IEEE *
DCOSS 2021
INTERNATIONAL CONFERENCE ON DISTRIBUTED COMPUTING IN SENSOR SYSTEMS

M&M: Recognizing Multiple Co-evolving Activities from Multi-source Videos

Lan Zhang, Mu Yuan, Daren Zheng, Xiang-Yang Li

University of Science and Technology of China

中国科学技术大学
University of Science and Technology of China

Lab for Intelligent Networking and Knowledge Engineering

Outline

- **Introduction**
- 3D Poses Reconstruction
- Co-evolving Activity Representation & Recognition
- Implementation & Evaluation

Introduction

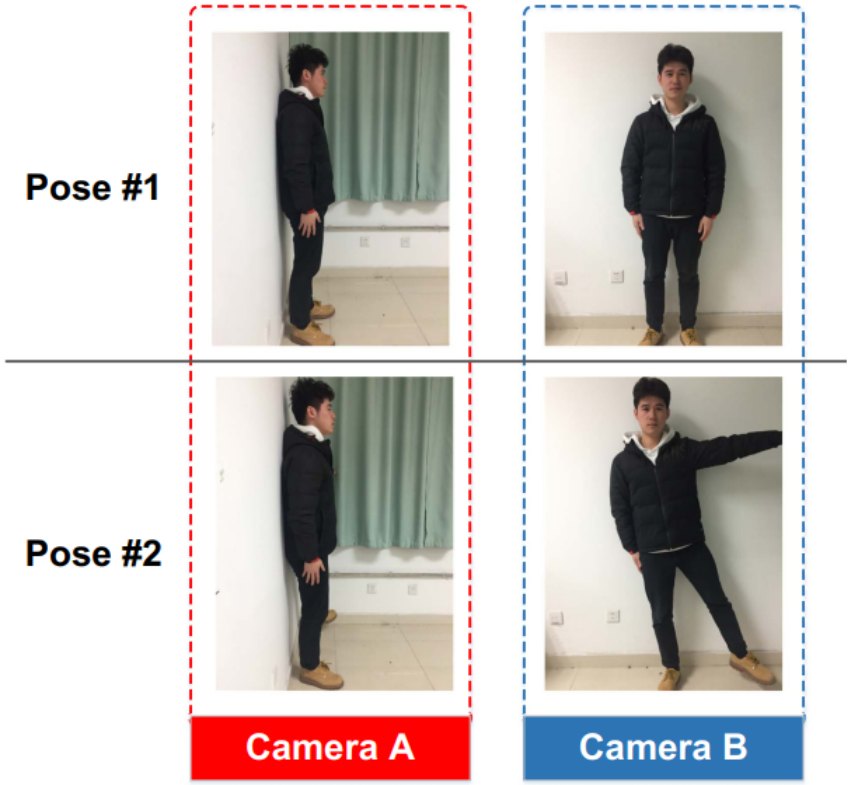
Video-based Activity Recognition

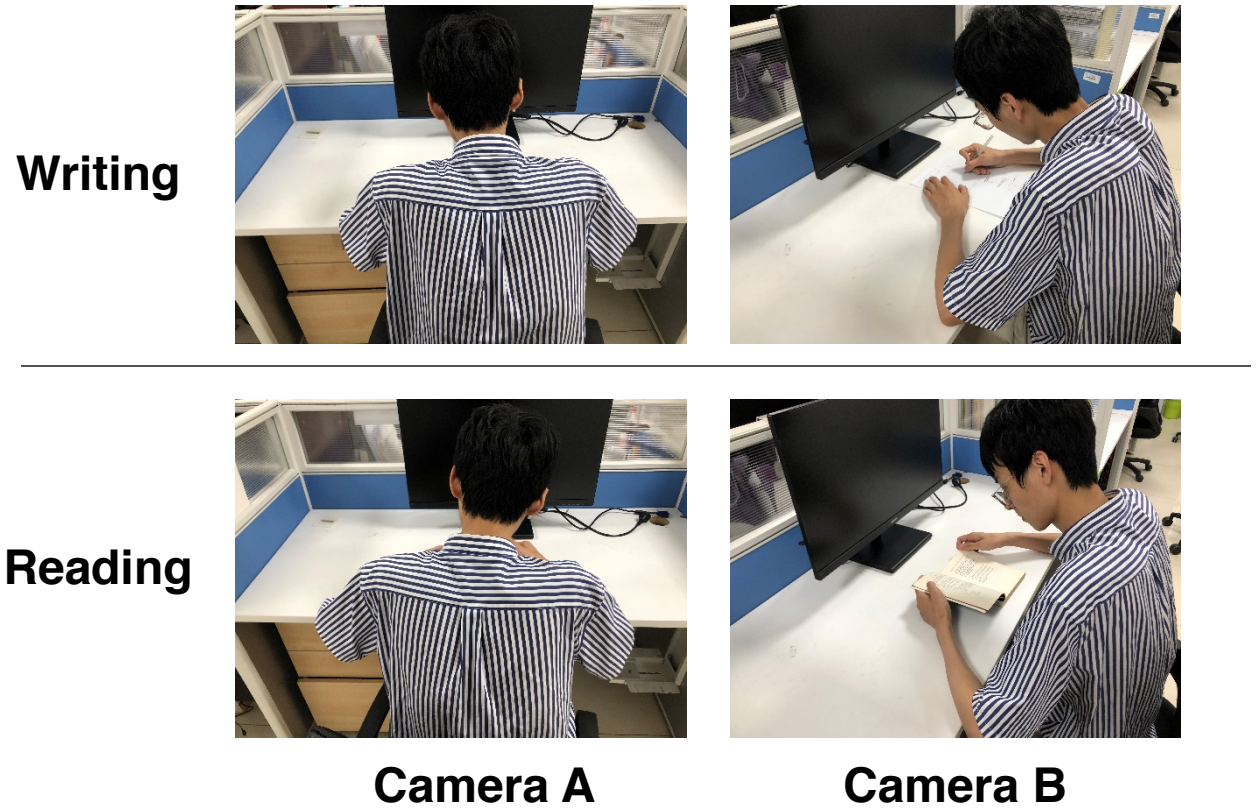
- Babysitting
- Elderly care
- Security
- …

Introduction

One-view Indistinguishability

- Human poses and actions could be indistinguishable from one view.

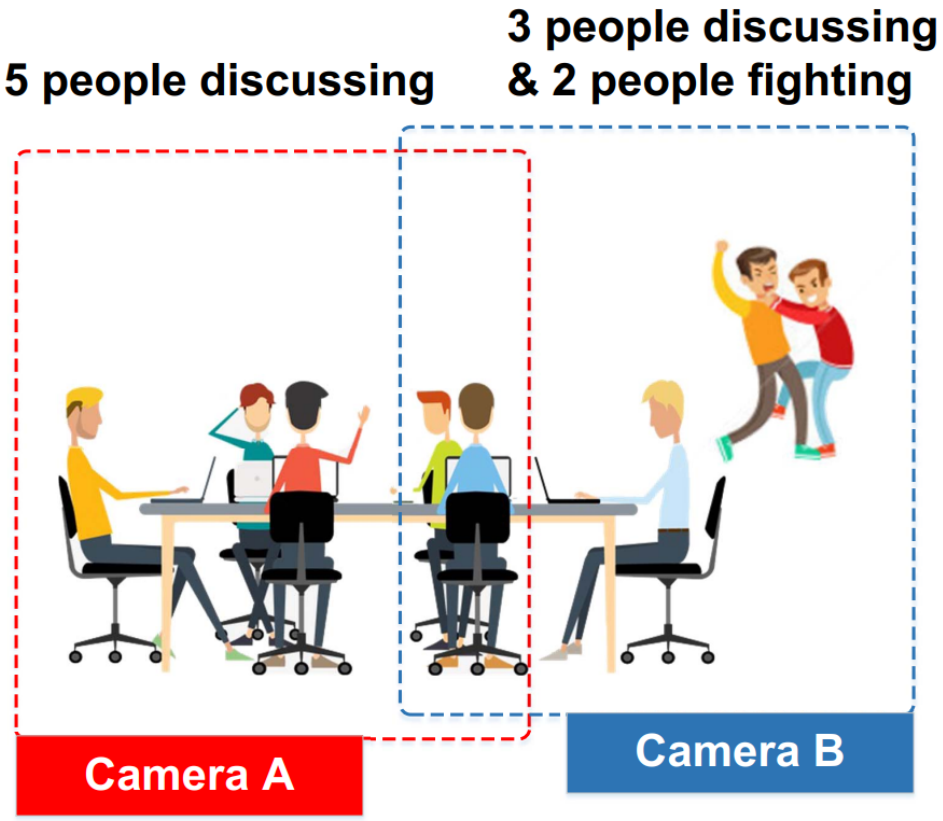


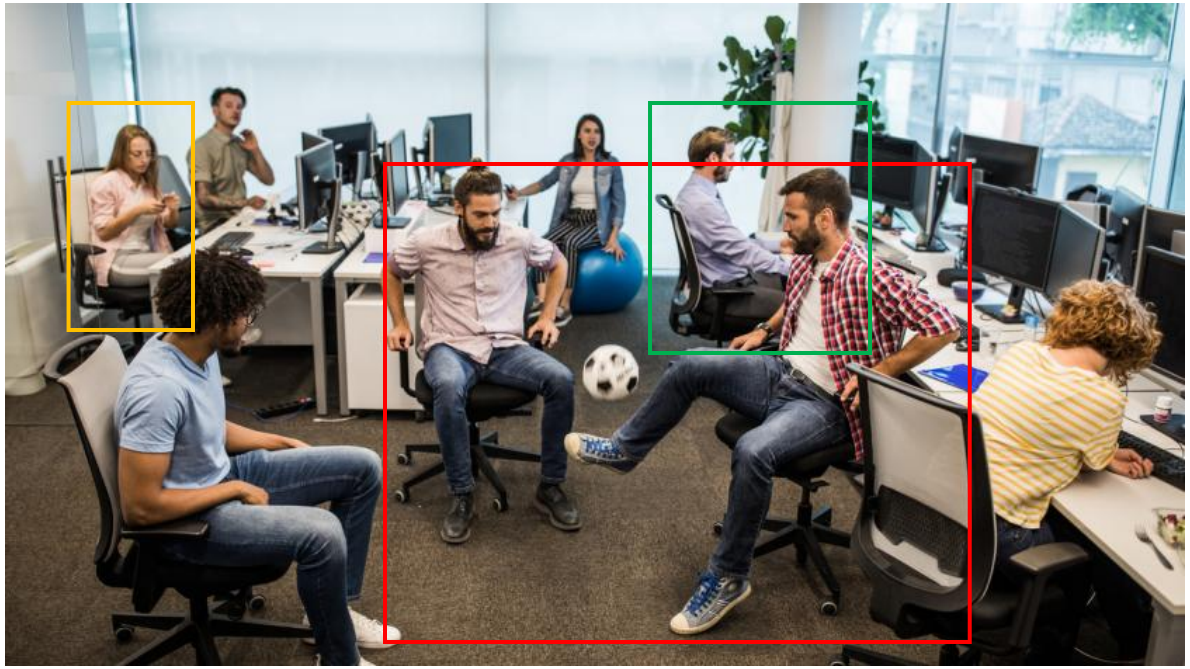


Introduction

Co-evolving Activities

- Multiple activities co-evolve in one scenario.



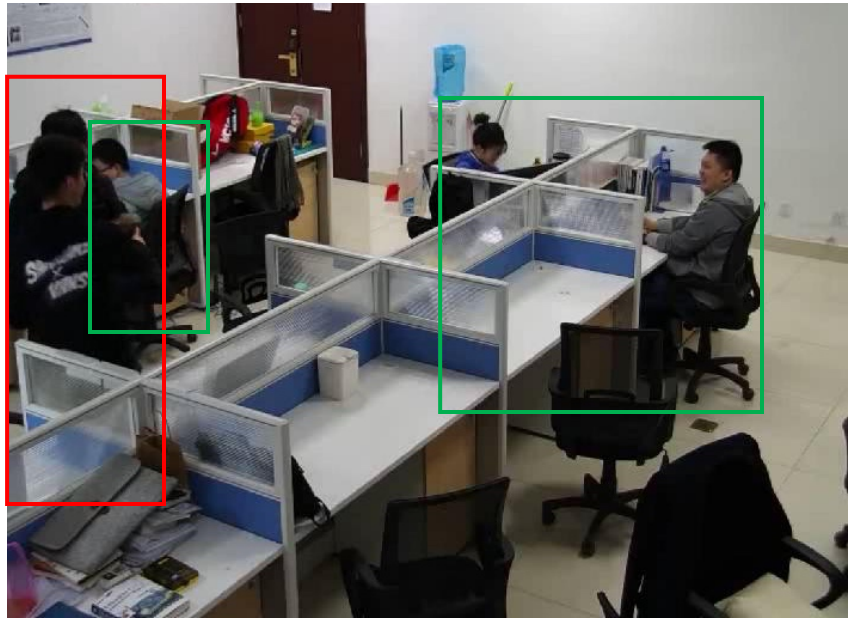


Introduction

Recognizing Cross-view & Co-evolving Activities

- In this work, we study the recognition task of cross-view and co-evolving activities.

Camera A



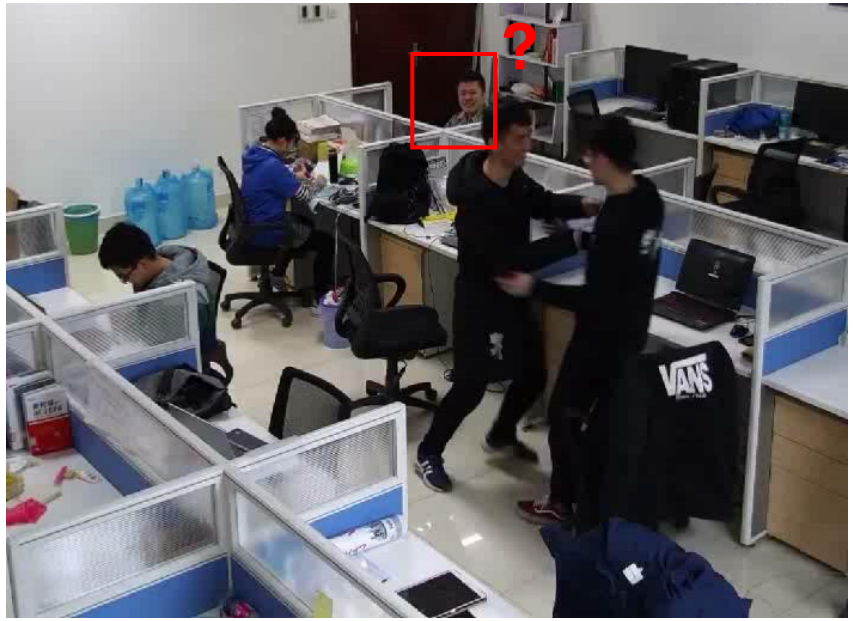
Camera B

2 students are fighting
3 students are working

Introduction

Challenges

- One-view indistinguishability



Camera A

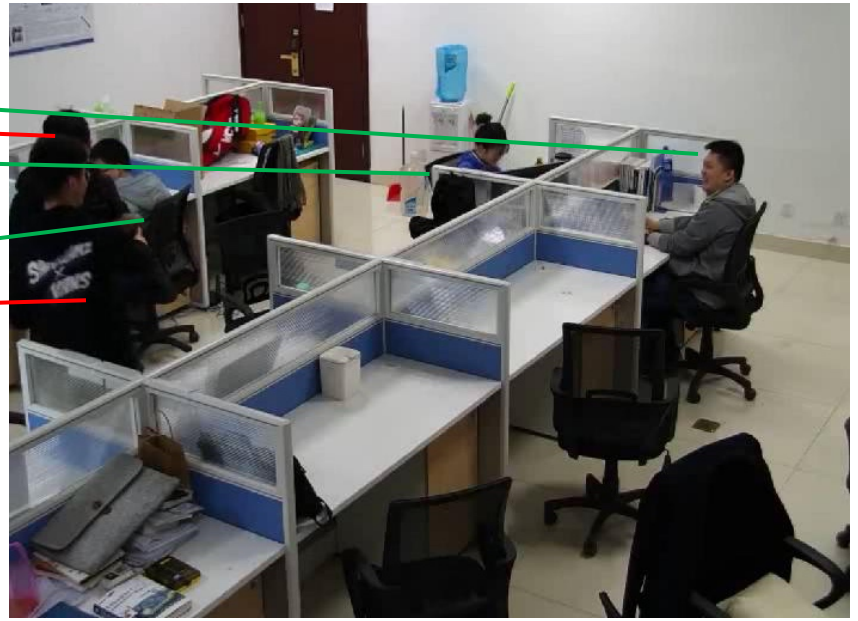
Camera B

Introduction

Challenges

- One-view indistinguishability
- Cross-view alignment

Camera A

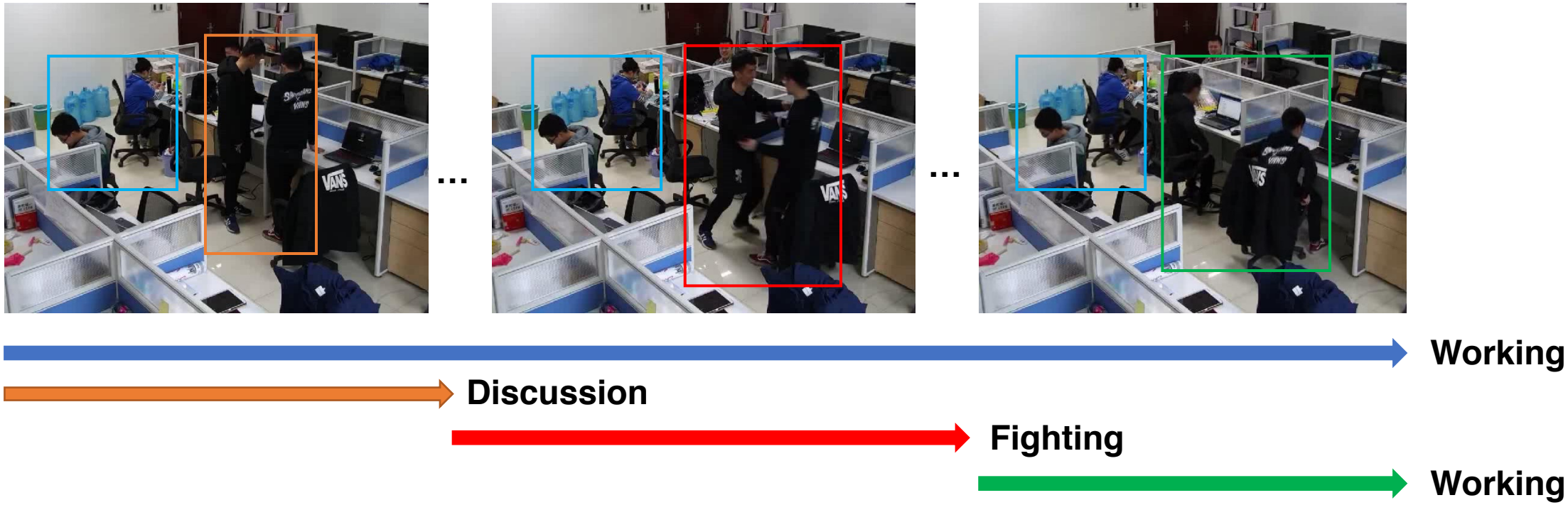


Camera B

Introduction

Challenges

- One-view indistinguishability
- Cross-view alignment
- Asynchronous activities



Introduction

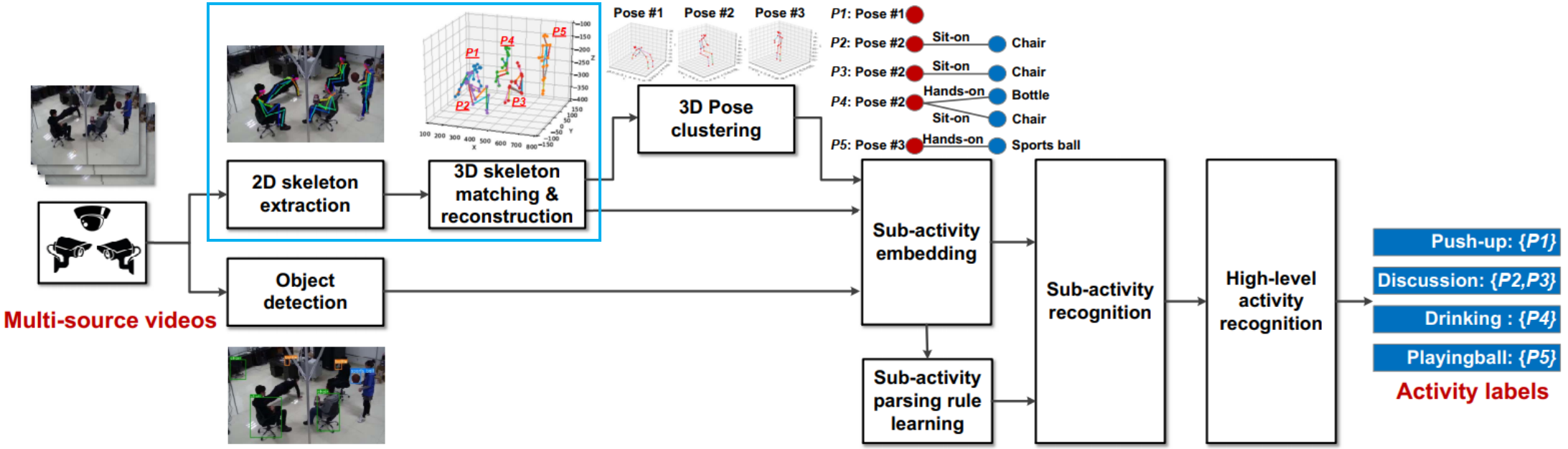
M&M

- The first framework for recognizing **M**ultiple co-evolving activities from **M**ulti-source videos.

Introduction

M&M

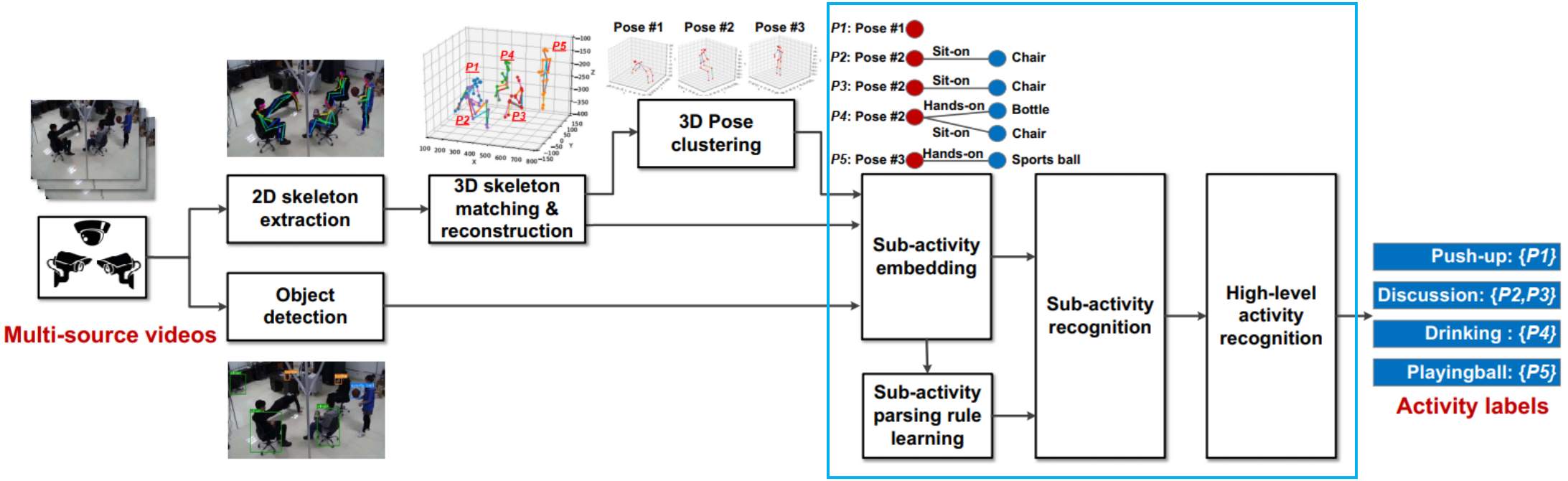
- The first framework for recognizing Multiple co-evolving activities from Multi-source videos.
- 3D pose reconstruction: eliminates one-view indistinguishability & aligns cross-view information.



Introduction

M&M

- The first framework for recognizing Multiple co-evolving activities from Multi-source videos.
- 3D pose reconstruction: eliminates one-view indistinguishability & aligns cross-view information.
- Graph-based Multi-activity embedding: asynchronous activities as sub-graphs



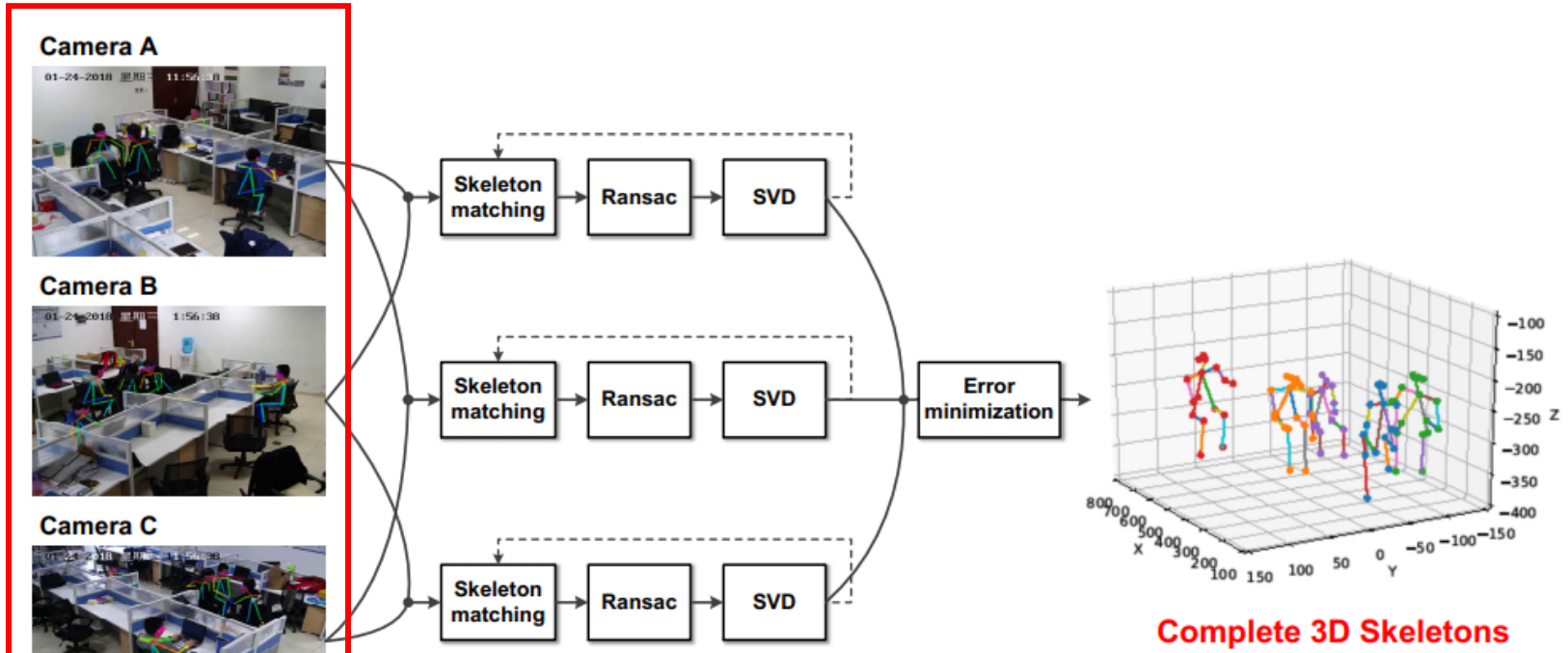
Outline

- Introduction
- **3D Poses Reconstruction**
- Co-evolving Activity Representation & Recognition
- Implementation & Evaluation

3D Poses Reconstruction

2D Skeleton Extraction

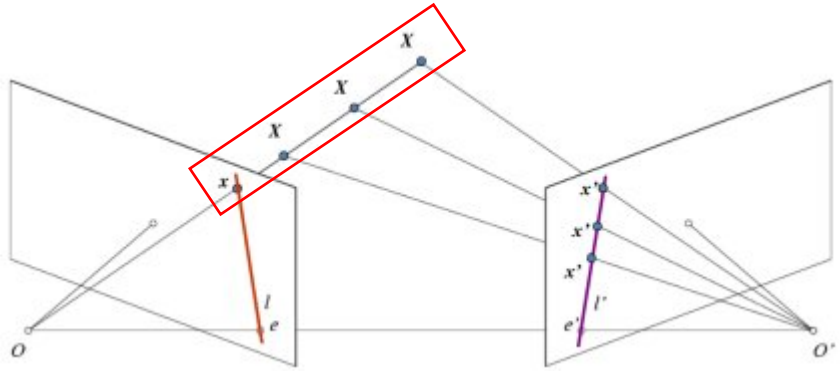
- pose estimation model



3D Poses Reconstruction

3D Skeleton Reconstruction

- Multi-view geometry



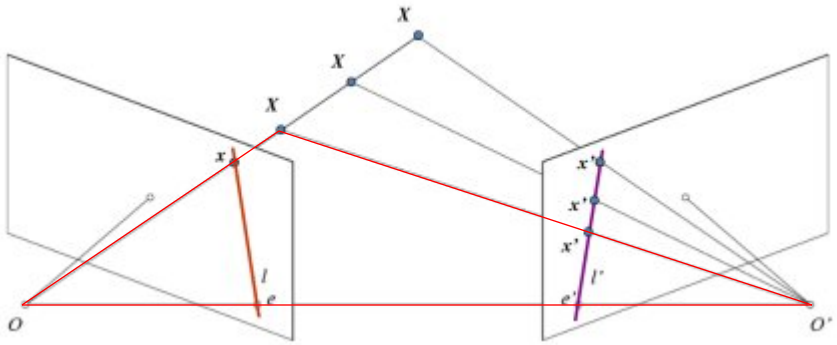
From only one view, we cannot find the 3D point of x.

https://docs.opencv.org/master/da/de9/tutorial_py_epipolar_geometry.html

3D Poses Reconstruction

3D Skeleton Reconstruction

- Multi-view geometry



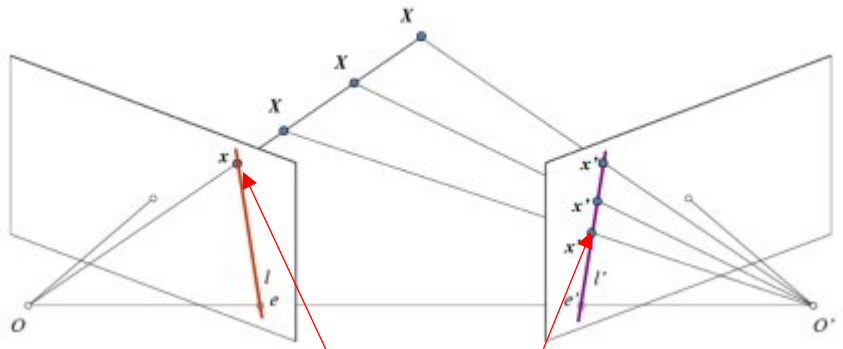
With the help of another view, we can triangulate the correct 3D point.

https://docs.opencv.org/master/da/de9/tutorial_py_epipolar_geometry.html

3D Poses Reconstruction

3D Skeleton Reconstruction

- Fundamental matrix



$$x'^{T} F x = 0$$

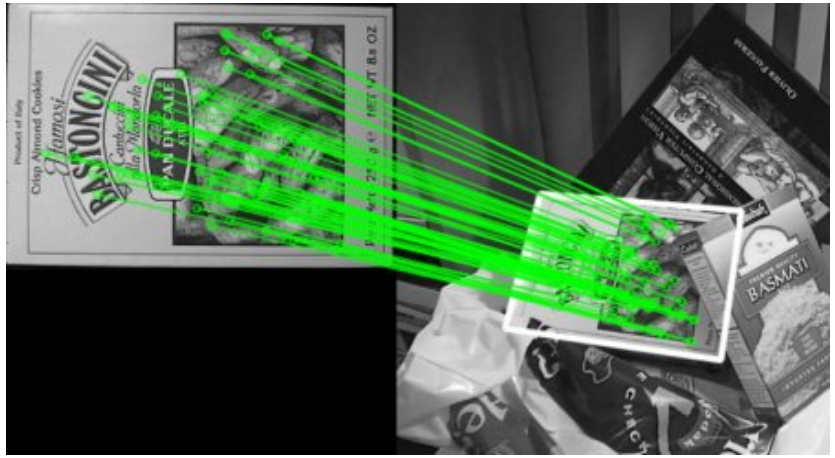
https://docs.opencv.org/master/da/de9/tutorial_py_epipolar_geometry.html

3D Poses Reconstruction

3D Skeleton Reconstruction

- Reconstruction from two views:
  I. Compute the fundamental matrix *F* from point correspondences
  II. Compute the camera matrices from *F*
  III. For each point correspondence, compute the 3D point

View#1 **View#2**



https://docs.opencv.org/4.5.2/d1/de0/tutorial_py_feature_homography.html

3D Poses Reconstruction

3D Skeleton Reconstruction

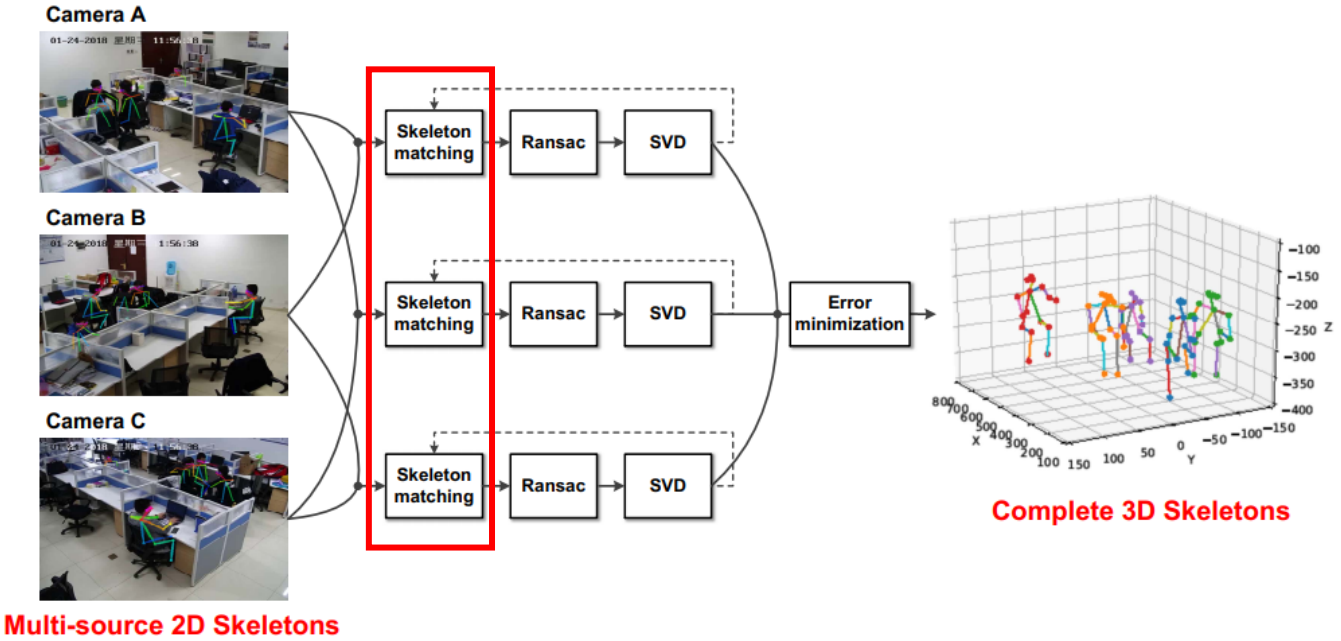
- Reconstruction from two views:
  I. Compute the fundamental matrix *F* from point correspondences
  II. Compute the camera matrices from *F*
  III. For each point correspondence, compute the 3D point

View#1 **View#2**

3D Poses Reconstruction

3D Skeleton Reconstruction

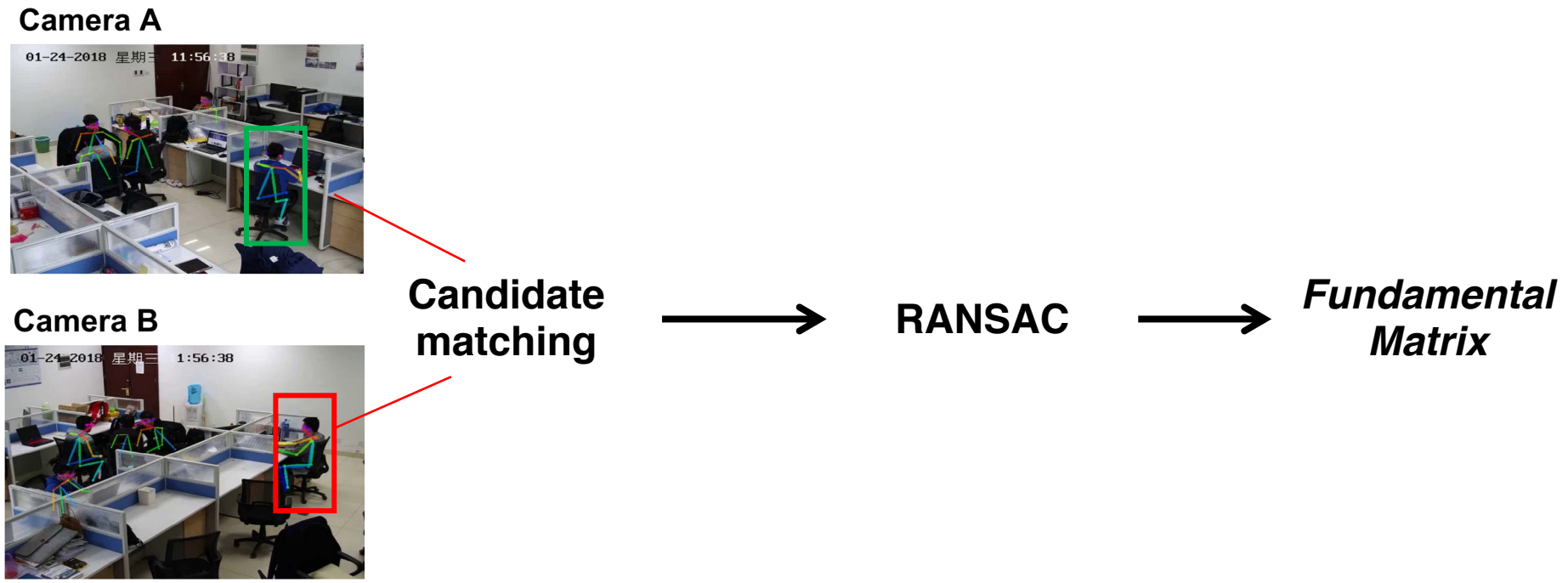
- Reconstruction from two views:
  I. Compute the fundamental matrix *F* from point correspondences
  II. Compute the camera matrices from *F*
  III. For each point correspondence, compute the 3D point



3D Poses Reconstruction

3D Skeleton Reconstruction

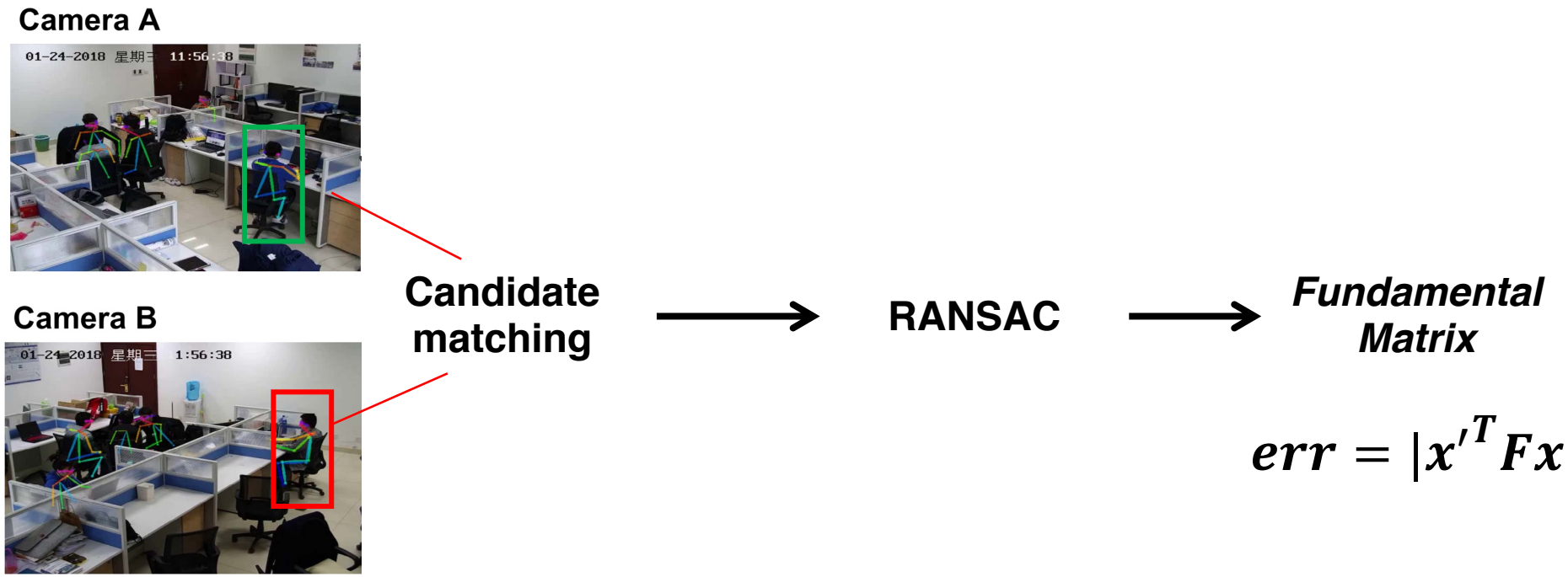
- Reconstruction from two views:
  I. Compute the fundamental matrix *F* from point correspondences
  II. Compute the camera matrices from *F*
  III. For each point correspondence, compute the 3D point



3D Poses Reconstruction

3D Skeleton Reconstruction

- Reconstruction from two views:
  I. Compute the fundamental matrix *F* from point correspondences
  II. Compute the camera matrices from *F*
  III. For each point correspondence, compute the 3D point



$$err = |x'^T F x|$$

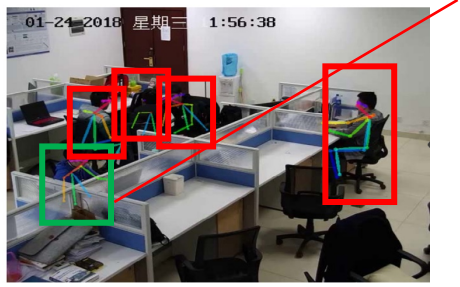
3D Poses Reconstruction

3D Skeleton Reconstruction

- Reconstruction from two views:
  I. Compute the fundamental matrix *F* from point correspondences
  II. Compute the camera matrices from *F*
  III. For each point correspondence, compute the 3D point

Camera A

Camera B

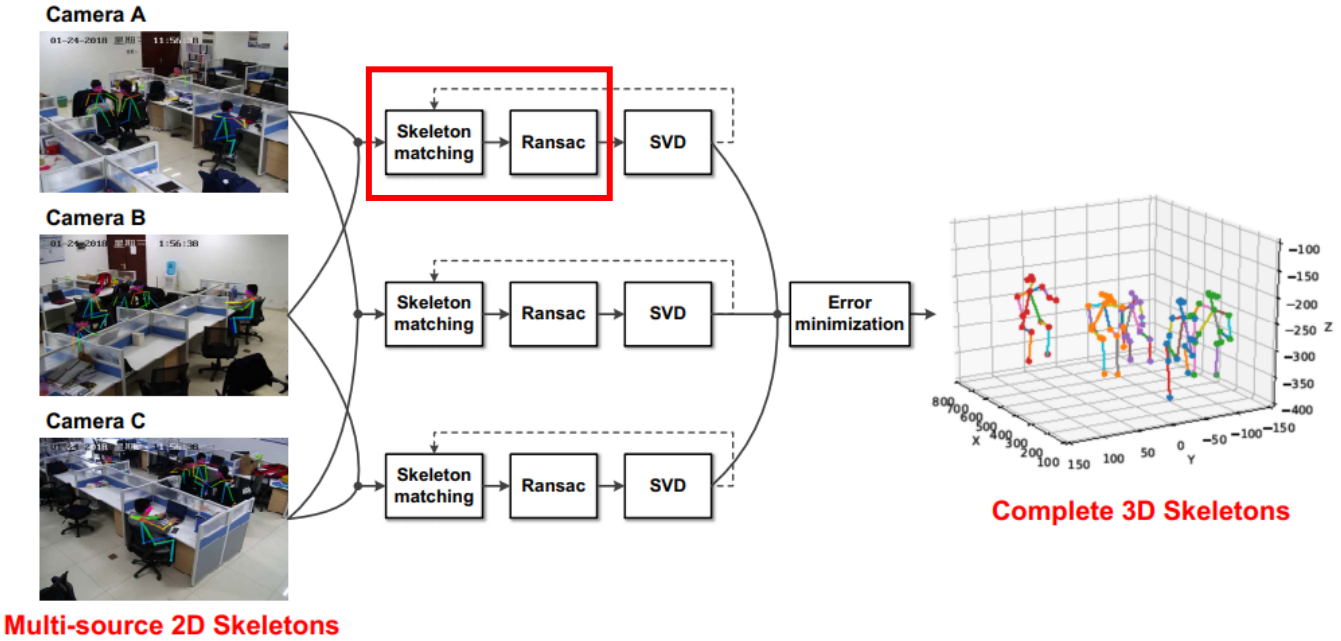


The matching with minimal error

3D Poses Reconstruction

3D Skeleton Reconstruction

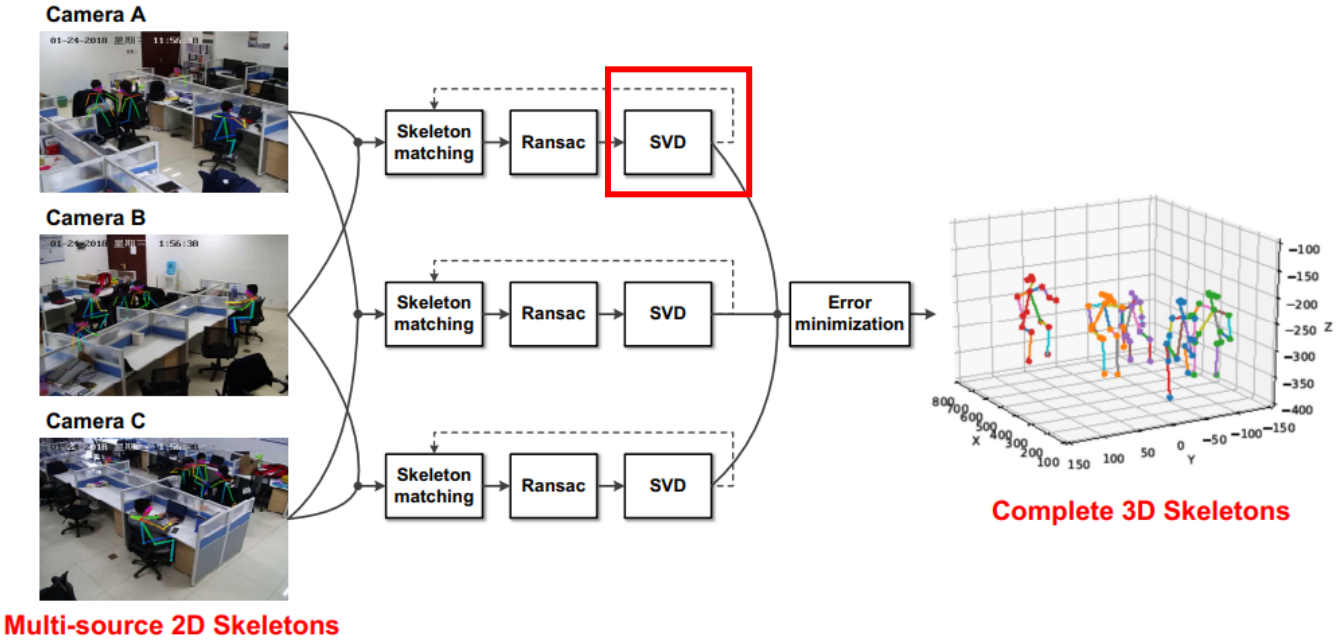
- Reconstruction from two views:
  I. Compute the fundamental matrix *F* from point correspondences
  II. Compute the camera matrices from *F*
  III. For each point correspondence, compute the 3D point



3D Poses Reconstruction

3D Skeleton Reconstruction

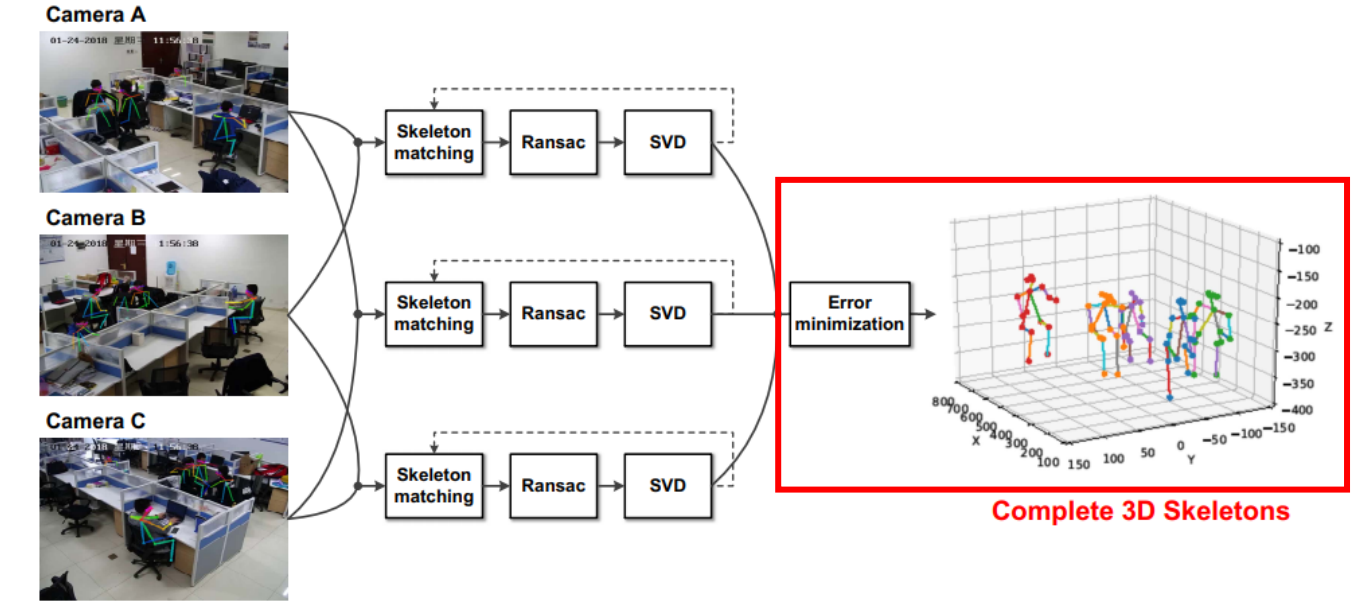
- Reconstruction from two views:
  I. Compute the fundamental matrix *F* from point correspondences
  II. Compute the camera matrices from *F*
  III. For each point correspondence, compute the 3D point



3D Poses Reconstruction

3D Skeleton Reconstruction

- Reconstruction from two views:
  I. Compute the fundamental matrix *F* from point correspondences
  II. Compute the camera matrices from *F*
  III. For each point correspondence, compute the 3D point



$$S = \sum_{i \neq j} \frac{c_{ij} S_{ij}}{\sum_{i \neq j} c_{ij}}$$

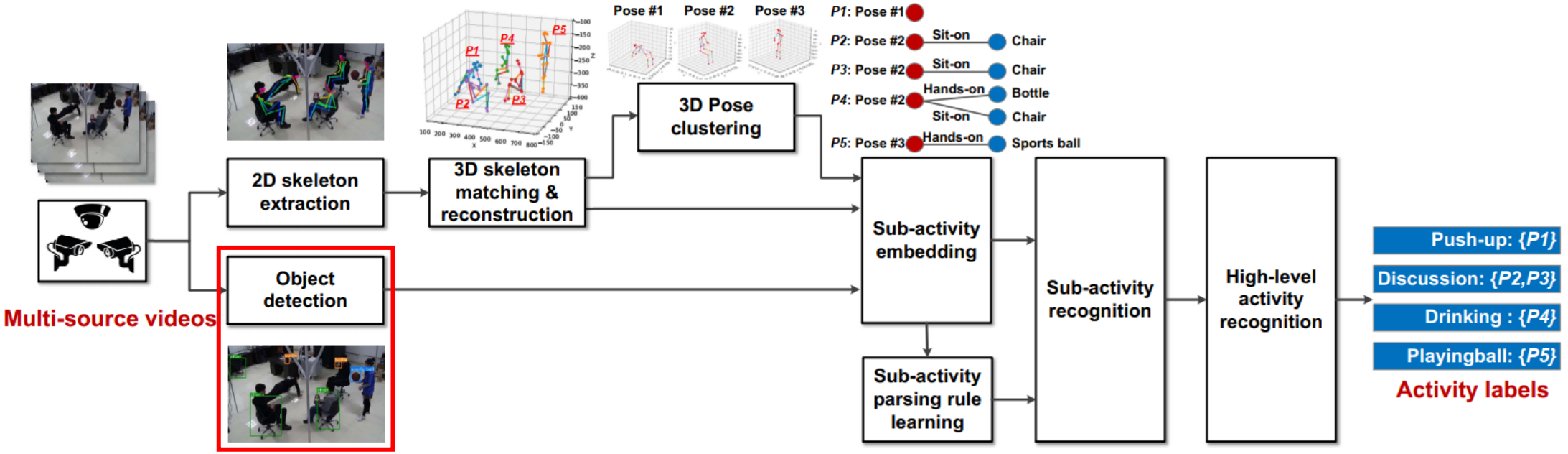
Outline

- Introduction
- 3D Poses Reconstruction
- **Co-evolving Activity Representation & Recognition**
- Implementation & Evaluation

Co-evolving Activity Representation

Object Detection

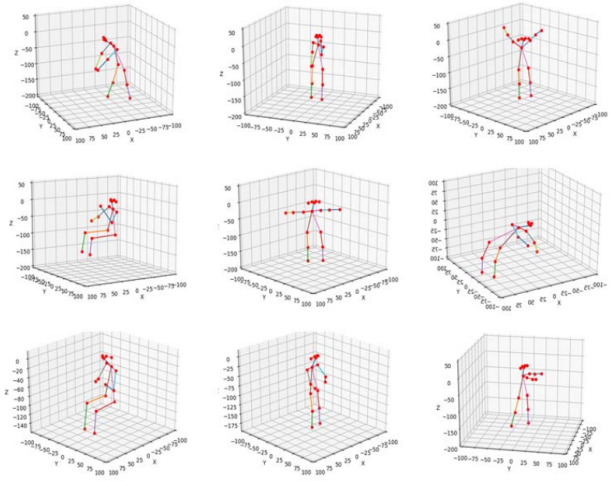
- Detect activity-related objects

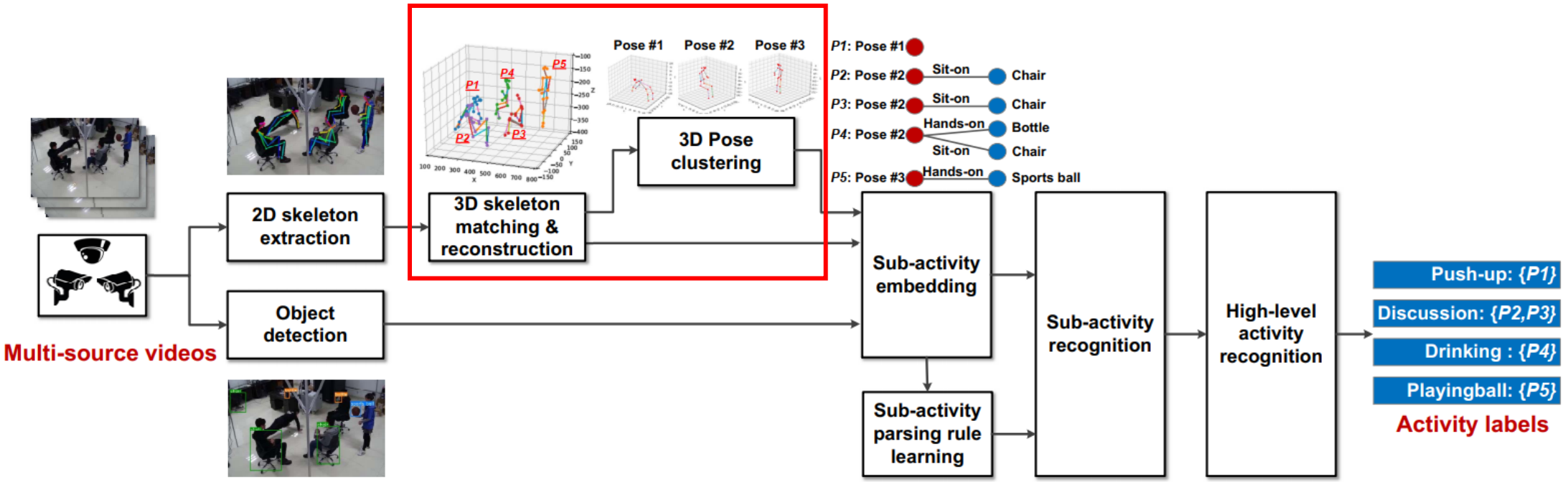


Co-evolving Activity Representation

3D Pose Clustering

- Decrease the feature dimension

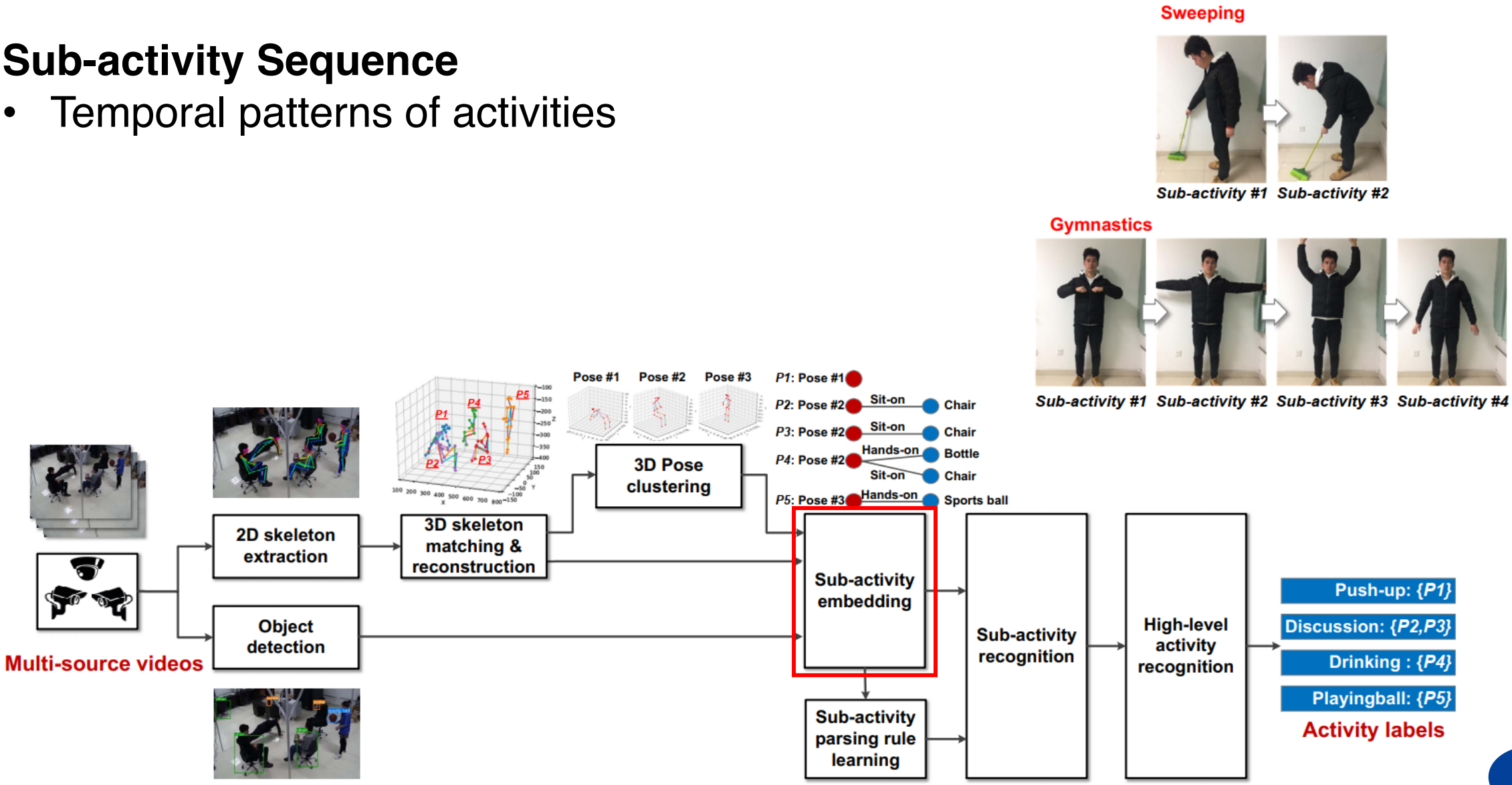




Co-evolving Activity Representation

Sub-activity Sequence

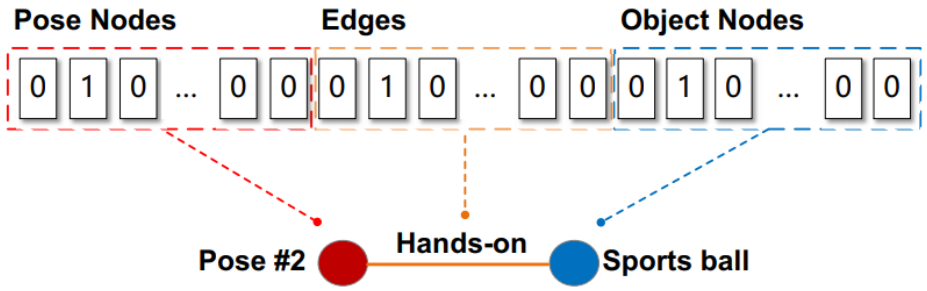
- Temporal patterns of activities

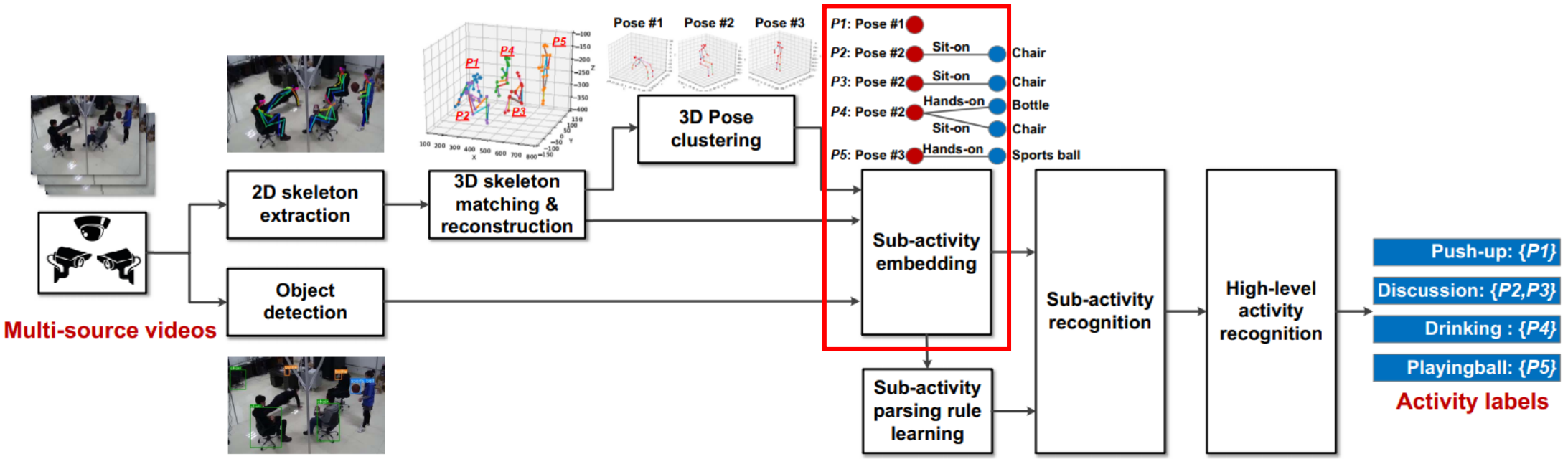


Co-evolving Activity Representation

Pose-Pose & Pose-Object Interaction

- Spatial relations-based rules

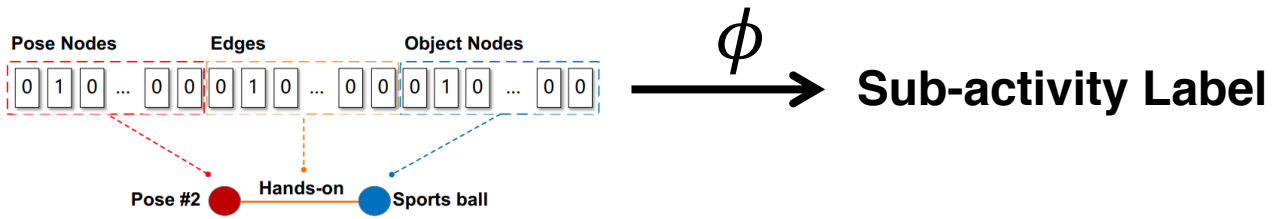


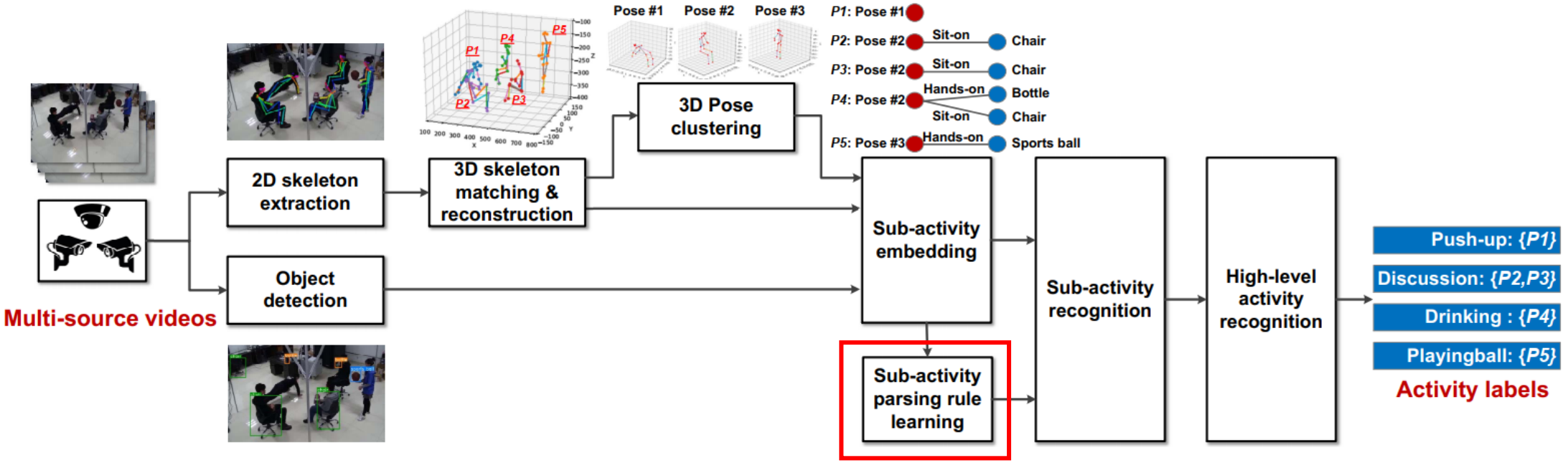


Co-evolving Activity Representation

Sub-activity Parsing

- $\phi: V \times E \times V \rightarrow S$
- Nearest neighbor classifier

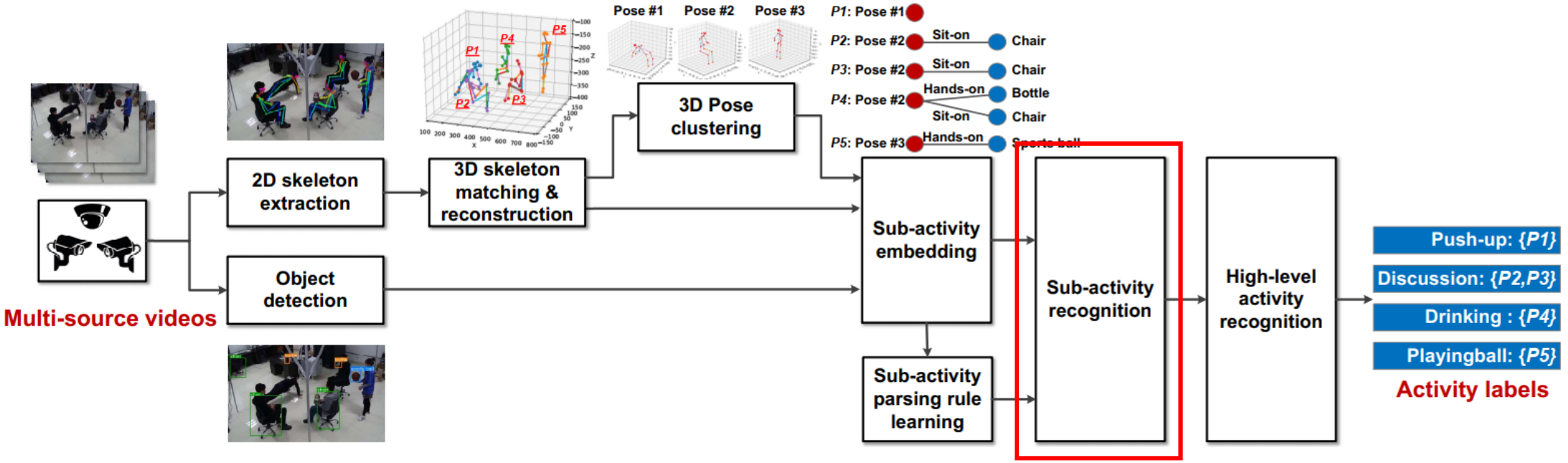




Co-evolving Activity Representation

Sub-activity Recognition

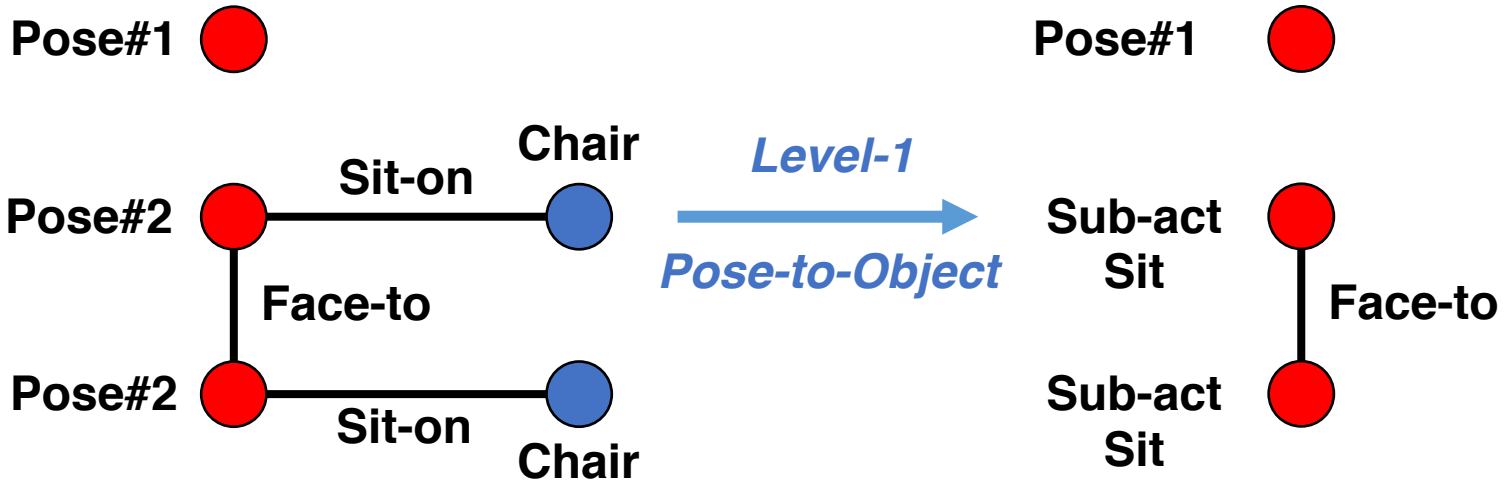
- Hierarchical graph coarsening algorithm
- 3-level: pose-to-object / pose-to-pose / pose only



Co-evolving Activity Representation

Sub-activity Recognition

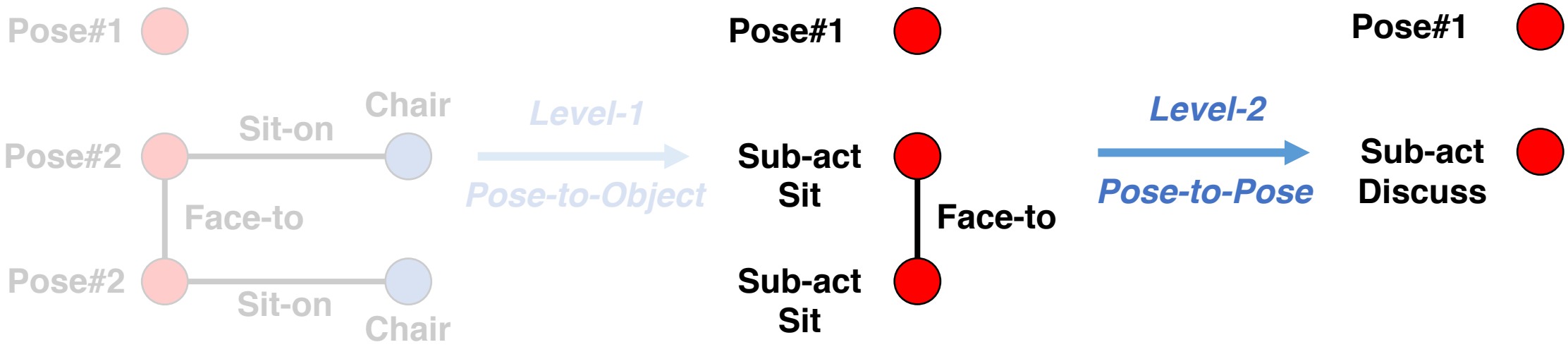
- Hierarchical graph coarsening algorithm
- 3-level: pose-to-object / pose-to-pose / pose only



Co-evolving Activity Representation

Sub-activity Recognition

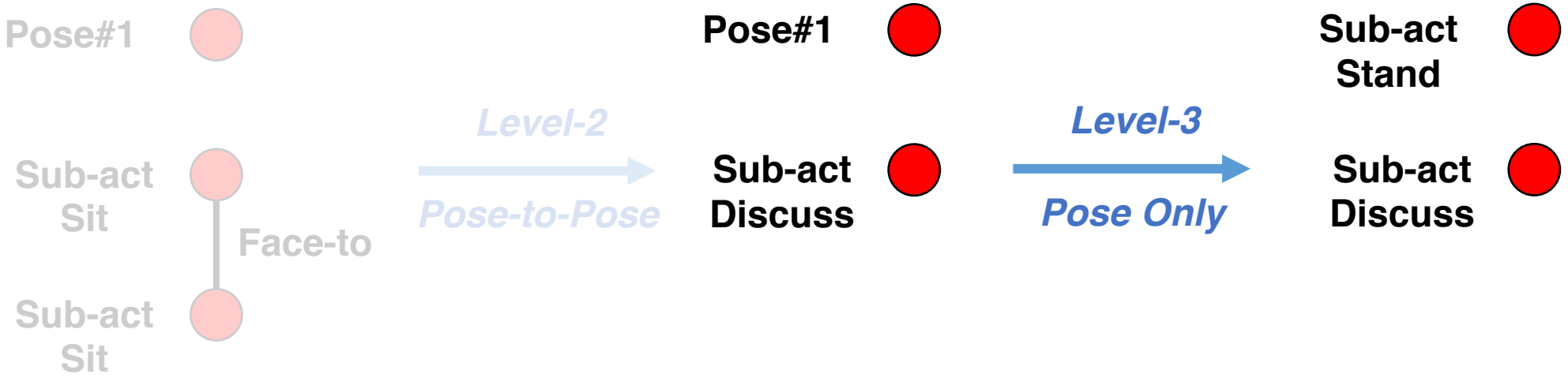
- Hierarchical graph coarsening algorithm
- 3-level: pose-to-object / pose-to-pose / pose only



Co-evolving Activity Representation

Sub-activity Recognition

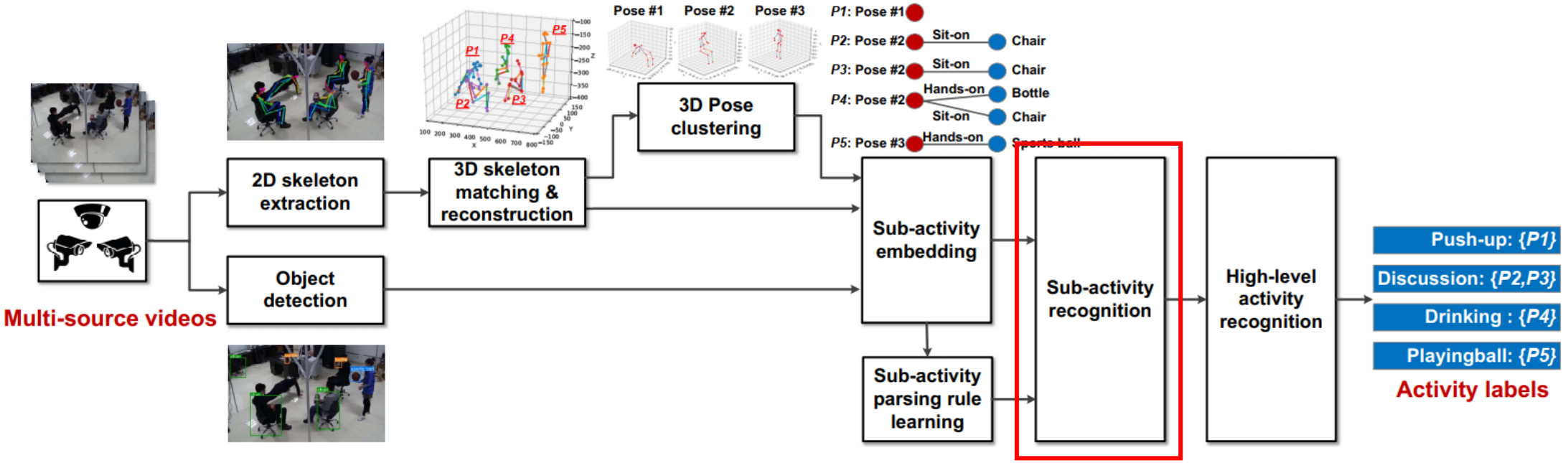
- Hierarchical graph coarsening algorithm
- 3-level: pose-to-object / pose-to-pose / pose only



Co-evolving Activity Representation

Sub-activity Recognition

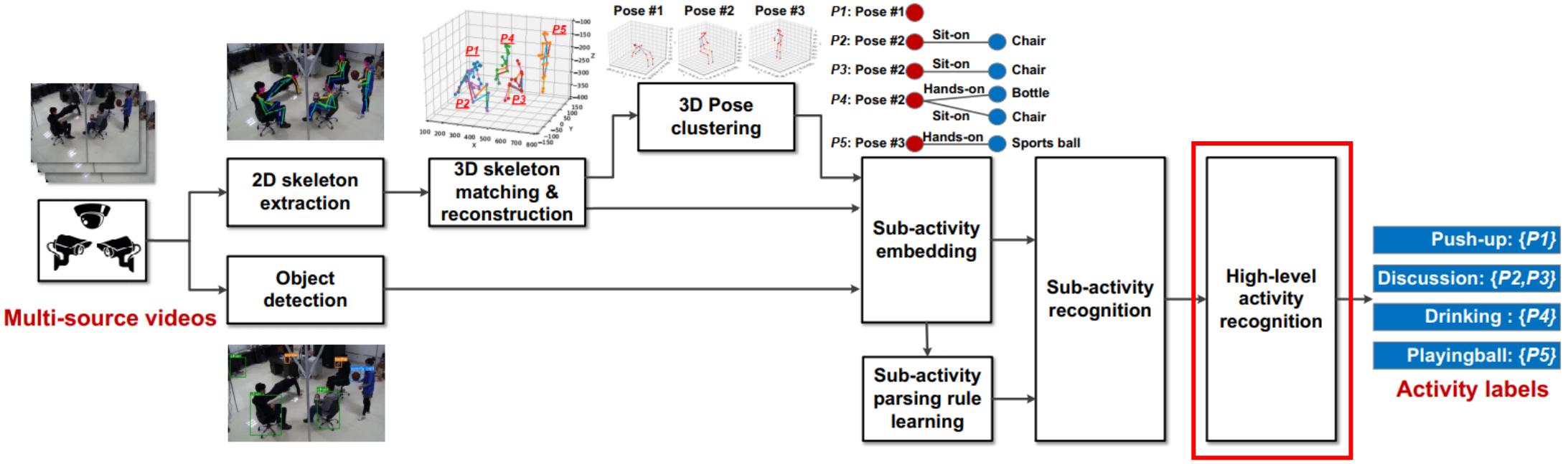
- Hierarchical graph coarsening algorithm
- 3-level: pose-to-object / pose-to-pose / pose only
- $O(|E|)$ parsing complexity



Co-evolving Activity Representation

High-level Activity Recognition

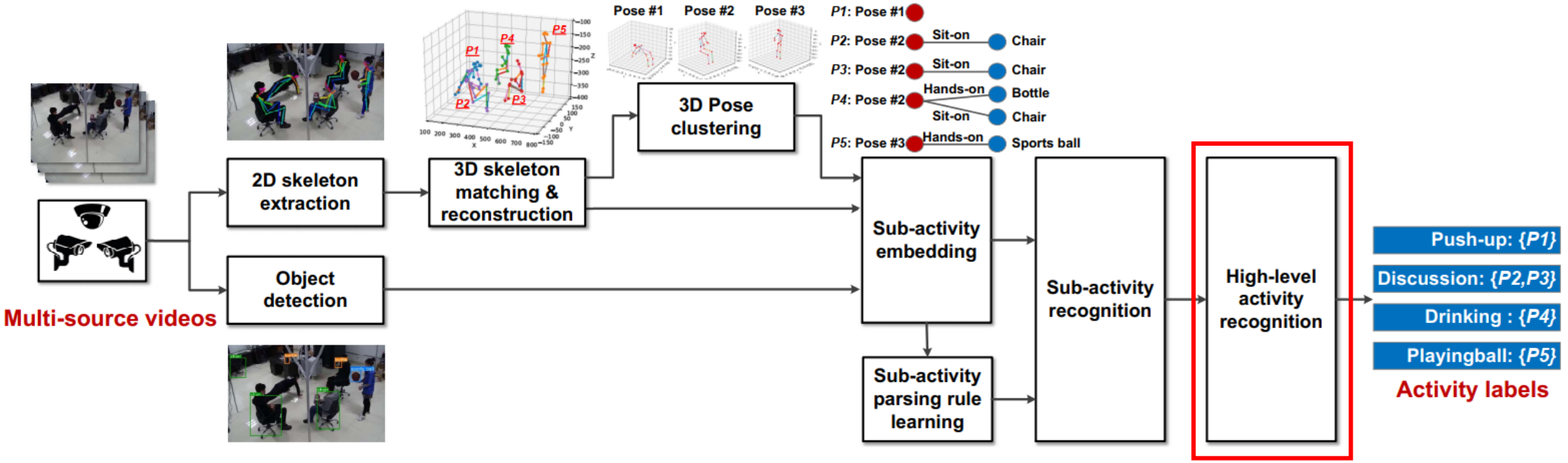
- Hidden Markov Models for temporal structure
- one HMM for scoring one high-level activity



Co-evolving Activity Representation

High-level Activity Recognition

- Hidden Markov Models for temporal structure
- Bayesian Model Merging algorithm



Outline

- Introduction
- 3D Poses Reconstruction
- Co-evolving Activity Representation & Recognition
- **Implementation & Evaluation**

Experiment Setup

Off-the-shelf Models

- YOLO-V3 object detector
- OpenPose 2D pose estimator

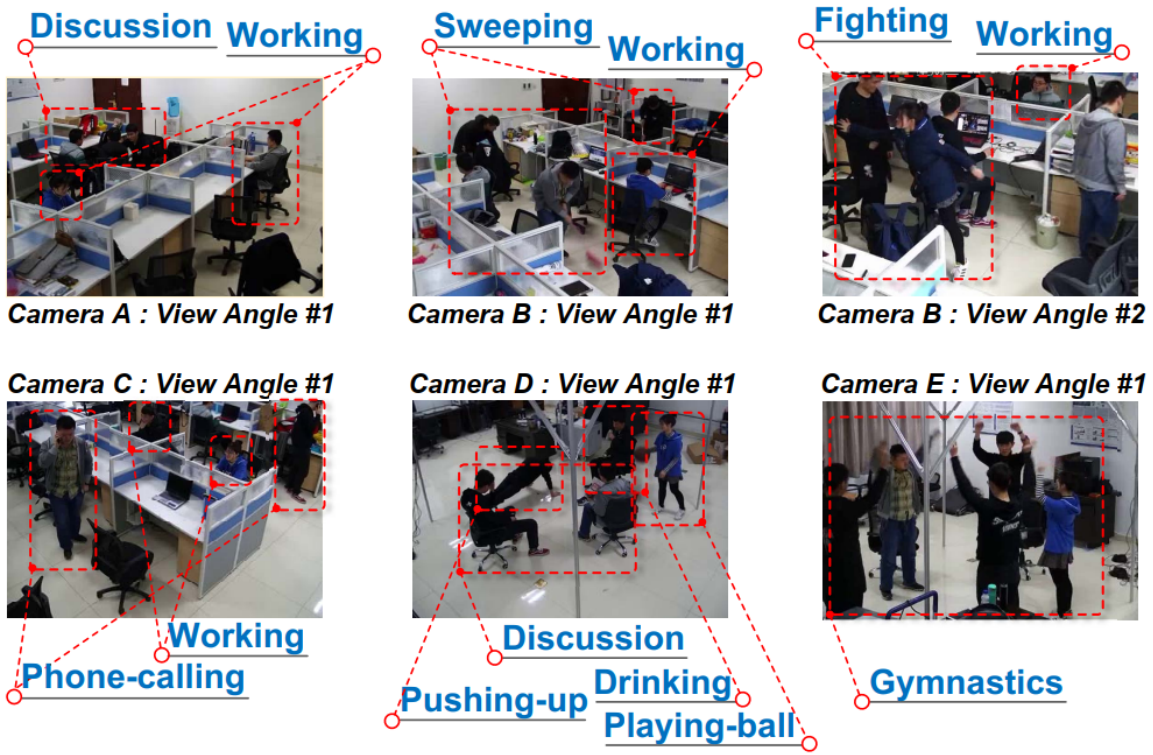
Dataset

- Office (3 cameras) & laboratory (2 cameras)
- 10 high-level activities

Experiment Setup

Dataset

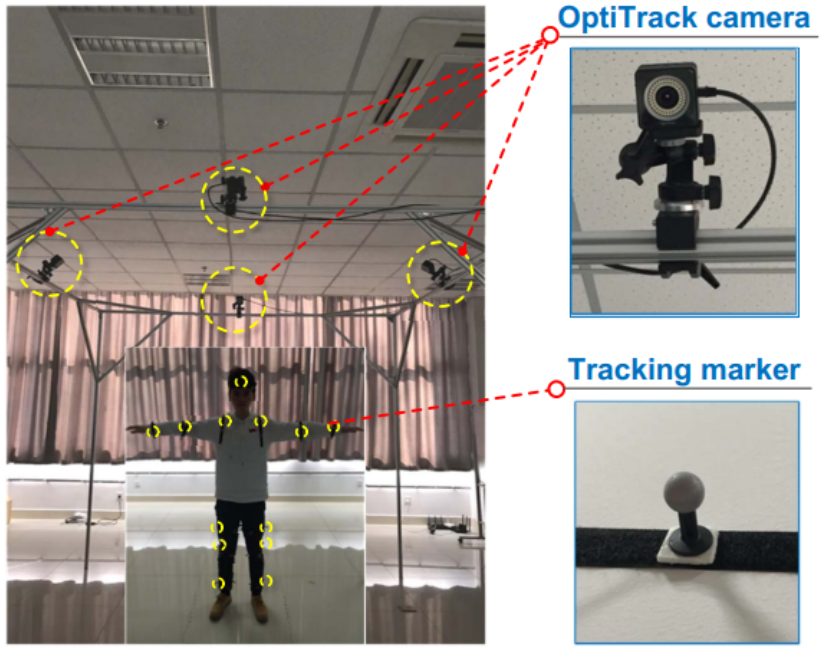
- Office (3 cameras) & laboratory (2 cameras)
- 704x576, 24 FPS, 45 min
- 10 high-level activities

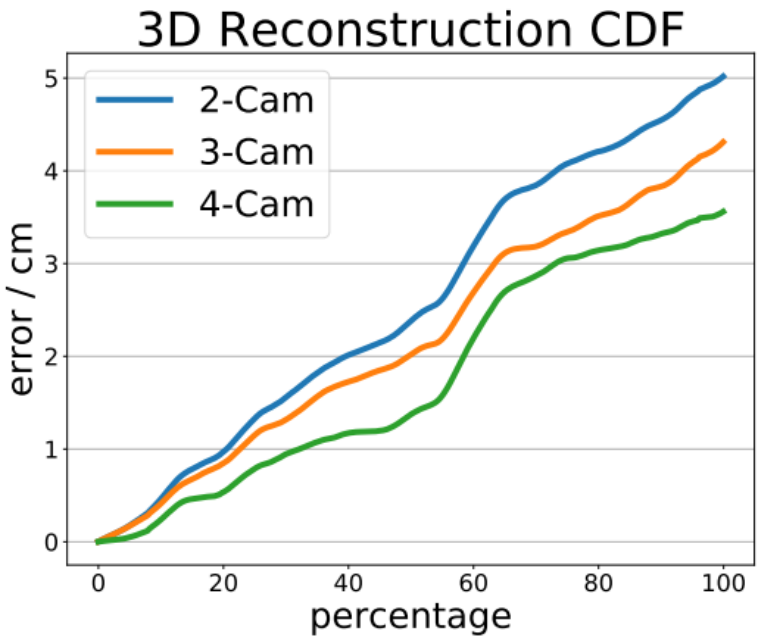


Evaluation

3D Pose Reconstruction

- Groundtruth: 4 OptiTrack cameras + 13 tracking markers

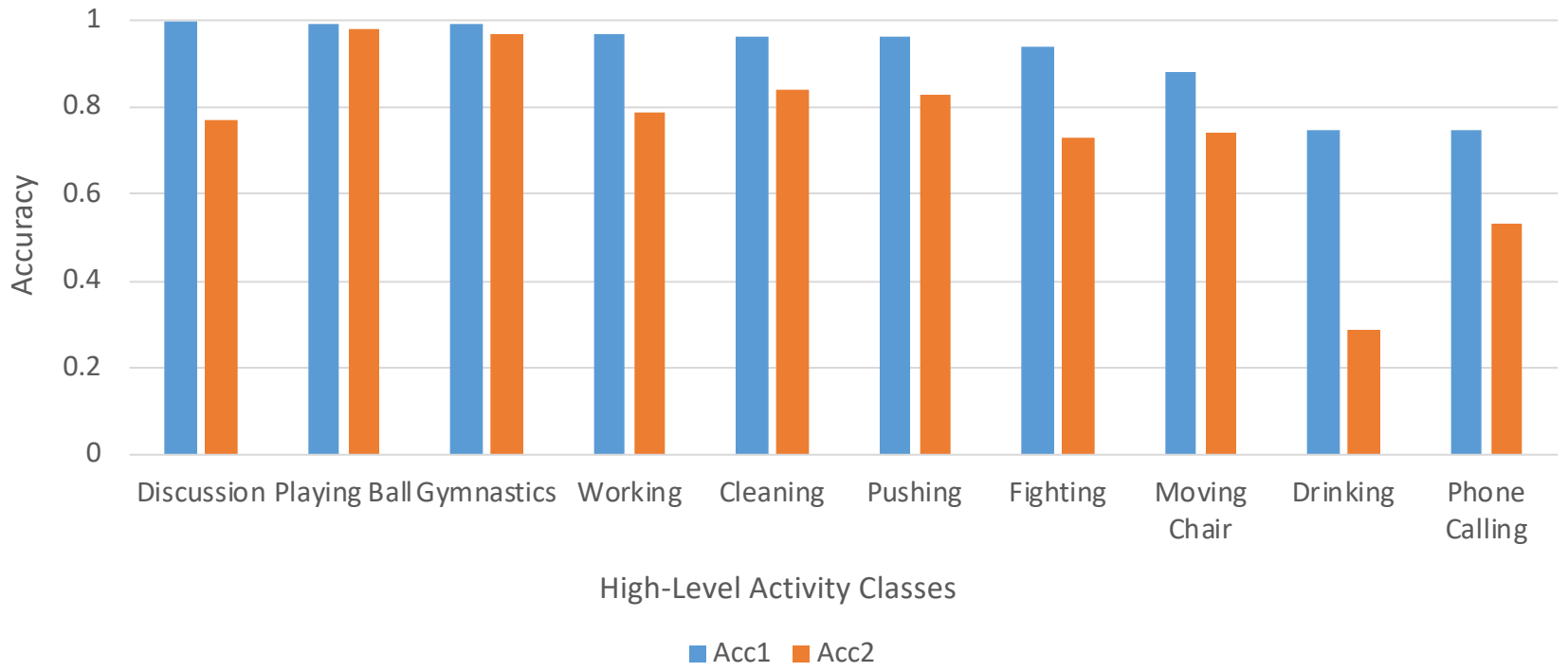




Evaluation

10-Class Activity Recognition

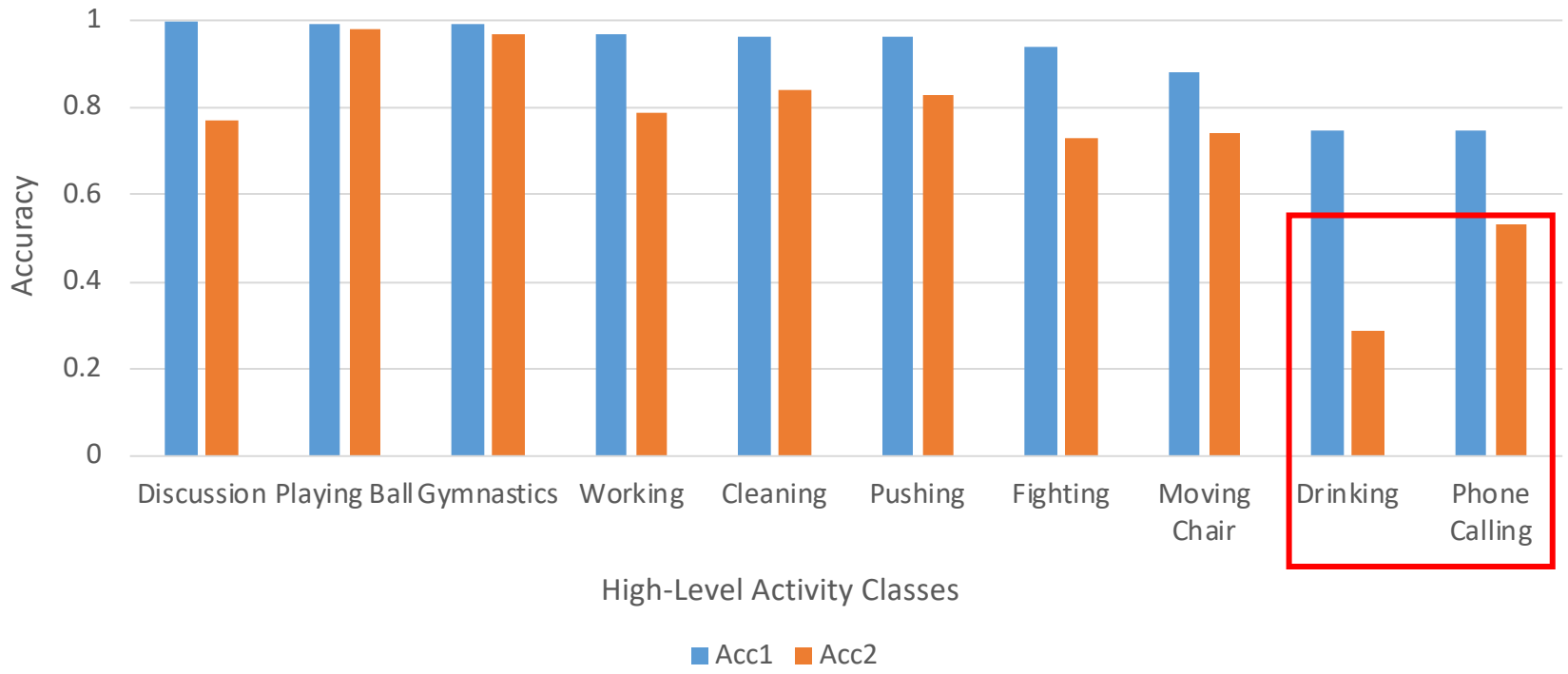
- Acc1: ratio of frames with correct high-level activity labels
- Acc2: ratio of frames with correct high-level activity labels and correct number of people involved
- Avg.Acc1=91.2% / Avg.Acc2=70.5%



Evaluation

10-Class Activity Recognition

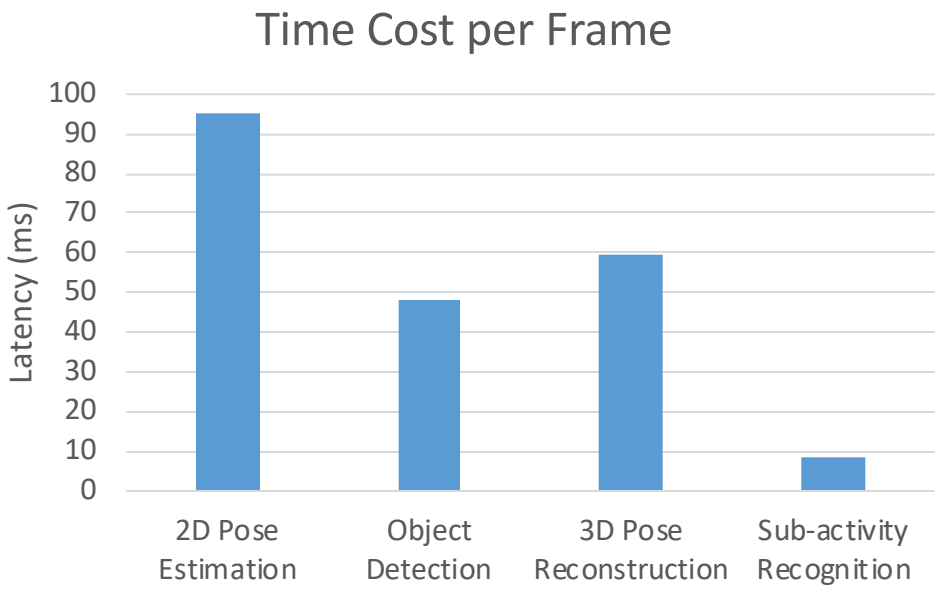
- Acc1: ratio of frames with correct high-level activity labels
- Acc2: ratio of frames with correct high-level activity labels and correct number of people involved
- Avg.Acc1=91.2% / Avg.Acc2=70.5%



Evaluation

M&M Overheads

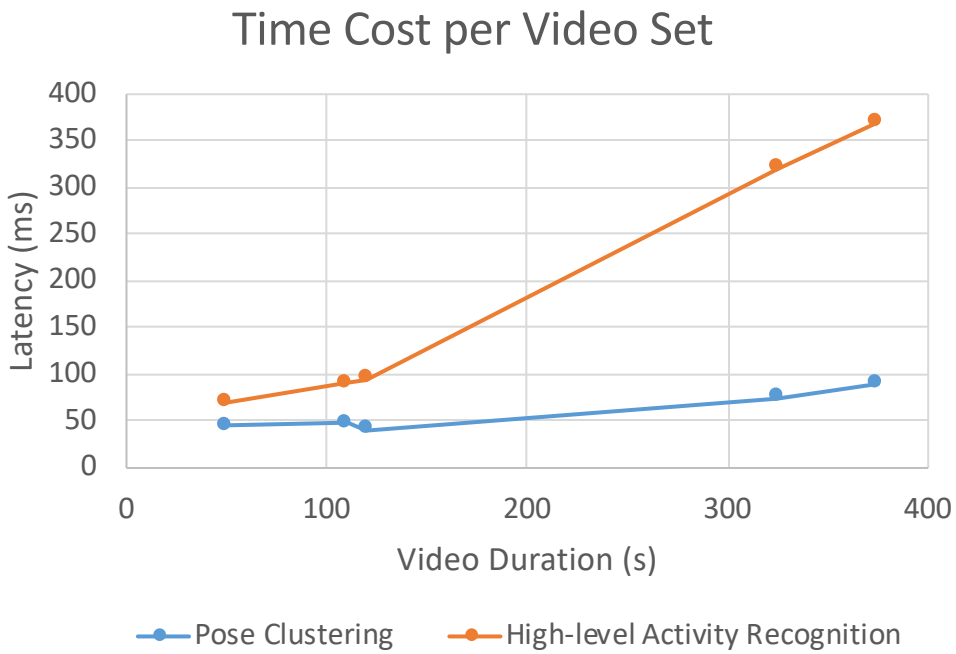
- Frame-level modules: 2D pose estimation, object detection, 3D pose reconstruction, sub-activity recognition



Evaluation

M&M Overheads

- Frame-level modules: 2D pose estimation, object detection, 3D pose reconstruction, sub-activity recognition
- Video-level modules: pose clustering, high-level activity recognition



Summary

- 2D human poses can be utilized to fuse multi-view information and obtain a view-invariant 3D description of cross-view activities.

- M&M enables efficient and accurate recognition of co-evolving activities from multi-source videos.

Lab for Intelligent Networking & Knowledge Engineering

12 Faculty Members, 2 Post-Docs, 3 Secretaries; 7 with PhD from abroad

XiangYang Li

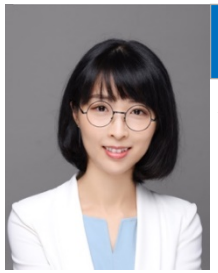
IEEE Fellow
ACM Fellow
ACM China Co-Chair

Panlong Yang

Wireless Network
Mobile Computing

Nikolaos M.Freris

USA NYU A.P.
CPS Algorithms
Distributed optimization
Machine learning



Lan Zhang

ACM China Doctoral Dissertation Award
Qingcheng Award
Data Understanding and Trading
Privacy Protection

Bei Hua

High-Performance Computing
Edge Computing

Yu Zhang

System Software
Software Optimization
Security
Quantum software

Hao Zhou

Wireless Network Resource Management

Yanyong Zhang

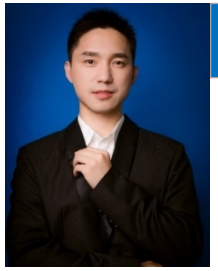
IEEE Fellow

Haisheng Tan

Cloud Computing
Algorithms Analysis

YuBo Yan

Wireless/Passive Network
IntelliSense
IoT
SDR



Xin He

Passive Network Theories of Information and Coding

Xing Guo

Edge Computing
Security of IoT

Xuerong Huang

Research Assistant

Ludi Xue

Research Assistant

M&M: Recognizing Multiple Co-evolving Activities from Multi-source Videos

WeChat Official

Lab Website
https://linke.ustc.edu.cn/

Mu Yuan
University of Science and Technology of China
School of Computer Science and Technology
ym0813@mail.ustc.edu.cn

Lan Zhang
University of Science and Technology of China
School of Computer Science and Technology
zhanglan@ustc.edu.cn