Surface Normal Integration for Convex Space-time Multi-view Reconstruction

Martin R. Oswald
martin.oswald@in.tum.de
Daniel Cremers
cremers@tum.de

Computer Vision Group,
Department of Computer Science,
Technische Universität München

We propose a convex variational approach to space-time reconstruction which estimates surface normal information and integrates it into the photoconsistency estimation as well as into an anisotropic spatio-temporal total variation regularization. As such the proposed method generalizes the works [4], [5]. Although [4] already studied anisotropic regularization they did not estimate normals but used the normals from [2]. The combination of these methods, [4] and [2], is more than 40 times slower than our method as [4] alone needs about 1h to compute a single frame. In contrast, our method only takes about 3 minutes per frame including normal estimation and temporal regularization due to the proposed efficient implementation. Moreover, the method by Kolev et al. [4] does not work well on the 4D data sets we consider, as shown in [5, Fig. 5]. With the estimated normals at hand, we further propose an improvement of the photoconsistency voting scheme by Hernández and Schmitt [1] resulting in superior accuracy especially for sparse camera setups.

We represent the space-time surface as a binary interior/exterior labeling function $u: V\times T \mapsto \{0,1\}$ and state the spatio-temporal 3D reconstruction task as a minimization problem of the following energy.

$$E(u) = \int_{V\times T} \Big[|\nabla_{\boldsymbol{x}} u|_{D_{\boldsymbol{x}}} + g_t |\nabla_t u| + \lambda\, f u \Big] \; d\boldsymbol{x}dt \qquad (1)$$

The energy consists of three terms. An **anisotropic spatial regularization term**, defined be the norm $|\boldsymbol{y}|_{D_{\boldsymbol{x}}} = \langle \boldsymbol{y}, D_{\boldsymbol{x}}\boldsymbol{y}\rangle^{1/2}$ and the anisotropic diffusion matrix $D_{\boldsymbol{x}}(\boldsymbol{x},t) = \rho(\boldsymbol{x},t)^2 \boldsymbol{n}\boldsymbol{n}^T + \boldsymbol{n}_0\boldsymbol{n}_0^T + \boldsymbol{n}_1\boldsymbol{n}_1^T$ which lowers smoothing in the surface normal $\boldsymbol{n} \in \mathbb{R}^3$ direction and favors smoothness along the corresponding tangential directions $\boldsymbol{n}_0$ and $\boldsymbol{n}_1 = \boldsymbol{n}\times\boldsymbol{n}_0$. Further, the **temporal regularization term** weighted by function $g_t(\boldsymbol{x},t) = \exp\big(-a|\nabla f(\boldsymbol{x},t)|\big)$ accounts for a motion-dependent temporal smoothing. The purpose of the temporal regularization is to reduce surface jittering in scene parts with slow motion. Lastly, the **data term**, represented by function $f: V\times T \mapsto \mathbb{R}$ and smoothness weight λ, avoids trivial solutions of the energy and gives local preferences for an interior or exterior label. Both, the photoconsistency measure $\rho(\boldsymbol{x})$ in $D_{\boldsymbol{x}}$ and f depend on a voting scheme based on surface normal-dependent normalized cross-correlation (NCC) scores, represented by $C_i(\boldsymbol{x},d)$ for each point defined by the ray from camera i through point $\boldsymbol{x}$ at distance d.

$$\rho(\boldsymbol{x}) = \exp\Big[-\mu \sum_{i\in C'} \underbrace{\delta\big(d_i^{\max} = \text{depth}_i(\boldsymbol{x})\big)\cdot C_i(\boldsymbol{x}, d_i^{\max})}_{\text{VOTE}_i(\boldsymbol{x})} \Big] \qquad (2)$$

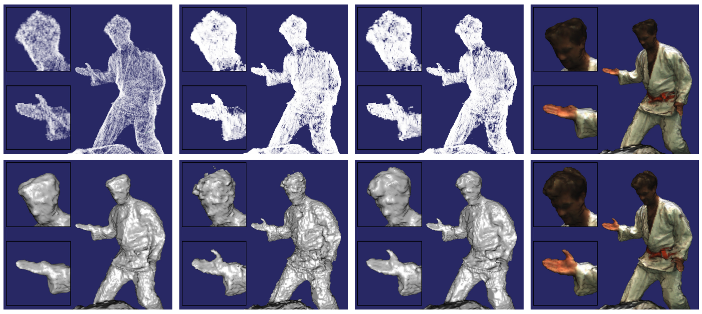
The original voting scheme [1] computes the best depth hypothesis per camera ray as $d_i^{\max} = \arg\max_d C_i(\boldsymbol{x},d)$ and does not enforce any spatial regularity of the votes, which we introduce by the following normal-dependent regularized voting scheme:

$$d_i^{\max} = \arg\max_d \int_{V_{\boldsymbol{x}}} C_i(\boldsymbol{x}-\boldsymbol{y},d)\; \mathcal{G}(\boldsymbol{y};\Sigma_{\boldsymbol{n}}) \;\; d\boldsymbol{y} \;\; , \qquad (3)$$

where $\mathcal{G}(\boldsymbol{y};\S igma_{\boldsymbol{n}})$ is a normal-aligned anisotropic 3D Gaussian. We use surface normals at three places within our method: (a) NCC score, (b) voting scheme regularization and (c) anisotropic surface regularization. To estimate normals, we run our algorithm in two passes (see Fig. 1):

- Pass 1: camera-to-point direction as normal for (a) and (b), isotropic surface regularization with high λ for (c)
- Pass 2: normals from the previous pass for (a),(b) and (c) with lower λ for surface smoothness as desired

Finally, we propose an efficient GPU-accelerated primal-dual optimization of energy (1) which allows for comparatively low computation times. Our model yields significantly improved results over [5] which also compare well to other state-of-the-art reconstruction methods (see Fig. 2).



without normal integration [5] | pass 1 | pass 2 (proposed) | comparison column 1 & 3

Figure 1: The photoconsistency measure $\rho(\boldsymbol{x})$ (top) and corresponding reconstructions (bottom) with and without the proposed normal integration. Pass 1 demonstrates the effect of the proposed normal-based voting regularization. Pass 2 adds anisotropic regularization and improved photoconsistency scores from pass 1.

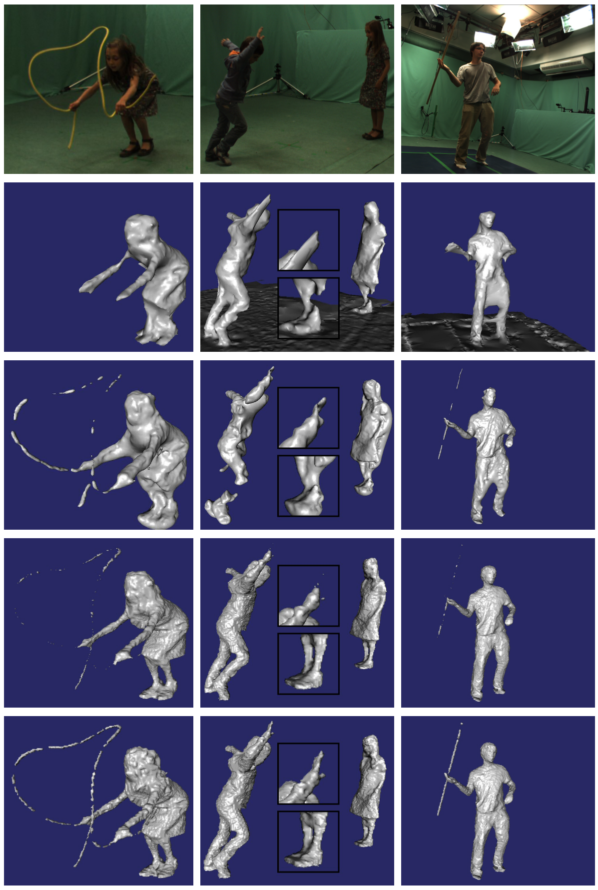


Figure 2: Reconstruction results of different methods. From top to bottom: input, Jancosek and Pajdla [3], PMVS+Poisson [2], Oswald and Cremers [5] and proposed. Our method recovers fine details and reduces temporal surface jittering.

[1] Carlos Hernández Esteban and Francis Schmitt. Silhouette and stereo fusion for 3d object modeling. *CVIU*, 96(3):367–392, Dec 2004.

[2] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE TPAMI*, 32(8):1362–1376, Aug 2010.

[3] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *CVPR*, pages 3121–3128, 2011.

[4] K. Kolev, T. Pock, and D. Cremers. Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. In *ECCV*, Heraklion, Greece, Sep 2010.

[5] Martin R. Oswald and Daniel Cremers. A convex relaxation approach to space time multi-view 3d reconstruction. In *ICCV - Workshop on Dynamic Shape Capture and Analysis (4DMOD)*, 2013.