MVA2019

Consolidating Segmentwise Non-Rigid Structure from Motion

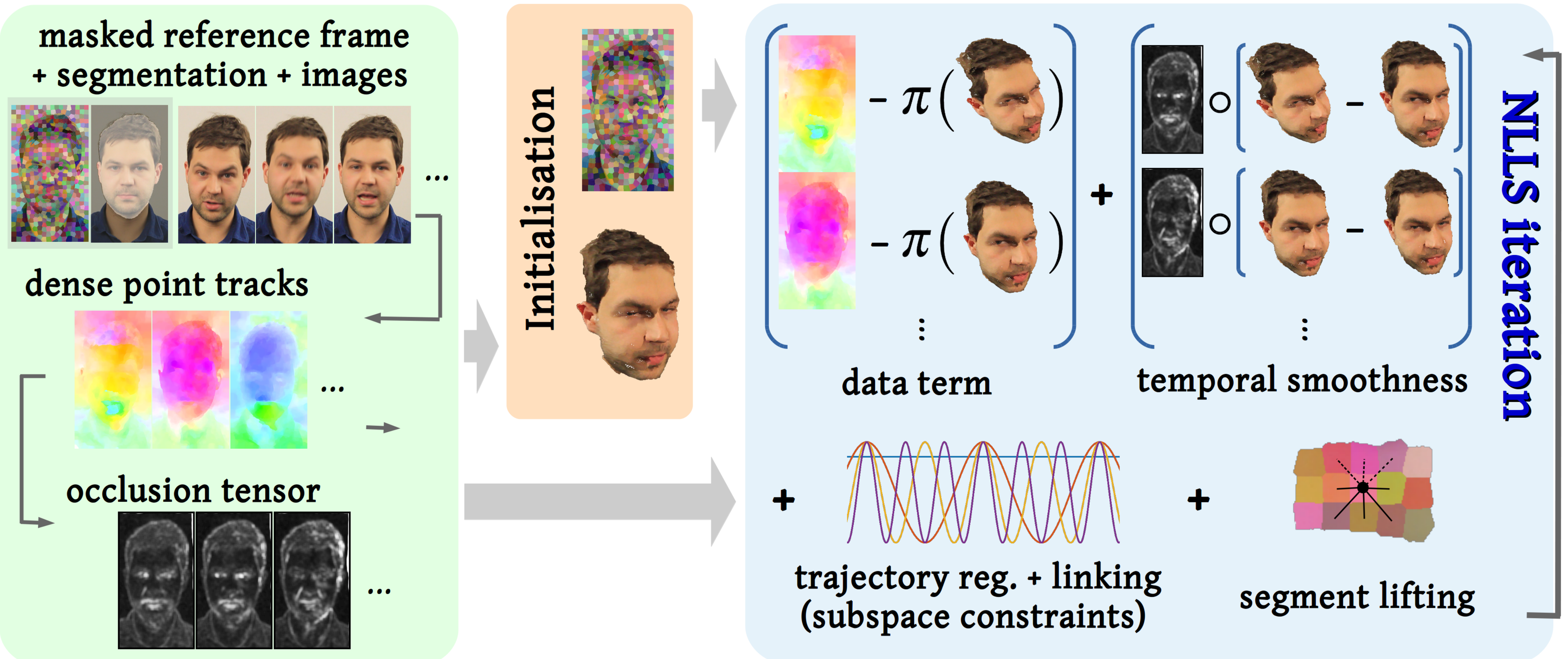
Vladislav Golyanik [1, 2] André Jonas [2] Didier Stricker [2, 3]

1 max planck institut informatik 2 TECHNISCHE UNIVERSITÄT KAISERSLAUTERN 3 DFKI

Abstract

We introduce a new **segmentwise technique** which **consolidates multiple principles for NRSfM** into **a single energy-based framework**. The energy functional of our **CMDR approach** is optimised by **NLLS** and includes terms allowing to define the deformation model and additional constraints **simultaneously in the metric and trajectory spaces**. CMDR achieves high accuracy on several tested sequences while providing **robustness** and **scalability** due to the **spatial scene segmentation** and the **new lifted spatial Laplacian term**.

Method Overview



Energy Functional and Optimisation

Segmentwise data term: (# of residual blocks: FP)

$$\mathbf{E}_{\text{data}}(\mathbf{R},\mathbf{T}) = \sum_{f=1}^{F} \left\| \mathbf{W}_f - \pi\Big(\mathbf{R}_f\, [g(\mathbf{T}_1^f,\mathbf{S}_1)\,\dots\, g(\mathbf{T}_L^f,\mathbf{S}_L)]\Big) \right\|_\epsilon^2$$

+ Segment trajectory smoothness: ($(F-1)L$)

$$\mathbf{E}_{\text{temp}}(\mathbf{T}) = \sum_{f=2}^{F}\sum_{l=1}^{L} \left\| \mathbf{\Phi}_f^l \circ (\mathbf{T}_l^f - \mathbf{T}_l^{f-1}) \right\|_\epsilon^2$$

+ Subspace constraints on segment trajectories: (FL)

$$\mathbf{E}_{\text{linking}}(\mathbf{S},\mathbf{A}) = \left\| \mathbf{\Psi} - (\Theta \otimes \mathbf{I}_3)_{3F\times 3K}\, \mathbf{A}_{3K\times L} \right\|_\epsilon^2$$

$$\text{where } \mathbf{\Psi} = \begin{bmatrix} g(\mathbf{T}_1^1,\mathbf{S}_1) & g(\mathbf{T}_2^1,\mathbf{S}_2) & \dots & g(\mathbf{T}_L^1,\mathbf{S}_L) \\ g(\mathbf{T}_1^2,\mathbf{S}_1) & g(\mathbf{T}_2^2,\mathbf{S}_2) & \dots & g(\mathbf{T}_L^2,\mathbf{S}_L) \\ & & \vdots & \\ g(\mathbf{T}_1^F,\mathbf{S}_1) & g(\mathbf{T}_2^F,\mathbf{S}_2) & \dots & g(\mathbf{T}_L^F,\mathbf{S}_L) \end{bmatrix} \text{ and } \Theta = \begin{pmatrix} \theta_{11} & \dots & \theta_{1K} \\ \vdots & \ddots & \vdots \\ \theta_{F1} & \dots & \theta_{FK} \end{pmatrix}$$

+ Trajectory coefficient regularisation: ($8L$)

$$\mathbf{E}_{\text{reg.}}(\mathbf{A}) = \sum_{l=1}^{L}\sum_{k=1}^{K} \mathbf{w}(l)\, \left\| \nabla \mathbf{A}_{k,l} \right\|_\epsilon^2$$

+ Segment coherency regulariser (segment lifting for detection of topological changes): ($8FL$)

$$\mathbf{E}_{\text{lifting}}(\mathbf{T},\mathbf{w}) = \sum_{f=1}^{F}\sum_{\forall w_{j,h}} \left(\zeta_1 \left\| w_{j,h}^2 (\mathbf{T}_j - \mathbf{T}_k) \right\|_2^2 + \zeta_2 \left\| (1 - w_{j,h}^2) \right\|_2^2 \right)$$

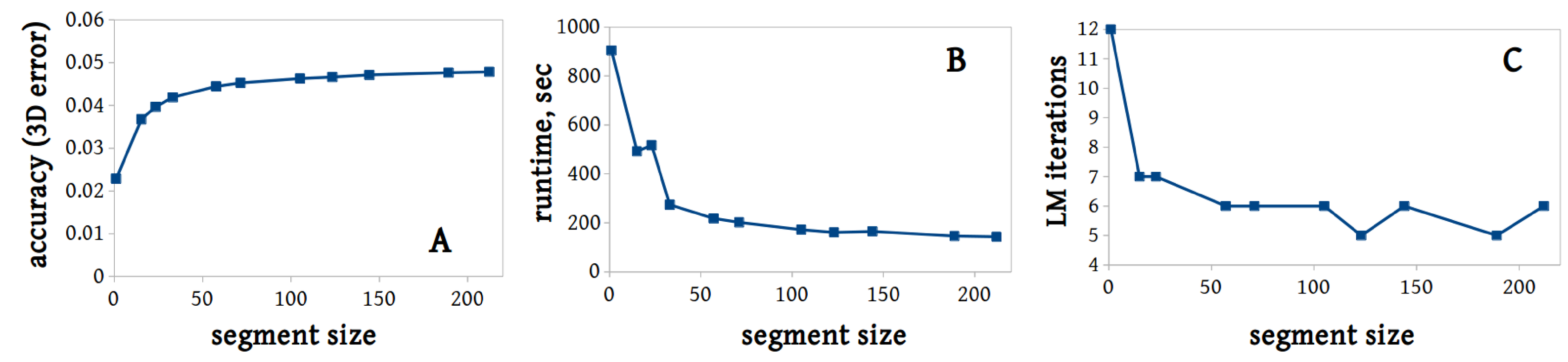
= **The energy functional of CMDR** $M = FP + L(10F + 7)$

$Q = 3F + 7LF + 3KL + 8L$ parameters:

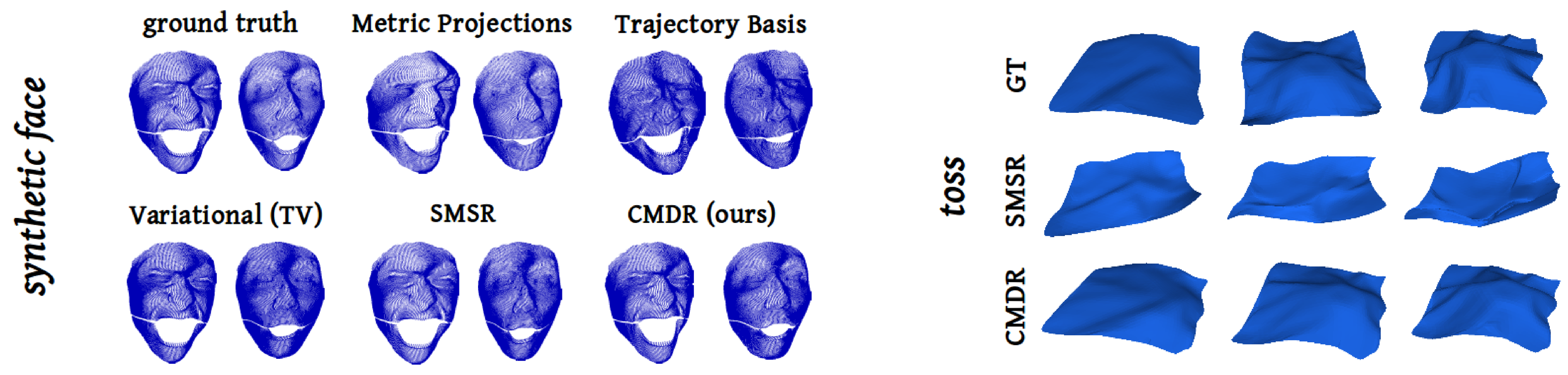
- $3F$ for global poses
- $7LF$ for segment orientations per every frame
- $3KL$ for DCT coefficients
- $8L$ for the lifting weights

$\mathbf{F}(\mathrm{x}) : \mathbb{R}^Q \rightarrow \mathbb{R}^M$ (optimised with Levenberg-Marquardt)

Experimental Results



3D error, runtime and number of Gauss-Newton solver iterations as the functions of the segment size, for the *actor mocap* sequence



	TB	MP	VA	DSTA	CDF	SMSR	GM	CMDR
Seq. 3	0.1252	0.0611	0.0346	0.0374	0.0886	**0.0304**	**0.0294**	**0.0324**
Seq. 4	0.1348	0.0762	0.0379	0.0428	0.0905	**0.0319**	**0.0309**	**0.0369**

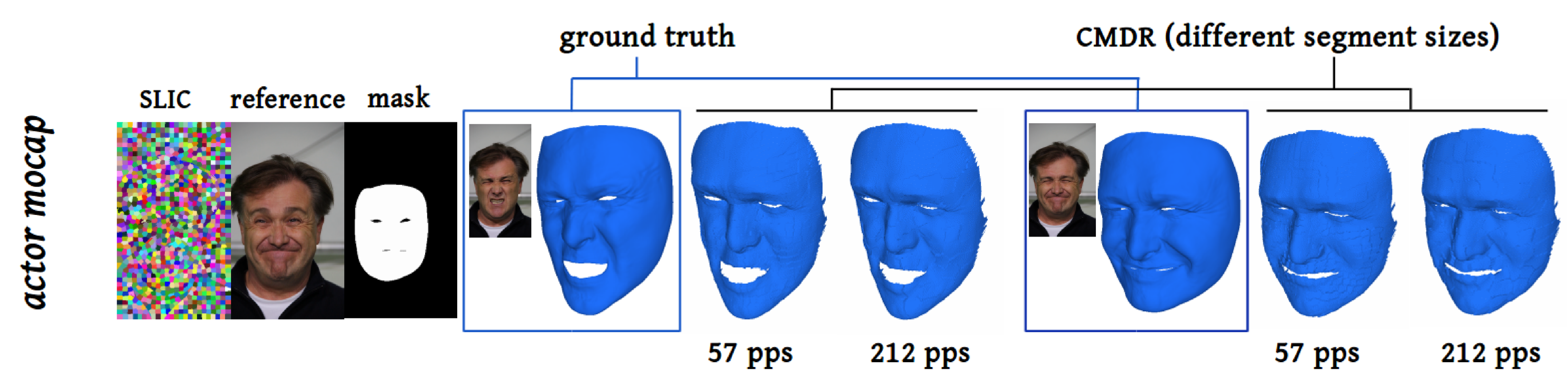
Comparison of methods on synthetic faces [3], observed by two different camera settings. CMDR achieves the third highest accuracy with a volatile difference to SMSR [6] and GM [7]

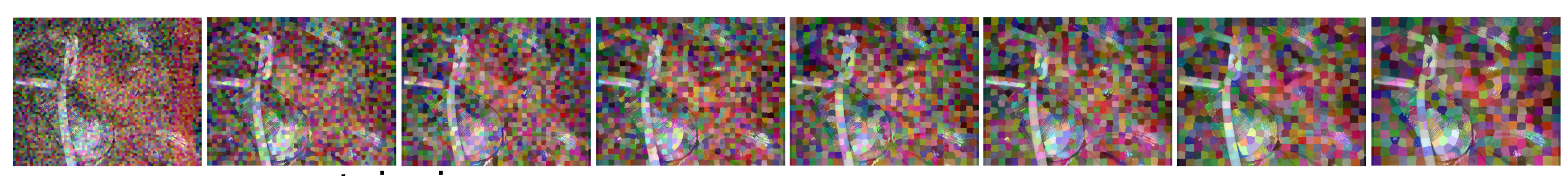
approach	*coin*	*toss*	*flag mocap*	*synth. flag*	*actor mocap*	*actor mocap #2*
SMSR [23]	0.2424	0.4003	0.196	0.1467	0.054	**0.0145**
CMDR	**0.0696**	**0.3064**	**0.0792**	**0.084**	**0.0257**	0.0228

SMSR [6] vs. CMDR (ours) on multiple sequences [3, 8, 9]

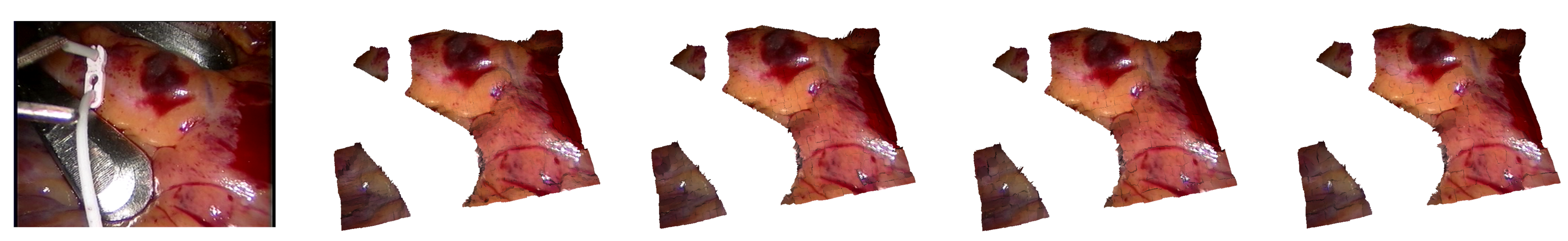
seq.	all terms	no $\mathbf{E}_{\text{reg.}}$	no $\mathbf{E}_{\text{lift.}}$	$\mathbf{E}_{\text{data}}$, $\mathbf{E}_{\text{temp.}}$ and $\mathbf{E}_{\text{lift.}}$	$\mathbf{E}_{\text{data}}$, $\mathbf{E}_{\text{temp.}}$
Seq. 3	0.0627	0.0616	0.0906	0.0622	0.0681
Seq. 4	0.0678	0.0678	0.0967	0.0682	0.0736

results of the ablation study (actop mocap)





~23 pps. segment size increases ~112 pps. ~211 pps.



reference frame ~23 pps. ~75 pps. ~112 pps. ~211 pps.

References

[1] Akhter *et al.*: *Trajectory space: A dual representation for nonrigid structure from motion*, 2011.
[2] Paladini *et al.*: *Optimal metric projections for deformable and articulated structure-from-motion*, 2012.
[3] Garg *et al.*: *Dense variational reconstruction of non-rigid surfaces from monocular video*, 2013.
[4] Dai *et al.*: *Dense non-rigid structure-from-motion made easy – a spatial-temporal smoothness based solution*, 2017.
[5] Golyanik *et al.*: *Introduction to coherent depth fields for dense monocular surface recovery*, 2017.
[6] Ansari *et al.*: *Scalable dense monocular surface reconstruction*, 2017.
[7] Kumar *et al.*: *Scalable dense non-rigid structure-from-motion: A grassmannian perspective*, 2018.
[8] White *et al.*: *Capturing and animating occluded cloth*, 2007.
[9] Valgaerts *et al.*: *Lightweight binocular facial performance capture under uncontrolled lighting*, 2012.