Global Co-occurrence Feature Learning and Active Coordinate System Conversion for Skeleton-based Action Recognition
Sheng Li, Tingting Jiang, Tiejun Huang, Yonghong Tian
NELVT, Department of Computer Science, Peking University, China

INTRODUCTION

Motivation & Background:
1) Most skeleton-based action recognition methods do not fully explore the spatio-temporal co-occurrence, and many methods use separate strategies for spatial and temporal feature extraction.
2) On the other hand, the current method lacks some common modules compatible with multiple methods.

Idea & Contribution:
1) We propose a new spatio-temporal co-occurrence feature enhancement method. It can extract feature across spatial-temporal domain.
2) We design an active coordinate system transformation that can better align the skeleton data for action recognition. And it is compatible with most CNN-based or GCN-based methods.

METHOD

Active Coordinate System Conversion (ACSC):
Step1 (pretrain):
Pre-trained ACSC so that it can fit the transformation process from Cartesian coordinates to Cylindrical coordinates.

Step2 (End-to-end training):
Use the ground truth of the action label to supervise ACSC to complete the final training.

Spatio-temporal-unit Feature Enhancement (STUFE):
First step a) is to calculate the distance between every two spatio-temp units. Then all the results can be obtained after the step c). Finally, the global co-occurrence feature enhancement is completed by self-attention mechanism according to the distance.

NTU-RGB+D:
Ablation study on the SBU Kinect Interaction dataset and NTU-RGB+D dataset CCS benchmark.

Training time comparison:
Top-1 accuracies on NTU-RGB+D and SBU Kinect Interaction benchmarks. Both methods achieved high performance.

Ablation study:
Ablation study on the SBU Kinect Interaction dataset and NTU-RGB+D dataset CCS benchmark. CCS refers to the cylindrical coordinate system.

Related Video Link: