IEEE ICHI 2026
June 1–3, 2026
Minneapolis, MN

Scientific Session 14 / System Track

A Scalable Dual-Camera Platform for Behavioral Assessment in Healthcare

Mostafa Habibi, Mehrdad Nourani
Department of Electrical and Computer Engineering
The University of Texas at Dallas

Dennis H. Sullivan, MD
Donald W. Reynolds Department of Geriatrics
University of Arkansas for Medical Sciences

Presenter: Mostafa Habibi

Outline

- Motivation
  - AI-Powered Solution
- Prior Works
- Proposed Methodology
  - Detection, Tracking, Identification, and Filtering
  - Behavior and Spatial Analysis
  - Interaction Assessment and Reporting
- Experimental Results
- Limitations, Future Work, and Conclusion

Motivation

IEEE ICHI 2026
June 1–3, 2026
Minneapolis, MN

- Hospitalized patients, on average, spend more than 85% of their time in bed
    - Immobility leads to DVT, pneumonia, pressure ulcers, infections, and increased mortality
- Early and progressive ambulation (EPA)
    - Proven to reduce risks, shorten stays, lower costs, and improve recovery outcomes
    - Based on nurse observations; thus, inconsistent, incomplete, and difficult to scale
- Camera-based ambulation monitoring systems offer
    - Non-intrusive, Automation, Continuous
- **These limitations motivate a camera-based approach for reliable monitoring in an occluded clinical environment**

AI-Powered Camera-Based HOI Monitoring

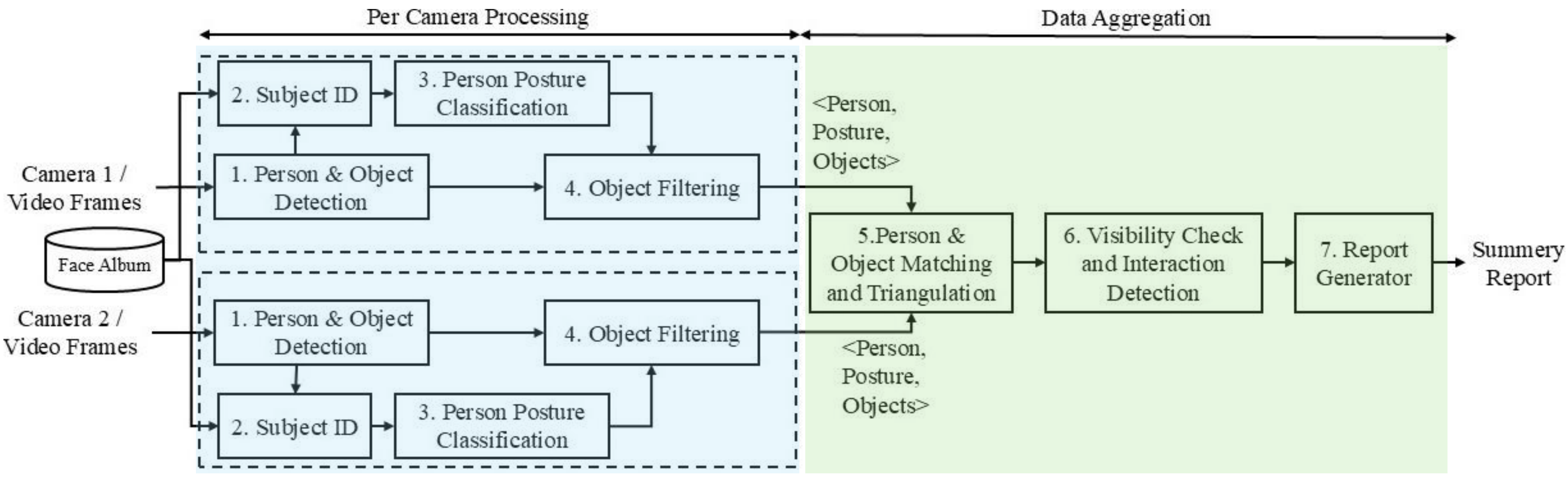
- Human-object interaction detects how people interact with objects
  - Not only detecting people and objects
- In healthcare, this helps monitor patient behavior and interaction
  - Patients may interact with beds, chairs, wheelchairs, walkers, canes, and IV poles
- Camera-based systems provide a passive and non-wearable way to monitor patients
  - Detect the patient
  - Detect clinically relevant objects
  - Estimate their spatial relationship
  - Summarize behavior in a clinically useful way

Prior Works

IEEE ICHI 2026
June 1–3, 2026
Minneapolis, MN

- Single-camera systems showed that automated patient monitoring is practical
  - [Gabriel et al.] used CNN-based object detection and MediaPipe for ambulation assessment.
  - [Habibi et al.] tracked human pose, object location, and interaction using one RGB camera.
- Single-camera systems have limitations
  - Occlusion
  - Limited field of view
  - Weak monocular depth estimation
  - Lower 3D localization accuracy
- Multi-camera systems improve visibility and spatial understanding
  - [Meratwal et al.] used YOLOv4, DeepSORT, and attention-based LSTM for indoor activity recognition
  - Wearable and multimodal systems, such as PAMS, can track movement and posture well
- Current solutions do not provide **human-object interaction analysis** for clinical monitoring

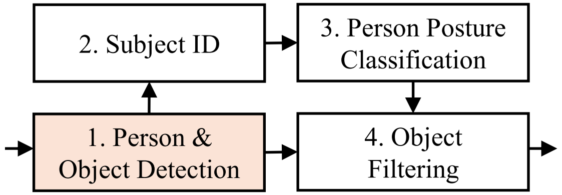
Proposed Methodology



IEEE ICHI 2026
June 1–3, 2026
Minneapolis, MN

Per Camera Processing

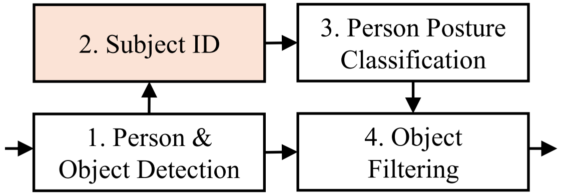
1. Person & Object Detection and Tracking



- Created a custom dataset of **23,704 annotated** images to retrain YOLOv9 for 14 clinically relevant hospital objects.
- Used **BoT-SORT** to assign unique IDs to detected patients and objects across video frames.
- Maintained temporal consistency for real-time tracking in dynamic hospital environments.

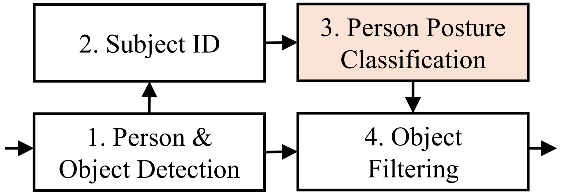
Objects	# of Annotations	Objects	# of Annotations
TV	2393	IV_Pole	1165
Bed	1780	Person	1301
Chair	1812	Sofa	2662
Table	456	Door	801
Cane	2309	Wheelchair	3534
Walker	1298	Crutch	1491
Food Tray	1542	Toilet sign	1050

2. Subject Identification

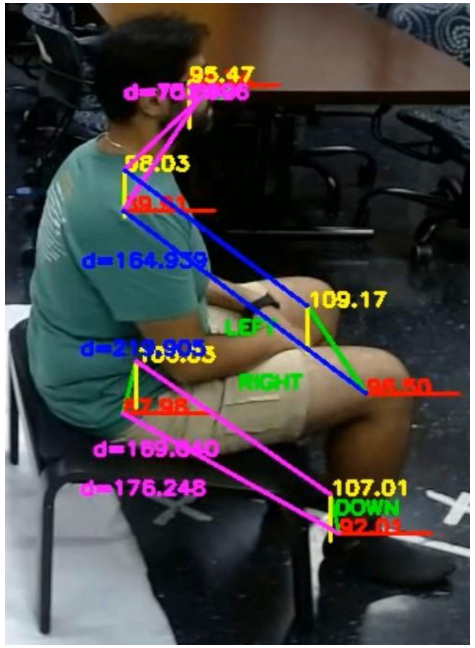


- Used **DeepFace** to assign consistent patient identification across frames and camera views.
- A small face album is created using a few reference images for each subject.
- DeepFace performs three main steps:
  - **Face detection and alignment**: Using CNN backbone
  - **Feature extraction**: 4096-dimensional identity embedding.
  - **Face matching**: Using cosine similarity
- If the face is occluded, BoT-SORT maintains the tracking ID across frames.
  - If recognized later, the name is assigned to previous frames with the same track ID.
  - If unmatched, the subject is labeled as Unknown–Track ID for review.

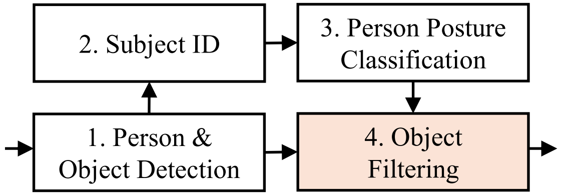
3. Person Posture Classification



- ➢ Used **MediaPipe** to extract **33 body landmarks** in real time.
  - ➢ Selected **11 key landmarks**: eyes, shoulders, hips, knees, ankles, nose.
  - ➢ Extracted **33 geometric features** (6 distances and 27 angles)
- ➢ Classified three clinically important postures
  - Sitting, Standing, and Lying
- ➢ Created a custom dataset from **40 videos** of **8 subjects**
  - 12,657 sitting, 10,597 standing, and 17,235 lying (40,489 annotated frames).
- ➢ Used a **Random Forest classifier** for robust posture recognition in clinical environments.



4. Objects Filtering



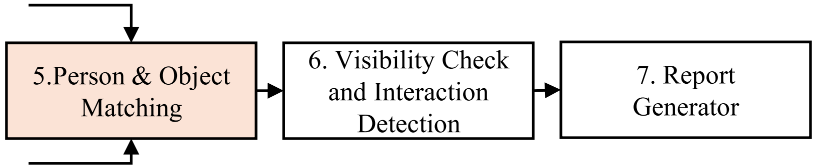
- ➢ Relational table links patient postures to meaningful object interactions

- ➢ Using logical real-world constraints
  - Example: A patient can lie in bed, but not in chair

Posture	Objects	Posture	Objects	Posture	Objects
Sit	Bed	Stand	Cane	Lie	Bed
	Chair		Walker		Sofa
	Sofa		Crutch		Table
	Table		IV Pole		
	Wheelchair				

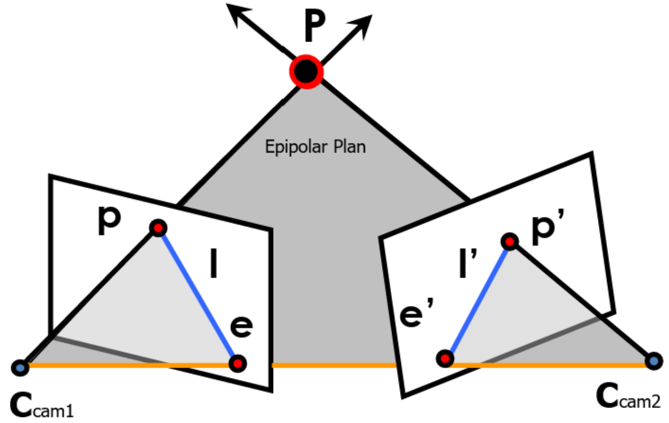
Data Aggregation

5. Person & Object Matching



5.1) Fundamental Matrix & Epipolar Geometry

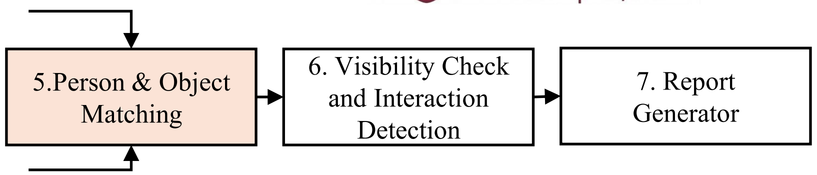
- F describes the geometric relationship between two camera views, where corresponding points p and p' satisfy $p'^T F p = 0$.
- For a point p in one image, its match in the second image lies on the epipolar line $l' = Fp$.



5.2) Feature Extraction & Matching with SIFT

- Candidate bounding boxes near the epipolar line are selected and SIFT extracts local keypoints with 128-dimensional descriptors.
- Brute-force matching compares descriptors between camera views, and matches are accepted using the nearest-neighbor ratio test.

5. Person & Object Matching (Cont.)



5.3) Camera Calibration and Projection

- Estimate intrinsic parameters (K) and extrinsic parameters (R,t) to project 3D points onto the 2D image plane using $P = k[R|t]$.

$$k = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

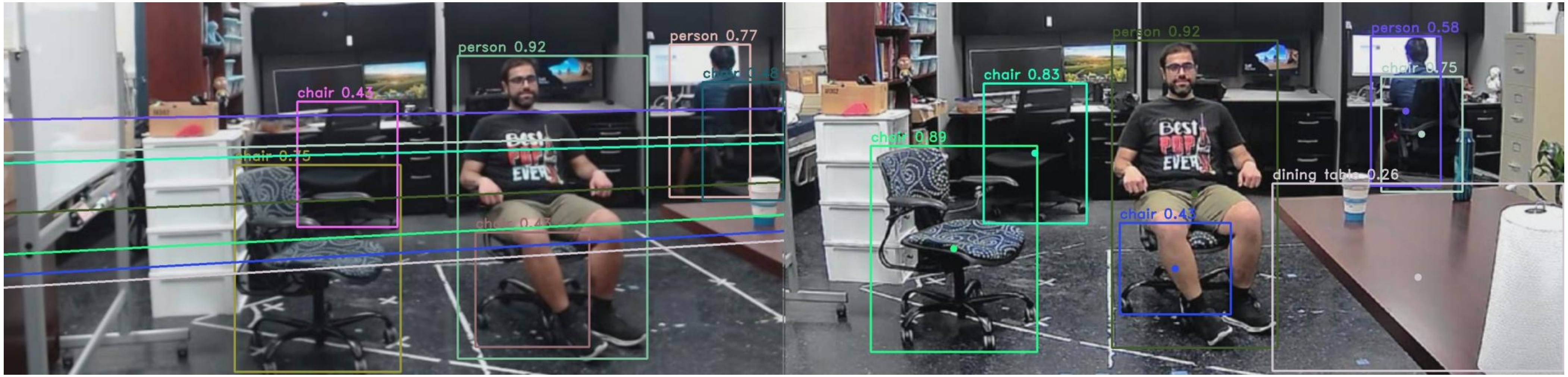
5.4) Triangulation & 3D Reconstruction

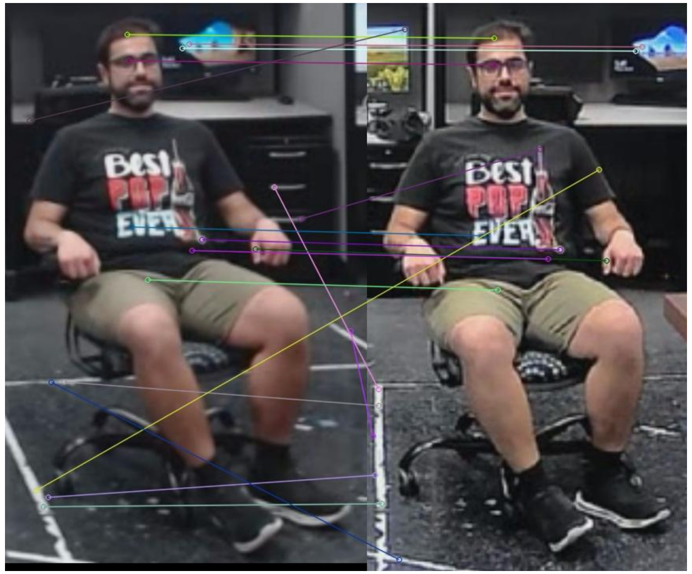
- For each 3D point X, dual-camera system provides a corresponding 2D projection

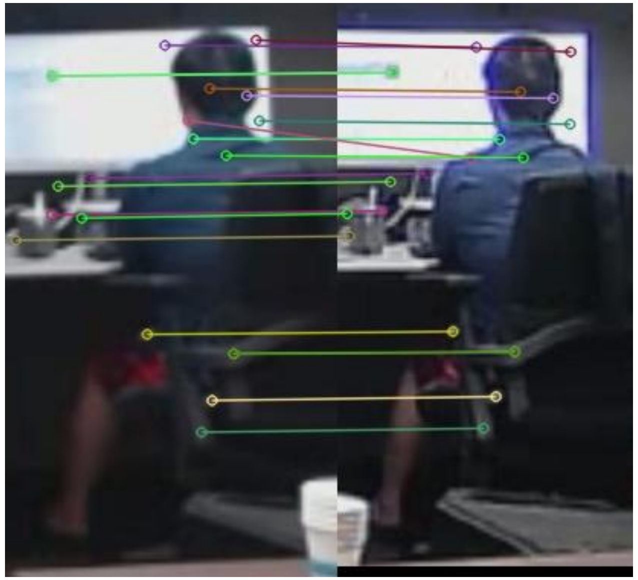
$$x_1 = P_1 X, x_2 = P_2 X$$

- Transform this system into linear equations and solve it to estimate the 3D pint X.

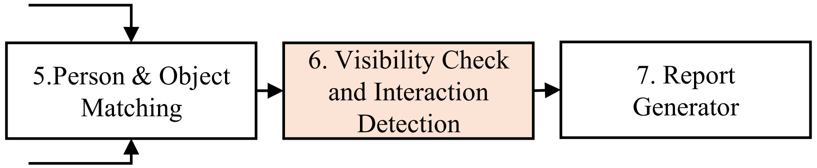
5. Person & Object Matching (Cont.)





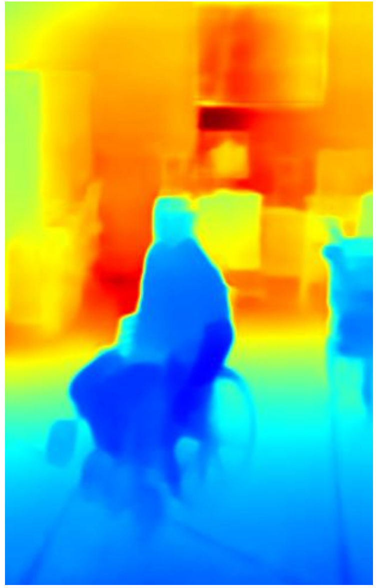


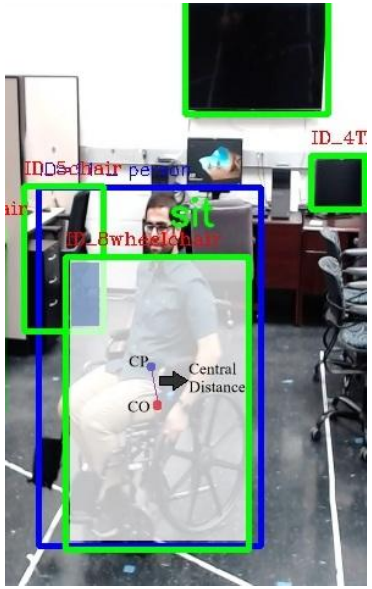
6. Visibility Check and Interaction Detection



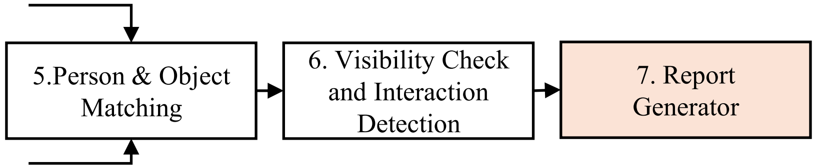
- Real-world clinical environments often include **occlusion** and **restricted visibility**.

- The system handles three visibility cases
  - **Full visibility:** Person and object are visible in both cameras
    - Triangulate 3D positions of the person and object
    - Analyze interaction using posture and real-world distance
  - **Partial occlusion:** Person or object is visible in only one camera
    - Estimate depth using disparity maps
    - Use 2D position, IoU, central distance, and depth cues
  - **Single view:** Both person and object are visible in only one camera
    - Use 2D position, IoU, and central distance
    - Apply monocular depth estimation, such as AdaBins

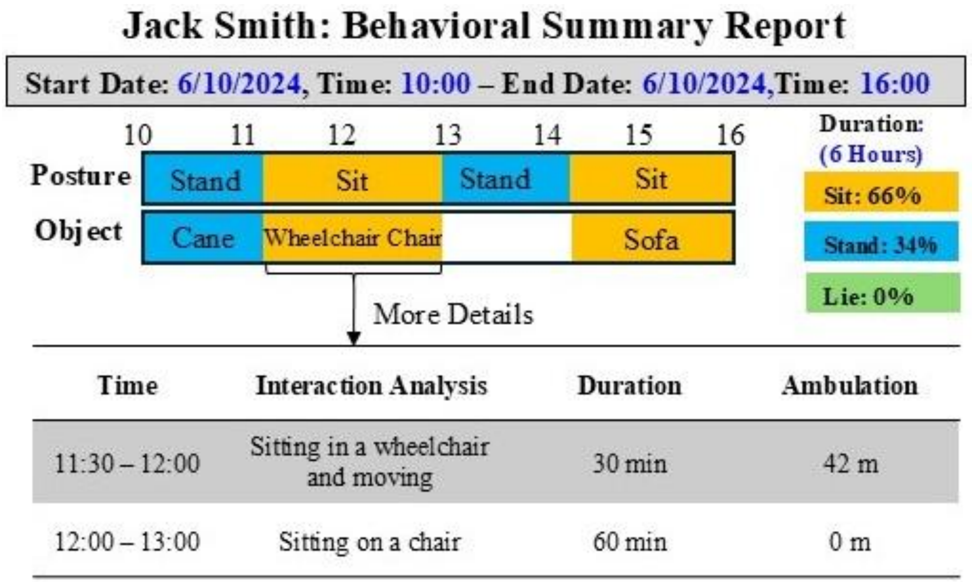




7. Report Generation



- Extracted frame-level patient status over time and recorded
  - Timestamp, Posture, and Interaction state
  - Segmented the video into distinct activity episodes
  - This helps clinicians monitor **mobility**, **detect inactivity**, and **ambulation**

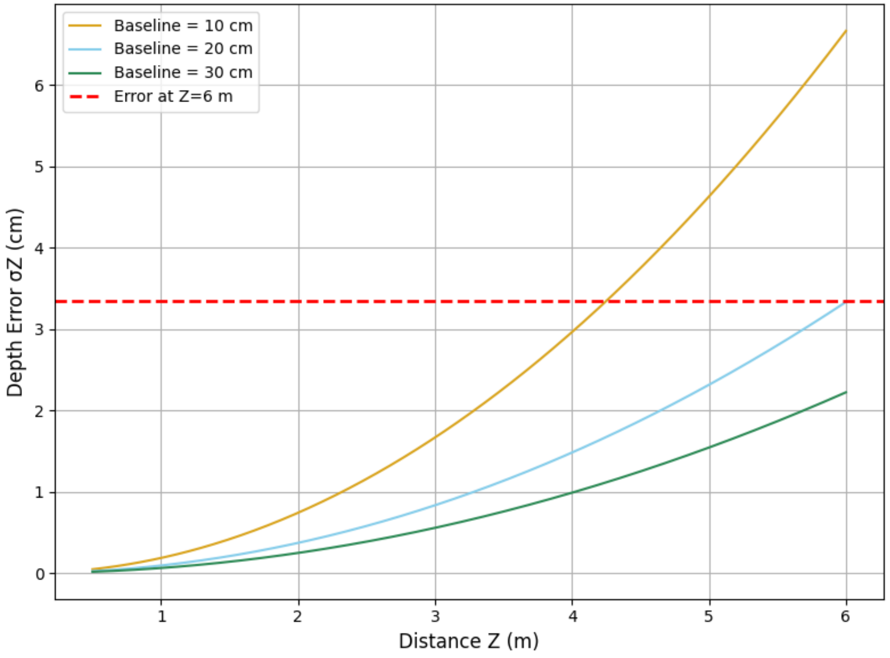


Experimental Results (1 of 3)

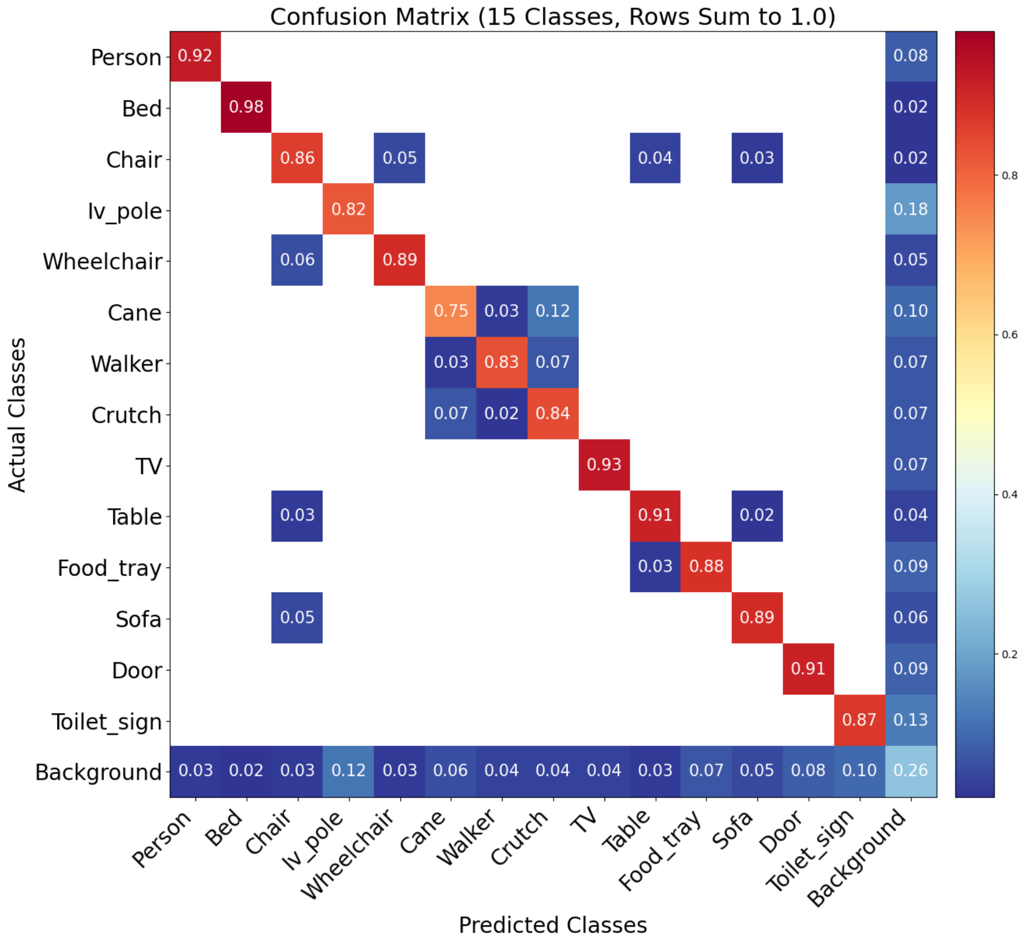
IEEE ICHI 2026
June 1–3, 2026
Minneapolis, MN

➢ System Setup

- Laptop: Intel i9-12900HK CPU, 32 GB RAM, NVIDIA RTX 3050
- A regular webcam Logitech C920S camera
- A 20 cm baseline provides sufficient depth accuracy for monitoring up to 6 m



➢ Detection and Tracking Performance



Experimental Results (2 of 3)

IEEE ICHI 2026
June 1–3, 2026
Minneapolis, MN

- Posture Classification
  - Random Forest achieved the best posture classification performance on the custom dataset

Classifier	Accuracy	Precision	Recall	F1 Score
Random Forest	**99**	**93**	**91**	**98**
SVM	88	89	87	88
K-NN (n=5)	83	84	82	83

  - Generalized well on the POLAR public dataset for sitting, standing, and lying postures

Posture	Frames	Accuracy
Sitting	5,101	91
Standing	4,656	89
Lying down	3,581	88

- Person & Object Matching and Triangulation
  - Evaluated SIFT matching on 1,200 dual-camera images
    - Accuracy **97%** for large objects, **93%** for small objects
  - Dual-camera triangulation improved 3D localization and ambulation tracking (5 participants)

Walking Speed	Single-Camera	Dual-Camera
Slow	95%	**98%**
Normal	92%	**96%**
Fast	88%	**94%**

Experimental Results (3 of 3)

- Visibility Check and Interaction Detection
  - Evaluated HOI detection with 15 participants across sitting, lying, and standing postures
  - A prediction was correct only when **posture**, **object**, and **interaction** were all correctly detected

Scenario	Posture	Test Frames	Correct Detections	Accuracy (%)
Full Visibility	Sitting	812	788	**97**
Partial Occlusion	Sitting	715	658	92
Single View	Sitting	620	527	85
Full Visibility	Lying	740	725	**98**
Partial Occlusion	Lying	650	611	94
Single View	Lying	550	490	89
Full Visibility	Standing	720	711	**99**
Partial Occlusion	Standing	650	618	95
Single View	Standing	500	440	88

Limitations, Future Work, and Conclusion

- Limitations
  - Tested in a controlled room with fixed layout and conditions
  - Real clinical settings may include lighting changes, clutter, and unpredictable patient behavior
- Future Work
  - Test on more patient population and real clinical environments
  - Use LLMs to convert detection events into clear clinical summaries
  - Reduce reliance on facial recognition using privacy-preserving re-identification
- Conclusion
  - Developed a proof-of-concept dual-camera platform for patient behavior monitoring
  - Combined object detection, multi-camera matching, triangulation, posture analysis, and interaction detection
  - Generated structured episodic reports to support clinical interpretation and decision-making

Thank you

Questions?

Mostafa Habibi
The University of Texas at Dallas
Mostafa.habibidehsheikhi@utdallas.edu

IEEE
ICHI 2026
June 1–3, 2026
Minneapolis, MN