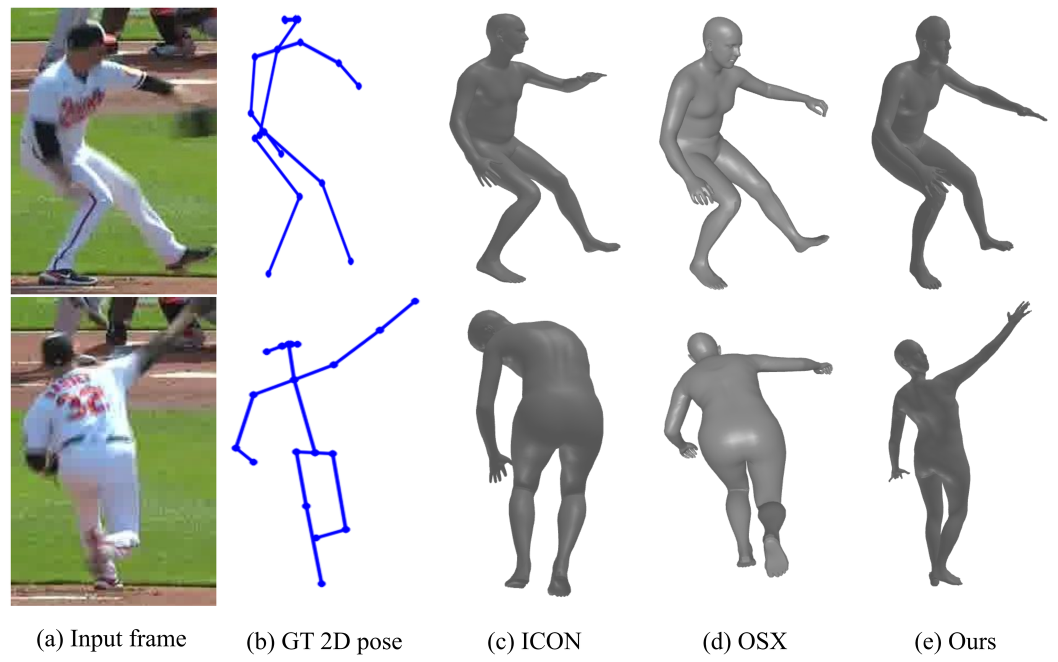
Mitigating Motion Blur for Robust 3D Baseball Player Pose Modeling for Pitch Analysis

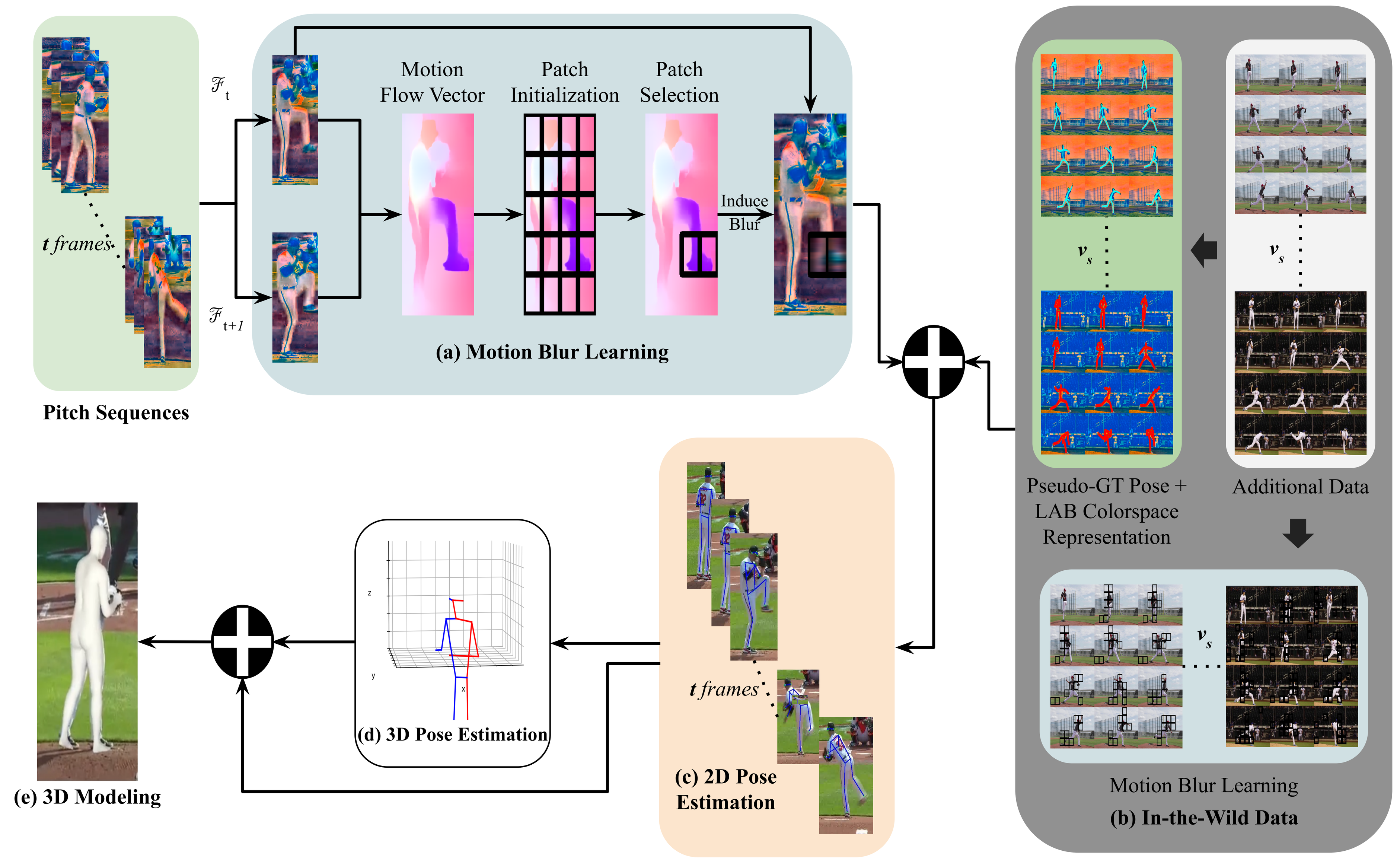
Jerrin Bright, Yuhao Chen and John Zelek
University of Waterloo, Waterloo, Ontario, Canada

Key Contributions

- **A focused augmentation strategy incorporating motion blur artifacts**, challenging conventional belief in pipelines.
- **Leveraging in-the-wild datasets**, aids in capturing the variability and complexity present in the data.
- **Improved performance of existing pose estimators with proposed framework incorporation**, where we demonstrate the substantial enhancement
- **Spatiotemporal cost reinforced by histogram representations**, to effectively align partially synchronized frames.



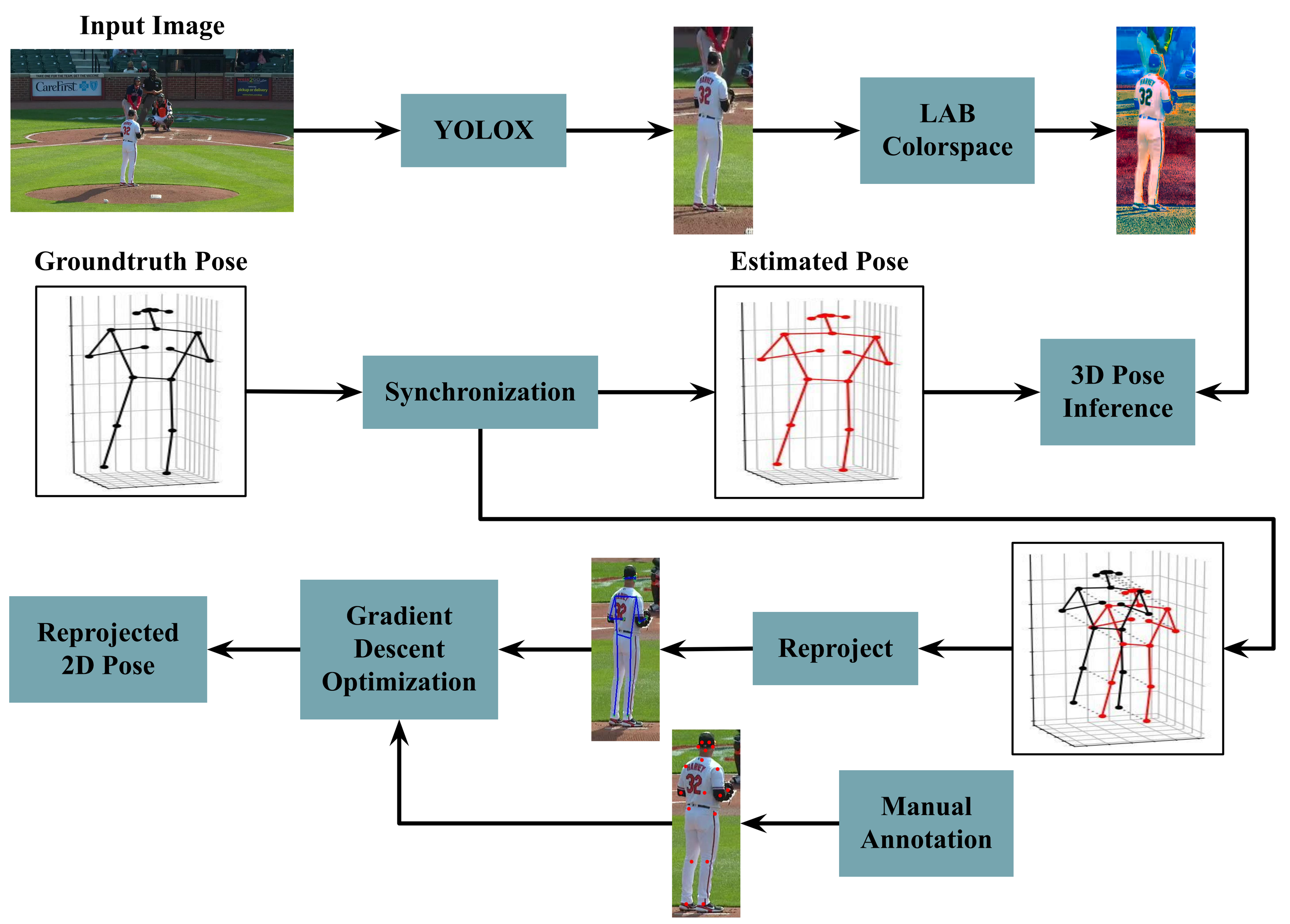
Methodology



The proposed approach comprises several key steps:

1. **Data Representation:** Each pitch sequence is represented as $\hat{\mathcal{P}} = \{\mathcal{F}_t : \mathcal{F}_t \in \mathbb{R}^{H\times W\times 3}\}_{t=1}^{t_n}$.
2. **Motion Blur Augmentation:** Motion flow (MF) between consecutive frames is analyzed by dividing each frame into k patches. The top $\mathcal{N}$ patches with the highest MF are selected as target regions for inducing blur.
3. **2D Pose Estimation:** In each frame $\mathcal{F}_t$, the 2D pose of the pitcher is estimated, resulting in $\mathcal{P}_{2D}^{(t)} \in \mathbb{R}^{\mathcal{J}\times 2}$.
4. **3D Pose Estimation:** Utilizing a receptive field of s consecutive 2D pose ($\mathcal{P}_{2D} \in \mathbb{R}^{s\times\mathcal{J}\times 2}$), the 3D pose of the pitcher is estimated, producing $\mathcal{P}_{3D} \in \mathbb{R}^{1\times\mathcal{J}\times 3}$.
5. **Concatenation:** The 2D and 3D poses are concatenated represented by $\mathcal{P}_{\text{concat}}^{(t)} \in \mathbb{R}^{1\times\mathcal{J}\times 5}$.
6. **Human Mesh Recovery:** The 3D body mesh represented by $\mathcal{H}_{3D} \in \mathbb{R}^{\mathcal{V}\times 3}$ is then modeled using spectral convolutional networks [1].

Dataset



Synchronization: Warping the time axis and minimizing the distance (cost) between the sequence. A one-to-one hard constraint was assigned with a weighted cost function ($\mathcal{G}$).

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

Camera Projection: Through a process of gradient descent optimization, we iteratively refine the initialized focal length (f_i), which will be used to reproject the 3D GT pose to 2D image coordinate.

$$\hat{f} = f_i - \alpha\Delta L(f_i) \quad (2)$$

Results

Performance of different SOTA 2D pose 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)

Results of the estimated pose with 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**

Study on the region size and frequency of blur effect

s_{patch} | $\mathcal{N}$	1	3	5	7	9
10	0.83	0.74	0.66	0.64	0.67
20	0.71	0.57	0.62	0.60	0.62
30	0.68	**0.55**	0.61	0.639	0.59
40	0.74	0.63	0.68	0.75	0.78

Loss Functions

The loss function leveraged for 2D and 3D pose estimators is the Euclidean distance between γ dimensions, defined as:

$$\mathcal{L}_{pose} = \frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}}\frac{1}{\mathcal{J}}\sum_{j=1}^{\mathcal{J}}\|kp_{pred}^{(ij)} - kp_{gt}^{(ij)}\|_\gamma \quad (3)$$

where,

$$\|\cdot\|_\gamma = \begin{cases}\|\cdot\|_2, & \text{if } \gamma = 2 \text{ (for } \mathcal{P}_{2D})\\ \|\cdot\|_3, & \text{if } \gamma = 3 \text{ (for } \mathcal{P}_{3D})\end{cases}$$

The loss function employed for human mesh recovery encompasses vertex, joint, normal, and edge loss, defined as:

$$\mathcal{L}_{mesh} = \lambda_v\mathcal{L}_v + \lambda_j\mathcal{L}_j + \lambda_n\mathcal{L}_n + \lambda_e\mathcal{L}_e \quad (4)$$

Conclusion

1. **Innovative Augmentation for Motion Blur:** The research introduces a unique technique to strategically enhance motion blur, improving the network's ability to handle this challenge during pose estimation.
2. **In-the-Wild Video Data Integration:** Incorporating in-the-wild video data, along with pseudo-groundtruth pose information, improves the network's performance under varying lighting and camera conditions.
3. **Significant Accuracy Improvement:** Substantial increase in SOTA pose estimation accuracy, particularly during pitching actions, underscores the importance of thoughtful augmentation to address motion blur.

Acknowledgement

Our work was supported by the Baltimore Orioles, MLB through the Mitacs Accelerate Program. We also acknowledge the Digital Research Alliance of Canada for their hardware support.

References

[1] Hongsuk Choi, Gyeongsik Moon, and Kyoung Mu Lee. Pose2mesh: Graph convolutional network for 3d human pose and mesh recovery from a 2d human pose. *ECCV 2020*, pages 769–787, 2020.

[2] Kaan Koseler and Matthew Stephan. Machine learning applications in baseball: A systematic literature review. *Applied Artificial Intelligence*, 31:1–19, 02 2018.