On the autonomous detection of coronal mass ejections in heliospheric imager data

S. J. Tappin,[1] T. A. Howard,[2] M. M. Hampson,[1,3] R. N. Thompson,[2,4] and C. E. Burns[2,5]

[1] We report on the development of an Automatic Coronal Mass Ejection (CME) Detection tool (AICMED) for the Solar Mass Ejection Imager (SMEI). CMEs observed with heliospheric imagers are much more difficult to detect than those observed by coronagraphs as they have a lower contrast compared with the background light, have a larger range of intensity variation and are easily confused with other transient activity. CMEs appear in SMEI images as very faint often-fragmented arcs amongst a much brighter and often variable background. AICMED operates along the same lines as Computer Aided CME Tracking (CACTus), using the Hough Transform on elongation-time J-maps to extract straight lines from the data set. We compare AICMED results with manually measured CMEs on almost three years of data from early in SMEI operations. AICMED identified 83 verifiable events. Of these 46 could be matched with manually identified events, the majority of the non-detections can be explained. The remaining 37 AICMED events were newly discovered CMEs. The proportion of false identification was high, at 71% of the autonomously detected events. We find that AICMED is very effective as a region of interest highlighter, and is a promising first step in autonomous heliospheric imager CME detection, but the SMEI data are too noisy for the tool to be completely automated.

1. Introduction

[2] Heliospheric imaging is a relatively new technique for imaging space weather phenomena using visible white light. While the ability to observe transient phenomena such as coronal mass ejections (CMEs) and corotating interaction regions (CIRs) in white light at large distances from the Sun was demonstrated by the Helios zodiacal light photometers [*Richter et al.*, 1982], the first wide field imager designed to detect and track solar wind disturbances did not appear until 2003 with the launch of the Coriolis spacecraft carrying the Solar Mass Ejection Imager (SMEI) [*Eyles et al.*, 2003]. SMEI observed many transient phenomena during its 8-1/2 year lifetime including CIRs [*Tappin and Howard*, 2009a], comets [*Kuchar et al.*, 2008; *Buffington et al.*, 2008] and CMEs [e.g.,

Tappin et al., 2004; *Howard et al.*, 2006; *Webb et al.*, 2006]. Around 400 coronal mass ejections (CMEs) have been confirmed by SMEI during its lifetime (reports including the lists of SMEI events can be found in *Webb et al.* [2006] and *Howard and Simnett* [2008]).

[3] Heliospheric imagers make use of the same physics as coronagraphs: they observe sunlight that has been Thomson scattered off free electrons in the corona and solar wind. Plasma density enhancements such as CMEs are observed as arcs of enhanced brightness moving away from the Sun. The intensity enhancement due to a CME is very much less than the background light which is dominated by the zodiacal light. The contrast also becomes disproportionally fainter as they move across the sky, and they also share the sky with much brighter astronomical objects such as stars and planets. Numerical techniques must therefore be employed to extract the CME from the background.

[4] The ability to detect and track CMEs across the sky has proven to have many useful applications, not least in space weather prediction. It is well known that CMEs are responsible for the most significant geomagnetic storms [e.g., *Wilson*, 1990; *Marubashi*, 2000; *Echer et al.*, 2005] with their many consequences, including spacecraft and ground power station damage, atmospheric drag increase on satellites, and radiation dosage increase for airline passengers and astronauts [e.g., *Baker et al.*, 2009]. A number of techniques have been developed that use heliospheric imagers to predict the arrival time and speed of CMEs at the Earth, with high

[1]National Solar Observatory, Sunspot, New Mexico, USA.
[2]Department of Space Studies, Southwest Research Institute, Boulder, Colorado, USA.
[3]Now at Laboratory for Atmospheric and Space Physics, University of Colorado at Boulder, Boulder, Colorado, USA.
[4]Now at Theoretical and Computational Epistemology Group, Department of Plant Sciences, University of Cambridge, Cambridge, UK.
[5]Now at Space Research, University of Michigan, Ann Arbor, Michigan, USA.

Corresponding author: S. J. Tappin, National Solar Observatory, PO Box 62, Sunspot, NM 88349, USA. (jtappin@nso.edu)
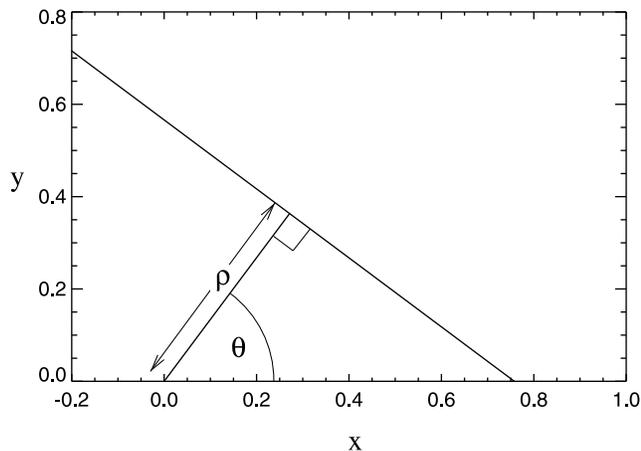
Figure 1. The parametrization of a line in terms of its perpendicular distance to the origin and the direction of that perpendicular.

degrees of accuracy [e.g., *Davis et al.*, 2009; *Howard and Tappin*, 2010].

[5] This paper reports on recent progress toward an autonomous detection tool for CMEs using heliospheric imagers. We have developed a tool that attempts to automatically detect CMEs in the SMEI data set, using concepts developed for the Computer Aided CME Tracking (CACTus) [*Robbrecht and Berghmans*, 2004; *Robbrecht et al.*, 2009] technique. We call this tool "Automatic Interplanetary CME Detection" (AICMED). The operation of this tool is described, and its results on almost three years of data from the early part of the SMEI mission are compared with manual measurements of CMEs for the same interval. We conclude with an evaluation of the tool's performance and on future implications for other heliospheric imagers.

2. Autonomous Detection of CMEs

[6] The last decade has witnessed efforts toward the automated detection of CMEs. While CMEs have been observed manually since they were first recognized in the 1970s [e.g., *Howard et al.*, 1975; *Munro and Sime*, 1985; *Hundhausen et al.*, 1984; *St. Cyr et al.*, 2000; *Yashiro et al.*, 2004] it was only with the vastly improved data quality of the SOHO coronagraph imagers that automatic detection of CMEs became feasible.

2.1. Automated CME Detection Using Coronagraphs

[7] The most widely used autonomous tool is CACTus [*Robbrecht and Berghmans*, 2004; *Robbrecht et al.*, 2009]. The CACTus CME list is maintained at http://sidc.be/cactus/ and is updated regularly. CACTus is applied to coronagraph data (originally SOHO/LASCO and now STEREO/COR as well), and works on the principle that CMEs travel with an approximately constant speed. This means that when their trajectory is plotted in a distance-time graph they display a roughly linear trend. Distance-time data are not immediately available from the coronagraph images, but can be produced via the following procedure:

[8] 1. Coordinate transform the coronagraph image from Cartesian coordinates (x, y) into polar coordinates (r, ϕ, where

r is distance from the Sun center and ϕ is position angle) and map it onto a new grid with resolution (δr, $\delta \phi$).

[9] 2. Stack a collection of images into a 3-D data cube (t, r, ϕ, where t is image time).

[10] 3. Choose a ϕ slice through the cube to generate an image in (t, r).

[11] 4. Plot the resulting 2-D image in (t, r) coordinates.

[12] The result is a plot of distance against time for each object that crosses the selected ϕ plane. An example of such a plot for LASCO images can be found in *Robbrecht and Berghmans* [2004, Figure 1].

2.2. The Hough Transform

[13] Once the (t, r) image is produced, straight lines can be extracted from it using the Hough Transform. The Hough transform [*Duda and Hart*, 1972] is a means by which straight lines are extracted from an image, by transforming Cartesian coordinates of points in real space into parameters describing the lines between the points. *Robbrecht and Berghmans* [2004] describe the parametrization as two variables (t_0, Δt), where t_0 is the coordinate of the intersection point with the t-axis and Δt is the distance along the t-axis corresponding to a distance $r_{max} - r_{min}$ in the r-direction. While this summarizes the effect of the Hough transform process, we think it is useful to describe it in rather more detail in the remainder of this section.

[14] Any line in a plane may be parameterized in terms of its perpendicular distance (ρ) from the origin and the direction (θ) of the perpendicular (Figure 1). In this formulation the equation of the line is

$$\rho = x\cos\theta + y\sin\theta. \tag{1}$$

This representation has computational advantages over the use of gradient and intercept as, for a finite region, both ρ and θ are bounded [*Duda and Hart*, 1972]. By convention, θ is allowed to vary from 0 to π, and ρ may be positive or negative (the sign is the same as that of the y-intercept). The parametrization of the infinite set of lines passing through the point at (x, y) also satisfies the relationship in equation (1), but now θ and ρ are variable and x and y are constant, this results in a sine curve in [θ, ρ] space. Figure 2 illustrates the coordinate systems involved. Thus to detect aligned points, it is simply a matter of summing these sine curves and looking for maxima. This leads to the formal definition of the continuous Hough transform of a function $A(x, y)$ as

$$H(\theta, \rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(x,y)\delta(\rho - x\cos\theta - y\sin\theta)\mathrm{d}x\mathrm{d}y, \tag{2}$$

where δ is the Dirac delta. In the analysis of digitized images, the values of x and y lie on a grid of discreet pixels, as will the θ and ρ parameters. For this reason, the continuous transform is converted to its discrete form:

$$H(\theta_i, \rho_j) = \sum_m \sum_n A(x_m, y_n)\delta_{j,j'}, \tag{3}$$
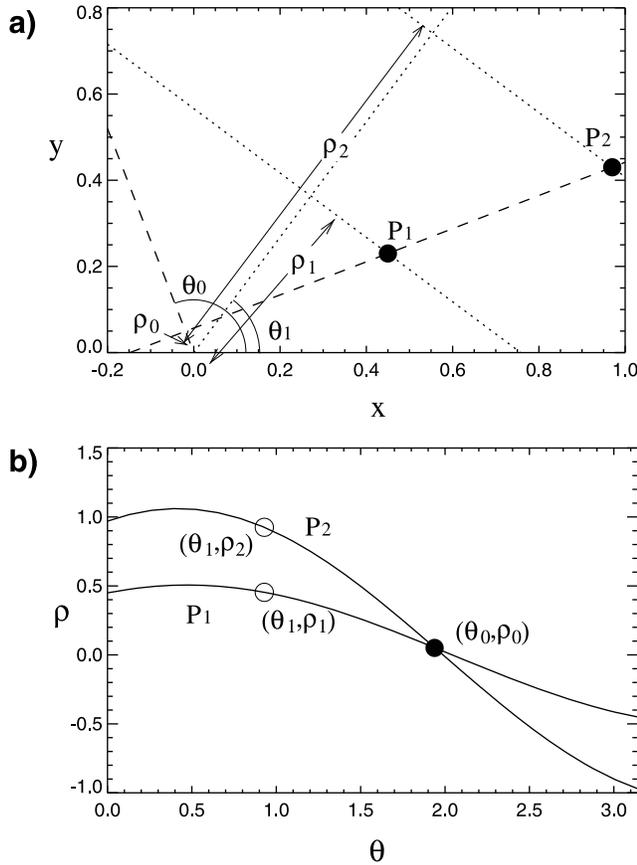
a)

b)

Figure 2. Illustration of the coordinates and concepts of the Hough transform. (a) The real-space configuration. Two points (P_1 and P_2) are defined, for each of these there are a range of values of ρ as a function of θ. Two values of θ are shown: θ_1 (dotted lines), where the two points produce different values of ρ (ρ_1 and ρ_2); and θ_0 (dashed lines) which is the perpendicular to the line joining the points, so that both points map to the same value of ρ (ρ_0). (b) The accumulator space. The sinusoids showing ρ as a function of θ for the two points from Figure 2a are plotted, the intersection of these is at (θ_0, ρ_0).

where δ is now the Kronecker delta and j' is the nearest index to the value of ρ determined from θ_i, x_m and y_n, and is given by

$$j' \equiv \text{round}\left(\frac{x_m \cos \theta_i + y_n \sin \theta_i - \rho_{\min}}{\Delta \rho}\right),$$

where $\Delta \rho$ is the output grid spacing in ρ and $\rho_{\min}$ is the lowest value in the output grid, and "round" denotes rounding to the nearest integer. In the discrete form, it is customary to refer to the [θ, ρ] space as the accumulator space, since it is implemented as a two-dimensional array into which the sinusoids are accumulated. The value at each point (θ_i, ρ_j) in the accumulator becomes the number of points in the original image lying on that line, weighted by A.

[15] The inverse transform or back projection, is obtained by summing all the accumulator space points that correspond to lines through a given point in real space; thus,

$$B_{mn} = \begin{cases} \sum_i \sum_j H\left(\theta_i, \rho_j\right) \delta_{n,n'} & \text{where } |\sin \theta_i| > 1/\sqrt{2} \\ \sum_i \sum_j H\left(\theta_i, \rho_j\right) \delta_{m,m'} & \text{where } |\sin \theta_i| \leq 1/\sqrt{2} \end{cases}, \quad (4)$$

where the pseudo-indices $n' \equiv \text{round}(am + b)$ and $m' \equiv \text{round}(a'n + b')$ are defined in terms of the gradients and intercepts:

$$a = -\frac{\Delta x \cos \theta_i}{\Delta y \sin \theta_i}, \quad b = \frac{\rho_j - x_{\min} \cos \theta_i - y_{\min} \sin \theta_i}{\Delta y \sin \theta_i},$$

$$a' = \frac{1}{a}, \quad b' = \frac{\rho_j - x_{\min} \cos \theta_i - y_{\min} \sin \theta_i}{\Delta x \cos \theta_i}.$$

Hence, straight lines can be identified by applying a threshold across accumulator space. For example, if one wished to extract only those lines that comprised of more than 100 points in Cartesian space one would apply a threshold in accumulator space with a minimum "z" value of 100 times the "typical" value of $A(x, y)$. This would reveal a number of points in accumulator space that, when applied to an inverse Hough Transform, would appear as lines in Cartesian space. In the case of a line of finite width or a not quite straight line, all lines of sufficient length that can be inscribed within the region are detected. This process is illustrated in Figure 3. It is seen that the loop whose longest contained straight distance is less than the threshold length is eliminated, while the line (although it is not exactly straight) is retained. The reconstruction from the complete accumulator space (Figure 3d) is considerably blurred relative to the original image. This is a consequence of the discretization of the accumulator space in equation (3). As well as the elimination of the loop while keeping the line, it is worth noting that the intersection of the band from the loop in accumulator space with the edges of the line's transform produces a small region of values above the chosen threshold near $\theta = 1.55$, $\rho = 40$ which produces a spurious line in the reconstruction—the elimination of such artifacts requires either tuning of the accumulator threshold or the application of a threshold to the reconstruction.

[16] It should also be noted here that the Hough transform is able to detect broken lines. For example, a 50% dashed line would appear in the accumulator space to be the same as a line of the same direction and distance from the origin but half the length.

[17] We use the Hough transform routine from IDL (Interactive Data Language) in the analysis presented in this paper.

2.3. Automated CME Detection Using Heliospheric Imagers

[18] Attempts to apply this Hough Transform technique to heliospheric imagers are hindered by the fact that measurements of "distance" in heliospheric imagers cannot be easily converted to true units of distance. This is because the assumptions of sky-plane localization and small angles that can be applied to coronagraph images break down at large angles from the Sun. Hence with heliospheric imagers we

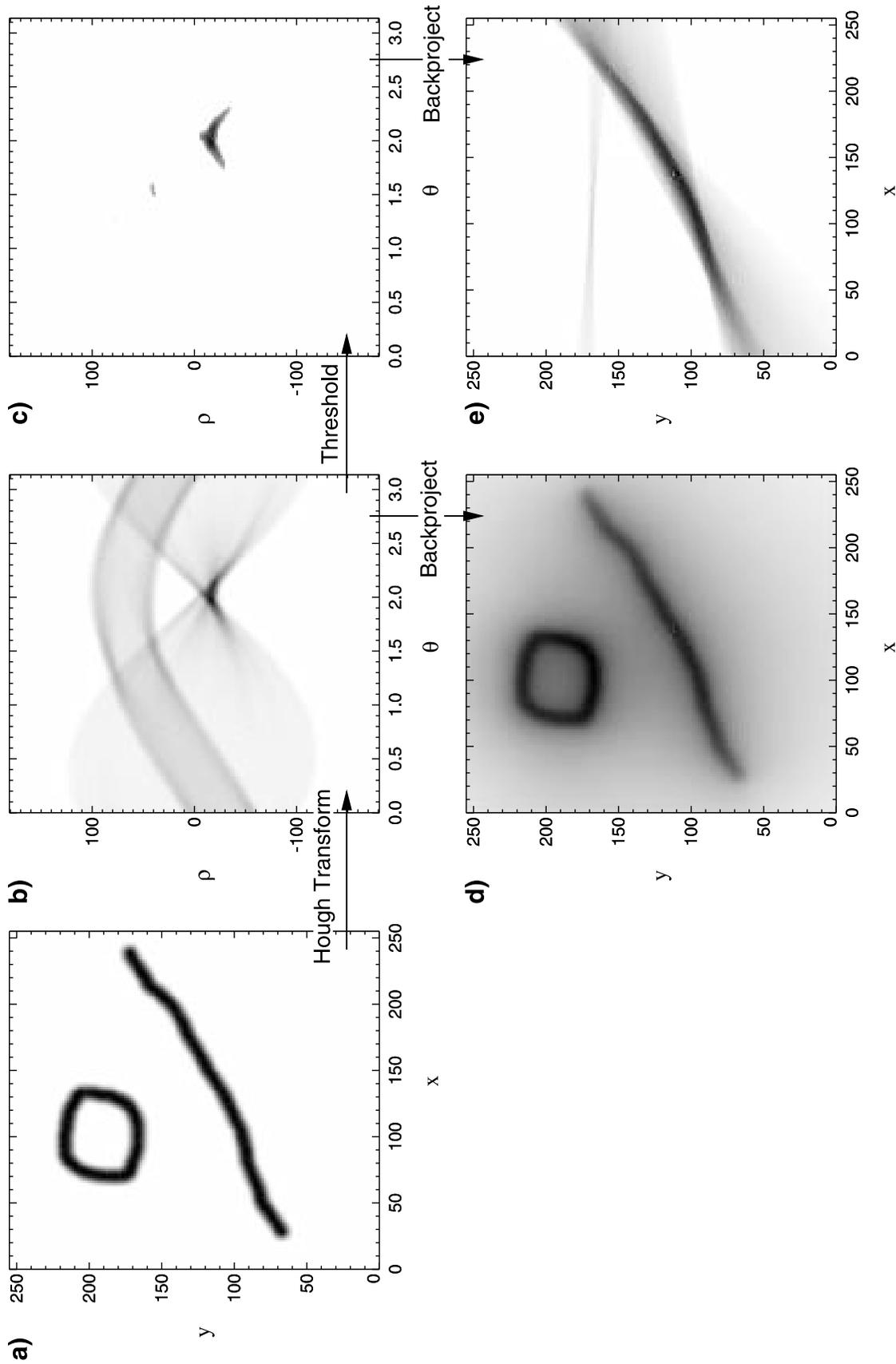**Figure 3.** A simple example of the use of the Hough transform to locate lines in an image. (a) The original image with a loop and an approximate line (all the panels in this figure are shown with high values dark), (b) the raw Hough transform of the image, (c) the Hough-transformed data after thresholding, (d) the back projection of the raw Hough transform, (e) the back projection of the thresholded Hough transform.
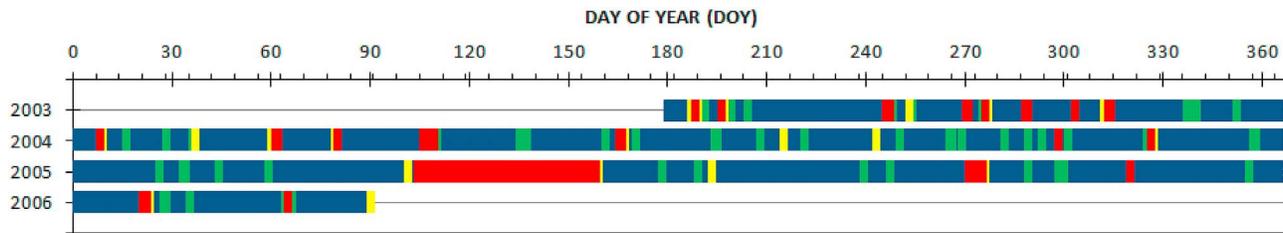
Figure 4. Representation of the SMEI duty cycle during the 33 months from 29 June 2003 until 31 March 2006. SMEI had an orbital period of ~100 min and produced one image per orbit and so when operating nominally produced 14–15 images per day. These numbers regard as valid any image for which at least one SMEI camera provided data. The color code is as follows: blue: >10 images/day; green: 5–10 images/day; yellow: 1–4 images/day; red: 0 images/day. The number of images shown here totals 12604.

must revert to the original units of elongation. Elongation (ε) is the angle between the observer-Sun vector and the vector from the observer through the measured point in the sky. Coordinate rotation and ϕ selection can still be applied to a stack of heliospheric images, except that the 2-D map is in (t, ε), rather than (t, r) coordinates. Such 2-D maps are now commonplace for heliospheric imager data and are generally referred to as "J-maps" [e.g., *Davies et al.*, 2009]. Many workers currently make extensive use of measurements from J-maps for heliospheric image CME identification [e.g., *Davis et al.*, 2009; *Liu et al.*, 2010].

[19] This means that CME tracks in (t, ε) coordinates do not form a straight line, even when traveling at a constant speed. This is discussed for the extreme case of a single point moving through space by *Howard and Simnett* [2008] and t-ε plots for points traveling with the same speed, but at different trajectories, are shown in their Figure 1. In other words, because the approximation $\sin\varepsilon \approx \varepsilon$ breaks down at the elongations at which heliospheric imagers observe (the approximation is accurate to better than 0.25% even at the outer edge of LASCO C3, whereas at $30°$ elongation the error is almost 5%), CME trajectories in heliospheric imager J-maps are not linear, and their linearity and gradients are influenced not only by their speed, but also the direction of their trajectory.

[20] While we therefore cannot reliably identify the speed of CMEs using the Hough Transform technique, we can still identify them in the heliospheric imager data sets if we apply one further assumption: That CME tracks in (t, ε) space are continuous. By breaking the CME tracks into intervals divided by overlapping sub-ranges of elongation (hereafter referred to as "parsing") such that the track of the CME through any sub-range is approximately linear, we can assign a straight line to each sub-range via the Hough Transform. The extracted lines can then be connected along the entire elongation range by making the assumption of continuity.

3. The SMEI Data Set

[21] We have applied the automatic detection tool to heliospheric images taken by the SMEI instrument. SMEI was launched in January 2003, and began making science-mode observations in February 2003. It was deactivated on 28 September 2011 after providing almost continuous data for approximately 8-1/2 years. Figure 4 provides a representation of the SMEI duty cycle during the time interval

selected for our study. Aside from a large data gap (red) from April-June 2005 (when Coriolis was in Sun-point mode and thus not properly oriented for SMEI to collect useful data) data are available for the vast majority of the time.

[22] Figure 5 shows a typical SMEI image generated by the NSO pipeline after the 3-day background model has been subtracted. A number of features are indicated, mostly saturating artifacts from energetic particles, hot pixels and the moon. The problems of saturation caused by these artifacts are discussed by a number of workers, including *Tappin et al.* [2004] and *Webb et al.* [2006]. A summary of the processing required to generate such maps from raw SMEI data is given in Appendix A.

4. Autonomous Interplanetary CME Detection (AICMED)

[23] AICMED applies the Hough transform process to SMEI data. It first prepares the data by rebinning the rectangular projection images to a resolution of $3°$ in position angle (ϕ) and $1°$ in elongation (i.e. the original resolution of 720×360 pixels is reduced to 120×180 bins). The region beyond $150°$ elongation is then removed, reducing the image size to 120×150 bins as CMEs are rarely, if ever, seen beyond that elongation, so little useful information can be obtained in that region. Two other things must be done to the data at this stage. First; there are occasional missing orbits in the SMEI data stream, these are simply filled in with blank images to maintain a constant cadence. Secondly; since the SMEI data are noisy, it is necessary to identify and exclude regions of noisy data as far as possible before the processing. Noise is identified in the J-map as areas where the brightness (positive or negative) exceeds twice the threshold described in the next paragraph and regions of high pixel-to-pixel variation as well as regions flagged by the original processing pipe as being suspect. These regions are set to zero.

[24] The images are then scaled and truncated such that values within the threshold range (normally ± 1.0 ADU) are rescaled into $0 \rightarrow 1$. The generation of J-maps from these images is simply a matter of stacking the images into a data cube in (t, ε, ϕ) coordinates, and then extracting slices at constant ϕ.

[25] The J-map for each ϕ slice is divided into five over-lapping elongation ranges. The program then performs the Hough transform on each of these ranges, conducts a number of Hough space filtering and masking processes, then
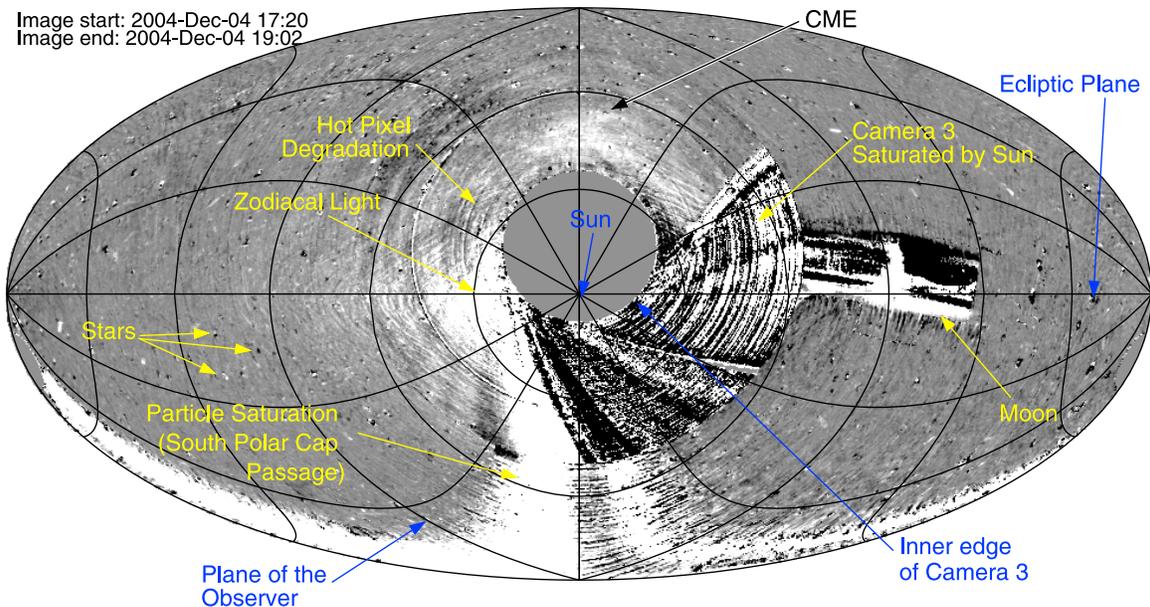
Figure 5. An all-sky SMEI image obtained during the orbit from 17:20–19:02UT on 4 December 2004. This is shown as a Hammer-Aitoff projection and has been processed through the NSO pipeline. A number of "noise" features have been indicated with yellow labels, including: energetic particle saturation, hot pixel degradation, stars, the moon, and a region when Camera 3 was saturated due to its pointing too close to the Sun. Features of the observation geometry are denoted by blue labels. A CME is visible too and is also labeled, but it is significantly fainter than the other features.

returns the data to Cartesian space with additional "post-Hough" masking and edge filtering. After this has been applied across each of the elongation ranges, it searches for continuity of these features and identifies those as CMEs.

[26] Figure 6 shows the steps of the process for a single J-map. Figure 6a shows a J-map for a selected sequence of running-difference SMEI images during May 2003, which was an early SMEI CME reported by *Tappin et al.* [2004] (this event occurred before the start of the interval which we shall consider in the remainder of this paper, but because it was a bright CME and has been extensively studied it is a good CME to use for the initial testing of the algorithms). The CME is the relatively bright feature around the middle of the panel. The first step is to parse the data into five overlapping elongation bands: 0–45°, 23–67°, 53–97°, 84–127°, and 105–150°. We then perform the Hough transform on each of these, revealing the accumulator space map in Figure 6b. Here a number of sinusoidal curves can be identified, but the dominant feature is the bright region corresponding to (θ, ϕ) coordinates along bright straight lines in the original J-map. We apply a threshold cutoff to this map at $z = 40$ revealing the accumulator map shown in Figure 6c (it should be noted that because CMEs are extended objects, the values in the accumulator space are a superposition of several lines and so give higher levels than the actual length of the line). Next, the inverse Hough transform is applied returning the data to the original J-map coordinates (Figure 6d).

[27] This J-map revealed in Figure 6d shows that the process has performed well in extracting the CME feature from the rest of the data. An edge detection routine is applied (we use Laplacian edge filtering) and the elongation strips are recombined, resulting in the overlay shown in Figure 6e. Finally, a threshold is applied to define whether a CME is present at the given elongation and time. This final result (Figure 6f) is a mask set to 1 (white) for those J-map pixels that are part of a CME, and to 0 (black) for those which are not. When this is applied to all of the position angle slices, a 3-dimensional array of points is produced with 1's where there is CME-like structure and 0's elsewhere. It should be noted that the curvature of the CME track is a result of the non-linearity of the geometry of heliospheric imager observations rather than a real acceleration.

[28] After the CME-like features have been identified, a number of further selections are applied to reduce the number of false positives. To be considered as a valid CME a feature must: extend through a minimum of 3 orbits, and have been identified in a minimum of 3 adjacent position angle bins, it must also include at least 40 bins in the J-map data cube. Features moving more slowly than 3.3°/day or faster than 66.2°/day (which correspond approximately to speeds of 100 km/s and 2000 km/s) are rejected as these are outside the range of speeds of the vast majority of CMEs in the solar wind. The CME pixels are then grouped into contiguous regions in the J-map data cube and labeled with an arbitrary serial number.

[29] In addition there is an optional filter that allows AICMED to exclude apparent CMEs that lie predominantly very close to noise features as these are likely to be detections of the quasi-systematic drift of those noise features.

[30] After the CMEs have been identified, further routines identify the leading edge of the structure for input to the Tappin-Howard model [*Tappin and Howard*, 2009b].
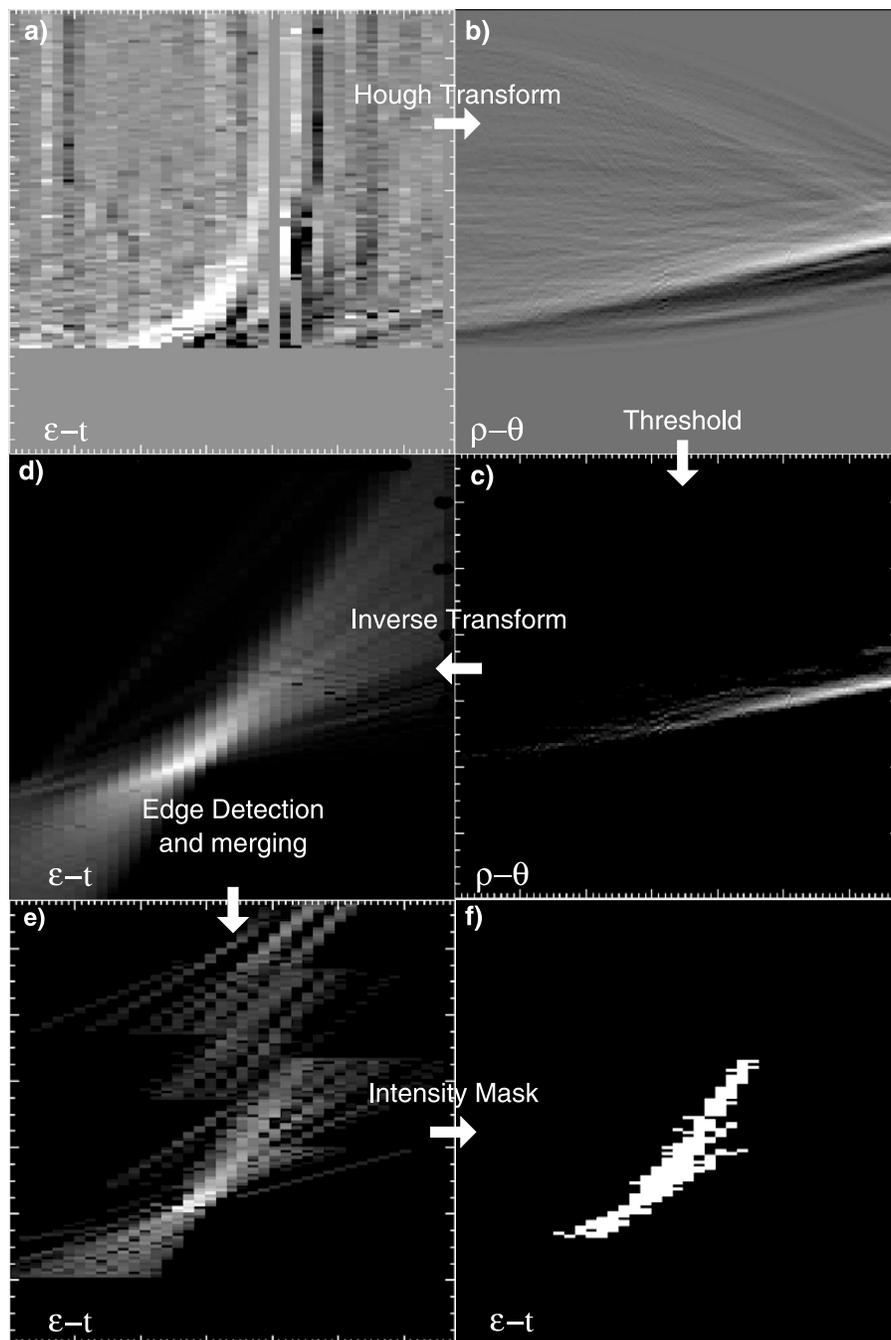
Figure 6. The stages of the AICMED process. (a) A SMEI J-map from 29–31 May 2003 showing a CME. (b) The same map as it appears transformed in accumulator space (for one of the elongation strips). (c) Accumulator space after the filtering is applied. (d) The new J-map after an inverse-Hough transform has been applied. (e) The map following an edge detection routine, and merging the elongation strips (N.B. This panel uses a square-root scaling to show the fainter features, all others are linear). (f) The final map (yes = 1, no = 0) following an intensity mask. The result is an elongation-time plot of all of the points from the original J-map that satisfy the AICMED requirements to be classified as part of a CME.

[31] Figure 7 shows a frame from the movie of one of the subintervals from the study presented in this paper (Movie S1 of the auxiliary material).[1] This is a "rectangular" projection map in which the axes are position angle and elongation, such that the Sun is the bottom edge of the image, and the antisolar direction is the top edge; North is at the center of the x-axis and South is at the edges. Thus a CME will appear to move approximately vertically up the map as it travels out from the Sun. The individual CME features identified by AICMED are
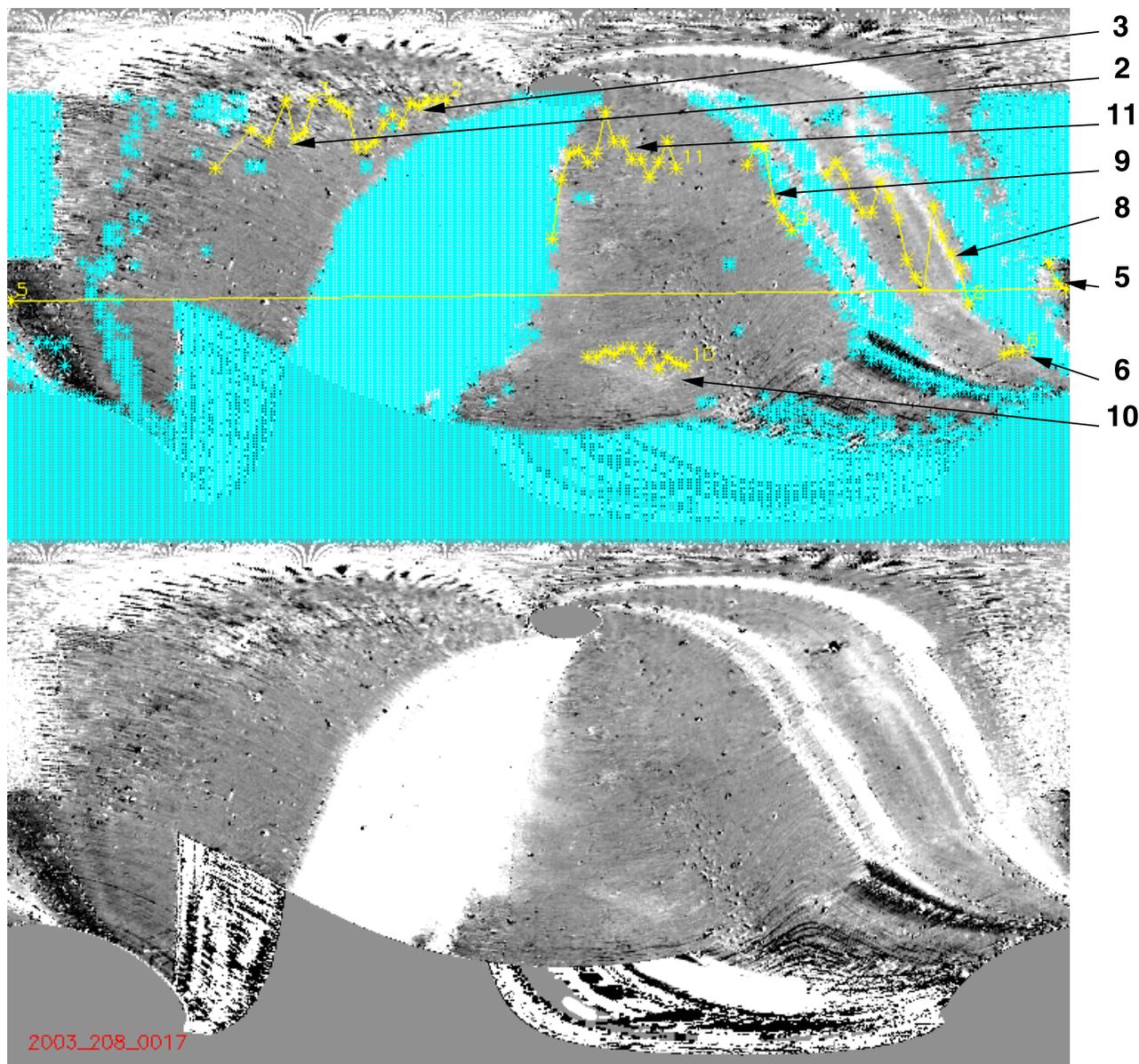
Figure 7. One selected frame from the SMEI movie for the 2003 days 204–209 interval. (top) Annotations showing noise regions and CME candidates identified by AICMED, (bottom) the same data without the annotations. The leading edges of features identified as a CME are in yellow and those identified as the noise are in cyan. The numbers on the CME features indicate the arbitrary serial number assigned by AICMED, which is then used in post-analysis (as the number on the original image are hard to see on the printed page, we have added indications of the location of each feature at the side of the figure). This image includes the CME candidates rejected by the automatic noise-proximity filter. The features present in this frame were identified as follows: 2–real, 3–real same thing as 2, 5 and 6–automatically rejected, 8–partly real (but not the part in this frame), 9–automatically rejected, 10–automatically rejected, but actually real, 11–partly real.

indicated and numbered on the movie so that the observer can identify and classify them.

5. Results

[32] To characterize and demonstrate the performance of AICMED we have selected 33 months of SMEI data starting from 29 June 2003 and ending on 31 March 2006. The starting date was selected because prior to that date there were frequent changes in the details of the observing mode and also a large number of calibration and characterization observations. The end date is that of the failure of the A-side Data Handling Unit (DHU) during a South Atlantic Anomaly passage. SMEI was switched to the redundant B-side DHU and resumed operating on 5 May 2006, and only minor anomalies followed until its deactivation more than

Table 1. Summary of the Identified CMEs From the AICMED Runs[a]

I	Run	N	Start Time	End Time	Dur	C-PA	Wid	Emin	Emax	ID	Cls	NB[b]
1	2003_180	1	2003/181 11:34	2003/182 4:30	16.9	307.5	18	53.5	95.5	No	G	1
2	2003_188	8	2003/190 19:6	2003/192 10:3	38.9	340.5	57	28.5	126.5	2003/191		
3	2003_192	1	2003/193 2:59	2003/193 23:18	20.3	316.5	12	53.5	87.5	2003/193?		
4	2003_200	1	2003/202 2:2	2003/204 9:55	55.9	55.5	15	51.5	77.5	No	P	1
5	2003_204	1	2003/205 4:33	2003/207 19:12	62.7	13.5	18	35.5	115.5	2003/205a		
6	2003_204	2	2003/205 4:33	2003/208 22:18	89.8	49.5	39	71.5	149.5			2
	2003_204	11	2003/205 4:33	2003/208 23:59	91.4	334.5	45	53.5	149.5			
	2003_204	3	2003/205 7:56	2003/208 18:55	83.0	94.5	39	83.5	149.5	2004/206b,		
	2003_204		2003/205 4:33	2003/208 23:59	91.4	33	162	53.5	149.5	2004/207bc		
7	2003_204	8	2003/205 4:33	2003/208 23:59	91.4	238.5	51	53.5	138.5	2003/205b		
8	2003_204	10	2003/205 13:1	2003/208 23:59	83.0	325.5	33	28.5	92.5	2003/206a		1
9	2003_208	1	2003/207 22:35	2003/210 23:24	72.8	325.5	27	53.5	126.5	2003/207a?		
10	2003_208	2	2003/208 20:36	2003/212 21:8	96.5	349.5	30	35.5	96.5	No	G	1
11	2003_220	3	2003/220 17:6	2003/224 5:46	84.7	304.5	9	46.5	94.5	2003/219?		3
12	2003_244	6	2003/245 3:1	2003/245 18:15	15.2	298.5	24	41.5	84.5	2003/245		
13	2003_256	4	2003/258 4:17	2003/260 22:20	66.0	334.5	42	36.5	149.5	No	G	
14	2003_284	1	2003/284 18:42	2003/287 21:12	74.5	109.5	15	83.5	148.5	2003/285		
15	2003_296	1	2003/296 10:6	2003/299 14:18	76.2	103.5	72	40.5	149.5			
	2003_296	2	2003/297 1:20	2003/300 20:47	91.4	325.5	57	63.5	149.5	2003/296,		
			2003/296 10:6	2003/300 20:47	106.7	38	203	40.5	149.5	2003/297a-c		2
16	2003_320	1	2003/322 0:48	2003/324 22:13	69.4	130.5	42	32.5	126.5			
	2003_324	1	2003/323 22:31	2003/324 18:50	20.3	142.5	24	83.5	129.5			1
			2003/322 0:48	2003/324 22:13	69.4	132	45	32.5	129.5	2003/323		
17	2003_332	2	2003/333 4:21	2003/335 2:4	45.7	238.5	60	51.5	127.5	No	P	1
18	2003_336	3	2003/337 23:30	2003/339 7:40	32.2	121.5	9	76.5	126.5	No	P	
19	2003_340	2	2003/341 0:18	2003/342 20:20	44.0	145.5	12	53.5	96.5	No	G	
20	2003_352	1	2003/353 18:48	2003/354 15:7	20.3	103.5	54	83.5	149.5		V	
	2003_352	5	2003/353 18:48	2003/354 15:7	20.3	139.5	15	83.5	126.5			
			2003/353 18:48	2003/354 15:7	20.3	112	70	83.5	149.5	No		
21	2004_011	1	2004/10 22:31	2004/11 20:32	22.0	127.5	9	53.5	95.5	2004/011		
22	2004_019	1	2004/21 10:49	2004/23 15:19	52.5	148.5	72	25.5	113.5	2004/021		
23	2004_023	1	2004/24 3:10	2004/27 20:55	89.7	151.5	9	53.5	86.5	2004/025		
24	2004_039	2	2004/38 22:46	2004/41 13:25	62.7	121.5	15	31.5	95.5	No	P	
25	2004_039	4	2004/41 18:30	2004/43 23:0	52.5	13.5	51	37.5	149.5			
	2004_043	1	2004/42 23:17	2004/44 22:42	47.4	19.5	42	83.5	149.5			
			2004/41 18:30	2004/44 22:42	76.2	14	52.5	37.5	149.5	2004/040		
26	2004_067	1	2004/67 5:48	2004/71 23:15	113.5	10.5	30	23.5	96.5	2004/068		
27	2004_083	1	2004/85 2:13	2004/87 21:56	67.7	145.5	15	65.5	122.5	No	G	
28	2004_091	2	2004/92 5:14	2004/95 22:59	89.7	13.5	54	33.5	96.5	2004/092?		
29	2004_095	2	2004/95 0:58	2004/97 2:4	49.1	355.5	12	26.5	66.5	2004/095a		
30	2004_103	1	2004/104 5:6	2004/105 14:58	33.9	7.5	9	35.5	73.5	2004/105		1
31	2004_119	1	2004/122 6:35	2004/123 23:14	40.6	10.5	12	53.5	93.5			
	2004_119	2	2004/122 6:35	2004/123 23:14	40.6	31.5	18	53.5	96.5			
	2004_123	2	2004/123 1:13	2004/126 20:39	91.4	43.5	12	56.5	149.5			
			2004/122 6:35	2004/126 20:39	110.1	27	45	53.5	149.5	2003/122		
32	2004_123	5	2004/123 1:13	2004/127 23:45	118.5	133.5	33	67.5	131.5			
	2004_123	4	2004/125 5:43	2004/126 20:39	38.9	106.5	9	83.5	126.5			
			2004/123 1:13	2004/127 23:45	118.5	130.5	48	67.5	131.5	2004/125		
33	2004_131	6	2004/130 22:52	2004/132 22:17	47.4	25.5	57	36.5	99.5			
	2004_131	5	2004/133 5:3	2004/135 11:14	54.2	16.5	42	54.5	149.5			
			2004/130 22:52	2004/135 11:14	108.4	25	91.5	36.5	149.5	2004/131		
34	2004_135	4	2004/138 1:54	2004/139 21:55	44.0	10.5	33	53.5	146.5	No	P	
35	2004_139	4	2004/140 23:19	2004/141 16:15	16.9	178.5	9	95.5	137.5	No	V	
36	2004_147	1	2004/148 5:44	2004/151 23:28	89.7	25.5	45	33.5	145.5			1
	2004_147	4	2004/148 14:12	2004/151 23:28	81.3	355.5	30	32.5	80.5	2004/148,		
			2004/148 5:44	2004/151 23:28	89.7	14	67.5	33.5	145.5	2004/149		2
37	2004_163	4	2004/162 23:38	2004/164 19:39	44.0	16.5	114	30.5	96.5	No	P	
38	2004_179	2	2004/178 22:19	2004/182 22:50	96.5	43.5	33	83.5	149.5		G	
	2004_179	4	2004/178 22:19	2004/182 22:50	96.5	346.5	42	32.5	138.5			
			2004/178 22:19	2004/182 22:50	96.5	15	94.5	32.5	149.5	No		
39	2004_187	2	2004/187 12:54	2004/189 22:28	57.6	55.5	9	53.5	74.5	No	P	
40	2004_191	2	2004/191 8:20	2004/193 19:36	59.3	88.5	48	83.5	148.5		V	
	2004_191	3	2004/191 11:44	2004/193 0:59	37.3	139.5	9	35.5	96.5			
			2004/191 8:20	2004/193 19:36	59.3	104	79.5	35.5	148.5	No		
41	2004_191	4	2004/191 23:35	2004/193 19:36	44.0	319.5	39	27.5	81.5	2004/192		1
42	2004_195	5	2004/197 8:16	2004/198 21:31	37.3	313.5	12	32.5	66.5			1
	2004_195	6	2004/197 11:39	2004/198 21:31	33.9	334.5	21	30.5	65.5			1
			2004/197 8:16	2004/198 21:31	37.3	326	37.5	30.5	65.5	2004/197		
43	2004_199	3	2004/199 4:18	2004/203 6:31	98.2	325.5	12	53.5	109.5	No	P	
44	2004_203	2	2004/203 11:35	2004/205 19:28	55.9	346.5	24	34.5	125.5	2004/203		1
45	2004_215	2	2004/217 7:29	2004/219 6:53	47.4	49.5	9	53.5	96.5	No	V	
46	2004_243	3	2004/246 15:53	2004/247 17:17	25.4	160.5	12	37.5	66.5	2004/246?		

Table 1. (continued)

I	Run	N	Start Time	End Time	Dur	C-PA	Wid	Emin	Emax	ID	Cls	NB[b]
47	2004_255	4	2004/257 4:11	2004/259 6:59	50.8	202.5	42	25.5	96.5	No	E	1
48	2004_311	4	2004/311 0:10	2004/315 21:0	116.8	112.5	54	53.5	149.5			
	2004_311	1	2004/310 22:29	2004/312 15:7	40.6	16.5	9	31.5	96.5			1
	2004_311	2	2004/311 1:52	2004/312 16:48	38.9	34.5	12	53.5	114.5	2004/310,		1
			2004/310 22:29	2004/315 21:0	118.5	54	129	31.5	149.5	2004/311		2
49	2004_323	3	2004/323 5:6	2004/325 9:36	52.5	112.5	54	64.5	149.5	No	G	
50	2004_323	6	2004/324 13:17	2004/325 9:36	20.3	337.5	27	35.5	80.5	No	V	
51	2004_327	2	2004/329 6:44	2004/331 23:4	64.3	226.5	15	54.5	96.5	No		1
52	2004_339	1	2004/338 22:42	2004/342 6:17	79.6	94.5	93	53.5	149.5			
	2004_339	4	2004/338 22:42	2004/342 6:17	79.6	355.5	54	35.5	149.5	2004/339a-c		
			2004/338 22:42	2004/342 6:17	79.6	35	172	35.5	149.5			2
53	2004_351	5	2004/351 10:25	2004/355 2:28	88.1	349.5	12	84.5	120.5			
	2004_351	1	2004/351 12:7	2004/355 17:42	101.6	16.5	9	83.5	125.5			
	2004_351	2	2004/353 8:8	2004/355 17:42	57.6	124.5	9	53.5	96.5			
	2004_355	1	2004/354 23:5	2004/355 17:42	18.6	136.5	45	83.5	106.5			
	2004_355	4	2004/354 23:5	2004/355 17:42	18.6	4.5	12	105.5	130.5			
	2004_355	5	2004/356 2:10	2004/357 8:39	30.5	4.5	12	130.5	149.5	2004/354		
			2004/351 10:25	2004/357 8:39	142.2	71	176	53.5	149.5			
54	2004_359	3	2004/359 1:17	2004/360 14:33	37.3	358.5	12	53.5	95.5			
	2004_359	2	2004/361 5:47	2004/363 22:8	64.3	127.5	42	37.5	96.5			
			2004/359 1:17	2004/363 22:8	116.9	70.5	156	37.5	96.5	2004/361		
55	2004_363	1	2004/365 1:13	2005/1 20:57	67.7	154.5	9	77.5	111.5	No	G	
56	2005_005	1	2005/7 12:25	2005/9 21:59	57.6	136.5	36	53.5	126.5	2005/004?		3
57	2005_009	2	2005/9 20:17	2005/11 23:5	50.8	115.5	33	37.5	96.5	2005/010?		
58	2005_021	1	2005/22 14:46	2005/25 22:21	79.6	133.5	18	26.5	82.5	2005/024?		
59	2005_025	1	2005/27 1:27	2005/29 22:52	69.4	127.5	66	25.5	149.5			
	2005_029	2	2005/28 23:10	2005/30 20:53	45.7	112.5	15	95.5	149.5			
	2005_029	3	2005/28 23:10	2005/31 0:16	49.1	139.5	24	105.5	149.5			
	2005_029	4	2005/29 0:51	2005/32 21:59	93.1	349.5	12	54.5	125.5	2005/027,		
			2005/27 1:27	2005/32 21:59	140.5	67.5	168	25.5	149.5	2005/028ab		2
60	2005_049	2	2005/51 6:14	2005/53 2:15	44.0	112.5	12	53.5	96.5	No	P	
61	2005_053	1	2005/54 5:21	2005/55 13:31	32.2	31.5	30	24.5	95.5			
	2005_053	3	2005/54 5:21	2005/55 13:31	32.2	127.5	9	53.5	82.5			
	2005_053	6	2005/55 16:55	2005/57 21:24	52.5	19.5	36	53.5	96.5			
	2005_053	4	2005/55 18:36	2005/57 11:14	40.6	127.5	9	72.5	96.5	2005/054?		
			2005/54 5:21	2005/57 21:24	88.0	67	130.5	24.5	96.5			
62	2005_061	3	2005/61 21:55	2005/65 8:53	83.0	16.5	63	22.5	126.5			
	2005_061	1	2005/65 13:58	2005/65 22:26	8.5	19.5	18	103.5	126.5			
	2005_061	2	2005/65 13:58	2005/65 22:26	8.5	40.5	9	105.5	126.5	2005/062,		
	2005_065	1	2005/65 13:58	2005/67 21:50	55.9	4.5	27	84.5	148.5	2005/063		
			2005/61 21:55	2005/67 21:50	143.9	16.5	63	22.5	148.5			2
63	2005_077	1	2005/77 7:3	2005/81 22:48	111.8	118.5	15	60.5	121.5	No	G	
64	2005_089	3	2005/89 5:12	2005/89 20:27	15.2	40.5	9	53.5	65.5		E	
	2005_089	1	2005/90 4:54	2005/93 17:34	84.7	28.5	45	53.5	148.5			
			2005/89 5:12	2005/93 17:34	108.4	28.5	45	53.5	148.5	No		
65	2005_097	7	2005/99 0:34	2005/100 20:35	44.0	16.5	60	31.5	106.5	No	P	
66	2005_161	4	2005/161 17:51	2005/165 21:45	99.9	112.5	45	65.5	149.5	2005/162		
67	2005_169	1	2005/171 11:31	2005/172 23:5	35.6	31.5	9	33.5	66.5	2005/171?		1
68	2005_173	1	2005/173 5:51	2005/177 21:36	111.8	82.5	114	57.5	149.5			
	2005_173	5	2005/173 9:14	2005/177 18:13	105.0	346.5	24	31.5	143.5			
			2005/173 5:51	2005/177 21:36	111.8	57	165	31.5	149.5	2005/173		
69	2005_177	4	2005/180 5:29	2005/181 23:49	42.3	7.5	33	31.5	91.5			
	2005_181	4	2005/180 22:25	2005/184 7:41	81.3	349.5	42	31.5	96.5			
			2005/180 5:29	2005/184 7:41	98.2	356	55.5	31.5	96.5	2005/180		
70	2005_197	5	2005/197 0:28	2005/200 21:36	93.1	265.5	63	53.5	149.5			
	2005_197	4	2005/197 3:51	2005/199 6:39	50.8	223.5	15	53.5	134.5			
			2005/197 0:28	2005/200 21:36	93.1	256.6	81	53.5	149.5	2005/197		
71	2005_205	2	2005/204 23:48	2005/207 22:55	71.1	58.5	27	53.5	96.5		P	
	2005_205	3	2005/208 5:42	2005/209 22:20	40.6	61.5	27	76.5	149.5			
			2005/204 23:48	2005/209 22:20	118.5	60	30	53.5	149.5	No		
72	2005_209	1	2005/212 16:22	2005/213 22:51	30.5	103.5	18	67.5	126.5			
	2005_213	1	2005/212 23:8	2005/214 22:33	47.4	64.5	15	53.5	114.5			
			2005/212 16:22	2005/214 22:33	54.2	85	55.5	53.5	126.5	2005/212b		
73	2005_241	2	2005/243 5:32	2005/244 18:47	37.3	121.5	12	105.5	149.5	No	G	
74	2005_277	1	2005/278 5:22	2005/281 23:7	89.7	85.5	84	77.5	149.5	No	P	
75	2005_293	1	2005/295 20:40	2005/297 21:47	49.1	10.5	15	61.5	126.5		P	
	2005_293	4	2005/296 23:46	2005/297 23:28	23.7	91.5	24	66.5	126.5			
			2005/295 20:40	2005/297 23:28	50.8	53	100.5	61.5	126.5	No		
76	2005_321	3	2005/322 9:19	2005/324 8:44	47.4	337.5	27	30.5	132.5		P	
	2005_325	6	2005/324 22:17	2005/326 21:41	47.4	355.5	18	83.5	139.5			
			2005/322 9:19	2005/326 21:41	108.4	344	40.5	30.5	139.5	No		
77	2005_341	2	2005/342 13:35	2005/345 21:10	79.6	352.5	42	83.5	149.5	No	P	

Table 1. (continued)

I	Run	N	Start Time	End Time	Dur	C-PA	Wid	Emin	Emax	ID	Cls	NB[b]
78	2005_349	2	2005/350 9:32	2005/353 22:11	84.7	118.5	24	53.5	126.5			
	2005_353	2	2005/352 22:29	2005/354 23:35	49.1	130.5	24	92.5	149.5			
			2005/350 9:32	2005/354 23:35	110.0	124.5	36	53.5	149.5	2005/352		
79	2005_357	2	2005/359 15:20	2005/361 23:13	55.9	118.5	27	83.5	130.5			
	2005_357	1	2005/360 13:21	2005/361 23:13	33.9	67.5	9	105.5	149.5			
			2005/359 15:20	2005/361 23:13	55.9	97.5	69	83.5	149.5	2005/359		
80	2006_016	1	2006/16 2:4	2006/17 6:51	28.8	43.5	9	105.5	149.5	No	P	
81	2006_048	1	2006/48 1:5	2006/50 15:44	62.7	46.5	9	65.5	126.5	No	P	
82	2006_064	1	2006/66 14:24	2006/68 23:58	57.6	37.5	39	53.5	149.5		G	
	2006_064	5	2006/66 14:24	2006/68 23:58	57.6	1.5	21	83.5	149.5			
	2006_068	1	2006/67 22:34	2006/69 10:8	35.6	34.5	21	105.5	149.5			
	2006_068	3	2006/67 22:34	2006/69 10:8	35.6	1.5	9	91.5	149.5	No		
			2006/66 14:24	2006/69 10:8	67.7	24	66	53.5	149.5			
83	2006_072	1	2006/74 5:16	2006/76 16:32	59.3	19.5	15	31.5	120.5		P	
	2006_072	3	2006/74 5:16	2006/76 23:18	66.0	4.5	9	53.5	125.5			
			2006/74 5:16	2006/76 23:18	66.0	13.5	27	31.5	125.5	No		

[a]The columns are: I–The event index, Run–the start time of the search run in which the feature was found, N–the feature number within that run (these two fields can be used to cross reference to Table S1 of feature identifications), Start Time, End Time–The start and end times of the event in Y/D H:M format, Dur–The duration of the event in hours, C-PA–The central position angle (degrees from North), Wid–The angular width of the event, Emin, Emax–The range of elongations over which the event was detected, ID–The identification(s) in the manual list ("No" indicates that there was no identification), Cls–For those events with no match in the manual list, this is an estimate of the confidence class of the event. , NB– Special features of the event or its identification, refers to the notes at the end of the table. Where several AICMED events appear to be part of the same CME, the individual components are followed by a summary line for the whole event.

[b]1) This event was rejected as noise by the automated filter, but determined on inspection to be real. 2) These events from the "Webb" list appear to be parts of a single event. 3) This "Webb" event appears to be the same as the AICMED event but seen at a different elongation range.

5 years later. We have concentrated our efforts on the earlier part of the mission for two reasons: first from late-2006 through 2010, solar activity was very low and so there were few CMEs to detect; and secondly the performance of the SMEI cameras, especially Camera 3 which looks nearest to the Sun, degraded significantly through the course of the mission.

[33] In order to run AICMED on this data set, it was divided into 5-day intervals with 1 day of overlap between the intervals so as to allow CMEs that crossed the boundaries of the intervals to be detected. When AICMED was run on a set of sample data, it was very clear that while it was successful detecting CMEs there were also a large number of false positives even after the automatic rejection of the edges of noise features. The majority of the false positives were regions of particle contamination which were below the threshold to be rejected as noise. These can move slightly from orbit to orbit in such a way as to produce an apparent outward motion on any position-angle slice. Because of this the movie for each sub-interval must be inspected by eye, and an assessment made of whether a feature is actually a CME or something else. At this stage we also decide (again by the judgment of an observer) whether features identified as separate CMEs are in fact part of the same structure divided up by noise features either in time or in position angle. A summary of all the detections and their status is presented in Table S1 of the auxiliary material.

[34] The raw output of AICMED identified 713 features that were considered to be CMEs. Of these almost half (335) were identified as being noise edges by the automatic filter (we subsequently found that of these 17 (5%) were in fact real CME features that had been rejected because they were too close to noise regions), leaving 378 features to be inspected manually. Of those we found that 268 were still false positives, leaving 110 (127 including the false negatives) features that were real CME structures. We also found

that several of these appeared to be a single structure divided into several parts, leaving a total of 83 separate CMEs identified by AICMED. We should note here that although we refer to all features as CMEs a CIR would most likely be detected as a very slow CME. However, since there were no confirmed CIRs in the data sample used we cannot be certain.

[35] The commonest causes of false positives were found to be noise edges (142) and artifacts such as rings generated by hot pixels in Camera 3 (75). In addition we also found detections of scattered light from the moon (8), of comet tails (3) and of aurora (4). There were also 32 features for which there was no obvious explanation. While this is an undesirably large number of false positives, all the adjustments of parameters that we have tried to reduce this have resulted in an unacceptably large number of false negatives. We consider that, since most of the false positives can be easily spotted by a human observer, this is preferable to failing to detect large numbers of real features. This does however lead us to conclude that AICMED cannot be used entirely autonomously on SMEI data.

5.1. Comparison Between AICMED and Manual CME Observations

[36] The manual list of SMEI events was compiled by D. Webb and is reported for the time period of interest by *Webb et al.* [2006] and *Howard and Simnett* [2008]. This list is maintained online via the NSO SMEI web page, linked from http://smei.nso.edu/smeilinks.html. Over the time interval considered in this paper, a total of 184 CMEs were reported from manual observations and measurements. This compares with the 83 identified by AICMED (although as we shall see it is probably more appropriate to compare this with the 127 features). In the manual list, CMEs are classified into five quality classes based on the ease with which the CME could be identified and tracked (E = excellent, V = very

Table 2. Reasons for Initial AICMED Non-detections

Reason	E	V	G	P/?
Unexplained	0	0	2	6
Spurious or misclassified	1	3	1	5
Zodiacal Light[a]	1	6	5	3
Other image problems[b]	1	1	0	0
Part of another[c]	3	1	1	0
Too close[d]	0	0	2	0
False noise rejection	2	5	0	2
User Error[e]	2	2	1	0

[a]Using the 3-day background model means that there is a significant fluctuation in the zodiacal light close to the Sun making events that do not pass about 35° elongation difficult to detect.

[b]E.g. problems with the background model which cause a large residual background to remain.

[c]In these cases it appears that features listed as separate events in the manual list are actually different parts of the same structure, but in the original identifications only one of the manual-list events was identified as matching the AICMED CME.

[d]If 2 CMEs follow one another closely and in the same PA range, they can be merged into one.

[e]Cases where on re-examination, it was found that the manual-list event did actually match with an AICMED event.

good, G = good, P = poor and ? = dubious), and so when we compare the performance of AICMED we look at the success in detecting CMEs in each class.

[37] For all the events detected by AICMED, we examined the CME parameters generated and compared them with those of the CMEs listed in the manual list. We then manually decided whether they were in sufficient agreement (in position angle, elongation and angular speed) to consider them the same event. The identifications are summarized in Table 1.

5.1.1. CMEs Not Detected by AICMED

[38] Since there are many more CMEs in the manual list than were detected by AICMED we must determine what is causing AICMED to fail to detect some CMEs. For this we looked individually at all the events of class E or V which were not detected. For events of class G and below, there are many more events in the manual list so for these we only examined the events in the first half of 2004 as a representative sample. In both cases, we looked at the movies made by AICMED as well as re-examining the Aitoff maps using a number of different background schemes. On the basis of this we assigned a reason for the non-detection if one could be determined, these reasons are summarized in Table 2.

[39] The most significant result here is that for events of classes E and V, there were no unexplained non-detections, and even for the class G events only a small fraction cannot be explained.

5.1.2. CMEs Not in the Manual List

[40] Of the 83 CMEs identified by AICMED and confirmed to be CMEs by a human observer, 37 could not be matched with a CME identified in the manual list. Most of these are relatively faint events and so it is probable that many were simply overlooked at the time, this is especially likely as the data processing in use at the time was less sophisticated than the version that we have used in this study. There are however a small number of CMEs that would be classified as V or even E that had not previously been identified. Other than the fact that most were faint, there does not seem to be any common characteristic of these newly identified CMEs. We do however notice that

there are a number of "clumps" of new CMEs close together in time; and since the manual list was compiled by collecting together reports from members of the SMEI team it is quite possible that some intervals were simply never examined.

6. Discussion

[41] Since manually identifying CMEs in heliospheric imager data is a somewhat time-consuming process and liable to errors (especially of omission) it is clearly desirable to have an automatic tool to do this. We have found that by using techniques similar to those employed by the CACTus program developed for coronagraph data it is possible to identify CMEs in SMEI images. There are however a number of qualifications that mean that this technique is not fully automatic and it appears to us that it will never be possible for the detection to be entirely automated for SMEI data.

[42] *False positives*: It is very apparent that as well as detecting CMEs, there are a large number of false positives in the AICMED features. The commonest cause of these is the presence of noise features (typically regions of intense particle hits in the polar caps and the South Atlantic Anomaly (SAA) which overwhelm the hit removal algorithm) that almost repeat from orbit to orbit but move enough to cause AICMED to identify them as a moving feature. In addition we noted a number of cases where AICMED had identified a real CME but conflated it with a noise feature. While the noise-proximity filter that we are currently using removes many of these false positives, many remain and making the filter stricter causes an unacceptably large number of real CMEs to be rejected.

[43] *Fragmentation*: AICMED works by identifying contiguous regions of enhanced brightness with properties consistent with those expected for CMEs. Therefore if a CME arc crosses a region of noise, or appears in orbits separated by a number of bad orbits, that CME will be marked as two or more separate features. At the present time the only means by which it is possible to connect these together is by the judgment of an observer.

[44] *Incomplete detection*: We also noted cases where although AICMED had detected CMEs it was clear that the real extent of the CME was considerably greater than that detected by AICMED. Clearly for any modeling of the CME it is desirable that the full extent be used.

[45] Despite these limitations we find that AICMED is a very promising tool to assist an observer in detecting CMEs. It is reliable for the strong CMEs that progress beyond about 45° elongation, which are the ones likely to have significant space weather consequences. It is not a time-consuming program, each 5-day subinterval was scanned in approximately 1 minute.

[46] Two main priorities exist in further developing this method: first adapting it to be able to work on data from the STEREO HI instruments [*Eyles et al.*, 2009], and; secondly developing a better method of rejecting the false positives without also rejecting real CMEs.

[47] Although AICMED does output leading edge and noise region files suitable for use in the TH model [*Tappin and Howard*, 2009b; *Howard and Tappin*, 2010], we have found that the problems outlined above mean that it is better to use AICMED to locate the CME and then measure it by hand and use those measurements as inputs to the model. In

other words AICMED, like CACTus, is best used as a region-of-interest identifier, with validation and scientific measurements performed manually.

7. Conclusions

[48] We have shown that it is possible to use automated techniques to detect CMEs in heliospheric imager data.

[49] With SMEI data there are a large number of false positives. This is due to the very noisy data set, and means that manual assessment of the identifications is necessary. With appropriate design and data-processing decisions this problem should be much reduced in future imagers.

[50] Despite the number of false positives, AICMED was able to detect the vast majority of those CMEs likely to have significant space weather consequences, and was also able to identify a considerable number of CMEs that had been overlooked in manual inspection of the data.

[51] For a fully automated system it will be necessary to drastically reduce the number of false positives. While some progress may be made by improvements in the algorithm (particularly the noise-proximity filter), it is also important to consider noise sources in the design of future imagers if automated CME detection for forecasting purposes is to be a goal. Most important among these is regions of particle hits, which account for more than half of the false positives. We suggest that the use of an instrument similar in concept to SMEI, but with a wide field of view in the scanning direction is the best solution as this would ensure that most if not all of the sky would be imaged at times away from the polar caps and SAA. The alternative would be to place the instrument well away from the geospace environment, but this carries a significant cost in terms of data latency which is a major problem for real-time use. We are currently developing a version of AICMED of use with the STEREO/HIs which will be reported in a later publication.

Appendix A: Data Processing

[52] In this appendix we provide a summary of the processes required to convert the raw data from SMEI into sky maps in which CMEs can be detected.

[53] SMEI maps the sky by taking images of narrow strips of the sky as it travels around its orbit. The generation of all-sky images from these raw image frames is a non-trivial process, and a number of different pipelines exist. The images used in this study were generated using the so-called NSO pipe. Since this pipeline has not been described in the literature we will describe the process briefly in the following paragraphs.

[54] 1. Determine the geographical location of the Coriolis spacecraft as a function of time and divide the data stream into intervals of one orbit. We choose to make the break between orbits as the spacecraft crosses $70°$S in a northerly direction–this was chosen as it generally occurs during the south-polar cap passage and thus makes the discontinuity in the image less apparent.

[55] 2. Find the raw image frames that are in the desired orbit.

[56] 3. Subtract the various background contributions from each frame. These contributions include: electronic bias (the level returned by the A-D converter when no signal

is applied), dark charge (noise in the detector when it is not illuminated) and glare (a form of stray light within the instrument optics).

[57] 4. Apply gain corrections to the individual pixels.

[58] 5. Identify and flag frames contaminated by bright objects (the Sun, the Moon, Venus, etc.).

[59] 6. Some pixels (especially in Camera 3, nearest the Sun) flip between a normal state and a hot state in which they show an elevated dark charge level. It is therefore necessary to attempt to correct any pixels that have flipped between states, and eliminate those that cannot be corrected.

[60] 7. Locate and eliminate pixels that are affected by particle hits, and flag frames with very high hit rates.

[61] 8. Determine the location of bright stars on the raw image frame and subtract them.

[62] 9. Map the raw image frame onto a Right Ascension (RA) and Declination (Dec) grid to produce an intermediate stage map. A separate map is generated for each of the three SMEI cameras.

[63] 10. The remaining steps are then carried out on these intermediate maps.

[64] 11. Subtract a faint star pattern that was generated from about two years of intermediate maps.

[65] 12. Match the overlap regions between the cameras.

[66] 13. Transform the RA and Dec. orbit map into the final projection in heliocentric ecliptic coordinates. Three different projections are routinely generated:

[67] *Hammer-Aitoff* : This projection maps the entire sky onto an elliptical image, and is the most widely used to display SMEI data in publications.

[68] *Zenithal-Equidistant* : Also commonly referred to as "Fisheye", shows the sky out to an elongation of $135°$. This is often used in manual CME searches as CMEs move radially outwards.

[69] *Rectangular* : A simple mapping of the entire sky into position angle (measured from ecliptic north) and elongation coordinates. While this is not a particularly easy projection for human observers, it is very convenient for automated studies such as that presented here. In this projection CMEs move vertically from bottom to top.

[70] To produce images in which CMEs are visible, further subtraction is needed to remove the contributions of the zodiacal light (F-corona) and also the Thomson-scattered light from the ambient solar wind. Regions of intense aurora are flagged as bad data. Since there are several ways to do this depending on what features are being sought, this process is the responsibility of the end-user display tools. A number of schemes are implemented. Among the frequently used methods are "running difference" where the previous orbit's image is subtracted from the current image and a 6-orbit base subtraction where the median of the three previous and three following orbits is subtracted. For this study we have used a running 3-day median, as this is long-enough that slowly moving features are not removed with the background (a common problem with running-difference subtraction), but short enough that the zodiacal light does not change significantly.

References

Baker, D. N., et al. (2009), *Severe Space Weather Events: Understanding Societal and Economic Impacts: A Workshop Report*, Natl. Acad. Press, Washington, D. C.

Buffington, A., M. M. Bisi, J. M. Clover, P. P. Hick, B. V. Jackson, and T. A. Kuchar (2008), Analysis of plasma-tail motions for comets C/2001 Q4 (NEAT) and C/2002 T7 (LINEAR) using observations from SMEI, *Astrophys. J.*, *677*, 798.

Davies, J. A., R. A. Harrison, A. P. Rouillard, N. R. Sheeley Jr., C. H. Perry, D. Bewsher, C. J. Davis, C. J. Eyles, S. R. Crothers, and D. S. Brown (2009), A synoptic view of solar transient evolution in the inner heliosphere using the Heliospheric Imagers on STEREO, *Geophys. Res. Lett.*, *36*, L02102, doi:10.1029/2008GL036182.

Davis, C. J., J. A. Davies, M. Lockwood, A. P. Rouillard, C. J. Eyles, and R. A. Harrison (2009), Stereoscopic imaging of an Earth-impacting solar coronal mass ejection: A major milestone for the STEREO mission, *Geophys. Res. Lett.*, *36*, L08102, doi:10.1029/2009GL038021.

Duda, R. O., and P. E. Hart (1972), Use of the Hough Transformation to detect lines and curves in pictures, *Comm. ACM*, *15*, 11.

Echer, E., M. V. Alves, and W. D. Gonzalez (2005), A statistical study of magnetic cloud parameters and geoeffectiveness, *J. Atmos. Sol. Terr. Phys.*, *67*, 839.

Eyles, C. J., G. M. Simnett, M. P. Cooke, B. V. Jackson, A. Buffington, N. R. Waltham, J. M. King, P. A. Anderson, and P. E. Holladay (2003), The solar mass ejection imager (SMEI), *Solar Phys.*, *217*, 319.

Eyles, C. J., et al. (2009), The heliospheric imagers onboard the STEREO mission, *Solar Phys.*, *254*, 387.

Howard, T. A., and G. M. Simnett (2008), Interplanetary coronal mass ejections that are undetected by solar coronagraphs, *J. Geophys. Res.*, *113*, A08102, doi:10.1029/2007JA012920.

Howard, T. A., and S. J. Tappin (2010), Application of a new phenomenological coronal mass ejection model to space weather forecasting, *Space Weather*, *8*, S07004, doi:10.1029/2009SW000531.

Howard, R. A., et al. (1975), Synoptic observations of the solar corona during Carrington rotations 1580–1596 (11 October 1971–15 January 1973), *Rep. UAG 48A*, World Data Cent. Sol.-Terr. Phys., Boulder, Colo.

Howard, T. A., J. C. Johnston, and D. F. Webb (2006), Coronal mass evolution in the inner heliosphere: Geometry and speed, *Eos Trans. AGU*, *87*(52), Fall Meet. Suppl., Abstract SH32A-04.

Hundhausen, A. J., C. B. Sawyer, L. House, R. M. E. Illing, and W. J. Wagner (1984), Coronal mass ejections observed during the solar maximum mission: Latitude distribution and rate of occurrence, *J. Geophys. Res.*, *89*, 2639.

Kuchar, T. A., A. Buffington, C. N. Arge, P. P. Hick, T. A. Howard, B. V. Jackson, J. C. Johnston, D. R. Mizuno, S. J. Tappin, and D. F. Webb (2008), Observations of a comet tail disruption induced by the passage of a CME, *J. Geophys. Res.*, *113*, A04101, doi:10.1029/2007JA012603.

Liu, Y., A. Thernisien, J. G. Luhmann, A. Vourlidas, J. A. Davies, R. P. Lin, Robert, and S. D. Bale (2010), Reconstructing coronal mass ejections with coordinated imaging and in situ observations: Global structure, kinematics, and implications for space weather forecasting, *Astrophys. J.*, *722*, 1762.

Marubashi, K. (2000), Physics of interplanetary magnetic flux ropes: Toward prediction of geomagnetic storms, *Adv. Space Res.*, *26*, 55.

Munro, R. H., and D. G. Sime (1985), White-light coronal transients observed from SKYLAB May 1973 to February 1974: A classification by apparent morphology, *Solar Phys.*, *97*, 191.

Richter, I., C. Leinert, and B. Planck (1982), Search for short term variations of zodical light and optical detection of interplanetary plasma clouds, *Astron. Astrophys.*, *110*, 115.

Robbrecht, E., and D. Berghmans (2004), Automated recognition of coronal mass ejections (CMEs) in near-real-time data, *Astron. Astrophys.*, *425*, 1097.

Robbrecht, E., D. Berghmans, and R. A. M. Van der Linden (2009), Automated LASCO CME catalog for solar cycle 23: Are CMEs scale invariant?, *Astrophys. J.*, *691*, 1222.

St. Cyr, O. C., et al. (2000) Properties of coronal mass ejections: SOHO LASCO observations from January 1996 to June 1998, *J. Geophys. Res.*, *105*, 18169.

Tappin, S. J., and T. A. Howard (2009a), Direct observation of a corotating interaction region by three spacecraft, *Astrophys. J.*, *702*, 862.

Tappin, S. J., and T. A. Howard (2009b), Interplanetary coronal mass ejections observed in the heliosphere: 2. Model and data comparison, *Space Sci. Rev.*, *147*, 55.

Tappin, S. J., et al. (2004), Tracking a major interplanetary disturbance with SMEI, *Geophys. Res. Lett.*, *31*, L02802, doi:10.1029/2003GL018766.

Webb, D. F., et al. (2006), Solar Mass Ejection Imager (SMEI) observations of coronal mass ejections (CMEs) in the heliosphere, *J. Geophys. Res.*, *111*, A12101, doi:10.1029/2006JA011655.

Wilson, R. M. (1990), On the behavior of the Dst geomagnetic index in the vicinity of magnetic cloud passages at Earth, *J. Geophys. Res.*, *92*, 215.

Yashiro, S., N. Gopalswamy, G. Michalek, O. C. St. Cyr, S. P. Plunkett, and R. A. Howard (2004), A catalog of white light coronal mass ejections observed by the SOHO spacecraft, *J. Geophys. Res.*, *109*, A07105, doi:10.1029/2003JA010282.