A Scalable Dual-Camera Platform for Behavioral Assessment in Healthcare

Mostafa Habibi and Mehrdad Nourani
Predictive Analytics & Technologies Lab
THE UNIVERSITY OF TEXAS AT DALLAS
Erik Jonsson School of Engineering and Computer Science
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Motivation

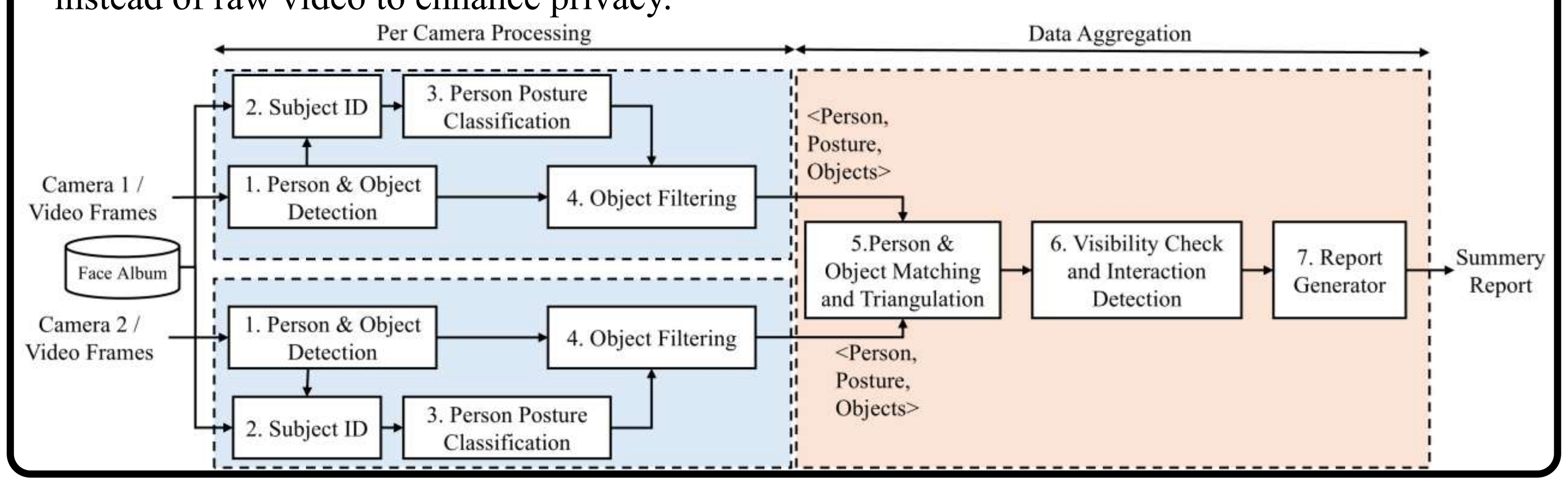
- Hospitalized patients, especially older adults spend more than 85% of time in bed, leading to functional decline and medical complications.
- Early ambulation (EPA) improves recovery and reduces hospital stay.
- Current ambulation monitoring relies on manual observation, which is subjective and inaccurate.
- Camera-based systems provide automated, continuous, non-intrusive tracking.
- Single-camera approaches face challenges such as occlusion and inaccurate 3D estimation.
- Integrating AI-driven HOI monitoring into clinical workflows can enhance decision-making, enable personalized recovery tracking, and detect early signs of deterioration.

Main Contribution

- Extending the detection pipeline to dual synchronized cameras, improving visibility and reducing occlusion errors.
- Integrating multi-view posture classification to align postures and object interactions across cameras.
- Computing 3D coordinates of key landmarks and bounding boxes using triangulation for accurate ambulation tracking.
- Generating activity summaries automatically to support clinical decision-making.
- Introducing the first system to combine multi-camera fusion with posture-aware interaction analysis in healthcare environments.
- Designing a modular architecture enabling flexibility, scalability, and reproducibility for future extensions and evaluations.

Proposed Platform and Methodologies

- **IRB-approved:** All data collection and analysis are conducted under IRB-approved protocols, ensuring ethical and privacy protection.
- **Data Security:** The platform follows hospital-grade standards with encryption, access control, and PII anonymization.
- **Privacy-Aware Design:** Operates in real time and can store only key extracted information instead of raw video to enhance privacy.



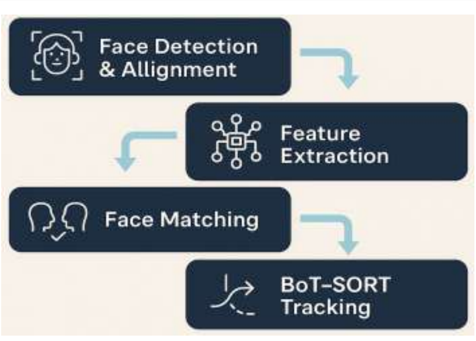
Step 1: Person & Object Detection and Tracking

- YOLOv9 is fine-tuned on a custom dataset (23,704 annotated images) representing real hospital environments.
- The architecture is trained to detect 14 clinically relevant objects across dual-camera views for improved accuracy and visibility.
- BoT-SORT is enabled multi-object tracking and consistent ID assignment across frames.

Object	Annotation	Object	Annotation
TV	2,393	IV Pole	1,165
Bed	1,780	Person	1,301
Chair	1,812	Sofa	2,662
Table	456	Wheelchair	3,534
Cane	2,309	Door	801
Walker	1,298	Crutch	1,591
Food Tray	1,542	Toilet Door Sign	1,050

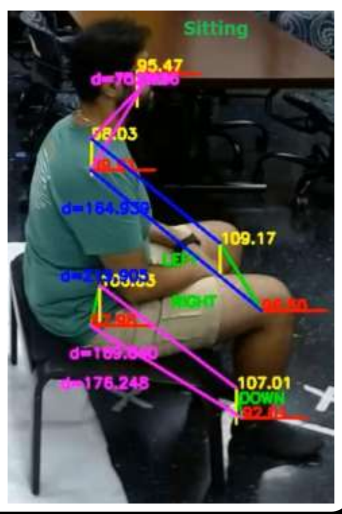
Step 2: Subject Identification

- DeepFace framework is used for reliable multi-view person identification using limited facial data.



Step 3: Person Posture Classification

- A vision-based system to classify postures [sitting, standing, and lying].
- **Dataset** includes 40 videos (8 subjects, 40,489 frames) with three posture classes collected.
- **Feature extraction** uses MediaPipe (33 landmarks) and selects 11 key points to compute 33 geometric features (6 distances and 27 angles).
- **Classification** applies a Random Forest model for posture recognition.



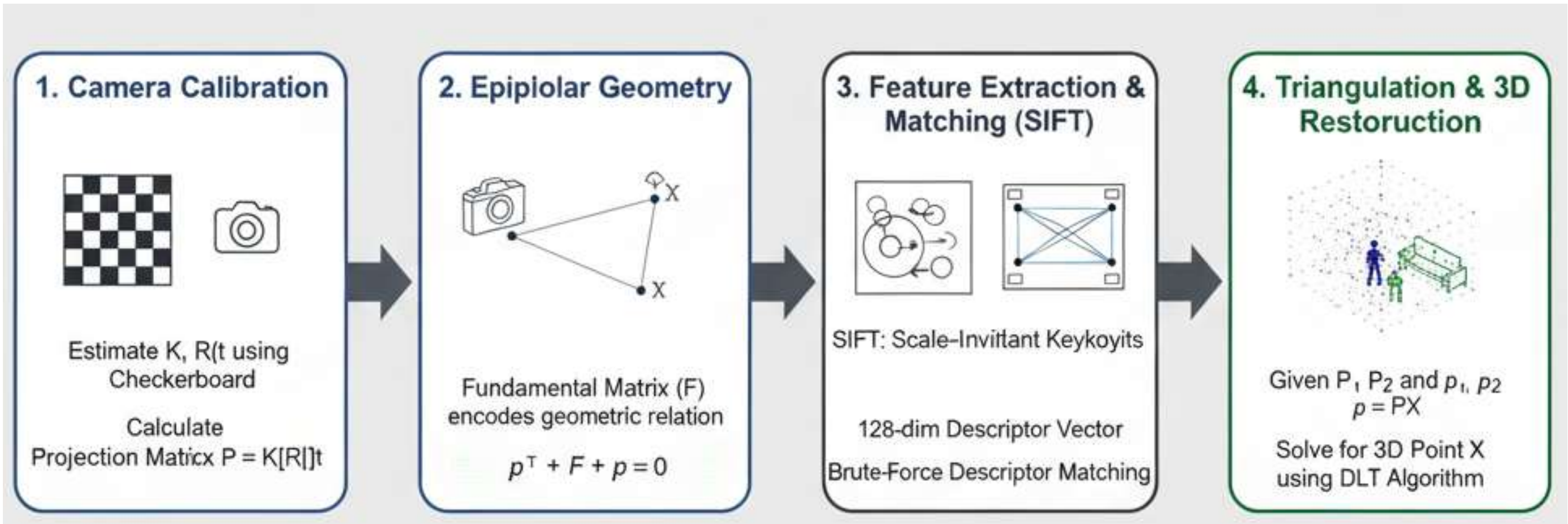
Step 4: Object Filtering

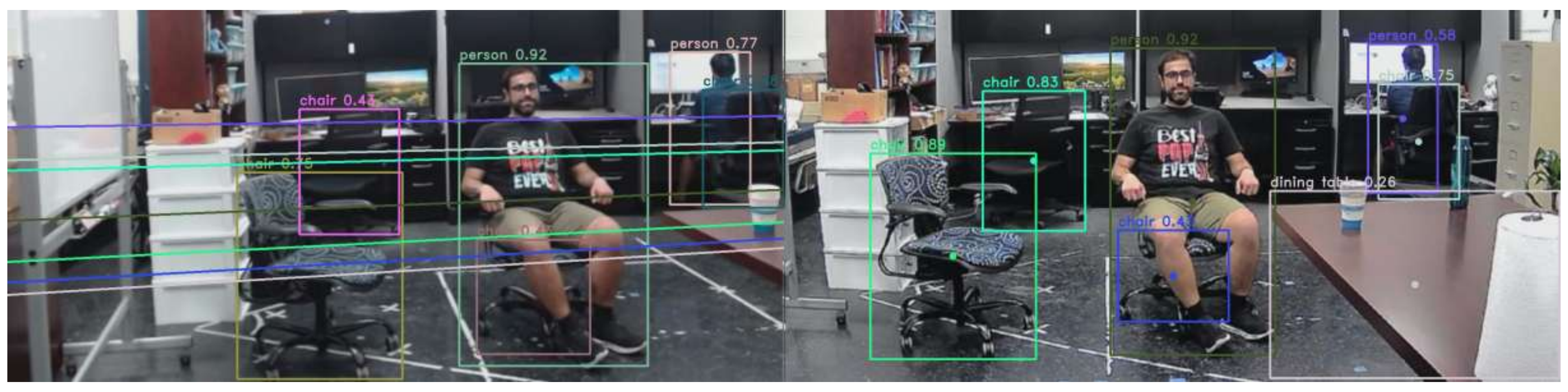
- Implements a rule-based filtering mechanism to validate realistic patient–object interactions based on posture.
- Applies logical constraints to ensure clinically valid relations (e.g., a patient may lie on a bed but not sit on the floor).
- This filtering step enhances interaction accuracy by removing implausible posture-object pairs.

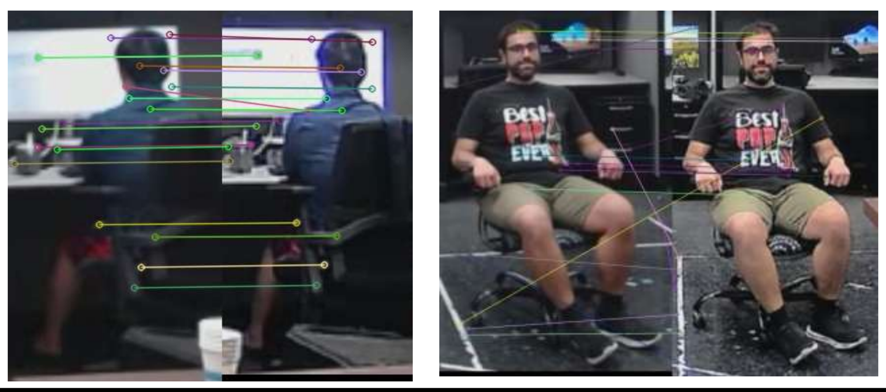
Posture	Relevant Objects
Sitting	Bed, Chair, Sofa, Table, Wheelchair
Standing	Cane, Walker, Crutch, IV Pole
Lying	Bed, Sofa, Table

Step 5: Person & Object Matching and Triangulation

This step focuses on accurately aligning detections (persons and objects) across multiple camera views and reconstructing their 3D spatial positions for precise analysis of movement and interaction.



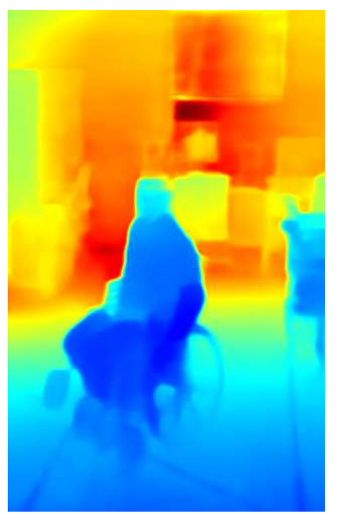


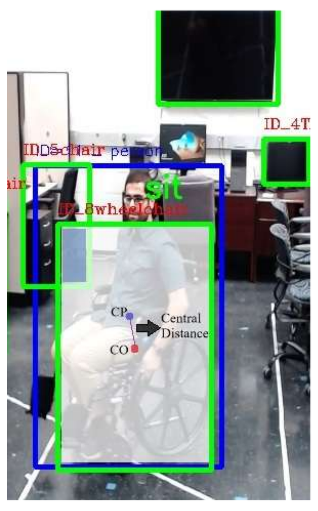


Step 6: Visibility Check and Interaction Detection

Three visibility scenarios:

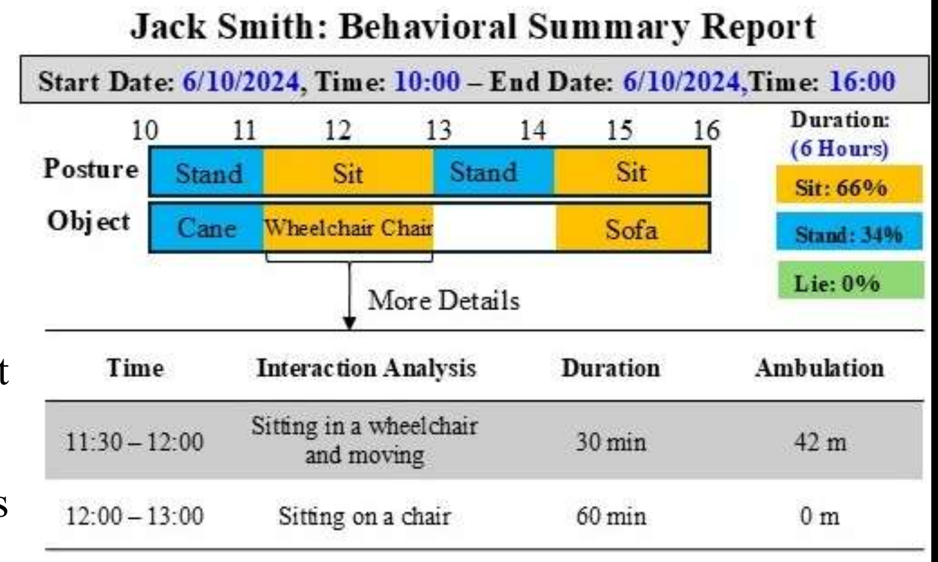
1. **Full Visibility:** Detects both person and object in both views, using triangulation to compute 3D positions and analyze posture-distance interactions.
2. **Partial Occlusion:** Detects one entity in only one view, using disparity-based depth maps, IoU, and central distance for interaction estimation.
3. **Single-View:** Detects both entities in a single view, relying on 2D spatial cues and monocular depth estimation (AdaBins) to maintain depth consistency.





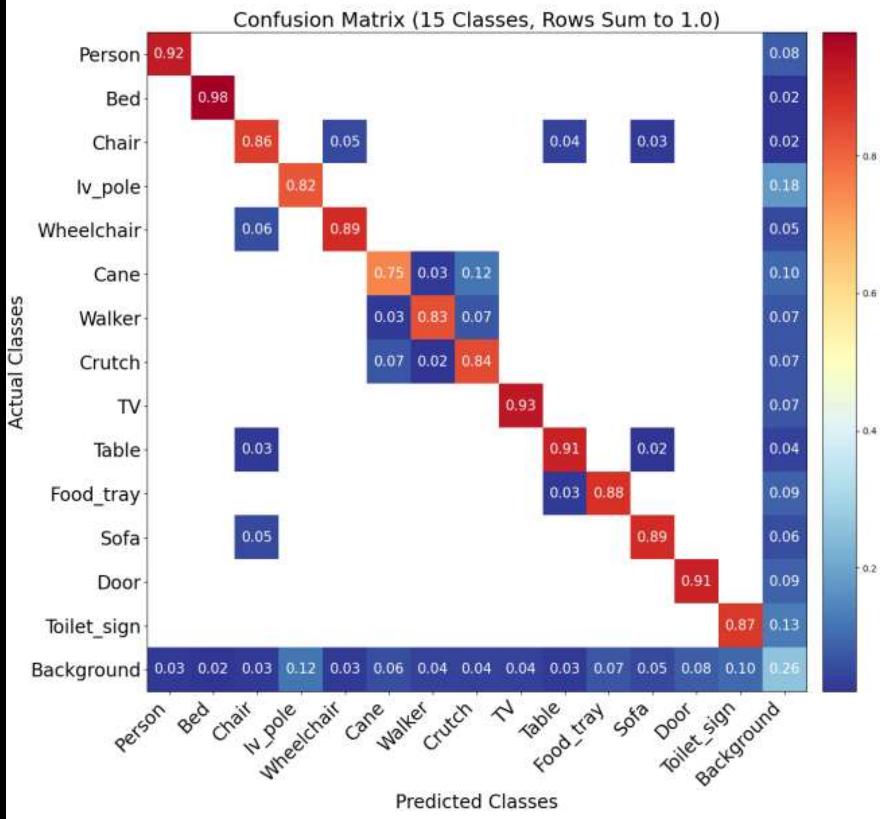
Step 7: Automated Report Generation

- Develops a frame-level reporting system to capture the temporal dynamics of patient–object interactions.
- Each frame records:
  - Timestamp (when the frame is captured)
  - Posture (lying, sitting, standing)
  - Interaction state (e.g., bed, chair, wheelchair)
- Detects transitions in posture or interaction to segment videos into distinct activity episodes.
- Aggregates episodes into clinician-friendly summaries that illustrate mobility patterns and object engagement.
- Enables clinicians to track ambulation, detect inactivity, and evaluate adherence to recovery protocols.



Experimental Results

- Detection and Tracking Performance



- Posture Classifier Performance

Model Performance on our dataset

Classifiers	Accuracy	Precision	Recall	F1 Score
Random Forest	**99**	**93**	**91**	**98**
SVM	88	89	87	88
K-NN	83	8	82	83

Model Performance on POLAR dataset

Posture	Frame #	Accuracy
Sitting	5,101	91
Standing	4,656	89
Lying Down	3,581	88

- Visibility Check and Interaction Detection

Scenario	Posture	# of Test Frames	# of 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

Conclusion

- Developes a proof-of-concept platform for automated patient behavior monitoring using computer vision and AI.
- Combines multi-camera fusion, object detection, posture recognition, and 3D triangulation for accurate interaction analysis.
- Uses regular cameras and custom hospital datasets for a cost-effective, scalable, and automated solution.
- Generates clinically interpretable summaries to support decision-making and monitor recovery trends.

Acknowledgements

The authors extend their gratitude to Dennis H. Sullivan, MD (Department of Geriatrics, University of Arkansas for Medical Sciences) for his clinical guidance.