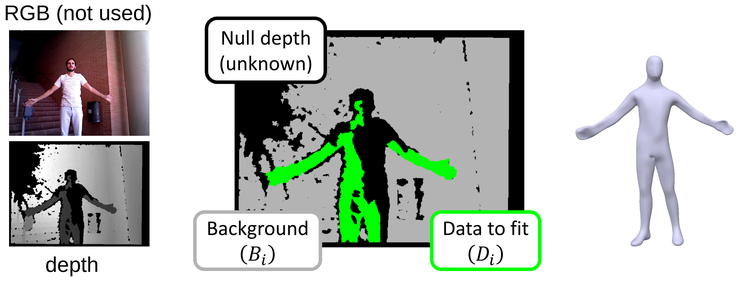
An Efficient Background Term for 3D Reconstruction and Tracking with Smooth Surface Models

Mariano Jaimez[1,2], Tom Cashman[3], Andrew Fitzgibbon[3], Javier González-Jiménez[2] and Daniel Cremers[1]
[1]TU Munich, [2]University of Málaga, [3]Microsoft

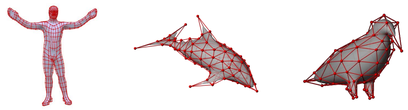
CVPR July 21-26 HONOLULU 2017

Problem Context: Reconstruct or track objects from depth images.



We employ **subdivision surfaces** to model 3D shapes [1]. Pros:

- They allow gradient-based optimization jointly over model and correspondences.
- Compact representation.

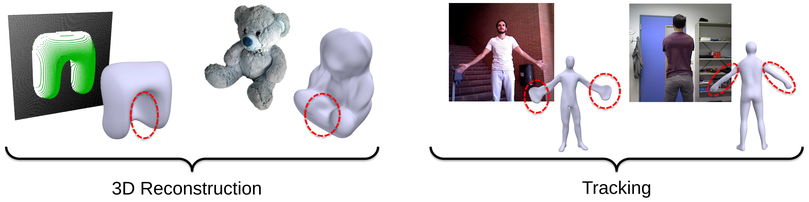


Standard approach: Obtain the control vertices X and the camera poses ξ_i that minimize the distance between the model and the foreground data (D_i) of one or multiple depth images i:

$$E_{ij}^{p}(X,\xi_i,u) = \left\| p_{ij} - T(\xi_i)\, s\left(u_{ij}|X\right)\right\|_T^2, \quad E_{ij}^{n}(X,\xi_i,u) = \lambda_n \left\| n_{ij} - R(\xi_i)\, s^{\perp}\left(u_{ij}|X\right)\right\|_T^2$$

$$\min_{X,\xi,\mathcal{U}} \left\{ \sum_i \sum_{j\in D_i} E_{ij}^{p}(X,\xi_i,u) + E_{ij}^{n}(X,\xi_i,u) \right\}, \qquad u_{ij}\text{: correspondences (opt. jointly)}$$

Problem: The 3D model spills over the background beyond the real object:



Background term: Exploit background data (B_i) to enforce silhouette consistency of the model.

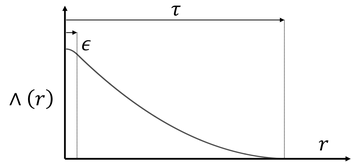
Existing solutions [2,3] include a distance-transform-based background term (DT). This term/formulation:

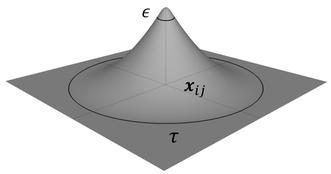
- fails to capture concavities.
- is highly affected by wrongly segmented pixels.
- does not handle uncertainty (null depth) properly.

Our contribution: New background term (SK) that formulates raycasting as a differentiable energy.

$$E_{ij}^{b}(X,\xi_i) \sim \begin{cases} 1 & \text{if } \min_u \left\| x_{ij} - \pi(s(u|X),\xi_i)\right\| = 0 \\ 0 & \text{otherwise} \end{cases},$$

$$E_{ij}^{b}(X,\xi_i) = \wedge\left(\min_u \left\| x_{ij} - \pi(s(u|X),\xi_i)\right\|\right) = \max_u \wedge\left(\left\| x_{ij} - \pi(s(u|X),\xi_i)\right\|\right)$$





Overall formulation:

$$\min_{X,\xi,\mathcal{U}} \left\{ \sum_i \sum_{j\in D_i} E_{ij}^{p}(X,\xi_i,u) + E_{ij}^{n}(X,\xi_i,u) + \sum_i \sum_{j\in B_i} E_{ij}^{b}(X,\xi_i) \right\}$$

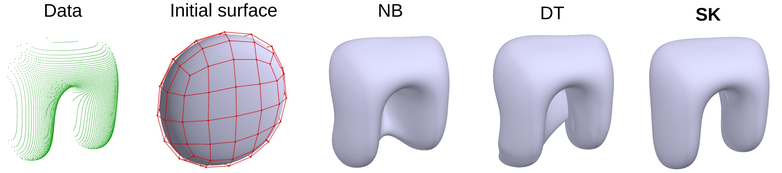
Distance Transform (DT)

Shrinking Kernel (SK)

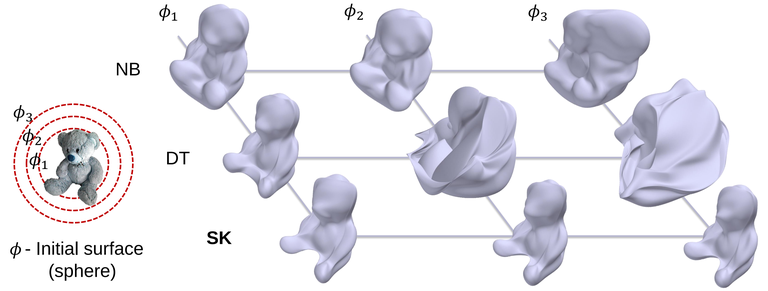
Experiments and Results:

NB: No background term
DT: Distance transform
SK: Shrinking kernel (ours)

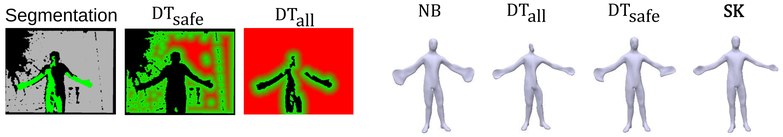
1) Modeling an arch



2) Multi-view reconstruction with wrongly segmented pixels



3) Tracking under poor illumination conditions (high uncertainty)



References:

[1] T. Cashman and A. Fitzgibbon. *What shape are dolphins? Building 3D morphable models from 2D images*. IEEE Trans. On Pattern Analysis and Machine Intelligence, 35(1), pp 232-244, 2013.

[2] V. Ganapathi, C. Plagemann, D. Koller and S. Thrun. *Real-time human pose tracking from range data*. ECCV, pp 738-751, 2012.

[2] S. Vicente and L. Agapito. *Balloon shapes: Reconstructing and deforming objects with volume from images*. 3DV, pp 223-230, 2013.