

# Clouds in The Cloud

Dmitry Veikherman, Amit Aides, Yoav Y. Schechner and Aviad Levis

Dept. of Electrical Engineering, Technion - Israel Inst. of Technology, Haifa, Israel

**Abstract.** Light-field imaging can be scaled up to a very large area, to map the Earth’s atmosphere in 3D. Multiview spaceborne instruments suffer low spatio-temporal-angular resolution, and are very expensive and unscalable. We develop sky light-field imaging, by a wide, 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. Such sensing can complement satellite imagery, be useful to meteorology, make aerosol tomography realizable, and give new, powerful tools to atmospheric and avian wildlife scientists.

## 1 Introduction

Plenoptic, light-field and integral imaging [1, 6, 8, 12, 23, 35, 29] sample the directional and spatial distribution of radiance. This imaging mode has been used in small-scale setups. However, it can be scaled up 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) [18, 21], the Airborne Multiangle SpectroPolarimetric Imager (AirMSPI) [19, 20] and POLDER [10, 14, 39]. These architectures have crude resolution spatially (up to kilometers per pixel), angularly ( $\approx 9$  angles per view) [47], or temporally (orbit takes several days to return to the same terrestrial spot). Furthermore, spaceborne instruments are extremely expensive and unscalable. We develop a complementing approach: the atmospheric light-field is captured from below, by wide-angle cameras looking upwards. This approach is a *scalable sensor network*, that captures images simultaneously over a very large area, densely.

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

wherever communication reaches. 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 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. We demonstrate this in real field experiments by building a small version Sky-Technion Array of Sensors (STARS). Such a network enables unprecedented 3D imaging of cloud fields, in high spatio-temporal resolution. This approach can complement multi-angular satellite imagery. It can make aerosol tomography [2, 5] realizable, offer new ways to study weather phenomena and avian wildlife, and aid electric power management [40].

## 2 Background

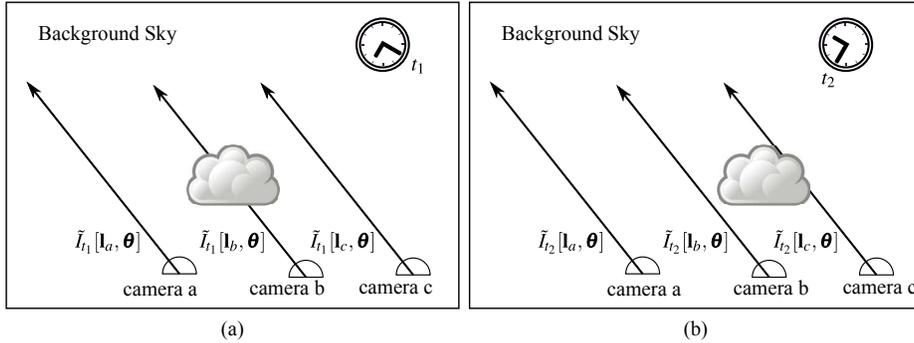
### 2.1 Monocular radiometric self-calibration

A large network should use low-cost camera units. Such cameras often have spatial and temporal radiometric inconsistencies. For example, spatial gain (e.g., by vignetting [27, 36]) is often modeled by a function  $M(\mathbf{x})$ , where  $\mathbf{x} = (x, y)$  is a camera pixel. The image irradiance at time  $t$  is  $\tilde{I}_t(\mathbf{x}) = M(\mathbf{x})I_t(O)$ , where  $I_t(O)$  is the pixel irradiance when  $M = 1$ , for observed object  $O$ . 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. Correspondence is established between modulated readouts, e.g. by aligning a pan sequence. Assuming brightness constancy, corresponding measurements yield constraints. Aggregating constraints over different pixels and frames recovers parameters of radiometric inconsistencies. This recovery makes monocular pixel readout spatially consistent. Sec. 4 expands this principle to a camera-array.

### 2.2 Avoiding Blooming, Lens-Flare in a Single Camera

In wide-angle sky-views, the sun is liable to frequently shine directly into the lens, creating blooming. Moreover, sun-rays create an additive, spatially varying lens-flare. Reducing flare was suggested [30, 42, 43, 48] using either a specialized detector array for nearby objects, or camera rotation during capture of a static scene. Both ways complicate the need for simple, low-cost units and operation.

Sky-observing wide-field cameras often have a *dynamic sun blocker*: 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 [41] that need to work year-around significantly complicate such camera units, making them very expensive. Sec. 7 explains that a large camera network inherently bypasses the problem, without a need to constantly move a Sun blocker.



**Fig. 1.** Regional stationarity: in a wide region, objects at infinity and the background sky should have the same color, for a common view angle  $\theta$ , e.g.,  $[\mathbf{l}_a, \theta]$  vs.  $[\mathbf{l}_c, \theta]$  at  $t_1$ . Nearby objects (clouds) result in pixel differences, e.g.,  $[\mathbf{l}_b, \theta]$  vs.  $[\mathbf{l}_c, \theta]$  at  $t_1$ . Nevertheless, the 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.

### 2.3 Current 3D Cloud Mapping

Existing research and operational sky-imaging systems<sup>1</sup> are few, relying on high quality components [3, 9, 16, 28, 37, 46]. 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 [45]. 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.

## 3 Regional Stationarity in a Camera Network

A network of sky-observing cameras is spread over a region. The location vector of camera  $c$  is  $\mathbf{l}_c$ . Any image pixel  $\mathbf{x}$  of camera  $c$  is back-projected to a ray at direction-angle vector (zenith, azimuth)  $\theta$  in a global coordinate system. The data is the radiance measured per location, direction and time,  $\tilde{I}_t[\mathbf{l}_c, \theta(\mathbf{x})]$ . 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  $\tilde{I}_t[\mathbf{l}_c, \theta]$  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

<sup>1</sup> There are also ground viewing webcams that happen to see sky parts [13, 26] and weather cameras that are too sparse to be integrated for recovery.

stationarity, a change of viewpoint has an effect similar to change in time: a cloud in  $\tilde{I}_t[\mathbf{l}_c, \boldsymbol{\theta}]$  is often not in  $\tilde{I}_t[\mathbf{l}_{c'}, \boldsymbol{\theta}]$ . Consequently, monocular algorithms can be extended to statistics gathered over both time and viewpoint (as done underwater in [4]). Regional stationarity is supported by meteorological research [7, 33]. Stationarity breaks down across large topographic discontinuities: a mountain ridge, coast line. These locations are known, and hence can be handled or avoided in stationarity-based analysis.

## 4 Self-Calibration in a Camera Network

Internal geometric and radiometric camera characteristics (including distortions, radiometric response) are calibrated in the lab using established monocular methods. However, once a camera is placed in the field, unknown parameters are introduced. External sources in the vicinity of a camera may create a weak lens glare, that *offsets* radiometric readings, in 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 by dirt accumulation on lenses. Similarly to Sec. 2.1, the solution relies on redundant measurements at corresponding points.

For correspondence, geometric calibration [44] is necessary. The internal parameters  $\Psi_c$  of camera  $c$  are pre-calibrated in the lab. In the field, the location vector  $\mathbf{l}_c$  is known by GPS but the orientation (yaw, pitch and roll angle vector  $\boldsymbol{\Theta}_c$ ) is loosely set. The orientation is calibrated by automatically detecting and tracking extra-terrestrial (XT) objects (Moon, planets, Sun) [32, 44], across night or day,<sup>2</sup> at pixel  $\mathbf{x}_{\text{measured}}^{\text{XT}}(t)$ . Using astronomical charts, an XT object is known<sup>3</sup> to be at angle vector (zenith, azimuth)  $\boldsymbol{\theta}^{\text{XT}}(t)$  relative to a global coordinate system. Given camera orientation  $\boldsymbol{\Theta}_c$ , a projection  $\Pi$  converts a ray direction  $\boldsymbol{\theta}^{\text{XT}}(t)$  to pixel  $\Pi(\boldsymbol{\theta}^{\text{XT}}(t); \boldsymbol{\Theta}_c, \Psi_c)$ .

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

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

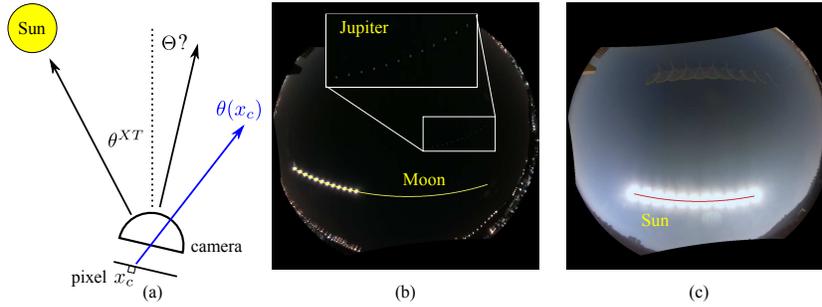
We solved it using exhaustive search or gradient descent from null initialization, with the same results. The orientation calibration is illustrated in Fig. 2.

Based on  $\hat{\boldsymbol{\Theta}}_c$ , 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

$$\boldsymbol{\theta}(\mathbf{x}) = \Pi^{-1}(\mathbf{x}; \hat{\boldsymbol{\Theta}}_c, \Psi_c). \quad (2)$$

<sup>2</sup> Manual tracking of a special flight and long exposures at night were used in [44].

<sup>3</sup> Higher than  $20^\circ$  above the horizon [11], errors caused by atmospheric refraction are smaller than  $0.05^\circ$ , much less than the angular size of each of our pixels,  $0.18^\circ$ .



**Fig. 2.** (a) To estimate the camera yaw-pitch-roll angle vector  $\Theta_c$ , we rely on image locations of extra-terrestrial objects, whose direction vector  $\theta^{XT}(t)$  is known  $\forall t$ . (b) Photo-montage of night sky images. It shows the Moon at different times, the expected trajectory based on the estimated  $\Theta_c$ , 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  $\Theta_c$  and lens-flares.

### Inter-camera Relative Radiometric Self-Calibration

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

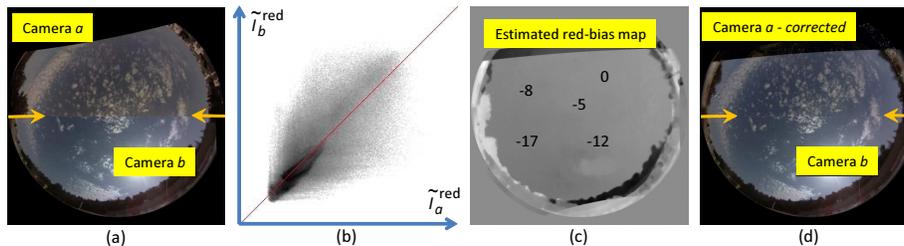
1. Different camera locations mean different observed objects. Momentarily, camera  $c$  may observe a cloud while  $c'$  observes a clear sky, or vice versa. Camera  $c'$  may momentarily observe a somewhat denser haze volume than  $c$ , etc.
2. Slight inter-camera radiometric inconsistencies, which 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  $\theta$ , 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 (STARS), detailed in Sec. 5. Figure 3a shows radiometric inconsistency between cameras  $a$  and  $b$ . Figure 3b is a scatter-plot of  $\tilde{I}_t[\mathbf{l}_a, \theta]$  vs.  $\tilde{I}_t[\mathbf{l}_b, \theta]$ ,  $\forall t, \theta$ , for the red-channel. From such plots, we hypothesized that camera  $a$  has a slight offset vs.  $b$ . We thus estimated, per color channel, the map of radiometric offset (across pixels, or ray-directions). A temporal median was used:

$$\hat{o}_{b-a}(\theta) = \text{median}_t\{\tilde{I}_t[\mathbf{l}_b, \theta] - \tilde{I}_t[\mathbf{l}_a, \theta]\}. \quad (3)$$

The map  $\hat{o}_{b-a}(\theta)$  was then spatially smoothed and used to correct  $\tilde{I}_t[\mathbf{l}_a, \theta]$ . As shown in Fig. 3d, the results have much better inter-camera consistency. A similar process was applied to other cameras, but they had negligible radiometric



**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  $\hat{I}_t[\mathbf{l}_a, \boldsymbol{\theta}]$  or  $\hat{I}_t[\mathbf{l}_b, \boldsymbol{\theta}]$ : inconsistencies in the line between the marked arrows are greatly diminished.

offsets with respect to camera  $b$ . The spatially varying offset in camera  $a$  was later found to be due to a nearby light source.

A similar process detects slight variations of gain (vignetting). Suppose there is no offset. In analogy to Eq. (3), the gain in  $b$  is higher than in  $a$  by a factor

$$\hat{M}_{b/a}(\boldsymbol{\theta}) = \text{median}_t \{ \tilde{I}_t[\mathbf{l}_b, \boldsymbol{\theta}] / \tilde{I}_t[\mathbf{l}_a, \boldsymbol{\theta}] \}. \quad (4)$$

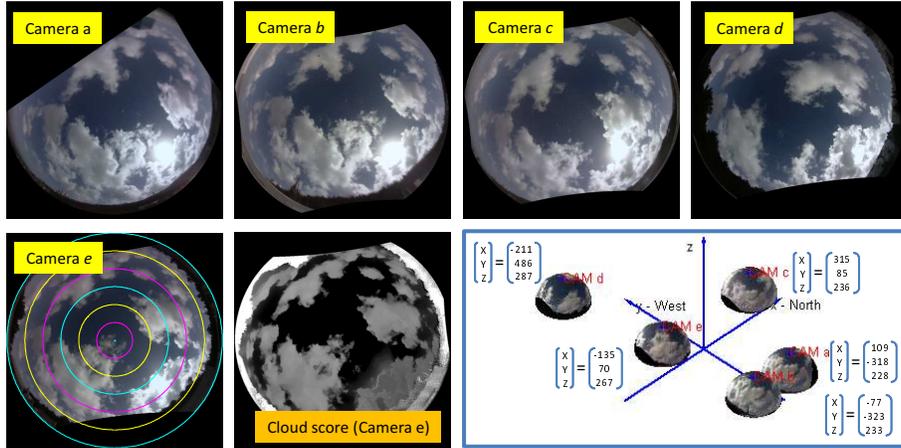
This way, all the network is radiometrically aligned to a single master camera. After radiometric corrections, the light-field samples are denoted  $\hat{I}_t[\mathbf{l}_b, \boldsymbol{\theta}]$ .

## 5 More Details About The Experimental Setup

Before proceeding with mathematical problems and solutions, we give more details about a small version STARS network. Each of the five nodes is built from a basic component set. Its core is a Raspberry-Pi computer and a 5MP Raspberry-Pi camera, whose gain, response and white-balance can be fixed, avoiding temporal radiometric variations. We manually mounted small fisheye lenses. Due to this coarse lens-to-chip alignment, each camera has a different peripheral dead-region, creating a missing part in the view-field and distinct vignetting (Fig. 4). As we explain, a network as-a-whole can inherently overcome these issues. Every 30 seconds, synchronously, all units automatically transmit image data to the internet (cloud-service). Each unit is solar powered. STARS operated for weeks from rooftops at the Technion, uploading data [17].

## 6 Network-assisted Background Estimation

In monocular settings, change-detection algorithms use temporal filtering to characterize the background: foreground dynamic objects are at *different lo-*



**Fig. 4.** Images taken simultaneously by a 5-node STARS 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 had dead-regions, due to rough lens alignment. Corresponding to the frame in camera  $e$ , a *cloud score* map (Eq. 8) has high values in cloud-pixels, diminishing outside them. [Bottom-right] The 3D setup of the small STARS, laterally spread over hundreds of meters, at somewhat different altitudes.

*cations 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.

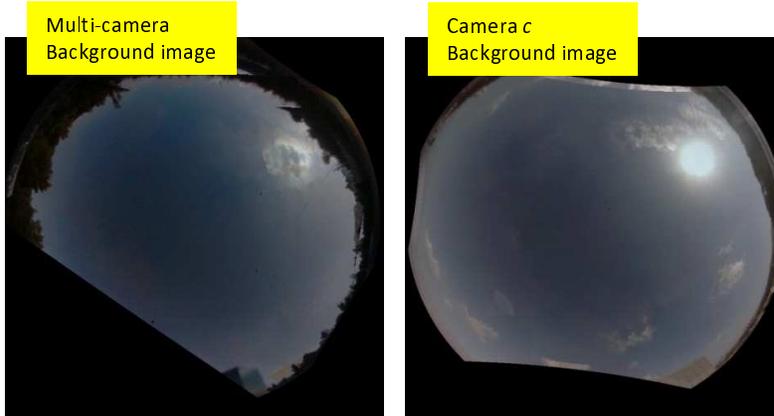
A light-field network enhances this principle, with more effective pruning-per-time. Recall regional stationarity (Sec. 3). A change of viewpoint has an effect similar to change in time: a cloud in  $\tilde{I}_t[\mathbf{l}_c, \boldsymbol{\theta}]$  is often not in  $\tilde{I}_t[\mathbf{l}_{c'}, \boldsymbol{\theta}]$ . Consequently, background sky values are obtained by data filtering over *both* time and viewpoint. This network-based principle can enhance arbitrary background estimation algorithms, which would otherwise be monocular. We demonstrate this using a simplistic, basic criterion. In broad daylight, clouds are brighter than the sky [24]. Hence, an estimator for the sky background can be, for example

$$\text{SKY}(\boldsymbol{\theta}) = \arg \min_{t,c} \tilde{I}_t[\mathbf{l}_c, \boldsymbol{\theta}] \quad (5)$$

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



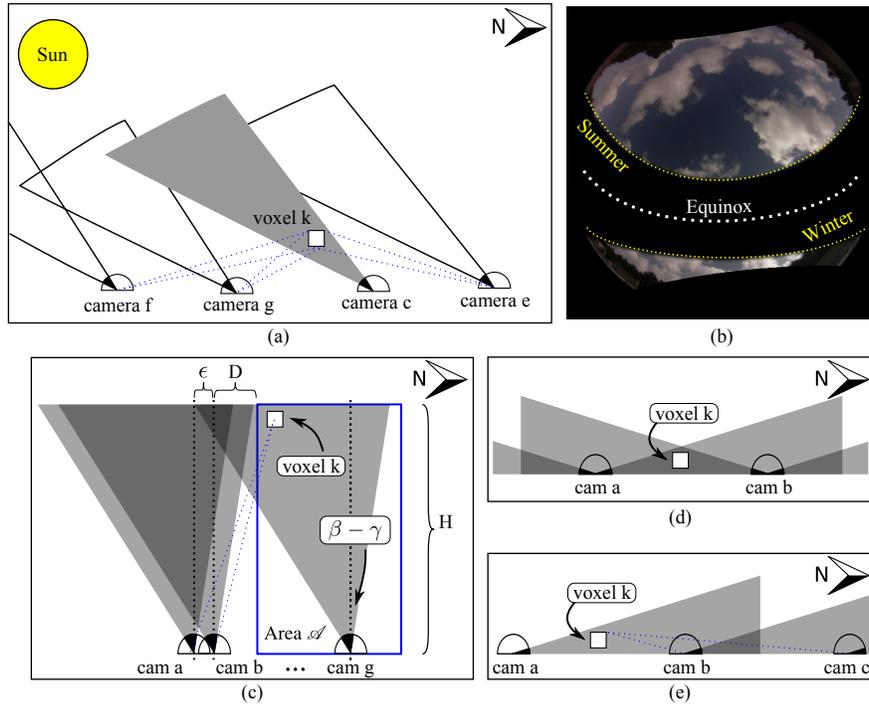
**Fig. 5.** [Left] Estimation of the sky background, using Eq. (5) based on five temporal instances and five viewpoints. [Right] Estimation of the sky background, using five temporal instances, but just a single viewpoint, resulting in more residual clouds.

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  $\Gamma_c$  of atmospheric voxels. These voxels, however, are generally visible at several other cameras, e.g., those indexed  $e, f, g$  in Fig. 6. Hence, a sufficiently wide network has no 3D blind spot, despite permanent sun-blocking. Voxels that are not obstructed in any view are better constrained than voxels in  $\Gamma_c$ .

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  $\gamma$ , thus, a permanent sun blocker spans zenith angles in the range  $\gamma \pm \beta$ . Earth is divided here to three region classes:

- In the tropics, the sky directly above a camera is blocked. Consider a small area  $\mathcal{A}$ , e.g., 1km wide. According to Fig. 6c, the sky above  $\mathcal{A}$  can efficiently be observed without a blind spot by cameras to its *south*. The network needs units extending to distance  $D = H \tan(\beta - \gamma) + \epsilon$  from  $\mathcal{A}$ , where  $\epsilon$  is a small distance, sufficient for triangulation at  $H$ . At the equator  $D \approx 7.4$ km. It can be shown that if  $\mathcal{A}$  is wider than  $2H[\tan(\beta - \gamma) + \tan(\beta + \gamma)]$ , the network can triangulate all the sky above it.
- As latitude increases towards the tropic circles,  $D$  decreases to zero. Thus the



**Fig. 6.** [a] Camera  $c$  has a blind-region, covering Sun directions at  $\mathbf{l}_c$ . The blind region corresponds to set  $\Gamma_c$  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  $\gamma$ , 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  $360^\circ$  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.

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

One application is estimation of the 3D cloud field above the network domain, and beyond. This can be done by the following steps: (A) Per time  $t$ , give a cloud score  $s$  to each ray  $[\mathbf{l}_c, \boldsymbol{\theta}]$ , as we explain below. (B) Perform a fuzzy version of space carving [25, 31].

We first describe a simple 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  $[\mathbf{l}_c, \boldsymbol{\theta}]$ . Voxel  $k$  projects to a subset of the rays  $\rho_k \subset \mathcal{R}$ , that reach  $\nu_k$  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

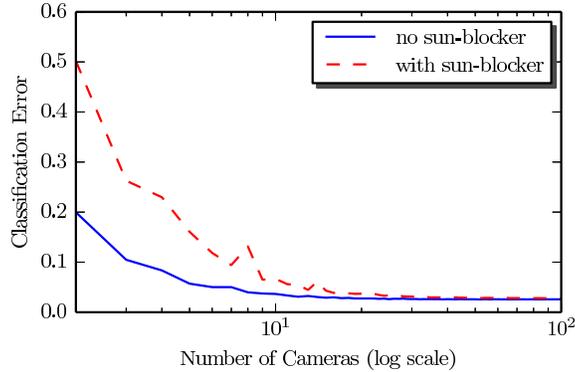
$$B_k = \left[ \prod_{r \in \rho_k} s(r) \right]^{1/|\rho_k|} \quad \text{if } \nu_k \geq 2. \quad (6)$$

This score 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 (6) carves-out voxels that contradict support for clouds.

Different cloud regions have signature appearances. Ignoring this would allow erroneous matching of, say, a darker cloud-bottom to a bright sun-lit side of a cloud. Thus, photometric and appearance consistency across viewpoints is incorporated (the photo-hull concept in space-carving [31]). From the images, a feature vector  $\mathbf{v}(r, t)$  is extracted for any measured ray  $r$ . We used SIFT descriptors [38] 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 [49] as

$$P_k = \exp \left( -\Sigma_q \{ \text{VAR}[\mathcal{V}_q(k, t)] \} / \sigma^2 \right), \quad (7)$$

where  $\sigma^2$  is a scale parameter. 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 can be pruned (carved) out. Voxel  $k$  can maintain its  $T_k$  if there are at least two camera viewpoints from which  $k$  is not occluded by other cloudy voxels. Pruning is achieved by sweeping [31] 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.



**Fig. 7.** Classification error rate vs.  $N^{\text{views}}$ , without or with a sun blocker.

### Basic cloud score

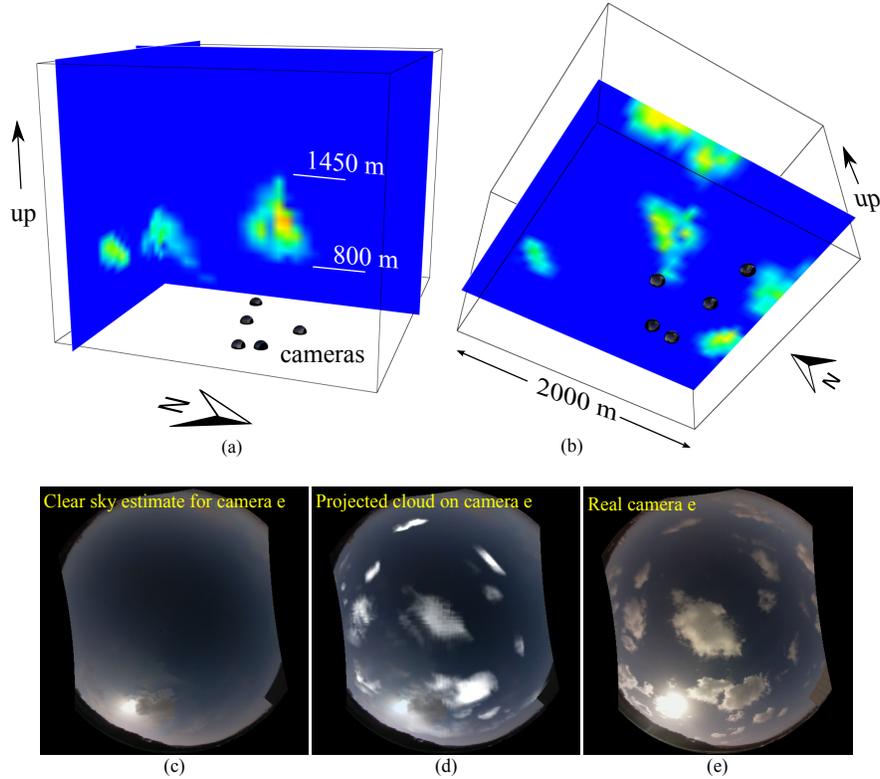
In the literature there are various functions [40] for a basic cloud score (step A). A ratio of image readout at the red/blue color channels,  $\tilde{I}^{\text{red}}/\tilde{I}^{\text{blue}}$ , is widely used [44, 50]. Overall, we found it effective in broad daylight: clouds are grey (unit red-blue ratio), and the cloudless sky is significantly biased to blue. Thus, for demonstrations in this paper, the cloud-score we used per ray (pixel) is

$$s(r) = \begin{cases} \frac{6[\tilde{I}^{\text{red}}(r)/\tilde{I}^{\text{blue}}(r)-0.8]}{0.2+\tilde{I}^{\text{red}}(r)/\tilde{I}^{\text{blue}}(r)} & \text{if } \tilde{I}^{\text{red}}(r)/\tilde{I}^{\text{blue}}(r) > 0.8 \\ 0 & \text{otherwise} \end{cases} . \quad (8)$$

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 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. Radiative transfer using the discrete ordinate spherical-harmonic method (SHDOM) [22] rendered images taken by 100 cameras 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 (Fluctuations are within the random sampling standard deviation). Even with a sun blocker, the algorithm is able to reconstruct the cloud formation, but, more cameras are needed in order to compensate for the limited view of each camera.



**Fig. 8.** 3D cumulus cloud recovery results. (a,b) Cross-sections of the recovered cloud-occupancy field  $\{\tilde{T}_k\}$ . The domain of the clouds is much larger than STARS. Cloud altitude is above sea-level. (c) Estimated sky-background image. Based on four view-points (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 overlaid on the estimated sky-background image. The resulting synthetic cloud-score image  $J[\mathbf{l}_e, \boldsymbol{\theta}]$  is shown in (d). This can be compared to the real captured image  $\hat{I}_t[\mathbf{l}_e, \boldsymbol{\theta}]$ , shown in (e).

### Cloud Reconstruction Experiment

We applied the approach on various captured scenes.<sup>4</sup> One scene had cumulus clouds. 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 STARS. Accounting for the altitude of STARS above sea-level, the clouds mainly reside between 800m to 1450m above sea-level. We used two indicators to validate the results. First, a balloon-based radiosonde measured the vertical humidity profile in Beit-Dagan. It is on a similar coastal strip, and roughly used by forecasters

<sup>4</sup> Sun blocker was not used here, since saturation and blooming do not impair cloud shape reconstruction.

for our Technion 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 a 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 created as follows. *Ray casting* [34] of field  $\{\tilde{T}_k\}$  is performed on a ray corresponding to  $[\mathbf{l}_e, \boldsymbol{\theta}]$ . Ray-casting aggregates  $\{T_k\}$  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  $\alpha$ -map to the estimated sky-background image (Eq. 5). The  $\alpha$ -map is

$$\alpha[\mathbf{l}_e, \boldsymbol{\theta}] = \begin{cases} 2w[\mathbf{l}_e, \boldsymbol{\theta}] & \text{if } 2w[\mathbf{l}_e, \boldsymbol{\theta}] < 1 \\ 1 & \text{otherwise} \end{cases} . \quad (9)$$

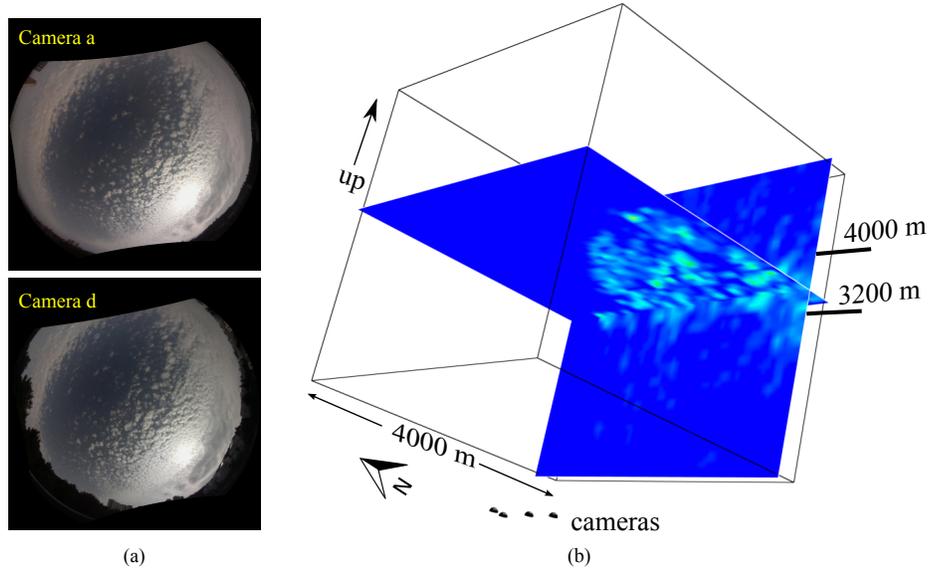
The rendered image is then  $J[\mathbf{l}_e, \boldsymbol{\theta}] = \alpha[\mathbf{l}_e, \boldsymbol{\theta}] + (1 - \alpha[\mathbf{l}_e, \boldsymbol{\theta}])\text{SKY}(\boldsymbol{\theta})$ . This image does not pretend to properly render clouds in their true shades and effect on the sky. It simply served to visualize the result (Fig. 8d), compared to the true corresponding image  $\hat{I}_t[\mathbf{l}_e, \boldsymbol{\theta}]$ , in Fig. 8e. Like sun-blocking, this rendering exploits network redundancy. Even if a viewpoint is blocked, much of its information can be derived using other viewpoints compounded with 3D recovery.

Another scene had a layer of altocumulus clouds. Figure 9 shows sample frames from this scene, and a cross-section of the recovered 3D cloud-occupancy field  $\{\tilde{T}_k\}$ . Accounting for the altitude of STARS, these estimated clouds mainly reside on a horizontal layer at  $\approx 3450 \pm 500\text{m}$  above sea-level. Here, the radiosonde 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.

## 9 Discussion

Scaling light-field imaging hugely to sense the sky, would use a large network of camera nodes, each having a wide field of view, deployed over a wide area. Such a network can reach anywhere communication exists. 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. This sensing approach poses new questions for computer vision and computational photography. These include network-based extensions to monocular tasks including network-based radiometric calibration and background estimation. Network redundancy offers the ability of by-passing saturation or blind spots, as those created by the sun, without moving parts.

To enable a massive network, each node should have very low-cost. To demonstrate this can work, units in the small STARS used very basic components and coarse alignment. This concept can spawn more interesting research. In the direction of the sun blocker, other sensors can be incorporated. Night-time operation is an interesting challenge. Furthermore, such a light-field system can be used for studying airborne animals (birds, bats [15], locust), in 3D time-lapses.



**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.

## 10 Acknowledgments

We are grateful to Pinhas Alpert, Daniel Rosenfeld, Orit Altaratz-Stollar, Nir Stav, Raanan Fattal, Arnon Karnieli, David Diner and Anthony Davis for useful discussions. We thank Mark Shenin and Technion building superintendents for experiment assistance. We thank Johanan Erez, Ina Talmon, Tamar Galateanu and Dani Yagodin for technical support. Yoav Schechner is a Lanadu Fellow - supported by the Taub Foundation. His research is supported in part by the Israel Science Foundation (ISF Grant 1467/12) and the Asher Space Research Institute. This work was conducted in the Ollendorff Minerva Center. Minerva is funded through the BMBF.

## References

1. Adelson, E., Wang, J.: Single lens stereo with a plenoptic camera. *IEEE Trans. PAMI* **14** (1992) 99–106
2. 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
3. Allmen, M.C., Kegelmeyer Jr., P.: The computation of cloud base height from paired whole-sky imaging cameras. *Machine Vision and Applications* **9** (1997) 160–165
4. Alterman, M., Schechner, Y.Y., Swirski, Y.: Triangulation in random refractive distortions. In: *Proc. IEEE ICCP* (2013) 1–10

5. Alterman, M., Schechner, Y.Y., Vo, M., Narasimhan, S.G.: Passive tomography of turbulence strength. In: Proc. ECCV (2014) 47–60
6. Alterman, M., Swirski, Y., Schechner, Y.Y.: STELLA MARIS: Stellar marine refractive imaging sensor. In: Proc. IEEE ICCP (2014) 1–10
7. Atkinson, B.W.: *Meso-Scale Atmospheric Circulations* Academic Press (1989), 19
8. Basha, T., Avidan, S., Hornung, a., Matusik, W.: Structure and motion from scene registration. Proc. IEEE CVPR (2012) 1426–1433
9. Baumgarten, G., Fiedler, J., Fricke, K.H., Gerding, M., Hervig, M., Hoffmann, P., Müller, N., Pautet, P.D., Rapp, M., Robert, C., Rusch, D., von Savigny, C., Singer, W.: The noctilucent cloud (NLC) display during the ECOMA/MASS sounding rocket flights on 3 August 2007: morphology on global to local scales. *Annales Geophysicae* **27** (2009) 953–965
10. Baxter, B., Hooper, B.A., Williams, J.Z., Dugan, J.P.: Polarimetric remote sensing of ocean waves. In: Proc. MTS/IEEE OCEANS (2009) 1–5
11. Bennett, G.G.: The calculation of astronomical refraction in marine navigation. *The Journal of Navigation* **35** (1982) 255–259
12. Bishop, T.E., Zanetti, S., Favaro, P.: Light field superresolution. Proc. IEEE ICCP **129** (2009) 1–9
13. Bradley, E.S., Toomey, M.P., Still, C.J., Roberts, D.A.: Multi-scale sensor fusion with an online application: Integrating GOES, MODIS, and webcam imagery for environmental monitoring. *IEEE Selected Topics in Applied Earth Obs. and Remote Sen.* **3** (2010) 497–506
14. 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
15. Breslav, M., Fuller, N., Sclaroff, S., Betke, M.: 3D Pose estimation of bats in the wild. In: Proc. IEEE WACV (2014)
16. Cazorla, a., Olmo, F.J., Aladosarboledas, L., Alados-Arboledas, L.: Using a sky imager for aerosol characterization. *Atmospheric Environment* **42** (2008) 2739–2745
17. Clouds in the Cloud webpage and data link  
[http://webee.technion.ac.il/~yoav/research/clouds\\_inthe\\_cloud.html](http://webee.technion.ac.il/~yoav/research/clouds_inthe_cloud.html)
18. 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 spectro-radiometer (MISR) instrument description and experiment overview. *IEEE Trans. Geoscience and Remote Sens.* **36** (1998) 1072–1087
19. Diner, D.J., Davis, A., Hancock, B., Gutt, G., Chipman, R.A., Cairns, B.: Dual-photoelastic-modulator-based polarimetric imaging remote sensing. *Appl. Opt.* **46** (2007) 8428–8445
20. 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. *Appl. Opt.* **49** (2010) 2929
21. Diner, D.J., Martonchik, J.V.: Atmospheric transmittance from spacecraft using multiple view angle imagery. *Appl. Opt.* **24** (1985) 3503–3511
22. Evans, K.F.: The spherical harmonics discrete ordinate method for three-dimensional atmospheric radiative transfer. *Journal of the Atmospheric Sciences* **55** (1998) 429–446
23. Horstmeyer, R., Euliss, G., Athale, R., Levoy, M.: Flexible multimodal camera using a light field architecture. In: Proc. IEEE ICCP (2009) 1–8

24. Hosek, L., Wilkie, A.: An analytic model for full spectral sky-dome radiance. *ACM TOG* **31** (2012) 95:1–95:9
25. Ihrke, I., Magnor, M.: Image-based tomographic reconstruction of flames. In: *Proc. ACM SIGGRAPH* (2004) 365–373
26. Jacobs, N., King, J., Bowers, D., Souvenir, R.: Estimating cloud maps from outdoor image sequences. In: *Proc. IEEE WACV* (2014)
27. Kang, S., Weiss, R.: Can we calibrate a camera using an image of a flat, textureless Lambertian surface? *Proc. ECCV* (2000) 640–653
28. Kassianov, E., Long, C., Christy, J.: Cloud-base-height estimation from paired ground-based hemispherical observations. *J. Applied Meteorology* **44** (2005) 1221–1233
29. Kim, J., Lanman, D., Mukaigawa, Y., Raskar, R.: Descattering transmission via angular filtering. In: *Proc. ECCV* (2010) 86–99
30. Koreban, F., Schechner, Y.Y.: Geometry by deflaring. *Proc. IEEE ICCP* (2009) 1–8
31. Kutulakos, K., Seitz, S.: A theory of shape by space carving. *IJCV* **38** (2000) 199–218
32. Lalonde, J.F., Narasimhan, S., Efros, A.: What do the sun and the sky tell us about the camera? *IJCV* **88** (2010) 24–51
33. Lensky, I., Rosenfeld, D.: The time-space exchangeability of satellite retrieved relations between cloud top temperature and particle effective radius. *Atmospheric Chemistry and Physics* **6** (2006) 2887–2894
34. Levoy, M.: Efficient ray tracing of volume data. *ACM TOG* **9** (1990) 245–261
35. Levoy, M., Ng, R., Adams, A., Footer, M., Horowitz, M.: Light field microscopy. *ACM TOG* **25** (2006) 924–934
36. Litvinov, A., Schechner, Y.Y.: Addressing radiometric nonidealities: a unified framework. In: *Proc. IEEE CVPR* (2005) 52–59 vol. 2
37. Long, C.N., Sabburg, J.M., Calbo, J., Pages, D., Calbó, J., Pagès, D.: Retrieving cloud characteristics from ground-based daytime color all-sky images. *J. Atmospheric and Oceanic Technology* **23** (2006) 633–652
38. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *IJCV* **60** (2004) 91–110
39. 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: *Proc. Airborne Imaging Spectroscopy Workshop* (2004)
40. Peng, Z., Yoo, S., Yu, D., Huang, D., Kalb, P., Heiser, J.: 3D cloud detection and tracking for solar forecast using multiple sky imagers. In: *Proc. ACM Sympos. Applied Computing* (2014) 512–517
41. Pust, N.J., Shaw, J.A.: Digital all-sky polarization imaging of partly cloudy skies. *Appl. Opt.* **47** (2008) H190–H198
42. Raskar, R., Agrawal, A., Wilson, C.a., Veeraraghavan, A.: Glare aware photography. *ACM TOG* **27** (2008) 56:1–56:10
43. Rouf, M., Mantiuk, R., Heidrich, W., Trentacoste, M., Lau, C.: Glare encoding of high dynamic range images. *Proc. IEEE CVPR* (2011) 289–296
44. Seiz, G., Baltsavias, E., Gruen, A.A.: Cloud mapping from the ground: Use of photogrammetric methods. *Photogrammetric Eng. and Remote Sensing* **68** (2002) 941–951
45. Seiz, G., Davies, R.: Reconstruction of cloud geometry from multi-view satellite images. *Remote Sensing of Environment* **100** (2006) 143–149

46. Seiz, G., Shields, J., Feister, U., Baltsavias, E.P., Gruen, A.: Cloud mapping with ground-based photogrammetric cameras. *Int. J. Remote Sens.* **28** (2007) 2001–2032
47. Schechner, Y.Y., Diner, D., Martonchik, J.: Spaceborne underwater imaging. In: *Proc. IEEE ICCP* (2011) 1–8
48. Talvala, E.V., Adams, A., Horowitz, M., Levoy, M.: Veiling glare in high dynamic range imaging. *ACM TOG* **26** (2007)
49. Veikherman, D., Aides, A., Schechner, Y.Y., Levis, A.: Clouds in the cloud: supplementary material for *Proc. ACCV* (2014)
50. Yamashita, M.: Cloud cover estimation using multitemporal hemisphere imageries. In: *Proc. XXth Congress of the Society for Photogrammetry and Remote Sensing* (2004) 818—821