Detecting Falls and Slips in Wheelchair Users Using Low-Resolution Thermal Imaging

ICCE Consumer Systems for Healthcare and Wellbeing (CSH) 12/1/2025

Shisei Nakamura †1, Masaaki Yamauchi †1, Miwa Sugita †2 †3, Yoshihiro Aso †4, Yuichi Ohsita †5 and Hideyuki Shimonishi †5

†1 Graduate School of Information Science and Technology, Osaka University,

†2 Graduate School of Engineering, Osaka University, †3 Montessori Care Association Japan,

†4 Advanced Network and DX Strategy Dept, NEC Corporation,

†5 D3 Center, Osaka University

大阪大学
OSAKA UNIVERSITY

INTRODUCTION

Aging Society and Privacy Concerns

Monitoring individuals in private rooms while protecting privacy is essential.

Falls of wheelchair uses

Falls are a major cause of fatal injuries especially wheelchair user.

Caregiver Workload & Non-Urgent Slips

- Monitoring nursing home rooms heavily burdens caregivers.
- Misclassifying slips as falls increases caregiver workload and risks overlooking real dangers.

⇨ **Fall detection system with protecting privacy.**

DEVICE CANDIDATES

	Privacy	Non-invasive	Easy of use	Affordability	Accuracy
Wearable sensor	✓	×	×	✓	✓
Wi-Fi UWB	✓	✓	×	✓	×
RGB camera	×	✓	✓	✓	✓
3D LiDAR	✓	✓	✓	×	×
Depth camera	✓	✓	✓	✓	×
Thermal camera	✓	✓	✓	✓	?

TECHNICAL ISSUES & CONTRIBUTIONS

- Existing video-based fall detection using CNNs.
- CNNs struggle to extract features from thermal images

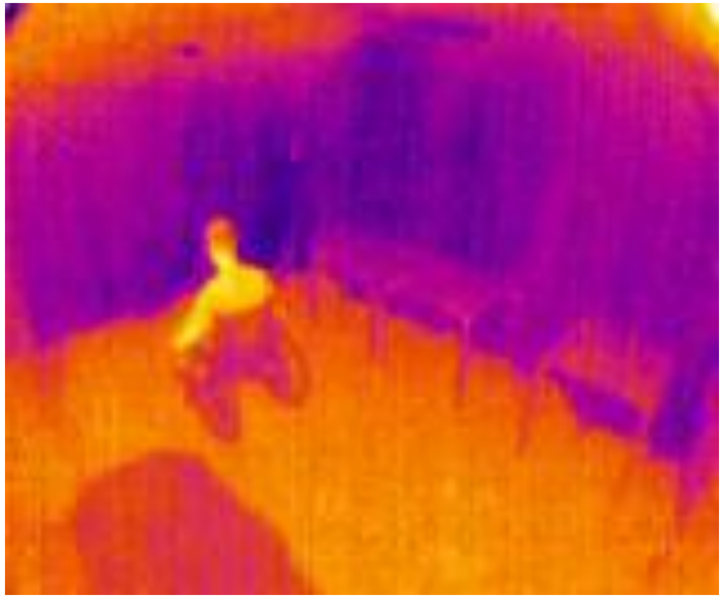
Thermal images

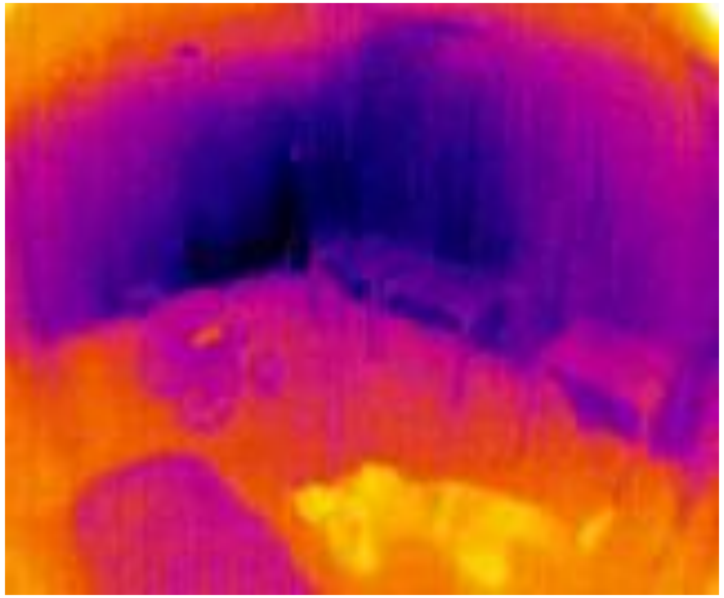
✓ Lack of fine textures

✓ Rough silhouettes

✓ Blurry body parts

✓ Temperature noise

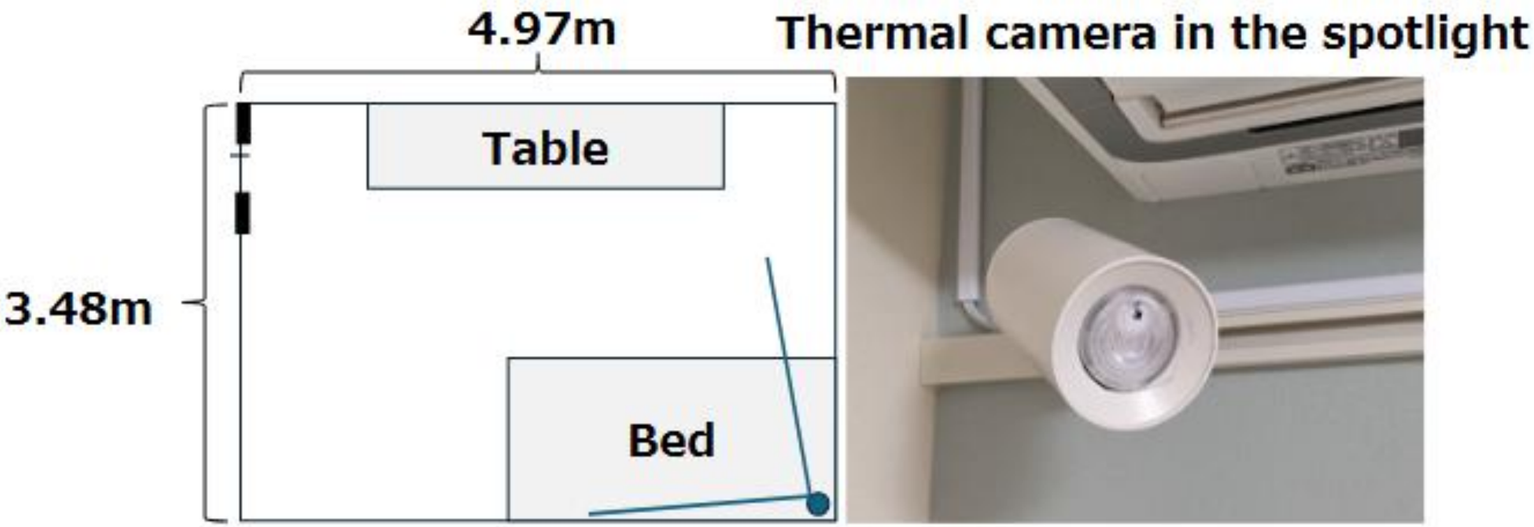




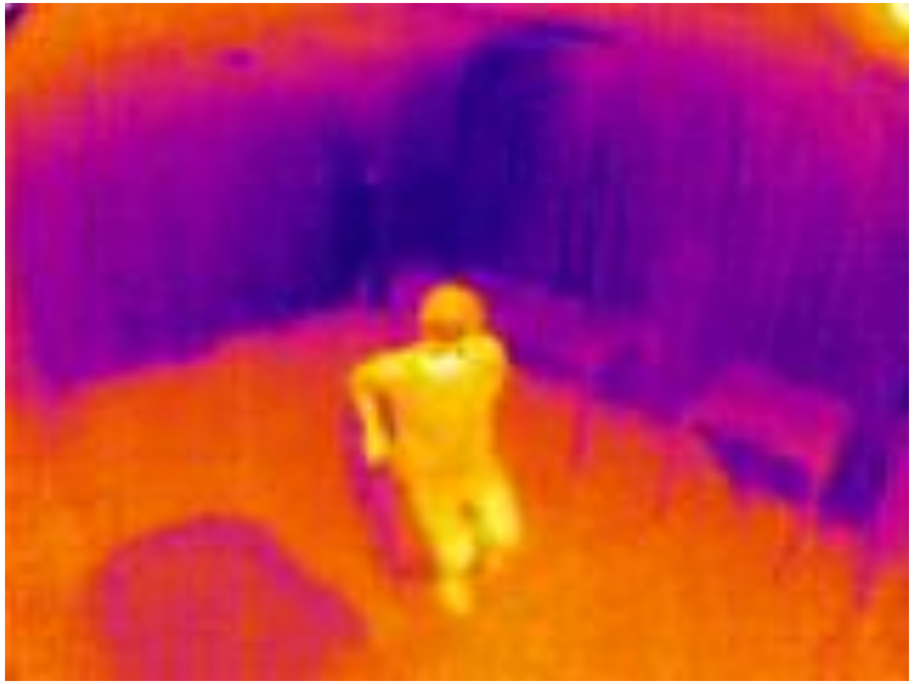
Fall Detection using CNN and Torso Features (FDCTF)

DATA COLLECTION

- Simulated private room environment.
- 10 to 30 seconds videos, fixed camera angle, one person
  - 39 fall events
  - 45 slip events
  - 23 daily activity videos



Simulated room layout and thermal sensor placement



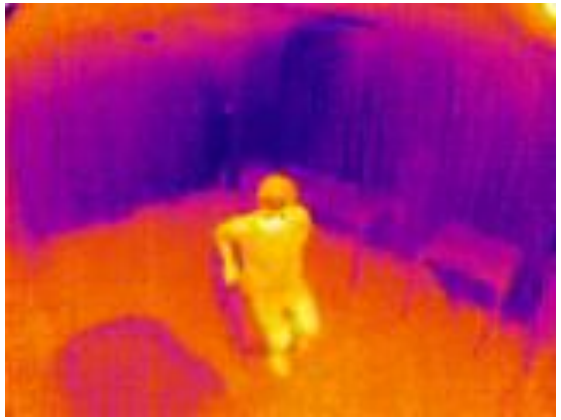
Captured original image

PROPOSED METHOD

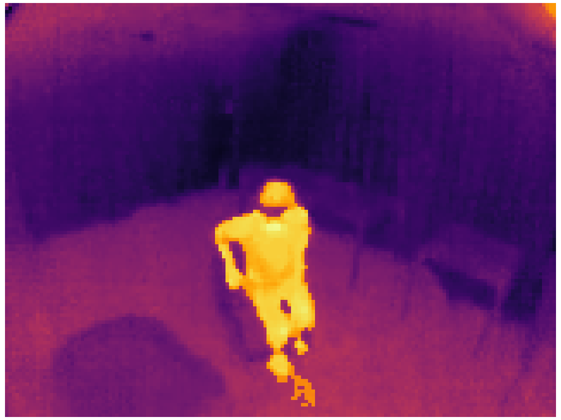
Preprocessing	Emphasize the human figure
Getting features using YOLO	Bboxes & posture probabilities from YOLOv8
Image Features Extraction	Getting torso features (Torso Angle Change)
Frame Level Prediction using LSTM	Frame → ‘usual’, ‘under falling’ or ‘after falling’
Event Level Prediction	Event → ‘caution’ or ‘emergency’

PREPROCESS & YOLO

Preprocessing



Original image

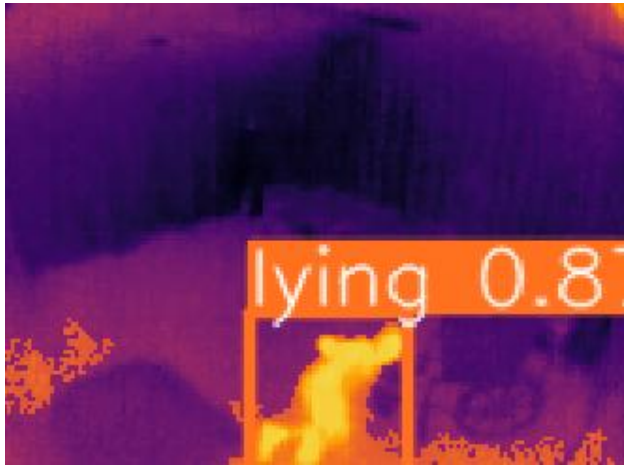


Enhancing human image

Training of YOLO model

- Detect person positions in images
- Output posture class probabilities: Sitting, Lying, Standing
- Train the model using the created dataset

Sitting & Head



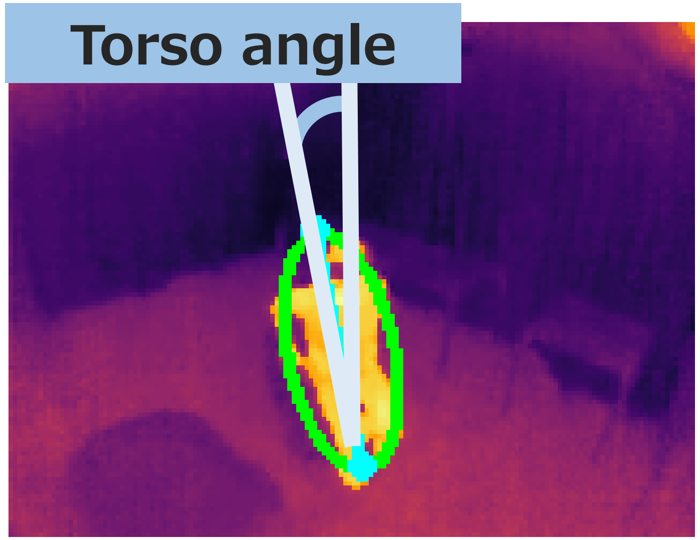
Lying

Torso Angle Changes

Crop and Binarize images

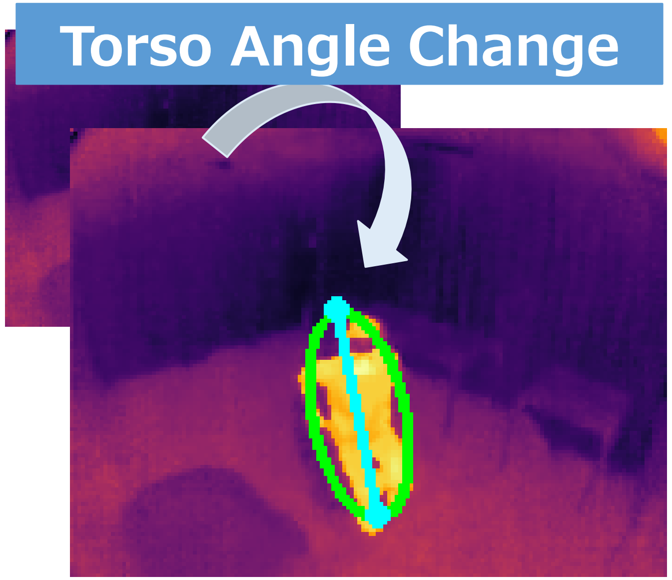
For better feature extraction using the bboxes.

Extract Torso Angle



An ellipse is fitted to the silhouette

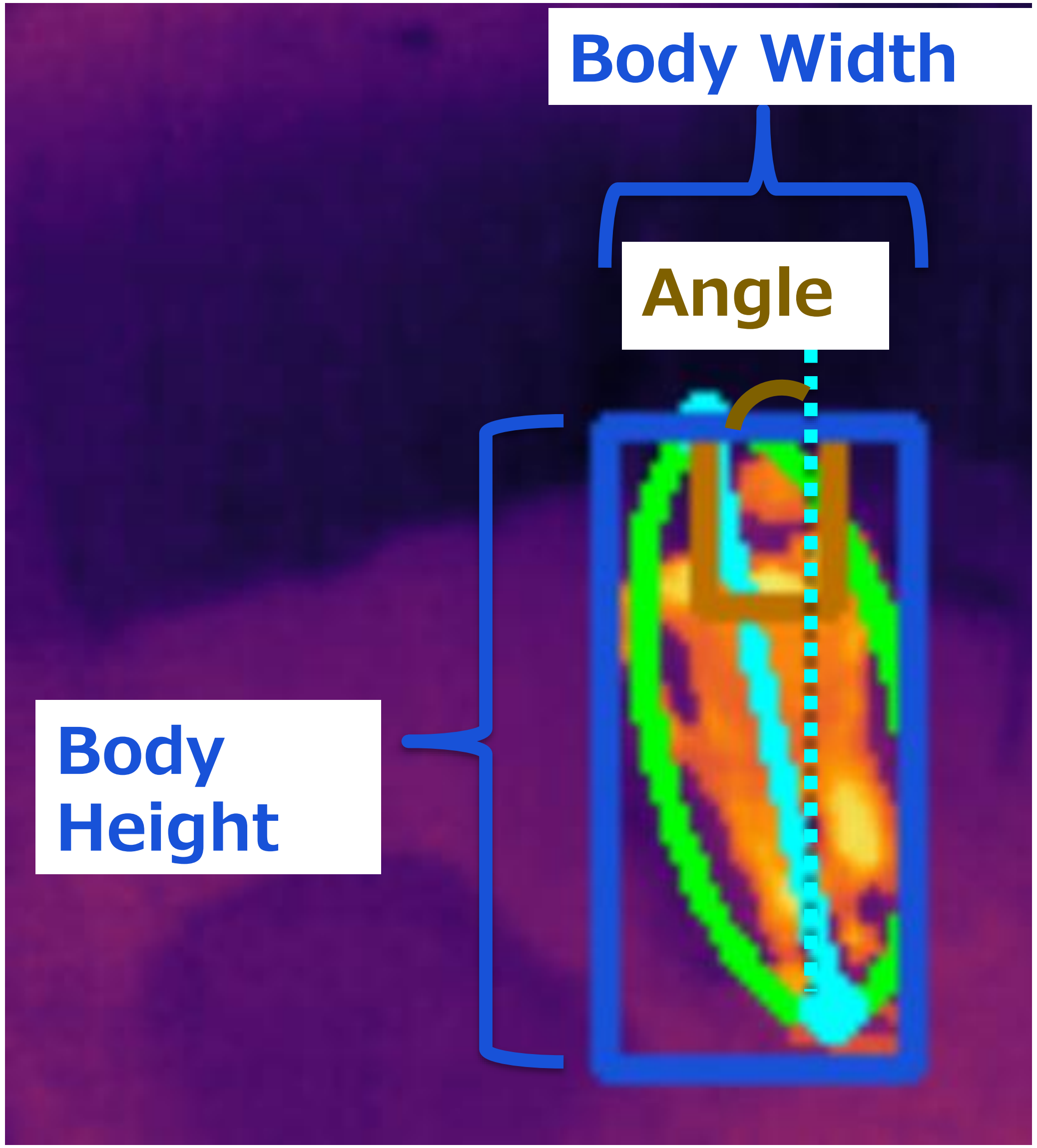
Calculate Torso Angle Change



Absolute difference in tilt angles

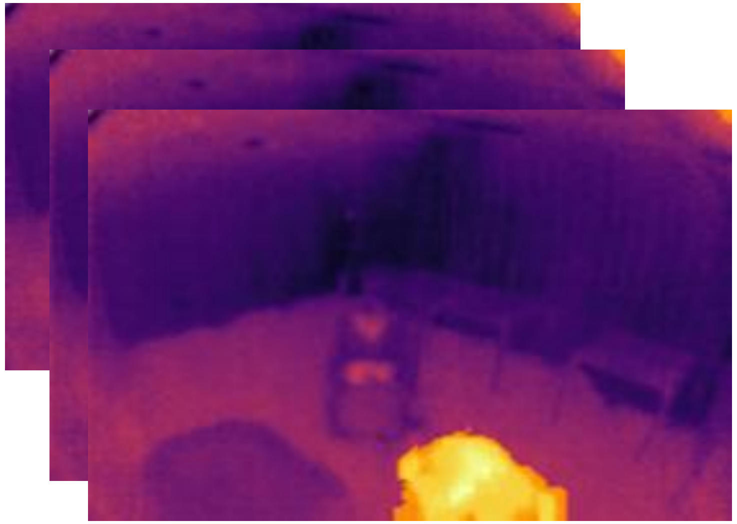
Other Torso features

Feature type	Feature	Abbreviation
Torso Features	Torso Speed	TS
	Head Z-coordinate	Z
CNN Features	CNN Feature Vector	CFV
	Posture Probabilities	PP



Frame Level Prediction

Split video into 4-second



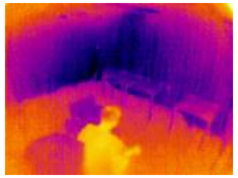
4-second video

Model

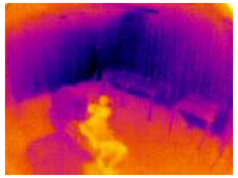
LSTM

Input: CNN & Torso feature

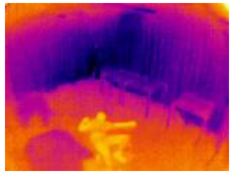
Prediction classification



'usual'



'under falling'



'after falling'

final frame label

Definition of TP/FN

1. 'after falling' frames (C) exceeding C_{max} → 'Pred Event.'
2. Differences(D) between the start frames (D_{max} = 5-seconds)
3. D < D_{max} → TP, D ≧ D_{max} → FN

Actual event1

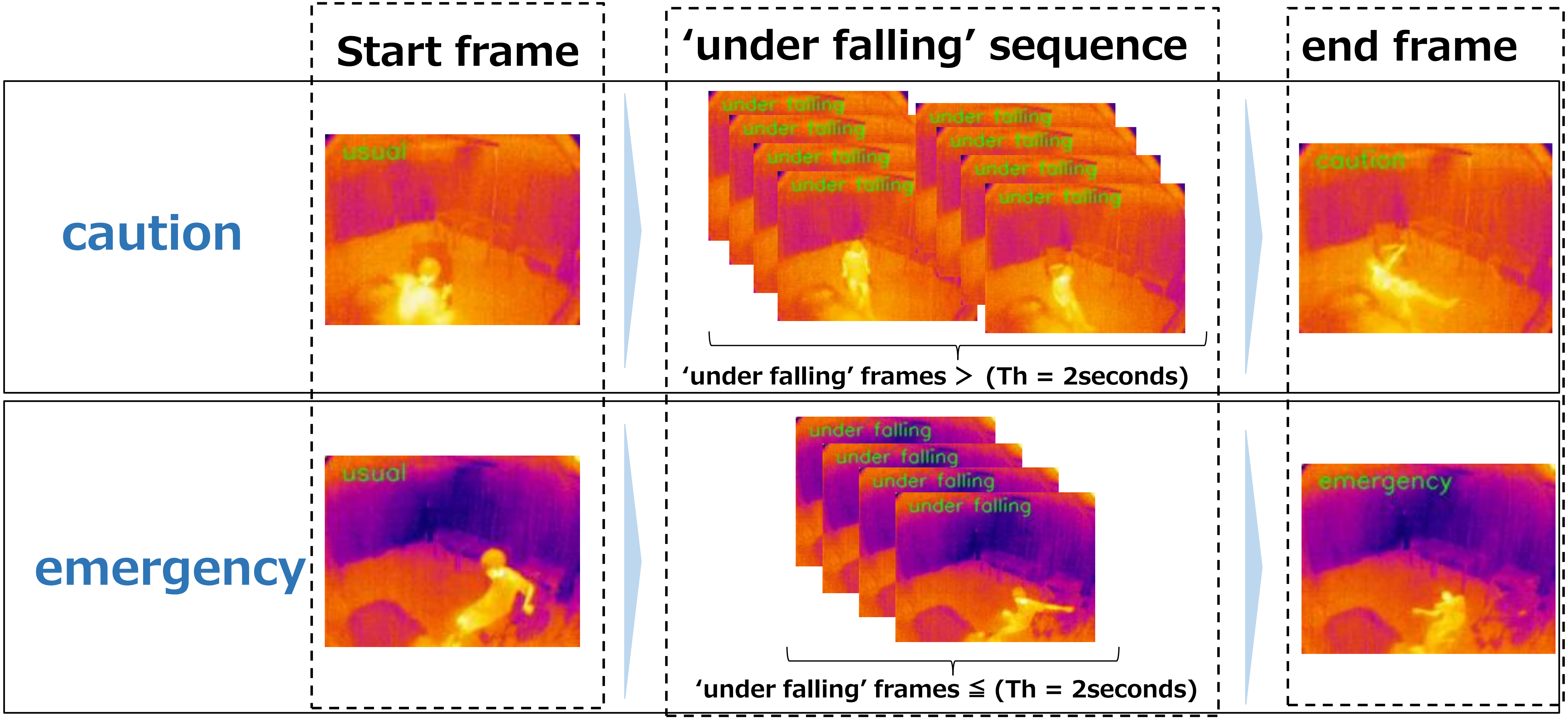
Pred event1

Ground truth	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
Frame level prediction	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1

D=5

C=8

Caution / Emergency



Single Environments Results

- CFV(raw images) >CFV(thermal image)
- FDCTF method > only CNN-based features
- CFV + PP +TAC is highest combination

LSTM input features		F1
CNN Features only	CFV (from raw images)	0.56
	CFV (from binarized thermal image)	0.67
	CFV+PP	0.91
FDCTF	CFV+PP+TAC	1.00
	CFV+PP+Z	0.95
	CFV+PP+TS	0.96

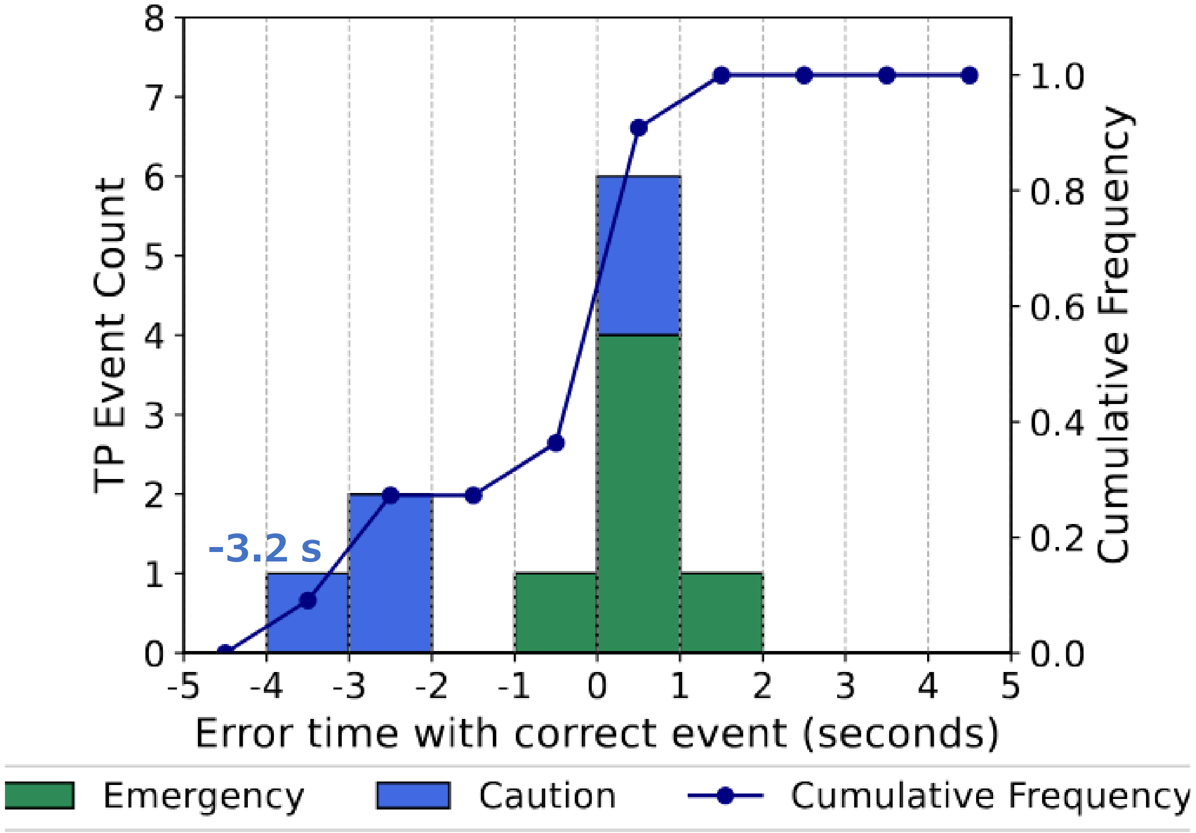
Multi Environments Results

- CFV(raw images) >CFV(thermal image)
- FDCTF method > only CNN-based features
- CFV + PP +TAC is highest combination

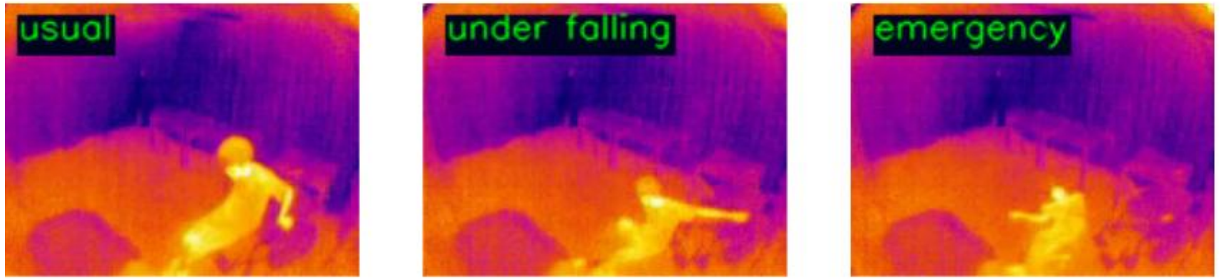
LSTM input features		F1
CNN Features only	CFV	0.56
	CFV+PP	0.67
FDCTF	CFV+PP+TAC	**1.00**
	CFV+PP+TAC +Multi(acc, tilt)	
	CFV+PP+TAC +Multi(acc, tilt)	0.95
	CFV+PP +Multi(acc, tilt)	0.96

CFV+PP+TAC RESULTS

- CFV+PP+TAC: F1 score of 1.0 (≦ 3.2s)
- 5.2 FPS (CPU i7-9700K) , 8 FPS (GPU RTX 2070)



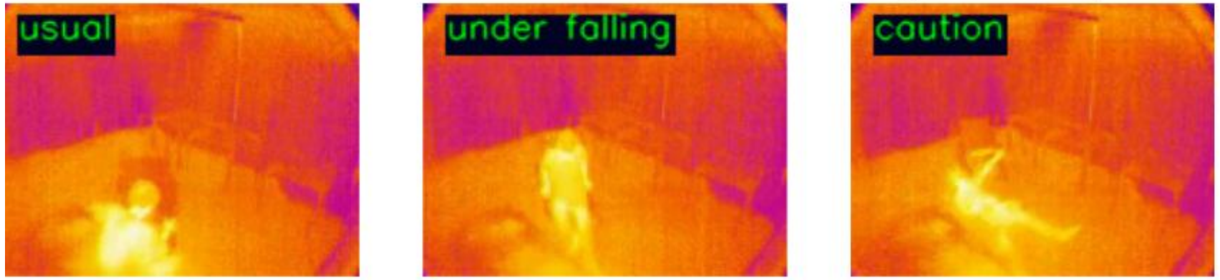
(a) to (c) = 'emergency', (d) to (f) = 'caution'.



(a) Usual sitting on a wheelchair

(b) Under falling 1 second

(c) Emergency (high-risk fall)



(d) Usual on a bed

(e) Under falling 7 seconds

(f) Caution (low-risk slip)

Video: Emergency

https://www.ane.cmc.osaka-u.ac.jp/wp-content/uploads/2024/11/Fall_Detection_using_CNN_and_Torso_Features_video.mp4

Video: Caution

https://www.ane.cmc.osaka-u.ac.jp/wp-content/uploads/2024/11/Fall_Detection_using_CNN_and_Torso_Features_video.mp4

CONCLUSION & FUTURE WORK

Conclusion

- FDCTF combines CNN and torso features to improve accuracy.
- FDCTF meets requirements.
  - Privacy, Non-invasive, Ease of use, Affordable, Accuracy
  - Realtime running

Future work

Enhanced robustness in various environments:

- Different camera positions and angles
- Diverse individuals

Appendix

Video

https://www.ane.cmc.osaka-u.ac.jp/wp-content/uploads/2024/11/Fall_Detection_using_CNN_and_Torso_Features_video.mp4

Others