

# Predicting solar irradiance with all-sky image features via regression

Chia-Lin Fu<sup>a</sup>, Hsu-Yung Cheng<sup>b,\*</sup>

<sup>a</sup> Industrial Technology Research Institute, No. 195, Sec. 4, Chung Hsing Rd., Chutung, Hsinchu 31040, Taiwan

<sup>b</sup> Department of Computer Science and Information Engineering, National Central University, No. 300 Zhongda Rd., Zhongli City, Taoyuan County 32001, Taiwan

Received 19 March 2013; received in revised form 28 August 2013; accepted 8 September 2013

Available online 2 October 2013

Communicated by: Associate Editor Christian Gueymard

## Abstract

To address the problem of forecasting solar irradiance for grid operators, the aim of this work is to automatically predict solar irradiance several minutes in advance. This work presents a solar irradiance prediction scheme that utilizes features extracted from all-sky images. To select a proper feature subset for prediction, various features are analyzed and compared. We propose to utilize the regression technique to predict clearness index and then to calculate the desired solar irradiance from the predicted clearness index. We validate the effectiveness of the proposed scheme using a challenging dataset collected at a coastal site. The experiments have shown that the designed clearness index prediction mechanism yields better prediction results than predicting solar irradiance directly. Also, irradiance prediction at 5 min in advance can be achieved with mean absolute error of around 22%. The results of this work could provide very useful information for grid operators to ensure greater efficiency of the renewable energy supply.

© 2013 Elsevier Ltd. All rights reserved.

**Keywords:** All-sky image; Regression; Solar irradiance prediction

## 1. Introduction

With the trend of sustainable and green energy, there is a growing demand for solar energy technology. To utilize solar energy effectively, integrated and large scale photovoltaic systems need to overcome the unstable nature of solar resource. Mechanisms of scheduling, dispatching, and allocating energy resources adaptively are desired by Photovoltaics grid operators. The ability to forecast surface solar irradiance is helpful for planning and deployment of electricity generated by different units. Numerical weather prediction information or satellite cloud observations are popular materials used for wide-range prediction (Perez et al., 2002, 2010; Remund et al., 2008). However, the resolution of prediction with respect to space and time

obtained by weather prediction information or satellite cloud images is relatively coarse compared to the resolution desired for photovoltaic grid operators. For more refined spatial and temporal resolution of irradiance prediction, researches that analyze images obtained from devices capturing skies have emerged. Sabburg and Wong have conducted an evaluation on a ground-based sky camera system to study the effect of clouds (Sabburg and Wong, 1999). Devices developed more recently include Whole Sky Imager developed by Scripps Institute of Oceanography at the University of California (Kassianov et al., 2005; Li et al., 2004), Whole Sky Camera developed by Spain's University of Girona (Long et al., 2006), All Sky Imager developed by Japanese Communications Research Laboratory (Kubota et al., 2003), and Total Sky Imager by Yankee Environmental Systems (Calbo and Sabburg, 2008; Long et al., 2006; Pfister et al., 2003). Among these instruments, more researches on extracting useful meteorological information are conducted using images captured

\* Corresponding author. Tel.: +886 3 4227151x35306; fax: +886 3 422 2681.

E-mail address: [chengsy@csie.ncu.edu.tw](mailto:chengsy@csie.ncu.edu.tw) (H.-Y. Cheng).

by Total Sky Imagers. Researchers have shown that within the 15-min time horizon, forecasting with sky image processing coupled with stochastic learning techniques demonstrated obvious improvement over satellite images (Marquez and Coimbra, 2013).

Cloud coverage, configurations and types are critical factors that influence the solar irradiance (Cazorla et al., 2008). Therefore, a category of research works are devoted to analyzing and classifying clouds. Cloud cover determination is an important topic that aims at deciding if a pixel belongs to a cloud. Typically, red blue ratio (RBR) of each pixel is used to indicate whether the dominant source of the pixel is from a clear sky or clouds (Chow et al., 2011; Johnson et al., 1989, 1991; Long et al., 2006; Shields et al., 2007, 2009). Then, a threshold is applied to RBR to determine cloud pixels in a sky image. The pixels whose RBRs are lower than the threshold are classified as clear sky and the pixels whose RBRs are higher the threshold are labeled as clouds. In addition to pure color characteristics, Roy et al. also tried a neural network approach with a wider range of variables for cloud segmentation (Roy et al., 2001). Under lower-visibility conditions, aerosol and thin clouds tend to cause errors in cloud determination. To improve the accuracy of the single threshold method, Huo and Lu proposed an integrated method for cloud determination under low visibility conditions (Huo and Lu, 2009). The integrated cloud-determination algorithm uses fast Fourier transform, symmetrical image features, and self-adaptive thresholds. Ghonima et al. improved the cloud detection process by employing a haze correction factor to account for the effects of variations in aerosol optical depth on RBR (Ghonima et al., 2012).

The relationships between cloud coverage and surface solar irradiance have been explored (Feister and Shields, 2005; Pfister et al., 2003). Pfister et al. quantified the cloud impact on surface irradiance by using estimates for the clear-sky surface irradiance and the ratio of measured to estimated clear-sky values. The research pointed out that although cloud fraction and surface irradiance are obviously negatively correlated, optically thin clouds tend to enhance the diffuse radiation significantly. Also, under broken cloud conditions, surface irradiance often exceeds expected clear-sky values because the direct component of solar radiation is nearly unaffected. Their research indicated that different cloud types would exhibit different impact on solar radiation. Therefore, in addition to estimating cloud coverage, classifying cloud types in all-sky images have gradually drawn researchers' attention. In the research by Martínez-Chico et al. (2011), the clouds are classified into different attenuation groups according to different levels of attenuation of the direct solar radiation reaching the surface. The authors also analyzed the annual and seasonal frequencies of each cloud group. For classification using image features, simple statistical texture features such as entropy, smoothness, and moments have been extracted from images to classify clouds (Calbo and Sabburg, 2008; Heinle et al., 2010). Heinle et al. used a

$k$ -nearest neighbor to classify the clouds into seven different types. Other features such as autocorrelation, co-occurrence matrices, edge frequency, Law's features and primitive length are also tested for cloud classification (Singh and Glennen, 2005). However, these pioneer researches have not conducted further prediction tasks after performing cloud classification.

Chow et al. developed a forecasting system that nowcasts cloud cover map with information extracted from 30 s to 5 min ahead of prediction time (Chow et al., 2011). However, the global vector determined from the inter-frame cross-correlation cannot describe the cloud motion well. Also, the relationship of the predicted cloud map and the solar irradiance are not clearly stated in this work. Wood-Bradley et al. tracked the cloud with optical flow (Wood-Bradley et al., 2012). However, they only forecast cloud movement, not solar irradiance. To address the problem of forecasting solar irradiance for grid operators, the aim of this work is to automatically predict solar irradiance several minutes in advance. In this paper, we explore the potentials and possibilities of utilizing the features extracted from all-sky images to predict solar irradiance. The systems modules are illustrated in Fig. 1. We extract relevant features based on our observations and train a regression model for the prediction purpose. We design a mechanism that predicts clearness index and then calculates the desired solar irradiance from the predicted clearness index. We will show that the designed clearness index prediction mechanism yields better results than predicting solar irradiance directly.

## 2. All-sky image feature extraction

In this work, the device used to capture the all sky images is the all sky camera manufactured by the Santa Barbara Instrument Group (SBIG). The longitude and latitude of the all sky camera are  $24^{\circ}46'34.00''\text{N}$  and  $121^{\circ}2'40.10''\text{E}$ . The detailed specifications of the all sky camera are listed in Table 1.

The device used to measure the ground truth solar irradiance is Delta OHM LP RYRA 03. It is a point sensor

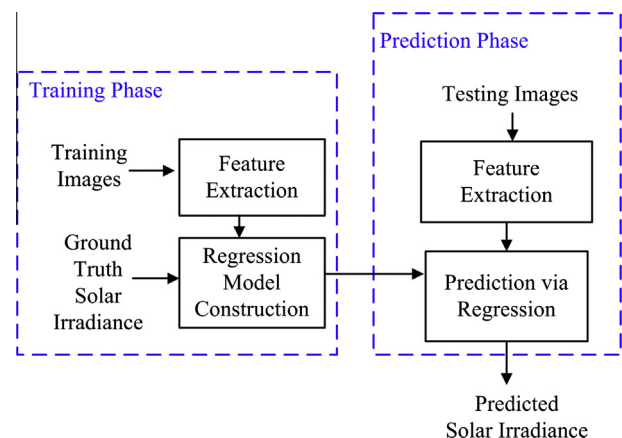


Fig. 1. System modules.

Table 1  
Specifications of the all sky camera used for experiments.

Specifications of the all sky camera	
Charge-coupled device (CCD)	Kodak KAI-0340 @ 7.4 microns square
Lens	Fujinon FE185C046HA-1
Field of view (FOV)	185°
Resolution	640 × 480 bitmap image
Focal length	1.4 mm
Focal ratio range	f/1.4–f/16
Weight	0.4 kg
Communication interface	RS-232
Power consumption	12VDC, 600 mA max (2.4 W)

that is located next to the all sky camera. The sampling interval is 10 s. The specifications of Delta OHM LP RYRA 03 are listed in Table 2.

Based on the dataset collected by the above mentioned devices, this work analyzes and compares several different types of all-sky image features that can be used for training the regression model for solar irradiance prediction. In the following sub-sections, we elaborate the all-sky image features considered in this work.

### 2.1. Number of cloud pixels

We apply the classic RBR thresholding method to decide cloud pixels. Let  $NC_t$  denote number of pixels whose RBR are larger than a threshold at frame  $t$  (Long et al., 2006). Fig. 2 displays selected examples of cloud detection results. The images in Fig. 2(a) are the original all-sky images, and Fig. 2(b) lists the corresponding detection results. The green pixels in the detection images are the detected cloud pixels with RBR are larger than 0.8 but smaller than 1. The number of cloud pixels  $NC_t$  in frame  $t$  is considered as a useful clue that would affect solar irradiance. We take the average and variance of  $NC_t$  over a time period  $T$  as features,  $MEAN_{NC}$  and  $VAR_{NC}$ . The equations of computing  $MEAN_{NC}$  and  $VAR_{NC}$  are shown in the following equations:

$$MEAN_{NC} = \frac{1}{T} \sum_{t=1}^T NC_t \quad (1)$$

$$VAR_{NC} = \frac{1}{T} \sum_{t=1}^T (NC_t - MEAN_{NC})^2 \quad (2)$$

### 2.2. Frame difference

We observe that large values in the difference image of frame  $t$  and frame  $t-1$  often correspond to the boundary

Table 2  
Specifications of delta OHM LP RYRA 03.

Measuring range	0–2000 W/m <sup>2</sup>
Viewing field	2π sr
Spectral field	335–2200 nm (95%)
Impedance	13.75 μV (W/m <sup>2</sup> )
Weight	0.45 kg

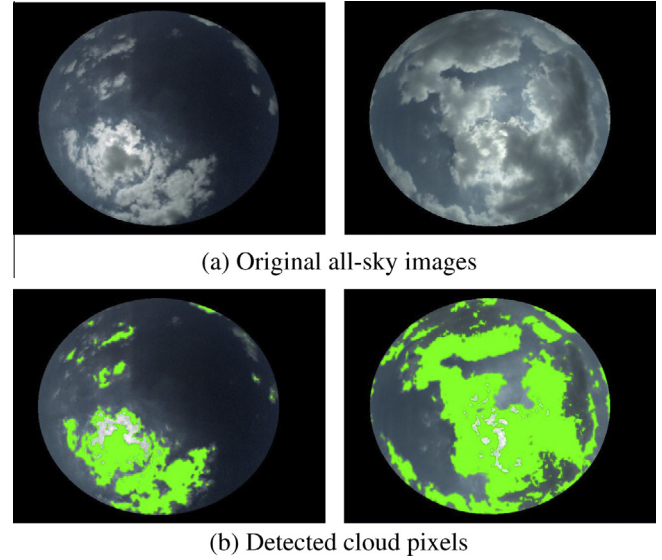


Fig. 2. Example of cloud detection: (a) original all-sky images; (b) detected cloud pixels.

areas of moving clouds. If the clouds are still, there are few pixels in the difference image after performing thresholding. Two types of features related to frame differences are considered. According to Eq. (3) listed below,  $\sum_{x \in \text{frame } t} DA_t(x)$  records the summation of absolute difference values that exceed the threshold in image frame  $t$ . According to Eq. (4),  $\sum_{x \in \text{frame } t} DC_t(x)$  records the count of the pixels whose absolute difference values exceed the threshold in image frame  $t$ .

$$DA_t(x) = \begin{cases} |I_t - I_{t-1}| & \text{if } |I_t - I_{t-1}| \geq Thr_D \\ 0 & \text{if } |I_t - I_{t-1}| < Thr_D \end{cases} \quad (3)$$

$$DC_t(x) = \begin{cases} 1 & \text{if } |I_t - I_{t-1}| \geq Thr_D \\ 0 & \text{if } |I_t - I_{t-1}| < Thr_D \end{cases} \quad (4)$$

Note that in Eqs. (3) and (4),  $I_t$  denotes the gray level image of frame  $t$ . The mean  $MEAN_{DA}$  and variance  $VAR_{DA}$  of  $\sum_{x \in \text{frame } t} DA_t(x)$  over a time period  $T$  are computed as features. The mean  $MEAN_{DC}$  and variance  $VAR_{DC}$  of  $\sum_{x \in \text{frame } t} DC_t(x)$  are computed similarly. Fig. 3 shows an example of performing frame differencing. Fig. 3(a and b) is two consecutive image frames. By performing frame differencing according to Eqs. (3) and (4) with  $Thr_D = 20$ , we can obtain difference images in Fig. 3(c and d), respectively. Note that in Fig. 3(c) the gray level values are normalized to 0–255 for better visualization.

### 2.3. Gradient magnitude

The gradient information corresponds to edges in an image. Edges are often related to cloud boundaries in all-sky images. We apply Sobel gradient operators (Gonzalez and Woods, 2002) to the image frames in order to detect edge pixels. The Sobel gradient operators of  $x$  direction ( $I_x$ ) and  $y$  direction ( $I_y$ ) are shown in Fig. 4(b). The gradient of  $x$  direction ( $G_x$ ) and  $y$  direction ( $G_y$ ) are obtained by

convolving the original all-sky image frame with  $I_x$  and  $I_y$ , respectively. And the gradient magnitude is defined as  $G = \sqrt{G_x^2 + G_y^2}$ . The mean and variance of the summation of gradient magnitude in an image frame are computed over a time period  $T$ . These two features are denoted as  $MEAN_{GM}$  and  $VAR_{GM}$ , respectively. In Fig. 4, the magnitude of Sobel gradient operator on each pixel in a cloudy scene is scaled to 0–255 for better visualization.

#### 2.4. Intensity level

We can observe from the all-sky image dataset that the brightness level of clear skies and cloudy skies are very different. Therefore, the mean and variance of the average intensity value of image frames over a period  $T$  are considered as features. These two features are denoted as  $MEAN_{IL}$  and  $VAR_{IL}$ , respectively.

#### 2.5. Accumulated intensity along the vertical line of sun

The “vertical” line of Sun is an artifact created by the CCD device as can be seen in Fig. 5. If the sunlight is strong, the line caused by the sun would cross the entire image. Also, the width of the line would be thick, as shown in Fig. 5(a). If the sun is partially occluded by the clouds, the line would be short and thin as shown in Fig. 5(b). When the sun is completely blocked, there would be no line at all, as shown in Fig. 5(c). The irradiance is highly related to the accumulated intensity along the vertical line. As can be demonstrated in Section 4, the correlation between the ground truth irradiance and the accumulated intensity along the vertical line is high. The procedure of accumulating the intensity is first applying Hough transform (Gonzalez and Woods, 2002) to detect the position of the line.

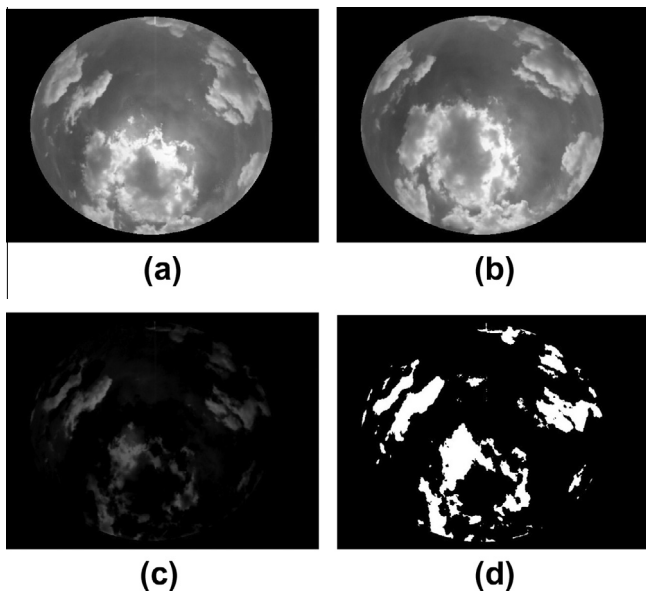


Fig. 3. Example of frame differencing: (a) gray scale image frame  $t-1$  (b) gray scale image frame  $t$  (c) gray level difference image  $DA_t$ . (d) Binary difference image  $DC_t$ .

Hough transform is used to detect the vertical lines of sun in the all-sky images. Since all the points on a line satisfy the same line equation, the procedure of detecting lines is equivalent to finding the coefficients of the line equations using a voting mechanism. This can be achieved by dividing the parameter space into grids. However, if the line equation  $y = mx + b$  is used, it would be very inefficient to explore the parameter space  $(m, b)$  due to the fact that possible values for the slope  $m$  ranges from minus infinity to infinity. Hough transform is an algorithm of determining the parameters of a line equation by changing the parameter space to  $\rho$  and  $\theta$ . The line equation is re-parameterized as  $x \cos \theta + y \sin \theta = \rho$ , as illustrated in Fig. 6(a). The grid for voting is illustrated in Fig. 6(b), where  $\rho_{\min}$  and  $\rho_{\max}$  can be set according to the size of the image. And  $\theta_{\min}$  and  $\theta_{\max}$  are  $-180^\circ$  and  $180^\circ$ , respectively.

An example of performing Hough transform on an all-sky image is displayed in Fig. 7. Edge detection needs to be performed first to generate a binary edge map. In Fig. 7(b), the vertical line of sun is preserved in the edge map. Applying Hough transform on the edge map, we can observe in Fig. 7(c and d) that there would be large accumulated values in the grids corresponding to detected lines. Fig. 8 demonstrates two selected image frames and their line detection results. After detecting the vertical line, a line window with a fixed width is established at the detected line. The pixels within the line window are thresholded according to their intensity values. The threshold we used in this work is 200. The thresholded intensities are then accumulated throughout the entire line window. The mean  $MEAN_{SI}$  and variance  $VAR_{SI}$  of the accumulated intensity over a period  $T$  are computed as features.

#### 2.6. Number of corners

Corners in all-sky images correspond to the details of the clouds that have edges with at least two directions in a local patch  $A$ . One of the common methods to detect corners is Harris corner detector (Shapiro and George, 2001). The gradients in vertical and horizontal directions,  $I_x$  and  $I_y$ , are computed for each pixel in the local patch  $A$ .

Then a covariance matrix  $C$  is computed using the gradients of all pixels in the local patch  $A$  according to Eq. (5). The covariance matrix  $C$  carries the information of how the gradients of local pixels are distributed. If there are two or more gradient orientations in the local patch, both the Eigen-values of the matrix  $C$  would be large. Therefore, the Eigen-values of  $C$  are used to check for corners.

$$C = \begin{pmatrix} \sum_{(x,y) \in A} I_x^2(x,y) & \sum_{(x,y) \in A} I_x(x,y)I_y(x,y) \\ \sum_{(x,y) \in A} I_x(x,y)I_y(x,y) & \sum_{(x,y) \in A} I_y^2(x,y) \end{pmatrix} \quad (5)$$

Fig. 9 displays selected examples of detected corners with local patch  $A$  of size  $7 \times 7$ . For the number of detected



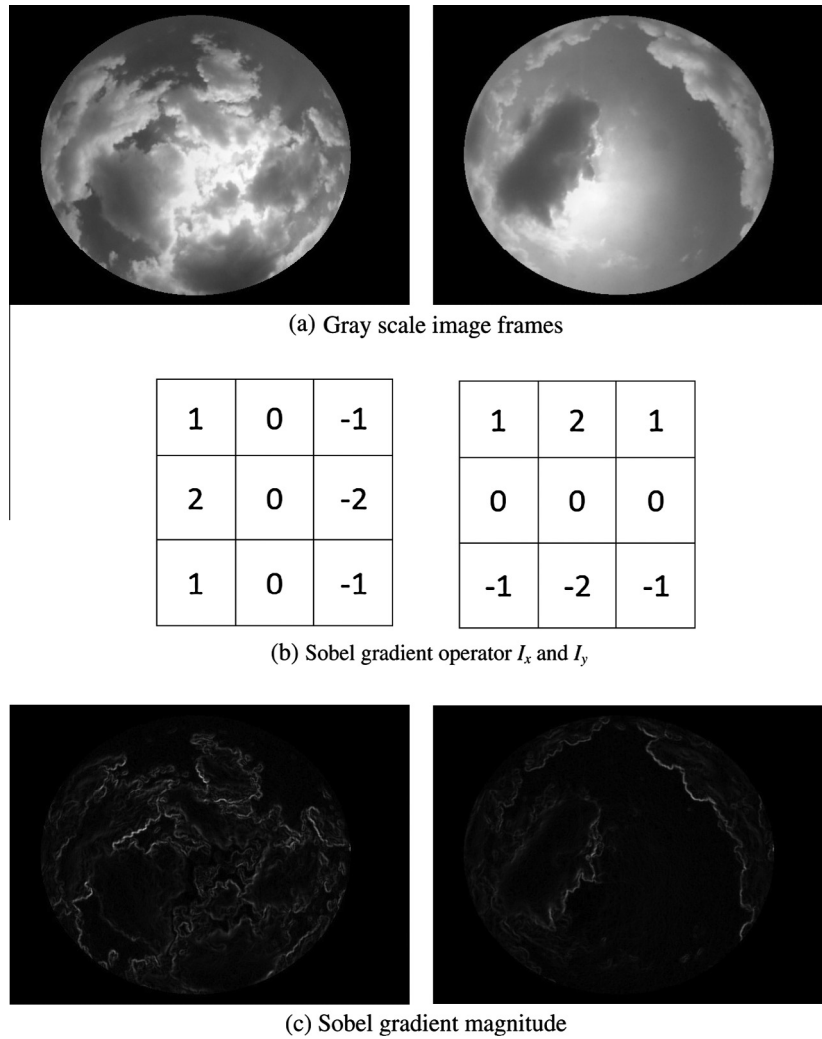


Fig. 4. Example of edge detection: (a) gray scale image frames (b) sobel gradient operator of x direction and y direction (c) sobel gradient magnitude.

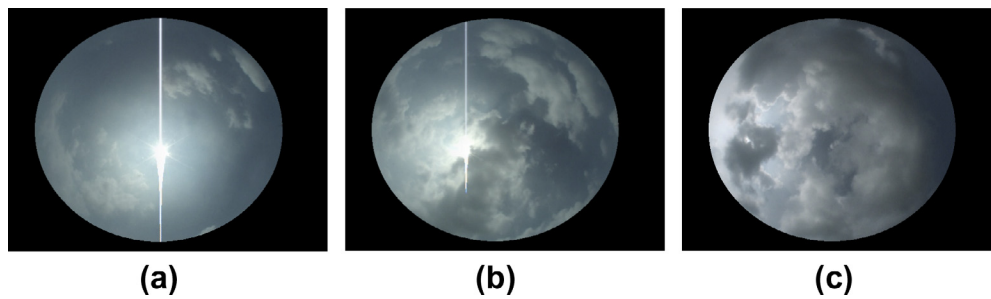


Fig. 5. Example of lines of the sun with different lengths and widths.

corners, its mean  $MEAN_{COR}$  and variance  $VAR_{COR}$  over a period  $T$  are computed as features.

### 3. Solar irradiance prediction via regression

We use regression techniques to achieve solar irradiance prediction. The goal is to establish the correlation between the extracted features and the desired solar irradiance prediction through a regression model. However, instead of predicting solar irradiance directly, we predict the solar

irradiance using clearness index and extraterrestrial solar irradiance, which is stated in the Perez conversion model (Perez et al., 1987). According to this model, the clearness index (CI) can be computed by Eq. (6), where GHI denotes global horizontal irradiance at surface. TSI denotes the total solar irradiance, which measures mean extraterrestrial solar irradiance received over a surface perpendicular to the solar beam. We use the long-term averaged TSI, which is known as the solar constant when applying Eq. (6) (Gueymard, 2004). More specifically, the value of TSI in

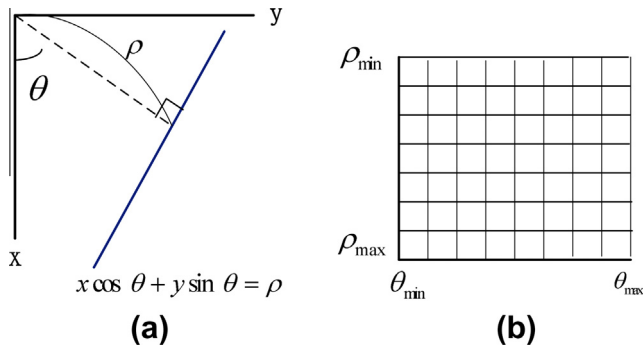


Fig. 6. Hough transform.

Eq. (6) is set as  $1367 \text{ W/m}^2$  in this work. The symbol  $r_0$  denotes mean earth–sun distance;  $r$  denotes the earth–sun distance at the time of interest; and SZA denotes the solar zenith angle at the time of interest, which can be computed according to the solar position algorithm by Reda and Andreas (2004). The denominator on the right hand side of Eq. (6) is the extraterrestrial solar irradiance. Therefore, the solar irradiance we would like to predict, which is GHI in Eq. (6), can be expressed by the clearness index multiplied by  $TSI \cdot (r_0/r)^2 \cos(SZA)$ .

$$CI = GHI / [TSI \cdot (r_0/r)^2 \cos(SZA)] \quad (6)$$

In the training phase, the features from training images and the ground truth clearness indices are used to construct the regression model in the training phase. In the prediction phase, the same features are extracted and serve as inputs of the constructed regression function. The regression function outputs the clearness index prediction. Afterwards, the clearness index is multiplied by the extra-terrestrial solar irradiance and converted back into the desired solar irradiance. Note that the term  $TSI \cdot (r_0/r)^2 \cos(SZA)$  can be determined in advance given the information of  $r_0$ ,  $r$ , and SZA (Reda and Andreas, 2004). For the regression model, a simple polynomial regressor is applied. The regression function can be expressed by  $f_{\text{regression}}(\mathbf{X}_i)$  where  $\mathbf{X}_i = (x_i^{(1)} x_i^{(2)} \dots x_i^{(D)})^T$  is a  $D$ -dimensional feature vector. Algorithm A and algorithm B summarize the prediction schemes with and without clearness index conversion mechanism, respectively. The features are computed over a time period of  $T$  minutes and the algorithm outputs the predicted irradiance of  $P$  minutes later. We will show in Section 4 that such conversion mechanism yields more stable and accurate prediction results.

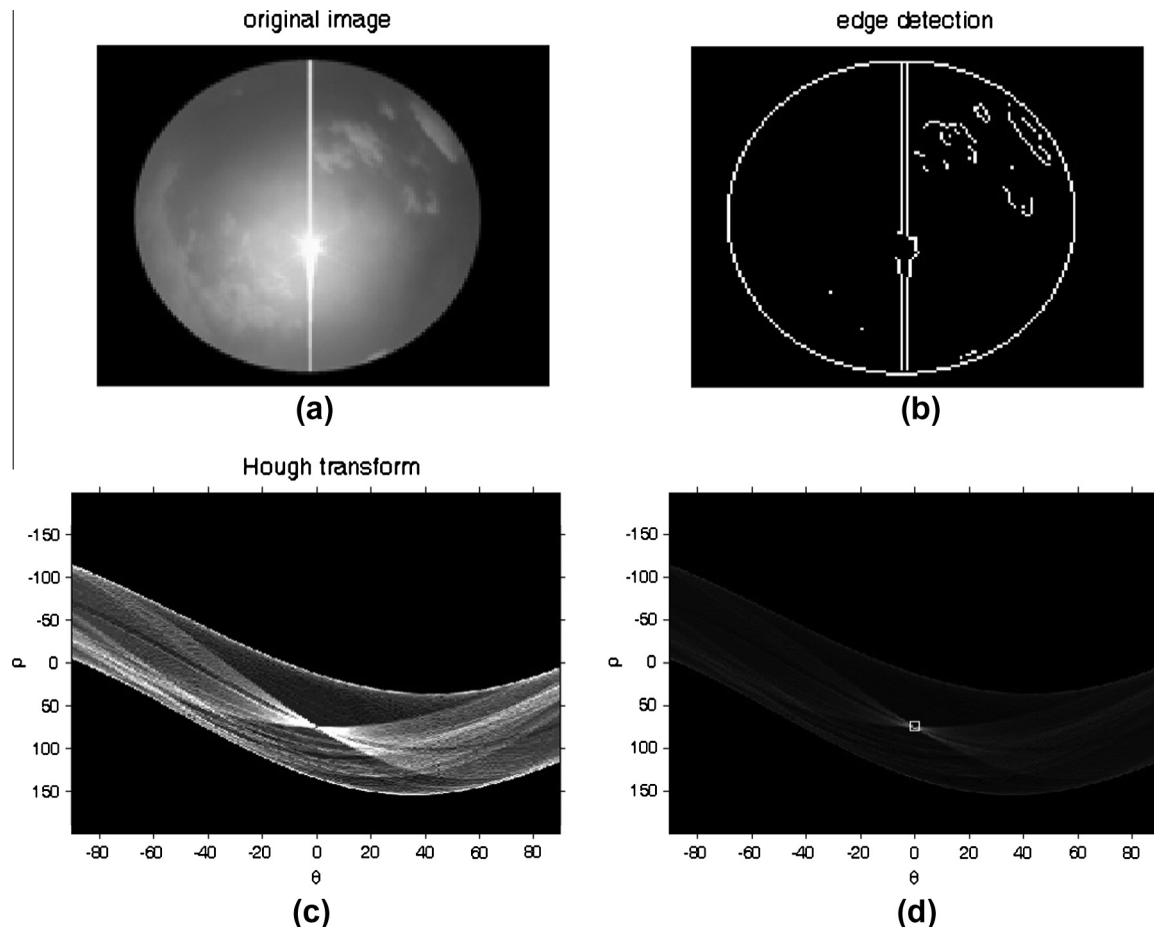


Fig. 7. Example of Hough transform: (a) original image; (b) edge detection result (c) Hough transform response (d) detected peak in the Hough parameter space.



Fig. 8. Example of line detection: (a) original all-sky images; (b) detected vertical lines.

---

**Algorithm A**


---

For each time instance at the  $i$ th minute  
 Compute the feature vector  $X_i$  from time interval  $[i - T, i]$   
 Compute the predicted clearness index of the  $(i + P)$ th minute  $\hat{C}I_{i+P} = f_{regression}^{(A)}(X_i)$   
 Compute the predicted solar irradiance  $\hat{Irr}_{i+P} = \hat{C}I_{i+P} \cdot [TSI \cdot (r_0/r)^2 \cos(SZA)]_i$   
 end

---



---

**Algorithm B**


---

For each time instance at the  $i$ th minute  
 Compute the feature vector  $X_i$  from time interval  $[i - T, i]$   
 Compute the predicted solar irradiance of the  $(i+P)$ th minute  $\hat{Irr}_{i+P} = f_{regression}^{(B)}(X_i)$   
 end

---

#### 4. Experimental results

The all-sky image device used in the experiments captured one all-sky image per minute. However, the device would fail to capture images at certain time instances, and some images might be missing occasionally. For the time instance that an image is missing, the features computed from the image nearest to that time instance is reused in the averaging process when obtaining the feature vector for the regression function. The four-week dataset is recorded during August 2011 in Taiwan. In our experiment, the data in the 4th week includes only 6 days because the 7th day of the 4th week is rainy and therefore not included in our experiment. The dataset is very challenging since the variation of weather and cloud cover is high with the marine climate at the costal site. The typical summer weather with tropical convection and fast cloud movement makes the prediction task more difficult.

##### 4.1. Feature selection

In Section 2, fourteen features are described, which are  $MEAN_{NC}$ ,  $VAR_{NC}$ ,  $MEAN_{DA}$ ,  $VAR_{DA}$ ,  $MEAN_{DC}$ ,

$VAR_{DC}$ ,  $MEAN_{GM}$ ,  $VAR_{GM}$ ,  $MEAN_{IL}$ ,  $VAR_{IL}$ ,  $MEAN_{SI}$ ,  $VAR_{SI}$ ,  $MEAN_{COR}$ , and  $VAR_{COR}$ . To examine the correlation between the  $d$ th feature  $x_i^{(d)}$  and the ground truth solar irradiance  $y_{i+P}$  at the prediction time, we compute the correlation coefficient using Eq. (7). In Eq. (7),  $n$  is the total number of data samples used for training.

$$coeff^{(d)} = \frac{\sum_{i=1}^n (x_i^{(d)} - \bar{x}^{(d)}) (y_{i+P} - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i^{(d)} - \bar{x}^{(d)})^2} \sqrt{\sum_{i=1}^n (y_{i+P} - \bar{y})^2}} \quad (7)$$

$$\text{where } \bar{x}^{(d)} = \frac{1}{n} \sum_{i=1}^n x_i^{(d)} \quad \text{and} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_{i+P} \quad (8)$$

Note that both positive correlation and negative correlation are helpful when constructing the regression model for prediction. Therefore, we inspect the absolute values of the correlation coefficients of different features. Fig. 10 plots the absolute value of the correlation coefficient between each feature and the ground truth solar irradiance to be predicted. In Fig. 10, we use the settings of  $P = 5$  and  $T = 15$ . It means that the features are computed over the 15 previous minutes and the prediction is made for 5 min later.

The features are ranked according to their absolute values of correlation coefficients. From Fig. 10, it is obvious that  $MEAN_{NC}$ ,  $VAR_{NC}$ ,  $MEAN_{GM}$ ,  $MEAN_{IL}$ , and  $MEAN_{SI}$  have higher absolute values of correlation coefficients. We plot the ground truth solar irradiance of the second week along with the top-ranked features in Fig. 11. Among these five features,  $MEAN_{GM}$ ,  $MEAN_{IL}$ , and  $MEAN_{SI}$  exhibit positive correlations with solar irradiance. On the contrary,  $MEAN_{NC}$  and  $VAR_{NC}$  exhibit negative correlations.

We use Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Bias Error (MBE) between the ground truth irradiance  $Irr_i$  and predicted irradiance  $\hat{Irr}_i$  to measure the accuracy of our prediction mechanism. The measurements RMSE, MAE, and MBE are computed according to Eqs. (9)–(11) respectively. In these equations,  $n$  denotes the number of prediction samples. Since the system gives one prediction output every minute,  $n = 11,340$  if we would like to compute the RMSE of the entire dataset, which we refer to as four-week RMSE hereafter. If hourly

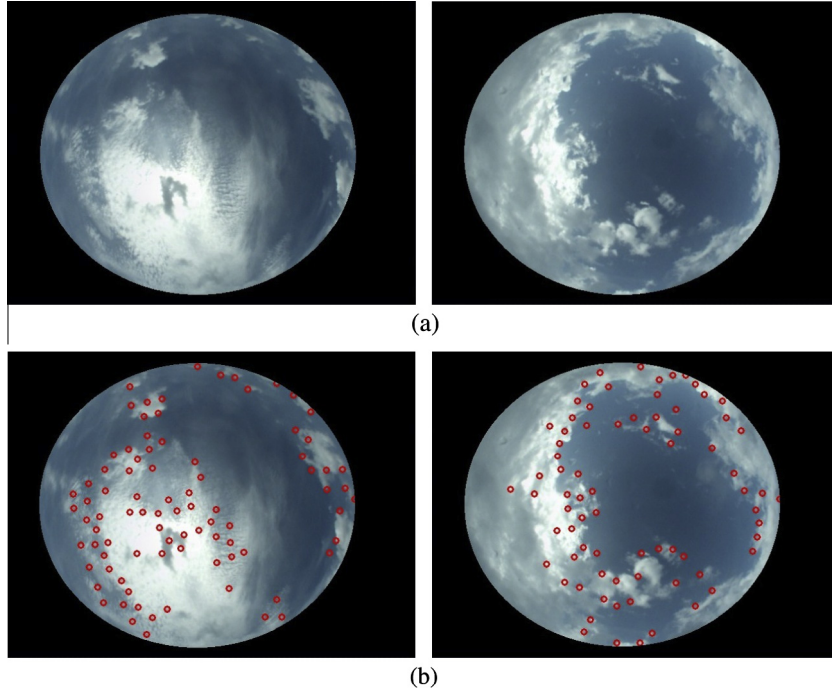


Fig. 9. Example of corner detection: (a) original all-sky image; (b) detected corners.

MAE is to be computed, then  $n$  would be 60. Cross validation is performed to compute the average RMSE, MAE and MBE. Cross validation is carried out by rotating the training data and testing data. We perform fourfold cross validation by dividing the dataset into four weeks. Data of each week is used as training data in turn.

$$RMSE = \sqrt{\frac{1}{n} \sum_i^n (\hat{Irr}_i - Irr_i)^2} \quad (9)$$

$$MAE = \frac{1}{n} \sum_i^n |\hat{Irr}_i - Irr_i| \quad (10)$$

$$MBE = \frac{1}{n} \sum_i^n (\hat{Irr}_i - Irr_i) \quad (11)$$

From the top ranked feature, one feature is added every time to perform regression and compute the cross validated four-week RMSE. From Fig. 12, it is observed that using the first three to five top-ranked features yields the best results. The top five ranked features all have  $|coeff^{(d)}| > 0.2$ . Therefore, in the rest of the experiments, we select features with  $|coeff^{(d)}| > 0.2$  to perform prediction, which are  $MEAN_{NC}$ ,  $VAR_{NC}$ ,  $MEAN_{GM}$ ,  $MEAN_{IL}$  and  $MEAN_{SI}$ .

#### 4.2. Prediction accuracy

Table 3 lists the cross validated four-week RMSE, MAE and MBE with different settings of  $P$  and  $T$ . Note that  $P$  denotes that prediction is performed  $P$  minutes ahead of time, and  $T$  denotes that the features are computed over  $T$  minutes. For different settings of  $P$  and  $T$ , the subset

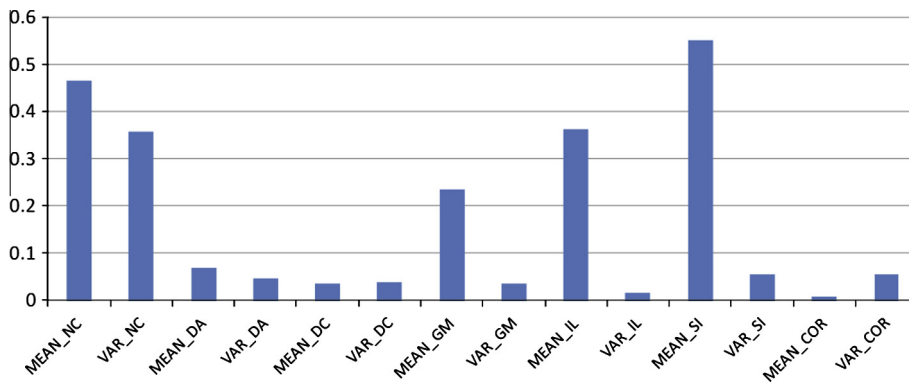


Fig. 10. Correlation coefficient between each feature and the solar irradiance to be predicted.



of features with  $|coeff^{(d)}| > 0.2$  remains the same. It is obvious that the RMSE, MAE and MBE increase as we attempt to predict solar irradiance farther away from the time when the features are computed. In Table 3, RMSE, MAE and MBE are also expressed in percentage to show the relative error. When RMSE, MAE and MBE are expressed in percentage, the results obtained by Eqs. (9)–(11) are divided by the ground truth irradiance.

To visualize the effectiveness of prediction, the predicted irradiance  $\hat{Irr}_i$  is plotted along with the ground truth irradiance  $Irr_i$  in Fig. 13. The prediction is performed

$P = 5$  min ahead of time with features computed over  $T = 5$  min. In Fig. 13, the data from the first week is used for training and the rest of the data for testing. The constructed regression expression obtained by training data from the first week is listed in equation.

$$\begin{aligned}\hat{CI}_{i+P} &= f_{\text{regression}}(X_i) \\ &= -0.0056MEAN_{NC} - 0.00104VAR_{NC} \\ &\quad + 0.000638MEAN_{GM} + 0.001013MEAN_{IL} \\ &\quad + 0.011841MEAN_{SI} + 0.313751\end{aligned}\quad (12)$$

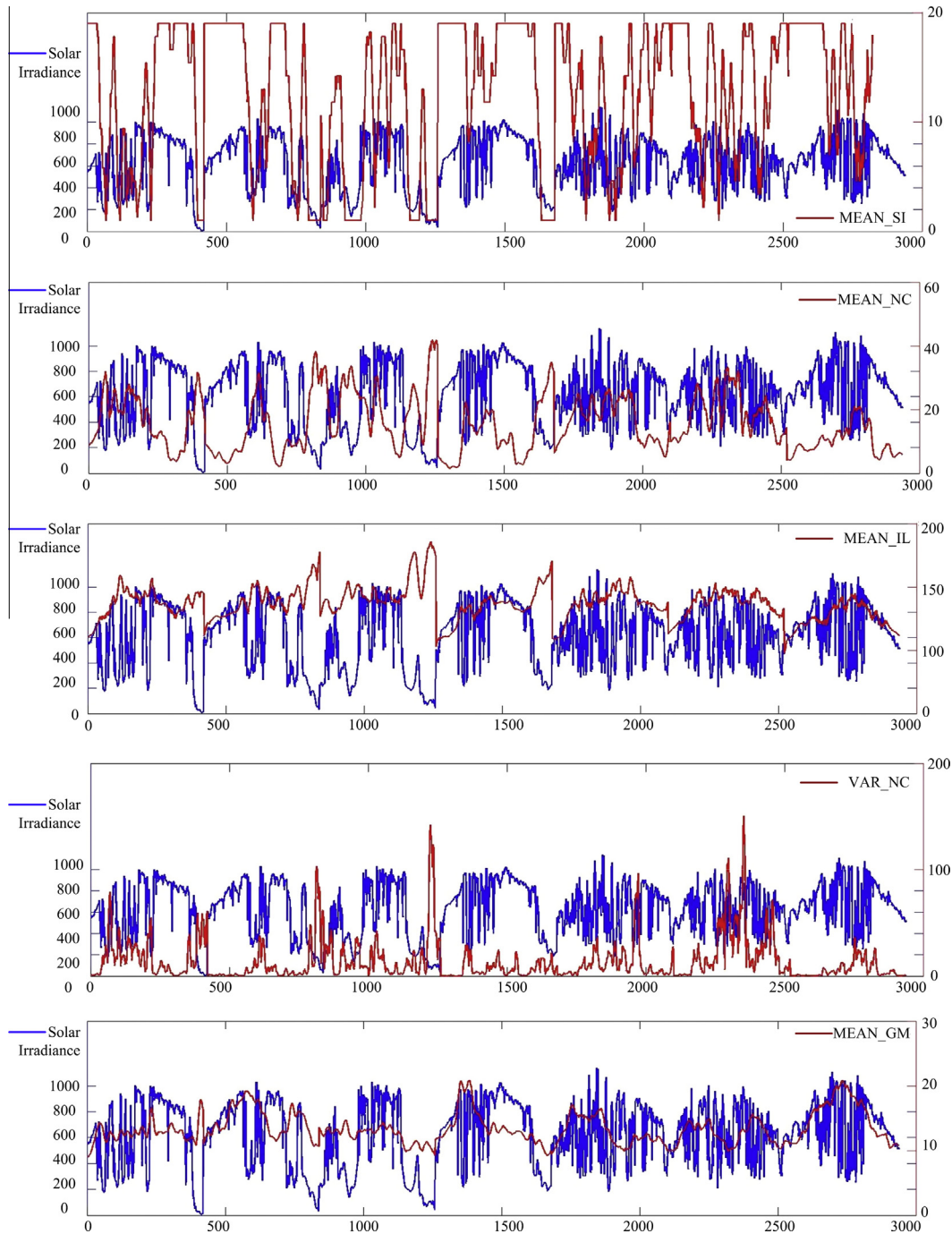


Fig. 11. Plots of five top-ranked features and the ground truth solar irradiance of the second week.

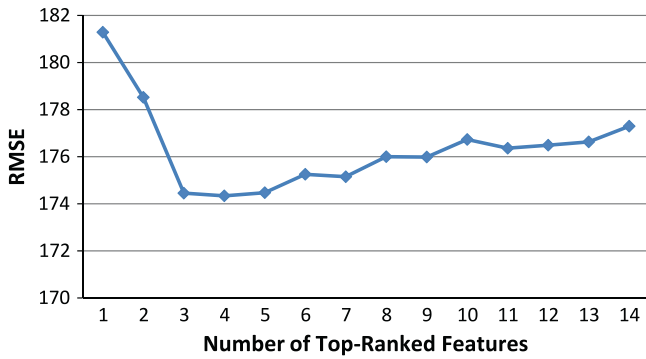


Fig. 12. Cross-validated four-week RMSE vs. number of top-ranked features used for regression.

The trend of the solar irradiance can be successfully captured by the constructed regression model. The irradiance prediction up to 15 min in advance can be achieved with mean absolute error less than 24% for this highly challenging dataset. However, for sudden surge or drop of solar irradiance, the prediction scheme still has some space for improvement.

Since four-week RMSE, MAE, and MBE cannot demonstrate the stability of short-term prediction, hourly MAE is computed to see how the prediction performs hour by hour. Fig. 14 plots the hourly MAE of 189 h in our dataset. We are interested in the number of hours whose hourly MAE is less than a threshold  $MAE_{Threshold}$ . The quality ratio is defined as the number of hours whose hourly MAE is less than  $MAE_{Threshold}$  divided by the number of total hours  $N$ , as stated in Eqs. (13) and (14). For example, out of 189 h in our dataset, there are 92 h whose hourly MAE are less than 25%, yielding a quality ratio of 48.68%. Fig. 15 shows the cumulative quality ratio as the threshold  $MAE_{Threshold}$  increases. For  $MAE_{Threshold} = 40\%$ , the quality ratio is 76.72%. For  $MAE_{Threshold} = 50\%$ , the quality ratio reaches 86.24%.

$$QR = \frac{1}{N} \sum_{i=1}^N B_i \times 100\% \quad (13)$$

$$B_i = \begin{cases} 1 & \text{if } \text{Hourly MAE} \leq MAE_{Threshold} \\ 0 & \text{if } \text{Hourly MAE} > MAE_{Threshold} \end{cases} \quad (14)$$

Also, we would like to show that the designed regression method using clearness index conversion performs better than predicting the solar irradiance directly. Fig. 15 compares the cumulative quality ratio of both schemes. The proposed prediction scheme demonstrates higher quality ratios. Fig. 16 displays the cross validated four-week RMSE, MAE, and MBE for prediction with and without the clearness index conversion scheme. It is clear that prediction with clearness index conversion scheme outperforms prediction without clearness index conversion scheme.

#### 4.3. Discussions

It is not adequate to compare the results of the proposed work and the existing works directly because the experimental results are obtained from different datasets at different locations. However, by looking at the prediction accuracy of the related works, we can have a general picture of how well the existing works can do and see if the proposed work is competitive. Most existing works have verified that the prediction accuracy for clear days would be much better than that for cloudy days. Also, it is established that the prediction accuracy would drop if the dataset contains more cloud motions. In the work of Heinemann et al., the reported prediction RMSE ranges from 26% to 54% for various forecast sources (Heinemann et al., 2006). According to the survey from Glassley et al. (2010), several prediction results are listed as follows. The GHI forecasting RMSE ranges from 5% for clear skies to 60% for all conditions using different forecast sources in the work of Schroedter-Homscheidt et al. (2009). The GHI forecasting RMSE ranges from 12% for clear skies to 85% for cloudy conditions in the work of Lorenz et al. (2009). For prediction methods incorporating sky imagery, Marquez and Coimbra integrated present and past irradiance, meteorological variables, and cloud cover statistics from images to train artificial neural networks for irradiance forecasting (Marquez and Coimbra, 2011). They reported RMSE ranging from 15% to 22% for different models constructed on a 13 month dataset. With the above mentioned statistics, we can see that the proposed work is

Table 3

Cross validated four-week RMSE, MAE, and MBE of the proposed regression scheme with different settings of  $P$  and  $T$  over 1 month.

		RMSE (W/m <sup>2</sup> )	RMSE (%)	MAE (W/m <sup>2</sup> )	MAE (%)	MBE (W/m <sup>2</sup> )	MBE (%)
$P = 5$	$T = 5$	169.28	24.69	139.43	21.91	15.75	2.44
	$T = 10$	173.11	25.26	141.08	22.14	10.15	1.54
	$T = 15$	174.47	25.48	141.58	22.31	17.71	2.76
$P = 10$	$T = 5$	180.42	26.33	147.15	23.16	16.40	2.56
	$T = 10$	184.11	26.88	148.16	23.29	12.16	1.88
	$T = 15$	183.42	26.81	147.60	23.31	18.98	2.98
$P = 15$	$T = 5$	188.90	27.59	152.02	23.97	18.70	2.93
	$T = 10$	190.71	27.87	152.65	24.04	13.15	2.05
	$T = 15$	189.32	27.72	152.02	24.07	19.03	3.00

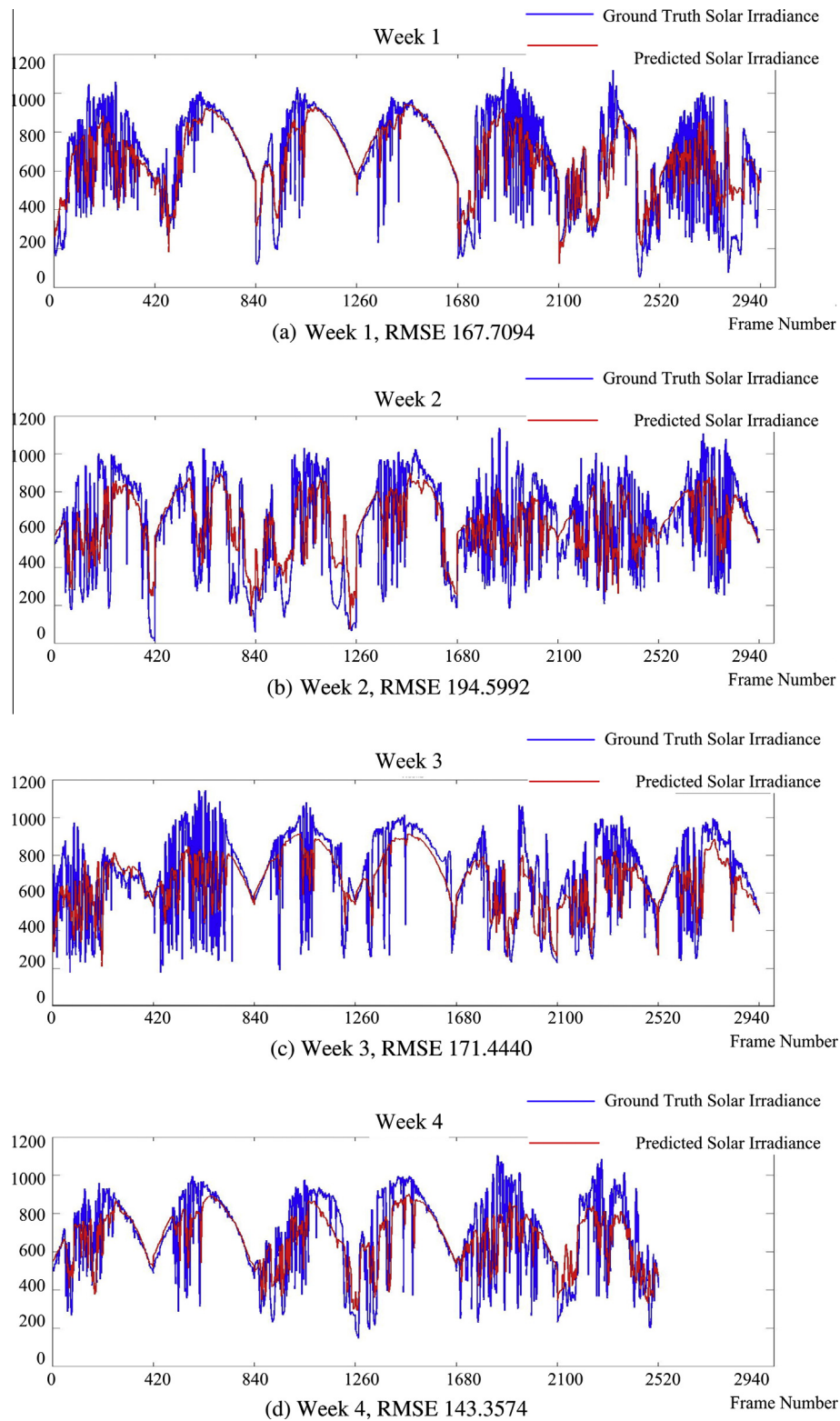


Fig. 13. Ground truth solar irradiance (blue lines) and predicted irradiance (red lines). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

very competitive given the fact that the experimental dataset includes a lot of cloudy conditions and frequent cloud motions. The proposed method is also suitable for

the initial stage of deployment at a certain location since the reported results only rely on training data of one week. Although there is much room for improvement, the

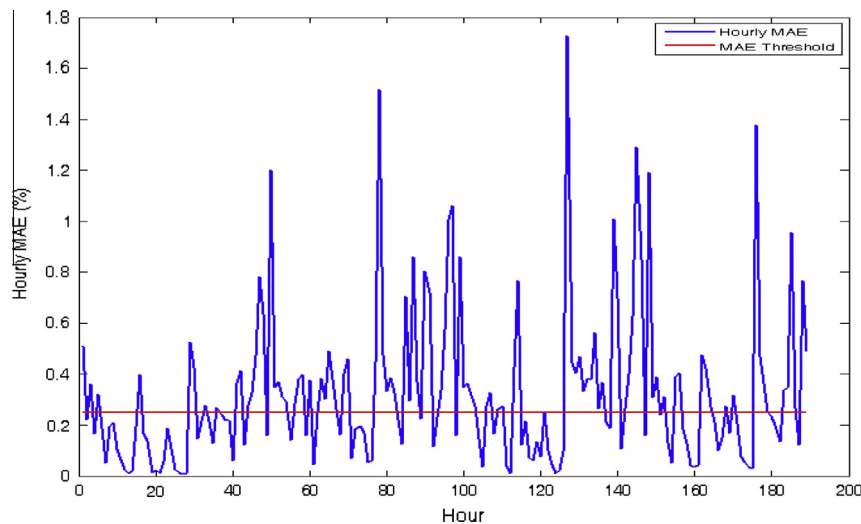


Fig. 14. Hourly MAE of 189 h.

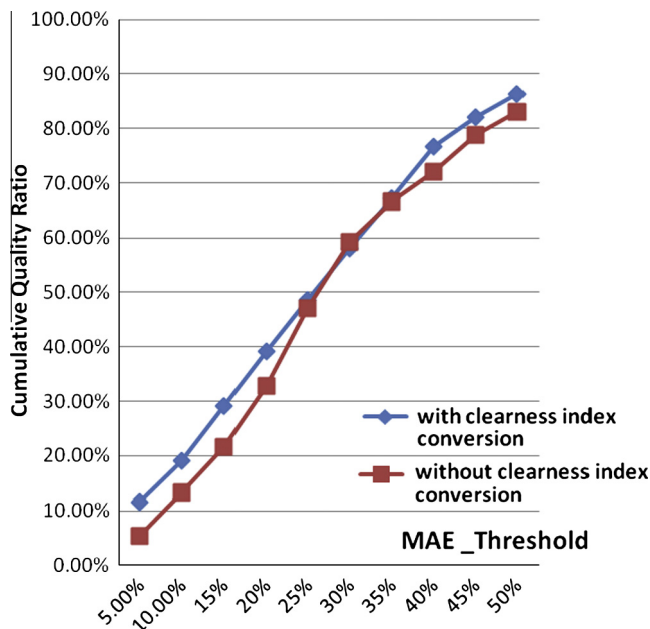
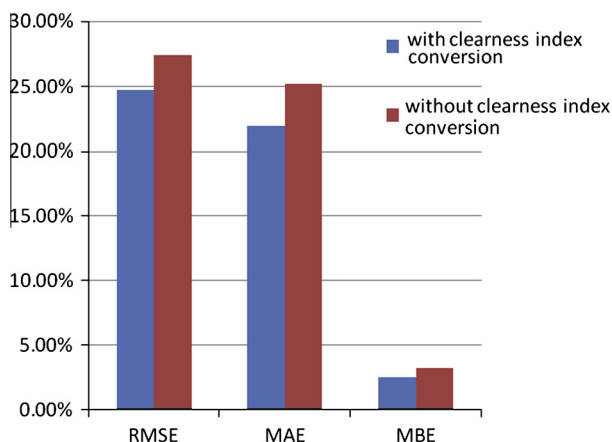
Fig. 15. Cumulative quality ratio as  $MAE_{Threshold}$  increases.

Fig. 16. Comparisons on cross validated four-week RMSE, MAE, and MBE for prediction with and without the clearness index conversion scheme.

enhancement that can be developed based on this work would be promising. We analyze potential ways to enhance this work in the Conclusion section.

## 5. Conclusions

In this work, we explore the potentials to utilize the features extracted from all-sky images to predict solar irradiance. By analyzing various image features, a feature subset with higher correlation with desired prediction target is selected. The selected features include mean and variance of number of cloud pixels, mean of gradient magnitude, mean of brightness level, and mean of accumulated intensity along the line of the sun. A regression model is trained using the selected features to perform prediction. The clearness index conversion technique is incorporated in the designed regression mechanism for more accurate prediction results. The predict horizon of this work ranges from 5 to 15 min. A challenging dataset collected during a month at a coastal site in Taiwan is used to validate the proposed method. The experiments show that prediction of 5 min ahead of time can achieve a monthly averaged MAE of 21.91%. There are 86.24% of hours that have hourly MAE less than 50%.

Implementing the solar irradiance prediction scheme would be helpful for grid operators to manage and allocate resources. For marine climate, the variable weather conditions would make the prediction task more challenging, especially with tropical convection and fast cloud movement in summer. Therefore, the prediction accuracy at locations with marine climate would usually be lower compared with locations with continental climate. There are several possible ways that the prediction scheme can be improved to achieve better prediction accuracy. The first potential improvement method is to classify different types of clouds. Since the height and the thickness of the clouds would vary for different types of clouds, their impact on the solar irradiance would also vary a lot. Therefore, cloud



classification can be performed first and different regression models could be constructed accordingly. Also, under conditions with frequent cloud movement, performing cloud tracking is expected to enhance the prediction. If a cloud region is tracked and predicted to be likely to approach the location of sun and cause the irradiance drop, the predicted solar irradiance could be adjusted. The amount of adjustment can be determined from both past data of prediction error and the current cloud type. If a dataset of a longer period can be collected at a certain location, it is also a good idea to refer to the model constructed by the data of the same season and the same weather condition when performing prediction. These potential improvement methods can be developed based on the proposed prediction framework via regression.

### Acknowledgement

The financial support provided by the Bureau of Energy of Taiwan is gratefully acknowledged.

### References

- Calbo, J., Sabburg, J., 2008. Feature extraction from whole-sky ground-based images for cloud-type recognition. *J. Atmos. Ocean. Tech.* 25, 3–14.
- Cazorla, A., Olmo, F.J., Alados-Arboledas, L., 2008. Development of a sky imager for cloud cover assessment. *J. Opt. Soc. Am. A* 25, 29–39.
- Chow, C.W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., Washom, B., 2011. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Sol. Energy* 85, 2881–2893.
- Feister, U., Shields, J., 2005. Cloud and radiance measurements with the VIS/NIR daylight whole sky imager at Lindenberg (Germany). *Meteorol. Z.* 14, 627–639.
- Ghonima, M.S., Urquhart, B., Chow, C.W., Shields, J.E., Cazorla, A., Kleiss, J., 2012. A method for cloud detection and opacity classification based on ground based sky imagery. *Atmos. Meas. Tech.* 5, 2881–2892.
- Glassley, W., Kleissl, J., Dam, C.P., Shiu, H., Huang, J., Braun, G., Holland, R., 2010. California renewable energy forecasting, resource data and mapping. Technical report of California institute for energy and environment.
- Gonzalez, R.C., Woods, R.E., 2002. *Digital Image Processing* 2nd Edition. Prentice Hall.
- Gueymard, C.A., 2004. The sun's total and spectral irradiance for solar energy applications and solar radiation models. *Sol. Energy* 76 (4), 423–453.
- Heinemann, D., Lorenz, E., Girodo, M., 2006. *Solar Irradiance Forecasting for the Management of Solar Energy Systems*, Solar 2006. Denver, CO., USA.
- Heinle, A., Macke, A., Srivastav, A., 2010. Automatic cloud classification of whole sky images. *Atmos. Meas. Technol.* 3, 557–567.
- Huo, J., Lu, D., 2009. Cloud determination of all-sky images under low visibility conditions. *J. Atmos. Ocean. Technol.* 26 (10), 2172–2180.
- Johnson, R., Hering W., Shields, J., 1989. Automated visibility and cloud cover measurements with a solid-state imaging system. Tech. Rep., University of California, San Diego, Scripps Institution of Oceanography, Marine Physical Laboratory, SIO Ref. 89-7, GL-TR-89-0061, 128.
- Johnson, R., Shields, J., Koehler, T., 1991. Analysis and interpretation of simultaneous multi-station whole sky imagery. Marine Physical Laboratory. Scripps Institution of Oceanography, University of California San Diego, SIO 91–3, PL-TR-91-2214.
- Kassianov, E., Long, C.N., Ovtchinnikov, M., 2005. Cloud sky cover versus cloud fraction: whole-sky simulations and observations. *J. Appl. Meteor.* 44, 86–98.
- Kubota, M., Nagatsuma, T., Murayama, Y., 2003. Evening corotating patches: a new type of aurora observed by high sensitivity all-sky cameras in Alaska. *Geophys. Res. Lett.* 30, 1612.
- Li, Z., Cribb, M.C., Chang, F.-L., Trishchenko, A.P., 2004. Validation of MODIS-retrieved cloud fractions using whole sky imager measurements at the three ARM sites. In: Proc. 14th ARMS science team meeting, Albuquerque, NM, Atmospheric Radiation Measurement Program 6, 2–6.
- Long, C.N., Sabburg, J., Calbo, J., Pagès, D., 2006. Retrieving cloud characteristics from ground-based daytime color all-sky images. *J. Atmos. Ocean. Tech.* 23, 633–652.
- Lorenz, E., Hurka, J., Heinemann, D., Beyer, H.G., 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems, IEEE J. of selected topics in applied earth observations and remote sensing 2, 2–10.
- Marquez, M., Coimbra, C.F.M., 2011. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Sol. Energy* 85, 746–756.
- Marquez, R., Coimbra, C.F.M., 2013. Intra-hour DNI forecasting based on cloud tracking image analysis. *Sol. Energy* 91, 327–336.
- Martínez-Chico, M., Batlles, F.J., Bosch, J.L., 2011. Cloud classification in a mediterranean location using radiation data and sky images. *Energy* 36 (7), 4055–4062.
- Perez, R., Seals, R., Ineichen, P., Stewart, R., Menicucci, D., 1987. A new simplified version of the perez diffuse irradiance model for tilted surfaces. *Sol. Energy* 39, 221–231.
- Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F., 2002. A new operational model for satellite-derived irradiances: description and validation. *Sol. Energy* 73, 307–317.
- Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., K., Renné, D., Hoff, T., 2010. Validation of short and medium term operational solar radiation forecasts in the US. *Sol. Energy* 84, 2161–2172.
- Pfister, G., McKenzie, R.L., Liley, J.B., Thomas, A., Forgan, B.W., Long, C.N., 2003. Cloud coverage based on all-sky imaging and its impact on surface solar irradiance. *J. Appl. Meteorol.* 42, 1421–1434.
- Reda, I., Andreas, A., 2004. Solar position algorithm for solar radiation applications. *Sol. Energy* 76, 577–589.
- Remund, J., Perez, R., Lorenz, E., 2008. Comparison of solar radiation forecasts for the USA. In: Proc. of the 23rd European PV Conference, Valencia, Spain, 1.9–4.9.
- Roy, G., Hayman, S., Julian, W., 2001. Sky analysis from CCD images: cloud cover. *Light. Res. Technol.* 33 (4), 211–222.
- Sabburg, J., Wong, J., 1999. Evaluation of a ground-based sky camera system for use in surface irradiance measurement. *J. Atmos. Ocean. Technol.* 16, 752–759.
- Schroedter-Homscheidt, M., Hoyer-Klick, C., Rikos, E., Tselepis, S., Pulvermüller, B., 2009. Nowcasting and forecasting of solar irradiance for solar energy electricity grid integration, In: International symposium on remote sensing of environment.
- Shapiro, L., George, C.S., 2001. *Computer Vision*. Prentice Books, Upper Saddle River.
- Shields, J., Karr, M., Burden, A., Johnson, R., Hodgkiss, W., 2007. Continuing support of cloud free line of sight determination including whole sky imaging of clouds, final report for ONR contract N00014-01-D-0043DO #13, Marine physical laboratory, Scripps Institution of Oceanography, University of California San Diego, Technical Note 273.
- Shields, J., Karr, M., Burden, A., Johnson, R., Mikuls, V., Streeter, J., Hodgkiss, W., 2009. Research toward Multi-Site Characterization of Sky Obscuration by Clouds, final report for grant N00244-07-1-009, Marine Physical Laboratory, Scripps Institution of

- Oceanography, University of California San Diego, Technical Note 274.
- Singh, M., Glennen, M., 2005. Automated ground-based cloud recognition. *Pattern Anal. Appl.* 8, 258–271.
- Wood-Bradley, P., Zapata, J., Pye, J., 2012. Cloud tracking with optical flow for short-term solar forecasting. In: 50th conference of the australian solar energy society, Melbourne, December.