Robust 3D Human Modeling for Baseball Broadcast Analysis

Jerrin Bright

MASc Graduate Student

Principal Investigator: Prof. John Zelek

Department of Systems Design Engineering

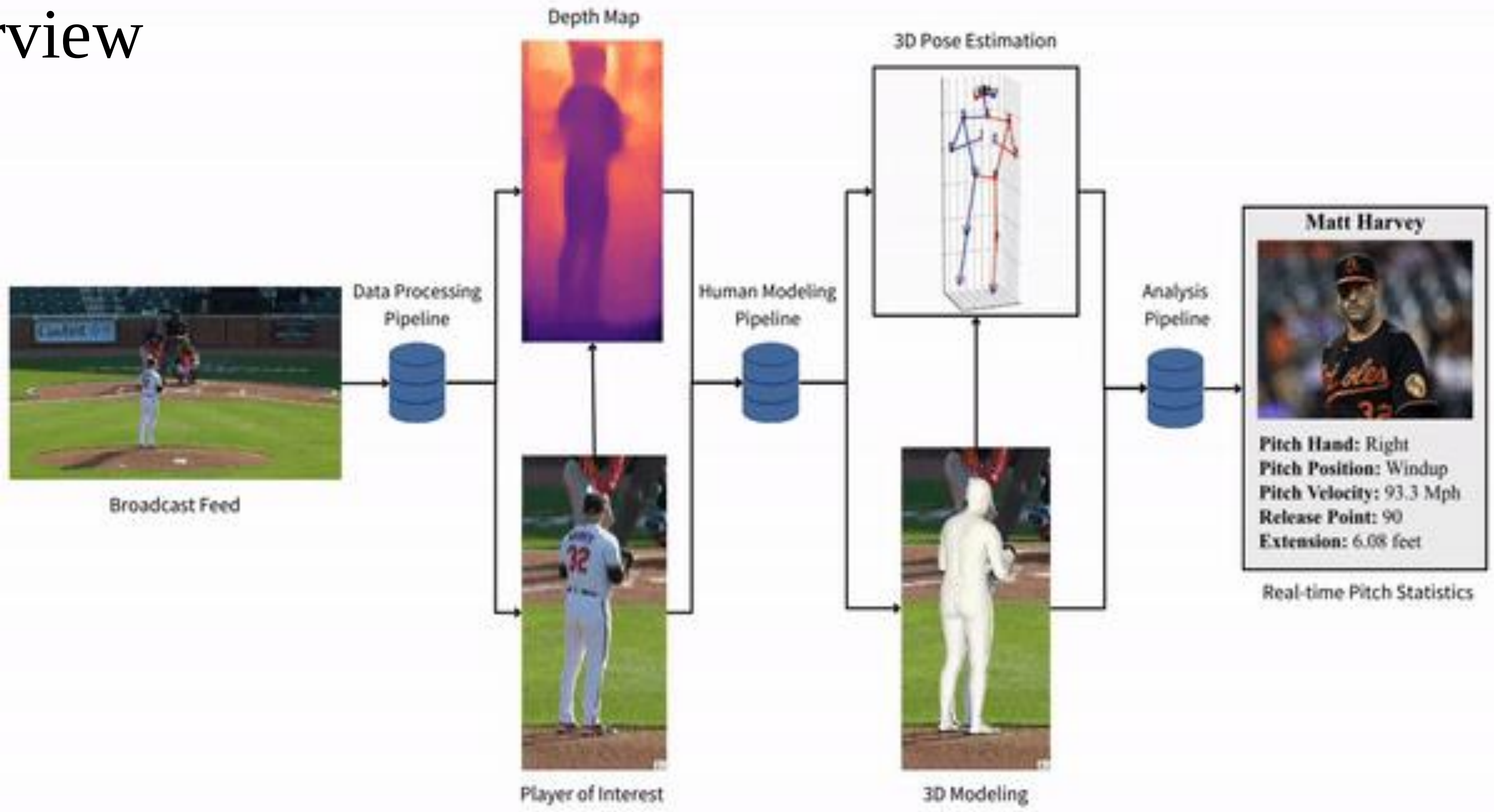
University of Waterloo, Waterloo, ON, Canada

Date: Friday, April 26th, 2024

Motivation

- Quantitative performance from a single camera (smart phone) from the stands

Overview

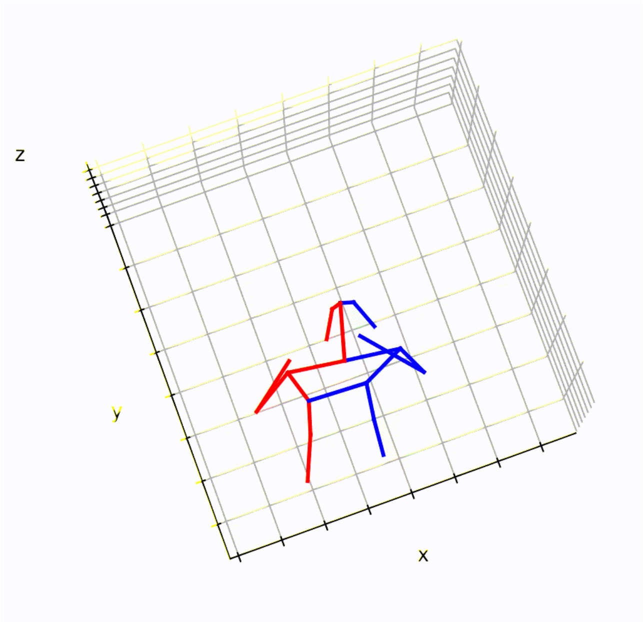


https://youtu.be/wKBOtDPPxws

Dataset Overview

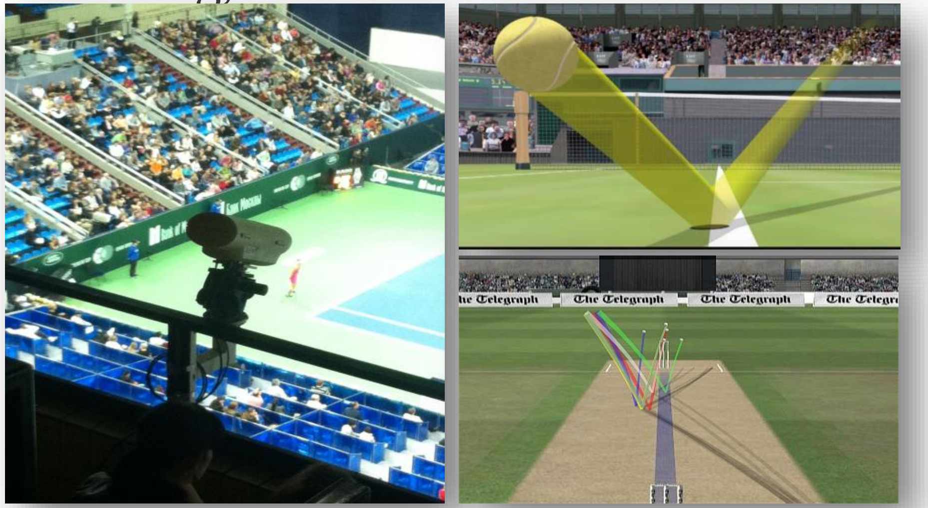
What we have?

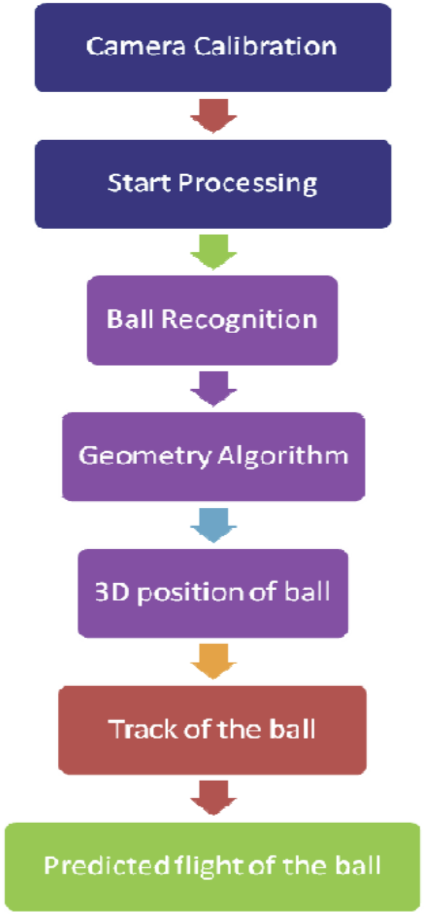
- 1000+ games
- 3D Hawk-Eye pose data
- Various pitch metrics



What is Hawk-Eye Camera System?

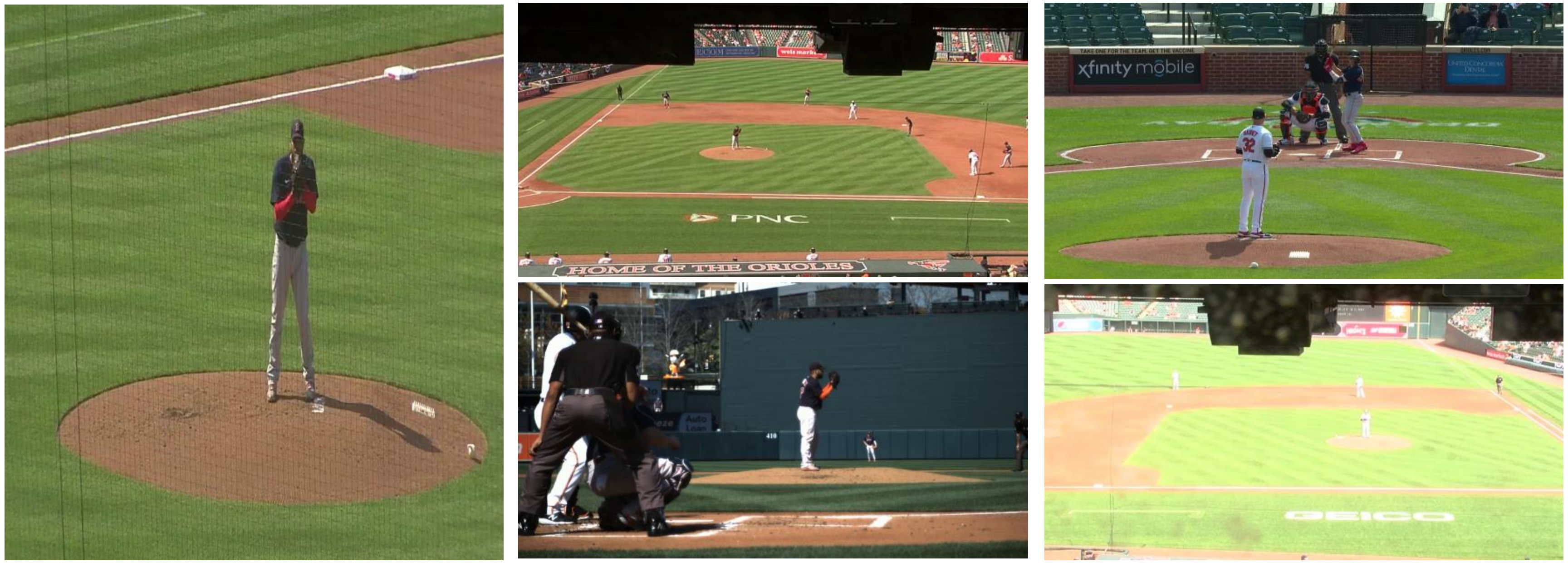
- Triangulation with many cameras around the playing area
- Applications include pose estimation, tracking, etc.



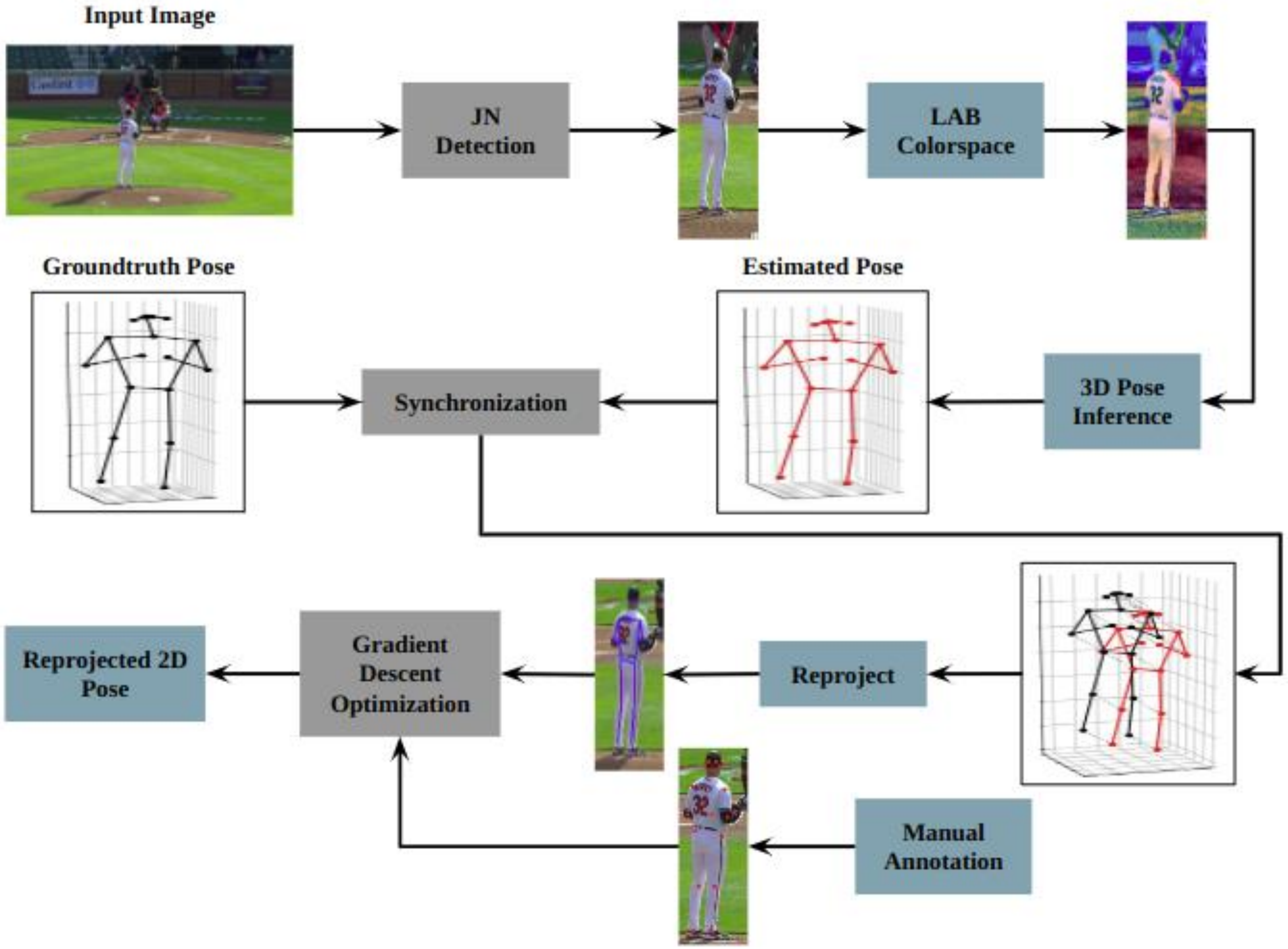


Dataset Challenges

- Multiview unsynchronized data (alignment and sampling rate)
- No 2D groundtruth pose
- No Hawk-Eye camera parameters

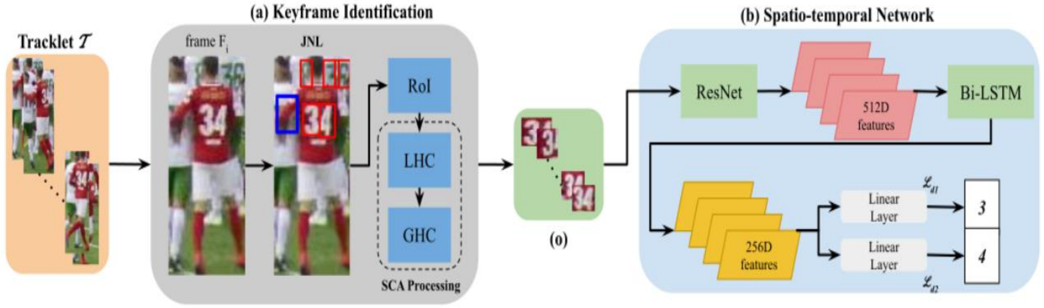


Dataset Processing

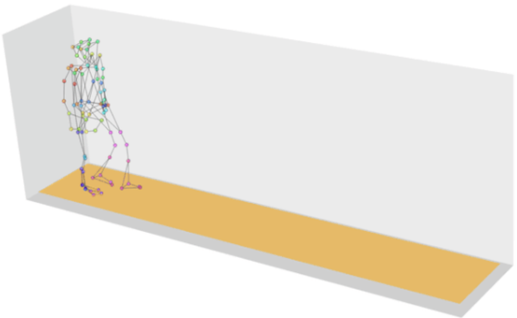


Architecture of the method adapted to solve the issues with dataset from behind-the-pitcher viewpoint

Jersey Number Identification



Data Synchronization

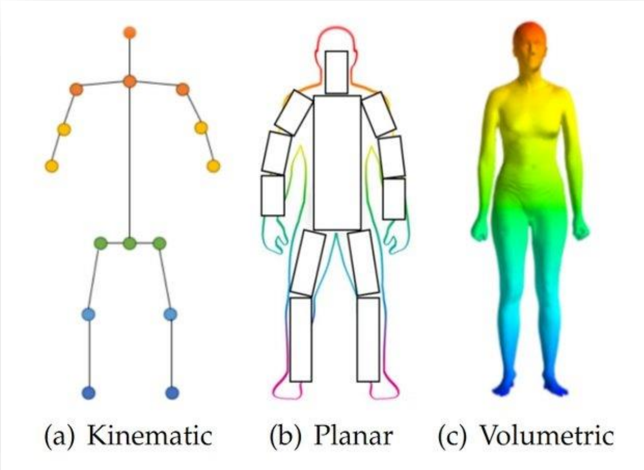


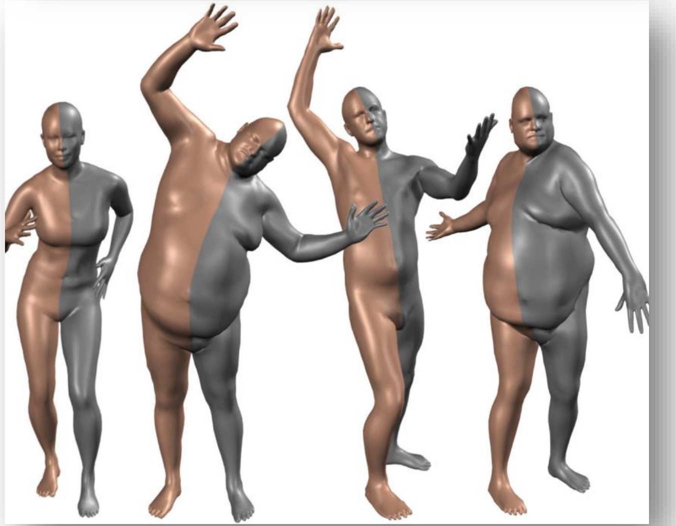
$$\mathcal{G} = g_s\left(\frac{1}{\mathcal{J}}\sum_{i=1}^{\mathcal{J}}\left(kp_{gt}^{(i)} - kp_{pred}^{(i)}\right)^2\right) + g_t\left(1 - \frac{\sum_{i=1}^{\mathcal{J}} kp_{gt}^{(i)} \cdot kp_{pred}^{(i)}}{\sqrt{\sum_{i=1}^{\mathcal{J}}(kp_{gt}^{(i)})^2} \cdot \sqrt{\sum_{i=1}^{\mathcal{J}}(kp_{pred}^{(i)})^2}}\right)$$

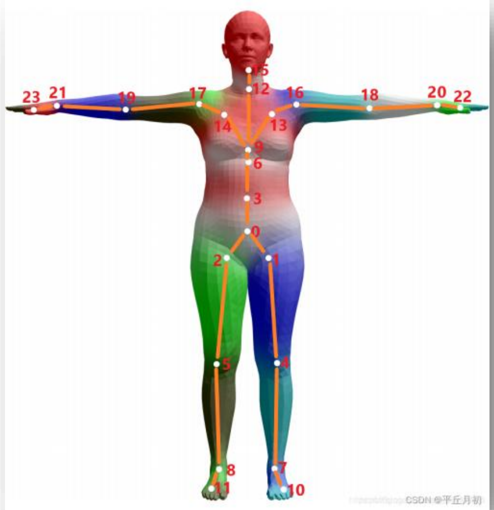
Background

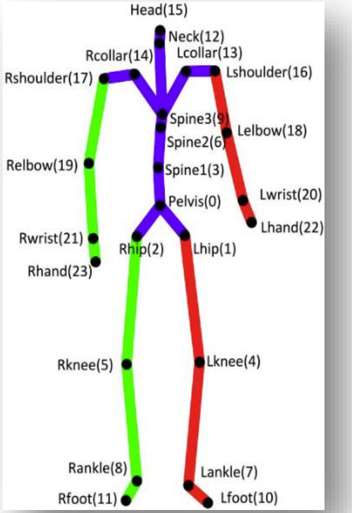
3D Human Modeling - SMPL

- Skinned **M**ulti-**P**erson **L**inear model[1].
- 72 joint and 10 shape parameters -> 6890 vertices.
- Learns pose from 1700 3D scans with 44 subjects.
- Learns shape from 4000 3D scans from Ceaser Dataset.









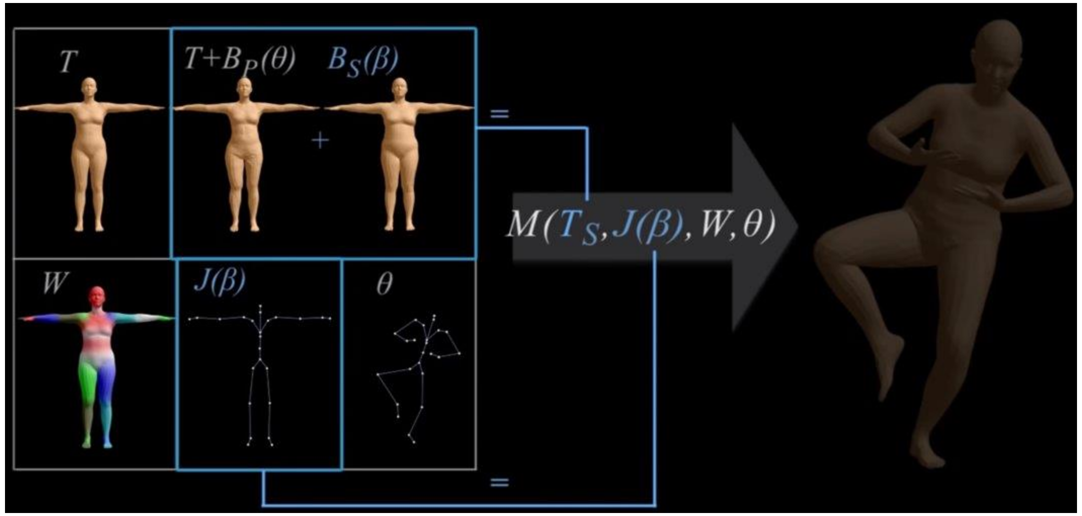
Credits:

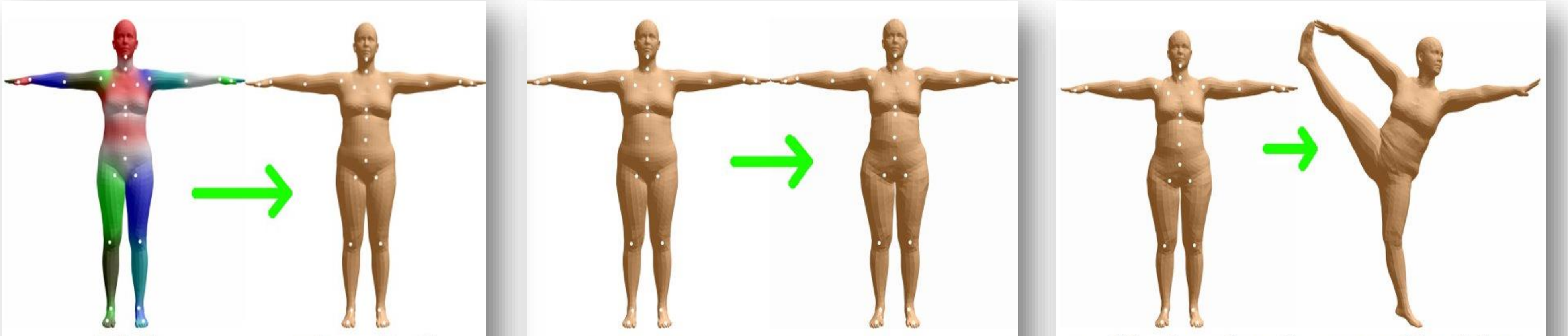
[1] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: a skinned multi-person linear model. ACM Transactions on Graphics, 2015.

Background

3D Human Modeling - SMPL

- Shape Blended Shape
- Pose Blended Shape
- Skinning





(a) $\bar{\mathbf{T}}, \mathcal{W}$ (b) $\bar{\mathbf{T}} + B_S(\vec{\beta}), J(\vec{\beta})$

(b) $\bar{\mathbf{T}} + B_S(\vec{\beta}), J(\vec{\beta})$ (c) $T_P(\vec{\beta},\vec{\theta}) = \bar{\mathbf{T}} + B_S(\vec{\beta}) + B_P(\vec{\theta})$

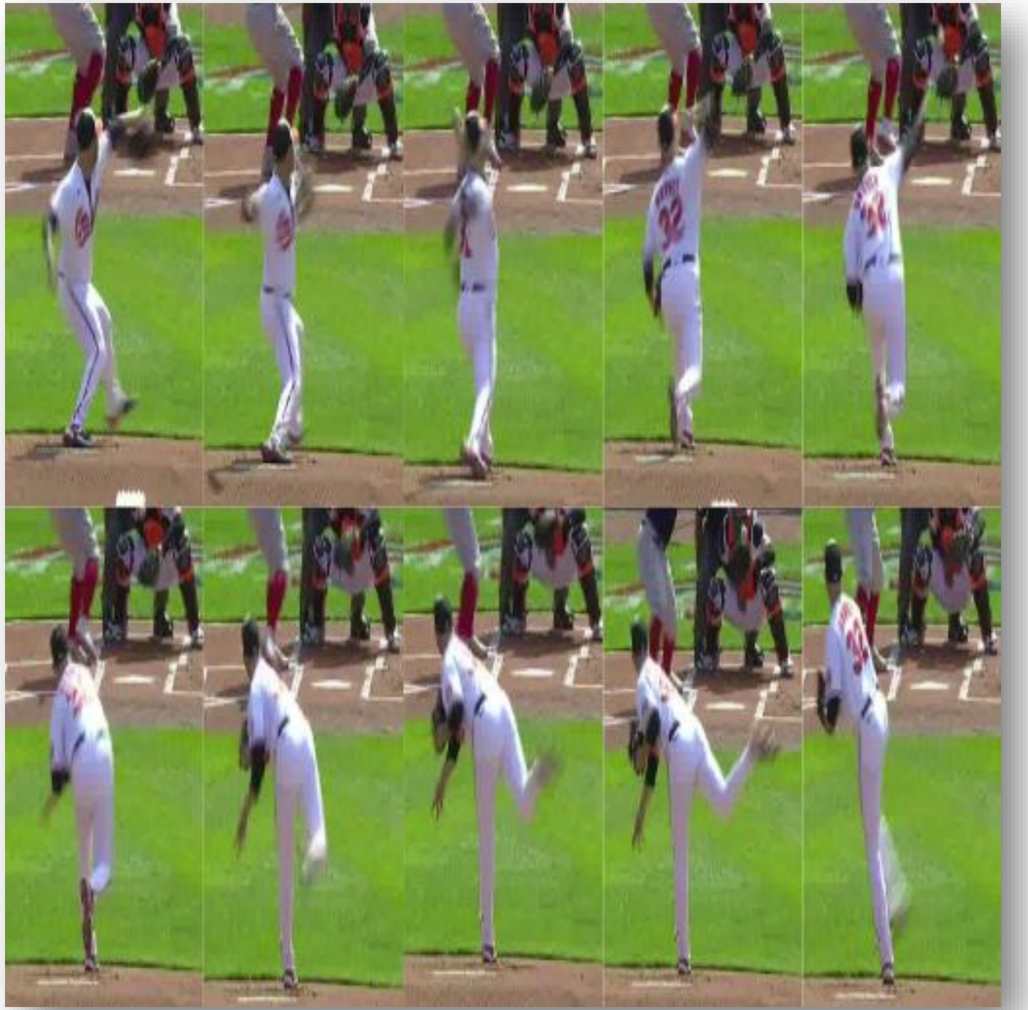
(c) $T_P(\vec{\beta},\vec{\theta}) = \bar{\mathbf{T}} + B_S(\vec{\beta}) + B_P(\vec{\theta})$ (d) $W(T_P(\vec{\beta},\vec{\theta}), J(\vec{\beta}), \vec{\theta}, \mathcal{W})$

How can we estimate robust 3D models?

3D Human
Modelling Pipeline

Challenges

Motion Blur

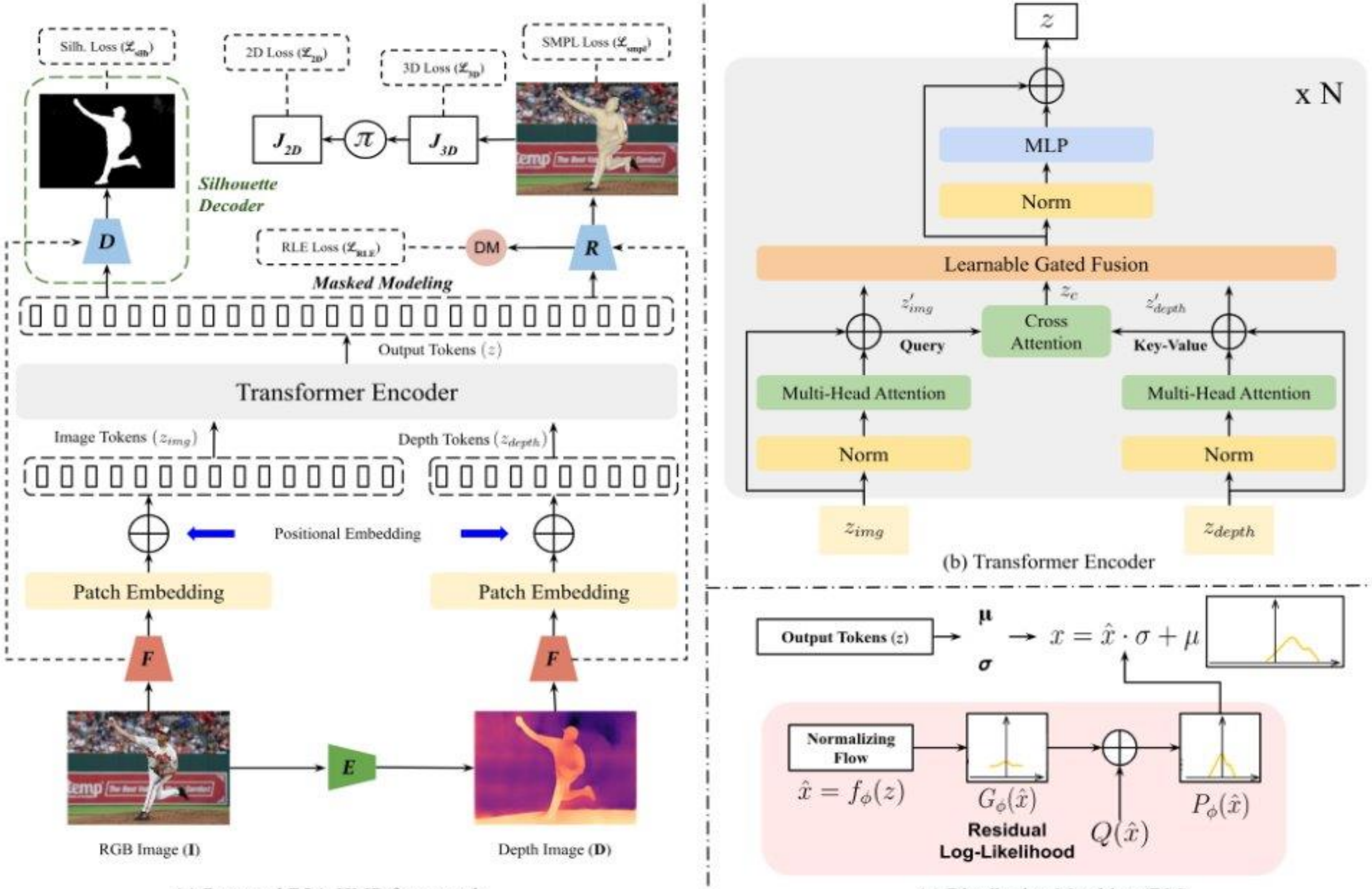


Self-Occlusion

Out-of-distribution

Human Mesh Recovery

Jerrin Bright, Bavesh Balaji, Harish Prakash, Yuhao Chen, David A Clausi, and John Zelek. 2024. Distribution and Depth-Aware Transformers for 3D Human Mesh Recovery. In 21st Conference on Robots and Vision - ORAL



(a) Proposed D2A-HMR framework

(b) Transformer Encoder

(c) Distribution Matching (DM)

Human Mesh Recovery

Jerrin Bright, Bavesh Balaji, Harish Prakash, Yuhao Chen, David A Clausi, and John Zelek. 2024. Distribution and Depth-Aware Transformers for 3D Human Mesh Recovery. In 21st Conference on Robots and Vision - ORAL

Quantitative on SOTA datasets

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

Quantitative on MLBPitchDB dataset

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 of depth and distribution modules

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

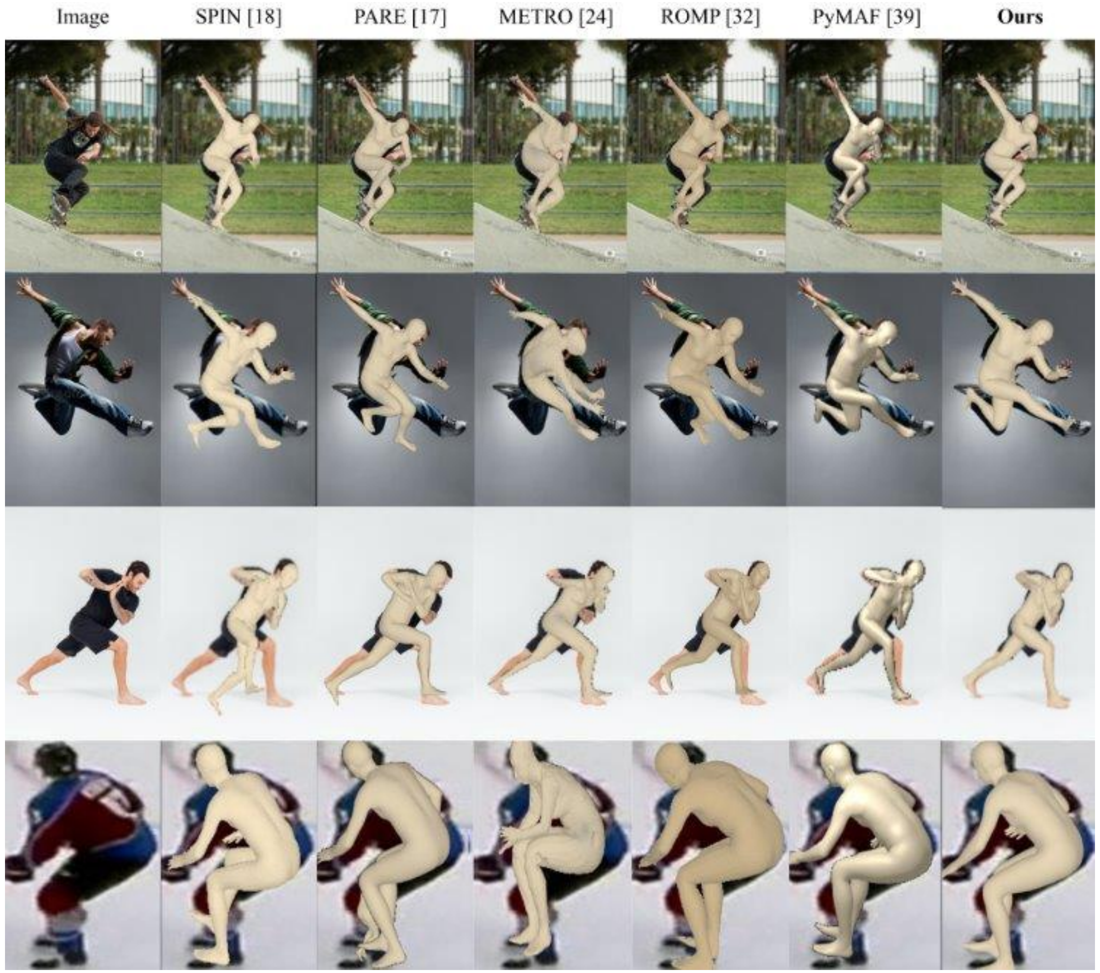
Ablation of silhouette and masked modeling modules

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

VIPlab Vision and Image Processing Research Group vip.uwaterloo.ca
UNIVERSITY OF WATERLOO | FACULTY OF ENGINEERING

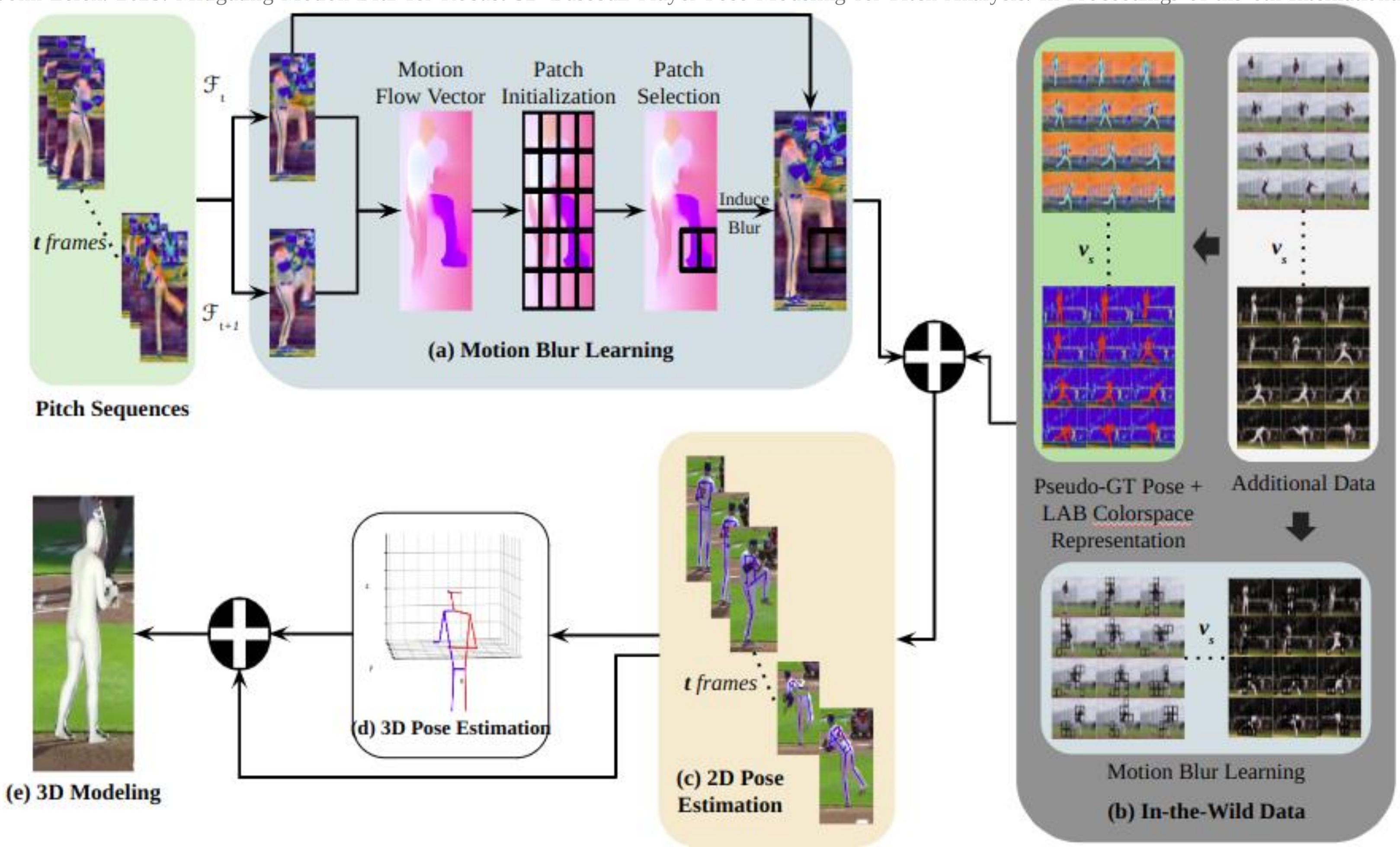
Human Mesh Recovery

Jerrin Bright, Bavesh Balaji, Harish Prakash, Yuhao Chen, David A Clausi, and John Zelek. 2024. Distribution and Depth-Aware Transformers for 3D Human Mesh Recovery. In 21st Conference on Robots and Vision - ORAL



Mitigating Motion Blur

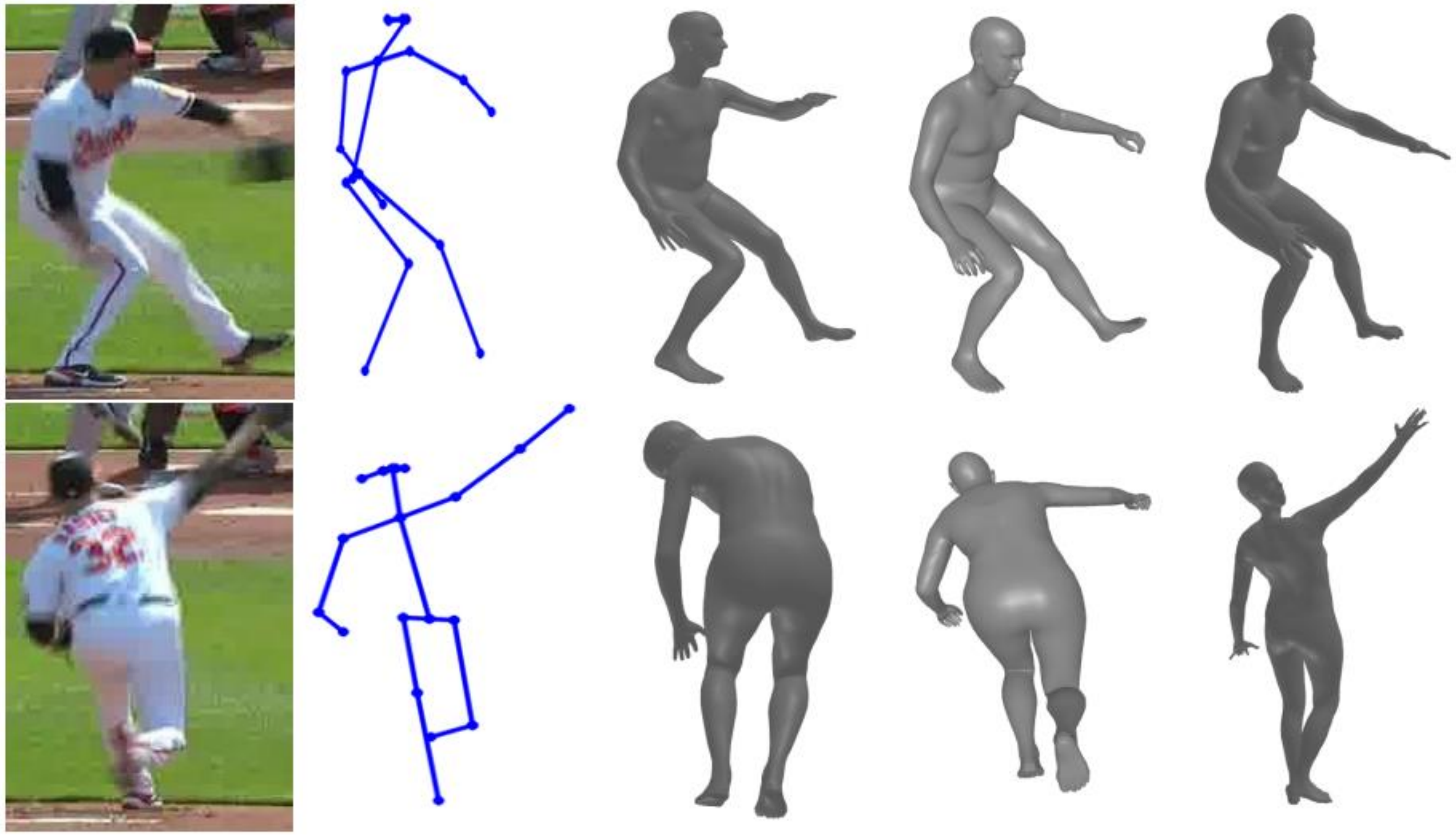
Jerrin Bright, Yuhao Chen, and John Zelek. 2023. Mitigating Motion Blur for Robust 3D Baseball Player Pose Modeling for Pitch Analysis. In Proceedings of the 6th International Workshop on Multimedia Content Analysis in Sports.



Architecture of the algorithm adapted to mitigate motion blur

Mitigating Motion Blur

Jerrin Bright, Yuhao Chen, and John Zelek. 2023. Mitigating Motion Blur for Robust 3D Baseball Player Pose Modeling for Pitch Analysis. In Proceedings of the 6th International Workshop on Multimedia Content Analysis in Sports.



(a) Input frame (b) GT 2D pose (c) ICON (d) OSX (e) Ours

Performance on SOTA Techniques

Method	Type	MB	Loss
Xu et al.	Heatmap		1.37
Ke et al.	Heatmap		1.46
Panteleris et al.	Regressor		1.15
Li et al.	Heatmap		1.83
Mao et al.	Regression		1.26
Xu et al.	Heatmap	✓	1.17 (+0.20)
Ke et al.	Heatmap	✓	1.21 (+0.25)
Panteleris et al.	Regressor	✓	0.55 (+0.60)
Li et al.	Heatmap	✓	1.46 (+0.37)
Mao et al.	Regressor	✓	0.61 (+0.65)

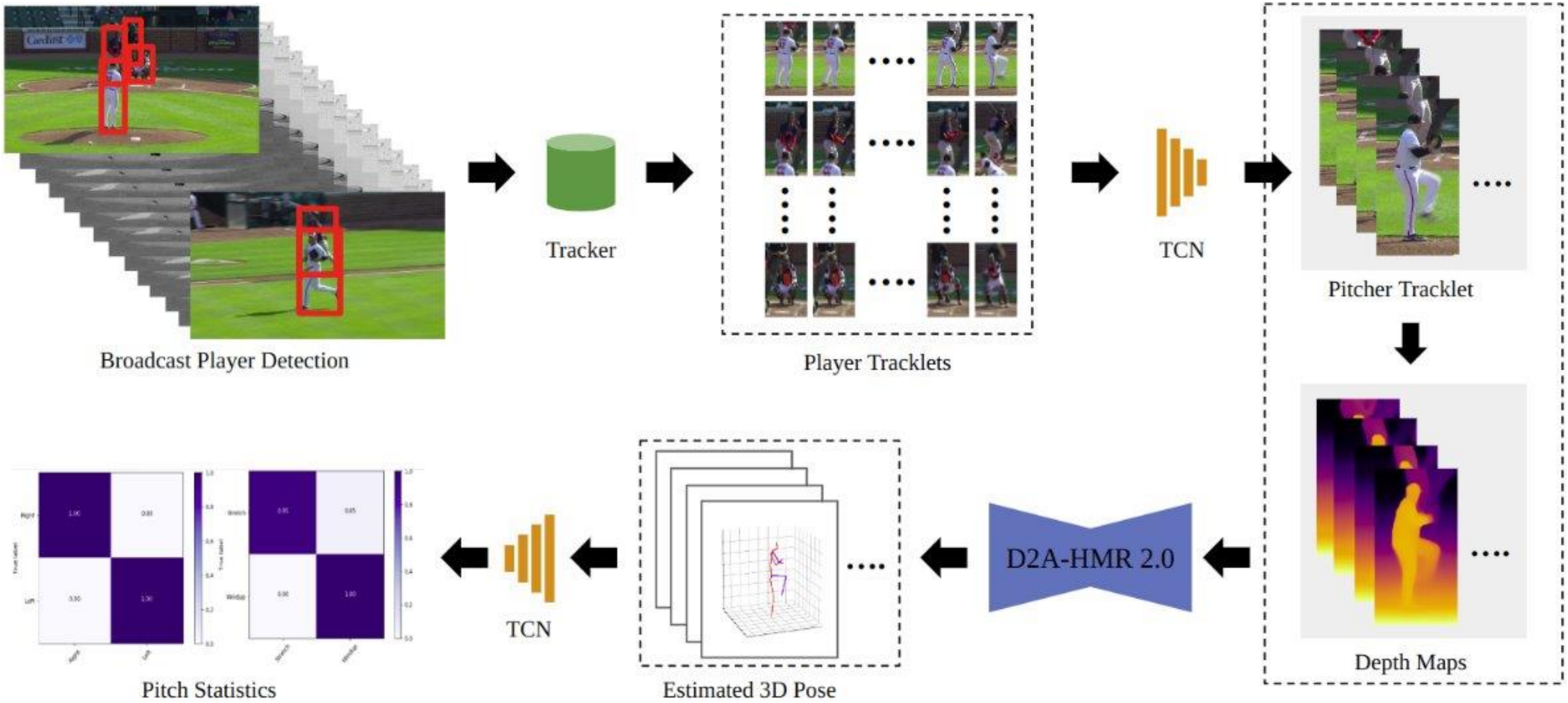
Impact of different modules

Base Model	ItW	MB	2D Loss	3D Loss
✓			1.05	1.93
✓	✓		0.88	1.61
✓		✓	0.55	1.47
✓	✓	✓	**0.48**	**1.23**

Vision and Image Processing Research Group
VIPlab
vip.uwaterloo.ca
UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

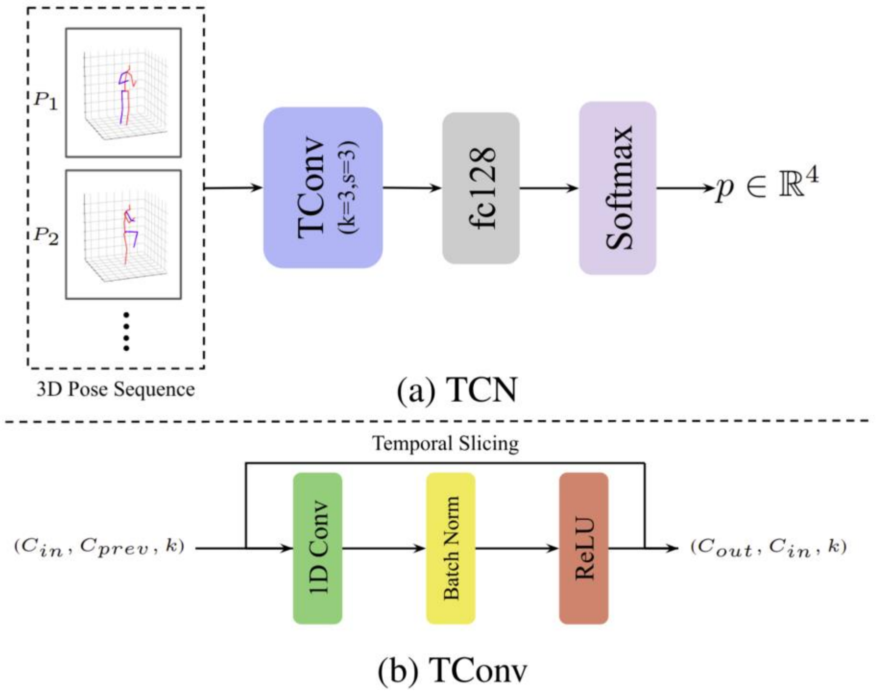
Pitch Analysis

Jerrin Bright, Bavesh Balaji, Yuhao Chen, David A Clausi, John Zelek. 2024. PitcherNet: Powering the Moneyball Evolution in Baseball Video Analytics. In IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops - ORAL



Pitch Analysis

Jerrin Bright, Bavesh Balaji, Yuhao Chen, David A Clausi, John Zelek. 2024. PitcherNet: Powering the Moneyball Evolution in Baseball Video Analytics. In IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops - ORAL



Overview of the Temporal Convolutional Network

Performance of our pitch statistics modules

(a) Handedness

	Accuracy ↑	F1 Score ↑	Precision ↑
LSTM	85.0	85.7	90.0
Ours (TCN)	**100.0**	**100.0**	**100.0**

(b) Pitch Position

	Accuracy ↑	F1 Score ↑	Precision ↑
LSTM	81.3	82.5	85.0
Ours (TCN)	**97.5**	**97.4**	**95.0**

(c) Release Point

	A_1 ↑	A_2 ↑	A_5 ↑
LSTM	31.3	46.4	63.5
TCN	43.4	51.5	77.6
Ours	**80.8**	**85.8**	**97.9**

(d) Pitch Velocity

	$A_{1\%}$ ↑	$A_{2\%}$ ↑	$A_{5\%}$ ↑
LSTM	5.1	13.1	22.2
TCN	10.1	18.1	48.4
Ours	**43.4**	**68.6**	**94.9**

(e) Release Extension

	$A_{5\%}$ ↑	$A_{8\%}$ ↑	$A_{10\%}$ ↑
LSTM	4.0	7.1	11.1
TCN	14.1	19.1	25.2
Ours	**24.2**	**31.3**	**37.3**

Comparison of the player identification techniques

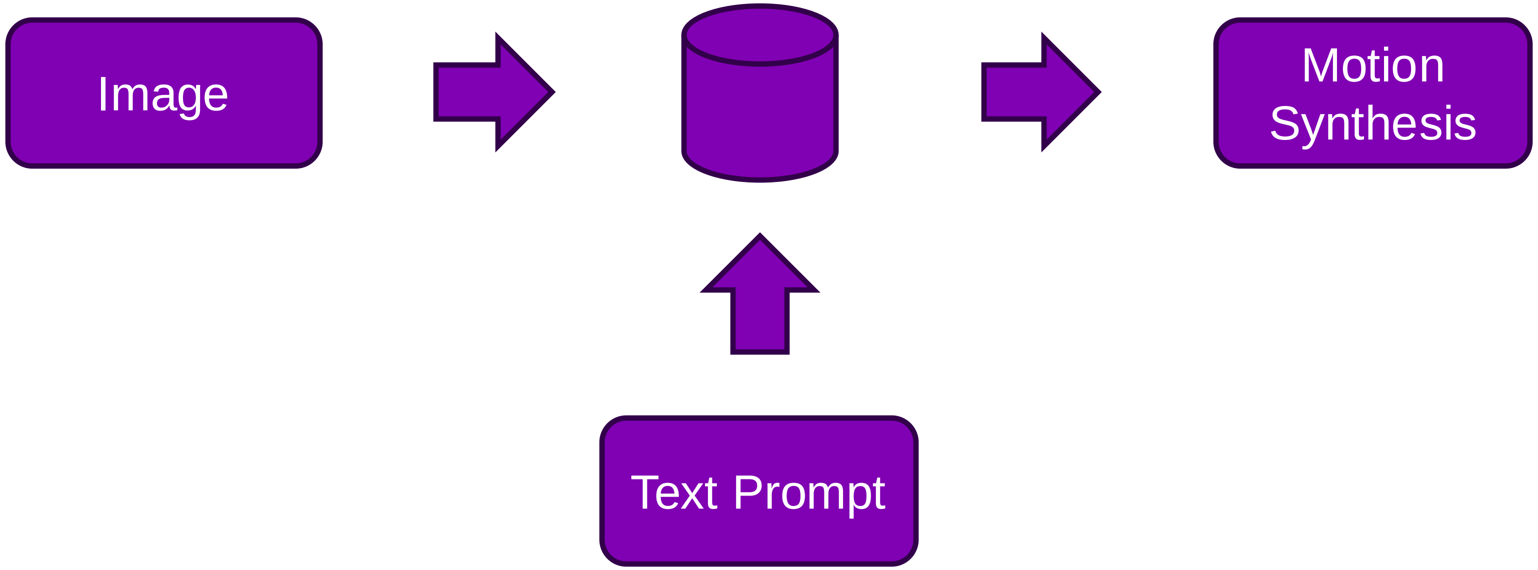
	Test Accuracy ↑
LSTM	85.55
Transformer	91.11
Ours	**96.66**

Pitch Analysis

Jerrin Bright, Bavesh Balaji, Yuhao Chen, David A Clausi, John Zelek. 2024. PitcherNet: Powering the Moneyball Evolution in Baseball Video Analytics. In IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops - ORAL

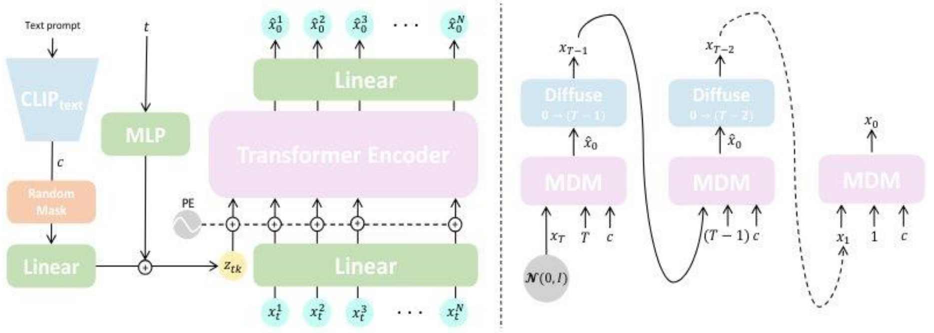
How to generalize for Multiview data?

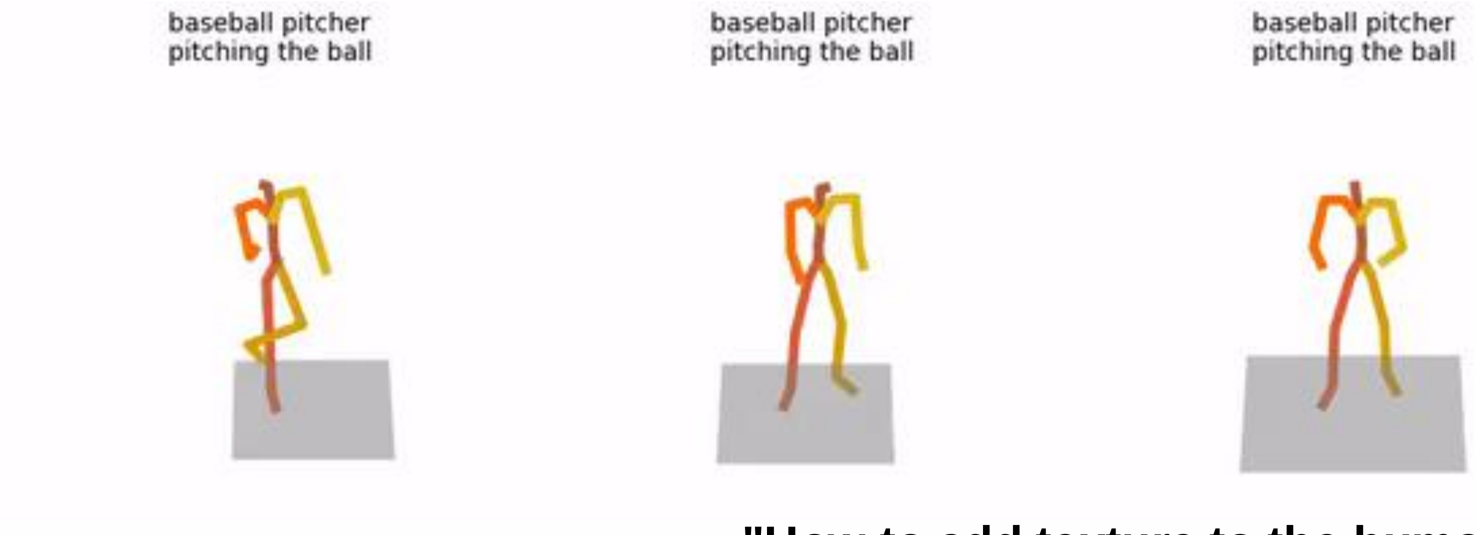
Objective: Synthesis Realistic 3D Human and Motion



Realistic 3D Human and Motion Synthesis

- Text-driven Motion Synthesis

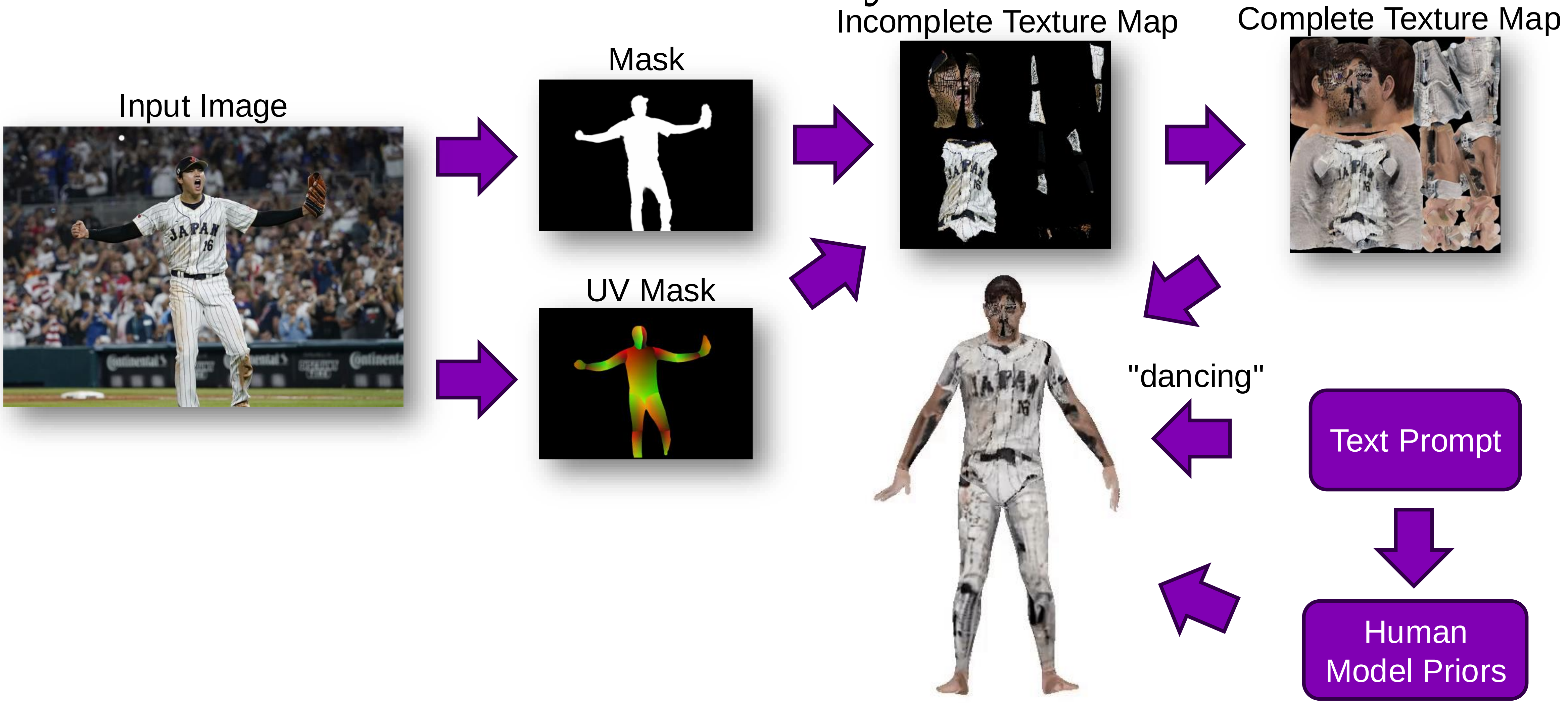




"How to add texture to the human model?"

Credits:
[1] Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohenor, Amit H. Bermano. MDM: Human Motion Diffusion Model International Conference on Learning Representations 2023.

Realistic 3D Human and Motion Synthesis

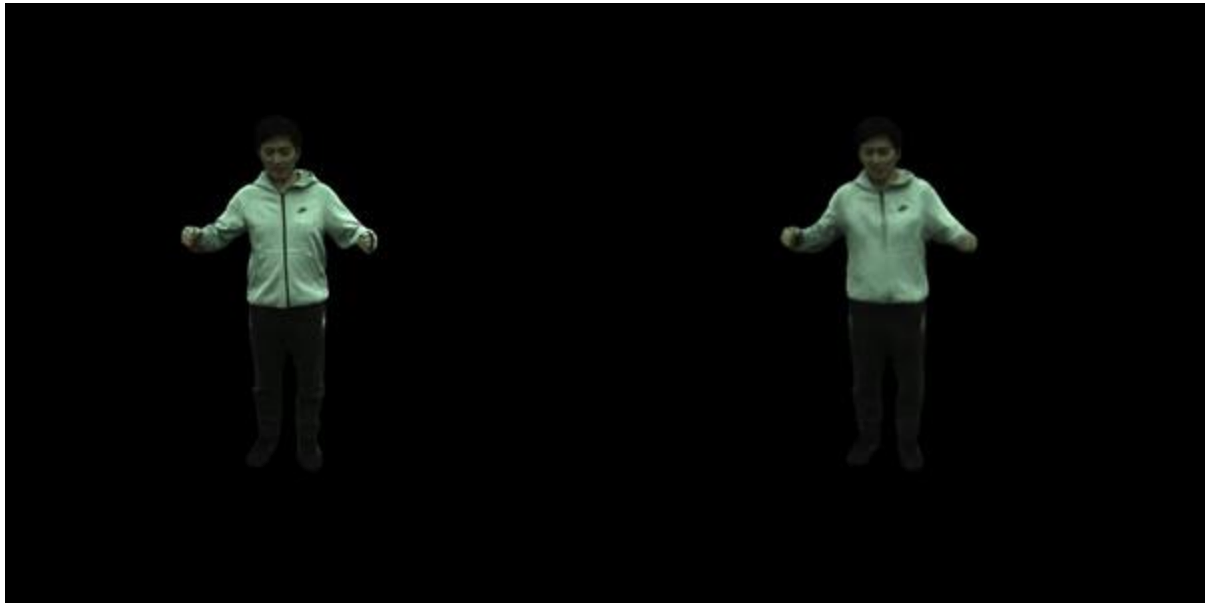


Credits:
[1] Casas, Dan and Comino-Trinidad, Marc. SMPLitex: A Generative Model and Dataset for 3D Human Texture Estimation from Single Image British Machine Vision Conference 2023.

Realistic 3D Human and Motion Synthesis

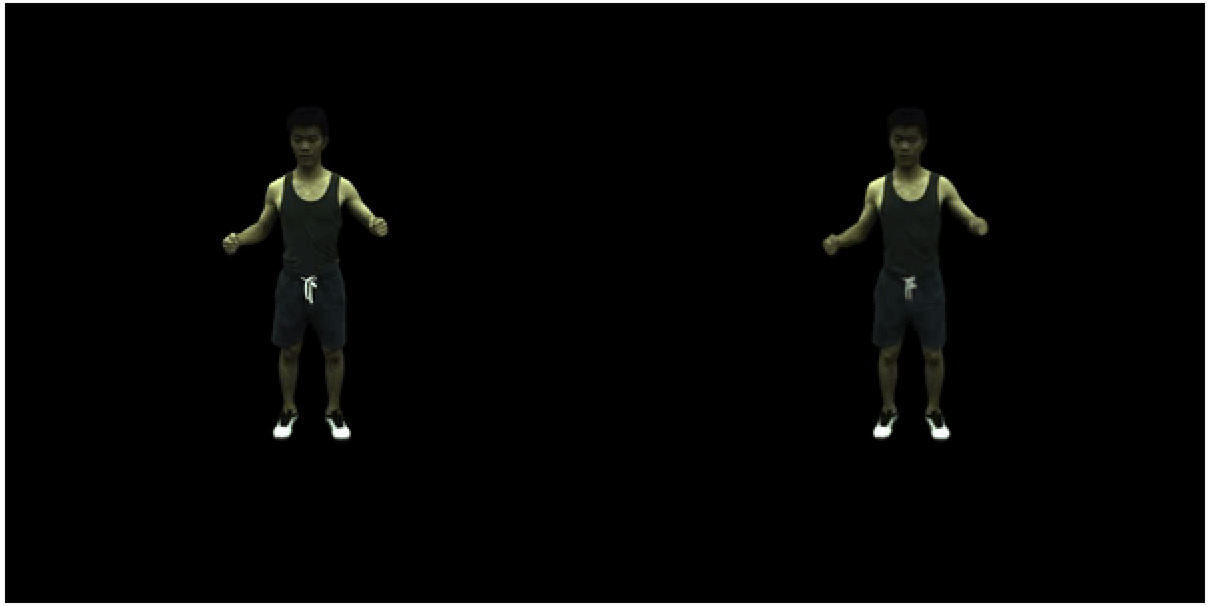
- Gaussian Splatting

GT Prediction



zju_mocap dataset: test #387

GT Prediction



zju_mocap dataset: test #377

Credits:
[1] Hu, Shoukang and Liu, Ziwei. GauHuman: Articulated Gaussian Splatting from Monocular Human Videos in EEE/CVF Conference on Computer Vision and Pattern Recognition 2024.

Summary

MAIN CONTRIBUTIONS

- Reliable pitch analysis driven by player kinematics and human model priors.
- Generalizable 3D human modeling with depth and distribution modeling.
- Realistic 3D human and motion generation with text guidance.

CURRENT CHALLENGES

- Severe motion blur
- 2D pose performance

TODO

- Motion representation
- One-stage optimization
- Diffusion-priors

Thank you!

Supported by: