REACTO: Reconstructing Articulated Objects from a Single Video

CVPR JUNE 17-21, 2024 SEATTLE, WA

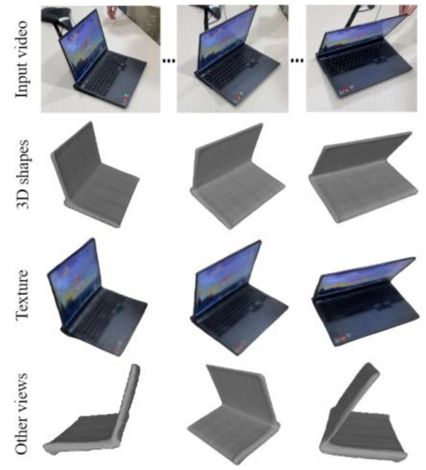
Chaoyue Song[1], Jiacheng Wei[1], Chuan-Sheng Foo[2], Guosheng Lin[1], Fayao Liu[2]
[1]Nanyang Technological University, [2]A*STAR

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
Agency for Science, Technology and Research SINGAPORE

Introduction

Motivation

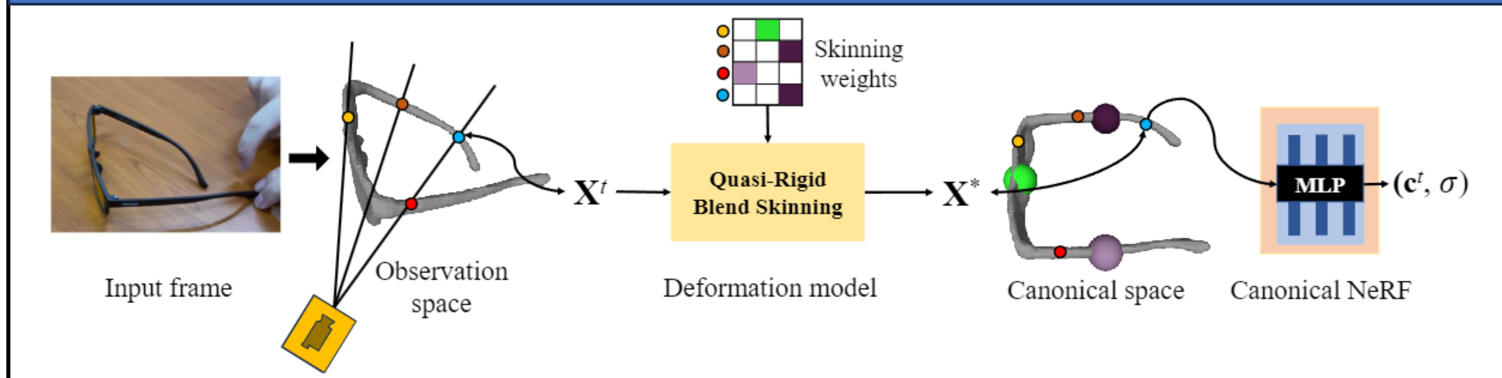
- Current research focuses on modeling humans or animals it's essential to give greater consideration to general articulated objects, which are prevalent in our daily lives.
- Part-rigidity nature of general articulated objects sets them apart from humans or animals, need to design rigging strategy and deformation model to describe their motion.



Task Description

- Given a single casual video capturing a general articulated object (e.g., stapler, laptop, etc.), we model the geometry, texture, and motions of the object.

Overview



We model an articulated 3D object using a shape and appearance model based on a canonical NeRF and a deformation model (Quasi-Rigid Blend Skinning) for transforming 3D points between the observation space and the canonical space. The colors in skinning weights signify the assigned bone for each point.

Method

Canonical NeRF for shape and appearance

$$\mathbf{c}^t = \mathbf{MLP}_{\text{color}}(\mathbf{X}^*, \mathbf{D}^t, \psi_a^t),$$

$$\sigma = \Phi_\beta(\mathbf{MLP}_{\text{SDF}}(\mathbf{X}^*)),$$

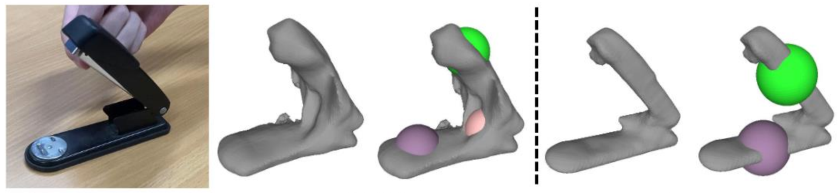
Quasi-rigid blend skinning for deformation

1. Motion representation.

$$\mathbf{X}^t = \mathcal{D}^{t,c\to o}(\mathbf{X}^*) = \mathbf{T}^t_{\text{global}}\mathbf{T}^{t,c\to o}_{\text{obj}}\mathbf{X}^*,$$

$$\mathbf{X}^* = \mathcal{D}^{t,o\to c}(\mathbf{X}^t) = \mathbf{T}^{t,o\to c}_{\text{obj}}(\mathbf{T}^t_{\text{global}})^{-1}\mathbf{X}^t,$$

2. Bone definition: rig on bones (ours) v.s. rig on joints.



3. Quasi-sparse skinning weights.

$$\mathbf{W}^s = \text{softmax}(\frac{d_M(\mathbf{X}) + \mathbf{W}_\Delta}{\gamma}).$$

4. Geodesic point assignment.

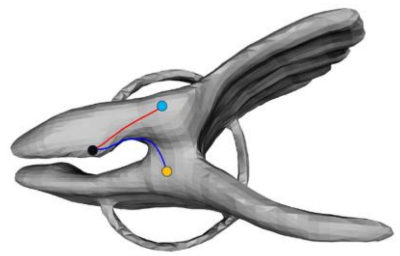
Algorithm 1 Geodesic point assignment

Input: Point assignment $\mathbf{M} = \mathbf{0} \in \mathbb{R}^B$, Mahalanobis distance d^i_M and d^j_M, geodesic distance d^i_G and d^j_G, bone index i and j, hyperparameters η, ζ.

Output: Updated assignment $\mathbf{M}$

if $d^i_M / d^j_M < 1 - \eta$ **then**
  $\mathbf{M}[i] \leftarrow 1$
else if $\frac{|d^i_G - d^j_G|}{\min(d^i_G, d^j_G)} < \zeta$ **then**
  $\mathbf{M}[i], \mathbf{M}[j] \leftarrow 1$ ▷ Assigning to joints
else
  $\mathbf{M}[\text{argmin}(d_G)] \leftarrow 1$
end if

$$\mathcal{L}_{sparse} = \frac{\sum \left\|\mathbf{W}^s \odot \bar{\mathbf{M}}\right\|^2}{\sum \bar{\mathbf{M}}},$$

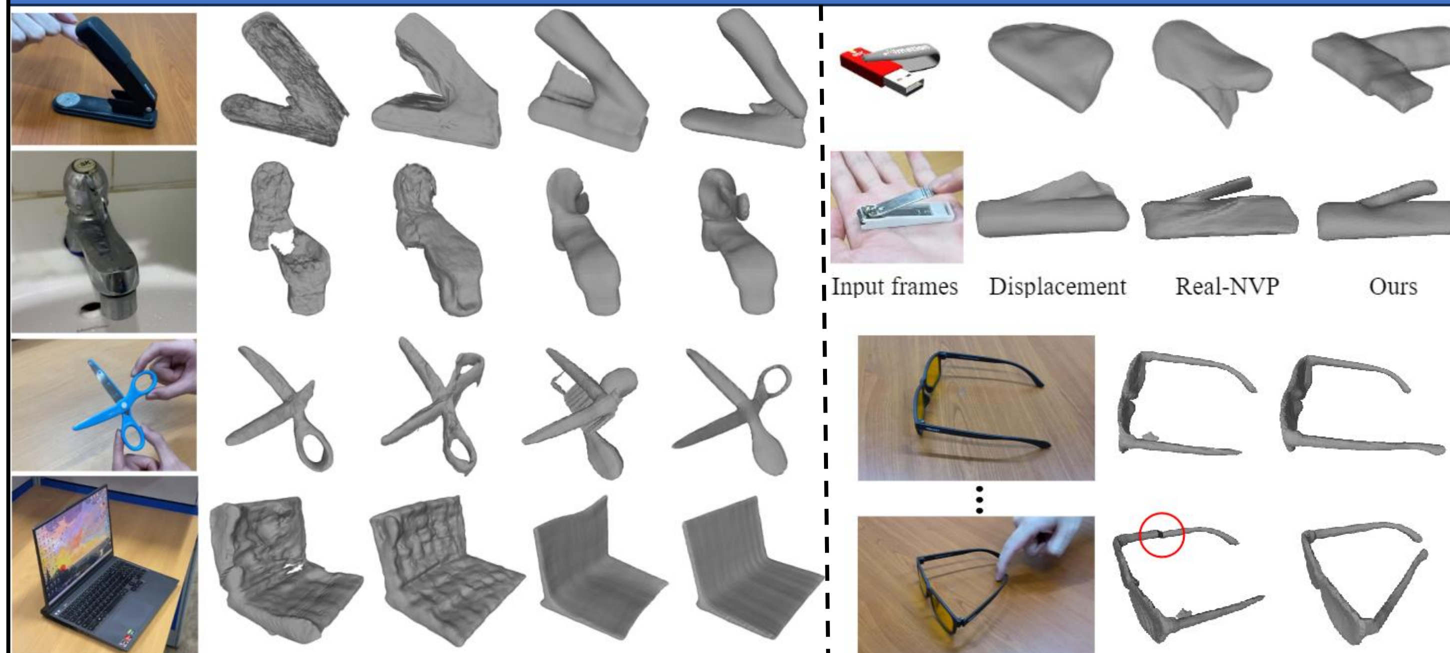


Geodesic distances between 3D point (black) and bones (blue and yellow). Shorter distances indicate stronger associations.

5. Quasi-rigid blend skinning.

$$\mathbf{X}(\psi_\mathbf{p}) = \mathbf{T}_{\text{obj}}\mathbf{X} = (\sum_{b=0}^{B-1} w^S_b \mathbf{T}_b)\mathbf{X},$$

Experimental Results



Method	USB			stapler			scissors		
	CD(↓)	F(10%, ↑)	F(5%, ↑)	CD(↓)	F(10%, ↑)	F(5%, ↑)	CD(↓)	F(10%, ↑)	F(5%, ↑)
BANMo	20.3	65.1	45.0	19.1	57.8	32.8	19.9	66.8	41.4
MoDA	17.1	74.9	49.5	18.8	64.2	40.3	14.8	77.7	42.3
PPR	20.7	65.7	38.9	16.8	67.5	40.0	16.1	71.4	39.9
Ours	**15.3**	**78.6**	**51.5**	**14.3**	**75.5**	**42.7**	**14.0**	**78.2**	**43.9**

We introduce REACTO, a groundbreaking method for reconstructing general articulated 3D objects from single casual videos, achieving enhanced modeling and precision by redefining rigging structures and employing Quasi-Rigid Blend Skinning. QRBS ensures the rigidity of each component while retaining smooth deformation on the joints by utilizing quasi-sparse skinning weights and geodesic point assignment.