Principles of stereo reconstruction of aerial objects using stationary cameras

DAVID M. ROMPS*

Department of Earth and Planetary Science, University of California, Berkeley, California, USA

Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA

ABSTRACT

An overview is given here of the principles and mathematics of stereo reconstruction of objects in the sky using stationary cameras with an emphasis on meteorological applications. Through its Atmospheric Radiation Measurement program, the Department of Energy has operated stereo-photogrammetric cameras since 2017 as part of an effort to measure the life-cycle properties of clouds. At the core of that technology is stereo reconstruction, which calculates the real-world position of an object from the location of the object's image in two cameras' photographs. Here, stereo reconstruction is stripped down to its basic elements and presented using conventions tailored to applications in atmospheric science. In addition, the resulting equations are used to illustrate the high sensitivity of reconstructed cloud positions to errors in the cameras' Euler angles. The interested reader will find here a self-contained guide to performing stereo reconstructions using distortion-corrected images from a pair of calibrated, stationary cameras, as well as a demonstration of the need for high accuracy in the measurement of camera properties and orientations.

1. Introduction

Stereo photogrammetry uses simultaneous images from two or more cameras to reconstruct the visible scene in three dimensions, i.e., to calculate the X, Y, and Z positions of objects in the cameras' common field of view. This technique has been used to study clouds since the late 1800s, when researchers began using multiple cameras to measure the altitudes of noctilucent clouds in the mesosphere (Foerster and Jesse 1892; Jesse 1896; Størmer 1933; Paton 1949; Witt 1962). Stereo photogrammetry has been applied to clouds in the troposphere from the 1950s up to the present day (e.g., Malkus and Ronne 1954; Kassander and

---

Corresponding author address: David M. Romps, 307 McCone Hall, University of California, Berkeley, CA
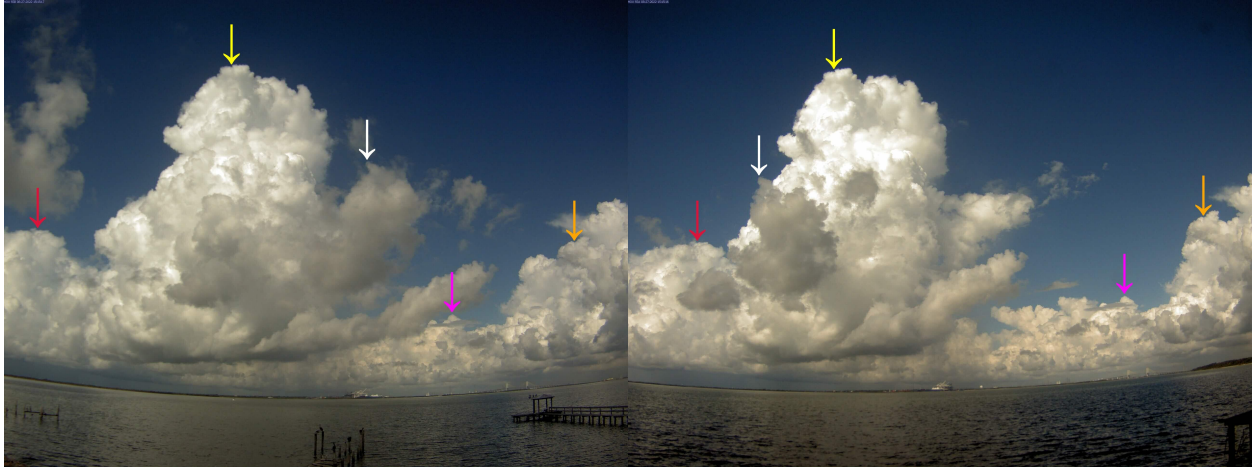
E-mail: romps@berkeley.edu

Fig. 1. Images from the (left) left and (right) right stereo cameras taken at 15:15:20 UTC on August 27, 2022 during the TRACER campaign. Colored arrows have been added to draw attention to a sample of five feature points (among hundreds) matched between the two images. Note that the horizon in these photos is warped; these images have not yet been corrected for radial distortion.

Sims 1957; Bradbury and Fujita 1968; Damiani et al. 2008; Beekmans et al. 2016; Romps and Öktem 2018). In particular, stereo cameras have been used to calculate the altitudes of cloud bases (Allmen and Kegelmeyer 1996; Seiz et al. 2002; Kassianov et al. 2005; Savoy et al. 2017), the heights of cumuliform clouds (Zehnder et al. 2007; Beekmans et al. 2016), and cloud-top vertical velocities (Orville and Kassander 1961; Zehnder et al. 2007). These techniques have led to several scientific advances, including the demonstration of: small entrainment rates in deep tropical convection (Malkus and Ronne 1954); vortical circulations in convective cloud tops (Warner et al. 1973); the dominance of drag in the momentum budget of rising clouds (Romps and Öktem 2015); the degree of plume-like versus bubble-like behavior in shallow moist convection (Romps et al. 2021); and a theory for the spacing of shallow clouds (Öktem and Romps 2021).

Today, the U.S. Department of Energy runs one of the most extensive and sustained efforts in tropospheric stereo photogrammetry, gathering data on clouds since 2017 as part of its Atmospheric Radiation Measurement (ARM) program. At the time of this publication, ARM has a total of eight stereo cameras, including six at the Southern Great Plains site in Oklahoma (Romps and Öktem 2018) and two as part of the first ARM Mobile Facility (AMF1). The AMF1 cameras have traveled to Argentina as part of the CACTI campaign (Varble et al. 2018) and to Houston, Texas as part of the TRACER campaign (Jensen et al. 2022). Figure 1 shows a pair of simultaneous images taken by the AMF1 cameras during the TRACER campaign. These cameras were facing roughly west and were separated by a north-south distance of about 1 km. The image on the left was taken by the southern (or left) camera, and the image on the right was taken by the northern (or right) camera.

Inspecting Figure 1, we see many cloud features that are present in both images, five of which are indicated by colored arrows. Identifying and matching those features is the first step in stereo reconstruction. In the early days, this feature matching was done manually, but

computers now perform this task using techniques like the Scale Invariant Feature Transform (SIFT; Lowe 1999) or block matching (Öktem and Romps 2015). For each feature point in a matched pair, we record the position of the feature point in the left camera's photograph. This position is represented by a pair of numbers giving, e.g., the distance (or number of pixels) up and to the right of the photograph's lower-left corner ($x^{l\prime}$ and $y^{l\prime}$). Likewise, we record the position of the feature point in the right camera's photograph ($x^{r\prime}$ and $y^{r\prime}$). The second step is to take those four numbers and calculate the position of the feature point in three-dimensional space (X, Y, and Z) using the methods described in section 2.

In practice, stereo photogrammetry involves far more than just stereo reconstruction. Considerable effort must be devoted to measuring the intrinsic (optical) properties of the cameras, deploying them in rigid orientations with suitable fields of view, measuring the orientation (i.e., Euler angles) of those deployed cameras, synchronizing the photographs taken by the cameras, correcting the resulting images for radial distortion, and using suitable algorithms for feature identification and matching. Only then can we reconstruct three-dimensional positions. And after we have done so, there are many additional post-processing steps that might be desired, such as using techniques of computer vision to identify edges and/or faces of clouds to make sense of the results (Romps and Öktem 2018). Despite all of this complexity, the step of stereo reconstruction – i.e., calculating the X, Y, and Z from the $x^{l\prime}$, $y^{l\prime}$, $x^{r\prime}$, and $y^{r\prime}$ – lies at the core of any stereo-photogrammetric effort. While stereo reconstruction is covered in many computer-vision textbooks (e.g., Forsyth and Ponce 2002; Hartley and Zisserman 2003), the goal here is to describe the basic principles of stereo reconstruction and its mathematics using conventions that are particularly well-suited to the application of stereo photogrammetry to atmospheric science, and to provide a guide for those interested in making meteorological measurements using stationary stereo cameras.

2. Stereo reconstruction

The objective of stereo reconstruction is to calculate the position of an object in Cartesian world coordinates using the location of the object's image in the image planes of two cameras. In this section, we will derive the equation for this stereo reconstruction. But, to do so, we must first solve the opposite problem: calculating the location of an object's image in a camera's image plane given the object's world coordinates. Only once we have solved that forward problem can we invert to perform stereo reconstruction.

Let us denote an object's world coordinates by $\boldsymbol{X} = (X, Y, Z)^{\mathrm{T}}$. Assuming we are far from Earth's poles, we will adopt the standard convention in atmospheric science in which X increases to the east, Y increases to the north, and Z increases upwards. The location of the object's image in a camera's image plane will clearly depend on many factors, including where the camera is and in what direction the camera is pointing. We will deal with each of these factors in turn, starting with the camera's location.

Despite the multiple lenses in modern cameras, photographs corrected for radial distortion can be interpreted as if they were generated by an idealized pinhole camera. In such

a model, light reaches the sensor by passing in a straight line from a source to the image plane through a small hypothetical pinhole in the camera's housing. For any camera, the location of this equivalent pinhole is termed the camera center (also known as the center of projection or the perspective center); this is where a pinhole would need to be to generate an identical image. Thus, we can model the image captured by a lens-based camera using a pinhole camera model with the pinhole located at the camera center.

Let us denote the location of the camera center as $\boldsymbol{X}_0 = (X_0, Y_0, Z_0)^{\mathrm{T}}$. We will want to subtract this from $\boldsymbol{X}$ to get the object's camera-relative position. Although we could write this as $\boldsymbol{X} - \boldsymbol{X}_0$, it is standard practice to write this using matrix multiplication. To this end, let us define $\mathbf{X} = (X, Y, Z, 1)^{\mathrm{T}}$ to be the object's homogeneous world coordinates. Throughout, $\boldsymbol{X}$ will be a three-component Cartesian vector and $\mathbf{X}$ will be the corresponding four-component homogeneous coordinates. Using this notation, the camera-relative location of the object is $\mathbf{TX}$, where

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & -X_0 \\ 0 & 1 & 0 & -Y_0 \\ 0 & 0 & 1 & -Z_0 \end{pmatrix} . \tag{1}$$

Thus, $\mathbf{TX} = (X - X_0, Y - Y_0, Z - Z_0)^{\mathrm{T}}$. Next, we need to account for the orientation of the camera, which we describe using Euler angles.

Unfortunately, Euler angles can be and have been defined in many different ways, leading to ample opportunity for confusion. Worse yet, these angles can be defined on a variety of different world coordinates. There are two such world coordinates that are commonly used in the broader field of computer vision. The first is a right-handed coordinate system in which Y points up, which contravenes the meteorological convention of Y pointing north. The second convention is a left-handed coordinate system in which Z points up, X points to the east, but Y points to the south. As a further complication, it is a standard convention in computer vision to define all rotations as right-handed. This is contrary to the meteorological definition of azimuth, in which an increase in azimuth rotates a vector around the Z axis in a left-handed fashion (i.e., clockwise as seen from above). These considerations motivate defining a set of Euler angles suited to meteorological applications.

To begin, we must first define what we mean by a camera's image plane. In a pinhole camera, the camera center is located at the pinhole and the image plane is a surface behind that pinhole, i.e., inside the camera. Similarly, for a lens-based camera, the camera center is a virtual point amidst the assembly of lenses and the image plane – occupied by an electronic sensor – is located behind camera center, i.e., deeper inside the camera. In both cases, the images that get projected onto the image plane are inverted. The resulting digital images, however, are always presented to the user inverted a second time. To avoid needing to think about this double inversion, we may imagine that our image plane is *in front* of the camera center, as depicted in Figure 2. In this way, objects project onto the image plane via the line that connects them to the camera center, and the resulting image is aligned with the real world: objects higher up from the camera's perspective appear higher up in the image,

Z and y

Y and z

image plane

y'

principal point (x'_p, y'_p)
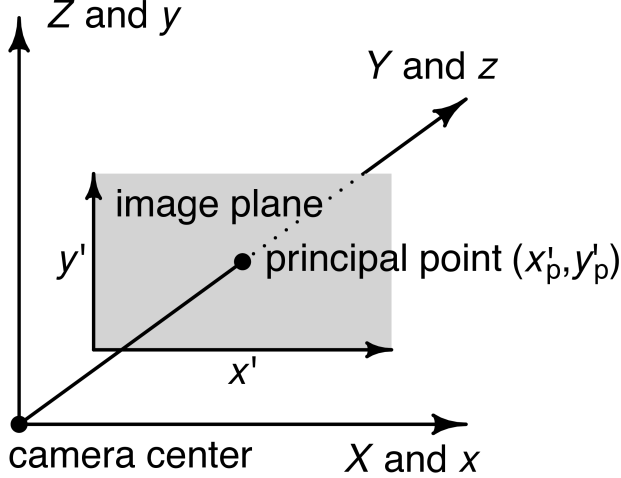
x'

camera center

X and x

FIG. 2. The location and orientation of the camera when the camera center is at the origin of the world coordinates and $\theta_a = \theta_p = \theta_r = 0$. In this orientation, the camera is pointing north with the camera's bottom facing down. The circle at the origin is the camera center and the shaded rectangle is the image plane, which has been placed in front of the camera center.

and objects that are further to the right from the camera's perspective appear further to the right in the image.

A camera carries two sets of axes with it as it rotates about its camera center. First, we have the three-dimensional coordinates $\boldsymbol{x} = (x, y, z)^{\mathrm{T}}$ in the frame of the camera. By convention, $x = y = z = 0$ is the camera center, z is in the direction the camera is pointing (from the camera center and normal to the image plane), x points out through the right side of the camera, and y points up through the top of the camera. Second, we have the coordinates x' and y' for the image plane. Note that we will reserve primes exclusively for image-plane coordinates. A common convention is to define $x' = y' = 0$ as the upper left corner of the image with either x' or y' increasing to the right and the other increasing downwards. Here, we will adopt a more intuitive convention by defining $x' = y' = 0$ as the lower-left corner of the image plane, with x' increasing to the right and y' increasing upwards.

Let us now define three angles: azimuth θ_a, which is zero when the camera is pointing north and which increases as the compass direction changes clockwise (i.e., pointing east is $\theta_a = \pi/2$); pitch θ_p, which is zero when pointing at the horizon and increases upward (i.e., pointing up is $\theta_p = \pi/2$); and roll θ_r, which is zero when the base of the camera is parallel with the ground and increases with a right-handed rotation around the pointing direction (i.e., the right side of the camera facing down corresponds to $\theta_r = \pi/2$). For $\theta_a = \theta_p = \theta_r = 0$, the camera points north and is level with the horizon. In this case, $\boldsymbol{x} = \mathbf{STX}$, where the swap matrix $\mathbf{S}$ is defined as

$$\mathbf{S} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{2}$$

which effectively swaps the Y and Z axes. Therefore, if $\theta_a = \theta_p = \theta_r = 0$, then an object at $\mathbf{X}$ is located, in the camera's frame, at $\boldsymbol{x} = \mathbf{STX} = (X - X_0, Z - Z_0, Y - Y_0)^{\mathrm{T}}$.

If the camera is oriented differently – i.e., for general θ_a, θ_p, and θ_r – the relationship between $\boldsymbol{x}$ and $\mathbf{X}$ gets modified by a rotation matrix. Here, let $c_\zeta = \cos(\theta_\zeta)$ and $s_\zeta = \sin(\theta_\zeta)$,

where $\zeta = $ a, p, or r. In this case, $\boldsymbol{x} = \mathbf{RSTX}$ with the rotation matrix $\mathbf{R}$ given by

$$\mathbf{R} = \begin{pmatrix} c_r & -s_r & 0 \\ s_r & c_r & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_p & -s_p \\ 0 & s_p & c_p \end{pmatrix} \begin{pmatrix} c_a & 0 & -s_a \\ 0 & 1 & 0 \\ s_a & 0 & c_a \end{pmatrix}. \tag{3}$$

If the angles are all zero, then each of these is an identity matrix. On the other hand, if the angles are non-zero, then these correspond to a rotation of θ_a around the camera's y axis (to set the azimuth), followed by a rotation of θ_p around the camera's new x axis (to set the pitch), following by a rotation of θ_r around the camera's new z axis (to set the roll). Also, note that these matrices do not transform the camera itself; instead, they transform the location of objects to the camera's new coordinate system. For example, the right-most matrix with $\theta_a = \pi/2$, which orients the camera eastward, transforms a point originally at $(10, 0, 0)^T$, which is ten units of distance to the camera's right when the camera points north, to $(0, 0, 10)^T$, which is ten units directly in front of the camera.

Finally, we can define the camera matrix $\mathbf{C}$, which projects objects in three-dimensional space onto the two-dimensional image plane. This projection is defined by the line connecting the object to the camera center (at $x = y = z = 0$); the image of that object is located where the line intersects the image plane. That location on the image plane can be given in terms of the coordinates x' and y', with x' increasing to the right, y' increasing upwards, and $x' = y' = 0$ corresponding to the lower-left corner of the image plane. As with world coordinates, we can define a Cartesian vector $\boldsymbol{x}' = (x', y')^T$ on the image plane and also homogeneous coordinates $\mathbf{x}' = (x', y', 1)^T$. For the purposes of projecting the object onto the image plane, it is easier to work with the homogeneous coordinates. In particular, $\mathbf{x}'$ is related to the camera-relative position $\boldsymbol{x}$ of the object by the equation $w\mathbf{x}' = \mathbf{C}\boldsymbol{x}$ for some real number w, where the camera matrix $\mathbf{C}$ is defined as

$$\mathbf{C} = \begin{pmatrix} f & 0 & x'_p \\ 0 & f & y'_p \\ 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

Here, f is the focal length and x'_p and y'_p are the coordinates of the principal point on the image plane. The principal point is the orthogonal projection of the camera center to the image plane. A camera configured to photograph distant objects in the sky will be focused at infinity, so the distance from the camera center to the principal point (called the principal distance) will be equal to the focal length f.

To see why $w\mathbf{x}' = \mathbf{C}\boldsymbol{x}$ is the correct equation, let us consider the case where the object is at $y = 0$ so that we can picture everything in the $y = 0$ plane. This case is equivalent to a camera in two dimensions. In this case, $\boldsymbol{x} = (x, 0, z)^T$ and $w\mathbf{x}' = \mathbf{C}\boldsymbol{x}$ can be written out as

$$wx' = fx + x'_p z \tag{5}$$
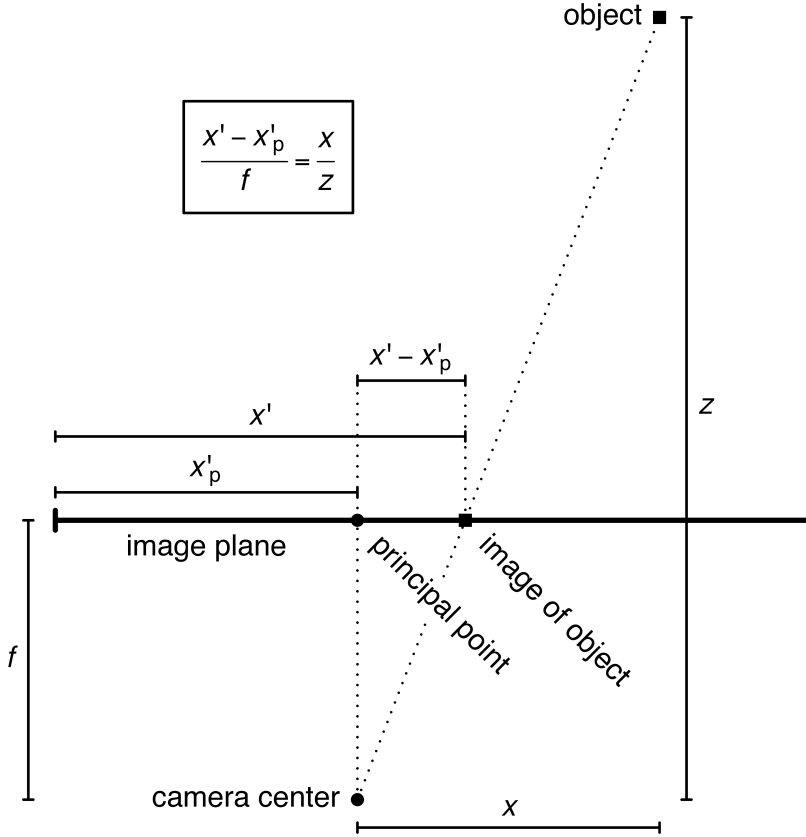$$wy' = y'_p z \tag{6}$$
$$w = z. \tag{7}$$

FIG. 3. A camera in two dimensions, illustrating the relationship between an object's camera-relative two-dimensional coordinates (x and z) and the position of the object's image in the image plane (x').

Using the last equation to eliminate w from the first two and then dividing both sides of those equations by z, we get

$$x' = fx/z + x'_{\mathrm{p}} \tag{8}$$

$$y' = y'_{\mathrm{p}}. \tag{9}$$

The second equation is telling us that the object's image will be at the same height in the image plane as the principal point; this is just a consequence of the object and principal point both being at $y = 0$. The first equation is more interesting: it gives the left-right position of the object's image in the image plane (x') as a function of the object's position (x and z). This equation can be rewritten as

$$\frac{x' - x'_{\mathrm{p}}}{f} = \frac{x}{z}. \tag{10}$$

This is a straightforward consequence of geometry, as illustrated in Figure 3. There, we see that there are two right triangles that are similar: one with legs of length $x' - x'_{\mathrm{p}}$ and f and the other with legs of length x and z. Since they are similar, the ratios of their leg lengths are equal. Therefore, we see that the camera matrix encodes this geometrical fact.

Combining $\boldsymbol{x} = \mathbf{RSTX}$ and $w\mathbf{x}' = \mathbf{C}\boldsymbol{x}$, we have

$$w\mathbf{x}' = \mathbf{PX}, \tag{11}$$

for some real number w, where the projection matrix $\mathbf{P}$ is defined as

$$\mathbf{P} = \mathbf{CRST}\,, \tag{12}$$

with $\mathbf{T}$, $\mathbf{S}$, $\mathbf{R}$, and $\mathbf{C}$ as defined in equations (1), (2), (3), and (4).

Now, imagine that we have two cameras, which we will refer to as left and right with subscripts l and r, respectively. Because they have different locations and possibly different orientations, focal lengths, and image sensors, they will have different projection matrices. Let us call them $\mathbf{P}^l$ and $\mathbf{P}^r$. Similarly, let the homogeneous coordinates of the image of $\mathbf{X}$ in the two image planes be $\mathbf{x}^{l\prime}$ and $\mathbf{x}^{r\prime}$. We now have two equations,

$$w^l \mathbf{x}^{l\prime} = \mathbf{P}^l \mathbf{X} \tag{13}$$
$$w^r \mathbf{x}^{r\prime} = \mathbf{P}^r \mathbf{X}\,, \tag{14}$$

for some real numbers w^l and w^r. These are six equations, which we can write out explicitly as

$$w^l x^{l\prime} = (\mathbf{P}^l \mathbf{X})_1 \tag{15}$$
$$w^l y^{l\prime} = (\mathbf{P}^l \mathbf{X})_2 \tag{16}$$
$$w^l = (\mathbf{P}^l \mathbf{X})_3 \tag{17}$$
$$w^r x^{r\prime} = (\mathbf{P}^r \mathbf{X})_1 \tag{18}$$
$$w^r y^{r\prime} = (\mathbf{P}^r \mathbf{X})_2 \tag{19}$$
$$w^r = (\mathbf{P}^r \mathbf{X})_3\,, \tag{20}$$

where the subscripts denote the component of the homogeneous coordinates. Using equations (17) and (20) to eliminate w^l and w^r, these become

$$x^{l\prime}(\mathbf{P}^l \mathbf{X})_3 = (\mathbf{P}^l \mathbf{X})_1 \tag{21}$$
$$y^{l\prime}(\mathbf{P}^l \mathbf{X})_3 = (\mathbf{P}^l \mathbf{X})_2 \tag{22}$$
$$x^{r\prime}(\mathbf{P}^r \mathbf{X})_3 = (\mathbf{P}^r \mathbf{X})_1 \tag{23}$$
$$y^{r\prime}(\mathbf{P}^r \mathbf{X})_3 = (\mathbf{P}^r \mathbf{X})_2\,. \tag{24}$$

We see that we have four homogeneous linear equations in $\mathbf{X}$. We can write them out explicitly as

$$x^{l\prime}(P^l_{31}X + P^l_{32}Y + P^l_{33}Z + P^l_{34}) = P^l_{11}X + P^l_{12}Y + P^l_{13}Z + P^l_{14} \tag{25}$$
$$y^{l\prime}(P^l_{31}X + P^l_{32}Y + P^l_{33}Z + P^l_{34}) = P^l_{21}X + P^l_{22}Y + P^l_{23}Z + P^l_{24} \tag{26}$$
$$x^{r\prime}(P^r_{31}X + P^r_{32}Y + P^r_{33}Z + P^r_{34}) = P^r_{11}X + P^r_{12}Y + P^r_{13}Z + P^r_{14} \tag{27}$$
$$y^{r\prime}(P^r_{31}X + P^r_{32}Y + P^r_{33}Z + P^r_{34}) = P^r_{21}X + P^r_{22}Y + P^r_{23}Z + P^r_{24}\,, \tag{28}$$

where P^l_{mn} is the element in row m and column n of the matrix $\mathbf{P}^l$, and likewise for P^r_{mn}. With a little algebra, this can be written as four linear equations in $\boldsymbol{X}$, which we can write as

$$\mathbf{A}\boldsymbol{X} = \mathbf{B}\,, \tag{29}$$

where $\mathbf{A}$ is a four-by-three matrix,

$$\mathbf{A} = \begin{pmatrix} x^{l\prime}P^l_{31} - P^l_{11} & x^{l\prime}P^l_{32} - P^l_{12} & x^{l\prime}P^l_{33} - P^l_{13} \\ y^{l\prime}P^l_{31} - P^l_{21} & y^{l\prime}P^l_{32} - P^l_{22} & y^{l\prime}P^l_{33} - P^l_{23} \\ x^{r\prime}P^r_{31} - P^r_{11} & x^{r\prime}P^r_{32} - P^r_{12} & x^{r\prime}P^r_{33} - P^r_{13} \\ y^{r\prime}P^r_{31} - P^r_{21} & y^{r\prime}P^r_{32} - P^r_{22} & y^{r\prime}P^r_{33} - P^r_{23} \end{pmatrix} , \tag{30}$$

and $\mathbf{B}$ is a four-by-one matrix,

$$\mathbf{B} = \begin{pmatrix} P^l_{14} - x^{l\prime}P^l_{34} \\ P^l_{24} - y^{l\prime}P^l_{34} \\ P^r_{14} - x^{r\prime}P^r_{34} \\ P^r_{24} - y^{r\prime}P^r_{34} \end{pmatrix} . \tag{31}$$

Since there are four equations and only three degrees of freedom in $\boldsymbol{X}$, this is an overdetermined problem. To understand how we ended up with an overdetermined problem, consider an object whose image appears at $\boldsymbol{x}^{l\prime}$ in the left camera's image plane. If that is all the information we have, then we only know that the object is located somewhere along the line connecting the left camera center and the point $\boldsymbol{x}^{l\prime}$ on the left camera's image plane. Next, imagine that we make that line visible to all using a powerful laser beam that reflects off the impurities in the air along its path. In the left camera's image plane, the reflected light from that laser beam would appear simply as a point in its image plane because the beam intersects its camera center. In general, however, the laser beam does not pass through the right camera center, so the illuminated path appears as a line in the image plane of the right camera. That line is called the "epipolar line," and the image of the object must lie on that line in the right camera's image plane. To resolve the ambiguity as to where the object is, we only need to know where it is on that epipolar line, which can be determined with only one piece of additional information: either $x^{r\prime}$ or $y^{r\prime}$. That would give us three pieces of information to determine the three world coordinates of the object.

In reality, we do not reconstruct positions that way. Instead, we use both $x^{r\prime}$ and $y^{r\prime}$. If both the cameras were perfect and our knowledge of the camera's projection matrices were perfect, then the four equations would be degenerate and we could solve for $\boldsymbol{X}$ using only three of them. In reality, however, neither the cameras nor our information is perfect. The way this manifests in the images is that $\boldsymbol{x}^{r\prime}$ does not lie exactly on the epipolar line corresponding to $\boldsymbol{x}^{l\prime}$. Mathematically, this manifests as there being no solution to $\mathbf{A}\boldsymbol{X} = \mathbf{B}$.

To deal with this, we estimate the object's position as the $\boldsymbol{X}$ that minimizes the sum of squared residuals

$$|\mathbf{A}\boldsymbol{X} - \mathbf{B}|^2 = (\mathbf{A}\boldsymbol{X} - \mathbf{B})^{\mathrm{T}}(\mathbf{A}\boldsymbol{X} - \mathbf{B}) . \tag{32}$$

Differentiating with respect to $\boldsymbol{X}$ and setting the result to zero, we get

$$\boldsymbol{X} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{B} . \tag{33}$$

This, then, is our estimate of the position of the object in world coordinates and our stereo reconstruction is complete.

3. Sensitivity to error

The above equations are straightforward to code in programming languages like R and Python. If attempting these calculations for the first time, a worthwhile exercise is to code up the projection matrix for a single camera with some arbitrarily chosen parameters, choose an arbitrary $\boldsymbol{X}$, and then calculate $\boldsymbol{x}'$ and confirm that it is roughly as expected. For this purpose, it can be helpful if the arbitrary parameters are chosen to be rather simple. A next step can be to vary $\boldsymbol{X}$ or some of the camera parameters and make sense of how $\boldsymbol{x}'$ varies in response. Finally, to put all of the equations through a sanity check, parameters for two cameras can be defined, two sets of $\boldsymbol{x}'$ calculated, and then those $\boldsymbol{x}'$ values, along with the two projection matrices, can be used to reconstruct $\boldsymbol{X}$. If everything has been coded properly, this will return the original position.

As mentioned above, there is always some amount of error, or uncertainty, in the values fed into this mathematical machinery. For example, the Euler angles of the cameras are never known exactly and even a small object in the field of view may appear smudged out over a few pixels in the digital images. Since reconstructions are sensitive to these errors, it is essential, in practice, to reduce the uncertainties as much as possible by, e.g., measuring the Euler angles as accurately as possible and using a camera with high-quality optics.

To illustrate the sensitivity of reconstructions to measurement errors, consider two cameras at ground level that are separated by a 1-km east-west baseline. Without any loss of generality, we may choose a world coordinate system that has the left camera at $\boldsymbol{X}_0^{\mathrm{l}} = (-0.5 \text{ km}, 0, 0)^{\mathrm{T}}$ and the right camera at $\boldsymbol{X}_0^{\mathrm{r}} = (0.5 \text{ km}, 0, 0)^{\mathrm{T}}$. Consider a feature point (e.g., on a cloud) located 10 km north of and 5 km above the middle of the cameras' baseline, i.e., $\boldsymbol{X} = (0, 10 \text{ km}, 5 \text{ km})^{\mathrm{T}}$. This configuration – with a 1-km baseline and a mid-tropospheric cloud 10-km away – is typical in the stereo photogrammetry of clouds. For simplicity, let us also assume that each of the cameras is pointed directly at that feature point with zero roll. To find the Euler angles that this implies for the left camera, we use $\boldsymbol{x} = \mathbf{RSTX}$ with $\mathbf{STX} = (0.5, 5, 10)^{\mathrm{T}}$ and $\boldsymbol{x} = (0, 0, \sqrt{0.5^2 + 10^2 + 5^2})^{\mathrm{T}}$ and, recalling that $\theta_{\mathrm{r}} = 0$, solve for θ_{p} and θ_{a} in $\mathbf{R}$. This gives $\theta_{\mathrm{a}}^{\mathrm{l}} = \tan^{-1}(0.05)$ and $\theta_{\mathrm{p}}^{\mathrm{l}} = \tan^{-1}\left(\sqrt{1.0025}/2.005\right)$. By symmetry of the setup, $\theta_{\mathrm{a}}^{\mathrm{r}} = -\theta_{\mathrm{a}}^{\mathrm{l}}$ and $\theta_{\mathrm{p}}^{\mathrm{r}} = \theta_{\mathrm{p}}^{\mathrm{l}}$.

In reality, there will be errors in the measurement of the cameras' Euler angles and in the measurement of the location of the feature point's image in the image plane. For the setup described here, a small error $\delta\theta_{\mathrm{p}}^{\mathrm{l}}$ in $\theta_{\mathrm{p}}^{\mathrm{l}}$ adds approximately the same error to the reconstruction as an error $\delta\theta_{\mathrm{p}}^{\mathrm{l}} f$ in $y^{\mathrm{l}\prime}$, and likewise for the right camera. Similarly, a small error $\delta\theta_{\mathrm{a}}^{\mathrm{l}}$ in $\theta_{\mathrm{a}}^{\mathrm{l}}$ adds approximately the same error to the reconstruction as an error $\cos(\theta_{\mathrm{p}}^{\mathrm{l}})\delta\theta_{\mathrm{a}}^{\mathrm{l}} f \approx 0.9\delta\theta_{\mathrm{a}}^{\mathrm{l}} f \approx \delta\theta_{\mathrm{a}}^{\mathrm{l}} f$ in $x^{\mathrm{l}\prime}$, and likewise for the right camera. Therefore, we can simultaneously emulate Euler-angle errors and feature-matching errors by adding independent Gaussian errors to the components of $\boldsymbol{x}^{\mathrm{l}\prime}$ and $\boldsymbol{x}^{\mathrm{r}\prime}$ (i.e., four independent errors, drawn from a normal distribution, added in turn to each of $x^{\mathrm{l}\prime}$, $y^{\mathrm{l}\prime}$, $x^{\mathrm{r}\prime}$, and $y^{\mathrm{r}\prime}$). When we then reconstruct $\boldsymbol{X}$ from these image-plane coordinates, we will not get the original value, but the original value plus some error.
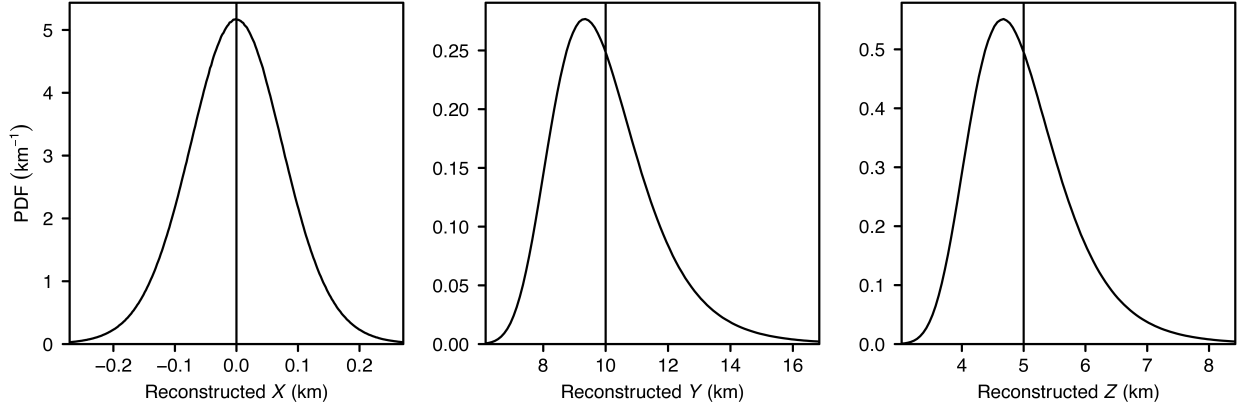
FIG. 4. Distributions of the components of the reconstructed $\boldsymbol{X}$ of the cloud given independent Gaussian noise in $x^{\mathrm{l}\prime}$, $y^{\mathrm{l}\prime}$, $x^{\mathrm{r}\prime}$, and $y^{\mathrm{r}\prime}$ with a standard deviation of $0.01f$, which is equivalent to errors in the Euler angles of pitch and azimuth of about 0.01 radians. Vertical lines denote the actual values.

To use a small-sounding error, let us choose the standard deviation of the Euler-angle error to be 0.01 radians, or about 0.57°, which is equivalent to a standard deviation of a feature-matching error of about $0.01f$. Adding independent errors drawn from a normal distribution with zero mean and a standard deviation of $0.01f$ to each of the components of $\boldsymbol{x}^{\mathrm{l}\prime}$ and $\boldsymbol{x}^{\mathrm{r}\prime}$, reconstructing $\boldsymbol{X}$, and then repeating ten million times, we get the distribution of reconstructions shown in Figure 4. We see that errors in the Euler angles of only 0.01 radians lead to substantial errors in the reconstruction of $\boldsymbol{X}$, manifesting as about ± 2 km error in the northward distance to the cloud and ± 1 km error in the height of the cloud. In practical applications of stereo photogrammetry to meteorology, the errors must be substantially less than this, which demands a painstaking characterization of the cameras' Euler angles and optics.

4. Summary

We have stepped through the principles of stereo reconstruction using stationary cameras, which have the benefit of time-invariant projection matrices. Note that we have assumed here that we are working with images that are consistent with a pinhole camera model, and so are not otherwise distorted by a camera's real optics. In practice, the images taken from a real camera need to be corrected for radial distortion before proceeding with any of the steps outlined here.

After measuring a camera's internal parameters (f, x'_{p}, and y'_{p}), its position in world coordinates (X_0, Y_0, and Z_0), and its azimuth, pitch, and roll (θ_{a}, θ_{p}, and θ_{r}), we can calculate its projection matrix using equation (12). We do the same for the second camera and label them "left" and "right," at which point we have their projection matrices $\mathbf{P}^{\mathrm{l}}$ and $\mathbf{P}^{\mathrm{r}}$, respectively. For any object (or cloud feature) that appears in both of the cameras' synchronized images, we can measure the position of that object in the left image ($x^{\mathrm{l}\prime}$ and $y^{\mathrm{l}\prime}$) and likewise for the right image ($x^{\mathrm{r}\prime}$ and $y^{\mathrm{r}\prime}$). Defining the matrix $\mathbf{A}$ and vector $\mathbf{B}$

using equations (30) and (31), we can then calculate (i.e., reconstruct) the three-dimensional location $\boldsymbol{X}$ of the object using equation (33).

While it is straightforward to code up these equations and experiment with toy examples, it is important to recognize how sensitive the stereo reconstruction is to real-world errors and uncertainty. Even an uncertainty in the Euler angles much less than one degree can still lead to unacceptably large errors, as shown in Figure 4. Of great importance to practical applications, therefore, is a careful measurement of Euler angles, as well as careful characterization of internal camera parameters and the selection of high-quality optics.

ORCID

David M. Romps `https://orcid.org/0000-0001-7649-5175`

Data availability

Observational data were obtained from the Atmospheric Radiation Measurement (ARM) user facility, a U.S. Department of Energy (DOE) Office of Science user facility managed by the Biological and Environmental research program. The data shown in Figure 1 comes from `houstereocamaS5.a1.20220827.151520.jpg` within `houstereocamaS5.a1.20220827.000000.jpg.tar` downloaded from `https://adc.arm.gov/discovery/#/results/id::houstereocamaS5.a1_jpg_macro_stereocam_cloud?dataLevel=a1&showDetails=true` and `houstereocambS5.a1.20220827.151520.jpg` within `houstereocambS5.a1.20220827.000000.jpg.tar` downloaded from `https://adc.arm.gov/discovery/#/results/id::houstereocambS5.a1_jpg_macro_stereocam_cloud?dataLevel=a1&showDetails=true`.

References

Allmen, M. C. and P. Kegelmeyer, 1996: The computation of cloud-base height from paired whole-sky imaging cameras. *Journal of Atmospheric and Oceanic Technology*, **13 (1)**, 97–113.

Beekmans, C., J. Schneider, T. Läbe, M. Lennefer, C. Stachniss, and C. Simmer, 2016: Cloud photogrammetry with dense stereo for fisheye cameras. *Atmospheric Chemistry and Physics*, **16 (22)**, 14 231–14 248.

Bradbury, D. L. and T. Fujita, 1968: Computation of height and velocity of mesospheric clouds from dual, whole-sky, time-lapse picture sequences. Satellite and Mesometeorology Research Project (SMRP) Paper No. 70, Department of Geophysical Sciences, University of Chicago, Chicago, Illinois.

Damiani, R., B. Geerts, J. Demko, S. Haimov, J. French, J. Zehnder, A. Razdan, J. Hu, J. Petti, M. Leuthold, and G. S. Poulos, 2008: The Cumulus, Photogrammetric, In Situ, and Doppler Observations experiment of 2006. *Bulletin of the American Meteorological Society*, **89 (1)**, 57–73.

Foerster, W. and O. Jesse, 1892: Aufforderung zu Beobachtungen der leuchtenden Nacht-wolken. *Astronomische Nachrichten*, **130 (3120)**, 425–430.

Forsyth, D. A. and J. Ponce, 2002: *Computer vision: a modern approach.* Prentice Hall Professional Technical Reference.

Hartley, R. and A. Zisserman, 2003: *Multiple view geometry in computer vision*, Vol. 2. Cambridge University Press.

Jensen, M. P., J. H. Flynn, L. M. Judd, P. Kollias, C. Kuang, G. Mcfarquhar, R. Nad-karni, H. Powers, and J. Sullivan, 2022: A succession of cloud, precipitation, aerosol, and air quality field experiments in the coastal urban environment. *Bulletin of the American Meteorological Society*, **103 (2)**, 103–105.

Jesse, O., 1896: Die Höhe der leuchtenden Nachtwolken. *Astronomische Nachrichten*, **140 (3347)**, 161–168.

Kassander, A. R. and L. L. Sims, 1957: Cloud photogrammetry with ground-located K-17 aerial cameras. *Journal of Meteorology*, **14 (1)**, 43–49.

Kassianov, E., C. N. Long, and J. Christy, 2005: Cloud-base-height estimation from paired ground-based hemispherical observations. *Journal of Applied Meteorology*, **44 (8)**, 1221–1233.

Lowe, D. G., 1999: Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, **2**, 1150–1157.

Malkus, J. S. and C. Ronne, 1954: On the structure of some cumulonimbus clouds which penetrated the high tropical troposphere. *Tellus*, **6 (4)**, 351–366.

Öktem, R. and D. M. Romps, 2015: Observing atmospheric clouds through stereo recon-struction. *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, Vol. 9393, 93 930H–1.

Öktem, R. and D. M. Romps, 2021: Prediction for cloud spacing confirmed using stereo cameras. *Journal of the Atmospheric Sciences*, **78 (11)**, 3717–3725.

Orville, H. D. and A. R. Kassander, 1961: Terrestrial photogrammetry of clouds. *Journal of Meteorology*, **18 (5)**, 682–687.

Paton, J., 1949: Luminous night clouds. *The Meteorological Magazine*, **78 (930)**, 354–357.

Romps, D. M. and R. Öktem, 2015: Stereo photogrammetry reveals substantial drag on cloud thermals. *Geophysical Research Letters*, **42 (12)**, 5051–5057.

Romps, D. M. and R. Öktem, 2018: Observing clouds in 4D with multiview stereophotogrammetry. *Bulletin of the American Meteorological Society*, **99 (12)**, 2575–2586.

Romps, D. M., R. Öktem, S. Endo, and A. M. Vogelmann, 2021: On the lifecycle of a shallow cumulus cloud: Is it a bubble or plume, active or forced? *Journal of the Atmospheric Sciences*, **78 (9)**, 2823–2833.

Savoy, F. M., S. Dev, Y. H. Lee, and S. Winkler, 2017: Stereoscopic cloud base reconstruction using high-resolution whole sky imagers. *2017 IEEE International Conference on Image Processing (ICIP)*, 141–145.

Seiz, G., E. P. Baltsavias, and A. Gruen, 2002: Cloud mapping from the ground: Use of photogrammetric methods. *Photogrammetric Engineering & Remote Sensing*, **68 (9)**, 941–951.

Størmer, C., 1933: Height and velocity of luminous night clouds observed in Norway 1932. Tech. Rep. 6, Norwegian Academy of Sciences.

Varble, A., S. Nesbitt, P. Salio, E. Zipser, S. van den Heever, G. McFarquhar, P. Kollias, S. Kreidenweis, P. DeMott, M. Jensen, J. R Houze, K. Rasmussen, R. Leung, D. M. Romps, D. Gochis, E. Avila, and C. W. nd P Borque, 2018: Cloud, Aerosol, and Complex Terrain Interactions (CACTI) Science Plan. Tech. Rep. DOE/SC-ARM-17-004, Department of Energy.

Warner, C., J. Renick, M. Balshaw, and R. Douglas, 1973: Stereo photogrammetry of cumulonimbus clouds. *Quarterly Journal of the Royal Meteorological Society*, **99 (419)**, 105–115.

Witt, G., 1962: Height, structure and displacements of noctilucent clouds. *Tellus*, **14 (1)**, 1–18.

Zehnder, J. A., J. Hu, and A. Razdan, 2007: A stereo photogrammetric technique applied to orographic convection. *Monthly Weather Review*, **135 (6)**, 2265–2277.