ICPR2020

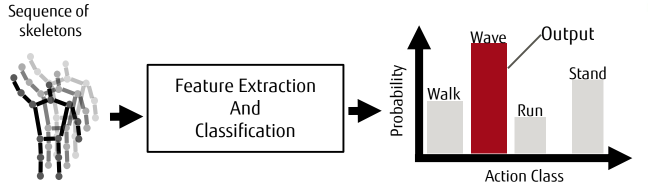
FUJITSU
shaping tomorrow with you

Temporal Extension Module for Skeleton-based Action Recognition

Yuya Obinata and Takuma Yamamoto (FUJITSU LABORATORIES LTD.)

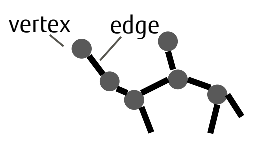
1. Overview

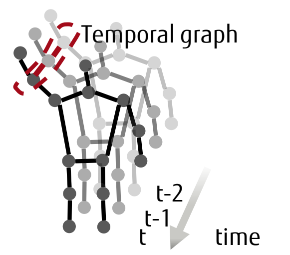
- Expectation of Action recognition
  - Applications in human behavior understanding and human social behavior
- Mainstream method in action recognition
  - Methods using RGB images
  - Methods using sequence of skeletons
    - Low-cost computing resources
    - Robust by noise from various illumination conditions



2. Motivation

- ST-GCN (AAAI 2018)
  - The well-known method using graph convolutional network
    - Represents human joints as vertices and their natural connections in human body as edges
    - Draws attention owing of high performance
  - Temporal graph
    - Adds edges to same vertex on temporal dimension
    - Extracts feature of trajectory of same joints
  - Drawback
    - Cannot extract feature of correlative movement between each joint on the inter-frame
      - Does not add edges to another vertex on temporal graph

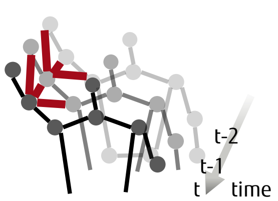




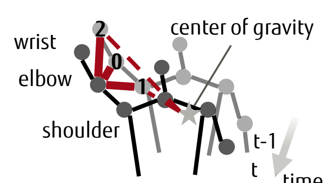
Temporal graph of elbow

3. Proposed Method

- Temporal Extension Module (TEM)
  - Extends of temporal graph on inter-frame
  - Adds edges to not only same vertex but also neighboring multiple vertices
  - Sets multiple subsets based on length to center of gravity
    - Adds info of kinematics correlation
  - Advantage
    - Extracts feature of correlative movement between each joint on inter-frame
    - Good for recognition of action that these adjacent joints often move together, such as "throw."
  - Apply to existing model easily



Temporal graph with TEM

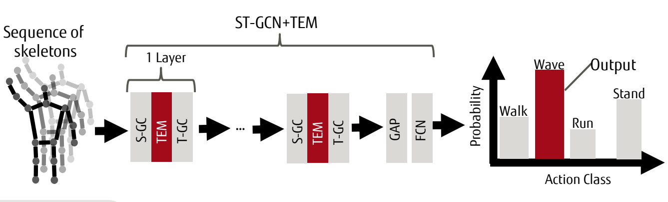


Subset on inter-frame

4. Implementation of TEM

- ST-GCN+TEM
  - We attach our module between *S-GC and T-GC.
    - To expand sampling area for temporal dimension gradually
  - We do not change the structure of conventional convolutions.
    - Our module can readily apply to many existing methods.

*Spacial Graph Convolution and Temporal Graph Convolution



5. Ablation Study

- Methods with TEM outperform without TEM.

TABLE I. Comparisons of the Recognition Accuracy the Models with TEM and without TEM

Methods	NTU-RGB+D		Kinetics-Skeleton	
	CS (%)	*CV (%)*	*Top-1 (%)*	*Top-5 (%)*
ST-GCN	82.6	88.7	32.5	54.9
ST-GCN+TEM	85.2	90.2	34.5	56.7
2s-AGCN	88.6	95.2	36.7	59.8
2s-AGCN+TEM	88.7	95.8	38.6	61.6
MS-AAGCN	90.3	96.1	37.4	60.6
MS-AAGCN+TEM	91.0	96.5	38.0	61.4

6. Comparison of State-of-the-art

- Best performance model with TEM achieves state-of-the-art performance.

TABLE II. Comparisons of the Recognition Accuracy with MS-AAGCN+TEM and Current State-of-the-art Methods on NTU RGB+D Dataset

Methods	NTU-RGB+D	
	CS (%)	*CV (%)*
ST-GCN [7]	81.5	88.3
2s-AGCN [9]	88.5	95.1
GCN-NAS [11]	89.4	95.7
DGNN [10]	89.9	96.1
MS-AAGCN [12]	90.0	96.2
BAGCN [22]	90.3	96.3
Sym-GNN [13]	90.1	96.4
MS-AAGCN+TEM(Ours)	**91.0**	**96.5**

TABLE III. Comparisons of the Recognition Accuracy with 2s-AGCN+TEM and Current State-of-the-art Methods on the Kinetics-Skeleton

Methods	Kinetics-Skeleton	
	Top-1 (%)	*Top-5 (%)*
ST-GCN [7]	30.7	52.8
2s-AGCN [9]	36.1	58.7
DGNN [10]	36.9	59.6
GCN-NAS [11]	37.1	60.1
Sym-GNN [13]	37.2	58.1
BAGCN [22]	37.3	60.2
MS-AAGCN [12]	37.8	61.0
2s-AGCN+TEM (Ours)	**38.6**	**61.6**