Video-Based Human-Object Interaction Analysis for Healthcare

M. Habibi[1], Z. Delaram[1], M. Nourani[1] and D. H Sullivan[2]

1 Predictive Analytics & Technologies Laboratory, The University of Texas at Dallas

2 Geriatric Research, Education and Clinical Center, Central Arkansas Veterans Healthcare System

IEEE BHI2024 IEEE-EMBS INTERNATIONAL CONFERENCE ON BIOMEDICAL AND HEALTH INFORMATICS November 10-13, 2024 | Houston, TX

Poster No. 80

INTRODUCTION

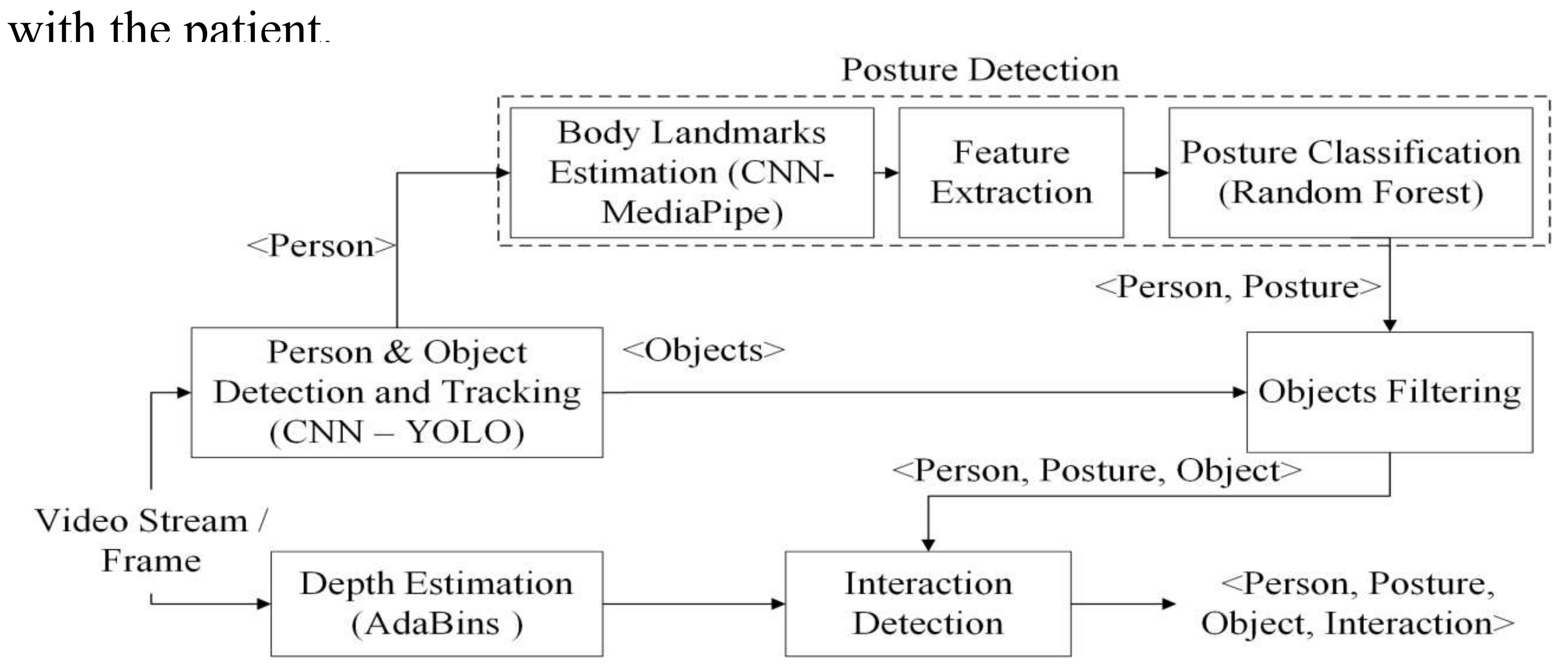
- Video-based platforms can be used to monitor patients in real-time or offline without using wearable devices enhancing patient comfort and care.
- We propose a systematic methodology for analyzing patient-object interactions using person/object identification, posture detection, and Intersection over Union (IoU), central distance, and depth estimation.
- Our methodology achieves a higher accuracy and a lower computational demand by customizing YOLO models (for person and object detection) and Random Forest (for posture classification) in hospital environments.

METHODOLOGY

1. **Person & Object Detection and Tracking:** We utilized YOLO algorithm for efficient real-time detection of patients and objects of interest in the hospital environment.
2. **Posture Classification:** We used Mediapipe to extract body landmarks, followed by features ranking and selection. These features were then used in a Random Forest model to classify postures.
3. **Object Filtering:** We employed object filtering to determine which objects are meaningful in interacting with the patient as related to their postures as shown in the following table:

Posture	Object$_{Sit}$	Posture	Object$_{Stand}$	Posture	Object$_{Lying}$
Sit	Bed	Stand	Cane	Lying Down	Bed
	Chair		Walker		Sofa
	Sofa		Crutch		Table
	Table		IV_Pole		
	Wheelchair				

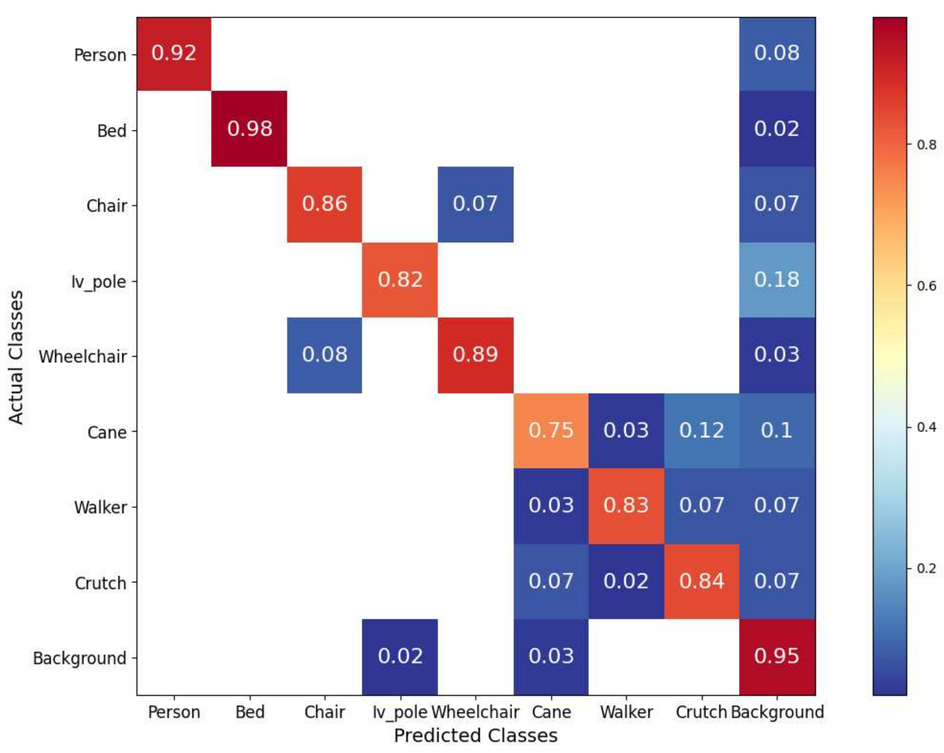
4. **Spatial Relationship Quantification:** We calculated IoU and central distance to quantify spatial relationships between detected objects and the patient.
5. **Depth Estimation:** We estimated the depth of objects within the frame to ensure that a detected object in the scene is truly relevant to the interaction with the patient.



RESULTS

1. **Person & Object Detection and Tracking:** We constructed a customized dataset that includes common objects (relevant to patient's status and behavior) typically found in a hospital environment shown in Table. Figure shows a confusion matrix for selected objects after YOLO is trained.

Object Name	Annotation #	Object Name	Annotation #
TV	2393	IV_Pole	1165
Bed	1780	Person	1301
Chair	1812	Sofa	2662
Table	456	Wheelchair	3534
Cane	2309	Door	801
Walker	1296	Crutch	1591
Food Tray	1542	Toilet Door sign	1050



2. Posture Classification: We created a dataset that includes three main postures: sitting (12657 frames), standing(10597 frames), and lying down (17235 frames). These frames were labeled and used for model training and evaluation (5-fold cross-validation; 80% for training and 20% for test). We have combined Mediapipe for body landmark estimation with feature extraction involving eleven specific landmarks and the calculation of 33 features. We compared the performance of four different classifiers on the same dataset.

RESULTS

Classifier	Accuracy	Precision	Recall	F1-Score
Random Forest	**0.99**	**0.93**	**0.91**	**0.98**
Logistic Regression	0.85	0.86	0.84	0.85
SVM	0.88	0.89	0.87	0.88
K-NN (k=5)	0.83	0.84	0.82	0.83

3. Spatial Relationship Quantification: We calculated parameters to discover the interaction between patient, posture, and objects.

a) **IoU** quantifies the overlap between two bounding boxes. It essentially measures how much two objects coincide in an image by dividing the area of their intersection by the area of their combined union.

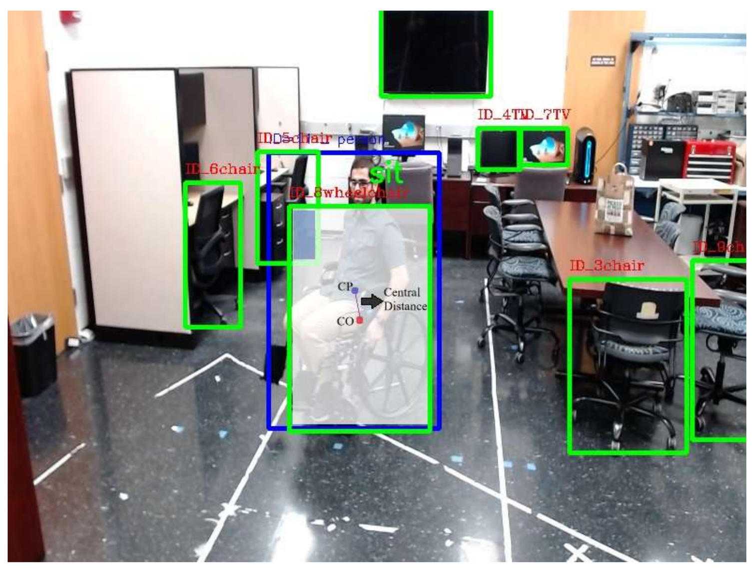
$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Object\ +\ Area\ of\ Patient\ -\ Area\ of\ Intersection}$$

b) **Central Distance:** This is a distance between the bounding boxes of detected objects and a patient and refers to the spatial proximity or relative position between these entities within an image. Central distance helps determine which objects are within close proximity to the patient.

$$Center_{Bounding\ Box}(X,Y) = \left(\frac{w-x}{2}, \frac{h-y}{2}\right)$$

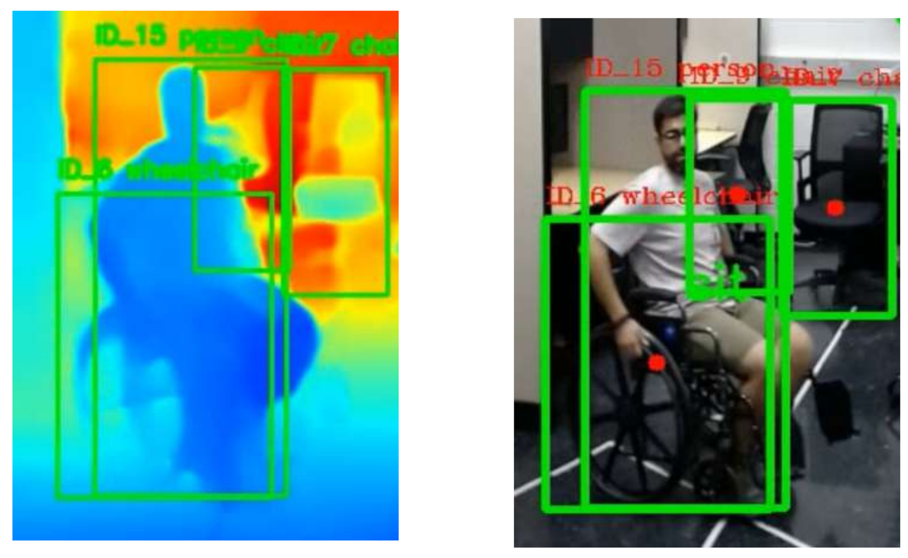
$$Distance = \sqrt{(X_{CP} - X_{CO})^2 + (Y_{CP} - Y_{CO})^2}$$

- In the example (figure) below, the white-shaded area is IoU between patient and wheelchair. CP (blue dot) is center of patient and CO (red dot) is center of object (wheelchair), based on this two points, the central distance calculated.



- The central distance metric encounters limitations in a 2D context.
  - Objects of varying sizes and distances from the camera can appear as similarly sized in a 2D projection leading to identical central distance measurements.
  - IoU metric is insufficient since when objects are equidistant from the patient, errors can arise in determining with which object the patient is interacting.
- To address these challenges, we introduce a third parameter: depth.

- **AdaBins for Depth Estimation**: Monocular depth estimation is the process of predicting depth information from a single image using computational techniques without the need for stereo or multi-view input as shown here:



- To evaluate our system, various patient-object interactions are recorded and analyzed. The average accuracy of recognizing objects, postures and interactions are tabulated below:

Posture	Object$_{Set}$	Frame #	Correct #	Accuracy(%)
Sitting	Object$_{Sit}$	2147	1783	85
Standing	Object$_{Stand}$	1870	1851	94
Lying Down	Object$_{Lying}$	1940	1938	89

CONCLUSION

In hospital settings with limited resources, our patient monitoring platform can be effective in automatically capturing relevant patient-object interactions. Our system can generate a comprehensive report to assist caregivers with monitoring and improve patient's recovery and their clinical outcomes.

UT DALLAS The University of Texas at Dallas

Contact Information : mostafa.habibidehsheikhi@utdallas.edu

Predictive Analytics & Technologies Lab, The University of Texas at Dallas, Richardson, TX 75080, USA