplest way to obtain an LBCC from the \( \mu \)-bins obtained in Section 3.2.2 is to measure the average intensity inside each of the \( \mu \)-bins over the time window \( I(\mu) \), and subtract the averaged intensities at the disk center bin \( I_0 \); i.e., directly applying the definition of the LBCC of Equation (1)). In Figure 5, we display example LBCCs computed in this manner for each of the three instruments. The curves are seen to be very similar in form to the single-temperature LBCCs of Equation (2). We also note that the LBCC (at least for the time chosen) of AIA differs noticeably from those of STA and STB. This is likely due to the differences in their respective response functions (also shown in Figure 5) as well as differences in the resolution and stray light profiles of the telescopes.

The data-derived LBCCs are still \( I_0 \)-independent, but they are expected to be more accurate than using the theoretical model of Section 3.2.1 because they capture effects not included in the analytical model (such as opacity, density variations, etc.). Also, because they represent the mean convolved limb-brightening rate of all of the multithermal structures, they eliminate the need to arbitrarily choose a mean coronal temperature as is required when using Equation (2). Despite these improvements, we wish to further enhance the technique to treat specific coronal structures independently (i.e., an \( I_0 \)-dependent correction). This can be achieved using the information present in the full histogram of intensities within each \( \mu \)-bin. In the next section, we explore how image transformations can be used to closely match these histograms.

3.2.4. Intensity-dependent Limb-brightening Correction

Up to now, the LBCCs have been applied as a transformation of the \( \log_{10} \) image described as

\[
I_0 = I(\mu) - L(\mu),
\]

i.e., given an image pixel at a \( \mu \) level, one subtracts the LBCC \( L(\mu) \) to get the pixel’s approximate value as if it were at the disk center. Thus, every pixel at the same \( \mu \) level is corrected with the same \( L \) value. To remove this constraint, we can instead use a nonlinear (linear in log-space) image transformation described as

\[
I_0 = \beta(\mu) I(\mu) + y(\mu).
\]

This transformation is equivalent to using a unique LBCC for each and every possible disk center intensity \( I_0 \) value \( (I_0 \)-dependent), which can be expressed by combining Equations (4) and (5) to yield

\[
L(\mu, I_0) = \left( \frac{1}{\beta(\mu)} - 1 \right) I_0 - \frac{y(\mu)}{\beta(\mu)}.
\]

The form of the transformation of Equation (5) is chosen primarily because it is the lowest-order nonlinearity that can modify both the position and the shape of the \( \mu \)-bin histograms to match the central histogram. However, we observed that the temperature dependence of the theoretical LBCCs in Figure 3 can be well approximated as a linear function of temperature for each \( \mu \) value. Assuming a correspondence between image intensity \( (I_0) \) and temperature (see Appendix), the linearly \( I_0 \)-dependent form of Equation (6) can be viewed as a first-order approximation of the temperature dependence of the theoretical LBCCs.

In order to find the optimal values of \( \beta(\mu) \) and \( y(\mu) \) over time, we perform a two-dimensional least-squares optimization between the normalized histograms at each \( \mu \) level (\( H(I; \mu) \)) to
the disk center normalized histogram \((H(\mu; \mu))\). The histograms are normalized by dividing the bin sample value by the total number of samples. It turns out that this is not a well-behaved optimization because there is a fairly large parameter space in which the histograms match very well. To illustrate this, we show the sum-of-squares error (SSE) for a typical histogram-matching optimization in Figure 6. We see that although there is a unique minimum, there is a large parameter space where the error is small and produces visibly satisfactory histogram matches. Thus, we expect our computed \(\beta(\mu)\) and \(y(\mu)\) curves to suffer from some local variations over time and over \(\mu\) (as can be seen in Figure 8), yet still produce consistent quality transformations. Due to the discrete and sometimes noisy nature of the histograms, many classic optimization algorithms that require smoothness have convergence problems. Therefore, we use the derivative-free Nelder-Mead Simplex Method, which is shown to have good convergence properties in two-dimensional problems such as ours (Lagarias et al. 1998).

In Figure 7, we show surface plots of the normalized one-year averaged intensity histograms \(H(I; \mu)\) for the STA disk images centered at 2011 February 03 before and after the transformation of Equation (5) over all of the \(\mu\) values. For comparison, we also show the result of merely shifting the histograms for each bin along the intensity axis to match the position of the center bin histogram (i.e., applying an \(I_0\)-independent LBCC of Equation (1)). Due to the amplitude and shape differences of the histograms over the \(\mu\)-bins, the \(I_0\)-dependent image transformation creates a much more

---

**Figure 6.** Left: an example of the sum-of-squares error (log) in the \(\beta - y\) parameter space for the difference between a normalized histogram \(H(I; \mu = 0.532)\) transformed by Equation (5) and the central normalized histogram at \(\mu = 1\). Although there is a unique minimum, there is a large parameter space where the error is small and produces visibly satisfactory histogram matches. Right: original and transformed histograms for the \(\beta - y\) values indicated in the left figure.

---

**Figure 5.** Left: LBCCs for 2011 February 03 for STA (black dashed), STB (blue dotted–dashed), and AIA (red solid) computed with average intensities over one year at 6 hr cadence. The curves were computed with the \(\mu\)-bin locations and latitude limit level shown in Figure 4. The theoretical curves of Equation (2) (green dashed lines) for \(T \in [0.5, 1.25, 2, 3.75]\) MK (top to bottom) are shown for comparison. Right: normalized response functions for STA/B and AIA at a number density of \(n_0 = 5 \times 10^9\) cm\(^{-3}\). The response functions are determined using calibration data available in SSW and synthetic spectra generated from the CHIANTI 7.1 database.
consistent histogram along $\mu$ than that of Equation (1), resulting in a better limb-brightening correction of the images.

3.2.5. Limb Brightening Results and Analysis

We have computed and made available (see Section 5) the optimal $\beta(\mu)$ and $y(\mu)$ values (based on one year running-average data) for use with the LBCC image transformation of Equation (5) from 2010 December 10 to 2014 February 17 at 6 hr cadence. An example of the $\beta(\mu)$ and $y(\mu)$ values over $\mu$ is shown for 2011 February 03 in Figure 8, while the computed values over the entire time period are displayed in Figure 9. Each instrument results in similar values of $\beta(\mu)$ and $y(\mu)$, but exhibits variation over time. These changes emphasize the need
for a data-driven approach in order to obtain optimal corrections that respect changes in the images over time. Since the LBCC factors are only formulated up to $r = R_\odot$, when applying the corrections to the EUV images, we extrapolate the curves to correct the pixels in the small region between $r = R_\odot$ and $r = 0$.

An illustration of the effect of applying the transformation of Equation (5) to an EUV image containing an extended CH near the limb is shown in Figure 10. A color plot of the difference between the original and corrected images (i.e., the effective LBCCs) is also shown. We see that, as desired, the CH in the image has a more consistent intensity level from the limb toward the disk center after the correction. Also, it is apparent that the correction is highly structure/intensity dependent, and that brighter structures are darkened less than darker structures, which increases the overall contrast (a desirable trait for CH detection). An advantage of this contrast enhancement over those of standard image processing techniques is that it is both
unit-preserving and respects the physical nature of the limb-brightening problem.

We now explore the ability of the $I_0$-dependent image transformation to represent LBCCs for various disk center intensities. In Figure 11, we show four generated LBCCs for 2011 February 03 computed with Equation (6) from the values of $\beta(\mu)$ and $\gamma(\mu)$ from Figure 8. The curves are generated using $I_0$ values chosen such that they match closely to theoretical LBCCs of selected temperatures. We see that the LBCCs for increasing values of $I_0$ are smoothly varied in the direction of the theoretical LBCCs computed with Equation (2) for increasing temperature, and that they closely match the shape and range of the curves. This qualitatively fits with the expectation that structures that vary in EUV intensity levels (low to high), such as CHs, quiet Sun, and ARs, are similarly ordered in temperature (low to high). This idea is shown visually in the right panel of Figure 11 where a corrected AIA EUV image is displayed with contours of the chosen $I_0$ values.

The general correlation of different temperature curves to different $I_0$ values suggests that one could possibly employ this technique to serve as an indirect statistical estimate of average coronal temperatures based on images at a single wavelength. We plan to explore this concept in the future.

### 3.3. Inter-instrument Intensity Transformations

After completing the steps described in Sections 3.1–3.2, the preprocessed EUV images are now suitable for performing CH detection. However, in order to make synchronic maps using multiple instruments, it is highly desirable to derive IITs which allow us to equate the intensities from one instrument to another. In addition to being a useful data product on its own, we find this helpful for two reasons: first, a proper scaling will allow one to employ the same thresholding parameters in the
CH detection algorithm, basically eliminating the need to find the separate “best” values for all three instruments. This also helps us to ensure that the algorithm will detect a CH with the same tolerance in each image. Second, scaling the images to each other allows one to create continuous synchronic EUV intensity maps from the three instruments, which helps for tracking time-dependent features as they rotate or propagate from one viewpoint to another. This is very useful when judging the quality of the CH detection, as well as tracking CH evolution across instruments (see Section 5).

Transforming one instrument’s EUV image so that it appears as if another instrument has viewed it is very similar to the task of the LBCC calculations in Section 3.2 performed on each \( \mu \) section. Therefore, the same techniques can be applied and we once again take a data-driven approach. In this case, there is no a priori reason to assume that a nonlinear transformation like Equation (5) is necessary, and perhaps a linear histogram shift (or average intensity comparison) is all that is needed. However, we have found that, similar to the LBCCs, the shapes of the histograms from different instruments do not match reliably and, consequently, the results based on linear histogram shifts are not ideal. We find once again that a transformation of the form

\[
I_a = \alpha I_b + x,
\]

where the subscripts \( a \) and \( b \) represent the log10 images of instrument \( a \) and instrument \( b \), provides a better match between the histograms of the multiple instruments (see Figure 12).

When using linear histogram shifts or average intensity differences, the order of operations of LBCC and inter-instrument scaling is arbitrary and each can be applied separately. However, due to the combination of the nonlinear nature of the transformations that we choose to use in each step, the LBCC and inter-instrument scaling cannot be applied independently. Therefore, one must apply one correction to the EUV images, and then use the resulting images to compute the second correction factors. We choose to first apply the limb-brightening corrections computed in Section 3.2 to the set of EUV images, and then use the resulting images to compute the proper \( \alpha \) and \( x \) correction factors for the inter-instrument corrections.

In order to ensure that the spacecraft are observing more or less the same portions of the Sun during a given averaging window, we shift our averaging window in time corresponding to the separation in Carrington longitudes of the spacecraft’s position. Additionally, we limit the latitude of the disk data based on the \( B_0 \) angles of each instrument such that each instrument sees the same area. For our data set, the limit needed is \( \pm 75.3 \) from the center.

Figure 12 shows an example of the histograms of one-year averaged data centered around 2011 February 03 for STA, STB, and AIA both before and after the correction of Equation (7) is applied. Also shown are selected quiet-Sun portions of each image at the same date. We see that the correction factors match the averaged histograms nearly exactly, and when applied to the EUV image, cause the three instruments’ image intensities, on average, to match very closely.

We have computed the IIT factors for transforming STB to STA and AIA to STA from 2010 December 21 to 2014 February 17 at a 6 hr cadence using a one-year averaging window. The factors are computed using the EUV images after the application of the limb-brightening correction of Equation (5) with the computed LBCC factors from Figure 9. The results are shown in Figure 13. We see that, like the LBCC factors in Figure 9, the IIT factors vary over time. Although the changes are small, they are within the sensitivity range of the threshold values used in our CH detection algorithm, re-enforcing the need for a data-driven approach.

### 3.4. Validation of Preprocessing Results

The overarching goal of our preprocessing is to transform multi-instrument EUV intensity measurements to a common vantage point and unit scaling as best we can (Sections 3.2 and 3.3). This is not perfectly achievable because of inherent geometric and observational limitations such as projection.

Figure 12. Example of the normalized histograms of one-year 6 hr cadence averaged data centered about 2011 February 03 for STA, STB, and AIA before (a) and after (b) the correction of Equation (7). The histograms use 400 intensity bins, a latitude limit of \( \pm 75.3 \) from center and an optimization relative tolerance of \( 10^{-4} \). The inset shows selected quiet-Sun portions of each instrument’s image at the central date before and after applying the correction.
effects, time synchronization differences, obscuration, variable apparent coronal radius, etc. However, we can still characterize the performance of our preprocessing by examining overlapping observations of the spacecraft.

We start by taking EUV images from each instrument and mapping them to a common heliographic grid (Section 4.2). Then, we determine the midpoint \( m \) value of the overlap region between two instruments, \( m_0 \), and select regions for inter-comparison by extracting overlapping data within a small strip defined by \( m_0 \pm \Delta m \). Choosing a small \( \Delta m \) restricts the variation from the mean line-of-sight angle and helps limit projection effects (we choose \( \Delta m = 0.05 \)). The value of \( m_0 \) changes naturally over time due to the variation of the orbital position and orientation of the spacecraft (see Figure 14). These variations affect our analysis by changing the mean line-of-sight angle in the overlap.

Denoting the mapped intensity data in each \( \Delta m \)-limited overlap region as \( J \), we first explore how average intensity values change before and after our final IIT preprocessing step. Defining the percent difference of the mean intensity as

\[
D_{pdm} = 100 \times \frac{T_m - T_b}{T_b}, \quad T = \frac{1}{N} \sum_i J(i),
\]

we compute \( D_{pdm} \) for every map and overlap pair in our data set. The average of \( D_{pdm} \) before and after applying the IIT for all pairs with \( m_0 > 0.4 \) is shown in Table 1. The corresponding standard deviations for these data are also shown.

We find that the IIT dramatically improves the average value of \( D_{pdm} \), reducing the percent difference from over 10% to less than 1% for the STA–AIA and STB–AIA pairs. For STA–STB, this difference drops from 4% to less than 1%. This is not surprising because the mean is mostly sensitive to the relative scaling between the instruments, but it does confirm that the IIT is serving its purpose of normalizing intensities between the spacecraft. The standard deviations of \( D_{pdm} \) do not change substantially but remain relatively low within \( \pm 6\% \). This indicates that a baseline level of variation is present in the data. The presence of variations is reasonable because the corona is an extended, optically thin medium, and each spacecraft will naturally have an oppositely directed perspective through it in these overlap regions.

We also compute one-year running averages of \( D_{pdm} \) over the entire data set and see a similar story. These curves are

![Figure 13. Double-y axis plots of the computed inter-instrument transformation factors \( \alpha \) and \( \psi \) of Equation (7) for transforming STB images to STA (blue dashed line) and AIA images to STA (red solid line) from 2010 December 21 to 2014 February 17. The computational parameters are the same as in Figure 12.](image-url)

![Figure 14. Midpoint \( m \) value of the overlap region, \( m_0 \), as a function of time for each instrument pair. The pairs are STA–AIA (black), STB–AIA (blue), and STA–STB (red).](image-url)

| Instrument Pair | Before IIT Mean | \( \sigma_{ad} \) | After IIT Mean | \( \sigma_{ad} \) |
|-----------------|----------------|----------------|----------------|----------------|
| AIA–STA         | −10.81\%       | 5.25\%         | AIA–STA        | 0.15           | 0.042          |
| AIA–STB         | −13.80\%       | 5.60\%         | AIA–STB        | 0.18           | 0.043          |
| STA–STB         | 4.03\%         | 6.13\%         | STA–STB        | 0.11           | 0.042          |

Table 1: The Mean and Standard Deviation (\( \sigma_{ad} \)) of \( D_{pdm} \) and NRMSD Computed in the \( J \) Overlap Regions for Each Instrument Pair with \( m_0 > 0.4 \) Over the Entire Data Set Before and After Applying the IIT.
shown for each instrument pair in the top row of Figure 15, with the dashed lines indicating where $\mu_0 < 0.4$ (poor viewing). Before applying the IIT, the mean curves all have their own non-zero offset and these offsets are removed afterwards. Even more importantly, we see a non-zero slope in some of the curves before applying the IIT (particularly relative to AIA) and find that these slopes are substantially reduced afterward. This indicates that the time-dependent calculation of the IIT is warranted and that the transformation is working properly.

Next, we examine the normalized root-mean-square-difference (NRMSD) between all of the pixels inside an overlap region:

$$\text{NRMSD} = \frac{1}{J_{\text{max}} - J_{\text{min}}} \sqrt{\frac{1}{N} \sum_{i} (J_a(i) - J_b(i))^2}. \quad (9)$$

The NRMSD can be thought of as a general quality indicator of the pixel-to-pixel match between instruments. For consistent normalization, we choose $J_{\text{max}} - J_{\text{min}} = 4$ to be the approximate total intensity range for all images.

Intrinsic differences due to opposed viewing projections through the corona are expected to contribute significantly to the NRMSD. This means that even with perfect limb brightening and intensity corrections, we do not expect the NRMSD to trend to zero. Despite this fact, the NRMSD can be a useful indicator of preprocessing quality by comparing the values of the three overlap pairs to each other (STA–STB, STA–AIA, STB–AIA), mainly because we should expect consistent results between pairs for a common $\mu_0$.

We give the complete time average and standard deviation of the NRMSD before and after applying the IIT in Table 1 (again for $\mu_0 > 0.4$), and show the corresponding one-year running averages in the bottom row of Figure 15. As with $D_{\text{pdm}}$, the average NRMSD values are reduced with the application of the IIT. Most importantly, we see that the mean of the NRMSD for all three instrument pairs becomes much more consistent with one another, suggesting that their variations also become similarly distributed.

This consistency can be explored further by looking at the one-year averages of NRMSD versus $\mu_0$. Before applying the IIT, the curves do not share a common value or slope, but afterward they converge to one another, especially in the range $\mu_0 = 0.4–0.6$. A slope is expected even after applying the IIT in this case because the change in mean projection angle with $\mu_0$ will influence the pixel-to-pixel variation even if the overall intensity scaling is identical. Unsurprisingly, the NRMSD is worse with a poor viewing angle and improves as $\mu_0$ increases. Finally, the difference between the STA–STB pair (red) and the STA/B–AIA pairs (blue/black) around $\mu_0 = 0.7–1.0$ is not unreasonable considering that these favorable elongations occur at different times, once near solar maximum for STA–STB (2013 onwards) and once near solar minimum for the pairs with AIA (2010–2011).

Despite the systematic limitations of the problem, we have shown with these diagnostics that our preprocessing pipeline produces a common intensity scaling for all of the spacecraft pairs, and that the relative scaling and measurement variance is small and consistent between them (at least in a statistical or average sense). With the preprocessing of the images...
completed, we can now move on to the next step in our data pipeline.

4. CORONAL HOLE MAP GENERATION

After the preprocessing steps of Section 3 are completed, the images are ready to be used with a CH detection algorithm. Although there are many segmentation algorithms available in the image processing community (as discussed in Section 1), they are often not suited for CH detection or are difficult to obtain and/or use. We therefore have written our own segmentation algorithm modular code called ezseg which utilizes a dual-threshold, region-growing-style algorithm with variable connectivity written in FORTRAN-OpenMP and C-CUDA. We also provide MATLAB interfaces to ezseg for fast and easy use. The resulting maps for each instrument are then merged to form the synchronous CH maps using MATLAB scripts.

4.1. Coronal Hole Detection Algorithm: ezseg

As discussed in Section 1, the simplest technique for image segmentation is intensity thresholding. Due to the transient nature of structures on the Sun in both intensity levels and abundances, automatic threshold selection algorithms such as the popular “graythresh” (Otsu 1979) algorithm are not useful in this context. Additionally, simple thresholding for CH detection itself is problematic due to the overlap of the tails of the intensity distributions of CHs with other features such as dark filament channels and dark regions of the quiet Sun. An example of this is shown in Kirk et al. (2009), where thresholding is used to capture polar CHs, and in the process detects much of the quiet Sun (in that study, the authors were only concerned with polar holes and used a unique tracing algorithm so quiet-Sun detection was not a problem). Some have relied on using magnetic field data to exclude such features (Lowder et al. 2014), but it is important that our technique not rely on such data because photospheric magnetic field observations are not made by the STEREO spacecraft, and a key requirement is that our technique takes advantage of the instantaneous (near) full-Sun EUV coverage provided by these spacecraft. We have found that the combination of the preprocessing steps of Section 3 with a region-growing style algorithm (described below) is typically sufficient to distinguish CHs from other structures without the need of magnetic data.

Our algorithm is a simplified region-growing algorithm using one threshold ($t_1$) for seed placement and a second value ($t_2$) for the stopping criteria. This is similar in spirit to the popular “flood-fill” or “magic-wand” tools that are common in image processing software. The use of region-growing avoids the false detection of disconnected dark areas of quiet Sun by only growing regions that are seeded with intensities corresponding to the darkest regions of CHs.

In our simplified implementation of region growing, we do not track/identify individual regions that then have to be merged, rather we only use local information to grow and stop growing the region. Therefore, the region average and detection history do not need to be stored in specialized data structures, and there is no order dependence allowing the algorithm to be formulated in an embarrassingly parallel manner. Although, for some cases, the form of the algorithm may be less efficient than standard methods, the parallelism allows us to implement the algorithm on both shared-memory processors and GPUs, allowing it to perform well for our needs. For example, a CH detection on the STA image ($2048 \times 2048$) on 2010 June 10 required 275 iterations, which took a total of 2.6 s using two threads on a Core i3 3.3 GHz CPU, and only 0.3 s using a GTX970 GPU (including data transfer time).

An important feature of the algorithm is to check the connectivity of a pixel in all eight neighboring pixels and enforce a connectivity condition that requires $n$ consecutive marked pixels for inclusion in the region. This helps to avoid false connectivity due to noise or “bleed-out” from small adjacent dark quiet-Sun pixel regions. Through testing, we have found that $n = 3$ yields the best detection results overall.

The algorithm starts with a blank CH map with pixel locations corresponding to missing/unused EUV data flagged. Each viable pixel location is independently checked to see if the EUV image at that location is below $t_1$, and if it is, then the pixel location in the CH map is marked. Then, in each subsequent iteration, each pixel location is checked and is marked as a CH pixel if it is above $t_1$, below $t_2$, and has the requisite number of consecutive neighboring CH pixels. This iterative procedure continues until no additional pixels are added to the CH map; Figure 16 shows a schematic depiction.

Figure 17 shows the advantage of using a region-growing approach versus single-value thresholding. The green areas are the result of using a threshold of $t_1$, the red areas of using only $t_2$, and the black areas (which include green areas) are the result of the two-threshold region-growing algorithm. We see that the region-growing algorithm captures the CH boundaries but avoids adding the darker regions of quiet Sun that would be detected when using a single threshold of $t_2$.

The main challenge in using the algorithm is choosing $t_1$ and $t_2$ such that a desired CH map is produced over a large number of sequential frames. It is not clear if/how the CH intensities vary over the solar cycle (in which case the thresholds would need to be modified over time), and the choice of $t_2$ is...
somewhat subjective (indeed, the use of different color tables to display the images can change the perceived detection quality). There are different strategies for choosing $t_1$ and $t_2$. For example, the SWAMIS code uses preselected $\sigma$ values of the histogram distribution to select the values. However, since there may be no CHs in an image and/or large bright ARs, such a selection is not ideal for automatic CH detection over longer periods of time. In order to address this issue in the simplest way possible, we have manually selected optimal values for $t_1$ and $t_2$ for multiple times in the range of interest. This subjective process yielded no distinct trends in the way $t_1$ and $t_2$ changed over time, and the results were approximately constant over the five year period. We therefore have chosen to employ a single average value for each threshold ($t_1 = 0.95$ and $t_2 = 1.35$) for use with the entire time period. These values apply to preprocessed images (as described in Section 3) and yield good detection results overall. They can easily be modified as needed for improving detection for a specific data set, time, or time-range of interest. Also, we note that the ezseg algorithm is written in such a way that the user can manually (or though other automatic means) specify seed points in the input array, eliminating the need to choose a $t_1$ intensity value.

4.2. Mapping

In order to avoid possible detection errors introduced by mapping the disk images (DeForest 2004), as well as allowing the detection algorithm to yield no distinct trends in the way $t_1$ and $t_2$ changed over time, and the results were approximately constant over the five year period. We therefore have chosen to employ a single average value for each threshold ($t_1 = 0.95$ and $t_2 = 1.35$) for use with the entire time period. These values apply to preprocessed images (as described in Section 3) and yield good detection results overall. They can easily be modified as needed for improving detection for a specific data set, time, or time-range of interest. Also, we note that the ezseg algorithm is written in such a way that the user can manually (or though other automatic means) specify seed points in the input array, eliminating the need to choose a $t_1$ intensity value.

Figure 17. Left: coronal hole map for AIA on 2011 February 03 produced by thresholding of $t_1$ (green), thresholding of $t_2$ (red), and the ezseg algorithm (black). Right: corresponding preprocessed AIA EUV image. The ezseg detection result is shown as the magenta contour, while the additional false detections by only using a threshold of $t_2$ is shown as red contours. The $t_1$ and $t_2$ values used are 0.95 and 1.35, respectively.

such as those used here, the apparent radius of the line-of-sight disk image is slightly larger than $R_{\odot}$ due to the bulk of the emitting plasma residing in the lower corona (as shown in Figure 18). This is an important effect to keep in mind, especially when projecting multiple disk images to the same map, as we have found that the difference between the photospheric and lower coronal radii is large enough to cause overlapping mapped images to display a noticeable shift relative to each other (see Figure 18). Therefore, we use an estimate of the effective lower coronal boundary ($R_0 = 1.01R_{\odot}$) for all of the preprocessing and mapping steps.

Since it is often the case that images from two (or more) instruments will overlap in the map, it is necessary to select or merge the data to be used. We have implemented two methods of doing this. The first is simply to use the data that has the highest $\mu$-value associated with it, since the higher the $\mu$, the more direct the viewing angle and the less distortion there will be from the coordinate transformation. This can sometimes produce “cutoff” lines through the images and CH maps that sweep through the structures over time. This may happen when the CH line of sight becomes obscured by nearby quiet Sun or ARs, in which case, another instrument might have a better view of the CH even though its line of sight is closer to the limb. Therefore, we have implemented a merge mapping where each disk image is first mapped to its own Carrington map, and then the three maps are merged by taking the lowest intensity values of the overlap. The CH map data is then set based on which data were used for the EUV map. We find that this method produces EUV and CH maps that tend to be more continuous for dark structures at the seams and work well for the present context. An example showing the difference between the two types of map merging is shown in Figure 19. Of course, there are instances where a CH becomes obscured by intervening structures from all viewpoints and therefore has portions drop out from the detection; this is an unavoidable systematic error inherent to mapping the EUV corona from limited viewpoints. We note that the methods of merging the overlap regions discussed here are only a selection criteria of
which data to use, i.e., they do not modify any of the image or CH disk data.

One problem with either of the two mapping methods described above is that data that is very close to the limb is often quite distorted, and it may be desirable to not include it in the map. This can be avoided by setting a minimum cutoff for $\mu$, denoted $\mu_{\text{cut}}$, which only allows for data with $\mu \geq \mu_{\text{cut}}$ to be used. We find that a value of $\mu_{\text{cut}} = 0.4$ is a good choice for avoiding distortions while still retaining most of the data. For cases where one is willing to tolerate distorted data past the $\mu$-cutoff if it is the only data available, we can use two $\mu$-cutoffs: one for the merging regions with overlap, and the other for non-merging regions. We choose to use the maximum amount of available data (using a merge cutoff of $\mu = 0.4$ and a non-merge of $\mu = 0$) so that we can create EUV and CH maps with the best merged data in the overlapped regions, while still allowing the use of disk data that extends all the way to the limb/poles (the merged map in Figure 19 used these parameters).

5. SAMPLE RESULTS

Here, we display several examples of CH maps generated by ezseg using STA, STB, and AIA data, after preprocessing with the techniques described in Section 3.

In Figure 20, we show an assortment of synchronic EUV maps with an overlay contour of the detected CH maps within the five-year time period where STEREO-A/B and AIA were all active (a full set of such maps at 6 hr cadence from 2010 June 10 to 2014 August 18 are made available; see Section 6). The data-driven preprocessing allows for the production of nearly seamless full-Sun synchronic maps, and the CH detection performs well throughout the data set, even with the use of constant values for the threshold parameters, $t_1$ and $t_2$.

Our automated pipeline allows us to observe the evolution of CHs over time for longer periods of time than a single instrument’s images would allow. In Figure 21, we show the evolution of two CHs; one from a quiet-Sun time (2011 January 22–2011 February 09) and the other from a more active time (2013 May 27–2013 June 18). Since the current form of ezseg occasionally finds minor CH areas at the borderline of $t_1$ that flicker in and out, we display smoother 66 hr persistence maps. A persistence map is a floating point mean of a time-series of CH maps (see Lowder et al. 2014, for details) and, just like a running average, a time-centered persistence map can be created using neighboring data for any time in our data set. In this case, both CHs evolve over the time periods shown, with the CH from 2013 May 27 to 2013 June 18 showing a great deal of shape transformation and the CH from 2011 January 22 to 2011 February 09 appearing to rotate. Note that both evolutions were detected with the same CH detection parameters even though the time periods are over two years apart. Once again, this is only possible due to the data-driven preprocessing.
Time-dependent analysis of these maps allows one to track the properties of the CHs over time. An example is given in Figure 22 where we show the computed surface area of the CHs over time. The area of each CH, in units of $R_0^2$, is computed by isolating the region containing the CH and multiplying the area factor $\Delta \sin(\zeta) \Delta \phi$ by the total number of CH pixels. The CH area for 2011 January 22–2011 February 09 seems to be fairly constant with only a slow overall rise in surface area, even though the CH shape itself is evolving. It also seems that the minimum-intensity form of merging the CH maps does a better job of tracking the CH area through instrument transitions than the maximum $\mu$ approach (see Section 4.2). This conclusion is reasonable because it is unlikely for the CH area to exhibit sudden jumps only during
overlap transitions. In contrast, both merging methods appear to have trouble tracking the CH area for the other time period. This can be seen in the right area plot during the first overlap, which has dramatic up and down spikes in the trend. Observing the evolution (shaded frames of Figure 21(b)), this oscillation in area does not appear to be true CH evolution, but rather is due to geometric obscuration by surrounding quiet-Sun structures, first from one side and then the other. This issue is compounded by the very small overlap of the instruments at this time, which worsens the mean projection angle and increases the likelihood of obscuration from both sides.

Finally, in an attempt to illustrate CH evolution with a single visualization, we introduce a method for constructing time-colored evolution maps based on persistence data. Given the...
series of CH maps $C(x, y, t)$, where pixels identified as belonging to a CH are set to $C = 1$ and otherwise $C = 0$, each pixel in the time–color evolution map is colored using an HSV color table with

$$
H' = \langle t \rangle / \Delta t,
S = P,
V = 1 - P,
$$

(10)

where $\Delta t = t_{\text{end}} - t_{\text{start}}$ is the length of time of the CH evolution, $\langle t \rangle$ is the integrated mean time of the detection, and $P(x, y) \in [0, 1]$ is the normalized persistence map. $\langle t \rangle$ and $P$ are computed as

$$
\langle t \rangle = \frac{1}{P} \int_{t_{\text{start}}}^{t_{\text{end}}} (t - t_{\text{start}}) C(x, y, t) dt,
$$

$$
P(x, y) = \frac{1}{\Delta t} \int_{t_{\text{start}}}^{t_{\text{end}}} C(x, y, t) dt,
$$

and $H' \in [0, 1]$ corresponds the chosen range of hue values ($H$). Since hue is cyclical, we truncate the full range of hue values to start at $H = 242$ (blue) backward through to $H = 297$ (pink) leaving out the pink-to-blue transition values. Thus, blue color in the map represents the CH area near the start of the evolution time while pink represents the CH area near the end time. The map becomes darker and more color saturated in areas of high persistence, and is brighter and has less color saturation in areas of low persistence. For example, a CH traveling from left to right in the time period of the persistence map would appear as a streak of color ranging in blue on the left to pink on the right. In Figure 23, we show examples of time–color evolution maps for both CH evolutions from Figure 21. We also show a standard grayscale persistence map for the CH from 2011 January 22 to 2011 February 09 to illustrate the advantage of the new colorized map. One can see that the time-colored evolution map allows one to visualize the time-evolution of the CHs in a single glance. For example, the map for 2011 January 22–2011 February 09 seems to indicate that the central core of the CH remains steady throughout the evolution, while the outer areas are undergoing differential rotation. This can be seen qualitatively by plotting the latitude where the Sun’s rotation rate matches the...
Carrington rotation rate (≈26°) and noting that the CH clearly shows left-sided (slower) motion above the Carrington latitude line and right-sided motion (faster) below. In contrast, the CH from 2013 May 27 to 2013 June 18 does not show any discernible signal of differential rotation. However, this equatorial hole was present with nearby active regions emerging, evolving, and sometimes partially obscuring the CH. Further analysis of CH evolutions are underway and will be discussed in future publications.

6. SUMMARY, PRODUCTS, AND FUTURE WORK

In this paper, we have described a process to generate synchronic CH evolution maps from multiple EUV images. We applied the method to EUV images from STEREO-A/B EUVI 195 Å and SDO AIA 193 Å from 2010 June 10 to 2014 August 18, which represents the majority of the time span when STEREO-A/B and AIA were all simultaneously functioning.

The ability to produce automatic synchronic CH maps using the techniques described here allows for the analysis of CH evolution. The surface areas of CHs were tracked over a longer period of time than the images from a single instrument could provide. We then produced persistence maps of the CHs using a new time-colored method allowing for the trends in CH evolution to be observed in a single generated image. For a CH in the beginning of 2011, we observed a compelling signal of differential rotation in the CH. Further study of CH evolution using the detection methods presented in this paper are planned.

An important element in this work is the development and application of several preprocessing steps. These include PSF deconvolution, limb-brightening correction, and inter-instrument transformations that help to sharpen the images and reduce scattered light, equalize intensity values from the disk center to the limb, and equalize intensity values across instruments, respectively. In addition to their practical image processing utility, the limb-brightening correction results illustrate their sensitivity to the temperature structure of the corona, which we plan to explore further. Although we have only shown preprocessing parameter data for specific wavelengths and instruments, the same processing steps can be applied to any available wavelength/instrument combination.

The preprocessing steps applied here allow us to use a simple segmentation algorithm to produce acceptable CH maps. The algorithm (denoted ezseg) is a dual-threshold-based region-growing algorithm with variable-connectivity constraints. It seeds CH map points based on an intensity threshold value, and then grows the regions iteratively until a growth intensity threshold value is reached or the connectivity falls below the required consecutive neighbor criteria. The choice of the seed and growth thresholds are currently not determined automatically and require manual selection. However, we have found that (due to the new preprocessing) a single set of values work surprisingly well over a long span of time. The CH maps for each instrument are then combined into a single sin(\(\zeta\))–\(\phi\) map where any overlapping regions of the images are merged by selecting the minimum-intensity values to bring out the CHs.

We have shown that the current ezseg algorithm produces very useful CH maps over time. However, the algorithm has some limitations whose improvement could enhance more detailed CH evolution studies. The time-dependent CH maps can suffer from a flickering effect where CHs (usually small ones or ones near instrument borders) will pop in and out of the produced maps. This effect is caused by using singular threshold values over time. Another common problem is “bleed-out” where dark quiet Sun happens to be near a CH, causing the algorithm to sometimes follow the path of the quiet Sun and mistakenly detect it as being attached to the CH. Our use of a connectivity requirement reduces this problem significantly, but more improvement is desired. These issues may be mitigated by the use of a more advanced detection algorithm, such as the seeded region-growing algorithm (SRG) of Mehnert & Jackway (1997) which, given seed locations, segments an image without the need of any tuning parameters. Further along those lines would be the SRG algorithm of Fan et al. (2005) which automatically chooses the seed locations, and auto-tracks the segmentation through time, which would be useful for studying CH evolution. To our knowledge, the SRG algorithm has not yet been utilized in any form in the context of solar images and looks to be a promising area to pursue.

Several data products are publicly provided to the community at a dedicated site.5 These include the following.

1. High-resolution synchronic EUV and CH maps from 2010 June 10 to 2014 August 18 at 6 hr cadence.
2. New preprocessing data for STEREO-A/B EUVI 195 Å and SDO AIA 193 Å for performing limb-brightening correction and inter-instrument intensity normalization for the time period 2010 December 21 to 2014 February 17.
3. A modular region-growing image segmentation algorithm ezseg (which can be linked into any desired driver code) implemented in both an OpenMP-enabled FORTRAN and GPU-accelerated CUDA-C version.
4. A ready-to-use implementation of our CH detection pipeline in the form of a MATLAB driver script (euv2chm) that reads in EUV images, applies our new preprocessing steps, performs CH detection using MATLAB’s MEX interface to call ezseg, and generates the merged EUV and CH maps. An SSW/IDL script to prepare image data for use in euv2chm from fits images is also included (euv2chm_data_prep.pro).
5. The MATLAB scripts used to generate the LBCC and IIT preprocessing data. These are included for advanced users who wish to use the procedure described in this paper with alternative data sets (such as data with alternate PSFs and/or other EUV channels).

Ultimately, we hope that the method and tools presented here can be useful for those wishing to study CH evolution as well as the full-Sun EUV corona and/or limb brightening in general. These methods are particularly relevant for looking at global EUV evolution in time and for applications where a careful treatment of center-to-limb and inter-instrument intensity variations is required.

The authors thank the anonymous reviewer for constructive comments and suggestions which served to greatly improve the manuscript and led to important additions to the methodology. This work was supported by AFOSR, AFRL, NASA’s LWS TR&T program, and NSF’s FESD program.

5 URL: www.predsci.com/chd
Figure 24. Coordinate system for computing limb-brightening correction curves. The center-to-limb angle $\theta$ is shown in a congruent angle position.

APPENDIX

Theoretical LBCCs

A theoretical LBCC can be computed by calculating the observed intensity as a function of $\mu$ by assuming an idealized hydrostatic atmosphere at a given coronal temperature. For example, Andretta et al. (2003) used a constant-gravity approximation for generating such a curve where it was used to estimate photon fluxes in EUV emission. Here, we illustrate the effect of limb brightening using a model that incorporates a $1/r^2$ gravitational fall-off for completeness.

In order to compute the emission intensities, the electron density as a function of height above the base of the corona ($R_0$) is required. Hydrostatic equilibrium for $r \geq R_0$ in the solar corona can be described as

$$\frac{dP}{dr} = -g_0 \left( \frac{R_0}{r} \right)^2,$$

where $g_0 = G M_\odot / R_0^2$, and $P$ is the pressure given by $P = n k_B T$, where $n = \rho/\mu m_p$ is the number density, where $\mu$ is the mean molecular weight (we use the coronal value of $\mu = 0.6$ (Priest 2014)), $k_B$ is the Boltzmann constant, and $m_p$ is the proton mass. Using an isothermal approximation, Equation (11) can be written as

$$\frac{d\rho}{dr} = -\rho \left( \frac{R_0}{r} \right)^2,$$

where we define $\lambda(T)$ as

$$\lambda(T) = \frac{k_B T}{g_0 \mu m_p}.$$

Assuming that all particle species are perfectly coupled, the solution to Equation (12) yields the following form for the electron density:

$$n_e = n_{e0} \exp \left[ -\frac{R_0}{r} \frac{R_0}{\lambda(T)} \right],$$

where $n_{e0}$ is the electron density at the base of the corona ($r = R_0$). The emissivity at any point is found by multiplying the instrument-dependent response function $R(T_e, n_e)$ (which includes the area factor, etc.) by the number density squared. To obtain the observed emission ($F$) for an observer at infinity as a function of $\mu = \cos \theta$, where $\theta$ is the center-to-limb angle, we integrate over a line of sight as shown in Figure 24, which we describe as

$$F(\mu; T_e) = \int_0^\infty R(T_e, n_e) n_e^2(r) dx,$$

where $r$ is a function of $x$ and $\mu$, parametrized by $R_0$, defined as

$$r(x; \mu) = \sqrt{x^2 + 2 x \mu R_0 + R_0^2}.$$

We ignore the weak number density dependence of $R(T_e, n_e)$ for the relevant ranges of density and temperature observed by STA, STB, and AIA, which allows us to choose a typical density $\bar{n}_e$ and pull $R(T_e, \bar{n}_e)$ out of the integral to yield

$$F(\mu; T_e) \approx R(T_e, \bar{n}_e) n_{e0}^2 \int_0^\infty \exp \left[ -\frac{2 R_0}{r} \frac{R_0}{\lambda(T_e)} \right] dx.$$

Unlike the case of constant gravity, the integrand of Equation (17) converges to a constant value as $x \to \infty$, causing the integral to diverge (a classical proof of the existence of the solar wind; e.g., Lang 2009). However, we have found numerically using typical coronal temperatures and a base coronal radius of $R_0 = 1.01 R_\odot$ that the percent difference in the computed integral using an integration limit of $x_{max} = 10 R_\odot$ versus $x_{max} = 10,000 R_\odot$ is $<0.1\%$. Thus, the divergence of the integral of Equation (17) is quite slow and, since most of the plasma that emits in the EUV bands used in this paper is very near the Sun, truncated computations of Equation (17) are a reliable means of obtaining approximate emission values. The LBCC is then defined (in log-space) as

$$L(\mu; T_e) = \log 10 \left[ \frac{F(\mu; T_e)}{F(1; T_e)} \right].$$

It is interesting to note that at the disk center ($\mu = 1$), the equivalent integral of Equation (17) for a constant-gravity model has an analytic solution given by

$$F_{\text{const}}(1, T_e) = \frac{n_{e0}^2 k_B}{2 g_0 \mu m_p} R(T_e) T_e,$$

which shows that under such an approximation, for a given density at the surface, the intensity is directly proportional to the temperature multiplied by the response function at that temperature. This implies a correspondence between image intensity and temperature within an EUV image. While such a direct correspondence is somewhat suspect, this could be justified (at least on average) by a heuristic, physics-based argument. Drawing from a long history of idealized coronal heating scaling laws (e.g., Aschwanden & Schrijver 2002 and references therein), if one assumes that there is on average some general power-law correspondence between number density and temperature (no matter how weak), then the strong dependence of emissivity on density implies that the intensity of a given structure will be positively correlated with temperature.

REFERENCES

Altschuler, M. D., Trotter, D. E., & Orrall, F. Q. 1972, SolPh, 26, 354
Andretta, V., Del Zanna, G., & Jordan, S. D. 2003, A&A, 400, 737
Antiochos, S. K., Mikic, Z., Titov, V. S., Lionello, R., & Linker, J. A. 2011, ApJ, 731, 112
Aschwanden, M. J. 2010, SolPh, 262, 235
Aschwanden, M. J., & Schrijver, C. J. 2002, ApJ, 142, 269
Bohlin, J. D. 1977, Coronal Holes and High Speed Wind Streams, ed. J. B. Zirker (Washington, DC: American Geophysical Union), 257
Bonettiini, S., Zanella, R., & Zanni, L. 2009, InvPh, 25, 015002
Cheung, M. 2015; in Presented at the NVIDIA GPU Technology Conf. (San Jose, CA)
