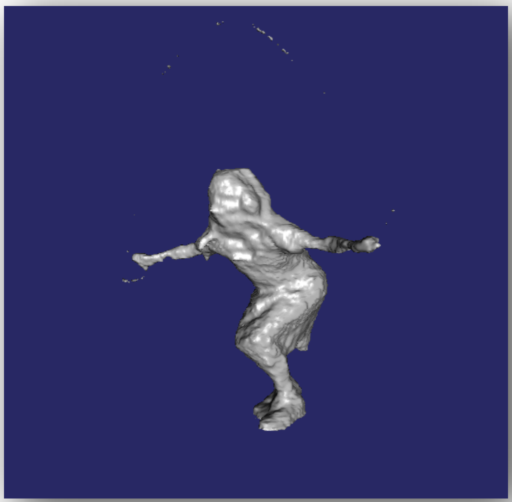
Computer Vision Group

Technische Universität München

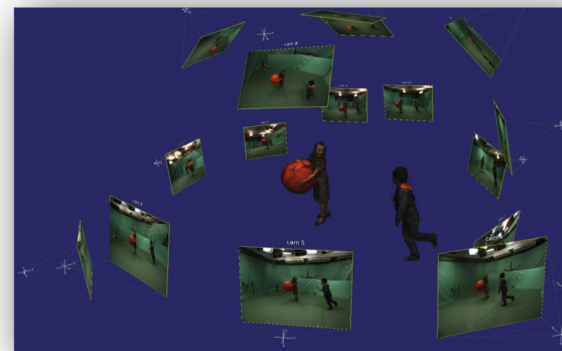
Spatio-temporal Multi-view 3D Reconstruction with Generalized Connectivity Constraints

Martin Oswald, Jan Stühmer and Daniel Cremers

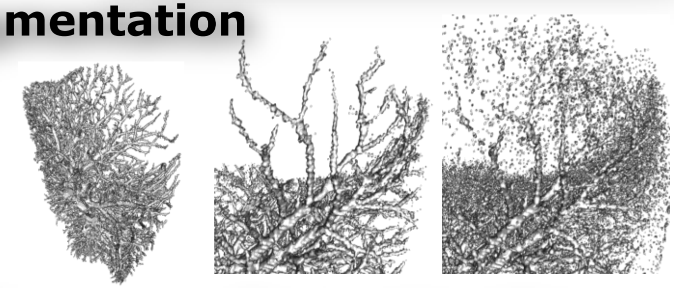


Overview

- **Spatio-temporal Multi-view Reconstruction**
  Oswald and Cremers, ICCV'13, 4DMOD Workshop

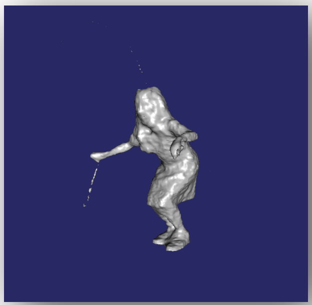


- **Review of Connectivity Constraints for Segmentation**
  Stühmer, Schröder, Cremers, ICCV'13

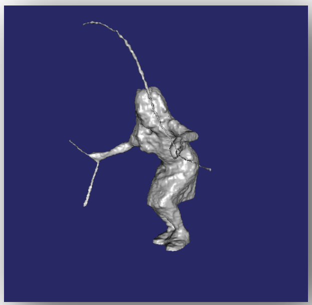


- **Generalized Connectivity Constraints for Multi-View Reconstruction**
  Oswald, Stühmer, Cremers

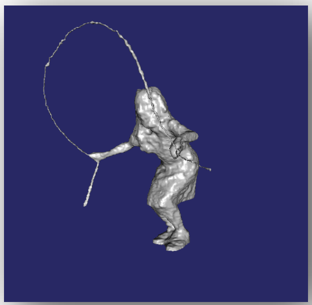
input



no connectivity



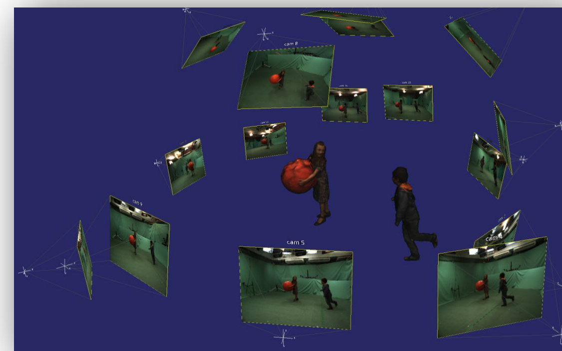
with connectivity



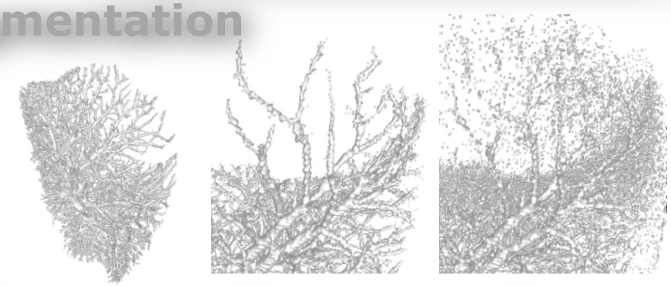
generalized connectivity

Overview

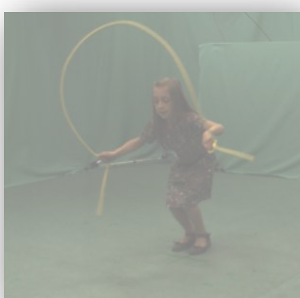
- **Spatio-temporal Multi-view Reconstruction**
  Oswald and Cremers, ICCV’13, 4DMOD Workshop



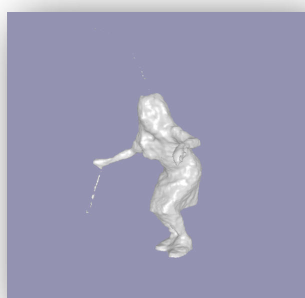
- **Review of Connectivity Constraints for Segmentation**
  Stühmer, Schröder, Cremers, ICCV’13



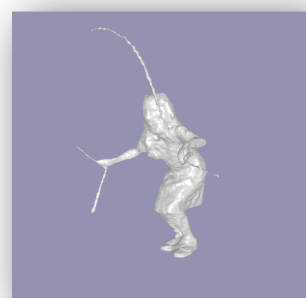
- **Generalized Connectivity Constraints for Multi-View Reconstruction**
  Oswald, Stühmer, Cremers



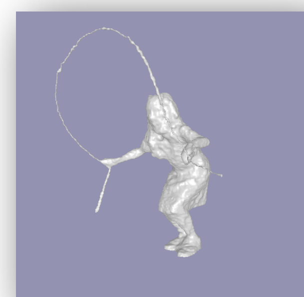
input



no connectivity



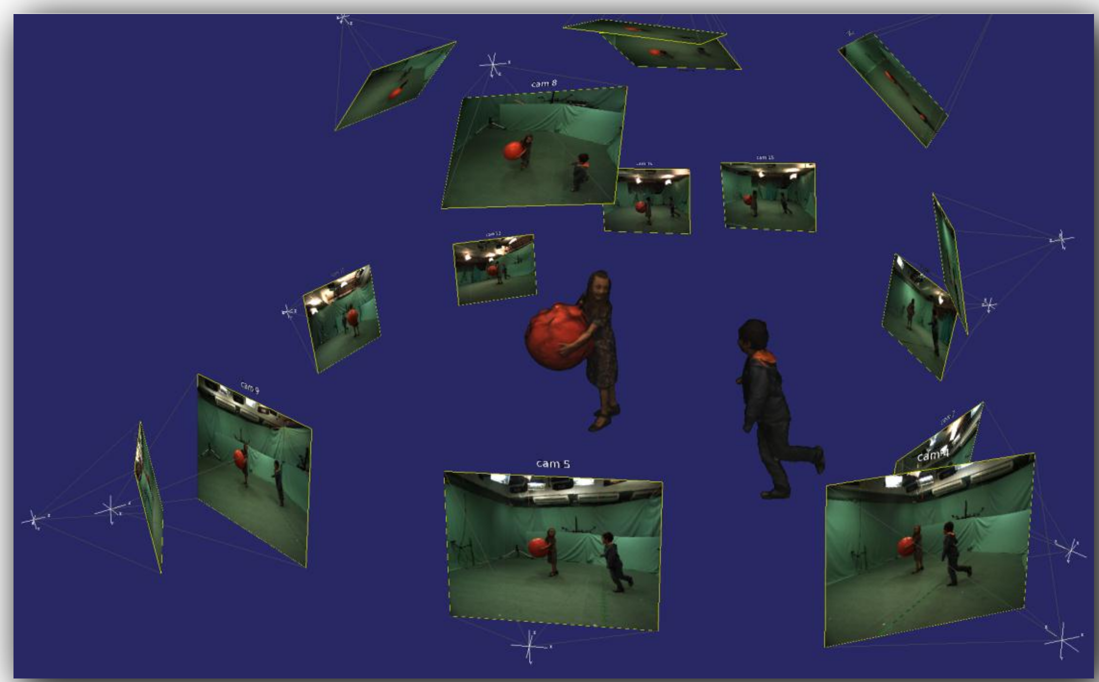
with connectivity



generalized connectivity

Spatio-temporal Multi-view Reconstruction

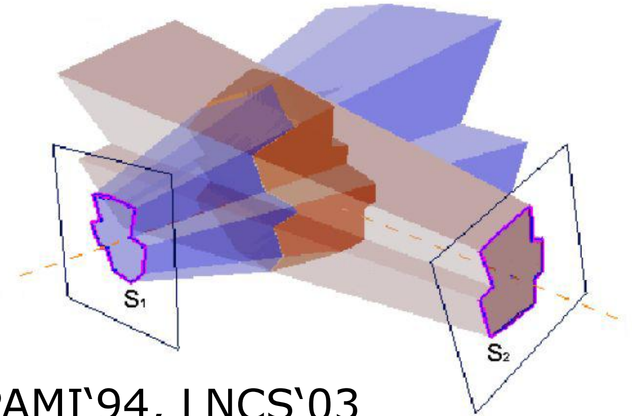
Input:
- images $\{I_i(t)\}_{i=1}^N$
- calibration $\{\pi_i\}_{i=1}^N$
- silhouettes $\{S_i(t)\}_{i=1}^N$



Output: space-time surface (mesh)

Visual hull

$$\mathcal{VH}(t) = \bigcap_{i=1}^{N} \pi_i^{-1}(S_i(t))$$



Baumgart, Stanford'74; Laurentini et al. TPAMI'94, LNCS'03

Applications

- Free Viewpoint TV
- Markerless Motion Capture
- Motion Analysis for Sports
- Special Effects for Movies
  e.g. „Bullet Time" Effect from „The Matrix"

The making of:

Related Work

3D Reconstruction

$$\min_u \int_V \rho\,|\nabla u|\,dx + \lambda \int_V f u\,dx$$

weighted TV + data term

Kolev et al. IJCV'09

$$u : V \mapsto \{0,1\}$$

Interior/exterior labeling

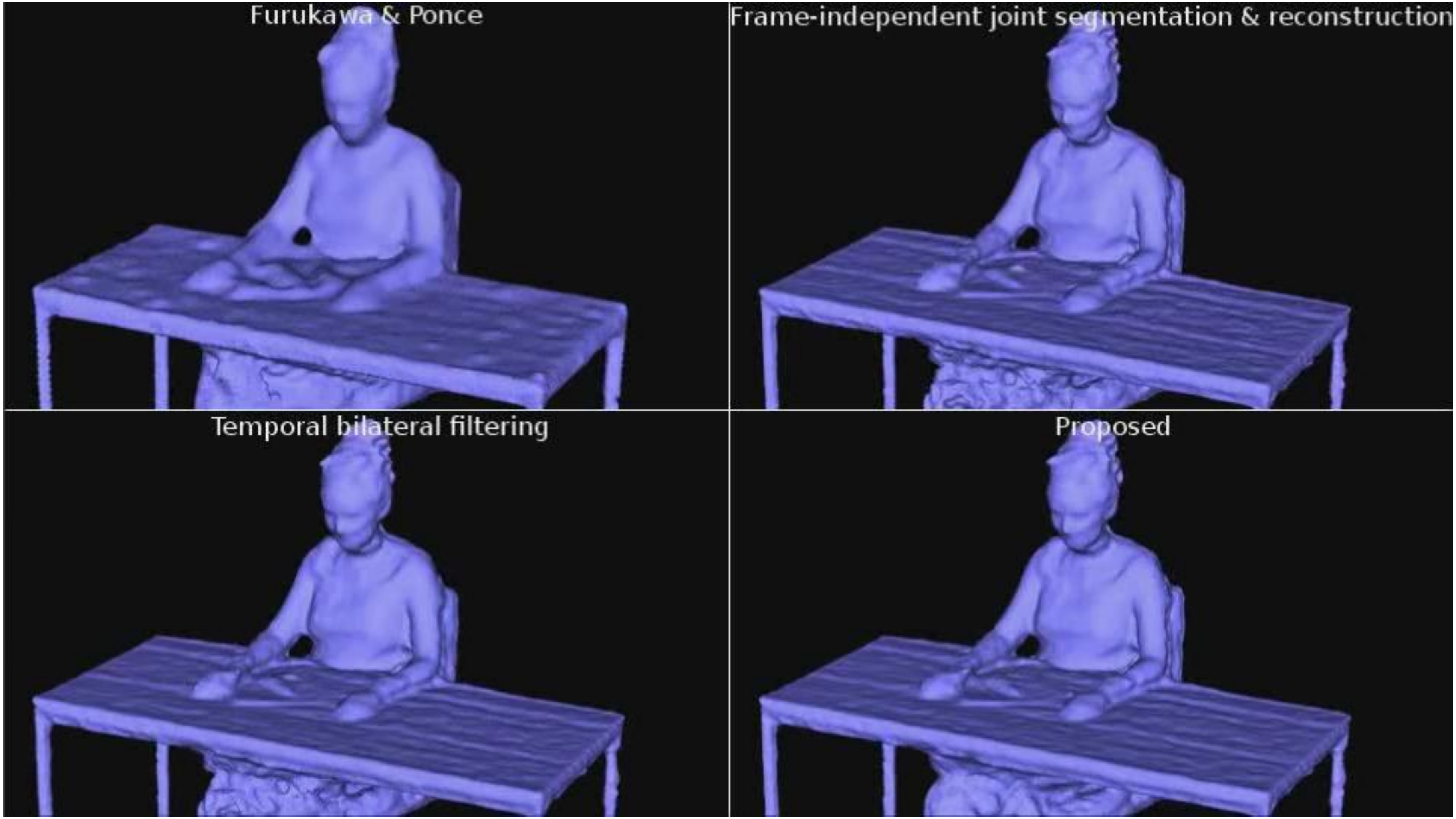
Pros

- topology changes are easily handled
- global optimization
- natural regularization in 3D space rather than in the image domain
- desirable interpolation/extrapolation behavior for redundant/missing data

Cons

- slow: runtimes of hours
- memory intensive: volume resolution limits reconstruction accuracy
- data terms are not well suited for sparse camera setups

Why 4D Reconstruction?



Guillemaut, Hilton, 3DIMPVT'12

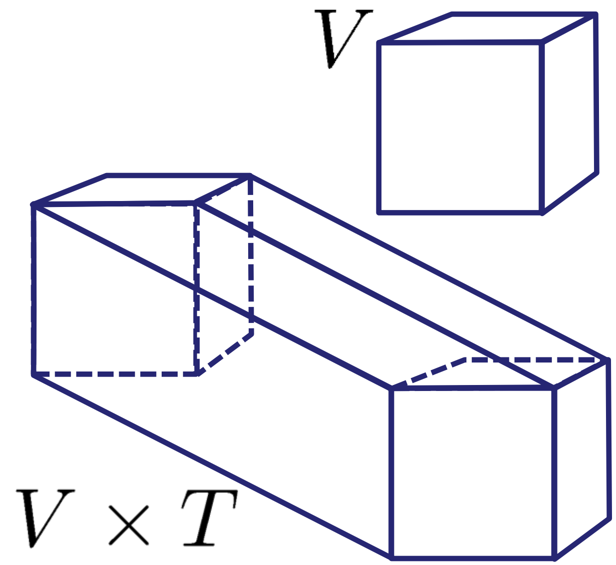
Noisy Input Images ⟶ Noisy 3D Models

Spatio-Temporal Multi-view Reconstruction

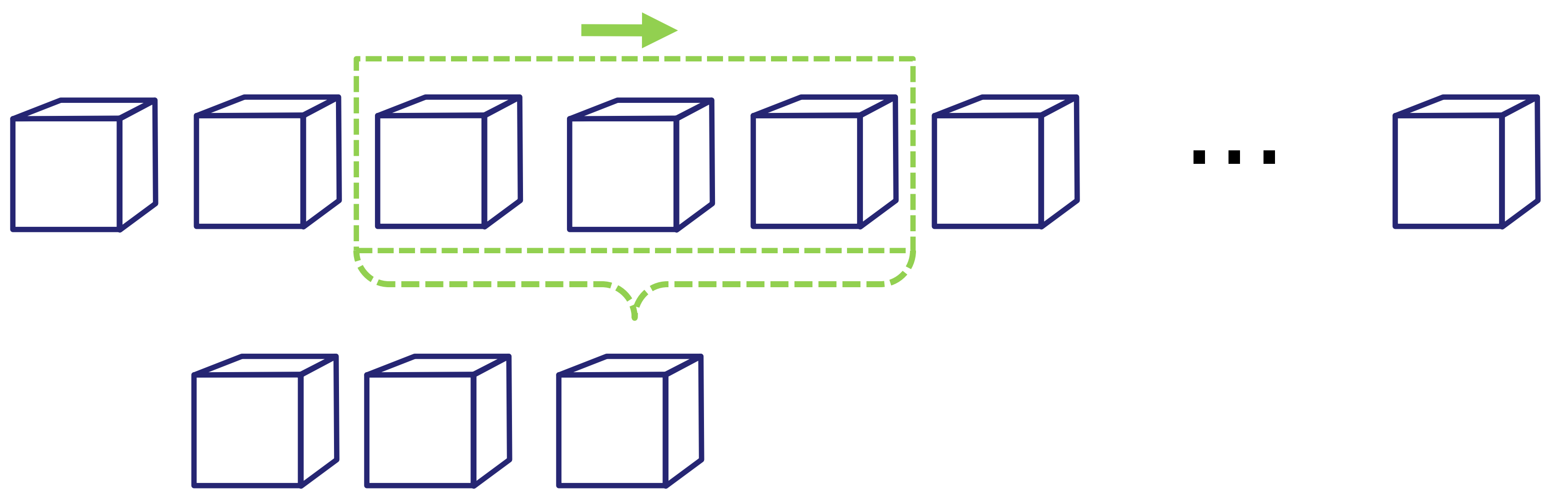
$$\min_u \int_{V\times T} (\rho\,|\nabla_x u| + g_t\,|\nabla_t u|)\,dx\,dt + \lambda \int_{V\times T} f u\,dx\,dt$$

spatial temporal
regularization term

data term



- interior/exterior labeling $u : V \times T \mapsto \{0, 1\}$
- sliding window optimization (for fixed window size |T|)



Photoconsistency

$$\min_u \int_{V\times T} (\rho\,|\nabla_x u| + g_t\,|\nabla_t u|)\,dx\,dt + \lambda \int_{V\times T} f\,u\,dx\,dt$$

$\rho : V\times T \mapsto [0,1]$

$\rho = 0$ perfect photometric match

$\rho = 1$ no photometric match

voting scheme by Hernández et al. CVIU'04:

$$C_i(x,d) = \sum_{j\in\mathcal{C}'\setminus i} w_i^j(x)\cdot \mathrm{NCC}\Big(\pi_i\big(r_i(x,d)\big), \pi_j\big(r_i(x,d)\big)\Big)$$

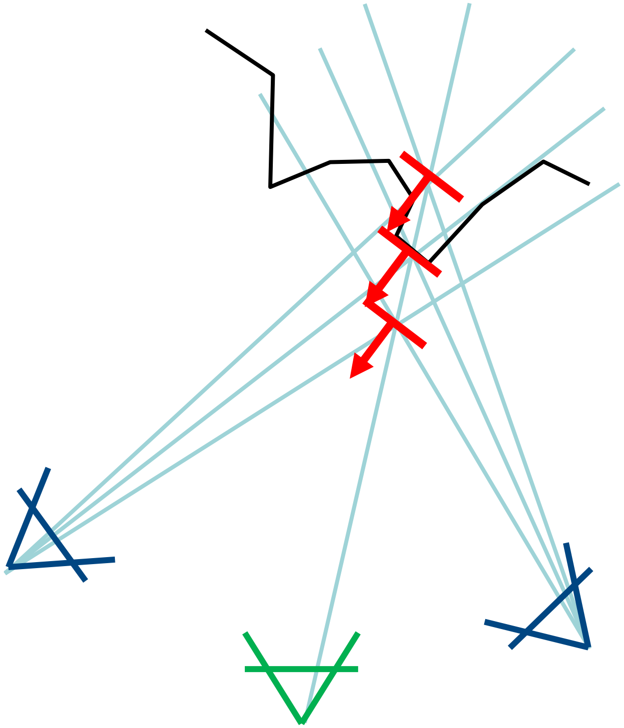
$$d_i^{\max} = \arg\max_d C_i(\boldsymbol{x}, d)$$

$$\rho(x,t) = \exp\Big[-\mu\sum_{i\in\mathcal{C}'} \underbrace{\delta\big(d_i^{\max}{=}\mathrm{depth}_i(x)\big)\cdot \bar{C}_i(x, d_i^{\max})}_{\mathrm{VOTE}_i(x)}\Big]$$

1 of 20 images

Kolev et al.

proposed



Proposed Data Term

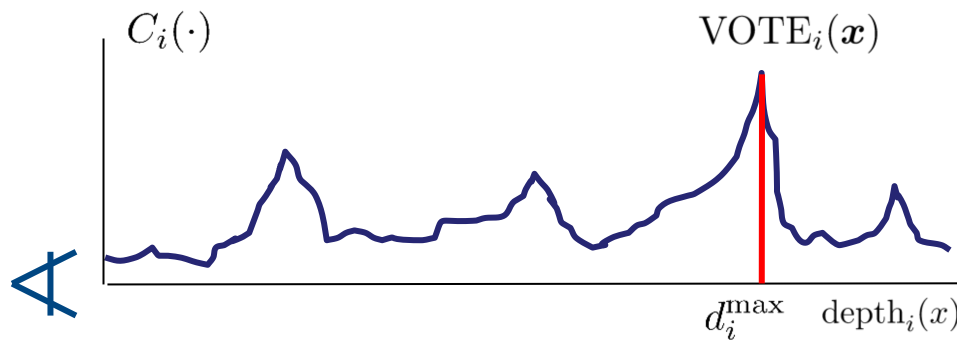
$$\min_u \int_{V\times T} \left(\rho\,|\nabla_x u| + g_t\,|\nabla_t u|\right) dx\,dt + \lambda \int_{V\times T} \boxed{f}\, u\, dx\,dt$$

$f : V\times T \mapsto \mathbb{R}$

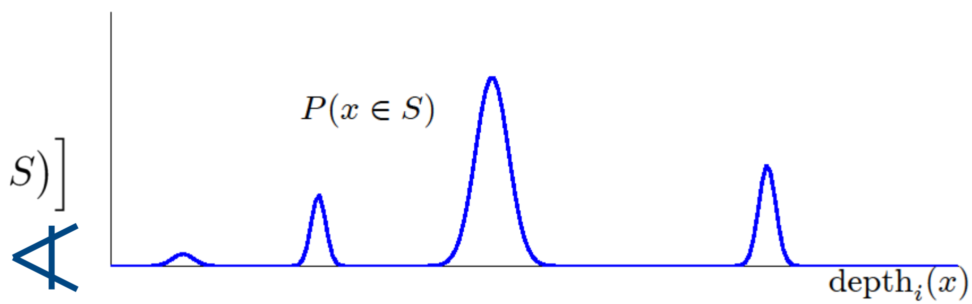
$f < 0$ favor interior labeling

$f = 0$ neutral – no preference

$f > 0$ favor exterior labeling

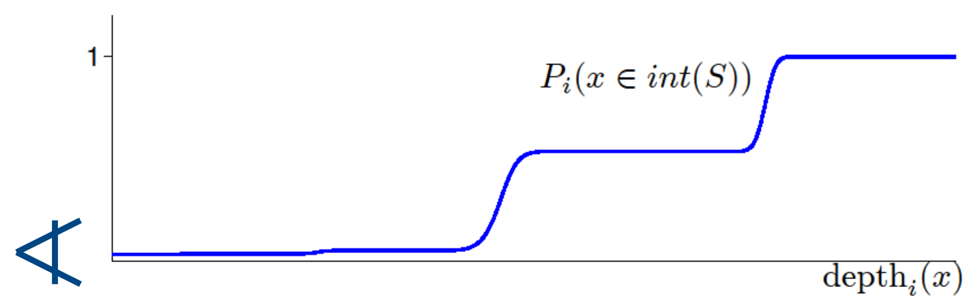


$$P_i(x \in S) = 1 - \frac{1}{Z}\exp\left[-\eta \cdot \mathrm{VOTE}_i(x)\right]$$



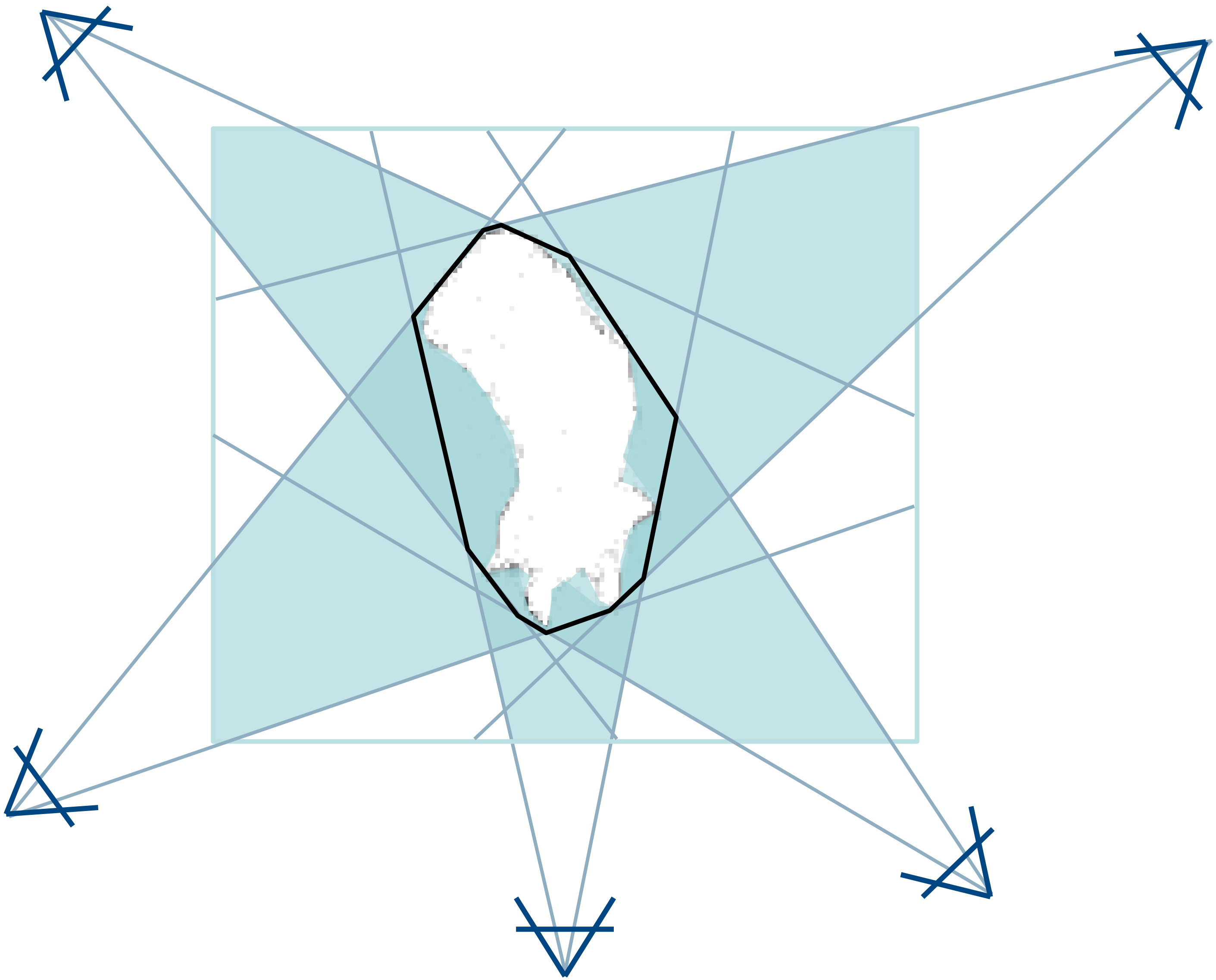
$$P_i\big(x \in int(S)\big) = \prod_{j=1}^{N} \prod_{\mathrm{depth}_i(x) < d \leq d_i^{\max}} \left[1 - P_j\big(r_i(x,d) \in S\big)\right]$$

$$P\big(x \in int(S)\big) = \prod_{i=1}^{N} P_i\big(x \in int(S)\big)$$



$$f(x,t) = -\ln\left(\frac{1 - P\big(x \in int(S)\big)}{P\big(x \in int(S)\big)}\right)$$

Proposed Data Term



Photoconsistency and Data Term

$$\min_u \int_{V\times T} (\rho \, |\nabla_x u| + g_t \, |\nabla_t u|) dx dt + \lambda \int_{V\times T} f u \, dx dt$$

input

photoconsistency $\rho : V\times T \mapsto [0,1]$

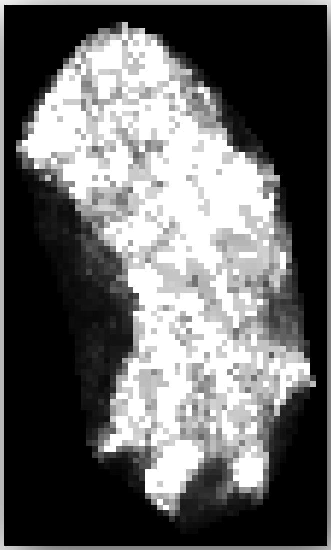
data term $f : V\times T \mapsto \mathbb{R}$

1 of 20 images

Kolev et al.

proposed

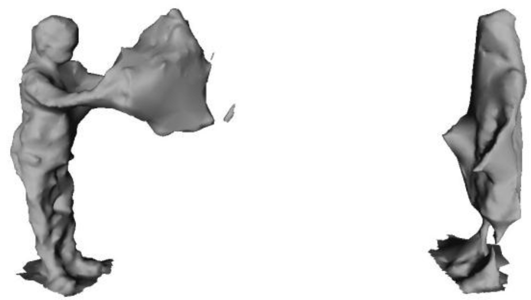
Kolev et al.



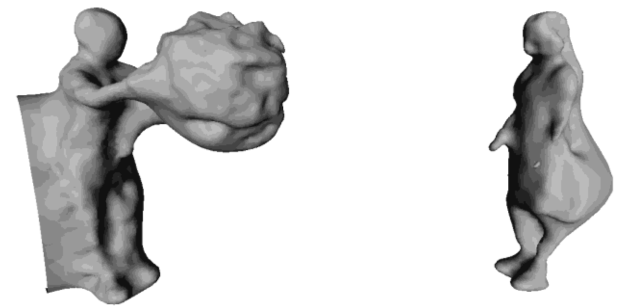
proposed

Photoconsistency and Data Term

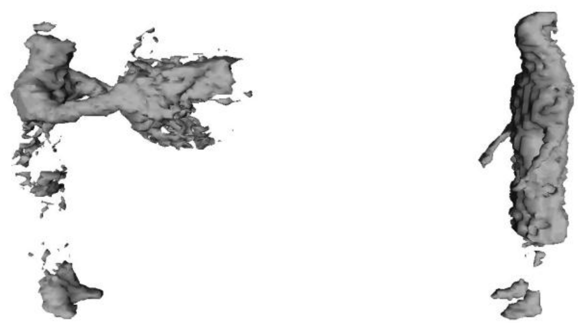
input



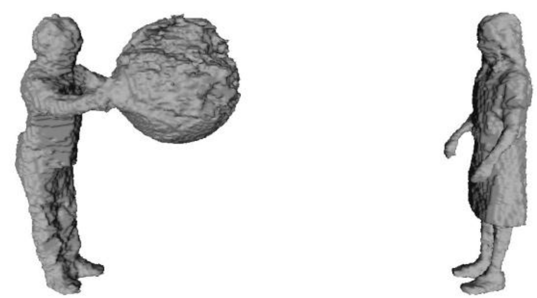
Jancosek and Pajdla, CVPR‘11



PMVS + Poisson
Furukawa et al. PAMI‘10



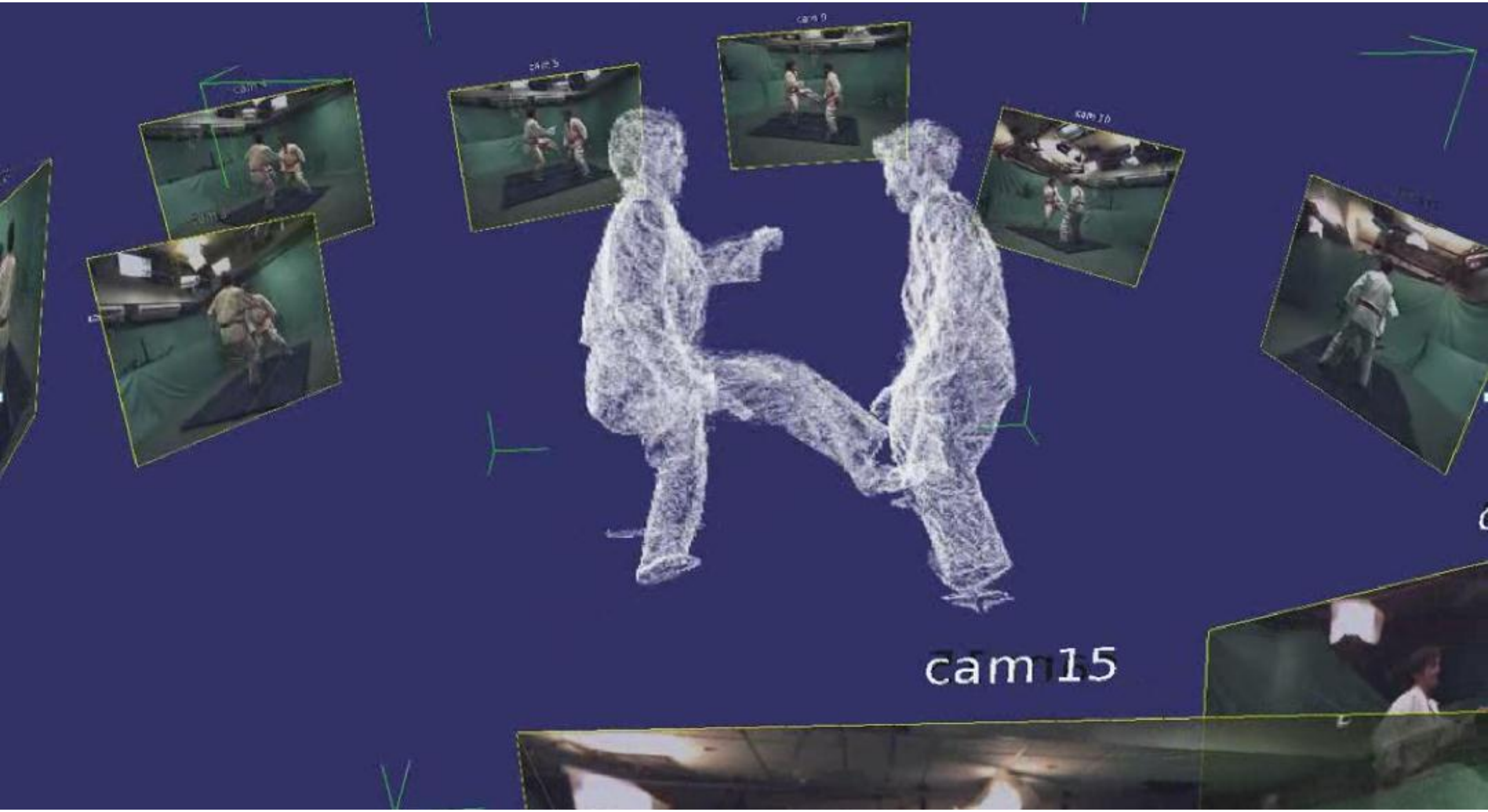
Kolev et al. IJCV‘09



proposed

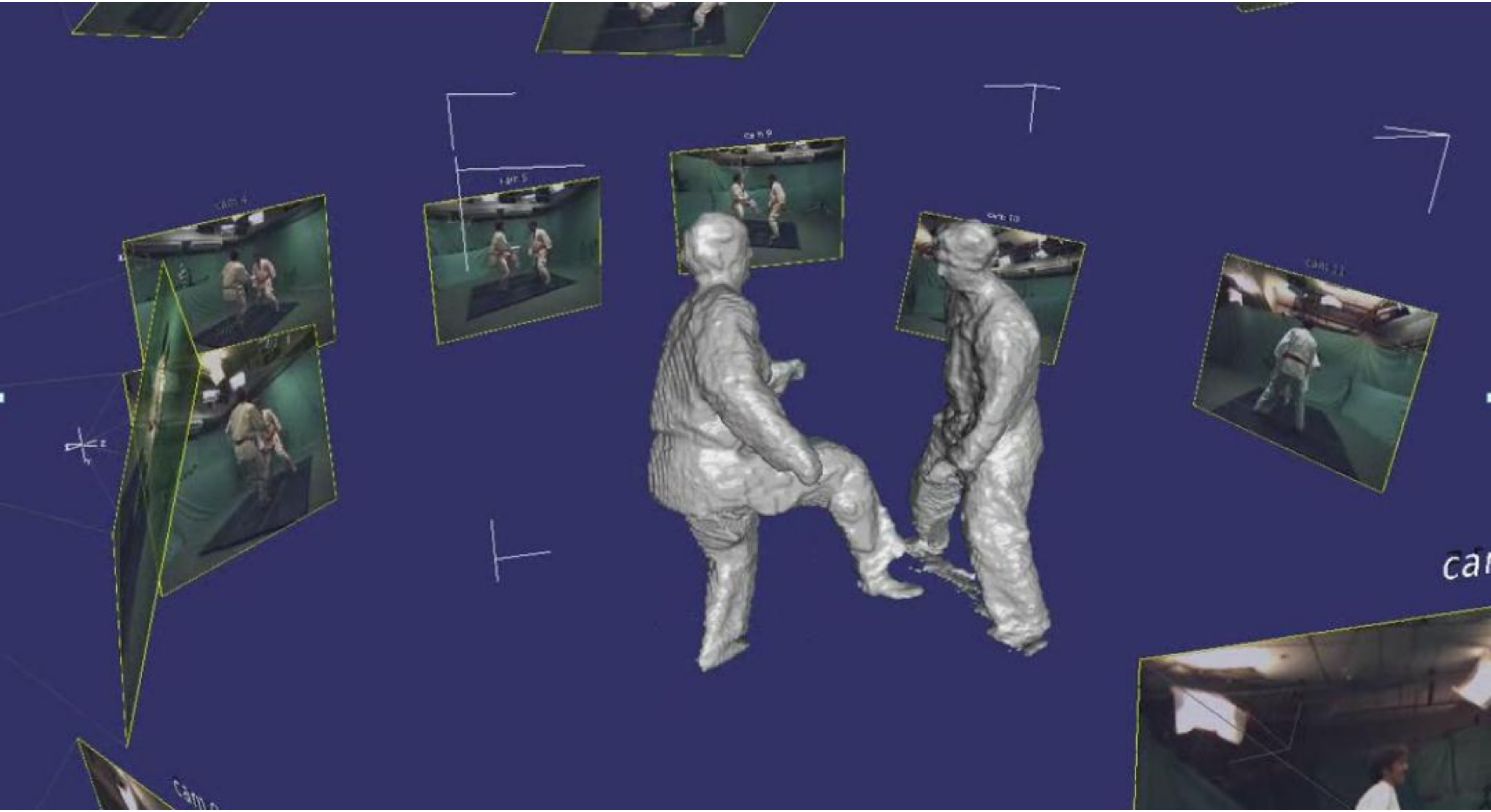
Photoconsistency and Data Term

$$\min_u \int_{V\times T} (\rho\,|\nabla_x u| + g_t\,|\nabla_t u|)\,dx\,dt + \lambda \int_{V\times T} f u\,dx\,dt$$



Reconstruction Result

$$\min_{u} \int_{V\times T} (\rho\,|\nabla_x u| + g_t\,|\nabla_t u|)dxdt + \lambda \int_{V\times T} f u\, dxdt$$

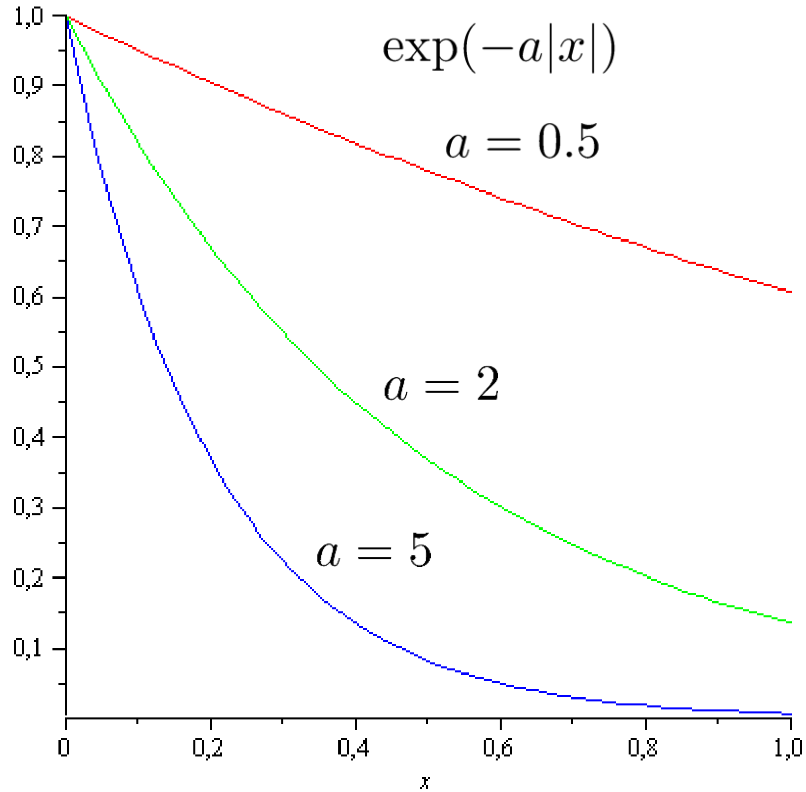


Temporal Regularization

$$\min_u \int_{V\times T} (\rho\,|\nabla_x u| + \boxed{g_t}\,|\nabla_t u|)\,dxdt + \lambda \int_{V\times T} f u\, dxdt$$

$g_t = \text{const}$ → doesn‘t work well with motion

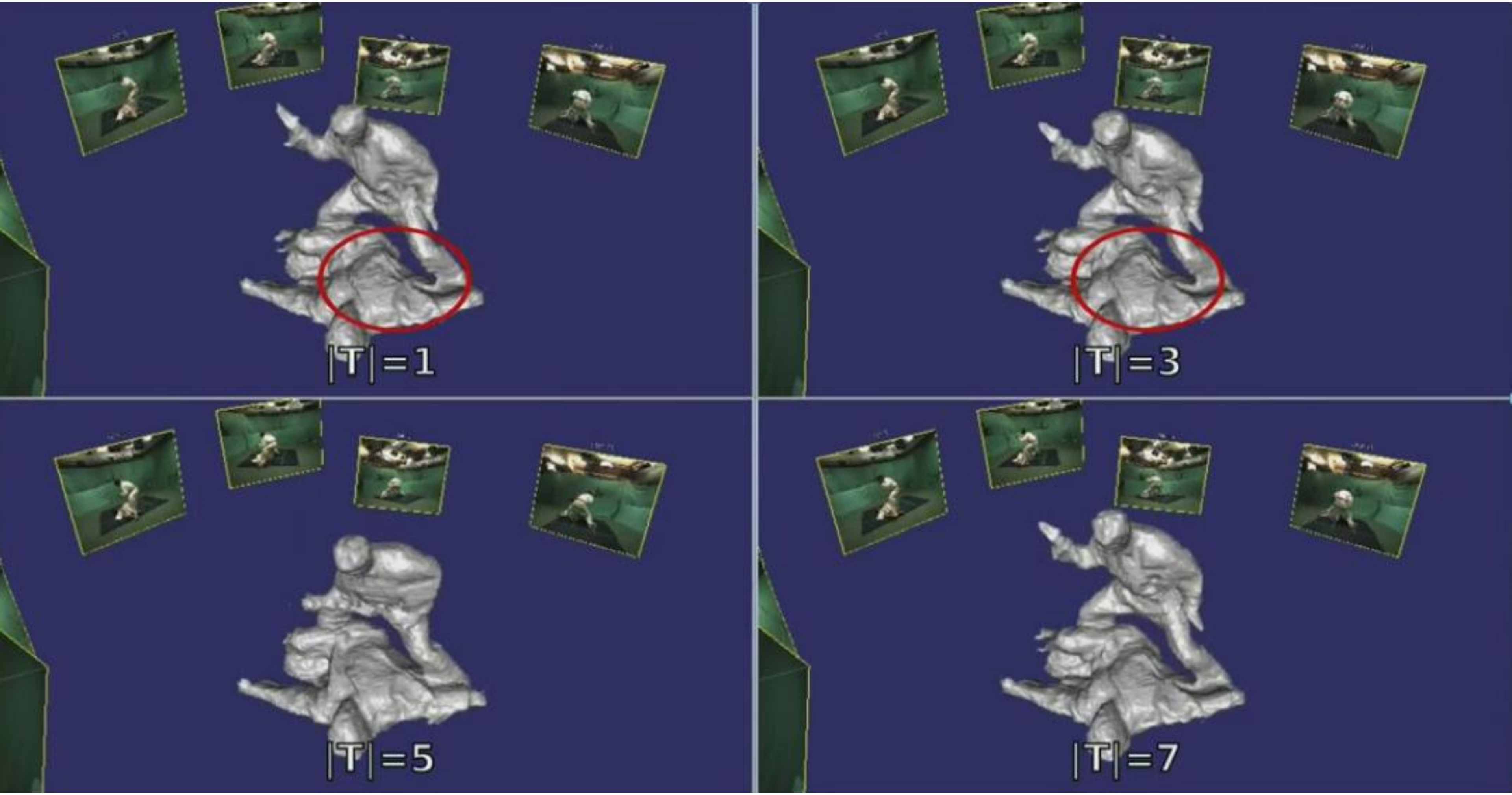
$$g_t(x,t) = \exp\big(-a|\nabla_t f(x,t)|\big)$$



Temporal Regularization

different window sizes |T| (a=1)

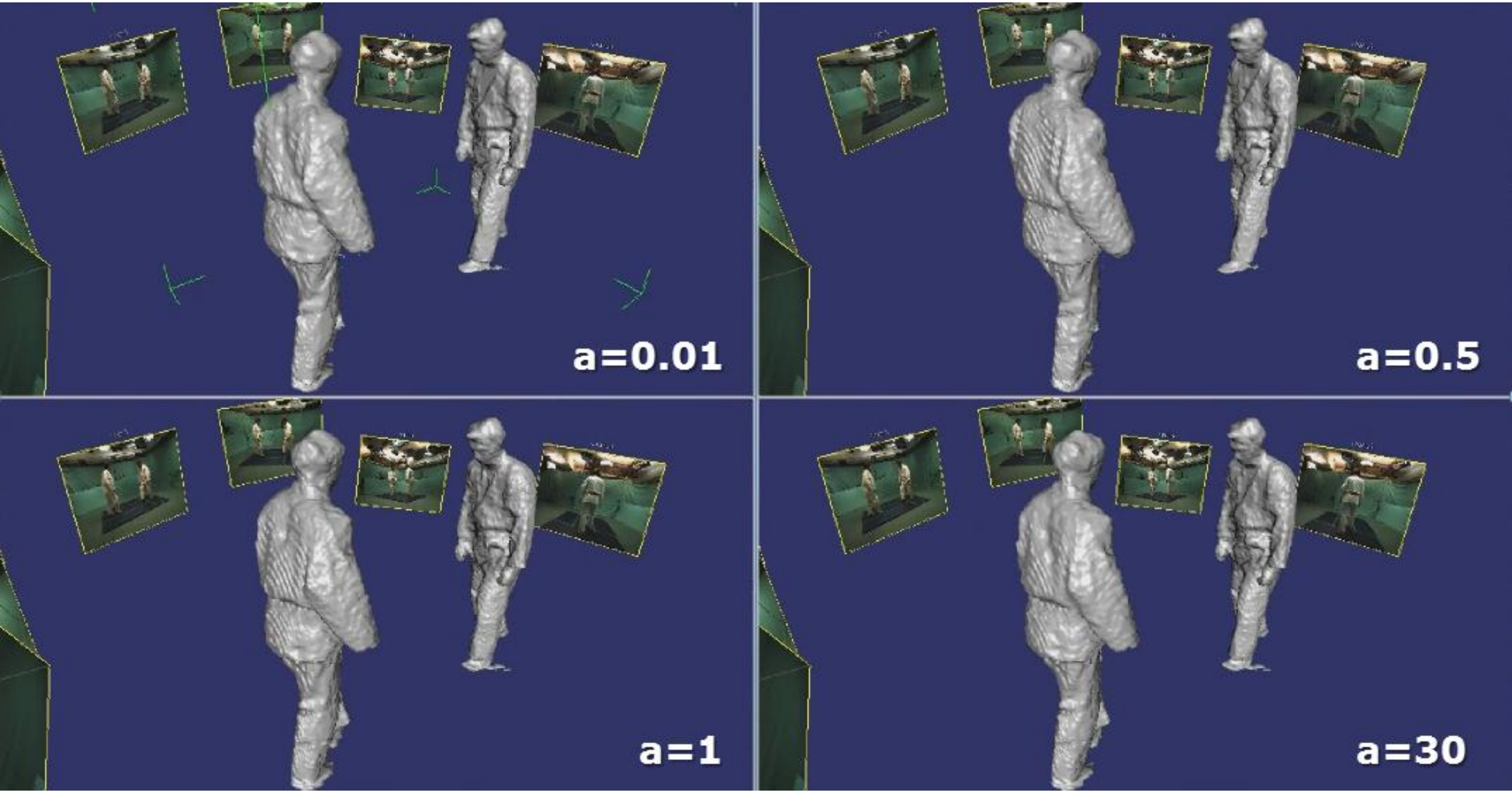
$$g_t(x,t) = \exp\big(-a|\nabla_t f(x,t)|\big)$$



Temporal Regularization

different smoothness values a (|T|=3)

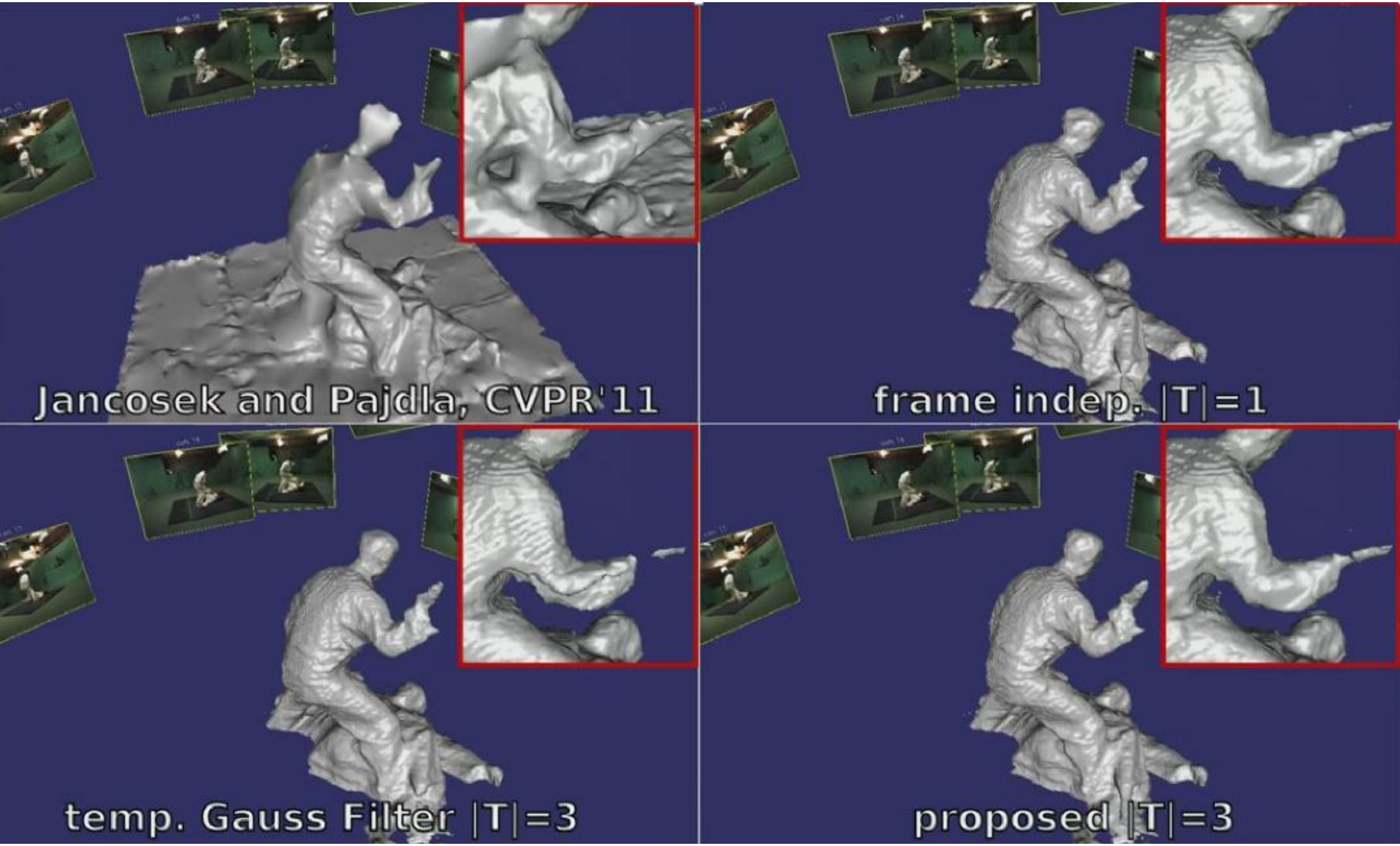
$$g_t(x,t) = \exp\big(-a|\nabla_t f(x,t)|\big)$$



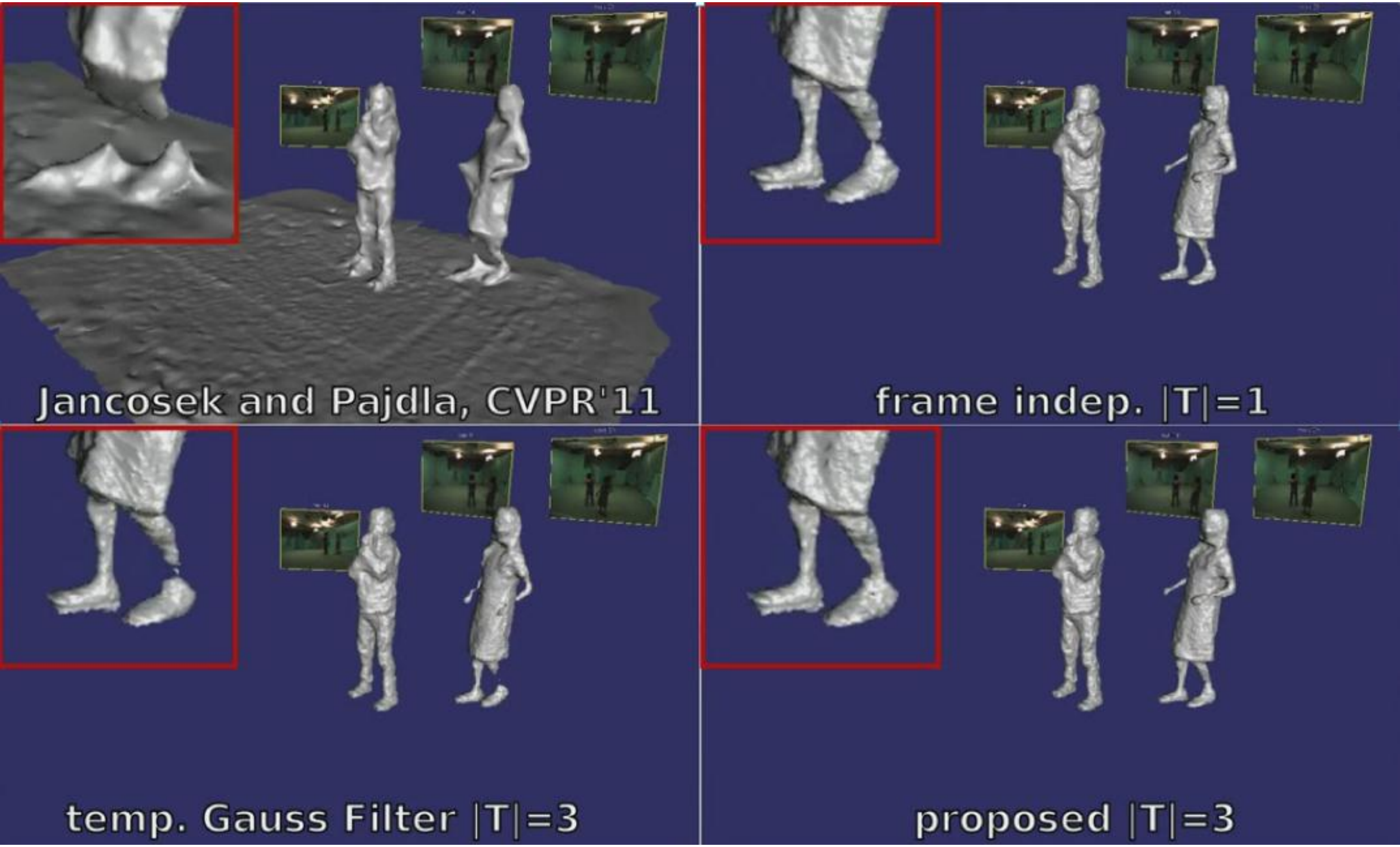
Method Comparison

- Jancosek and Pajdla, CVPR‘11
- Furukawa et al. PAMI’10: PMVS + Poisson Surface reconstruction
- baseline (|T|=1, temp. indepentend)
- baseline (|T|=1, temp. indepentend) + temp. Gaussian filtering (|T|=3)
- proposed (|T|=3)

Method Comparison

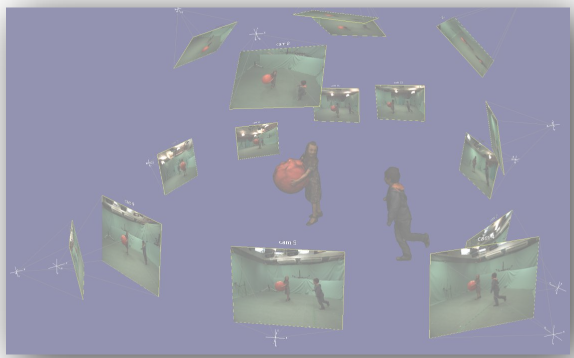


Method Comparison

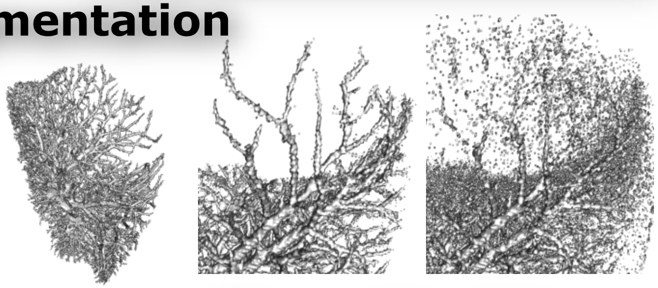


Overview

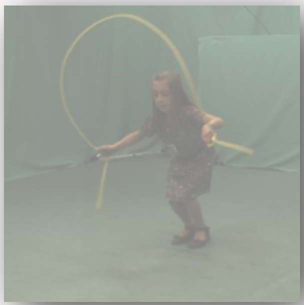
- **Spatio-temporal Multi-view Reconstruction**
  Oswald and Cremers, ICCV’13, 4DMOD Workshop



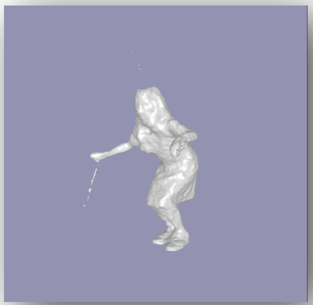
- **Review of Connectivity Constraints for Segmentation**
  Stühmer, Schröder, Cremers, ICCV’13



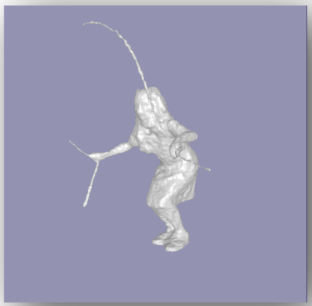
- **Generalized Connectivity Constraints for Multi-View Reconstruction**
  Oswald, Stühmer, Cremers



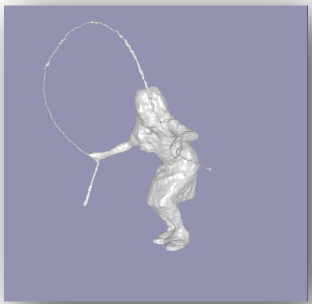
input



no connectivity

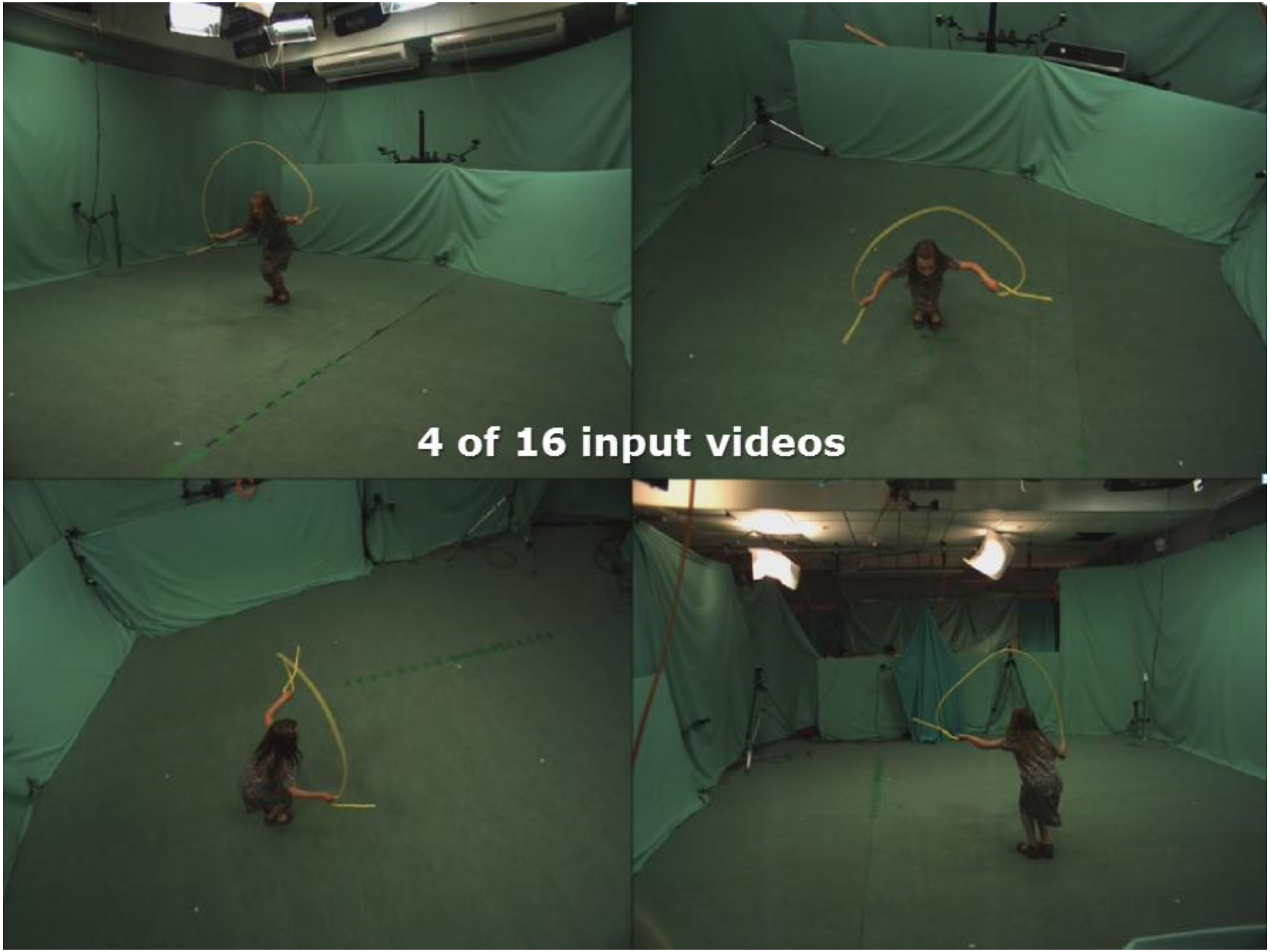


with connectivity

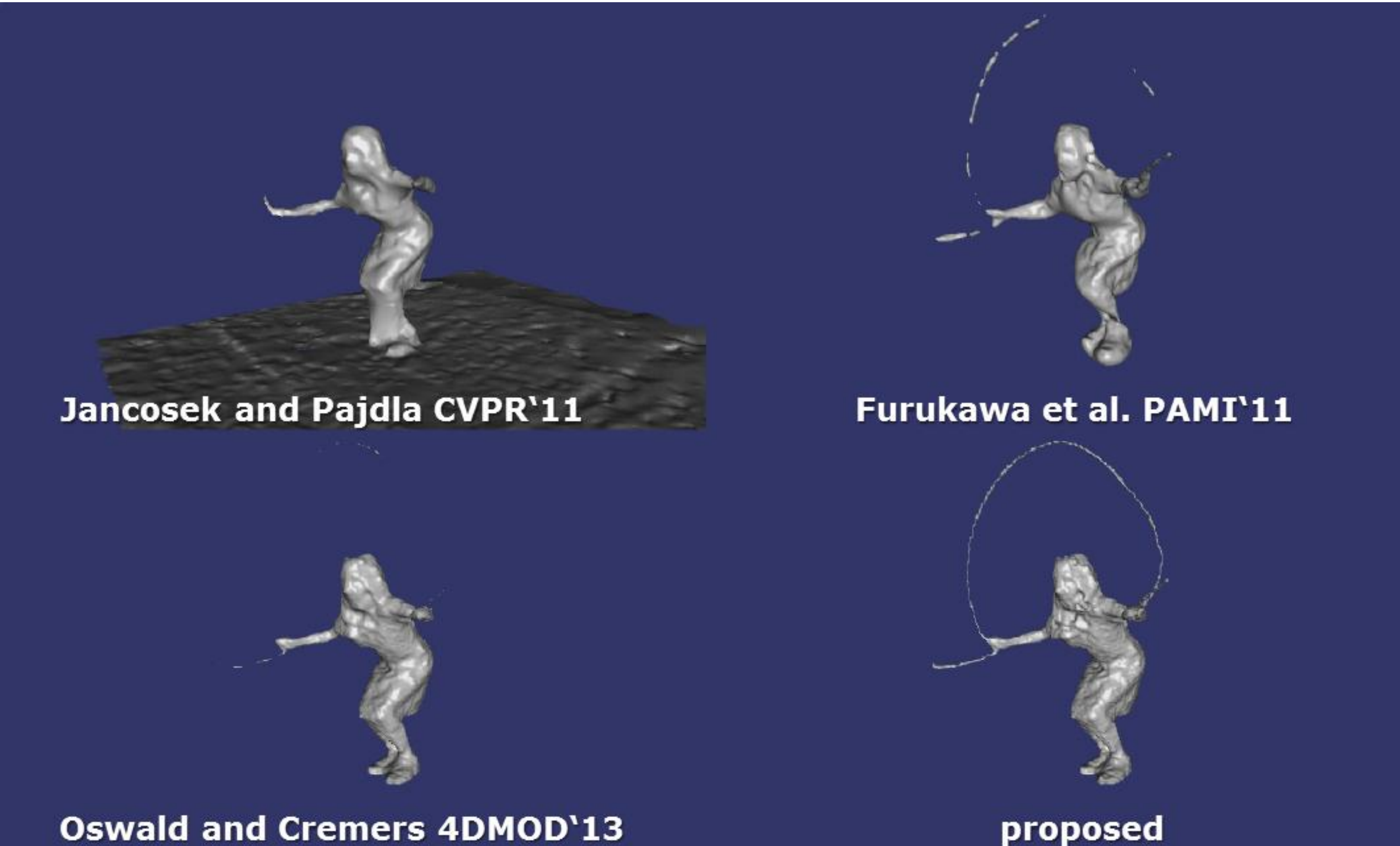


generalized connectivity

Connectivity Constraints - Motivation



Connectivity Constraints - Motivation



Connectivity Constraints

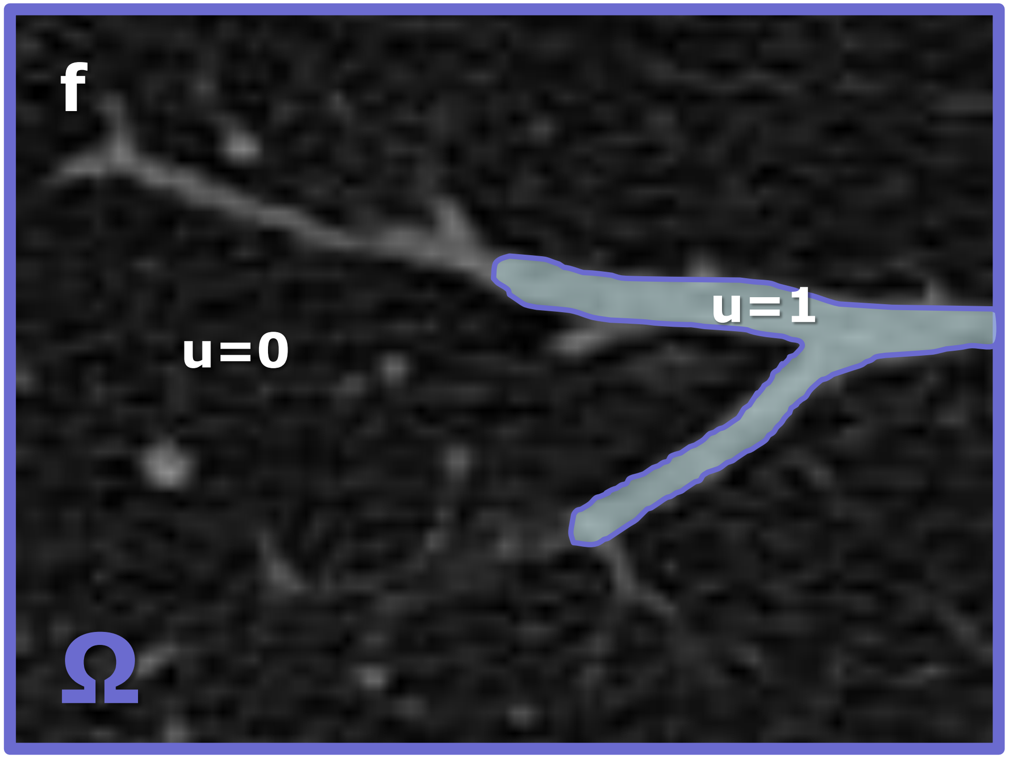
Image Segmentation

$f : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R} \qquad u : \Omega \mapsto \{0, 1\}$

$f < 0$ favor interior labeling

$f = 0$ neutral – no preference

$f > 0$ favor exterior labeling

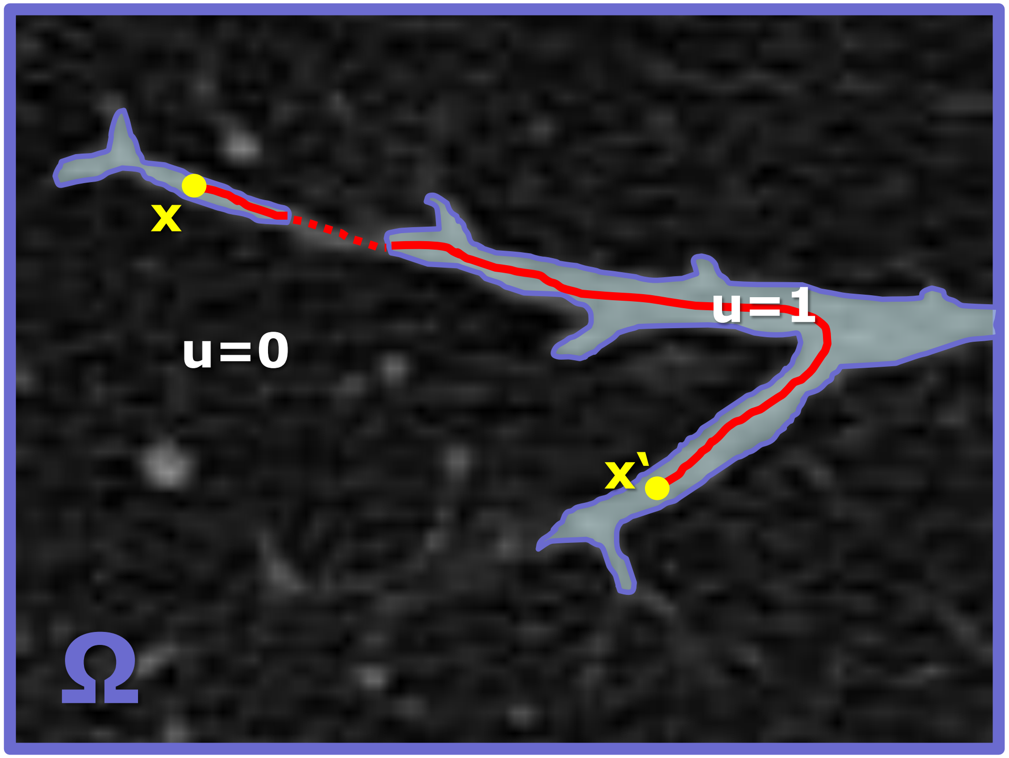


$$\min_u \int_\Omega |\nabla u| \, dx + \lambda \int_\Omega f u \, dx$$

Connectivity Constraints

Image Segmentation

$$f : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R} \qquad u : \Omega \mapsto \{0,1\}$$



$$\min_u \int_\Omega |\nabla u| \, dx + \lambda \int_\Omega f u \, dx$$

s.t.

$$\forall x, x' \in \Omega_{u=1} : \exists C_x^{x'} \subset \Omega_{u=1}$$

Theorem (Vicente et al. 2008)**:**
This problem is NP hard.

Connectivity Constraints

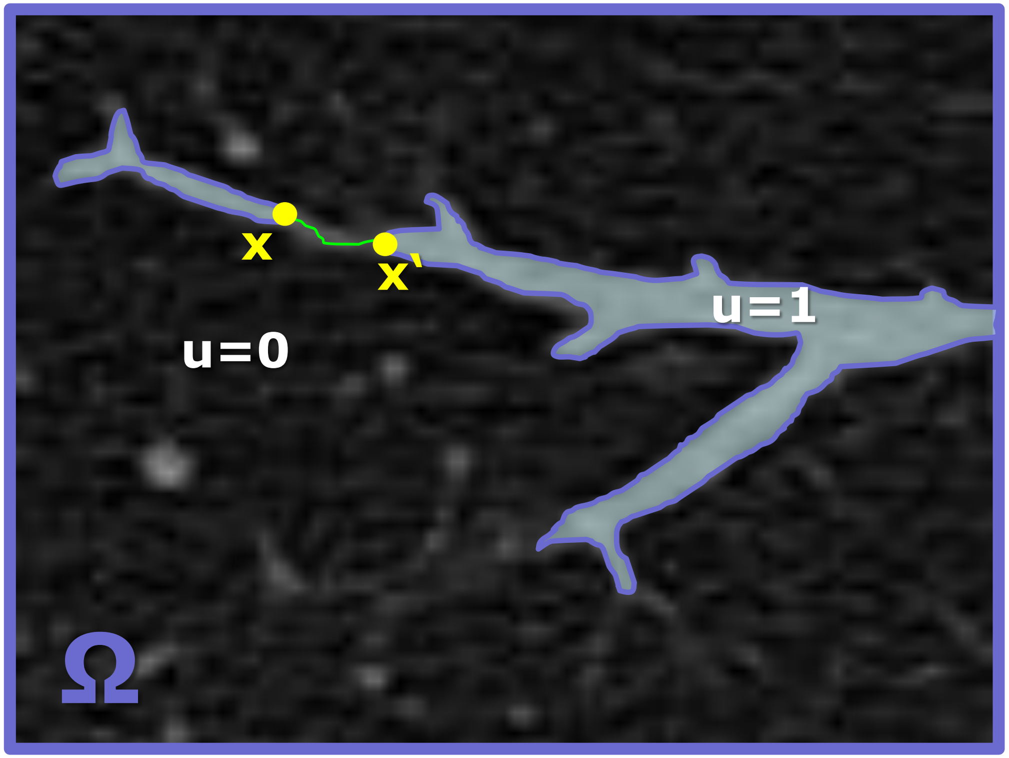
Image Segmentation

$f : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R} \qquad u : \Omega \mapsto \{0,1\}$

shortest geodesic path wrt. f

$$\min_{C_x^{x'}} \ell(C_x^{x'})$$

$$\ell(C_x^{x'}) = \int_0^1 f^+(C_x^{x'}(r))dr$$



$$\min_u \int_\Omega |\nabla u|\, dx + \lambda \int_\Omega f u\, dx$$

s.t.

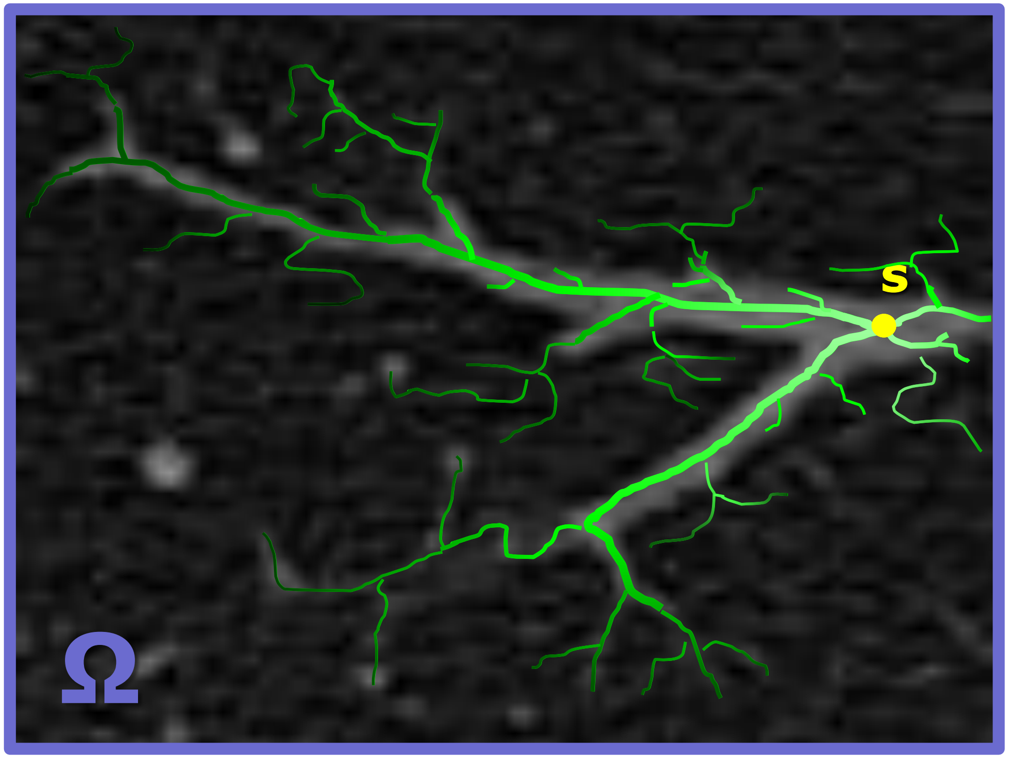
$$\forall x, x' \in \Omega_{u=1} : \exists C_x^{x'} \subset \Omega_{u=1}$$

Connectivity Constraints

Image Segmentation

$f : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R} \qquad u : \Omega \mapsto \{0,1\}$

$\mathcal{G}_s = (\mathcal{V}, \mathcal{E}) \quad s \in \Omega$



$$\min_u \int_\Omega |\nabla u| \, dx + \lambda \int_\Omega f u \, dx$$

s.t.

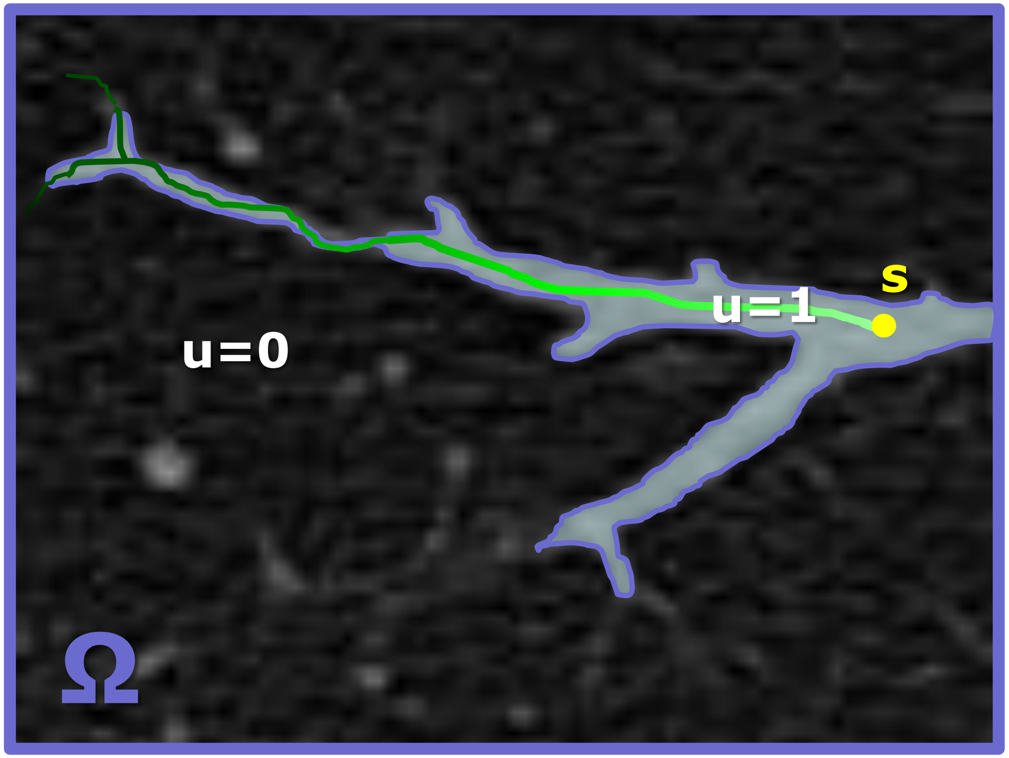
$$\nabla_e u \leq 0 \quad \forall e \in \mathcal{E}$$

Connectivity Constraints

Image Segmentation

$f : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R} \qquad u : \Omega \mapsto \{0, 1\}$

$\mathcal{G}_s = (\mathcal{V}, \mathcal{E}) \quad s \in \Omega$



$$\min_u \int_\Omega |\nabla u| \, dx + \lambda \int_\Omega f u \, dx$$

s.t.

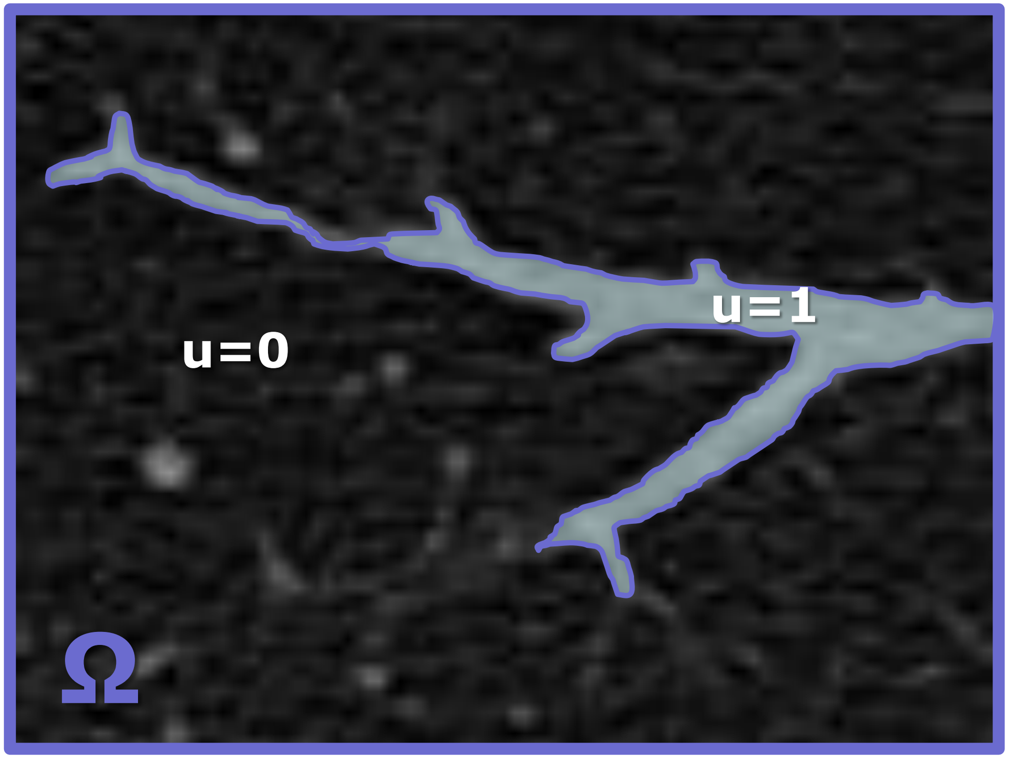
$$\nabla_e u \leq 0 \quad \forall e \in \mathcal{E}$$

Connectivity Constraints

Image Segmentation

$f : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R} \qquad u : \Omega \mapsto \{0, 1\}$

$\mathcal{G}_s = (\mathcal{V}, \mathcal{E}) \quad s \in \Omega$

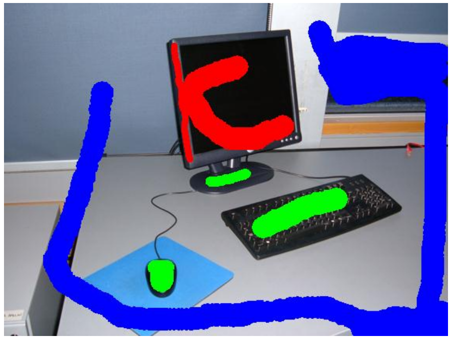


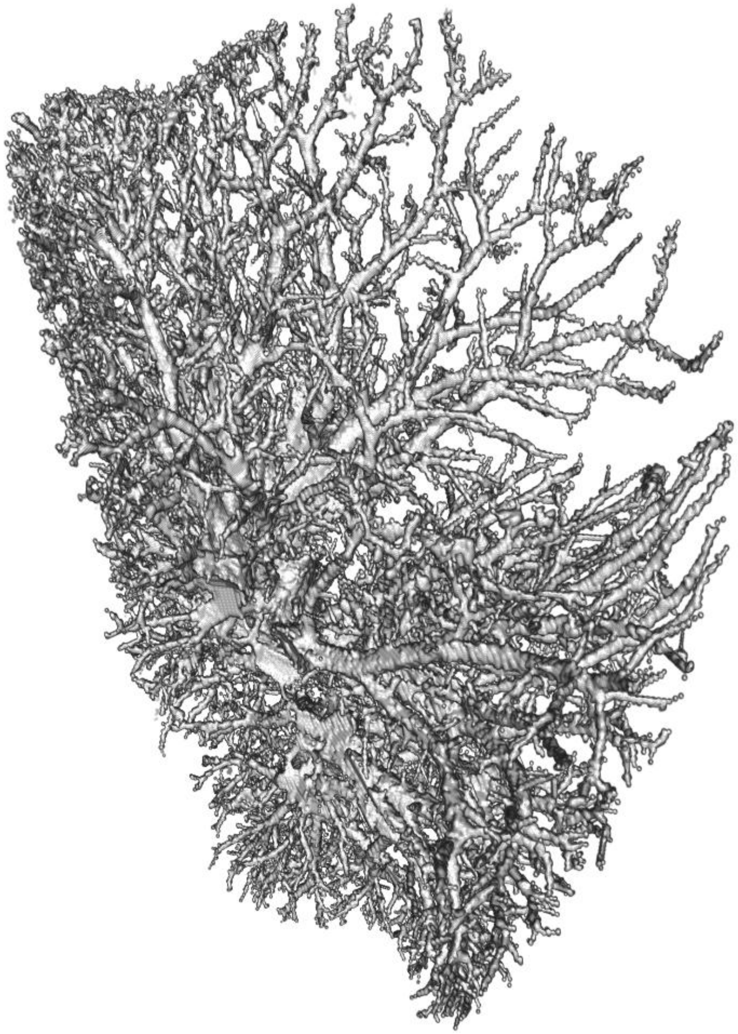
$$\min_u \int_\Omega |\nabla u| \, dx + \lambda \int_\Omega f u \, dx$$

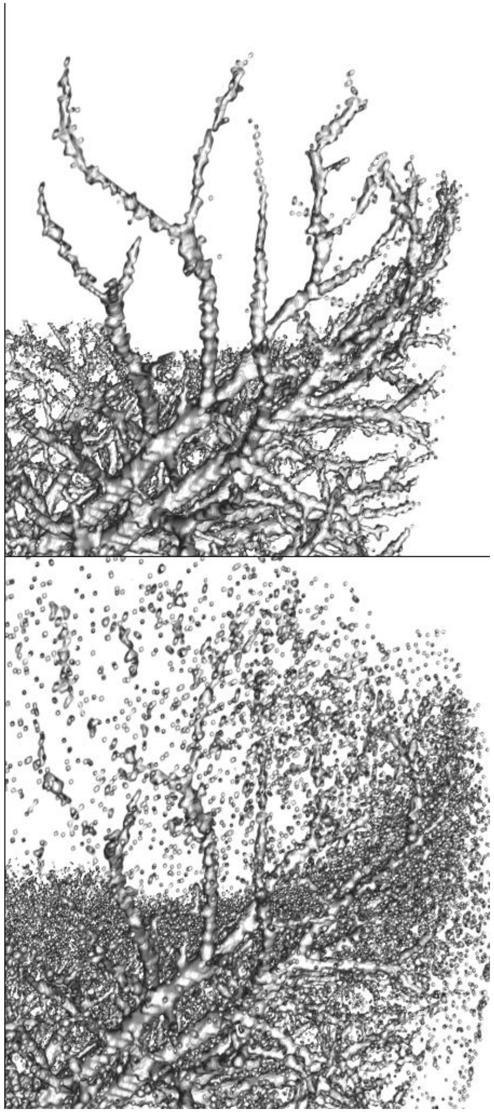
s.t.

$$\nabla_e u \leq 0 \quad \forall e \in \mathcal{E}$$

Connectivity for Image Segmentation







Stühmer, Schröder, Cremers, ICCV 2013

Summary

Spatio-temporal multi-view 3D reconstruction

- semi-local data term performs better in wide-baseline setups
- temporal regularization reduces surface jitterering
- faster photoconsistency computation
- efficient parallel implementation

2-3min/frame
competitive computation time

Generalized Connectivity Constraints

- integrated connectivity constraints into spatio-temporal reconstruction
- guaranteed loop preservation
- generalized connectivity constraints that respect the data term
- low computational time overhead additionally ~1-2min/frame