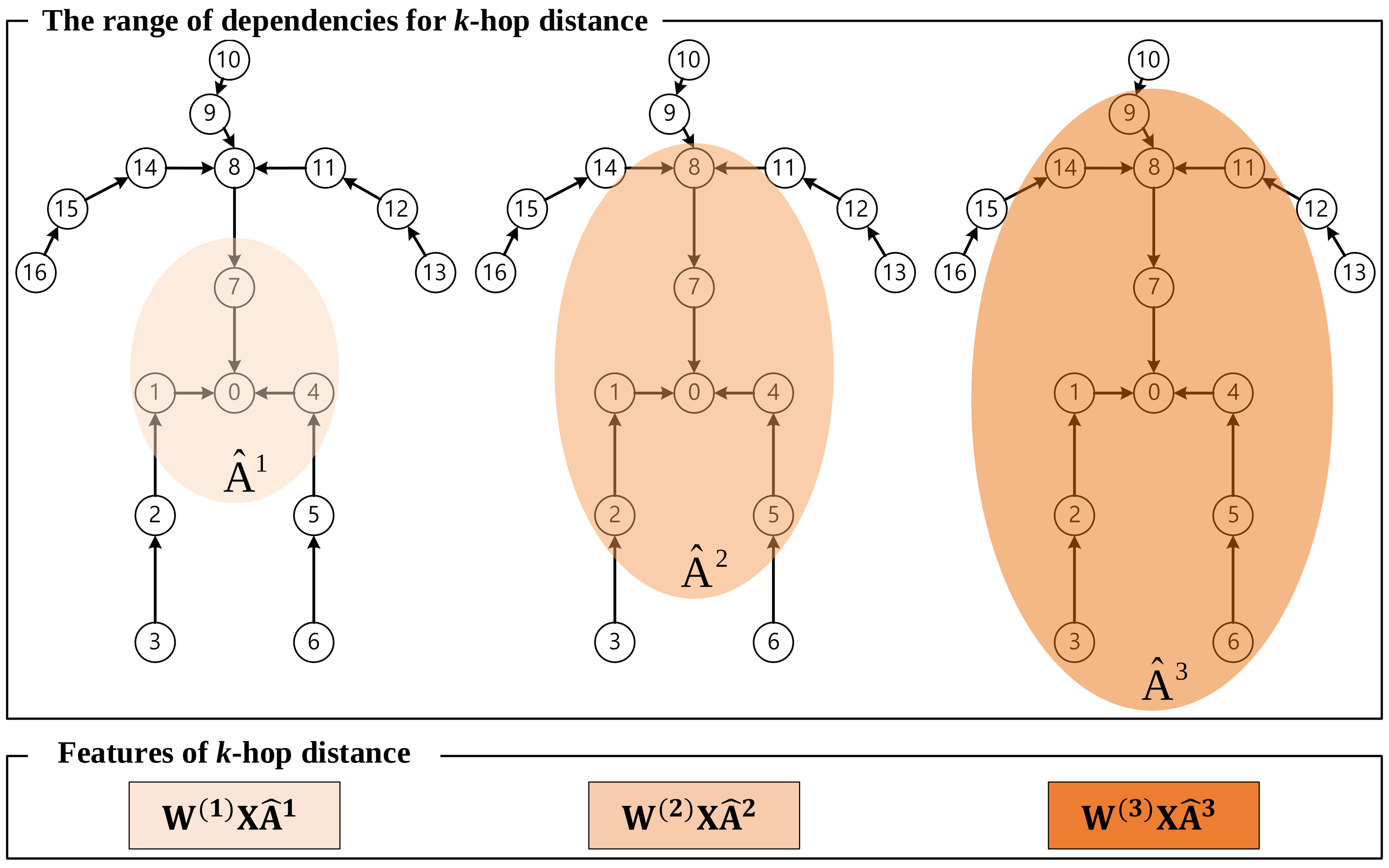
Multi-hop Modulated Graph Convolutional Networks for 3D Human Pose Estimation

Jae Yung Lee, I Gil Kim
AI2XL, Institute of Convergence Technology, KT, Korea

Motivation



(a) 1-hop distance (b) 2-hop distance (c) 3-hop distance

Figure 1. High-order GCN based 3D HPE

- Adjacency Matrix
  - The range of dependencies for each hop distance have high correlation with each other. Thus, it makes difficult to merge features of all *k*-hop distance.
- Aggregate Method
  - *summation* and *concatenation* are insufficient to model the relationships between the aggregate features of each hop distance.
  - *Unshared* weight matrix $W^{(k)}$ and *concatenation* increases number of model parameters significantly.

Proposed Model

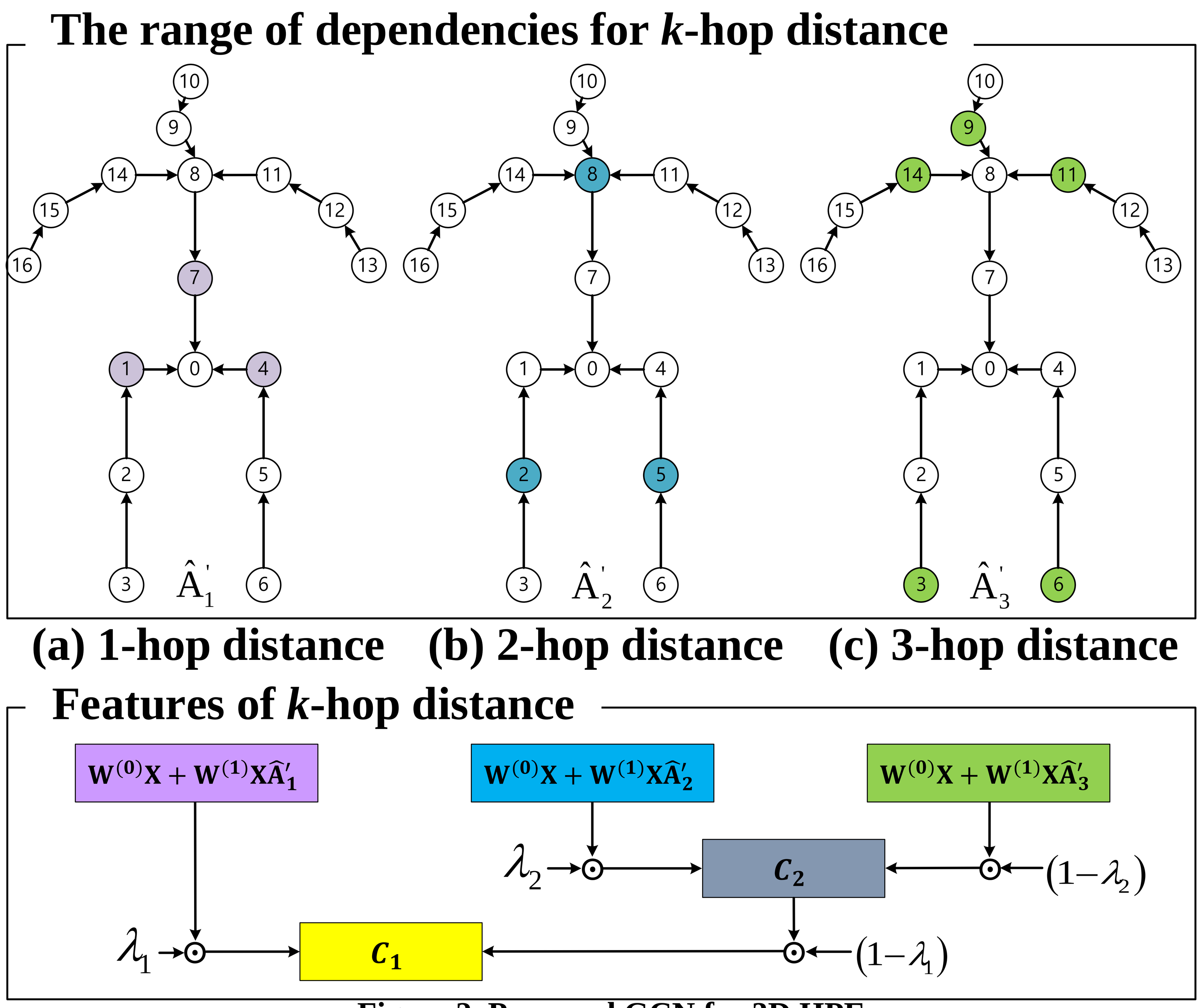


Figure 2. Proposed GCN for 3D HPE

- Adjacency Matrix
  - The proposed adjacency matrix $\hat{A}'_k$ represents the relationships between neighbouring joints, except middle joints up to a distance of *k*-hops, the relationships between the adjacency matrices $\hat{A}'_k$ of each hop have low correlations with each other.
- Aggregate Method
  - $\lambda_k \in R^{D' \times N}$ is a learnable modulation matrix to model the relationships between the features of the k-hop distance and the merged features up to the *(k+1)*-hop distance C_{k+1}.

$$C_k = \lambda_k \odot \left(W^{(0)}H + W^{(1)}H\hat{A}'_k \right) + \left(1-\lambda_k\right) \odot C_{k+1}$$

$$H' = \sigma\left(\lambda_1 \odot \left(W^{(0)}H + W^{(1)}H\hat{A}'_1 \right) + (1-\lambda_1) \odot C_2 \right)$$

  - The MM-GCN is designed such that the features become more heavily weighted as the hop distance becomes shorter.

- Network Architecture

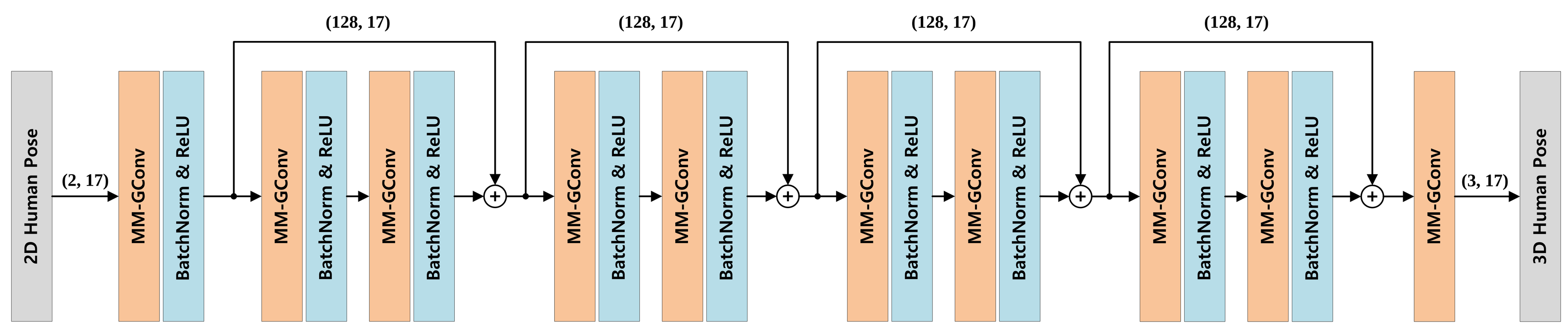


Figure 3. Network architecture of proposed MM-GCN for 3D HPE

- Network architecture of proposed MM-GCN for 3D HPE. (D, N) indicates feature channels and number of body joints, respectively.

Experiments & Results

- Experiments Setting
  - **Dataset** : We evaluate our approach on the Human3.6M dataset. To demonstrate the generalizability of our model quantitatively, we evaluated our model on the testing set of MPI-INF-3DHP after the model was trained on Human3.6M.
  - **Implementation Detail** :

	2D ground truth	2D pose detection
# of Channels	128	384
Non-local layer	X	O
Bath size	1024	
# of Epochs	200	

- Ablation Study

Method	# of Channels	# of Parameters	MPJPE	P-MPJPE
SemGCN	128	0.27M	42.14	33.53
SemGCN w/ Non-local	128	0.43M	40.78	31.46
Modulated GCN w/o AM	128	0.27M	38.83	30.35
Modulated GCN	128	0.29M	38.25	30.06
Ours (2-hop) w/ AM	128	0.31M	38.36	29.64
Ours (3-hop) w/ AM	128	0.33M	37.41	29.31
Ours (4-hop) w/ AM	128	0.36M	36.10	28.76
Ours (5-hop) w/ AM	128	0.38M	35.63	27.55

Table 1. Performance comparison of proposed MM-GCN and various GCN-based methods.

- Qualitative Results

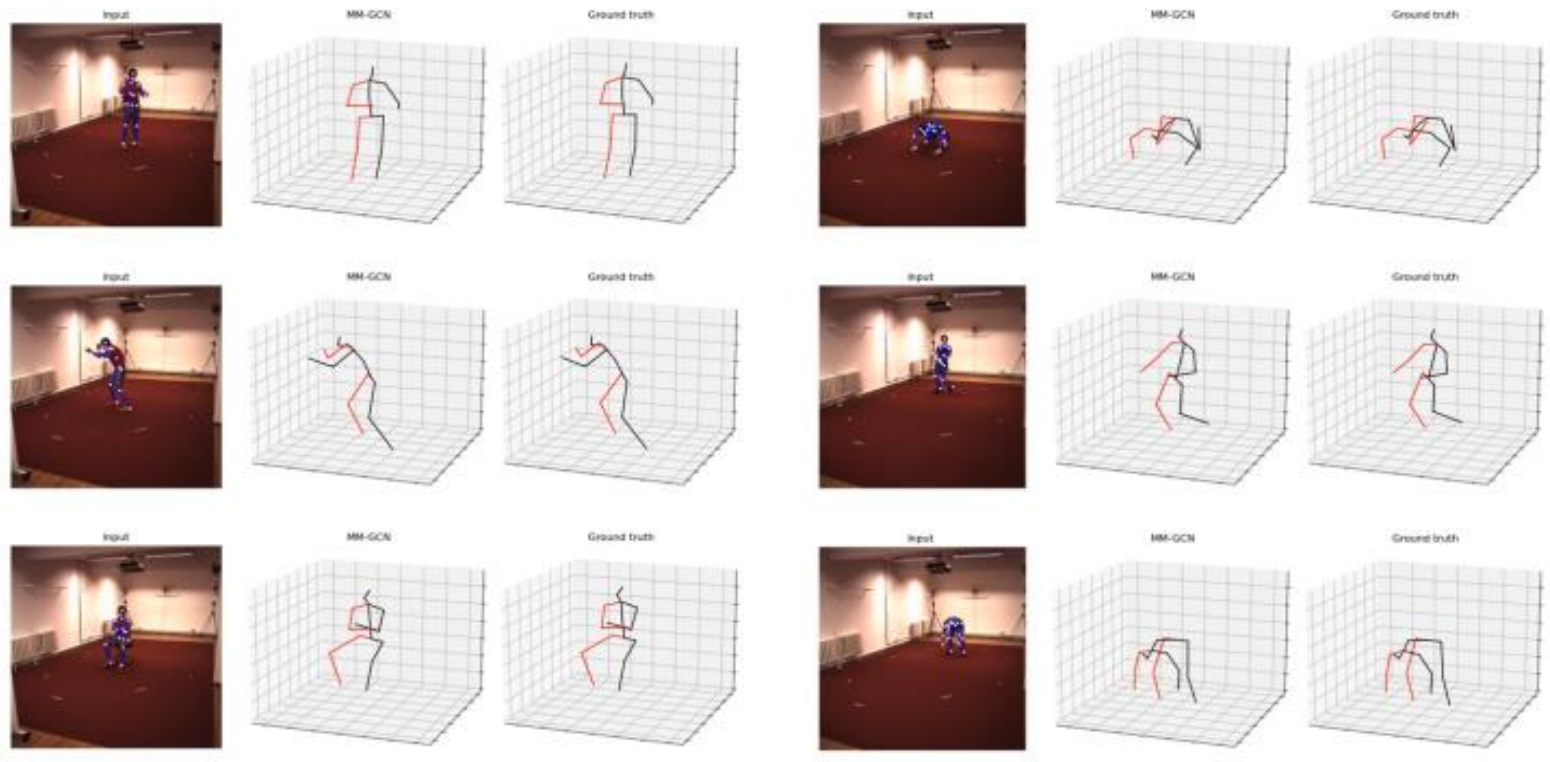


Figure 4. Qualitative results obtained by our MM-GCN on the Human3.6M test set

- Quantitative Results

Method	Dire.	Disc.	Eat	Phon.	…	WalkT.	Avg.
Martinez	51.8	56.2	58.1	69.5	…	52.4	62.9
Sun	52.8	54.8	54.2	61.8	…	53.4	59.1
Yang	51.5	58.9	50.4	62.1	…	47.7	58.6
Fang	50.1	54.3	57.0	66.6	...	50.6	60.4
Pavlakos	48.5	54.4	54.4	59.4	…	47.8	56.2
Zhao	47.3	60.7	51.4	61.1	…	45.3	57.6
Sharam	48.6	54.5	54.2	62.2	…	49.7	58.0
Zou	49.0	54.5	52.3	59.2	…	45.4	55.6
Quan	47.0	53.7	50.9	57.8	…	45.3	54.8
Liu	**46.3**	52.2	47.3	55.5	…	43.7	52.4
Zou&Tang	48.2	51.6	47.8	53.1	…	42.6	52.4
Ours(3-hop)	46.8	**51.4**	**46.7**	**52.5**	…	**42.2**	**51.7**

Table 2. Quantitative comparisons on Human3.6M under MPJPE.

Method	Dire.	Disc.	Eat	Phon.	…	WalkT.	Avg.
Martinez	39.5	43.2	46.4	51.0	…	43.1	47.7
Sun	42.1	44.3	45.0	51.5	…	44.8	48.3
Fang	38.2	41.7	43.7	48.5	…	41.7	45.7
Pavlakos	34.7	39.8	41.8	42.5	...	36.5	41.8
Hossain&Little	35.7	**39.3**	44.6	47.2	…	39.4	44.1
Zou	38.6	42.8	41.8	44.6	…	37.9	43.7
Quan	36.9	42.1	40.3	43.7	…	37.8	42.9
Liu	35.9	40.0	38.0	42.5	…	36.2	41.2
Zou&Tang	36.6	40.1	37.7	41.0	…	34.9	41.0
Ours(3-hop)	**35.7**	39.6	**37.3**	**40.0**	…	**34.7**	**40.3**

Table 3. Quantitative comparisons on Human3.6M under P-MPJPE.

Method	3DPCK	AUC
Yang	69.0	32.0
Pavlakos	71.9	35.3
Habibie	70.4	36.0
Wang	71.9	35.8
Quan	72.8	36.5
Liu	79.3	47.6
Ours	81.6	50.3

Table 4. Quantitative comparisons on MPI-INF-3DHP.

- The proposed MM-GCN performed the best on most of the tasks, and on average, under both MPJPE and P-MPJPE.

Conclusions

We introduced an MM-GCN for 3D HPE to effectively model long-range dependencies between each body part and its distant neighbours. We performed experiments to demonstrate its competitive performance in comparison with state-of-the-art methods for 3D HPE.