Second order statistics for the analysis of textural properties of coronal holes and filament channels in SDO/HMI and SDO/AIA 193Å images

M. Reiss[1], S.J. Hofmeister[1], M. Temmer[1], A.M. Veronig[1], R. De Visscher[2], V.Delouille[2], T. Rotter[1]

[1]University of Graz, Institute of Physics, IGAM-Kanzelhöhe Observatory, NAWI Graz, Graz, Austria
[2]Royal Observatory of Belgium, Brussels, Belgium

e-mail: martin.reiss@uni-graz.at

Abstract

A way to forecast the velocity of high speed solar wind streams is to empirically relate the area of coronal holes (CH) to the solar wind speed at 1AU with a lead time of about 4 days. This requires an automated method for the extraction of CH regions. We use intensity-based thresholding on SDO/AIA 193A images. Such intensity-based detection method cannot distinguish filament channels (FC) from CHs however. Here we propose to investigate the benefits of using Haralick's textural features to analyze the intrinsic texture information contained with CHs and FCs in AIA and HMI images. In combination with first order statistic and shape measures we tested several classifiers to find the most suitable decision rule for a differentiation between CHs and FCs. First results reveal that Support Vector Machine, Random Forest Classifier, and Decision Tree provide good results in general, although the exact performance may vary a lot from one data set to another.

Introduction

Motivation: Based on a detection method to obtain fractional CH areas from AIA 193Å images we provide automated forecasts of high speed solar wind streams arriving at Earth 4 days in advance [1].

Problem: To improve the forecast algorithm we need to distinguish FCs from CHs.

Methods: Image statistics and shape analysis was carried out in order to achieve reliable parameters for CH and FC classification. First order as well as second order image statistical parameters from AIA and HMI images together with shape descriptors [2] were calculated. These parameters were interpreted as attributes for data mining investigations in order to classify CHs and FCs.

Dataset: Based on visual inspection of H_α images over the time range 2011−2013, we extracted a set of 348 manually labelled CHs and 61 FCs.

2nd Order Statistics

Texture provides intrinsic information about the structural arrangement of pixel values in images. We present 2nd order statistics for the calculation of textural features of AIA and HMI images.

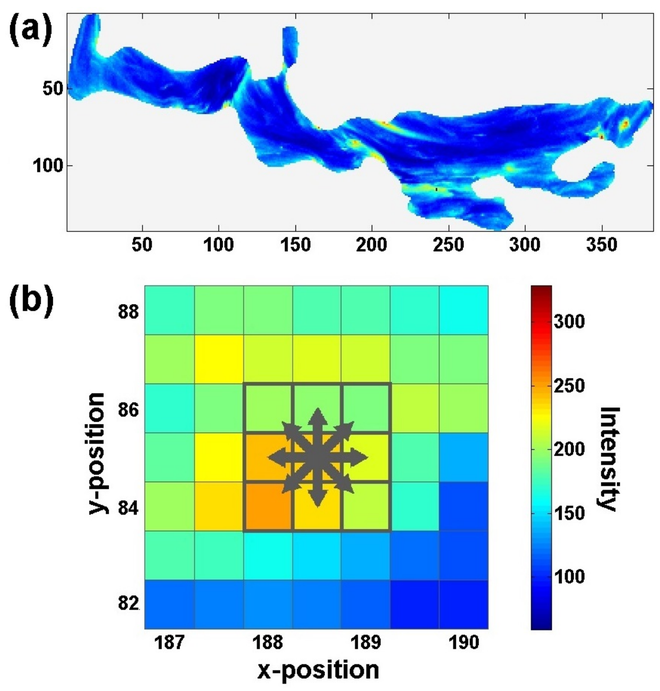

Figure 1: (a) FC from 2011−05−03 in AIA 193Å. (b) Spatial relationship of neighbouring pixels.

Textural Features

The probabilities of co-occurrences of pixel values can be determined with the co-occurrence matrix $p(i, j)$.

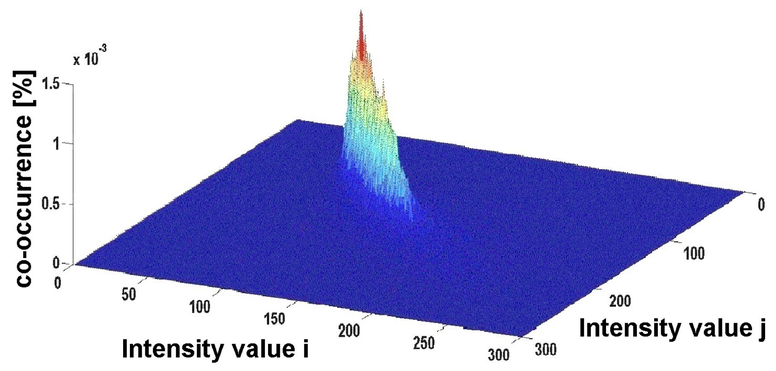

Figure 2: Co-occurrence matrix $p(i, j)$ for the FC shown in Fig.1. $p(i, j)$ represents the probability of neighbouring occurrences of pixel value i with pixel value j.

We adapted the proposed textural features from Haralick to open value ranges. This enables us to characterize magnetic field configuration with textural image features. We used 13 equations which define a set of textural features. For illustrative propose, we want to mention two textural features:

- Energy: $f_1 = \sum_i \sum_j p(i, j)^2$
- Contrast: $f_3 = \sum_i \sum_j (i - j)^2 p(i, j)$

The f_1 is a measure of homogeneity of the pixel pair distribution. It is high for a small number of different pixel pair combinations and low for a large number of different combinations.

The contrast feature f_2 is a measure of the amount of local variations presented in an image. It is high for large differences of pixel pair values and low for small differences.

First Results - Multivariate Analysis

Support Vector Machine (SVM), Random Forest Classifier, and Decision Trees algorithms were tested in order to differentiate between CHs and FCs. The set of first order as well as second order statistical parameter calculated from AIA and HMI images, together with shape measures from binary images, were used as attributes of CHs and FCs.

(a)

Predicted	Observed: Coronal Hole (CH)	Filament Channel (FC)
CH	True Positive (TP)	False Positive (FP)
FC	False Negative (FN)	True Negative (TN)

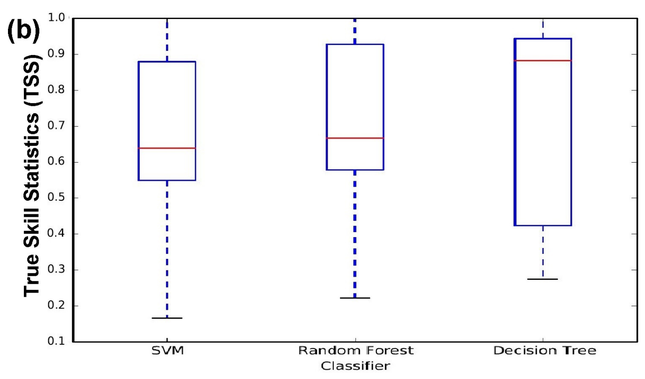

Figure 3: (a) Confusion Matrix, (b) Box plot for the 3 algorithms, computed over the 10-fold cross validation. The red line is the median value, the box gives the inter-quartile interval, and the whiskers goes up to the min and max value.

The original data set was split into two: the first subset, called *development set* was used to find optimal values for the hyperparameters of the three algorithms. The second subset, was used for *training and validation* as follows: 90% of the data are used for training the model and estimating the boundary decision, and the remaining 10% were used to measure the accuracy of the model. Such training and validation is made 10 times doing 10-fold cross-validation, that is, taking for the first run the first 10% for the validation set, then for the second run the following 10%, and so on. This allows us to estimate the confusion matrix for this data set, as is shown in Fig. 3.a. The True Skill Statistics (TSS) is a performance measure going [-1,1] that is independent of the proportion of CH and FC in the data set. A TSS of 0 indicates that the algorithm cannot distinguish between CH and FC. TSS is defined as the proportion of correctly predicted CH minus the proportion of wrongly predicted CH:

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN}$$

Fig. 3.b shows the box plot for the TSS for the three algorithms tested. The median TSS for the Decision tree is the highest, but the TSS performance of this algorithm varies the most. The SVM and Random Forest Classifier on the other hand have a lower median TSS, but their inter-quartile range is also lower.

Conclusion & Outlook

This study presents a new approach for the differentiation of CHs and FCs. First results reveal that the proposed methods decrease CH classification errors. In a next step we want to implement the developed algorithms in form of an automated real time decision tool that avoids wrong classifications of CHs as FCs in our forecast algorithm.

References

[1] Rotter et al. *Solar Phys.* **281**, 793-813, 2012
[2] Reiss et al. *CEAB*, arXiv:1408.2777, 2014
[3] Haralick et al. *IEEE Trans.Syst.* **3**, 610-621, 1973
[4] Duda et al. *Pattern Classification*, 2000