State-of-the-art 3D reconstruction pipelines for photogrammetry in 2025

Your current HLOC + SuperPoint + LightGlue + GLOMAP pipeline represents an **excellent foundation** that aligns with cutting-edge practices—but targeted upgrades can yield meaningful quality improvements. The most impactful changes are switching to **ALIKED for feature extraction** (better geometric invariance, permissive licensing), adopting **MASt3R for challenging image pairs** (30% improvement on wide-baseline matching), and using **GOF or PGSR for mesh extraction** from trained Gaussian splats. For end-to-end scenarios, **InstantSplat** now enables 30x faster reconstruction without traditional SfM when COLMAP fails.

Feature extraction: ALIKED outperforms SuperPoint with faster inference

SuperPoint remains a solid baseline, but **ALIKED** (CVPR 2022, updated through 2024) delivers superior performance through its Sparse Deformable Descriptor Head, which provides geometric invariance via learned deformable sampling. On the KITTI visual odometry benchmark, ALIKED achieves the best Relative Pose Error while maintaining **125+ FPS** at VGA resolution—roughly 20x faster than SuperPoint's 6-7 FPS. The BSD-3-Clause license GitHub eliminates SuperPoint's restrictive non-commercial terms.

XFeat (CVPR 2024) offers the best speed-accuracy tradeoff when real-time performance matters, achieving 150+ FPS on GPU with 64-dimensional descriptors github that still match SuperPoint's accuracy on MegaDepth benchmarks. For maximum keypoint density, **DeDoDe v2** (CVPRW 2024) benefits from extracting up to **30,000 keypoints**—useful for high-density reconstruction scenarios.

Scenario	Recommended Extractor	Configuration
Drone/aerial footage	ALIKED + LightGlue	8,000 keypoints, resize_max 1600
Fisheye/wide-angle	ALIKED (deformable heads handle distortion)	Undistort first if severe
360° equirectangular	ALIKED via cubemap projection	Process 6 faces separately
iPhone/smartphone	XFeat or ALIKED-T	4,096 keypoints, works on CPU
Large datasets (6000+)	ALIKED + NetVLAD retrieval	Top 50 neighbors, not exhaustive

For HLOC integration, LightGlue directly supports ALIKED via `features='aliked'`. The sub-pixel accuracy upgrade from ECCV 2024 ("Learning to Make Keypoints Sub-Pixel Accurate") adds only ~7ms overhead and works with SuperPoint, ALIKED, and XFeat.

Feature matching: LightGlue excels for speed while MASt3R dominates challenging cases

LightGlue remains the optimal choice for standard workflows—its adaptive early-exit mechanism achieves **35-40ms per pair** while maintaining competitive accuracy (56.4 AUC@5° on MegaDepth). For your 6,000+ image datasets, this speed advantage compounds significantly.

However, **MASt3R** (ECCV 2024) represents a paradigm shift for challenging scenarios. By treating matching as a 3D task using pointmap regression, it arXiv achieves **93% VCRE AUC** on the Map-free localization benchmark—a 30% absolute improvement arXiv over LoFTR. MASt3R handles extreme viewpoint changes up to 180° Ecva and eliminates the need for RANSAC in pose estimation. The tradeoff is speed: ~300ms per pair versus LightGlue's 35ms.

RoMa v2 (November 2025) sets the new state-of-the-art across all six standard evaluation datasets, leveraging DINOv3 features with a decoupled matching-then-refinement pipeline. It's 1.7x faster than RoMa v1 with reduced memory usage through custom CUDA kernels.

For Gaussian splatting preparation, dense matchers (RoMa, DKM, MASt3R) produce superior results in **texture-poor regions** where sparse matching fails. The Dense-SfM approach (CVPR 2024) combines dense matching with Gaussian Splatting-based track extension, significantly improving reconstruction completeness on ETH3D benchmarks—though at 5-10x slower SfM times.

Recommended hybrid strategy: Use LightGlue for primary matching, then apply MASt3R selectively for image pairs that fail initial matching or show low match counts. The **MP-SfM** (CVPR 2025) configuration `sp-mast3r` combines SuperPoint keypoints with MASt3R dense features for robust reconstruction with precise localization.

SfM backends: GLOMAP validated as optimal for large datasets

Your GLOMAP choice is validated by benchmarks. On ETH3D SLAM sequences, GLOMAP achieves **8% higher recall** than COLMAP with **+9 AUC points** at the 0.1m threshold. The speed advantage is transformative: GLOMAP completes the south-building dataset in 89.8 seconds GitHub versus COLMAP's multi-hour runtime—roughly **10-50x faster** for large scenes.

GLOMAP's joint global positioning of cameras and 3D points eliminates drift accumulation inherent to incremental SfM. This matters especially for sequential video captures and nearly co-linear camera motion, where traditional translation averaging fails.

VGGSfM (CVPR 2024, Facebook Research) offers a fully differentiable alternative that ranked first in the CVPR24 IMC Challenge for camera pose estimation. It outputs COLMAP-compatible formats and handles up to 1,000+ frames in video mode with sliding windows. VGGSfM excels when you need dense depth export (via integrated Depth-Anything-V2) or dynamic object filtering through masks.

MASt3R-SfM (3DV 2025) represents the most radical alternative—a drop-in COLMAP replacement that scales to 1,000+ images with quasi-linear complexity. It eliminates Bundle Adjustment entirely through gradient-based 3D+2D optimization. Consider this for scenes where traditional feature matching consistently fails.

For optimal GLOMAP configuration with large datasets:

```bash
glomap mapper \
    --database_path /path/to/database.db \
    --image_path /path/to/images \
    --output_path /path/to/sparse \
    --TrackEstablishment.max_num_tracks 6000000 \  # ~1000 × n_images
    --GlobalPositioning.max_num_iterations 100 \
    --BundleAdjustment.max_num_iterations 100
```

After reconstruction, run `colmap point_triangulator` for denser point cloud initialization if targeting Gaussian splatting.

End-to-end pipelines bypassing traditional SfM show promise for sparse views

InstantSplat combines MASt3R geometry priors with Gaussian Splatting to achieve **30x acceleration** over traditional SfM + 3DGS pipelines. On sparse-view benchmarks, it improves SSIM from 0.3755 to 0.7624 compared to traditional pipelines, with reconstruction completing in under one minute on A100. InstantSplat supports 2DGS, 3DGS, and Mip-Splatting backends.

Splatt3R enables zero-shot Gaussian splatting from uncalibrated image pairs at 4 FPS—useful for quick previews before committing to full reconstruction. **DUSt3R** provides foundational geometry but struggles with scale ambiguity and is limited to ~32 images at 512×384 resolution due to quadratic memory scaling.

These methods excel when **COLMAP consistently fails**: sparse-view scenarios, challenging lighting, texture-less surfaces, or uncalibrated casual captures. For archival/heritage reconstruction with adequate overlap, traditional HLOC+GLOMAP still produces superior metric accuracy.

Mesh extraction: GOF and PGSR deliver highest quality from Gaussian splats

Gaussian Opacity Fields (GOF) from SIGGRAPH Asia 2024 currently **surpasses all 3DGS-based methods** in surface reconstruction quality. It defines an opacity field as the minimum opacity over all training views, enabling direct surface extraction via Marching Tetrahedra without requiring Poisson reconstruction or TSDF fusion. GOF handles unbounded scenes with adaptive mesh resolution, completing in ~24 minutes for Tanks & Temples scenes.

PGSR (TVCG 2024) achieves state-of-the-art Chamfer distance of **0.47** on DTU through planar-based Gaussian compression with unbiased depth rendering. Zju3dv arXiv Training completes in ~30 minutes—roughly 100x faster than NeRF-based methods—with direct UV texture extraction. arXiv

Method	DTU Chamfer	Training Time	Unbounded	Best For
GOF	Excellent	~45 min	✓	Complex outdoor scenes
PGSR	0.47 (SOTA)	~30 min	✓	High-fidelity textures
2DGS	Very good	~30 min	✓	View-consistent geometry
SuGaR	Good	Minutes	Limited	Blender/Unity editability

2DGS (SIGGRAPH 2024) collapses 3D volumes into 2D oriented planar Gaussian disks, providing inherently view-consistent depth and normal maps. Surfsplatting This approach produces better thin surface recovery than 3DGS and integrates with TSDF fusion for mesh extraction.

Upstream SfM quality critically affects mesh results: camera pose errors cause multi-view inconsistency in Gaussians, producing noisy depth maps. Target sub-degree rotation error and sub-centimeter translation error. Enable **depth regularization** for untextured regions and **exposure compensation** for varying lighting conditions.

Multi-camera pipeline design for mixed sensors

For mixed camera types (fisheye, pinhole, smartphone), organize images by camera in subfolders and use COLMAP's folder-based intrinsic sharing:

```bash
colmap feature_extractor \
  --database_path project/database.db \
  --image_path project/images \
  --ImageReader.single_camera_per_folder 1
```

Fisheye cameras require fisheye-specific models—standard pinhole models cannot represent fisheye distortion. COLMAP Use `OPENCV_FISHEYE` for known calibration or `RADIAL_FISHEYE` for auto-calibration:

```bash
colmap feature_extractor \
  --ImageReader.camera_model OPENCV_FISHEYE \
  --ImageReader.camera_params "fx,fy,cx,cy,k1,k2,k3,k4"
```

For 360° multi-camera rigs, COLMAP 3.12+ supports native rig configuration via JSON: github

```json
[{
  "cameras": [
    {"image_prefix": "cam0/", "ref_sensor": true},
    {"image_prefix": "cam1/", "cam_from_rig_rotation": [0.7071, 0, 0.7071, 0]}
  ]
}]
```

iPhone 16 Pro Max and iPad Pro: Standard HEIC/JPEG at 12MP produces more consistent results than 48MP ProRAW for photogrammetry. Lock exposure and focus during capture sessions using manual camera apps (Blackmagic Cam, Halide). COLMAP reads EXIF focal length automatically. For LiDAR integration, capture with Polycam or Record3D, export PLY, then align with COLMAP reconstruction via ICP or GPS-based geo-registration. OpenTopography

Rolling shutter from smartphone video can be compensated using **3dgs-deblur** (SpectacularAI) which works with COLMAP poses and VIO velocity data. Ecva

Cross-platform installation and GPU optimization

HLOC installation (both platforms):

```bash
git clone --recursive https://github.com/cvg/Hierarchical-Localization/
cd Hierarchical-Localization && pip install -e .
git submodule update --init --recursive
```

GLOMAP on Ubuntu 24.04 has Eigen/GCC 13 compatibility issues. Build with `-DCMAKE_CXX_FLAGS="-Wno-error"` to bypass warnings-as-errors. Install cuDSS for GPU Bundle Adjustment acceleration.

GLOMAP on Windows 11 pre-compiled releases lack CUDA support—compile from source with vcpkg for GPU acceleration:

```powershell
cmake .. -DCMAKE_TOOLCHAIN_FILE=vcpkg/scripts/buildsystems/vcpkg.cmake `
  -DVCPKG_TARGET_TRIPLET=x64-windows-release -DCMAKE_CUDA_ARCHITECTURES=all-major
```

RTX 3090 Ti (24GB) and A6000 (48GB) both handle large reconstructions well. For feature extraction and dense stereo, enable multi-GPU: github

```bash
colmap feature_extractor --SiftExtraction.gpu_index 0,1
colmap patch_match_stereo --PatchMatchStereo.gpu_index 0,1
```

For 6,000+ image datasets, use vocab tree matching instead of exhaustive matching to avoid O(n^2) complexity:

```bash
colmap vocab_tree_matcher --VocabTreeMatching.vocab_tree_path vocab_tree_256K.bin
```

Emerging methods worth monitoring in 2025

VGGT (CVPR 2025) processes arbitrary-length image sets within memory, jointly predicting camera poses, point maps, depth maps, and 3D point tracks—a potential unified replacement for separate SfM and MVS stages.

MASt3R-SLAM (CVPR 2025) enables real-time dense SLAM with 3D reconstruction priors, bridging the gap between offline photogrammetry and live capture.

CUT3R provides continuous updating for video and photo collections, while **Fast3R** implements global fusion for simultaneous multi-view processing without iterative pair-wise computation.

RoMa v2 (November 2025) with DINOv3 integration sets new benchmarks across all standard evaluation datasets with 1.7x speed improvement.

Recommended pipeline upgrades

Based on this research, here are prioritized upgrades to your current HLOC + SuperPoint + LightGlue + GLOMAP stack:

1. **Switch to ALIKED** for feature extraction—better accuracy, 20x faster inference, permissive license. Use 8,000 keypoints for high-density reconstruction.
2. **Keep LightGlue** for primary matching, but add **MASt3R as fallback** for pairs with low match counts or wide baselines.
3. **Keep GLOMAP**—it's validated as optimal for large datasets. Run `colmap point_triangulator` after reconstruction for denser splat initialization.
4. **Use GOF or PGSR** for mesh extraction from trained Gaussian splats, depending on whether you prioritize unbounded scene handling (GOF) or texture quality (PGSR).
5. **For challenging scenes where SfM fails**, evaluate **InstantSplat** or **MASt3R-SfM** as end-to-end alternatives.
6. **For iPhone/iPad capture**, use standard JPEG with locked exposure, organize by device in subfolders, and use `SIMPLE_RADIAL` camera model. COLMAP

Key repositories and resources

Component	Repository
ALIKED	github.com/Shiaoming/ALIKED
XFeat	github.com/verlab/accelerated_features
LightGlue	github.com/cvg/LightGlue
MASt3R	github.com/naver/mast3r
RoMa v2	github.com/Parskatt/romav2
HLOC	github.com/cvg/Hierarchical-Localization
GLOMAP	github.com/colmap/glomap
VGGSfM	github.com/facebookresearch/vggsfm
gsplat	github.com/nerfstudio-project/gsplat
GOF	github.com/autonomousvision/gaussian-opacity-fields
PGSR	github.com/zju3dv/PGSR
2DGS	github.com/hbb1/2d-gaussian-splatting
SuGaR	github.com/Anttwo/SuGaR
InstantSplat	instantsplat.github.io
MP-SfM	github.com/cvg/mpsfm