Readme File

Singapore Whole sky IMaging CATegories (SWIMCAT) Database

May 19, 2015

Contact:
Stefan Winkler
Stefan.Winkler@adsc.com.sg

1 Introduction

The SWIMCAT database is introduced and used in the following paper:

S. Dev, Y. H. Lee, and S. Winkler: "Categorization of cloud image patches using an improved texton-based approach." In *Proc. IEEE International Conference on Image Processing (ICIP)*, Québec City, Canada, Sep. 27-30, 2015.

If you use or adapt any part of this dataset, please cite the above paper.

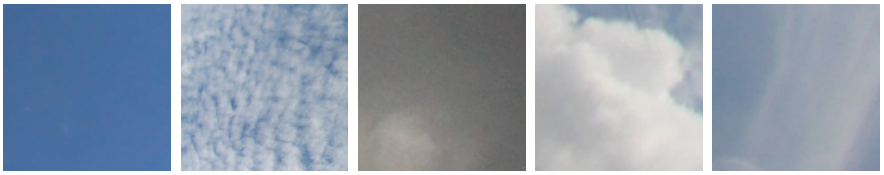
2 Description

The SWIMCAT dataset contains 784 images of sky/cloud patches, categorized in 5 distinct categories, as specified in Table 1:

Category	**Number of Images**	**Type**
A	224	Clear Sky
B	89	Patterned clouds
C	251	Thick dark clouds
D	135	Thick white clouds
E	85	Veil clouds
Total	784	

Table 1: Different categories of sky/cloud images patches.

All images have a dimension of 125×125 pixels, and were captured using WAHRSIS, a calibrated ground-based whole sky imager. The imager is located at Nanyang Technological University in Singapore (1.34°N, 103.68°E). The images were captured over a period of 17 months from January 2013 till May 2014. They were selected based on visual characteristics, and were categorized after consultation with experts from Singapore Meteorological Services. Sample representative images from the 5 categories are shown below in Fig. 1.



(a) (b) (c) (d) (e)

Figure 1: Sample images from the database: (a) Clear sky (b) Patterned clouds (c) Thick dark clouds (d) Thick white clouds (e) Veil clouds.

3 Database Content

The images and metadata of the SWIMCAT dataset are contained in five folders, corresponding to the five image categories:

- A-sky
- B-pattern
- C-thick-dark
- D-thick-white
- E-veil

Each of the cloud category folders have a folder called "images" and a comma-delimited .csv file. The "images" folder contains all the image patches of the particular category. The comma-delimited .csv file contains all the related metadata information pertaining to the individual images.

The image files are named as follows: <CategoryID>_<ImageNumber>img.png, where <CategoryID>$\in \{A, B, C, D, E\}$ according to the image category as per Table 1, and <ImageNumber>$= \{1, 2, \ldots\}$ is the serial number of the image for a particular category, for example, `B_3img.png`.

The contents of the .csv metadata files are as follows.

1. **File**: Name of the corresponding sky/cloud image file in the "images" subfolder.
2. **Date**: Capture date of the image in YYYYMMDD format.
3. **Hour**: Hour of capture of the image, ranges from 0 to 23.
4. **Min**: Minute of capture of the image, ranges from 0 to 59.
5. **Sec**: Second of capture of the image, ranges from 0 to 59.
6. **Fnumber**: F-number of the camera.
7. **ShutterSpeed**: Shutter speed (in milliseconds) of the camera.
8. **ISO**: ISO setting of the camera.
9. **Elevation**: Elevation angle of the virtual camera direction.
10. **Azimuth**: Azimuth angle of the virtual camera direction.

4 Licensing Information

The dataset is released under a Creative Commons license (`https://creativecommons.org/licenses/by-nc/4.0/`). You are free to:

- Share – copy and redistribute the material in any medium or format
- Adapt – remix, transform, and build upon the material

subject to the following terms:

- Attribution – You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- NonCommercial – You may not use the material for commercial purposes.

Full details can be found in the licensing file.

5 Additional Files

- license.html: This is the licensing file that we recommend you read before using or sharing this dataset.
- readme.pdf: This file.