



Ana Maria Aires Carreira

Licenciada em Ciências de Engenharia Eletrotécnica
e de Computadores

Detection and Tracking of Coronal Holes in Solar Images

Dissertação para obtenção do Grau de Mestre em
Engenharia Eletrotécnica e de Computadores

Orientador: Professor Doutor André Teixeira Bento Damas Mora,
Professor Auxiliar, FCT-UNL

Júri:

Presidente: Professor Doutor José António Barata de Oliveira,
Professor Associado, FCT-UNL

Arguentes: Professor Doutor José Manuel Matos Ribeiro da Fonseca,
Professor Associado, FCT-UNL

Vogais: Professor Doutor André Teixeira Bento Damas Mora,
Professor Auxiliar, FCT-UNL



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março, 2020

Detection and Tracking of Coronal Holes in Solar Images

Copyright © Ana Maria Aires Carreira, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa. A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

To my parents.

Acknowledgements

First and foremost, I have to thank my dissertation supervisor, Dr. André Mora. Without his assistance, encouragement and commitment in every step of this work, this dissertation would have never been accomplished. It is not possible to put into words my gratitude for all the patience, effort and dedication that Dr. André Damas Mora provided me. Thank you for the dedication and knowledge that was shared with me.

I also have to thank all members of the CA3 research group, who among both serious and more relaxed moments, always offered their support and aid into finding solutions to the obstacles that were presented to me during the realization of this work. Thank you all for the aid you so kindly offered me.

I am grateful to Dr. Ivan Dorotovič from Slovak Central Observatory in Hurbanovo, Slovakia, for taking some of his time to provide me with knowledge and feedback when doubts of what steps to take emerged.

I would like to thank *Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa* and DEE for everything that I learned, academically and otherwise, and that has shaped me into becoming what I am today.

Lastly, I have to give thanks to the two people who have been my supervisors of life: my parents. None of what I have achieved today would have been possible without their determination and sacrifices they willingly put themselves through to push me another step forward. Mother and Father, I will be eternally thankful to you both.

Abstract

The study of solar activity and its effects on space weather is of great interest to humankind. Whether to study the dynamic of the star itself or the resulting phenomena and associated consequences from it, every different feature of the Sun provides valuable data to perform these studies. Features of the Sun are, for the most part, studied individually. However, studying different events collectively may result in new conclusions and findings that can be of as much interest as the individual studies.

The objectives for this dissertation is to complement a Coronal Bright Points (CBPs) tracking algorithm, previously developed by (Pires, 2018), with an additional feature: detection of Coronal Holes (CHs) and classification of CBPs regarding whether they are located inside or outside of CHs.

The proposed methodology is fully performed in Python language. Different image processing operations are applied in order to obtain a good detection result. The pre-processing stage involves an automatic image intensity normalization. The CHs detection uses a simple blurring before a fixed-value threshold segmentation. A last post-processing step includes performing adjustments to the detection results, using a closing morphologic operator, filling holes and an object detection.

The data gathered by both tools is at the end consolidated, so that a result on the classification of each CBP is obtained and lastly added to the database.

Keywords: Coronal Holes, Coronal Bright Points, threshold segmentation, image processing, solar images, solar features, space weather

Resumo

O estudo da actividade solar e dos seus efeitos na meteorologia espacial é de grande interesse para a humanidade. Seja para estudar a dinâmica da estrela em si ou os fenómenos resultantes e respectivas consequências, cada característica do Sol providencia dados importantes para efetuar estes estudos. As características do Sol são, na sua maioria, estudados individualmente. No entanto, o estudo de diferentes eventos na coletiva podem levar a novas conclusões e descobertas que podem ser de igual interesse aos estudos individuais.

O objectivo estipulado para esta dissertação é complementar um programa de seguimento de *Coronal Bright Points (CBPs)*, previamente desenvolvido por (Pires, 2018), com uma componente adicional: detecção de *Coronal Holes* e classificação de *Coronal Bright Points* referente a se estão dentro ou fora de *Coronal Holes*.

A metodologia proposta foi plenamente realizada em linguagem Python. Diferentes operações de processamento de imagem são aplicadas de modo a obter um bom resultado na detecção. A fase de pré-processamento envolve normalização automática de intensidade de imagem. A detecção de CHs usa um filtro de média simples antes de uma segmentação por *threshold* de valor fixo. Uma última etapa de pós-processamento inclui a realização de ajustes nos resultados da detecção, nomeadamente, aplicando uma operação morfológica de fecho, preenchendo buracos e uma detecção de objetos.

Os dados obtidos pelas ferramentas de detecção de CBPs e CHs são no final consolidados, para que seja obtido um resultado de classificação de cada CBP como estando dentro ou fora de um CH, e finalmente guardar essa informação na base de dados do CBPTracker.

Palavras-chave: Buracos Coronais, Pontos Brilhantes Coronais, segmentação por *threshold*, processamento de imagem, imagens solares, características solares, meteorologia espacial

Table of Contents

1. INTRODUCTION	1
1.1. SCOPE AND MOTIVATION	1
1.2. THE SUN	3
1.2.1. SOLAR STRUCTURE.....	3
1.2.1.1. INTERNAL STRUCTURE	4
1.2.1.2. SOLAR ATMOSPHERE.....	5
1.2.2. SOLAR FEATURES	5
1.2.3. SOLAR ACTIVITY	6
1.2.4. THE SOLAR CYCLE.....	6
1.3. RESEARCH PROBLEM AND CONTRIBUTIONS	7
1.4. DISSERTATION OUTLINE	7
2. STATE-OF-THE-ART	9
2.1. SOLAR DATA-ANALYSIS SOFTWARE.....	9
2.1.1. SUNPY	9
2.1.2. ASTROPY	10
2.2. CBPTRACKER - "A WEB TOOL TO DETECT AND TRACK SOLAR FEATURES FROM SDO IMAGES"	10
2.3. CORONAL HOLES DETECTION ALGORITHMS	13

2.3.1.	THE SPOCA-SUITE SOFTWARE	13
2.3.2.	A MODIFIED SHIFTED MEANS BASED SEGMENTATION APPROACH.....	14
2.4.	WEBSITES.....	15
2.4.1.	SPACE WEATHER LIVE	15
2.4.2.	HELIOVIEWER.....	16
2.5.	GENERAL CONCLUSIONS	17
3.	IMPLEMENTATION.....	19
3.1.	FITS IMAGE DATASET	19
3.2.	SYSTEM REQUIREMENTS:.....	20
3.2.1.	PACKAGES	20
3.2.2.	IMAGE PROCESSING MODULES.....	21
3.3.	PROPOSED ARCHITECTURE FOR CORONAL HOLE DETECTION	21
3.3.1.	DIRECTORIES	22
3.4.	IMAGE RETRIEVAL	23
3.4.1.	PRE-PROCESSING	24
3.4.1.1.	IMAGE PREPARATION.....	24
3.4.1.2.	INTENSITY NORMALIZATION.....	25
3.5.	CHS DETECTION	30
3.6.	POST-PROCESSING	34
3.7.	INTEGRATION WITH THE CORONAL BRIGHT POINT TRACKER.....	38
4.	RESULTS ASSESSMENT.....	41
4.1.	SMALL SOLAR ACTIVITY	42
4.2.	MEDIUM SOLAR ACTIVITY	45
4.3.	HIGH SOLAR ACTIVITY	47
4.4.	INTEGRATION WITH CORONAL BRIGHT POINT TRACKER.....	49
4.4.1.	LOW SOLAR ACTIVITY	50
4.4.2.	MEDIUM SOLAR ACTIVITY	52
4.4.3.	HIGH SOLAR ACTIVITY	54
5.	CONCLUSIONS AND FUTURE WORK.....	57

5.1.	FUTURE WORKS	57
6.	REFERENCES.....	59

Table of Figures

FIGURE 1.1 - LEFT: FULL DISK SDO/AIA IMAGE IN THE FE XII 193 Å CHANNEL DURING THE MINIMUM OF SOLAR CYCLE 24 ENTIRELY DOMINATED BY THE QUIET SUN AND CORONAL HOLES. A WHITE SQUARE OUTLINES A FIELD-OF-VIEW THAT CONTAINS TWO BRIGHT POINTS. RIGHT: ENLARGED FIELD-OF-VIEW OF THE OUTLINED REGION IN THE LEFT PANEL IMAGE (SIZE OF 110" × 100") (MADJARSKA, 2019)	2
FIGURE 1.2 - SDO/AIA 193 CHANNEL IMAGE WITH THREE CORONAL HOLES (OBTAINED FROM JHELIOVIEWER)....	2
FIGURE 1.3 - "THE SOLAR ONION", FROM (VITA-FINZI, 2008).....	4
FIGURE 2.1 - EXAMPLE OF A PLOT OF AN AIAMAP FROM SUNPY (SUNPY COMMUNITY <i>ET AL.</i> , 2015).....	10
FIGURE 2.2 - CIRCULAR DIAGRAM ILLUSTRATING THE PROPOSED TOOL STEPS SEQUENCE – ADAPTED FROM (PIRES, 2018)	11
FIGURE 2.3 – CORONAL BRIGHT POINTS DETECTION IN A SINGLE SDO/AIA IMAGE.....	12
FIGURE 2.4-“ WORKFLOW OF THE PROPOSED METHOD”, FROM (SURESH, DUBE AND GLENN, 2012)	14
FIGURE 2.5 - SPACEWEATHERLIVE WEBSITE.....	16
FIGURE 2.6 - HELIOVIEWER WEB TOOL	17
FIGURE 3.1 - CONTRIBUTIONS TO PREVIOUS CBPTRACKER ALGORITHM – CBPTRACKER IN LIGHT BLUE BLOCKS; PROPOSED ADD-ON IN DARK BLUE BLOCKS.....	22
FIGURE 3.2 -DIRECTORY STRUCTURE	23
FIGURE 3.3: THREE EXAMPLE SDO IMAGES WITH DIFFERENT WAVELENGTHS: A) AIA 193; B) AIA 171; C) AIA 211	24
FIGURE 3.4: A) DOWNLOADED IMAGE THROUGH THE PROGRAM; B) SAME IMAGE, TAKEN MANUALLY FROM SDO....	25
FIGURE 3.5: SOLAR IMAGE INTENSITY NORMALIZATION: A) NO MUL_VALUE; B) MULVALUE = 10.....	26
FIGURE 3.6: IMAGES WITH THE SAME MUL_VALUE APPLIED: A) YEAR 2014 (ACTIVE SUN); B) YEAR 2016; C) YEAR 2018 (QUIET SUN).....	26
FIGURE 3.7: HISTOGRAM OF AN IMAGE WITH MUL_VALUE = 10 APPLIED	27
FIGURE 3.8: MEAN AVERAGE TRENDING LINE.....	29
FIGURE 3.9 - AIA 193 IMAGES FOR THREE SOLAR ACTIVITY PERIODS: A) HIGH ACTIVITY (28-10-2014); B) DATE: 28-10-2016; C) DATE: 28-10-2018.....	29
FIGURE 3.10: SOLAR DISK RADIUS DIFFERENCE. A) IMAGE CAPTURED ON 01-01-2018; B) IMAGE CAPTURED ON 30-06-2018	31

FIGURE 3.11 - IMAGE TAKEN MANUALLY FROM SDO (COURTESY OF SDO)	32
FIGURE 3.12 – NEGATIVE DETECTION MASKS: A) THRESHOLD = 20; B) THRESHOLD = 30, C) THRESHOLD = 40	33
FIGURE 3.13: EFFECT OF BLURRING THRESHOLD PROCESS: A) C) WITHOUT BLURR; B) D) WITH 21X21 MEAN FILTER	34
FIGURE 3.14: CLOSING OPERATION: A) C) WITHOUT CLOSING; B) D) WITH CLOSING	35
FIGURE 3.15: OBJECT DETECTION AND DELETION: A) C) WITHOUT FILTER; B) D) WITH MINIMUM SIZE FILTER	36
FIGURE 3.16: OBJECT DETECTION CORRECTION MASK	37
FIGURE 3.17: A) C) DETECTION BEFORE OBJECT DELETION; B) D) DETECTION AFTER OBJECT DELETION	38
FIGURE 4.1 - THREE CHOSEN CAPTURES FROM YEAR 2018: A) JANUARY 1ST; B) MAY 1ST; C) NOVEMBER 1ST	42
FIGURE 4.2 - DETECTION RESULTS: A), C) AND E) SPACE WEATHER LIVE; B), D) AND F) PROPOSED TOOL	43
FIGURE 4.3 - DETECTION RESULTS: A), C) AND E) SPoCA; B), D) AND F) PROPOSED TOOL (PLEASE PAY ATTENTION THAT THE IMAGE HAVE SLIGHTLY DIFFERENT RESOLUTIONS – SOLAR RADIUS IS BIGGER IN HELIOVIEWER)....	44
FIGURE 4.4: THREE CHOSEN CAPTURES FROM YEAR 2016: A) JANUARY 1ST; B) MAY 1ST; C) NOVEMBER 1ST.....	45
FIGURE 4.5: DETECTION RESULTS: A), C) AND E) SPoCA; B), D) AND F) PROPOSED TOOL (PLEASE PAY ATTENTION THAT THE IMAGE HAVE SLIGHTLY DIFFERENT RESOLUTIONS – SOLAR RADIUS IS BIGGER IN HELIOVIEWER)....	46
FIGURE 4.6: THREE CHOSEN CAPTURES FROM YEARS 2010 AND 2011: A) NOVEMBER 1ST, 2010; B) DECEMBER 1ST, 2010; C) JANUARY 1ST, 2011	47
FIGURE 4.7: DETECTION RESULTS: A), C) AND E) SPoCA; B), D) AND F) PROPOSED TOOL (PLEASE PAY ATTENTION THAT THE IMAGE HAVE SLIGHTLY DIFFERENT RESOLUTIONS – SOLAR RADIUS IS BIGGER IN HELIOVIEWER)....	48
FIGURE 4.8 - A), C) AND E) SDO IMAGES; B), D) AND F) RESPECTIVE IMAGES WITH CORONAL HOLE DETECTION AND CORONAL BRIGHT POINT CLASSIFICATION	51
FIGURE 4.9 -A), C) AND E) SDO IMAGES; B), D) AND F) RESPECTIVE IMAGES WITH CORONAL HOLE DETECTION AND CORONAL BRIGHT POINT CLASSIFICATION	53
FIGURE 4.10 –A), C) AND E) SDO IMAGES; B), D) AND F) RESPECTIVE IMAGES WITH CORONAL HOLE DETECTION AND CORONAL BRIGHT POINT CLASSIFICATION	55

Table of Tables

TABLE 3.1 – SOFTWARE APPLICATIONS AND THEIR SPECIFIC VERSIONS 20

TABLE 3.2 – ADEQUATE MUL_VALUE PER YEAR, OBTAINED EMPIRICALLY 28

TABLE 3.3 – CALCULATION OF WEIGHTED AVERAGE..... 28

TABLE 3.4 – SOLAR DISKS RADIUS AND DIFFERENCE..... 32

Acronyms

AIA – Atmospheric Image Assembly

AR – Active Region

CBP – Coronal Bright Point

CH – Coronal Hole

EUV – Extreme Ultraviolet

FITS – Flexible Image Transport System

GPL – Gradient Path Labelling

IDL – Interface Definition Language

JPEG – Joint Photographic Experts

NASA – National Aeronautics and Space Administration

PNG – Portable Network Graphics

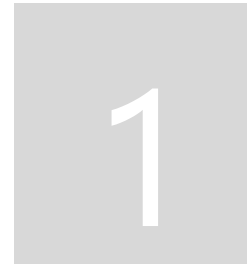
RGB – Red, Green and Blue

SDO – Solar Dynamic Observatory

SPoCA – Spatial Possibilistic Clustering Algorithm

UV – Ultraviolet

VSO – Virtual Solar Observatory



1. Introduction

The Sun not only has great influence on Earth's weather and climate system (Hanslmeier, 2009), but its activity can also affect our planet's magnetic field to the point of causing serious disruptions in it (Pattichis *et al.*, 2015). As a consequence, grid power systems on Earth, radio communications, air and spacecrafts and orbiting satellites can all be negatively affected (Mobberley, 2008; Kempton, Schuh and Angryk, 2018).

That being said, the study, monitoring and prediction of the Sun's phenomena in real time is of utmost importance, for it gives us the possibility of acting beforehand and reducing damages significantly.

The Solar Dynamics Observatory (SDO) was the first mission launched by NASA in 2010 with the intent of gathering data about the Sun's features and phenomena (Kempton, Schuh and Angryk, 2018). This information is crucial for visualizing and analyzing the Sun's activities and consequential effects.

1.1. Scope and Motivation

Coronal Bright Points (CBPs) are features commonly used for the study of Solar rotation. With a simple structure and relatively short lifetimes (Madjarska, 2019), along with the fact that CBPs remain numerous regardless of the stages of the Solar Cycle, they reveal to be more reliable for measurements of solar rotation than the formerly used Sunspots (Coelho, 2017).

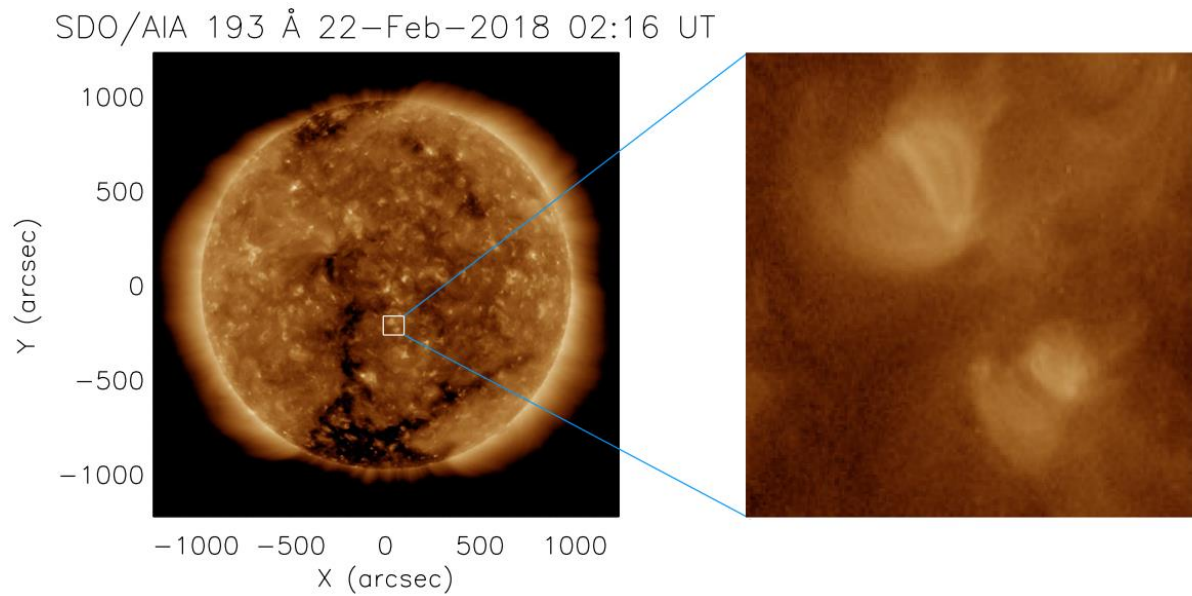


Figure 1.1 - Left: full disk SDO/AIA image in the Fe xii 193 Å channel during the minimum of solar cycle 24 entirely dominated by the quiet Sun and coronal holes. A white square outlines a field-of-view that contains two bright points. Right: enlarged field-of-view of the outlined region in the left panel image (size of 110'' × 100'') (Madjarska, 2019)

Coronal Holes, for their part, are observed as large and dark areas in the frequencies they are observable in, and highly associated with the variations of Solar Wind, which is the Solar activity that directly and most heavily affects Earth's climate, magnetic field and humankind.

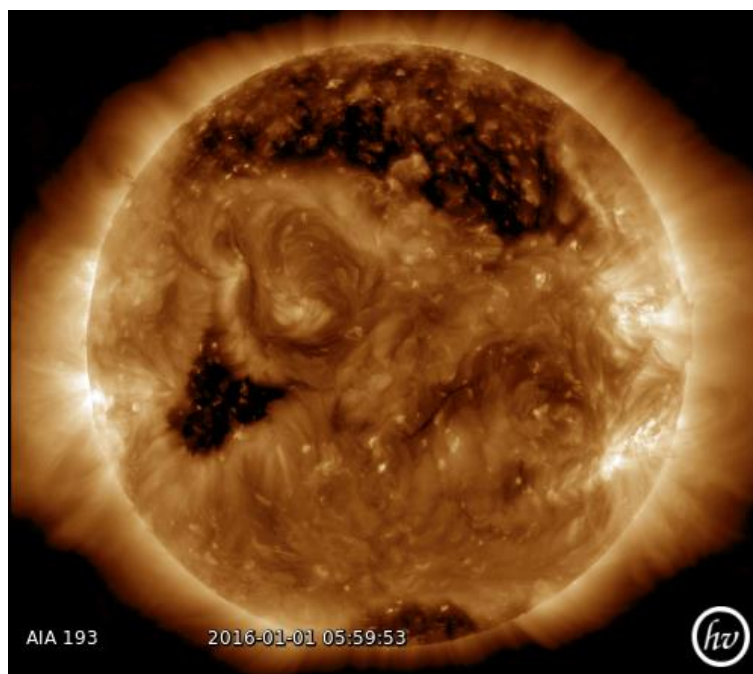


Figure 1.2 - SDO/AIA 193 channel image with three Coronal Holes (obtained from JHelioviewer)

While there are several studies for CBPs and CHs, from the most recent to the ones going as far back as the 1970s, these studies are mostly done on only one individual phenomenon rather than the two collectively. The only work found in this research that focused on an integration of that sort was (Karachik and Pevtsov, 2011), where it was specifically researched whether CPBs had any direct influence on Solar Wind speed or not.

The additional tool proposed in this dissertation would facilitate in obtaining data for more studies of the kind that might be performed in the future.

1.2. The Sun

The Sun is an essential source of energy, providing us with a vast amount of radiation, including light and warmth, which are crucial for sustaining life on Earth.

Located about 150 million kilometers away from Earth, and with a radius of about 7×10^5 km, it is the only star on which detailed observation of features and events on the surface is possible (Hanslmeier, 2009).

To better understand the influence the Sun has on our planet, it is relevant to study this star's structure and features, which will be the focus of this chapter. A simple description of the overall properties of the Sun is given in subsections 1.2.1 through 1.2.4.

1.2.1. Solar Structure

The Sun is composed of various different layers, as can be observed in Figure 1.3. These layers are grouped in two main sections:

- **The internal structure:** consists of the core, radiative zone and convection zone
- **The atmosphere:** composed of the photosphere, the chromosphere and the corona

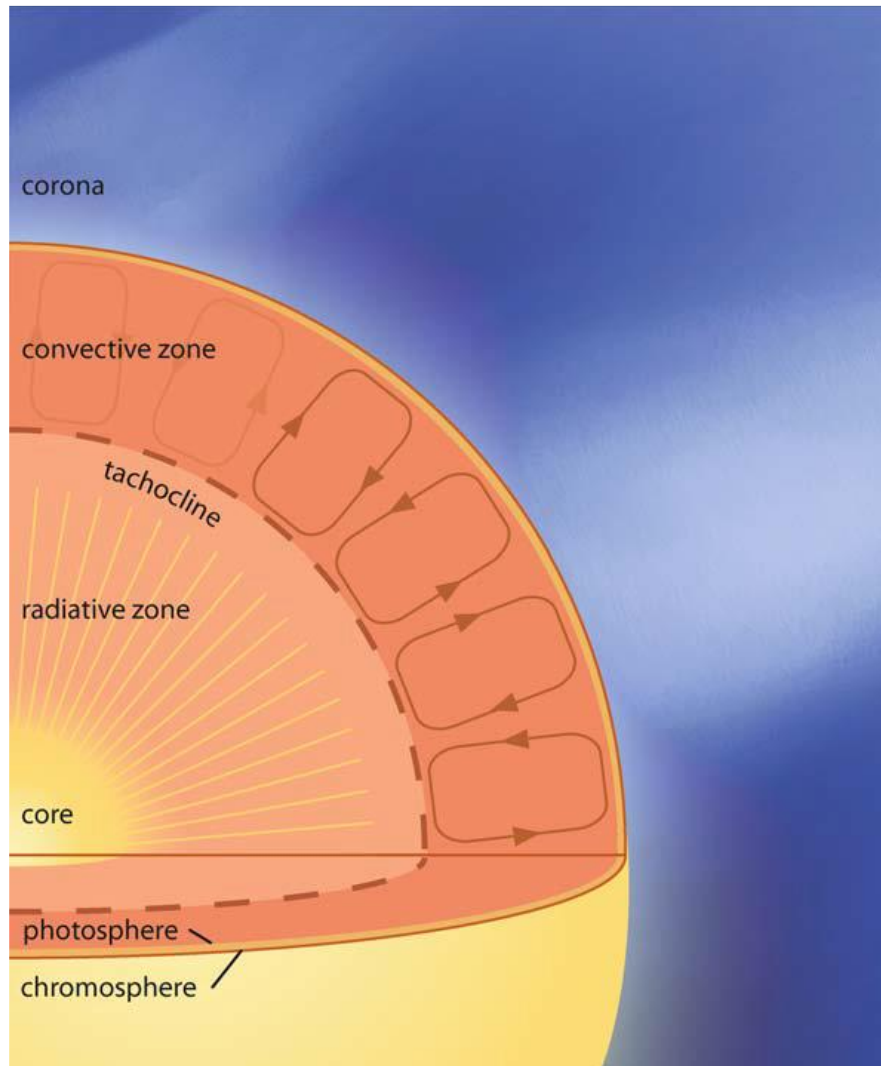


Figure 1.3 - "The Solar Onion", from (Vita-Finzi, 2008)

1.2.1.1. Internal Structure

The internal structure is divided in three layers: the core, the radiative zone and the convective zone. The extension of each of these three layers is around 1/3 of the total radius they form. (Hanslmeier, 2009)

The core is the most central zone in the Sun, where the temperature is about 1.4×10^7 K. It is the reactor of the star, for it is in there that the energy is initially produced, through the conversion of hydrogen into helium. Only a very small percentage of the resulting thermal radiation from this fusion manages to escape (Vita-Finzi, 2008).

The radiative zone is the layer that surrounds the core, where the density and temperature start to decrease. Due to deflections, absorption, re-radiation, among other phenomena, the radiation that emanates outwards from the core is slowed down on its journey, to the point where it can take millions of years to reach the Sun's surface (Saha, 2008).

The convective zone, starting at about 200,000 km beneath the surface, is where convection occurs: the difference in temperatures between the limits of the convective layer makes the more heated material rise outwards, release the accumulated energy, thus cooling down, and then sink back. This leads to the cellular-like pattern commonly observed in solar images, called granulation (Hanslmeier, 2009).

1.2.1.2. Solar Atmosphere

The Sun's atmosphere consists of three layers: photosphere, chromosphere and corona.

The photosphere is the first layer surrounding the convective zone of the sun. It extends to about 500 km, which, considering the large size of the sun, gives the illusion of being a very thin layer. The majority of radiation in the visible spectrum is originated from this layer, making it the deepest layer whose observation is possible.

The chromosphere is the layer on top of the photosphere, extending to around 2000 km. The strong emission of hydrogen in this layer is what causes it to be red rather than white like the photosphere. The temperature in the chromosphere rises when compared to the previous layers, going up to 10000 K.

Lastly, above the chromosphere is the corona. This layer extends deep into space, yet cannot be seen with the naked eye due to having lower light intensity than the photosphere, except in the occurrence of a total solar eclipse or with the aid of specific instruments for the purpose (for example, coronagraphs). The energy emitted from the corona is mainly in the UV and X-ray wavelengths.

1.2.2. Solar Features

It is important to understand the features the Sun has on its surface, as well as distinguish them. The three most important features for this thesis are described below.

- **Active Regions:** ARs are the areas on the Sun's surface with strong magnetic fields (Suresh, Dube and Glenn, 2012). When displayed in extreme ultra-violet images, these regions appear bright, contrary to the Coronal Holes. These regions are commonly observed containing sunspots, and are associated with events like flares and coronal mass ejections.
- **Coronal Holes:** CHs are regions of lower density and temperature, and have the lowest emission in UV and X-ray images, thus appearing as dark patches in said wavelengths. In a way, they are the opposite of Active Regions. They are the least active regions in the Sun's surface and are associated with the acceleration of the high-speed solar wind (Cranmer, 2009), which is influenced by their magnetic field.
- **Sunspots:** Sunspots are dark areas in the photosphere, due to their temperature being lower (about 2000 degrees) than their surroundings (Hanslmeier, 2009). Intense magnetic fields are the reason behind this thermal difference; they impede

the convective transport of hot plasma to the surface, thus keeping the areas they cover at cooler temperatures.

- **Coronal Bright Points:** CBPs are small, bright structures observed in EUV and X-ray frequencies. With a lifetime of about 8 hours and area of around $2 \times 10^8 \text{ km}^2$ (Coelho, 2017), they can be found in all latitudes of the solar corona, including the poles. CBPs. Are associated with bipolar magnetic features, and are also commonly used in studies that focus on the rotation of the Sun.

1.2.3. Solar Activity

It is not only relevant to know the features of the Sun, it is also important to know about the events associated with them. The most important ones are described below:

- **Solar Wind:** this event consists in a stream of electrically-charged particles that travel with enough speed to deviate themselves from the Sun's surface, even having the possibility of breaking out of its gravitational field. Because the solar wind is electrically conductive, the Sun's magnetic field lines are carried along with it (Mobberley, 2008), which can affect the Earth's own magnetic field.
- **Solar Flares:** flares are violent outbursts on the surface of the Sun, due to the release of a vast amount of energy in the period of a few minutes. They arise near active regions like sunspots.
- **Coronal Mass Ejections (CME):** although similar to flares, CMEs are much more massive bubbles of gas that get ejected from the Sun for several hours, produced by unstable magnetic field (Mobberley, 2008; Hanslmeier, 2009). These more dramatic events can cause damage to satellites and, if directed at Earth, can affect the power lines due to electrical currents induced in them by the electrically-charged particles that manage to break through Earth's magnetic field's protection. (Mobberley, 2008)

1.2.4. The Solar Cycle

The periodic cyclic behavior of the Sun's activity is referred to as Solar cycle, Schwabe cycle (name given after its discoverer, Schwabe) or Sunspot cycle (Hanslmeier, 2009). It is an 11-year cycle characterized by the fluctuations of the number and surface area of sunspots (Hathaway, 2015) and defined as occurring between two sunspot minima (Kane, 2008). Most solar phenomena, such as the total solar irradiance, the magnetic field, flares and coronal mass ejections, change in association with the cycle (Hathaway, 2015).

The most recent Solar cycle, labeled cycle 24, started in 2008 according to an issued consensus by a panel of international scientists (Kane, 2008) and is believed to end in 2019/2020.

1.3. Research Problem and Contributions

As a follow-up of previous work developed by (Pires, 2018), where the focus is on the detection and tracking of CBPs, this dissertation will focus on the creation of an algorithm to detect and track another feature: Coronal Holes.

The purpose for this work is to, at the end, blend the results with the data gathered by CBPs detection algorithm and take conclusions about whether a CBP is located inside or outside of a Coronal Hole.

The CBP tracking webtool saves all the data in a database. So needed details, such as the location and name of the file the Coronal Bright Point occurred, can easily be accessed by the program. The objective is to use that already stored data, consolidate it with the data from the proposed tool, obtain the result and add it to the database.

This thesis contributions are as follows:

- Image intensity normalization: appropriate brightness and contrast enhancement to ensure more accurate results
- Image pre-processing: filter application to aid in the exclusion of unnecessary details in the detection
- CHs detection through threshold segmentation and final adjustments to reduce inaccuracies
- Integration with the CBPTracker program from (Pires, 2018)

1.4. Dissertation Outline

This dissertation features 5 chapters that provide a more in depth description of the developed work. The first two chapters focus on the theory, current knowledge and existing works with either the same or similar objectives. The other three chapters address the proposed tool itself.

Chapter 1: Introduction hosts the preface of the work, including the motivations and objectives for the project, as well as a more detailed description of physical phenomena and consequences that further emphasize the motives for this work.

Chapter 2: State-of-the-Art reviews some already existing algorithms and their approach to the problem, both as a background research study and as tools that can aid in the validation of the results from this work.

Chapter 3: Implementation presents all the proposed methods, applied techniques, obstacles and respective solutions that were found along the development of the work. An initial review of the requirements for the work is firstly presented, followed by a detailed description of each stage of the work. The proposed methodology can be divided into three sections: image pre-processing, segmentation and post processing. The explication in each section is sustained with visual aid of data and results of each step for further insight.

Chapter 4: Results summarizes and displays the obtained results, their comparison with other authors' works and integration with the data from the CBP tracker.

Chapter 5: Conclusions and Future Work evaluates the results obtained and proposes improvements and topics for future developments, additions and optimizations.



2. State-of-the-Art

A review of state of the art technologies needs to be done before fully getting into the details of the proposed tool. Firstly, the existing software libraries for solar analysis are introduced. The CBP tracking tool proposed in (Pires, 2018), is also be presented since it will have to be integrated at the end. Details of the webtools architecture and procedures will be described, since the CHs detection tool proposed in this dissertation will be integrated in it.

Other related works on CHs detection that were considered to be relevant for this research will be depicted and explained. The main objective of this search and review is to gather information about what methods can be used and for comparison purposes.

Lastly, a few useful space weather related websites will also be presented, where coronal holes and other solar events are published periodically.

2.1. Solar data-analysis software

2.1.1. SunPy

SunPy (SunPy Community *et al.*, 2015) is a free and open-source, community-developed Python software package for solar data analysis. The aim of the SunPy project was to supply a package that, among other features, would be robust, easy to use and that conformed to modern engineering standards.

SunPy also employs packages Numpy, Matplotlib and Astropy, which makes this package highly compatible with most scientific projects developed in Python language.

An example of a very useful feature from SunPy is the creation of maps from SDO/AIA FITS files (Figure 2.1).

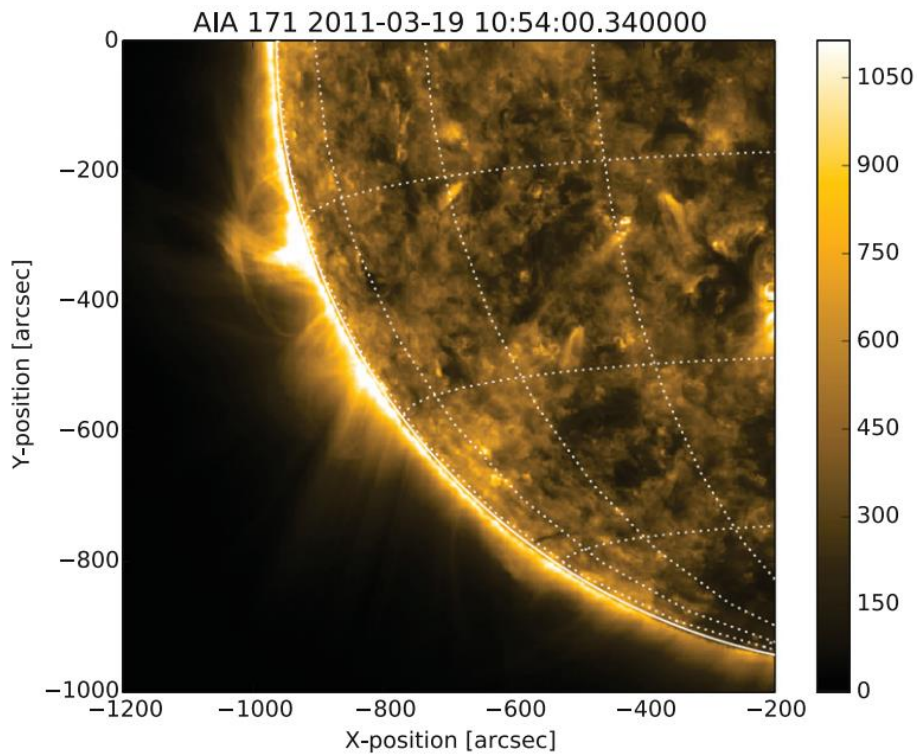


Figure 2.1 - Example of a plot of an AIAMap from SunPy (SunPy Community *et al.*, 2015)

2.1.2. Astropy

Just like SunPy, Astropy is also a community-developed package containing key functionalities and tools for astronomy and astrophysics studies in Python language.

The functionalities of this package are aimed at professionals in the field of astronomy and astrophysics, but can also be used in any astronomy related software development. Much like SunPy, Astropy is free and open-source.

Astropy is also compatible with other packages. It includes affiliated packages, i.e., astronomy-related python packages that, although not developed by the core source development team, share the same goals and resort to the package.

2.2. CBPTracker - "A web tool to detect and track Solar features from SDO images"

The CBPTracker proposed in (Pires, 2018) is a webtool with the objective of retrieving, pre-processing, detecting and tracking CBPs in solar images, in order to facilitate search and visualization of their solar information via database and for solar rotation profile estimation.

The technologies used in the creation of this webtool were SunPy, Astropy and OpenCV. Additionally, there was also used a Gradient Path Labelling by (Mora *et al.*, 2011) and MySQL to handle the results database.

The steps of the algorithm are presented in Figure 2.2.

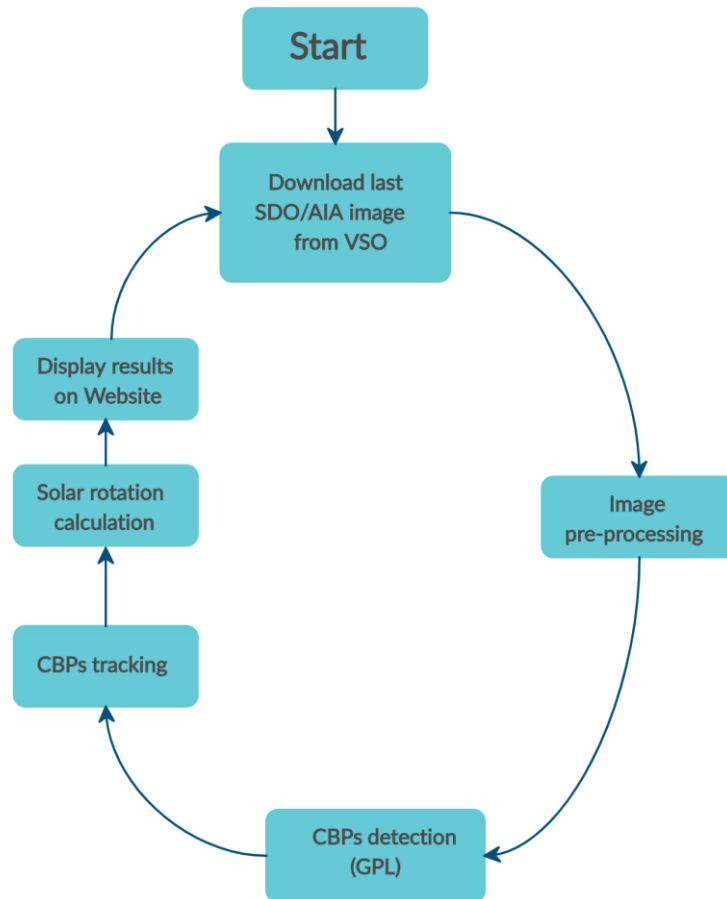


Figure 2.2 - Circular diagram illustrating the proposed tool steps sequence – adapted from (Pires, 2018)

The algorithm starts by downloading the most recent solar images from the Virtual Solar Observatory (VSO) and pre-processing them. These image come in lower resolution (LRi) rather than high (HRi) due to the fact that LRi has a much faster availability (about 1 hour after capture) than HRi (about 5 days after capture), along with LRi displaying faster results with the pre-processing stage.

The pre-processing starts by correcting the data for the effect of the instrument PSF function. Following that, the date is normalized to values between 0 and 1 through min-max method, and then resampled into 4096x4096 pixels size through bilinear interpolation.

After the pre-processing is done, Gradient Path Labelling (GPL) is used for the segmentation of the images. The method relies on image gradient as the basis for pixel labelling, grouping ascending paths that belong to the same regional maximum.

A post-processing algorithm is applied along with GPL to merge neighbour regions with similar amplitudes together. This is done by use of an active regions mask, and results in an image segmented in various regions based on intensity. Said regions are then filtered to match the intended solar features.

The images are split into 16 parts and the GPL is performed on each one separately for faster processing. The neighbor merging procedure is reapplied to correct over-segmentation from the GPL.

Lastly, CBPs close to the solar limb are discarded due to being more imprecise.

The result of the detection is presented in Figure 2.3.

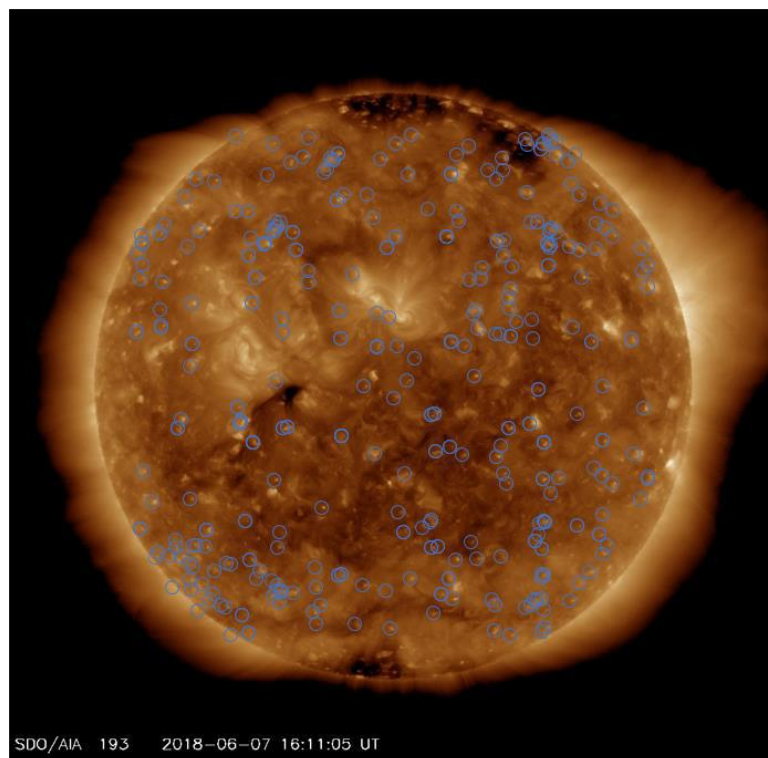


Figure 2.3 – Coronal Bright Points detection in a single SDO/AIA image

After the detection, the algorithm proceeds onto the tracking phase. First, the location of each detected CBP is given by its brightest point. This information is provided by the GPL. Next, the tracking of active CBPs is performed over the data of the database. For that, the location of the CBPs is converted, by means of a Sunpy method, from pixels to heliographic Stonyhurst system to ensure more accurate results. The distance between the newly detected CBP (dCBP) and the active ones already existing in the database (aCBP) is calculated. The time span between image captures is also taken into account in the process. A maximum distance is imposed to determine whether a dCBP and an aCBP are the same or not.

If a CBP is new, it gets added to the database. If it was an already active CBP, it gets its data updated.

After processing the image, the tool check through all the active CBPs to determine if any should be set as non-active. If the CBP has stopped being detected for 60 minutes, the tool classifies it as non-active.

With the tracking completed, the program uses the newly stored data to compute the Solar Rotation calculations.

The webtool of this work is available at the website of the Space-Planetary Interactions monitoring and forecasting Laboratory of the Geophysics and Astronomy Observatory, Coimbra, Portugal (SPINLab): http://www.mat.uc.pt/~obsv/SPINLab/SPINLab_solSDO.php

2.3. Coronal Holes detection algorithms

There are not as many algorithms for the detection of Coronal Holes as there are for Coronal Mass Ejections or Flares, and a quite a few of the ones found rely on manual annotation, which is not only extremely time consuming (thus making it inadequate for a project with the purpose of working in real time) but also difficult to quantify due to the observer's qualitative assessments of the images (Kirk *et al.*, 2009).

Nevertheless, there were a few approaches that showed potential as guides to the process of this development and that are described next.

2.3.1. The SPoCA-suite software

This software, last updated in 2018, is a set of algorithms developed to extract Coronal Holes, Active Regions and Quiet Sun on Extreme Ultra-Violet images provided by SDO (Delouille *et al.*, 2018).

"The SPoCA-suite implements three types of fuzzy clustering algorithms: the Fuzzy C-means (FCM), a regularized version of FCM called Possiblistic C-means (PCM) algorithm, and a Spatial Possiblistic Clustering Algorithm (SPoCA) that integrates neighboring intensity values." (Delouille *et al.*, 2018)

Before implementing the fuzzy clustering algorithms, a pre-processing has to be performed in order to reduce noise in the images. This is revealed to be especially important for the detection of Coronal Holes, for they are affected by Poisson noise.

The pre-processing consists on calibration of the images using solar software, followed by a conversion of Poisson noise into gaussian noise, and lastly a correction of limb brightening effect.

Once the pre-processing is done, the algorithm proceeds to the segmentation, having at its disposal a set of decision rules to obtain segmented maps.

Following the segmentation, the program performs the region extraction and post-processing, obtaining the classified regions.

After classification, the tracking is performed, with special attention to changes to the regions, such as splitting or disappearing.

2.3.2. A Modified Shifted Means Based Segmentation Approach

Suresh, Dube and Glenn, 2012, proposed a different approach, using spatial neighborhood information from images with different wavelengths, to reduce the time complexity of fuzzy logic based methods, and not lose accuracy in the process.

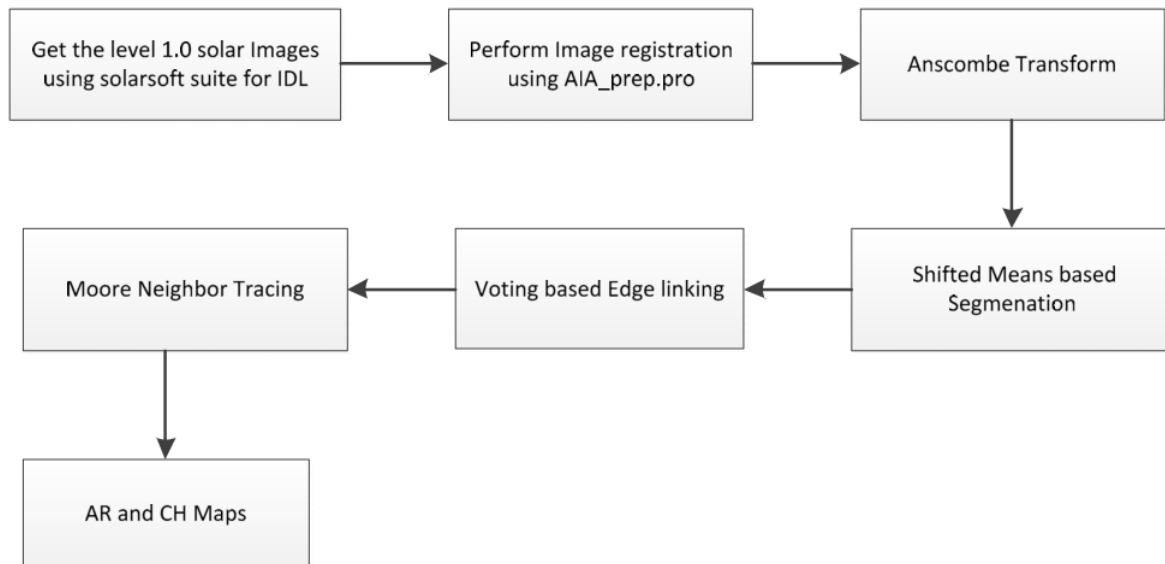


Figure 2.4- "Workflow of the proposed method", from (Suresh, Dube and Glenn, 2012)

The algorithms starts with the download of the level 1.0 images from one of the servers. The images are afterward registered using a common center and, with the aid of an IDL program, limb darkening corrections are performed.

Then, to clear the Poisson noise the images display, the Anscombe transform is applied, "which is equivalent to taking the square root of the intensity" (Suresh, Dube and Glenn, 2012). Only the solar disk is selected, to avoid classifications to the outer regions.

The mean shift segmentation algorithm is applied on the 19.3nm (X-ray wavelength) and 335nm (EUV wavelength) images, with the intent of separating the dark from the bright regions.

“Voting based edge linking” is performed to join any remaining fragments. (Suresh, Dube and Glenn, 2012) use “the Edge Detection and Image Segmentation system, a C++ source code implementation provided online which is fast as compared to other fuzzy segmentation schemes”.

Once the image is properly segmented into the classes, the boundaries of the regions are traced in separate, with the use of a Moore-Neighbor Tracing algorithm.

The regions that are less than .02 percent of the solar disc are discarded. (Suresh, Dube and Glenn, 2012)

2.4. Websites

A few existing websites that provide space weather information are listed in this section. While most are mainly focused in displaying the final results relative to the consequential Solar Wind rather than the Coronal Holes themselves, there are a couple that provide their own results of Coronal Holes detection.

2.4.1. Space Weather Live

The Space Weather Live website (<https://www.spaceweatherlive.com/en>) displays real-time results and data of both Auroral and Solar activity, as well as news from the missions and technologies at work. From current measures of the Solar Wind, Flares and CMEs to predictions of disturbance storms, this website provides a lot of information to the public.

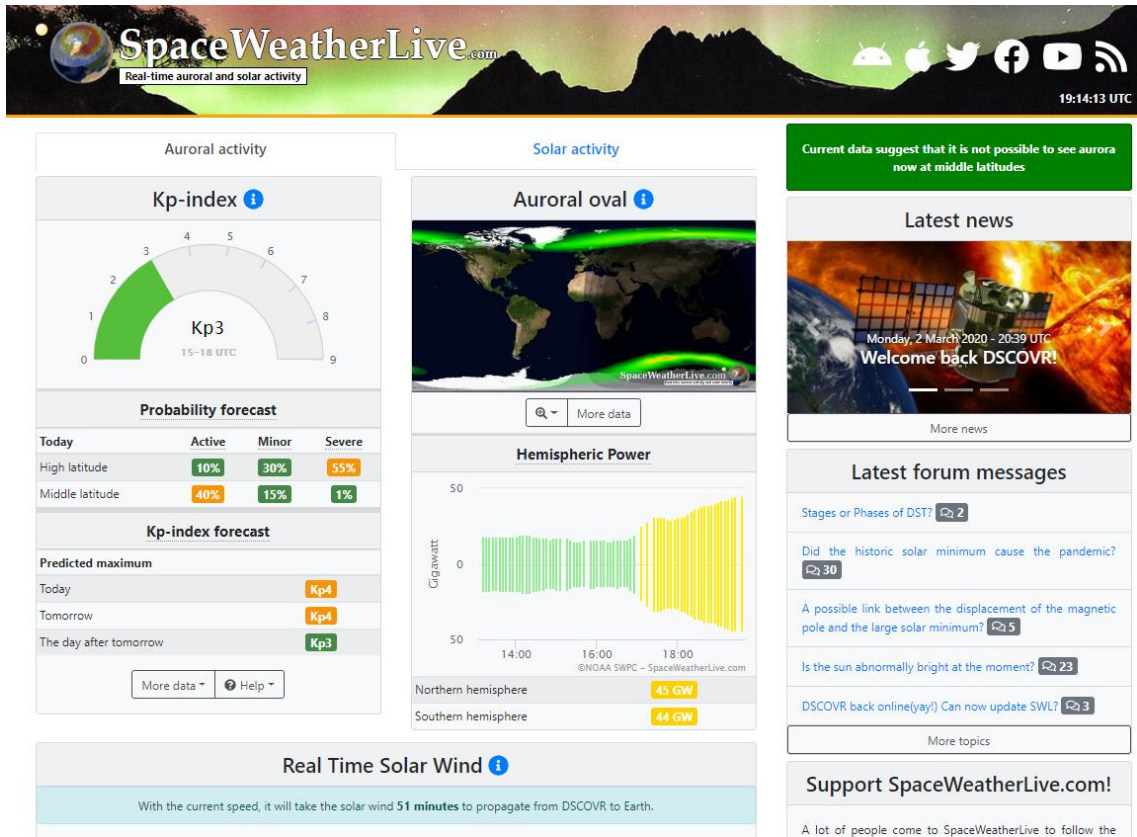


Figure 2.5 - SpaceWeatherLive website

2.4.2. Heliviewer

The Heliviewer project seeks to provide tools that can aid researchers perform their studies on the Sun in a more intuitive way. Their website (<https://heliviewer.org/>) is a web-based application where a lot of information is shared, including results from different works. It also provides the ability to overlay more than two images and experiment with different solutions, with the visual result changing in real time and allowing for more intuitive testing.

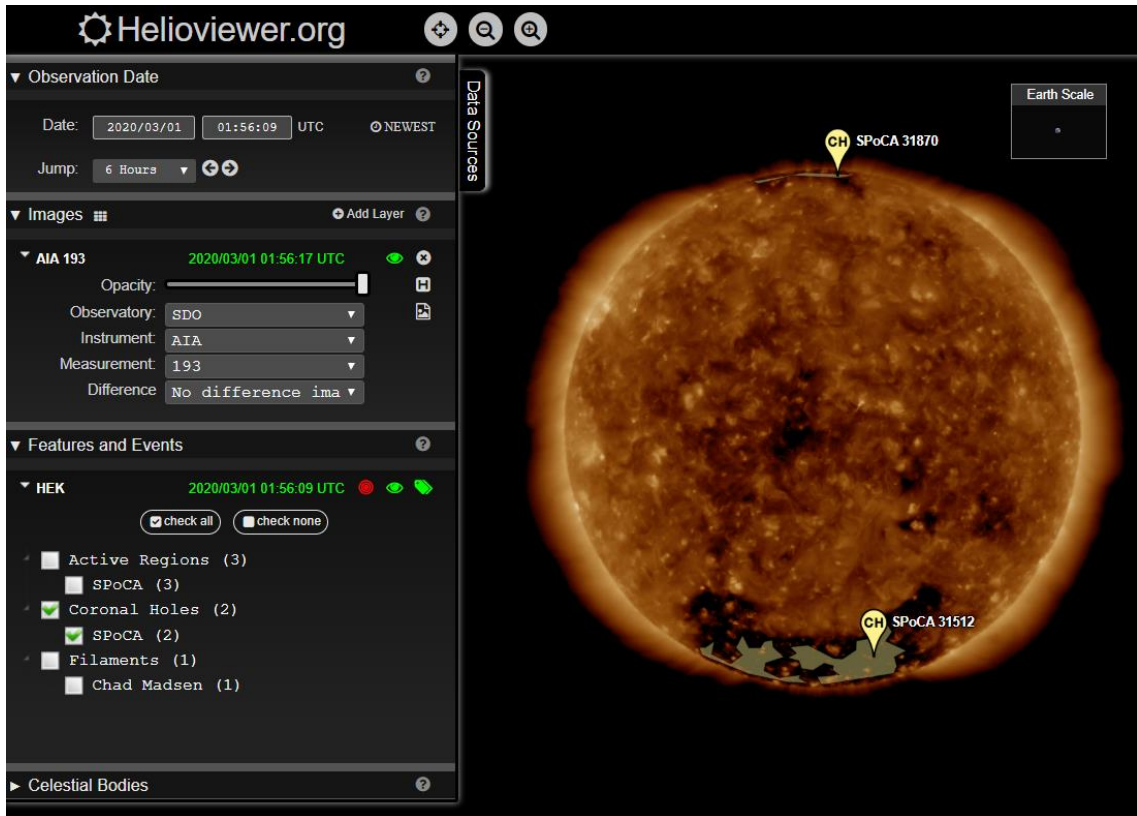


Figure 2.6 - Helioviewer web tool

2.5. General Conclusions

The analysis the approaches presented in this chapter gave an initial idea of what methods could be applied and a first notion of the structure that the program should take. Based on what was done in previous works, it is clear that the solution could reveal itself quite simple, or quite complicated.

For this work the CBP tracker will be the basis for this dissertation in order to keep the results of the two compatible.

3

3. Implementation

In this chapter the implementation procedures, thought processes, challenges and respective solutions are described. As the cornerstone of this dissertation, this section depicts the entire process in detail for easier understanding and acquainting the reader to the scientific and technical approach to the research problem and proposed solutions.

3.1. FITS Image Dataset

Flexible Image Transport System (FITS) is a file format widely used in astronomy related software. It is the standard format astronomers use to transport, analyze and archive scientific data files.

Originally, FITS in its most basic form was designed merely for the transfer of images, and was later expanded to accommodate other, more complex formats.

A FITS file is structured as follows, according to the latest standard document:

- Primary header and data unit (HDU)
- Conforming Extensions (optional)
- Other special records (optional, restricted).

In the development of this work, only the information from the primary HDU is needed, so the FITS files to be used are in the most basic form.

The primary HDU contains the primary header, which may or may not be followed by the primary data array. The primary header contains all the necessary keyword and respective information about the image.

3.2. System requirements:

A few requirements are needed for the correct response of the CHs detection application, both due to incompatibilities between technologies and upgrades in the coding program.

3.2.1. Packages

Throughout the development of the project, the programs and modules had various upgrades installed. Table 3.1 presents the versions of all technologies at the moment of completion of this project, which are functioning together correctly.

Table 3.1 – Software applications and their specific versions

	Version	Description
Python	3.6.7	Interpreted, object-oriented, high-level programming language used in the work
Anaconda	4.8.3	Python distribution platform
Spyder	4.0.1	Scientific Python Development Environment (included in the Anaconda platform)
SunPy	1.0.6	Software package for solar physics
Astropy	4.0	Core Python package for astronomy
OpenCV	3.4.1	Computer vision and machine learning software library
Numpy	17.3	Fundamental package for scientific computing in Python (required for use of OpenCV in Python)
mysql-connector-python	8.0.18	API that allows data stream compression between Python and the MySQL database server
GPL – Gradient Path Labelling	1.0	Segmentation algorithm that applies a labelling procedure to segment regions (Mora <i>et al.</i> , 2011)

3.2.2. Image Processing Modules

In Python there are two image processing modules that can be used to perform the intended image processes: Pillow and CV2.

Pillow is a fork from the Python Imaging Library (PIL). It's a library that offers image processing capabilities to Python interpreters and extensive support to a large number of formats. It is compatible with most GUI toolkits and interfaces as well.

CV2 is the OpenCV module for Python. When it comes to image processing, CV2 includes several algorithms such as linear and non-linear filtering, geometrical image transformations, color space conversion, histograms, and a lot more utilities.

During the creation of the program, both modules were tested to perform the required tasks. The Pillow module was dismissed for the following reasons:

- CV2 was the module used in the previous work
- The modules have low compatibility with each other and require conversion between their matrices, which is not practical
- It was observed during development that CV2 not only had better performance of the tasks than Pillow, it was also more time-efficient

3.3. Proposed Architecture for Coronal Hole Detection

The proposed CHs detection application is a continuous process that retrieves images, analyses them and stores the information on a database for being published online. This work is an add-on for the CBPTracker proposed by (Pires, 2018), being integrated on its processing pipeline. Figure 3.1 displays the update of the cyclic process presented in Figure 2.2, with the CBPTracker webtool represented in lighter blue blocks and the proposed add-on in darker blue ones.

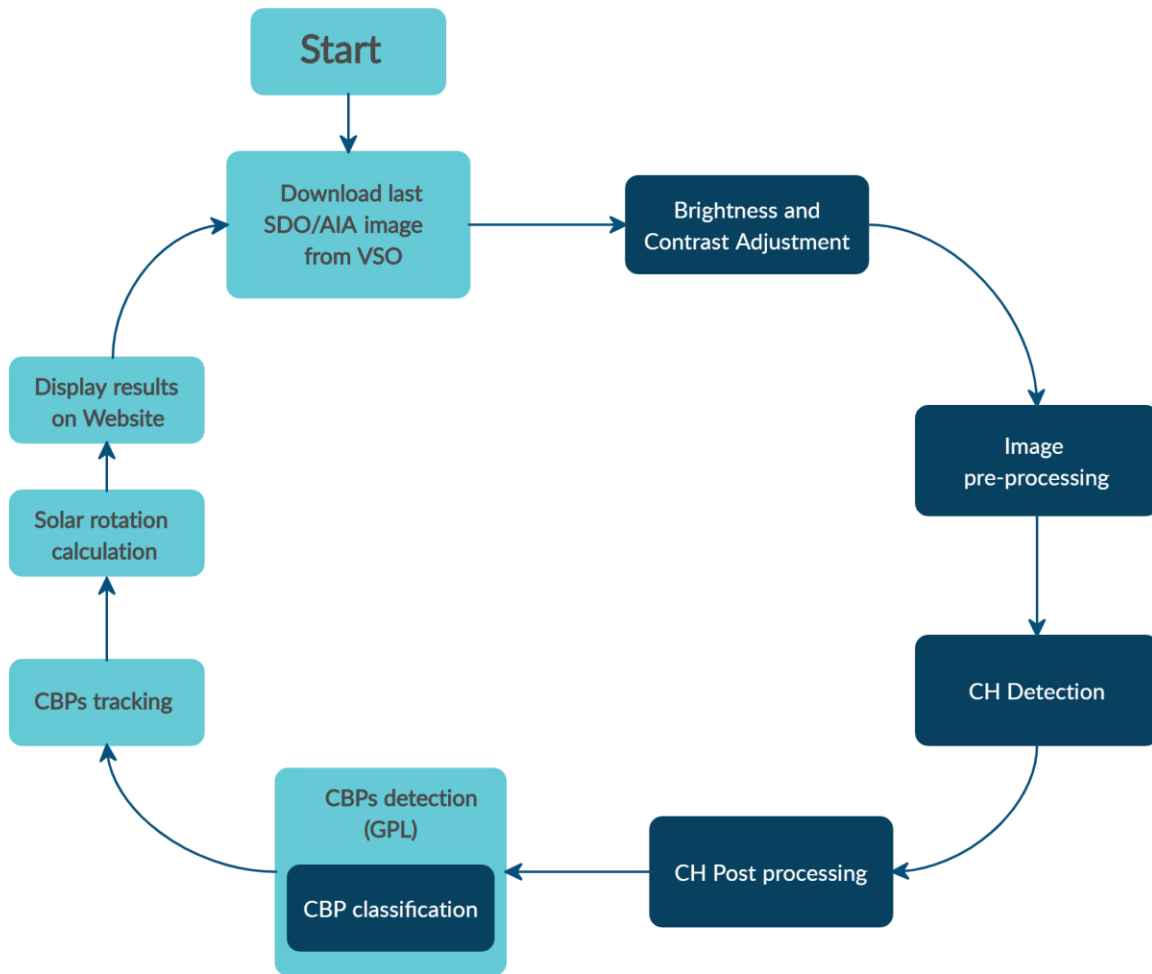


Figure 3.1 - Contributions to previous CBPTracker algorithm – CBPTracker in light blue blocks; proposed add-on in dark blue blocks

The created program is composed of four scripts:

- **CH_cons.py:** contains all the paths for directories and files needed in the program.
- **CH_functions.py:** holds all the functions created for the CH detection program. Each function performs an individual process.
- **CBPTracker_main.py:** contains the entire algorithm of the CBP tracker, now with the CH detection added. The functions from the CH_functions.py file are called to perform the CH detection, and the obtained results are afterwards incorporated with the data on CBPs to draw the final results.

3.3.1. Directories

To facilitate the access to all data and files in the procedure, this program is prepared to automatically handle all the needed directories. The hierarchy of the directories is displayed in Figure 3.2.

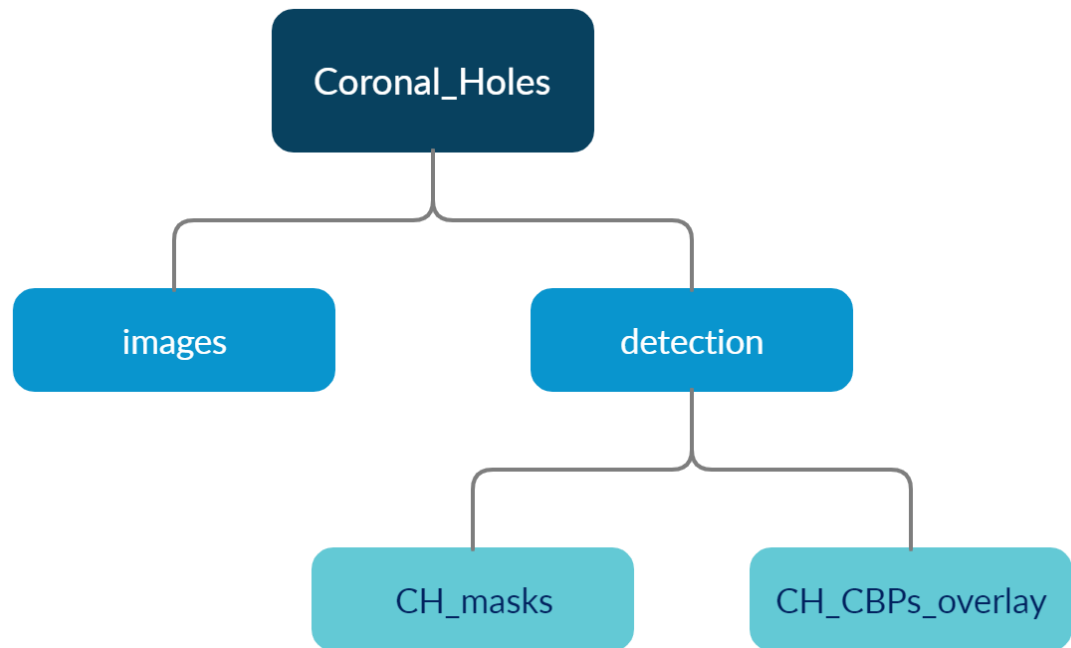


Figure 3.2 -Directory Structure

First, if no location is specified or it simply doesn't exist, the "Coronal_Holes" folder is created in the same directory the code files are placed in. If the user wants the program to create these files in a specific directory, all they have to do is, in the cons.py file, change the variable *path* to their preferred location, as is exemplified in Listing 3.1.

```

13 myPath = os.path.join("G:", os.path.sep, "University", "Thesis")
14 if os.path.isdir(myPath):
15     path = myPath
16 else:
17     path = os.getcwd()
  
```

Listing 3.1 - Specification of folders' location

If the folders do not exist already, the program creates them in the location specified in Listing 3.1, in the set hierarchical order. Note that, for simplicity in this process, it is only tested if the first folder, "Coronal_Holes", exists.

The advantage of this is that the user does not have to worry about looking through the code and creating all the folders accordingly. The program does that all by itself.

3.4. Image Retrieval

The first step of the algorithm is to retrieve the images from the Solar Dynamics Observatory. This process is performed by reusing the Python code previously developed in CBP Tracker.

A few changes had to be performed in the code, both to adapt it to the purpose of this work and to keep up with the updates in Python and FITS file handling.

In this stage, the first step is to specify the parameters for the images:

- Starting date and time
- Ending date and time
- Cadence between images (12 minutes)
- Wavelength (193 Angstrom)

These parameters are then provided to Sunpy's search tool, Fido, which then obtains the FITS files. It is important to note that the parameters are the same as the ones used in the CBP tracker. While the wavelength parameter does not obligatorily need to remain the same in the two processes, it was determined that the value used in CBP tracker tool was also the most adequate for the CHs detection due to the higher contrast in the CH areas when compared with other wavelengths (Figure 3.3).

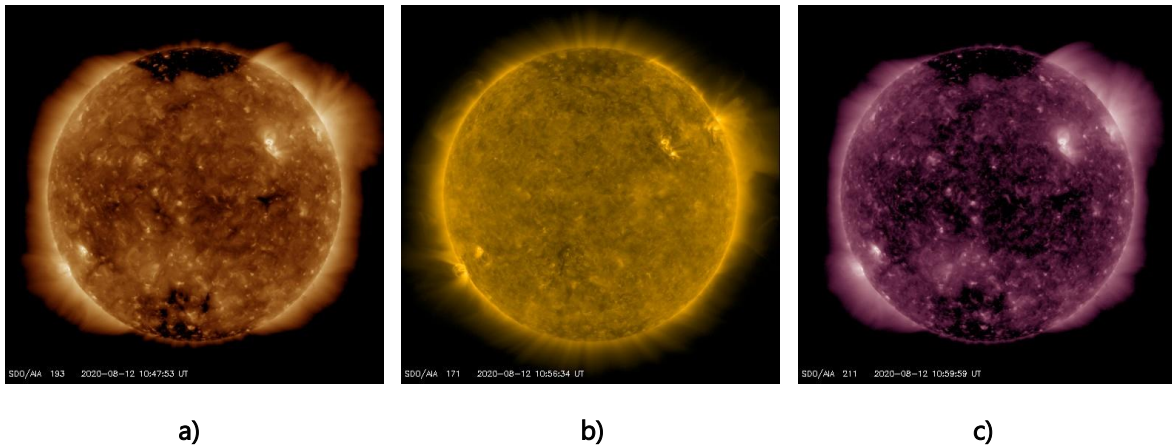


Figure 3.3: Three example SDO images with different wavelengths: a) AIA 193; b) AIA 171; c) AIA 211

3.4.1. Pre-processing

3.4.1.1. Image preparation

The image pre-processing remains practically the same as developed for CBPTracker. The program starts by correcting the raw images with the use of a PSF function, which had been provided at the beginning of the work. When it comes to the following intensity normalization, however, a change has been made to fit the current project. Instead of using a min-max method between 0 and 1, as was done in the CBP tracker, the normalization is done with a constant factor, because depending on the solar activity the solar events can be unnoticed or saturated.

Finally, the image is resized into 4096x4096 pixels and the program proceeds on to the next phase.

3.4.1.2. Intensity normalization

Running the code presented in the previous section will result in images as can be seen in Figure 3.4a). As can be observed, the resulting images are noticeably too dark and have too little contrast to be able to correctly differentiate between what is a Coronal Hole and what is not. However, as can be concluded by Figure 3.4b), the difference should be clear. Because of this issue, an image enhancement needs to be performed.

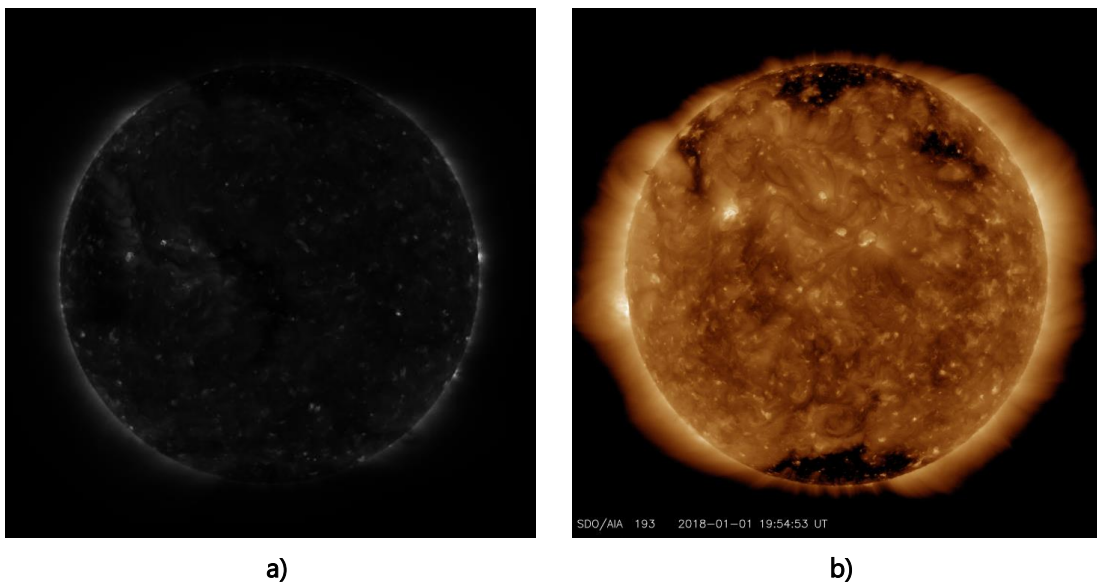


Figure 3.4: a) Downloaded image through the program; b) Same image, taken manually from SDO

The enhancement process consists in applying a multiplication factor, mul_value , to the intensity value. This enhances the contrast in the images, meaning that while bright colored pixels become even brighter, the darkest colored pixels remain dark. The result allows better performance for the CHs segmentation.

Initially it was assumed that the same mul_value would suffice for all images. By trial and error and observation of the resulting image, it was determined that a mul_value of value 10 was an adequate multiplication factor (Figure 3.5).

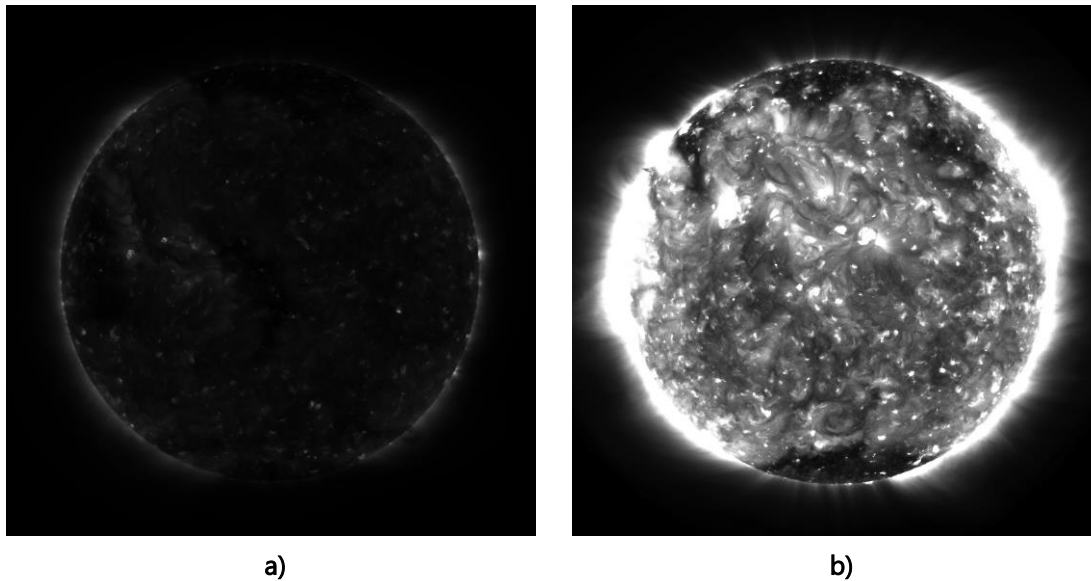


Figure 3.5: Solar image intensity normalization: a) no `mul_value`; b) `mulvalue = 10`

However, it was also observed that the stage of the Solar cycle also influences significantly the fixed multiplication factor results. Images taken during the higher activity years (Active Sun) were far brighter than the ones taken during the Quiet Sun, as can be observed in Figure 3.6. This difference was not trivial enough to be ignored, and thus a new solution had to be found to resolve this issue.

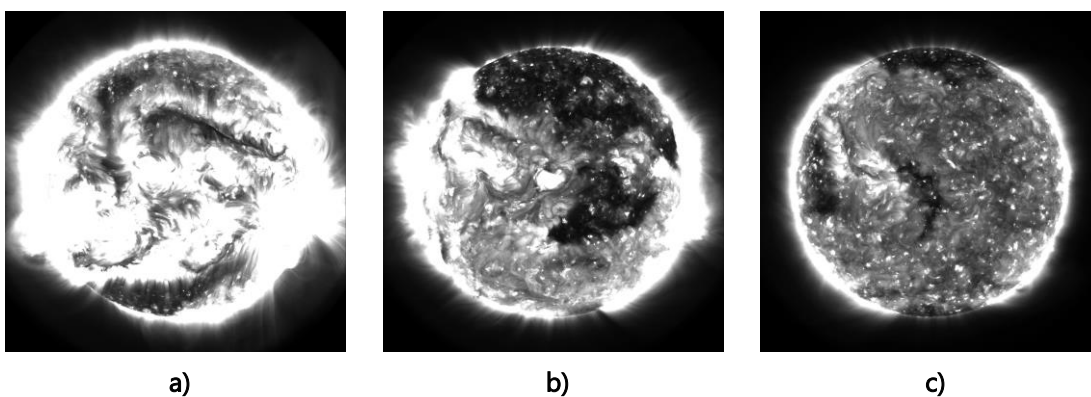


Figure 3.6: Images with the same `mul_value` applied: a) Year 2014 (Active Sun); b) Year 2016; c) Year 2018 (Quiet Sun)

A possible solution for this problem was to estimate the average intensity of the solar surface without disturbances (middle intensity pixels), by calculating the weighted average of the images through their histograms.

The weighted average of each set of values was determined by the multiplication of each bin by the respective number of occurrences of that bin. The sum of all these values is calculated and at the end divided by the total amount of accounted pixels, which, in this case, would simply be the image size, 4096 x 4096 pixels.

So the average intensity equation would be:

$$avg_w = \frac{\sum_{i=0}^{255} (i \times hist(i))}{4096^2} \quad (1)$$

However, by observation of the images, it was noted that due to the fact that a large amount of pixels were located outside of the solar disk, the result of the equation could be heavily influenced. The amount of darker pixels outside the solar disk was considerable enough to alter the precision of the result.

By analyzing the histogram of the images after the brightness adjustment was performed, a solution was proposed. The histograms, for their most part, had practically all the same form: a first histogram peak at the origin, followed by a second peak representing the middle intensity pixels.

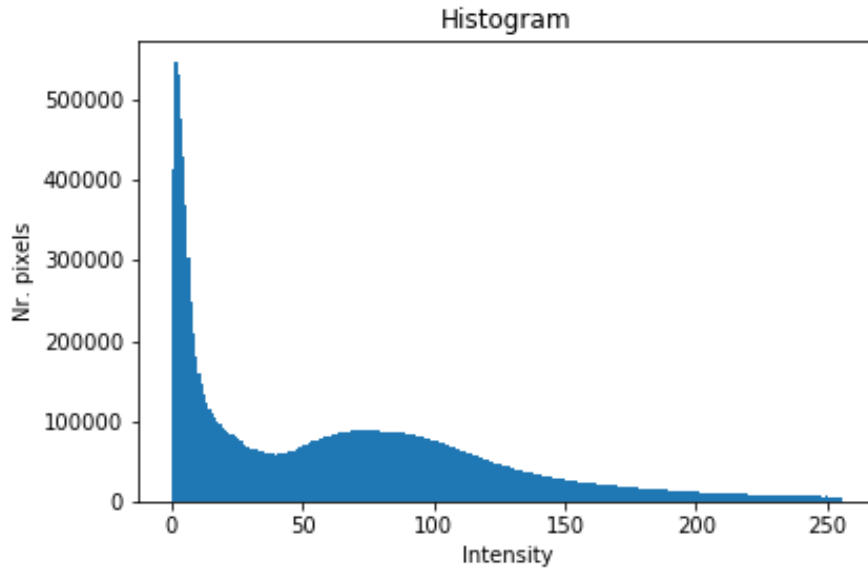


Figure 3.7: Histogram of an image with mul_value = 10 applied

So, to balance the results, it was decided that only a determined range of bins should enter the equation. It was discerned that the second maximum mostly held the values of the pixels within the solar surface. So it would be of interest to ensure that the range of bins included that maximum. It was important to make sure the range always included that value regardless of the image, so it was decided that a big interval should be considered.

The range selected is between the values 50 and 200.

Applying that change, the equation is then:

$$avg_w = \frac{\sum_{i=50}^{200} (i \times hist(i))}{\sum_{i=50}^{200} hist(i)} \quad (2)$$

A few *mul_value*s had to be manually experimented, to find the best fit for each set of images and determine a pattern.

- 2014: brightest
- 2016: medium
- 2018: darkest

The values presented in Table 3.2 were the ones found to be adequate for the three referenced years.

Table 3.2 – Adequate *mul_value* per year, obtained empirically

Year	Mul_value (manual)
2014	5.0
2016	7.5
2019	10.0

A set of five images for each year were selected and their respective weighted average obtained (Table 3.3). Merging with the manual *mul_value*s indicated in Table 3.2, these weighted average values, along with the previous equation, allowed for a trending line to be obtained (Figure 3.8), and thus an equation for the *mul_value* itself to be found.

Table 3.3 – Calculation of weighted average

	<i>avg_w</i>	Mul_value (manual)
Brightest	84.193	5.0
	84.082	5.0
	85.481	5.0
	89.458	5.0
	88.792	5.0
Medium	80.131	7.5
	77.257	7.5
	76.559	7.5
	74.405	7.5
	80.014	7.5
Darkest	60.805	10.0
	63.111	10.0
	64.456	10.0
	75.322	10.0
	65.080	10.0

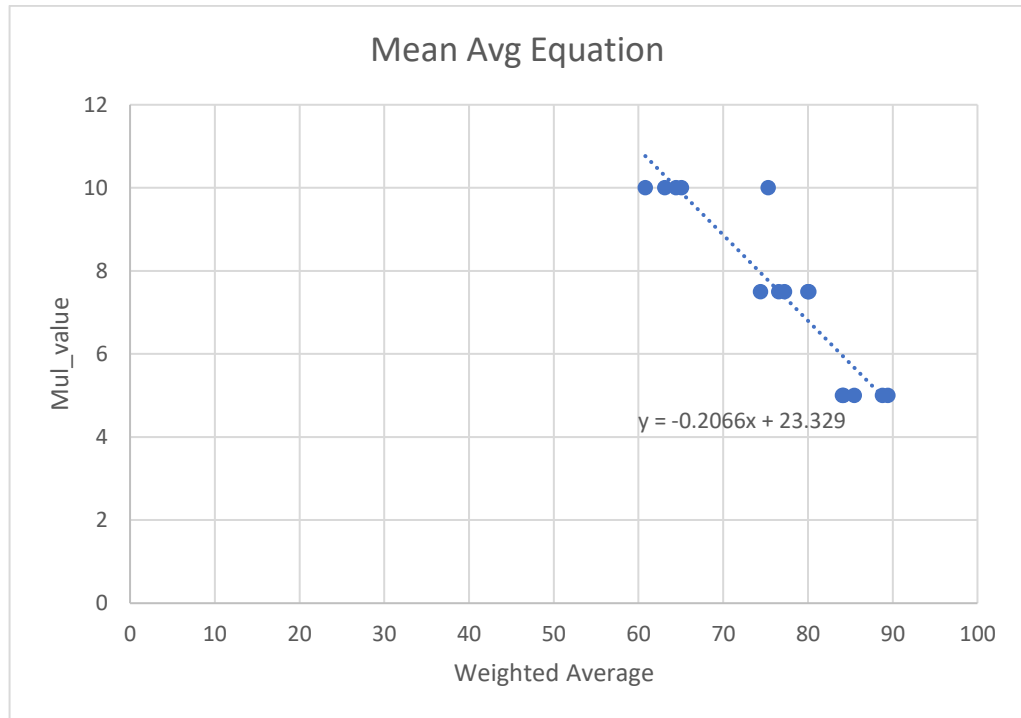


Figure 3.8: Mean Average trending line

The equation found for the trending line is then

$$mul_{value} = -0.21 \times avg_w + 23.33 \quad (3)$$

The result of applying this equation for the mul_value is displayed in Figure 3.9. The images are now much more similar in terms of intensity, and therefore can all have the same threshold applied to them in the detection.

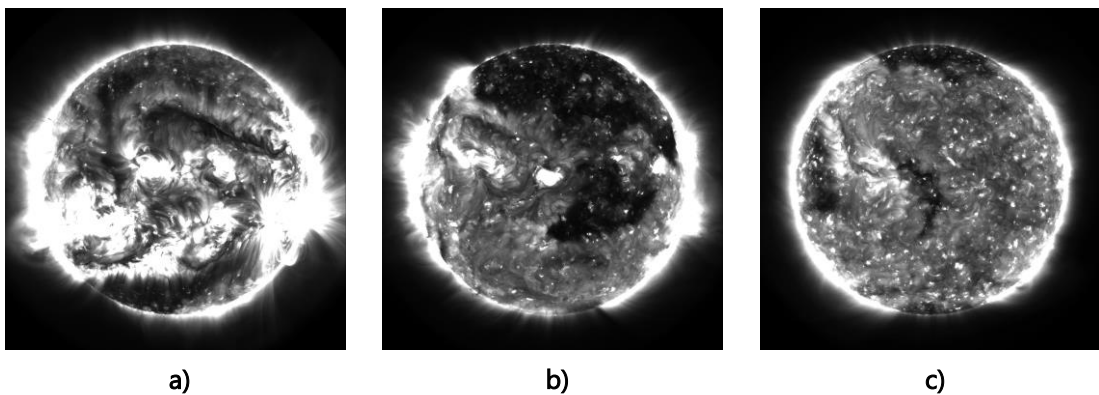


Figure 3.9 - AIA 193 images for three solar activity periods: a) High activity (28-10-2014); b) Date: 28-10-2016; c) Date: 28-10-2018

3.5. CHs Detection

The first detail to take into account for the CHs detection was that it should only be performed within the Solar disk. Initially, only a constant solar radius value was taken into account, but looking at the solar radius within the FITS file meta-information, it was found that it varies and is not the same in every image. This is due to SDO's geosynchronous orbit (a detail that is also pointed out in (Pesnell, 2014)). The satellite orbits around the Earth, which orbits around the Sun. The problem is that Earth's orbit around the Sun is not circular, but elliptical.

Figure 3.10 shows an example of this disparity. A set of images was taken throughout the year 2018, with a cadence of 30 days. For each image, the respective FITS file was accessed and the radius and center of the Sun obtained and saved in a small database.

From all the images, two were picked: the one with the largest radius and the one with the smallest. Both of them are shown in Figure 3.10a) and Figure 3.10b), respectively. As can be observed, the difference between the size of their solar disks is not easy to discern.

A mask was created for each of them, with the background in black and the area of the solar disk in white (Figure 3.10c) and Figure 3.10d)). However, the difference in size remains difficult to perceive in the human eye. So, to make it more perceptible, the two masks were merged with each other and their difference highlighted (Figure 3.10e)).

Considering that the exemplified difference is between the two most extreme cases (largest and smallest), the first thought while visually analyzing Figure 3.10e), was that the difference could be considered small enough to be discarded. However, a few details should be taken into account:

- The images are large in size (4096 x 4096 pixels)
- It is of interest to keep precision at the highest level possible
- The pixels that are either discarded or over accounted for might negatively influence the result

Table 3.4 displays the radii of the two masks represented in Figure 3.10, as well as the difference between them. As pointed out previously, although visually it doesn't seem significant, the difference is far from trivial.

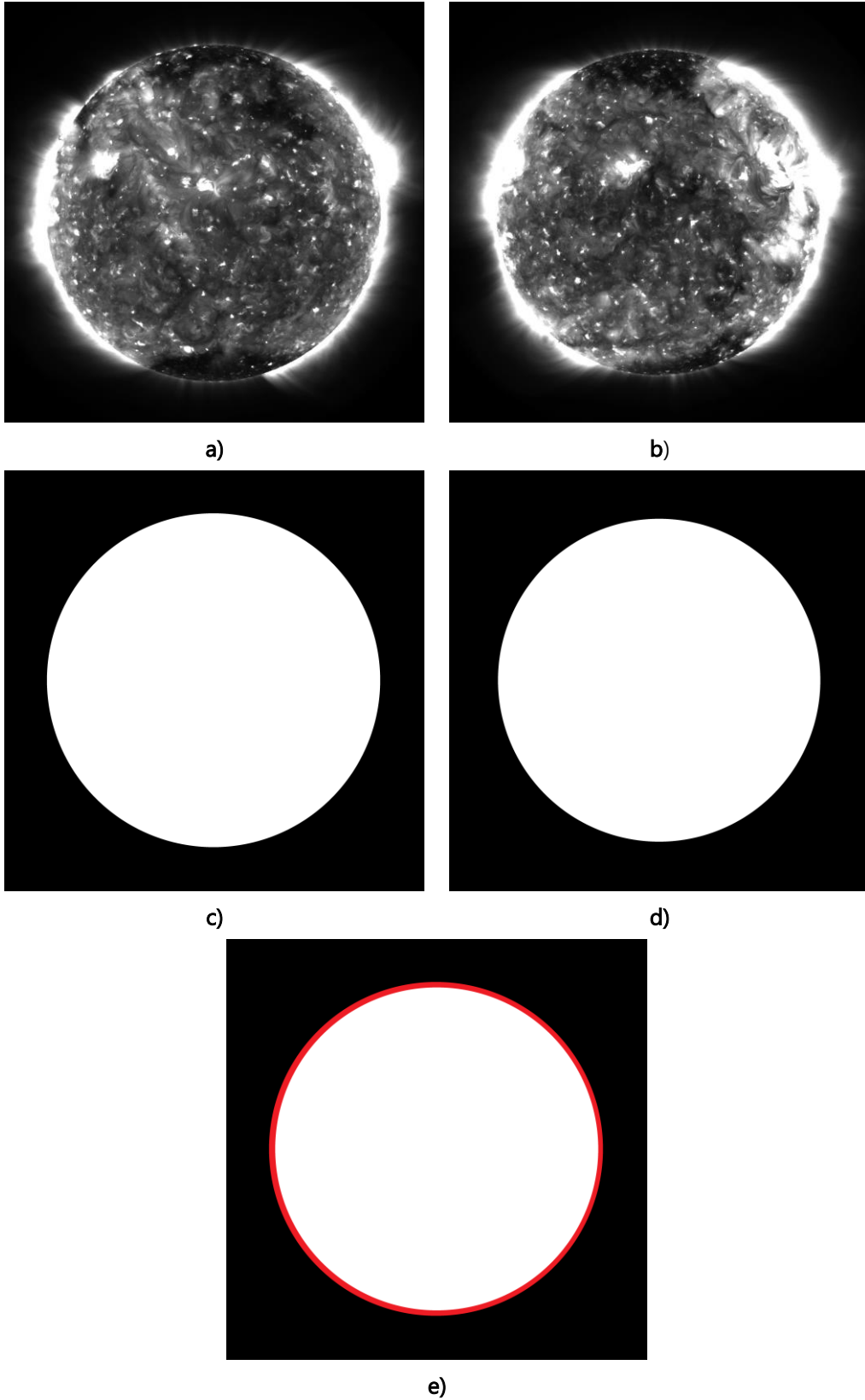


Figure 3.10: Solar disk radius difference. a) Image captured on 01-01-2018; b) Image captured on 30-06-2018
c) mask of a); d) mask of b); e) Difference between masks c) and d)

Table 3.4 – Solar disks radius and difference

	Value (pixels)
Radius in 01/01/2018 (Figures a and c)	1624.810
Radius in 30/06/2018 (Figures b and d)	1571.587
Radius Difference	53.223

Every solar disk mask is created right after each FITS file is retrieved. Although the masks are saved in a python dictionary for use in a future stage of the work, it was decided that said masks should also be saved in memory, for later use. All the masks are saved in the folder named 'masks' (see Figure 3.2). That way, if another process requires their use, there won't be a need to redo the entire retrieval process again.

With the masks created and ready for use, it was then proceeded on to the detection. To make the images lighter and maintain the algorithm's performance at a higher level, the detection results are saved as a mask rather than a colored image. Even though in the CV2 module it is not possible to create binary images, and thus the masks inevitably have to be 8-bit images (shades of grey), it is still a better option than to convert the detections into an RGB image, which would become heavier and take much longer to create.

The threshold segmentation was then performed. Figure 3.11 serves as a visual guide to provide better judgement of what areas the algorithm is supposed to detect. It is easily discernible that there are four dark areas that the algorithm ought to detect.

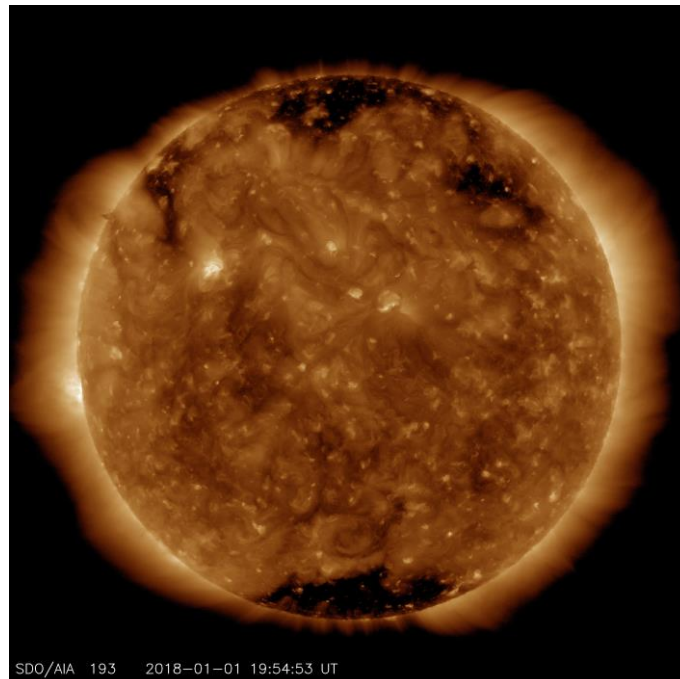


Figure 3.11 - Image taken manually from SDO (courtesy of SDO)

Different thresholds were tested, all only within the white area of the image's respective mask, and the result is presented in Figure 3.12, being white the pixels below the threshold and black the opposite. By observation, it was determined that a threshold of value 30 is the most adequate for the process.

A threshold of 40 was found to detect an excessive amount of areas and isolated pixels that, from Figure 3.11, clearly should not be detected.

A threshold of 20 manages to detect practically the four regions it is meant to, but barely, with a large portion of the intended regions being discarded.

Therefore, in the range considered, a threshold of 30 is considered to be the best to perform the detection. (It is noteworthy, however, that this value might change due to changes at the source of the images. While at the time these results were tested the threshold of 30 showed the best results, changes in the images could alter that factor.)

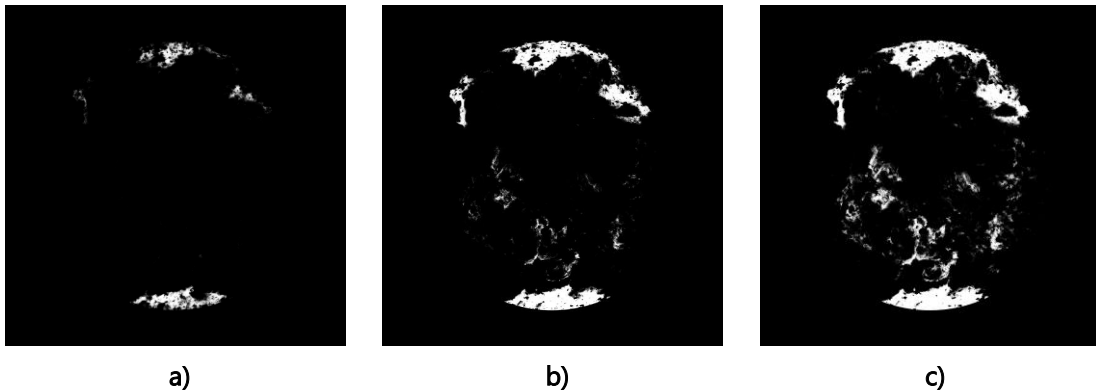


Figure 3.12 – Negative detection masks: a) threshold = 20; b) threshold = 30, c) threshold = 40

Because the detection result still holds a large amount of false positives a mean filter is applied to the image before performing the threshold segmentation. The module CV2 revealed itself to be very time efficient in this process, regardless of the size of the kernel used. So a kernel of 21x21 was chosen to smooth the image.

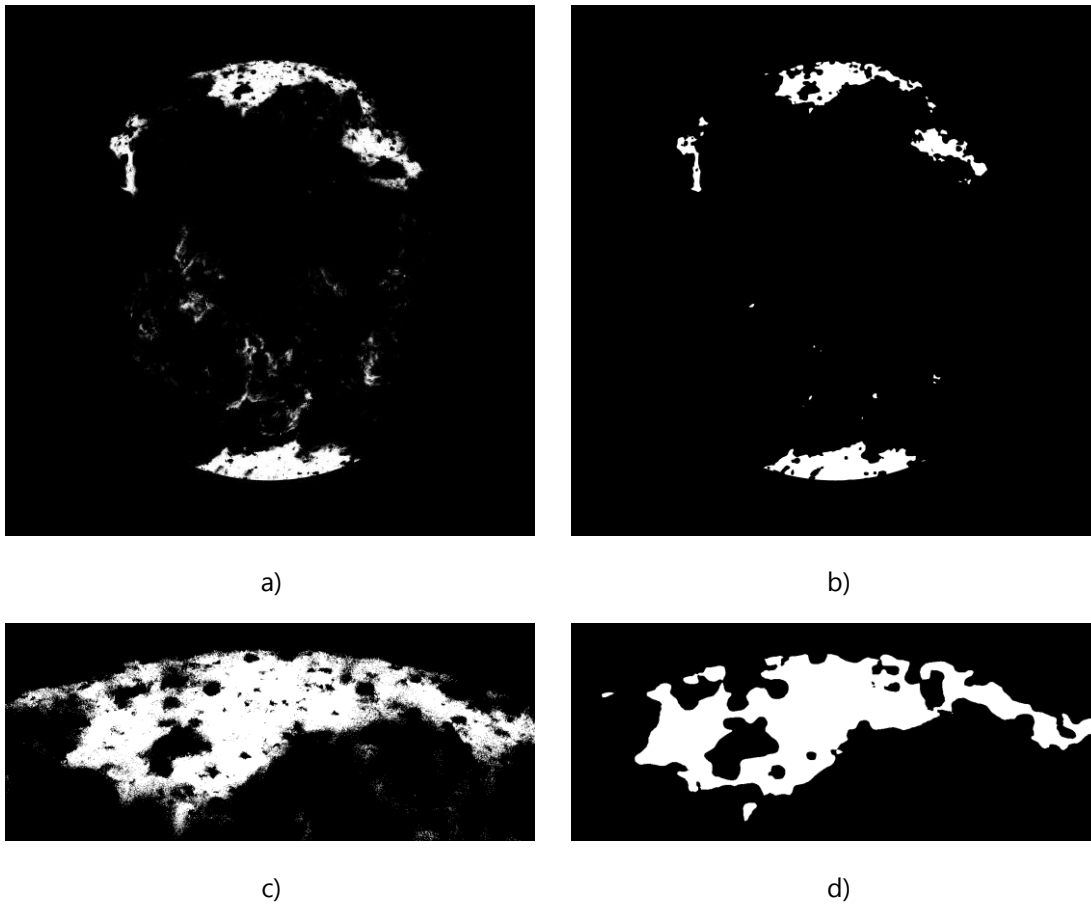


Figure 3.13: Effect of blurring threshold process: a) c) without blurr; b) d) with 21x21 mean filter

3.6. Post-processing

After the segmentation is complete, it was noted that there were still many small black patches inside the white areas of the masks. This is due to CBPs that are located inside the CHs. Their higher intensity leads to them being dismissed when the segmentation is performed, and therefore not included in the mask. As a result, the mask ends up with small “holes” in the Coronal Holes’ detection.

To attempt to correct this, the morphological operation of closing was applied. Once again, CV2 showed good time efficiency when performing this operation, no matter the size of the kernel used. So, a 21x21 kernel was applied. But, to ensure that the result of the operation remained smooth in the edges, a circular kernel was used rather than a rectangular one. CV2 allows for this kernel to be easily created with its structuring elements function:

```
cv2.getStructuringElement(cv2.MORPH_ELLIPSE, size_of_kernel)
```


The result of the closing operation does not show much progress. A few gaps did indeed disappear, but there are still plenty that remained, as can be observed in Figure 3.14.

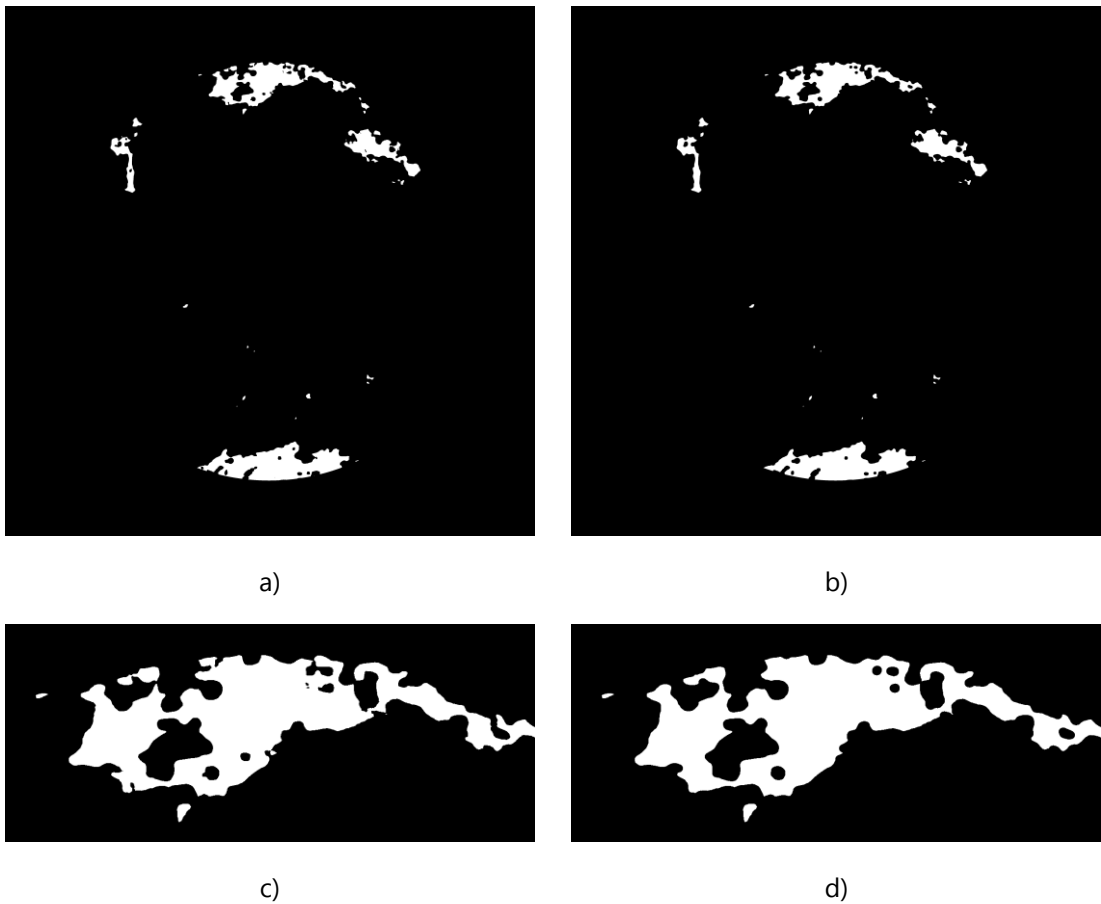


Figure 3.14: Closing operation: a) c) without Closing; b) d) with Closing

So, another solution was put in practice. An object detection was a promising alternative. The process consists in detecting all the objects in the mask, then eliminating those that had an area under a given value.

The maximum area specified for this operation was 10000 pixels, and it was done to both black and white objects. Every single object smaller than 10000 pixels is discarded (Figure 3.15).

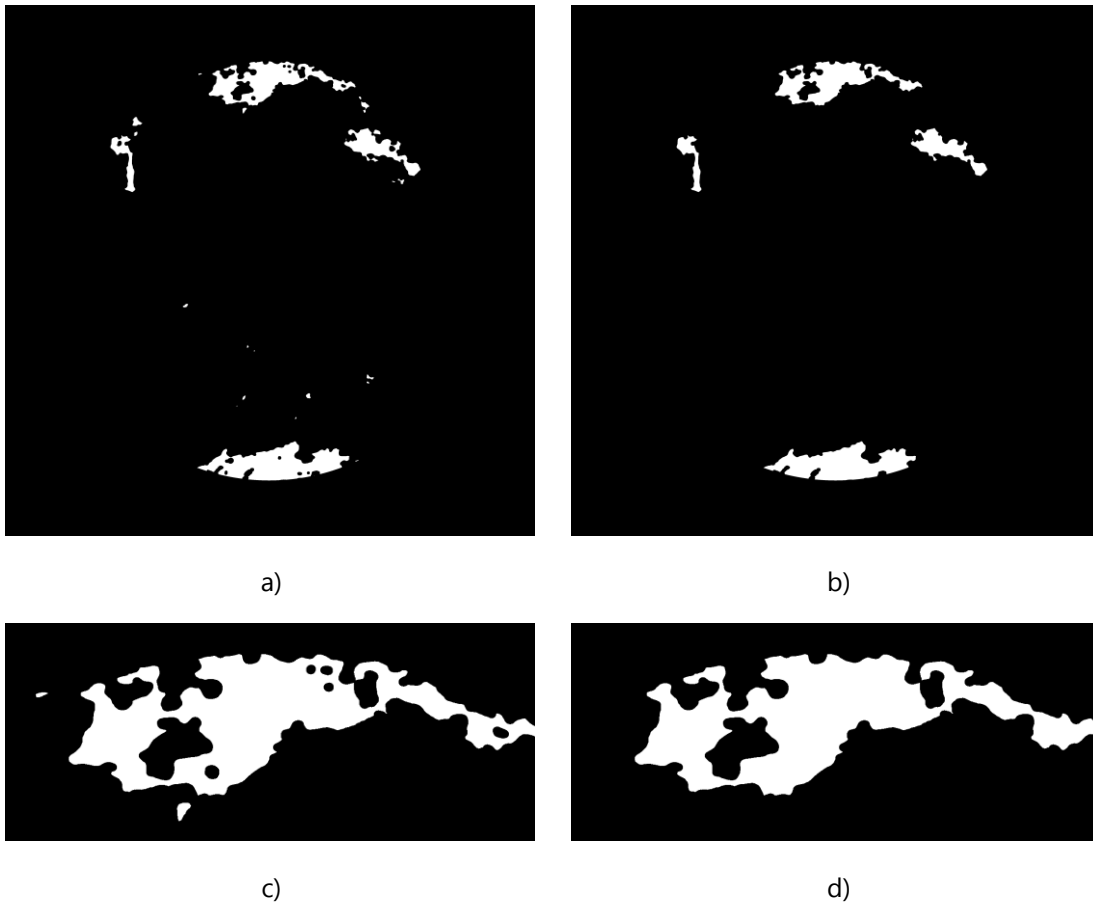


Figure 3.15: Object detection and deletion: a) c) without filter; b) d) with minimum size filter

A detail that was noticed in the result is that, even after the small objects deletion method, a few false negatives remained in the mask, due to being located at the limb and therefore considered part of the object outside the solar disk.

Unfortunately, the function `cv2.findcontours` does not include the option of adding a mask. So the solution was to “trick” the algorithm into separating between the small objects at the limb and the object outside of the Solar disk.

The disk masks created in the first stage of the work were reused for this. All the pixels inside the solar disk remain unaltered. The outside of the solar disk, however, is painted in a third color (cannot be black nor white) to successfully isolate the limb CBP areas (Figure 3.16).

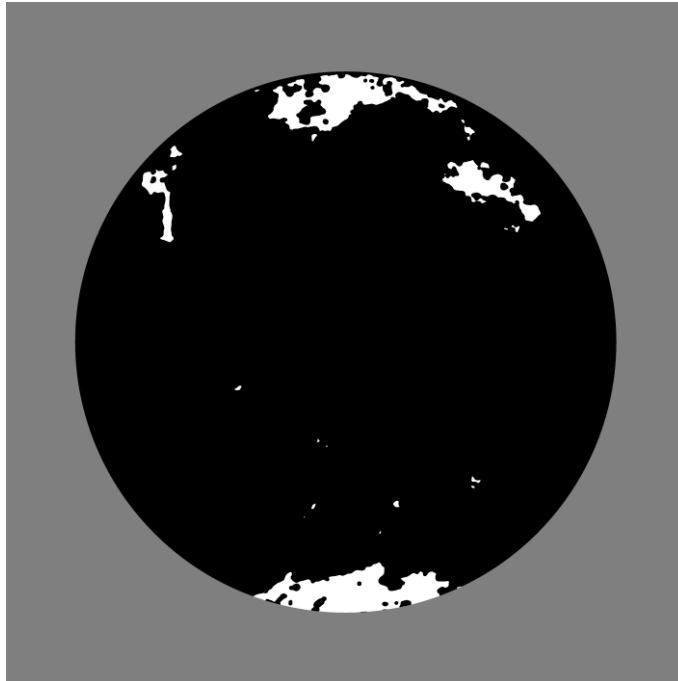


Figure 3.16: Object detection correction mask

The object detection is applied again, which successfully eliminates a few of those small areas located at the limb (Figure 3.17).

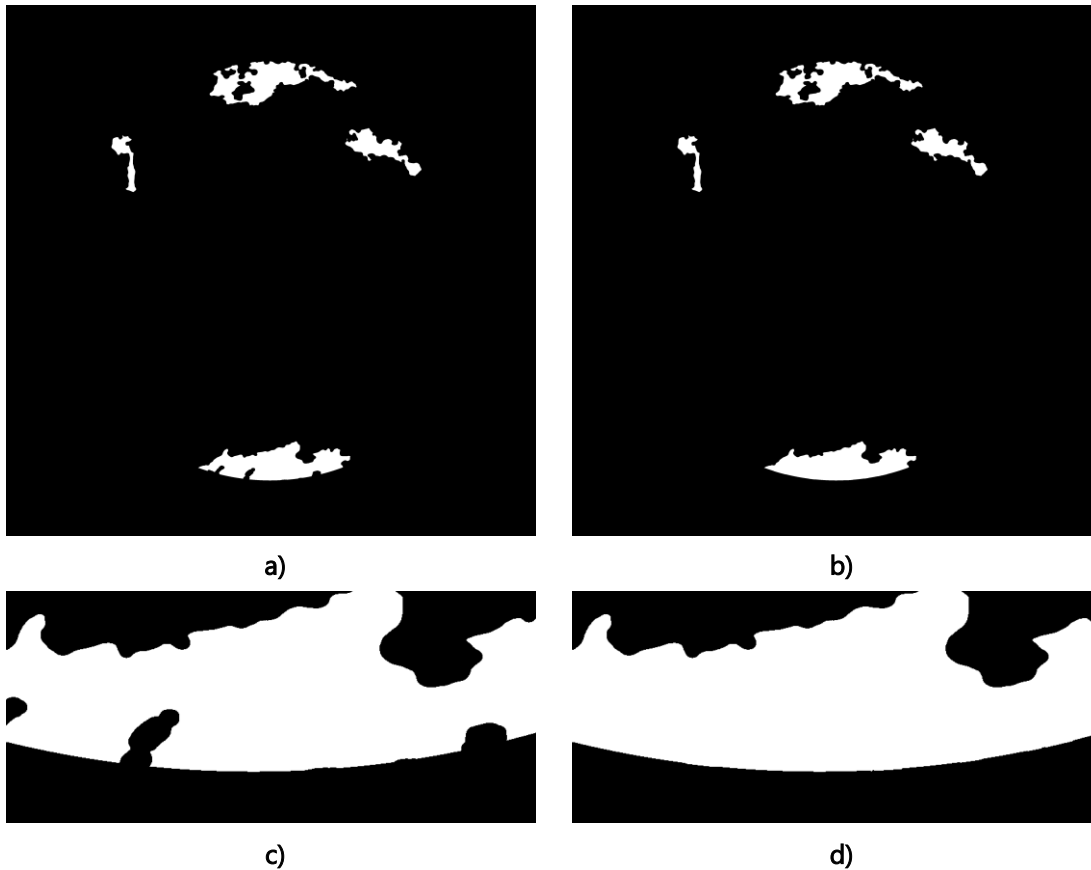


Figure 3.17: a) c) Detection before object deletion; b) d) Detection after object deletion

3.7. Integration with the Coronal Bright Point Tracker

The CBP tracker detects CBPs and acquires several information that it stores in a database. Of all that data on each CBP, the CH detector only needs three parameters: the name of the analyzed image and the two coordinates of the CBP.

The name of the analyzed image lets the detector know which of the masks of the detection of CH to use. Both the image being analyzed and the detection mask have the same name. They are simply in different folders. So the algorithm only has to look for the file with the same name as the image, in the folder where all CH detection masks are kept in. Once found, the program uses the mask to determine whether a CBP is inside a CH or not.

For that result to be reached, the two central coordinates of each CBP, (x, y) , are needed. The algorithm reads the coordinates, opens the mask file and reads the value of the pixel in those same coordinates.

Because the masks only have two colors (black or white), the method for obtaining a result on whether a CBP is inside of a CH or not is simple:

- If the value of the mask at point (x, y) is equal to 255 (i.e., white), then the CBP is inside a Coronal Hole;
- If the value of the mask at point (x, y) is equal to 0 (i.e., black), then the CBP is outside of a Coronal Hole.

This result is afterwards saved in the variable `in_CH`, which is then updated in the CBP's data in the database.

In order to better display the results, the algorithm proceeds to drawing the results in the original image, both for CBP locations and CH areas.

Coronal Hole areas are drawn onto the image in a shade of green, with some slight transparency so the features located inside those areas are still visible. The CBPs, for their part, are signaled with a circumference around them. The circumference is painted in red when the CBP is located outside of a Coronal Hole, or orange when it is located inside one (the orange color was chosen to better contrast with the green shade of the CH areas).



4. Results Assessment

For the purpose of confirming the final results' reliability and identifying improvements to be made, the results of the program were compared with the results of two sources: Space Weather Live (SWL) and SPoCA. This comparison is only made through visual observation of the images, since it was not possible to download the other tool's results in a way that they would be compatible in terms of size, resolution and type.

Because the Sun changes over the course of its 11-year cycle, it is of interest to perform this comparison in different stages of it. In this case, it was determined that three stages would suffice:

- High Solar Activity (years 2010/2011)
- Medium Solar Activity (year 2016)
- Low Solar Activity (year 2018)

Space Weather Live currently only displays their results from the year 2017 and forward, which means that the comparison with their results can only be done with reference to the low Solar activity stage. They also only provide one result per day. Fortunately, the displayed images have the date and time included in them, so it was possible to pick a detection closer to that time to prevent errors in the comparison.

SPoCA, on the other hand, has the results pertaining to all years available in the Helioviewer website, so the comparison with them can be done for all stages.

4.1. Small Solar Activity

The results from three captures from 2018 were chosen to compare (Figure 4.1).

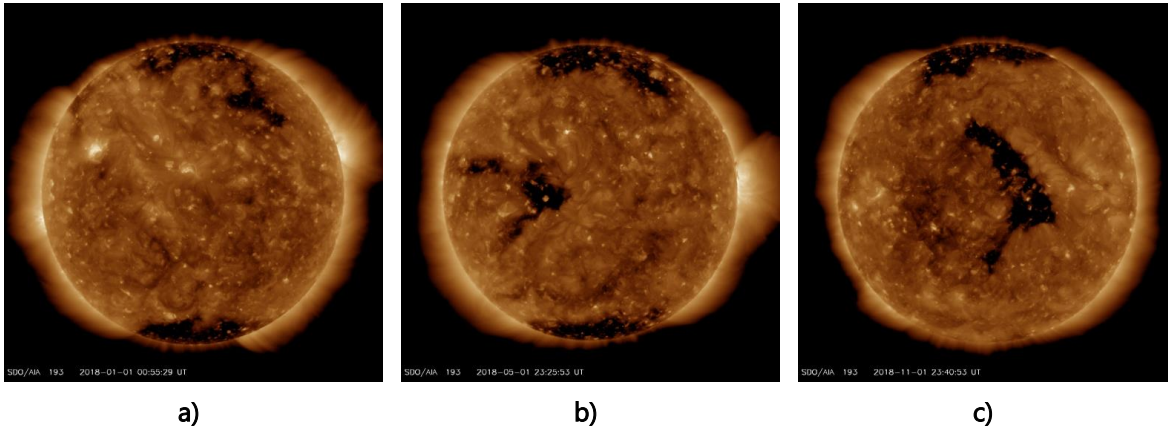


Figure 4.1 - Three chosen captures from year 2018: a) January 1st; b) May 1st; c) November 1st

The comparison between the results from SWL and the CHs detection tool are presented in Figure 4.2 and between SPoCA and the CHs detection tool are presented in Figure 4.3

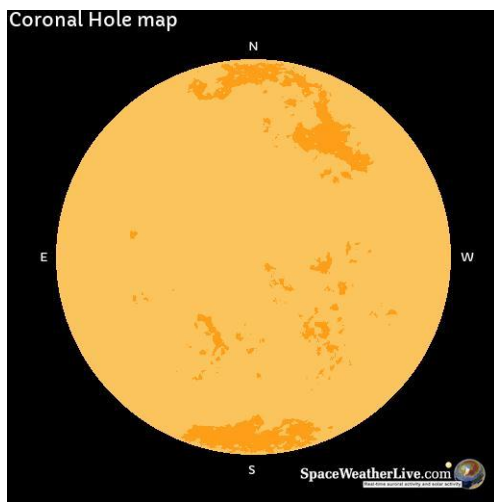
Starting with the comparison with SWL (Figure 4.2), it is noticeable that the proposed tool manages to filter out false positives more efficiently than SWL. When it comes to false negatives, however, both tools are at about the same level.

Overall, this CHs detection tool shows more accuracy in the final product than SWL.

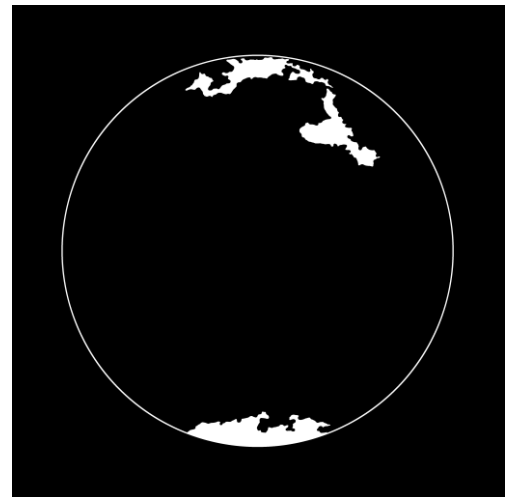
Moving on to the comparison with SPoCA detection, in Figure 4.3, the first detail to note is that the results between SPoCA and the CHs detection tool are much more similar.

While in some situations SPoCA shows more success in dealing with false negative detection (Figure 4.3e)), in most of the cases the CHs algorithm shows better performance in detecting the CHs regions.

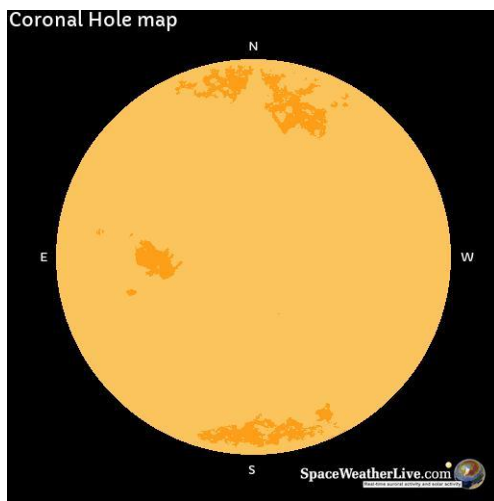
As can be seen in Figure 4.3a) through d), there are smaller areas that SPoCA discards and which, by observation of the images in Figure 4.1, clearly should not be dismissed, while the CHs detection tool maintains them in the result.



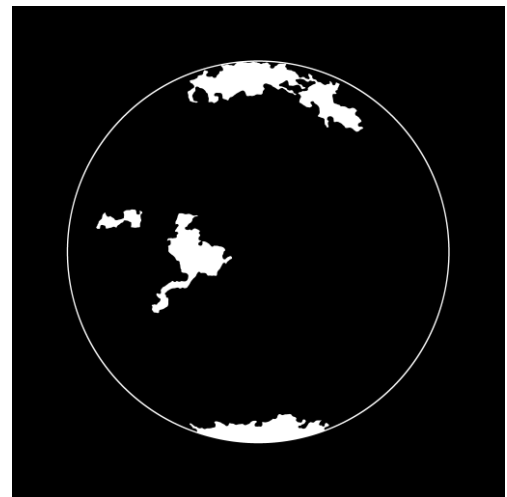
a)



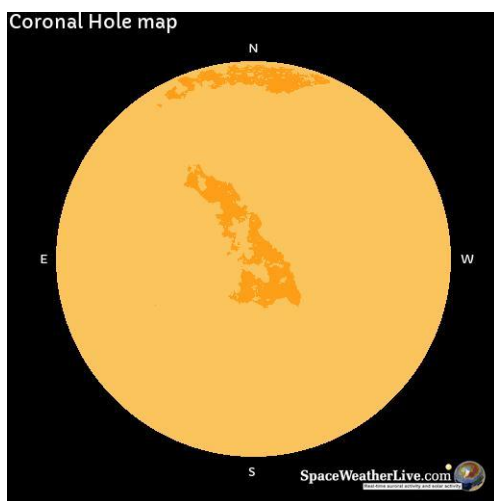
b)



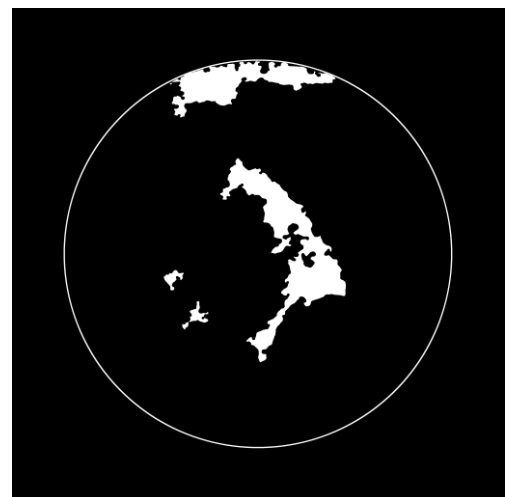
c)



d)

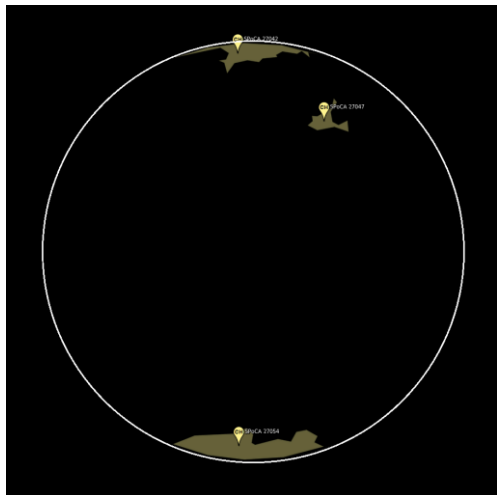


e)

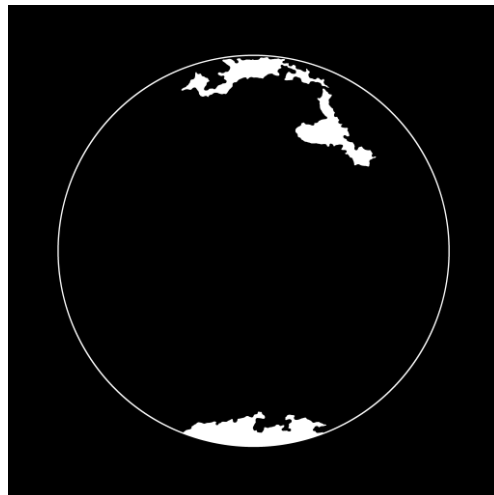


f)

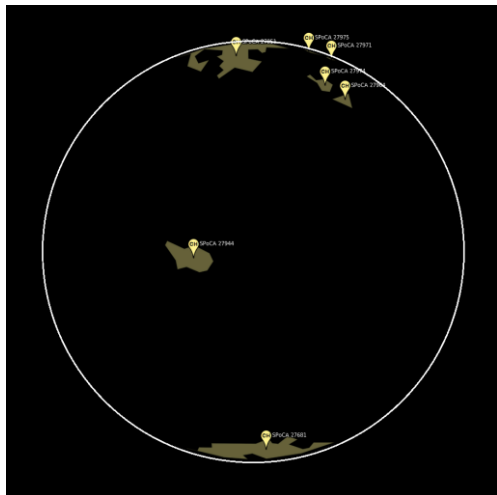
Figure 4.2 - Detection results: a), c) and e) Space Weather Live; b), d) and f) Proposed Tool



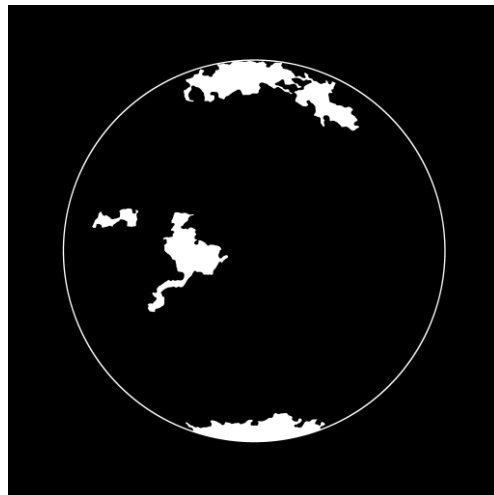
a)



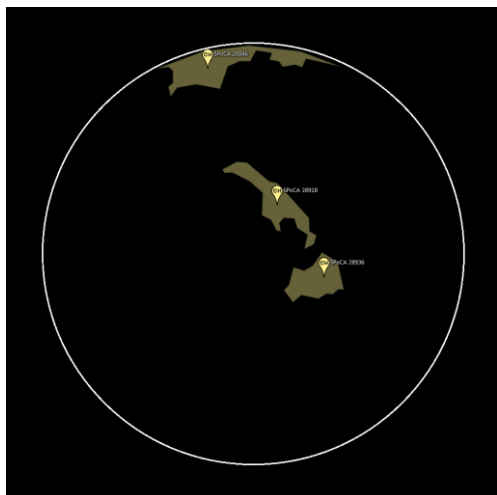
b)



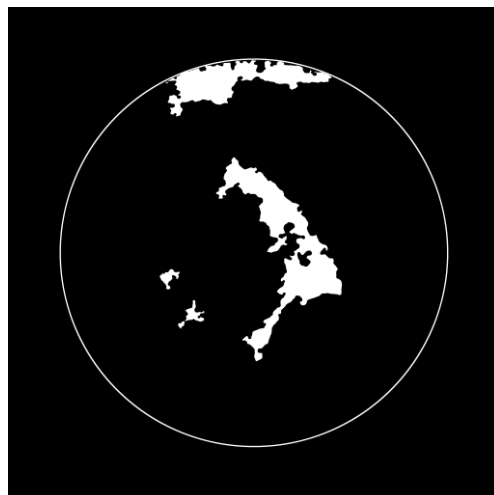
c)



d)



e)



f)

Figure 4.3 - Detection results: a), c) and e) SPOCA; b), d) and f) Proposed Tool (Please pay attention that the image have slightly different resolutions – solar radius is bigger in Heliviewer)

4.2. Medium Solar Activity

For this stage of the Solar Cycle, and because Space Weather Live is only available since year 2017 and forward, we just used SPoCA results for comparison.

Following the same logic as was presented in the previous section, three captures from the year 2016 were chosen and are presented in Figure 4.4.

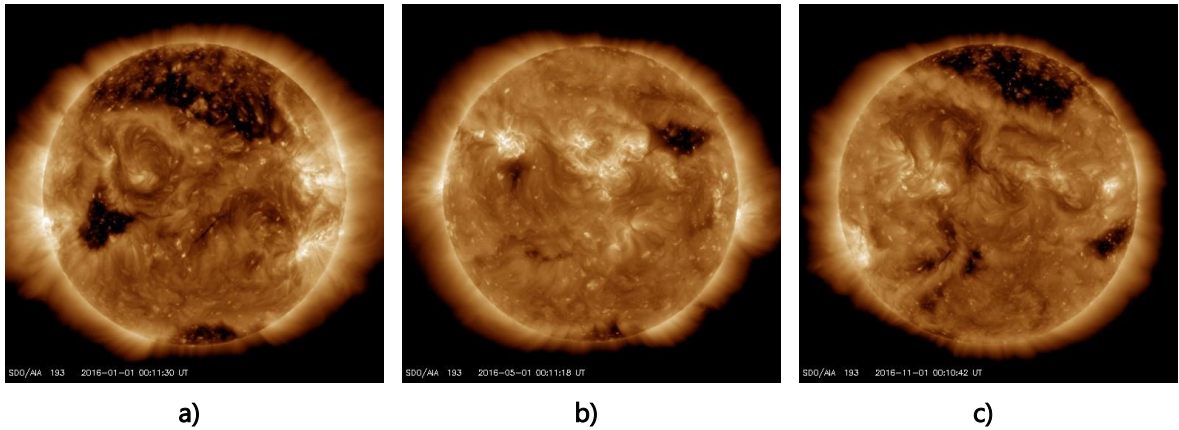
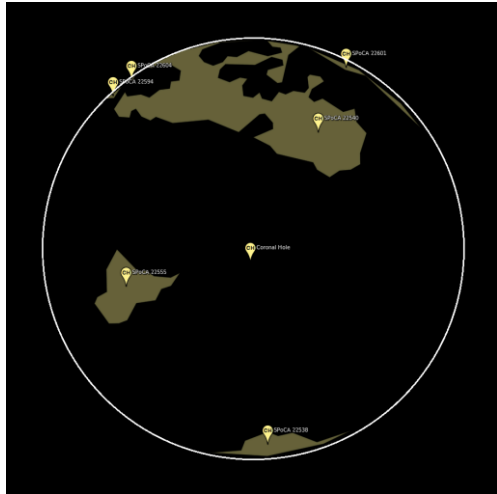


Figure 4.4: Three chosen captures from year 2016: a) January 1st; b) May 1st; c) November 1st

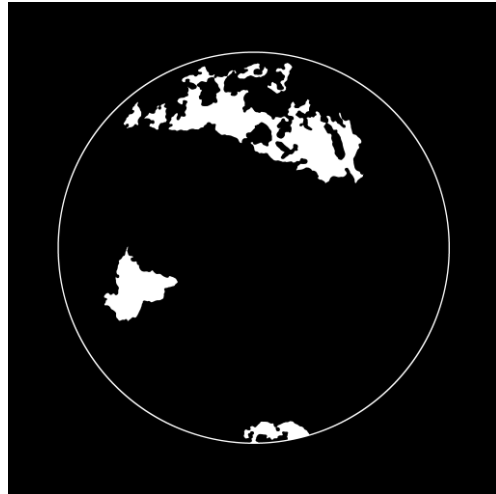
In this stage of the Solar cycle, the results seem to differ from each other a bit more than they did with lower activity images.

SPoCA does not manage to discard events such as filaments, as can be observed in Figure 4.5a). The CHs detection algorithm is also not prepared to be able to differentiate between one and the other, and thus might end up considering bigger sized filaments as a CH as well. However, if the filaments are small, as is the case present in Figure 4.4a), they get discarded by the tool.

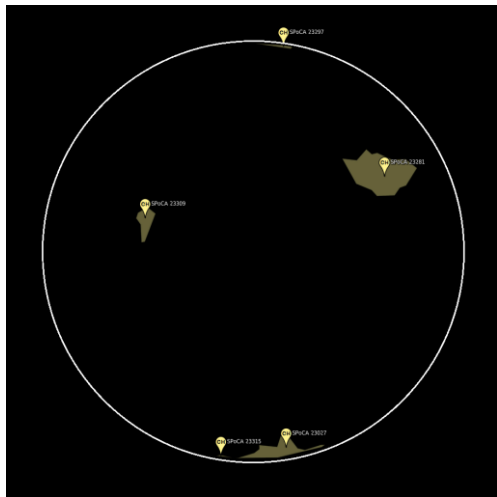
SPoCA shows better results than the CHs algorithms in regards to CHs who are smaller in size and located at the limb. When it comes to larger CHs, however, the CHs detection tool still shows better results, mainly in the shape of the areas.



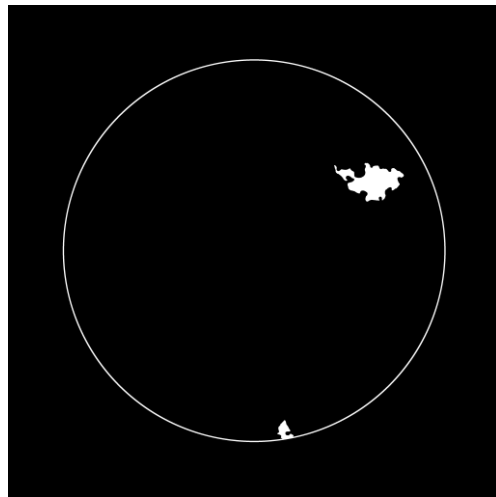
a)



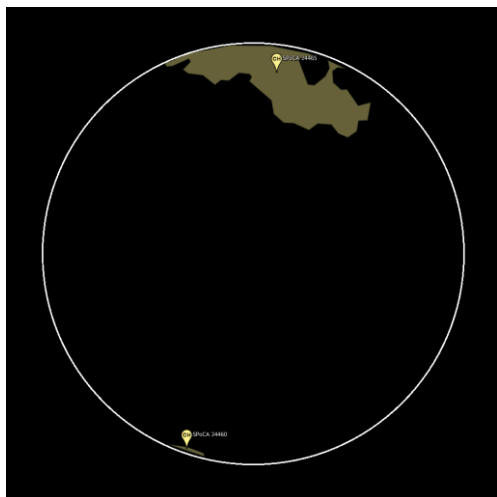
b)



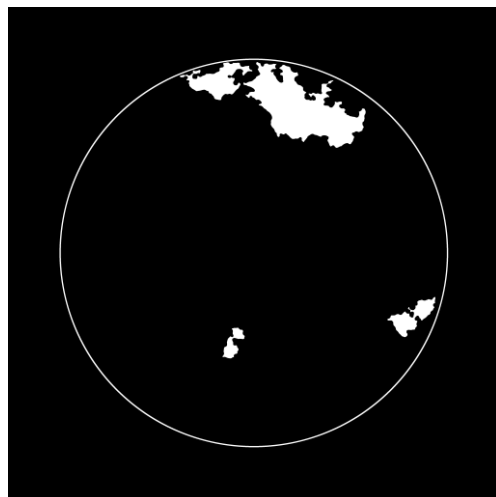
c)



d)



e)



f)

Figure 4.5: Detection results: a), c) and e) SPoCA; b), d) and f) Proposed Tool (Please pay attention that the image have slightly different resolutions – solar radius is bigger in Helioviewer)

4.3. High Solar Activity

Similarly as Medium Solar Activity, for this stage of the Solar cycle, only SPoCA results are available for comparison.

Following the same logic as was presented in the previous sections, three captures from the end year 2010 and beginning of the year 2011 were chosen and are presented in Figure 4.6.

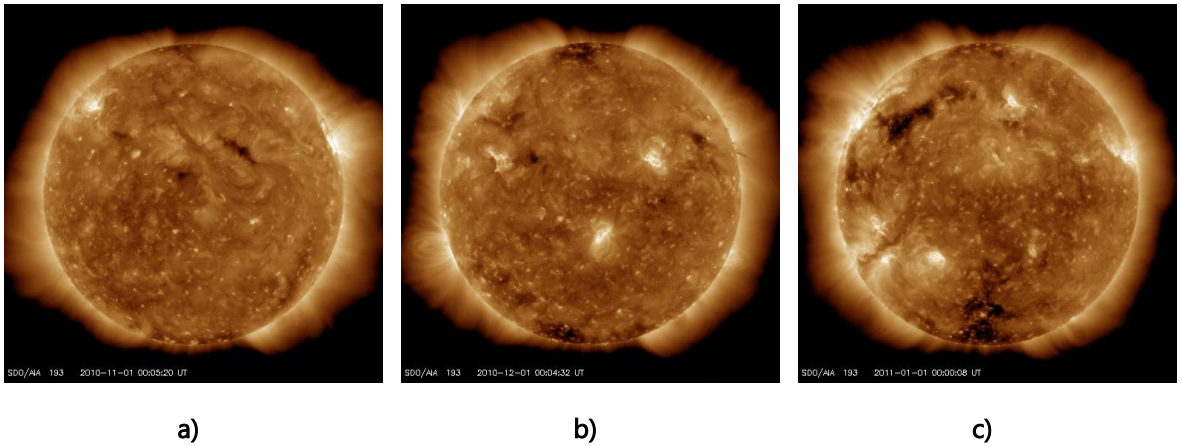
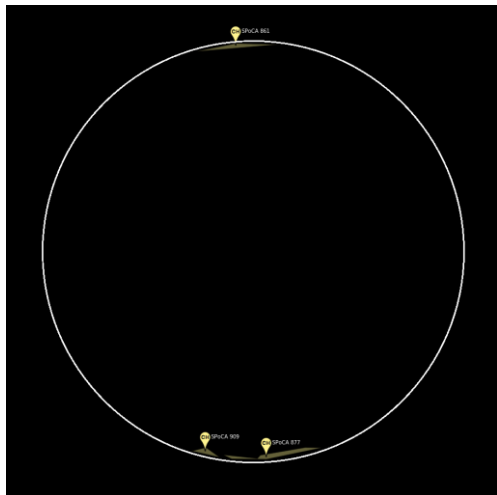


Figure 4.6: Three chosen captures from years 2010 and 2011: a) November 1st, 2010; b) December 1st, 2010; c) January 1st, 2011

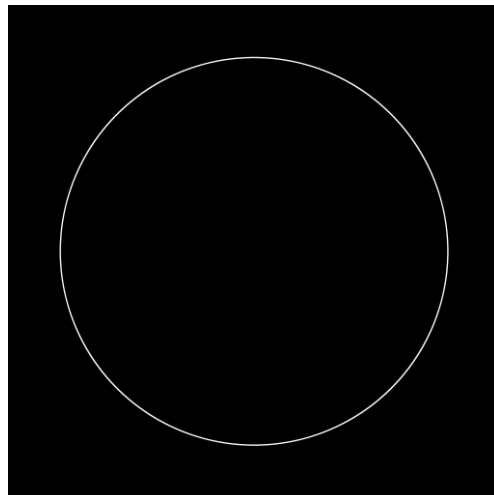
In this stage, both tools have far more difficulty in detecting CHs than in other stages. However, SPoCA displays much better results than the CHs algorithm.

This is especially apparent when it comes to detecting CHs at the limb. SPoCA manages to not only detect them, but also delimit them with more precision.

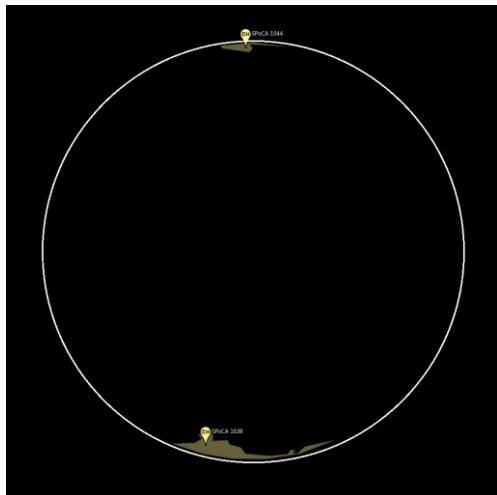
It is clear that the CHs algorithm had more difficulty handling the images of this stage, not being able to detect CHs as well in these cases.



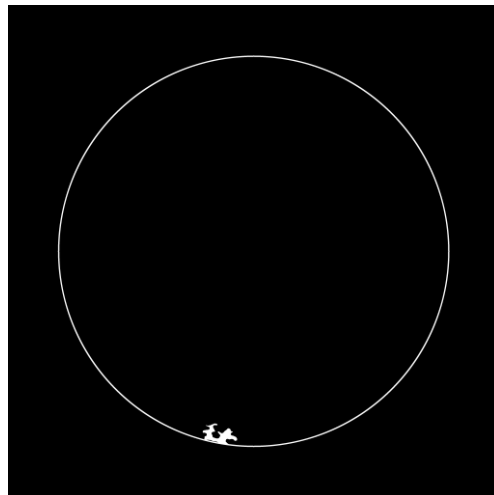
a)



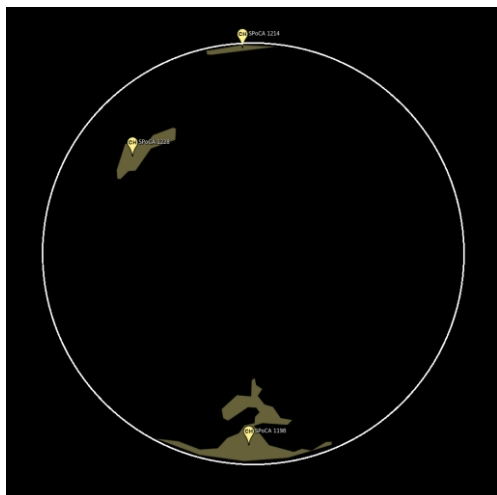
b)



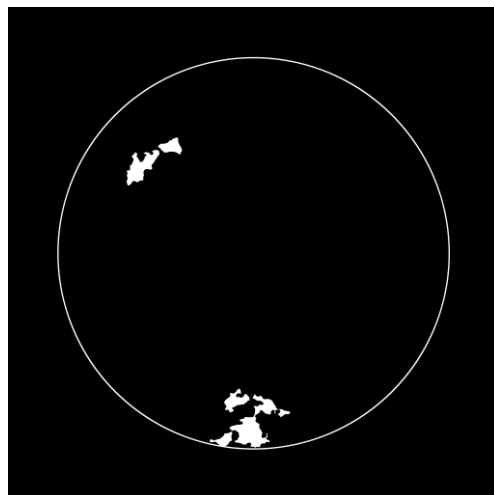
c)



d)



e)



f)

Figure 4.7: Detection results: a), c) and e) SPoCA; b), d) and f) Proposed Tool (Please pay attention that the image have slightly different resolutions – solar radius is bigger in Helioviewer)

4.4. Integration with Coronal Bright Point Tracker

The final stage of the work consisted in using the results of the CH detection (as are exemplified in sections 4.1, 4.2 and 4.3) to determine whether the CBPs are located inside or outside a CH. The examples given in the previous sections will be used in this one as well. A few factors must first be pointed out, for it is of importance that they are acknowledged.

The CBP tracker only detects CBPs until a specified latitude and longitude (95% of the solar disk). Because the CHs are most commonly located at the poles, this leads to a large number of CBPs located inside CHs not being detected at all, and thus not classified. However, there are cases where these polar CHs are large enough to expand into the CBPs detection zone, thus making it possible to detect a few CBPs within the extended areas.

Although CHs are most commonly located at the poles, there can still be instances where a CH is located on the more equatorial area, thus having its entire area inside the CBP detection zone and not including the previously mentioned issue.

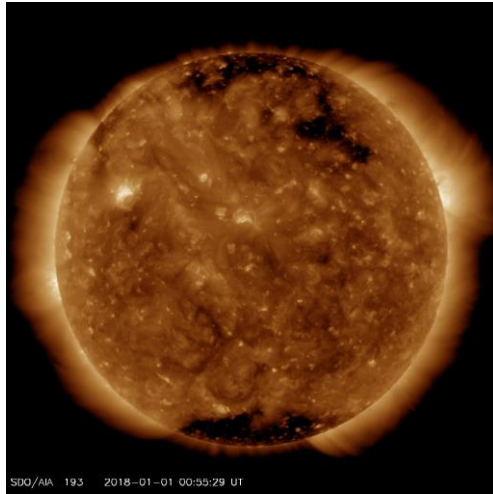
4.4.1. Low Solar Activity

Using the same images that were used in Section 4.1, the masks obtained in Figure 4.2 lead to the following results observed in Figure 4.8, along with the results of the CBP tracker.

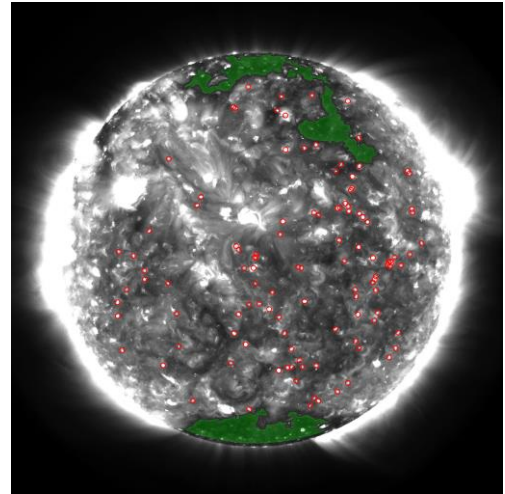
In this stage of the Solar Cycle CHs are bigger and tend to be darker, which facilitates the precision in their detection.

However, there are a few noticeable issues that influence the results:

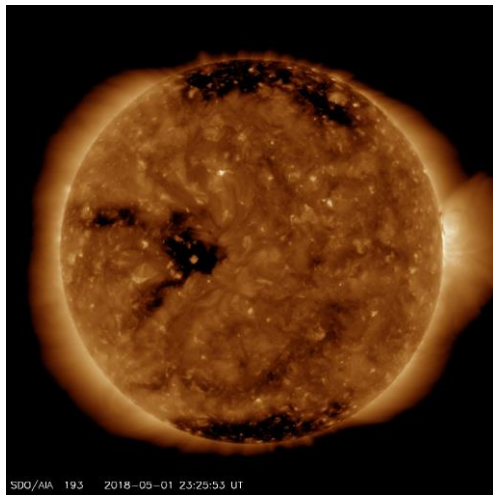
- By observation of Figure 4.8a), Figure 4.8c) and Figure 4.8e), there are noticeably CBPs located inside the CBP detection zone that were not detected by the algorithm;
- The areas of the CHs still contain various gaps either within them or at their borders. Many of these gaps are caused by CBPs themselves, mainly due to their differences in size, intensity and the fact that a few are so close to one another that they form clusters. While a few of these gaps might be easily solved by altering some parameters in the object detection and deletion stage, other cases would not be solved as easily, such as CBPs located at the border of CH areas (as is observable in the equatorial CH in Figure 4.8c)).



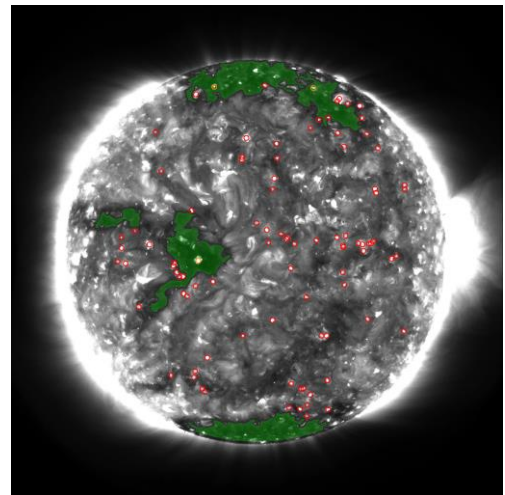
a)



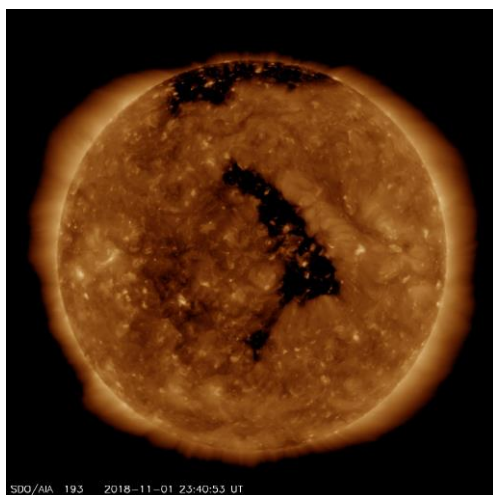
b)



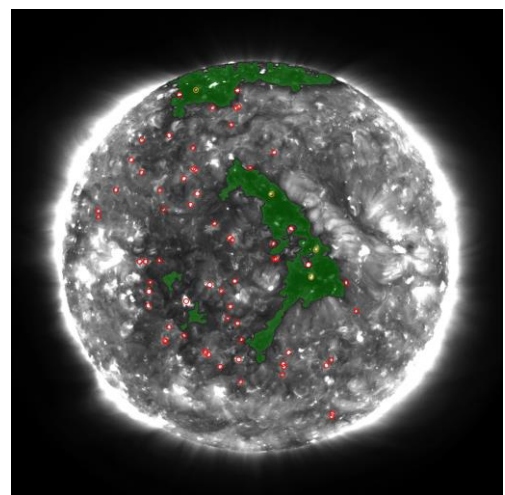
c)



d)



e)



f)

Figure 4.8 - a), c) and e) SDO images; b), d) and f) respective images with Coronal Hole detection and Coronal Bright Point classification

4.4.2. Medium Solar Activity

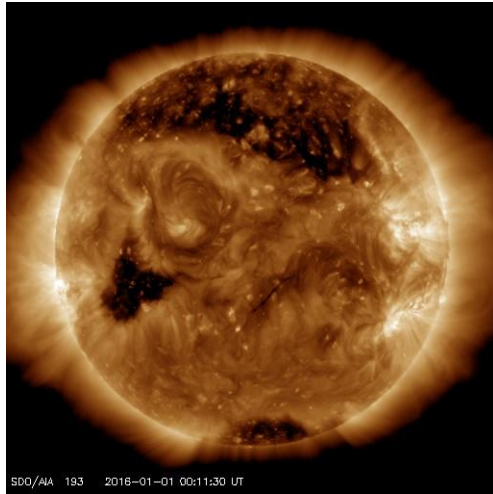
Using the same images that were used in Section 4.2, the masks obtained in Figure 4.5 lead to the following results observed in Figure 4.9, along with the results of the CBP tracker.

In comparison with the Low Solar Activity stage in Section 4.4.1, these images show an increase in brightness. This leads to an observable increase in the range of color within the CH areas, which also leads to uncertainty if certain areas ought to be considered CH or not (as is the example of the upper polar CH in Figure 4.9a)).

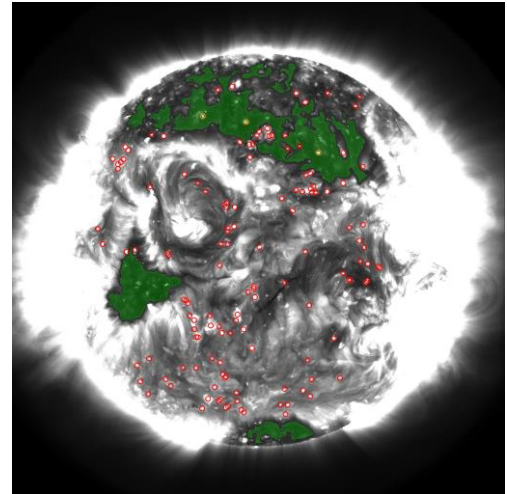
The difference in brightness in these images also adds difficulty in detecting some CBPs, as well as create some uncertainty on a few cases as well.

There are also noticeable instances where CBPs inside CHs ought to be detected, but were not.

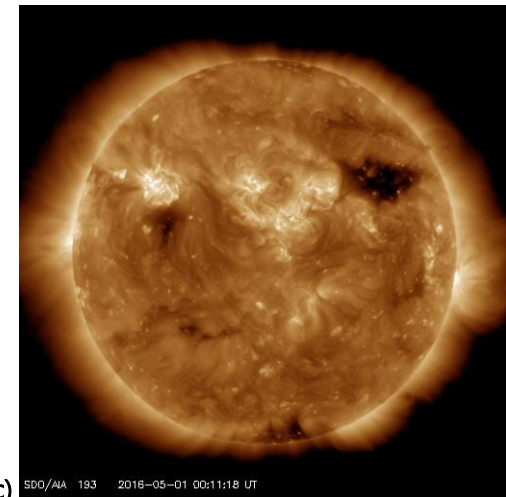
Lastly, the existence of gaps as described in the previous section can also be discerned in these results. Some CBPs, as well as some clusters, influence the results of the CHs detection.



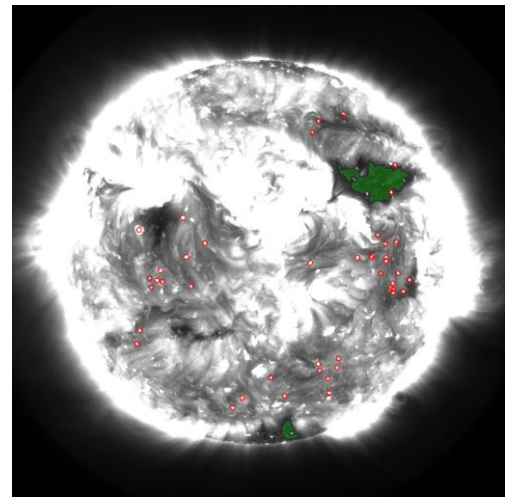
a)



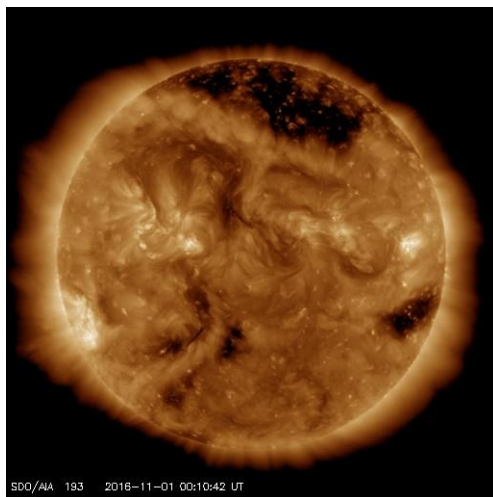
b)



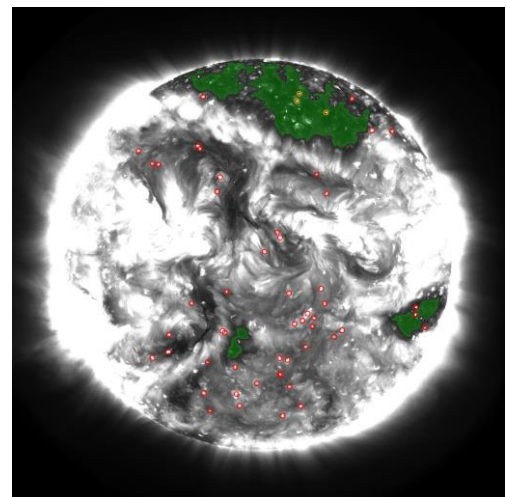
c)



d)



e)



f)

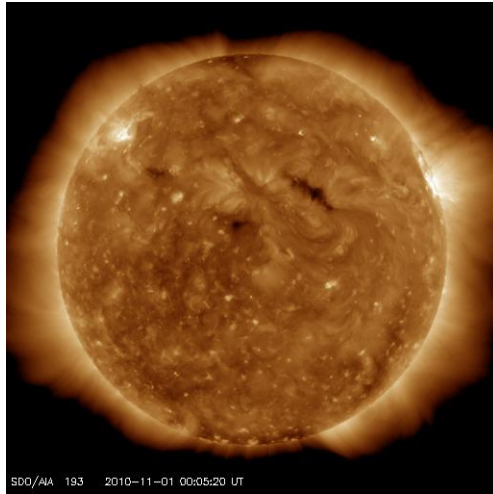
Figure 4.9 -a), c) and e) SDO images; b), d) and f) respective images with Coronal Hole detection and Coronal Bright Point classification

4.4.3. High Solar Activity

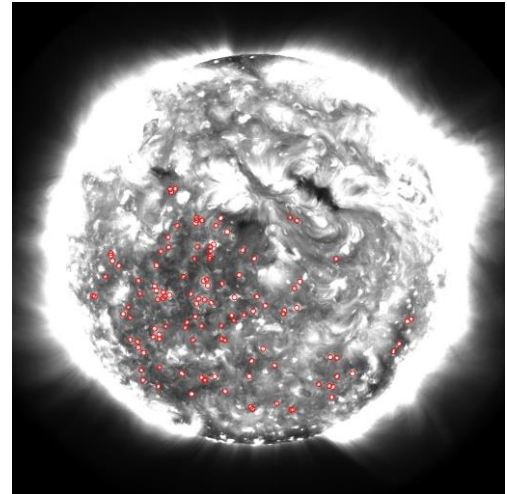
Using the same images that were used in Section 4.3, the masks obtained in Figure 4.7 lead to the following results observed in Figure 4.10, along with the results of the CBP tracker.

Due to being of the stage of highest Solar Activity, the increased brightness of these images further adds to the algorithm's difficulty in detecting CBPs in some areas.

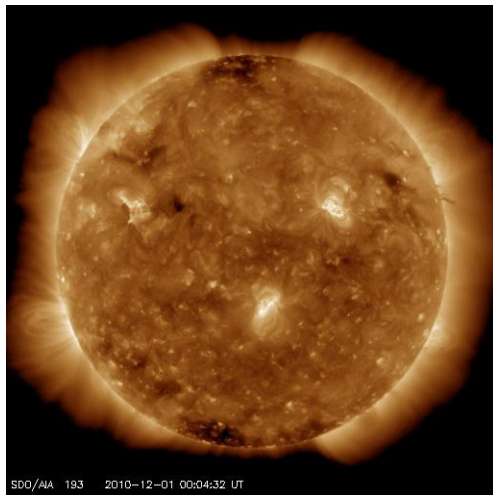
CHs are also much smaller in size, in some instances not being detected at all. The existence of gaps caused by CBPs is also present in the images where the CHs were detected.



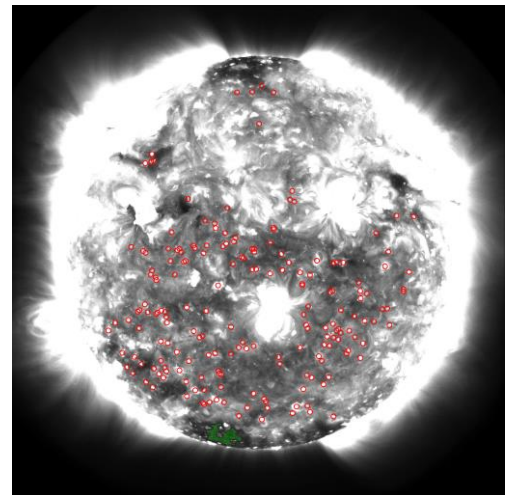
a)



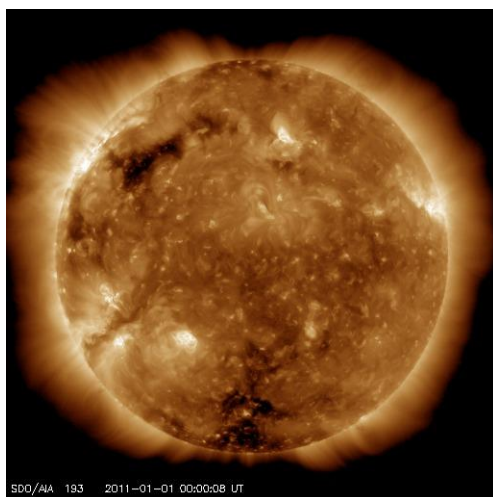
b)



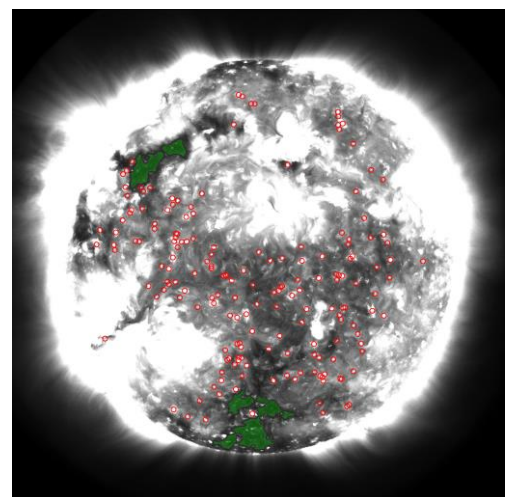
c)



d)



e)



f)

Figure 4.10 –a), c) and e) SDO images; b), d) and f) respective images with Coronal Hole detection and Coronal Bright Point classification



5. Conclusions and Future Work

The algorithm proposed in this dissertation had the objective of correctly detecting Coronal Holes in solar images from the SDO/AIA instruments, and it performs, in the majority of cases, with good results.

Overall, the tool's final product shows sufficient precision for achieving the goals the integration in the CBP tracker. The solutions that were used to resolve the problems were rather simple, yet very effective. One great detail that must be pointed out about this tool is its time efficiency. The newly added elements of CHs detection revealed to be the most time effective in the whole process.

The results also had good performance when compared with the data that is available from other similar tools. However, there is still room for further improvements.

5.1. Future Works

The algorithm for automatic detection of Coronal Holes still needs several improvements. Although it proves to be able to handle images of different intensities, it still has trouble with the detection in some of the most extreme cases.

A good improvement would be the optimization of the pre-processing and retrieval of the images, for this procedure takes longer to perform than all the other stages of the algorithm together.

Another important improvement is related to handling false negatives and other types of events, where the algorithm still has issues in. One great possibility is the addition of another wavelength to aid in the detection. While doing some experiments in available interactive tools, it was noted that the fusion of images of different wavelengths, with different weights applied to them, could lead to better results.

Lastly, the integration between the two tools (CHs detection and CBP tracker) is worthy of attention. One possibility that was pondered was that, in terms of image pre-processing procedures, each tool might require a different process to further improve the precision of the overall results. In other words, each original image would be manipulated in two different ways, thus creating two new versions: one specifically for the CBP tracker's process and another only for the CH detection process. The existing data on the CBPs could also possibly aid to improve the precision in CH area demarcation.



6. References

- Coelho, A. M. F. (2017) *Solar rotation speed by detecting and tracking of Coronal Bright Points*. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.
- Cranmer, S. R. (2009) 'Coronal Holes Imprint / Terms of Use', *Living Reviews in Solar Physics*, 6. Available at: <http://adsabs.harvard.edu/abs/2009LRSP....6....3C>.
- Delouille, V. *et al.* (2018) 'The SPOCA-suite : a software for extraction and tracking of Active Regions and Coronal Holes on EUV images', pp. 1–15.
- Hanslmeier, A. (2009) 'Habitability and Cosmic Catastrophes', pp. 55–79. doi: 10.1007/978-3-540-76945-3.
- Hathaway, D. H. (2015) 'The solar cycle', *Living Reviews in Solar Physics*, 12(1). doi: 10.1007/lrsp-2015-4.
- Kane, R. P. (2008) 'Prediction of Solar Cycle Maximum Using Solar Cycle Lengths', *Solar Physics*, 248(1), pp. 203–209. doi: 10.1007/s11207-008-9125-8.
- Karachik, N. V. and Pevtsov, A. A. (2011) 'Solar wind and coronal bright points inside coronal holes', *Astrophysical Journal*. Institute of Physics Publishing, 735(1). doi: 10.1088/0004-637X/735/1/47.
- Kempton, D. J., Schuh, M. A. and Angryk, R. A. (2018) 'Tracking Solar Phenomena from the SDO', *The Astrophysical Journal*. IOP Publishing, 869(1), p. 54. doi: 10.3847/1538-4357/aae9e9.
- Kirk, M. S. *et al.* (2009) 'Automated detection of EUV polar coronal holes during solar cycle 23', *Solar Physics*, 257(1), pp. 99–112. doi: 10.1007/s11207-009-9369-y.
- Madjarska, M. S. (2019) 'Coronal bright points', *Living Reviews in Solar Physics*. Springer International Publishing, 16(1), pp. 1–79. doi: 10.1007/s41116-019-0018-8.
- Mobberley, M. (2008) 'Probing the Mysteries of our Nearest Star...', in, pp. 140–149.
- Mora, A. D. *et al.* (2011) 'Automated drusen detection in retinal images using analytical modelling algorithms', *BioMedical Engineering Online*, 10. doi: 10.1186/1475-925X-10-59.
- Pattichis, M. S. *et al.* (2015) 'Detecting coronal holes for solar activity modeling', *Conference Record - Asilomar Conference on Signals, Systems and Computers*, 2015-April, pp. 89–93. doi: 10.1109/ACSSC.2014.7094403.

Pesnell, W. D. (2014) 'Solar Dynamics Observatory (SDO)', *Handbook of Cosmic Hazards and Planetary Defense*, pp. 1–15. doi: 10.1007/978-3-319-02847-7_16-1.

Pires, R. M. P. (2018) *A web tool to detect and track Solar features from SDO images*. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

Saha, K. (2008) 'The Sun And Its Radiation', in *The Earth's Atmosphere*. Springer, Berlin, Heidelberg, pp. 89–98. doi: 10.1111/j.1444-0938.1961.tb00396.x.

SunPy Community *et al.* (2015) 'SunPy — Python for solar physics SunPy — Python for solar physics', *Computational Science and Discovery*. IOP Publishing, 8, pp. 0–23. doi: 10.1088/1749-4699/8/1/014009.

Suresh, S., Dube, R. and Glenn, C. (2012) 'A Modified Shifted Means Based Segmentation Approach to Detect Active Regions and Coronal Holes in the Solar Dynamics Observatory Images', 8399, pp. 1–6. doi: 10.1117/12.919136.

Vita-Finzi, C. (2008) 'The Sun', in, pp. 25–42. doi: 10.1007/978-1-4020-6881-2.