3DBubbles: an experimental dataset for model training and validation

Baodi Yu[1,2], Qian Chen[1,2], Yanwei Qin[1], Xiaohui Su[1,2], Sunyang Wang[1,2], Fanyong Meng[1,*].
[1]*State Key Laboratory of Mesoscience and Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing, 100190, China*
[2]*School of Chemical Engineering, University of Chinese Academy of Sciences, Beijing, 100049, China*
[*]*To whom correspondence may be addressed. Email: fymeng@ipe.ac.cn*

Abstract:
Accurate measurement and validation of bubble structures are pivotal for elucidating the mechanisms of gas-liquid two-phase flow. With advancements in measurement techniques, computational fluid dynamics (CFD), and digital image analysis (DIA), there is a growing demand for high-quality benchmark datasets of bubbles, which are challenging to obtain through traditional experimentation. This study introduced a novel methodology, yielding a comprehensive bubble dataset containing over 10,000 three-dimensional (3D) bubbles coupled with two-dimensional (2D) images rendered from full-angle virtual projection, called 3Dbubbles. These bubbles were generated within a static gas-liquid flow field phantom designed to simulate real bubbly flow, and their 3D structures were scanned and reconstructed using an in-house developed high spatial resolution X-ray computed tomography (CT) system. 3DBubbles integrated reconstruction, structural characterization, data compression based on spherical harmonic functions, and 2D images rendered via virtual projection. 3Dbubbles aims to establish an ImageNet-like benchmark with its scale, accuracy and reliability to provide a rigorous bubble dataset embedded with rich structural information for bubbly flow investigation.

Keywords: Bubbly flow, Dataset, Deep learning, Benchmark, Computed tomography, Spherical harmonic

Introduction:
The accurate 3D structures of the bubbles are essential for calibrating measurement techniques, validating CFD models, and enhancing DIA. Various methods, including multiphase flow measurement techniques [1], numerical simulations [2], computer graphics (CG), and deep generative models (DGM) [3], have been employed to explore the internal structures and evolutionary mechanisms of bubbly flow. However, due to the intricate dynamics of structural evolution, obtaining accurate 3D bubble structures with high spatial and temporal resolution remains challenging. To address these, this study proposes a method to experimentally build bubble datasets from bubbly flow phantoms and elucidates potential application scenarios.

Experimental Setup:
Within the static gas-liquid flow field phantom, hydrogel served as the liquid phase and air as the gas phase, generating bubbles that were measured and reconstructed with high accuracy using an in-house developed X-ray CT system. Fig. 1 illustrates the static phantom and data processing workflow, yielding over 10,000 3D bubbles.

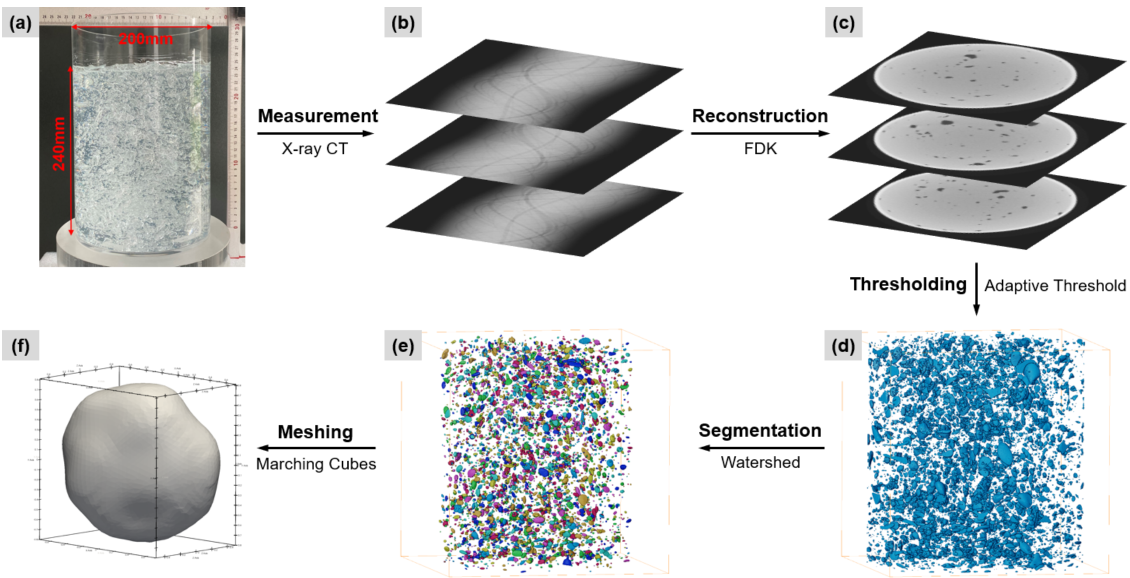

Figure 1: Experimental and data processing workflow. (a) Static bubbly flow phantom; (b) Sinogram; (c) 2D section of bubbly flow field; (d) 3D binary bubbly flow field; (e) Dispersed bubbles; (f) Individual 3D bubble.

Results and discussion:
Fig. 2(a) visualizes the reconstructed bubble point cloud closer to the target bubble in the lower right as the spherical harmonic degree N increases. When $N = 15$, the reconstructed volume (V_r) and reconstructed surface area (S_r) are closer to the volume (V) and surface area (S) of the target bubble. Fig. 2(b-c) quantifies the relationship between the three types of

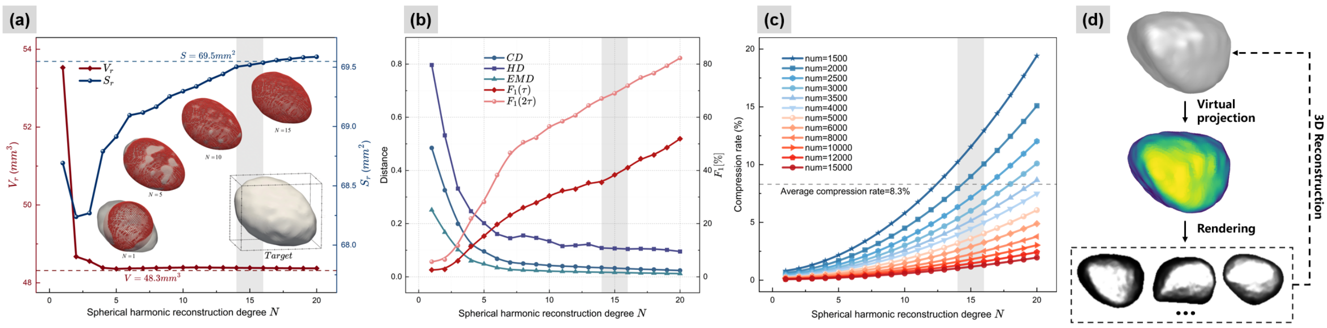

Figure 2: 3D reconstruction and 2D projections. (a) Reconstructed volume (V_r), reconstructed surface area (S_r) trends, and reconstructed point cloud visualization; (b) Chamfer distance (CD), Hausdorff distance (HD), Earth Mover's distance (EMD) [4] of the reconstructed point cloud to the target bubble and $F_1(\tau = 0.001)$; (c) Data compression rate; (d) Virtual projection and rendering.

reconstruction distances, the F_1-score, the data compression rate versus the degree of spherical harmonic, achieving an average data compression rate of 8.3% for 3Dbubbles at $N = 15$. To ensure the consistency and reliability of the structural characterization, principal component analysis (PCA) and spherical harmonic function were used to calculate 3D structural parameters, including aspect ratio, roundness, sphericity, convexity, and angularity. The results indicate that the dataset extensively covers the feature distribution of bubbles under the actual operating conditions. Fig. 2(d) demonstrates the virtual projection of a 3D structure and rendering of 2D images, simulating the attenuation of light through a bubble. In turn, coupling multiple 2D projections with corresponding 3D structures can extend the application of 3DBubbles in image processing through customized 3D reconstruction models. When predicting gas holdup from images, reconstructing bubble shapes and postures in 3D space using 2D projections can replace traditional methods such as contour reconstruction or ellipse fitting. Meanwhile, the broadly distributed 2D projection can supplant the concentric circular arrangement (CCA) [5] or deep generative models (DGM) [6] methods for the synthesis of bubbly flow images, serving as a training dataset for object detection algorithms and reducing the cost of manual annotation and verification.

Conclusion:

This study proposed a method for constructing a 3D bubble dataset, and experimentally measured a hydrogel-based static gas-liquid flow field phantom simulating real bubbly flow using an in-house developed high spatial resolution X-ray CT. Based on this method, a dataset containing 10,000 3D bubbles and full-angle 2D projections called 3DBubbles was constructed. The 3D bubble structure was characterized using a spherical harmonic function of degree 15, and the average EMD of the point cloud was 9.57×10^{-3}, with an average data compression rate of 8.3%. Virtual projection and rendering methods bridge the gap between 2D projections and 3D structures, providing benchmarks in scenarios such as gas holdup monitoring and accurate prediction of bubble structures.

Acknowledgments:

This study was financially supported by the National Natural Science Foundation of China (22178355, 62050226), and Strategic Priority Research Program of the Chinese Academy of Sciences (XDA0390501).

References

[1] Fanyong Meng. *Computed tomography in process engineering.*, Chemical Engineering Science. **252**, 117272 (2022).

[2] Sheikh Md Shakeel Hassan & Arthur Feeney. *BubbleML: A Multi-Physics Dataset and Benchmarks for Machine Learning.*, 37th Conference on Neural Information Processing Systems. Zenodo (2023).

[3] Chaoyue Gong & Yuchen Song. *BubDepth: A neural network approach to three-dimensional reconstruction of bubble geometry from single-view images.*, International Journal of Multiphase Flow. **152**, 104100 (2022).

[4] Nanyang Wang & Yinda Zhang. *Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images.*, Computer Vision – ECCV 2018, Lecture Notes in Computer Science. **11215**, 55-71 (2018).

[5] Y.M. Lau & Yuchen Song. *Development of an image measurement technique for size distribution in dense bubbly flows.*, Chemical Engineering Science. **94**, 20-29 (2013).

[6] Zhong Xiang & N.G. Deen. *Advanced Deep Learning-Based Bubbly Flow Image Generator under Different Superficial Gas Velocities.*, Industrial & Engineering Chemistry Research. **61**, 1531-1543 (2022).