Internet of Things Platform for Photovoltaic Maintenance Management: Combination of Supervisory Control and Data Acquisition System and Aerial Thermal Images

Isaac Segovia Ramirez¹,* and Fausto Pedro García Márquez¹,**

¹Ingenium Research Group, Universidad Castilla-La Mancha, 13071 Ciudad Real, Spain

Abstract. Suitable maintenance management plants of solar photovoltaic plants are required for global energy demands. The volume and variety of data acquired by thermographic cameras carried by unmanned aerial vehicles and Supervisory Control and Data Acquisition Systems increase the complexity of fault detection and diagnosis. The maintenance industry is requiring novel fault detection techniques that can be implemented in Internet of Thing platforms to automate the analysis and increase the suitability and reliability of the results. This paper presents a novel platform built with PHP, HTML, CSS and JavaScript for the combined analysis of data from Supervisory Control and Data Acquisition Systems and thermal images. The platform is designed. A real case study with thermal images and time series data from the same photovoltaic plant is presented to test the viability of the platform. The analysis of thermal images showed a 97% of accuracy for panel detection and 87% for hot spot detection. Shapelets algorithm is selected for time series analysis, providing an 84% of accuracy for the pattern selected by user. The platform has proven to be a flexible tool that can be applied for different solar plants through data upload by users. Keywords: Maintenance management, Machine learning, Photovoltaic, IoT platform.

1 Introduction

New energy sources and maintenance advances are required to achieve current global energy demands and decrease high global electricity prices. Renewable energies are expected to play a major role due to high CO₂ levels and the effects of climate change, with an exponential growth in the coming years, accounting for almost 95% of the increase in global electricity capacity through to 2026 and representing more than 80% of global energy generation in 2050 [1].

The price of large-scale photovoltaic (PV) solar power today is currently cheaper than conventional power generation methods, being one of the most relevant renewable energies. Despite the overall increment of prices for silicon and PV cell and modules, PV solar has presented an important growth in 2022. Three markets will produce more than 20 GW in 2023: China with 94.3 GW, United States with 37.4 GW and India with 20.6 GW, accounting for the 55% of the global PV production between 2024-2026, see figure 1, and it is expected to reach 1583 GW in 2030 [2, 3].

*e-mail: isaac.segovia@uclm.es
**e-mail: faustopedro.garcia@uclm.es

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
The efficiency of PV panels is affected by the connection of cells and hard environmental conditions that may cause different faults and degradation mechanisms, e.g., deposition of dust, corrosion, delamination, cracks or hot spots, among others, reducing operating performance of PV panels around 20% [4–6]. Fault detection and diagnosis (FDD) techniques are applied to develop preventive and predictive maintenance plans, avoiding PV performance reduction with lower operation and maintenance (O&M) costs [7, 8]. Supervisory Control and Data Acquisition (SCADA) acquires data, e.g., irradiation, power production, performance ratios, voltage and current, among others, from different types of condition monitoring systems (CMS) about the real state of PV panels [9]. SCADA data is considered a time series and several statistical analysis techniques are proposed for fault detection and pattern recognition. Zhao et al. [10] developed different rules for outlier detection based on 3-sigma, Hampel identifier and boxplots. Yang et al. [11] considered the irradiance data as time series data. The authors applied principal component and biplot for outlier detection and string fault detection, demonstrating that these techniques can be implemented in real-time analysis. Betti et al. [12] presented a key performance indicator for prediction of generic fault detection, achieving an accuracy of 95%. However, statistics-based methods produce high false alarm rate due to the application of adaptive thresholding [13, 14] and different surface defects, e.g., the influence of dirt or partial reductions in energy production, are usually undetected by SCADA systems. For these reason, novel inspection methods with new CMS are required [15, 16].

Infrared thermography provides temperature data through the transformation of the acquired infrared energy emitted by bodies. Thermal cameras provide images with a predefined resolution where the temperature value of each pixel is determined. However, this technique has a low acquisition range in PV maintenance, and it is usually implemented by technicians. This technology is integrated in Unmanned Aerial Vehicles (UAVs) to increase the data acquisition rate and monitor several PV panels in the same thermal image. The implementation of thermography and UAVs is widely applied in the current state of the art [17]. Addabbo et al. [18] presented a review about the state of the art of UAVs in PV plants, analysing different systems and configurations. Niccolai et al. [19] developed an advanced system to digitalize PV plants with light UAVs to increase the reliability of O&M tasks. Overall aerial PV measurements were presented in reference [20] to increase the reliability of the measure-
ment process and optimize the flights. It has been demonstrated that the UAV positioning
system has a relevant influence in the measurement process of PV panels, being Real Time
Kinematic one of the most reliable positioning system [21].

The analysis of images acquired by aerial inspections with thermal cameras is a challenge
in the current state of the art, being necessary new reliable techniques for object detection
and region identification that allow the detection of hot spots, delamination or faults with
visual patterns. Machine Learning (ML) algorithms have emerged as reliable solutions for
fault detection in images to detect Regions of Interest (ROIs) [22]. Artificial Neural Net-
works (ANNs) are novel processing systems based on biological nervous systems applied
for classification and feature extraction, being capable of learning with complex nonlinear
input-output connections through training processes. ANNs are defined by the combination
of different neurons with specific activation functions, being widely applied for forecasting of
solar radiation, modelling and PV performance evaluation achieving accuracies higher than
99%. However, this algorithm is the most applied technique for image processing where
bounding boxes define ROIs and classify PV panels into two classes: healthy and damaged.
Convolutional Neural Network (CNN) is one of the most relevant and applied neural network
for object detection in the current state of the art [23]. CNNs are widely applied for the class-
ification and detection of ROIs in images due to high accuracy around 90%. Region-CNN
(R-CNN) proposes 2000 candidate regions selected by a selective search algorithm based on
visual parameters, e.g., size, colour intensity, distribution, etc. The two main phases of R-
CNN are the proposal of regions and classification process, being defined by different types
of layers, e.g., convolutional, fully connected, pooling, among others [24]. Faster-RCNN
algorithm improve the performance of R-CNN avoiding selective search and applying an ad-
ditional network to propose regions, reducing the computational costs and operating periods.
Huerta et al. [25] combined several C-NNs for hot spot detection The approach was validated
in a real PV plant with 92% accuracy. Pathak et al. [26] proposed a fault detection approach
with R-CNN for aerial images. The results were compared with different techniques, such as
statistical analysis or support vector machine, and R-CNN obtained the best results with 99%
accuracy [26]. It is concluded that automatic fault detection in PV panels with ANNs is still a
complex and computationally expensive task that requires new advances and improvements.

Internet of Things (IoT) is an advanced system based on an open-source network with
the ability of acquiring large volume and variety of data in real time from intelligent sensors
or machines. IoT allows a comprehensive range of opportunities in several industries, e.g.,
healthcare, security, production management, among others, through data visualization and
cloud computing. The main challenges of this technology are energy optimization of devices,
tracking resources, data management, privacy and security, among others [27]. The combina-
tion of SCADA data in PV plants and IoT platforms has been studied by several authors. Qays
et al. [28] implemented a SCADA system into IoT platform for PV panels to monitor electrical
performance. The data was visualized in ThingSpeak open-source platform. Garcia et al.
[29] presented a SCADA monitoring architecture with low-cost IoT sensors. The authors also
implemented a predictive fault diagnosis method based on the analysis of current and voltage
parameters, and different scenarios simulating faults were presented, achieving suitable re-
sults Garcia and Segovia [30] developed and IoT structure to analyze infrared data carried by
UAV. The objective is dust detection, where the healthy and dirty states are detected with high
accuracy. The same approach was applied in reference [31] for dust detection. Aghenta and
Iqbal [32] designed a Thinger.IO local IoT platform for SCADA data analysis with current
and voltage parameters, defining the architecture and showing the graphical possibilities for
data visualization.

The main novelties of this paper are resumed as follows:
• The combination of ML techniques for object detection in images and time series analysis algorithms of SCADA data for fault detection and identification. Both methodologies are widely implemented separately, and the is a novelty in the state of the art. Therefore, the implementation of shapelets algorithm or ANNs is not the main novelty, but on the whole approach that analyzes SCADA and visual patterns associated to failures in PV panels.

• The imaging detection container combines two sequential networks, the first one initially detects panels and subsequently, the second network performs hot spot detection. This methodology programmed in Python ensures high reliability and avoid the definition of wrong regions as hot spots.

• The main novelty is implementation of both techniques in an IoT platform constructed on containers with different functions, to increase the suitability and effectiveness of the analysis. This is a relevant novelty for PV maintenance management, and only limited number references have been obtained.

This paper is organized in the following sections: Section 2 presents the main basis of approach, combining IoT platform requirements and architecture of the ANNs; a real case study is developed in Section 3, where SCADA data and thermographic images have been analyzed to determine the reliability of the IoT platform; Section 4 summarizes the conclusions.

2 Approach

The methodology presented in this work develops a novel IoT platform that combines the analysis of thermal images acquired by UAVs for hot spot detection and SCADA data in PV solar plants. The interface of the platform is established with CakePHP and HTTP, and the main structure of the platform is formed by three independent containers with different capabilities and functions. Python is one of the most applied and suitable programming languages, and all the algorithms are developed in this language [33]. The users can upload CSV files and thermal images to create MySQL databases that will be accessed for further analysis. The visualization of SCADA data and thermal images is performed by Python and JavaScript. The diagram of the IoT platform is shown in figure 2, where the connection between containers and the requirements of each phase are defined.

The user container controls the access, security levels of the platform and the system interface with the HTTP requests, see figure 2 (colour red). It contains HTML, CSS and JavaScript code that provides structure and format to the platform. The web interface allows the configuration of the type of analysis (hot spot or SCADA), and the definition of the dataset for further processing. The platform is designed to efficiently manage datasets, being possible to upload a directory of images or CSV data saved in the platform to perform the analysis several times. The selection of the user is transferred as an HTTP query in form of event that defines the container and the algorithms for the analysis. A queue control system visualized with RabbitMQ is also included to manage the processing periods and avoid the overload and blocking of the platform.

The hot spot container detects faults in thermal images as input dataset, and it is determined the connections and operations for hot spot detection, see figure 2 (colour blue). The fault detection process has been divided into two stages: the first stage is focused on panel detection while the second one uses the detected panels as inputs for the recognition of hot spots, ensuring that the identified faults are on the PV panels and reducing the number of false hot spots. The training and labelling of panels and hot spots for both ANNs are externally performed and implemented on the platform to simplify data processing and reduce computational costs. It is important to ensure that the training has been performed with a certain number of hot spots to increase the reliability of the analysis. The final output of the platform
for this container is the information on the number of hot spots, overall accuracy, location of each spot and the thermal image with bounding boxes for both hot spots and panels.

SCADA container applies series methodologies with SCADA input dataset in CSV format, as it is observed in figure 2 (colour green). Several types of time series algorithms can be applied but for this particular case study, it is proposed a time series analysis with low computational costs based on shapelets. As input for the analysis, it is required the definition of a period time where the algorithms will search for similar patterns.

2.1 Time Series Data Analysis

A shapelet is a time series subsequence that describes a class. Shapelets algorithm is a data mining technique applied for time series classification with high reliability for the identification of patterns. The definition of the interest period is required to extract the main characteristic of the pattern, and the shapelets extracted from the training are applied for the recognition of shapelets that is present in most of the time series data. Pearson correlation coefficient is applied to obtain a training dataset of labelled normal and abnormal curves. The quantification of the similarity between the shapelet and the pattern is applied with the distance between them.

The Euclidean distance described in Eq. (1) is widely implemented, being \( x_{(t,t+l)} \) the subsequence of the pattern \( x \) at time index \( t \).

\[
d(x, s) = \min_t \| x_{(t,t+l)} - s \|_2.
\]
2.2 ANNs for Object Detection in Thermal Images

R-CNN creates candidates to be ROIs analyzing different main characteristics, e.g., form, distribution or type of colours. This algorithm reduces the number of image regions by employing an additional ANN to predict ROIs, obtaining reduced performance periods, leading to suitable implementation in online applications. The sliding window is a 3×3 window that can navigate within the feature map to perform an evaluation of a set of predefined bounding boxes with changing sizes known as anchor boxes. The goal is the minimization of CNN training. R-CNN uses a 3×3 sliding window that allows a reliable navigation into the feature map to estimate predefined bounding boxes with variable sizes.

The training is a critical phase that requires a minimization of the loss function $L$ through continuous adjustments of weight between layers, as it is shown in Eq. (2).

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{loc}(t^u, v). \tag{2}$$

Being $p$ and output vector, $u$ provides the labels, $t^u$ the decision parameter and $v$ the vector of transformation. The classification losses $L_{cls}(p, u)$ are shown in Eq. (3) and $L_{loc}(t^u, v)$ is determined in Eq. (4) is the loss function of the candidate box.

$$L_{cls}(p, u) = -\log(p_u). \tag{3}$$

$$L_{loc}(t^u, v) = \sum_{i \in x,y,w,h} S_L(t^u - v_i). \tag{4}$$

3 Case Study

3.1 Introduction of the Case Study

The case study is formed by thermal images acquired by UAV and SCADA data from the same real 8MW PV plant to test the reliability of the platform. The operators of the PV plant are not currently implementing any combination of these two types of data, demonstrating the novelty of the methodology developed in this work. The image dataset is formed by 20 images with a resolution of 640×512 pixels for testing phase and 80 images for validation. The time series dataset consists of more than 300 records because it is focused only on the same month when the image dataset was acquired in order to synchronize the results. The power generation of the plant is the selected variable for the study. The platform is located in a workstation with i7 CPU and 32GB RAM for performing both analysis.

3.2 Hot Spot Detection in Thermal Images

Faster-RCNN has been selected for both panel and hot spot detection because of short computational times and high accuracy, although it is possible to upload any type of neural network. It is considered that the images were acquired at the same aerial condition and the PV panels do not present variations. The training is performed externally to the platform and uploaded to the platform directory, ensuring that the losses are less than 1. The IoT platform performs the analysis with the previously determined image dataset, providing the thermal images with
Figure 3. Results of the IoT platform: thermal images with the boundary boxes and possibility of download the results in CSV format

Table 1. Confusion matrix for hot spot detection

<table>
<thead>
<tr>
<th>Real Class</th>
<th>Healthy</th>
<th>Faulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Faulty</td>
<td>12</td>
<td>52</td>
</tr>
</tbody>
</table>

the boundary boxes, and the results in a CSV file with the number of panels and host spots detected, the score of the detection process and the GPS position of the image, as it is observed in figure 3.

The objective of this section is to test the applicability of this type of ANN on the IoT platform. The testing of 20 different images for this case study has shown that this network has an average execution period of 140 seconds with an average reliability of 97% for panels and 87% for hot spots, although different hot spots were detected with accuracies higher than 93%. The confusion matrix of hot spot detection is used to contrast all the results achieved by Faster-RCNN with the real classes for this type of PV plant, see table 1. Two classes are defined based on PV panel condition: healthy with no faults and faulty with active hot spots.

Table 1 show the results for panel and hot spot detection performed in the IoT platform.

The results provided by the IoT platform shows that 52 hot spots were accurately detected, demonstrating the reliability of the training. There are no false positives and, therefore, none of the healthy regions were classified as faulty. However, the network did not detect 12 regions that actually had a hot spot, and it will be required further improvements in the trainings.

As future work, it is proposed more case studies with real data from PV plants to test the reliability of the methodology. It is also proposed the connection between the SCADA data and the hot spot detection to obtain a robust tool with higher accuracy in the PV maintenance management.

3.3 Shapelets for SCADA Data

The Shapelets implementation process is divided into different sub-processes designed in interactive specific scripts with the required libraries. The first step is the definition of the
A shapelets model is used to prepare the data for the correct operation of the shapelets algorithm using different libraries. The input data is loaded in the IoT platform by the user, and for this case study, the energy production data provided by SCADA of the PV plant is analyzed. The next phase is the definition of the reference pattern by selecting a time range with incidents, failures, or maintenance problems. This information is obtained from the maintenance data provided by the operator. This period is also defined by the user in the IoT platform, see figure 4(a), and the platform generates the pattern, as it is observed in figure 4(b). Shapelets produces sub-time series of the same length as the pattern from the time series data, classifying the patterns as similar or not similar with respect to the reference pattern using the Pearson’s correlation coefficient. Fig. 4(c) shows several candidates of failures detected by the algorithm.

Figure 4. Results provided by the shapelets algorithm (a) Energy production of the selected day; (b) Pattern associated to failure; (c) Candidates detected by the algorithm using correlation coefficient.

The training defined by the user generates a shapelets model that can be applied for the detection of previously determined patterns. For this case study, the shapelets model defined in the IoT platform in this execution provided more than 20 candidates with an accuracy of 0.838 with a standard deviation of 0.019, demonstrating the reliability of the approach. All the results can be downloaded from the platform in CSV format. The detected patterns by the shapelets model are compared with O&M operations to determine the real cases of the decrease in energy generation. It is determined that these candidates are clearly associated to string faults in specific inverters.

4 Conclusions

The maintenance of solar photovoltaic plants requires new fast and suitable tools that allow the use of advanced algorithms to reach the competitiveness in global energy market. Supervisory Control and Data Acquisition systems provide data acquisition related to energy consumption, performance ratios or irradiation, among others, although it is not possible to detect surface defects, e.g., dirt or hot spots. Infrared thermography with drones is an advanced methodology, although this technology also has limitations, requiring the application of advanced algorithms with complex trainings for object detection. This paper proposes a robust Internet of Thing platform based on the combination of artificial neural networks and analysis of time series data. The novelty lays in the combination of two types of algorithms in Python: shapelets algorithms for time series analysis and Faster-Recurrent Neural Network for thermal images, being implemented in an online platform defined with CakePHP and HTTP. It is presented a real case study that analyzes aerial images and time series data from the same photovoltaic plant. Aerial thermograms analysis provided a 97% of accuracy for panel detection and 87% for hot spot detection. Shapelets algorithm analyses time series data.
data for pattern detection with an accuracy of 84%. The platform designed for this work has proved to be a robust tool for suitable photovoltaic maintenance management.

The work reported herewith has been financially supported by the Spanish Ministerio de Economía y Competitividad, under the Research Grant RTC2019-007364-3.

References