Superpixel Meshes for Fast Edge-Preserving Surface Reconstruction

András Bódis-Szomorú, Hayko Riemenschneider, Luc Van Gool

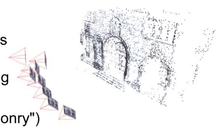
CVL Computer Vision Lab — ETH zürich — erc

www.varcity.eu

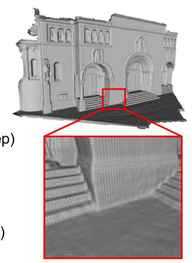
1 Motivation

Goal: fast, lightweight surface modeling from (street-side) imagery
Input: Structure-from-Motion (SfM) data + source images

SfM point clouds
- sparse, inhomogeneous, noisy, outliers
- unstable normals (and clustering)
- unreliable detailed multi-structure fitting

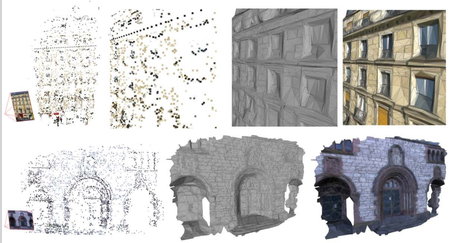


Dense Multi-View Stereo("heavy weaponry")
- volumetric MVS (3D)
  - excellent quality, great detail
  - poor scalability
  - typically slow
- depthmap-based MVS (2.5D)
  - less natural for fusion
  - can be scalable
  - can be real-time (e.g. GPU plane-sweep)
  - plane-sweep relies on priors from SfM
- tedious photoconsistency evaluations
- textureless areas difficult (prior)
- parametric surfaces over-sampled
- detail can be a burden (e.g. city modeling)
- often extra meshing, simplification



2 Our Approach

Idea: Single-view image-aligned direct mesh fitting to a 3D point cloud
Input: images + SfM (or LiDAR or MVS) points + visibility
Output: a triangle mesh per view

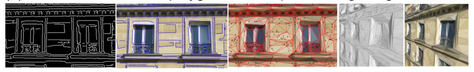


Highlights / Contributions
- directly obtain a mesh (vs. dense MVS first - then simplify)
- mesh edges image-aligned: compactness & simplified rendering
- continous/watertight (where needed) unlike e.g. superpixel-stereo
- 2nd-order smoothness (curvature penalty vs. fronto-parallel prior)
- sparse linear formulation, fast linear solver
- allows full per-view parallelization

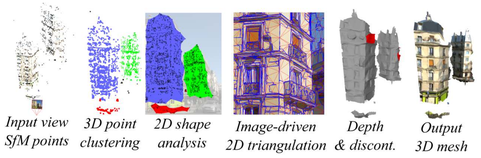
3 Obtaining a 2D Base Mesh

(1) Delaunay Triangulation (DT) over corner-like keypoints (e.g. Harris)

(2) Constrained DT over polygonized, simplified image edges

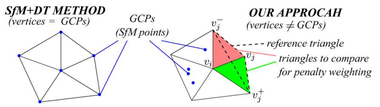


(3) Constrained DT over polygonized, simplified superpixel boundaries



Input view SfM points — *3D point clustering* — *2D shape analysis* — *Image-driven 2D triangulation* — *Depth & discont.* — *Output 3D mesh*

4 Depth reconstruction



$\mathbf{d} = (d_1, d_2, \ldots, d_N)^T$ known depths of GCPs $\mathcal{P} = \{p_i\}_{i=1}^{N}$
$\hat{\mathbf{d}} = (\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_V)^T$ unknown depths of vertices $\mathcal{V} = \{v_i\}_{i=1}^{V}$

energy formulation $E(\hat{\mathbf{d}}) = E_{fit}(\hat{\mathbf{d}}; \mathbf{d}) + \lambda E_{smooth}(\hat{\mathbf{d}})$

fitting term (soft) $E_{fit}(\hat{\mathbf{d}}; \mathbf{d}) = (\mathbf{d} - \mathbf{A}\hat{\mathbf{d}})^T \mathbf{\Sigma}^{-1} (\mathbf{d} - \mathbf{A}\hat{\mathbf{d}})$

curvature penalty $E_{smooth}(\hat{\mathbf{d}}) = \hat{\mathbf{d}}^T \mathbf{B}^T \mathbf{W}^2 \mathbf{B} \hat{\mathbf{d}}$

sparse linear system to solve (fast)

$$(\mathbf{A}^T\mathbf{A} + \lambda \mathbf{B}^T \mathbf{W}^2 \mathbf{B})\hat{\mathbf{d}} = \mathbf{A}^T \mathbf{d}$$

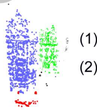
depth interpolation (fitting term)

$p_i = \alpha_{pi} v_p + \alpha_{qi} v_q + \alpha_{ri} v_r \qquad \tilde{d}_i = \alpha_{pi} \hat{d}_p + \alpha_{qi} \hat{d}_q + \alpha_{pi} \hat{d}_r$

linear formulation of curvature penalty

$\hat{d}_{ij} = \beta_1^{ij} \hat{d}_j + \beta_2^{ij} \hat{d}_j^+ + \beta_3^{ij} \hat{d}_j^- \qquad E_{ij} = w_{ij}^2 (\hat{d}_i - \hat{d}_{ij})^2 = w_{ij}^2 (\mathbf{b}_{ij}^T \hat{\mathbf{d}})^2$

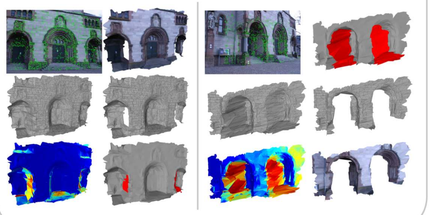
5 Handling discontinuities



(1) Per-view 3D point clustering and 2D α-shapes
(2) Detecting residual discontinuities via graph-cuts

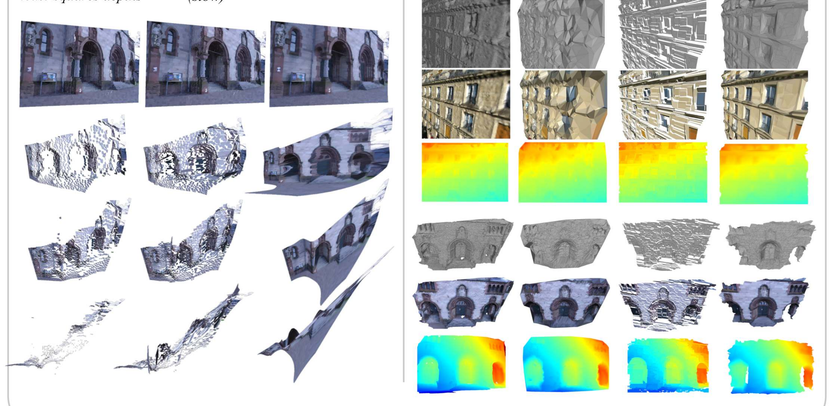
$$E(\mathcal{L}) = \sum_{i=1}^{F} E_i^{disc}(l_i) + \lambda_{disc} \cdot \sum_{\{f_i, f_j\}} e_{ij} \mathbb{I}[l_i \neq l_j]$$

$$\phi_i = \exp\{-\tfrac{1}{2\sigma^2}(\mathbf{v}_i^T \mathbf{n}_i)^2)\}$$



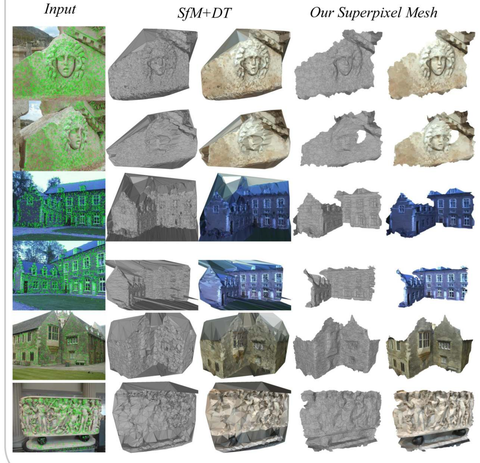
6 Qualitative comparison to other methods

Fronto-parallel superpixels least-squares depths — *Slanted-plane optimization (slow)* — *Our Superpixel Mesh (over full image)* — *Dense Multi-View Stereo (CMP-MVS)* — *SfM+DT* — *FS+Fronto+Soft* — *Superpixel Mesh*

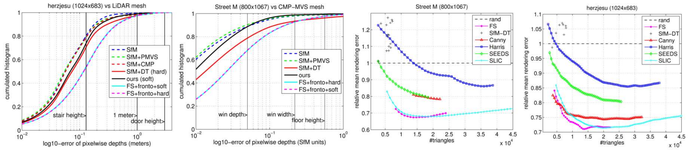


7 Results

Input — *SfM+DT* — *Our Superpixel Mesh*



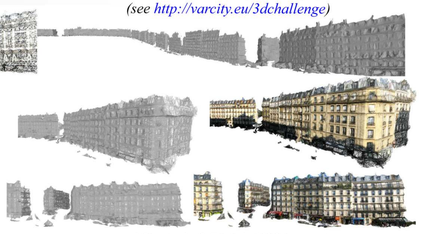
Reconstruction Accuracy



Rendering Quality vs. Base Mesh Methods

Dataset	Images(±)	Resolution	SfM #pts	SfM #tri/im	Ours #pts	PMVS2 #tri	CMP-MVS #tri	Input Match	Input SfM	8× PMVS2	8× CMP-MVS	GPU Ours	GPU Par
Street Z	630 (630)	800×1200	238.9k	15.4k	1 620k	3 927k	16051s	207s	1624*	24061s+	1171s	1.9s
Street P	428 (10)	800×1067	365k	13.4k	1 630k	4 697k	710s	811s	1290s*	16143s+	843s	2.0s
Street M	26 (26)	800×1067	19.5k	14.2k	93.3k	1 253k	45s	4s	49s*	1083s+	50.9s	2.0s
LeuvenCastle	28 (28)	800×600	16.2k	9.1k	29.6k	586.0k	69s	3s	28s*	825s+	32.3s	1.2s
Medusa	15 (15)	800×640	16.3k	9.4k	30.8k	536.1k	19s	1s	16s*	470s+	18.0s	1.2s
Fountain	11 (11)	1024×683	13.5k	10.1k	44.5k	707.9k	8s	1s	20s*	468s+	16.3s	1.5s
HerzJesu	8 (8)	1024×683	8.3k	10.4k	35.2k	492.5k	7s	1s	16s*	334s+	10.6s	1.3s
Dionysos	8 (8)	800×600	4.1k	7.3k	17.4k	243.3k	7s	1s	16s*	229s+	10.0s	1.3s
MertonVI	6 (6)	1024×768	2.1k	13.3k	10.8k	180.5k	2s	1s	9s*	123s+	9.9s	1.7s

Rue Monge, Paris (428 images)
(see http://varcity.eu/3dchallenge)



Münsterhof, Zürich (630 images)