AUTOMATED DETECTION OF CMES IN LASCO DATA

D. Berghmans[1,2], B. H. Foing[2], and B. Fleck[2]

[1]Royal Observatory of Belgium, Ringlaan 3, B-1180 Brussels, Belgium
[2]ESA Research and Scientific Support Department, ESTEC, 2201 AG Noordwijk, The Netherlands

ABSTRACT

We have developed software that autonomously detects CMEs in image sequences from LASCO. The crux of the software is the detection of CMEs as bright ridges in [height, time] maps using the Hough transform. The output is a list of events, similar to the classic catalogs, with principle angle, angular width and velocity estimation for each CME. In contrast to catalogs assembled by human operators, these CME detections by software can be faster, which is especially important in the context of space weather, and possibly also more objective, as the detection criterion is written explicitly in a program. In this paper we describe the software and validate its performance by comparing its output with the visually assembled CME catalogs. We discuss its present success rate (about 75%) and prospects for improvement. Finally, we show that the software can also reveal CMEs that have been not been listed in the catalogs. Such unreported cases might be of influence on CME statistics and prove that also the present catalogs do not have a 100% success rate.

Key words: Sun: corona, Sun: particle emission.

1. INTRODUCTION

Over the past 6 years, coronal mass ejections (CMEs) have been detected routinely by visually checking each image from the Large Angle Spectrometric Coronagraph (LASCO, Brueckner et al. (1995)) on-board SOHO (Domingo et al., 1995). Event catalogs have been assembled continuously and are made publicly available (http://lasco-www.nrl.navy.mil/cmelist.html, but also http://cdaw.gsfc.nasa.gov/CME_list). Each observed CME is listed with its time of appearance, the angle of its central axis, its angular span, velocity and acceleration estimates and a short morphological description. These catalogs are used as a reference and form a valuable resource for further statistical analysis on the nature of CMEs (St. Cyr et al., 2000).

The visual detection of CMEs in the flood of incoming new data is a labor intensive task. It is up till now essentially the human eye that detects a CME occurrence and a scientist that collects all the CME parameters in the catalogs. With the future coronagraphs on the 2 STEREO spacecraft or on the SDO mission this will become a big investment of man power. Meanwhile near real time alerts for halo CMEs are needed by the space weather community. Although halo CMEs take a few days to arrive to the Earth, their detection is timely as CPU-time intensive 3D MHD simulation are required to estimate their geoeffectiveness. This implies that CME halo alerts should be issued 24 hours per day.

Moreover, the subjective interpretation by a human operator or scientist makes it doubtful whether this visual CME detection is stable over a solar cycle, as the operator gains experience or personnel is replaced. There is probably not much confusion for big, well structured events but small and/or weak events might be arbitrarily detected or not.

For all these reasons, we have developed a software package that detects CMEs in coronagraphic images. In this paper we will present the inner machinery of this package. A first preprocessing module (section 2) merges the C2 and C3 images, cleans, rebins and reformats them with every step optimized for improving the CME contrast. The second image-recognition module (section 3) then extracts motion patterns and groups the patterns in CMEs. In section 4, we estimate the performance of the package and in section 5 we list our main conclusions.

2. THE PREPROCESSING MODULE

The default processing of the LASCO images (see eg latest images on http://sohowww.nascom.nasa.gov) is not optimized for CME detection. The CME signature is convolved with quasi-static K-corona streamer structures and with slowly moving stars, planets and comets as well as the instrumental stray light and F-corona backgrounds. Also, towards the edge of the field-of-view (FOV), the CME contrast diminishes, which makes it difficult to follow outmoving features.

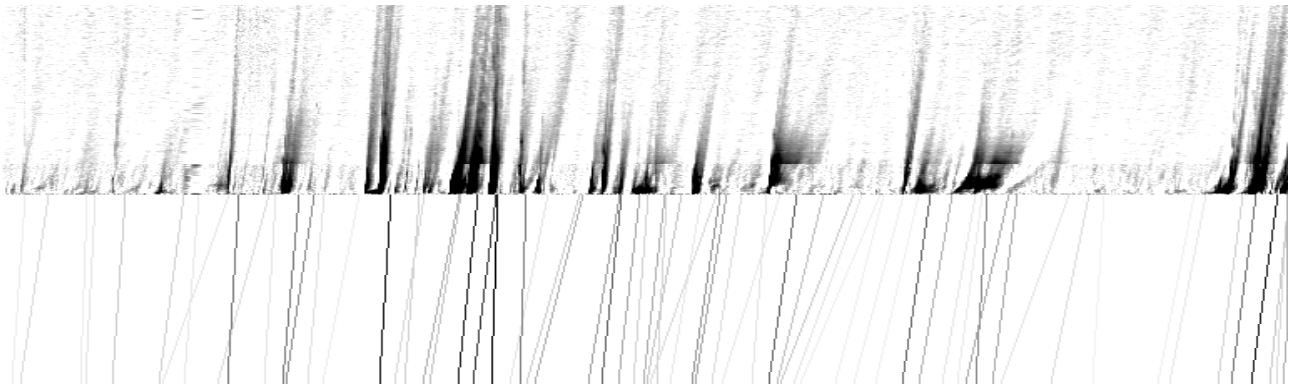

Figure 1. Example of a [time,height] slice through the datacube (top) and the ridges detected in it with the Hough transform (bottom). The inclination angle of the ridges corresponds to the propagation velocity. The horizontal range corresponds to the month May 1998. The vertical range corresponds to the combined C2/C3 field of view.

The images are relatively large with a spatial resolution far beyond what is needed for CME detection. A typical CME is only a relatively weak variation in intensity and only visible in a few subsequent images. All this means that the 'CME signal' is only very scarcely present in the huge amount of incoming data. Straight application of image recognition techniques on the usual 1024x1024 images would therefore result in a giant computational overhead. Finally, the different spatial and temporal resolution of C2 and C3 data make a combined analysis difficult. To avoid all these complications, a preprocessing module is applied that reformats the input images:

- Each level 0.5 image from LASCO/C2 and C3 is read in. Exposure time normalization is applied and bright point like sources (cosmic ray hits, but also planets and stars) are removed.
- A polar transformation is applied: the $[x, y]$ FOV becomes a $[r, \theta]$ FOV (Figure 5), with θ the poloidal angle around the Sun and r the distance from the limb. By choosing the r-range appropriately, the dark occulter and corners are easily avoided. While transforming we rebin, from 1024x1024 pixels for the $[x, y]$ FOV to a 200x360 pixels $[r, \theta]$ FOV. This enhances the signal to noise ratio significantly, especially far from the disc, as the size $r\Delta\theta\Delta r$ of the 'footprint' of a $[r, \theta]$ pixel in $[x, y]$ images grows as r.
- The $[r, \theta]$ images originating from C2 and C3 are combined in a single composite image by rescaling/matching the different spatial and temporal resolution of the two coronagraphs. Since the LASCO C2 FOV is much smaller than that of C3, this step essentially comes down to adding a thin C2 strip at the bottom of the $[r, \theta]$ C3 images. To take into account the different observation times of the C2 and C3 images, a cubic spline interpolation is applied so that C2 images are matched to the default 1 hour cadence of C3.
- Stacking the composite $[r, \theta]$ images in a $[r, \theta, t]$ datacube, we derive a background as a running average over 1 day. For each $[r, \theta]$ pixel, a CME passage results in a short lived, positive deviation from the running average. In an iterative procedure, such deviations are identified and removed for the next iteration of the background calculation. After a few iterations, the resulting background contains only the variability on timescales larger than 1 day. In what follows, we consider relative deviations from the original $[r, \theta, t]$ datacube with respect to this background. This effectively removes the dust corona but also streamers that rotate in and out of the FOV.

The output of all this is a $[r, \theta, t]$ datacube which is much smaller than the total of the original input data, and in which most of the non-CME signal is removed or strongly attenuated.

3. THE IMAGE RECOGNITION MODULE

CMEs are seen as bright features moving outward from the Sun. It turned out not to be feasible to identify in each separate image the location of individual CMEs by segmentation techniques. The CMEs are too variable in appearance, they are often too weak to identify their extension (especially their trailing edge), and they easily get merged with one another.

Instead of trying to detect CMEs in each $[r, \theta]$ image, we looked at $[t, r]$ slices (Figure 1, top) for each θ in the $[r, \theta, t]$ datacube. If a $[t, r]$ slice cuts through a CME, an inclined ridge is seen in the $[t, r]$ slice. Detecting CMEs in $[t, r]$ slice was first introduced by Sheeley et al. (1999). Thanks to the preprocessing module however, our $[t, r]$ slices, have a much better contrast and contain less noise. Working with $[t, r]$ slice has the advantage that all CMEs look the same (inclined ridges) and that even weak CMEs show up with a clear signature. Finally, as a bonus, the propagation speed of the CME can be determined from the inclination angle of the CME.

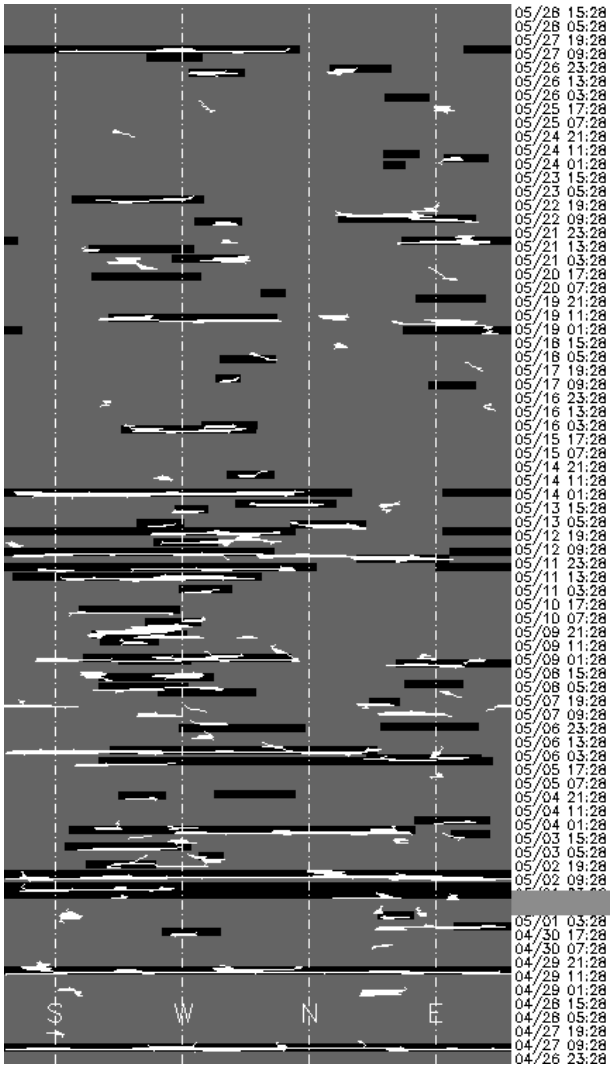

Figure 2. Comparison of the catalog CMEs (dark) and those found by the software (white). Time runs vertically over most of May, 1998. The poloidal angle runs counterclockwise from left (near C3 pylon) to right. The catalog time of appearance corresponds to the bottom of the dark boxes. The 'thickness' of the boxes is arbitrarily set to 8 hours.

A well known technique for detecting ridges is the so-called Hough transform (Jähne, 1997). We make an Hough transformation of every $[t, r]$ slice. In transformed space, we select the most significant signals, which after inversion correspond to the required ridges (Figure 1, bottom). Each ridge R in a $[t, r]$ slice at an angle θ_R is defined by its onset time t_R, its speed v_R and its brightness b_R. We build up a datacube $[v, \theta, t]$ by setting for each ridge $[v_R, \theta_R, t_R] = b_R$. Since a CME is a large scale structure, the onset time and velocity will only slightly differ from angle to angle. This means that a CME is represented in the $[v, \theta, t]$ datacube as a dense cluster of datapoints. The problem of detecting CMEs has thus been reduced to identifying clusters in a 3D scatter plot. We simply integrate the $[v, \theta, t]$ cube along the v-direction and identify the location of clusters in the resulting $[\theta, t]$ map (Figure 2).

4. VALIDATION OF THE SOFTWARE

We applied the software to the LASCO data from April, 27, 1998 to May, 27, 1998. This month is among the latest months for which a 'final CME catalog' exists (http://lasco-www.nrl.navy.mil/cmelist.html). For the period mentioned, the catalog lists 71 CMEs of which 4 were halo CMEs. In Figure 2 we show the angular span and time of occurrence of these catalog CMEs as dark boxes. The software found 95 events, which are shown as white elongated regions in Figure 2. The overal distribution in (angle, time) space is very similar and the number of CMEs found is of the same order of magnitude. Comparing the two sets in more detail is a delicate exercise. The 'success rate' of our software obviously depends on the tolerance allowed on the deviations.

Of the 71 catalog CMEs, 19 (27%) are reproduced with nearly identical time of appearance and angular location. Allowing for a reasonable tolerance on the time of appearance (within 3 hours) and on the angular span (at most 50%), the number of reproduced CMEs increases to 53 (75%). In this success rate we also included cases in which the software merged events that were listed as separate CMEs. At the other hand, about 10 catalog CMEs (14%) are missed completely. The remaining 11% are disputable detections eg when a CME is detected but the time of appearance deviates more than 3 hours from the catalog value. For space weather applications it is important to note that out of the 4 catalog halo CMEs, 2 are indeed reproduced as halo CMEs. The remaining 2 are missed because of a data gap (grey zone in time bar of Figure 2).

The software found 95 cases whereas there are only 71 CME entries in the catalog. Part of the difference between the two numbers is due to cases in which different parts of a CME are erroneously detected as separate events. Yet, we found about 15 (21%) detections which are 'far from' any catalog CME. Some of these are due to false alerts generated by fast streamer evolution. In at least some cases, our software has found 'unreported CMEs'. An example

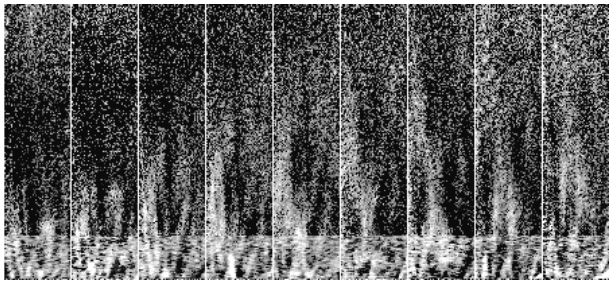

Figure 3. Example of an unreported CME, lifting off at 19h28 on April 27, 1998 with a speed of about 200 km/s. The subfield shown is 50 degrees wide, centered around the South direction. There is a 3 hours lapse between the different subfields.

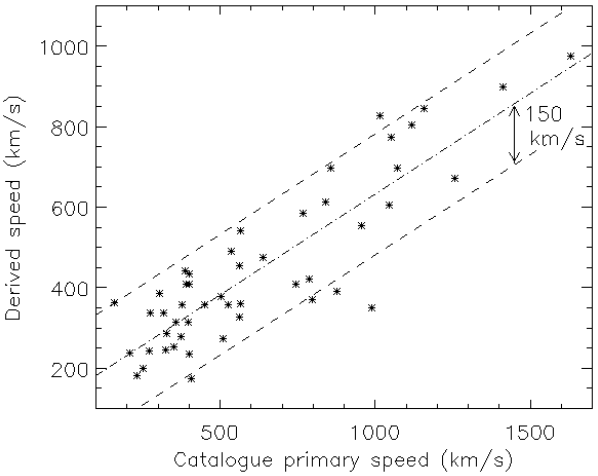

Figure 4. Comparison of the CME primary speeds listed in the catalog (horizontally) and a derived proxy for it (vertically) for those events which are well recovered by the software.

of such a case is shown in Figure 3. This means that also the catalogs do not have a 100% success rate.

Each CME detection is collected from ridge profiles in $[r,t]$ slices along adjacent θ angles in the $[r,\theta,t]$ datacube. This means that we have the velocity at each point along the detected CME front (Figure 5). The catalogs only give 1 velocity value called the 'primary speed' of the CME. This primary speed is typically derived from tracking the leading edge. In contrast, our software tracks the brightest features at each angle θ. We have tried to reproduce the primary speed by calculating a 'derived' speed as the average speed (weighted over brightness) of the ridges along the central 50% of the CME. In Figure 4 we show that the 'derived' speed (vertically) correlates well (linear Pearson correlation coefficient 0.873) with the catalog primary speed (horizontally). Given the good correlation, the derived speed can be transformed in an estimated primary speed

$$v_{primary} = 1.98 * (v_{derived} - 132)$$

The catalog also lists the acceleration of the CME. At the present time, this is not possible yet with the software since the Hough transform detects ridges in the $[r,t]$ slices as straight lines. This implicitly assumes constant velocity CMEs.

5. CONCLUSIONS

This paper shows that it is possible to automatically detect CMEs in coronagraphic images by software. The current version of our program processes 1 month of LASCO data in a few hours of CPU time. In such a run, about 75% of the CMEs listed in the catalogs are recovered. Note that the 'final' or 'version 2' catalogs which we compared with, are compiled from various sources and have undergone several iterations. In addition, the software did detect

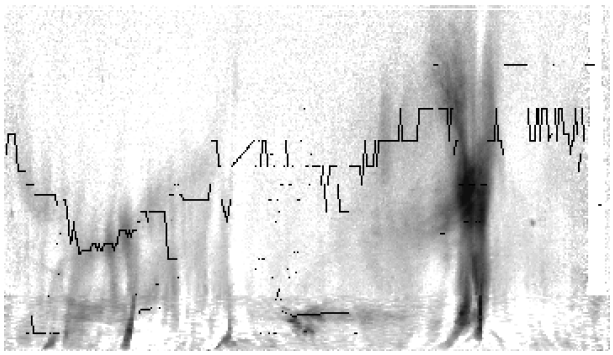

Figure 5. $[r,\theta]$ image showing the halo CME of April 29. The overplotted, dark, broken line is the detected CME front. Note that different parts of the CME propagating at different speeds are correctly tracked.

weak CMEs which were missing in the catalog, so also the human operators do not have a 100% success rate.

Several improvements in the preprocessing module can still be envisioned and there is hope to reach a 90% success rate. We assume that the remaining 10% are due to cases of 'intelligent interpretation' by the human operator eg with data gaps or with partially corrupted images. Future work will be based on the generalised Hough transform using parabolas instead of straight lines to detect the ridges. This will allow to additionally estimate the acceleration of the CMEs. We also plan to develop a (near) real time version of the software that outputs the detected events directly on the web at http://sidc.oma.be.

ACKNOWLEDGEMENTS

SOHO is a project of international cooperation between ESA and NASA. We acknowledge the LASCO consortium that made the instrument possible. This work is part of an ESA/PRODEX and a Belgian OSTC project on space weather and image recognition software. It is a pleasure to acknowledge stimulating discussions with J.-F. Hochedez and E. Verwichte on the latter subject. We appreciated comments by R. Van der Linden and C. St. Cyr.

REFERENCES

Brueckner G.E., Howard R.A., Koomen M.J., et al., 1995, Sol. Phys., 162, 357

Domingo V., Fleck B., Poland A.I., 1995, Sol. Phys., 162, 1

Jähne B., 1997, Digital Image Processing, 463, Springer-Verlag

Sheeley N.R., Walters J.H., Wang Y.M., Howard R.A., Nov. 1999, J. Geophys. Res., 104, 24739

St. Cyr O.C., Howard R.A., Sheeley N.R., et al., Aug. 2000, J. Geophys. Res., 105, 18169