Fusing Wearable IMUs with Multi-View Images for Human Pose Estimation: A Geometric Approach

Zhe Zhang[1,2], Chunyu Wang[2], Wenhu Qin[1], Wenjun Zeng[2]
[1]Southeast University, [2]Microsoft Research Asia

Motivation

The Task

Recovering absolute 3D human pose in world coordinate system by fusing Wearable IMUs and Multi-View Images

Main Challenges

It is nontrivial to deeply and effectively incorporate IMUs in the existing image processing pipeline

Contribution

We present an approach to fuse IMUs with images for robust pose estimation even when occlusion occurs

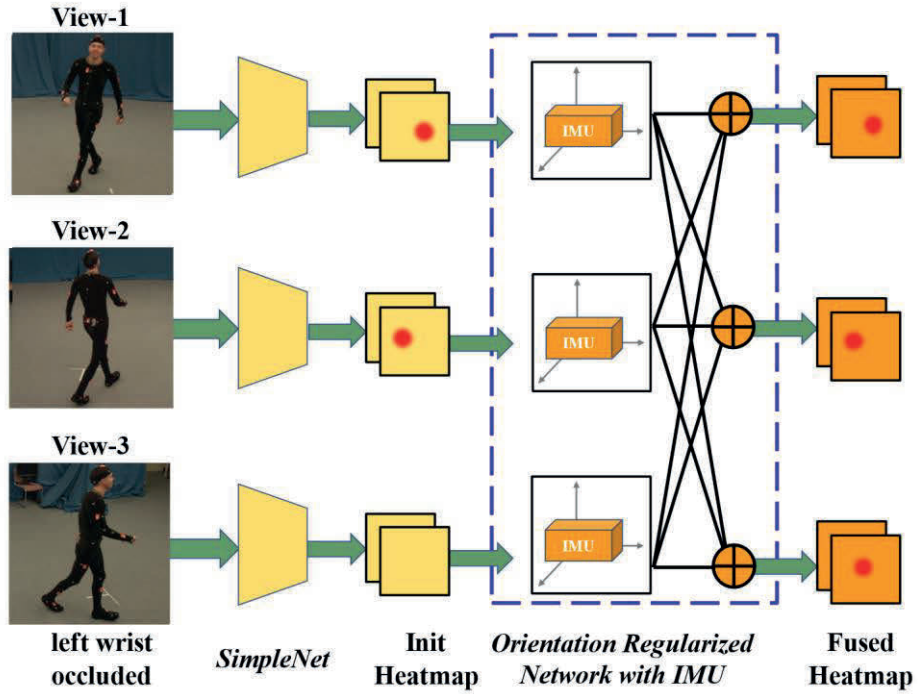
Cross-Joint-Fusion in both **2D & 3D** pose estimation

- *Orientation Regularized Network* (ORN)
  - IMU orientations as a structural prior
  - mutually fuse the image features of each pair of joints linked by IMUs
- *Orientation Regularized Pictorial Structure Model* (ORPSM)
  - an orientation prior that requires the limb orientations of the 3D pose to be consistent with the IMUs

SOTA Results

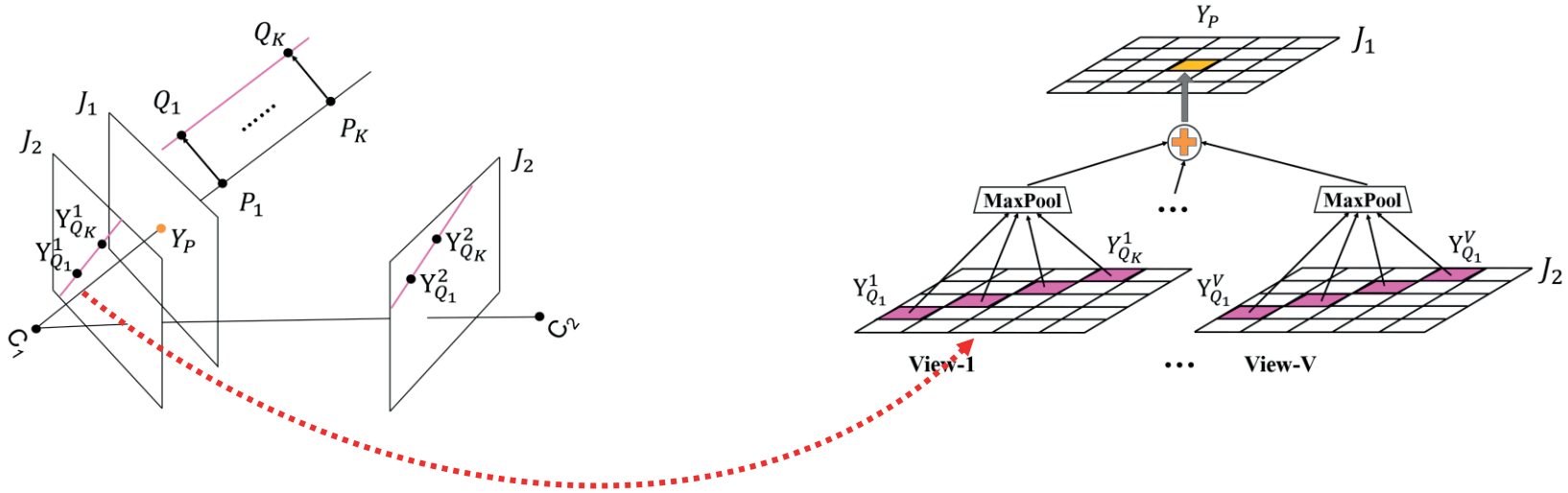
Orientation Regularized Network (ORN)

- It firstly takes multi-view images as input and estimates initial heatmaps independently for each camera view.
- Then with the aid of IMU orientations, it mutually fuses the heatmaps of the linked joints across all views.



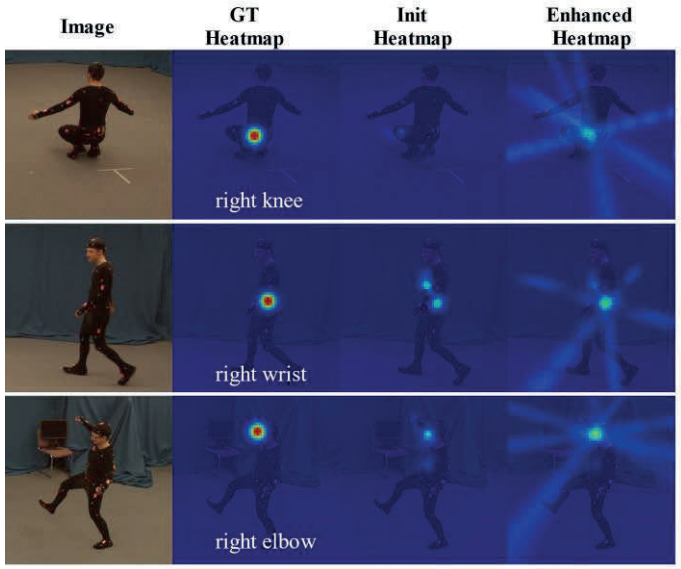
Orientation Regularized Network (ORN)

- The main challenge is to determine the relative positions between each pair of joints (Y_P and Y_Q) in the images
- Since depth is an ambiguity, we find all the possible Y_Q corresponding to Y_P in a line by adding limb offset (*orientation * length*)
- We then enhance Y_P's heatmap by adding maximum response of Y_Q line in each view to it



Orientation Regularized Network (ORN)

- The correct location will be enhanced most, though some irrelevant locations are also enhanced



Orientation Regularized PSM (ORPSM)

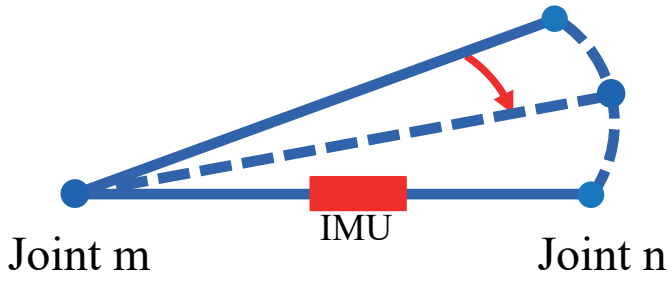
- We introduce a novel limb orientation prior based on IMUs into PSM
- It works as a soft constraint to let limb comply to IMU orientations

Pictorial Structure Model (PSM)

$$p(\mathcal{J}|\mathcal{F}) = \frac{1}{Z(\mathcal{F})} \prod_{i=1}^{M} \phi_i^{\text{conf}}(J_i, \mathcal{F}) \prod_{(m,n)\in\mathcal{E}_{limb}} \psi^{\text{limb}}(J_m, J_n) \prod_{(m,n)\in\mathcal{E}_{IMU}} \psi^{\text{IMU}}(J_m, J_n),$$

Orientation Potential

$$\psi^{\text{IMU}}(J_m, J_n) = \frac{J_m - J_n}{\|J_m - J_n\|_2} \cdot o_{m,n}$$



Experimental Results

- ORN notably improves 2D pose estimation accuracy especially when one joint is occluded.

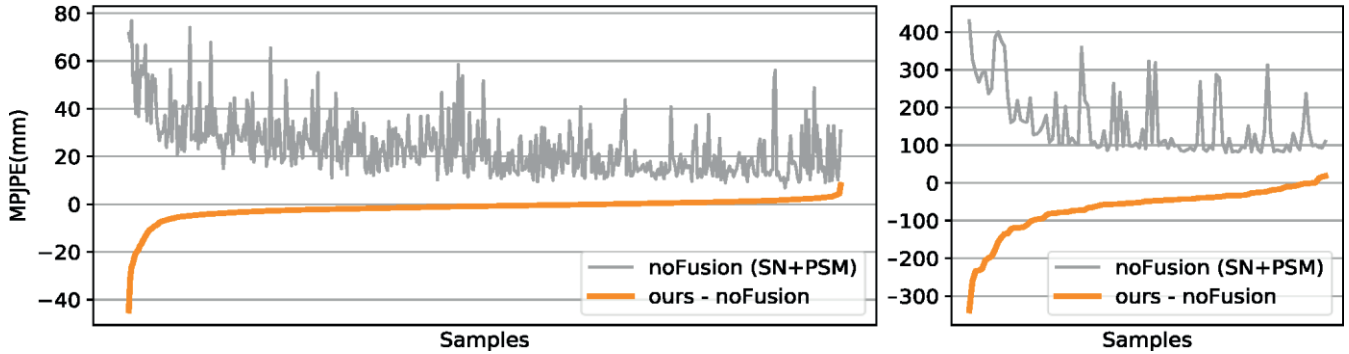
Table 1. The 2D pose estimation accuracy (PCKh@t) on the Total Capture Dataset. "SN" means SimpleNet which is the baseline. ORN^{same} and *ORN*, respectively, represent that the same-view and cross-view fusion are used. "Mean (six)" is the average result over the six joint types. "Others" is the average result over the rest of the joints. "Mean (All)" is the result over all joints.

Methods	PCKh@	Hip	Knee	Ankle	Shoulder	Elbow	Wrist	*Mean (Six)*	Others	Mean (All)
SN	1/2	99.3	98.3	98.5	98.4	96.2	95.3	97.7	99.5	98.1
ORN^{same}	1/2	99.4	99.0	98.8	98.5	97.7	96.7	98.3	99.5	98.6
ORN	1/2	**99.6**	**99.2**	**99.0**	**98.9**	**98.0**	**97.4**	**98.7**	99.5	98.9
SN	1/6	97.5	92.3	92.5	78.3	80.8	80.0	86.9	95.4	89.1
ORN^{same}	1/6	97.2	94.0	93.3	78.1	83.5	82.0	88.0	95.4	89.9
ORN	1/6	**97.7**	**94.8**	**94.2**	**81.1**	**84.7**	**83.6**	**89.3**	95.4	90.9
SN	1/12	**87.6**	67.0	68.6	47.4	50.0	49.3	61.7	78.1	65.8
ORN^{same}	1/12	81.2	70.1	68.0	43.9	51.6	50.1	60.8	78.1	65.2
ORN	1/12	85.3	**71.6**	**70.6**	**47.7**	**53.2**	**51.9**	**63.4**	78.1	67.1

Experimental Results

Table 2. 3D pose estimation errors (mm) of different variants of our approach on the Total Capture dataset. "Mean (six)" is the average error over the six joint types. "Others" is the average error over the rest of the joints. "Mean (All)" is the average error over all joints.

2D	3D	Hip	Knee	Ankle	Shoulder	Elbow	Wrist	*Mean (Six)*	Others	Mean (All)
SN	PSM	**17.2**	35.7	41.2	50.5	54.8	56.8	37.1	20.3	28.3
ORN	PSM	17.4	29.9	35.2	49.6	44.2	45.1	32.8	20.4	25.4
SN	***ORPSM***	18.3	25.8	34.0	44.8	44.2	49.8	32.1	19.9	25.5
ORN	***ORPSM***	18.5	**24.2**	**30.1**	**44.8**	**40.7**	**43.4**	**30.2**	19.8	24.6



- The grey line shows the 3D MPJPE error of the noFusion approach.
- The orange line shows the error difference between our method and noFusion.
- If the orange line is below zero, it means our method has smaller errors.
- We split the testing samples into two groups according to the error scale of noFusion.

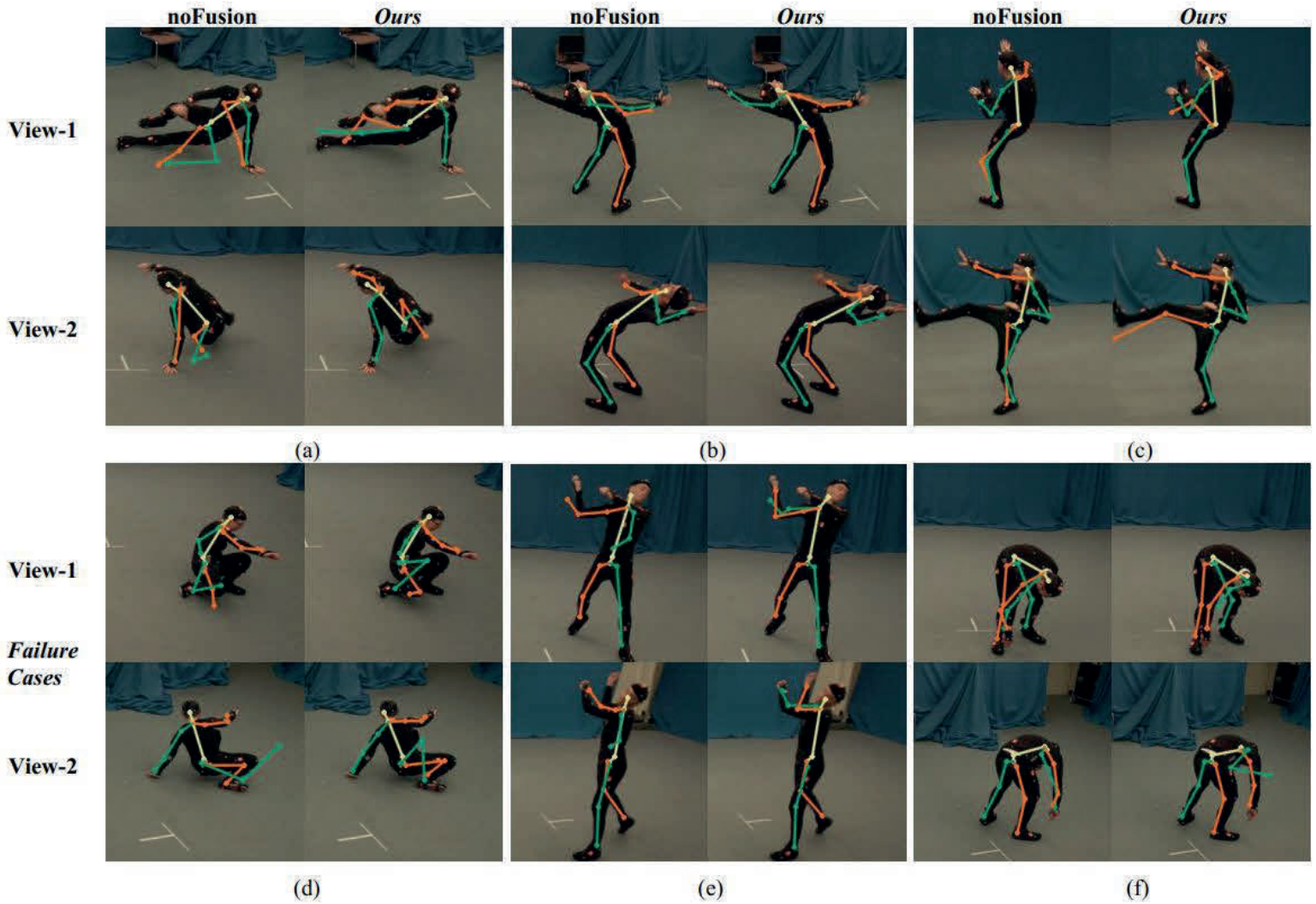
Experimental Results

3D MPJPE comparison with SOTAs on Total Capture dataset

Table 3. 3D pose estimation errors MPJPE (mm) of different methods on the Total Capture dataset. "Aligned" means whether we align the estimated 3D poses to the ground truth poses by Procrustes.

Approach	IMUs	Temporal	Aligned	Subjects(S1,2,3)			Subjects(S4,5)			Mean
				W2	A3	FS3	W2	A3	FS3	
PVH [27]				48.3	94.3	122.3	84.3	154.5	168.5	107.3
Malleson *et al.* [15]	✓	✓		-	-	65.3	-	64.0	67.0	-
VIP [28]	✓	✓	✓	-	-	-	-	-	-	26.0
LSTM-AE [26]		✓		**13.0**	23.0	47.0	**21.8**	40.9	68.5	34.1
IMUPVH [6]	✓	✓		19.2	42.3	48.8	24.7	58.8	61.8	42.6
Qiu *et al.* [19]				19.0	21.0	28.0	32.0	33.0	54.0	29.0
SN + PSM				14.3	18.7	31.5	25.5	30.5	64.5	28.3
SN + PSM			✓	12.7	16.5	28.9	21.7	26.0	59.5	25.3
ORN + ORPSM	✓			14.3	**17.5**	**25.9**	23.9	**27.8**	**49.3**	**24.6**
ORN + ORPSM	✓		✓	12.4	14.6	22.0	19.6	22.4	41.6	20.6

Qualitative Results



Code

aka.ms/imu-human-pose