

SkyGAN: Realistic Cloud Imagery for Image-based Lighting

Martin Mirbauer^{†1} , Tobias Rittig^{†1} , Tomáš Iser¹ , Jaroslav Křivánek^{1,2} , and Elena Šikudová¹ 

¹ Charles University, Faculty of Mathematics and Physics, Czech Republic

² Chaos Czech a.s., Czech Republic

1. Introduction

This is the supplemental document to the paper published in Computer Graphics Forum. In here we describe our capturing setup (Section 2), the processing of photographs (Section 3) and the dataset (Section 4) in more detail.

2. Capturing Setup

Capturing the sky seems simple at first glance but bears some unexpected difficulties. We have solved many of them over the years and thus would like to provide our insights here.

Camera and lens We used a Canon 5D MkII at first, later a Canon 6D MkII together with a Canon 8-15mm f/4.0L USM fisheye lens set to 8mm (circular fisheye). The fact it is a zoom lens is a drawback when you point the lens to the sky for hours on end and gravity overcomes the friction in the zoom mechanism. This creates a unwanted slow zooming effect in the time-lapse. We overcome this problem with a rubber band over the zoom ring and attached to the tripod head rail that is under constant tension and keeps the lens at its minimum focal length. A fixed circular fisheye lens would be preferable.

HDR Capturing an extended dynamic range with limited range sensors usually involves exposure bracketing where you take multiple exposures with different exposure times in quick succession. In post-processing, the HDR image is then reconstructed by fusing all exposures together.

With the Canon 5D camera we used the alternative firmware Magic Lantern to capture 9 exposure stacks starting from the lowest exposure time $\frac{1}{8000}$ s with aperture 22. The 6D has built-in HDR mode however only with 7 exposures and a slower exposure time of only $\frac{1}{2000}$ s.

The small aperture necessary to capture the most of the sun's brightness, but unfortunately still not enough to depict it unclipped. For that one would require ND (neutral density) filters to be placed behind the fisheye lens, which in turn increase exposure times to multiple seconds for the brightness where clouds and sunsets are visible. In effect the clouds would be heavily blurred with higher wind speeds which is very undesirable so we have not opted for an ND filter.



Figure 1: Our capturing setup courtesy of [Hoj19].

Tripod For stable positioning we use a sturdy Manfrotto aluminum three-stage tripod equipped with a Novoflex VR-System PRO II tripod head as shown in Figure 1. The head allows us to position the camera perfectly vertical and rotate it around the vertical axis to have the flash-mount point north. This is our chosen extrinsic calibration that we try to keep as consistent as possible. One could fix some deviations in post-processing, but it is always better to avoid them in the first place.

As the camera is usually positioned in an open field or on a rooftop wind becomes an issue for longer exposure times. We stabilize the tripod further by hanging a heavy bag in the center of it which increases its inertia.

Battery As the goal is to capture as long as possible time-lapse images per day there are only two technical limitations (besides the weather): First, the battery of the camera and second the storage capacity of the SD card. With cards exceeding 256GB available at an affordable price point storage is less of an issue, if the card is emptied once per day.

Supplying the camera with power is a bigger issue. We have solved this with a cheap, third-party battery adapter that converts 2x USB-A plugs to the required 7.2 V. Together with standard 74 Wh power-banks, the camera can run for multiple days on the same charge. The heavier power bank and attached electronics are also placed in the waterproof bag hanging from the tripod with only one cable running to the camera on top.

3. Dataset Processing

We have developed a pipeline for automatic processing of RAW image files into brightness, colour, and geometry-calibrated OpenEXR images. The combination of time-lapse photography (every 30 s) and HDR stacking (7 exposures/image) creates 840 images/h (up to 25 GB/h) that all need to be processed.

With the amount of data in the order of terabytes, input/output (IO) performance and disk space become an issue. We thus perform processing in a streaming fashion using a RAM disk for storing intermediate files. The pipeline is designed for a human-in-the-loop workflow, as it is able to recover from failures in consecutive runs (e.g. incompletely exported sessions) or performing a dry-run analysis. This allows the user to fix any inconsistencies in the data retroactively by manipulating the input, deleting the corresponding false output, and rerunning the pipeline.

Assumptions imposed on the input data are kept to a minimum: one capturing session is simply a folder with RAW files all taken with the same camera settings. Otherwise, the EXIF data is used to infer required processing parameters automatically. We first sort the images by capture time, then group multiple exposures together, and finally validate that the camera settings did not change (which would require a session split). Operations are performed one exposure stack at a time (typically using 8 threads), but multiple processes can be run in parallel on many-core machines.

For conversion of the camera’s RAW files, the pipeline relies on *RawTherapee* [HR05], an open-source raw image processor. We use a minimalistic processing profile, avoiding any introduction of error while still taking advantage of the program’s demosaicing, denoising, and colour-processing algorithms. The built-in lens profile is used for the correction of lens distortion and vignetting. White-balancing the images is challenging, as the illumination changes constantly. We opt for a fixed white balance to stay consistent throughout the dataset and set the value to 5400 K. The intermediate images are exported as 16-bit TIFF files into a RAM disk, where they are accessible for the next step.

We take advantage of the computationally-heavy RAW processing being done once and export multiple panorama projections from the same intermediate files. Stacking the TIFF files into HDR images is done simultaneously with reprojection using *PTGui Pro* [New01], a commercial panorama stitching software. We typically export a 1024×1024 stereographic image pointing to the

zenith and a full-resolution equirectangular image (8192×4096). A consistent brightness throughout the dataset is achieved by computing an exposure offset per session. From the exposure time, ISO, and aperture value found in the EXIF metadata, we compute the exposure value (EV) and shift it such that the low dynamic range (LDR) range corresponds to 14 on the EV_{100} scale. Because *PTGui* assumes the central exposure to have a (relative) $EV = 0$, our global EV shift has to be applied to all exposures equally.

Allowing for visual verification, we generate a preview video per session from the stereographic images using *FFmpeg*. Hence, the user can check if the sun trajectory is smooth and if there are no flickering or unusual rotations (e.g. caused by the camera not pointing north or the accelerometer rotating the images) present.

4. Dataset Description

We are providing the dataset consisting of 39,000 images in three different formats:

- The RAW camera images are 6264×4180 in .CR2 format and can be obtained on request. Their size and quantity (7× more) prevents us from hosting them publicly on the internet.
- One processed version are images in stereographic (up) projection with a resolution of 1024×1024 in JPG and EXR format.
- The second processed version are images in equirectangular projection (“latlong”) with a resolution of 8192×4096 in JPG and EXR format. The bottom half of the image is black as our dataset only covers the hemispherical sky.

The images were taken in 16 different locations, of which 11 are located in the Czech Republic, three in Germany and two in California. The durations of the shooting sessions are plotted in [Figure 2](#).

Our dataset consists primarily of partly cloudy or clear sky photographs, but also contains cloudy and overcast sky. Notably, apart from a few cases, it lacks storm/rain clouds – because the camera is not waterproof, we were often forced to interrupt our shooting session due to the weather conditions. A set of example images from our dataset is shown in [Figure 3](#).

An up-to-date description of the dataset and download instructions can be found at <https://github.com/CGGMFF/SkyGAN> [MRI*23].

References

- [Hoj19] HOJDAR, ŠTĚPÁN. “Using neural networks to generate realistic skies”. MA thesis. Charles University, Faculty of Mathematics and Physics, 2019. URL: <http://hdl.handle.net/20.500.11956/1091311>.
- [HR05] HORVÁTH, GÁBOR and RAWTHERAPEE CONTRIBUTORS. *RawTherapee*. 2005. URL: <https://www.rawtherapee.com/2>.
- [MRI*23] MIRBAUER, MARTIN, RITTIG, TOBIAS, ISER, TOMÁŠ, et al. *CGGMFF/SkyGAN*. 2023. URL: <https://github.com/CGGMFF/SkyGAN2>.
- [New01] NEW HOUSE INTERNET SERVICES B.V. *PTGui: Photo Stitching Software*. 2001. URL: <https://ptgui.com/2>.

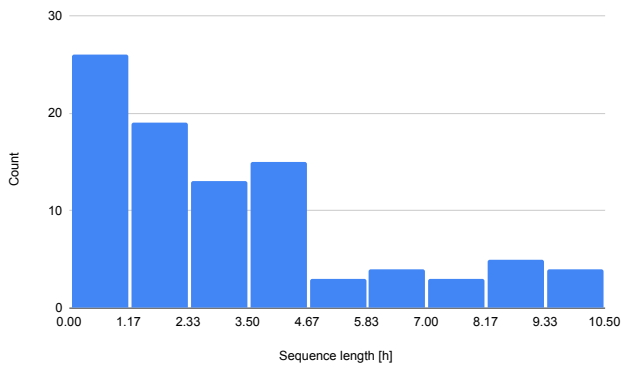


Figure 2: Histogram of the duration of our 92 shooting sessions showing the range of consecutive time-lapse data. In total this sums up to almost 300 h of capturing with an average length of 192 min per session.

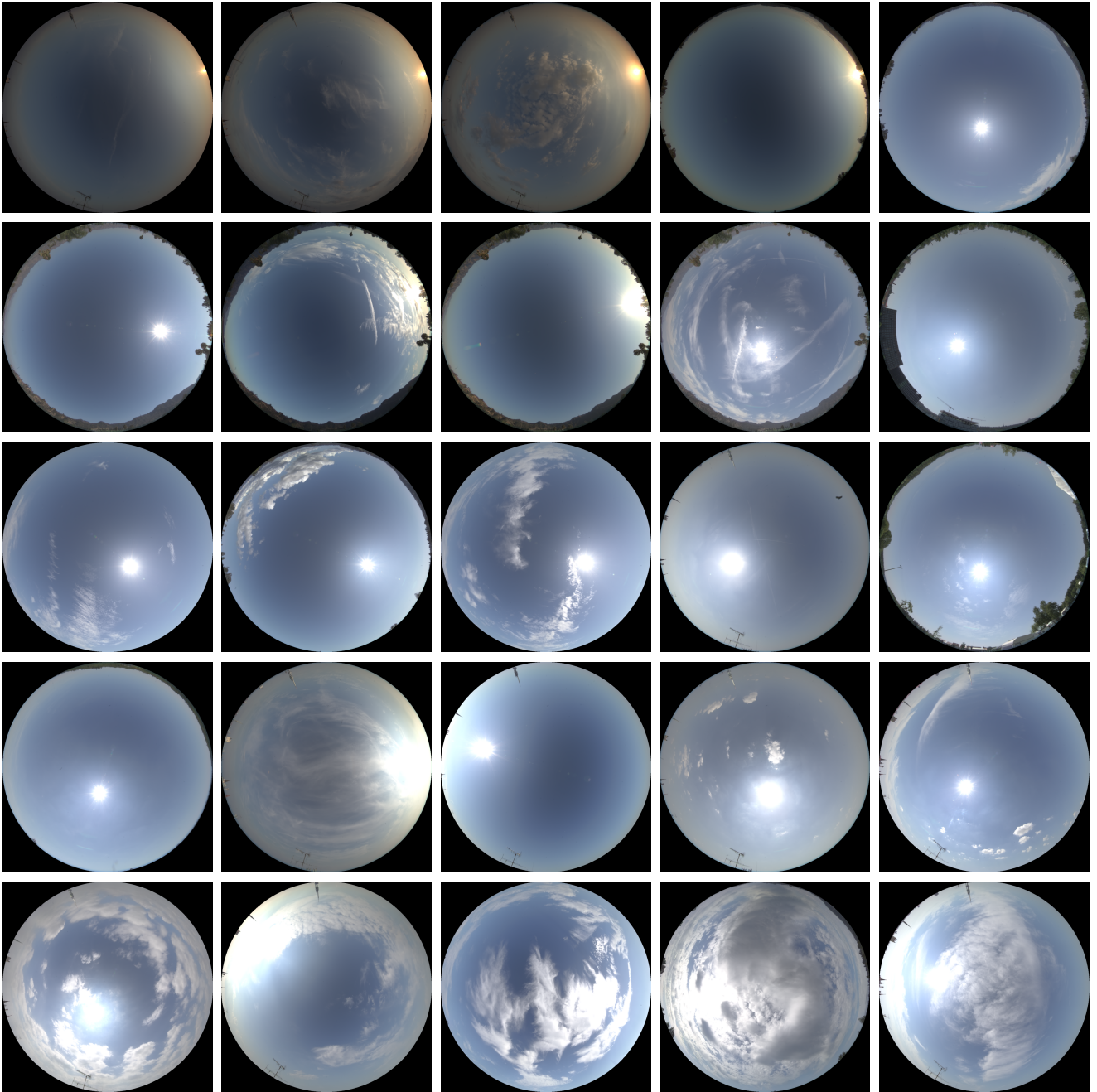


Figure 3: A curated preview of our dataset showing the variety of cloud patterns we have captured. The images are processed, hemispherical HDR images in stereographic projection. North is pointing upwards in the images with sunrise (East) being on the left and sunset (West) on the right.