

Stereophotogrammetry of Oceanic Clouds

RUSEN ÖKTEM*

*Department of Earth and Planetary Science, University of California, Berkeley, and Earth Sciences Division,
Lawrence Berkeley National Laboratory, Berkeley, California*

PRABHAT

*Department of Earth and Planetary Science, University of California, Berkeley, and Computational Research Division,
Lawrence Berkeley National Laboratory, Berkeley, California*

JAMES LEE AND AARON THOMAS

Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California

PAQUITA ZUIDEMA

Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, Miami, Florida

DAVID M. ROMPS

*Department of Earth and Planetary Science, University of California, Berkeley, and Earth Sciences Division,
Lawrence Berkeley National Laboratory, Berkeley, California*

(Manuscript received 29 October 2013, in final form 10 March 2014)

ABSTRACT

This study extends ground-based stereophotogrammetry of clouds to oceanic settings, where there are often none of the landmarks used in traditional camera calibration. This paper introduces a zero-landmark calibration technique and tests it with two off-the-shelf digital cameras situated about 1 km apart facing Biscayne Bay in Miami, Florida. The precision of the stereo reconstruction is studied theoretically, and the accuracy of the reconstructions is validated against lidar and radiosondes. The stereo cameras are able to accurately reconstruct a histogram of cloud-base heights from a single-image pair, a task that requires tens of minutes of observation from a cloud lidar. The stereo cameras are also able to accurately reconstruct horizontal winds in cloud layers with a temporal resolution in the range of 30 s to 5 min, compared to once every 12 h for a typical radiosonde launch site.

1. Introduction

Stereophotography has the potential to provide a unique window into the behavior of clouds, enabling the calculation of four-dimensional trajectories of cloud surfaces in exquisite detail. The fine spatial resolution and

high frame rate allow for the tracking of individual clouds and convective turrets, providing a perspective on cloud life cycles that cannot be replicated with other existing technologies. Radar, with its ability to measure the condensates and motions within clouds, comes closest to this capability. For example, vertically pointing W-band radar can map the distribution and vertical velocity of shallow, nonprecipitating clouds (Kollias et al. 2001, 2003, 2007; Kollias and Albrecht 2010; Ghate et al. 2011), but its “soda straw” view of the atmosphere cannot provide information on the life cycle of any individual cloud. To measure cloud life cycles, individual clouds must be imaged repeatedly, which would require the use of a scanning radar. The downside of a scanning radar is its

* Current affiliation: Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, California.

Corresponding author address: Rusen Öktem, Department of Earth and Planetary Science, University of California, Berkeley, 455 McCone Hall, Berkeley, CA 94720.
E-mail: roktem@lbl.gov

coarse spatial and temporal resolution, caused by typical beamwidths on the order of 1° and the time it takes for the radar to mechanically scan the sky. For example, scanning weather radar are typically used to generate cloud maps with a spatial and temporal resolution of ~ 1 km and ~ 10 min, respectively (e.g., Davies-Jones 1979; Yuter and Houze 1995; Collis et al. 2013), which is too coarse to track the life cycle of cloud turrets with either size ≤ 1 km or updraft speed ≥ 10 m s $^{-1}$.

To make a more apples-to-apples comparison between a scanning radar and a digital camera, we should consider a scanning mode that covers an atmospheric volume similar to a camera’s field of view. For example, the boundary layer range–height indicator (BL-RHI) scan, as employed by the W-band scanning radar at the Atmospheric Radiation Measurement Program (ARM) sites (Mather and Voyles 2013; Ackerman and Stokes 2003), covers 90° of elevation (horizon to zenith) and 80° of azimuth. This is comparable to the fields of view for the cameras used in this study: the horizontal fields of view are equal to 76° and 67° for the left and right cameras, respectively. With the BL-RHI scan, the sky is mapped with a resolution of 1° in elevation and 2° in azimuth once every 5 min (Fielding et al. 2013; Kollias et al. 2014). By comparison, an off-the-shelf digital camera has an angular resolution of about 0.1° , and it can capture images at frame rates greater than 1 Hz. Consider a convective cloud at a distance 20 km rising at 10 m s $^{-1}$. A BL-RHI scan will image the cloud with a transverse resolution of 700 m and a vertical displacement of the cloud between scans of 3 km. A camera with a modest 0.1-Hz frame rate will image the cloud with a resolution of about a few tens of meters and a vertical displacement between images of 100 m. With this spatial and temporal resolution, stereophotogrammetry can provide detailed cloud life cycle data.

Stereophotogrammetry dates back more than 100 years, with early cloud studies making use of theodolites (Koppe 1896). Analog photographs were used in the stereophotogrammetry of clouds from at least the 1950s to obtain cloud positions and velocities (Malkus and Ronne 1954; Kassander and Sims 1957; Orville and Kassander 1961; Bradbury and Fujita 1968; Warner et al. 1973). More recently, digital photographs have been used (Allmen and Kegelmeyer 1996; Kassianov et al. 2005; Zehnder et al. 2007; Damiani et al. 2008), which opens up the possibility of using feature-matching algorithms to automate the reconstruction process (Seiz et al. 2002; Zehnder et al. 2006).

In all of these studies, the essential first step is the careful calibration of camera positions and orientations. Orientation of the cameras—that is, the determination of the camera’s three Euler angles—is often the trickiest

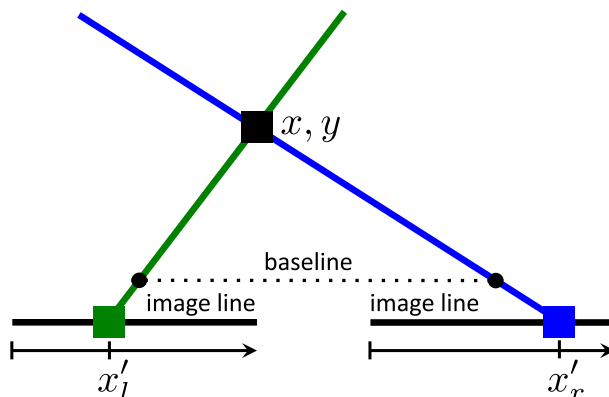


FIG. 1. An example of stereo reconstruction in two dimensions. Each camera is represented by a pinhole (black dot) and a projection screen (black line). The object of interest (black square) projects onto the left and right cameras at the positions of the green square and blue square, respectively. The left camera knows that the black square lies along the green line, and the right camera knows that the black square lies along the blue line. The position of the black square is calculated as the intersection of these two lines.

step. Previous approaches have used the positions of known landmarks, such as topographic features (e.g., Hu et al. 2009) or the positions of stars (e.g., Seiz et al. 2002). In this study, we describe how to calibrate a pair of cameras in the absence of landmarks, as needed for the case of daytime images of land-free scenes. This method is applied to images collected by a pair of cameras looking out over Biscayne Bay in Miami, Florida, during April 2013. Section 2 reviews the concept of three-dimensional reconstruction with stereo cameras. Section 3 discusses the setup and calibration of the two cameras. Section 4 calculates the precision of the stereo reconstruction. The accuracy of the reconstructions is validated by comparison with lidar in section 5 and radiosondes in section 6. Section 7 summarizes the results.

2. Calibration without landmarks

Given two contemporaneous photographs from two cameras with known positions and orientations, we can reconstruct the position of any object that appears in both images. This “stereo reconstruction” of an object’s three-dimensional position is simply a matter of triangulation. An example in two dimensions is given in Fig. 1. Each of the two cameras (left and right) is represented by a pinhole through which light is admitted (black circle) and an image line where a photosensitive array would be located (black line). Each camera reports a single number: the position of the object in the camera’s image plane. We can write these positions as x'_l and x'_r , which we can think of as pixel indices in digital image sensor arrays. The disparity d is defined as

the difference in these two image-plane positions: $d \equiv x'_l - x'_r$. Using triangulation, these two numbers allow us to determine the location of the object (x and y) in two-dimensional space. The left camera tells us that the object lies somewhere on the locus of points denoted by the green line, for all objects on that line will project onto x'_l . The right camera tells us that the object lies somewhere on the locus of points denoted by the blue line, for all objects on that line will project onto x'_r . The true location of the object is given by the intersection of these two lines.

The line passing through the two pinholes is referred to as the baseline. In Fig. 1, the two cameras are drawn with parallel baseline and image lines. In some applications, this type of arrangement is both desirable and practical. For example, if a short baseline is appropriate, then the two cameras can be physically attached to one another, thereby guaranteeing that the baseline and image lines are parallel. In some applications, however, the required baseline is too long for this approach to be feasible. Baselines are chosen to be about one or two orders of magnitude shorter than the distance from the cameras to the objects being imaged (Gallup et al. 2008). This reflects a trade-off between 1) the greater disparity with a larger baseline, which makes the reconstruction more precise; and 2) the increased difficulty of identifying matching points (and a greater occurrence of occlusions) with a larger baseline, which reduces the amount of data. If we wish to capture clouds with a resolution of 100 m or better with a 4-mm focal-length lens and a sensor that is 4 mm wide with 1000 pixels across, then the basic pinhole model geometry (Forsyth and Ponce 2003) tells us that we are restricted to looking at clouds within 100 km. These distances are best sampled with a baseline on the order of 1–10 km, which is far too large to have the cameras attached to each other by a rigid structure.

Therefore, for stereophotography of clouds, the two cameras must be sited and oriented independently of one another. This leads to a potential problem: errors in the assumed position and orientation of the cameras will generate errors in the reconstructed positions of objects. In Fig. 1, the reconstructed x and y will be erroneous if we believe that the two image lines are parallel with the baseline when, in fact, one of them is not. To mitigate this source of error, careful attention must be paid to the calibration (i.e., measurement) of camera position, internal optical parameters, and camera orientation. Camera position can be determined by using the global positioning system (GPS) or by collocation of the camera with known landmarks. Internal optical parameters can be obtained from the manufacturer or from controlled laboratory tests of the camera. Orientation,

on the other hand, must be measured in the field upon deployment.

A standard approach to calibrating orientation is to use landmarks with known positions. For example, in the two-dimensional case, at least one landmark is needed in the field of view to determine the one angle of orientation. In three dimensions, there are three Euler angles (roll, pitch, and yaw), which can be measured with at least two landmarks. This is the approach taken by recent stereophotogrammetric studies (e.g., Zehnder et al. 2007; Hu et al. 2009). But, what if there are fewer than two landmarks or no landmarks at all? In the study of maritime clouds, cameras looking out over the ocean will see, in general, only ocean and sky, with no fiducial landmarks in the field of view. Here, we report on a calibration technique for this scenario, that is, a calibration method that does not require the use of landmarks.

To explain calibration without landmarks, let us return, for the moment, to Fig. 1. In two dimensions, there is almost always a solution to the intersection of two lines of sight (the rare exception being when the lines are supposedly parallel). Therefore, the cameras will report an unambiguous position for almost any object, regardless of errors in calibration. This can be stated mathematically as follows: for supposedly parallel baseline and image lines, any pair of $x'_l \neq x'_r$ will yield an unambiguous prediction for x and y . While this may sound like a good thing, it can cause problems by masking errors in the calibration: stereo cameras will report unambiguous reconstructions whether they are well calibrated or not.

In three dimensions, however, this is not the case. Generally, any error in the calibration will cause the lines of sight to fail to intersect in three-dimensional space. This is illustrated in Fig. 2, which shows the image planes for the left and right cameras. In the left camera, the object's image is denoted by the green square, which has a position in the image plane denoted by x'_l and y'_l . Given the image location (x'_l , y'_l) and the assumed position and orientation of the left camera, we can calculate the line in three-dimensional space on which the object must lie (i.e., the set of all points in three-dimensional space that would project onto the green square). Given the position and orientation of the right camera, we can then draw the theoretical image of that line as the green line in the image plane of the right camera; this is called the epipolar line. If the position and orientation of the cameras were known with perfect accuracy, then this epipolar line would pass through the object in the image plane. In practice, however, an error in the calibration of the cameras leads to an error that manifests itself as some amount of displacement of the image from the epipolar line. Similarly, the object's image

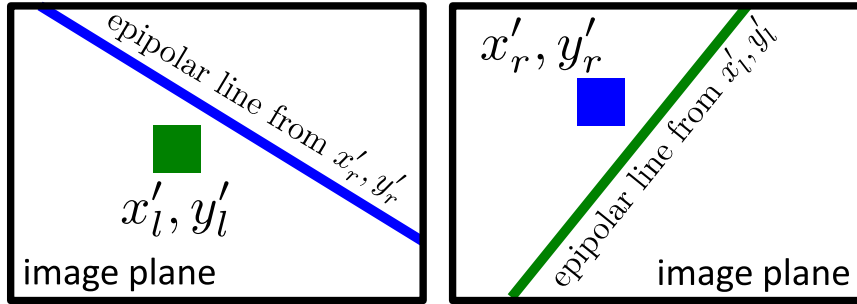


FIG. 2. In practice, errors in the calibration of camera position and orientation lead to failure of the two loci to intersect in three dimensions. In a camera's image plane, this manifests itself as a displacement of the object's image from the epipolar line predicted by the other camera.

in the left camera's image plane is also displaced from the epipolar line coming from the right camera.

In the absence of landmarks, we can use this mismatch to calibrate the cameras. The idea is simple in principle. Given a large number of points in the field of view (spots on clouds, ships in the ocean, birds in the sky), define an objective function as the sum (over all points and both image planes) of the deviation (in the image plane) between the image and the associated epipolar line. Then, using an optimization algorithm, find the set of calibration parameters that minimizes this objective function. In practice, this is an optimization problem of high dimensionality (six position parameters, six Euler angles, and several internal optical parameters), and so is numerically intractable without a good initial estimate of the parameters. In the case described in section 3c, estimates are obtained using Google Earth for the camera positions (GPS would also work for this purpose), the position of the sun for yaw (a compass would work just as well, correcting for magnetic declination), and the ocean horizon for pitch and roll.

3. Camera setup and calibration

Stereo reconstruction is an implementation of projective geometry, where the projection of a 3D point in space onto a 2D image frame is related through a projection matrix \mathbf{P} . Let $\mathbf{X} = (x, y, z, 1)$ denote the homogenous coordinates of a point in 3D space. In homogeneous coordinates, $(\alpha x, \alpha y, \alpha z, \alpha)$ defines the same point for all $\alpha \neq 0$. Similarly, let $\mathbf{X}' = (x', y', 1)$ be the homogeneous coordinates of the image in the image plane. The definitions and units of the notations used throughout the paper are listed in Table 1. For a pinhole camera, the relation between \mathbf{X} and \mathbf{X}' is expressed as (Forsyth and Ponce 2003)

$$\mathbf{X}' = \mathbf{P}\mathbf{X}. \tag{1}$$

The projection matrix is the product of a camera matrix \mathbf{C} , rotation matrix \mathbf{R} , and translation matrix \mathbf{T} :

$$\mathbf{P} = \mathbf{C}\mathbf{R}\mathbf{T}. \tag{2}$$

The camera matrix \mathbf{C} is given by

$$\mathbf{C} = \begin{bmatrix} fk_x & 0 & c_x \\ 0 & fk_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where f is the focal length; k_x (k_y) is the density of pixels, that is, number of pixels per distance, in the image plane's x' (y') direction; and c_x (c_y) is the principal point coordinate of the image plane. The rotation matrix \mathbf{R} aligns the world coordinates (x, y, z) into the coordinate system aligned with the camera (x' and y' in the image plane and z' perpendicular to the image plane in the direction of the line of sight). Letting θ_1 , θ_2 , and θ_3 denote the three Euler angles of pitch, yaw, and roll, respectively, \mathbf{R} can be expressed as

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & \sin\theta_1 \\ 0 & -\sin\theta_1 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} \cos\theta_2 & 0 & -\sin\theta_2 \\ 0 & 1 & 0 \\ \sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix} \\ \times \begin{bmatrix} \cos\theta_3 & -\sin\theta_3 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

The translation matrix transforms an object's world coordinates into camera-relative coordinates by subtracting the camera's world coordinate. Denoting the camera's world coordinates by x_0 , y_0 , and z_0 , \mathbf{T} is given by

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & -x_0 \\ 0 & 1 & 0 & -y_0 \\ 0 & 0 & 1 & -z_0 \end{bmatrix}. \tag{5}$$

TABLE 1. List of variables and their units.

Variable	Definition	Unit
$x_{0,l}, y_{0,l}, z_{0,l}$	World coordinates of the left camera	m
$x_{0,r}, y_{0,r}, z_{0,r}$	World coordinates of the right camera	m
x, y	Position to the east and to the north of the right camera	m
z	Height relative to the sea level	m
$\theta_1, \theta_2, \theta_3$	Euler angles of pitch, yaw, and roll	rad
x', y'	Pixel indices in the image plane	pixel
f	Focal length	mm
k_x, k_y	Pixel density	mm^{-1}
c_x, c_y	Principal point coordinate	pixel
k_1, k_2, k_3	Lens distortion coefficients	—
a_r	Intercept in horizon line fitting	pixel
m_r	Slope in horizon line fitting	—
u, v	Eastward and northward wind speeds	m s^{-1}

Note that all the expressions in this text assume a Cartesian coordinate system.

The projections of an object onto the two image planes provides two sets of equations,

$$\mathbf{X}'_l = \mathbf{P}_l \mathbf{X} \quad \text{and} \quad \mathbf{X}'_r = \mathbf{P}_r \mathbf{X}, \quad (6)$$

where subscripts l and r refer to the left and right cameras, respectively. Stereo reconstruction is the process of calculating \mathbf{X} from \mathbf{X}'_l and \mathbf{X}'_r using these two equations (Hartley and Zisserman 2003). Of course, in order for this procedure to work, the projection matrices must be known; that is, the cameras must be calibrated. Camera calibration will be discussed below after the description of the camera setup.

a. Camera setup

The two cameras used in this study are located on Virginia Key in Miami, where they overlook Biscayne

Bay (25.6°N, 80.2°W; see Fig. 3). One of the cameras (a 5-megapixel Stardot NetCam SC) is mounted on the roof of the Marine and Science Technology (MAST) Academy. This camera will be referred to as the right camera throughout the text. The left camera (a 3-megapixel Stardot NetCam XL) is mounted on the roof of the Rosenstiel School of Marine and Atmospheric Science (RSMAS) as part of South Florida's Cloud–Aerosol–Rain Observatory (CAROB; <http://carob.rsmas.miami.edu>). It is located 296 m to the east and 822 m to the south of the right camera. The right and left cameras face 6.56° and 18.19° west of south, respectively. The right and left cameras capture images at 1296 × 960 and 1024 × 768 pixel resolution, and 30- and 3-s intervals, respectively. The left camera was acquired in 2011 as part of CAROB, but the right camera was acquired in 2012 to launch the stereo reconstruction research; hence, it has a higher resolution than the left camera. The two cameras are connected to separate servers and so are controlled independently from each other, but they are synchronized to a common time server. Both cameras have wide-angle lenses, which cause noticeable radial distortion that cannot be ignored as in previous works (such as Hu et al. 2009).

b. Camera lens distortion

Although the theory of stereophotogrammetry is formulated for pinhole cameras, real camera lenses do not behave as pinholes. Wide-angle lenses, in particular, suffer from a significant amount of radial distortion (i.e., deviation of the projection from that generated by a pinhole). The cameras used here have wide angles of view—67° and 76°—and the radial distortion is clearly visible (see Fig. 4). This distortion is corrected by a mapping based on the Brown distortion model (Brown 1966),

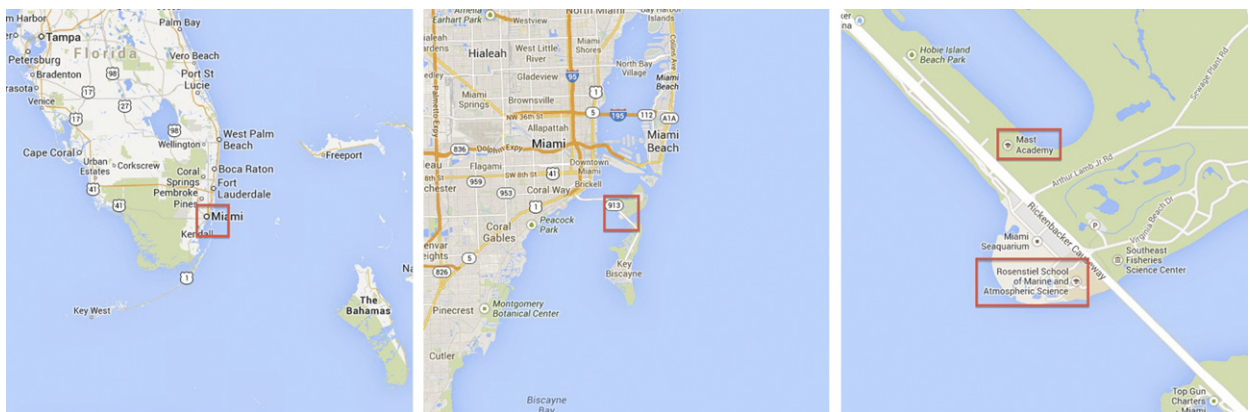


FIG. 3. (left to right) The locations of the two cameras shown as maps with increasing zoom. The red squares in (left) and (middle) indicate the regions displayed in (middle) and (right), respectively. The red rectangles in (right) indicate the locations of the two cameras.



FIG. 4. Images (top) before and (bottom) after barrel-distortion correction for the (left) and (right) right cameras.

$$\tilde{x} = (x - c_x)(1 + k_1r^2 + k_2r^4 + k_3r^6) + c_x, \quad (7a)$$

$$\tilde{y} = (y - c_y)(1 + k_1r^2 + k_2r^4 + k_3r^6) + c_y, \quad (7b)$$

$$r^2 = (x - c_x)^2 + (y - c_y)^2, \quad (7c)$$

where (\tilde{x}, \tilde{y}) is the observed (distorted) position of a point in the image plane, and (x, y) is the corrected position (i.e., where on the image plane the point would lie if the lens were replaced by a pinhole). Equations 7a and 7b establish a relation between the observed and correct pixel positions in terms of unknown radial distortion parameters k_i . A checkerboard pattern is printed out and held in front of the camera at various angles and positions. The distorted corner-point positions are extracted from the captured images by an automatic corner-extraction algorithm in OpenCV, an open source computer-vision library (Bradski and Kaehler 2008). A lens correction algorithm from the same OpenCV library is then used to estimate the k_i using Levenberg–Marquardt optimization (Press et al. 2007). With the k_i estimated in this way, the mapping of Eqs. (7a)–(7c) is applied to compensate for the distortion. Samples of images before and after this correction are displayed in

Fig. 4. OpenCV lens distortion correction algorithm also gives an estimate of fk , c_x , and c_y , together with k_i .

c. Camera calibration

Camera “calibration” refers to the estimation of intrinsic parameters (given by the elements of \mathbf{C}) and extrinsic parameters (given by the elements of \mathbf{R} and \mathbf{T}). The elements of \mathbf{C} can be extracted from the camera’s technical specifications if the lens model is known. The cameras used here are equipped with varifocal lenses, with focal length f varying between 4 and 8 mm. This sets a lower and upper bound on f , but the precise value is unknown a priori. The principal point coordinate, denoted by (c_x, c_y) , is often assumed to be at the center of the image plane, but it may deviate from the center depending on the lens used.

Camera extrinsic parameters include the world coordinate of the camera center and the three angles of the camera’s yaw, roll, and pitch. These parameters must be determined with high accuracy to achieve meaningful reconstructions. The Google Earth mapping software provides the coordinates of both cameras in 3D. The sun provides a way of estimating the yaw when it is low enough to be captured by the cameras in January and February. By recording the time of day when the sun is

centered horizontally on the image plane, the azimuth of the sun at that time (obtained from <http://www.esrl.noaa.gov/gmd/grad/solcalc>) gives the absolute yaw of the right camera and an initial estimate of the relative yaw between the two cameras. This and other parameters are then calculated using the epipolar constraint and the horizon constraint, described below.

1) THE EPIPOLAR CONSTRAINT

Given the homogeneous world coordinate \mathbf{X} of some object, the projection matrix \mathbf{P}_l maps \mathbf{X} onto the homogeneous image coordinate \mathbf{X}'_l in the left camera's image plane, as discussed above. In other words, $\mathbf{X}'_l = \mathbf{P}_l \mathbf{X}$. Likewise, for the right camera, $\mathbf{X}'_r = \mathbf{P}_r \mathbf{X}$. As discussed in section 2, the epipolar line in the right camera's image plane is the image of all world coordinates that would project onto \mathbf{X}'_l in the left camera's image plane. Likewise, the epipolar line in the left camera's image plane is the image of all world coordinates that would project onto \mathbf{X}'_r in the right camera's image plane. If \mathbf{P}_l and \mathbf{P}_r are known perfectly, then \mathbf{X}'_r will lie on the epipolar line in the right camera's image plane and likewise for \mathbf{X}'_l in the left camera's image plane. On the other hand, if \mathbf{P}_l and \mathbf{P}_r contain error, then images will not, in general, coincide with the epipolar lines. In particular, it is error in the internal parameters (focal length and principal point) and in the relative displacement, relative pitch, relative yaw, and relative roll of the two cameras that causes these deviations. Note that changes in displacement, pitch, yaw, and roll that are applied equally to both cameras cannot cause images to deviate from epipolar lines; only relative errors in these quantities manifest as deviations of images from epipolar lines.

This deviation of images from epipolar lines can be quantified using the "fundamental matrix" \mathbf{F} , which is constructed from \mathbf{P}_l and \mathbf{P}_r (for details, see Hartley and Zisserman 2003; Forsyth and Ponce 2003). The elements of $\mathbf{F}\mathbf{X}'_r$ define the epipolar line in the left camera's image plane; that is, if $\mathbf{F}\mathbf{X}'_r = (a, b, c)^T$, then the epipolar line in the left image plane is equal to $(x'_l, y'_l: ax'_l + by'_l + c = 0)$. If the calibration of \mathbf{P}_l and \mathbf{P}_r is perfect, then

$$\mathbf{X}'_l{}^T \mathbf{F}\mathbf{X}'_r = 0, \quad (8)$$

for all image pairs \mathbf{X}'_r and \mathbf{X}'_l . Writing $\mathbf{F}\mathbf{X}'_r = (a, b, c)^T$, this equation reduces to

$$(x'_l \quad y'_l \quad 1) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = ax'_l + by'_l + c = 0,$$

which simply implies that \mathbf{X}'_l is on the epipolar line.

The fundamental matrix is a 3×3 matrix with rank 2. Equation (8) implies that the fundamental matrix can be estimated from point correspondences, in the least squares error sense. Note that, although \mathbf{P}_l and \mathbf{P}_r uniquely define the fundamental matrix, the reverse is not true. The fundamental matrix determines \mathbf{P}_l and \mathbf{P}_r up to an overall displacement, pitch, yaw, and roll; in other words, \mathbf{F} gives the relative displacement, relative pitch, relative yaw, and relative roll of the left camera with respect to the right camera. The absolute position, pitch, yaw, and roll of the right camera must be determined by other means. For absolute position of the right camera, we use Google Earth, but readings from a GPS unit would have worked just as well. For yaw, we use the image of the sun, but we could have just as easily determined the camera's yaw with a compass, correcting for magnetic declination. Note that any errors in the absolute position and yaw only add an overall displacement and azimuth to the reconstructions; they do not affect the stereographically reconstructed heights, distances, or velocities. Errors in absolute pitch and absolute roll, however, will affect reconstructed heights, distances, and velocities, so they must be measured accurately.

2) THE HORIZON CONSTRAINT

To accurately calibrate the right camera's absolute pitch and absolute roll, we use the ocean horizon. Consider a camera located at world coordinates $(x, y, z) = (0, 0, h)$ using the standard meteorological convention of x increasing to the east, y increasing to the north, and z increasing with altitude. To use the horizon for calibration, we must first determine where the horizon is located in world coordinates. Let us denote the horizontal displacement of the horizon from the camera by d_H and the vertical displacement of the horizon from the camera's sea level by \tilde{h} , as depicted in Fig. 5. The horizon is the set of world coordinates where lines emanating from the camera touch the earth's surface as a tangent. Using the two triangles in Fig. 5, we obtain the similarity equation $(h + \tilde{h})/d_H = d_H/(R_e - \tilde{h})$, where R_e is the radius of the earth. This implies that

$$d_H^2 = (R_e - \tilde{h})(h + \tilde{h}). \quad (9)$$

Using the Pythagorean theorem, we also have

$$R_e^2 = d_H^2 + (R_e - \tilde{h})^2. \quad (10)$$

Substituting Eq. (9) for d_H^2 in Eq. (10) results in

$$\tilde{h}(R_e + h) = R_e h. \quad (11)$$

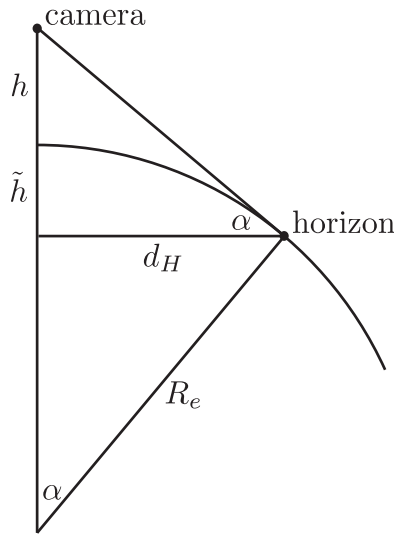


FIG. 5. For a camera at height h above sea level, the horizon is located a horizontal distance d_H away from the camera and a distance \tilde{h} below the camera's sea level. Using the trigonometry described in the text, it can be shown that $h \approx \tilde{h}$ and $d_H \approx \sqrt{2R_e \tilde{h}}$, where R_e is the radius of the earth.

Approximating $R_e + h \approx R_e$ for $h \ll R_e$, $\tilde{h} \approx h$ can be derived. In other words, the depth of the horizon equals the height of the camera. Note that $h - \tilde{h}$ becomes only 16 cm even when h is 1 km, so $\tilde{h} \approx h$ is a safe approximation for most practical cases. Then, the distance of the camera to the horizon can be calculated from Eq. (9) as

$$d_H = \sqrt{2R_e h}. \tag{12}$$

Having found expressions for \tilde{h} and d_H , we know that the horizon for the right camera at (x_r, y_r, h_r) can be described by the set of 3D points

$$\begin{aligned} [x_{H_r}, y_{H_r}, z_{H_r} | (x_{H_r} - x_r)^2 + (y_{H_r} - y_r)^2 \\ = 2R_e h_r, z_{H_r} = -h_r]. \end{aligned} \tag{13}$$

Figure 6 shows the horizon in the camera image plane for heights of $h = 1, 10,$ and 100 m with zero roll angle. As an observer moves up off the surface, the curvature of the earth increasingly manifests itself as a bending of the horizon (until, in outer space, the horizon is clearly a circle). As seen in Fig. 6, however, the projection of the horizon (i.e., its outline in the photograph) deviates from a straight line by less than one pixel for camera heights up to 100 m. Hence, we can assume that the projection of the horizon is a line in the right camera's image plane, which can be described by the set of pixels

$$(x'_{H_r}, y'_{H_r} | y'_{H_r} = m_r x'_{H_r} + a_r), \tag{14}$$

for some constants m_r and a_r . The values of m_r and a_r can be obtained easily by regressing a line on the image of the horizon. We know, therefore, that the projection matrix \mathbf{P}_r must satisfy

$$(\mathbf{P}_r \mathbf{X}_{H_r})_2 = m_r (\mathbf{P}_r \mathbf{X}_{H_r})_1 + a_r,$$

where the \mathbf{X}_{H_i} denotes elements of Eq. (13) and the subscripts 1 and 2 denote the first and second components, respectively, of the image vector. This equation implicitly defines the right camera's pitch and roll.

3) CALIBRATION AND RECONSTRUCTION

To calibrate the cameras, we can use the fundamental matrix to identify the intrinsic camera parameters up to a scale and relative camera pose, and we can use the visible horizon to identify absolute pitch and roll and the scale. Let us define a cost function C as

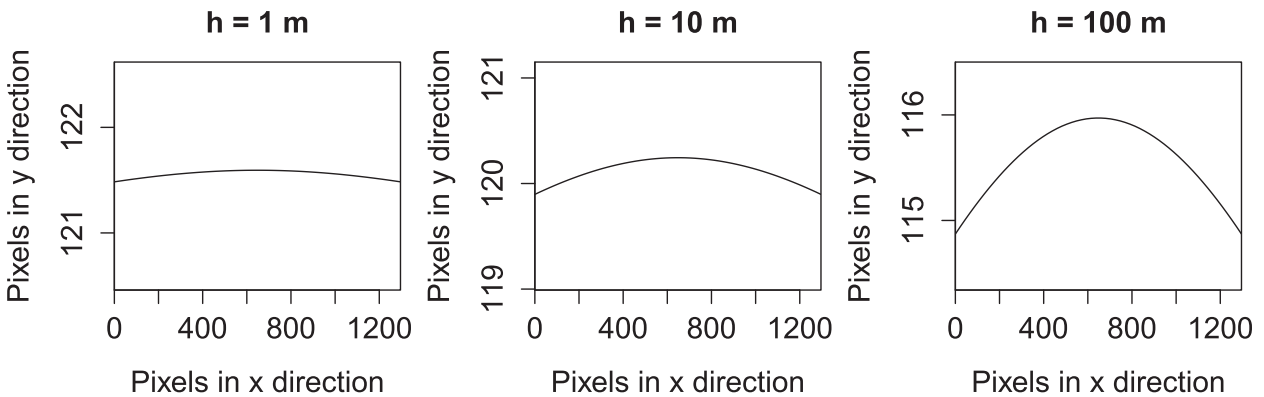


FIG. 6. (left to right) The location of the horizon in the camera's image plane for $h = 1, 10,$ and 100 m. Note that, to within one pixel, the projection of the horizon forms a straight line across the image.

$$C = \frac{1}{N_1} \sum_{i=1}^{N_1} \frac{|\mathbf{X}'_i{}^T \mathbf{F} \mathbf{X}'_i|^2}{(\mathbf{F} \mathbf{X}'_i)_1^2 + (\mathbf{F} \mathbf{X}'_i)_2^2} + \frac{1}{N_2} \sum_{i=1}^{N_2} |m_r (\mathbf{P}_r \mathbf{X}_{H_i})_1 + a_r - (\mathbf{P}_r \mathbf{X}_{H_i})_2|^2. \quad (15)$$

This cost function jointly penalizes the deviations from the epipolar constraint and the horizon constraint, and it can be expressed solely in terms of the left and right camera projection matrices. The first term can be understood as follows. Writing the components of $\mathbf{F} \mathbf{X}'_i$ as $(a, b, c)^T$, the epipolar line in the left image plane is defined by $ax'_i + by'_i + c = 0$. The vector (a, b) points perpendicular to this epipolar line. Therefore, $|d_e|$ is the minimum distance of (x'_i, y'_i) from the epipolar line if adding

$$\frac{d_e}{\sqrt{a^2 + b^2}}(a, b)$$

to (x'_i, y'_i) gives a coordinate that lies on the epipolar line. Therefore, d_e is defined by

$$a \left(x'_i + d_e \frac{a}{\sqrt{a^2 + b^2}} \right) + b \left(y'_i + d_e \frac{b}{\sqrt{a^2 + b^2}} \right) + c = 0,$$

which can be rearranged to give

$$d_e = -\frac{ax'_i + by'_i + c}{\sqrt{a^2 + b^2}} = -\frac{\mathbf{X}'_i{}^T \mathbf{F} \mathbf{X}'_i}{\sqrt{(\mathbf{F} \mathbf{X}'_i)_1^2 + (\mathbf{F} \mathbf{X}'_i)_2^2}}.$$

The first term of Eq. (15) is, therefore, the sum of squared distances of images from their epipolar lines. The objects used in this sum can be any set of objects in the field of view of both cameras: clouds, sea gulls, ships, buoys, etc. In the second term of (15), the sum is over a set of world coordinates chosen from Eq. (13). The expression $m_r (\mathbf{P}_r \mathbf{X}_{H_i})_1 + a_r$ is the actual y'_i position of the horizon at $x'_i = (\mathbf{P}_r \mathbf{X}_{H_i})_1$, while $(\mathbf{P}_r \mathbf{X}_{H_i})_2$ is the expected position of the horizon at that x'_i given the projection matrix \mathbf{P}_r . Therefore, the second term in the cost function is the sum of squared distances (in the y' direction) of the actual horizon from the projected horizon. Levenberg–Marquardt optimization (Press et al. 2007) is used to minimize the cost function and, thereby, solve for the unknown parameters of the projection matrices.

Focal lengths, principal point coordinate, pitch, roll, and relative yaw are used as the free parameters in this optimization. The initial values for intrinsic parameters are obtained from the lens distortion correction algorithm and the initial pitch and roll angles are estimated

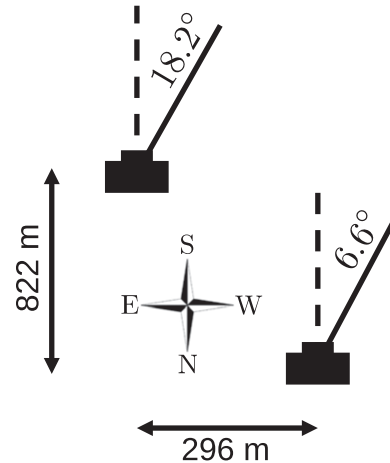


FIG. 7. Relative camera location and yaw (not drawn to scale).

from the horizon image. Once the projection matrices are constructed, 3D reconstruction is performed with the direct linear transformation (DLT) method (Hartley and Zisserman 2003). DLT looks for the solution \mathbf{X} , which is in the null-space of $\mathbf{A} \mathbf{X} = \mathbf{0}$, where

$$\mathbf{A} = \begin{pmatrix} x'_r \mathbf{P}_r^{(3)} - \mathbf{P}_r^{(1)} \\ y'_r \mathbf{P}_r^{(3)} - \mathbf{P}_r^{(2)} \\ x'_l \mathbf{P}_l^{(3)} - \mathbf{P}_l^{(1)} \\ y'_l \mathbf{P}_l^{(3)} - \mathbf{P}_l^{(2)} \end{pmatrix} \quad (16)$$

and $\mathbf{P}^{(i)}$ refers to the i th row of the projection matrix.

4. Estimating the reconstruction precision

In this study, the baseline between the two cameras is 296 m in the east–west direction and 822 m in the north–south direction. The left camera is pointed 18.19° west of south, and the right camera is pointed 6.56° west of south. This configuration is depicted in Fig. 7. The yaw angles given here are obtained from the sun’s position, but the relative yaw angle between the cameras is obtained by the epipolar constraint. The fields of view of the left and right cameras are 76° and 67° , respectively. By coincidence, the leftmost viewing angle of the left camera coincides nearly perfectly with the baseline.

This configuration poses some unique challenges. Consider a line of sight from one of the cameras that is collinear with the baseline. In this case, the two lines of sight are collinear. As a result, there is no information that can be gathered about the depth of any object on this line: all positions along on this line will project onto the same pair of image-plane pixels. Therefore, we see that the reconstruction must fail for objects near the baseline. In most applications of stereophotogrammetry,

the baseline is not within the field of view of both cameras, so this is not an issue. In our configuration, however, we must be careful to quantify the reconstruction error near the baseline.

To get a feeling for the magnitude and spatial distribution of this reconstruction error, let us derive the error analytically for the case of two identical cameras in 2D with the same relative displacement and relative yaw as the Miami cameras. To this end, let us derive the disparity $d \equiv x'_r - x'_l$ as a function of an object's position relative to the left camera. Let x_l and z_l denote the coordinates of the object in the left camera's reference frame (i.e., with the origin at the left camera, the positive z_l axis pointing in the direction of the left camera's line of sight, and with x_l increasing to the right of the left camera). Similarly, let x_r and z_r denote the object's coordinates in the right camera's reference frame. Let $(x_l, z_l) = (x_b, z_b)$ be the location of the right camera in the left camera's reference frame; here, the subscript b refers to the baseline. Likewise, $(x_r, z_r) = (-x_b, -z_b)$ is the location of the left camera in the right camera's reference frame. In addition, let the right camera's image plane be tilted counterclockwise at an angle θ relative to the left camera's image plane.

Figure 8 depicts this geometry, with the perspectives from the two cameras overlaid on a common image plane. The horizontal line at the bottom of the figure is the image plane, and the two diagonal lines are the two lines of sight, which intersect each other at the pinhole. In the left camera's reference frame, the object is located at (x_l, z_l) , which is at the end of the line of sight that slopes up and to the right. In the right camera's reference frame, the object is located at (x_r, z_r) , which is at the end of the line of sight that slopes up and to the left. The dashed lines aid the eye in seeing the trigonometric relationships to be used in the derivation below. As depicted in Fig. 8, $x_b > 0$, $z_b < 0$, and $x_r < 0$.

By trigonometric similarity,

$$z_l/L = f/d \tag{17}$$

The distance L is given by

$$L = \Delta x - (-\Delta z x_r/z_r), \tag{18}$$

where $\Delta x = x_l - x_r$, $\Delta z = z_r - z_l$, and x_r and z_r are given by

$$\begin{pmatrix} x_r \\ z_r \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_l - x_b \\ z_l - z_b \end{pmatrix}. \tag{19}$$

Combining these into an expression for the disparity as a function of x_l and z_l , we obtain

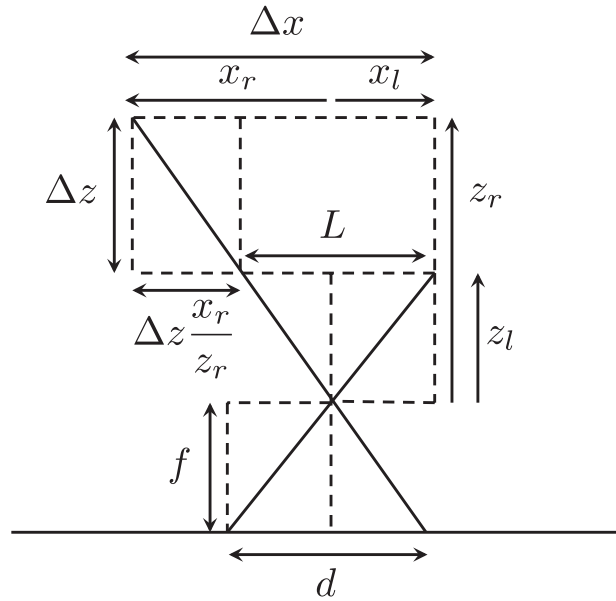


FIG. 8. This sketch, for the derivation of reconstruction error in 2D, shows the perspective of the two cameras overlaid on a common image plane and pinhole. See the text for more details.

$$d(x_l, z_l) = f[(x_l - x_r) - (z_r - z_l)x_r/z_r]/z_l, \tag{20}$$

where x_r and z_r are defined in terms of x_l and z_l by Eq. (19).

For simplicity, assume that there is no error in the measurement of the image location in the left camera's image plane. This measurement fixes the ratio x_l/z_l . Now, imagine that the measurement of the image in the right camera is uncertain. This introduces uncertainty into the disparity d , which generates an uncertainty in the values of x_l and z_l . Let us assume that the error in the right camera's image plane is on the order of one pixel. We would like to know how much fractional error that generates in x_l and z_l (the fractional error will be the same for the two, since we assume their ratio is known exactly).

To get a sense of the potential error in the reconstruction from the Miami cameras, we will assume that each of the two cameras has the average properties of the two Miami cameras. In particular, we use $kf = 817$ in place of f in Eq. (20) to give d in pixels ($kf = 817$ is the average of 651 for the left camera and 983 for the right camera). As stated above, the relative position and relative yaw of the two cameras is chosen to be the same as for the Miami cameras. Figure 9 plots

$$1 \left/ \frac{\partial}{\partial \alpha} d(\alpha x_l, \alpha x_r) \right|_{\alpha=1}, \tag{21}$$

which gives the fractional error in x_l and z_l per pixel of error in the position of the object in the right camera's

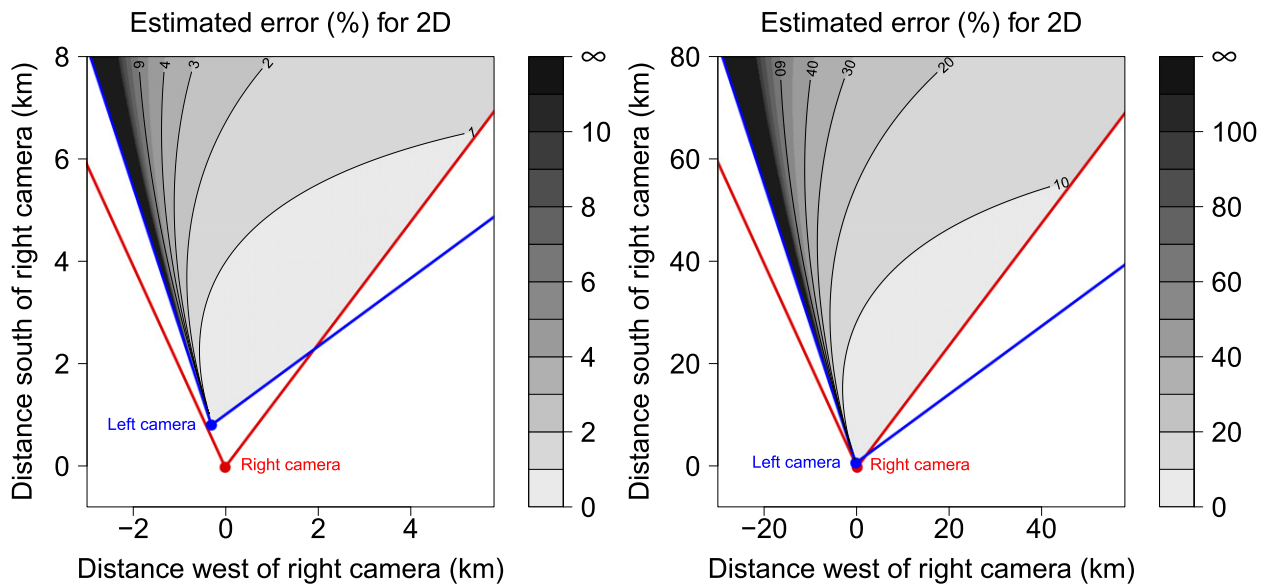


FIG. 9. For identical cameras in 2D, the theoretical error (%) of the reconstructed depth due to one pixel of error in the measurement of the image position in the right camera's image plane. The error is plotted for ranges of (left) 8 and (right) 80 km. The red (blue) circles and red (blue) lines denote the right (left) camera and its field of view. Note that the error blows up near the baseline, which happens to coincide with the leftmost field of view. Nevertheless, the reconstruction error is $<2\%$ over the majority of the field of view out to 8 km, and $<20\%$ out to 80 km.

image plane. As seen in Fig. 9, the error in the reconstructed depth of the object goes to infinity at the baseline, which just happens to coincide with the leftmost view from the left camera. Despite this fact, the vast majority of the field of view out to 8 km has an estimated reconstruction error of less than 2% (see the left panel), and the vast majority of the field of view out to 80 km has an estimated reconstruction error of less than 20% (see the right panel).

In 3D, the behavior of the reconstruction error is more nuanced. For cameras at sea level, the lines of sight to an object can be collinear with the baseline only if the object is at sea level. With the Miami cameras, which are located near sea level, the lines of sight to clouds are always pitched upward. As a result, the lines of sight to clouds are never parallel to the baseline, even though the baseline is in the field of view. This motivates a hypothesis that the reconstruction error diminishes near the leftmost field of view as the cloud height increases. To check this, we can calculate maps of estimated reconstruction error using the full 3D projection matrices for the Miami cameras. These error maps are shown in Fig. 10 for hypothetical objects at altitudes of 2, 6, and 12 km. The estimated error in these maps is the fractional change (in percent) of the reconstructed depth due to a one-pixel error in the disparity. Since the image planes are two dimensional, the plotted error is the root-sum-square of the errors generated by a one-pixel error in the x' and y' directions. The white regions are

those areas that are not visible by the two cameras; note that the object's altitude affects where it enters the field of view. With regard to the hypothesis, we see that the error in the leftmost field of view decreases as the height of the reconstructed object increases, as predicted.

5. Validation against lidar

The preceding analysis gives some sense for the precision of the stereo reconstructions. To assess accuracy, however, the reconstructions must be compared against meteorological data from independent observations. In this section, reconstructed cloud bases are compared against data from lidar. In the next section, the reconstructed horizontal motions of clouds are compared against wind profiles from radiosondes.

For the validation of reconstructed heights, cloud bases are calculated from three pairs of still images taken at 2025 UTC 8 April 2013, 1344 UTC 4 April 2013, and 1120 UTC 9 April 2013. At these times, we observed stratocumulus (Sc), altocumulus (Ac), and cirrocumulus (Cc) at heights of 2, 6, and 12 km, respectively. For each pair of images, authors Öktem and Romps manually identified pairs of matching cloud-base feature points in the two images. A mouse-controlled graphical user interface (GUI) was designed for manual identification of feature points and automatic reconstruction. Once a point is selected in one image, the GUI displays the corresponding epipolar line in the other image to serve

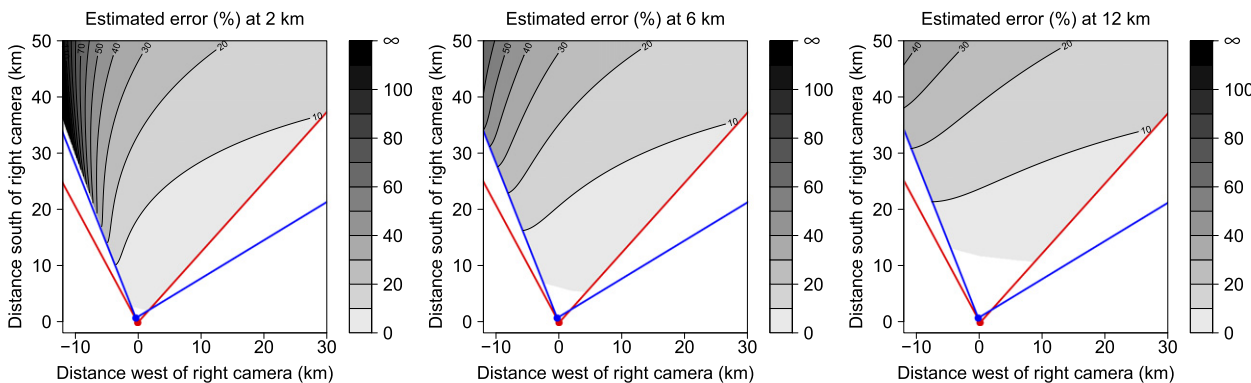


FIG. 10. For the Miami cameras in 3D, the theoretical error (%) of the reconstructed depth due to one pixel of error (in both x' and y') in the measurement of the image position in the right camera's image plane. The error is plotted for hypothetical objects at altitudes of (left) 2, (middle) 6, and (right) 12 km. The red (blue) circles and red (blue) lines denote the right (left) camera and its field of view. Note that the error blows up near the baseline, which happens to coincide with the leftmost field of view. White areas denote regions that are not within the field of view of both cameras at that altitude. Note that, as predicted, the reconstruction error at the leftmost field of view decreases with increasing height of the reconstructed object.

as a guide for identifying the matching point. For each pair of image points (i.e., \mathbf{X}'_l and \mathbf{X}'_r), the method of direct linear transformation is used to calculate the corresponding world coordinate (i.e., \mathbf{X}). For each of the three cases, the two histograms of reconstructed heights obtained from the data of Öktem and Romps were compared and found to be very similar, suggesting that any human error was either negligible or consistent. The analysis below is performed on the union of these two datasets for each case, resulting in 440, 416, and 307 stereo image pairs for 8, 4, and 9 April, respectively.

To validate that the stereo cameras are giving accurate cloud-base altitudes, the histogram of stereophotogrammetric heights at the one instant in time are compared against lidar data collected over a 20-min period from 10 min prior to 10 min after the photographs were taken. For the stratocumulus and altocumulus, cloud-base heights were obtained with a Vaisala CL31 ceilometer that is collocated with the left camera as part of the CAROb instrumentation. This ceilometer reports cloud-base heights up to 7.5 km at 10-m resolution every 15 s. For the cirrocumulus, cloud-base heights were obtained with a Sigma MPL-4B-IDS-532 micropulse lidar also collocated with the left camera, which reports backscattered intensity up to 15 km at 30-m

resolution every 15 s. The micropulse lidar samples only zenith view with a 1.2-m footprint at 12 km, associated with 100- μ rad pulse width. From visual inspection, the clouds in the lidar data were clearly above 11 km, so twice the maximum logarithm of backscattered intensity between 9 and 11 km was used as the threshold to identify cloud bases. Since the thickness of the cirrus layer is about 1 km, practically any method for identifying cloud bases gives the same answer; in fact, using all cloudy points—not just cloud bases—produces a very similar distribution of height. Table 2 lists the mean and standard deviation for the observed cloud bases, obtained from both the stereo cameras and the lidars.

On 8 April 2013, when the view was dominated by stratocumulus, 440 pairs of feature points were sampled from the pair of photographs taken at 2025 UTC. The top panels of Fig. 11 show the locations of the feature points in the two images. The reconstructed heights included three outliers: one data point at 828 m and two data points at an altitude of around 8 km. Those points were sampled from distant cumulus and cirrus (at depths from the cameras exceeding 20 km), which are not captured by the ceilometer at that time. Those three samples are discarded from the results in Table 2 and Fig. 11.

TABLE 2. Cloud-base statistics from the stereo cameras and ceilometers (Vaisala CL31 for Sc and Ac, and Sigma MPL for Cc).

Date	Time (UTC)	Cloud type	Distance from right camera (km)		Avg height (m) stereo camera	Avg height (m) ceilometer	Std height (m) stereo camera	Std height (m) ceilometer
			Min	Max				
8 Apr	2025	Sc	2.2	13	1804	1805	56	42
4 Apr	1344	Ac	5.6	34	5904	5913	174	173
9 Apr	1120	Cc	11	39	11 895	11 512	690	159

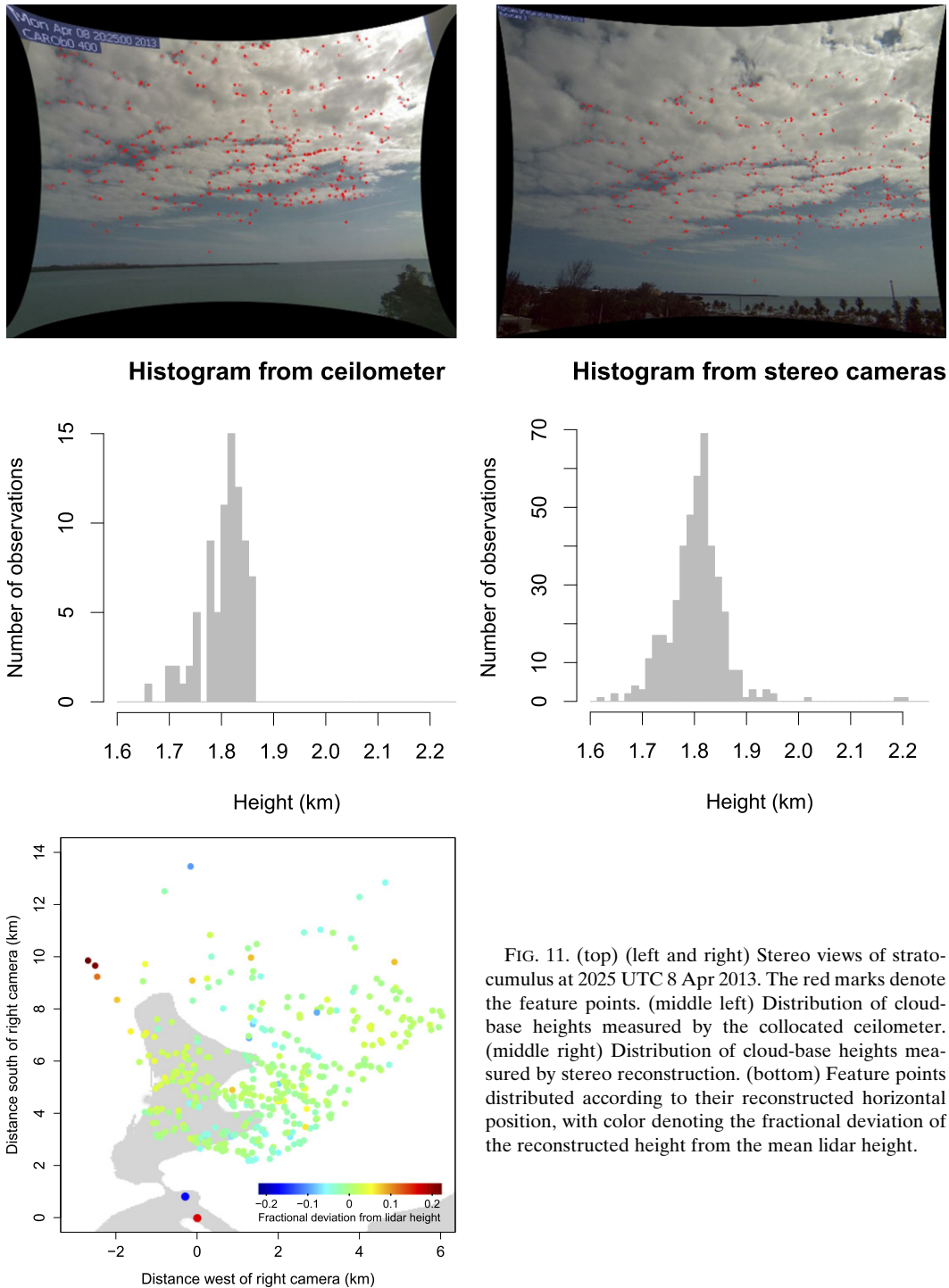


FIG. 11. (top) (left and right) Stereo views of strato-cumulus at 2025 UTC 8 Apr 2013. The red marks denote the feature points. (middle left) Distribution of cloud-base heights measured by the collocated ceilometer. (middle right) Distribution of cloud-base heights measured by stereo reconstruction. (bottom) Feature points distributed according to their reconstructed horizontal position, with color denoting the fractional deviation of the reconstructed height from the mean lidar height.

The samples of the stratocumulus layer were located 2.2–13 km south of the right camera. As seen in the middle row of Fig. 11, the histogram of stereo-camera cloud-base heights closely resembles the histogram of ceilometer cloud-base heights. The means of the two distributions are 1804 and 1805 m, respectively. The

standard deviations are 56 and 42 m, respectively. The bottom panel of Fig. 11 shows the spatial distribution of reconstructed heights: the color of each point gives the fractional deviation of that point's height relative to the mean of the ceilometer heights. These deviations have no discernable spatial pattern other than the clump

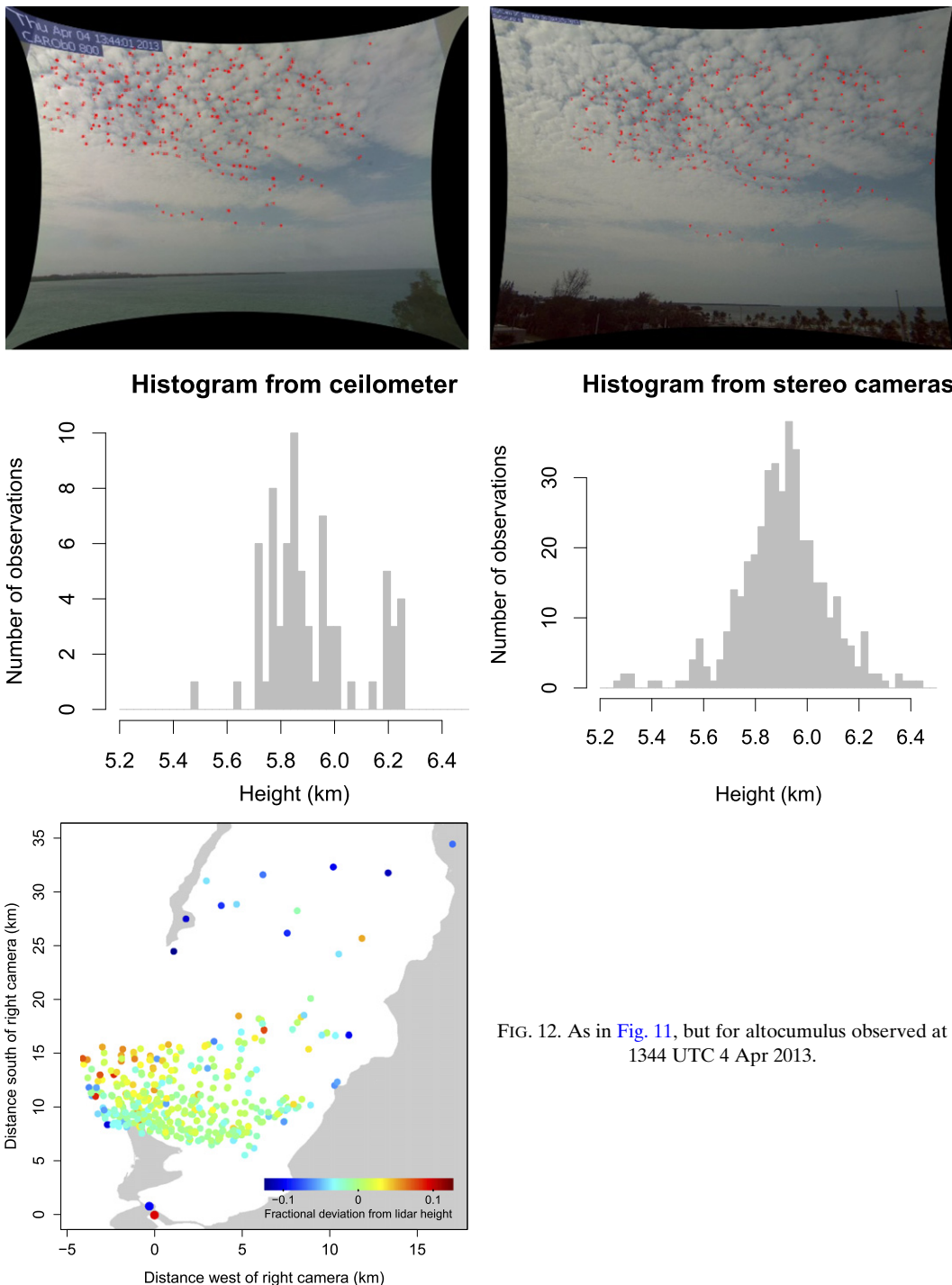


FIG. 12. As in Fig. 11, but for altocumulus observed at 1344 UTC 4 Apr 2013.

of three high values near the leftmost field of view, where we expect the reconstruction to be the poorest.

For 4 April 2013, when the view was dominated by altocumulus, 416 pairs of feature points were sampled from the pair of photographs taken at 1344 UTC. The top panels of Fig. 12 display the feature points for this pair of images. The altocumulus feature points range in

distance from 5.6 to 34 km south of the right camera. As seen in the middle row of Fig. 12, the histogram of stereo-camera cloud-base heights closely resembles the histogram of ceilometer cloud-base heights. As is true for all three cases, the number of points collected from the instantaneous pair of stereophotographs is significantly larger than the number of data points collected by

lidar during the surrounding 20-min interval. This leads to a superior sampling of the distribution of cloud-base heights by the stereo cameras. Nevertheless, the means and standard deviations of the two distributions are very similar: the stereo cameras and the ceilometer report means of 5904 and 5913 m, and the standard deviations of 174 and 173 m, respectively. The bottom panel of Fig. 12 suggests that the variance of reconstructed heights increases somewhat up and to the left, as anticipated by Fig. 10. It also appears that the altocumulus at distances greater than 20 km south of the cameras is lower in altitude than the altocumulus closer to the cameras.

For 9 April 2013, when the view was dominated by cirrocumulus, 307 pairs of feature points were sampled from the pair of photographs taken at 1120 UTC. The images used for this analysis, as well as the locations of feature points, are shown in Fig. 13. The cirrocumulus feature points range in distance from 11 to 39 km south of the right camera. Unlike the previous two cases, the histograms from the lidar and the stereo cameras have some noticeable differences. As seen in the middle row of Fig. 13, the histogram of stereo-camera cloud-base heights is broader, extending to larger altitudes. At the low end, the 10th percentile of the lidar and stereo-camera distributions is quite similar: 11.3 and 11.1 km, respectively. The 90% quantiles, on the other hand, are quite different: 11.7 km for the lidar and 12.7 km for the stereo cameras. The bottom panel of Fig. 13 suggests a reason for this difference: the cirrocumulus appears to slope upward in the southward direction. Since there is no indication from the previous two cases of a positive height bias for distant clouds, it is quite plausible, although difficult to verify by other means, that the cirrocumulus was tilted in this way.

In stereo reconstruction, there are two main sources of error: image-identification error and parametric uncertainty. When identifying the pixel location of images in the image plane (either manually or algorithmically), some amount of error is to be expected, typically on the order of one pixel. These errors generate a disparity error, which produces a reconstruction error proportional to the distance from the camera, as depicted in Fig. 10. A one-pixel inaccuracy in image-point locations results in 15, 133, and 565 m of error in the mean of the measured cloud-base heights of 8, 4, and 9 April, respectively; that is, less than 1%, 3%, and 5% of the lidar-measured cloud-base averages for each of the three cases, respectively. With regard to parametric uncertainty, Table 3 lists the uncertainties in the relative camera positions ($x_{0,r} - x_{0,l}$ and $y_{0,r} - y_{0,l}$), camera heights ($z_{0,r}$ and $z_{0,l}$), horizon line-fitting parameters (a_r and m_r), and lens distortion correction parameters (k_1 , k_2 , and k_3), and the corresponding errors that these uncertainties would generate in the

mean reconstructed cloud-base heights. Uncertainty in the absolute yaw angle, which is estimated from the sun, does not affect the height reconstruction. The right camera is mounted on the outer rails of a cylindrical tower, and the left camera is mounted on the roof of a building. Both of these sites can be clearly identified in the aerial and satellite photographs provided by Google Earth. The camera elevations are measured with Google Earth's 3D view enabled, whereas the x and y positions are measured with 3D view disabled. The relative positions ($x_{0,r} - x_{0,l}$ and $y_{0,r} - y_{0,l}$) vary by no more than 3 m when accounting for the range of possible camera locations in the Google Earth images. Similarly, an uncertainty of 0.5 m is estimated for the heights ($z_{0,r}$ and $z_{0,l}$). The horizon parameters a_r and m_r are calculated by identifying two points close to the right and left ends of the visible ocean horizon in the right camera view, and by deriving the equation of the line passing through these two points. A one-pixel uncertainty in the location of these two points leads to a one-pixel uncertainty in a_r and a 0.002 uncertainty in the slope. To estimate the error bound in lens distortion correction parameters, we re-executed the distortion-correction algorithm several times by omitting one of the calibration pattern images from the training set in each execution.

Table 3 lists these uncertainties and the typical magnitudes of the errors that they generate for reconstructed cloud-base heights on 8 April (Sc), 4 April (Ac), and 9 April (Cc). Note that the disparity error (i.e., the image-identification error) is a random error, whereas the uncertainty in the parameters leads to a bias. When averaging over many reconstructed cloud-base heights, the disparity error has a negligible effect on the mean. The total error in this table is the square root of the sum of all of the squared errors above (including the disparity error), and so it should be thought of as the expected error for the reconstructed height of any one feature; the expected error for the mean of the cloud-base heights is smaller (i.e., it is obtained by excluding the disparity error). Note from Table 3 that these uncertainties have a negligible impact on the height reconstructions for low-altitude clouds: the total error for the height of an individual feature on a shallow cumulus is only about 30 m (less than 2% of the height). As the cloud altitudes increase, the errors increase due to the larger distances to the clouds being measured: the total error for the height of a feature on a cirrocumulus cloud is about 900 m (8% of the height).

6. Validation against radiosondes

To compare velocities against radiosonde measurements, the horizontal speeds of stratiform clouds in the



Histogram from lidar

Histogram from stereo cameras

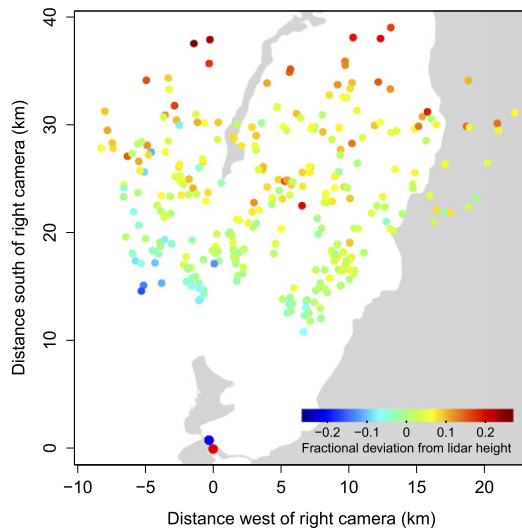
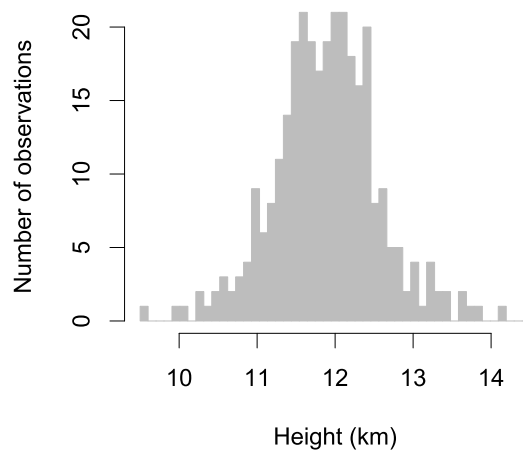
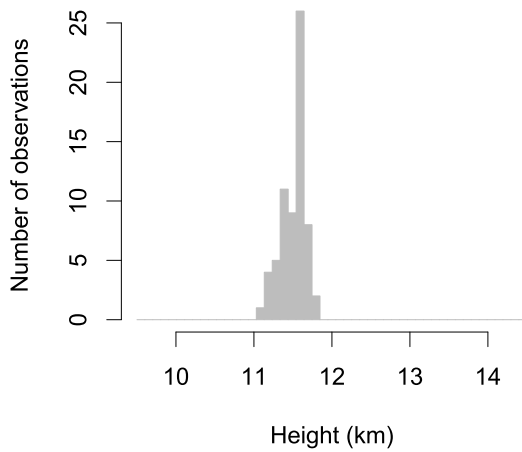


FIG. 13. As in Fig. 11, but for cirrocumulus at 1120 UTC 9 Apr 2013.

east and north directions (u and v , respectively) are calculated as

$$u = \frac{1}{N} \sum_{i=1}^N \frac{x_i(t + \delta t) - x_i(t)}{\delta t} \quad (22)$$

$$v = \frac{1}{N} \sum_{i=1}^N \frac{y_i(t + \delta t) - y_i(t)}{\delta t}, \quad (23)$$

where $[x_i(t), y_i(t)]$ is the horizontal position of feature point i at time t . For each of the four cloud layers studied

TABLE 3. The impacts of parametric uncertainties (first seven data rows) and image-identification error (disparity) on the reconstructed cloud-base heights. See Table 1 for the definition of the listed parameters. The total error for the reconstructed height of any one feature is given by the root-sum-square of all of the errors, and is rounded to the first significant digit. The relative error in the last row is obtained by dividing by the mean heights from Table 2 and rounding to the first significant digit.

Parameter	Uncertainty	Absolute error (m)		
		Apr 8	Apr 4	Apr 9
$x_{0,r} - x_{0,l}$	3 m	5	46	215
$y_{0,r} - y_{0,l}$	3 m	9	58	219
$z_{0,r}$	0.5 m	7	53	226
$z_{0,l}$	0.5 m	6	47	196
a_r	1 pixel	11	84	345
m_r	0.002	12	64	251
k_1, k_2, k_3	0.001, 0.005, 0.001	9	48	252
Disparity	1 pixel	15	133	565
Total error (m)		30	200	900
Total error/mean height (%)		2	3	8

(one each on 4 and 8 April, and two on 9 April), N independent cloud features were identified by Öktem from several frames over a given time interval, and those same features were identified in images δt later. The number of features N used for each case is listed in Table 4. For each cloud layer, a fixed δt was used.

The interval δt was chosen subjectively for the following reasons. Recall from section 4 that the error in the reconstructed positions increases with distance between the cameras and the observed feature. When δt is small and the distance to the feature is large, the position-estimation error can exceed the distance the cloud feature travels in δt and results in low accuracy in velocity estimation. On the other hand, if δt is too large, then the cloud feature may exit the field of view, be obscured by another cloud, or evaporate in this time period. Hence, it was not possible to define a general rule for setting δt . A low value of $\delta t = 30$ s was used for the low-level clouds of 8 and 9 April, whose feature points were in the range 2–5 km from the cameras. For the high-altitude clouds of 4 and 9 April, whose feature points were at distances exceeding 8 km from the cameras, a longer time interval of $\delta t = 300$ s was used.

The results are listed in Table 4 for the altocumulus layer of 4 April, the stratocumulus layer of 8 April, and the cirrocumulus and shallow cumulus layers of 9 April. For each of the four cloud layers, this table reports the mean of the height, u , and v of the sampled feature points, as well as the standard deviations of those distributions. These data are plotted in Fig. 14 along with radiosonde measurements from the nearby Miami airport (obtained from the website of the Department of Atmospheric Science at the University of Wyoming, <http://weather.uwyo.edu/upperair/sounding.html>). The location of the radiosonde releases is about 10 km inland from the cameras. The radiosondes are launched twice daily, at 1105 UTC (local morning) and 2305 UTC (local evening). Figure 14 displays the morning soundings for 4 and 9 April, and the evening sounding for 8 April. It is not always possible to observe clouds contemporaneously with radiosonde releases, so we picked the times as close to the radiosonde release times as we could. These radiosonde releases are about 2.5 h before the stereo data on 4 April, 2.5 h after the stereo data on 8 April, and nearly contemporaneous with the stereo data on 9 April. In spite of the fact that the radiosonde data are collected at a somewhat different time and location than the stereo data, the agreement between the reconstructed winds and the radiosonde winds is quite good.

The error bars in Fig. 14 show the standard deviations in the measurements; that is, they correspond to the two rightmost columns of Table 4. Table 5 lists the expected errors in the horizontal velocities caused by parametric uncertainties and image-identification errors. From the last row of this table, we see that the relative errors are significantly larger for the v speeds than for the u speeds. This can be understood as follows. Since the cameras face approximately south, the v speeds correspond to motion roughly parallel with the lines of sight, whereas the u speeds correspond to motion roughly perpendicular to the lines of sight. With a single camera, we can readily detect motions that are transverse to the line of sight, but we cannot detect any motion of a point moving along the line of sight. With a second camera, the motion of a point moving along the first camera's line of sight can be detected, but the displacement in

TABLE 4. Horizontal velocities of cloud bases measured by stereo reconstruction.

Date	Cloud type	Time interval (UTC)	δt (s)	N	Avg height (m)	Avg u (m s^{-1})	Avg v (m s^{-1})	Std height	Std u	Std v
Apr 4	Ac	1334–1353	300	20	5916	21.9	-2.8	135	0.8	1.1
Apr 8	Sc	2015–2036	30	30	1817	2.6	-5.6	32	0.9	1.3
Apr 9	Cu	1111–1134	30	20	741	-5.9	3.5	54	0.7	2.9
Apr 9	Cc	1110–1124	300	20	11344	35.0	-1.7	522	2.1	2.4

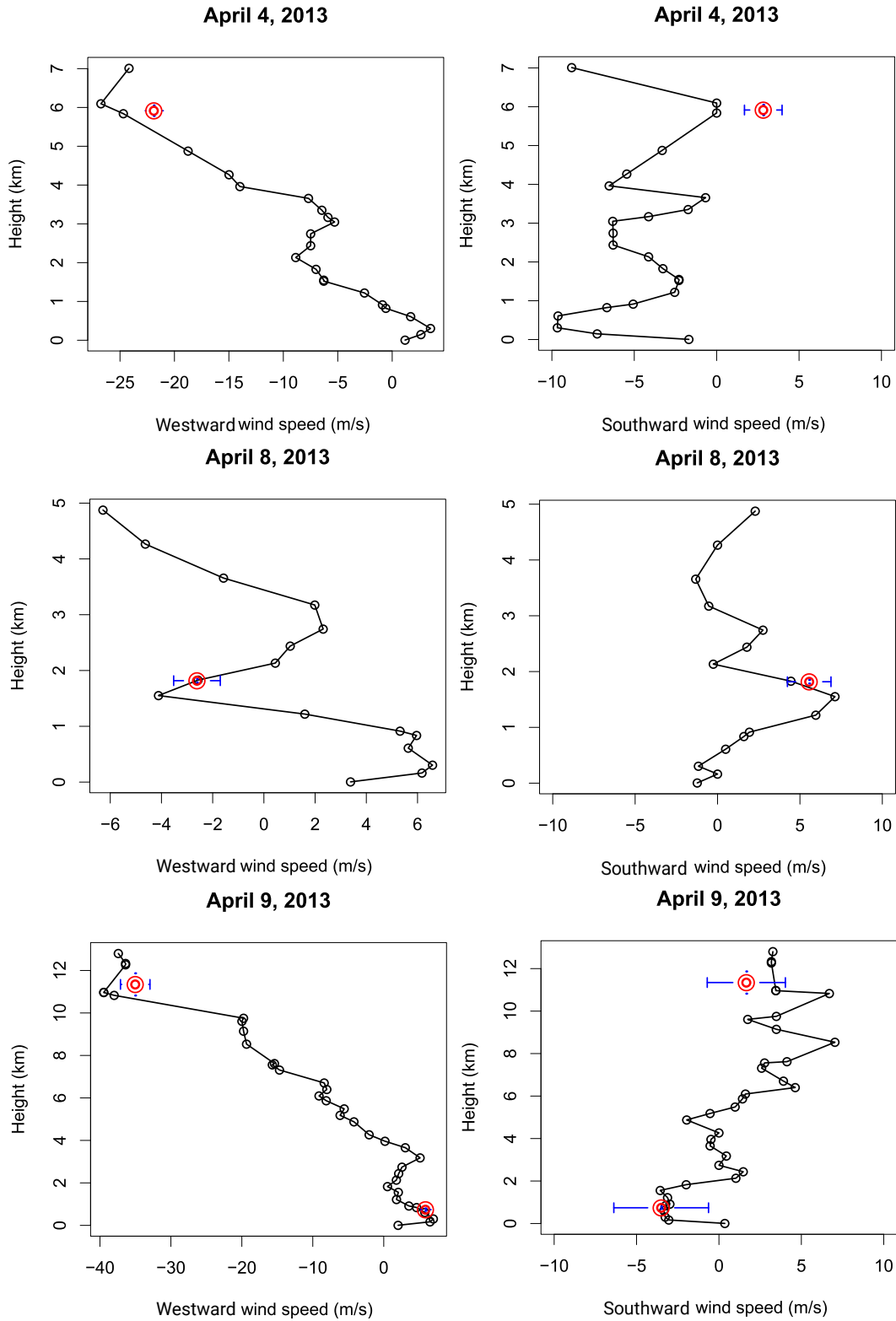


FIG. 14. Reconstructed horizontal velocities from Table 4 (red circles) compared with radiosonde wind profiles from the nearby Miami airport (black circles). The blue error bars denote the standard deviations of height and velocity among the sampled feature points.

TABLE 5. As in Table 3, but for horizontal velocities. The relative error in the last row is obtained by dividing by the mean speeds from Table 4 and rounding to the first significant digit.

Parameter	Uncertainty	Absolute error (m s^{-1})							
		Apr 4 Ac		Apr 8 Sc		Apr 9 Cu		Apr 9 Cc	
		u	v	u	v	u	v	u	v
$x_{0,r} - x_{0,l}$	3 m	0.13	0.19	0.01	0.02	0.01	0.02	0.42	0.59
$y_{0,r} - y_{0,l}$	3 m	0.17	0.09	0.01	0.04	0.02	0.02	0.51	0.24
$z_{0,r}$	0.5 m	0.19	0.04	0.01	0.02	0.01	0.02	0.58	0.07
$z_{0,l}$	0.5 m	0.16	0.03	0.01	0.03	0.01	0.03	0.53	0.07
a_r	1 pixel	0.26	0.04	0.01	0.05	0.02	0.03	0.87	0.11
m_r	0.002	0.21	0.04	0.01	0.04	0.02	0.04	0.67	0.10
k_1, k_2, k_3	0.001, 0.005, 0.001	0.25	0.31	0.11	0.18	0.02	0.27	0.75	0.95
Disparity	1 pixel	0.32	0.87	0.32	1.10	0.11	0.42	1.42	3.01
Total error (m s^{-1})		0.6	1.0	0.3	1.0	0.1	0.5	2.0	3.0
Total error/mean speed (%)		3	30	10	20	0	10	6	200

the image plane is small compared to the displacement of a point moving with the same speed transverse to the first camera's line of sight. This reduces the signal-to-noise ratio when calculating speeds from displacements in the image planes. Therefore, the speeds of objects moving in the north–south direction are harder to measure accurately in our setup than the speeds of objects moving in the east–west direction.

Comparing the last two columns of Table 4 to the second-to-last row of Table 5, we see that the standard deviations of the measured speeds for the 9 April Cu are significantly larger than the total expected errors for that case. The video sequence of 9 April shows that the sampled cumulus clouds are active, and that the features are not very sharp due to the early-morning light not yet illuminating those clouds (see the low-altitude and gray-looking cumuli in Fig. 13). These two conditions deteriorate the performance of feature identification and matching between contemporaneous and consecutive frames, and this leads to the large standard deviation of velocity measurements for the 9 April cumuli.

7. Summary

This paper presents a method for performing stereophotogrammetry of clouds over the open ocean. By eliminating the requirement of landmarks in the field of view, this approach widens the potential uses of stereophotogrammetry to the study of tropical oceanic convection. The key to calibrating without landmarks is to tune the calibration parameters to minimize the sum of squared distances (in the image plane) of 1) image points from their respective epipolar line for a collection of random objects (clouds, birds, ships, etc.), and 2) the actual horizon from the projected horizon.

In the field, the location of cameras is often dictated by practical constraints, such as the availability of electrical

power and Internet connectivity. This can result in the baseline—the line connecting the two cameras—being close to or within the field of view. As shown here, the precision of the reconstruction suffers the closer a cloud feature is to the baseline. Precision also degrades with distance from the camera, as expected.

Despite these limitations, the accuracy of the stereo reconstruction is quite good. The stereo-reconstructed and lidar-measured cloud-base heights show good agreement, as do the stereo-reconstructed and radiosonde-measured winds. An error analysis leads to the conclusion that the uncertainty in the reconstructed heights of shallow clouds is less than 2%, with the uncertainty increasing with altitude to values as large as 8% for cirrocumulus. Although the comparisons with lidar and radiosondes did not reveal any obvious biases in the reconstructions, a more detailed comparison is precluded by the fact that the lidar is not collocated with the imaged clouds, and the radiosonde releases are neither collocated nor contemporaneous. A more definitive statement on the accuracy will require new measurements, such as images collected by a third camera or contemporaneous measurements by a scanning cloud radar.

Like any observational technique, stereophotogrammetry has its limitations. For example, data collection can be impeded when the view of the clouds being investigated is blocked by other clouds. This occlusion can be an occasional spoiler when trying to track a cloud continuously in time. Environmental light can also impact the quality of photogrammetric measurements. Low light can result in a high signal-to-noise ratio and excess light can result in overexposure of images, both of which deteriorate the accuracy of feature matching. Nevertheless, even with basic off-the-shelf cameras, stereophotogrammetry can provide a large amount of instantaneous data for clouds in a wide field of view. For example, the development of a deep-convective

turret 20 km away can be imaged with a few tens of meters resolution and a 0.1-Hz frame rate. Although manual tracking of features over image sequences is labor intensive and time consuming, algorithms from the field of computer vision should make it possible to automate the feature identification and matching process, enabling the processing of vast amounts of data in seconds.

Acknowledgments. This work was supported initially by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under the U.S. Department of Energy Contract DE-AC02-05CH11231 and, subsequently, by the Hellman Fellows Fund. The Marine and Science Technology (MAST) Academy hosted one of the cameras, and many thanks are due to the school administration and technical staff, without whom this project would not have been possible.

REFERENCES

- Ackerman, T. P., and G. M. Stokes, 2003: The atmospheric radiation measurement program. *Phys. Today*, **56**, 38–44, doi:10.1063/1.1554135.
- Allmen, M. C., and P. Kegelmeyer, 1996: The computation of cloud-base height from paired whole-sky imaging cameras. *J. Atmos. Oceanic Technol.*, **13**, 97–113, doi:10.1175/1520-0426(1996)013<0097:TCOCBH>2.0.CO;2.
- Bradbury, D. L., and T. Fujita, 1968: Computation of height and velocity of clouds from dual, whole-sky, time-lapse picture sequences. SMRP Paper 70, Dept. of Geophysical Sciences, University of Chicago, 34 pp.
- Bradski, G., and A. Kaehler, 2008: *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, 555 pp.
- Brown, D. C., 1966: Decentering distortion of lenses. *Photogramm. Eng.*, **32**, 444–462.
- Collis, S., A. Protat, P. T. May, and C. Williams, 2013: Statistics of storm updraft velocities from TWP-ICE including verification with profiling measurements. *J. Appl. Meteor. Climatol.*, **52**, 1909–1922, doi:10.1175/JAMC-D-12-0230.1.
- Damiani, R., and Coauthors, 2008: The Cumulus, Photogrammetric, In Situ, and Doppler Observations Experiment of 2006. *Bull. Amer. Meteor. Soc.*, **89**, 57–73, doi:10.1175/BAMS-89-1-57.
- Davies-Jones, R. P., 1979: Dual-Doppler radar coverage area as a function of measurement accuracy and spatial resolution. *J. Appl. Meteor.*, **18**, 1229–1233, doi:10.1175/1520-0450-18.9.1229.
- Fielding, M. D., J. C. Chiu, R. J. Hogan, and G. Feingold, 2013: 3D cloud reconstructions: Evaluation of scanning radar scan strategy with a view to surface shortwave radiation closure. *J. Geophys. Res. Atmos.*, **118**, 9153–9167, doi:10.1002/jgrd.50614.
- Forsyth, D. A., and J. Ponce, 2003: *Computer Vision: A Modern Approach*. Prentice Hall, 689 pp.
- Gallup, D., J.-M. Frahm, P. Mordohai, and M. Pollefeys, 2008: Variable baseline/resolution stereo. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2562–2569, doi:10.1109/CVPR.2008.4587671.
- Ghate, V. P., M. A. Miller, and L. DiPreto, 2011: Vertical velocity structure of marine boundary layer trade wind cumulus clouds. *J. Geophys. Res.*, **116**, D16206, doi:10.1029/2010JD015344.
- Hartley, R., and A. Zisserman, 2003: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 655 pp.
- Hu, J., A. Razdan, and J. A. Zehnder, 2009: Geometric calibration of digital cameras for 3D cumulus cloud measurements. *J. Atmos. Oceanic Technol.*, **26**, 200–214, doi:10.1175/2008JTECHA1079.1.
- Kassander, A. R., and L. L. Sims, 1957: Cloud photogrammetry with ground-located K-17 aerial cameras. *J. Meteor.*, **14**, 43–49, doi:10.1175/0095-9634-14.1.43.
- Kassianov, E., C. N. Long, and J. Christy, 2005: Cloud-base-height estimation from paired ground-based hemispherical observations. *J. Appl. Meteor.*, **44**, 1221–1233, doi:10.1175/JAM2277.1.
- Kollias, P., and B. A. Albrecht, 2010: Vertical velocity statistics in fair-weather cumuli at the ARM TWP Nauru Climate Research Facility. *J. Climate*, **23**, 6590–6604, doi:10.1175/2010JCL3449.1.
- , —, R. Lhermitte, and A. Savtchenko, 2001: Radar observations of updrafts, downdrafts, and turbulence in fair-weather cumuli. *J. Atmos. Sci.*, **58**, 1750–1766, doi:10.1175/1520-0469(2001)058<1750:ROODA>2.0.CO;2.
- , —, and F. D. Marks Jr., 2003: Cloud radar observations of vertical drafts and microphysics in convective rain. *J. Geophys. Res.*, **108**, 4053, doi:10.1029/2001JD002033.
- , E. E. Clothiaux, M. A. Miller, B. A. Albrecht, G. L. Stephens, and T. P. Ackerman, 2007: Millimeter-wavelength radars: New frontier in atmospheric cloud and precipitation research. *Bull. Amer. Meteor. Soc.*, **88**, 1608–1624, doi:10.1175/BAMS-88-10-1608.
- , N. Bharadwaj, K. Widener, I. Jo, and K. Johnson, 2014: Scanning ARM cloud radars (SACR's). Part I: Operational sampling strategies. *J. Atmos. Oceanic Technol.*, **31**, 569–582, doi:10.1175/JTECH-D-13-00044.1.
- Koppe, C., 1896: *Photogrammetrie und Internationale Wolkenmessung*. Braunschweig Verlag, 108 pp.
- Malkus, J. S., and C. Ronne, 1954: On the structure of some cumulonimbus clouds which penetrated the high tropical troposphere. *Tellus*, **6A**, 351–366, doi:10.1111/j.2153-3490.1954.tb01130.x.
- Mather, J. H., and J. W. Voyles, 2013: The ARM Climate Research Facility: A review of structure and capabilities. *Bull. Amer. Meteor. Soc.*, **94**, 377–392, doi:10.1175/BAMS-D-11-00218.1.
- Orville, H. D., and A. R. Kassander Jr., 1961: Terrestrial photogrammetry of clouds. *J. Meteor.*, **18**, 682–687, doi:10.1175/1520-0469(1961)018<0682:TPOC>2.0.CO;2.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2007: *Numerical Recipes: The Art of Scientific Computing*. Cambridge University, 1256 pp.
- Seiz, G., E. P. Baltasvias, and A. Gruen, 2002: Cloud mapping from the ground: Use of photogrammetric methods. *Photogramm. Eng. Remote Sens.*, **68**, 941–951.
- Warner, C., J. Renick, M. Balshaw, and R. Douglas, 1973: Stereo photogrammetry of cumulonimbus clouds. *Quart. J. Roy. Meteor. Soc.*, **99**, 105–115, doi:10.1002/qj.49709941910.
- Yuter, S. E., and R. A. Houze, 1995: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part I: Spatial distribution of updrafts, downdrafts, and precipitation. *Mon. Wea. Rev.*, **123**, 1921–1940, doi:10.1175/1520-0493(1995)123<1921:TDKAME>2.0.CO;2.
- Zehnder, J. A., L. Zhang, D. Hansford, A. Radzan, N. Selover, and C. Brown, 2006: Using digital cloud photogrammetry to characterize the onset and transition from shallow to deep convection over orography. *Mon. Wea. Rev.*, **134**, 2527–2546, doi:10.1175/MWR3194.1.
- , J. Hu, and A. Razdan, 2007: A stereo photogrammetric technique applied to orographic convection. *Mon. Wea. Rev.*, **135**, 2265–2277, doi:10.1175/MWR3401.1.