Distribution and Depth-Aware Transformers for 3D Human Mesh Recovery

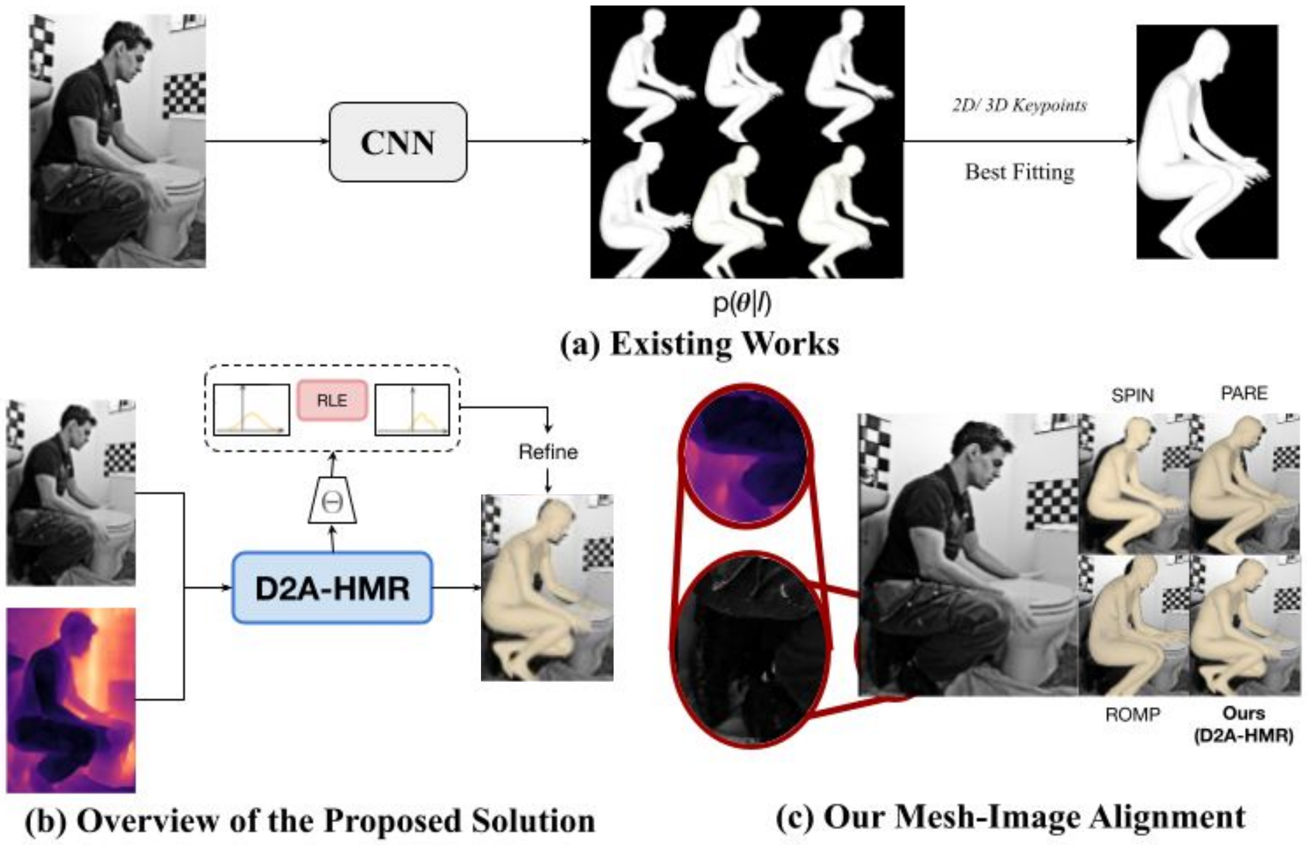
Jerrin Bright, Bavesh Balaji, Harish Prakash, Yuhao Chen, David Clausi, John Zelek
Vision and Image Processing Lab
Systems Design Engineering Department
University of Waterloo
Waterloo, Ontario, Canada

Overview

Objective:

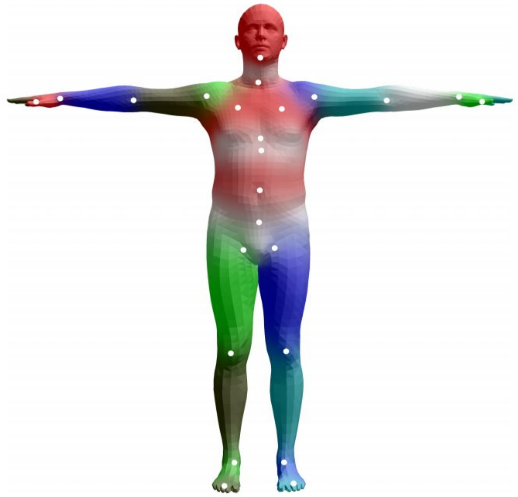
- ❖ Generalizable 3D human modeling from monocular images.
- ❖ Robust alignment in diverse conditions.

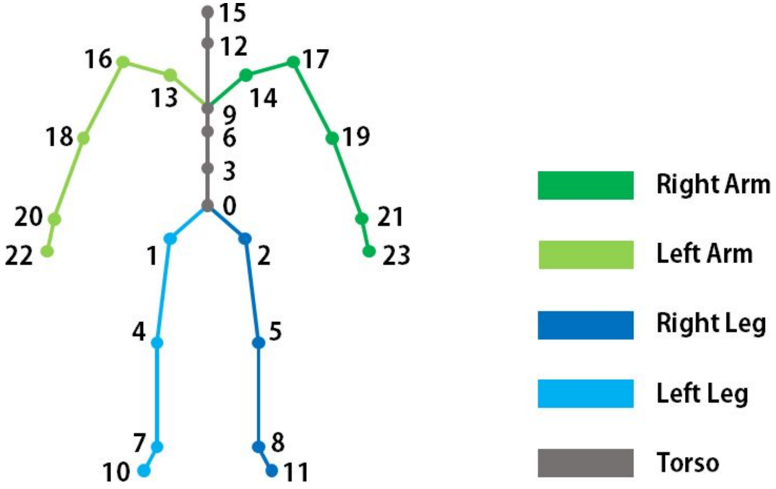
Model the depth and distribution

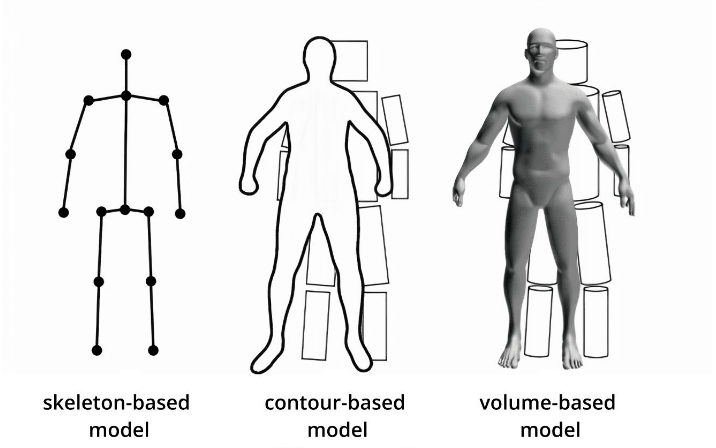


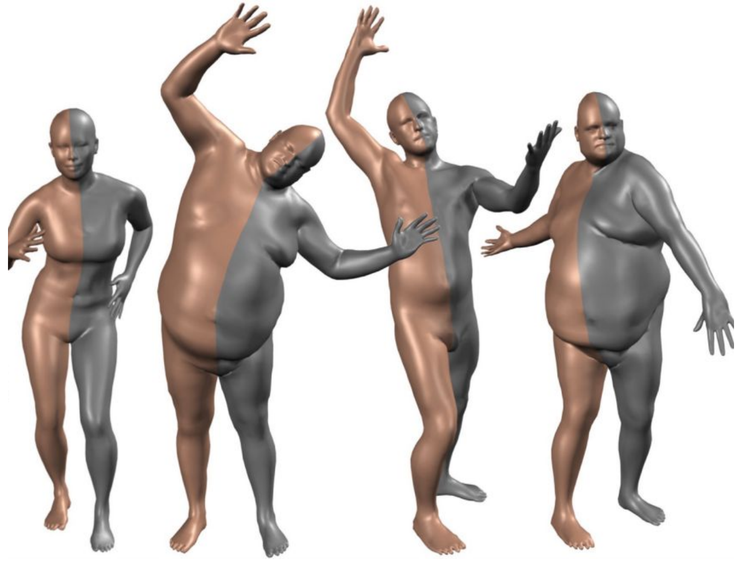
Background - SMPL

- ❖ Skinned Multi-Person Linear model [1].
- ❖ 72 joint and 10 shape parameters -> 6890 vertices.



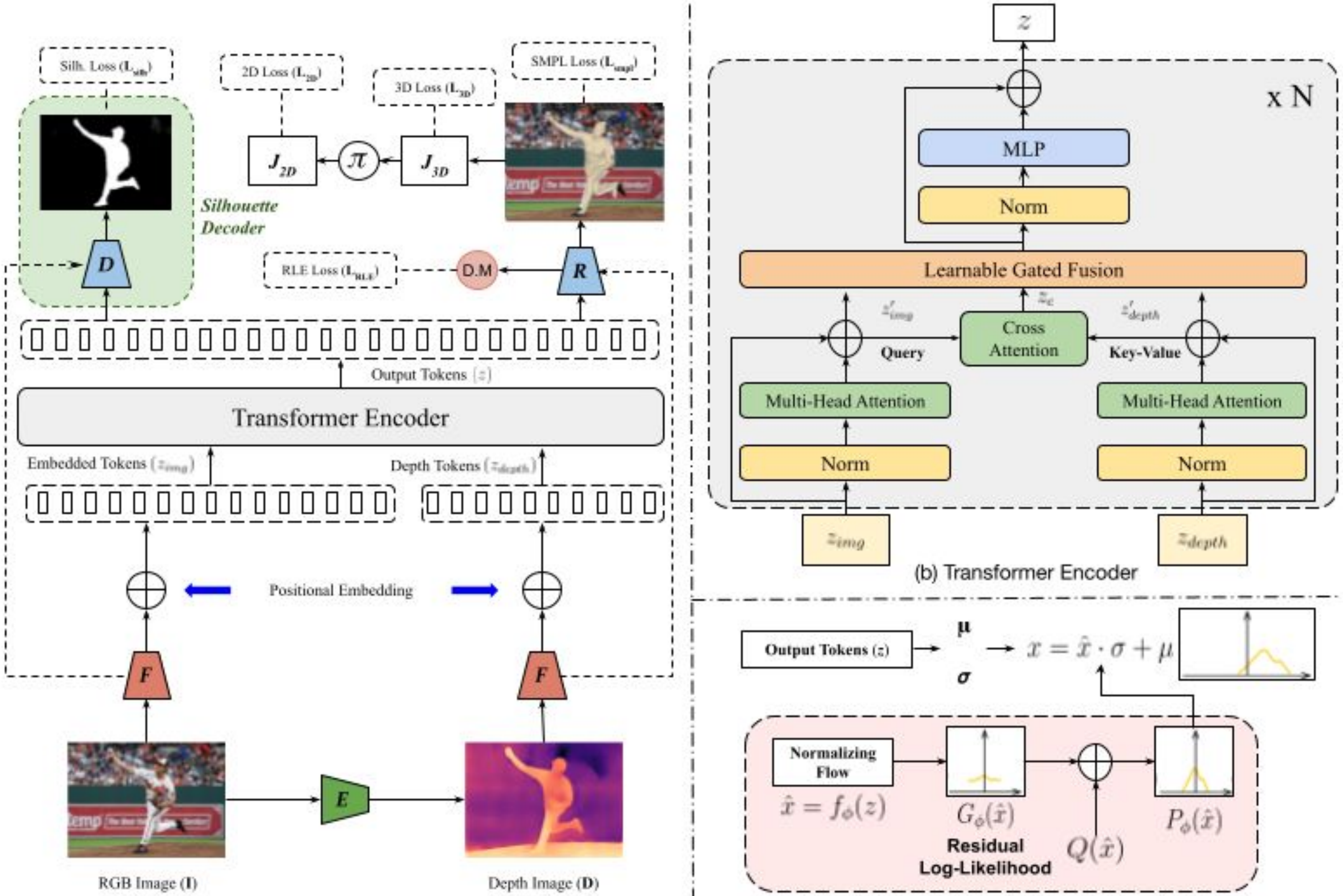






[1] Loper, Matthew and Mahmood, Naureen and Romero, Javier and Pons-Moll, Gerard and Black, Michael J. *SMPL: A Skinned Multi-Person Linear Model*. ACM Trans. Graphics (Proc. SIGGRAPH Asia), 2015

Our Work



(a) Proposed D2A-HMR framework

(b) Transformer Encoder

(c) Distribution Matching (D.M)

Objective Functions

- **Distribution Loss:**

$$\mathcal{L}_{RLE} = -\log Q(\bar{\mu}_g) - \log G_\phi(\bar{\mu}_g) - \log c + log\ \sigma$$

- **2D Pose Loss:**

$$\mathcal{L}_{2D} = |J_{2D} - J^g_{2D}|$$

- **3D Pose Loss:**

$$\mathcal{L}_{3D} = |J_{3D} - J^g_{3D}|$$

- **Overall Objective:**

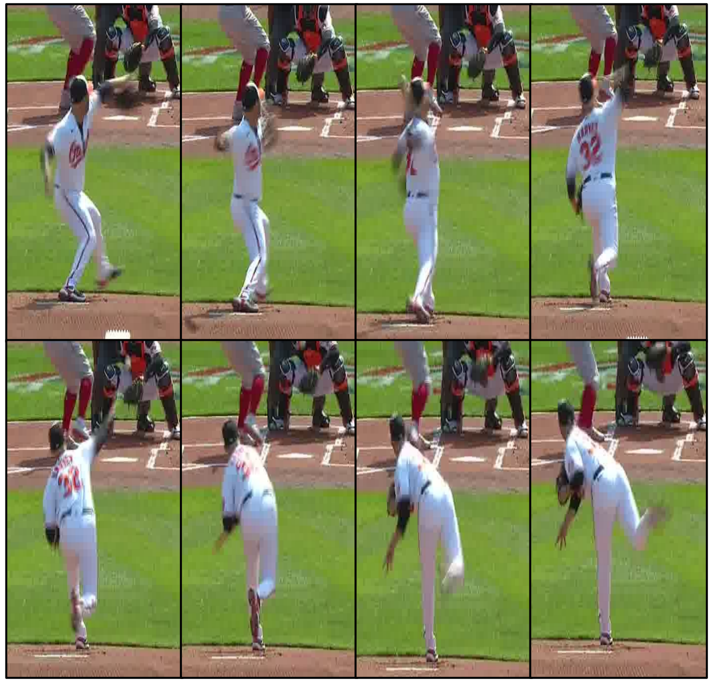
$$\mathcal{L} = \lambda_d \mathcal{L}_{RLE} + \lambda_v \mathcal{L}_v + \lambda_{3D} \mathcal{L}_{3D} + \lambda_{2D} \mathcal{L}_{2D} + \lambda_s \mathcal{L}_{silh}$$

Datasets

3D Poses in the Wild

Human3.6M

MLBPitchDB



Benchmarked Datasets

In-house Sports Dataset

Quantitative Comparison (Benchmarked Datasets)

Table I: Comparison to state-of-the-art 3D pose reconstruction approaches on 3DPW and Human3.6M datasets. **Bold**: best; <u>Underline</u>: second best.

	Method	Human3.6M		3DPW		
		mPJPE ↓	PA-mPJPE ↓	mPVE ↓	mPJPE ↓	PA-mPJPE ↓
Video	HMMR [5]	-	58.1	139.3	116.5	72.6
	TCMR [33]	62.3	41.1	111.5	95.0	55.8
	VIBE [9]	65.6	41.4	99.1	93.5	56.5
Model-based	HMR [4]	88.0	56.8	-	130.0	81.3
	SPEC [34]	-	-	118.5	96.5	53.2
	SPIN [10]	62.5	41.1	116.4	96.9	59.2
	PyMAF [35]	57.7	40.5	110.1	92.8	58.9
	ROMP [36]	-	-	105.6	89.3	53.5
	HMR-EFT [37]	63.2	43.8	98.7	85.1	52.2
	PARE [11]	76.8	50.6	97.9	82.0	50.9
Model-free	ProHMR [12]	-	41.2	109.6	95.1	59.5
	I2LMeshNet [22]	55.7	41.1	-	93.2	57.7
	Pose2Mesh [7]	64.9	47.0	-	89.2	58.9
	METRO [8]	<u>54.0</u>	<u>36.7</u>	**88.2**	**77.1**	**47.9**
	D2A-HMR (Ours)	**53.8**	**36.2**	<u>88.4</u>	<u>80.5</u>	<u>48.4</u>

Qualitative Results (Benchmarked Datasets)

Comparison (In-house Sports Dataset)

Table II: Comparison of D2A-HMR on the MLBPitchDB baseball dataset [41]. **Bold**: best; Underline: second best; Double Underline: third best.

Method	Acc. ↑	mPJPE ↓
HMR [8]	65.9	61.3
SPIN [10]	84.7	32.1
ProHMR [8]	76.1	48.2
ROMP [8]	77.4	48.9
METRO [8]	81.5	37.8
PARE [11]	84.0	33.7
D2A-HMR (Ours)	**87.9**	**30.6**

Ablation Studies

Table III: Ablation study on pseudo-depth and distribution modeling for D2A-HMR evaluated on 3DPW dataset.

Depth	**Distribution**	**mPJPE** ↓	**PA-mPJPE** ↓
✓		92.7	61.8
	✓	90.0	56.9
✓	✓	**80.5**	**48.4**

Table IV: Ablation study on the impact of depth modeling for D2A-HMR evaluated on 3DPW dataset.

	mPJPE(z) ↓	**PA-mPJPE(z)** ↓
w/o depth modeling	69.1	58.3
w/ depth modeling	**65.4**	**53.6**

Table V: Ablation study on the silhouette decoder and masked modeling evaluated on 3DPW dataset.

Silhouette	**Masked Modeling**	**mPJPE** ↓	**PA-mPJPE** ↓
✓		89.5	62.2
	✓	84.7	51.4
✓	✓	**80.5**	**48.4**

Table VI: Different input representations as the backbone for D2A-HMR evaluated on 3DPW dataset.

Backbone	**mPJPE** ↓	**PA-mPJPE** ↓
ResNet50	91.1	59.9
ResNet101	89.5	55.8
HRNet-w40	85.2	52.1
HRNet-w64	**80.5**	**48.4**

Conclusion

- A novel image-based HMR model named **D2A-HMR** that adeptly models the underlying distributions and integrates pseudo-depth priors for efficient and accurate mesh recovery.

- By leveraging **residual log-likelihood** approach, we refine the model by learning the disparity between the underlying predicted and ground truth distribution.

- **Validation** of the enhanced performance through the integration of pseudo-depth and distribution-aware modules in HMR, particularly in complex human pose scenarios.

Thank you!

Supported by: