Intrinsic Dynamic Shape Prior for Dense Non-Rigid Structure from Motion

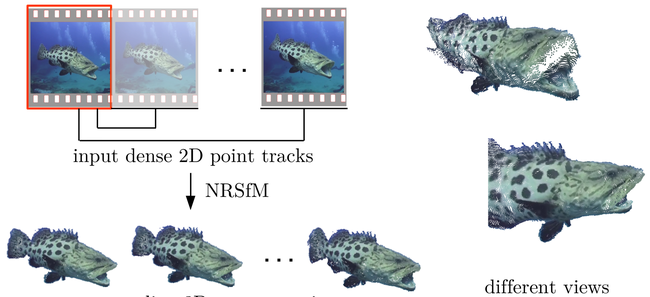
Vladislav Golyanik (max planck institut informatik) André Jonas (DFKI) Didier Stricker (DFKI) Christian Theobalt (max planck institut informatik)

3DV 2020

Overview/Contributions

i/ a new ***DSPR*** method for monocular non-rigid 3D reconstruction operating on point tracks and using a ***DSP*** (a set of states) as a prior
ii/ a new **dense NRSfM technique** for representative sequences **D-CMDR**
iii/ a new **synthetic dataset** with ground-truth 3D shapes, rendered images and several variants of optical flow
iv/ ***evaluation of both components*** on several synthetic and real datasets

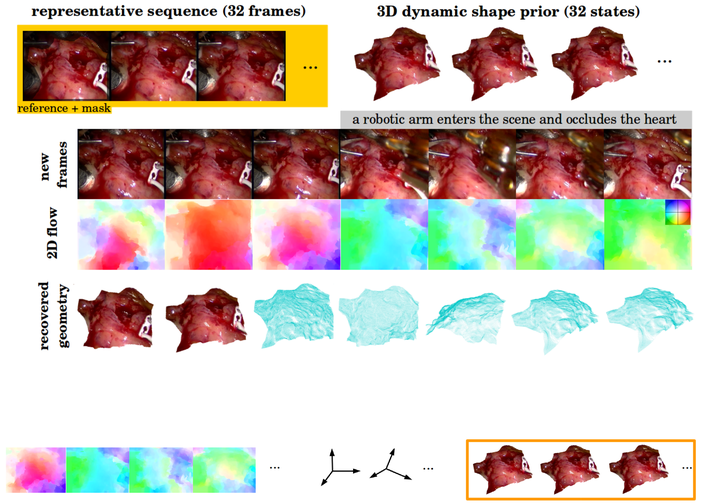
Introduction: Dense NRSfM



input frame — surfaces recovered by different NRSfM methods

results by Sidhu *et al.*, ECCV'20, applied to different objects (the follow-up to this work)

Our DSPR Approach



$$\mathbf{E}(\mathbf{S}_f = \mathbf{D}_i, \mathbf{R}_f) = \alpha \left\| \mathbf{W}_f - \mathbf{I}_{2\times 3}\mathbf{R}_f \mathbf{D}_{i:\lambda_i=1} \right\|_{\mathcal{F}} + \beta \left\| \mathbf{D}_{i:\lambda_i=1} - \mathbf{S}_{f-1} \right\|_{\mathcal{F}} + \gamma \left(\|\boldsymbol{\lambda}\|_0 - 1 \right)^2$$

α **data term** + β **shape similarity** + γ **DSP indicator**

Repeat until converged:

> Fix $\mathbf{R}_f$, release $\mathbf{D}_{i:\lambda_i=1} = \mathbf{S}_f$:
>
> k MSGD seeds (DSP initialisations)
> find k local minima
> find the best fit out of k
>
> $$\frac{\partial \mathbf{E}_{\mathbf{R}_f}(\mathbf{S})}{\partial \mathbf{S}} \frac{\partial \mathbf{S}}{\partial \boldsymbol{\lambda}} = 0$$
>
> Fix $\mathbf{S}_f$, release $\mathbf{R}_f$:
>
> affine update projected to the SO(3) group

Postprocessing of DSP:

+ build a map of pairs $\chi = (\|\mathbf{S}_l^\sharp\|_{\mathcal{F}}, \mathbf{S}_l^\sharp)$
+ arrange the shapes according to the increasing $\|\mathbf{S}_l^\sharp\|_{\mathcal{F}}$
+ prune similar shapes if the norm difference between the current $\mathbf{S}_l^\sharp$ and the latest included $\mathbf{D}_i$ exceeds μ

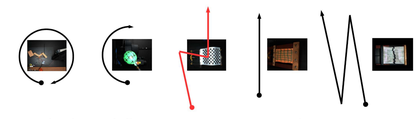
Experimental Results

Shape metric:
m. RMSE $e_{3D} = \frac{1}{F}\sum_f \frac{\|\mathbf{S}'_f - \mathbf{S}_f\|_{\mathcal{F}}}{\|\mathbf{S}'_f\|_{\mathcal{F}}}$

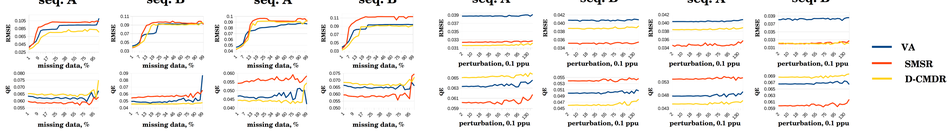
Camera pose metric:
m. QE $e_q = \frac{1}{F}\sum_f |\mathbf{q}'_f - \mathbf{q}_f|$ (quaternion norm)
corresponds to
$\min_{\mathbf{R}^z} \sum_f \left\| \mathbf{R}'_f - \mathbf{R}^z \mathbf{R}_f \right\|_\epsilon$ ($\epsilon = 1.0$)

TB[1]	MP[2]	VA[3]	DSTA[4]	CDF[5]	SMSR[6]	GM[7]	JM[8]	D-CMDR
0.1252	0.0611	0.0346	0.0374	0.0886	0.0304	0.0294	0.280	0.0324
0.1348	0.0762	0.0379	0.0428	0.0905	0.0319	0.0309	0.327	0.0369

trajectory ↓ \ scene →	articulated	balloon	paper	stretch	tearing	*mean RMSE* (over camera trajectories)
circle	75.823	55.664	81.181	92.017	80.873	77.112
flyby	44.603	36.293	56.873	53.716	44.139	47.125
line	65.914	35.784	45.796	44.625	29.103	44.244
semi-circle	46.741	32.41	57.67	55.567	53.656	49.209
tricky	62.559	34.77	34.509	58.012	43.888	46.748
zigzag	37.544	31.298	37.017	41.101	36.525	36.697
mean RMSE (over sequence types)	55.53	37.703	51.486	57.506	48.031	**50.19**

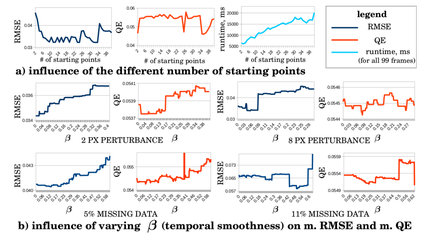


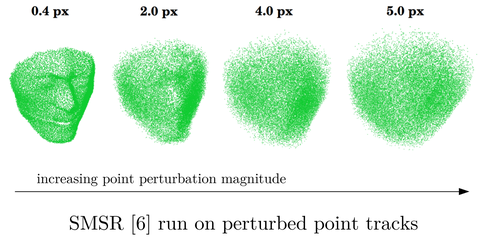
+ five sequences
+ six camera trajectories



METHOD		MISSING DATA 1%	3%	11%	17%	23%
SMSR	RMSE	0.1001	0.1778	0.3365	0.4143	0.4849
	QE	0.2972	0.2973	0.2968	0.2968	0.2975
D-CMDR (ours)	RMSE	0.1001	0.1777	0.3365	0.4143	0.4849
	QE	0.0663	0.0663	0.0662	0.0663	0.0663
DSPR (ours)	RMSE	**0.0327**	**0.03578**	**0.0754**	**0.0962**	**0.0994**
	QE	**0.0602**	**0.05984**	**0.0584**	**0.0585**	**0.0581**

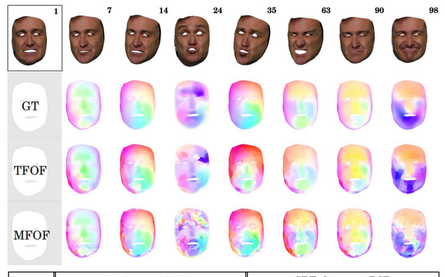
METHOD		PERTURBED DATA 0.4 px	1.2 px	1.6 px	2.0 px	3.0 px	4.0 px	5.0 px
SMSR	RMSE	0.0455	0.0962	0.1243	0.1536	0.2232	0.2956	0.3885
	QE	0.2434	0.2999	0.3287	0.2450	0.3068	0.2280	0.3510
D-CMDR (ours)	RMSE	0.0646	0.1918	0.2541	0.2867	0.3571	0.4056	0.4522
	QE	0.0689	0.1077	0.1514	0.1711	0.4617	0.4578	0.4506
DSPR (ours)	RMSE	**0.0324**	**0.0324**	**0.0324**	**0.0324**	**0.0324**	**0.0324**	**0.0325**
	QE	**0.0602**	**0.0601**	**0.0600**	**0.0601**	**0.0603**	**0.0600**	**0.0603**





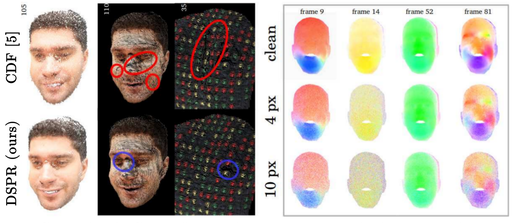
SMSR [6] run on perturbed point tracks

Actor Mocap: New Dataset



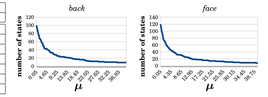
	GT shapes as DSP: GT flow	TFOF	MFOF	CDF shapes as DSP: GT flow	TFOF	MFOF
RMSE	0.00027	0.00042	0.00041	0.00525	0.00527	0.00527
QE	0.09136	0.09382	0.09271	0.12977	0.15017	0.13508

Applications in Shape Compression



C = 3 (top), C = 8 (bottom) — perturbed point tracks

μ	MSGD seeds	face DSP	face C	back DSP	back C
1.5	▲20	▲73	1.64	▲67	2.23
2.5	20	59	2.03	57	2.63
5	20	42	2.85	40	3.75
10.0	12	25	4.8	25	6
20.0	8	15	8	16	9.375
30.0	5	10	12	11	13.63
▼40.0	4	8	▼15	9	▼16.6



Acknowledgements

ERC Consolidator Grant 4DReply (770784)
Project VIDETE (01IW18002)
Adapted by permission from Springer Nature Customer Service Centre GmbH (license number 4926380741574)

References

[1] I. Akhter *et al.* Trajectory space: A dual representation for nonrigid structure from motion. TPAMI, 2011.
[2] M. Paladini *et al.* Optimal metric projections for deformable and articulated structure-from-motion. IJCV, 2012.
[3] R. Garg *et al.* Dense variational reconstruction of non-rigid surfaces from monocular video. In *CVPR*, 2013.
[4] Y. Dai *et al.* Dense non-rigid structure-from-motion made easy – a spatial-temporal smoothness based solution. In *ICIP*, 2017.
[5] V. Golyanik *et al.* Introduction to Coherent Depth Fields for Dense Monocular Surface Recovery. In *BMVC*, 2017.
[6] M. D. Ansari *et al.* Scalable Dense Monocular Surface Reconstruction. In *3DV*, 2017.
[7] S. Kumar *et al.* Scalable Dense Non-rigid Structure-from-Motion: A Grassmannian Perspective. In *CVPR*, 2018.
[8] S. Kumar. Jumping Manifolds: Geometry Aware Dense Non-Rigid Structure from Motion. In *CVPR*, 2019.