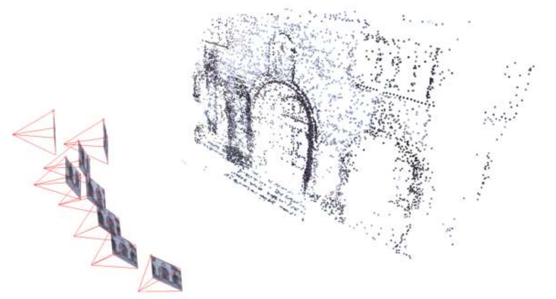
Fast, Approximate Piecewise-Planar Modeling Based on Sparse Structure-from-Motion and Superpixels

András Bódis-Szomorú, Hayko Riemenschneider, Luc Van Gool

www.varcity.eu

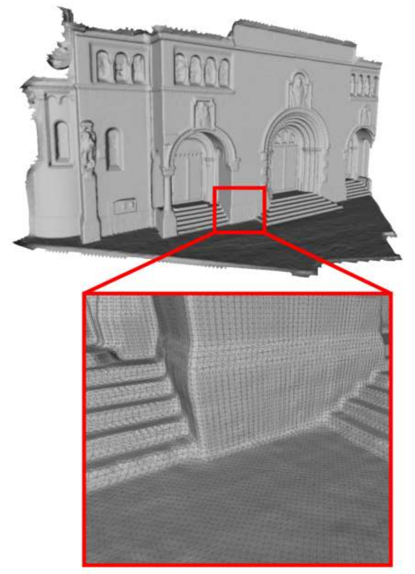
1 Problem Statement

Goal: fast, lightweight surface modeling of man-made scenes from images
Input: Structure-from-Motion (SfM) data & source images



SfM point clouds typically too sparse for:
- reliable normal extraction (and clustering)
- direct planar region growing
- robust sequential fitting
- global robust multi-structure fitting
- capturing more than some major planes

Drawbacks of re-using images "sparsely":
- normals via photoconsistency imprecise
- vanishing directions not always possible

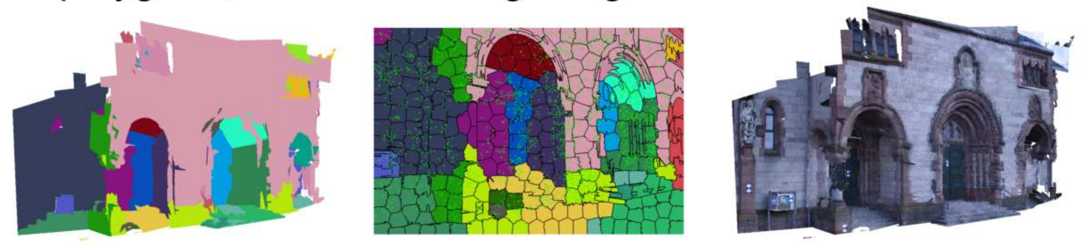


Problems with Dense Multi-View Stereo:
- enforcing photoconsistency (slow)
- difficulties with textureless areas (aggregation requires priors)
- often time-consuming, poor scalability
- Manhattan assumption (not always enough + prior orientations needed, see above)
- non-parametric, redundant sampling
- often needs post-processing, e.g. segmentation, parametric fitting

2 Proposed Idea

Idea: SfM data & superpixels for multi-view surface reconstruction
Assumption: piecewise-planar scene
Inputs: SfM with visibility & source images

Outputs: 3D polygons, multi-view image segmentation

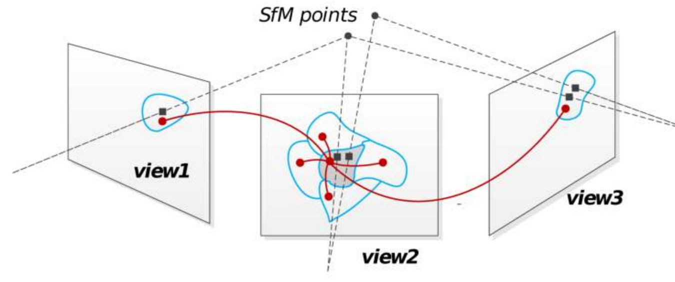


Contributions:
- combining SfM & superpixels for multi-view surface optimization
- novel joint multi-view MRF/energy formulation
- criterion for measuring plane stability
- dense 3D output as polygons

3 Multi-view Optimization

Input: superpixels, 3D points, cameras, visibility, plane hypotheses
Task: assign all superpixels (from all views) to global plane hypotheses

$\mathcal{S} = \{s_1, s_2, \ldots, s_S\}$ set of superpixels (from all views)
$\Pi = \{\pi_1, \pi_2, \ldots, \pi_L\}$ set of plane hypotheses
$l_i \in \{1, 2, \ldots, L\}$ possible assignments of superpixel s_i to a plane
$\mathcal{L} = \{l_1, l_2, \ldots, l_S\}$ assignment of each superpixel in each view to a plane



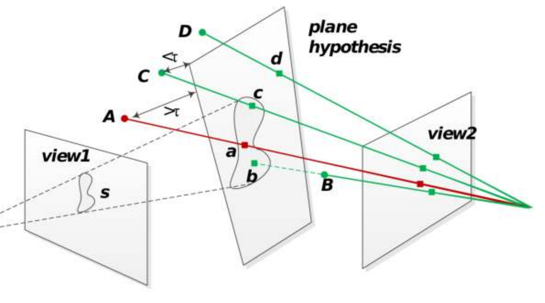
Graph formulation $\mathcal{G} = \{\mathcal{V} \equiv \mathcal{S}, \mathcal{E}^b \cup \mathcal{E}^w\}$
- vertices are superpixels
- $\mathcal{E}^w$ within-view edges between adjacent superpixels
- $\mathcal{E}^b$ between-view edges generated by common SfM points

$$E(\mathcal{L}) = \underbrace{\sum_{i=1}^{S} D_i(l_i)}_{\text{unary terms}} + \underbrace{\sum_{(i,j)\in\mathcal{E}^w} V_{ij}^w(l_i, l_j)}_{\text{within-view pairwise}} + \underbrace{\sum_{(i,j)\in\mathcal{E}^b} V_{ij}^b(l_i, l_j)}_{\text{between-view pairwise terms}}$$

Optimization: $\alpha-$expansion (graph cuts)

I. Unary terms $D_i(l_i) = D_i^{fit}(l_i) + D_i^{rays}(l_i) + D_i^{angle}(l_i)$

- **Fitting** term: encourage planes that fit well to points seen in a superpixel
- **Visibility** term: penalize free space violations
- **Angle** term: penalize planes seen in a sharp angle through a superpixel



Free-space violations

II. Pairwise terms

a) Within views: $$V_{ij}^w = (\underbrace{\alpha C_{ij}}_{\text{color term}} + \underbrace{\beta G_{ij}}_{\text{gradient term}}) \cdot \underbrace{\omega_{ij}^w}_{\text{adjacency weights}} \cdot \mathbb{I}[l_i \neq l_j]$$

- **Color** term: neighboring superpixels with similar color to the same plane
- **Gradient** term: weakly separated superpixels to the same plane
- **Weighting**: neighbors sharing a shorter r. boundary affect each-other less

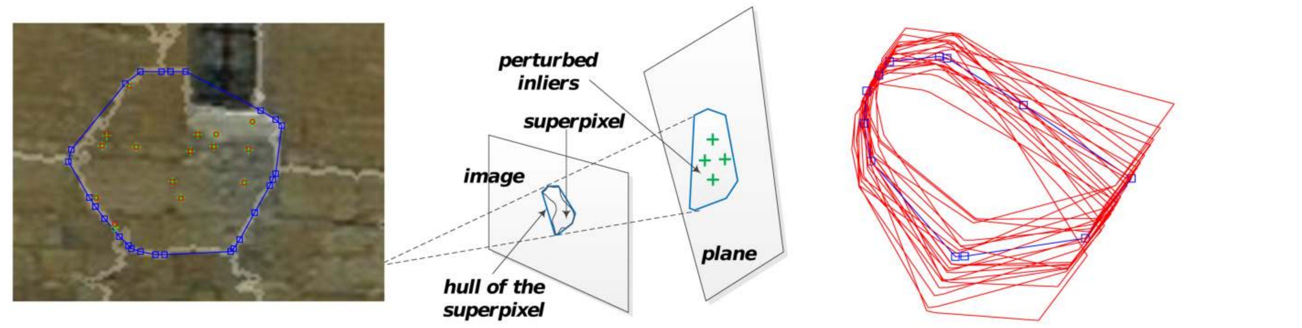
b) Between views: $$V_{ij}^b = \gamma \underbrace{\omega_{ij}^b}_{\text{overlap weights}} \underbrace{C_{ij}}_{\text{color term}} \cdot \mathbb{I}[l_i \neq l_j]$$

- **Color** term: encourage superpixels in different views with similar color to belong to the same plane
- **Weighting**: increases with the number of SfM points jointly observed

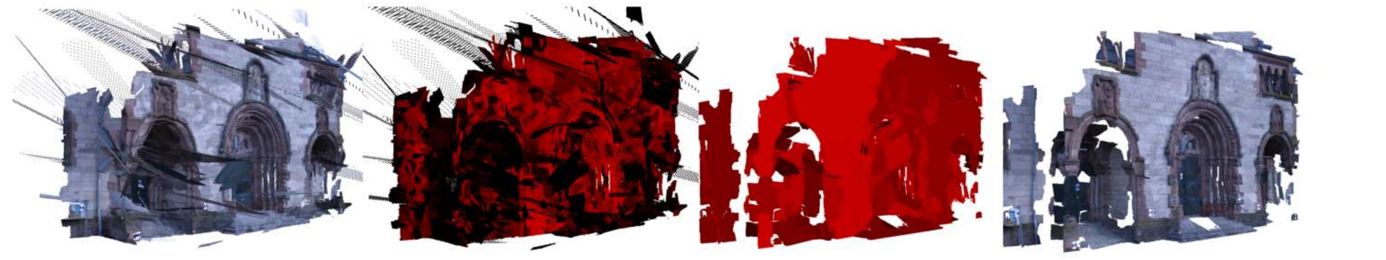
4 Initialization: Plane Hypotheses

Input: superpixels, 3D points, cameras, visibility

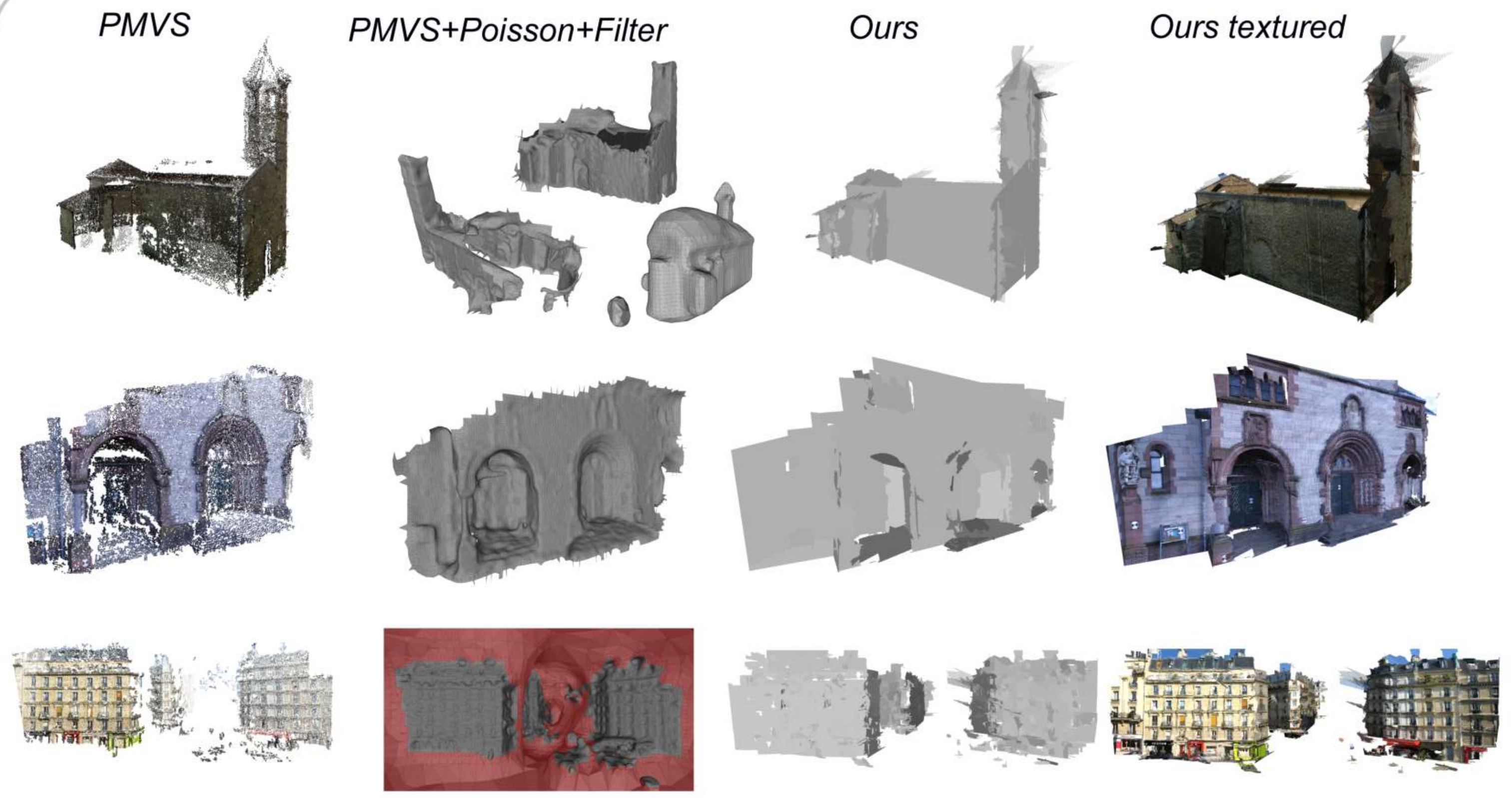
I. Robust plane fitting to observed SfM points per superpixel (local)
II. Plane filtering: stability measure via Monte-Carlo experiments



III. Global plane merging: greedy, merge if all inliers explained

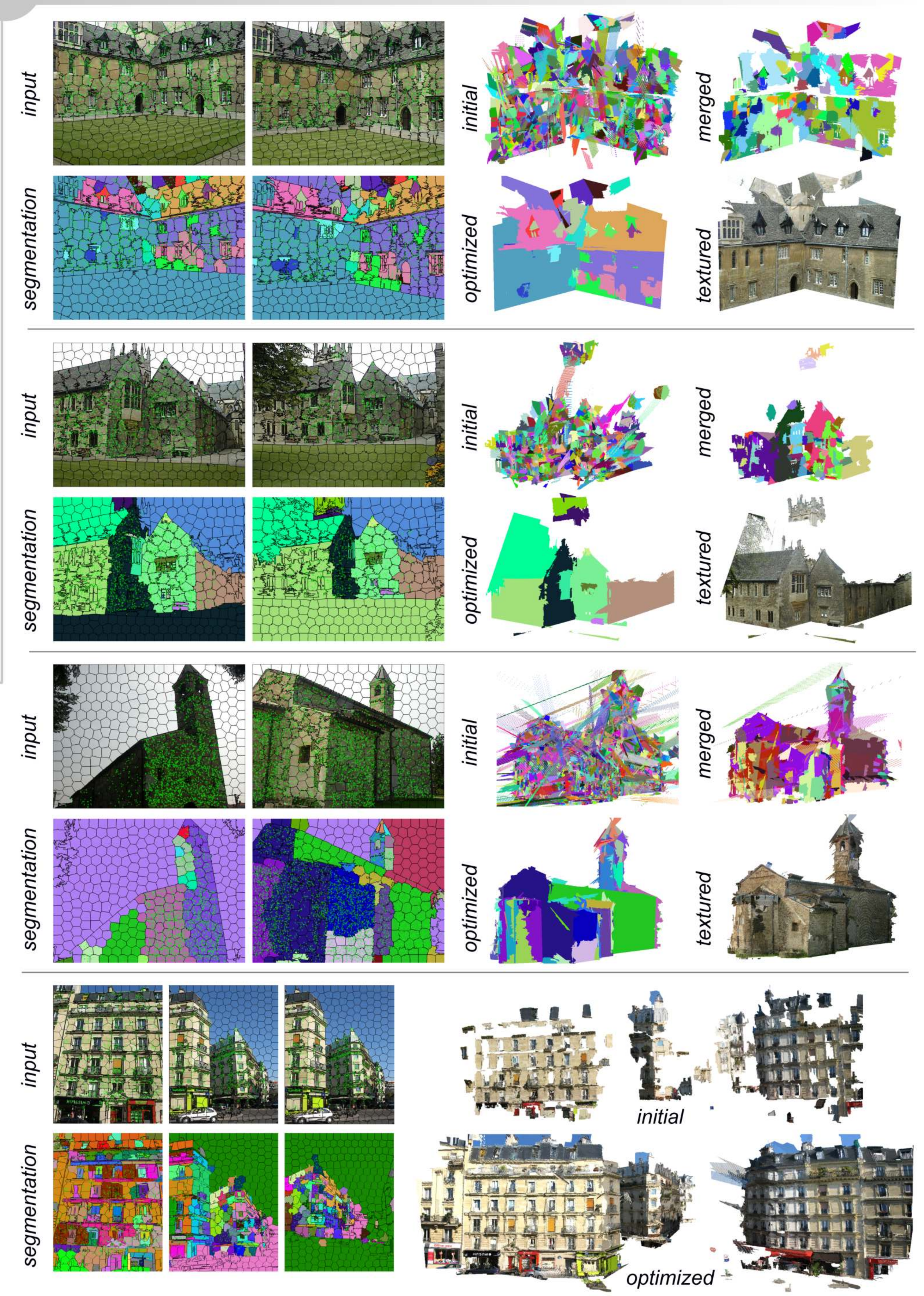


Robust fitting | Stability Score | Filtered & Merged | Textured



Advantages of our method
- dense & lightweight output
- detailed boundaries from images
- fast: no pixelwise photoconsistency computations
- highly parallelizable (superpixels, energy terms)
- copes with textureless areas
- more than just principal planes captured
- no Manhattan assumption
- not required: sparse normals & clustering, vanishing points, or dense depth maps

5 Results



merging step very effective — minutes instead of hours

	Input data			Superpixels / MRF			Planes				Timing*		
Dataset	imgs	pts	rays	sp	sp(data)	pts/sp	ini	filt	merge	gco	sp	3D	gco
Merton I	3	2.9k	6.7k	1.6k	60.0%	7.2	623	409	69	19	27 sec	10 sec	0.14 sec
Merton III	3	2.2k	5.0k	1.4k	47.9%	7.4	474	317	55	13	25 sec	7 sec	0.10 sec
HJ-P8	8	8.3k	25.4k	3.0k	76.6%	11.1	1883	1193	64	29	54 sec	29 sec	0.30 sec
Mirbel	26	19.5k	66.0k	16.9k	57.0%	6.9	6068	1426	292	185	4.4 min	2.9 min	10.6 sec
Pozzo	53	38.6k	135k	21.4k	50.5%	12.5	8152	5481	80	58	7.0 min	4.1 min	4.5 sec

*Timing: seconds in Matlab, intel Core i7 3.4 GHz CPU, on a single core