Superpixel Meshes for Fast Edge-Preserving Surface Reconstruction

András Bódis-Szomorú[1], Hayko Riemenschneider[1], Luc Van Gool[1,2]
[1]Computer Vision Lab, ETH Zurich Switzerland, [2]PSI-VISICS, KU Leuven, Belgium.

Most Multi-View-Stereo (MVS) methods aim to find a photoconsistent 3D representation of the scene at the highest detail possible. However, such detail is often not required, and the amount of data generated can pose serious difficulties in storage and further processing at larger scales (e.g. city reconstruction), crying for a summarization. Instead of spending time on a pixelwise reconstruction to then simplify and mesh it, we directly fit meshes from single views to sparse Ground Control Points (GCPs) (Fig. 1), without assuming dominant orientations, piecewise planarity, Manhattan-world, 2.5D height maps, or having to rely on the computation and clustering of sparse normals or multi-plane fitting, which are hard to stabilize over sparse data. We used sparse Structure-from-Motion (SfM) points, but our method is not specific to the GCP type, LiDAR or dense RGB-D can also be used.

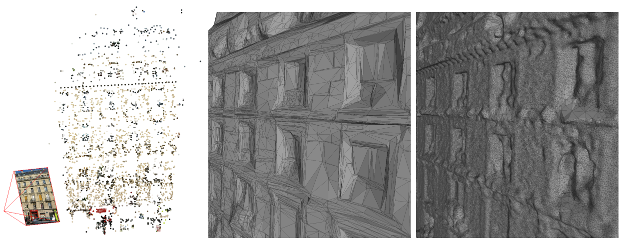

Figure 1: Given only sparse SfM data and a single view (left), our method produces a mesh edge-aligned to image gradients (middle), to be compared with a state-of-the-art CMP-MVS reconstruction produced orders of magnitude slower (right).

Our approach is inspired by others exploiting *sparse* GCPs to approximate a surface. For example, [3] initializes dense MVS by a 3D mesh fit to SLAM points, [1] computes a 2D Delaunay-triangulation over sparse matches to reduce the stereo search space, [4] applies GCPs as hard constraints to interpolate stereo disparity maps, while the recent video summarization work [2] fits a proxy mesh to SfM points for image-based rendering.

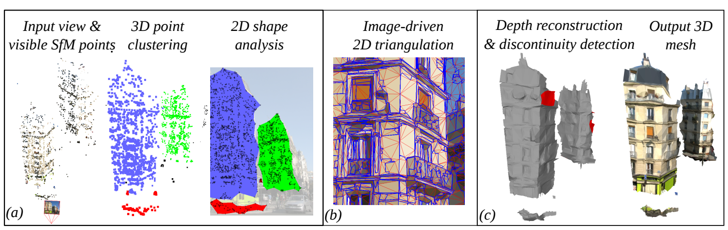

Figure 2: Our approach for single-view mesh reconstruction from sparse SfM data. (a) 3D/2D point cloud analysis identifies areas populated by SfM points, (b) 2D base mesh extraction, (c) fast depth reconstruction and discontinuity segmentation.

Unlike these works, our method (see Fig. 2) operates over a triangular mesh rather than pixels, derives a 2D base mesh aligned to gradients in the image (Fig. 3), and then lifts this mesh to 3D by performing an efficient depth optimization at the vertices using the sparse GCPs as soft constraints.

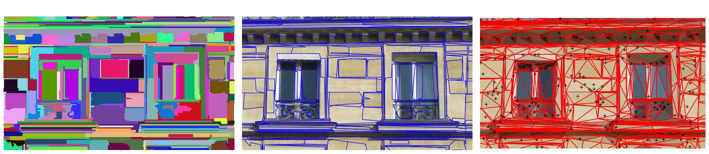

Figure 3: One way of obtaining a 2D base mesh: superpixelization (left), boundary polygonization (middle), constrained Delaunay-triangulation (right).

Our solution has the following highlights:

- *Watertightness*: our meshes are inherently and exactly watertight (where needed), while it is difficult to achieve even approximate watertighness in piecewise models, e.g. in slanted-plane stereo methods.
- *Edge-alignment*: mesh edges are aligned with image gradients (Fig. 3), and likely geometric edge locations, ensuring that the surface is sampled by vertices where actually needed (compactness), and enables good quality visualizations with a lightweight textureless rendering by per-face flat coloring (no need for complicated texture atlases).
- *Smoothness*: our smoothness term penalizes curvature (2nd-order constraint) in the mesh, whereas numerous other methods apply 1st-order penalties favoring local fronto-parallellity (e.g. [2]).
- *Quality*: our evaluatuation of depth errors vs ground-truth shows that the quality of our models is better than single-view baseline methods, and comparable to state-of-the-art MVS methods.
- *Discontinuities*: To avoid undesirable watertightness at depth discontinuities (occlusions), our mesh reconstruction operates only on areas populated by SfM points, as identified by a preliminary 3D point clustering and α-shape analysis step (Fig. 2a). An additional graph-cut mesh segmentation step identifies self-occlusions (Fig. 2c).
- *Speed*: Our energy for depth optimization (including the curvature penalty) allows for an efficient sparse linear solver. This renders our depth reconstruction already fast on CPU giving a significant speed improvement compared to pixelwise MVS methods.
- *Parallelization*: our approach operates on individual views, hence, there is no need for view pairing, pairwise rectification or stereo depth estimation, and it is naturally parallelizable per view.

Figure 4: Some of our single-view (bottom) and 428-image large-scale (top) results. All visualizations are lightweight textureless renderings with per-face mean coloring.

Overall, our method avoids the computational bottlenecks of dense MVS, hence, we believe it is a good alternative to do lightweight modeling where there is no need and/or time to compute an extremely detailed pixelwise 3D model. In particular, it makes large-scale ground-level urban modelling practically feasible, as we can reconstruct meshes of entire street scenes with CityGML LoD-3 details from ground-level images in a matter of minutes.

[1] Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient large-scale stereo matching. In *ACCV*, 2010.

[2] Johannes Kopf, Michael F. Cohen, and Richard Szeliski. First-person hyper-lapse videos,. In *SIGGRAPH*, 2014.

[3] Richard A. Newcombe and Andrew J. Davison. Live dense reconstruction with a single moving camera. In *CVPR*, 2010.

[4] Liang Wang and Ruigang Yang. Global stereo matching leveraged by sparse ground control points. In *CVPR*, 2011.

This is an extended abstract. See the full paper at the Computer Vision Foundation webpage.
This work was supported by the European Research Council (ERC) project VarCity (#273940)