Patchwork Stereo: Scalable, Structure-aware 3D Reconstruction in Man-made Environments

Anonymous WACV submission #179

Abstract

In this paper, we address the problem of Multi-View Stereo (MVS) reconstruction of highly regular man-made scenes from calibrated, wide-baseline views and a sparse Structure-from-Motion (SfM) point cloud. We introduce a novel patch-based formulation via energy minimization which combines top-down segmentation hypotheses using appearance and vanishing line detections, as well as an arrangement of creased planar structures which are extracted automatically through a robust analysis of available SfM points and image features. The method produces a compact piecewise planar depthmap and a mesh which are aligned with the scene's structure. Experiments show that our approach not only reaches similar levels of accuracy w.r.t state-of-the-art pixel-based pipelines while using much fewer images, but also produces a much more compact, structure-aware mesh in a considerably shorter runtime by several of orders of magnitude.

1 Overview

The present document presents additionnal results and theoretical details about the approach under the following form:

1. A qualitative comparison of our Patchwork Stereo (PWS) method with state-of-the-art Super-pixel Stereo [?, ?] on a common dataset.
2. Additionnal illustrations with qualitative results on datasets presented in the paper.
3. Theoretical details about the definition of the pairwise potential $\Psi_{pq}^{\mathsf{Connectivity}}$ we introduce in the energy-based formulation.

2 Qualitative comparison with prior work

In the absence of publically available implementations of prior work, we present a qualitative side-by-side comparison of our method (PWS) against a state-of-the-art superpixel modelling pipeline, Super-pixel Stero (SPS) [?, ?] on the GMU-building dataset provided by the authors [?, ?].

As seen in fig.??, our methods produces 3D meshes with much less triangle faces, which are aligned with the dominant vanishing directions (VDs) and image-level strong edges which also translates into more pleasing colour texturing and rendering. SPS uses bottom-up superpixels [?] which are agnostic of the scenes' VD-aligned edges.

3 Additional Qualitative Results

4 Theoretical details – Pairwise potential

We initially defined our pairwise regularization term as follows:

$$\Psi_{pq}^{\mathsf{Connectivity}}(y_p, y_q) = \begin{cases} 0 & : \text{if } y_p = y_q, (y_p, y_q) \in \mathcal{T}_{continuity} \\ \lambda_1 & : \text{elseif } (y_p, y_q) \in \mathcal{T}_{crease} \\ \lambda_2 & : \text{elseif } (y_p, y_q) \in \mathcal{T}_{occlusion_1} \\ \lambda_3 & : \text{elseif } (y_p, y_q) \in \mathcal{T}_{occlusion_2} \\ \lambda_4 & : \text{otherwise} \end{cases} \tag{1}$$

where $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4$ are the respective costs for neighboring patches to: (zero cost) lie on the same plane, (λ_1) form a crease junction, (λ_2) lie at a depth discontinuity (first case of occlusion, where $\vec{e}_{pq}$ is consistent with the

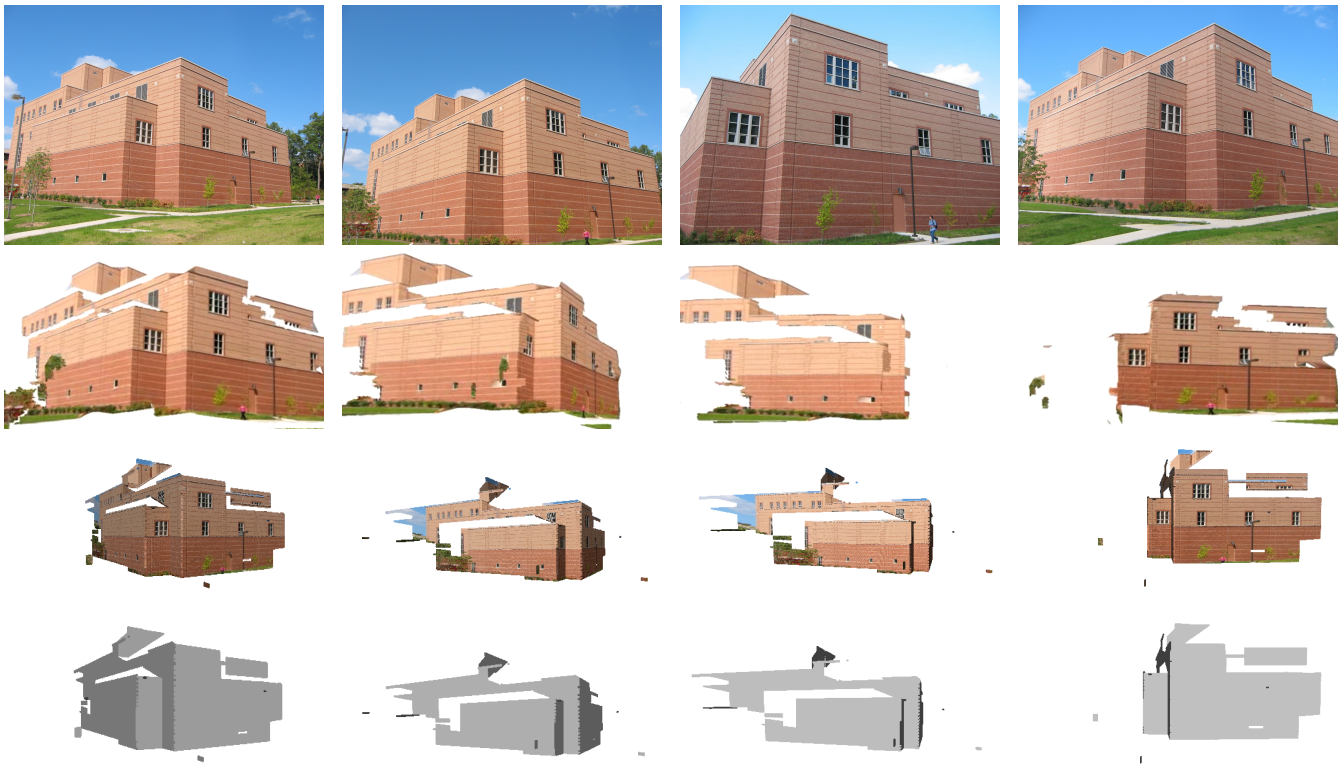

Figure 1: GMU-building dataset [?] – First row, images from the given dataset. Second row, results from Superpixel Stereo [?]. Third row, our results using the same set of images and the corresponding sparse SfM point cloud. Fourth row, our results uncoloured. Few sky pixels have been removed to facilitate visualization.

orientations of both p and q), and (λ_3) the second case of acceptable occlusion is met when $\vec{e}_{pq}$ is consistent only with the occluding (fronting) patch, as seen in Figure 4). All other configurations are given a prohibitive penalty λ_4.

Depending on how these $\lambda_{1...4}$ are set, the smoothness function can either be a metric, or a semi-metric. The metric case allows a safer inference as it guarantees the solution to be at a known factor from global optimum, but is more restrictive in expressive power [?]. In all our experiments, we adopt the semi-metric case by setting these parameters to $\{\alpha, \beta, \gamma, \lambda, \lambda_1, \lambda_2, \lambda_3, \lambda_4\} = \{1, 0.5, 0.4, 30, 0, 0.6, 3.8, 50\}$. The final energy can hence be optimized using swap-based graph-cut moves. In practice, we found the alpha-expansion [?] inference to give better results even in the semi-metric case although there is no theoretical guarantee to be close the optimum.

We now define the pairwise potential with additional details:

$$\Psi_{pq}^{\text{Connectivity}}(y_p, y_q) = \begin{cases} 0 & : \text{if } y_p = y_q \\ \lambda_1 & : \text{elseif } \theta_{pq}(y_p, y_q) \ \wedge \ \chi_{pq}^{3D}(y_p, y_q) \ \wedge \ \vec{n_p} \neq \vec{n_q} \\ \lambda_2 & : \text{elseif } \overline{\theta_{pq}(y_p, y_q)} \ \wedge \ \chi_{pq}^{3D}(y_p, y_q) \\ \lambda_3 & : \text{elseif } \theta_{pq}(y_p, y_q) \ \wedge \ \overline{\chi_{pq}^{3D}(y_p, y_q)} \ \wedge \ \phi_{pq}(y_p, y_q) \\ \lambda_4 & : \text{Otherwise} \end{cases} \tag{2}$$

where θ_{pq} is a 3D tightness predicate which is true when patches touch in 3D[1], χ_{pq}^{3D} means $\vec{n_p}$ and $\vec{n_q}$ share a common vanishing point (i.e., relate to some $\vec{n_{ij}}$ and $\vec{n_{ij'}}$, hence such oriented surfaces could form a junction pointing towards $\vec{v_i}$ which they have in common). ϕ_{pq} indicates whether the orientation of the common boundary, $\vec{e_{pq}}$, belongs to the hypothesis of the fronting reconstructed 3D patch which is a case of plausible occlusion.
The top bar notation designates predicate negation.

$$\theta_{pq}(y_p, y_q) = [\![\ max\{ \ \rho(\vec{e_{pq}}(1), y_p, y_q), \ \rho(\vec{e_{pq}}(2), y_p, y_q) \ \} \leq \varepsilon \ ]\!] \tag{3}$$

[1] Along the whole common linear boundary, up to an $\epsilon = 10^{-5}$

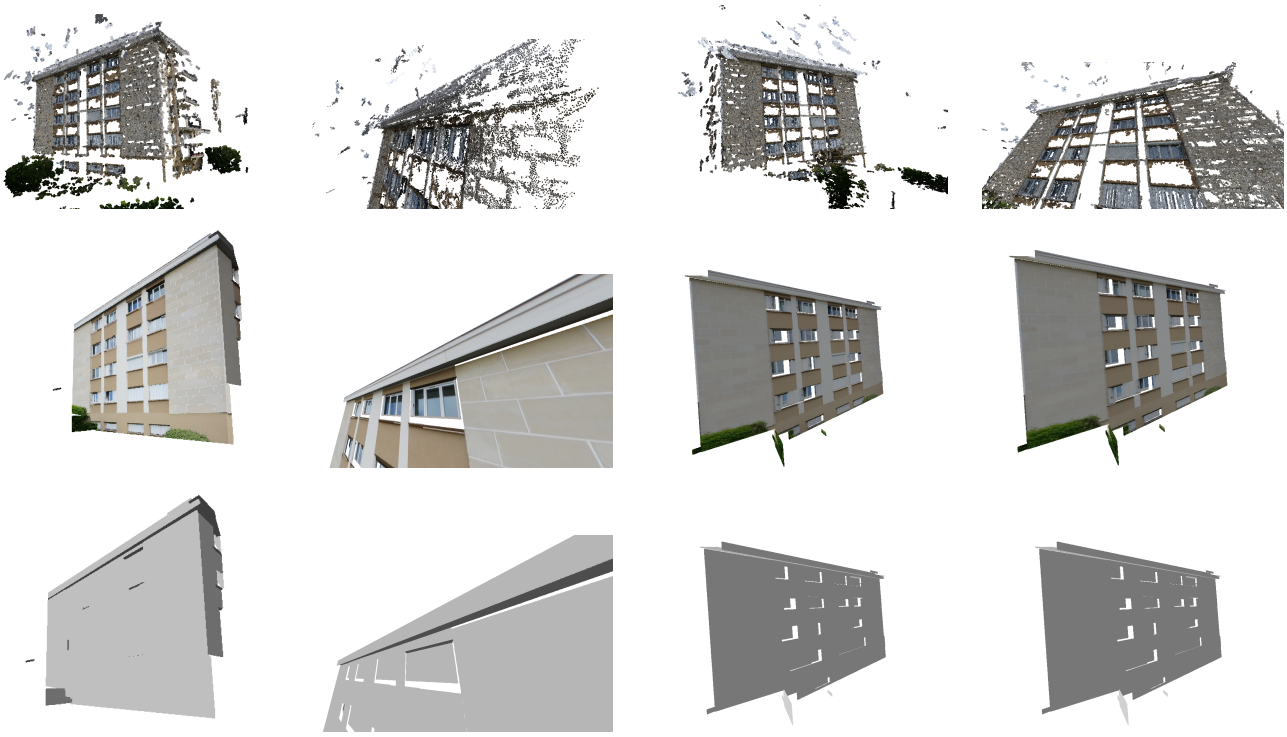

Figure 2: Additional qualitative results. Comparison between the baseline PMVS2 [?] in the first row and our method in the second row (coloured) and in third row (uncoloured).

where $e_{pq}^{\rightarrow}(1)$ and $e_{pq}^{\rightarrow}(2)$ refer to the two vertices delineating the common edge boundary between neighbor patches[2]. ρ is the relative 3D reconstruction error of a pixel x w.r.t planar hypotheses y_p and y_q:

$$\rho(x, y_p, y_q) = \frac{||\mathbf{X}_{y_p} - \mathbf{X}_{y_q}||}{2 \max\{ ||\mathbf{X}_{y_p}||, ||\mathbf{X}_{y_q}|| \}} \tag{4}$$

$$\begin{aligned} \phi_{pq}(y_p, y_q) = [\![\ (e_{pq}^{\rightarrow} \in \{i, j\} \ \wedge \ ||X_{p_1}|| < ||X_{q_1}|| \ \wedge \ ||X_{p_2}|| < ||X_{q_2}||) \\ \vee (e_{pq}^{\rightarrow} \in \{i', j'\} \ \wedge \ ||X_{p_1}|| > ||X_{q_1}|| \ \wedge \ ||X_{p_2}|| > ||X_{q_2}||) \ ]\!] \end{aligned} \tag{5}$$

where X_{p_1} (resp. X_{p_2}, X_{q_1}, X_{q_2}) is a shortcut notation to designate the 3D reconstruction of pixel $e_{pq}^{\rightarrow}(1)$ (resp. $e_{pq}^{\rightarrow}(2)$) via y_p (resp. y_q)). i, j, (resp. i', j') are line orientations corresponding to vanishing points $\vec{v}_i$, $\vec{v}_j$ (resp. $\vec{v}_{i'}$, $\vec{v}_{j'}$)

References

[1] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[2] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(8):1362–1376, 2010.

[3] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 508–515. IEEE, 2001.

[4] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):147–159, 2004.

[5] B. Mičušík and J. Košecká. Piecewise planar city 3d modeling from street view panoramic sequences. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2906–2912. IEEE, 2009.

[6] B. Mičušík and J. Košecká. Multi-view superpixel stereo in urban environments. *International journal of computer vision*, 89(1):106–119, 2010.

[2] $[\![\ ]\!]$ stand for the *Iverson* bracket.