The authors would like to thank the reviewers and the editor for their valuable comments to help improve the quality of the manuscript. The reviewers' comments are addressed in this document.

Referee #1:

Very nice work! A couple questions:

1) Where was the camera images acquired? Was this part of another observational campaign?

The images were acquired in National Central University, Taiwan.

2) In general, please state your equations in the text (with the full equation and equation number) BEFORE discussing them in paragraph form

Thank you for the comment. We have modified the manuscript according to your comment.

3) You have done a very nice analysis of multiple classification schemes to improve pixel segmentation into clear and cloudy regions. However, the reasons for your choices are not always clear. Why did you pick these four classifiers, and not others? Why do you select a 3x3 neighborhood at each Level? Why do you select 5 levels, with sizes 5, 10, 15, 20, and 25 pixels? Wouldn't it be better to have an odd number of pixels, so the pixel of interest can be in the center of the tile? (like in your Figure 3.

Thank you for the comment. For the motivation of choosing classifiers, we chose some of the popular classifiers (including random forest, SVM, and Bayesian classifier) to test their classification performance on this application. RBR thresholding is also used because it is a classic way to detect clouds. And we have discovered that combining the knowledge of different experts can yield better classification results.

There are indeed many other classifiers and we did not try every existing classifier. But the proposed framework can be extended when the classification results of more classifiers are added in the final voting decision process. When results from more classifiers are added, only the parameter N_v needs to be adjusted in the framework.

For multiscale information, at each level, a 3x3 neighborhood can cover the neighborhood at its own level and the level below it and above it. It is an effective way to construct multiscale information at a specific level.

As for the number of levels is set as 5, we can observe from Fig. 6 that when the patch sizes are too large or too small, the classification accuracy at that level decreases. Therefore we feel that we should not include more levels.

To test the number of levels required to yield better detection results, we plot the detection accuracy using different number of levels in Fig. 8. Note that for the 6th level and 7th level, the size of the local image patch is $L_6 = 30 \times 30$ and $L_7 = 35 \times 35$. We can observe that using 4 or 5 levels results in better detection accuracy. When involved with levels with image sizes that are too large, the detection accuracy drops.

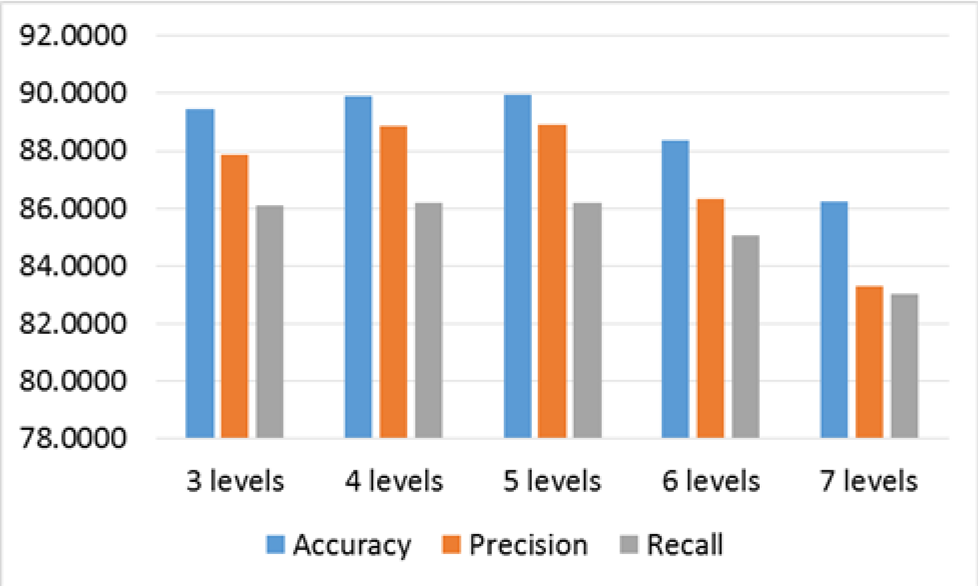

Fig. 8 Detection accuracy uing different number of levels

Some of the sizes of the patches are not odd numbers. The pixel of interest not being in the center of the patch would not have large impact on feature extraction. We had also performed experiments using patches of 5x5, 11x11, 15x15, 21x21, and 25x25, and the final performance is very similar. And we decided to use fixed increment on patch sizes.

4) Are you advocating that people adopt this procedure when segmenting sky pixels? Figure 9 appears to try to make this point.

Thank you for your good work and nice manuscript.

Thank you for the comment. Yes, the proposed method exhibits better performance for segmenting sky and cloud pixels.

Referee #2

- Previous works dealing with the same matter (cloud identification in sky images) by the same author (HY Cheng) should be more clearly reviewed and results compared with the present approach.

Thank you for the comment. The previous work on clouds focused on cloud classification. The clouds were classified into six classes (Cirrus, Cirrostratus, Scattered cumulus or altocumulus, Cumulus or cumulonimbus, Stratus, Clear sky) based on blocks in the previous work. The previous work did not perform cloud detection. In other words, it did not indicate which pixel in the all-sky image belongs to cloud or sky.

As indicated in the previous work, the current cloud detection methods need to be improved if we want to perform component based cloud classification. Therefore, this is one of the motivations that we propose an improved cloud detection method in this work.

- References should be checked throughly; at least on reference (Heinle et al 2010) is not mentioned in the text.

Thank you for the comment. We have checked and corrected the references and the corresponding citations thoroughly.

Referee #2

Received and published: 12 September 2016

General comment.

This paper addresses a relevant question as it is the automatic processing of All-sky images to get information about the sky conditions. Specifically, and unlike other paper by the same authors, this paper focuses on distinguishing between cloudy and clearsky pixels in an image. The title and the abstract are explicative enough. The paper does not present very novel concepts, but it introduces a combination of classifying methods that may be useful for the community of researchers (and company engineers) that are dealing with the issue of cloud detection in ground based sky images. In this sense, the paper concludes that the suggested method overcomes other earlier methods from the literature, in particular the broadly used red-to-blue (RBR) method.

In general, the scientific methods and assumptions are valid and clearly outlined, and the results are sufficient to support the conclusions. The traceability of results is somewhat complicated, as some parts of the paper are not fully understandable, from my point of view (see my specific comments below). Overall, however, the language is good enough, and the content is correctly structured. Mathematical symbols and abbreviations could be simplified sometimes (see comments below).
The authors give credit to related work and indicate their original contribution, although some previous works could be cited in a different way (see my specific comments below).
The paper can be published at AMT, provided that the authors consider the comments below in a revised version of the paper.

Specific comments.
Abstract. When the authors say "The classic method for cloud detection is based on thresholding of the red blue ratio of an image pixel", they could add that other methods have been suggested.

> Thank you for the comment. We have modified the abstract in the revised manuscript.

Introduction, first paragraph. At the end, the authors talk abobut "recent developments" while in fact, all devices mentioned are more than 10 years old (as the references are).
Right now there are several sky imagers (either commercially available or not) that are more recent. Authors must keep these references as pioneering works, but should add that new, more recent developments exist. In addition, some of these references could be placed below, when the authors refer to works devoted to classifying clouds, as this was the goal of most of those papers.

> Thank you for the comment. We have modified the manuscript as follows.
> "Devices developed to monitor the sky presented in some of the pioneering works include Whole Sky Imager (Kassianov et al., 2005; Li et al., 2004), Whole Sky Camera (Long et al., 2006), All-Sky Imager (Kubota et al., 2003), and Total Sky Imager (Pfister et al., 2003). More recent commercial products include all-sky cameras by Eko Instruments, Oculus, SBIG, etc."
> Since the newer camera models are mostly described in commercial web pages instead of academic papers, we did not add new references of new devices.
> Also, some of the references about cloud detection and cloud classification have been placed below, where we refer to works of detecting and classifying clouds.

Introduction, second paragraph. It is pretty obvious that cloud cover is critical for solar irradiance assessment. You don't need to add the reference Cazorla et al (2008) here, since that work is focused on sky imaging. A more general reference (from a text book or similar) or no reference at all would be ok.

Thank you for the comment. We have removed the reference of Cazorla et al. (2008) in the context.

Introduction, third paragraph. Please write "red to blue ratio (RBR)" instead of "red blue". In addition, you should mention that Long et al (2006) already suggested using several thresholds depending on the relative cloud/sun/horizon position.

Thank you for the comment. We have corrected the manuscript using "red to blue ratio" instead of "red blue". We have also added that the work by Long et al. (Long et al., 2006) suggested that different thresholds should be selected depending on the relative position of the pixel being classified in contrast to the positions of sun and horizon.

Introduction, last paragraph. Although they are defined later, I think that RGB, HSV, and YCbCr should be defined here, which is the first time these acronyms appear. Contrarily, the authors don't need to define again RBR in the first paragraph of section 2.

Thank you for the comment. We have added the full component names for color models RGB and HSV. However, YCbCr is the name of a color model, which is not an acronym. Therefore, we still defined YCbCr in Section 2.2. We have also removed the redundant definition of RBR in the first paragraph of section 2.

Section 2.1. How w1 and w2 are determined? Are the values "a priori" fixed or are they a result also from the Hough line transform?

Thank you for the comment. The values of w1 and w2 are determined depending on the size of the all-sky images. In our experiments, we set w1 and w2 as 60 and 12 pixels, respectively. We have added the information in the manuscript.

Section 2.2. It is not clear to me which color model use the "original" or "raw" images. Are they codified in the three color models? Or do your transform from one model to the others? Are these color model independent, or can they be derived from each other? The authors must think that readers of the geophysical community are not necessarily familiar with color models, so more explanation is needed here.

Thank you for the comment. The color models are not completely independent. There are equations that can transform RGB color components into HSV and YCbCr color components. We have added explanations and references of color space transformation in the manuscript.

Section 2.6. The authors should make an effort to improve the explanations in this section, as well as to simplify mathematical symbols. From my point of view, this section is difficult to follow, and Figure 3, that should help, is not explanatory.

Thank you for the comment. We have revised Section 2.6 and added an example with number of votes in Fig. 3.

Section 3, first paragraph. So the same 250 images are used for training and for assessing the classifier? Please state this clearly. What do the authors mean by "10-fold cross validation"? Definition of symbols TP, TN,: : : could go after Equations 5, 6,7.

Thank you for the comment. Ten-fold cross validation means that the dataset is divided into 10 none-overlapping subsets. Nine subsets are used for training, and the remaining one subset is used for testing. Then the training subsets and testing subsets are rotated for 10 times. The average classification rate of these 10 experiments is the 10-fold cross validated classification accuracy. Therefore, different images were used for the classifier training and testing.
We have moved the definition of symbols TP, TN, … after Equations 5, 6, 7.

Section 3, second paragraph. From my point of view, this paragraph should not start with "According to Long (Long et al., 2006), the recommended RBR threshold is 0.6." In my opinion, the authors should start this paragraph by explaining that the RBR method will be used as baseline (as they say later). In addition, the sentence referring to Long et al (2006) could be rewritten, since the value of 0.6 was recommended for one system (the WSC) in that paper, while for the TSI several thresholds are used depending on the relative position between pixel/sun/horizon.

Thank you for the comment. We have revised the manuscript.
In this work, the RGR thresholding method proposed by Long (Long et al., 2006) will be used as the baseline method for comparison. In Long's work, an RBR threshold is recommended for the Whole Sky Camera and several thresholds are suggested to be used for the Total Sky Imager. Since the desired threshold varies due to different devices and weather conditions, we perform an experiment to test the best threshold for our all-sky camera. Also, to avoid false positive detection at highlighted regions around the sun, we employ an upper bound threshold.

Section 3, paragraph regarding Thr_PCA. I don't understand why the accuracy gets worse for higher Thr_PCA. I understand the use of PCA, as this means that feature vector reduce its dimension and the whole process is accelerated. But it seems to me that a higher Thr_PCA means including more information in the classifier, so why results deteriorate?

Thank you for the comment. When the dimensionality is high, reducing feature dimensions would often improve the performance of the machine learning models. A proper Thr_PCA usually locates between 90%~99% and is selected empirically. Typically, the accuracy of classification would increase as Thr_PCA goes from 100% (which means no dimensionality reduction at all) to 99%. The accuracy of classification would continue increasing until Thr_PCA reaches a certain value (caused by the benefit of dimensionality reduction). After that, the accuracy of classification would start to decrease due to too much information loss. We have added the explanation in the manuscript.

Figure 8. For these or for other examples, the authors could also show the result of the RBR method, to illustrate the differences.

Thank you for the comment. We have added the results of RBR method to illustrate the differences.

Conclusions. I wouldn't say that the RBR is not "feasible". This method produces somewhat poorer results, but is easily applied, so I would recommend a different wording here. Finally, the authors should explain further how the new method can be implemented to other cameras. Do other researchers need to select a number of images, set "ground truth" and train the method? Are the values that are set her for thresholds and voting (Thr_PCA, Nv, etc.) adequate for all cases, or must be "tuned" for each site?

Thank you for the comment. We have revised the conclusions. The experimental images are captured at the same site but the taken in different seasons and under various lighting conditions. The experimental images are captured by only one all-sky camera, so we cannot validate that how well the trained classifier would work on images captured by different devices. However, we would expect that adding images captured by various cameras to the training set can improve the detection robustness of the detection model over different cameras. For the selection of parameters Thr-PCA and Nv, if the number of levels and feature length are fixed, the desired parameters should not be seriously affected even if the training samples are changed.