Motion-Aware Animatable Gaussian Avatars Deblurring

Muyao Niu[1,2,*] Yifan Zhan[1,2] Qingtian Zhu[1] Zhuoxiao Li[1] Wei Wang[2]
Zhihang Zhong[3,†] Xiao Sun[2] Yinqiang Zheng[1]

[1]The University of Tokyo [2]Shanghai Artificial Intelligence Laboratory [3]Shanghai Jiao Tong University

Abstract

The creation of 3D human avatars from multi-view videos is a significant yet challenging task in computer vision. However, existing techniques rely on high-quality, sharp images as input, which are often impractical to obtain in real-world scenarios due to variations in human motion speed and intensity. This paper introduces a novel method for directly reconstructing sharp 3D human Gaussian avatars from blurry videos. The proposed approach incorporates a 3D-aware, physics-based model of blur formation caused by human motion, together with a 3D human motion model designed to resolve ambiguities in motion-induced blur. This framework enables the joint optimization of the avatar representation and motion parameters from a coarse initialization. Comprehensive benchmarks are established using both a synthetic dataset and a real-world dataset captured with a 360-degree synchronous hybrid-exposure camera system. Extensive evaluations demonstrate the effectiveness of the model across diverse conditions. Codes Available: https://github.com/MyNiuuu/MAD-Avatar

1. Introduction

Motion blur arises when scene changes occur during camera exposure, degrading image quality and perceptual clarity. Despite advances in camera technology, it remains prevalent due to the unpredictable nature of object motion. Consequently, many methods have been proposed to restore sharp details from blurry captures [3, 62, 69, 79, 87, 90].

In the field of 3D reconstruction, creating high-quality 3D human avatars holds immense potential for the industry [4, 5, 13, 14, 18, 34, 38, 50, 68, 75, 81, 84]. Cutting-edge methods employing 3D Gaussian Splatting [8, 9, 21, 40] and the Skinned Multi-Person Linear (SMPL) model [41] have achieved notable success. These techniques [12, 13, 34, 57] typically rely on video data with SMPL parameters from calibration techniques like EasyMocap [55].

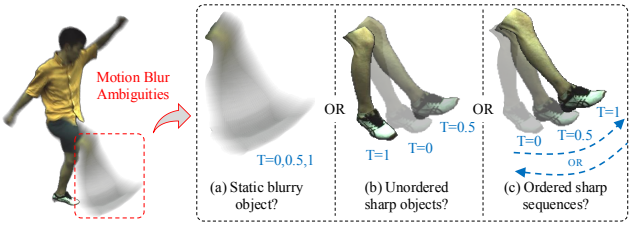

Figure 1. The ambiguity brought by motion blur. When reconstructing sharp 3DGS avatars from blurry frames, motion-induced blur introduces challenging ambiguities in motion interpretation.

While these models have achieved remarkable results with well-calibrated and high-quality video frames, their performance can be significantly degraded by blur effects caused by human motion. In practical scenarios, such motion-induced blur is often unavoidable due to the unpredictable speeds and movements of subjects. In particular, blur effects can adversely affect the performance of existing models in two key ways. First, blurry captures may cause the 3DGS model to learn a distorted 3D representation, arising from the inherent ambiguity of motion blur (Fig. 1). Such ambiguity hinders the accurate recovery of structural information and texture details. Second, even though static cameras can be calibrated before recording, blurry captures still lead to erroneous estimation of SMPL parameters.[1]

A practical strategy to mitigate this issue is to implement a two-stage baseline approach. First, 2D deblurring techniques [37, 53, 80, 89] are applied to restore sharp video sequences. Then, the deblurred frames are used to train the 3DGS avatar model. While this baseline method improves visual quality by partially resolving motion blur ambiguity through 2D deblurring, it overlooks intrinsic 3D scene information. This omission often leads to inconsistencies across multiple views during the deblurring process, compromising the performance of the 3DGS avatar model.

Based on the settings of existing gaussian avatar mod-

*Part of this work was done during the author's internship at the Shanghai Artificial Intelligence Laboratory.

†Corresponding author.

[1]While certain approaches, such as GauHuman [13], integrate pose refinement modules, they often yield suboptimal results because they lack a rigorous mechanism for modeling blur-aware motion.

els [12, 13, 57], this paper introduces the first model for reconstructing sharp, animatable 3D human avatars directly from blurry video frames. By exploiting the distinctive characteristics of human avatar models, we expand the conventional physics-based 2D image blur process to a 3D-aware blur formation model. This formulation decomposes the inherently ill-posed deblurring problem into two primary tasks: optimizing sub-frame motion representations and constructing the canonical sharp 3DGS avatar model. To mitigate the sub-frame motion ambiguity introduced by motion-induced blur, a 3D-aware human motion model based on the SMPL framework is incorporated. This model enables the joint optimization of human motion and the sharp 3DGS avatar. Using the optimized motion and canonical 3DGS, motion-blurred frames are synthesized by averaging a sequence of "virtual" sharp images, and the loss is computed against the observed blurry frames.

Extensive evaluations are conducted on various datasets, baselines, and settings to confirm the superiority and robustness of the proposed method. As there are no blur-aware avatar benchmarks for qualitative and quantitative evaluation, a synthetic dataset is developed based on the widely used ZJU-MoCap dataset [55], complemented by a high-quality real-world dataset captured using a 360-degree hybrid-exposure system. A do-it-yourself style demo with iPhone 16 Pro is also provided. Codes and datasets will be publicly released to facilitate relevant research in the future.

2. Related Work

Video Deblurring. In recent years, deblurring has shifted from traditional methods [16, 30, 59, 73] to learning-based models that directly predict sharp videos [2, 6, 17, 26, 27, 35, 47, 48, 52, 60, 65, 70, 86, 88, 91, 92]. Recently, the advent of NeRF [45] and 3DGS [21] has promoted multi-view deblurring. Specifically, methods like [28, 43, 67, 85] focus on deblurring static scenes affected by defocus or camera movement, while some approaches [1, 42, 61, 64] deblurs dynamic scenes. However, none of these methods could reconstruct sharp and animatable avatars from blurry videos.

3D Human Avatars. Since the advent of Neural Radiance Fields [45], research on neural human rendering has flourished [11, 19, 20, 33, 51, 54, 56, 63, 68, 72, 74, 78]. Recently, the superior performance of 3DGS [21] has spurred extensive research into using 3D Gaussian representations for dynamic human reconstruction [12, 18, 25, 29, 34, 39, 46, 76, 93]. However, these methods rely on high-quality input images and fail when applied to blurry captures.

3. Method

3.1. 3D Blur Formation Model

The physical mechanism of image formation involves capturing photons during the camera's exposure period. This process can be mathematically represented in the 2D camera coordinate space as the integration of a sequence of conceptual sharp images over the exposure duration:

$$\mathbf{I}^B(\mathbf{u}) = \int_0^\tau \mathbf{I}_t^S(\mathbf{u})\,\mathrm{d}t, \tag{1}$$

where $\mathbf{I}^B(\mathbf{u}) \in \mathbb{R}^{3\times H\times W}$ denotes the captured blurry image, with H and W representing the image height and width, respectively. The variable $\mathbf{u} \in \mathbb{R}^2$ indicates the pixel location, τ is the exposure time, and $\mathbf{I}_t^S(\mathbf{u}) \in \mathbb{R}^{3\times H\times W}$ represents the virtual sharp image at time t within the exposure period. This continuous integration process can be discretely approximated as:

$$\mathbf{I}^B(\mathbf{u}) \approx \frac{1}{n}\sum_{i=0}^{n-1} \mathbf{I}_i^S(\mathbf{u}), \tag{2}$$

where the blurred image $\mathbf{I}^B(\mathbf{u})$ is estimated by averaging n virtual intermediate sharp images $\mathbf{I}_i^S(\mathbf{u})$. The blur formation process is redefined from the perspective of 3D human avatar modeling, extending beyond the constraints of the 2D camera coordinate space. Specifically, the motion of a 3DGS avatar during the exposure period is modeled using a set of K 3D Gaussians $G_k(\mathbf{x})_{k=0}^{K-1}$ defined in the canonical space, together with a sequence of T SMPL [41] parameters $\{\mathcal{S}_t\}_{t=0}^{T-1} = \{\boldsymbol{\Theta}_t, \boldsymbol{\beta}_t, \mathcal{B}_t\}_{t=0}^{T-1}$. These parameters dynamically deform the 3D Gaussians into the observation space at each discrete time step t. Consequently, the resulting blurry image $\mathbf{I}^B$ can be expressed as:

$$\mathbf{I}^B = \frac{1}{T}\sum_{t=0}^{T-1} \mathcal{R}(\mathcal{W}(\{G_k(\mathbf{x})\}_{k=0}^{K-1}, \mathcal{S}_t), \mathbf{R}, \mathbf{K}), \tag{3}$$

where $\mathbf{x}$ denotes the coordinates of the 3D Gaussians, and $\mathbf{R}$ and $\mathbf{K}$ represent the camera's extrinsic and intrinsic parameters, respectively. The operator $\mathcal{R}$ denotes the rasterization process of the 3DGS model, while $\mathcal{W}$ represents the warping of 3D Gaussians from the canonical space to the observation space, governed by the SMPL parameters $\mathcal{S}_t$.

3.2. 3D Human Motion Model

As illustrated in Fig. 1, motion-induced blur introduces significant ambiguities in motion that are readily observable in the pixel space. To tackle this challenge, a 3D-aware human motion model is proposed to effectively estimate both sub-frame motion within each exposure period and inter-frame global motion across consecutive frames.

Sub-frame rigid sequential pose model. The pose parameters $\boldsymbol{\Theta}_t$, defined in the SMPL model as $\boldsymbol{\Theta}_t \in \mathbb{R}^{24\times 3}$, represent the rotations of 24 joints at each time step t, expressed in $\mathbf{SO}(3)$. Considering the inherent continuity of joint motion, De Boor–Cox formulation for B-splines [7, 32, 49, 58, 66] is employed to interpolate intermediate poses.

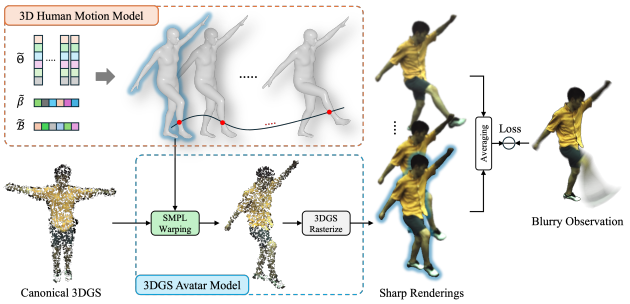

Figure 2. Brief illustration of the pipeline. The sub-frame motion for each blurry frame is modeled using the SMPL representation, followed by warping the canonical 3DGS according to the estimated motion parameters. The final blurry image is synthesized by averaging the sequence of rendered virtual sharp images.

For each joint j in the SMPL model, P control parameters are defined as $\tilde{\mathbf{\Theta}}^j = \{\tilde{\mathbf{\Theta}}^j_p\}_{p=0}^{P-1} \in \mathbb{R}^{P\times 3}$, where P denotes the predefined number of control knots. These learnable parameters are initialized from the coarse estimation, and optimized during the training process. At each time step t, the interpolation process begins with the computation of the timestep basis $\mathbf{B}(t) \in \mathbb{R}^{1\times P}$:

$$\mathbf{B}(t) = [1, \frac{t}{T}, (\frac{t}{T})^2, \ldots, (\frac{t}{T})^{P-1}], \tag{4}$$

where T denotes the total exposure time. Based on this basis, the interpolated pose parameters for the j-th joint at time step t are computed as:

$$\hat{\mathbf{\Theta}}^j_t = \mathbf{B}(t) \cdot \mathcal{M}^P \cdot \tilde{\mathbf{\Theta}}^j, \tag{5}$$

where $\mathcal{M}^P \in \mathbb{R}^{P\times P}$ is the interpolation matrix [58]:

$$\mathcal{M}^P_{i,j} = C^{P-1}_{P-1-i} \sum_{s=j}^{P-1} (-1)^{s-j} C^P_{s-j} (P-s-1)^{P-1-i}, \tag{6}$$

where $C^n_k = \frac{n!}{k!(n-k)!}$ denotes the binomial coefficient.

Pose deformation model. While B-spline interpolation captures basic pose trajectories, it remains limited in representing non-rigid, high-frequency pose variations. To overcome this, a pose displacement Δ^j_t is introduced for each joint j at every time step t within the exposure period:

$$\Delta^j_t = G_{disp}(\hat{\mathbf{\Theta}}^j_t; \theta_{disp}), \tag{7}$$

$$\mathbf{\Theta}^j_t = \hat{\mathbf{\Theta}}^j_t + \Delta^j_t, \tag{8}$$

where G_{disp} denotes the CNN designed to estimate fine-grained pose variations, and θ_{disp} represents its set of learnable parameters. This formulation allows the model to more accurately capture complex pose dynamics, restoring more realistic motion within each blurred frame.

Inter-frame motion regularization. Although the L1 loss can yield satisfactory photometric results at the midpoint of the exposure period ($t = 0.5$), it may suffer from directional ambiguity, as illustrated in Fig. 1(c). This ambiguity occurs because motion in either direction can produce a similarly plausible blurry image, potentially leading to inaccurate motion direction estimation and misaligned renderings at non-midpoint timesteps. To mitigate this issue, a regularization term is introduced based on the inter-frame continuity commonly observed in video sequences. This term measures the Geodesic distance between the pose parameters at the final timestep of the current exposure period and those at the initial timestep of the subsequent exposure period:

$$\mathcal{L}_{reg} = \frac{1}{24\cdot(N_e-1)} \sum_{n=0}^{N_e-2} \sum_{j=0}^{23} \left| \hat{\mathbf{\Theta}}^j_{n,T-1} - \hat{\mathbf{\Theta}}^j_{n+1,0} \right|_{\mathrm{G}}, \tag{9}$$

where $\hat{\mathbf{\Theta}}^j_{n,t}$ denotes the estimated pose parameters of the j-th joint at timestep t for the n-th exposure period, and N_e represents the total number of exposure periods. $|\cdot|_{\mathrm{G}}$ denotes the Geodesic distance. This regularization term is designed to enhance the inter-frame temporal coherence of joint movements, ensuring that the motion remains consistent and progresses naturally from one frame to the next.

Shape estimation. The shape parameters $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{10}$ of the SMPL model are initialized from the coarse estimation, then optimized as parameters during the training process.

Linear Blend Skinning (LBS) weights estimation. The SMPL model [41] provides pre-trained LBS weights, denoted as $\tilde{\boldsymbol{\mathcal{B}}}$. Consistent with prior studies, identical LBS weights $\hat{\boldsymbol{\mathcal{B}}}$ are employed across all time steps. Training begins with the initial pre-trained SMPL weights, which are further refined by introducing an LBS offset δ:

$$\hat{\boldsymbol{\mathcal{B}}} = \tilde{\boldsymbol{\mathcal{B}}} + \delta, \tag{10}$$

where δ is predicted by a simple CNN that takes the coordinates of all 3D Gaussians as input, enabling dynamic refinement of the LBS weights.

3.3. Optimization

Pipeline. A brief illustration for the optimization pipeline of the proposed model is provided in Fig. 2. The process begins with the estimation of sub-frame motion using the dedicated motion model described in Sec. 3.2. The SMPL parameters are initialized using the coarse estimation from blurry frames, and gradually optimized within the 3D-aware framework. The canonical 3D Gaussian Splatting (3DGS) representation is then transformed according to the estimated motion parameters. Subsequently, sharp virtual images are rendered for each intermediate timestep. These images are averaged to produce the final blurry image used for loss computation.

Loss function. The loss function includes the L1 loss between the synthesized blurry frame $\hat{\mathbf{I}}^B$ and the observed

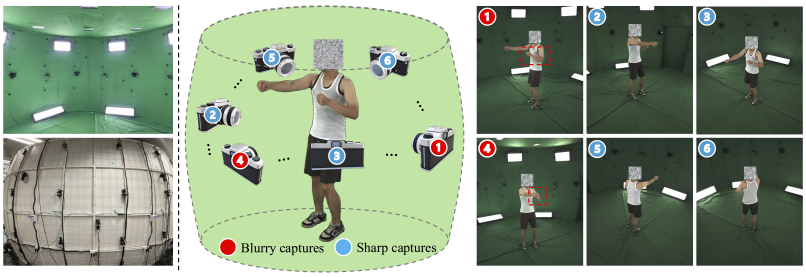

Figure 3. 360-degree hybrid-exposure camera system. Left: the inner side and outer side of the capture cage. Right: illustration of the system and samples of captures.

Capture type	Blur	Sharp
View number	4	8
Exposure	50 ms	3.125 ms
Resolution	2448 × 2048	
Camera model	BFS-U3-51S5C	
Scene number	8	
Picture format	PNG	

Table 1. Configuration of the captured real data.

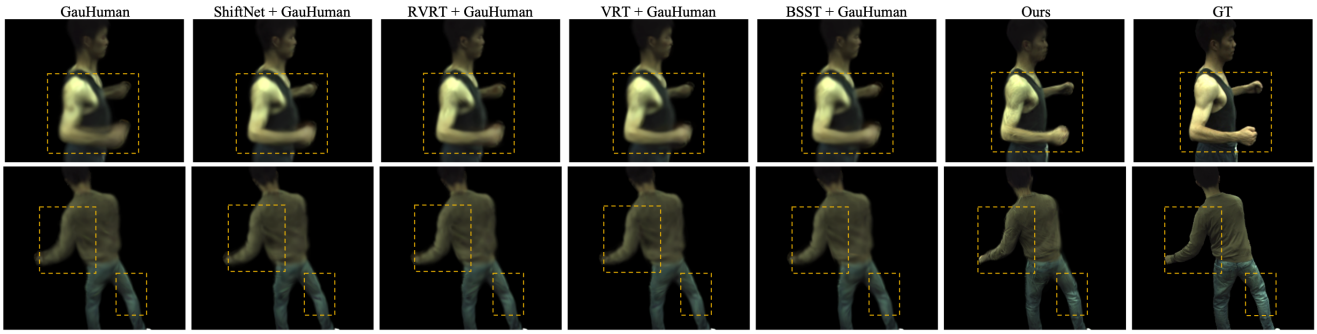

Figure 4. Qualitative comparison results on synthetic dataset. Zoom in for the best view.

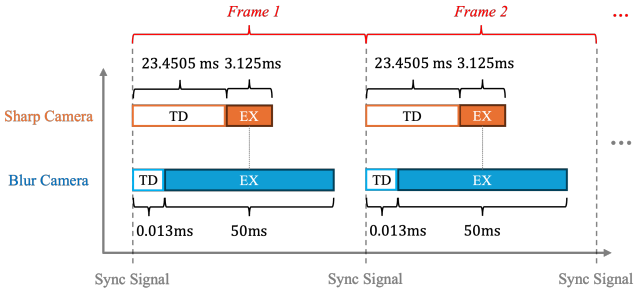

Figure 5. Time synchronization of the camera system. 'TD' and 'EX' stand for "Trigger Delay" and "Exposure".

blurry frame $\mathbf{I}^B$, and the inter-frame motion regularization loss as decribed in Sec. 3.2:

$$\mathcal{L} = ||\hat{\mathbf{I}}^B - \mathbf{I}^B||_1 + \mathcal{L}_{reg}. \tag{11}$$

The total loss enforces both accurate avatar reconstruction and the estimation of realistic, temporally coherent motion.

4. Experiments

4.1. Benchmarks

Synthetic dataset. The synthetic data is constructed based on the ZJU-MoCap [55] dataset since it is widely adopted by many 3D human avatar research work [10, 13, 22, 57, 72, 77]. Following their practice, six human subjects (IDs: 377, 386, 387, 392, 393, 394) are selected for experiments. Each blurry frame was synthesized by averaging K_{blur} consecutive sharp frames To mitigate discontinuities caused by direct averaging, the RIFE model [15] is employed to interpolate 16 intermediate frames between adjacent sharp frames. Blurry frames from Cameras 04, 10, 16, and 22 are used for training, while the sharp frames from the remaining 18 cameras are reserved for novel-view test. SMPL parameters are calibrated using the blurry video with EasyMoCap [55], and human masks are obtained using SAM [24].

Real dataset. The dataset consists of real-captured blurry videos and synchronously captured sharp videos from

Table 2. Quantitative comparison results on two datasets. We colorize results as best, second best, and third best.

Methods	Synthetic Dataset			Real Dataset		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
GauHuman	23.080	0.7660	0.2277	25.602	0.8044	0.2380
ShiftNet + GauHuman	23.089	0.7695	0.2219	25.549	0.8043	0.2347
RVRT + GauHuman	23.078	0.7697	0.2218	25.547	0.8065	0.2343
VRT + GauHuman	23.074	0.7696	0.2205	25.553	0.8067	0.2345
BSST + GauHuman	23.081	0.7698	0.2212	25.568	0.8068	0.2342
Ours	25.546	0.8290	0.1476	27.010	0.8271	0.1668

Table 3. Quantitative ablation results on two datasets. We colorize results as best, second best, and third best.

Models	Synthetic Dataset			Real Dataset		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
w/o interp.	24.009	0.8053	0.1620	25.825	0.8140	0.1729
w/o pose deform	25.301	0.8229	0.1545	26.426	0.8184	0.1743
w/o LBS opt.	25.394	0.8261	0.1486	26.821	0.8233	0.1697
w/o shape opt.	25.529	0.8284	0.1481	26.964	0.8261	0.1669
Ours (full model)	25.546	0.8290	0.1476	27.010	0.8271	0.1668

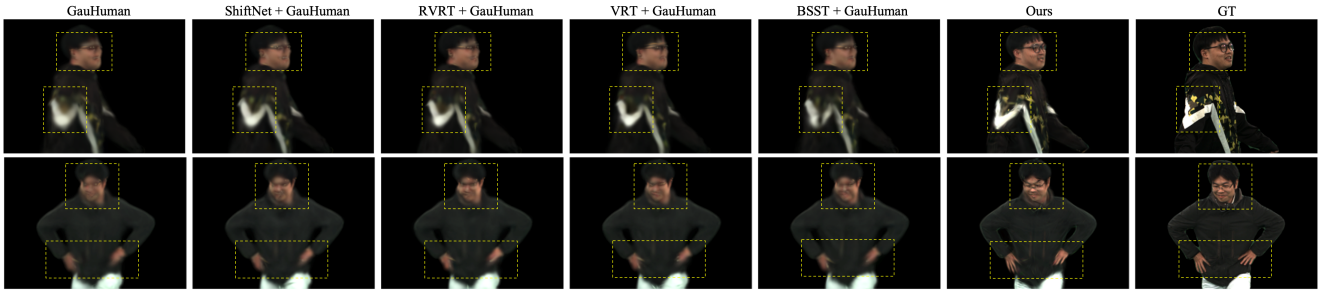

Figure 6. Qualitative comparison results on real data. Zoom in for the best view.

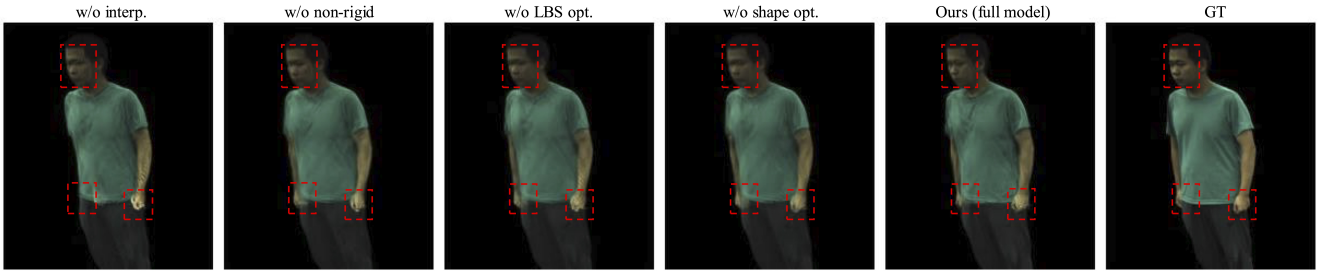

Figure 7. Qualitative results for ablation studies. Zoom in for the best view.

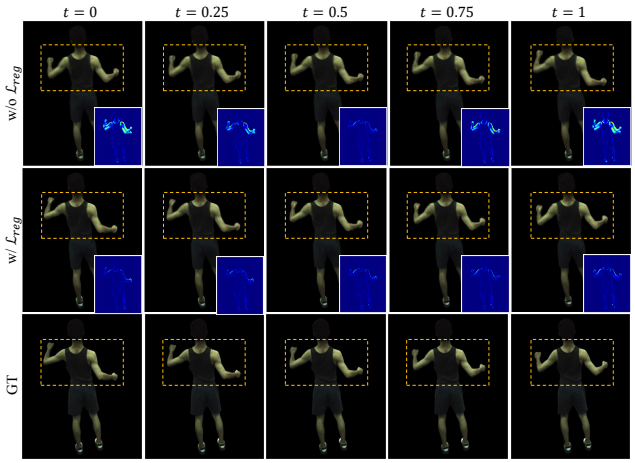

Figure 8. Qualitative ablation results for $\mathcal{L}_{reg}$. Model w/o $\mathcal{L}_{reg}$ gives worse result for non-middle timesteps due to incorrect motion direction estimates. Error maps are visualized for each output.

novel viewpoints. The capture system employs 12 time-synchronized Blackfly BFS-U3-51S5C cameras arranged in a 360-degree spatial configuration, as illustrated in Fig. 3. 4 cameras record blurry videos, while the remaining 8 capture sharp videos for novel-view evaluation. Fig. 5 illustrates the time synchronization of two types of cameras. To ensure consistent pixel intensity, the blurry cameras are equipped with a $1/16$ ND filter. Detailed configuration are reported in Tab. 1. SMPL parameters are calibrated using EasyMoCap, and human masks are generated using SAM [24].

Implementation details. Adam optimizer [23] is used with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rates and decay schedules follow the original 3DGS [21]. Following the protocol in [13], the input resolution is set to 512×512 for the synthetic dataset and 612×512 for the real dataset. All models are trained on a single NVIDIA GeForce RTX 4090 GPU. Since the images contain large black regions, each image is cropped to the human bounding box for quantitative evaluation.

Table 4. Quantitative ablation results for the inter-frame motion regularization loss $\mathcal{L}_{reg}$. We colorize better result of each sector.

K_{blur}	Models	Middle Timestep			Non-middle Timesteps		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
5	w/o $\mathcal{L}_{reg}$	25.567	0.8296	0.1478	24.421	0.8111	0.1601
	w/ $\mathcal{L}_{reg}$	25.546	0.8290	0.1476	25.417	0.8269	0.1485
7	w/o $\mathcal{L}_{reg}$	25.113	0.8197	0.1585	23.737	0.7950	0.1758
	w/ $\mathcal{L}_{reg}$	25.155	0.8200	0.1557	25.036	0.8179	0.1567
9	w/o $\mathcal{L}_{reg}$	24.628	0.8114	0.1680	23.198	0.7825	0.1888
	w/ $\mathcal{L}_{reg}$	24.680	0.8126	0.1636	24.605	0.8111	0.1645
11	w/o $\mathcal{L}_{reg}$	24.241	0.8009	0.1786	22.926	0.7717	0.1998
	w/ $\mathcal{L}_{reg}$	24.353	0.8039	0.1725	24.317	0.8031	0.1726

4.2. Comparisons

The proposed model is compared against several baseline approaches to assess its effectiveness. The simplest baseline involves training a SOTA avatar model (*e.g.*, GauHuman [13]) using blurry video frames. A two-stage baseline is then considered, where 2D deblurring methods (RVRT [36], ShiftNet [31], VRT [37], and BSST [83]) are first applied to restore sharp frames, then used to train the avatar model.

Quantitative results. Quantitative comparisons on two

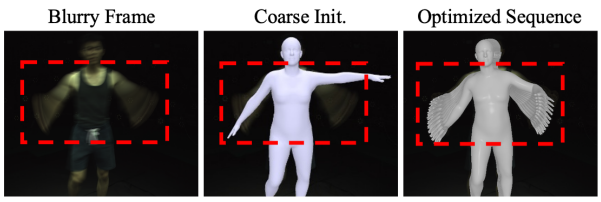

Figure 9. Visualization of the initial estimated SMPL by EasyMocap and optimized SMPL sequence by the proposed model.

Table 5. Qualitative ablation results on trajectories representations. We colorize result as best, second best, and third best.

Models	Synthetic Dataset			Real Dataset		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Linear	25.516	0.8289	0.1483	26.899	0.8257	0.1680
Slerp	25.539	0.8288	0.1481	27.013	0.8264	0.1674
B-Spline	25.546	0.8290	0.1476	27.010	0.8271	0.1668

Table 6. Ablation results for control knot number P.

P	Synthetic Dataset			Real Dataset		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
2	25.516	0.8289	0.1480	26.899	0.8257	0.1680
3	25.482	0.8284	0.1485	27.003	0.8266	0.1671
4	25.546	0.8290	0.1476	27.010	0.8271	0.1668
5	25.540	0.8288	0.1490	26.977	0.8260	0.1673

datasets are presented in Tab. 2. The two-stage baselines exhibit suboptimal performance, with noticeably lower quantitative metrics. This limitation arises because 2D deblurring models fail to maintain multi-view consistency across different viewpoints and lack explicit modeling of the scene's intrinsic 3D structure. They only slightly outperform the direct 3DGS avatar model, highlighting that inconsist 2D deblurring constrains the 3DGS model's ability to produce accurate reconstructions. In contrast, the proposed method outperforms baselines by explicitly modeling the 3D motion-induced blur formation process.

Qualitative results. Qualitative comparison results are shown in Fig. 4 and Fig. 6. While integrating 2D deblurring techniques with the 3DGS avatar model provides minor improvements in reconstruction quality (*e.g.*, the arm region in Fig. 4), prominent visual artifacts remain, including residual blur and loss of detail around body contours. These artifacts arise from inconsistencies in deblurring across multiple viewpoints, as 2D deblurring methods do not enforce multi-view consistency. The proposed approach incorporates a physics-based blur formation model, allowing accurate learning of the 3D representation and enabling simultaneous deblurring and improved avatar reconstruction.

4.3. Ablation Study

Model components. Ablation studies are conducted to evaluate the contributions of different model components.

- **w/o interp.**: Independently optimizing the pose at each timestep instead of using pose interpolation.
- **w/o pose deform**: Using only the rigid sequential pose model, excluding pose deformation model.
- **w/o LBS opt.**: Employing pre-trained LBS weights without additional optimization.
- **w/o shape opt.**: Directly using calibrated SMPL shape parameters without further refinement.

The quantitative and qualitative results are presented in Tab. 3 and Fig. 7. Independently optimizing poses for each intermediate timestep leads to unordered motion estimations (the case illustrated in Fig. 1 (b)), leading to misaligned deblurring results. Excluding the pose deformation model introduces additional artifacts, as B-spline interpolation alone cannot adequately capture complex human motion. The performance also degrades when LBS weights and SMPL shape parameters are not refined, emphasizing the necessity of these components.

Inter-frame regularization loss. Evaluations on $\mathcal{L}_{reg}$ are summarized in Tab. 4 and Fig. 8. It improves performance especially for non-middle timesteps by enabling more accurate estimations for inter-frame motion directions. Quantitative results on real data are excluded due to the lack of GT frames for non-middle timesteps.

Trajectory representations. Different trajectory representations are evaluated in the proposed model, including B-spline interpolation, linear interpolation, and spherical linear interpolation (Slerp). The corresponding results are presented in Tab. 5. It can be observed that B-spline interpolation generally achieves the best performance.

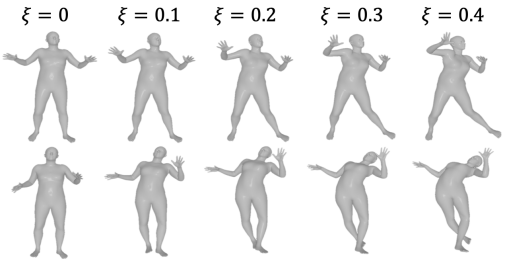

Figure 10. Visualizing SMPL under different perturbations.

Table 7. Ablation results for virtual sharp image number T.

T	Synthetic Dataset			Real Dataset		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
3	25.528	0.8276	0.1503	26.837	0.8253	0.1695
5	25.546	0.8290	0.1476	27.010	0.8271	0.1668
7	25.532	0.8281	0.1499	26.993	0.8269	0.1670
9	25.522	0.8279	0.1501	26.965	0.8261	0.1676

Table 8. Quantitative results with different perturbations.

ξ	0	0.1	0.2	0.3	0.4
PSNR	25.55	25.51	25.44	25.39	25.14
SSIM	.8290	.8284	.8271	.8258	.8225
LPIPS	.1476	.1476	.1493	.1504	.1524

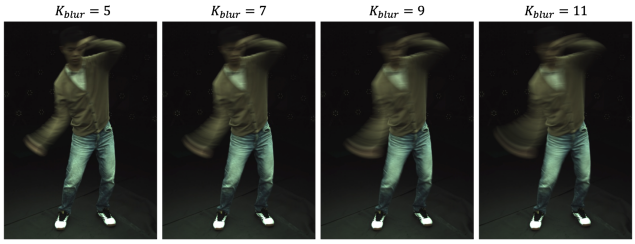

Figure 11. Visualization for different blur magnitudes K_{blur}.

Table 9. Quantitative result w/o optimizing SMPL parameters.

Methods	Synthetic Dataset			Real Dataset		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
w/o SMPL opt.	21.63	.7269	.2421	25.14	.7855	.2423
Ours	25.55	.8290	.1476	27.01	.8271	.1668

Table 10. Quantitative comparisons with different K_{blur}.

K_{blur}	Metrics	GauH	ShiftNet + GauH	RVRT + GauH	VRT + GauH	BSST + GauH	Ours
7	PSNR ↑	22.983	23.003	22.992	22.972	23.009	25.155
	SSIM ↑	0.7599	0.7648	0.7650	0.7654	0.7652	0.8200
	LPIPS ↓	0.2378	0.2314	0.2301	0.2299	0.2294	0.1557
9	PSNR ↑	22.693	22.699	22.692	22.693	22.732	24.680
	SSIM ↑	0.7532	0.7585	0.7587	0.7588	0.7590	0.8126
	LPIPS ↓	0.2497	0.2426	0.2411	0.2412	0.2409	0.1636
11	PSNR ↑	22.635	22.632	22.644	22.647	22.652	24.353
	SSIM ↑	0.7549	0.7554	0.7555	0.7564	0.7557	0.8039
	LPIPS ↓	0.2552	0.2502	0.2477	0.2480	0.2482	0.1725

Hyperparameters. The proposed model is evaluated with varying numbers of control knots, $P = 2, 3, 4, 5$, for B-spline sequential pose interpolation. As reported in Tab. 6, the best performance is obtained with $P = 4$. However, the performance differences remain relatively small, as the pose deformation model enhances robustness for smaller P values by effectively modeling complex motions, thereby reducing performance gaps across different settings of P. The influence of virtual sharp image number T is examined in Tab. 7. Optimal performance is achieved when $T = 5$.

4.4. Robustness and Generality Evaluation

Influence of the accuracy of SMPL estimation. SMPL estimations from blurry data could be inaccurate. However, the proposed approach does not depend on precise SMPL estimates but only coarse initializations. Fig. 9 visualizes the initial and optimized SMPL parameters, showing that the method refines inaccurate initializations and recover accurate sub-frame SMPL poses. To further assess robustness against inaccurate initialization, random perturbations $\epsilon \sim U[-\xi, \xi]$ are added to the initializations. Fig. 10 visualizes the effects of varying ξ values. The quantitative results in Tab. 8 confirm the method's resilience to even large perturbations. Finally, experiments are conducted without optimizing SMPL (*pose, shape, LBS*), using only the initial estimates. The results in Tab. 9 proves the necessity of SMPL optimization.

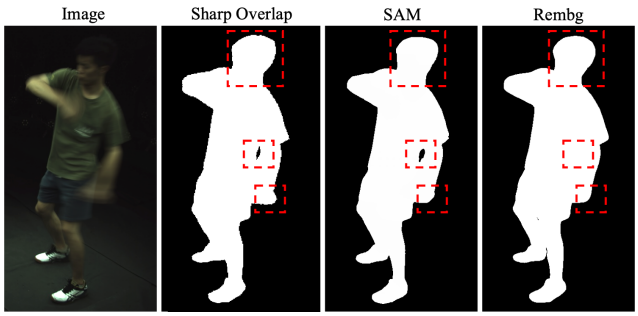

Figure 12. Visualization of different mask parsing results.

Table 11. Quantitative ablation results for different K_{blur}.

K_{blur}	Metrics	w/o interp.	w/o pose deform	w/o LBS opt.	w/o shape opt.	Ours (full model)
7	PSNR ↑	23.611	24.813	25.050	25.153	25.155
	SSIM ↑	0.7962	0.8127	0.8180	0.8199	0.8200
	LPIPS ↓	0.1710	0.1640	0.1563	0.1561	0.1557
9	PSNR ↑	22.800	24.239	24.596	24.683	24.680
	SSIM ↑	0.7811	0.8035	0.8109	0.8127	0.8126
	LPIPS ↓	0.1850	0.1734	0.1638	0.1639	0.1636
11	PSNR ↑	22.350	23.754	24.282	24.345	24.353
	SSIM ↑	0.7695	0.7924	0.8022	0.8037	0.8039
	LPIPS ↓	0.1945	0.1852	0.1725	0.1726	0.1725

Table 12. Quantitative results with different mask estimations.

Mask	PSNR ↑	SSIM ↑	LPIPS ↓
sharp	25.61	.8309	.1465
SAM	25.55	.8290	.1476
rembg	25.49	.8281	.1488

Different blur magnitudes K_{blur}. To evaluate the robustness of the proposed method under varying blur intensities, additional datasets were synthesized with blur magnitudes $K_{blur} = 7, 9, 11$. Visualization of different blur magnitudes K_{blur} is provided in Fig. 11. As reported in Tab. 10, the proposed method consistently outperforms all baseline approaches across varying blur level. Furthermore, corresponding ablation results in Tab. 11 confirm the contribution of each model component under different blur intensities.

Different view numbers N_{view}. To assess robustness with respect to the view numbers, the proposed model is evaluated using varying numbers of training views $N_{view} = 2, 3, 4$. As shown in Tab. 13, the proposed method consistently outperforms baseline approaches with different N_{view}, demonstrating its strong generality. Furthermore, the ablation results under different N_{view} in Tab. 14 further confirm the effectiveness of each component.

Mask estimation. Additional experiments are conducted to evaluate the model's robustness to inaccurate masks due to blurriness. Since the synthetic dataset is generated from sharp frames, an "ideal" case is defined in which the masks are obtained by overlapping estimated masks from the cor-

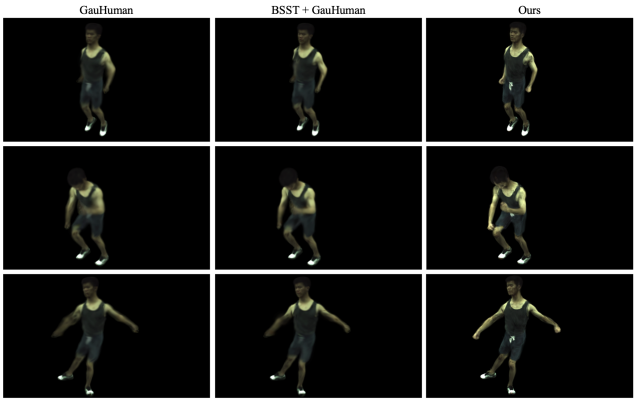

Figure 13. Novel pose animation results using out-of-domain motion from the AMASS [44] dataset.

Table 13. Quantitative comparisons with different N_{view}.

N_{view}	Metrics	GauH	ShiftNet + GauH	RVRT + GauH	VRT + GauH	BSST + GauH	Ours
2	PSNR ↑	21.844	21.805	21.805	21.799	21.801	23.331
	SSIM ↑	0.7501	0.7533	0.7533	0.7534	0.7532	0.7712
	LPIPS ↓	0.2506	0.2436	0.2431	0.2435	0.2431	0.1891
3	PSNR ↑	22.514	22.580	22.594	22.574	22.590	24.979
	SSIM ↑	0.7545	0.7584	0.7580	0.7580	0.7581	0.8134
	LPIPS ↓	0.2317	0.2225	0.2219	0.2225	0.2229	0.1587
4	PSNR ↑	23.080	23.089	23.078	23.074	23.081	25.546
	SSIM ↑	0.7660	0.7695	0.7697	0.7696	0.7698	0.8290
	LPIPS ↓	0.2277	0.2219	0.2218	0.2205	0.2212	0.1476

responding sharp frames. Though such masks cannot be accessed in applications, by comparing to this idealized setting, the robustness of the model to mask estimation can be evaluated. Furthermore, alternative mask tools like *rembg* are adapted. As summarized in Tab. 12, the proposed method performs comparably to the ideal case, and the variation across different mask models remains minimal. Fig. 12 visualizes different mask estimations.

Novel-pose evaluation. Novel poses from the AMASS dataset [44] are used to animate the reconstructed avatars for qualitative comparison. Fig. 13 shows that the proposed model achieves the best result.

4.5. Further Discussions

4.5.1. Demo captures from smartphones

To further demonstrate the generality of the proposed method in real-world scenarios, monocular video is captured using an iPhone 16 Pro, and SMPL parameters are estimated with TRAM [71], an open-sourced human motion estimation model. Fig. 14 confirm that the method generalizes well to real-world data and exhibits strong robustness to monocular SMPL estimations.

4.5.2. Limitations

While the proposed method successfully estimates motion and generates sharp 3DGS avatars, it has certain limitations.

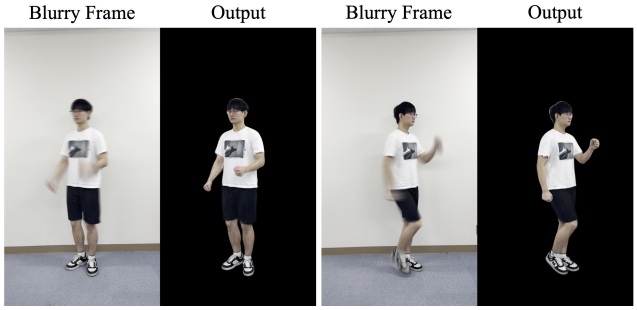

Figure 14. Qualitative results on captures from an iPhone 16 Pro.

Table 14. Quantitative ablation results with different N_{view}.

N_{view}	Metrics	w/o interp.	w/o pose deform	w/o LBS opt.	w/o shape opt.	Ours
2	PSNR ↑	22.661	23.310	23.321	23.411	23.331
	SSIM ↑	0.7584	0.7704	0.7718	0.7737	0.7712
	LPIPS ↓	0.1978	0.1919	0.1896	0.1893	0.1891
3	PSNR ↑	23.927	24.808	24.899	24.921	24.979
	SSIM ↑	0.7973	0.8081	0.8115	0.8129	0.8134
	LPIPS ↓	0.1683	0.1656	0.1594	0.1589	0.1587
4	PSNR ↑	24.009	25.301	25.394	25.529	25.546
	SSIM ↑	0.8053	0.8229	0.8261	0.8284	0.8290
	LPIPS ↓	0.1620	0.1545	0.1486	0.1481	0.1476

For example, since the model is built upon 3DGS representation, it struggles to accurately recover the geometry of the person, *e.g.*, the surface normal or the BRDF. In addition, real-world blur occurs as light integrates in the linear radiance space before the ISP process. Averaging non-linear sRGB values directly can lead to inaccuracies, particularly in high-contrast regions where linear summation is essential for physical fidelity. Future work aims to explore these aspects. Also, the proposed model relies on SMPL to represent sub-frame motion for 3D-aware deblurring. However, when the subject has handheld objects or wears loose garments, it fails to recover the motion of these accessories because SMPL does not have corresponding "joints". Future work could explore adaptive joints [82] with non-rigid modeling under blurriness to support more versatile avatar deblurring.

5. Conclusion

In this paper, a novel approach is presented for reconstructing sharp 3D human avatars from blurry video frames. By extending the traditional 2D motion blur formulation into a 3D-aware blur formation model, the proposed method jointly optimizes sub-frame motion representations while learning a canonical 3DGS human avatar model. To establish a benchmark for this task, both a synthetic dataset derived from the ZJU-MoCap and a real-captured dataset are provided. Extensive evaluations demonstrate that the proposed model consistently outperforms existing baselines.

Acknowledgment

This research was supported in part by the Tateisi Science and Technology Foundation, JSPS KAKENHI Grant Numbers 24KK0209 and 22H00529, and Advanced AI Talent Development to Lead the Next-Generation AI for Intelligent Society (BOOST NAIS) of The University of Tokyo.

References

[1] Minh-Quan Viet Bui, Jongmin Park, Jihyong Oh, and Munchurl Kim. Moblurf: Motion deblurring neural radiance fields for blurry monocular video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[2] Mingdeng Cao, Yanbo Fan, Yong Zhang, Jue Wang, and Yujiu Yang. Vdtr: Video deblurring with transformer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):160–171, 2022.

[3] Mingdeng Cao, Zhihang Zhong, Yanbo Fan, Jiahao Wang, Yong Zhang, Jue Wang, Yujiu Yang, and Yinqiang Zheng. Towards real-world video deblurring by exploring blur formation process. In *European Conference on Computer Vision*, pages 327–343. Springer, 2022.

[4] Sihun Cha, Kwanggyoon Seo, Amirsaman Ashtari, and Junyong Noh. Generating texture for 3d human avatar from a single image using sampling and refinement networks. In *Computer graphics forum*, pages 385–396. Wiley Online Library, 2023.

[5] Yutong Chen, Yifan Zhan, Zhihang Zhong, Wei Wang, Xiao Sun, Yu Qiao, and Yinqiang Zheng. Within the dynamic context: Inertia-aware 3d human modeling with pose sequence. In *European Conference on Computer Vision*, pages 491–508. Springer, 2024.

[6] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017.

[7] Gerald E Farin. *Curves and surfaces for CAGD: a practical guide*. Morgan Kaufmann, 2002.

[8] Yuanyuan Gao, Hao Li, Jiaqi Chen, Zhengyu Zou, Zhihang Zhong, Dingwen Zhang, Xiao Sun, and Junwei Han. Citygs-x: A scalable architecture for efficient and geometrically accurate large-scale scene reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 27187–27196, 2025.

[9] Yuanyuan Gao, Yuning Gong, Yifei Liu, Li Jingfeng, Dingwen Zhang, Yanci Zhang, Dan Xu, Xiao Sun, and Zhihang Zhong. Proxy-gs: Unified occlusion priors for training and inference in structured 3d gaussian splatting, 2026.

[10] Chen Geng, Sida Peng, Zhen Xu, Hujun Bao, and Xiaowei Zhou. Learning neural volumetric representations of dynamic humans in minutes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8759–8770, 2023.

[11] Chen Guo, Tianjian Jiang, Xu Chen, Jie Song, and Otmar Hilliges. Vid2avatar: 3d avatar reconstruction from videos in the wild via self-supervised scene decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12858–12868, 2023.

[12] Liangxiao Hu, Hongwen Zhang, Yuxiang Zhang, Boyao Zhou, Boning Liu, Shengping Zhang, and Liqiang Nie. Gaussianavatar: Towards realistic human avatar modeling from a single video via animatable 3d gaussians. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 634–644, 2024.

[13] Shoukang Hu, Tao Hu, and Ziwei Liu. Gauhuman: Articulated gaussian splatting from monocular human videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20418–20431, 2024.

[14] Yangyi Huang, Hongwei Yi, Yuliang Xiu, Tingting Liao, Jiaxiang Tang, Deng Cai, and Justus Thies. Tech: Text-guided reconstruction of lifelike clothed humans. In *2024 International Conference on 3D Vision (3DV)*, pages 1531–1542. IEEE, 2024.

[15] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In *European conference on computer vision*, pages 624–642. Springer, 2022.

[16] Tae Hyun Kim and Kyoung Mu Lee. Generalized video deblurring for dynamic scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5426–5434, 2015.

[17] Tae Hyun Kim, Kyoung Mu Lee, Bernhard Scholkopf, and Michael Hirsch. Online video deblurring via dynamic temporal blending network. In *Proceedings of the IEEE international conference on computer vision*, pages 4038–4047, 2017.

[18] Rohit Jena, Ganesh Subramanian Iyer, Siddharth Choudhary, Brandon Smith, Pratik Chaudhari, and James Gee. Splatarmor: Articulated gaussian splatting for animatable humans from monocular rgb videos. *arXiv preprint arXiv:2311.10812*, 2023.

[19] Tianjian Jiang, Xu Chen, Jie Song, and Otmar Hilliges. Instantavatar: Learning avatars from monocular video in 60 seconds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16922–16932, 2023.

[20] Wei Jiang, Kwang Moo Yi, Golnoosh Samei, Oncel Tuzel, and Anurag Ranjan. Neuman: Neural human radiance field from a single video. In *European Conference on Computer Vision*, pages 402–418. Springer, 2022.

[21] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

[22] Jaehyeok Kim, Dongyoon Wee, and Dan Xu. Motion-oriented compositional neural radiance fields for monocular dynamic human modeling. In *European Conference on Computer Vision*, pages 476–493. Springer, 2024.

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

[24] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer White-

head, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.

[25] Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, and Anurag Ranjan. Hugs: Human gaussian splats. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 505–515, 2024.

[26] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018.

[27] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8878–8887, 2019.

[28] Dogyoon Lee, Minhyeok Lee, Chajin Shin, and Sangyoun Lee. Dp-nerf: Deblurred neural radiance field with physical scene priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12386–12396, 2023.

[29] Jiahui Lei, Yufu Wang, Georgios Pavlakos, Lingjie Liu, and Kostas Daniilidis. Gart: Gaussian articulated template models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 19876–19887, 2024.

[30] Anat Levin. Blind motion deblurring using image statistics. *Advances in neural information processing systems*, 19, 2006.

[31] Dasong Li, Xiaoyu Shi, Yi Zhang, Ka Chun Cheung, Simon See, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. A simple baseline for video restoration with grouped spatial-temporal shift. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9822–9832, 2023.

[32] Moyang Li, Peng Wang, Lingzhe Zhao, Bangyan Liao, and Peidong Liu. Usb-nerf: Unrolling shutter bundle adjusted neural radiance fields. *arXiv preprint arXiv:2310.02687*, 2023.

[33] Ruilong Li, Julian Tanke, Minh Vo, Michael Zollhöfer, Jürgen Gall, Angjoo Kanazawa, and Christoph Lassner. Tava: Template-free animatable volumetric actors. In *European Conference on Computer Vision*, pages 419–436. Springer, 2022.

[34] Zhe Li, Zerong Zheng, Lizhen Wang, and Yebin Liu. Animatable gaussians: Learning pose-dependent gaussian maps for high-fidelity human avatar modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 19711–19722, 2024.

[35] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021.

[36] Jingyun Liang, Yuchen Fan, Xiaoyu Xiang, Rakesh Ranjan, Eddy Ilg, Simon Green, Jiezhang Cao, Kai Zhang, Radu Timofte, and Luc V Gool. Recurrent video restoration transformer with guided deformable attention. *Advances in Neural Information Processing Systems*, 35:378–393, 2022.

[37] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *IEEE Transactions on Image Processing*, 33:2171–2182, 2024.

[38] Tingting Liao, Xiaomei Zhang, Yuliang Xiu, Hongwei Yi, Xudong Liu, Guo-Jun Qi, Yong Zhang, Xuan Wang, Xiangyu Zhu, and Zhen Lei. High-fidelity clothed avatar reconstruction from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8662–8672, 2023.

[39] Yang Liu, Xiang Huang, Minghan Qin, Qinwei Lin, and Haoqian Wang. Animatable 3d gaussian: Fast and high-quality reconstruction of multiple human avatars. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 1120–1129, 2024.

[40] Yifei Liu, Zhihang Zhong, Yifan Zhan, Sheng Xu, and Xiao Sun. Maskgaussian: Adaptive 3d gaussian representation from probabilistic masks. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 681–690, 2025.

[41] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 851–866. 2023.

[42] Yiren Lu, Yunlai Zhou, Disheng Liu, Tuo Liang, and Yu Yin. Bard-gs: Blur-aware reconstruction of dynamic scenes via gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 16532–16542, 2025.

[43] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V Sander. Deblur-nerf: Neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12861–12870, 2022.

[44] Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. Amass: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5442–5451, 2019.

[45] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[46] Arthur Moreau, Jifei Song, Helisa Dhamo, Richard Shaw, Yiren Zhou, and Eduardo Pérez-Pellitero. Human gaussian splatting: Real-time rendering of animatable avatars. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 788–798, 2024.

[47] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017.

[48] Seungjun Nah, Sanghyun Son, and Kyoung Mu Lee. Recurrent neural networks with intra-frame iterations for video

deblurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8102–8111, 2019.

[49] Muyao Niu, Tong Chen, Yifan Zhan, Zhuoxiao Li, Xiang Ji, and Yinqiang Zheng. Rs-nerf: Neural radiance fields from rolling shutter images. In *European Conference on Computer Vision*, pages 163–180. Springer, 2024.

[50] Muyao Niu, Mingdeng Cao, Yifan Zhan, Qingtian Zhu, Mingze Ma, Jiancheng Zhao, Yanhong Zeng, Zhihang Zhong, Xiao Sun, and Yinqiang Zheng. Anicrafter: Customizing realistic human-centric animation via avatar-background conditioning in video diffusion models. *arXiv preprint arXiv:2505.20255*, 2025.

[51] Atsuhiro Noguchi, Xiao Sun, Stephen Lin, and Tatsuya Harada. Neural articulated radiance field. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5762–5772, 2021.

[52] Jinshan Pan, Haoran Bai, and Jinhui Tang. Cascaded deep video deblurring using temporal sharpness prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3043–3051, 2020.

[53] Jinshan Pan, Boming Xu, Jiangxin Dong, Jianjun Ge, and Jinhui Tang. Deep discriminative spatial and temporal network for efficient video deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22191–22200, 2023.

[54] Sida Peng, Junting Dong, Qianqian Wang, Shangzhan Zhang, Qing Shuai, Xiaowei Zhou, and Hujun Bao. Animatable neural radiance fields for modeling dynamic human bodies. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14314–14323, 2021.

[55] Sida Peng, Yuanqing Zhang, Yinghao Xu, Qianqian Wang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9054–9063, 2021.

[56] Sida Peng, Zhen Xu, Junting Dong, Qianqian Wang, Shangzhan Zhang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Animatable implicit neural representations for creating realistic avatars from videos, 2023.

[57] Zhiyin Qian, Shaofei Wang, Marko Mihajlovic, Andreas Geiger, and Siyu Tang. 3dgs-avatar: Animatable avatars via deformable 3d gaussian splatting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5020–5030, 2024.

[58] Kaihuai Qin. General matrix representations for b-splines. In *Proceedings Pacific Graphics' 98. Sixth Pacific Conference on Computer Graphics and Applications (Cat. No. 98EX208)*, pages 37–43. IEEE, 1998.

[59] Wenqi Ren, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Video deblurring via semantic segmentation and pixel-wise non-linear kernel. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1077–1085, 2017.

[60] Hyeongseok Son, Junyong Lee, Jonghyeop Lee, Sunghyun Cho, and Seungyong Lee. Recurrent video deblurring with blur-invariant motion estimation and pixel volumes. *ACM Transactions on Graphics (TOG)*, 40(5):1–18, 2021.

[61] Colton Stearns, Adam Harley, Mikaela Uy, Florian Dubost, Federico Tombari, Gordon Wetzstein, and Leonidas Guibas. Dynamic gaussian marbles for novel view synthesis of casual monocular videos. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024.

[62] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1279–1288, 2017.

[63] Shih-Yang Su, Frank Yu, Michael Zollhöfer, and Helge Rhodin. A-nerf: Articulated neural radiance fields for learning human shape, appearance, and pose. *Advances in neural information processing systems*, 34:12278–12291, 2021.

[64] Huiqiang Sun, Xingyi Li, Liao Shen, Xinyi Ye, Ke Xian, and Zhiguo Cao. Dyblurf: Dynamic neural radiance fields from blurry monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7517–7527, 2024.

[65] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8174–8182, 2018.

[66] Michael Unser, Akram Aldroubi, and Murray Eden. B-spline signal processing. i. theory. *IEEE transactions on signal processing*, 41(2):821–833, 2002.

[67] Peng Wang, Lingzhe Zhao, Ruijie Ma, and Peidong Liu. Bad-nerf: Bundle adjusted deblur neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4170–4179, 2023.

[68] Shaofei Wang, Katja Schwarz, Andreas Geiger, and Siyu Tang. Arah: Animatable volume rendering of articulated human sdfs. In *European conference on computer vision*, pages 1–19. Springer, 2022.

[69] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019.

[70] Yusheng Wang, Yunfan Lu, Ye Gao, Lin Wang, Zhihang Zhong, Yinqiang Zheng, and Atsushi Yamashita. Efficient video deblurring guided by motion magnitude. In *European Conference on Computer Vision*, pages 413–429. Springer, 2022.

[71] Yufu Wang, Ziyun Wang, Lingjie Liu, and Kostas Daniilidis. Tram: Global trajectory and motion of 3d humans from in-the-wild videos. In *European Conference on Computer Vision*, pages 467–487. Springer, 2024.

[72] Chung-Yi Weng, Brian Curless, Pratul P Srinivasan, Jonathan T Barron, and Ira Kemelmacher-Shlizerman. Humannerf: Free-viewpoint rendering of moving people from monocular video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern Recognition*, pages 16210–16220, 2022.

[73] Jonas Wulff and Michael Julian Black. Modeling blurred video with layers. In *European conference on computer vision*, pages 236–252. Springer, 2014.

[74] Hongyi Xu, Thiemo Alldieck, and Cristian Sminchisescu. H-nerf: Neural radiance fields for rendering and temporal reconstruction of humans in motion. *Advances in Neural Information Processing Systems*, 34:14955–14966, 2021.

[75] Wangze Xu, Yifan Zhan, Zhihang Zhong, and Xiao Sun. Sequential gaussian avatars with hierarchical motion context. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13592–13603, 2025.

[76] Keyang Ye, Tianjia Shao, and Kun Zhou. Animatable 3d gaussians for high-fidelity synthesis of human motions. *arXiv preprint arXiv:2311.13404*, 2023.

[77] Xiaoting Yin, Hao Shi, Kailun Yang, Jiajun Zhai, Shangwei Guo, Lin Wang, and Kaiwei Wang. Event-guided 3d gaussian splatting for dynamic human and scene reconstruction. *arXiv preprint arXiv:2509.18566*, 2025.

[78] Zhengming Yu, Wei Cheng, Xian Liu, Wayne Wu, and Kwan-Yee Lin. Monohuman: Animatable human neural field from monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16943–16953, 2023.

[79] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021.

[80] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022.

[81] Yifan Zhan, Wangze Xu, Qingtian Zhu, Muyao Niu, Mingze Ma, Yifei Liu, Zhihang Zhong, Xiao Sun, and Yinqiang Zheng. R3-avatar: Record and retrieve temporal codebook for reconstructing photorealistic human avatars. *arXiv preprint arXiv:2503.12751*, 2025.

[82] Yifan Zhan, Qingtian Zhu, Muyao Niu, Mingze Ma, Jiancheng Zhao, Zhihang Zhong, Xiao Sun, Yu Qiao, and Yinqiang Zheng. Towards explicit exoskeleton for the reconstruction of complicated 3d human avatars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14259–14269, 2025.

[83] Huicong Zhang, Haozhe Xie, and Hongxun Yao. Blur-aware spatio-temporal sparse transformer for video deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2673–2681, 2024.

[84] Jingbo Zhang, Xiaoyu Li, Qi Zhang, Yanpei Cao, Ying Shan, and Jing Liao. Humanref: Single image to 3d human generation via reference-guided diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1844–1854, 2024.

[85] Lingzhe Zhao, Peng Wang, and Peidong Liu. Bad-gaussians: Bundle adjusted deblur gaussian splatting. In *European Conference on Computer Vision*, pages 233–250. Springer, 2024.

[86] Zhihang Zhong, Ye Gao, Yinqiang Zheng, and Bo Zheng. Efficient spatio-temporal recurrent neural network for video deblurring. In *European conference on computer vision*, pages 191–207. Springer, 2020.

[87] Zhihang Zhong, Yinqiang Zheng, and Imari Sato. Towards rolling shutter correction and deblurring in dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9219–9228, 2021.

[88] Zhihang Zhong, Xiao Sun, Zhirong Wu, Yinqiang Zheng, Stephen Lin, and Imari Sato. Animation from blur: Multimodal blur decomposition with motion guidance. In *European Conference on Computer Vision*, pages 599–615. Springer, 2022.

[89] Zhihang Zhong, Mingdeng Cao, Xiang Ji, Yinqiang Zheng, and Imari Sato. Blur interpolation transformer for real-world motion from blur. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5713–5723, 2023.

[90] Zhihang Zhong, Ye Gao, Yinqiang Zheng, Bo Zheng, and Imari Sato. Real-world video deblurring: A benchmark dataset and an efficient recurrent neural network. *International Journal of Computer Vision*, 131(1):284–301, 2023.

[91] Shangchen Zhou, Jiawei Zhang, Jinshan Pan, Haozhe Xie, Wangmeng Zuo, and Jimmy Ren. Spatio-temporal filter adaptive network for video deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2482–2491, 2019.

[92] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9308–9316, 2019.

[93] Wojciech Zielonka, Timur Bagautdinov, Shunsuke Saito, Michael Zollhöfer, Justus Thies, and Javier Romero. Drivable 3d gaussian avatars. In *2025 International Conference on 3D Vision (3DV)*, pages 979–990. IEEE, 2025.