Learning non-rigid surface reconstruction from spatio-temporal image patches

Matteo Pedone, Abdelrahman Mostafa and Janne Heikkilä

Center for Machine Vision Research and Signal Analysis, University of Oulu, Finland

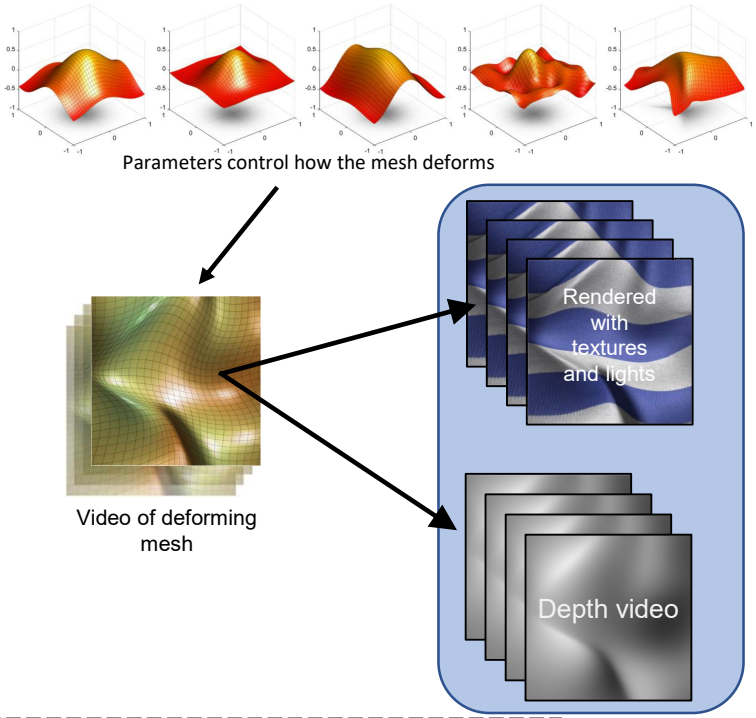
Abstract. *We present a method to reconstruct depth videos of non-rigidly deformable objects directly from a video sequence. The estimation of depth is performed locally on spatio-temporal patches of the video, and then the full depth video of the entire shape is recovered by combining them together. We artificially generate a database of small deforming rectangular meshes rendered with different material properties and light conditions, along with their corresponding depth videos and use such data to train a convolutional neural network based on the 3D U-Net architecture.*

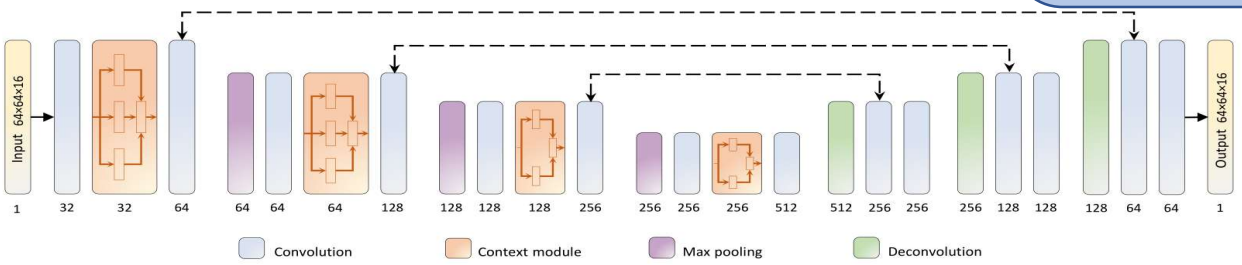
Motivation: Avoid point tracking and explicit non-rigid structure from motion (NRSfM).

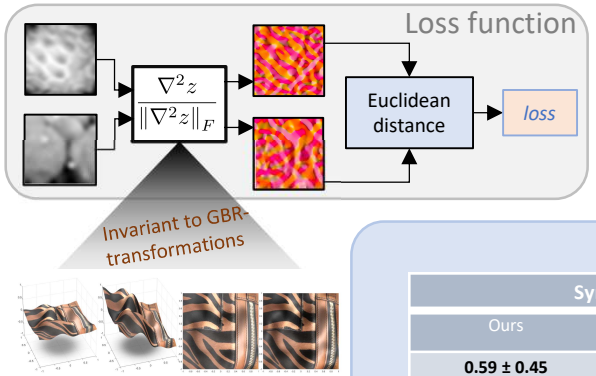
How: Train a network to infer shape directly from the video sequence:
Input = video sequence
Output = depth video

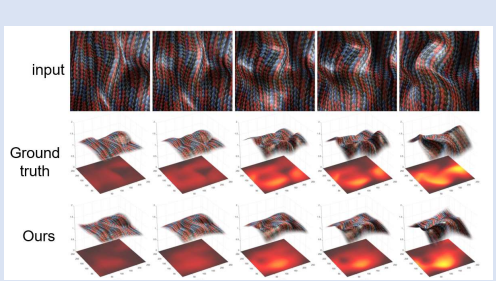
Problem: What is the training data? Too many factors to take into accont: type of objects, type/speed of motion, lighting conditions, etc…

Proposed Solution: Do it locally…in small neighborhoods deforming surfaces look less complex.









Synthetic data (1000 videos)		
Ours	CSF2	KSTA
0.59 ± 0.45	0.87 ± 0.63	0.87 ± 0.63

Avg. and std. of *spatially normalized MAE*

Kinect sequences		
Ours	CSF2	KSTA
3.7 mm	4.6 mm	4.3 mm

Average *MAE*