\Orchestrating a brighter world NEC

2021 IEEE International Conference on Image Processing (ICIP)

A Hybrid Two-stream Approach For Multi-person Action Recognition in Top-view 360^0 Videos

Karen Stephen[1], Jianquan Liu[1], Vivek Barsopia[2]

[1]Biometrics Research Laboratories

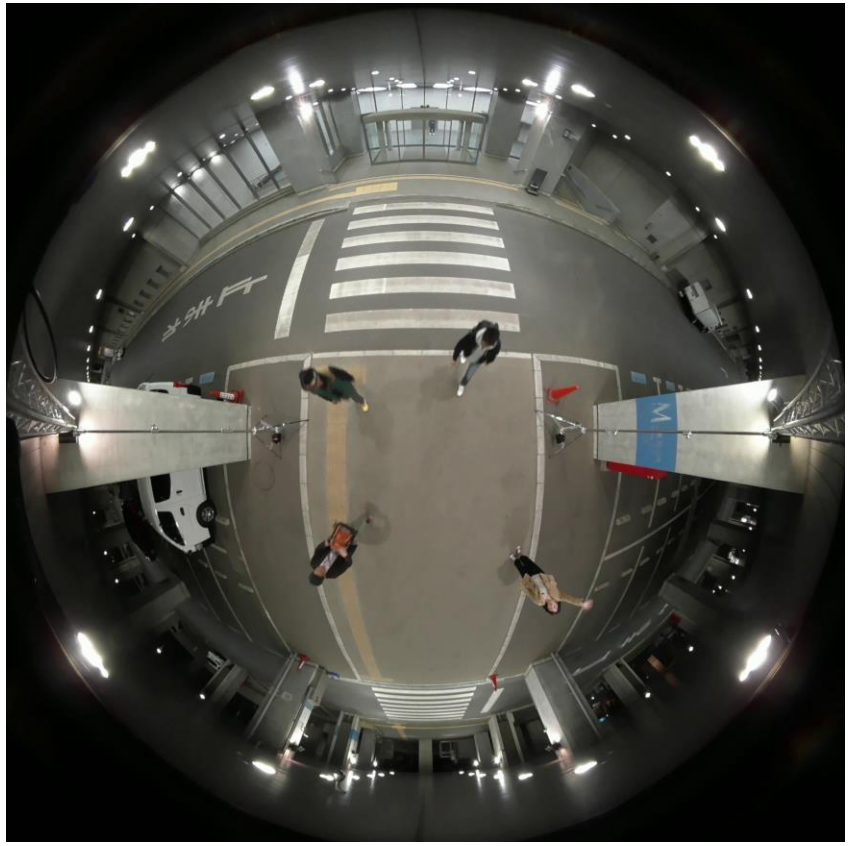
[2]Data Science Research Laboratories

NEC Corporation, Japan

Contact: stephenkaren@nec.com

Introduction

Introduction



360^0 surveillance video frame

- 360^0 cameras:
  - Popular in safety and surveillance applications
  - Wide field-of-view enables monitoring of a large area
  - Eliminates need for multiple perspective cameras and synchronization

Problem definition

Multi-person action recognition in top-view 360^0 videos

Input: 360^0 video

Drink water
Wear jacket
Walk upstairs
Play with phone

Output: Action labels

Existing works

- Large scale 360^0 action datasets unavailable
- Initial works used limited data and hand crafted features
- Recent work* followed a deep learning approach

Input 360^0 video frame

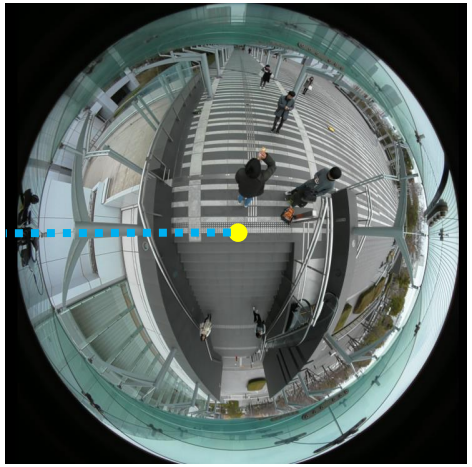
Panorama frame

Pre-trained network for perspective videos

Output actions

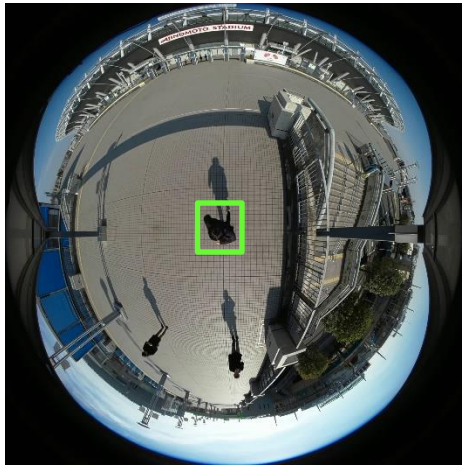
* Junnan Li, Jianquan Liu, Yongkang Wang, Shoji Nishimura, and Mohan S Kankanhalli, "Weakly-supervised multi-person action recognition in 360° videos," in Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2020, pp. 497–505.

Existing work: drawback

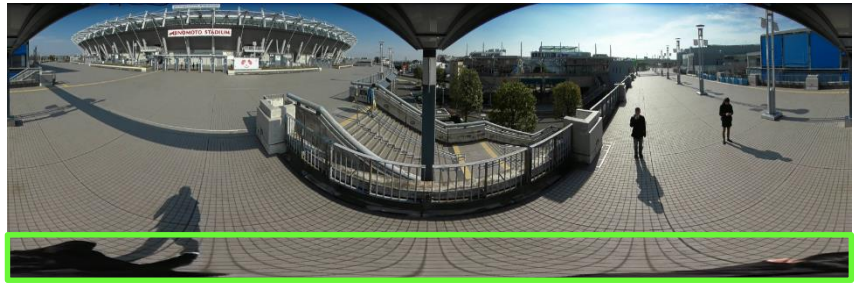


Input 360^0 video frame

Panorama frame



Input 360^0 video frame



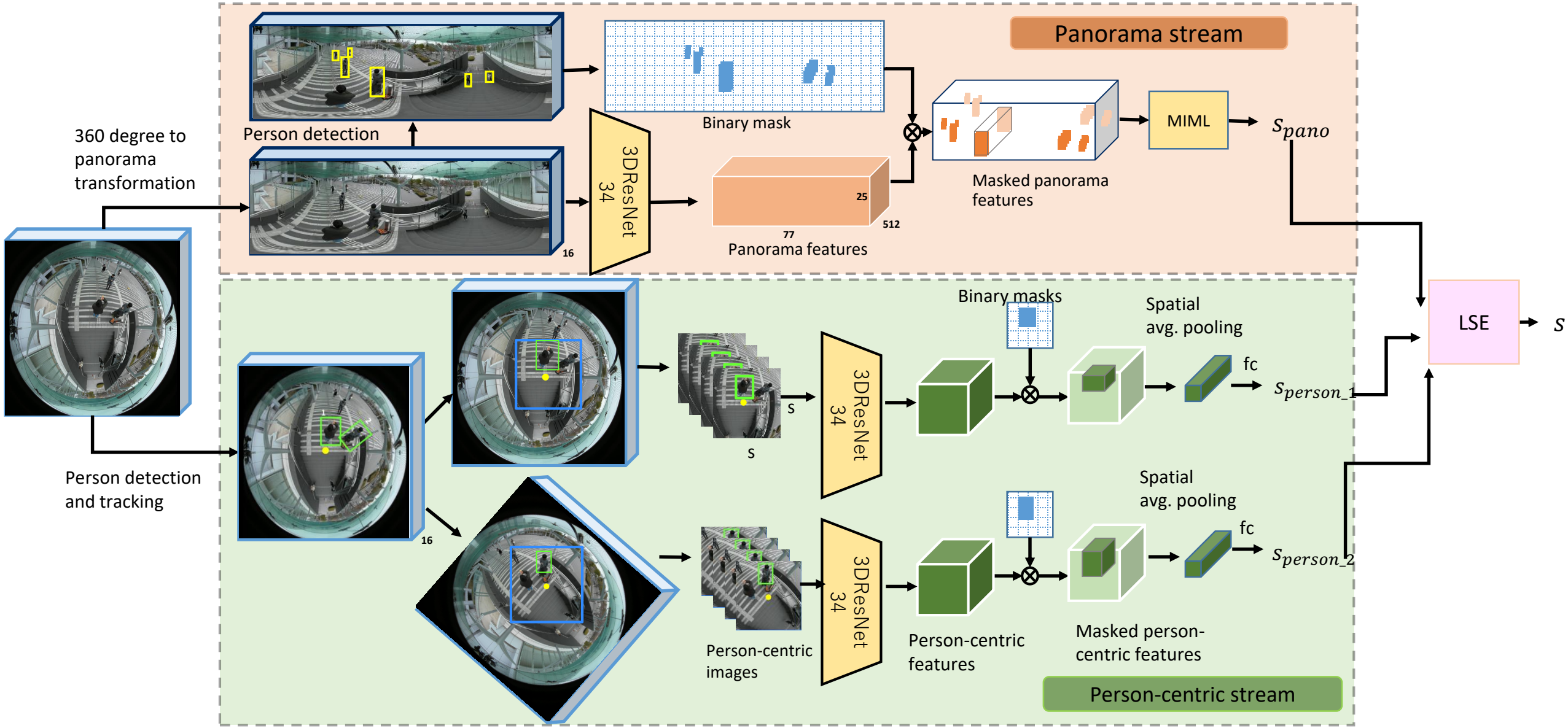
Panorama frame

- Persons near the center distorted in the panorama frames
- Difficult to recognize actions happening near the center

Proposed method

A Hybrid two-stream approach

Proposed method

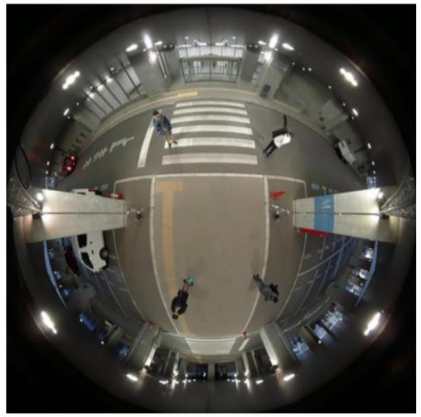


Experiments and Results

360 Action dataset


Stadium Yard

Stadium Gate


Carpark Night


Carpark Day


Lobby


Train Station Gate


Train Station Platform


Yard 1


Yard 2


Convenience Store

Experiments and Results

Comparison with state-of-the-art methods on 360Action dataset:

Method	mAP %
Collective [T. Bagautdinov et al. CVPR'17]	61.27
3D ResNet [K. Hara et al. ICCV'17]	61.95
R-C3D [H. Xu et al. ICCV'17]	58.74
MiCT [Y. Zhou et al. CVPR'18]	62.18
Panorama 3D-ResNet [J. Li et al. WACV'20]	70.12
Hybrid two-stream (Ours)	**72.40**

Ablation study:

Method	Panorama	Person-centric	mAP (%)
Panorama 3D-ResNet [J. Li et al. WACV'20]	✓	-	70.12
Person-centric only(Ours)	-	✓	67.0
Hybrid two-stream(Ours)	✓	✓	**72.40**

For more details, please visit our poster

Thank you!