The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W5-2025
13th International Conference on Mobile Mapping Technology (MMT 2025)
"Mobile Mapping for Autonomous Systems and Spatial Intelligence", 20–22 June 2025, Xiamen, China

HyGS-TDOM: A Hybrid Gaussian Splatting Famework for generating TDOMs from both dense and sparse views

Xiang Wang, Yiwei Xu, Wendi Zhang, Xin Wang*, Zongqian Zhan

Wuhan University, School of Geodesy and Geomatics, 430079 Wuhan, China PR
(xiangwang, xywjohn_sgg2020, wdzhang)@whu.edu.cn, (xwang, zqzhan)@sgg.whu.edu.cn

Keywords: True Digital Orthophoto Maps (TDOM), 3D Gaussian Splatting, Sparse View, Orthogonal Splatting.

Abstract

The True Digital Orthophoto Map (TDOM) possesses both map geometric accuracy and image characteristics, serving as an essential product for digital twins and Geographic Information Systems (GIS). Traditional TDOM generation methods typically involve a series of intricate geometric processing steps, which often result in computational inefficiency, high costs, and error accumulation. More recently, 3DGS-based methods were developed to generate TDOM in more efficient manner, yet they show some degenerated rendering performance on sparse view scenarios, which is naturally common when dealing with boundary area of photogrammetric UAV images. To address the above issues, we introduce a hybrid method that integrates 3DGS with Few-Shot Gaussian Splatting (FSGS, Zhu et al. (2024)). Specifically, our method first partitions the UAV images into dense and sparse view scenarios based on image overlapping degree. Then, two specific 3DGS training solutions are employed: in dense-view scenarios, the standard 3DGS optimization is applied, in sparse-view scenarios, the FSGS framework is adopted, which incorporates a proximity-guided Gaussian unpooling strategy and monocular depth supervision, thereby enhancing adaptive density control and geometric guidance through improved constraints on Gaussians. Third, two trained Gaussians are merged. Finally, by substituting the perspective projection with the orthogonal projection, our method directly generates TDOM while eliminating the requirement for explicit Digital Surface Model (DSM) and occlusion detection. Extensive experimental results demonstrate that our method outperforms existing commercial software in several aspects while achieving superior orthophoto quality compared to 3DGS in sparse-view scenarios. Project Web: https://walterwang2024.github.io/HyGS-TDOM/

1. Introduction

True Digital Orthophoto Map (TDOM) has been a pivotal topic in the field of photogrammtery and remote sensing, which are widely applied in geometric quality assessment (Wang et al., 2017), land management(Szostak et al., 2014), building monitoring (Qin et al., 2016), environmental conservation (Akbari et al., 2003). The main step of conventional methods for TDOM generation is occlusion detection and compensation. Heretofore, many impressive methods have been proposed, including Z-Buffer-based, angles, heights, vector polygons, texture synthesis and object-oriented methods (Shin and Lee, 2021). However, in the field of photogrammetry, these methods typically involves multiple complex computational stages: feature extraction and matching, image orientation with sparse point cloud generation, dense matching, mesh model construction, digital surface model (DSM) generation, occlusion detection and compensation (Schonberger and Frahm, 2016a). Although deep learning-based TDOM generation methods simplify the intricate geometric processing pipeline and achieve superior results in most scenarios, they struggle to handle complex environments (e.g., vegetation) and have weak generalization (Shin and Lee, 2021). In recent years, Neural Radiance Fields (NeRF) (Mildenhall et al., 2021), which use implicit 3D representation to encode intricate scene geometry and lighting with high fidelity, have made impressive strides in 3D scene reconstruction and new view synthesis. Although NeRF has been successfully applied in TDOM generation (Lv et al., 2024), their adoption remains constrained by computationally intensive training processes and stringent real-time rendering requirements.

As a recent advancement, 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) shares the objective of high-fidelity 3D reconstruction with NeRF, but replaces NeRF's implicit neural representations with explicit Gaussians. This innovation endows 3DGS with both efficient training capabilities and real-time rendering performance exceeding 100 frames per second (FPS), establishing it as a highly promising alternative solution. PG-SAG (Wang et al., 2025b) further enhances 3DGS by introducing semantic-aware grouping and parallel optimization for large-scale urban reconstruction, enabling fine-grained results without sacrificing image resolution. Tortho-Gaussian (Wang et al., 2024) successfully applies 3DGS to the generation of TDOM, significantly improving production efficiency. Moreover, this method achieves high-quality orthophoto results in complex environments. However, their method exhibits geometric distortions and texture blurring in sparse-view scenarios, which is naturally common when dealing with boundary area of photogrammetric UAV images.

In response to the degradation issues in sparse-view scenarios, 3DGS-derived methods have spawned another noteworthy research focus: few-shot reconstruction. To address the inadequate Gaussian optimization caused by insufficient image coverage, FSGS (Zhu et al., 2024) introduces a proximity-guided gaussian unpooling strategy for densification, coupled with monocular depth supervision to enforce geometric constraints on newly generated gaussians. This framework ultimately achieves high-fidelity reconstruction under extremely sparse view conditions.

To address the degenerated rendering performance of 3DGS-based methods in sparse view scenarios, we present a novel TDOM generation method, named HyGS-TDOM, a Hybrid Gaussian Splatting Famework for generating TDOMs from

* Corresponding Author

both dense and sparse views. It integrates 3DGS and FSGS through a unified pipeline, achieving efficient TDOM generation while maintaining high orthophoto quality in both dense-view and sparse-view scenarios. Similar to the original 3DGS pipeline, HyGS-TDOM likewise follows the three-stage workflow of sparse point cloud input, Gaussian optimization, and differentiable rendering, but adopts a partitioning optimization mechanism. More specifically, the proposed method is mainly composed of five steps: image coverage dentification, 3DGS optimization in dense-view scenarios, FSGS optimization in sparse-view scenarios, 3D Gaussian field fusion, and orthophoto rendering. In the first step, We determine dense-view and sparse-view scenarios based on overlapping degree, then assign corresponding 3D sparse images and points. In the second step, Gaussians in dense-view scenarios are optimized by 3DGS method. In the third step, Gaussians in sparse-view scenarios are optimized by FSGS method to enhance reconstruction quality in boundary area. In the fourth step, the Gaussians trained by both methods are merged within the same coordinate system. The final step replaces the perspective matrix in 3DGS with an orthogonal projection matrix to generate orthophoto rendering results. Our main contributions can be summarized as twofold:

1. To address the boundary degradation issues in 3DGS-based TDOM methods, we present the first integration of few-shot reconstruction (FSGS) into TDOM generation.
2. We propose a hybrid TDOM generation framework that implements region-specific optimization. By differentially processing dense-view scenarios with standard 3DGS and sparse-view scenarios with FSGS, our method achieves comprehensive high-quality reconstruction across full scenes.

2. Related Work

A brief overview of TDOM generation methods in existing studies is given in this section, including occlusion detection-based methods, deep learning-based methods, differentiable rendering-based methods. In addition, three innovative methods on 3DGS for sparse reconstruction are introduced.

2.1 Occlusion Detection-Based TDOM

To detect occlusion, various strategies have been developed, including Z-buffer, angles, heights, vector polygons, texture synthesis, and object-oriented methods.

The Z-buffer technique, initially adapted from computer graphics, becomes foundational for TDOM occlusion handling. Amhar et al. (1998) pioneered its photogrammetric use by rasterizing 3D building models against a DTM, but artifacts near building edges were attributed to rasterization and model simplification. Zhou et al. (2005) introduced a multi-level Z-buffer with adaptive DSM smoothing to reduce urban artifacts, though their per-pixel depth sorting incurred high computational cost in complex scenes. However, these methods are sensitive to the DSM sampling interval, which may lead to incorrect occlusion and visibility assessments.

Habib et al. (2007) introduced an occlusion detection technique that calculates off-nadir angles between the lines of the perspective centers and the DSM pixels. However, this approach necessitates repeated scanning of object points and angle comparisons, thereby decreasing efficiency. The height-based approach is an alternative methodology for true orthophoto generation using satellite imagery, which determines whether the vision of a camera is occluded according to the elevation (Bang et al., 2007). Oliveira and Galo (2013) proposed a height-gradient-based occlusion detection method, in which Radial height gradients are computed to identify occlusion starting points, which are then projected onto the DSM for final occlusion detection.

Vector polygons-based methods, exemplified by Sheng (2007), employ a novel strategy of processing image pixels as vector blocks instead of discrete points to mitigate artifacts inherent in conventional Z-buffer techniques. Wang et al. (2018) matched line segments across multi-view images to determine seamlines and reconstructed a high-precision TIN (Triangulated Irregular Network) that reduce facade misalignments in the true orthophoto mosaicking process. Li et al. (2019) proposed a fusion algorithm based on the pulse-coupled neural network (PCNN) model. Zhou and Wang (2016) proposed a object-oriented occlusion detection method that uses a seed growth algorithm to identify ghosting artefacts caused by building occlusions in urban scenes.

2.2 Deep Learning-Based TDOM

Deep learning has increasingly been employed to mitigate occlusions and improve geometric fidelity in TDOM generation. Shin et al. (2020); Shin and Lee (2021) proposed true orthophoto generation based on a generative adversarial network (GAN) with a Pix2Pix model using DSM and LiDAR data, but their performance is highly dependent on the quality of the LiDAR data. Ebrahimikia and Hosseininaveh (2022) focused on geometric correction using reconstructed edge points from UAV stereo images, aligning building outlines through edge-guided mesh resampling. Most recently, Ebrahimikia et al. (2024) developed urban-SnowflakeNet, a specialized CNN that incorporates structural priors and edge saliency cues to enhance façade integrity and spatial coherence. However, these methods face challenges in generalizing to unseen urban geometries and often require task-specific training data.

2.3 Neural Rendering-based TDOM

Implicit Reconstruction: The rapid advancements in NeRF have profoundly transformed the 2D image reconstruction methods. Lv et al. (2024) are the first to employ implicit neural representations for the rapid generation of digital orthophotos, exploring the potential of implicit reconstruction techniques. Wei et al. (2024) extended NeRF to large-scale urban scenes by integrating scene-block training strategy and multiview consistency. Chen et al. (2024) proposed Ortho-NeRF, a UAV-oriented pipeline that introduces orthogonal rendering constraints and hierarchical sampling to better preserve structural integrity during orthophoto projection. Qu et al. (2024) generated more large area TDOM by inputting satellites images. Despite these advances, current NeRF-based pipelines face limitations in efficiency and generalization.

Explicit Reconstruction: Recently, 3DGS, as a differentiable rendering method that incorporates explicit spatial structural information, has started to supplant NeRF. Wang et al. (2024) introduced Tortho-Gaussian, the first framework to leverage 3DGS for TDOM generation. The method introduces an orthogonal splatting technique designed for rendering scale-uniform images and enhances the 3D Gaussian representation through a fully anisotropic Gaussian. Wang et al. (2025a) proposed a 2D Gaussian splatting approach for efficient spatial reconstruction without requiring full 3D modeling. Yang et al.

(2025) introduced Ortho-3DGS, in which Gaussians are optimized with depth supervision and gradient-based refinement. While achieving real-time rendering and high geometric precision, these methods remain sensitive to sparse view conditions.

2.4 Sparse View Reconstruction

Zhang et al. (2024) proposed CoR-GS to address overfitting and inconsistency in sparse-view by co-training two Gaussian fields and using their disagreement for regularization and pruning, significantly improving reconstruction quality. Xiong (2024) introduced SparseGS, which integrates depth priors and a novel Unseen Viewpoint Regularization (UVR) to mitigate floaters and background collapse, enabling real-time rendering from as few as 3 images. Zhu et al. (2024) developed FSGS, which grows Gaussians via a proximity-guided unpooling strategy, aided by synthetic views and monocular depth priors to ensure high-quality reconstruction. Therefore, the derivative methods of sparse view 3DGS make it feasible to address the TDOM quality degradation issue in boundary areas encountered in this work.

3. Methodology

This section provides detailed explanations of our HyGS-TDOM on generating TDOM, which contains three technical parts: preliminaries of 3DGS and FSGS, the framework of our hybrid optimization method, and orthophoto rendering.

3.1 preliminary

To make this paper more self-contained, we next outline basics of 3DGS and FSGS.

3.1.1 3DGS The 3DGS pipeline reconstructs a radiance field from multi-view images by representing the scene as a dynamic set of anisotropic 3D Gaussian ellipsoids which are initialized by sparse points using the SfM Schönberger and Frahm, 2016b algorithm. Each ellipsoid is defined by a set of attributes, including position, covariance, opacity, and color. During the training step, the screen is divided into 16×16 pixel tiles, with each tile exclusively processing ellipsoids located within the view frustum. Each ellipsoid's position and covariance matrix are projected onto the image plane and assigned unique identifiers, followed by GPU-accelerated radix sorting. For each pixel, contributions are computed in depth order and composited using alpha blending to ensure coherent rendering. The rendered 2D image is then compared with the original input images to compute the loss function, which drives gradient-based optimization of Gaussian ellipsoid parameters.

A Gaussian ellipsoid G_x, centered at μ, with a covariance matrix given by Σ is represented by a 3D anisotropic Gaussian distribution.

$$G_i(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right) \tag{1}$$

where x and μ are 3×1 vectors, while Σ is a 3×3 covariance matrix. To guarantee the positive semi-definiteness of Σ, it is further parameterized by a rotation matrix R and a scaling matrix S:

$$\boldsymbol{\Sigma} = \mathbf{RSS}^\top \mathbf{R}^\top \tag{2}$$

where, the scaling matrix and rotation matrix are represented by a scaling factor and a quaternion, respectively, allowing them to be updated during the training process. According to the affine transformation propagation theorem, the Jacobian matrix is introduced:

$$\Sigma' = JW\Sigma W^T J^T \tag{3}$$

where W represents the viewpoint transformation matrix, while J represents the Jacobian matrix associated with the projective transformation whin its affine approximation. After depth sorting, the radiance field employs an alpha blending algorithm for pixel rendering which progressively accumulates colors based on the opacity of Gaussian ellipsoids:

$$C = \sum_{n=1}^{N'} c_n \alpha_n^t T_n, \tag{4}$$

where

$$\alpha_n^t = \alpha_n \times \exp\left(-\frac{1}{2}(\mathbf{x}'-\boldsymbol{\mu}_n^t)^\top (\boldsymbol{\Sigma}_n^t)^{-1}(\mathbf{x}'-\boldsymbol{\mu}_n^t)\right), \tag{5}$$

$$T_n = \prod_{i=1}^{n-1}(1-\alpha_i^t). \tag{6}$$

c_i represents the predicted color. The final splatting opacity α_n^t is obtained by multiplying the predicted opacity α_n with the splatted 2D Gaussian distribution. x^t and μ_n^t are coordinates splatted in 2D images plane.

3.1.2 FSGS To address degenerated reconstruction in sparse view scenarios, FSGS introduces two enhancements that extend the original 3DGS framework to operate robustly with 2–5 input views. The core idea of FSGS is to improve both the density control and geometry supervision of Gaussian ellipsoids in scenarios where multi-view consistency is not available. This is achieved through two key mechanisms: (1) a proximity-guided unpooling strategy, which enhances the adaptive density control mechanism in 3DGS; (2) monocular depth supervision from pseudo-novel views, which supplements the photometric loss in 3DGS with geometry guidance, thereby improving reconstruction fidelity in sparse-view scenarios.

Proximity-guided Gaussian Unpooling: 3DGS is initialized from SfM points, and its performance is strongly dependent on quality of the initialized points. FSGS constructs a directed graph, referred to as the proximity graph, to connect each existing ellipsoid with its nearest K neighbors by computing the proximity

$$D_i^K = K - min(d_{ij}) \tag{7}$$

where, d_ij is calculated via $d_{ij} = \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|$, representing the Euclidean distance among the centers of ellipsoid G_i and ellipsoid G_j. The assigned proximity score P_i to Gaussian G_i is calculated as the average distance to its K nearest neighbors:

$$P_i = \frac{1}{K}\sum_{j=1}^{K} D_i^K \tag{8}$$

If the proximity score of a Gaussian exceeds the threshold t, new ellipsoids will grow at the center of each edge, connecting the "source" and "destination" ellipsoids.

Geometry Guidance with Pseudo Views and Depth Supervision: In contrast to 3DGS that employs only original input images, FSGS employs pseudo views from unobserved perspectives for augmentation. Specifically, it synthesizes novel views

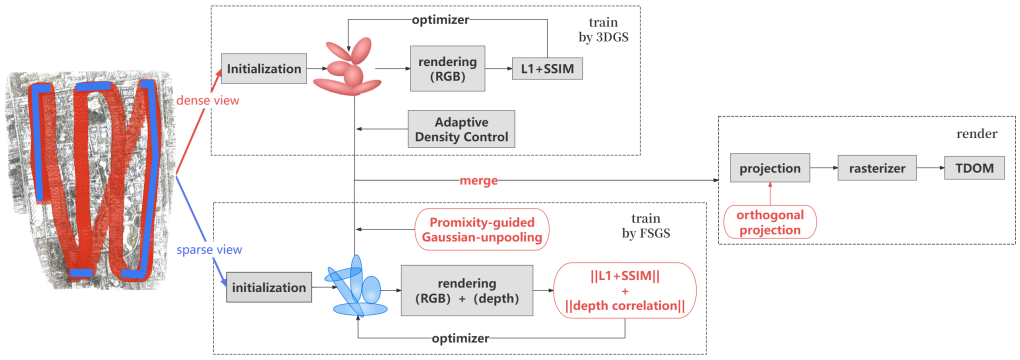

Figure 1. Overview of our HyGS-TDOM. Through two specific training solutions, our method can generate TDOM that maintain high quality in both dense-view and sparse-view scnearios.

by sampling between the two nearest cameras: first computing the average orientation of these cameras, then interpolating a pseudo view with intermediate viewing direction. Compared to the loss function in 3DGS, the FSGS framework incorporates an additional monocular depth loss to strengthen geometric constraints, which is implemented using a pre-trained depth estimator. For each input view, the system generates pseudo-depth maps (D_est) using the pre-trained monocular depth model, while the rendered depth (D_ras) of the Gaussian field is computed via z-buffer accumulation. Depth loss function is as follows:

$$\mathrm{Corr}(\hat{\boldsymbol{D}}_{\mathrm{ras}}, \hat{\boldsymbol{D}}_{\mathrm{cst}}) = \frac{\mathrm{Cov}(\hat{\boldsymbol{D}}_{\mathrm{ras}}, \hat{\boldsymbol{D}}_{\mathrm{cst}})}{\sqrt{\mathrm{Var}(\hat{\boldsymbol{D}}_{\mathrm{ras}})\mathrm{Var}(\hat{\boldsymbol{D}}_{\mathrm{cst}})}} \tag{9}$$

Combing all together, the training loss :

$$\begin{aligned}\mathcal{L}(G, C) = &\lambda_1 \underbrace{\left\| C - \hat{C} \right\|_1}_{\mathcal{L}_1} + \lambda_2 \underbrace{\text{D-SSIM}(C, \hat{C})}_{\mathcal{L}_{\text{psim}}} \\ &+ \lambda_3 \underbrace{\left\| \mathrm{Corr}(D_{\mathrm{ras}}, D_{\mathrm{cst}}) \right\|_1}_{\mathcal{L}_{\text{regularization}}}\end{aligned} \tag{10}$$

where L_1, and L_{ssim} stands for the photometric loss term between predicted images and input images.

3.2 Overview of the Proposed Framework

Our HyGS-TDOM aims to generate high-fidelity TDOM by integrating 3DGS and FSGS under a region-specific optimization framework.

As illustrated in Fig. 1, upon receiving the sparse point cloud input, we divide it into two subsets: one subset retains a complete copy of the point cloud and undergoes standard 3DGS pipeline processing to optimize Gaussian ellipsoids in dense-view regions; the other subset undergoes view transformation to extract point clouds from boundary areas, which are then processed using the FSGS framework incorporating proximity-guided unpooling and monocular depth supervision to address boundary degenerated issues. Following separate optimization procedures, our method then merges the resulting Gaussian sets to construct a unified 3D scene representation. For orthophoto generation, we replace the default perspective projection with an orthographic projection during the rendering stage. Our work enables pixel-level rendering that maintains vertical-view geometric consistency, ultimately producing TDOM results with images of various overlapping degree and occlusion-aware capabilities.

3.2.1 3D Gassuian optimzation To address the prevalent issue of uneven spatial view coverage in UAV images, especially in edge flight strips, where features often have only 2-4 image coverages, we employ a manual partitioning strategy to separately process dense and sparse regions.

Specifically, we first train the 3DGS pipeline using the complete sparse point cloud to obtain well-reconstructed central areas. For the sparse-view region, we manually select a subset of input images that correspond to edge-covering views. Then, during the "create-from-cloud" function of 3D Gaussian initialization, we filter the sparse point cloud using the image-to-point associations recorded in point3D document, extracting only those 3D points observed by the selected edge-view images. These filtered points are then instantiated as Gaussian ellipsoids and passed exclusively to the FSGS pipeline for training.

3.2.2 Fusion of ellipsoids from original 3DGS and FSGS To integrate the results of the partitioning optimization, a spatial fusion of the Gaussian ellipsoids generated independently by the 3DGS and FSGS pipelines is conducted. Due to the fact that both ellipsoids originate from the same SfM point cloud and share a unified global coordinate system, they can be directly aligned. After the two sets of ellipsoids are independently optimized, they are directly merged by concatenation, without requiring any spatial registration.

The process begins by identifying a central camera, selected as the one closest to the centroid of all camera centers. Its viewpoint transformation matrix is used to define the canonical orientation for merging. Using this reference, the optimized ellipsoids produced by both pipelines are loaded and rotated into the same coordinate system. With the ellipsoids aligned, a rectangular cropping box is defined around the central view, representing the dense-view region confidently handled by the 3DGS branch. We then apply spatial filtering: remove 3DGS ellipsoids outside the box and FSGS ellipsoids inside it. This complementary partitioning avoids overlap and guarantees a smooth transition between dense and sparse regions. Once filtered, the remaining ellipsoids are merged into a unified point cloud. The result is a spatially coherent representation that retains high-fidelity detail in the central area and effectively extends reconstruction to edge regions with sparse coverage.

3.3 Orthographic Projection for TDOM Rendering

To produce a geometrically accurate TDOM, the final Gaussian ellipsoids is rendered using orthogonal projection, replacing the default perspective projection in the 3DGS pipeline in Fig. 2. This type of projection accurately eliminates building facades, fulfilling the fundamental requirements for TDOM.

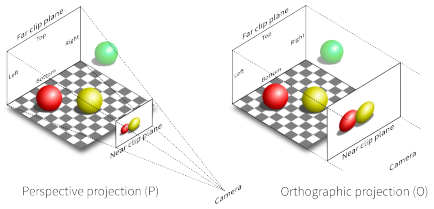

Figure 2. perspective projection and orthogonal projection (Shirley et al., 2009)

In orthophoto rendering, both the mean and the covariance of each ellipsoid need to be projected appropriately. For the mean, the standard perspective transformation is replaced with an orthogonal projection matrix P, defined as:

$$P_o = \begin{pmatrix} \frac{2}{r-1} & 0 & 0 & -\frac{r+l}{l-u} \\ 0 & \frac{2}{r-u} & 0 & -\frac{t+b}{l-u} \\ 0 & 0 & \frac{2}{z_l-z_n} & \frac{z_l+z_n}{z_l-z_n} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{11}$$

This matrix maps 3D world coordinates into an orthogonal clip space bounded by the parameters l, r, b, t, z_n, z_f, corresponding to the left, right, bottom, top, near, and far planes of the viewing box. Unlike perspective projection, this mapping maintains parallel lines and constant scale regardless of depth. Since the orthogonal projection is affine, we can linearly approximate its effect on the covariance using the Jacobian matrix. The 2D projected covariance becomes:

$$\Sigma' = J\Sigma J^T \tag{12}$$

Under orthogonal conditions, J becomes much simpler than in the perspective case and remains constant for all points in view, simplifying computation. Differentiating projected coordinates, we can obtain:

$$J_0 = \begin{pmatrix} \frac{2}{r-l} & 0 & 0 \\ 0 & \frac{2}{t-b} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{13}$$

With both mean and covariance correctly transformed, the renderer can blend overlapping ellipsoids on the 2D plane. This orthogonal rendering strategy enables the accurate generation of TDOM outputs from the fused 3D representation.

4. Experiments

In this section, comprehensive experimental results are reported to demonstrate the efficacy of the proposed HyGS-TDOM. We conduct qualitative and quantitative evaluations along with ablation studies on multiple datasets.

4.1 Experiment settings

Experimental Data: In this study, we conduct experiments using the NPU DroneMap dataset(Bu et al., 2016), containing 1920×1080 resolution aerial videos (30Hz) captured across diverse regions in China (Shaanxi/Henan/Hubei) with varying land cover types and terrain characteristics.

To better simulate the sparse view, We manually conducted edge-image sparsification on the npu dataset, ensuring that peripheral objects were covered by only 2–4 views.

Experiments protocols: The experimental evaluation in this paper is conducted from three perspectives: qualitative assessment, quantitative assessment, and ablation study. First, for qualitative assessment, comprehensive visual comparisons are conducted between several traditional photogrammetric software, including ContextCapture, MetaShape, and Pix4DMapper. The comparison focuses on multiple aspects, including building edges, building facades, weak textures, slender structures, and vegetation areas. Second, for the quantitative assessment, we evaluate the relative accuracy using MetaShape and Pix4DMapper as reference. Finally, we conduct ablation studies to analyze the effect of critical parameters in our framework. (1) Sampling resolution: Experiments are conducted under varying spatial resolutions to examine their influence on the level of detail in the generated TDOM. (2) Overlap: By incrementally reducing the number of input images, we evaluate the reconstruction performance of 3DGS and FSGS under varying overlap conditions.

Experimental Detail: In the 3DGS training stage, the number of iterations was set to 30,000. For FSGS training, 10,000 iterations were used. For hyperparameters in the FSGS pipeline, depth-pseudo-weight was set to 0.03, and the sample-pseudo-interval was configured to 50 steps. During fusion, for the NPU dataset, a rectangular bounding box was applied with the range of (–4, 2) along the x-axis and (–4, 4) along the y-axis. All experiments were conducted on four NVIDIA GeForce RTX 4090 GPUs.

4.2 Qualitative Evaluation

As shown in Fig. 3, detailed TDOM patches generated by the four methods are presented.

Building Edges: Building edges are expected to conform to the actual geometric structure of buildings, without irregular distortions, and the seams between adjacent buildings should be accurately aligned. As illustrated in Fig. 3 (columns (a) and (b)), the TDOMs generated by ContextCapture, Metashape, and Pix4DMapper exhibit varying degrees of structural distortion and misalignment at building boundaries (highlighted by red boxes). In contrast, our HyGS-TDOM method maintains more complete structures in the same regions, with sharper edges and higher overall geometric consistency.

Building facades: The extent to which building facades can be fully resolved reflects the effectiveness of occlusion detection in the true orthophoto generation process. As shown in Fig. 3 (columns (c) and (d)), ContextCapture and MetaShape fail to completely eliminate the façades in the orthophoto generation process, resulting in varying degrees of façade visibility in the final images. In contrast, our HyGS-TDOM method effectively remove building façades. The results strictly preserve the top-down view, closely aligning with the intended characteristics of TDOM representations.

Weak Textures: In weak-texture regions such as water surfaces, as illustrated in column (e), ContextCapture suffer from

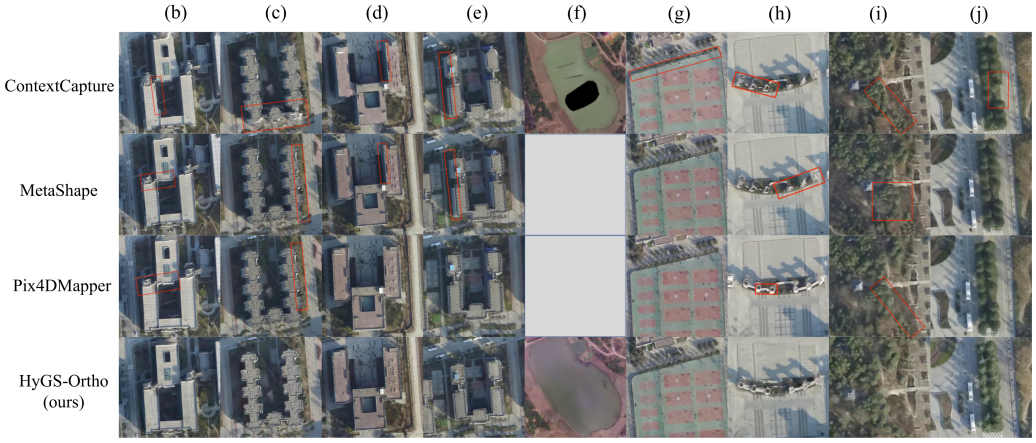

Figure 3. Qualitative comparison of TDOMs generated by commercial softwares and our method on NPU DroneMap dataset.

texture blurring and geometric drift resulting in visible holes. In contrast, our method demonstrates greater stability in weak-texture regions. The resulting TDOM show no significant geometric distortion or drift, thereby ensuring improved visual coherence and representational quality.

Slender Structures: Slender structures, such as railings and stair edges, are typically difficult to represent accurately in TDOM due to their narrow geometry. As shown in columns (f) and (g), distortions appear in the railings in ContextCapture output. Both Metashape and Pix4DMapper struggle to preserve the original architectural curvature and continuity in the statue sections of the statue. In contrast, our method maintains the geometric continuity and boundary clarity of these fine-scale structures during the training process.

Vegetation Area: In vegetation areas, the non-rigid nature often leads to displacement and deformation during UAV flights. As illustrated in columns (h) and (i), ContextCapture produce noticeable color discontinuities along the edges of trees. The other two methods demonstrate blurring effects of varying severity. Our method produces more natural imagery. The transitions to surrounding regions are smooth, especially at the boundaries between the edge of the tree and the surface of the ground.

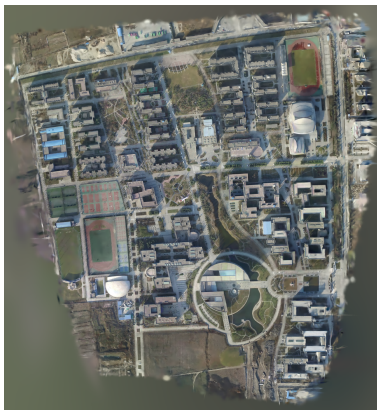

Figure 4. The final TDOM by HyGS-TDOM

4.3 Quantitative Evaluation

Due to the absence of 3D ground control points, we evaluated the relative accuracy of the TDOMs using Metashape and Pix4DMapper as reference baselines. On the generated TDOMs, we randomly selected eight sets of line segments for measurement. The first six sets were located in dense-view regions, while the last two were selected from boundary regions. Each set contains a pair of line segments defined by building corners, and the ratio of their lengths is used as an indicator of TDOM consistency. Ideally, this ratio should remain stable across all TDOMs.

We then calculated the absolute and relative errors of these length ratios, as summarized in Table 1. The average relative and absolute errors were 0.418% and 0.003113 between HyGS-TDOM and Metashape, and 0.570% and 0.005152 between HyGS-TDOM and Pix4DMapper. It is noteworthy that Metashape exhibited significant geometric distortions in edge1 and edge2, resulting in notably larger relative and absolute errors. Therefore, the corresponding measurements in these regions were excluded. These findings demonstrate that the proposed HyGS-TDOM framework achieves a level of geometric accuracy comparable to commercial software, confirming its reliability for precise reconstruction tasks.

4.4 Ablation Studies

This section conducts ablation studies from two perspectives: spatial resolution and overlap.

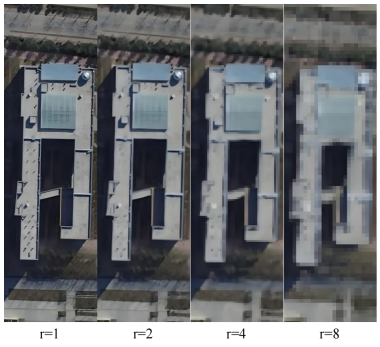

Figure 5. Results of TDOM using different Spatial Resolutions. From left to right, the spatial resolution decreases.

4.4.1 spatial resolution: Varying the spatial resolution influences both the final resolution of the generated TDOM and the number of ellipsoids present within each tile. To assess this effect, experiments were conducted at four different spatial resolution scales: 1, 1/2, 1/4, and 1/8.

Table 1. Accuracy Evaluation Based on MetaShape and Pix4DMapper

ID	Ours	MetaShape			Pix4DMapper		
	Ratio	Ratio	Relative Error(%)	Absolute Error	Ratio	Relative Error(%)	Absolute Error
1	0.99500	1.00000	0.50000	0.005000	1.00297	0.79438	0.007968
2	0.81868	0.82255	0.47080	0.003873	0.82120	0.30701	0.002521
3	0.29167	0.29296	0.44071	0.001291	0.29268	0.34722	0.001016
4	1.48958	1.49051	0.06192	0.000923	1.48640	0.21384	0.003178
5	0.77083	0.77750	0.85745	0.006667	0.77647	0.72601	0.005637
6	0.51623	0.50531	0.18002	0.000928	0.52153	1.01573	0.005297
Edge1	1.79302	~~1.82150~~	~~1.07490~~	~~0.019483~~	1.80147	0.46924	0.008453
Edge2	1.04437	~~1.01777~~	~~2.61377~~	~~0.026602~~	1.03722	0.68913	0.007148
Mean	-	-	**0.41848**	**0.003113**	-	**0.57032**	**0.005152**

As shown in Fig. 5, the proposed method successfully captures ground objects at all tested resolution levels, indicating its capability to produce TDOM products at multiple levels of detail. A clear improvement in image fidelity and structural detail is observed as the spatial resolution increases toward the original image resolution. This improvement can be attributed to the behavior of ellipsoids during the alpha-blending process.

4.4.2 overlap: The number of views is used to express the percentage of forward overlap. To investigate the performance of 3DGS and FSGS in boundary areas, a specific area was selected from the dataset. All 16 images containing this region were identified, and eight subsets were constructed, containing 2, 4, 6, 8, 10, 12, 14, and 16 images, respectively. Each subset was used to train and render with both the 3DGS and FSGS methods. Visual and quantitative comparisons were performed.

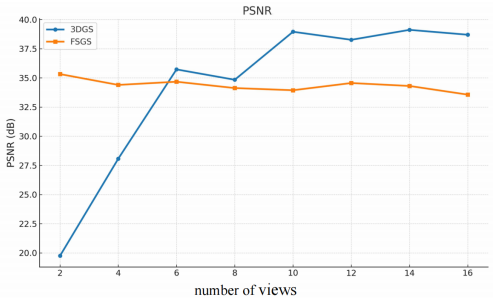

Figure 6. Comparison under different numbers of views.

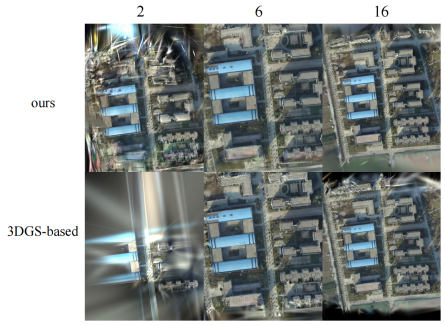

Figure 7. Visualization effects for different number of views.

As shown in Fig. 6, under extremely sparse conditions (number of views < 6), the PSNR of the FSGS reaches approximately 35, whereas that of the 3DGS remains around 20, indicating that FSGS significantly outperforms 3DGS in sparse view scenarios. Under moderately sparse conditions ($6 <$ number of views < 10), both methods achieve comparable PSNR values around 35, suggesting similarly high reconstruction quality. When sufficient views are available (more than 10), 3DGS surpasses FSGS. This observation suggests that the pseudo-depth self-supervised framework positively contributes to scene reconstruction in sparse. However, in dense view, the input views already offer sufficient structural cues, incorporating pseudo-depth may introduce cumulative errors stemming from depth estimation noise, ultimately leading to a decline in reconstruction accuracy. In addition, Fig. 7 demonstrates the visualization effects.

5. Conclusion

This paper proposes a hybrid training framework, HyGS-TDOM, that integrates 3D Gaussian Splatting (3DGS) and Few-Shot Gaussian Splatting (FSGS) for the generation of True Digital Orthophoto Map (TDOM). By leveraging the superior performance of FSGS in extremely sparse-view scenarios,our hybrid method can address the quality degradation that often occur in boundary area with limited views using conventional 3DGS-based methods. Experimental results demonstrate that HyGS-TDOM achieves accuracy and visual quality comparable to commercial software. In the future, two promising directions may be explored. First, the selection of edge-strip images in this study was performed manually. Automated identification of edge regions could improve the overall processing efficiency. Then, while all experiments in this paper are conducted in an offline setting, 3DGS possesses the potential for real-time rendering. Therefore, enabling real-time TDOM generation represents a valuable and practical extension of this work.

Acknowledgment

This work was jointly supported by the National Natural Science Foundation of China (42301507) and the Natural Science Foundation of Hubei Province, China (2022CFB727).

References

Akbari, H., Rose, L. S., Taha, H., 2003. Analyzing the land cover of an urban environment using high-resolution orthophotos. *Landscape and urban planning*, 63(1), 1–14.

Amhar, F., Jansa, J., Ries, C. et al., 1998. The generation of true orthophotos using a 3D building model in conjunction with a conventional DTM. *International Archives of Photogrammetry and Remote Sensing*, 32, 16–22.

Bang, K., Habib, A. F., Shin, S., Kim, K., 2007. Comparative analysis of alternative methodologies for true ortho-photo generation from high resolution satellite imagery. *ASPRS ANNUAL*, 2007.

Bu, S., Zhao, Y., Wan, G., Liu, Z., 2016. Map2dfusion: Real-time incremental uav image mosaicing based on monocular slam. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 4564–4571.

Chen, S., Yan, Q., Qu, Y., Gao, W., Yang, J., Deng, F., 2024. Ortho-NeRF: generating a true digital orthophoto map using the neural radiance field from unmanned aerial vehicle images. *Geo-spatial Information Science*, 1–20.

Ebrahimikia, M., Hosseininaveh, A., 2022. True orthophoto generation based on unmanned aerial vehicle images using reconstructed edge points. *The Photogrammetric Record*, 37(178), 161–184.

Ebrahimikia, M., Hosseininaveh, A., Modiri, M., 2024. Orthophoto improvement using urban-SnowflakeNet. *Applied Geomatics*, 16(2), 387–407.

Habib, A. F., Kim, E.-M., Kim, C.-J., 2007. New methodologies for true orthophoto generation. *Photogrammetric Engineering & Remote Sensing*, 73(1), 25–36.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4), 139–1.

Li, X., Yan, H., Yang, S., Niu, L., 2019. A fusion algorithm of multispectral remote sensing image and aerial image. *Remote Sensing Information*, 34(4), 11–15.

Lv, J., Jiang, G., Ding, W., Zhao, Z., 2024. Fast digital orthophoto generation: A comparative study of explicit and implicit methods. *Remote Sensing*, 16(5), 786.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), 99–106.

Oliveira, H., Galo, M., 2013. Occlusion detection by height gradient for true orthophoto generation, using LiDAR data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 275–280.

Qin, R., Tian, J., Reinartz, P., 2016. 3D change detection–approaches and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122, 41–56.

Qu, Y., An, X., Chen, S., Deng, F., 2024. Satellite true digital orthophoto map generation without elevation data: a new NeRF-based method. *Remote Sensing Letters*, 15(3), 258–269.

Schonberger, J. L., Frahm, J.-M., 2016a. Structure-from-motion revisited. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4104–4113.

Schönberger, J. L., Frahm, J.-M., 2016b. Structure-from-motion revisited. *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Sheng, Y., 2007. Minimising algorithm-induced artefacts in true ortho-image generation: a direct method implemented in the vector domain. *The Photogrammetric Record*, 22(118), 151–163.

Shin, Y. H., Hyung, S. W., Lee, D.-C., 2020. True orthoimage generation from LiDAR intensity using deep learning. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, 38(4), 363–373.

Shin, Y. H., Lee, D.-C., 2021. True Orthoimage Generation Using Airborne LiDAR Data with Generative Adversarial Network-Based Deep Learning Model. *Journal of Sensors*, 2021(1), 4304548.

Shirley, P., Ashikhmin, M., Marschner, S., 2009. *Fundamentals of computer graphics*. AK Peters/CRC Press.

Szostak, M., Wezyk, P., Tompalski, P., 2014. Aerial orthophoto and airborne laser scanning as monitoring tools for land cover dynamics: A case study from the Milicz Forest District (Poland). *Pure and Applied Geophysics*, 171(6), 857–866.

Wang, M., Cheng, Y., Chang, X., Jin, S., Zhu, Y., 2017. On-orbit geometric calibration and geometric quality assessment for the high-resolution geostationary optical satellite GaoFen4. *ISPRS Journal of Photogrammetry and Remote Sensing*, 125, 63–77.

Wang, Q., Yan, L., Sun, Y., Cui, X., Mortimer, H., Li, Y., 2018. True orthophoto generation using line segment matches. *The Photogrammetric Record*, 33(161), 113–130.

Wang, Q., Zhan, Z., He, J., Tu, Z., Zhu, X., Yuan, J., 2025a. High-Quality Spatial Reconstruction and Orthoimage Generation Using Efficient 2D Gaussian Splatting. *arXiv preprint arXiv:2503.19703*.

Wang, T., Wang, X., Hou, Y., Xu, Y., Zhang, W., Zhan, Z., 2025b. PG-SAG: Parallel Gaussian Splatting for Fine-Grained Large-Scale Urban Buildings Reconstruction via Semantic-Aware Grouping. *PFG(2025)*. https://doi.org/10.1007/s41064-025-00343-0.

Wang, X., Zhang, W., Xie, H., Ai, H., Yuan, Q., Zhan, Z., 2024. Tortho-Gaussian: Splatting True Digital Orthophoto Maps. *arXiv preprint arXiv:2411.19594*.

Wei, J., Zhu, G., Chen, X., 2024. NeRF-Based Large-Scale Urban True Digital Orthophoto Map Generation Method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Xiong, H., 2024. *SparseGS: Real-time 360° sparse view synthesis using Gaussian splatting*. University of California, Los Angeles.

Yang, J., Cai, Z., Wang, T., Ye, T., Gao, H., Huang, H., 2025. Ortho-3DGS: True Digital Orthophoto Generation from Unmanned Aerial Vehicle Imagery Using the Depth-Regulated 3D Gaussian Splatting. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Zhang, J., Li, J., Yu, X., Huang, L., Gu, L., Zheng, J., Bai, X., 2024. Cor-gs: sparse-view 3d gaussian splatting via co-regularization. *European Conference on Computer Vision*, Springer, 335–352.

Zhou, G., Chen, W., Kelmelis, J. A., Zhang, D., 2005. A comprehensive study on urban true orthorectification. *IEEE Transactions on Geoscience and Remote sensing*, 43(9), 2138–2147.

Zhou, G., Wang, Y., 2016. Occlusion detection for urban aerial true orthoimage generation. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 3009–3012.

Zhu, Z., Fan, Z., Jiang, Y., Wang, Z., 2024. Fsgs: Real-time few-shot view synthesis using gaussian splatting. *European conference on computer vision*, Springer, 145–163.