



# **Cloud Photogrammetry with Dense Stereo for Fisheye Cameras**

Christoph Beekmans<sup>1</sup>, Johannes Schneider<sup>2</sup>, Thomas Läbe<sup>2</sup>, Cyrill Stachniss<sup>2</sup>, and Clemens Simmer<sup>1</sup> <sup>1</sup>Meteorological Institute, University of Bonn <sup>2</sup>Institute of Geodesy and Geoinformation, University of Bonn *Correspondence to:* Christoph Beekmans (cbeekmans@uni-bonn.de)

**Abstract.** In this paper, we present our approach for dense 3D cloud reconstruction using two hemispheric sky imagers with fisheye lenses in a stereo setup. Fisheye lenses follow a different projection function than classical pinhole-type cameras, which provide a large field of view with a single image, but also renders the computation of dense 3D information more complicated, such that we cannot rely entirely on standard implementations for dense 3D stereo reconstruction.

- 5 In this work, we examine the epipolar rectification model, which allows the use of dense matching algorithms designed for classical perspective cameras to search for disparity information at every pixel. Together with an appropriate camera calibration, which includes internal camera geometry and global position and orientation of the stereo camera pair, we can use the disparity information for dense 3D stereo reconstruction of the a cloud and thus estimate its shape. From the obtained 3D shapes, cloud dynamics, size, motion, type and spacing can be derived and used e.g. for radiation closure under cloudy conditions.
- We implemented and evaluated the proposed approach using real world data and present two case studies. In the first case, we validate the quality and accuracy of the method by comparing the stereo reconstruction of a stratocumulus layer with the reflectivity observations measured by a cloud radar and the cloud base height estimated from a Lidar-ceilometer. The second case analyzes a rapid cumulus convection in the presence of strong wind shear.

# 1 Introduction

- 15 Ground-based photogrammetry has the great potential to complement current cloud observations due to its high spatial and temporal resolution. So-called hemispheric sky imagers record images with high temporal and spatial resolution and provide a complete hemispheric view of the cloudy sky. Up to now, such imagers are predominantly used only for the derivation of cloud cover or cloud type classification. The derivation of additional information related to cloud size and extension including their temporal development, especially of convective clouds, would provide valuable information for radiation closure studies under
- 20 cloudy conditions and can be used for validation of LES-scale cloud simulations, e.g. like the new ICON model (Zängl et al. (2015)).

Current ground-based cloud observation are made with cloud radars, lidars, lidar-ceilometers and infrared and microwave radiometers, all of which usually only sense clouds along a pencil beam, but are not able to view the whole cloud over time intervals of seconds during which clouds do not change significantly. Recent work show that stereo photogrammetry of clouds

25 may help to close this gap to some extent.





In this paper we investigate the potential of deriving the 3D structure of clouds with two hemispheric sky imagers, cf. to Figure 1. Two cameras with spatial displacement and simultaneous time of exposure provide the necessary information to perform a 3D reconstruction. Such stereo techniques are a well studied field in photogrammetry and computer vision and early approaches of cloud photogrammetry date back to the late 19th century (Koppe, 1896).

- 3D stereo reconstruction is based on triangulation. Knowing the two camera orientations and the direction vector (baseline) between the cameras, each pair of corresponding image points can be back-projected into ray directions which intersect in the mapped 3D point. This requires accurately known parameters for the interior orientation, e. g. focal length and lens distortion, and also accurate knowledge of the exterior orientation, namely the position and orientation of the cameras in space, which both need to be determined by a calibration procedure.
- 10 The main contribution of this paper is the combination of a dense stereo correspondence technique and a large scale stereo setup of hemispheric cameras in order to derive dense 3D cloud geometries, especially for convective clouds. In contrast to regular feature-based methods used in previous studies, dense methods seek a correspondence for every pixel in the stereo images, while at the same time imposing spatial consistency constraints, which is an important aspect in our application due to the lack of contrast in the appearance of clouds.
- 15 As a prerequisite, we employ an epipolar rectification technique on the recorded images, that is designed for fisheye stereo cameras and is needed to apply the dense stereo correspondence algorithm. The obtained correspondence information can then be used for triangulation and 3D reconstruction.

The paper is organized as follows. In Sec. 2, we discuss previous studies and their contributions to the field of cloud photogrammetry. In Sec. 3, we describe the fisheye camera model, the camera calibration techniques, as well as the epipolar rectification method and triangulation. Sec. 4 introduces the employed dense stereo algorithm to obtain corresponding image points. Sec. 5 presents our stereo setup, a geometric uncertainty analysis and two case studies of our reconstructions. One case shows a reconstructed stratocumulus layer which serves as a validation for the achieved geometric accuracy including comparisons with lidar-ceilometer and cloud radar observations. The second case analyzes the cloud development within strong convection and wind shear and illustrates the quality of the cloud morphology reconstruction.

# 25 2 Related Work

Recent studies employed cameras with wide-angle lenses in a stereo setup to recover 3D information of clouds. For example, Seiz (2003) exploited a stereo pair of consumer cameras with a 800 m baseline to derive the cloud base height within a field of view of about 100° with errors well below 5%. Hu et al. (2010) used a stereo camera system with a spatial offset of 1.5 km oriented towards the Santa Catarina Mountains near Tuscon, Arizona to reconstruct the evolution of orographic convection over

30 the diurnal cycle in three dimensions. Recently, Öktem et al. (2014) used a stereo camera setup for the observation of maritime clouds near Biscayne Bay, Florida, with a distance of 873 m between the cameras. Their results show a good agreement with heights obtained with a lidar, yielding errors of mostly below 2% for shallow clouds and up to 8% for high cirrocumulus





5

clouds. They also compared the derived cloud motion speed of the individual cloud layers with wind speed measurements from radiosondes.

The studies of Seiz (2003) and Öktem et al. (2014) show that an accurate cloud reconstruction is possible with a stereo camera system. To the best of our knowledge, only Hu et al. (2010) used stereo vision to reconstruct a convective cloud. Experiments involving sky imagers focused on the derivation of the cloud base height.

Allmen and Kegelmeyer (1996) used two Whole-Sky-Imagers (WSI) to derive cloud base heights. They experienced that a standard ordinary stereo matching method failed due to the rather large distance between the cameras of 5 km. They report that more than 50% of the estimated cloud heights deviated with less than 5% from heights obtained by the lidar-ceilometer. Kassianov et al. (2010) compared cloud base heights derived from two virtual and two real fisheye cameras located at the ARM

10 site in the Southern Great Planes, with a spatial offset of the cameras of 540 m. They used stochastic simulations to create a virtual cloud field and used the virtual fisheye projections for stereo vision. Comparisons with micropulse lidar showed that typical errors were about 10% for low-level clouds up to 2 km high.

In our work, we use two hemispheric sky imagers in a stereo setup and apply dense stereo methods to recover a dense 3D cloud geometry. We evaluate our results by comparing them with cloud base height observations from a lidar-ceilometer and 15 reflectivity profiles of a cloud radar.

### 3 Camera Calibration and Stereo Reconstruction

In this section, we describe the projection model for fisheye cameras and we formulate the geometric relationship between two sky imagers, which is required for 3D stereo reconstruction. We also introduce an epipolar image rectification scheme for fisheye stereo cameras that allows to identify corresponding image points between the two images using a dense stereo matching algorithm. Finally, we describe the employed camera calibration methods and the triangulation of 3D points using corresponding image points in the epipolar rectified images.

20

## 3.1 Interior Orientation of a Fisheye Camera

The interior orientation of a camera describes the camera specific projection of light onto the image plane. In case of a central projection each sensed ray passes through a certain projection center in one particular direction. When the geometry of the camera is known, that is, when the camera is calibrated, one can compute this direction for each measured image point.

The camera model contains a projection function, which should be close to the projection of light in the optics. The interior orientation consist of the camera calibration parameters of this model, describing the camera specific projection on the image plane. The projection can be split up into a mapping of a 3D point P to a 2D point  $\mathbf{x}'$  on the model image plane, and a mapping of  $\mathbf{x}'$  to  $\mathbf{x}$  to the actual pixel coordinates on the sensor plane. While most cameras follow the pinhole camera model (Sonka

30 et al., 1999; Stockman and Shapiro, 2001), fisheye cameras have lenses with a different projection function and follow the omnidirectional camera model (Kannala and Brandt, 2006; Bakstein and Pajdla, 2002). Figure 2 visualizes the omnidirectional





5

camera model. Each projection ray passes through the projection center and intersects the image sphere in the point  $\mathbf{x}''$ , which can be considered as the ray direction. The *optical axis* intersects the image plane in the *principal point*  $\mathbf{x}_0$ .

The ray direction  $\mathbf{x}''$  can be mapped to  $\mathbf{x}'$  using e.g. one of used projection functions  $r(\theta)$  provided by Abraham and Förstner (2005). Each symmetric projection function  $r(\theta)$  defines the distance between  $\mathbf{x}'$  to the principal point  $\mathbf{x}_0$  only as a function of the zenith angle  $r(\theta)$  between the incoming projection ray and the optical axis as depicted in Figure 2 (a). Accordingly, the pixel coordinates of  $\mathbf{x}'$  are given by

$$\mathbf{x}' = \begin{bmatrix} \cos(\varphi)r(\theta) \\ \sin(\varphi)r(\theta) \end{bmatrix}.$$
(1)

The mapping of  $\mathbf{x}'$  in Cartesian image coordinates to  $\mathbf{x}$  in pixel coordinates is usually described as an affine transformation

$$\mathbf{x} = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} c & s & u_0 \\ 0 & c & v_0 \end{bmatrix} \begin{bmatrix} \mathbf{x}' \\ 1 \end{bmatrix}$$
(2)

10 with the parameter of the interior orientation, namely the principal distance c, the shear factor s and the principal point  $\mathbf{x}_0 = (u_0, v_0)^{\top}$ , i.e. the projection center in pixel coordinates. Note that the origin of the sensor coordinate system lies in the upper left corner of the image as depicted in Figure 2 (b).

Due to lens imperfections real camera projections do not follow a projection model perfectly. In our case the lenses of the sky imagers are shielded by an additional glass dome which additionally refracts the light before it enters the lens. Radial symmetric distortions result in either a barrel or pillow like stretching or bending of the image with increasing distance from the projection center. Such distortions can be compensated by adding even powered polynomials to the radial distance function

the projection center. Such distortions can be compensated by adding even-powered polynomials to the radial distance function following Brown (1971)

$$\Delta u = L(\hat{r})\hat{u} = A_1 \,\hat{u} \,\hat{r}^2 + A_2 \,\hat{u} \,\hat{r}^4 + A_3 \,\hat{u} \,\hat{r}^6 \tag{3}$$

$$\Delta v = L(\hat{r})\hat{v} = A_1\,\hat{v}\,\hat{r}^2 + A_2\,\hat{v}\,\hat{r}^4 + A_3\,\hat{v}\,\hat{r}^6 \tag{4}$$

20 with  $\hat{u} = u - u_0$ ,  $\hat{v} = v - v_0$  and  $\hat{r} = \sqrt{\hat{u}^2 + \hat{v}^2}$ .

In summary, we formulate the mapping into (distorted) image point coordinates  $\tilde{\mathbf{x}}$  on the sensor plane as

| $\tilde{u} = c \cos \phi r(\theta) + s \sin \phi r(\theta) + u_0 + \Delta u$ |                    | (5) |
|--|--------------------|-----|
| $\tilde{v} = c \sin \phi r(\theta)$  | $+ v_0 + \Delta v$ |     |

and the reverse mapping of image point x in pixel coordinates to the 3D direction vector  $\mathbf{x}^{''}$  with unit length is given by

| $\begin{bmatrix} x \end{bmatrix}$   |   | $\cos\phi\sin\theta$                    |   | $\left[\frac{x'}{r}\cos r\right]_{u'}$          |
|-------------------------------------|---|---|---|---|
| $\begin{bmatrix} y\\z\end{bmatrix}$ | = | $\frac{\cos\phi\cos\theta}{\sin\theta}$ | = | $\left[\frac{\frac{g}{r}\cos r}{\sin r}\right]$ |





20

25

where x' and y' are normalized image coordinates

$$x' = \frac{u - \Delta u - u_0 - s(u - u_0)}{c}$$
$$y' = \frac{v - \Delta v - v_0}{c}$$

and  $r = \sqrt{x'^2 + y'^2}$  the respective value of the radial projection function  $r(\theta)$ . The equidistant projection  $r(\theta) = \theta$  fits the 5 projection of our sky imagers best. We will describe the calibration procedure to determine the parameters of interior orientation with distortion parameters as well as the calibration of a stereo camera pair in Sec. 3.5.

# 3.2 Exterior Orientation and Epipolar Geometry

The omnidirectional camera model refers to the local camera coordinate system with the projection center as the origin and the sensor plane defining its orientation. The exterior orientation of a camera, which consist of three rotation angles and three

10 translation shifts, is described in a common world reference system  $\Omega_W$  and allows to derive the geometric relationship between two or more cameras. We choose one camera as the reference camera, which is considered as the left camera and the other as the right camera, which simplifies the following notation and avoids misconceptions. The choice of the reference camera has no impact on the reconstruction results.

Figure 3 illustrates the principal stereo configuration with hemispheric cameras, making the world reference system  $\Omega_W$  and 15 the two camera reference systems  $\Omega_L$  and  $\Omega_R$  explicit. Let  $\mathbf{C}_{\mathbf{L}}$  be the world coordinates of the left camera and  $\mathbf{P}_{\mathbf{L}}$  an object point in the left camera reference frame. The transformation of  $\mathbf{P}_{\mathbf{L}}$  into world coordinates reads then as

$$\mathbf{P} = (R_L \, \mathbf{P_L}) + \mathbf{C_L} \tag{6}$$

with the rotation matrix  $R_L = R_x(\alpha_L) R_y(\beta_L) R_z(\gamma_L)$  and  $\mathbf{C}_L \in \mathbb{R}^3$ . Here  $\alpha_L, \beta_L$  and  $\gamma_L$  are the Eulerian angles (roll, pitch, yaw) and  $R_x(\alpha_L), R_y(\beta_L)$  and  $R_z(\gamma_L)$  the respective rotation matrices. Considering  $R_L$  and  $R_R$  as the rotation matrices and  $\mathbf{C}_L$  and  $\mathbf{C}_R$  as the world coordinates of the left and right camera, we obtain the relative orientation from the left to the right camera with rotation matrix R and the baseline vector  $\mathbf{t}$  with

$$R = R_L^{\top} R_R \quad \text{and} \quad \mathbf{t} = R_L^{\top} \mathbf{C}_{\mathbf{R}} - \mathbf{C}_{\mathbf{L}} \,. \tag{7}$$

The determination of an accurate relative pose is crucial, as errors in the estimated exterior orientations may sum up to larger errors in the relative orientation which compromises the image correspondence algorithm and the triangulation of 3D point coordinates as investigated by Hirschmüller and Gehrig (2009).

The two camera centers  $C_L$ ,  $C_R$  and the object point P span the epipolar plane. This geometry can be expressed with the coplanarity equation and holds when

$$\mathbf{x}_{\mathbf{L}}^{''} E \mathbf{x}_{\mathbf{R}}^{''} = 0 \quad \text{with} \quad E = [\mathbf{t}]_{\times} R,$$
(8)

where E is the essential matrix obtained by a matrix multiplication with the skew symmetric matrix  $[t]_{\times}$  of t.





5

20

Given two direction rays  $\mathbf{x}_{\mathbf{L}}^{''}$  and  $\mathbf{x}_{\mathbf{R}}^{''}$  of corresponding image points  $\mathbf{x}_{\mathbf{L}}$  and  $\mathbf{x}_{\mathbf{R}}$  in the left and the right camera images,  $E^{\top}\mathbf{x}_{\mathbf{L}}^{''}$  defines the normal vector of the epipolar plane spanned by  $\mathbf{x}_{\mathbf{L}}^{''}$  and  $\mathbf{t}$ , and hence requires its correspondence  $\mathbf{x}_{\mathbf{R}}^{''}$  to lie on the intersection circle between the epipolar plane and the image sphere. The same holds for  $E\mathbf{x}_{\mathbf{R}}^{''}$  and  $\mathbf{x}_{\mathbf{L}}^{''}$ . In case of deviations from this constraint, Eq. (8) defines the cosine of the angle between the two epipolar planes spanned by  $\mathbf{t}$  together with  $E^{\top}\mathbf{x}_{\mathbf{L}}^{''}$ and  $E\mathbf{x}_{\mathbf{R}}^{''}$  respectively, which we are using as an error measure to estimate E using image point correspondences.

The essential matrix E can be expressed with five parameters, therefore has five degrees of freedom (two for the baseline direction t and three for the rotation angles). We use an adapted version of the direct method of Longuet-Higgins (1981) to compute E with eight correspondences, which exploits that the eigenvalues of E are  $\lambda_1 = \lambda_2 = 1$  and that we use spherically normalized ray directions.

- The estimated essential matrix E can be decomposed to obtain the relative rotation R and baseline vector  $\mathbf{t}$ , which can be used for triangulation and epipolar image rectification, as described in Sec. 3.3. However, a 3D reconstruction based on R and  $\mathbf{t}$  alone takes place only in the coordinate system of the reference camera and only up to scale because of the scale ambiguity. A meaningful reconstruction in world coordinates requires the absolute length of the baseline (distance between the cameras) and the absolute orientation of the reference camera. Sec. 3.5 presents methods to estimate both, the parameters of exterior and  $\mathbf{t}$  interior estimates and  $\mathbf{t}$  interior estimates of the reference camera.
- 15 interior orientation.

# 3.3 Epipolar Rectification

Once the epipolar geometry and the interior orientation is known, the input images can be transformed so that corresponding image points lie on the same image row, which reduces the search for corresponding image points from two dimensions (image) to one (image row). The epipolar rectification allows to rectify a fisheye image over a broad spectrum of the angle  $\theta$ , which allows to use the complete image content of a fisheye image. However, epipolar rectification leads to lower accuracies at the margins as the image content is stretched in these areas, cf. to Schneider et al. (2016). The epipolar rectification scheme following Abraham and Förstner (2005) is shown in Figure 4.

Note, that the following formulas with respect to  $\beta$  and  $\psi$  are only valid for an epipolar rectified image pair. For each real camera we can define a virtual camera, such that the virtual cameras are in a canonical stereo setup, i.e. have a common *x*-axis, equal orientation  $R_{L,V} = R_{R,V} = I$  and are only shifted along the virtual *x*-axis  $\mathbf{t}_V = (\|\mathbf{t}\|, 0, 0)^{\top}$ . An object point  $\mathbf{P}_{\mathbf{V}}$  in the virtual world coordinate system is then defined by the angle  $\beta$  which denotes the respective epipolar plane, and the two angles  $\psi_L$  and  $\psi_R$  that define the angle of the projection ray within the epipolar plane, see Figure 4 (b). Based on this geometry, we are able to define a rectification scheme, see Figure 4 (d), to cover the whole 3D space: the image rows correspond to the angle  $\beta$ , which represents the rotation of the epipolar plane, while the image columns represent the respective angles  $\psi_L$  and  $\psi_R$  in 30 the rotated epipolar plane

$$\mathbf{x'}_{V} = \begin{bmatrix} \psi \\ \beta \end{bmatrix} = \begin{bmatrix} \operatorname{atan2}(z_{V}, y_{V}) \\ \operatorname{atan2}(x_{V}, y_{V}) \end{bmatrix} \quad \text{with} \quad \mathbf{x'}_{\mathbf{V}} = (x_{V}, y_{V}, z_{V})^{\top}.$$





5

Let  $\mathbf{x}_{\mathbf{L}}^{''}$  and  $\mathbf{x}_{\mathbf{R}}^{''}$  be the projection rays of an object point **P** in the left and the right camera coordinate system  $(\Omega_L, \Omega_R)$ , and  $\mathbf{x}_{\mathbf{V},\mathbf{L}}^{''}$  and  $\mathbf{x}_{\mathbf{V},\mathbf{R}}^{''}$  the corresponding projection rays in the virtual coordinate systems  $(\Omega_{L,V}, \Omega_{R,V})$ . The mapping between the real and virtual coordinate system follows a two-step procedure: In the first step,  $\mathbf{x}_{\mathbf{L}}^{''}$  and  $\mathbf{x}_{\mathbf{R}}^{''}$  are mapped from the local camera coordinate systems  $(\Omega_L, \Omega_R)$  to the world coordinate system  $\Omega_W$ . If we do not have knowledge about the world coordinate system  $\Omega_W$ , we choose  $\Omega_W = \Omega_L$ . Given the essential matrix E, we are able to extract the rotations  $R_L = I$  and  $R_R = R$ , which map from camera coordinates to world coordinates. This leads to an equal coordinate system orientation, see step 1 in Figure 4 (a). In the second step, we have to construct an appropriate rotation matrix  $R_V$  in order to align each systems x-axis

with the baseline t, see step 2 in Figure 4 (a).

Since the matrix columns of  $R_V$  are the images of the base vectors, the first column is the normalized baseline vector. We 10 can freely choose the other two coordinate axes as long as they form an orthonormal system, because each realization aligns the x-axis with the baseline. This means, that the rectification scheme is defined up to a rotation about the baseline t, which corresponds to a shift of the range of the angle  $\beta$  and a vertical translation in the rectified image. We define the virtual y-axis in the x-y-plane of the world coordinate system, which also determines the virtual z-axis.

Thus we finally get

$$\mathbf{e_{1}} = \mathbf{t} \| \mathbf{t} \|^{-1},$$

$$R_{V}^{-1} = [\mathbf{e_{1}}, \mathbf{e_{2}}, \mathbf{e_{3}}] \quad \text{with} \quad \mathbf{e_{2}} = (-y_{T}, x_{T}, 0)^{\top} \| (-y_{T}, x_{T}, 0) \|^{-1},$$
15 
$$\mathbf{e_{3}} = \mathbf{e_{1}} \times \mathbf{e_{2}}.$$

Given the angular information  $\beta$ ,  $\psi_L$  and  $\psi_R$  as well as principal distance c, we get the final rectified image coordinates

 $\begin{array}{ll} u_V &= c \, \psi + u_{0,V} \\ v_V &= c \, \beta + v_{0,V} \end{array} \hspace{1.5cm} \text{with} \hspace{1.5cm} u_{0,V} = c \, \pi/2 \hspace{1.5cm} \text{and} \hspace{1.5cm} v_{0,V} = c \, \pi/2 \, . \end{array}$ 

The reverse mapping, from rectified image coordinates to world coordinates is given by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R_V \begin{bmatrix} \sin(u_V^*) \\ \cos(u_V^*) \sin(v_V^*) \\ \cos(u_V^*) \cos(v_V^*) \end{bmatrix} \quad \text{where} \quad \begin{array}{l} u_V^* = (u_V - u_{0,V})/c \\ v_V^* = (v_V - v_{0,V})/c. \end{array}$$
(9)

#### 20 3.4 Triangulation for 3D-Reconstruction

Having corresponding image points  $\mathbf{x}_{\mathbf{L}}$  and  $\mathbf{x}_{\mathbf{R}}$  identified in the image rows of the epipolar rectified images,  $\mathbf{x}_{\mathbf{L}}^{''}$  and  $\mathbf{x}_{\mathbf{R}}^{''}$  can be directly derived from  $\mathbf{x}_{\mathbf{L},\mathbf{V}}$  and  $\mathbf{x}_{\mathbf{R},\mathbf{V}}$  using the reverse mapping of the rectification scheme of Eq. (9). Due to the rectification the ray directions are guaranteed to lie in the 3D epipolar plane and do intersect. Considering the geometry shown in Figure 5 we identify the relation  $s \cdot \sin(\psi_R) = b \cdot \sin(\gamma)$ . With  $\gamma = \psi_L - \psi_R$  and  $\mathbf{P} = s \cdot \mathbf{x}_{\mathbf{L}}^{''}$  we have

$$\mathbf{P} = b \left( \frac{\sin(-\psi_R)}{\sin(\psi_L - \psi_R)} \right) \mathbf{x}''_{\mathbf{L}}.$$
(10)



15

20



## 3.5 Calibration and Parameter Estimation

In the following we present our procedure to estimate the parameters of the interior and exterior orientation, introduced in Sec. 3.1 and 3.2.

# 3.5.1 Parameters of the Interior Orientation

5 For the estimation of the parameters of the interior orientation in Eq. (5) we employ a test field with markers that encode a geometric relationship. Such a test field can be a sophisticated setup in a laboratory (Seiz, 2003) or - as in our case - a pattern printed or fixed on a plane or inside an open half-cube as depicted in Figure 6. The calibration generally proceeds in two steps: The first step provides sample images of the pattern in different poses covering the field of view. In each image an image processing routine detects and extracts the image coordinates of the pattern geometry. In the second step, the extracted image 10 points are used to estimate the optimal parameters of the camera model with an adjustment procedure.

We employ a software developed by Abraham and Hau (1997), that accepts input images of a calibration cube with a fixed white dotted pattern. Each inner cube side has a fingerprint pattern to make sure the detected dots are properly identified as lying on the x-, y- or z-plane, which determines their corresponding absolute 3D coordinates with respect to the cube reference system. The extraction stage results in a set of correspondences between 2D image points and 3D cube points, which then are used in a nonlinear bundle adjustment that iteratively minimizes the reprojection error between the observed image points and

the reprojections of the 3D points of the pattern using the respective parameter estimation according to Eq. (5) and (6).

### 3.5.2 Parameters of the Exterior Orientation

First, we describe how to estimate the absolute location and orientation of each camera in the world reference system. This information can then be used to derive a first estimate of the essential matrix  $E^*$ , which will then be iteratively refined using point-feature correspondences obtained from the stereo images according to the epipolar constraint in Eq. (8).

Employing a satellite navigation system like GPS allows us to derive the geographic position of the cameras with an accuracy of about 2–3 m. Accuracies in the range of centimeters can be achieved by using additional correction information broadcasted by terrestrial reference stations (D-GPS). The obtained coordinates can be mapped from the global reference system, e.g. WGS-84, to a local reference system using a suitable projection in order to get the exact baseline length and vector.

A more challenging task is the estimation of the orientation of the cameras. Hu et al. (2010) uses geographic landmarks with known coordinates, Öktem et al. (2014) use the horizon and Seiz (2003) exploits stars as geometric references. As Seiz, we use sensed stars in the images, see Figure 7, as observations to estimate the absolute camera orientations.

This requires a set of reference stars which can be observed by the cameras in the local night sky. The coordinates of the stars can be obtained from a star catalog like Stellarium or from online sources, e.g. of the NAOJ<sup>1</sup>. The coordinates are usually provided with the (north-aligned) azimuth angle  $\varphi_n$  and altitude angle  $\theta$  and have to be converted to 3D unit vectors according

<sup>&</sup>lt;sup>1</sup>http://eco.mtk.nao.ac.jp/cgi-bin/koyomi/cande/horizontal\_rhip\_en.cgi, last accessed April 2016.





to

5

$$\mathbf{x}_{\mathbf{s}} = \begin{bmatrix} \varphi_n \\ \theta \end{bmatrix} \longrightarrow \mathbf{x}''_s = \begin{bmatrix} \cos(\varphi_n + \pi/2)\cos(\theta) \\ -\sin(\varphi_n + \pi/2)\cos(\theta) \\ \sin(\delta) \end{bmatrix}.$$

In order to get the coordinates of the each reference star in the recorded image, we first take several long-exposure night sky images, compute the median image and subtract the median image from the original night sky images. As a result, only the moving stars are left and we can compute the respective centroid coordinates. The correct identification of the stars in the image is currently done manually by adjusting the rotation angles  $\alpha$ ,  $\beta$  and  $\gamma$  until the projections are close enough to the centroids to be attributed. After the conversion of the stars image coordinates and catalog coordinates to 3D unit vectors ( $x_s''$  and  $x_{cat}''$ ), they can be used to estimate the rotation  $R_{abs}$  of the camera by a Levenberg-Marquardt minimization (Madsen et al., 2004) of the angular error

10 
$$\sum_{i \in stars} (1 - (\mathbf{x}_{cat}^{''} (R_{abs} \mathbf{x}_{s}^{''})))^2 \longrightarrow \text{minimize},$$

where  $R_{abs}$  can be parametrized as a unit quaternion or as axis-angle representation.

From the absolute location and orientation of the cameras we can derive the relative orientation using Eq. (7) and a first estimate  $E^*$  of the essential matrix can be composed according to Eq. (8).

For a further refinement of the essential matrix E\*, we collect SIFT-point-feature correspondences (Lowe (2004)) that are consistent with the epipolar constraint in Eq. (8): For each detected feature in the left image, we can select all features in the right image that are consistent with E\* up to a predefined error threshold, e.g. ∠(E\*<sup>T</sup>x<sup>''</sup><sub>L</sub>, E\*x<sup>''</sup><sub>R</sub>) < 2°, and then find the best match via K-Nearest-Neighbor using the SIFT feature descriptor. The same is done in the other direction, i.e. from the right to the left image, so that only mutually consistent matches are selected.</p>

A couple of image pairs are enough to collect plenty of evenly distributed correspondences, as is shown in Figure 8. Since this set of correspondences will contain mismatches that would lead to a flawed refinement of  $E^*$ , we use the robust parameter estimation technique RANSAC (Fischler and Bolles, 1981) to filter out those likely mismatches.

Finally, we employ Levenberg-Marquardt minimization of the cost function

$$\sum_{i \in \text{inliers}} \sin^2 \left( \measuredangle(\mathbf{x}_{\mathbf{L}}^{''}, \hat{\mathbf{x}}_{\mathbf{L}}^{''}) \right) + \sin^2 \left( \measuredangle(\mathbf{x}_{\mathbf{R}}^{''}, \hat{\mathbf{x}}_{\mathbf{R}}^{''}) \right) \longrightarrow \text{minimize} \,.$$
(11)

Because the observations  $\mathbf{x}''_{\mathbf{L}}$  and  $\mathbf{x}''_{\mathbf{R}}$  are always subject to some measurement error, they will not lie exactly within an epipolar

25 plane.  $\hat{\mathbf{x}}_{\mathbf{L}}^{''}$  and  $\hat{\mathbf{x}}_{\mathbf{R}}^{''}$  denote the estimated true locations of  $\mathbf{x}_{\mathbf{L}}^{''}$  and  $\mathbf{x}_{\mathbf{R}}^{''}$  that do lie exactly within an epipolar plane and at the same time are closest to the observations in an angular sense. As the estimations and the observations are unit vectors and lie on the image sphere, Eq. (11) formulates a meaningful angular error measure and its minimization provides an optimal maximum likelihood solution, see Oliensis (2002).





25

## 4 Stereo Matching

To calculate the 3D information of a point  $\mathbf{P}$ , we need to know the coordinates of the projected point on the both images planes. Only if such correspondences are known, the 3D information can be computed. The aim of stereo matching algorithms is to compute such correspondence information.

- 5 The visual appearance of a scene point P in each camera determines if stereo matching is successful or not. Automatic stereo matching is likely to fail if there are occlusions, specular reflections, varying illumination or large scale and pose differences between the images, so that either corresponding object points are not visible in both images or differ significantly in their appearance with respect to shape and size. Also objects may lack sufficient texture or contrast, or a unique surface does not exists that has a consistent visual appearance when observed from different perspectives. Especially the latter poses a problem 10 in cloud photogrammetry, and depending on the cloud situation stereo reconstruction has limitations.
- In practice, one aims at finding the correspondences between distinct points in the images (sparse stereo) or between all pixels (dense stereo). A good overview is given in Scharstein and Szeliski (2002).

We only employ sparse stereo during the estimation of the essential matrix (Figure 8) as described in Sec. 3.5.2.

- Dense stereo is advantageous when dealing with complex geometries, but also effectively delivers reasonable results for 15 image regions with low-contrast. In these cases, information from high-contrast regions is propagated into the low-contrasts regions assuming similar depth at nearby pixels with similar intensity or structural values. To simplify the search for correspondences, dense methods usually require epipolar rectified images, see Sec. 3.3. As a result of that, corresponding pixels are restricted to lie on the same image row, which reduces the search space from 2D to 1D.
- The correspondence information is stored in the so-called *disparity map* D, that contains for each pixel in the rectified reference image the horizontal sub-pixel distance d to its corresponding image point shifted in the same row in the other image, see Figure 9. Hence, for the two corresponding points  $\mathbf{x}_{L,V}$  in the left and  $\mathbf{x}_{R,V}$  in the right image, we have for each pixel position in the disparity map  $D(\mathbf{x}_{L,V}) = |u_L - u_R|$  and therefore have the relation  $\mathbf{x}_{R,V} = (u_R, v_R)^\top = (u_L - d, v_R)^\top$ .

In our current approach, we rely on a dense matching algorithm that is based on the Semi-Global Matching (SGM) proposed by Hirschmüller (2005) and is called Semi-Global Block-Matching (SGBM). It produces accurate results while being deterministic and fast to compute. In this work we use the implementation provided in OpenCV.

For a detailed algorithmic description, we refer to the original paper by Hirschmüller (2005) or to the OpenCV documentation. Here, we present only a short summary.

The basic problem of finding an optimal disparity map can be formalized as an energy minimization problem involving an energy functional and an appropriate minimization technique (Scharstein and Szeliski, 2002). A good disparity map should satisfy at least the following two aspects:

- 1. Two corresponding pixels should have similar intensity or structural values (data consistency).
- Neighboring pixels with similar intensity or structural values should have similar disparity values (smoothness assumption).





Both aspects can then be combined to form the global energy  $E(D) = E_{data} + E_{smooth}$  and are realized in SGBM as follows: To achieve data consistency (aspect 1),  $E_{data}(\mathbf{x})$  is computed for each pixel  $\mathbf{x}$  independently using a window-based similarity measure such as sum of absolute/squared differences or normalized cross-correlation (NCC), yielding one matching cost per pixel for each valid disparity value  $d \in [d_{min}, d_{max}]$ . Note, that using a larger window will smooth the disparity map since

5

20

small details have a smaller influence on the measure. This causes fine structures to disappear, but also reduces errors caused by image noise. However, relying only on a minimum in  $E_{data}$  will cause mismatches, especially in regions with low- contrast or repeating patterns, which results in a flawed disparity map. This can be encountered by introducing an additional smoothness term  $E_{smooth}$ , that penalizes larger disparity differences of neighboring image points across eight linear paths.

Figure 9 illustrates the computation of  $E_{data}$  using epipolar rectified images and the final disparity map after incorporating 10 also  $E_{smooth}$ . One can clearly see the difference in the disparity values between clouds that are closer (high disparity) and more distant clouds (smaller disparity).

In our application, we use a window-size of  $11 \times 11$  pixels. To achieve a successful matching in larger low-contrast regions and reduce the variability in the reconstruction due to the noisy image signal, we furthermore scale the input images to one quarter size. This causes an oversmoothing near cloud boundaries, but this way we obtain smoother cloud surfaces.

# 15 5 Stereo Setup and Results

In this section we present our stereo setup at the Forschungszentrum Jülich GmbH, Germany. We also give a geometric uncertainty analysis of the current setup, discuss common error sources and how they affect the calibration and reconstruction results, with a focus on asynchronous recordings. Finally, we present our experimental results of two case studies. The first case shows that our dense stereo approach is able to achieve a geometric accuracy that is comparable with those of previous studies using sparse stereo methods, like Seiz (2003) and Öktem et al. (2014). The second case illustrates the capability of our

approach to reconstruct the complex 3D cloud structure and dynamics of convective clouds.

### 5.1 Camera Setup

We exploit two sky imagers installed at the Forschungszentrum Jülich GmbH, Germany, which also hosts the Jülich Observatory for Cloud Evolution (JOYCE), cf. to Löhnert et al. (2014) for more information. For the evaluation of our results we use observations from a local lidar-ceilometer and cloud radar (Figure 10). The first cameras is located at 50.90849° N, 6.41342° E and the second at 50.90613° N, 6.41144° E, resulting in a baseline length of approximately 300 m. This rather short distance results in higher geometric uncertainty of the estimated 3D points on the clouds, but reduces occlusions and enhances the ratio of mutually visible cloud regions in both images. Furthermore, the short baseline increases the similarity of the cloud appearance in both images, which is beneficial for stereo matching.

30 Both cameras are IDS network-cameras of type uEye GigE UI-2280SE with a 2/3" CCD sensor consisting of 2448×2048 pixels and have a Fujinon FE185C057HA-1 C-Mount Fisheye adapter, providing a 185° field of view, and have a fixed focus. The cameras are mounted in a box with and point towards the sky. An acrylic glass dome protects the cameras against environmen-





tal effects. A power supply and a fan distribute heat to prevent the condensation of water on the glass dome. Each camera is connected to a small computer that hosts a self-developed camera control application, which bases on the IDS C++ SDK (IDS, 2013). and allows us to control the cameras remotely, e.g. for scheduled recordings with settings as exposure time, recording interval (e.g. 15 seconds) or modes like long-exposure (night mode) or High Dynamic Range (HDR). The images are currently saved locally and transferred if needed. Synchronization is done by frequent requests to a local NTP service.

5

20

### 5.2 Geometric Uncertainty

First, we compare the spatial resolution of a sky imager with a wide-angle camera used in previous studies.

Fisheye lenses cover a substantially larger field of view than normal perspective lenses, but at an reduced effective angular resolution. As a consequence, the stereo depth resolution is lower for fisheye lenses compared to perspective ones. This

- 10 drawback limits the effective range for a high quality reconstruction, especially for distant clouds. A comparison of our fisheye cameras with one of the wide- angle cameras from Öktem et al. (2014) highlights the differences: While our camera has a field of view (FOV) of about 180° and the circular view field covers a 3.5 Megapixel region (from a 5 Megapixel sensor), the wide-angle camera has a FOV of 67° and takes 1-Megapixel images from a 5-Megapixel sensor. To compare the view fields, the respective solid angle  $\Omega_{fish}$  for the sky imager and  $\Omega_{wa}$  for the wide angle camera given in steradians must be
- 15 derived. Assuming a field of view of  $180^{\circ}$  for the sky imager and  $67^{\circ}$  for the wide-angle camera leads to  $\Omega_{fish} = 6.28$  and  $\Omega_{wide} = 1.04$ .

Further, the solid-angle per pixel is  $6.28/3.5 \cdot 10^{-6} = 1.8 \cdot 10^{-6}$  for the fisheye and  $1.04/1.0 \cdot 10^{-6} = 1.04 \cdot 10^{-6}$  for the wide angle camera, resulting in a 43 % lower spatial resolution of the fisheye camera. Using the full resolution of the wide-angle camera with 5 Megapixel, the ratio would be 11 %. Hence, one must use a sensor almost 10 times higher resolution to compete with the wide-angle camera in this respect. Depending on the fisheye projection function and the location in the image, the

spatial resolution will of course vary due to the different degree of distortion.

Next, we demonstrate the spatial variability of the 3D reconstruction error based on our real stereo setup by assuming a constant error in the disparity map. Based on the distance between our cameras, we can estimate the uncertainty of the reconstruction using the triangulation method presented in Sec. 3.4. Figure 11 shows the reconstruction error for a virtual

- 25 cloud layer at 1500 m and 3000 m height assuming an error of 1 pixel in the disparity map after the matching phase, cf. Sec. 4, which corresponds to a directional error of approximately  $0.1^{\circ}$ . The values represent absolute errors within the epipolar plane. Therefore, depending on the horizontal distance to the cameras, the error has a larger vertical component (small distance) or a larger horizontal component (larger distance). The error grows larger in two cases: First, with increasing distance to the cameras, and second, with increasing co-linearity of the object point and the camera centers. In both cases, the angle  $\gamma$  between
- 30 the projection rays becomes very small, yielding larger triangulation errors. Thus, the sky imagers do not provide a hemispheric 3D reconstruction with homogeneous accuracy. However, the accuracy can be ameliorated by employing a third camera in a triangle configuration.

The imaging process of a sensor adds a random noise signal, which can be limited, but not avoided. In principle, this also effects parameter estimation, because both localization and measurement are disturbed. Given a large number of measurements





for the calibration, the signal noise can be compensated in a maximum likelihood estimation as the redundancy is high. The stereo matching is also effected by the noisy image signal and causes a disturbed 3D reconstruction as shown in Figure 11.

The following analysis is designed to investigate the effects of an asynchronous recording of the stereo images during the observation of a dynamic cloud scene.

- 5 Despite frequent requests to an NTP service, we sometimes experience asynchronous system times on the local computers in the range of a few seconds. Consequently, the whole cloud scene shifts between the single shots and thus causes a displacement in the images, which leads to a biased or flawed disparity map. We investigate the effects of a cloud field displacement of  $\Delta = \pm 15$  m along the baseline (x-direction) and perpendicular to the baseline (y-direction) in a virtual sky imager setup together with a virtual cloud layer at 3 km, using the 3D rendering software Blender citepblender. The virtual sky imagers have identical
- 10 internal camera geometries comparable to the real ones, and a relative pose of  $R_L = R_R = I$  and  $\mathbf{t} = (300, 0, 0)^{\top}$ m. Figure 12 shows the cross-sections of the respective reconstruction along the baseline (x-axis) and perpendicular to the baseline (y-axis). A displacement along the x-axis results in a lower (2875 m for  $\Delta = -15$  m) or a higher (3183 m for  $\Delta = +15$  m) cloud base compared to the unaffected reconstruction (3025 m for  $\Delta = 0$  m), while a displacement along the y-axis just causes an overall higher standard deviation of the reconstruction without a systematic error in the mean base height ( $\sigma(\Delta = +15 \text{ m}) = 45 \text{ m}$ ,
- 15  $\sigma(\Delta = -15 \text{ m}) = 48 \text{ m}$  and  $\sigma(\Delta = 0 \text{ m}) = 35 \text{ m}$ ). The results confirm that a displacement in x-direction is equivalent to a change in the length of the baseline t. Hence, the reconstructed cloud base  $\hat{h}$  compared to the real cloud base h can be derived according to  $\hat{h} = h \cdot [\|\mathbf{t}\|/(\|\mathbf{t}\| + \Delta)]$ .

### 5.3 Evaluation of the 3D Reconstruction

The following two case studies are designed to evaluate our approach using observations from a lidar-ceilometer and a cloud 20 radar, and to show its capability to capture the complex 3D shapes and dynamics of convective clouds.

We compare some of the reconstructions with observations of cloud base heights from a lidar-ceilometer and reflectivity measurements by a cloud radar, both deployed in the vicinity of the cameras at the JOYCE observation site. The lidar-ceilometer is a Vaisala CT25K which was operated with a range between 60 m and 7500 m and an angular resolution of 0.84°, which corresponds to 4.1 m at 3000 m height. The cloud radar is a Metek polarimetric Doppler Radar (MIRA) operating at 35 GHz

- 25 with a similar angular resolution over a range between 150 m and 15 km and a maximum sensitivity of -45 dBZ at 5 km. For our comparisons we average observations over an integration time of 0.15 seconds. For the empirical evaluation, we use the cloud base height observations from the lidar-ceilometer. Since the lidar-ceilometer offers only point measurements every 15 seconds, we use observations from a 10 minute period around the measurement time of the cameras to get a meaningful comparison. We compare the mean reconstructed height values from a near-zenith rectangular area of 9 km<sup>2</sup>. The cloud radar
- 30 offers a direct comparison between the reflectivity signal from an RHI cross-section scan and the respective cross-section of the 3D cloud shape derived by the stereo method. In order to cover the region of best geometric accuracy of the stereo method, we orientate the radar scans towards a almost perpendicular direction to the baseline vector, compare Figure 10 and 11). After image acquisition, the input images are preprocessed, which includes a resampling to one quarter size and a contrast enhancement using the *Contrast Limited Adaptive Histogram Equalization* (CLAHE). For stereo matching with SGBM we use





a window-size of  $11 \times 11$  pixels and a cloud mask to remove some artifacts. After triangulation of the 3D point cloud, we create a cloud surface mesh using methods from the open-source *Point Cloud Library* (Library, 2016).

### 5.3.1 Analysis of the 3D Reconstruction of Stratocumulus-layer Clouds

We present two cases with stratocumulus clouds, which we will use to evaluate our result by comparing it against observations 5 from the lidar-ceilometer and the cloud radar.

Figure 13 shows the results from the 11<sup>th</sup> of August, 2014, at 14:12:00 UTC, right after a shower moved over the JOYCE site with a trailing stratocumulus layer, that was also captured by the cloud radar. The mean cloud-base heights from the reconstructions are 2881 m while the lidar-ceilometer observations result in 2897 m. Note also, that the cloud radar is – due to its measuring frequency in the microwaves – not as sensitive as the lidar-ceilometer or the camera to the visible outer cloud

10 boundaries, with the consequence that especially the boundaries of a cloud might not be detectable by the radar, but by the camera or the lidar.

The second case has been recorded on the 5<sup>th</sup> of August, 2014, which is interesting for two reasons: First, a typical boundary layer evolution has been observed, starting with small to medium size cumulus convection along with a steady rise of the cloud base. Second, from early noon on the cloud scenery shifted towards stratocumulus layers, resulting in less convection

15 and thus providing again good conditions for a comparison against observations from the lidar-ceilometer and the cloud radar. The results shown in Figure 14 are organized as described in the previous case. Comparing the reconstruction result with the cloud radar indicates a reasonable estimation of the bases of the higher cloud layer and the cumulus cloud below it. However, differences between 100 to 150 m are noticeable and the lidar-ceilometer observations report a cloud-base mean height of 2922 m compared with the height of 2766 m derived from stereo reconstruction. We assume a systematic error in the

20 reconstruction, which is discussed in Sec. 5.3.3.

# 5.3.2 Analysis of the 3D Reconstruction of Cumulus Clouds

The 24<sup>th</sup> of July, 2014 showed strong convection with rapid cloud development and decline, providing excellent conditions to apply our dense stereo reconstruction and to capture their complex shapes and dynamics. Figure 15 shows a cumulus mediocris forming approximately 3 km away from the stereo camera pair. One can observe the convective updraft and the rising cumulus

25 turret. While reaching a height of about 4000 m, the turret enters the higher wind field, resulting in a skew shape of the cloud due to the wind shear. Figure 16 shows a cross-section of the reconstructed 3D cloud surface at 12:07:00 UTC together with the cloud base measured with the the lidar-ceilometer.

Figure 17 shows another example of a smaller convective cloud, but also with a rather complex morphology from 11:28:00 UTC to 11:32:00 UTC on the same day. One observes how the developing convective turret covers parts of the cloud, which results

30 from its increasingly concave shape. The temporally closest cloud base height values of the lidar-ceilometer are between 12:14:38 UTC and 12:17:08 UTC and report 1487 m on average. Figure 18 shows a cross-section of the reconstruction the same way as in Figure 16.





#### 5.3.3 Discussion of the Experimental Results

Our cloud reconstructions show an overall good agreement with the cloud base height observations from the lidar-ceilometer and the cloud radar. Mean near-zenith cloud base heights for the stratocumulus cases are within 1% for the 11<sup>th</sup> of August and 5% for the 5<sup>th</sup> of August of lidar-ceilometer mean values, and the stereo method is able to capture the geometric shapes
of the cloud bases as in Figure 14. A possible explanation for the deviation from the 5<sup>th</sup> of August is a shift between the computers system time as is described in Sec. 5.2. Although errors in the orientation parameters cannot be excluded, the relative orientation estimation shows a standard deviation of 0.04° and thus is comparably accurate. On the other hand, a bias in the computers system time could be observed at times and confirms this assumption. Based on the observations from the nearby wind-lidar, which reported a wind speed of about 5 m s<sup>-1</sup> in a direction almost collinear to the baseline, a time
difference of about +3 seconds between the left and the right camera would lead to a scaled reconstructed cloud base of 2922 m · (302 m/317 m) = 2783 m which is close to the actual reconstructed cloud base height of 2766 m.

The results from the 24<sup>th</sup> of July show that the dense stereo method is able to almost fully reconstruct the visible outer shape of a convective cloud. In both cases the concave and increasingly skew shape of the cloud is nicely captured by the dense stereo method, as is illustrated in the cross-sections in Figure 16 and Figure 18. Also the cloud base is clearly visible and

15 matches the lidar-ceilometer value of 1495 m quite well, considering that temporally close measurements are only available from 12:12:07 UTC to 12:17:08 UTC ranging from 1448 m to 1585 m.

# 6 Conclusions

In this paper, we investigated the problem of computing a 3D reconstruction from fisheye cameras using dense stereo approaches from photogrammetry. We presented a complete approach for stereo cloud photogrammetry using hemispheric sky imagers. Our approach combines calibration, epipolar rectification, and block-based correspondence search for dense fisheye stereo reconstruction for clouds. We showed that cloud photogrammetry is able to compute the cloud envelope geometry and demonstrated the potential of such methods for the analysis of detailed cloud morphologies. By applying an epipolar rectification together with a dense (semi-)global stereo matching algorithm instead of using the original images, we are able to compute clouds shapes that are more complete and contiguous than reconstructions relying on regular feature-based methods. Once the cameras are calibrated, the method can be fully automated to deliver real-time information of the cloud scene.

Our proposed technique requires accurate camera calibration parameters and synchronously triggered cameras. Although the validation of our results with cloud radar observations should be extended to convective clouds, the reconstructions have shown to be stable over time, yielding robust cloud base and motion estimates. The validation for stratiform clouds show acceptable deviations from the lidar-ceilometer and radar measurements.

30 Future work will focus on the combination of cloud photogrammetry with different sensors. Furthermore we want to use two additional hemispheric sky imagers outside the research center to obtain more closed surfaces by fusing different views.





5

*Acknowledgements.* This study has been performed within the framework of the High Definition Clouds and Precipitation for advancing Climate Prediction (HD(CP)<sup>2</sup>) project funded by the Federal Ministry for Education and Research in Germany (BMBF). We want to thank Dr. Emiliano Orlandi and Dr. Kerstin Ebell from the University of Cologne for their efforts to provide the observation data from the cloud radar at JOYCE, and Dr. Birger Bohn from the Forschungszentrum Jülich GmbH and our technical staff Martin Lennefer for organization, technical implementation and maintenance.





## References

5

- Abraham, S. and Förstner, W.: Fish-eye-stereo calibration and epipolar rectification, ISPRS Journal of Photogrammetry and Remote Sensing (JPRS), 59, 278–288, 2005.
- Abraham, S. and Hau, T.: Towards Autonomous High-Precision Calibration of Digital Cameras, in: Proc. of SPIE Videometrics, pp. 82–93, 1997.

Allmen, M. C. and Kegelmeyer, W. P.: The Computation of Cloud-Base Height from Paired Whole-Sky Imaging Cameras, Journal of Atmospheric and Oceanic Technology, 1, 97–113, 1996.

Bakstein, H. and Pajdla, T.: Calibration of a fish eye lens with field of view larger than 180°, in: Proc. of the Computer Vision Winter Workshop (CVWW), pp. 276–285, 2002.

10 Brown, D. C.: Close-range Camera Calibration, Photogrammetric Engineering, 37, 855–866, 1971.

Fischler, M. A. and Bolles, R. C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Communications of the ACM, 24, 381–395, 1981.

Hirschmüller, H.: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information, in: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 807–814, 2005.

- 15 Hirschmüller, H. and Gehrig, S.: Stereo Matching in the Presence of Sub-Pixel Calibration Errors, in: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 437–444, 2009.
  - Hu, J., Razdan, A., and Zehnder, J. A.: Geometric Calibration of Digital Cameras for 3D Cumulus Cloud Measurements, Journal of Athmospheric and Oceanic Technology, 26, 200–214, 2010.

IDS: manual for uEye cameras, http://de.ids-imaging.com/, [Online; accessed March-2016], 2013.

20 Kannala, J. and Brandt, S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses, IEEE Trans. on Pattern Analalysis and Machine Intelligence (TPAMI), 28, 1335–1340, 2006.

Kassianov, E., Long, C., and Christy, J.: Cloud-Base-Height Estimation from Paired Ground-Based Hemispherical Observations, Journal of Applied Meteorology, 44, 1221–1233, 2010.

Koppe, C.: Photogrammetrie und Internationale Wolkenmessung, Friedrich Vieweg und Sohn, 1896.

- 25 Library, P. C.: Point Cloud Library, http://pointclouds.org, [Online; accessed March-2016], 2016.
- Löhnert, U., Schween, J. H., Acquistapace, C., Ebell, K., Maahn, M., Barreraverdejo, M., Hirsikko, A., Bohn, B., Knaps, A., O'connor, E., Simmer, C., Wahner, A., and Crewell, S.: JOYCE: Jülich Observatory for Cloud Evolution, Bulletin of the American Meteorological Society, 96, 1157–1174, 2014.

Longuet-Higgins, H.: A computer algorithm for reconstructing a scene from two projections, Nature, 293, 133–135, 1981.

30 Lowe, D. G.: Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60, 91–110, 2004.
Madsen, K., Nielsen, H. B., and Tingleff, O.: Methods for non-linear least squares problems, http://www.ltu.se/cms\_fs/1.51590!/nonlinear\_least\_squares.pdf, 2nd Edition, 2004.

Öktem, R., Prabhat, Lee, J., Thomas, A., Zuidema, P., and Romps, D. M.: Stereophotogrammetry of Oceanic Clouds, Journal of Atmospheric and Oceanic Technology, 31, 1482–1501, 2014.

35 Oliensis, J.: Exact two-image structure from motion, IEEE Trans. on Pattern Analalysis and Machine Intelligence (TPAMI), 24, 2002, 2002. Scharstein, D. and Szeliski, R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, Intl. Journal of Computer Vision (IJCV), 47, 7–42, 2002.





Schneider, J., Stachniss, C., and Förstner, W.: On the Accuracy of Dense Fisheye Stereo, IEEE Robotics and Automation Letters, 1, 227–234, 2016.

Seiz, G.: Ground- and Satellite-Based Multi-View Photogrammetric Determination of 3D Cloud Geometry, Ph.D. thesis, ETH Zurich, 2003. Sonka, M., Hlavac, V., and Boyle, R.: Image Processing, Analysis, and Machine Vision, Brooks/Cole Publishing Company, 2nd edn., 1999.

 5 Stockman, G. and Shapiro, L. G.: Computer Vision, Prentice Hall, London, 1st edn., 2001.
 Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, Quarterly Journal of the Royal Meteorological Society, 141, 563–579, 2015.







**Figure 1.** We employ two hemispheric sky imagers (left) in a stereo setup with a baseline of 300 m to derive dense and detailed geometric information such as height and morphology of clouds using the simultaneously recorded image pairs. The derived heights (right) of an exemplary recorded cloud (middle) using the fisheye images, as shown in Figure 4 (c), of the two sky imagers.



Figure 2. (a) In the omnidirectional camera model the 3D object point **P** is mapped to  $\mathbf{x}'$  on the image plane. Several radial-symmetric projection functions  $r(\theta)$  can be used, which define the distance to the projection center  $\mathbf{x}_{\mathbf{C}}$ ; (b) The final projection  $\mathbf{x}$  in pixel coordinates is determined by the camera calibration parameters and additional distortion coefficients.



Figure 3. The two hemispheric cameras are located at  $\mathbf{C}_{\mathbf{L}}$  and  $\mathbf{C}_{\mathbf{R}}$  with independent orientation in the world coordinate system  $\Omega_W$ . The projections  $\mathbf{x}''_{\mathbf{L}}$  and  $\mathbf{x}''_{\mathbf{R}}$  of a 3D point  $\mathbf{P}$  on the image hemisphere in each camera system  $\Omega_L$  and  $\Omega_R$  span together with the baseline  $\mathbf{t}$  the epipolar plane and can be used to reconstruct  $\mathbf{P}$  via triangulation.







**Figure 4.** Epipolar rectification for omnidirectional cameras: The two-step rotational mapping between real and virtual cameras (a) results in a canonical camera setup of virtual cameras during rectification (b). A fisheye image (c) is rectified such that the angles  $\beta$  and  $\psi$  correspond to the lines and rows of the image (d).



**Figure 5.** Two corresponding ray directions  $\mathbf{x}''_{\mathbf{L}}$  and  $\mathbf{x}''_{\mathbf{R}}$  are defined by the angles  $\psi_L$  and  $\psi_R$  within the epipolar plane. With baseline length *b* the distance *s* between left camera and 3D Point **P** can be derived, which allows to determine 3D point coordinates  $\mathbf{P} = s \mathbf{x}''_{\mathbf{L}}$ .







**Figure 6.** Camera calibration with cube: the pattern on the inside of the cube defines a set of 3D points with respect to the cube coordinate system and can be used as reference data to solve for the fisheye projection parameters.



Figure 7. Absolute orientation estimation via stars from long-exposure images. In case of an accurate orientation estimation, the projected coordinates from the star catalog should match the detected stars in the image.







Figure 8. Detected interest points can be matched across the stereo images and are marked by the same color. At least 5 correspondences are needed to compute the relative orientation between the cameras and each also provides one 3D point.



Figure 9. Illustration of dense stereo matching using epipolar rectified images. The correspondence information is stored in a disparity map. Each disparity  $D(\mathbf{x}) > 0$  then defines a correspondence between two image points  $\mathbf{x}_L$  and  $\mathbf{x}_R$  and can be used for triangulation, which results in a dense 3D point cloud.







Figure 10. Camera setup at the Research Center Jülich and RHI-scanning directions of the cloud radar (map source: OpenStreetMap).



Figure 11. Errors in the disparity values cause a directional deviation of the projection rays within the epipolar plane. Assuming an error of  $\Delta \psi_R = 0.1^\circ$ , we can compute the absolute geometric error within the epipolar plane for a hypothetical cloud layer at 1.5 km and 3 km height in an area of 5 km around the cameras.







**Figure 12.** Simulation of an asynchronous recording by the stereo cameras. A virtual cloud layer at 3 km height was displaced between the two recordings of the stereo cameras by  $\pm 15$  m (a) along the baseline (x-direction) and (b) in perpendicular direction (y-axis). Plots show for each case the respective cross-sections in x- and y-direction of the reconstruction.







**Figure 13.** Comparison of the reconstruction with cloud radar (cross-section) and lidar (near-zenith cloud base heights) at 11 August 2014, 14:12:00 UTC. The original fisheye image showing the respective direction of the cross-sections (top left). The cross-section of the reconstruction compared with the reflectivities from the cloud radar (top right). Histogram of the cloud base heights observed by the stereo camera (bottom left) and the lidar-ceilometer (bottom right).







Figure 14. Comparison of the reconstruction with cloud radar and lidar at 5 August 2014, 11:31:30 UTC, as in Fig.13







**Figure 15.** 3D reconstruction of cumulus mediodcris at 24 July 2014: The left column shows a subsection of the images obtained from Camera 1. The central column visualizes the reconstruction as an untextured triangulated surface mesh. The right column shows the color-coded height of the reconstruction in meters with contour lines (right). Results are shown for 12:03:00 UTC (top row), 12:05:0 UTC (middle row) and 12:07:00 UTC (bottom row).







Figure 16. Cross-section of the reconstruction from 24 July 2014, 12:07:00 UTC, and highlighted cloud base height from the lidar-ceilometer.



Figure 17. 3D reconstruction of cumulus cloud from 24 July 2014, organized as in Figure 15. Results are shown for 11:28:00 UTC (top row), 11:30:00 UTC (middle row) and 11:32:00 UTC (bottom row)







Figure 18. Cross-section of the reconstruction from 24 July 2014, 11:32:00 UTC.