

# CLOUD SEGMENTATION AND CLASSIFICATION TO IMPROVE THE SHORT-TERM PV FORECASTING USING ALL SKY IMAGER

Dhia Eddine Melliti, Duy Long HA\*, Philippe Marechal, Sylvie Ghalila-Sevestre  
Univ Grenoble Alpes, CEA, LITEN, DTS, SMSP, INES, F-38000 Grenoble, France  
E-mail : Duy-Long.Ha@cea.fr

Jerome Lehaire, Thomas Capelle, Guillaume Tremoy  
Steadysun, 17 Avenue du Lac Léman, 73370 Le Bourget-du-Lac

**ABSTRACT:** Accurate prediction of photovoltaic energy remains a challenge, as PV production is dependent on fluctuating weather conditions, such as solar irradiance which relies on cloudy conditions. In this context, our work is based on a specific “All Sky Imager” (ASI), integrating “fish-eye” concave lenses with a 180° field of view.

This paper presents a novel approach of cloud segmentation using ASI to improve - PV production forecasting. First, we improve the identification of various components in the images (clouds, sun, noise, etc.). Then, we study the very short-term impact of clouds through a sky images segmentation and tracking processing. Finally, we correlate the sky-level segmented image to fluctuations in the actual on-site solar irradiance measurements.

The results show that the segmentation is efficient in clear and overcast sky conditions. However, high precision irradiance accuracy in partially cloudy sky conditions is difficult to obtain due to the chaotic impact of the circumsolar region and different cloud opacities.

Keywords: short-term forecasting; all sky imager; image processing; cloud segmentation

## 1 INTRODUCTION

Steadysun [1] develops and operates an advanced solar power production forecasting technology at intra-hour time horizon named Steady-Eye. This unique application delivers one of the most accurate forecasting solutions available on the market to anticipate solar energy variability. This work comes from a R&D cooperation within CEA-INES [2] to meet the needs of Steadysun to improve very short-term forecasting of PV production technology.

In this paper, we propose a segmentation algorithm to detect clouds on the images produced by Steadysun’s ASI. The algorithm is able to filter noise, flares around the sun. It also classifies clouds in two types: opaque and thin. These results will be used later to predict solar irradiance over a ten-minute to half-hour horizon by Steadysun.

## 2 METHODOLOGY

### 2.1 Camera calibration

The camera model is obtained beforehand with standard checkerboard based calibration method. This step is crucial as the camera model parameters are very important to understand the space as it is perceived by the camera then, transpose it (un-distortion) into the real world.



**Figure 1:** Checkerboard method for camera calibration

Scaramuzza et al. [3] toolbox is well known as a camera calibration tool. This toolbox uses black and white equal-square checkerboard images placed at different positions and tilt angles around the camera (figure 1). Using the known dimensions of the black and white square checkerboard pattern, the model estimates the desired parameters of the camera model.

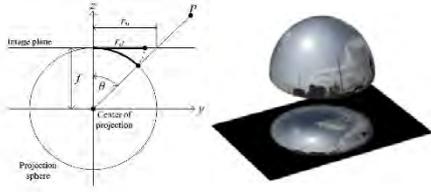
In this paper, we propose a different approach than the

one mentioned previously avoiding extra image acquisition. Indeed, this approach is based on the sun position to estimate the camera model parameters as mentioned in [4]. As a matter of fact, the sun position  $X_s=(x_s,y_s)$  (in pixels) is tracked on each image. In addition, the angular solar position  $\Phi_s=(\theta_s,\phi_s)$  is estimated using the NREL solar position algorithm [5]. The NREL algorithm takes the terrestrial coordinates (latitude, longitude, altitude) and time of the observation location as inputs, and outputs the solar zenith angle  $\theta_s$  and the solar azimuth angle  $\phi_s$ . Then, a linear transfer function between the solar coordinates  $X_s=(x_s, y_s)$  and the angular solar coordinates  $\Phi_s=(\theta_s, \phi_s)$  is calculated. In general, these parameters must be calculated for the camera at the calibration step and stored as initialization to the model to use them for later calculations.

### 2.2 Fish-eye lens model

Sky images in our model are acquired by a whole-sky imager installed locally at the research sites considered. The camera is located 200m from the solar irradiation measuring equipment. The imaging system consists of a high-resolution camera (3-megapixel RGB CMOS sensor) with an equisolid angle projection fish-eye lens (1.27mm focal length) attached to it, which provides a 187° Horizontal Field Of View (HFOV) and 168 ° Vertical Field Of View (VFOV). The optical system is enclosed in a waterproof box with a plastic dome above the lens. The box protects the camera from temperature, humidity, variations in wind and other environmental factors. Images are acquired at a resolution of (2048 x 1536) pixels and are saved at a sampling rate of one frame per minute.

Optical systems introduce some unwanted effects on the image, rendering the perception of the environment inaccurate. Generally, in most cameras, straight lines in the real world are mapped as straight lines in the image captured. However, the use of the fisheye lens introduces a radial distortion effect. This phenomenon is noticeable in the border region of the image, where the levels distortion level is extremely noticeable. This radial distortion causes pixels on the image plane to move from their ideal position in the standard camera model, along a radial axis in the fisheye image plane.



**Figure 2:** Geometric representation of the fish-eye model used and projection of the dome image on the plan

The model we used to simulate the behavior of the fisheye lens as a function of the focal length  $f$  of the camera and the angle  $\theta$ , inspired by [10], is given by the following equation:

$$r_d = k_1 * \sin(\theta/2) + k_2 \quad (1)$$

The relationship between the radial distance  $r_d$  from the center of the image and the angle  $\theta$  made by a pixel on the dome with respect to the vertical axes allows us to project a captured image onto the dome [6]. The parameters  $k_1$  and  $k_2$  of equation (1) were estimated using a linear regression in the practical case by calculating the apparent position of the sun on the test images using the polar coordinates  $X_{s,p} = (r_s, \phi_s)$ .

The last step in the calibration is to find the orientation of the camera relative to the north-south axis of the Earth. The parameters we considered are:

$\alpha$  : camera offset from the North-South axis.

$\gamma$  : azimuth angle between a point on the image and the North-South axis.

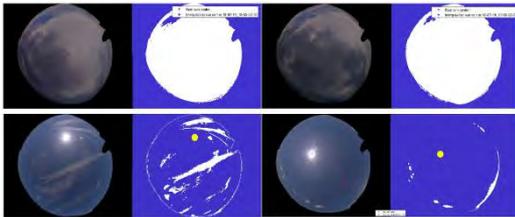
By calculating the coordinates  $(\phi_s, \gamma_s)$  of the sun, we will use these to find the offset of the North-South axis, through another linear regression of equation (2).

$$\gamma = \alpha - \phi \quad (2)$$

### 2.3 Clear Sky Library (CSL)

The binary segmentation is based on Clear Sky Library (CSL) similar to the approach described in [7, 8]. The CSL is a database that provides a reference with respect to the RBR (Red to Blue ratio  $R/B$ ) values of the RGB image for each pixel according to the zenith and azimuth angle of the image calculated from historical images in clear sky conditions.

The RBR level is higher in cloudy regions in the image, as well as in the solar and circumsolar regions. The RBR is the highest near the sun and decreases further away from the center of the sun. The RBR also increases near the horizon for large values of the zenith angle due to optical distortion and higher aerosol concentrations towards the horizon. The segmentation results based on the CSL and Otsu thresholding, which will describe in details in the next section, are presented in figure 3.

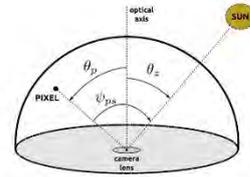


**Figure 3:** Segmentation results for days with clear and cloudy skies (INES)

For the construction of the CSL, a polynomial model using the RBR values of the clear sky images is estimated. The second degree polynomial model (9 coefficients) is calculated through the RBR image according to the zenith angle (in pixels) and the azimuth angle of the sun (in pixels) according to the following model:

$$RBR(\theta_p, \psi_{ps}) = \sum_{i=0}^2 \sum_{j=0}^2 p_{ij} * \theta_p^i * \psi_{ps}^j \quad (3)$$

For each clear sky image, the coefficients of the polynomial model (equation 3) are calculated and stored in the CSL with the corresponding zenith, azimuth angles and date/time of day. The representation of the angles  $\theta_p, \psi_{ps}$  and  $\theta_z$  is described in figure 4 with  $\theta_z$  the solar zenith angle, fixed for each image interpolation by the polynomial model.



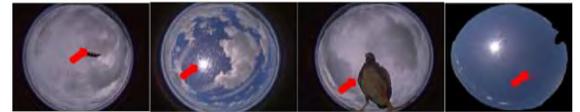
**Figure 4:**  $\theta_p, \psi_{ps}$  and  $\theta_z$  angles for the CSL model

## 3 IMAGE SEGMENTATION

This section will be reserved for the describing of the segmentation procedure for cloud detecting and the different techniques implemented ensure a good segmentation quality.

### 3.1 Image filtering

Before we start the segmentation process, we filter the dataset to remove outliers. Some examples of outliers are presented in figure 5.



**Figure 5:** Examples of anomalies in the images (insects, scratches, birds, flare)

The images with abnormalities are removed from the dataset through visual inspection or treated with specific algorithms, which will be described later on. Otherwise, they will affect the final cloud-detection decision image.

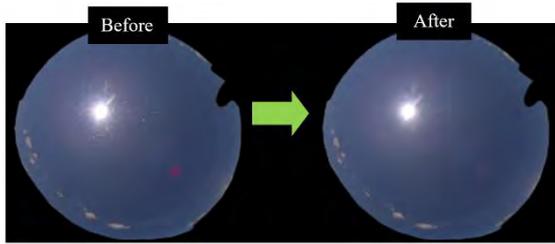
### 3.2 Image cleaning

The surface of the dome during real operational conditions is exposed to dirt or scratches which will result in additional scattering of light with specific patterns that change with the position of the sun and the time of day. Also, the reflections from the sun hitting the image sensor which become visible at high solar altitudes (for low values the zenith angle or the direct beam is almost orthogonal to the image plane) result in a red circular spot in the camera dome (flare). The impacts of this stray light were qualitatively assessed by visually inspecting clear sky images such as those in Figure 6.



**Figure 6:** Raw images artefacts and flares

These artefacts have consequences later on in the CSL construction and image segmentation. In fact, in the case of CSL, these spots present false high RBR values which biases the estimation of the polynomial model of the CSL. In addition, if left untreated, these spots will result in false detections later in the segmented image or be interpreted as false clouds.



**Figure 7:** Flares and artefacts elimination

To solve this problem, we apply a Gaussian filter with parameters  $\mu$  and  $\sigma$  in the specific regions where these artefacts and flares exist. An illustration of the resulting image after treatment is presented in figure 7.

### 3.3 Segmentation parameters

As previously exposed, the algorithm to detect clouds is based on the RBR. There are 4 cases where the detection procedure applied is different. Each of these cases is determined from the state of the sun (hidden, clear) and the circumsolar region (hidden, clear). Two different indicators are calculated, the Sun Disk Factor SDF and the Circumsolar Factor CF.

The circumsolar factor is defined as the ratio of clouds detected in the circumsolar zone compared to the total surface of the considered circumsolar region. The method used for detecting clouds in this region is based on the RBR of the considered image. According to the values of CF, we distinguish two cases:

$$\begin{aligned} \text{if } CF < \text{threshold}_{CF} : & \text{circumsolar clear} \\ \text{else} & : \text{circumsolar occulted} \end{aligned}$$

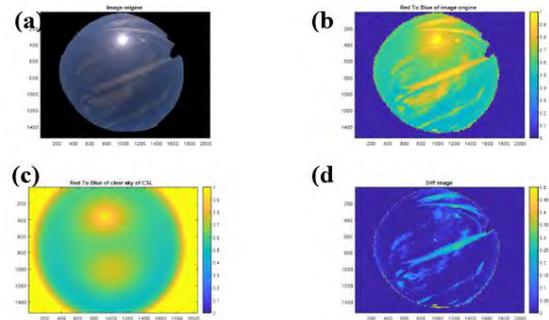
For the SDF, we calculate the area of the saturated region of the image ( $\text{Mean}(\text{RGB}) > 240$ ) in the solar perimeter (a region of radius 300 pixels for example). Then, we divide this area by a disk surface of radius 50 pixels (fixed sun radius). This will give us a variable solar ratio factor SDF that we will later use to determine if the sun is occulted.

$$\begin{aligned} \text{if } SDF < \text{seuil}_{SDF} : & \text{solar disk clear} \\ \text{else} & : \text{solar disk occulted} \end{aligned}$$

Considering a fixed sun radius simplifies the problem as the real sun radius in the image does not vary much in the considered region of interest of zenith angle  $[-85^\circ, 85^\circ]$ . A similar approach was proposed in [8] where this hypothesis was tested and validated.

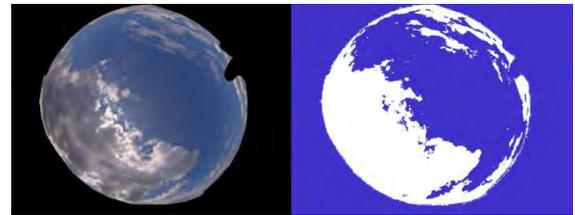
### 3.4 Segmentation steps

At this stage, we can now start the process of cloud detection. At first, we consider the RGB image in question (Figure 8-a). We start by eliminating the artefacts and flares from the image as mentioned in (4.1) and (4.2). After this step, considering the position of the sun in the image, we calculate the corresponding zenith and azimuth angles. Then, we look in the CSL data base (which we prepared beforehand) for a similar day (or closest) in terms of zenith and azimuth angles. The RBR image from the database for the corresponding values of zenith and azimuth is then generated using formula (3) and the polynomial coefficients found in the CSL (Figure 8-c). After subtracting the RBR of the CSL image (Figure 8-c) from the RBR of the raw image (Figure 8-b), we obtain a difference image of the RBR values (figure 8-d).



**Figure 8:** Cloud segmentation steps with CSL

By comparing the pixels of the difference image (figure 8-d) to an Otsu threshold calculated through the whole difference image, the final resulting image which indicates the presence of clouds or not, is obtained. If the pixel values in the difference image is above the Otsu threshold, the method classifies these pixels as cloudy ones which indicates the presence of clouds in that region. Otherwise, the pixels are bright (no clouds).



**Figure 9:** Cloud Detection with fixed global threshold

However, inspired by [8], we introduce a subtle correction in the form of an exponential Otsu threshold to improve detection and avoid misclassification especially in the closest vicinity of the sun (circumsolar region), which affects the very short-term solar forecasting in general and specifically in our model. The exponential Otsu threshold proposed in [8] is in the form:

$$\delta_{diff}(\psi_{ps}) = \delta_{diff,0} \left( 1 + \exp\left(\frac{-\psi_{ps}}{200}\right) \right) \quad (4)$$

With  $\psi_{ps}$  in degrees,  $\delta_{diff,0}$  is the Otsu threshold of the difference image and 200 is a scaling factor, which has been fixed through manually testing a set of batch images. Using the same logic as before, we then compare the calculated  $\delta_{diff}$  with the the difference image by classifying the pixel as cloudy if the pixel value is above  $\delta_{diff}$  and clear otherwise. However, we have exploited this result but with some modifications and improvements described in the following table:

**Table 1:** Different threshold settings for cloud detection

Case	Sun	Circumsolar	CSL	Threshold	Expression
1	Occluded	Occluded	No	Fixed threshold	Threshold = 0.7
2	Clear	Occluded	yes	$\delta_{diff,0} = \text{Threshold}_{\text{Otsu}}$	$\delta_{diff}(\psi_{ps}) = \delta_{diff,0}(1 - \delta_{diff,0} * \exp(\frac{-\psi_{ps}}{200}))$
3	Occluded	Clear	yes	$\delta_{diff,0} = \text{Threshold}_{\text{Otsu}}$	$\delta_{diff}(\psi_{ps}) = \delta_{diff,0}(1 + \exp(\frac{-\psi_{ps}}{200}))$
4	Clear	Clear	yes	$\delta_{diff,0} = 0.17$	$\delta_{diff}(\psi_{ps}) = \delta_{diff,0}(1 + 0.5 * \exp(\frac{-\psi_{ps}}{200}))$

These settings are site independent and are the same for all studied sites.

## 4 RESULTS AND VALIDATION WITH FIELDS TESTS

### 4.1 Cloud detection validation

The first validation step is visual. In fact, we inspect and compare the segmented images with the real images of the camera in order to check the quality of cloud detection especially in the circumsolar region. This first validation is carried out on the entire database available on the various studied sites.

Second, once the clouds are detected, we verify whether the sky situation described by the mask generated by our model is in accordance with the irradiance measured on the ground. In other to confirm the matching of generated image and the actual GHI measurements of the site, we apply a certain logic described by the following table:

**Table 2.** Comparison logic of the GHI measurements to the model

Case	Sun	Circumsolar	Condition	Solar irradiance
1	Occluded	Occluded	$G_m \leq G_{cs}$	Occluded
2	Clear	Occluded	$G_m > G_{cs}$	Diffused
3	Occluded	Clear	$G_m < G_{cs}$	Occluded
4	Clear	Clear	$G_m = G_{cs}$	Transmitted

$$G_m: \text{Irradiance measured} - G_{cs}: \text{Irradiance in clear sky condition}$$

In the case of occluded sun, the measured value should be lower than the theoretical value of the clear sky model. Also, if the sun is clear, the location of the circumsolar zone determines whether the irradiance reaching the sensor is transmitted or diffused. In the case of transmission, the measured irradiance is equal to that of clear sky conditions, otherwise in the case of scattering, the measured irradiance is much higher.

To validate these hypotheses, we compare in another step the values of irradiance measured on site to those of clear sky irradiance from a model, which generates these values taking into account the longitude, altitude and latitude of the site [9]. The comparison is made according

to the formula:

$$Diff_{irradiance}(\%) = \frac{\text{Irradiance}_{\text{measured}} - \text{Irradiance}_{\text{clear sky}}}{\text{Irradiance}_{\text{clear sky}}} \quad (5)$$

After that:

If  $|Diff_{irradiance}| < 12\%$  : sun and circumsolar clear

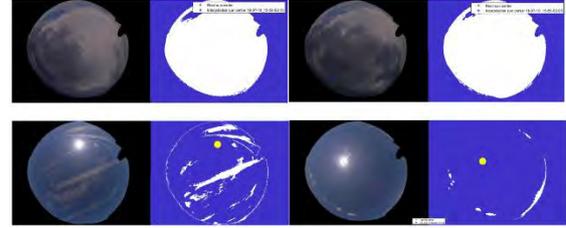
If  $Diff_{irradiance} < 12\%$  : sun occulted

else : sun and circumsolar occulted

This logic will be used subsequently for the validation of the segmentation results.

### 4.2 Results for cloudy and clear sky days

The validation of the segmentation quality was first done visually (Figure 12) and on the entire database of test sites.



**Figure 12:** Segmentation results for overcast and clear sky days.

The model shows good performance for days with overcast and clear sky. First, by eliminating flares and artefacts around the sun thoroughly, secondly in terms of the deployment of the CSL and its usefulness for days with clear sky conditions and thirdly, the effectiveness of the fixed threshold for overcast skies (Table 1 - Case 1) and the adaptive threshold (Table 1 - Case 4) especially in the circumsolar region which mainly impacts the GHI forecast in the very short term.

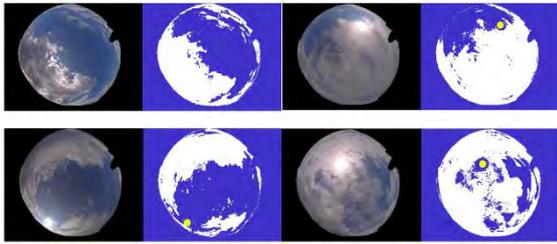
For the GHI validation of our cloud detection method, we only have GHI measurements for the INES site. The GHI validation results are summarized in the following table:

**Table 3:** GHI validation results for clear and overcast days

Site	Data available	Clear sky	Overcast	Precision [%]
INES	2018 (20 days over 12 months)	2 days	13 days	Clear sky 87 - 98
				Overcast 98 - 100

### 4.3 Results for days with partly cloudy skies

The proposed model is more beneficial to be applied in partially cloudy conditions. In fact, the variations of measured GHI are very fluctuating and thus intriguing in this case. This underlines the importance following the evolution of clouds especially in the circumsolar region, to better understand and correlate the impact of clouds on very short-term GHI fluctuations. A better understanding of this phenomenon will subsequently make it possible to provide a better, reliable forecast of solar irradiance.



**Figure 13:** Segmentation results for a partly cloudy sky (INES)

The results for the partly cloudy data set are displayed in table 4, as well as for the visual validation part in Figure 13.

**Table 4.** GHI validation results for partially cloudy days

Site	Data available	Partly cloudy	Precision [%]
INES	2018 (20 days over 12 months)	5 days	51 - 71

The model accuracy in partially cloudy days (51% - 71%) versus clear and overcast ones (87% - 98%) and (98% -100%) respectively ) illustrates the variability of cloud impact on GHI measurements and the difficulty to pinpoint with precision it's impact on all the data used, in the case of partially cloudy days. This large variability leads to greater uncertainties later on, over the assessments made on the future value of GHI. This is due to the different types of clouds (opaque, thin, permeable ... etc.) as we have concluded.

The validation using in-site irradiance measurements shows expectable results for clear sky and cloudy conditions. However, the segmentation should be improved specially for partially cloudy days, which will be the focus of our future work.

#### 4 CONCLUSION

In this paper, we have presented method for cloud detection and segmentation using preexisting techniques, which we have adapted locally for our study site and improved, based on data driven approaches.

The validation of our method shows good results for clear sky and overcast days. For partially cloudy sky conditions, there is still some improvements to work on to distinguish between different types and shapes of clouds. Indeed, based on the results obtained on partially cloudy days, we noticed that the opacity of the clouds approaching the circumsolar region and the sun in particular, influence the transmissibility of solar radiation to different degrees. The impact is measurable as soon as the clouds cross the circumsolar region, where measurements start to become chaotic. This is why a classification of the clouds according to the opacity level is essential to more precisely distinguish the transmissivity of clouds and subsequently, to increase the precision of the model.

One approach is a cloud classification procedure based on sky images histogram analysis. The impact of cloud opacity on irradiance transmissivity will be developed in future works. Another perspective is to distinguish between different types of clouds by a classification procedure based on color (RGB) and texture (gradient for

example). This step should help identifying the impact of clouds on solar irradiance and subsequently, improve the forecast.

Finally, the main finding is that there is no "one best model" for predicting solar irradiance, but some models give better results than others do in specific sky and meteorological conditions. It would be interesting to find a way to hybridize the different models we have according to the forecast horizon, different site conditions and optical systems in order to try to keep the forecast error as low as possible.

#### 6 REFERENCES

- [1] <https://www.steady-sun.com/>
- [2] <https://www.ines-solaire.org/>
- [3] D. Scaramuzza, A. Martinelli, and R. Siegwart, 'A Toolbox for Easily Calibrating Omnidirectional Cameras', in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, Oct. 2006, pp. 5695–5701, doi: 10.1109/IROS.2006.282372.
- [4] B. Urquhart, B. Kurtz, and J. Kleissl, 'Sky camera geometric calibration using solar observations', Atmospheric Measurement Techniques (Online), vol. 9, no. 9, Sep. 2016, doi: 10.5194/amt-9-4279-2016.
- [5] I. Reda and A. Andreas, 'Solar Position Algorithm for Solar Radiation Applications (Revised)', NREL/TP-560-34302, 15003974, Jan. 2008. doi: 10.2172/15003974.
- [6] C. Hughes, P. Denny, E. Jones, and M. Glavin, 'Accuracy of fish-eye lens models', Applied optics, vol. 49, pp. 3338–47, Jun. 2010, doi: 10.1364/AO.49.003338.
- [7] J. Du, Q. Min, P. Zhang, J. Guo, J. Yang, and B. Yin, 'Short-Term Solar Irradiance Forecasts Using Sky Images and Radiative Transfer Model', Energies, vol. 11, p. 1107, May 2018, doi: 10.3390/en11051107.
- [8] M. Caldas and R. Alonso-Suárez, 'Very short-term solar irradiance forecast using all-sky imaging and real-time irradiance measurements', Renewable Energy, vol. 143, pp. 1643–1658, Dec. 2019, doi: 10.1016/j.renene.2019.05.069