

Clouds in The Cloud

Anonymous ACCV 2014 submission

Paper ID 215

Abstract. Light-field imaging can be scaled up to continental size, to map the Earth's atmosphere in 3D. Multiview spaceborne instruments suffer low spatio-temporal and angular resolution, and are extremely expensive and unscalable. We develop lightfield imaging of the sky, by a scalable network of wide-angle cameras looking upwards, which upload their data to the cloud. This new type of imaging-system poses *new computational vision and photography problems*, some of which generalize prior monocular tasks. These include radiometric self-calibration across a network, overcoming flare by a network, and background estimation. On the other hand, network redundancy offers *solutions* to these problems, which we derive. Based on such solutions, the light-field network enables unprecedented ways to measure nature. We demonstrate this experimentally by 3D recovery of clouds, in high spatio-temporal resolution. It is achieved by space carving of the volumetric distribution of semi-transparent clouds. This sensing approach can complement satellite imagery, be useful to aviation meteorology and study of birds. The radiometric capabilities can make aerosol tomography realizable, and give new, powerful tools to atmospheric scientists.

1 Introduction

Following the introduction of plenoptic imaging to computer vision [1], light-field imaging [2–6] has had a significant impact on sensing. This modality samples the radiance distribution in location and direction. It has so far been used mainly in small-scale setups. However, it can be scaled up to continental size, to map the Earth's atmosphere in 3D. Sampling the atmospheric radiance spatio-angularly is achieved by a few spaceborne and airborne instruments, including the Multiangle Imaging SpectroRadiometer (MISR) [7, 8], the Airborne Multiangle SpectroPolarimetric Imager (AirMSPI) [9, 10] and POLDER [11–13]. However, these imaging architectures suffer from several drawbacks. They have crude resolution spatially (effectively several kilometers per pixel), angularly (≈ 7 angles per view), and temporally (orbit takes several days to return to the same terrestrial spot). Furthermore, spaceborne instruments are extremely expensive and unscalable. Hence, in this work we develop a counter approach: the atmospheric lightfield is captured from below, by wide-angle cameras looking upwards. Moreover, instead of expensive one-off instruments, our approach is a *scalable sensor network* that captures images simultaneously over a very large area, densely.

Creating and exploiting such a light-field imaging-network poses several requirements: low-cost units, communications and *tailored computational photography algorithms*. The first two requirements are met thanks to cellular-network infrastructure, cellular-based low-cost cameras and *cloud computing* services. Hence, we can deploy

solar-powered cameras at will, nearly anywhere. By wireless, they upload their sky-images to the “cloud”, from which the light-field data can be analyzed. However, this new type of imaging-system gives rise to new problems and algorithms, part of which we deal with in this paper. In a sense, the network generalizes some computer vision problems that had been posed for monocular setups a decade ago. On the other hand, network redundancy offers *solutions* to these problems.

The computational photography problems include radiometric self-calibration across a network of cameras, background estimation and overcoming saturation and flare by a network. Consequently, the light-field network enables unprecedented 3D imaging of clouds, in high spatio-temporal resolution. We demonstrate the approach using real field experiments. As a result, we believe this approach can complement multi-angular satellite cloud imagery, perhaps make aerosol tomography [14] realizable, and offer new capabilities to study weather phenomena and birds.

2 Background

2.1 Monocular radiometric self-calibration

A large network should use low-cost camera units. Such cameras often have spatial and temporal radiometric inconsistencies. For a single camera, consistent readouts can be obtained in the field by self-calibration. The strongest methods rely on redundant images, taken at modulated settings. Images taken during slow panning yield spatial non-uniformity correction (NUC) of offset and gain (e.g., by vignetting [15, 16]). For example, spatial gain is often modeled by a function $M(\mathbf{x})$, where $\mathbf{x} = (x, y)$ is a camera pixel. Thus the image pixel irradiance at time t is

$$\tilde{I}_t(\mathbf{x}) = M(\mathbf{x})I_t(O), \quad (1)$$

where $I_t(O)$ is the pixel irradiance when $M = 1$, for observed object O . Let the camera be panned, so that the object O projects to pixel \mathbf{x}' in time t' . *Correspondence* is established between the two pixels, by image alignment. This yields a redundant measurement $\tilde{I}'_t(\mathbf{x}') = M(\mathbf{x}')I'_t(O)$. Assuming brightness constancy, $I_t(O) = I'_t(O)$, the two measurements yield a linear constraint

$$\log M(\mathbf{x}) - \log M(\mathbf{x}') = \log \tilde{I}_t(\mathbf{x}) - \log \tilde{I}'_t(\mathbf{x}'). \quad (2)$$

Aggregating such constraints over different pixels and frames recovers $M(\mathbf{x})$, up to a global exponential ambiguity. This recovery makes pixel readout spatially consistent. Similar formulations exist for an additive offset. In Sec. 4.2 we expand this principle to a camera-array.

2.2 Avoiding Blooming and Lens-Flare in a Single Camera

In a wide-angle sky-viewing camera, sun-rays are liable to frequently shine directly into the lens. This can create blooming. Moreover, sun-rays create an additive, spatially varying lens-flare. Reducing lens flare was suggested [17–20], using either a specialized

detector array for nearby objects, or by camera rotation during imaging of a static scene. Both approaches complicate the need for simple, low-cost units and operation.

Consequently, sky-observing wide-field cameras often have a *dynamic sun blocker*. It is essentially an opaque object raised above the camera optics, blocking the Sun from view. There are various configurations, but all of them *move*, as the Sun direction changes during the day and across the year. Motorized solutions [21] that need to work year-around significantly complicate such camera units, making them very expensive. In Sec. 7 explains that a large camera network inherently bypasses the problem, without a need to constantly move a Sun blocker.

2.3 Current 3D Cloud Mapping

Existing sky-imaging systems used for research and operational meteorology¹ are few, relying on high quality cameras and other parts [24–29]. Due to their complexity and costs, they were only used to estimate *cloud-base* over narrow regions right above a narrow-baseline camera pair. Satellite-based estimation of 3D *cloud-tops* has been proposed by MISR [30]. It takes several minutes for MISR to capture multiple viewpoints of a region, during which the clouds generally move. Weather radars sense *raindrops*, which are much larger than cloud-drops and ice crystals. Hence radar does not sense clouds that do not produce rain.

3 Regional Stationarity in a Camera Network

Consider a network of sky-observing cameras, spread over a region. The location vector of camera c is \mathbf{l}_c . Any image pixel \mathbf{x} of camera c is backprojected to a ray at direction-angle vector (zenith, azimuth) $\boldsymbol{\theta}$ in a global coordinate system. The sampled light-field is the radiance measured per location, direction and time, thus given by $I_t[\mathbf{l}_c, \boldsymbol{\theta}(\mathbf{x})]$.

An interesting assumption that can be made is *regional stationarity*. In a region containing the cameras, the chance of a cloud, clear sky, or haziness affecting $I_t[\mathbf{l}_c, \boldsymbol{\theta}]$ is *independent* of c . Thus, inter-camera image variations due to atmospheric conditions are *random and unbiased*. This is illustrated in Fig. 1.

Some monocular algorithms tend to rely on gathering statistics over time, thus assuming temporal stationarity. Nevertheless, *simultaneous* images captured by *different camera nodes* are generally different from each other. Due to regional stationarity, a change of viewpoint has an effect similar to change in time: a cloud in $I_t[\mathbf{l}_c, \boldsymbol{\theta}]$ is often not in $I_t[\mathbf{l}_{c'}, \boldsymbol{\theta}]$. Consequently, monocular algorithms can be extended to statistics gathered over both time and viewpoint, as we show.

4 Self-Calibration in a Camera Network

Before a camera is deployed, its internal geometric and radiometric characteristics (including distortions, radiometric response) are calibrated in the lab. This is done using

¹ There are also ground viewing webcams that happen to see sky parts [22, 23] and weather cameras that are too sparse to be integrated for recovery.

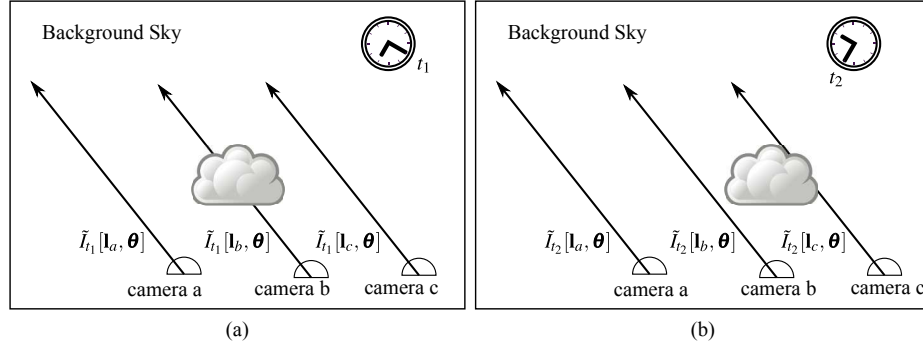


Fig. 1. Regional stationarity means that in a region, objects at infinity and the background sky should have the same radiance, for a common viewing angle θ , e.g., $[l_a, \theta]$ vs. $[l_c, \theta]$ at t_1 . Nearby objects (clouds) result in pixel differences, e.g., $[l_b, \theta]$ vs. $[l_c, \theta]$ at t_1 . Nevertheless, the light-field statistics (spatio-temporal variations and correlations) are stationary across viewpoints. This enables statistical processing across viewpoints and time. Residual measured bias is attributed to slight inter-camera radiometric inconsistencies.

established monocular methods. However, once the camera is placed in the field, we need to estimate unknown parameters. External light sources in the vicinity of a camera may create a weak lens glare, that *offsets* radiometric readings, in a way that varies both spatially and across viewpoints. Moreover, residual *gain* variations may be between different cameras, despite lab calibration. This may be exacerbated in the field due to dirt accumulated on the lenses. A large light-field network would need to use many cameras. Estimating and compensating for inter-camera differences is helpful for subsequent computer vision algorithms. Consistency is necessary for radiometry-based tasks, such as tomography [14]. In this section we describe a solution. Similarly to Sec. 2.1, the solution relies on redundant measurements at *corresponding* points. For correspondence, geometric calibration [31] is a first necessity, which is treated in Sec. 4.1.

4.1 Geometric calibration

The internal parameters Ψ_c of camera c are pre-calibrated in the lab. In the field, the location vector l_c is known by GPS. However, the orientation (yaw, pitch and roll angle vector Θ_c) is loosely set in the field. The orientation is calibrated by automatically detecting and tracking extra-terrestrial (XT) objects (Moon, planets, Sun) [31, 32], across night or day² in the hemispherical field of view. Using astronomical charts per time t and l_c , an XT object is known to be at angle vector (zenith, azimuth) $\theta^{XT}(t)$ relative to a global coordinate system. Given camera orientation Θ_c , a ray from $\theta^{XT}(t)$ should project to pixel:

$$\mathbf{x}_{\text{model}}^{XT}(t) = \Pi(\theta^{XT}(t); \Theta_c, \Psi_c), \quad (3)$$

where Π denotes projection (converting ray angles to image pixels). The actual measured pixel $\mathbf{x}_{\text{measured}}^{XT}(t)$ has uncertainty, due to the finite angular size of the XT object

² In [31], calibration relied on manual tracking of a special flight and long exposures at night.

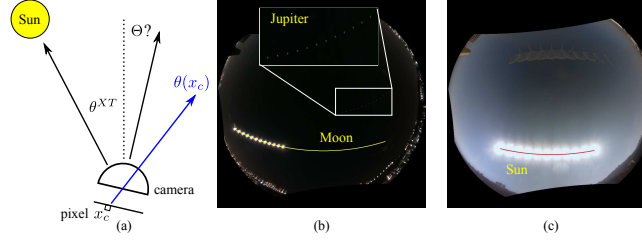


Fig. 2. (a) To estimate the camera yaw-pitch-roll angle vector Θ_c , we rely on image locations of extra-terrestrial objects, whose direction vector $\theta^{XT}(t)$ is known for all time t . (b) Photo-montage of night sky images. It shows the Moon at different times, the expected trajectory based on the estimated Θ_c , and a close-up on the corresponding sampled images of Jupiter. (c) Photo-montage of the daylight sky. It shows the Sun at different hours, the expected trajectory based on the estimated Θ_c and lens-flares.

and blooming. However, during the course of a day or night, the number of frames N^{frames} is $\mathcal{O}(100)$, leading to a simple optimization formulation:

$$\hat{\Theta}_c = \arg \min_{\Theta_c} \sum_{t=1}^{N^{\text{frames}}} \|\Pi(\theta^{XT}(t); \Theta_c, \Psi_c) - \mathbf{x}_{\text{measured}}^{XT}(t)\|^2. \quad (4)$$

We solved it using exhaustive search or gradient descent from null initialization, with the same results. The orientation calibration is illustrated in Fig. 2. Based on $\hat{\Theta}_c$, all captured images $\tilde{I}_{c,t}(\mathbf{x})$ taken by camera c are aligned to the global coordinate system: the backprojected ray has direction vector

$$\theta(\mathbf{x}) = \Pi^{-1}(\mathbf{x}; \hat{\Theta}_c, \Psi_c). \quad (5)$$

4.2 Inter-camera Relative Radiometric Self-Calibration

Consider a *fixed* view direction θ observed by several cameras. The light-field actual measured values $\{\tilde{I}_t[\mathbf{l}_c, \theta]\}_c$ correspond to readouts of parallel rays, back-projected from all cameras in the network. The values in this set generally differ from each other, e.g., $\tilde{I}_t[\mathbf{l}_c, \theta] \neq \tilde{I}_t[\mathbf{l}_{c'}, \theta]$. There are two causes for this difference:

1. Different camera locations mean different observed objects. Momentarily, camera c may observe a cloud while c' observes a clear sky, or vice versa. Camera c' may momentarily observe a somewhat denser haze volume than c , or vice versa.
2. Slight inter-camera radiometric inconsistencies, which we need to estimate.

Cause 1 is usually dominant. We need to overcome it, in order to analyze cause 2. Here we rely on the regional stationarity described in Sec. 3. Per camera c and view angle θ , bias is due to cause (2). We easily detect and characterize the bias by capturing *statistics over time*.

We performed experiments, with a small field-deployed network, further detailed in Sec. 5. Consider, for example, Fig. 3a. It shows radiometric inconsistency between cameras a and b . The figure then shows a scatter-plot of $\tilde{I}_t[\mathbf{l}_a, \theta]$ vs. $\tilde{I}_t[\mathbf{l}_b, \theta]$, $\forall t, \theta$,

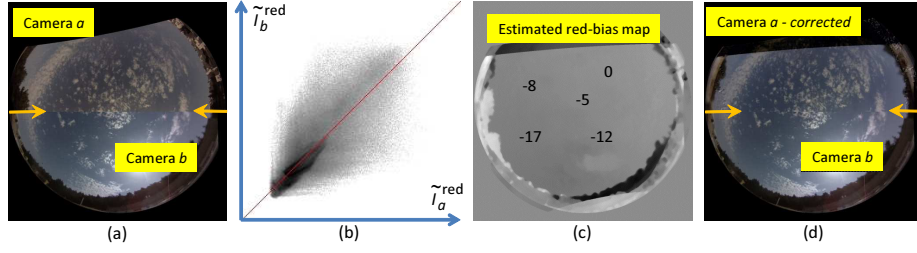


Fig. 3. [a] Splitting the field of view to upper/lower halves, to pixels corresponding respectively to either $\tilde{I}_t[\mathbf{l}_a, \boldsymbol{\theta}]$ or $\tilde{I}_t[\mathbf{l}_b, \boldsymbol{\theta}]$. In the line between the marked arrows, radiometric inconsistency shows-up as a seam across which colors slightly change (**please view on a color computer screen**). [b] Scatter-plot of $\tilde{I}_t[\mathbf{l}_a, \boldsymbol{\theta}]$ vs. $\tilde{I}_t[\mathbf{l}_b, \boldsymbol{\theta}]$, $\forall t, \boldsymbol{\theta}$, red-channel. [c] The estimated offset map $\hat{o}_{b-a}(\boldsymbol{\theta})$, red channel. It is derived based on a set of images taken during several hours. [d] Splitting the field of view in half, to *corrected* pixels from either $\tilde{I}_t[\mathbf{l}_a, \boldsymbol{\theta}]$ or $\tilde{I}_t[\mathbf{l}_b, \boldsymbol{\theta}]$: inconsistencies in the line between the marked arrows are greatly diminished.

for the red-channel. From this scatter plot and those of the other color channels, we hypothesized that camera *a* has a slight offset, particularly in the red channel, relative to camera *b*. We thus estimated, per color channel, the map of radiometric offset (across pixels, or ray-directions). A temporal median was used.

$$\hat{o}_{b-a}(\boldsymbol{\theta}) = \text{median}_t \{ \tilde{I}_t[\mathbf{l}_b, \boldsymbol{\theta}] - \tilde{I}_t[\mathbf{l}_a, \boldsymbol{\theta}] \}. \quad (6)$$

The map $\hat{o}_{b-a}(\boldsymbol{\theta})$ was then spatially smoothed and used to correct $\tilde{I}_t[\mathbf{l}_a, \boldsymbol{\theta}]$. As shown in Fig. 3d, the results have much better inter-camera consistency. This is apparent in images taken throughout that day. A similar process was applied to other cameras, but they had negligible radiometric offsets with respect to camera *b*. Later, we realized that the spatially varying offset in camera *a* was due to a nearby red light source.

The process was then repeated to detect slight local variations of gain (vignetting). Suppose there is no offset. In analogy to Eq. (6), the gain in *b* is higher than in *a* by a factor

$$\hat{M}_{b/a}(\boldsymbol{\theta}) = \text{median}_t \{ \tilde{I}_t[\mathbf{l}_b, \boldsymbol{\theta}] / \tilde{I}_t[\mathbf{l}_a, \boldsymbol{\theta}] \} \quad (7)$$

Under this scheme, all the cameras in the network are radiometrically aligned to a single master camera. After radiometric corrections, the light-field samples are denoted $\tilde{I}_t[\mathbf{l}_b, \boldsymbol{\theta}]$.

5 More Details About The Experimental Setup

Before proceeding with mathematical problems and solutions, we give more details about a small experimental network, which we built to test the hypotheses. To show that it can work with very low-cost camera nodes (units), each of the 5 units was built from basic components, and coarse alignment tolerances. Its core is a Raspberry-Pi computer used in DIY electronics, and a 5MP Raspberry-Pi camera. In this camera, the gain, response and white-balance can be fixed, avoiding temporal radiometric variations. We cut-off its OEM lens, and manually mounted a small fisheye lens. Due to this coarse

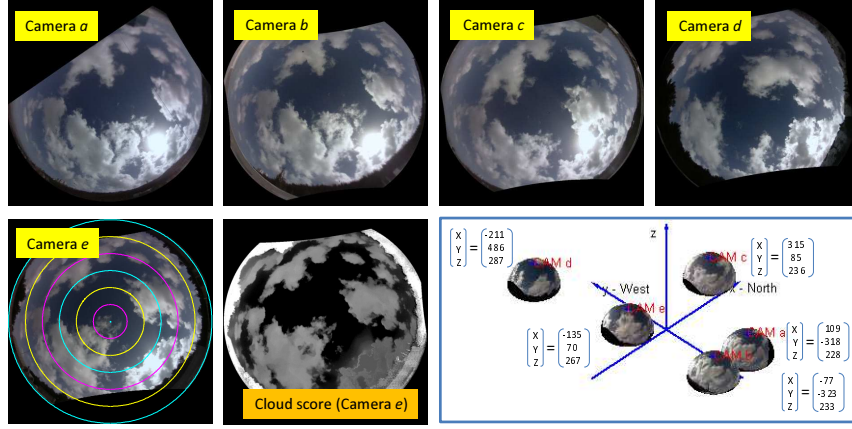


Fig. 4. Images taken simultaneously by a 5-node network. They are geometrically aligned to the zenith and north, and resampled to *polar azimuthal equidistant* projection in this global system. Equidistant zenith angle circles are overlaid on $\tilde{I}_t[\mathbf{l}_e, \boldsymbol{\theta}]$ (camera e). Each camera has dead-regions, since the lens-axis had only been roughly centered on the tiny DIY boards. Corresponding to the frame in camera e , a *cloud score* map (Eq. 11) has high values in cloud-pixels, diminishing outside them. [Bottom-right] The 3D setup of the experimental network. It is laterally spread over several hundreds of meters, on rooftops at somewhat different heights.

lens-to-chip alignment, *each camera has a different peripheral dead-region*, creating a missing part in the hemispheric view-field and vignetting distinct to each camera (seen in Fig. 3). As derived below, a network *as-a-whole* can inherently overcome these issues. Every 30 seconds, synchronously, all units automatically transmit image data using the cellular network to the internet (cloud-service). Each unit is solar powered. The units were placed on several rooftops (Fig. 4) and ran for weeks, capturing 70GB of sky imagery.

6 Network-assisted Background Estimation

It is often helpful to characterize the background, as motivated by change-detection computer vision algorithms. In monocular settings, such algorithms use temporal filtering to characterize the background³: foreground dynamic objects are at *different locations at different times* and are thus pruned. In our case this translates to stating that a cloud in $\tilde{I}_t[\mathbf{l}_c, \boldsymbol{\theta}]$ is often not in $\tilde{I}_{t'}[\mathbf{l}_c, \boldsymbol{\theta}]$, when $t' \neq t$. However, if clouds move slowly, while illumination gradually changes, temporal filtering may be insufficient. This is illustrated in Fig. 5.

A light-field network enhances this principle, with more effective pruning-per-time. Recall regional stationarity (Sec. 3). A change of viewpoint has an effect similar to

³ Physics-based sky appearance models [33] require several atmospheric parameters that are unknown in our case, e.g. the aerosol distribution.

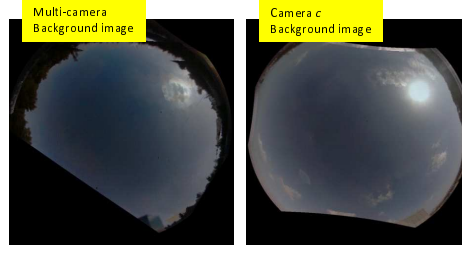


Fig. 5. [Left] Estimation of the sky background, using Eq. (8) based on five temporal instances and five viewpoints. [Right] Estimation of the sky background, using five temporal instances, but just a single viewpoint, resulting in more residual clouds.

change in time: a cloud in $\tilde{I}_t[l_c, \theta]$ is often not in $\tilde{I}_{t'}[l_c, \theta]$. Consequently, background sky values are obtained by data filtering over *both* time and viewpoint.

This network-based principle can generally enhance any state-of-the-art (or future) algorithm of background estimation, which would otherwise be monocular. For a quick demonstration, we show results using a simplistic, basic criterion. In broad daylight, clouds are brighter than the blue sky. Hence, as an example, an estimator for the sky background can be:

$$\text{SKY}(\theta) = \arg \min_{t,c} \tilde{I}_t[l_c, \theta] \quad (8)$$

where $t \in [1 \dots N^{\text{frames}}]$ and $c \in [1 \dots N^{\text{views}}]$. This is illustrated in Fig. 5.

7 Bypassing the Sun through a Camera Network

As Sec. 2.2 explains, in existing full sky-imagers, effects of direct sunlight are often mitigated by a small dynamic sun-blocker, which complicates the system and its cost, while having a blind-region. The network offers a different solution, which can be radical, yet simple. On each camera, the sun-blocker is *static*, and has no moving part. The blocker can be large, covering the entire range of directions the Sun may occupy during the year or part of it. In this configuration, each camera unit has a large blind area (See Fig. 6). Nevertheless, the entire *network has no blind spot*, when viewing the atmosphere. This remarkable property is a result of network-redundancy, as we explain.

A static year-round sun blocker on camera c permanently obstructs a set Γ_c of atmospheric voxels. These voxels, however, are generally visible at several other cameras, e.g., those indexed e, f, g in Fig. 6. Consequently, a sufficiently wide network has no 3D blind spot, despite permanent sun-blocking. Voxels that are not obstructed from any camera have more viewpoints to constrain their content, than voxels in Γ_c .

We now quantify the implication of this approach to the network extent, referring to the northern hemisphere without loss of generality. Nearly all weather phenomena are under the tropopause, whose altitude H above sea level is typically 17km at the equator, and decreasing with latitude. The solar seasonal angle amplitude is $\beta \approx 23.5^\circ$. At latitude γ , thus, a permanent sun blocker spans zenith angles in the range $\gamma \pm \beta$.

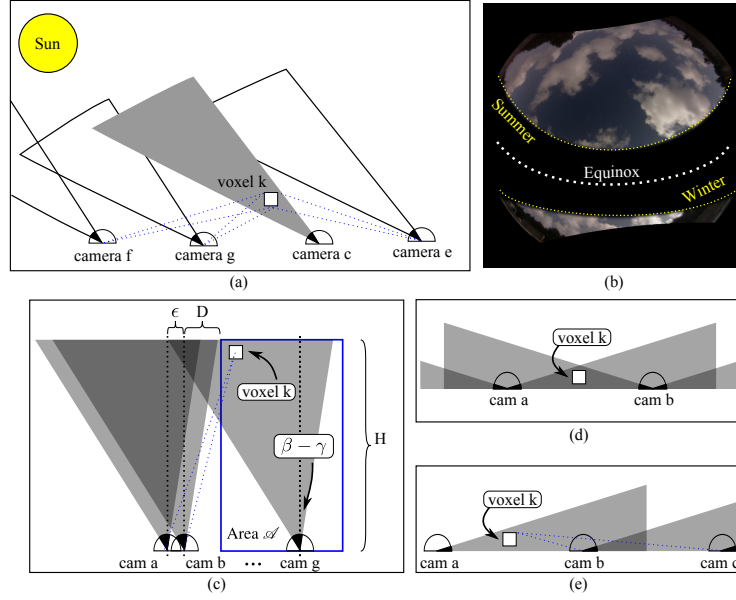


Fig. 6. [a] Camera c has a blind-region, covering Sun directions at l_c . The blind region corresponds to set Γ_c of atmospheric voxels not sensed by camera c . The network as a whole still has coverage of voxels $k \in \Gamma_c$, as they are observed by cameras e, f, g . [b] Simulation of a whole sky image (polar azimuthal equidistant projection), blocking all solar directions during the year, at a mid-latitude. [c] In the tropics, the network must have nodes at distance D outside surveyed area \mathcal{A} , if \mathcal{A} is narrow. The distance D depends on the latitude γ , while $\beta \approx 23.5^\circ$. [d] In the arctic, the blind region is adjacent to the horizon, in all azimuth angles. Fixed blocking of the Sun over 360° blocks low-altitude voxel k . [e] Arctic cameras fitted with a fixed north-facing sun blocker create a network that operates 12 hours a day. An adjacent camera at each node has a fixed south-facing sun blocker, for imaging during complementing hours.

Earth is divided here to three region classes:

- In the tropics, the sky directly above a camera is blocked. Consider a small tropical area \mathcal{A} , e.g., 1km wide, whose sky needs to be monitored. According to Fig. 6c, \mathcal{A} can efficiently be observed without a blind spot by cameras to its *south*. The network needs to have camera units extending to distance $D = H \tan(\beta - \gamma) + \epsilon$ from \mathcal{A} , where ϵ is a small distance, sufficient for triangulation at H . At the equator $D \approx 7.4\text{km}$. It can also be shown that if \mathcal{A} is wider than $2H[\tan(\beta - \gamma) + \tan(\beta + \gamma)]$, the network can observe and triangulate all the sky right above it.
- As latitude increases towards the tropic circles, D decreases to zero. Thus the network can observe and triangulate all the sky right above it, anywhere outside the tropics, in the mid-latitudes.
- In the arctic and antarctic summer, the Sun can appear in all azimuth angles over the day. A single 24-hour fixed sun-blocker blocks the horizon. So as shown in Fig. 6d, voxel k is not observed. One solution would be to mount two cameras, side by side, in each network node. Each camera in a node has a fixed sun blocker covering half of

the azimuth angles. One camera operates in the *polar daytime* (local 6AM to 6PM), as it has a south-oriented fixed blocker. The other camera operates in the complementing time (Fig. 6e), having a north-oriented fixed blocker. This way, the network never has a blind spot.

8 3D Clouds by a Camera Network

Network-based acquisition and the processes described above, yield a calibrated sampled light-field of the sky. Using this light-field, one application is to estimate the 3D cloud structure above the network domain, and beyond. This is done by the following steps: (A) Per time t , give a cloud score s to each ray $[l_c, \theta]$, as we explain below. (B) Perform a fuzzy version of space carving [34, 35] (reminiscent of back-projection in tomography).

We first describe a simple yet effective method to implement (B). The set of all sampled light-field rays is \mathcal{R} , where $|\mathcal{R}| = N^{\text{rays}}$. A ray is indexed by r , and it corresponds to a specific $[l_c, \theta]$. Voxel k projects to a subset of the rays $\rho_k \subset \mathcal{R}$, that reach ν_k viewpoints. Suppose a ray $r \in \mathcal{R}$ has a cloud-score $s(r) \in [0, 1]$, where $s = 0$ means there is definitely no cloud on the ray, while $s = 1$ means there is confidently a cloud there. Per voxel k , define a back-projected score

$$B_k = \left[\prod_{r \in \rho_k} s(r) \right]^{1/\rho_k} \quad \text{if } \nu_k \geq 2. \quad (9)$$

The back-projected score of voxel k is null, if k is not observed by at least two viewpoints. This score is null also if $s(r) = 0$ for any $r \in \rho_k$. If all $r \in \rho_k$ have same score s , then $B_k = s$. Equation (9) carves-out voxels that contradict support for clouds.

Different cloud regions have signature appearances. Ignoring this would wrongly allow matching of, say, a darker cloud-bottom to a bright sun-lit side of a cloud, or a smooth cloud region to a textured one. Thus, photometric and appearance consistency across viewpoints is incorporated (the photo-hull concept in space-carving [34]). Clouds are diffuse, cooperating with this requirement. From the images, a feature vector $\mathbf{v}(r, t)$ is extracted for any measured ray r . We used SIFT descriptors [36] and the radiance in each color channel. Element q of $\mathbf{v}(r, t)$ is $v_q(r, t)$. The values of this element, for all rays that intersect voxel k , is $\mathcal{V}_q(k, t) \equiv \{v_q(r, t)\}_{r \in \rho_k}$. Across viewpoints, the measured variance in this set is $\text{VAR}[\mathcal{V}_q(k, t)]$. Define an appearance consistency score [37] as

$$P_k = \exp \left(-\Sigma_q \{[\mathcal{V}_q(k, t)]\} / \sigma^2 \right), \quad (10)$$

where σ^2 is a scale parameter. Overall, the total cloud-score of a voxel is $T_k = B_k P_k$.

The resulting 3D field $\{T_k\}$ is a volumetric estimate of cloud occupancy. It is biased to yield clouds larger than they really are: high-altitude voxels occluded by the cloud-base from all viewpoints are interpreted as being cloudy, since for them T_k is high. This is a realization of a basic ambiguity: if a voxel is occluded from all viewpoints, then there is no way of telling if it is cloudy or not, unless auxiliary or prior knowledge is available. Incorporating a visibility prior favors smaller clouds that explain the data. If

voxel k is completely occluded by other cloudy voxels, then it is pruned (carved) out. Voxel k can only maintain its T_k if there are at least two camera viewpoints from which k is not occluded by other clouded voxels. This pruning is achieved by sweeping [34] the field $\{T_k\}$ iteratively. The pruned 3D cloud occupancy field is denoted $\{\tilde{T}_k\}$. We can maintain the non-binary (fuzzy) nature of $\{\tilde{T}_k\}$. This way, it possesses the inherent semi-transparency and subtle ambiguity of clouds.

Basic cloud score

For a basic cloud score (step A), there are various functions in the literature. We considered several of them, including using the background derived in Sec. 6. There is certainly room for increased sophistication, including sky modeling and machine learning, but we choose to stick with a simple measure. The ratio of image readout at the red/blue color channels, $\tilde{I}^{\text{red}}/\tilde{I}^{\text{blue}}$, is widely used in the literature [38, 31]. Overall, we found it to be an effective feature in broad daylight: clouds are grey (unit red-blue ratio), and the cloudless sky is significantly biased to blue (below ≈ 0.8 red-blue ratio). Thus, for demonstrations in this paper, the cloud-score we used per ray (pixel) is

$$s(r) = \begin{cases} \frac{6[\tilde{I}^{\text{red}}(r)/\tilde{I}^{\text{blue}}(r)-0.8]}{0.2+\tilde{I}^{\text{red}}(r)/\tilde{I}^{\text{blue}}(r)} & \text{if } \tilde{I}^{\text{red}}(r)/\tilde{I}^{\text{blue}}(r) > 0.8 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

Here $s \in [0, 1]$, where either bound is achieved at gray clouds or blue sky, respectively. An example of applying this operator on an image is shown in Fig. 4.

Simulations

Quantitative assessments used atmospheric-science gold standard simulators. An atmosphere over $8 \times 8\text{km}$ was produced using off-the-shelf large eddy simulation (LES), creating clouds between heights of 500m to 1500m. Lighting conditions were consistent with Copenhagen (unrelated to the anonymous authors). Radiative transfer using the discrete ordinate spherical-harmonic method (SHDOM) [39] rendered images taken by 100 cameras placed in a $2.5 \times 2.5\text{km}^2$ domain. Recovery simulations used random subsets of the network, where the whole network is either with or without a sun blocker. In the LES, a voxel is occupied by cloud if its water-mass parameter is not null. In the recovery, voxel k is classified as cloud if $T_k > 0.01$. We measured the classification error rate, across all voxels. The results are plotted in Fig. 7. As expected of space carving, results improve fast from 2 to 10 cameras. Even when a sun blocker is applied, the algorithm is able to reconstruct the cloud formation. Then, as expected, more cameras are needed in order to compensate for the limited view of each camera. In our setup the cameras observe the clouds only from below. This is the classical limited angle problem known from visual hull and tomography algorithms. Therefore the reconstruction error does not converge to zero.

Cloud Reconstruction Experiment

Our experiments showed that the blooming and saturation by the Sun did not affect cloud reconstruction. Therefore we did not use the sun blocker. We applied the approach

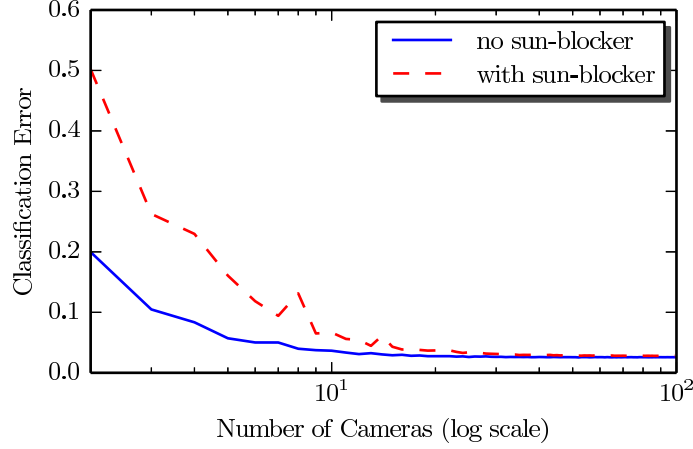


Fig. 7. Classification error rate as a function of the number of cameras. [Solid Blue] Without sun blocker. [Dashed Red] With sun blocker. Fluctuations are within the random sampling standard deviation.

described in Sec. 8 on various captured scenes. One scene had cumulus clouds, similar to those seen in Fig. 4. Cross-sections of the recovered 3D cloud-occupancy field $\{\tilde{T}_k\}$ are shown in Fig. 8. The lateral domain of the clouds is much larger than the camera network. Accounting for the altitude of the cameras above sea-level, the clouds mainly reside between 800m to 1450m above Sea-level. We used two indicators to validate the results. First, there are worldwide balloon-based radiosonde measurements including quantitative vertical humidity profiles. The radiosonde nearest to our network is 88km away, on a similar coastal strip, and used by forecasters for our location. It indicated a layer of high humidity that can yield clouds in the range [770, 1881]m above sea-level, consistent with our clouds.

Second, we cross-validated 3D recovery *with missing field of view*. We used four cameras (indexed a, b, c, d) out of five, for 3D estimation. Camera e was ignored. Then, we projected the estimated 3D cloud distribution into viewpoint e , and compared to the ground truth. The rendered image is as follows. *Ray casting* [40] of field $\{\tilde{T}_k\}$ is performed on a ray corresponding to $[\mathbf{l}_e, \boldsymbol{\theta}]$. Ray-casting aggregates $\{\tilde{T}_k\}$ on all voxels intersected by the ray. The result is a cloud-score image $w[\mathbf{l}_e, \boldsymbol{\theta}]$. To visualize $w[\mathbf{l}_e, \boldsymbol{\theta}]$, we used it as α -map to the estimated sky-background image (Eq. 8). The alpha-map is

$$\alpha[\mathbf{l}_e, \boldsymbol{\theta}] = \begin{cases} 2w[\mathbf{l}_e, \boldsymbol{\theta}] & \text{if } 2w[\mathbf{l}_e, \boldsymbol{\theta}] < 1 \\ 1 & \text{otherwise} \end{cases} . \quad (12)$$

The rendered image is then $J[\mathbf{l}_e, \boldsymbol{\theta}] = \alpha[\mathbf{l}_e, \boldsymbol{\theta}] + (1 - \alpha[\mathbf{l}_e, \boldsymbol{\theta}])\text{SKY}(\boldsymbol{\theta})$. This image does not pretend to properly render clouds in their true shades and effect on the sky. It simply served to visualize the result, as shown in Fig. 8d. It can be compared to the true corresponding image $\hat{T}_t[\mathbf{l}_e, \boldsymbol{\theta}]$, in Fig. 8e. This rendering relies on the redundancy of the

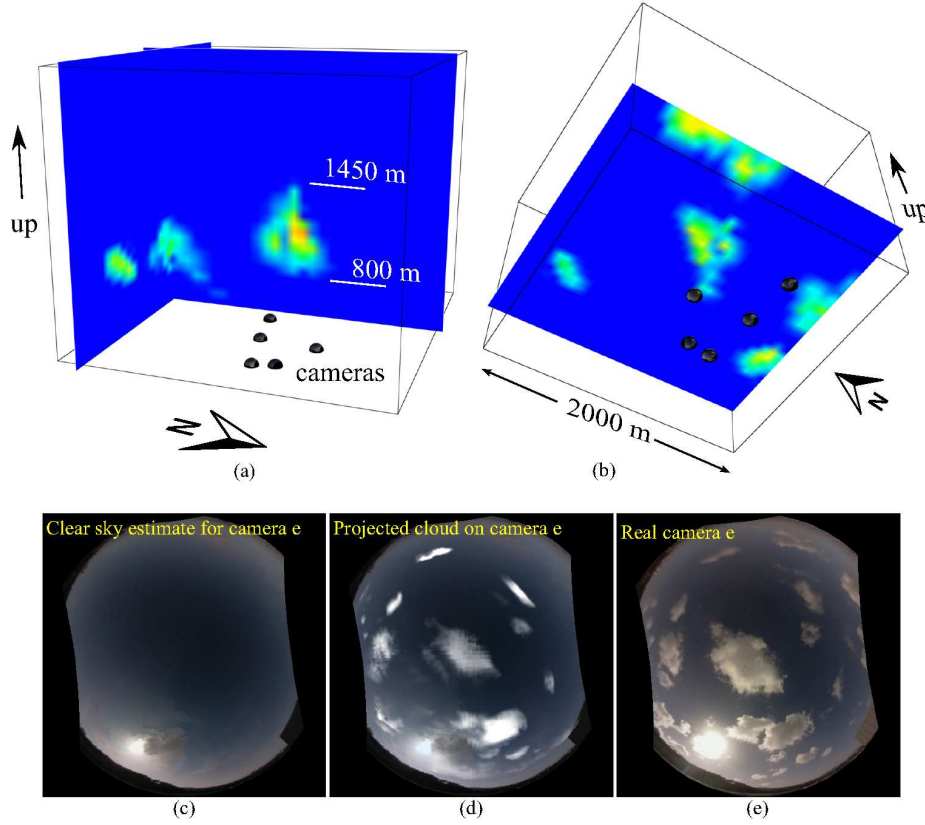


Fig. 8. 3D cumulus cloud recovery results. (a,b) Cross-sections of the recovered cloud-occupancy field $\{\tilde{T}_k\}$. Note how the domain of the clouds is much larger than the camera network. Cloud altitude is above sea-level. (c) Estimated sky-background image. Based on four viewpoints (indexed a, b, c, d), the 3D volumetric cloud-occupancy field $\{\tilde{T}_k\}$ was derived. The field $\{\tilde{T}_k\}$ was projected to viewpoint e , and overlayed on the estimated sky-background image. The resulting synthetic cloud-score image $J[l_e, \theta]$ is shown in (d). This can be compared to the real captured image $\hat{I}_t[l_e, \theta]$, shown in (e).

network. Even if a viewpoint is blocked, much of its information can be derived using other viewpoints compounded with 3D recovery. This is consistent with the proposed approach for handling static partial blocking of viewfields.

Another scene had a layer of alto-cumulus clouds. Figure 9 shows sample frames from this scene, and cross-sections of the recovered 3D cloud-occupancy field $\{\tilde{T}_k\}$. Accounting for the altitude of the cameras above sea-level, these estimated clouds mainly reside on a horizontal layer 3450m above sea-level, and the layer is about a kilometer thick. This time, the radiosonde readout indicated a layer of high humidity that can yield clouds in height range [3072, 4180]m above sea-level. This is in strong agreement with our results. More results are in [37].

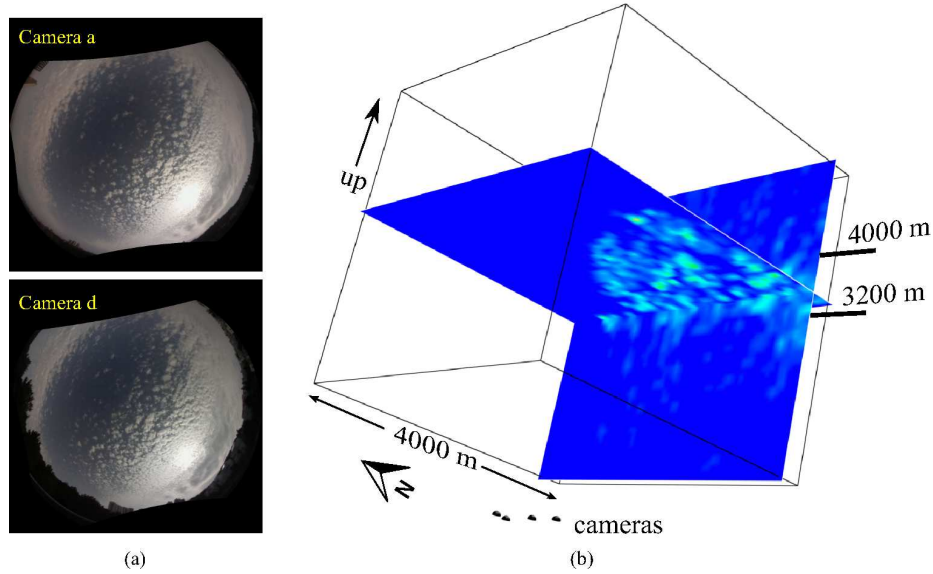


Fig. 9. 3D altocumulus cloud recovery results. (a,b) Sample frames. (c,d) Cross-sections of the recovered cloud-occupancy field $\{\tilde{T}_k\}$. Cloud altitude is above sea-level.

9 Discussion

The paper used common simple monocular features for cloud or sky detection in images. They demonstrated a general message: a network amplifies performance. More advanced classification features should improve results. This sensing approach poses new questions for computer vision and computational photography. These include network extensions to monocular tasks including handling radiometric inconsistencies, background estimation and exploitation of network redundancy.

Scaling light-field imaging hugely to sense the sky would use a large network of camera nodes, each having a hemispherical field of view, deployed over a wide area. Such a network can reach anywhere wireless communication exists, and in principle can be a continent-wide light-field system. This sensing approach offers significant advantages over existing technologies (experimental and operational) of atmospheric sensing, particularly 3D imaging, and doing so in high spatio-temporal resolution.

To enable a massive network, each node should have very low-cost. To demonstrate this can work, we built the units in our experimental small-network using very basic components and coarse alignment tolerances. In a long term, mass producing consistent units would ensure proper alignment performance (non dead-regions) and operational robustness over long periods. Still, this concept opens the door to more interesting research. In the direction of the sun blocker, a diffraction grating can disperse direct sunlight (currently blocked) or skylight, to measure its spectrum. This principle is already used by Aeronet [41] to characterize aerosols. Night-time operation is another interesting challenge: if a known planet and star appears missing on ray r , this indicated a cloud on r . Furthermore, such a light-field system can be used for studying bird, bats [42] and locust in 3D time-lapses.

References

1. Adelson, E., Wang, J.: Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14** (1992) 99–106
2. Basha, T., Avidan, S., Hornung, a., Matusik, W.: Structure and motion from scene registration. *2012 IEEE Conference on Computer Vision and Pattern Recognition (2012)* 1426–1433
3. Bishop, T.E., Zanetti, S., Favaro, P.: Light field superresolution. *IEEE ICCP* **129** (2009) 1–9
4. Horstmeyer, R., Euliss, G., Athale, R., Levoy, M.: Flexible Multimodal Camera Using a Light Field Architecture. In: *ICCP*. (2009) 1–8
5. Levoy, M., Ng, R., Adams, A., Footer, M., Horowitz, M.: Light field microscopy. *ACM Transactions on Graphics (TOG)* **25** (2006) 924–934
6. Kim, J., Lanman, D., Mukaigawa, Y., Raskar, R.: Descattering transmission via angular filtering. In: *Proc. ECCV' 10, Berlin, Heidelberg, Springer-Verlag* (2010) 86–99
7. Diner, D.J., Martonchik, J.V.: Atmospheric transmittance from spacecraft using multiple view angle imagery. *Appl. Opt.* **24** (1985) 3503–3511
8. Diner, D.J., Beckert, J.C., Reilly, T.H., Bruegge, C.J., Conel, J.E., Kahn, R.A., Martonchik, J.V., Ackerman, T.P., Davies, R., Gerstl, S.A.: Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview. *IEEE Transactions on Geoscience and Remote Sensing* **36** (1998) 1072–1087
9. Diner, D.J., Davis, A., Hancock, B., Gutt, G., Chipman, R.A., Cairns, B.: Dual-photoelastic-modulator-based polarimetric imaging concept for aerosol remote sensing. *Appl. Opt.* **46** (2007) 8428–8445
10. Diner, D.J., Davis, A., Hancock, B., Geier, S., Rheingans, B., Jovanovic, V., Bull, M., Rider, D.M., Chipman, R.A., Mahler, A.B., McClain, S.C.: First results from a dual photoelastic-modulator-based polarimetric camera. *Applied Optics* **49** (2010) 2929
11. Baxter, B., Hooper, B.A., Williams, J.Z., Dugan, J.P.: Polarimetric remote sensing of ocean waves. In: *OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges*. (2009) 1–5
12. Brdon, E.M., Bréon, F.M.: An analytical model for the cloud-free atmosphere/ocean system reflectance. *Remote Sensing of Environment* **43** (1993) 179–192
13. Mol, B.V., Ruddick, K., van Mol, B., K.Ruddick: The Compact High Resolution Imaging Spectrometer (CHRIS): the future of hyperspectral satellite sensors. *Imagery of Oostende coastal and inland waters*. In: *Proceedings of the Airborne Imaging Spectroscopy workshop, Brugge*. Number October (2004)
14. Aides, A., Schechner, Y.Y., Holodovsky, V., Garay, M.J., Davis, A.B.: Multi sky-view 3D aerosol distribution recovery. *Opt. Express* **21** (2013) 25820–25833
15. Litvinov, A., Schechner, Y.Y.: Addressing radiometric nonidealities: a unified framework. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Volume 2. (2005) 52–59 vol. 2
16. Kang, S., Weiss, R.: Can we calibrate a camera using an image of a flat, textureless Lambertian surface? *Computer VisionECCV 2000* (2000) 640–653
17. Koreban, F., Schechner, Y.Y.: Geometry by deflaring. *2009 IEEE International Conference on Computational Photography (ICCP)* (2009) 1–8
18. Raskar, R., Agrawal, A., Wilson, C.a., Veeraraghavan, A.: Glare aware photography. *ACM Transactions on Graphics* **27** (2008) 1
19. Talvala, E., Adams, A.: Veiling glare in high dynamic range imaging. *ACM Transactions on Graphics (TOG)* **26** (2007) 1–10
20. Rouf, M., Mantiuk, R., Heidrich, W., Trentacoste, M., Lau, C.: Glare encoding of high dynamic range images. *CVPR 2011* (2011) 289–296

- 675 21. Pust, N.J., Shaw, J.a.: Digital all-sky polarization imaging of partly cloudy skies. *Applied* 675
676 *optics* **47** (2008) H190–8 676
- 677 22. Bradley, E.S., Toomey, M.P., Still, C.J., Roberts, D.A.: Multi-Scale Sensor Fusion With an 677
678 Online Application: Integrating GOES, MODIS, and Webcam Imagery for Environmental 678
679 Monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote* 679
680 *Sensing* **3** (2010) 497–506 680
- 681 23. Jacobs, N., King, J., Bowers, D., Souvenir, R.: Estimating Cloud Maps from Outdoor Image 681
682 Sequences. In: *IEEE Winter Conference on Applications of Computer Vision (WACV).* 682
683 (2014) (acceptance rate=40%). 683
- 684 24. Allmen, M.C., Kegelmeyer Jr., P.: The computation of cloud base height from paired whole- 684
685 sky imaging cameras. *Machine Vision and Applications* **9** (1997) 160–165 685
- 686 25. Baumgarten, G., Fiedler, J., Fricke, K.H., Gerding, M., Hervig, M., Hoffmann, P., Müller, N., 686
687 Pautet, P.D., Rapp, M., Robert, C., Rusch, D., von Savigny, C., Singer, W.: The noctilucent 687
688 cloud (NLC) display during the ECOMA/MASS sounding rocket flights on 3 August 2007: 688
689 morphology on global to local scales. *Annales Geophysicae* **27** (2009) 953–965 689
- 690 26. Cazorla, a., Olmo, F.J., Aladosarboledas, L., Alados-Arboledas, L.: Using a Sky Imager for 690
691 aerosol characterization. *Atmospheric Environment* **42** (2008) 2739–2745 691
- 692 27. Long, C.N., Sabburg, J.M., Calbo, J., Pages, D., Calbó, J., Pagès, D.: Retrieving cloud 692
693 characteristics from ground-based daytime color all-sky images. *Journal of Atmospheric* 693
694 *and Oceanic Technology* **23** (2006) 633–652 694
- 695 28. Seiz, G., Shields, J., Feister, U., Baltsavias, E.P., Gruen, A.: Cloud mapping with ground- 695
696 based photogrammetric cameras. *Int. J. Remote Sens.* **28** (2007) 2001–2032 696
- 697 29. Kassianov, E., Long, C., Christy, J.: Cloud-base-height estimation from paired ground-based 697
698 hemispherical observations. *Journal of Applied Meteorology* **44** (2005) 1221–1233 698
- 699 30. Seiz, G., Davies, R.: Reconstruction of cloud geometry from multi-view satellite images. 699
700 *Remote Sensing of Environment* **100** (2006) 143–149 700
- 701 31. Seiz, G., Baltsavias, E., Gruen, A.A.: Cloud mapping from the ground: Use of photogram- 701
702 metric methods. *Photogrammetric engineering and remote sensing* **68** (2002) 941–951 702
- 703 32. Lalonde, J.F., Narasimhan, S.G., Efros, A.a.: What Do the Sun and the Sky Tell Us About 703
704 the Camera? *International Journal of Computer Vision* **88** (2009) 24–51 704
- 705 33. Hosek, L., Wilkie, A.: An analytic model for full spectral sky-dome radiance. *ACM Trans-* 705
706 *actions on Graphics (TOG)* **31** (2012) 95 706
- 707 34. Kutulakos, K., Seitz, S.: A theory of shape by space carving. *International Journal of Com-* 707
708 *puter Vision* **38** (2000) 199–218 708
- 709 35. Ihrke, I., Magnor, M.: Image-based tomographic reconstruction of flames. In: *ACM SIG-* 709
710 *GRAPH 2004 Sketches. SIGGRAPH '04, New York, NY, USA, ACM* (2004) 16– 710
- 711 36. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal* 711
712 *of computer vision* **60** (2004) 91–110 712
- 713 37. Authors: Supplementary material (2014) 713
- 714 38. Yamashita, M.: Cloud cover estimation using multitemporal hemisphere imageries. In: 714
715 *Proc. XXth Congress of the International Society for Photogrammetry and Remote Sensing* 715
716 *(ISPRS04).* (2004) 818–821 716
- 717 39. Evans, K.F.: The Spherical Harmonics Discrete Ordinate Method for Three-Dimensional 717
718 Atmospheric Radiative Transfer. *Journal of the Atmospheric Sciences* **55** (1998) 429–446 718
- 719 40. Levoy, M.: Efficient ray tracing of volume data. *ACM Transactions on Graphics* **9** (1990) 719
245–261
41. Holben, B.N., Eck, T.F., Slutsker, I., Tanré, D., Buis, J.P., Setzer, A., Vermote, E., Rea-
gan, J.A., Kaufman, Y.J., Nakajima, T., Lavenue, F., Jankowiak, I., Smirnov, A., Tanre, D.:
AERONET-A federated instrument network and data archive for aerosol characterization.
Remote Sensing of Environment **66** (1998) 1–16

720	42. Theriault, D.H., Fuller, N.W., Jackson, B.E., Bluhm, E., Evangelista, D., Wu, Z., Betke, M.,	720
721	Hedrick, T.L.: A protocol and calibration method for accurate multi-camera field videogra-	721
722	phy. The Journal of Experimental Biology (2014)	722
723		723
724		724
725		725
726		726
727		727
728		728
729		729
730		730
731		731
732		732
733		733
734		734
735		735
736		736
737		737
738		738
739		739
740		740
741		741
742		742
743		743
744		744
745		745
746		746
747		747
748		748
749		749
750		750
751		751
752		752
753		753
754		754
755		755
756		756
757		757
758		758
759		759
760		760
761		761
762		762
763		763
764		764