We propose a convex variational approach to space-time reconstruction which estimates surface normal information and integrates it into the photometric consistency 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 work well on the 4D data sets we consider, as shown in [5, Fig. 5]. With

The energy consists of three terms. An anisotropic spatial regularization term, defined be the norm \(|y|_{D_λ} = (y^TD_λy)^{1/2}\) and the anisotropic diffusion matrix \(D_λ(x,t) = \rho(x,t)^2n^2 + m_λ^2n_0^2 + n_1^2n_0^2\) which lowers smoothing in the surface normal \(n \in \mathbb{R}^3\) direction and favors smoothness along the corresponding tangential directions \(n_0\) and \(n_1 = n \times n_0\). Further, the temporal regularization term \(E_t\) weighted by function \(g_t(x,t) = \exp(-a|\nabla f(x,t)|)\) 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 \(E_f\) represented by function \(f : V \times T \rightarrow \mathbb{R}\) and smoothness weight \(\lambda\), avoids trivial solutions of the energy and gives local preferences for an interior or exterior label. Both, the photoconsistency measure \(\rho(x)\) in \(D_λ\) and \(f\) depend on a voting scheme based on surface normal-dependent normalized cross-correlation (NCC) scores, represented by \(C_i(x,d)\) for each point defined by the ray from camera \(i\) through point \(x\) at distance \(d\).

\[
E(u) = \int_{V \times T} \left[ |\nabla_{x,t}u|_{D_λ} + g_t |\nabla f| + \lambda f^2 \right] \, dx dt
\]

(1)

The energy consists of three terms. An anisotropic spatial regularization term, defined be the norm \(|y|_{D_λ} = (y^TD_λy)^{1/2}\) and the anisotropic diffusion matrix \(D_λ(x,t) = \rho(x,t)^2n^2 + m_λ^2n_0^2 + n_1^2n_0^2\) which lowers smoothing in the surface normal \(n \in \mathbb{R}^3\) direction and favors smoothness along the corresponding tangential directions \(n_0\) and \(n_1 = n \times n_0\). Further, the temporal regularization term \(E_t\) weighted by function \(g_t(x,t) = \exp(-a|\nabla f(x,t)|)\) 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 \(E_f\), represented by function \(f : V \times T \rightarrow \mathbb{R}\) and smoothness weight \(\lambda\), avoids trivial solutions of the energy and gives local preferences for an interior or exterior label. Both, the photoconsistency measure \(\rho(x)\) in \(D_λ\) and \(f\) depend on a voting scheme based on surface normal-dependent normalized cross-correlation (NCC) scores, represented by \(C_i(x,d)\) for each point defined by the ray from camera \(i\) through point \(x\) at distance \(d\).

\[
\rho(x) = \exp \left( -\mu \sum_{t \in T} \delta \left( |\nabla_{x,t}u|_{D_λ} - \text{VOTE}(x)\right) \right)
\]

(2)

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

\[
d_i^{\text{max}} = \max_d \int_{V_i} C_i(x-y,d) \mathcal{G}(y;\sigma_a) \, dy
\]

(3)

where \(\mathcal{G}(y;\sigma_{\text{normal}})\) 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 \(\lambda\) for (c)

Pass 2: normals from the previous pass for (a), (b) and (c) with lower \(\lambda\) 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).