Supplementary Material: Semantic 3D Reconstruction with Continuous Regularization and Ray Potentials Using a Visibility Consistency Constraint

Nikolay Savinov, Christian Häne, Ľubor Ladický and Marc Pollefeys
ETH Zürich, Switzerland
{nikolay.savinov,christian.haene,lubor.ladicky,marc.pollefeys}@inf.ethz.ch

1. Introduction

First, we provide the proof of Lemma 2 from the main text. Then we give additional experimental evaluation which did not fit into the main text of the paper: the additional experiments are shown for semantic 3D reconstruction as well as for classic 3D reconstruction. Afterwards, we give an intuition why the convex formulation, which was introduced in Section 3.2 of the main text, provides a very weak solution. Eventually, we show convergence experiments for our algorithm.

2. Proof of Lemma 2

Proof. For readability we drop the iteration index (n). First we note the following. If we fix k and s each $(z_s^{\ell m})_k$ only appears in two linear equation systems with L equations.

$$x_s^\ell = \sum_m \left(z_s^{\ell m}\right)_k, \quad x_{s+e_k}^\ell = \sum_m \left(z_s^{m\ell}\right)_k \quad \forall \ell \in \mathcal{L} \quad (1)$$

Hence, this constraints can be written in the form $A\mathbf{w}_\mathbf{s}^\mathbf{k} = b$ for each k and s, where $\mathbf{w}_\mathbf{s}^\mathbf{k}$ is a vector containing the variables $(z_s^{m\ell})_k \ \forall \ell, m$. b contains the values of x_s^ℓ and $x_{s+e_k}^\ell$. The variables $\tilde{\mathbf{z}}$ are initialized by projecting the variables $\mathbf{z}$ to the affine space defined by the equation system $A\mathbf{w}_\mathbf{s}^\mathbf{k} = b$ for each s,k combination individually. To also ensure that the non-negativity constraints on the $z_s^{\ell m}$ are fulfilled, the following substitution is applied until there are no more negative $\tilde{z}_s^{\ell m}$. Assuming $\tilde{z}_s^{\ell',m'} < 0$, from $x_s^\ell \geq 0$ it follows that there are $\tilde{z}_s^{\ell',m''} > 0$ and $\tilde{z}_s^{\ell'',m'} > 0$. Hence, we update

$$\tilde{z}_s^{\ell',m'} \leftarrow \tilde{z}_s^{\ell',m'} + \epsilon \qquad \tilde{z}_s^{\ell'',m''} \leftarrow \tilde{z}_s^{\ell'',m''} + \epsilon \quad (2)$$

$$\tilde{z}_s^{\ell',m''} \leftarrow \tilde{z}_s^{\ell',m''} - \epsilon \qquad \tilde{z}_s^{\ell'',m'} \leftarrow \tilde{z}_s^{\ell'',m'} - \epsilon. \quad (3)$$

Note that this substitution does not affect the original constraints if we choose ϵ such that non-negative variables stay non-negative. The above substitution is iteratively applied until no more non-negative variables are left. By always choosing ϵ as big as possible, meaning such that either the non-positive variable $\tilde{z}_s^{\ell',m'}$ or one of the positive variables gets 0, the number of iterations of the algorithm is bounded by $O(L^2)$. This holds because for each negative variable there is a maximum of $O(L)$ steps that can be made to increase it. □

3. Semantic 3D Reconstruction: Additional Results

Additional reconstructions are shown in Fig. 1. We refer the reader to the supplementary video where renderings of our models can be found.

4. Dataset "Head"

We test our algorithm on a challenging specular "Head" dataset from [1]. It was shown in that paper that the results of traditional dense 3D reconstruction methods can be improved by utilizing the silhouette information. This information was included in their formulation as energy over rays. We show even more improvement by using our non-convex ray potential formulation in Fig. 2.

5. Middlebury: Additional Analysis

We provide accuracy (Acc) and completeness (Comp) plots for Dino Ring dataset in Fig. 3. We also show additional renderings of reconstructions in Fig. 4.

Overall, besides being accurate (as shown in the paper), our algorithm produces reconstructions with very high completeness: for 5 out of 6 datasets our reconstructions have completeness above 99.5%.

6. Why is Convex Formulation so Weak?

In this section we give a small intuitive example why the convex relaxation gives a solution which is far from binary. We give this example for a 2-label problem without regularization and use the following notation for the labels: o means occupied, f means free-space. Consider one ray of the length $N = 3$ with costs $c_0^o = -2$, $c_1^o = -3$, $c_2^o = -2$ and the rest of the costs are 0. This is a realistic example since it corresponds to allowing the uncertainty around the

estimated depth position $i = 1$ (for example, camera sees the wall and stereo matching provides an estimate of depth, but this estimate is noisy in practice, so the uncertainty window along the ray is very desirable). Since we only consider single ray, the ray index r is omitted and the voxel space indexing function s_i simplifies to just position i along the ray. The exact problem, which we are solving, would be (as a reminder, y^f_{-1} is always set to be 1):

$$\begin{aligned} \psi = &- 2y^o_0 - 3y^o_1 - 2y^o_2 \rightarrow \min_{\mathbf{x},\mathbf{y}} \\ \text{s.t. } & y^o_i \leq y^f_{i-1},\ y^f_i \leq y^f_{i-1}, \\ & y^o_i \leq x^o_i,\ y^f_i \leq 1 - x^o_i, \\ & x^o_i \in [0,1],\ \ \forall i. \end{aligned} \tag{4}$$

The desired solution to this problem would be

$$\begin{aligned} & x^o_0 = 0,\ x^o_1 = 1,\ x^o_2 = 0, \\ & y^o_0 = 0,\ y^o_1 = 1,\ y^o_2 = 0, \\ & y^f_0 = 1,\ y^f_1 = 0,\ y^f_2 = 0. \end{aligned} \tag{5}$$

This means taking the best position in the uncertainty window. This solution has the cost $c_{binary} = -3$. Unfortunately, the solution where all the variables above take value 0.5 has a better cost: $c_{0.5} = -3.5$.

Our preliminary investigations indicate that the "all-0.5" solution will always be the optimal solution to the convex relaxation as long as the best cost $c^o_{\min} = \min_i c^o_i$ is larger than the sum of other occupied costs (as it is the case in the example above, -3 versus -4).

7. Convergence Analysis

In this section we analyze the convergence behavior of our method.

First, we evaluate how fast the algorithm converges using different minimization intervals in between the majorization steps. In Fig. 5 we can see that a frequent execution of the majorization step has a very beneficial effect on the convergence. Additionally, we see that for a broad range of values we reach similar (in energy) critical points of our cost function. This is a strong indication that our method is robust against bad solutions.

Second, we analyze tie handling in eq. 14 of the main text. As a reminder, this equation describes linear majorizer as

$$\begin{aligned} & g(x^f_{s_i}, y^f_{i-1} | x^{f,(n)}_{s_i}, y^{f,(n)}_{i-1}) \\ & \qquad = \begin{cases} 0 & \text{if } y^{f,(n)}_{i-1} \leq x^{f,(n)}_{s_i} \\ y^f_{i-1} - x^f_{s_i} & \text{if } y^{f,(n)}_{i-1} > x^{f,(n)}_{s_i} \end{cases} \end{aligned} \tag{6}$$

In that equation the tie case is $y^{f,(n)}_{i-1} = x^{f,(n)}_{s_i}$ and it is possible to choose any of the two branches in this case: 0 or $y^f_{i-1} - x^f_{s_i}$. Our experiment in Fig. 6 shows that the difference in final energies between these two choices is very small, 0.25% of their values.

References

[1] D. Cremers and K. Kolev. Multiview stereo and silhouette consistency via convex functionals over convex domains. *Transactions on Pattern Analysis and Machine Intelligence*, 2011. 1, 3

[2] C. Häne, C. Zach, A. Cohen, R. Angst, and M. Pollefeys. Joint 3D scene reconstruction and class segmentation. In *Conference on Computer Vision and Pattern Recognition*, 2013. 3

[3] C. Hernandez, G. Vogiatzis, and R. Cipolla. Probabilistic visibility for multi-view stereo. In *Conference on Computer Vision and Pattern Recognition*, 2007. 3

[4] K. Kolev, M. Klodt, T. Brox, and D. Cremers. Propagated photoconsistency and convexity in variational multiview 3D reconstruction. In *IN WORKSHOP ON*, 2007. 3

[5] N. Savinov, L. Ladicky, C. Häne, and M. Pollefeys. Discrete optimization of ray potentials for semantic 3D reconstruction. In *Conference on Computer Vision and Pattern Recognition*, 2015. 3

[6] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'2006)*, volume 1, pages 519–526. IEEE Computer Society, June 2006. 4

[7] G. Vogiatzis, P. H. Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts. In *Conference on Computer Vision and Pattern Recognition*, 2005. 3

Input Image | Häne et al. 2013 [2] | Savinov et al. 2015 [5] | Proposed Method

Figure 1: Semantic 3D Reconstructions.

Figure 2: Rendering of the results on the "Head" dataset. The columns from two to four are reported by [1]. It has been shown in [1] that ray information can help in reconstructing the thin pole on which the head is mounted. Our algorithm successfully reconstructs this pole as well.

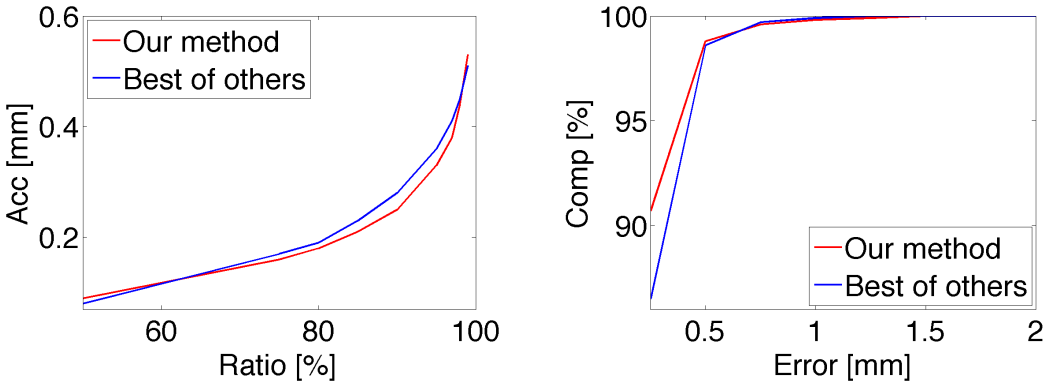

Figure 3: Acc vs. Ratio (lower curve better) and Comp vs. Error (higher curve better) plots for the Dino Ring dataset of the Middlebury benchmark (for details on these plots see [6]).

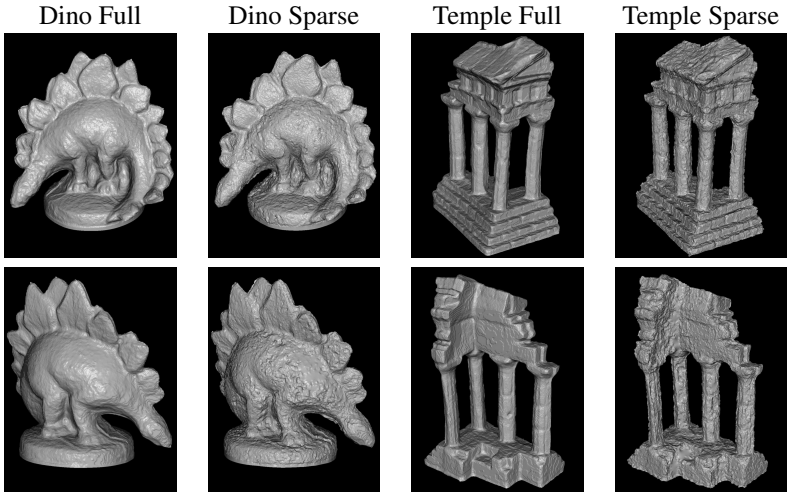

Figure 4: Rendering of Middlebury results.

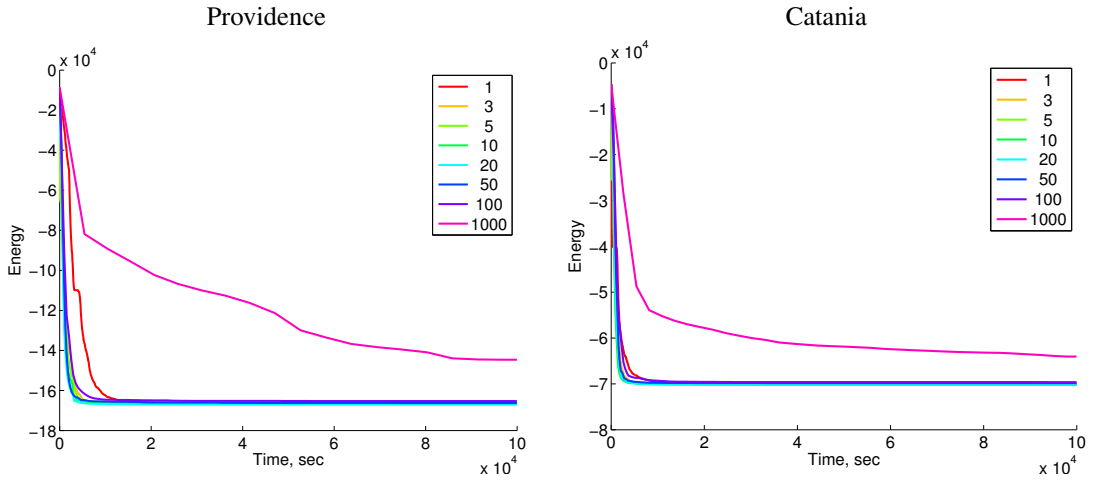

Figure 5: Evolution of the energy over time for different numbers of iterations the convex minimization algorithm is run in between the execution of the majorization step.

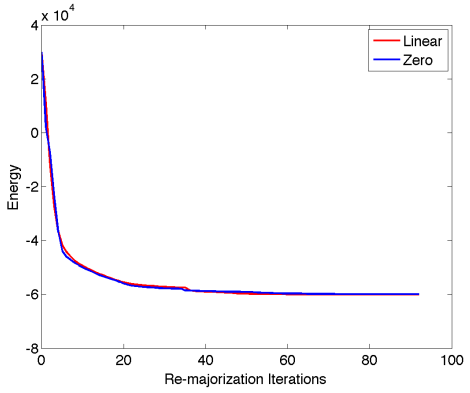

Figure 6: Evolution of the energy over iterations for two different re-majorization strategies. ”Linear” means that the tie case is handled with the linear branch, ”zero” means that constant branch with 0 value is taken.