InteractionFusion: Real-time Reconstruction of Hand Poses and Deformable Objects in Hand-object Interactions

Hao Zhang, Zi-Hao Bo, Jun-Hai Yong, Feng Xu*

School of Software, Tsinghua University

清華大學
Tsinghua University

thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

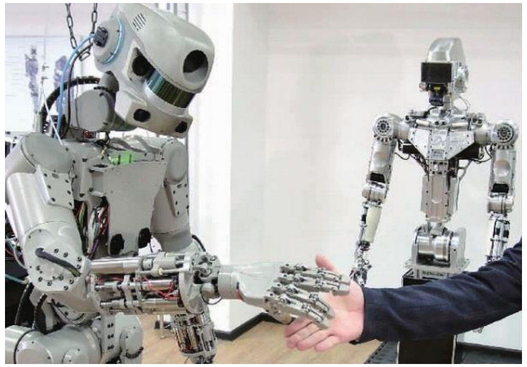
Outline

- Background
- Overview
- LSTM-based Pose Prediction
- Joint Hand-Object Motion Tracking
- Experiments & Results
- Limitations & Future Work
- Conclusion

Background

- ➢ Hand tracking has many applications

HCI



Robots

VR/AR

- ➢ Human hand often interacts with objects

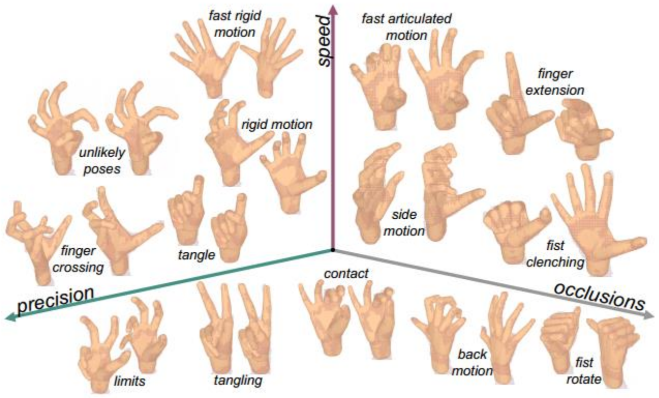
Hand-Object Interaction Reconstruction

清華大學
Tsinghua University

Background

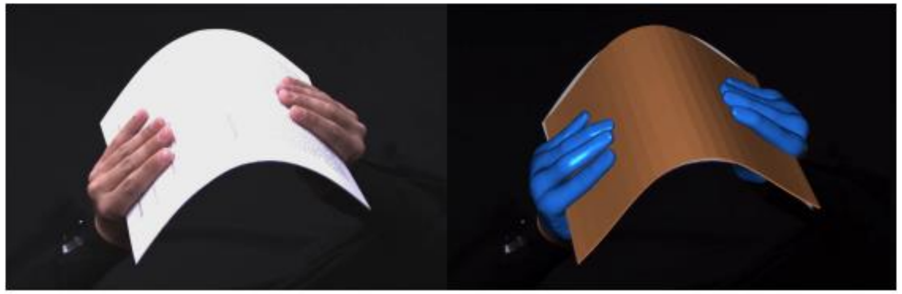
Challenges

- Isolated Hand Tracking
  - complex motions
  - lack of geometry/texture features
  - self-occlusion



[Tkach et al. 2016]

- Hand-Object Interaction
  - more occlusions in interactions
  - high dimensional solution space
  - physical plausibility

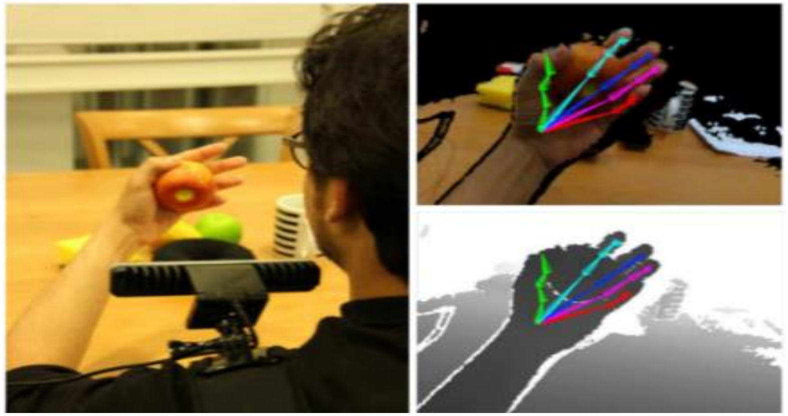


[Tzionas et al. 2016]

Background

- ➢ Hand tracking in interactions

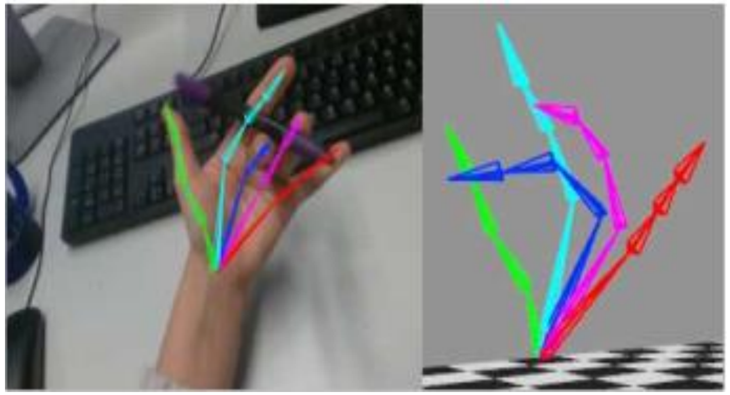
No Object In Output



[Mueller et al. 2017]

[Taylor et al. 2017]

[Simon et al. 2017]

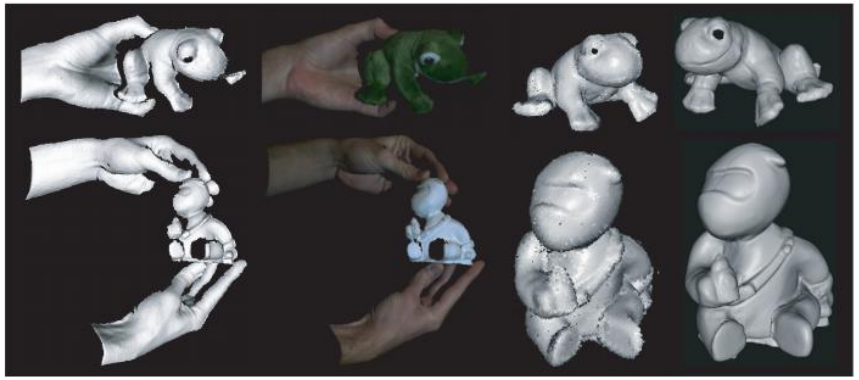


[Mueller et al. 2018]

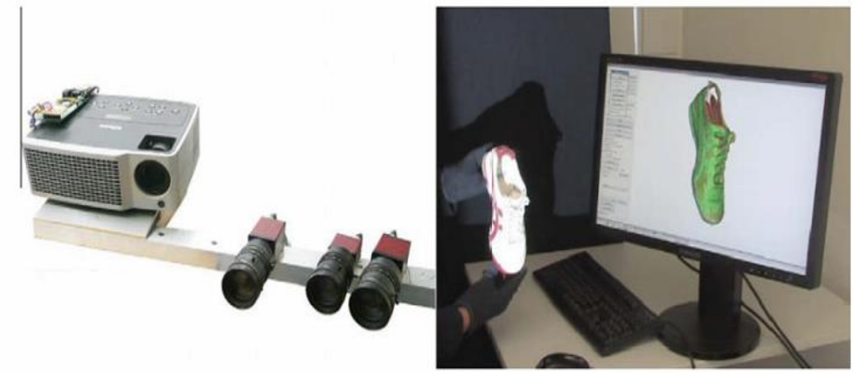
Background

- In hand reconstruction

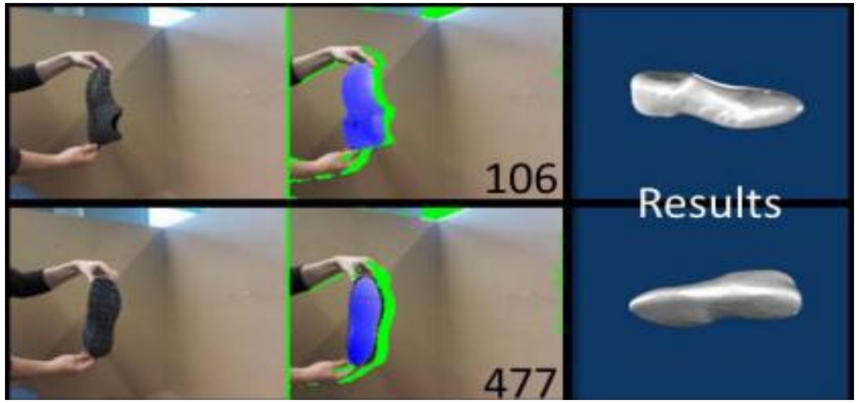
No Hand In Output



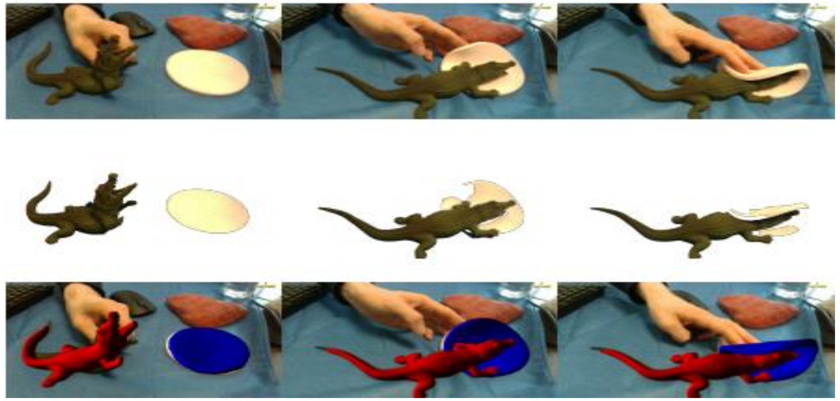
[Weise et al. 2008]



[Weise et al. 2011]



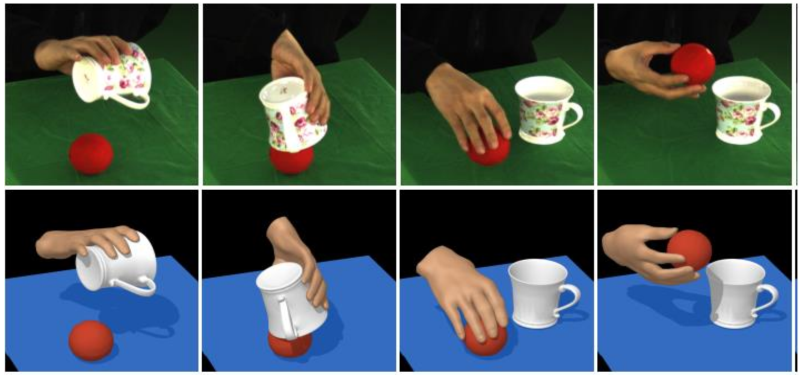
[Yuheng Ren et al. 2013]



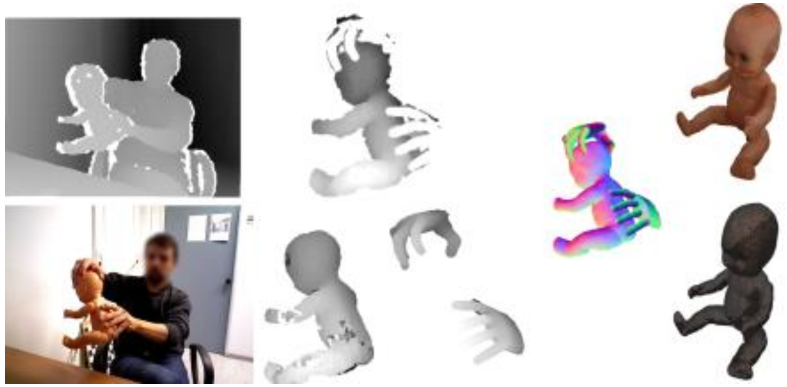
[Petit et al. 2018]

Background

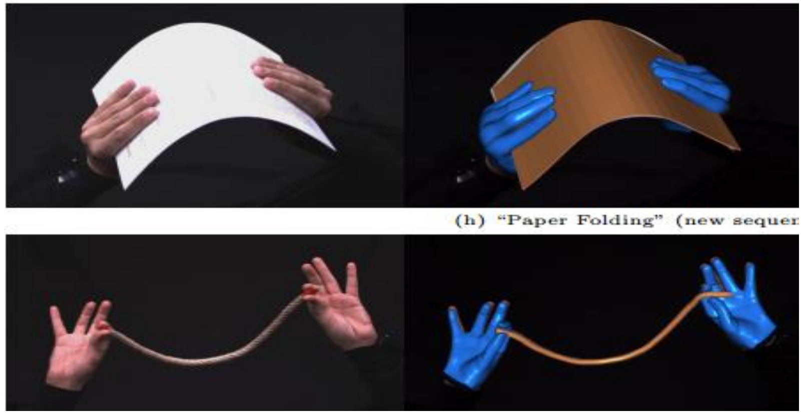
- ➢ Joint hand-object reconstruction Require initial template Rigid object



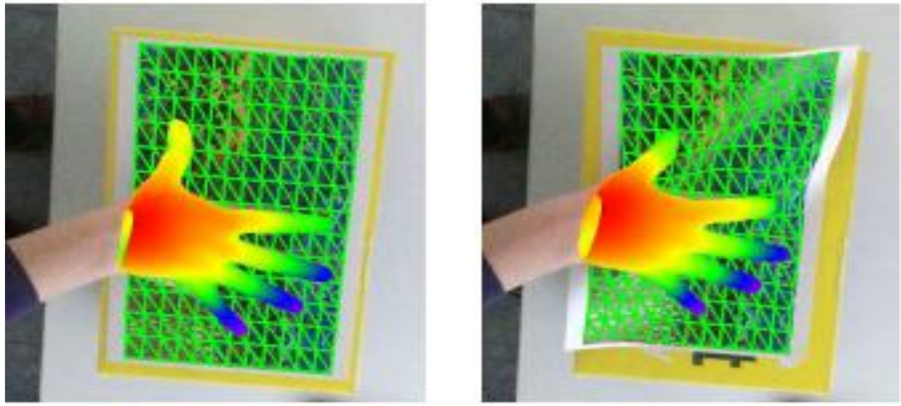
[Wang et al. 2013]



[Panteleris et al. 2015]



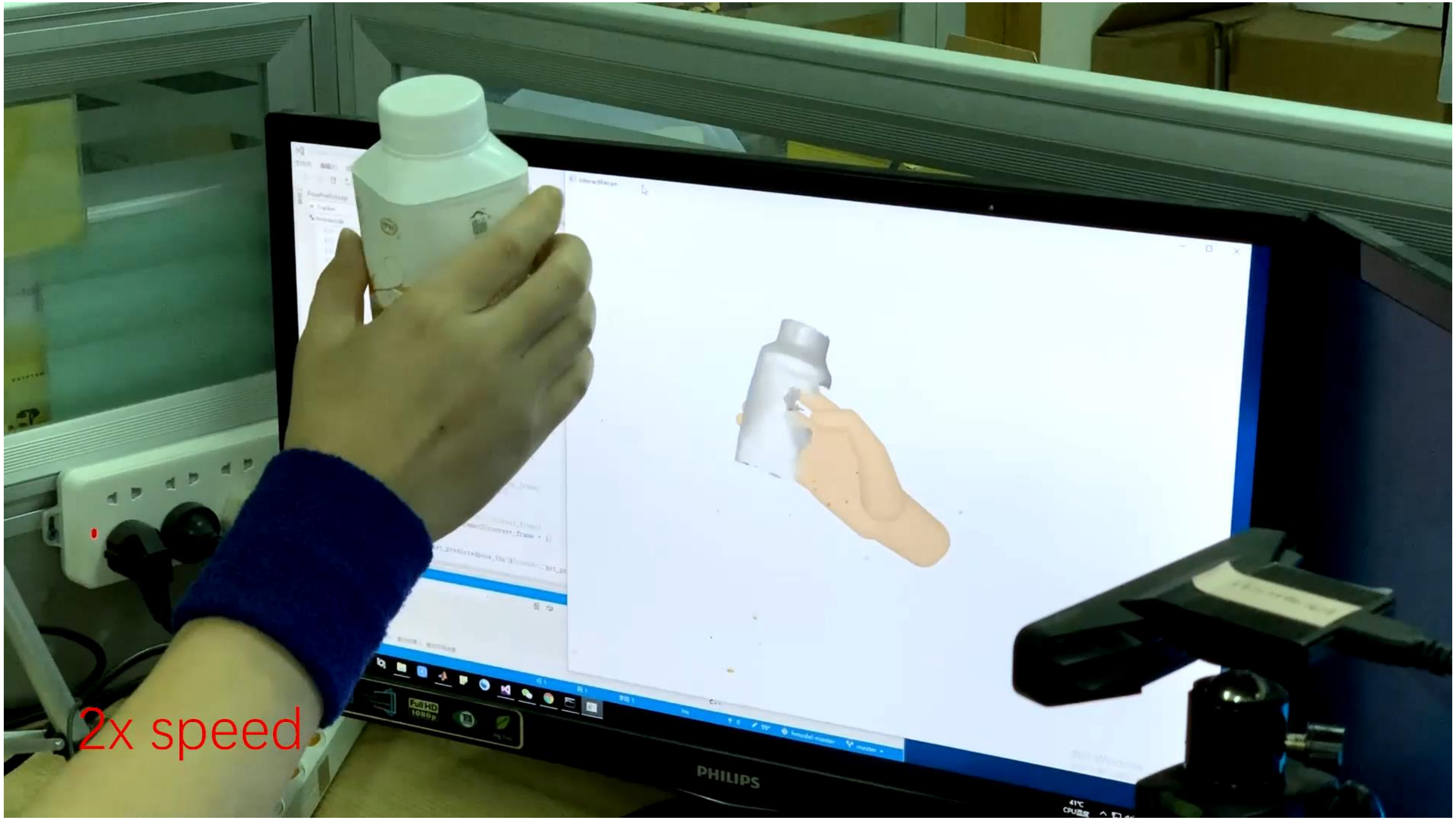
[Tzionas et al. 2016]



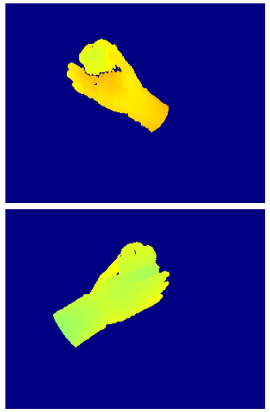
[Tsoli et al. 2018]

Our Work

- Reconstruct hand pose, object model and deformation in real-time

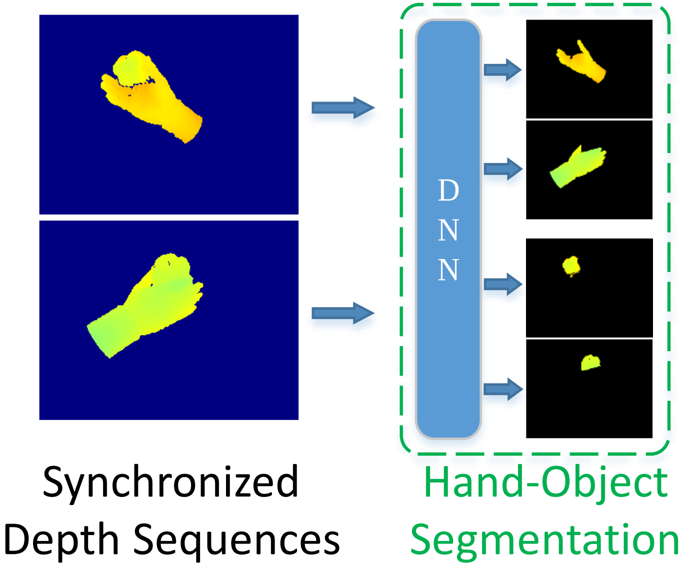


Overview

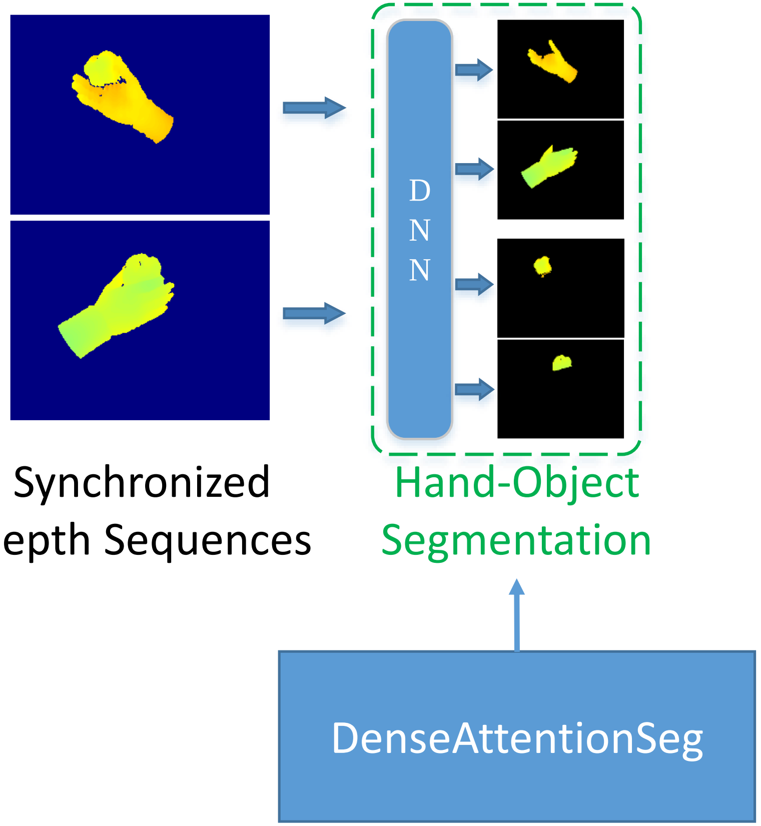


Synchronized
Depth Sequences

Overview


D
N
N
Synchronized
Depth Sequences
Hand-Object
Segmentation

Overview



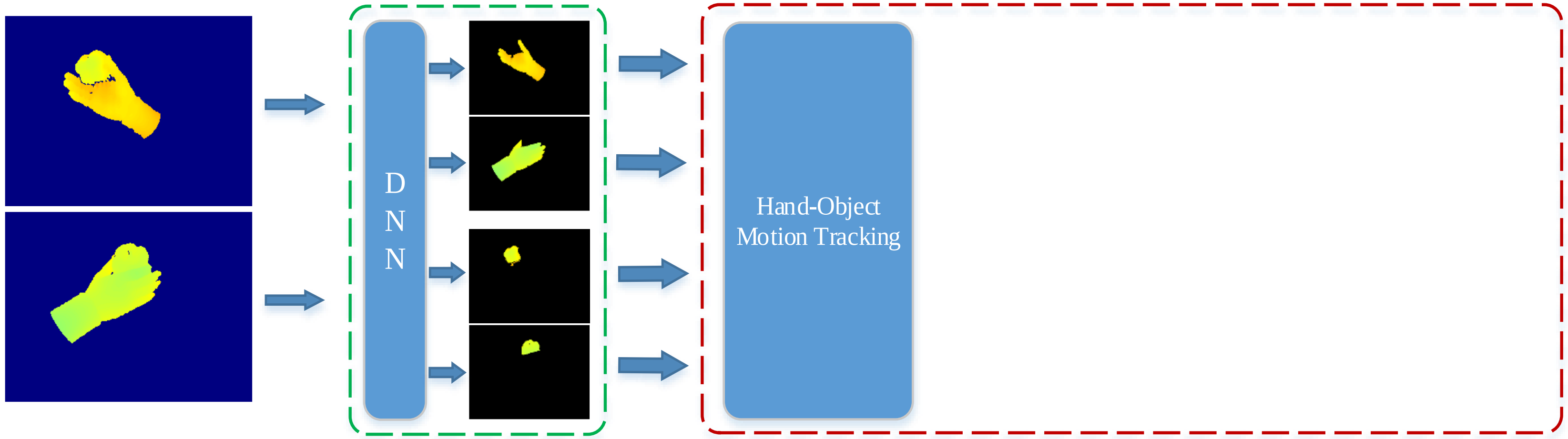
Synchronized Depth Sequences

Hand-Object Segmentation

DenseAttentionSeg

DenseAttentionSeg: Segment Hands from Interacted Objects Using Depth Input. arXiv preprint arXiv:1903.12368 (2019)

Overview

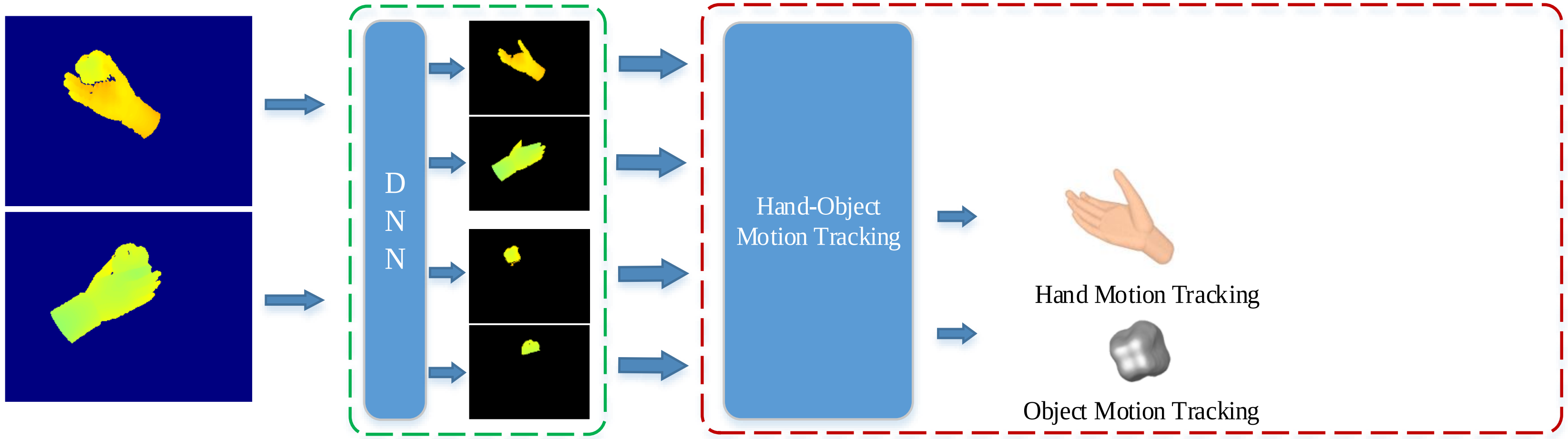


Synchronized
Depth Sequences

Hand-Object
Segmentation

Joint Hand-Object Motion Tracking and
Model Fusion

Overview

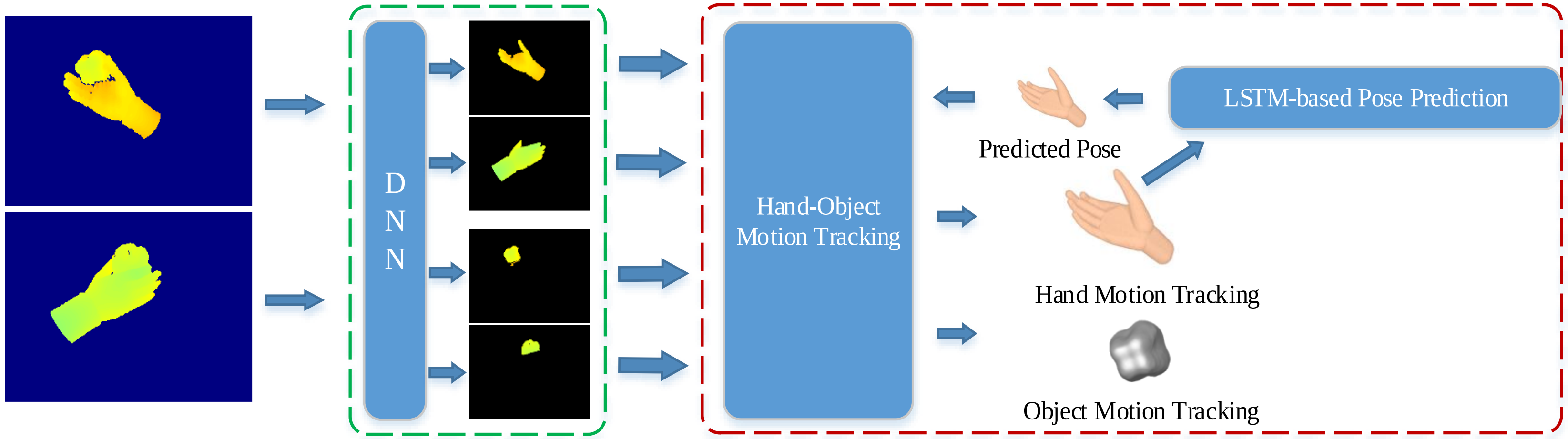


Synchronized
Depth Sequences

Hand-Object
Segmentation

Joint Hand-Object Motion Tracking and
Model Fusion

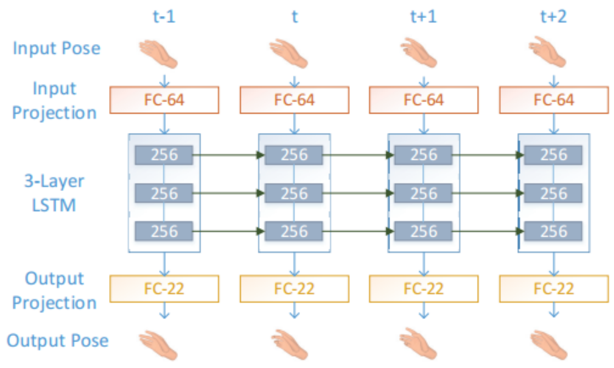
Overview



Synchronized Depth Sequences

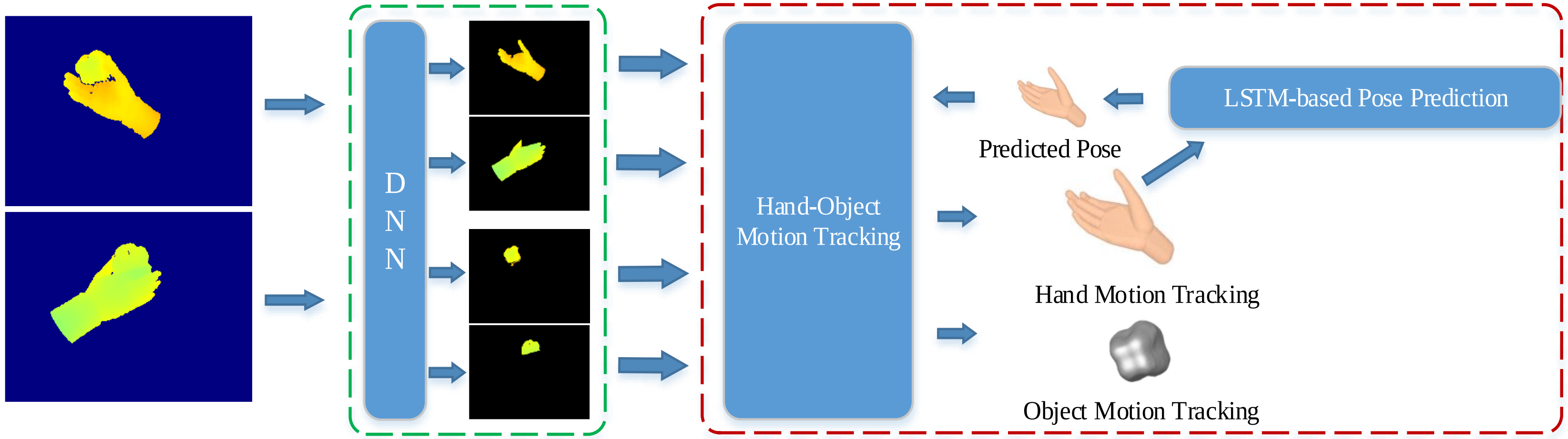
Hand-Object Segmentation

Joint Hand-Object Motion Tracking and Model Fusion



LSTM Model

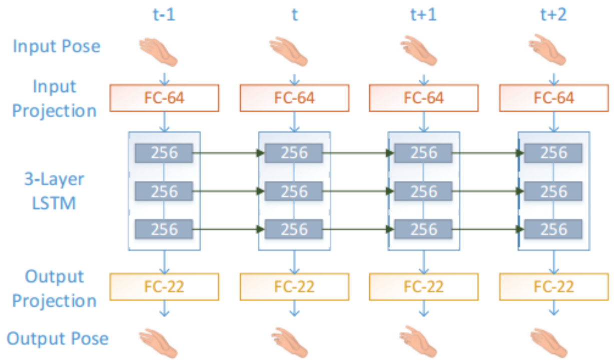
Overview



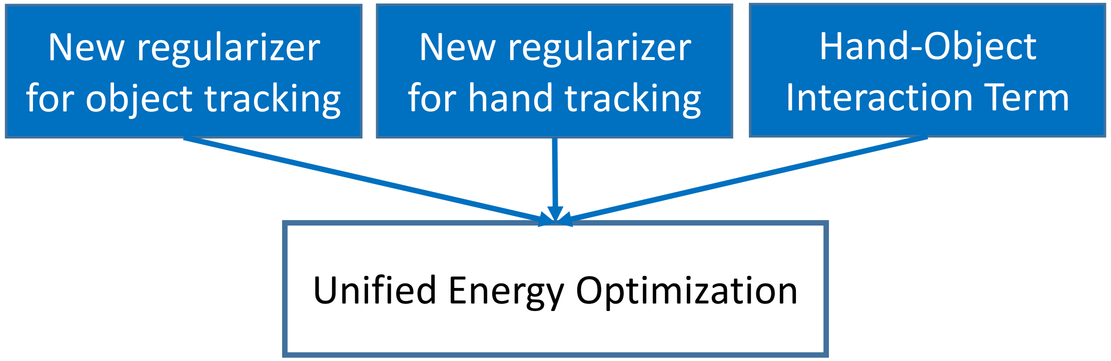
Synchronized Depth Sequences

Hand-Object Segmentation

Joint Hand-Object Motion Tracking and Model Fusion



LSTM Model



New regularizer for object tracking

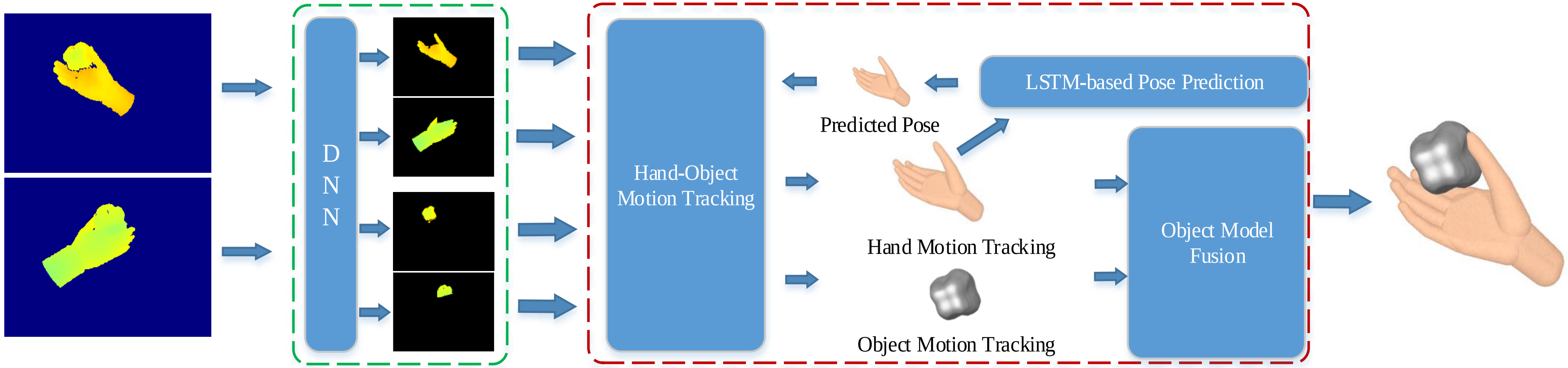
New regularizer for hand tracking

Hand-Object Interaction Term

Unified Energy Optimization

Joint Hand-Object Motion Tracking

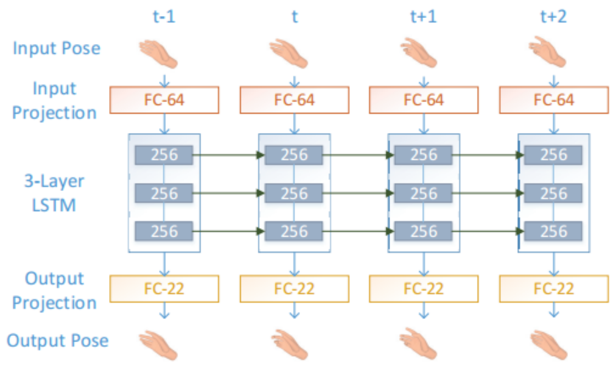
Overview



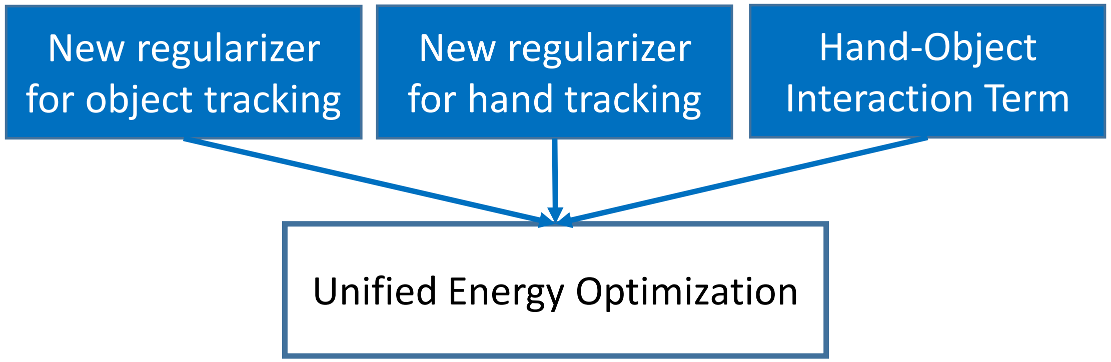
Synchronized Depth Sequences

Hand-Object Segmentation

Joint Hand-Object Motion Tracking and Model Fusion



LSTM Model


New regularizer for object tracking
New regularizer for hand tracking
Hand-Object Interaction Term
Unified Energy Optimization

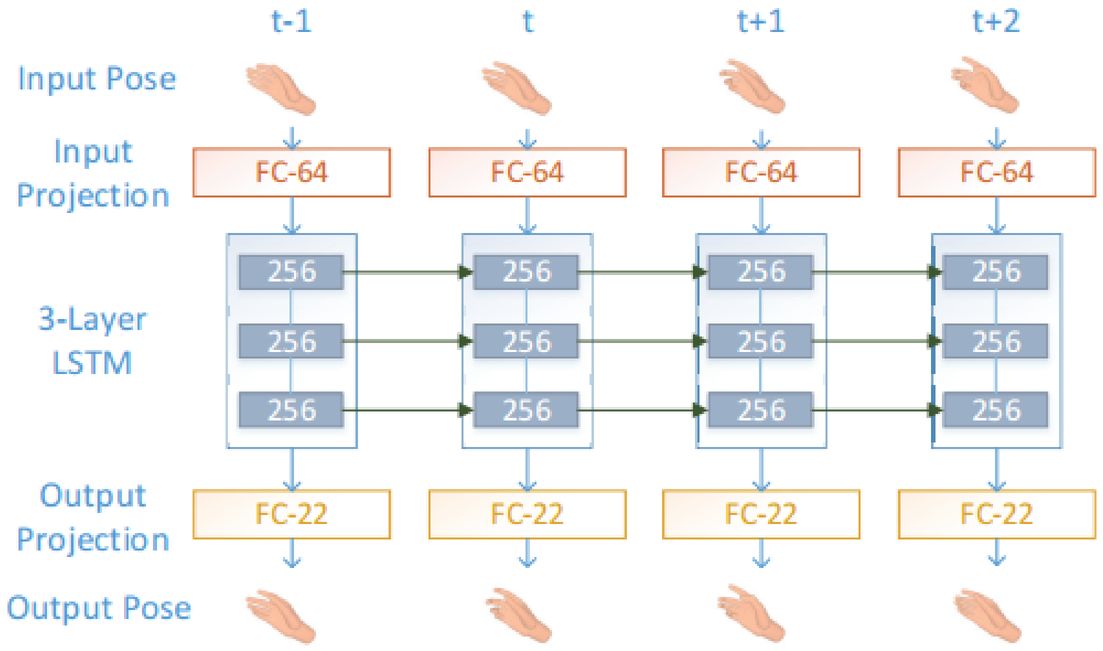
Joint Hand-Object Motion Tracking

LSTM-based Pose Prediction

Aim：

- Learning the hand motion pattern in interactions
- Improving the hand tracking accuracy in interactions

Structure



Input: 22 DoFs of Hand Pose

Output: 22 DoFs of Hand Pose

LSTM-based Pose Prediction

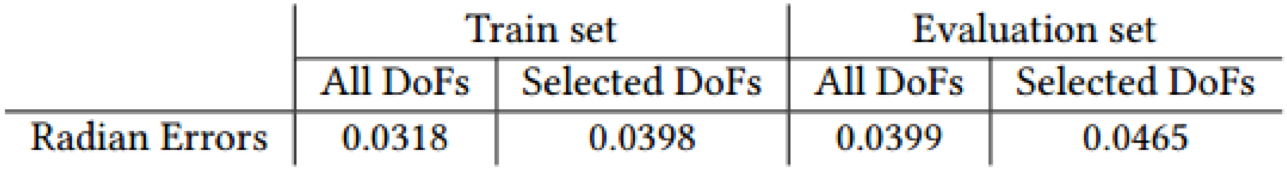
Dataset & Training

- 34 interaction sequences with about 20K frames.
- 90% as the training set, 10% as the evaluation set.
- Select no more than 3 DoFs in each frame to add large Gaussian noise.
- 100 epochs using Adam optimizer with learning rate of 0.001.

Mean Standard Deviation in input

Selected DoFs	Other DoFs
0.45 rad	0.042 rad

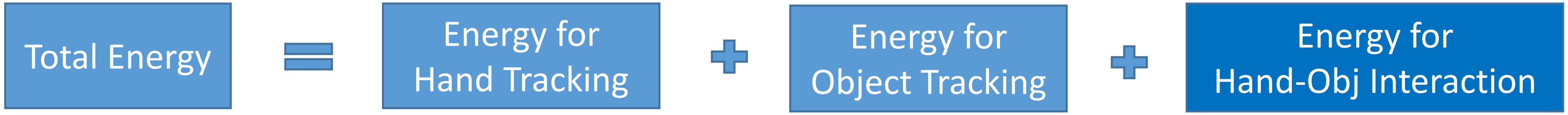
Test of LSTM



	Train set		Evaluation set	
	All DoFs	Selected DoFs	All DoFs	Selected DoFs
Radian Errors	0.0318	0.0398	0.0399	0.0465

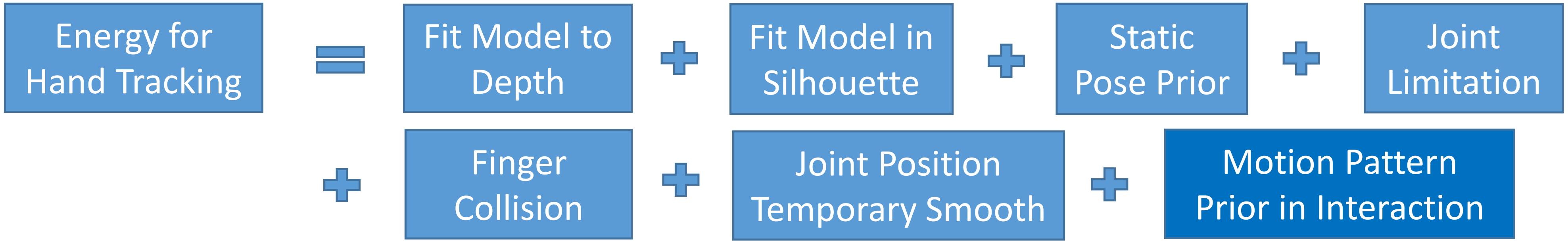
Joint Hand-Object Motion Tracking

- **Unified Energy**



Total Energy = Energy for Hand Tracking + Energy for Object Tracking + Energy for Hand-Obj Interaction

- **Energy for Hand Tracking**

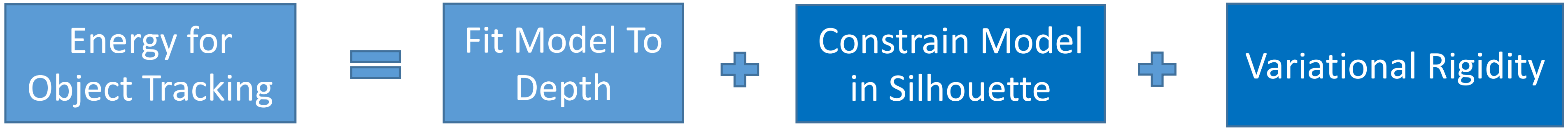


Energy for Hand Tracking = Fit Model to Depth + Fit Model in Silhouette + Static Pose Prior + Joint Limitation + Finger Collision + Joint Position Temporary Smooth + Motion Pattern Prior in Interaction

Sphere-meshes for realtime hand modeling and tracking. Anastasia Tkach, et al.TOG2016

$$E_{\mathrm{lstm}}(\theta^t) = \|\theta^t - \underset{\substack{\uparrow \\ \textbf{Output of LSTM}}}{\theta_p^t}\|_2^2$$

- **Energy for Object Tracking**

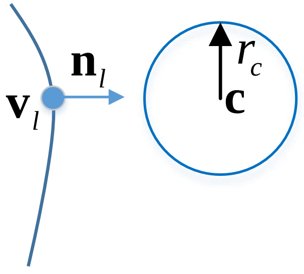


Energy for Object Tracking = Fit Model To Depth + Constrain Model in Silhouette + Variational Rigidity

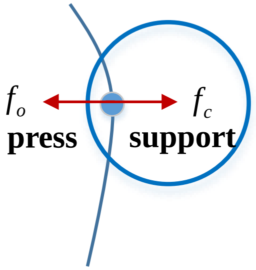
Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. Richard A Newcombe et al. CVPR2015

Joint Hand-Object Motion Tracking

➢ hand-object interaction



Object Surface Sphere of Hand



$$d_i(\mathbf{v}_l,\mathbf{c}) = r_c + \mathbf{n}_l(\mathbf{v}_l - \mathbf{c})$$

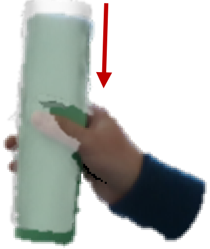
$$E = \tau(d_i)d_i^2$$

$$\tau(d_i) = \begin{cases} 1 & d_i > 0 \\ 0 & else \end{cases}$$

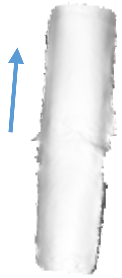
➢ model to silhouette

Reference Color

Reconstructed Object



with



without

➢ variational rigidity

Area far from Contact point
Large Rigidity

Area near Contact point
Small Rigidity

Experiments & Results

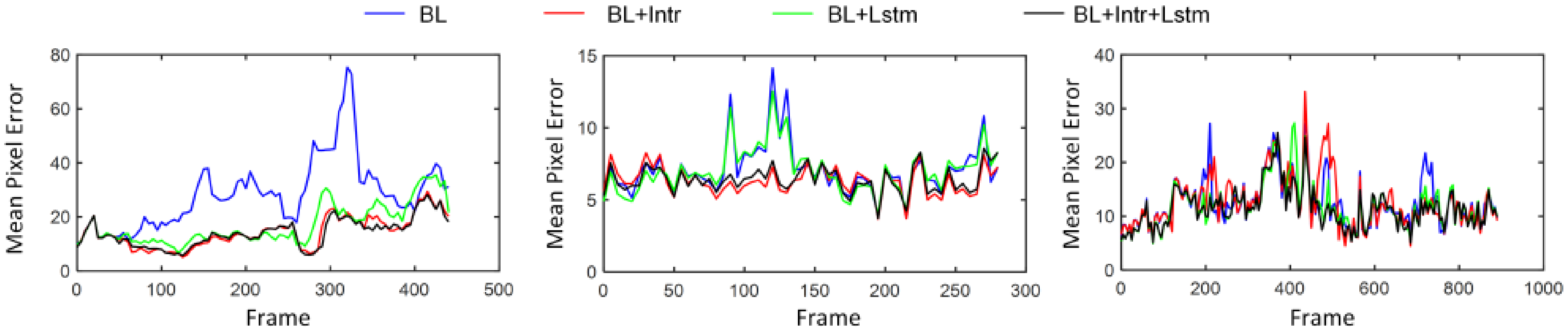
Evaluations

➢ **Ablation Study for Hand Tracking**

Sequence	Frames
RotatePepper	440
PourBottle	280
ReconstructCat	890

Mean Pixel Error

	BL	BL+Intr	BL+Lstm	BL+Intr+Lstm
RotatePepper	28.4	14.4	17.1	14.1
PourBottle	7.1	6.3	7.0	6.4
ReconstructCat	12.5	12.7	11.8	11.5



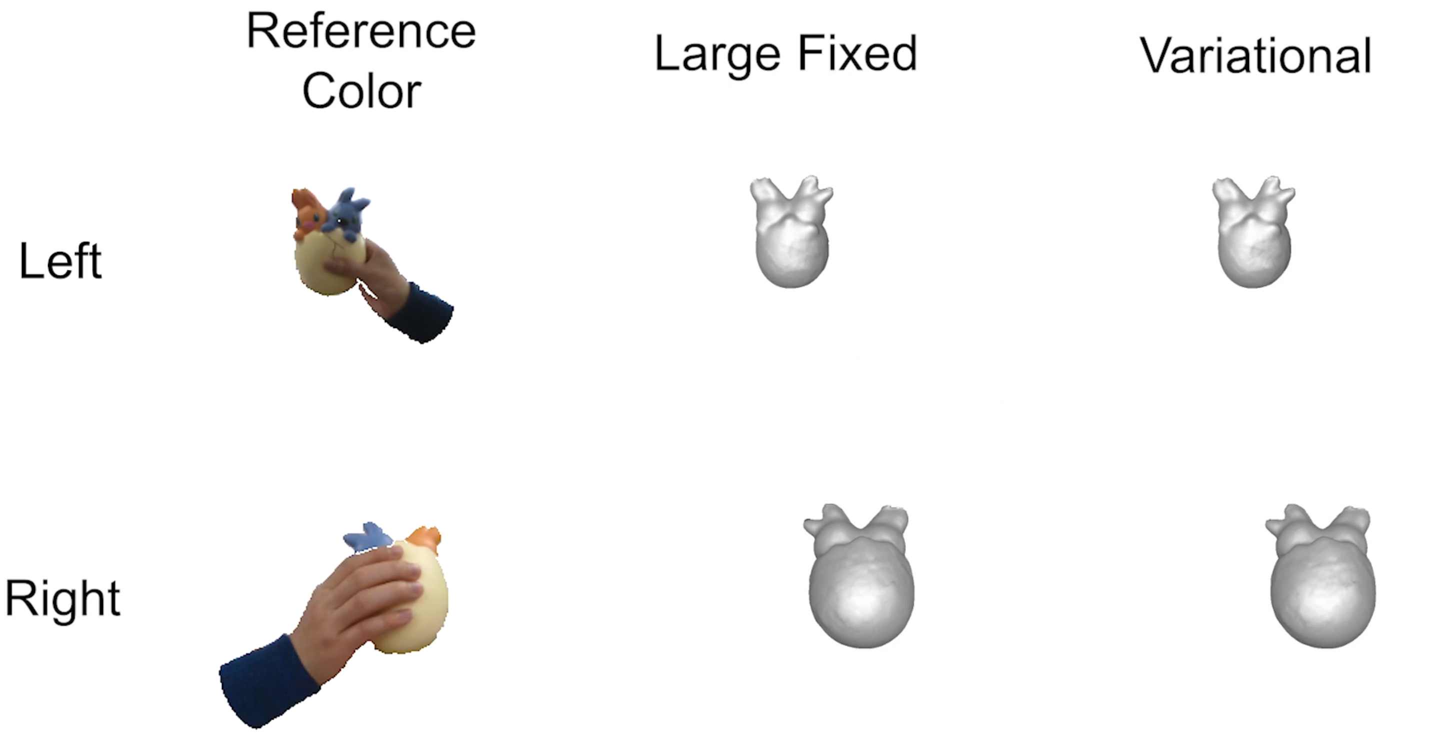
(a) RotatePepper (b) PourBottle (c) ReconstructCat

BL baseline **Intr** Interaction term **Lstm** Lstm based pose prediction

Evaluations

- **Ablation Study for Object Tracking**

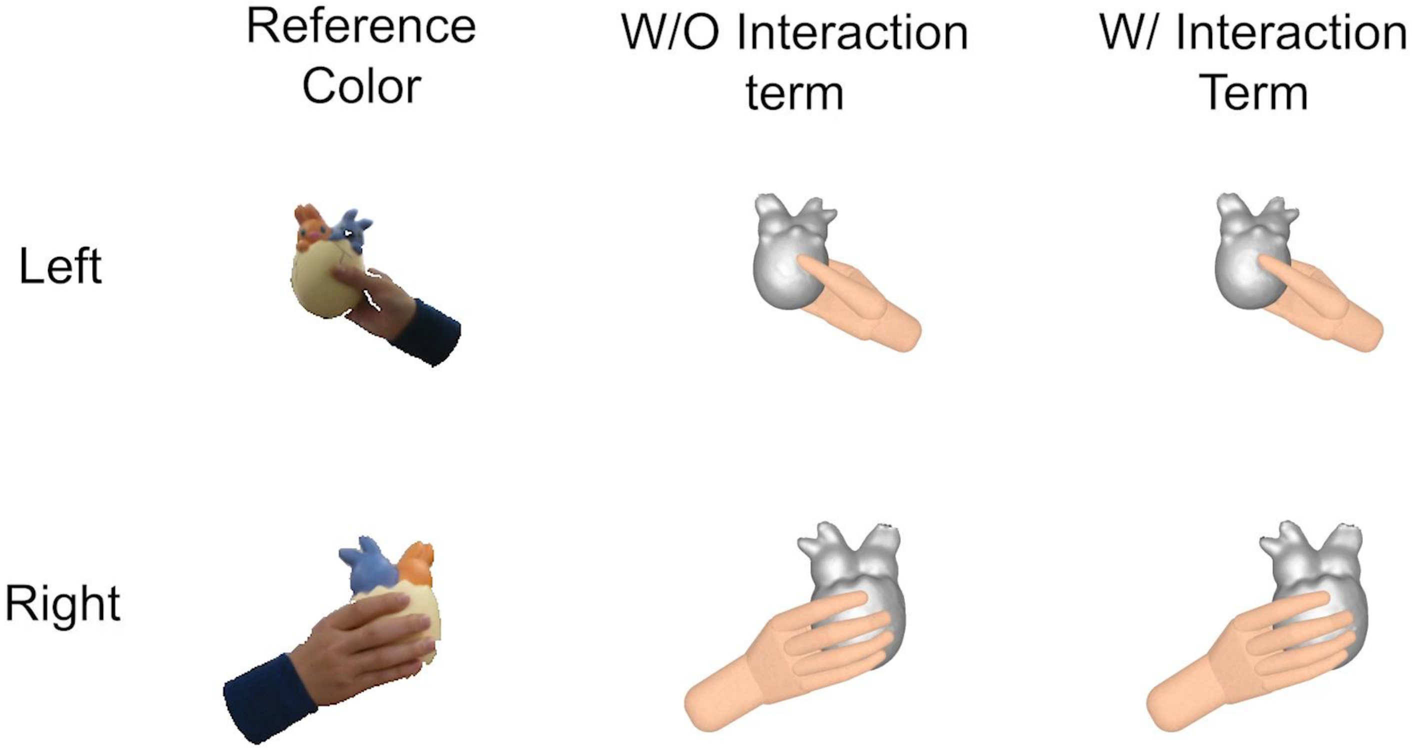
(a) Variational Rigidity



Evaluations

- **Ablation Study for Object Tracking**

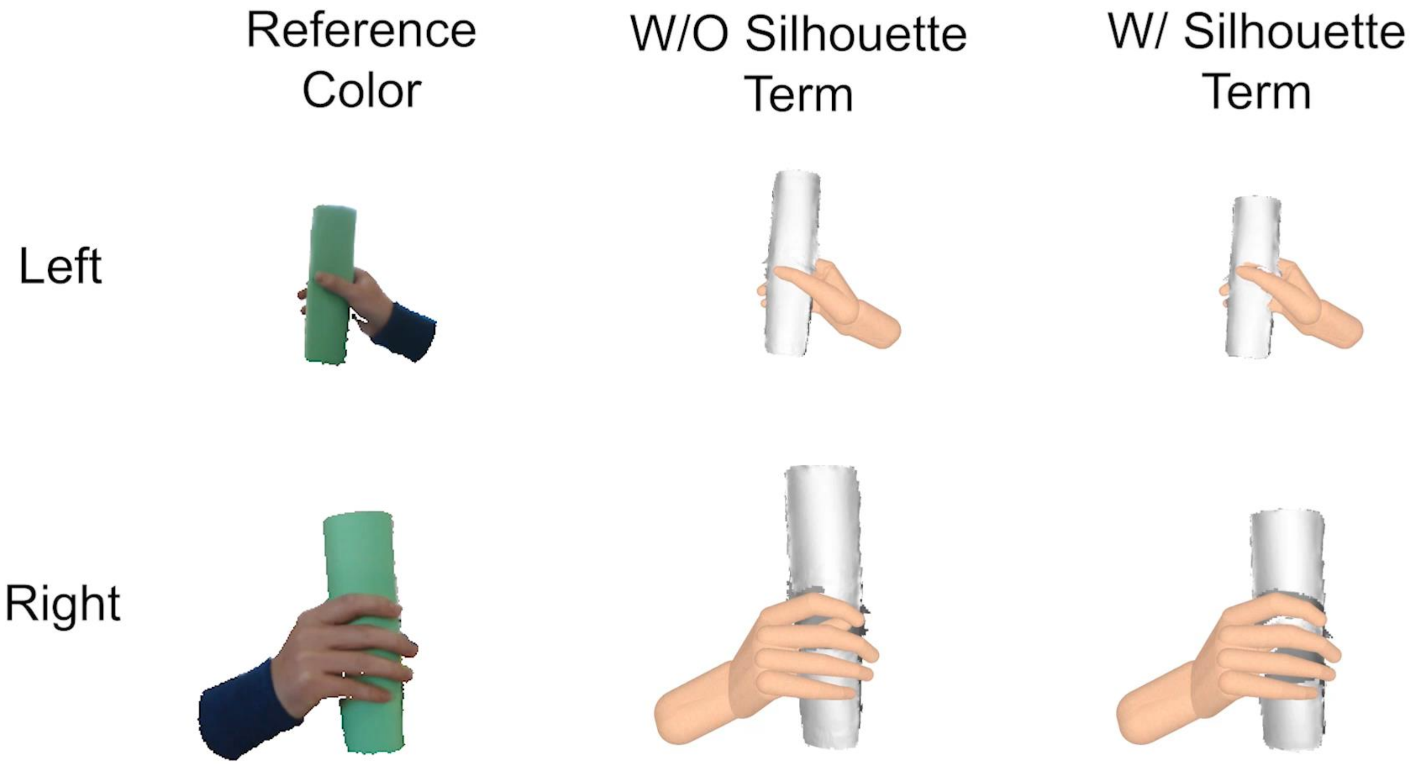
(b) Interaction Term



Evaluations

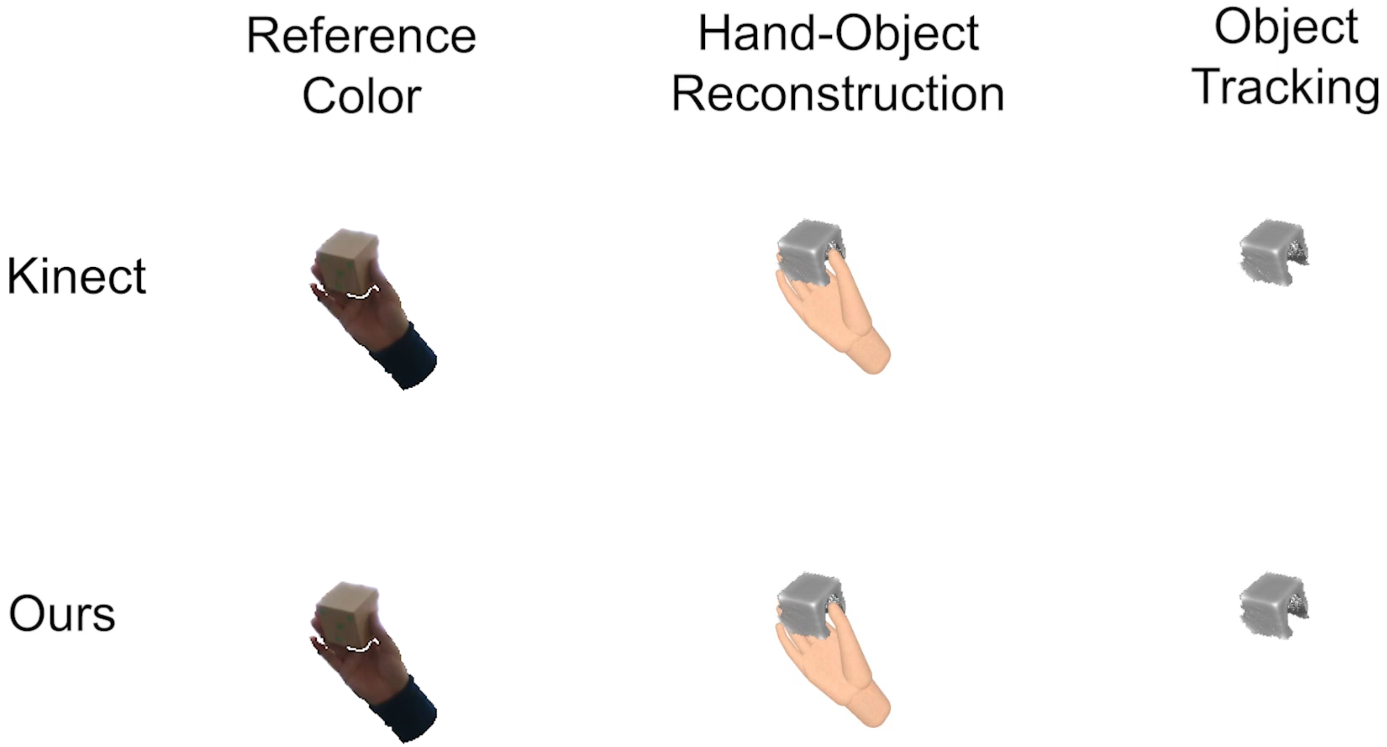
- **Ablation Study for Object Tracking**

(c) Silhouette Term



Qualitative Comparison

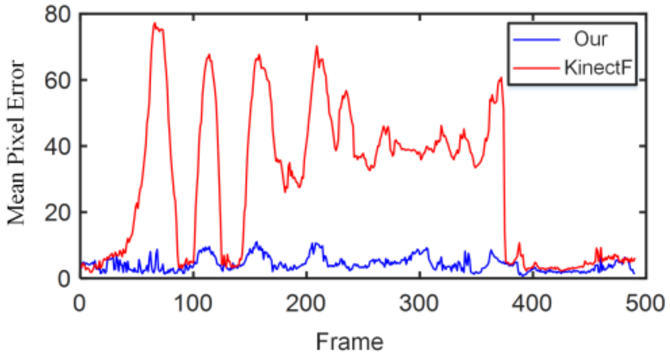
- **Comparison With KinectFusion**

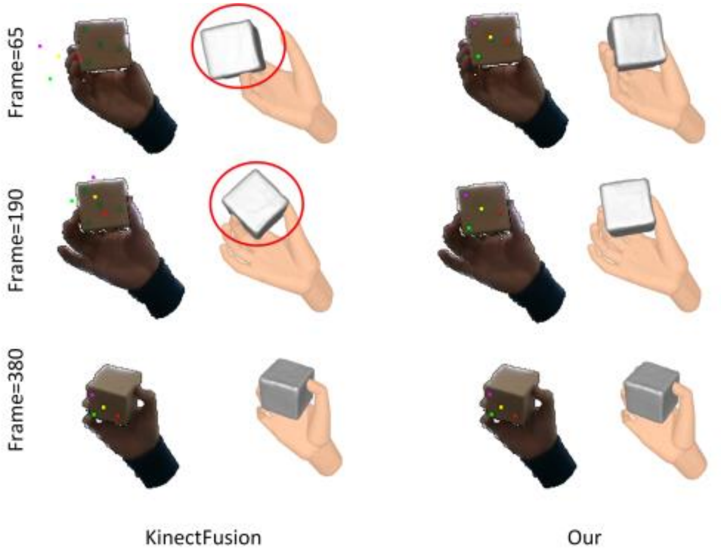


Experiments & Results

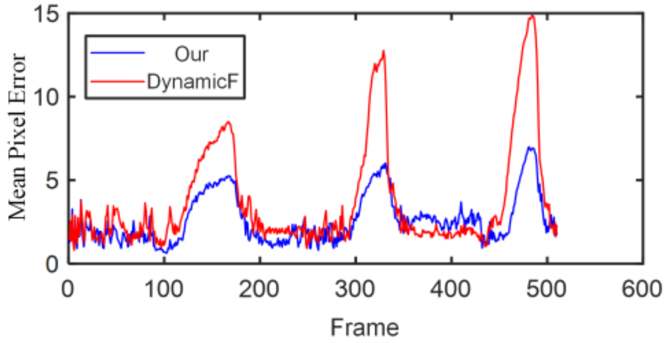
Quantitative Comparison

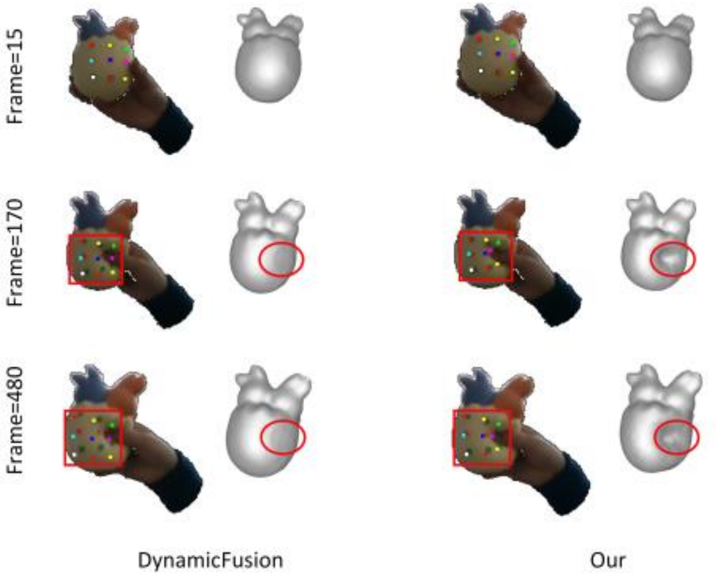
- Comparison With KinectFusion





- Comparison With DynamicFusion





Limitations & Future Work

➢ Limitations

- No color information in object tracking
- Only consider contact constraints
- Only one hand and one object
- Cannot handle topology change of object

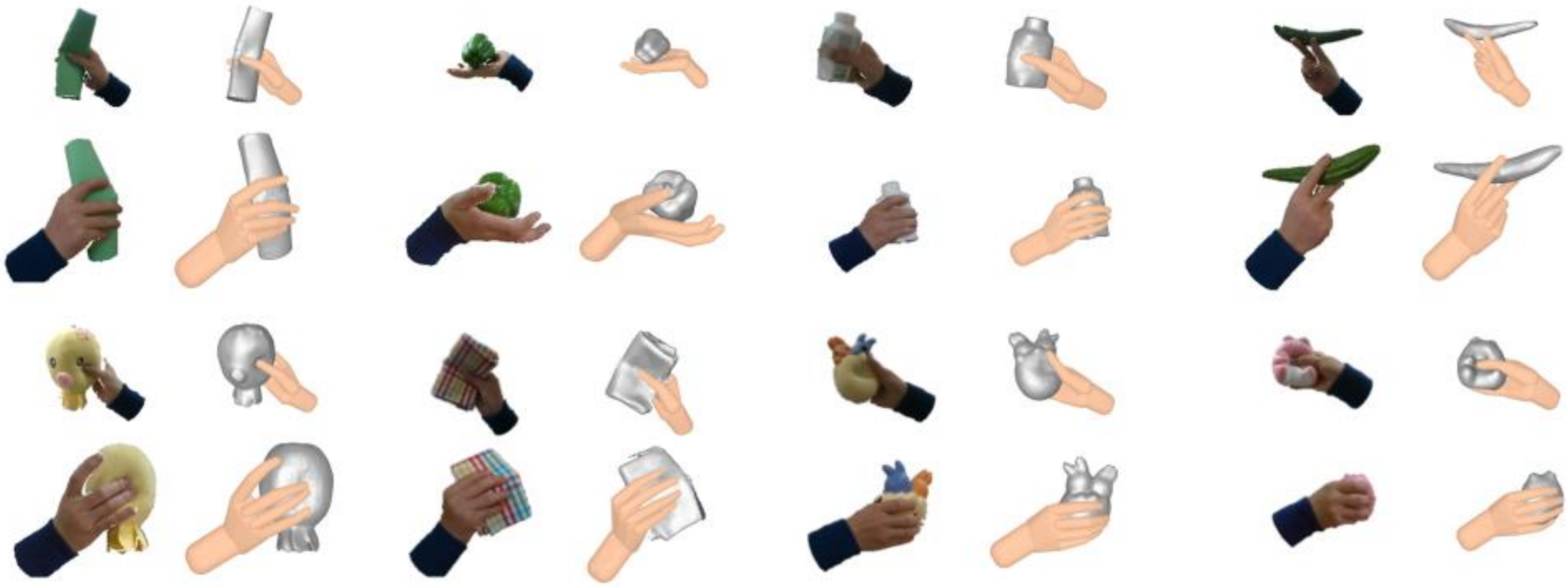
➢ Future Work

- Achieve more realistic interaction reconstruction
  color information, two hands with multi-objects, topology-change

- Reduce equipment requirement
  use one RGB-D camera

Conclusions

- ➢ An LSTM-based predictor, a novel interaction term, and variational rigidity
- ➢ A unified framework integrating segmentation information, pose prediction and new regularizers
- ➢ A system simultaneously achieving hand tracking, object fusion and nonrigid object tracking in real-time



Conclusions

Thanks for Your Attention!