Cloud-Stereo: A Dataset and Benchmark for Reconstructing Atmospheric Clouds from Stereo Images

Jacob Lin[1]
jacob.lin@cs.ox.ac.uk
Edward Gryspeerdt[2]
e.gryspeerdt@imperial.ac.uk
Ronald Clark[1]
ronald.clark@cs.ox.ac.uk

[1] Department of Computer Science
University of Oxford
Oxford, UK

[2] Department of Physics
Imperial College London
London, UK

Abstract

Obtaining accurate measurements of clouds is a critical problem in atmospheric physics, as accurate modeling of cloud properties allows us to better understand and predict climate change. Stereo camera networks have shown promise in obtaining such measurements, being able to reconstruct detailed cloud fields over multi-km^2 domains. However, previous studies on cloud stereo depth estimation have been limited to using traditional (non-learned) matching techniques, due to the absence of suitable training datasets for this challenging domain. In this work, we present a novel dataset (Cloud-Stereo) specifically tailored for cloud depth estimation. The Cloud-Stereo dataset includes: 1) a synthetic dataset for training, comprising 3000 stereo pairs and simulated dense LiDAR depth data, and 2) a high-accuracy real-world dataset consisting of $\approx 120k$ frames acquired from a stereo camera and Doppler Aerosol LiDAR for testing. Using our dataset we benchmark existing learning and non-learning based stereo depth estimation approaches, and demonstrate that fine-tuning on our dataset can lead to significant accuracy improvement for learned methods. We believe this dataset will enable the training of future, more accurate, methods for cloud field reconstruction, enhancing a unique measurement capability for developing and evaluating atmospheric models. The dataset is available at `https://cloud-stereo.jacob-lin.com/`.

1 Introduction

Clouds are a central component of the Earth's climate system. They help control the Earth's temperature by affecting the sunlight we receive and the heat we lose into space [38], as well as playing a vital part in the water cycle. Clouds are controlled by effects across a wide range of spatial scales, from weather systems (100s of km) to cloud droplet formation (micrometers). This makes them difficult to model accurately, such that clouds lead to the largest uncertainties in our understanding of how humans influence the climate system [7, 11] and how it responds to rising temperatures [46].

© 2025. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

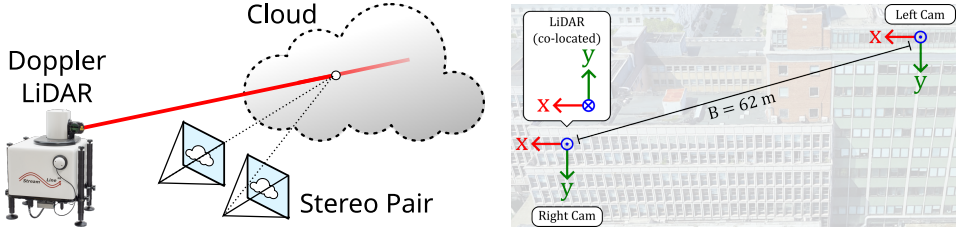

Figure 1: (**a**) We introduce a dataset that can be used for evaluating stereo cloud depth estimation methods, useful for the study of cloud heights, morphology, and dynamics. The ground truth cloud depths are obtained using a Doppler LiDAR. As part of Cloud-Stereo we also provide simulated cloud data that can be used to train stereo models. (**b**) The cameras are upwards pointing so that the bottom of the camera frames are aligned with the horizon, and the LiDAR scan patterns are aligned with the cameras.

One important factor in building accurate cloud models is understanding how clouds mix with drier, surrounding air. This mixing determines how long a cloud lasts, and it is often the cause of a cloud's eventual disappearance through evaporation [25]. It also controls the response of clouds to particulate pollution [1, 9, 15] and is responsible for much of the uncertainty in how greenhouse gases affect the climate [37]. Even in models that can accurately represent a typical small cloud, there is still a lot of uncertainty in how these clouds behave due to variations in the model design [36, 40]. We therefore need better observations of clouds at a high temporal and spatial resolution to better understand processes like mixing.

The strength of these processes can be inferred from the 3D shape of the cloud, with the geometrical thickness of the cloud, the cloud size, and the motion of the cloud top all being related to how a cloud mixes with the surrounding dry air [7]. However, current observing systems are typically unable to observe the 3D structure of clouds with the necessary resolution. While the latest generation of satellites can achieve high temporal resolution [34], they do not have the spatial resolution to observe small clouds and they primarily observe cloud tops. Ground-based scanning radars and LiDARs, with beamwidths of less than 0.5° [e.g. 14], can reconstruct clouds in 3D [10], but are limited in their repeat time as they require physically scanning a dish or mirror across a scene [41].

On the other hand, stereo cameras show great potential for obtaining cloud range, shape, and size at the high spatial and temporal resolution necessary to reconstruct clouds in 3D and measure cloud mixing processes. This requires accurate matching of corresponding points from multiple views of a scene. This is a challenging problem for clouds, as they are volumetric entities without a well-defined surface which is further compounded by effects such as glare, lighting changes, and occlusion.

Existing approaches to estimate correspondences across multiple images are typically based on block matching [16] to estimate correspondences across multiple viewpoints. However, these methods suffer from significant noise and accuracy issues when dealing with clouds. Deep learning-based stereo vision algorithms [45] have great promise to address the accuracy and robustness issues encountered in traditional stereo photogrammetry for clouds. However, there are a number of significant challenges associated with creating an effective deep learning based cloud depth estimation approach as this requires an accurately calibrated stereo camera and LiDAR data. As far as we are aware there is currently no dataset that can be used to train or evaluate supervised ground-based cloud depth estimation models.

In this paper, we directly address these challenges and present a dataset specifically designed for cloud depth estimation. Specifically, our contributions include:

1. A real-world dataset with sparse LiDAR measurements for benchmarking. Along with stereo camera images, this dataset includes synchronized sparse depth measurements collected using a steerable Doppler Aerosol Lidar. We propose this to be used as a benchmark for cloud photogrammetry.

2. A synthetic dataset for training. To facilitate the training of deep networks in this domain, we also introduce a synthetic dataset that provides accurate dense depth information, simulating the depth that would be recorded by the LiDAR.

3. Finally, we evaluate several existing (non-learned, pre-trained, and fine-tuned) stereo depth estimation methods on this dataset, demonstrating significant advantages when training methods on the synthetic data.

2 Background

In this section we first introduce general systems used for cloud reconstructions, then we present related work on stereo depth estimation approaches.

2.1 Cloud Field Reconstruction

In computer-vision terms, clouds can be treated as dynamic, volumetric objects that are reconstructed via classic multi-view stereo and voxel-carving pipelines. The earliest efforts relied on manual identification and correspondence of salient features across overlapping photographs [21]. These workflows were then automated: keypoint detectors and descriptors (akin to SIFT/SURF) were used to establish correspondences, followed by triangulation to recover 3D cloud-top geometry [19].

On the satellite side, instruments like MISR operate as a calibrated multi-camera rig in space, capturing simultaneous views at different angles. Matched image patches are triangulated to yield high-resolution cloud-top point clouds [24]. Geostationary constellations extend this paradigm by providing near-continuous, multi-angle coverage which enables both static height retrieval and dense motion fields via temporal stereo [5, 27]. Upcoming missions continue to refine sensor geometry and radiometric precision, improving 3D accuracy under a wider range of illumination and optical thickness [6].

Ground-based setups follow the same principles but trade altitude for finer angular baselines. For example, the COGS (Clouds Optically Gridded by Stereo) [29] project deploys three synchronized camera pairs on the ground. By combining their epipolar-based reconstructions with voxel-carving constraints, COGS reduces surface ambiguity and yields more robust volumetric cloud reconstructions than shape-from-silhouette approaches [29, 47]. These methods demonstrate that, with careful calibration and multi-view correspondence, ground-based setups are able to estimate detailed cloud shapes.

More recent approaches attempt to reconstruct the full internal distribution of the cloud volume, retrieving the cloud's internal structure has been referred to as the "holy grail" of cloud imaging [39, 42]. The scattering of light by liquid water droplets depends not only on the optical depth but on the droplet size and the droplet composition. Approaches such as 4D-cloud scattering tomography (4D-CST) [30], 3DeepCT [35] and VIP-CT [31] utilize

multiple satellite images and the sun as a light source and then reconstruct the volume as an inverse problem. Our dataset compliments these approaches by enabling the training of priors to help constrain the inverse problem, and help validate such methods.

2.2 Stereo Depth Estimation

In this work, we focus on stereo depth methods for 3D reconstruction, due to their wide applicability to problems in this domain and ease of deployment across existing ground-based and satellite systems. There are two main classes of matching algorithms.

Traditional stereo matching: Traditional stereo matching approaches have a long history in computer vision, aiming to use pixel or patch similarity to find corresponding points between two stereo images to estimate depth information. Block matching, which divides the image into small blocks to search for the best match, has been a fundamental technique [32]. Feature-based matching identifies and matches key features based on descriptors [23], while graph cuts formulate the problem as energy minimization [18]. Dynamic programming finds optimal matches by considering cost and constraints [26], and belief propagation is a probabilistic method that iteratively refines matching [9]. Semi-Global Matching (SGM) minimizes the global energy function by considering local matching costs and global smoothness constraints [17]. This has been one of the most popular stereo matching approaches because of its good balance between computational efficiency, and performance.

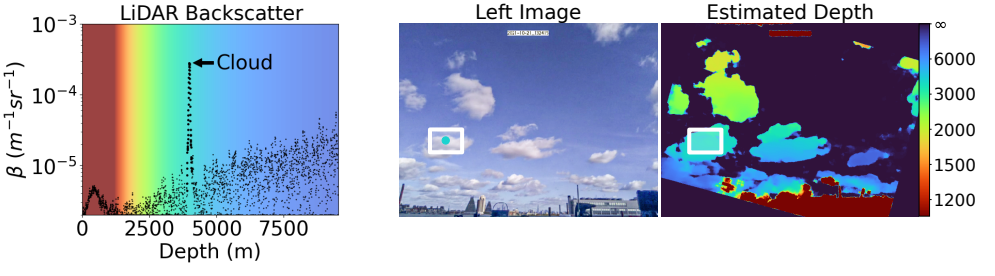

Figure 2: **Example LiDAR backscatter.** The highlighted marker in the left camera image shows the LiDAR ray point on the cloud projected onto the camera. The cloud depth is defined to be at the peak of the attenuated backscatter (denoted by β).

Learned stereo methods: As in many other areas of computer vision, deep learning has led to significant improvements in stereo depth estimation. Early methods like [45] used convolutional neural networks to learn the matching of image patches, leading to improved robustness. Subsequent works integrated traditional stereo algorithms into deep learning frameworks [44], or leveraged semantics and geometry, as in Pseudo-LiDAR [28]. Attention to real-time performance led to innovations like DeepPruner [8], which reduced computational requirements. Self-supervised and unsupervised methods have also been proposed such as MonoDepth [13], extending capabilities without the need for ground-truth data. Other approaches have focused on iterative refinement. RAFT-Stereo [20] is an extension of the RAFT algorithm, specifically designed for stereo matching. It leverages a recurrent neural network architecture that iteratively refines disparity estimation between stereo images, constructing a 3D cost volume to capture pixel similarity at different disparity levels, and allowing for end-to-end training. GMStereo [43] aims to provide a unified framework for various matching tasks, such as flow, disparity, and depth. It utilizes attention layers

for matching, emphasizing robustness to challenges like occlusions and efficiency. For this reason, we use GMStereo in our benchmark.

2.3 Overview of the Cloud-Stereo Dataset

Our Cloud-Stereo dataset is composed of two complementary parts: a large synthetic stereo subset created for model training, and a real-world LiDAR-camera subset designed for evaluation and benchmarking. Together, these provide both dense ground truth for learning and highly accurate measurements for assessing real-world performance.

2.4 Synthetic Stereo Data

To generate dense, per-pixel depth supervision, we render photorealistic cloud scenes using Terragen's volumetric engine. We simulate three canonical cloud types (low-level cumulus, mid-level altocumulus, and high-level cirrus) by sampling 1,000 combinations of key atmospheric parameters: cloud fraction, optical density, base altitude, vertical thickness, sun azimuth, and sun zenith angle. For each configuration, we produce a pair of stereo images along with their exact depth maps. Depth is computed via an opacity-weighted average of inverse ray distance, which approximates the peak LiDAR backscatter depth, using the volume rendering equation. From the inverse depth ξ we derive physical depth, $z = \frac{1}{\xi}$ and corresponding stereo disparity: $d = f B \xi$, where f is the focal length and B the camera baseline. To ensure viewpoint robustness, every parameter set is rendered from three elevation angles (0°, 45°, and 90° relative to the horizon). Finally, we provide a fixed, reproducible 90/10 train/test split (indices included) so that training and evaluation can be conducted under identical conditions.

2.5 Real-World LiDAR-Camera Data

Between October 11 and November 12, 2021, we collected 120,780 synchronized stereo image pairs and LiDAR scans over central London. Our imaging system consisted of two Raspberry Pi V2.1 cameras with fixed exposure and white balance, capturing frames every five seconds during daylight. The cameras were mounted 62 m apart, with the left 10 m above the right, and synchronized to under 0.1 s via a local NTP server. Absolute camera orientations were determined by matching daytime aircraft transponder data and nighttime star-field photographs.

Ground-truth distances were provided by a HALO Photonics Streamline 1.5 steerable Doppler LiDAR (1.5 μm), scanning a roughly $70° \times 25°$ field in a raster pattern with 2° ray spacing (four interleaved patterns offset by 0.5°). Each full scan took about ten minutes. To align the LiDAR and camera coordinate frames, we performed high-resolution raster scans (0.25° spacing) of nearby buildings and additional distant-tower scans at 0.1° spacing.

Raw LiDAR backscatter profiles $\beta(r)$ were corrected for focus and optics, then analyzed to locate the strongest return peak in each beam, which corresponds to the cloud edge. Those points were projected into the left-camera images to produce sparse depth maps. To suppress noise and aerosol returns, we discarded any LiDAR ray whose peak backscatter fell below one standard deviation above the mean at that range. Each synchronized frame pair is timestamped with sub-second accuracy.

2.6 Annotations and Metadata

Beyond depths, our dataset includes several key annotations. Cloud-base heights are derived from LiDAR ranges and extrinsic calibration. A meteorologist visually labeled 300 randomly selected frames per day, categorizing each into one of eleven types (e.g., cumulus, stratocumulus, cirrus, cumulonimbus). We also compute a "cloud coverage" metric for each full LiDAR scan, defined as the fraction of rays classified as cloud, capturing sky "cloudiness" over time. Example images and histograms illustrating these factors appear in Figure 3, with further calibration visuals and sample scenes provided in the supplementary material.

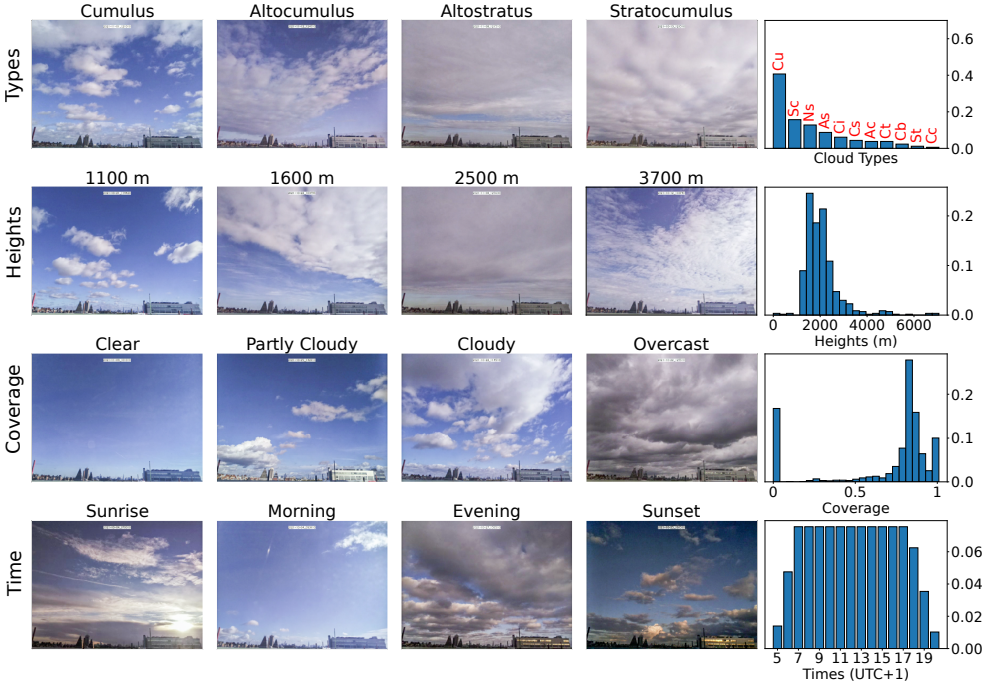

Figure 3: **Example images from the LiDAR dataset.** Each row highlights a factor of variation in our dataset. The first four columns show example images, while the last column shows a histogram that summarises the statistics of the relevant factor of variation. See main text for details.

2.7 Baselines

We use stereo depth estimation as an illustrative task to show how our dataset can be used for training and benchmarking cloud stereo depth estimation approaches. We select the following three methods for our benchmark.

a. Semi-Global Block Matching [16]: SGBM is a traditional stereo matching technique that is still widely used for cloud photogrammetry. It therefore serves as an important baseline to evaluate the effectiveness of existing cloud photogrammetry methods. We use the open-source implementation available in the OpenCV library [3]. We provide the hyperparameters used in the supplementary.

b. GMStereo [43] (pre-trained): We choose GMStereo as the representative method for learning-based approaches, as it has demonstrated state-of-the-art performance, while also training faster than other common learned matchers such as RAFT-Stereo [20]. Methods like GMStereo are generally trained on large synthetic datasets [12, 22, 33] that are composed of solid textured objects. Such objects greatly differ from clouds which are volumetric and often lacking in texture. For our evaluation, we use the released 4.7 M parameter model that has been trained on the scene flow datasets [22].

c. GMStereo [43] (fine-tuned): We fine-tune the pre-trained GMStereo mentioned above on our synthetic dataset. In doing so, the model can learn in a domain-specific manner, potentially learning to match cloud features better, while also learning to handle cloud-specific problems such as glare. For our fine-tuned model we introduce an opacity predictor on top of the transformer features. The opacity supervision allows for further control of the predicted depth maps through possibilities such as simple opacity thresholding, while potentially also implicitly helping the model learn how to separate the clouds and the sky. We fine-tune GMStereo with the following optimization objective:

$$\mathcal{L} = \mathcal{L}_{\text{disparity}} + \lambda_{opacity}\mathcal{L}_{\text{opacity}} \tag{1}$$

where we use a smooth L_1 loss for the disparity and a L_1 loss for the opacity. As with GMStereo (pre-trained), we initialize our model using the released 4.7 M parameter model that has been pre-trained on the scene flow datasets [22]. The model is then fine-tuned for 10000 steps, with a batch size of 8. All other training parameters follow the configuration by the original authors [43].

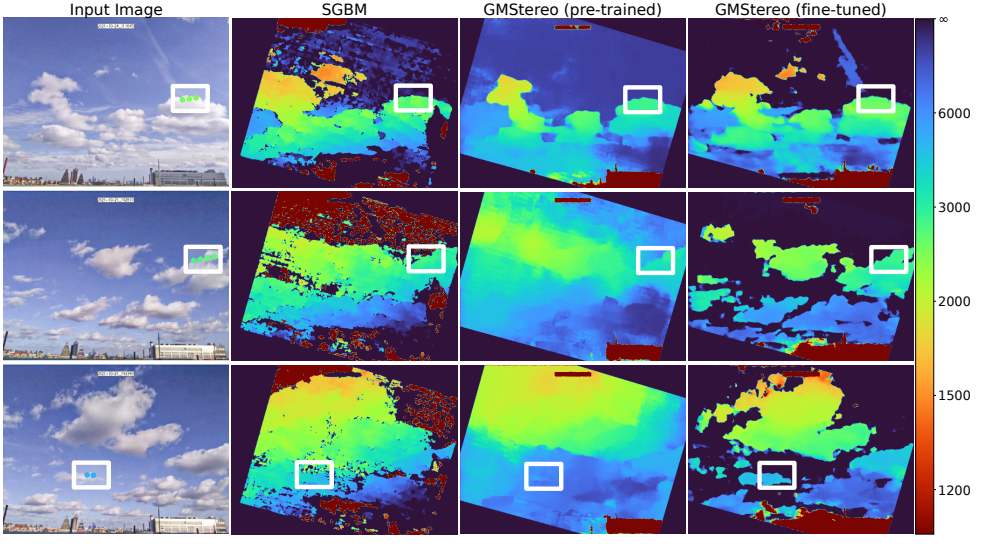

Figure 4: **Qualitative results for depth estimation.** The highlighted colored markers in the first column show the projected LiDAR/cloud intersections, where the color of the markers is sampled from the depth colormap. The fine-tuned method shows the best results, capturing the depth of both smaller clouds and contrails, while also being able to mask out the sky.

Metrics: Following previous work, we consider common error metrics, specifically: Absolute Relative difference (Abs Rel), Squared Relative difference (Sq Rel), and Root Mean

Square Error (RMSE). However, unlike in existing depth estimation domains such as autonomous driving and robotics, our depth values span a significant range. Specifically, the empty sky, which has very low disparity, usually fills a large portion of the image. This can lead to very large depth values (and hence errors). Therefore we clip all disparities to a lower bound of 1 px and use the *median* to compute the statistics across samples. To ensure we still measure the outliers we also include the accuracy at a threshold of $\delta < 1.25$.

3 Evaluation

In this section, we aim to: 1) benchmark the three selected approaches (non-learned, pre-trained, and fine-tuned) on our real dataset. 2) Determine how much training on the synthetic dataset helps improve accuracy.

3.1 Quantitative Results.

Fine-tuned GMStereo gives the most accurate and robust depth estimation. As illustrated in Figure 5, the fine-tuned GMStereo demonstrates a consistent performance where three-quarters of the LiDAR points are contained within a disparity error of 3 pixels, irrespective of the range of LiDAR distances. This underscores the model's robustness and precision in measuring across varying distances, which is also evident in Table 1 where the fine-tuned GMStereo outperforms the other methods.

Non-learned SGBM suffers from a large number of outliers and false matches. From Figure 5 we see SGBM exhibits a substantial number of outliers surpassing the 3 pixel disparity error. These discrepancies may stem from the inherent limitations of SGBM's fixed parameters, notably the block size. For instance, the block size that has been optimized for larger depths may underperform when faced with closer clouds. Closer clouds typically exhibit less distinctive features, challenging the system's ability to accurately map them based on a fixed block size.

Furthermore, while the configuration we used allows SGBM to search up to disparities corresponding to minimum depths of 1 km (which is lower than the evaluated minimum LiDAR depth), the system's maximum disparity can still be a constraining factor. This is because an incorrect but closer match could potentially be identified at higher disparities if the search space were larger. However, increasing the maximum disparity could also lead to more false matches. This risk is further amplified given that clouds tend to lack features, making them more susceptible to mismatches.

Consequently, this susceptibility to matching failures in closer, feature-less clouds leads to a higher frequency of outliers, as evidenced by Figure 5. This indicates the potential drawbacks associated with a non-learned method and the advantages of training a learned model on our data.

3.2 Qualitative Results

Fine-tuned GMStereo gives more details on small and thin clouds. In Figure 4 we compare the three methods using example images of cumulus clouds at varying depths. Unlike the other benchmarked methods, the fine-tuned GMStereo model is capable of identifying small and thin clouds such as the wisps and contrails seen in the first row of Figure 4. It is also

Table 1: **Quantitative results for depth estimation on clouds.** Due to the high number of outliers caused by predicted near-zero disparities, we use median values across samples for the Abs Rel, Sq Rel, RMSE, and RMSE log. See Section 2.7 for more detail on this choice.

Dataset	Method	Abs Rel ↓	Sq Rel ↓	RMSE ↓	$\delta < 1.25$ ↑
Synthetic	SGBM	0.071	8.785	122.9	0.634
	GMStereo (pre-trained)	0.038	1.866	47.97	0.876
	GMStereo (fine-tuned)	**0.013**	**0.197**	**15.26**	**0.975**
Real	SGBM	0.401	535.9	1278	0.382
	GMStereo (pre-trained)	0.134	48.00	356.8	0.684
	GMStereo (fine-tuned)	**0.112**	**33.61**	**294.3**	**0.716**

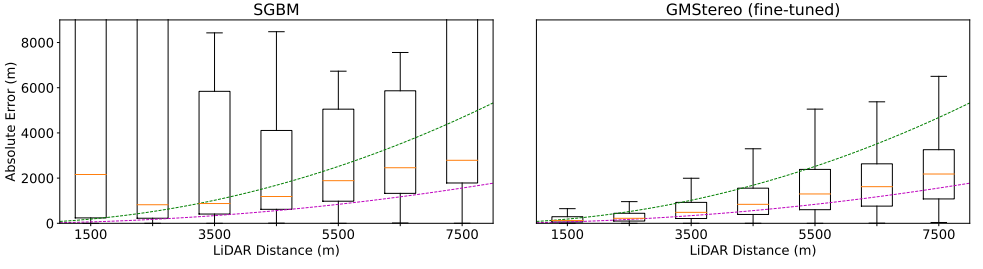

Figure 5: **Comparison of depth prediction errors between SGBM and GMStereo (fine-tuned) for different LiDAR ranges.** The magenta dashed line and the green dashed line shows the 1 px and the 3 px disparity error respectively. SGBM exhibits a high number of outliers at closer distances due to the non-learned comparison of raw pixel values. See the discussion in Section 3 for more detail.

better able to estimate the cloud's shape and silhouette. From both the quantatitve and qualitatitive results we can see that GMStereo fine-tuned on our synthetic dataset reaches the best accuracy, indicating that our synthetic dataset provides great value in improving real-world stereo cloud depth estimation accuracy.

4 Conclusion

In this paper, we present Cloud-Stereo, the first dataset and benchmark for cloud depth estimation. This dataset consists of two parts. We first create a synthetic dataset, where we have generated 3000 pairs of RGB stereo images with pixel-level depth info. We use this dataset to train learning-based approaches. We also collect 120780 pairs of real-world stereo image pairs with sparse depth measured from a LiDAR sensor. We use this real-world dataset to benchmark three widely used approaches for cloud depth estimation.

Our benchmark results show that the GMStereo fine-tuned on our synthetic dataset reaches the best accuracy. This indicates that our synthetic dataset allows learning-based approaches to perform better in the cloud domain. Furthermore, our benchmark enables a fair and thorough comparison between different approaches and allows highlighting the weaknesses of each approach. This is key to developing better methods in the future. We believe by enabling more accurate estimates of cloud distances, the Cloud-Stereo dataset will guide better

3D reconstructions of clouds and in doing so will enable new measurements of key processes in the atmosphere, and will support the development of more accurate and robust models for atmospheric research.

Acknowledgements: The authors would like to acknowledge the Atmospheric Measurement and Observation Facility (AMOF) a Natural Environment Research Council (UKRI-NERC) funded facility (NE/Y005376/1), for providing the Doppler Aerosol LiDAR. This work was supported by an ESPRC scholarship (2922572), the Natural Environment Research Council (grant nos.NE/X012255/1, NE/X018539/1), ARIA (Next-CAM, SCOP-SE01-P06), and the Royal Society (grant no. URF/R1/191602).

References

[1] Andrew S. Ackerman, Michael P. Kirkpatrick, David E. Stevens, and Owen B. Toon. The impact of humidity above stratiform clouds on indirect aerosol climate forcing. *Nature*, 432:1014, 12 2004. doi: 10.1038/nature03174. URL http://adsabs.harvard.edu/abs/2004Natur.432.1014A.

[2] N. Bellouin, J. Quaas, E. Gryspeerdt, S. Kinne, P. Stier, D. Watson-Parris, O. Boucher, K. S. Carslaw, M. Christensen, A.-L. Daniau, J.-L. Dufresne, G. Feingold, S. Fiedler, P. Forster, A. Gettelman, J. M. Haywood, U. Lohmann, F. Malavelle, T. Mauritsen, D. T. McCoy, G. Myhre, J. Mülmenstädt, D. Neubauer, A. Possner, M. Rugenstein, Y. Sato, M. Schulz, S. E. Schwartz, O. Sourdeval, T. Storelvmo, V. Toll, D. Winker, and B. Stevens. Bounding Global Aerosol Radiative Forcing of Climate Change. *Rev. Geophys.*, 58(1):e2019RG000660, 1 2020. doi: 10.1029/2019RG000660.

[3] Gary Bradski. The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11):120–123, 2000.

[4] C. S. Bretherton, P. N. Blossey, and J. Uchida. Cloud droplet sedimentation, entrainment efficiency, and subtropical stratocumulus albedo. *Geophys. Res. Lett.*, 34(3): L03813, 1 2007. doi: 10.1029/2006GL027648.

[5] James Carr, Dong Wu, Michael Kelly, and Jie Gong. MISR-GOES 3D Winds: Implications for Future LEO-GEO and LEO-LEO Winds. *Remote Sens.*, 10(12): 1885, 1 2018. doi: 10.3390/rs10121885. URL https://mdpi-res.com/d_attachment/remotesensing/remotesensing-10-01885/article_deploy/remotesensing-10-01885-v2.pdf.

[6] Paolo Dandini, Céline Cornet, Renaud Binet, Laetitia Fenouil, Vadim Holodovsky, Yoav Y. Schechner, Didier Ricard, and Daniel Rosenfeld. 3D cloud envelope and cloud development velocity from simulated CLOUD (C3IEL) stereo images. *Atmospheric Measurement Techniques*, 15(20):6221–6242, 10 2022. doi: 10.5194/amt-15-6221-2022. URL https://doi.org/10.5194%2Famt-15-6221-2022.

[7] S. Drueke, D. J. Kirshbaum, and P. Kollias. Evaluation of Shallow-Cumulus Entrainment Rate Retrievals Using Large-Eddy Simulation. *J. Geophys. Res.*, 124(16):9624–9643, 1 2019. doi: 10.1029/2019JD030889.

[8] Shivam Duggal, Shenlong Wang, Wei-Chiu Ma, Rui Hu, and Raquel Urtasun. Deeppruner: Learning efficient stereo matching via differentiable patchmatch. In *ICCV*, pages 4384–4393, 2019.

[9] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 70:41–54, 2006.

[10] Mark D. Fielding, J. Christine Chiu, Robin J. Hogan, and Graham Feingold. A novel ensemble method for retrieving properties of warm cloud in 3-D using ground-based scanning radar and zenith radiances. *J. Geophys. Res.*, 119(18):10,912–10,930, 9 2014. doi: 10.1002/2014jd021742.

[11] P Forster, T Storelvmo, K. C. Armour, W. D. Collins, J.-L. Dufresne, D. J. Frame, Daniel J. Lunt, Thorsten Mauritsen, Matthew D. Palmer, M. Watanabe, M. Wild, and H. Zhang. *The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1 2021. doi: 10.1017/9781009157896.009.

[12] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[13] Clément Godard, Oisin Mac Aodha, and Gabriel J Brostow. Unsupervised monocular depth estimation with left-right consistency. In *CVPR*, pages 270–279, 2017.

[14] J.W.F. Goddard, M. Thurai, and J.D. Eastment. The Chilbolton Advanced Meteorological Radar: a tool for multidisciplinary atmospheric research. *Electronics & Communication Engineering Journal*, 6(2):77–86, 4 1994. doi: 10.1049/ecej:19940205. URL https://doi.org/10.1049%2Fecej%3A19940205.

[15] Edward Gryspeerdt, Tom Goren, Odran Sourdeval, Johannes Quaas, Johannes Mülmenstädt, Sudhakar Dipu, Claudia Unglaub, Andrew Gettelman, and Matthew Christensen. Constraining the aerosol influence on cloud liquid water path. *Atmos. Chem. Phys.*, 19(8):5331–5347, 1 2019. doi: 10.5194/acp-19-5331-2019. URL https://www.atmos-chem-phys-discuss.net/acp-2018-885/acp-2018-885.pdf.

[16] Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE TPAMI*, 30(2):328–341, 2007.

[17] Heiko Hirschmuller and Daniel Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE TPAMI*, 31(9):1582–1599, 2008.

[18] Li Hong and George Chen. Segment-based stereo matching using graph cuts. In *CVPR*, 2004.

[19] Garry E. Hunt, Jan-Peter Müller, and Philip Gee. Convective growth rates of equatorial features in the jovian atmosphere. *Nature*, 295(5849):491–494, 2 1982. doi: 10.1038/295491a0.

[20] Lahav Lipson, Zachary Teed, and Jia Deng. Raft-stereo: Multilevel recurrent field transforms for stereo matching. In *International Conference on 3D Vision (3DV)*, pages 218–227, 2021.

[21] F. H. Ludlam. Planning cloud seeding research. *Archiv für Meteorologie, Geophysik und Bioklimatologie Serie A*, 8(3):283–292, 4 1955. doi: 10.1007/bf02247287.

[22] N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. URL http://lmb.informatik.uni-freiburg.de/Publications/2016/MIFDB16. arXiv:1512.02134.

[23] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE TPAMI*, 27(10):1615–1630, 2005.

[24] C. Moroney, R. Davies, and J.-P. Muller. Operational retrieval of cloud-top heights using MISR data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(7): 1532–1540, 7 2002. doi: 10.1109/tgrs.2002.801150. URL https://doi.org/10.1109%2Ftgrs.2002.801150.

[25] Johannes Mülmenstädt, O. Sourdeval, J. Delanoë, and J. Quaas. Frequency of occurrence of rain from liquid-, mixed-, and ice-phase clouds derived from A-Train satellite retrievals. *Geophys. Res. Lett.*, 42(15):6502–6509, 1 2015. doi: 10.1002/2015GL064604.

[26] Yuichi Ohta and Takeo Kanade. Stereo by intra-and inter-scanline search using dynamic programming. *IEEE TPAMI*, (2):139–154, 1985.

[27] Simon R. Proud, Andrew T. Prata, and Simeon Schmauß. The January 2022 eruption of Hunga Tonga-Hunga Ha'apai volcano reached the mesosphere. *Science*, 378(6619): 554–557, 11 2022. doi: 10.1126/science.abo4076. URL https://doi.org/10.1126%2Fscience.abo4076.

[28] Rui Qian, Divyansh Garg, Yan Wang, Yurong You, Serge Belongie, Bharath Hariharan, Mark Campbell, Kilian Q Weinberger, and Wei-Lun Chao. End-to-end pseudo-lidar for image-based 3d object detection. In *CVPR*, pages 5881–5890, 2020.

[29] David M Romps and Ruşen Öktem. Observing clouds in 4d with multiview stereophotogrammetry. *Bulletin of the American Meteorological Society*, 99(12), 2018.

[30] Roi Ronen, Yoav Y. Schechner, and Eshkol Eytan. 4d cloud scattering tomography. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021.

[31] Roi Ronen, Vadim Holodovsky, and Yoav Y Schechner. Variable imaging projection cloud scattering tomography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[32] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47:7–42, 2002.

[33] Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nesic, Xi Wang, and Porter Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In Xiaoyi Jiang, Joachim Hornegger, and Reinhard Koch, editors, *GCPR*, volume 8753 of *Lecture Notes in Computer Science*, pages 31–42. Springer,

2014. ISBN 978-3-319-11751-5. URL http://dblp.uni-trier.de/db/conf/dagm/gcpr2014.html#ScharsteinHKKNWW14.

[34] Timothy J. Schmit, Paul Griffith, Mathew M. Gunshor, Jaime M. Daniels, Steven J. Goodman, and William J. Lebair. A Closer Look at the ABI on the GOES-R Series. *Bulletin of the American Meteorological Society*, 98(4):681–698, 4 2017. doi: 10.1175/bams-d-15-00230.1.

[35] Yael Sde-Chen, Yoav Y. Schechner, Vadim Holodovsky, and Eshkol Eytan. 3deepct: Learning volumetric scattering tomography of clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[36] George Spill, Philip Stier, Paul R. Field, and Guy Dagan. Effects of aerosol in simulations of realistic shallow cumulus cloud fields in a large domain. *Atmos. Chem. Phys.*, 19(21):13507–13517, 1 2019. doi: 10.5194/acp-19-13507-2019. URL https://www.atmos-chem-phys.net/19/13507/2019/acp-19-13507-2019.pdf.

[37] D. A. Stainforth, T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A. Spicer, A. J. Thorpe, and M. R. Allen. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, 433:403, 2005. doi: 10.1038/nature03301. URL http://adsabs.harvard.edu/abs/2005Natur.433..403S.

[38] Graeme L. Stephens, Juilin Li, Martin Wild, Carol Anne Clayson, Norman Loeb, Seiji Kato, Tristan L'Ecuyer, Paul W. Stackhouse, Matthew Lebsock, and Timothy Andrews. An update on Earth's energy balance in light of the latest global observations. *Nat. Geosci.*, 5(10):691–696, 1 2012. doi: 10.1038/NGEO1580.

[39] J. Warner, J. F. Drake, and P. R. Krehbiel. Determination of Cloud Liquid Water Distribution by Inversion of Radiometric Data. *Journal of Atmospheric and Oceanic Technology*, 2(3):293–303, 9 1985. doi: 10.1175/1520-0426(1985)002<0293:doclwd>2.0.co;2.

[40] Bethan White, Edward Gryspeerdt, Philip Stier, Hugh Morrison, Gregory Thompson, and Zak Kipling. Uncertainty from the choice of microphysics scheme in convection-permitting models significantly exceeds aerosol effects. *Atmos. Chem. Phys.*, 17(19):12145–12175, 1 2017. doi: 10.5194/acp-17-12145-2017. URL https://www.atmos-chem-phys.net/17/12145/2017/acp-17-12145-2017.pdf.

[41] K Widener, N Bharadwaj, and K Johnson. Scanning ARM Cloud Radar Handbook. Technical report, ARM, 6 2012. URL https://doi.org/10.2172%2F1043296.

[42] W.J. Wiscombe. Scales, Tools and Reminiscences. In *3D Radiative Transfer in Cloudy Atmospheres*, pages 3–92. Springer-Verlag, Berlin/Heidelberg, 2005. doi: 10.1007/3-540-28519-9_1. URL https://doi.org/10.1007%2F3-540-28519-9_1.

[43] Haofei Xu, Jing Zhang, Jianfei Cai, Hamid Rezatofighi, Fisher Yu, Dacheng Tao, and Andreas Geiger. Unifying flow, stereo and depth estimation. *IEEE TPAMI*, 2023.

[44] Hongwei Yi, Zizhuang Wei, Mingyu Ding, Runze Zhang, Yisong Chen, Guoping Wang, and Yu-Wing Tai. Pyramid multi-view stereo net with self-adaptive view aggregation. In *ECCV*, pages 766–782, 2020.

[45] Jure Zbontar and Yann LeCun. Computing the stereo matching cost with a convolutional neural network. In *CVPR*, pages 1592–1599, 2015.

[46] Mark D. Zelinka, Timothy A. Myers, Daniel T. McCoy, Stephen Po-Chedley, Peter M. Caldwell, Paulo Ceppi, Stephen A. Klein, and Karl E. Taylor. Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophys. Res. Lett.*, 47(1), 1 2020. doi: 10.1029/2019GL085782.

[47] Rusen Öktem, Prabhat, James Lee, Aaron Thomas, Paquita Zuidema, and David M. Romps. Stereophotogrammetry of Oceanic Clouds. *J. Atmos. Ocean. Tech.*, 31(7): 1482–1501, 1 2014. doi: 10.1175/JTECH-D-13-00224.1.