MNRAS 000, 1–5 (2022) Preprint 23 September 2022 Compiled using MNRAS LaTeX style file v3.0

An Image Processing approach to identify solar plages observed at 393.37 nm by Kodaikanal Solar Observatory

Sarvesh Gharat,[1★] and Bhaskar Bose[2]
[1] Centre for Machine Intelligence and Data Science, Indian Institute of Technology Bombay, 400076, Mumbai, India
[2] Tata Consultancy Services, 560067, Bangalore

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Solar Plages are bright chromospheric features observed in Ca II K photographic observations of the sun. These are regions of high magnetic field concentration thus tracer of magnetic activity of the Sun and are one of the most important features to study long term variability of the Sun as Ca II K spectroheliograms are recorded for more than a century. . However, detection of the plages from century-long databases is a non-trivial task and need significant human resources for doing it manually. Hence, in this study we propose an image processing algorithm which can identify solar plages from Ca II K photographic observations. The proposed study has been implemented on archival data from Kodaikanal Solar Observatory. To ensure that the algorithm works, irrespective of noise level, brightness and other image properties, we randomly draw a samples of images from data archive to test our algorithm.

Key words: Sun: chromosphere – Sun: faculae, plages – techniques: image processing

1 INTRODUCTION

Solar Plages are the regions of high magnetic field concentration, which can help in tracing magnetic activity of the Sun (Shine & Linsky 1974) (Azariadis & Guesnerie 1986) (OLSON et al. 1978) (Neidig 1989) (Mackay et al. 2008) (Canfield et al. 2000)(Shine & Linsky 1972). Kodaikanal Solar Observatory has been actively observing Sun in Ca K wavelength since 1907 (Chatterjee et al. 2016) (Hasan et al. 2010), hence generating tremendous amount of data. It is almost impossible for professional solar astronomers to manually identify Solar Plages (Barata et al. 2018) (Benkhalil et al. 2003). Hence, in this study we propose an image processing algorithm to identify plages from Ca II spectroheliograms from Kodaikanal Observatory. The required data for this study has been collected from KSO data archive. To ensure that we cover all sort of images irrespective of varying brightness, noise, contrast and multiple image distortions, we randomly sample 900 + images from data archive.

Previously, lot of study has been done in identifying different solar features including sun spots, plages, filaments, etc (Barata et al. 2018) (Benkhalil et al. 2003) (Aschwanden 2010) (Qahwaji & Colak 2005) (Aboudarham et al. 2008) (Scholl & Habbal 2008). However, there's no exclusive study done for Kodaikonal observatory, hence we find a need to propose an image processing algorithm to automate this process. The data generated from this study along with some modifications can also be used to train a segmentation based algorithm such as U-net.

In (Barata et al. 2018) Barata et. al. make use of morphological transformations to segment data from Coimbra Observatory spectroheliograms. The authors make use of various transformations such as dilation, erosion and top hat before thresholding. However, considering the variation of data along with varying quality of images from KSO data archive, we can't make use of that algorithm in this case.

In (Benkhalil et al. 2003) Benkhalil et. al. make use of thresholding along with basic morphological operations, similar to (Barata et al. 2018) to detect active regions from Meudon Observatory in Hα and Calcium K images. Considering the difference in type of images and generalising for all 3 levels, all these algorithms won't be a good choice to use in case of Kodaikonal Observatory.

In (Aschwanden 2010) the authors provide an extensive review on use of Image Processing to identify different solar features along with the time dependency. The author's have also elaborated on use of Neural Networks to identify different features. However this study is limited to the features, wherein we have large amount of labelled data unlike plages. In (Qahwaji & Colak 2005) Qahwaji and Colak, focus on identification of plages along with providing a classification algorithm to differentiate plages and filaments. The proposed algorithm makes use of morphological classification along with hole filling on data from Meudon Observatory.

Similarly in (Aboudarham et al. 2008) (Scholl & Habbal 2008) the authors make use of different morphological transformations to identify multiple solar features.

However, as discussed earlier due to variation in image properties such as varying brightness, contrasts, artificial artifacts, etc we find a need to propose a novel algorithm to identify plages in generalised images (i.e all 3 levels are covered). We make use of OpenCv (Bradski 2000) to implement the proposed algorithm.

★ E-mail: sarveshgharat19@gmail.com

2 METHODOLOGY

This section focuses on the methodology proposed in this study.

2.1 Data Collection

An important part of any study involving data is the collection of data. In this study, we collect data from KSO data archive.
The collected data comprises of randomly sampled level 0 (raw), 1 (Basic Calibrated) and 2 (Limb Darkening Corrected) images with Ca K filter. More information on level of images can be viewed on official website of KSO data archive (https://kso.iiap.res.in/new/data).
The collected data is further pre-processed using multiple image processing algorithms before identifying plages. However, due to presence on uneven brightness, there are many outliers which are mistakened as plages. Hence to remove those, we make use of outlier detection techniques.
More details on the algorithm and the corresponding techniques is provided in subsequent subsections.

2.2 Algorithm

In this subsection, we elaborate the algorithm used in this study.
An overview of the same can be seen in Algorithm 1. The algorithm starts with:

Algorithm 1

```
Start
while Images are present do
    Median Blur with Kernel Size 3
    Image conversion to B/W image
    CLAHE with clip limit 3
    2 Iterations of Erosion
    1 Iteration of Dilation
    Thresholding with threshold of 180
    1 Iteration of Erosion
    1 Iteration of Dilation
    Find area of every bright spot
    Calculate mean area
    Remove all bright spots having area greater than 3 sigma
end while
Stop
```

2.2.1 *Median Blur*

The data due to increase in resolution and varying brightness level has a wide component of salt and pepper noise. Hence, to remove the same, we make use of median blur (Arce & McLoughlin 1987). Kernel Size (Horiuchi et al. 2019) acts as an important hyper parameter contributing to reduction of noise. Hence, to tune this hyper parameter we make use of hit and trial method. This filter is similar to other averaging filters. However, unlike average filters here we replace central pixel of kernel by median of pixels in the kernel area.

2.2.2 *CLAHE*

Histogram equalization (Pizer et al. 1987) is one of the most popular and widely used technique to improve contrast of any image (Abdullah-Al-Wadud et al. 2007). As pixel intensities surrounding the plages are not consistent, along with varying uneven brightness, use of global histogram equalization techniques is not possible. Hence in this study, we make use of Contrast Limited Adaptive Histogram Equalization (CLAHE) (Reza 2004) which extends the original technique by focusing separately on different regions within the image. Additionally, a clipping factor (Sundaram et al. 2011) (Liu et al. 2019) is also defined that clips the histogram at a certain point which prevents the over-amplification of noise as seen with a previous version. This clipping factor is one of the parameters that needs to be experimentally tuned to give the best possible results.

2.2.3 *Erosion and Dilation*

Erosion and Dilation (Soille 2004) are the basic transforms particularly used to enhance and remove small contrasting spots. Erosion is mainly responsible for removing the regions which are smaller than structuring element. Similarly, dilation joins two regions wherein the distance between those is less than or equal to structuring element. In this algorithm, erosion and dilation plays an important role in noise reduction

2.2.4 *Thresholding*

Image thresholding (Chowdhury & Little 1995) plays an important role in identifying plages. Using this technique, we segment the image by replacing all the pixel values greater than threshold with 255 and 0 for those less than threshold.
In our case, the threshold used is 180. A constant value of threshold works due to use of CLAHE (Reza 2004) which is one of the local histogram equalization technique as stated in previous subsection.
Please refer Algorithm 2 to get a rough idea on how the algorithm works.

Algorithm 2

```
Start
threshold = 180
for pixel in image do
    if pixel value > threshold then
        pixel value = 255
    else
        pixel value = 0
    end if
end for
Stop
```

2.2.5 *Outlier detection*

After thresholding and performing basic transformations (erosion and dilation), we get our segmented image. However due to artificial artifacts such as uneven bright spots, even these regions are segmented and can be misidentified as plages. Hence, to avoid this outliers (Ben-Gal 2005) we make use of Z-score (Aggarwal et al. 2019). In this method we first calculate area of all plages. Further based on mean and standard deviation of area of plages, we remove the regions which don't satisfy 3σ criteria. Hence, it ends up removing these outliers.

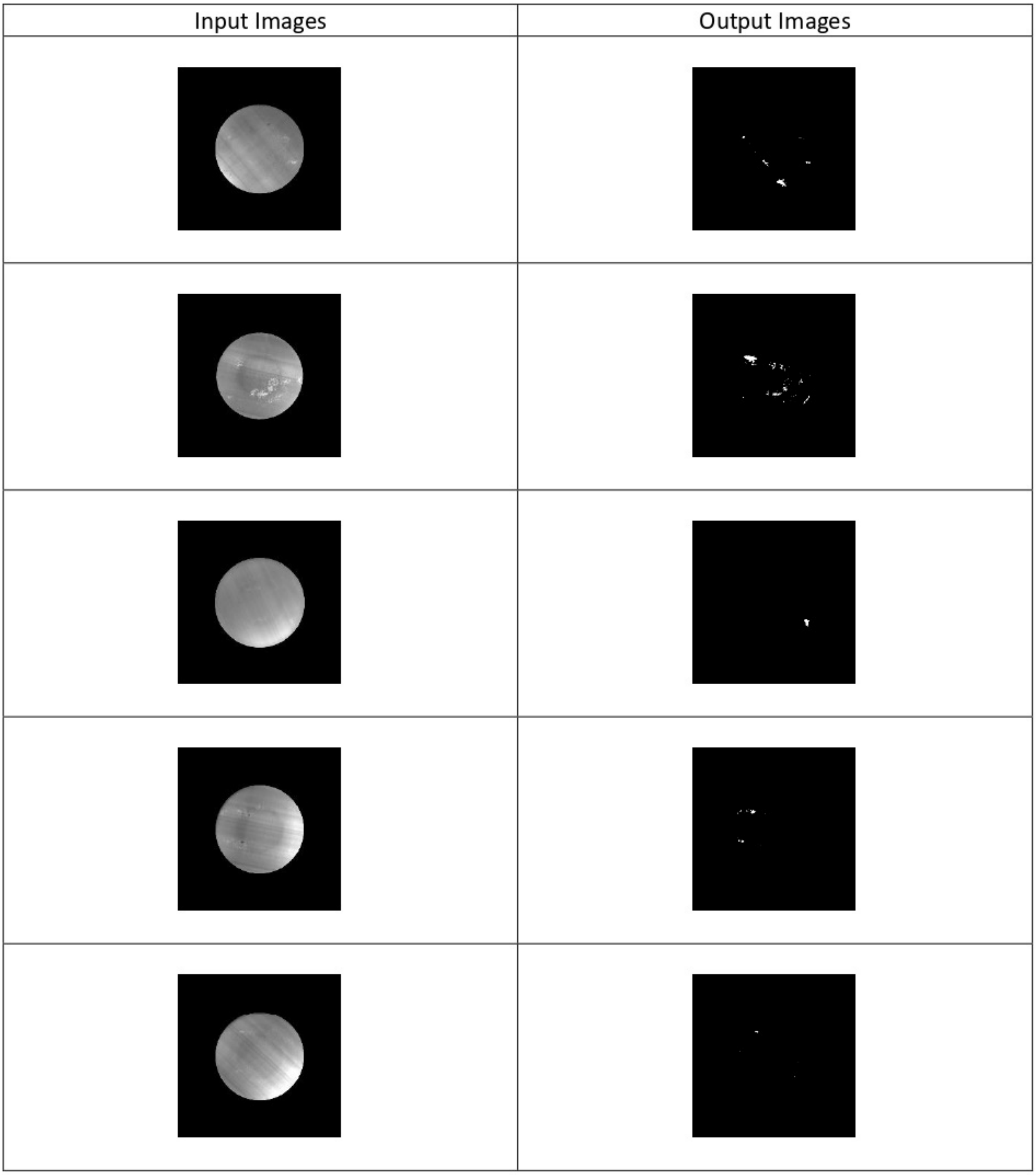

Figure 1. Randomly Sampled Input and Output Images

3 RESULTS AND DISCUSSION

In Figure 1 we see randomly sampled input and output images. The artificial artifacts along with varying brightness level can be seen in multiple images as opposed to the images which were used in previous literature. Hence, as discussed earlier there happens to be a need to have a generalized algorithm which can identify and segment plages.

Figure 2 represents output after every major transformation. From left to right we have input image → output of CLAHE → output of thresholding → output image. As we see in leftmost figure, we

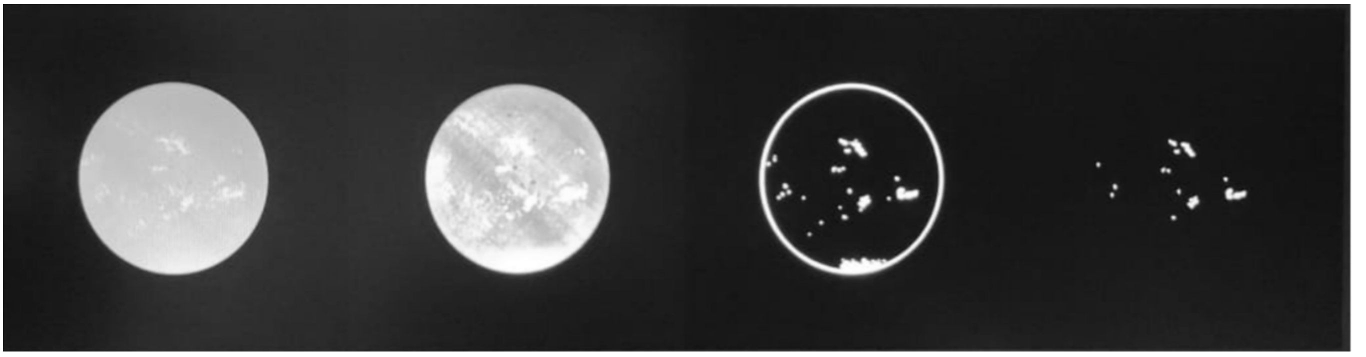

Figure 2. Output after every major transformation

can see that along with plages, there is a bright region at bottom. The second image which is output of CLAHE basically brightens all the white spots including the artificial artifact present in image. As CLAHE is a local histogram equalization technique, it ensures that there isn't much effect of varying uneven brightness present in image as seen in multiple input images in 1.
Next, we have thresholded image wherein all the pixels having pixel values less than 180 are replaced by 0, and remaining are replaced by 255. Doing CLAHE before thresholding helps us to decide with a constant value of threshold which wouldn't had been the case with global histogram equalization techniques. However, in 3^{rd} image we see that along with plages, the artificial artifact present in input image is also segmented. Hence, to remove that we make use of Z-Score outlier detection as discussed in above section. As seen in output image, use of outlier detection technique, helps us in removing the artificial artifact as seen in third image.

4 CONCLUSION

In this study we propose a generalised algorithm to identify plages in Calcium K images as observed by Kodaikanal Solar Observatory. The proposed algorithm works on all levels of images (i.e level 0, level 1 and level 2). This is one of the initial works done on automatic identification of plages on data from KSO data archive. Previously, this process was done manually after pre-processing the raw data. As our algorithm is generalized to work on any level of data, it also helps in omitting the initial preprocessing steps which involves correction of limb darkening, removing bias, etc.

ACKNOWLEDGEMENTS

We want to thank Kodaikanal solar observatory, a facility of Indian Institute of Astrophysics, Bangalore, India for making the archival data available. This data is available for public use at http://kso.iiap.res.in through a service developed at IUCAA under the Data Driven Initiatives project funded by the National Knowledge Network
Additionally, we also thank Mr. Abhimanyu Borthakur for his initial inputs and suggestions.

DATA AVAILABILITY

The input data used in this study is freely available in KSO data archive.
All the codes, used in this study are made available in our GitHub repository (https://github.com/SarveshVGharat/Plages-Identification). The output images generated using this algorithm will be made available on request after 2 years.

REFERENCES

Abdullah-Al-Wadud M., Kabir M. H., Dewan M. A. A., Chae O., 2007, IEEE Transactions on Consumer Electronics, 53, 593
Aboudarham J., Scholl I., Fuller N., Fouesneau M., Galametz M., Gonon F., Maire A., Leroy Y., 2008, in Annales Geophysicae. pp 243–248
Aggarwal V., Gupta V., Singh P., Sharma K., Sharma N., 2019, in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). pp 788–790
Arce G., McLoughlin M., 1987, IEEE transactions on acoustics, speech, and signal processing, 35, 60
Aschwanden M. J., 2010, Solar Physics, 262, 235
Azariadis C., Guesnerie R., 1986, The review of economic studies, 53, 725
Barata T., Carvalho S., Dorotovič I., Pinheiro F. J., Garcia A., Fernandes J., Lourenço A. M., 2018, Astronomy and Computing, 24, 70
Ben-Gal I., 2005, in , Data mining and knowledge discovery handbook. Springer, pp 131–146
Benkhalil A., Zharkova V., Ipson S., Zharkov S., 2003, in Proceedings of the AISB. pp 66–73
Bradski G., 2000, Dr. Dobb's Journal of Software Tools
Canfield R. C., Hudson H. S., Pevtsov A. A., 2000, IEEE transactions on plasma science, 28, 1786
Chatterjee S., Banerjee D., Ravindra B., 2016, The Astrophysical Journal, 827, 87
Chowdhury M. H., Little W. D., 1995, in IEEE pacific Rim conference on communications, computers, and signal processing. Proceedings. pp 585–589
Hasan S., Mallik D., Bagare S., Rajaguru S., 2010, in , Magnetic Coupling between the Interior and Atmosphere of the Sun. Springer, pp 12–36
Horiuchi K., Jiang W., Yamagishi S., Zhang X., 2019, in International Workshop on Advanced Image Technology (IWAIT) 2019. pp 7–12
Liu C., Sui X., Liu Y., Kuang X., Gu G., Chen Q., 2019, Journal of Modern Optics, 66, 1590
Mackay D. H., Gaizauskas V., Yeates A. R., 2008, Solar Physics, 248, 51
Neidig D. F., 1989, in , Solar and Stellar Flares. Springer, pp 261–269
OLSON R. H., ROBERTS W. O., PRINCE H. D., HEDEMAN E., 1978, Nature, 274, 140
Pizer S. M., et al., 1987, Computer vision, graphics, and image processing, 39, 355
Qahwaji R., Colak T., 2005, International Journal of Imaging Systems and Technology, 15, 199

Reza A. M., 2004, Journal of VLSI signal processing systems for signal, image and video technology, 38, 35
Scholl I. F., Habbal S. R., 2008, Solar Physics, 248, 425
Shine R. A., Linsky J. L., 1972, Solar Physics, 25, 357
Shine R. A., Linsky J. L., 1974, Solar Physics, 39, 49
Soille P., 2004, in , Morphological Image Analysis. Springer, pp 63–103
Sundaram M., Ramar K., Arumugam N., Prabin G., 2011, in 2011 International conference on signal processing, communication, computing and networking technologies. pp 842–846

This paper has been typeset from a TEX/LATEX file prepared by the author.