TOKAI UNIVERSITY 東海大学 情報技術センター

Futher attempts and method developments of EarthCARE MSI Cloud flag validation using whole-sky camera images

Minrui WANG[1], Takashi Y. NAKAJIMA[1]

1. Tokai University Research & Information Center

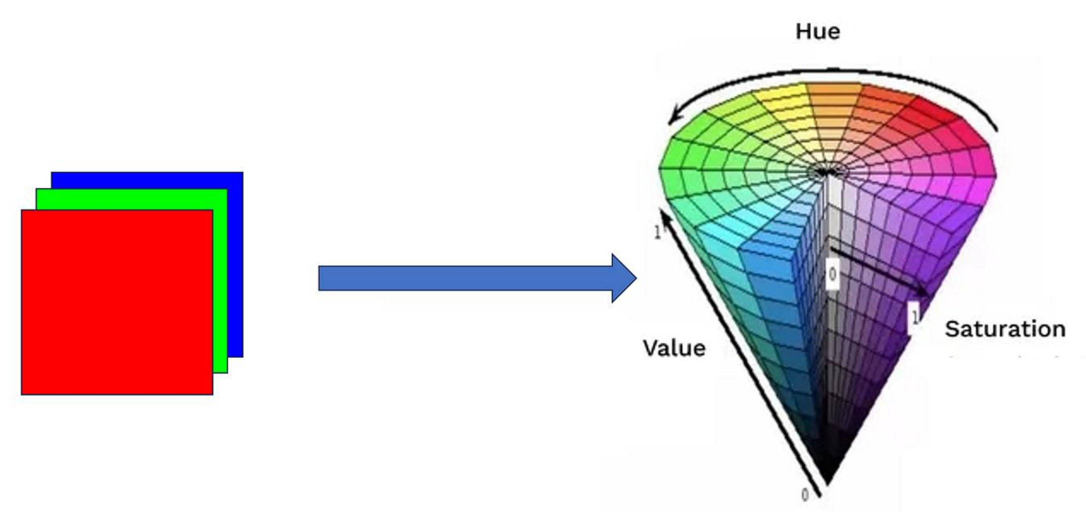
Introduction

Although satellite observations are effective for monitoring clouds over wide areas, their accuracy may decrease over bright surfaces such as urban areas or vegetation, and they may have difficulty detecting small or low-altitude clouds. Therefore, validation using ground-based camera observations is necessary.

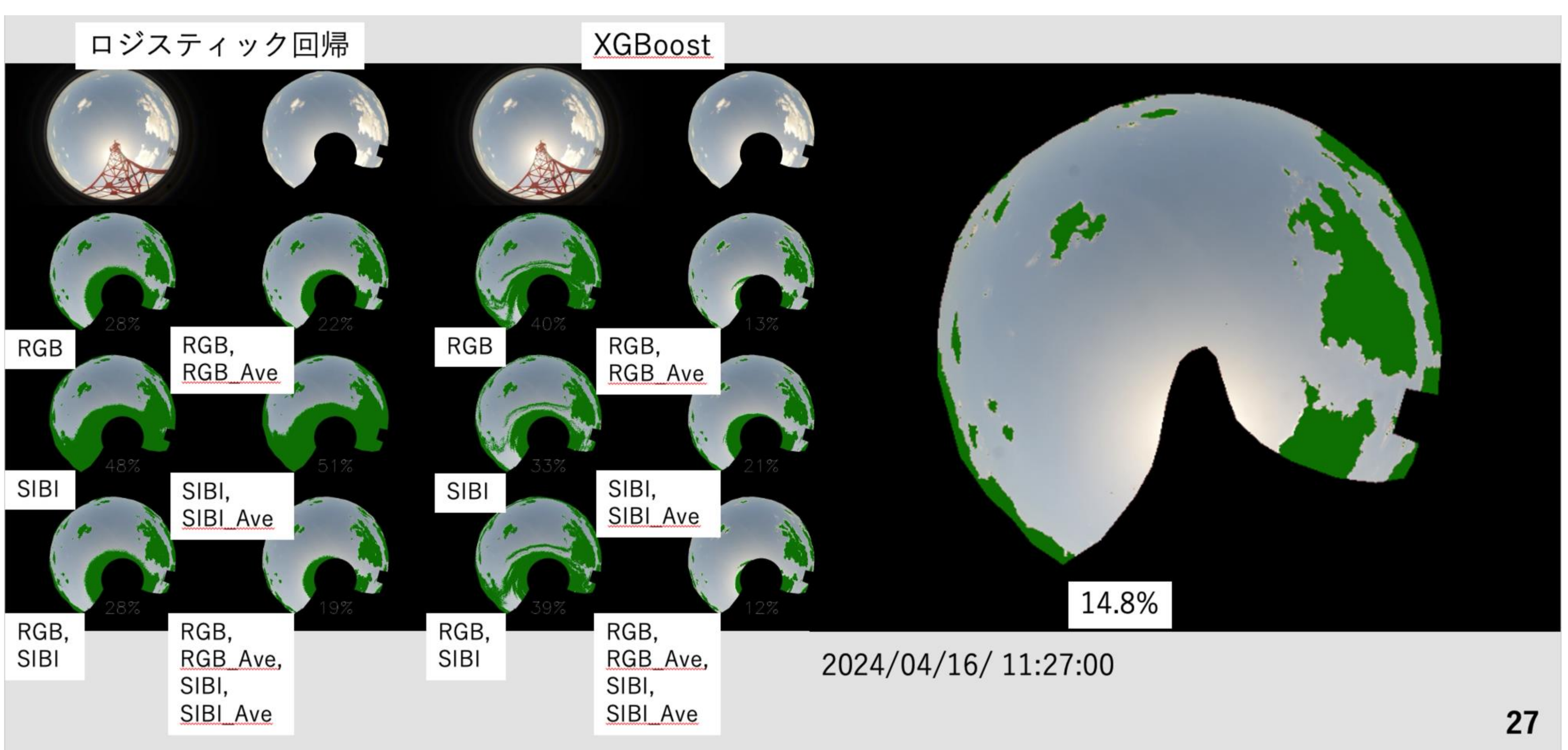
Previous studies estimated cloud areas using the SI–BI method, which classifies each pixel based on RGB information, and later automated the threshold setting by applying machine learning methods such as logistic regression and XGBoost. However, these methods still had difficulties under strong solar influence, especially around the sun, where halation and scattered light can be mistakenly classified as clouds. As a result, ground-based cloud fraction sometimes showed nonzero values even when satellite cloud fraction was zero.

HSV Method in cloud fraction validation

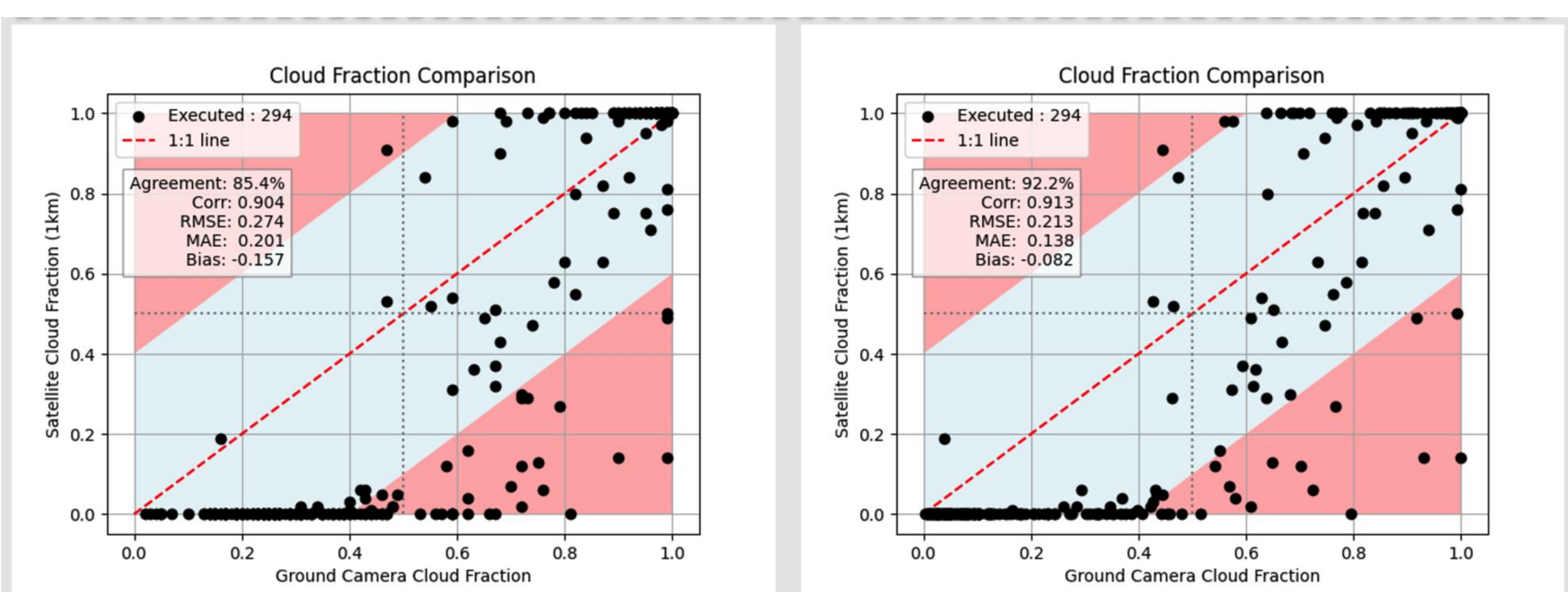
To solve this problem, this study introduces a new method based on HSV color-space conversion, similar-color region segmentation, feature extraction, and supervised machine learning using a Random Forest classifier. HSV color space is used because it separates brightness from hue and saturation, making it more robust against changes in illumination near the sun. Instead of classifying only "cloud" and "sky," the proposed method classifies image regions into five classes: sky, cloud, sun, cloud affected by solar influence, and scattering caused by water vapor or haze. Training data were created by manually labeling segmented regions from whole-sky camera images taken in 2023 and 2024.



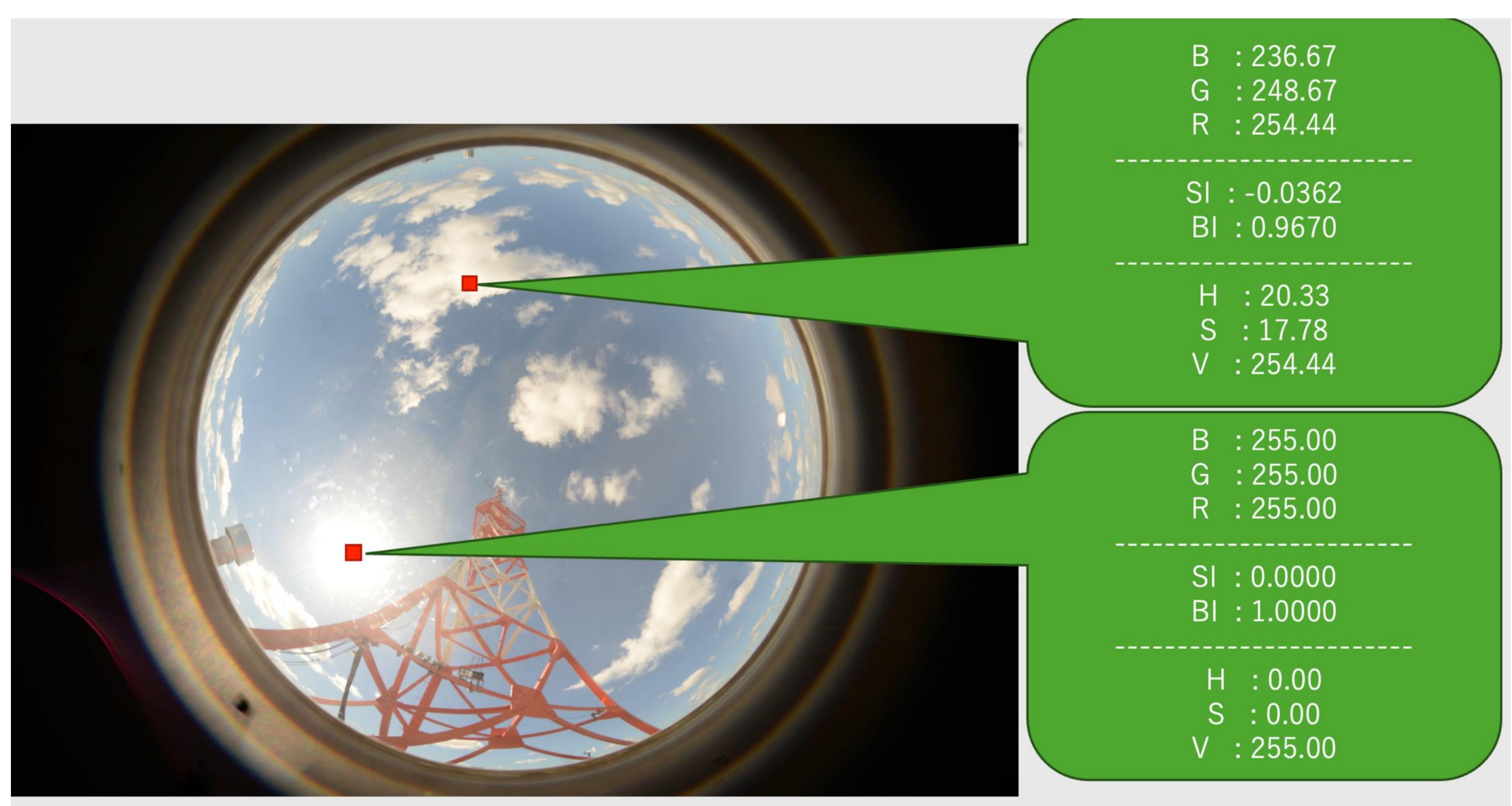
Result



The proposed method was then compared with previous methods in difficult classification cases. In one example, previous methods estimated cloud fraction values ranging from 19% to 51%, while the proposed HSV-based Random Forest method estimated 14.8%, which better represented the actual sky condition. In addition, solar position calculations were used to overwrite classification results near the sun based on probability, further reducing false cloud detection caused by solar effects.



Finally, cloud fractions calculated from ground-based whole-sky camera images were compared with satellite-derived cloud fractions based on Cloud_Flag data calculated by the CLAUDIA cloud detection algorithm. Scatter plots were created using data from the same location and time, and several evaluation metrics were used, including agreement rate, correlation coefficient, RMSE, MAE, and bias. Compared with the previous method, the proposed method improved the agreement with satellite data. For example, the agreement rate increased from 85.4% to 92.2%, RMSE decreased from 0.274 to 0.213, MAE decreased from 0.201 to 0.138, and bias improved from -0.157 to -0.082.



Future work includes increasing the training dataset, retraining the model using images from additional times of day, and applying the method to comparisons with other satellites such as GCOM-C, EarthCARE, and Himawari.