Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Human from Blur

Human Pose Tracking from Blurry Images

Yiming Zhao[1] Denys Rozumnyi[1] Jie Song[1] Otmar Hilliges[1] Marc Pollefeys[1,3] Martin R. Oswald[1,2]
ETH Zürich[1], Universiteit van Amsterdam[2], Microsoft[3]

ICCV23 PARIS

I. Introduction

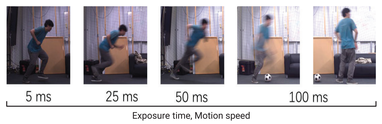
Motivation

Shape-from-Blur [1] and Motion-from-Blur [2] tackle the **rigid body** tracking and reconstruction from blurry images.

How to address the same problem for the **soft body**? Best application case: Human body motion blur.

Challenge

- Dataset almost unavailable
- Many factors determine blur amount



Real Scenario

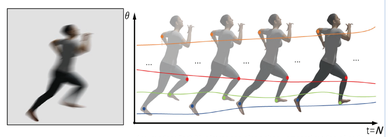
Blur Formation

$$\hat{I} = \underbrace{\left(1 - \frac{1}{N}\sum_{i=1}^{N}\mathcal{R}_S(\Theta_i)\right)}_{\text{Inverse alpha channel}} \cdot \underbrace{B}_{\text{Background}} + \underbrace{\frac{1}{N}\sum_{i=1}^{N}\mathcal{R}_S(\Theta_i)\cdot\mathcal{R}_F(\Theta_i)}_{\text{Blurred foreground (human)}}$$

$\mathcal{R}_S(\Theta_i)$ Differentiable rendered soft-silhouette

$\mathcal{R}_F(\Theta_i)$ Differentiable rendered foreground object

$\Theta_i = SMPL(\theta_{t=i}, \beta) + T_{t=i}$ Mesh driven by SMPL [3]



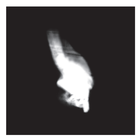
II. Setup

Input:

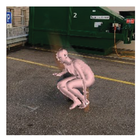
$\boldsymbol{B}$ Clear Background

$\boldsymbol{I}$ Blur Input

Initialize:

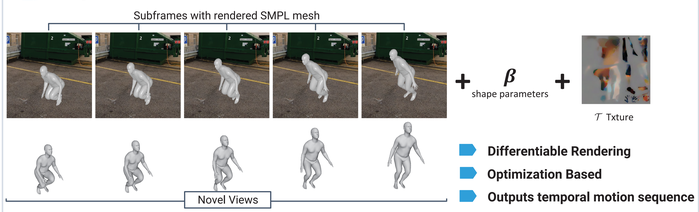


$\boldsymbol{\alpha}$ Extracted Matting [4]



$\boldsymbol{\theta}_{\text{ini}}$ Initial Pose Estimation METRO[5]+Registration[14]

Output:



- **Differentiable Rendering**
- **Optimization Based**
- **Outputs temporal motion sequence**

Target: Retrieve the motion in subframes for timestamps t from 1 to N

Blur rate:

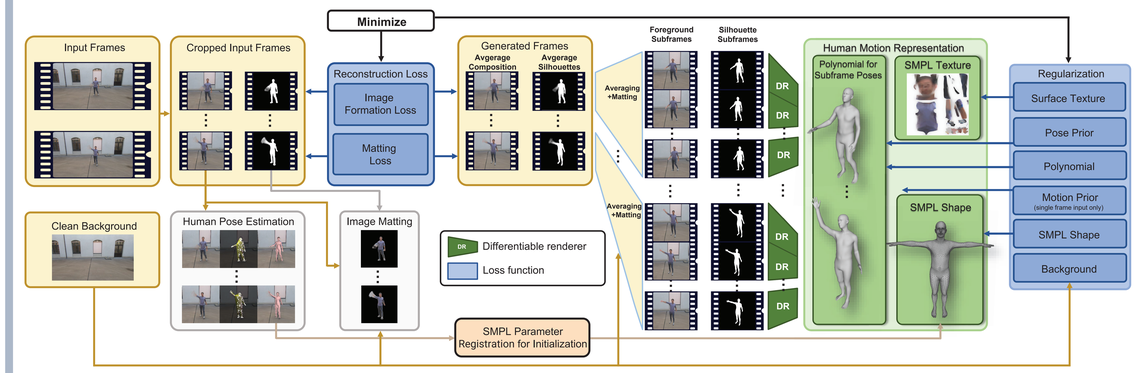
$$\text{blur rate} = \frac{|\bigcup_{t=1}^{N} S_i|_1}{|S_1|_1} - 1$$

We define a scale to quantify the level of the blur amount.
The union of all projected sub-frame silhouettes and divide it by the first silhouette.

Dataset:

- Synthetic Dataset: Generated from AMASS[6] , CAPE[7], BG-20K[8], SURREAL[9]
- Semi-real Dataset: B-AIST++[10]

III. Pipeline

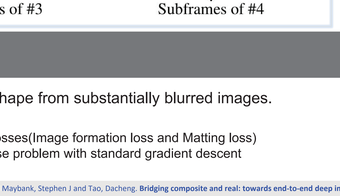


Addressing Ambiguity of Joint Motion

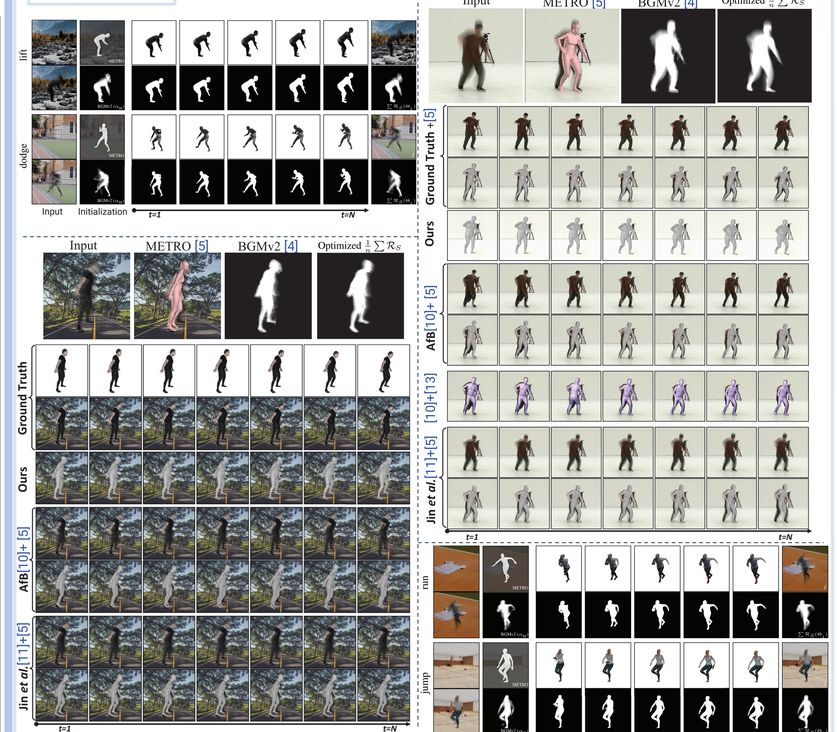
Single Frame
Short Temporal Motion Prior as a loss term during optimization

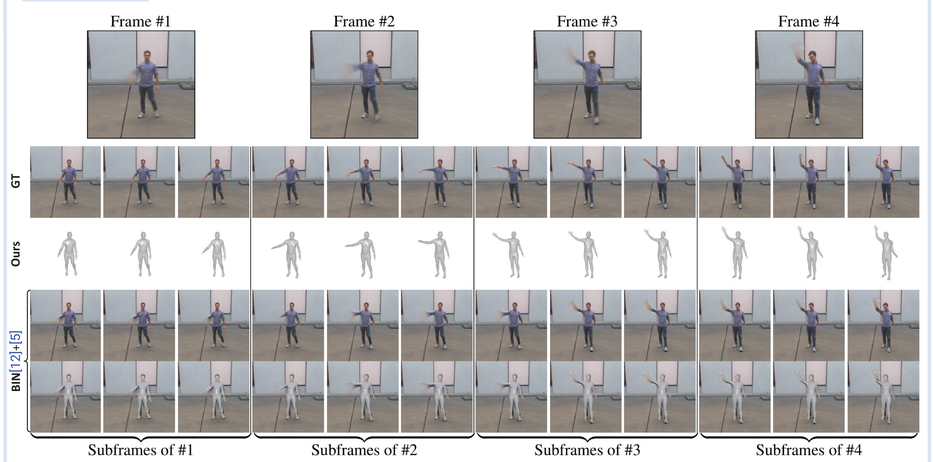
$$\mathcal{L}_A = |G(D(\boldsymbol{C}), \boldsymbol{C}) - \boldsymbol{C}|_1$$

Multiframes:
Minimize the difference of retrieved joints' position $\boldsymbol{J}$ and joints' rotation $\boldsymbol{\theta}$ and their first order derivatives between t =N of current blur frame and t = 1 of the next blur frame



IV. Experiments





V. Conclusion

We propose the first method to reconstruct sub-frame human motion and textured shape from substantially blurred images.

➔ **Key idea:**
Approaching the problem from a generative viewpoint and describing a fully differentiable forward process to generate blurry images from a given 3D human model.

➔ **Core of method:**
The render reconstruction losses(Image formation loss and Matting loss) that allow to solve the inverse problem with standard gradient descent methods.

[1] Denys Rozumnyi, Martin R. Oswald, Vittorio Ferrari, and Marc Pollefeys. **Shape from blur: Recovering textured 3d shape and motion of fast moving objects**. In NeurIPS 2021.
[2] Denys Rozumnyi, Martin R. Oswald, Vittorio Ferrari, and Marc Pollefeys. **Motion-from-blur: 3d shape and motion estimation of motion-blurred objects in videos**. In CVPR 2022
[3] Loper, Matthew and Mahmood, Naureen and Romero, Javier and Pons-Moll, Gerard and Black, Michael J. **SMPL: A Skinned Multi-Person Linear Model**. In SIGGRAPH Asia 2015
[4] Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian Curless, Steve Seitz, and Ira Kemelmacher Shlizerman. **Real-time high-resolution background matting**. In CVPR 2021
[5] Kevin Lin, Lijuan Wang, and Zicheng Liu. **End-to-end human pose and mesh reconstruction with transformers**. In CVPR 2021
[6] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. **AMASS: Archive of motion capture as surface shapes**. In ICCV 2019
[7] Q Ma et al. **Learning to Dress 3D People in Generative Clothing**. In CVPR 2020
[8] Li, Jizhizi and Zhang, Jing and Maybank, Stephen J and Tao, Dacheng. **Bridging composite and real: towards end-to-end deep image matting**. IJCV 2022
[9] Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J. Black, Ivan Laptev, and Cordelia Schmid. **Learning from synthetic humans**. In CVPR 2017
[10] Zhihang Zhong, Xiao Sun, Zhirong Wu, Yinqiang Zheng, Stephen Lin, and Imari Sato. **Animation from blur: Multimodal blur decomposition with motion guidance**. In ECCV 2022
[11] M Jin et al. **Learning to extract a video sequence from a single motion-blurred image**. In CVPR 2018
[12] Wang Shen, Wenbo Bao, Guangtao Zhai, Li Chen, Xiongkuo Min, and Zhiyong Gao. **Blurry video frame interpolation**. In CVPR 2020
[13] Hongwen Zhang*, Yating Tian*, Xinchi Zhou, Wanli Ouyang, Yebin Liu, Limin Wang, Zhenan Sun. **PyMAF: 3D Human Pose and Shape Regression with Pyramidal Mesh Alignment Feedback Loop**. In ICCV 2021
[14] Bhatnagar, Bharat Lal and Sminchisescu, Cristian and Theobalt, Christian and Pons-Moll, Gerard. **Combining Implicit Function Learning and Parametric Models for 3D Human Reconstruction**. In ECCV 2020