



TRƯỜNG ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Multi-Camera Tracking for Employee Behavior Monitoring

Trần Quốc Lập – 20194443
July 27, 2023

ONE LOVE. ONE FUTURE

Introduction and Objective

Proposed Method and Evaluation

Application System Development

Conclusion and Future Work

INTRODUCTION & OBJECTIVE

Multi-camera tracking (MCT) aims to track people **across** cameras.

MCT has **various applications**. In employee management, MCT can track and **monitor employee behaviors**.

Example: Amazon's worker surveillance

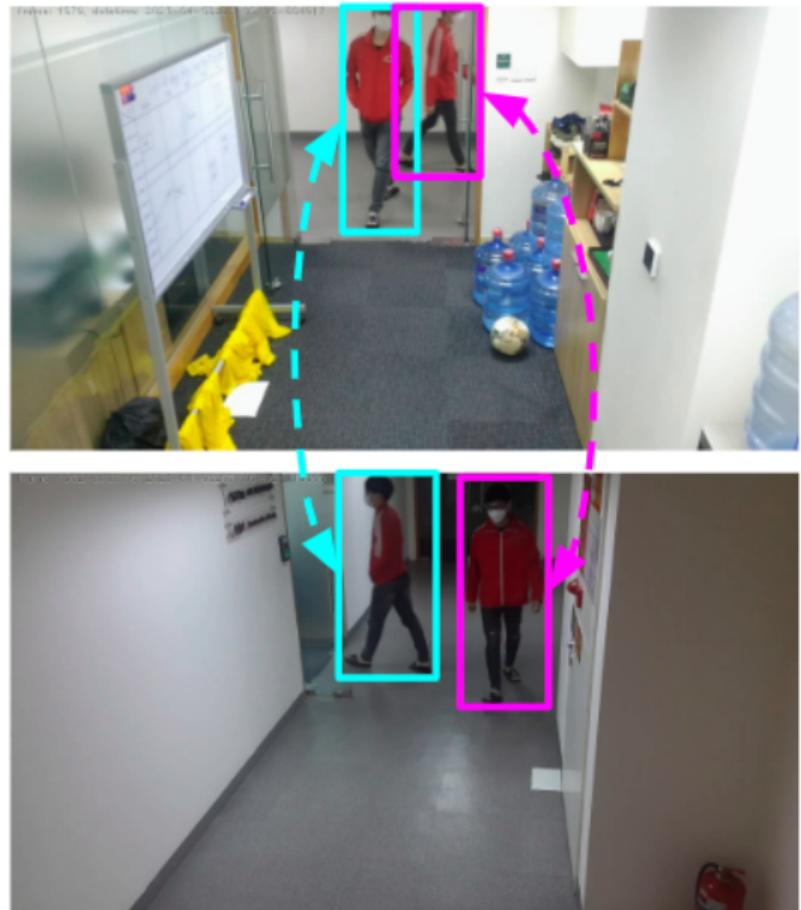


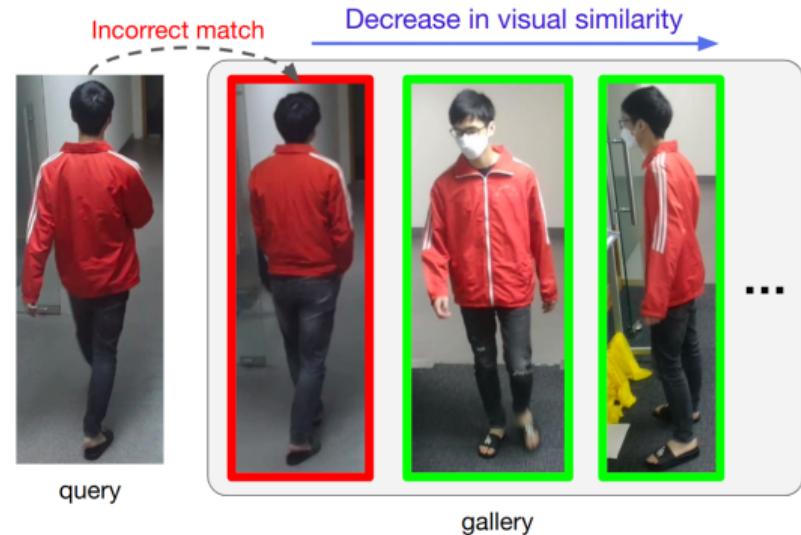


Figure 1: General pipeline for multi-camera tracking.

1. Person detection: Detect person in single frame (e.g. YOLO [3, 2]).
2. Single camera tracking: Match boxes to form tracks (e.g. SORT [1], ByteTrack [4]).
3. Multi-camera tracking: Match tracks across cameras.

To match tracks across cameras, popular methods compare people **visual appearance**, notably Re-ID.

Issue of Re-ID: easily matches **incorrectly** if people (e.g. employees) wear **uniforms**.



Thesis work: develop a MCT solution that can work well when people have **similar appearance**.

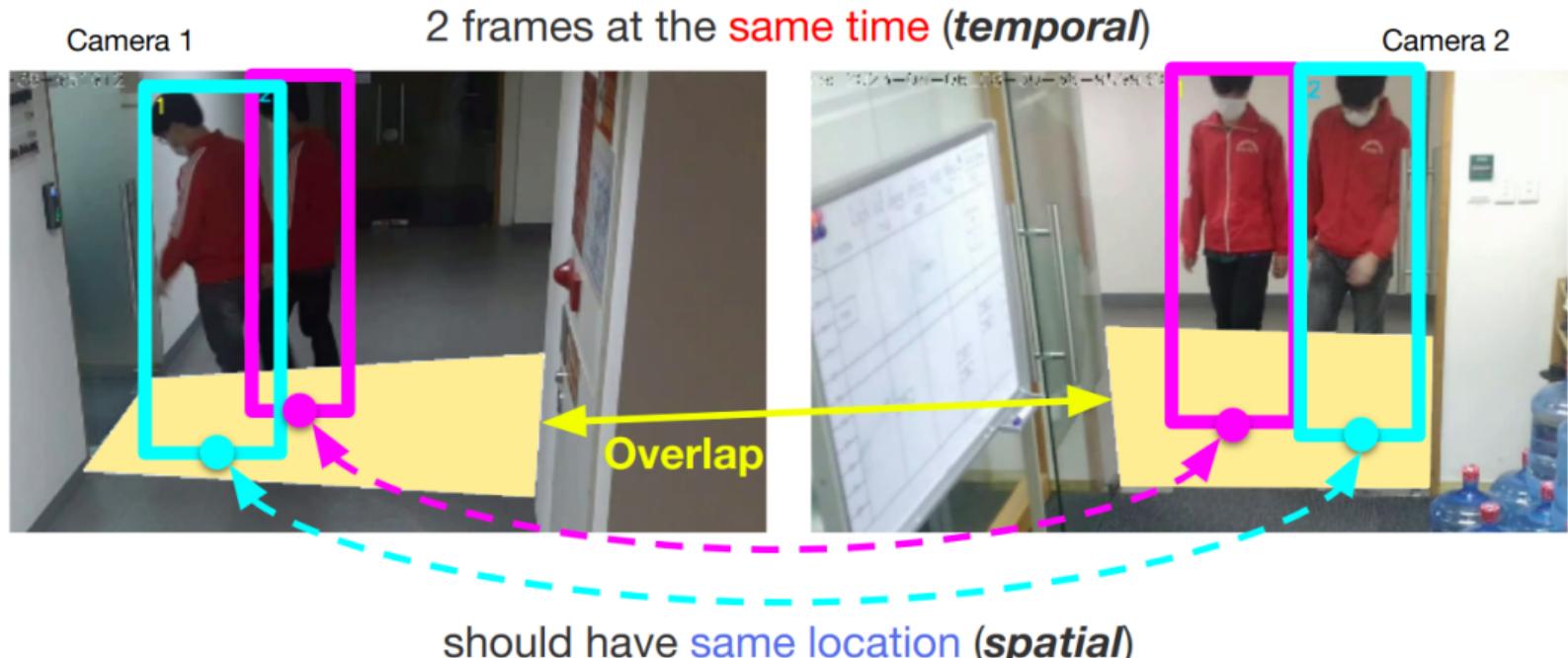


Figure 2: Proposed: MCT based on **Spatio-Temporal Association (STA)** with **overlapping** area.

Scope:

- ◊ Small overlap & synchronized cameras.
- ◊ Track employees wearing uniforms.
- ◊ Maximum 4 people.
- ◊ 3 cameras.
⇒ no public dataset
- ⇒ collect 36 videos, split into 3 sets:
Easy, **Medium**, **Hard**.



Figure 3: Camera setup and FOV

PROPOSED METHOD & EVALUATION

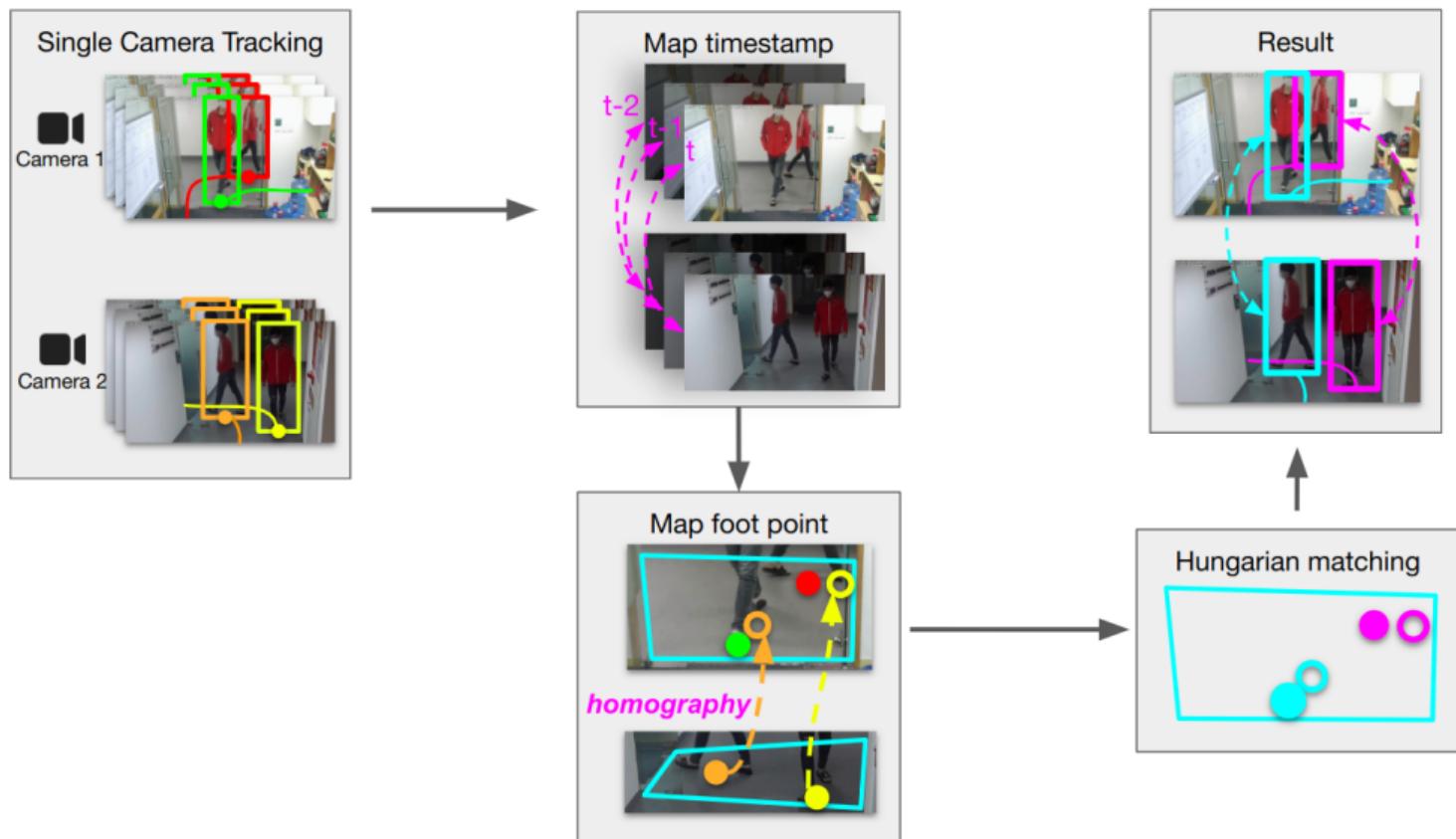


Figure 4: Overview of the proposed STA method.

Evaluation metric:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Video set	#TP	#FP	#FN	F1
Easy	511	7	7	0.986
Medium	662	46	22	0.951
Hard	966	144	41	0.913
Total	2139	197	70	0.941

Table 1: **TP**: correctly matched pair. **FP**: incorrectly matched pair. **FN**: missed pair.

- ◊ **Promising** but decreases with complexity.
- ◊ **#FP >> #FN**, especially with **hard** and **medium** sets.

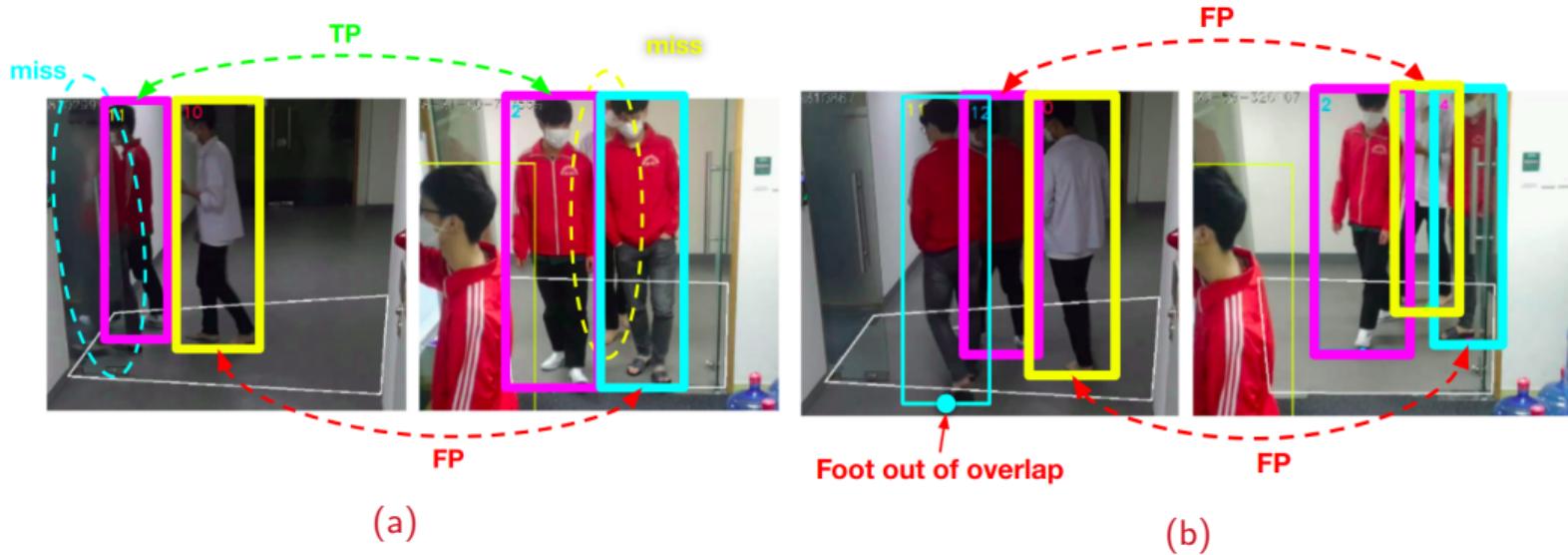


Figure 5: Two main causes of FP and FN:

- a) missing detection.
- b) incorrect foot interpolation.

Assumption: FP due to missing detections have larger spatial distance than TP.
⇒ treated those FP as outliers in the distance distribution.

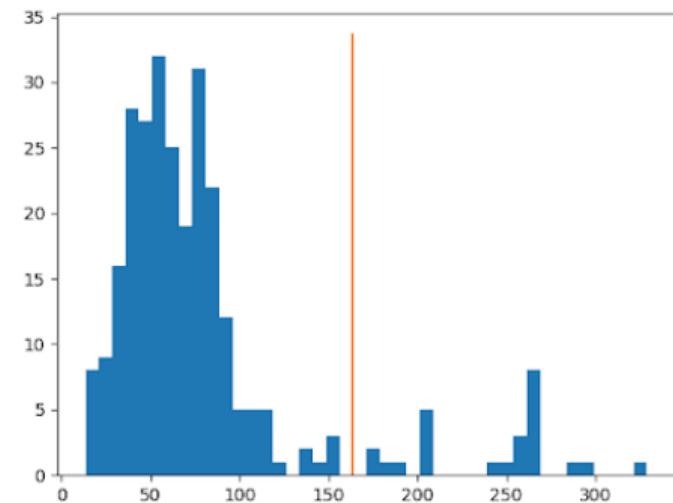
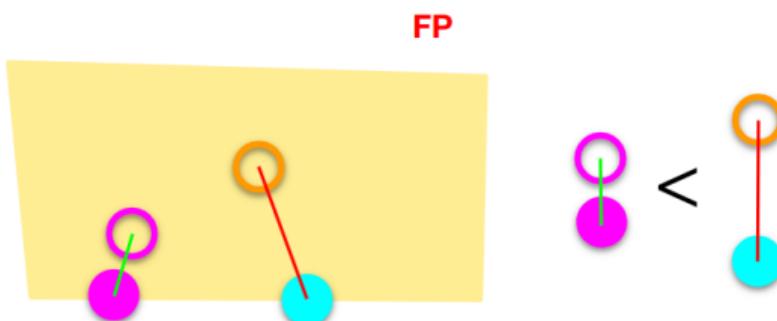
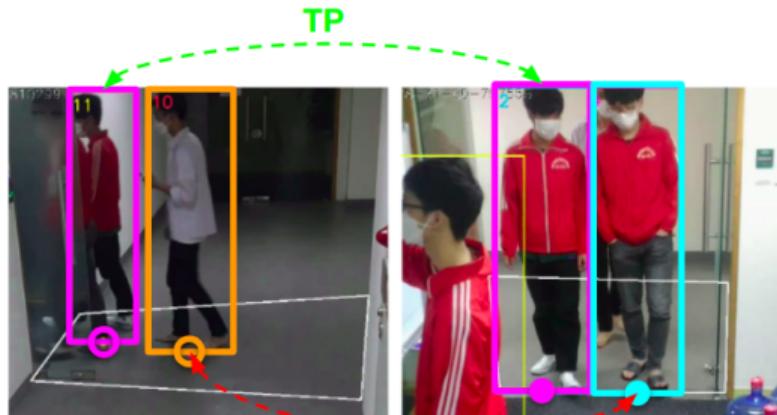


Figure 6: Distance distribution of matched pairs. x-axis: distance. y-axis : #matched pairs. Seam: upper bound by $IQR(25, 75)$.

Assumption: Sometimes a mismatch (FP and FN) can be solved by taking matches in previous and next timestamps into account
⇒ average distance over neighboring frames as an input cost for the Hungarian.

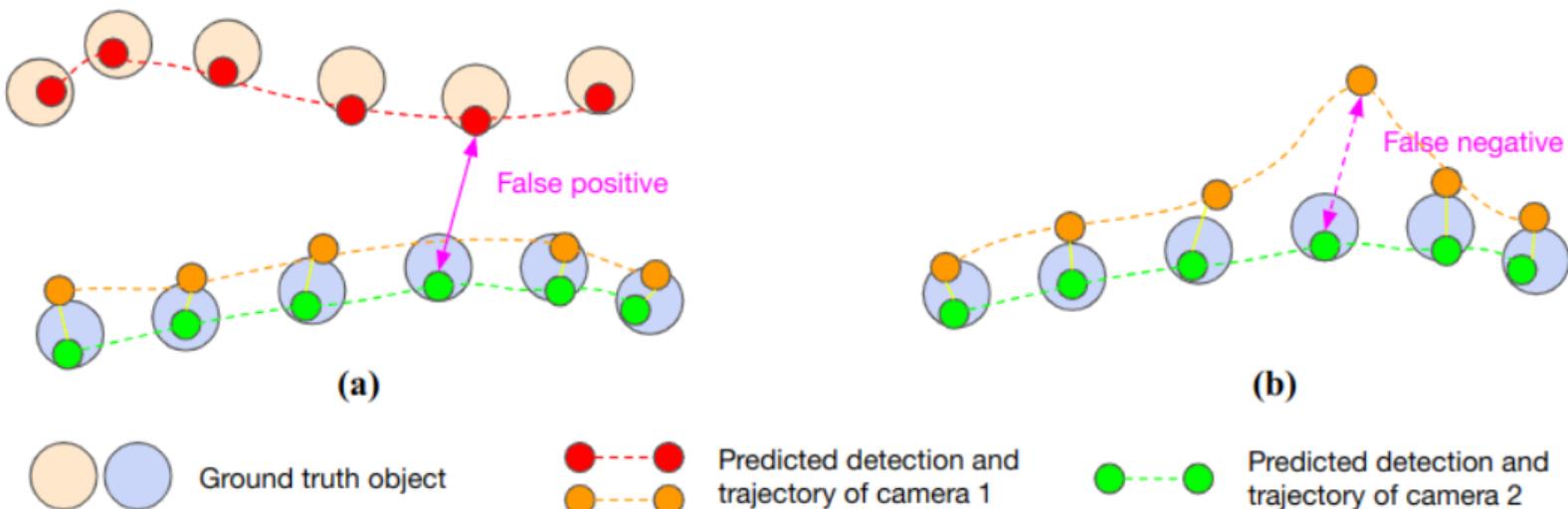


Figure 7: window-based mapping may reduce a) FP and b) FN.

Issue with rectangular bounding box: Even if it fits the body well, the foot point (midpoint of bottom edge) may not be accurate. E.g: when legs apart.

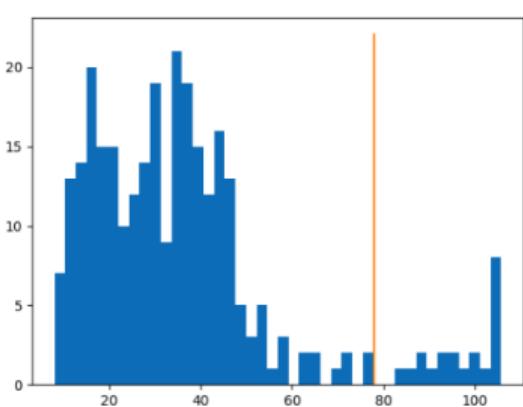
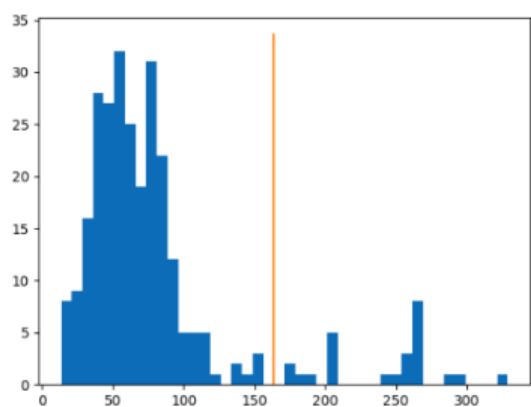


Figure 8: Distance distribution. **a)** using box. **b)** using pose.

Experiments prove that using pose:

- ◊ gives more accurate foot points.
- ◊ must combine with FP filtering, or else it will be worse than using box.

Set	Baseline	IQR (20, 80)	size = 15	IQR(20,80) size = 11	IQR(25, 75) size = 7 pose
Easy	0.986 (511,7,7)	0.983 (507,7,11)	0.994 (515,3,3)	0.986 (509,5,9)	0.985 (657,3,17)
Medium	0.951 (662,46,22)	0.958 (658,32,26)	0.955 (665,43,19)	0.969 (662,20,22)	0.991 (886,8,8)
Hard	0.913 (966,144,41)	0.927 (959,103,48)	0.921 (975,135,32)	0.938 (963,83,44)	0.960 (1179,36,63)
Total	0.941 (2139,197,70)	0.949 (2124,142,85)	0.948 (2155,181,54)	0.959 (2134,108,75)	0.976 (2722,47,88)

Table 2: Each cell format is $F1(\#TP, \#FP, \#FN)$ evaluated at frame-level.

In general, all the extensions have **greater and positive** impact on **complex** cases than on **easy** cases.



Figure 9: The matching results using Re-ID.

Video set	Re-ID	STA
Easy	0.5 (32 - 64 - 0)	1.0 (32 - 0 - 0)
Medium	0.348 (57 - 211 - 2)	0.982 (57 - 0 - 2)
Hard	0.380 (54 - 176 - 0)	0.991 (53 - 0 - 1)

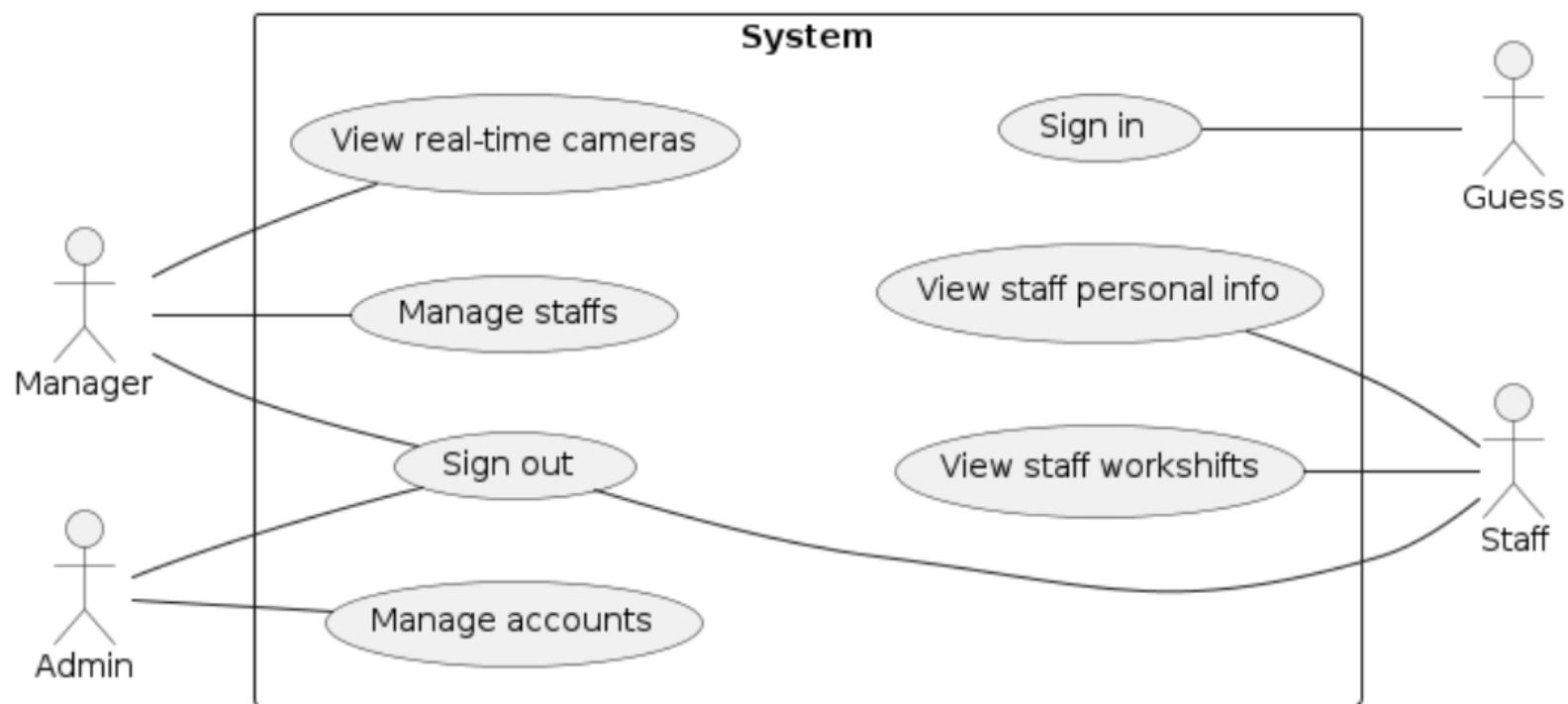
Table 3: Re-ID vs. the proposed STA method evaluated at track-level.

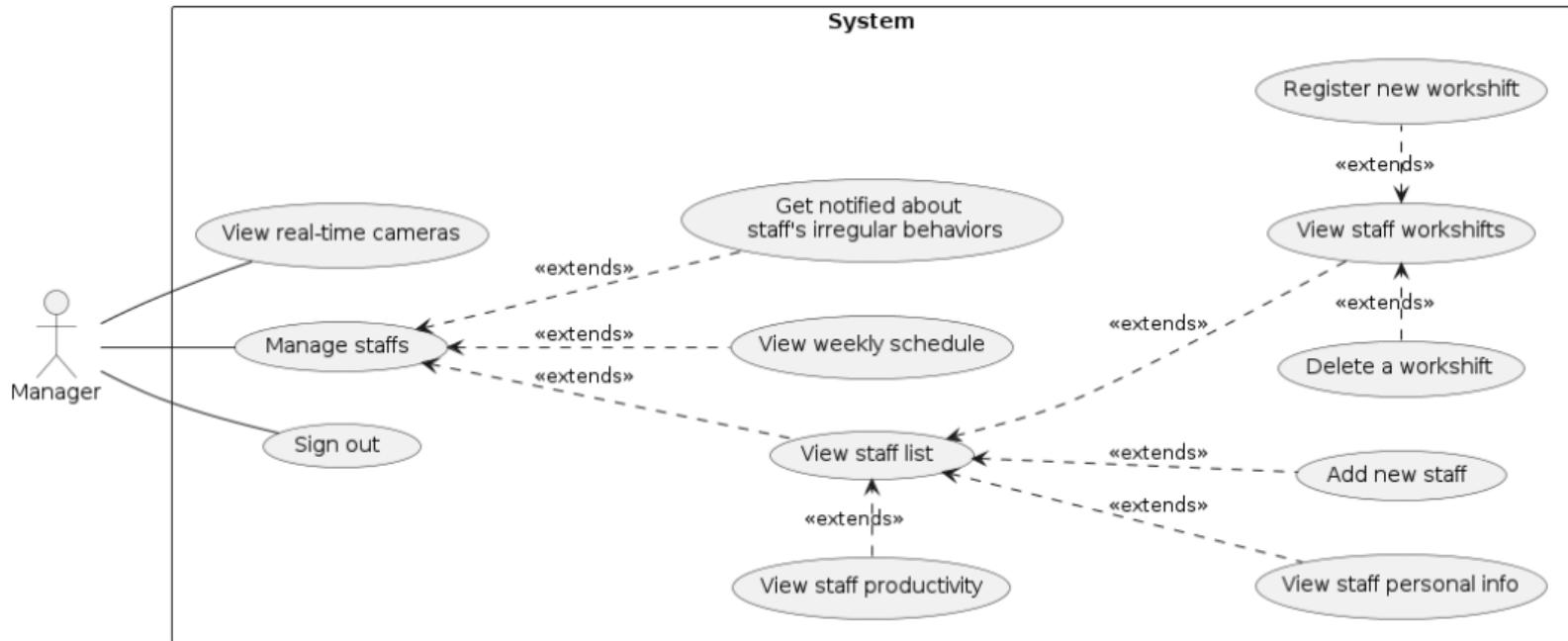
With Re-ID, all individuals wearing the same uniform are considered the same person.

- ◊ The proposed Spatio-Temporal Association can associate people in multiple cameras when they have similar appearance.
- ◊ Issues:
 - missing detection: improve with FP filtering and window-based mapping.
 - inaccurate foot point interpolation: improve with Pose, but need FP filtering to be applied.
- ◊ Those extensions have more impact on complex cases and can be combined to produce a more impressive result.

APPLICATION SYSTEM DEVELOPMENT

Application: Develop a software system to showcase the applicability and demand for the proposed solution.



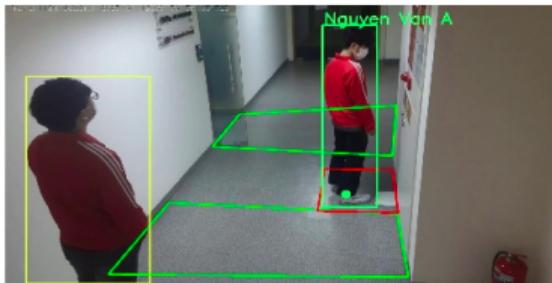
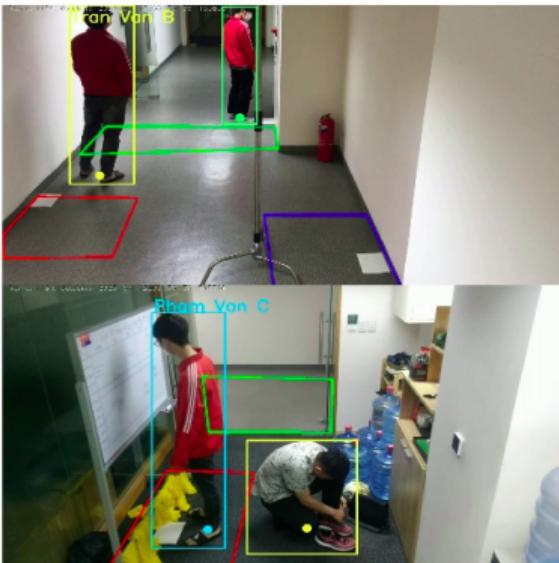


Activities Google Chrome ▾ 14:49 27 Thg 7

Employee Management + Not secure | 0.0.0.0:5555/view_cameras

Employee Management Home Manage staff View cameras View weekly schedule Messages 10 Logout

Camera view



Activities Google Chrome ▾ 15:02 27 Thg 7

Employee Management + ↻ ↺ ⚡ Not secure | 0.0.0.0:5555/productivity/intern

Employee Management Home Manage staff View cameras View weekly schedule Messages Logout

Productivity report: Nguyen Van A

Day	Date	Day shift	Start time	End time	Arrival	Staying time
Wednesday	12/04/2023	morning	08:30:10	08:31:30	Not yet	
Tuesday	11/04/2023	morning	08:30:10	08:31:30	08:30:11 (a few seconds late)	68.9% (0:00:55)
Monday	10/04/2023	morning	08:30:10	08:31:30	08:30:19 (a few seconds late)	56.4%
Friday	07/04/2023	morning	08:30:10	08:31:30	08:30:02	61.4% (0:00:49)

The screenshot shows a web browser window for 'Employee Management' on a local server at 0.0.0.0:5555/messages. The page title is 'Messages'. The interface includes a top navigation bar with links for 'Manage staff', 'View cameras', 'View weekly schedule', 'Messages', and 'Logout'. Below this, the main content area displays a list of messages in a conversational format:

- a few seconds ago*
Tran Van B was absent from work area since 08:31:04
- a few seconds ago*
Tran Van B is back to work area at 08:30:58
- a few seconds ago*
Tran Van B was absent from work area since 08:30:34
- a minute ago*
Pham Van C is back to work area at 08:30:38
- a minute ago*
Pham Van C was absent from work area since 08:30:29
- a minute ago*
Tran Van B is back to work area at 08:30:33
- a minute ago*
Pham Van C arrived back at 08:30:00

Conclusion:

- ◊ A Spatio-Temporal Association for MCT was proposed to replace Re-ID when people have similar appearance.
- ◊ Extensions were incrementally introduced to improve the proposed method.
- ◊ A software system was developed to showcase the applicability of the proposed method.

Improvement:

- ◊ Improve limitation in previous step of MCT (person detection, single camera tracking) such as ID switch, missing detection.
- ◊ Optimize software system.

-  Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft.
Simple online and realtime tracking.
In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3464–3468, 2016.
-  Joseph Redmon and Ali Farhadi.
Yolov3: An incremental improvement, 2018.
-  Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao.
YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.
arXiv preprint arXiv:2207.02696, 2022.
-  Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang.
Bytetrack: Multi-object tracking by associating every detection box.
2022.