Mining Coronal Loops in Solar Images

Nurcan Durak and Olfa Nasraoui
Web Mining and Knowledge Discovery Lab,
Department of Computer Engineering and Computer Science, University of Louisville
{nurcan.durak, olfa.nasraoui}@louisville.edu

Motivation
• Coronal loops are immense arches of hot gas on the surface of the Sun, thought to be jets of hot plasma flowing along in the alleys between the strong coronal magnetic fields.

• These loops are essential in studies of the Coronal Heating Problem which is one of the longest standing unsolved mysteries in astrophysics. The exact properties of temperature distribution along coronal loops help in understanding and modeling the Coronal Heating Problem.

• Astrophysicists have to prepare appropriate image data sets to complete studies of the solar loop properties. They need to acquire solar images with coronal loops.

• Astrophysicists use solar images taken by the Extraextreme Ultraviolet Imaging Telescope (EIT) on board the SOHO satellite.

• EIT monitors the sun and has been taking six solar images per day since 1996. That means there are 26820 EIT images in online NASA solar image databases.

• Astrophysicists have spend a lot of time:
  • to download images from NASA solar image databases,
  • to look into every solar image to detect which ones have loops, and
  • to determine the exact position of coronal loops.

Objective
The main objective of this study is:
• to design a system that can automatically discover the rare but interesting images with coronal loops. Such an automated system can help Astrophysicists save time and effort in acquiring the relevant data required for further analysis.

ACKNOWLEDGMENTS
This work is supported by the National Aeronautics and Space Administration under Grant No. NNB-05-0077-039 issued through the Office of Space Sciences and by National Science Foundation grant 04352435.

Methodology

1. Statistical Features: These features are extracted from intensity level blocks.

2. Hough-transform based features: Binary blocks are used to compute the Hough transform and the following edge-related features.

3. Hough-transform based EHD features: Instead of classical EHD (Edge Histogram) features, we define EHD features considering angle intervals.

4. Features capturing the spatial distribution of edges: The edge distribution in 4 horizontal bands are separated into 4 different features.

5. Curvature Based Features: We calculate the curvature degree to measure how curly the edges are.

• Training Classifiers:
  • We resorted to a supervised learning strategy, where a training data set with examples of blocks with and without loops are used to build a prediction model that can detect the occurrence of new loops based on computed features.
  • We used 150 solar images previously marked by experts. From these images, we acquired 403 LOOP blocks and 725 NO-LOOP blocks.
  • To validate classifier models, we performed 10-fold cross-validation on the data set.

After selecting the best feature set from all features, the AdaBoost classifier gave the best results: 0.643 precision and 0.596 recall.

Image Based Testing
To retrieve solar images containing loop shapes from online EIT image repositories, we subjected the unmarked images to the following process:

Testing takes 1.5 seconds per image

Experiments:
• 100 images (44 with loop and 56 without any loops) were subjected to the image based testing part of our solar loop mining system.
• Prediction results: 57 images have a loop and 43 images have no loops.

Precision = 72% Recall = 93%

Examples of system output:
Images with loops (Note: loop alignment is done automatically)
Images without loops

Conclusion
• Challenging problem: Loop shapes can be very diverse in their shape, size, and direction, and in some cases, are very hard to distinguish from other solar phenomena, such as coronal mass ejections (CME) and solar flares.

• Loop blocks require specially designed features to learn relevant loop shape characteristics in the training phase.

• While block-based precision and recall values were not high, image-based testing results were satisfying.