Supplementary Materials for LASR: Learning Articulated Shape Reconstruction from a Monocular Video

Table 1. Choices of hyper-parameters for training.

Name	Value
Optimization parameters	
Network architecture of ϕ_w	ResNet-18 (ImageNet-pretrained) [1]
Optimizer	Adam [2]
Learning rate for ϕ_w	1×10^{-4}
Learning rate for other params.	5×10^{-3}
Batch size	8 image pairs
Loss weight $\{\beta_1, \ldots, \beta_4\}$	$\{0.5, 0.5, 2, 5 \times 10^{-3}\}$
Measurement pre-processing	
Crop center	Center of object bounding box
Crop size	$1.2 \times$ longest edge
Resized to	256×256

1. Implementation details

Training details: We include details of the hyper-parameters used for training in Tab. 1.

Video pre-processing: We provide details for video pre-processing. To ensure enough object motion between adjacent frames, we use a heuristic rule that skips the next frame when the average magnitude of measured flow within the object silhouette is lower than 0.05 in the clip space.

2. Notations

A summary of the notations is listed in Tab. 2.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 1

[2] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

Table 2. Table of notations used in this work.

Symbol	Description
	Numbers
T	Number of frames in the input video
M	Number of faces in the mesh
N	Number of vertices in the mesh
B	Number of bones for LBS
	Measurements
I_t	Input RGB image at time t
S_t	Input or measured object silhouette image at time t
$\mathbf{u}_t^+$	Input or measured forward optical flow map from time t to $t+1$
$\mathbf{u}_t^-$	Input or measured backward optical flow map from time t to $t-1$
$\mathbf{Y}^*$	Union of all measurements $\{I_t, S_t, \mathbf{u}_t^+, \mathbf{u}_t^-\}$
	Renderings
$\hat{I}_t$	Rendered color image of the object at time t
$\hat{S}_t$	Rendered object silhouette image at time t
$\hat{\mathbf{u}}_t^+$	Rendered forward optical flow map of the object from time t to $t+1$
$\hat{\mathbf{u}}_t^-$	Rendered backward optical flow map of the object from time t to $t-1$
$\mathbf{Y}$	Union of all renderings $\{\hat{I}_t, \hat{S}_t, \hat{\mathbf{u}}_t^+, \hat{\mathbf{u}}_t^-\}$
	Variables
f_t	Focal length of the camera at time t
$\mathbf{K_t}$	Intrinsic matrix of a simple pinhole camera (with zero skew and square pixel) at time t
$\mathbf{R_{0,t}}$	Object-to-camera rotation matrix $\in SO(3)$ at time t
$\mathbf{T_{0,t}}$	Object-to-camera translation vector at time t
$\mathbf{G_{0,t}}$	Object-to-camera transformation at time t, $\mathbf{G_{0,t}} = \left(\mathbf{R_0} \mid \mathbf{T_0}\right)_\mathbf{t}$
$\mathbf{R_{1\ldots B,t}}$	Bone rotations from the rest pose to time t
$\mathbf{T_{1\ldots B,t}}$	Bone rotations from the rest pose to time t
$\mathbf{G_{1\ldots B,t}}$	Bone transformations from the rest pose to time t, $\mathbf{G_{i,t}} = \left(\mathbf{R_i} \mid \mathbf{T_i}\right)_\mathbf{t}, i \in \{1\ldots, B\}$
$\mathbf{D_t}$	Union of camera and bone transformations $\{\mathbf{G_{0,t}}, \ldots, \mathbf{G_{B,t}}\}$
$\mathbf{P_t}$	Projection matrix of the camera at time t, $\mathbf{P_t} = \mathbf{K_t G_{0,t}}$
$\Delta\mathbf{V_t}$	Vertex motion from the rest shape to time t
	Parameters
(p_x, p_y)	Principal point of the camera
$\bar{\mathbf{V}}_i$	Position of the i-th vertex of the mesh in the rest pose (or mean shape)
$\bar{\mathbf{C}}_i$	Color of the i-th vertex of the mesh
$\mathbf{S}$	Union of all mesh parameters, $\mathbf{S} = \{\bar{\mathbf{V}}, \bar{\mathbf{C}}, \mathbf{F}\}$
$\mathbf{J}_b$	Position of the center of the b-th bone (or Gaussian component)
$\mathbf{Q}_b$	Precision matrix of b-th bone (or Gaussian component)
$\mathbf{W}$	Skinning weights matrix, **W**=$\{\mathbf{J}, \mathbf{Q}\}$
ϕ_w	Weights of the convolutional camera and pose network
$\mathbf{n}^*$	Normal vector of the symmetry plane in the canonical frame
$\mathbf{X}$	Union of all Parameters
	Constants
$\mathbf{F}$	Faces of the mesh
$\boldsymbol{\beta}$	Weights between the losses
$\mathbf{H}$	Householder transformation matrix describing reflection about the y-z plane
$\mathbf{n}_0$	Unit vector towards to the x axis
	Others
S0	Training stage 0: optimize for $\{\phi_w(f_t, \mathbf{G_{0,t}}), p_x, p_y, \mathbf{n}^*, \bar{\mathbf{V}}, \bar{\mathbf{C}}\}$
S1-3	Training stage 1 to 3: optimize for $\{\phi_w(f_t, \mathbf{G_{0\ldots B,t}}), p_x, p_y, \mathbf{n}^*, \bar{\mathbf{V}}, \bar{\mathbf{C}}, \mathbf{J}, \mathbf{Q}\}$