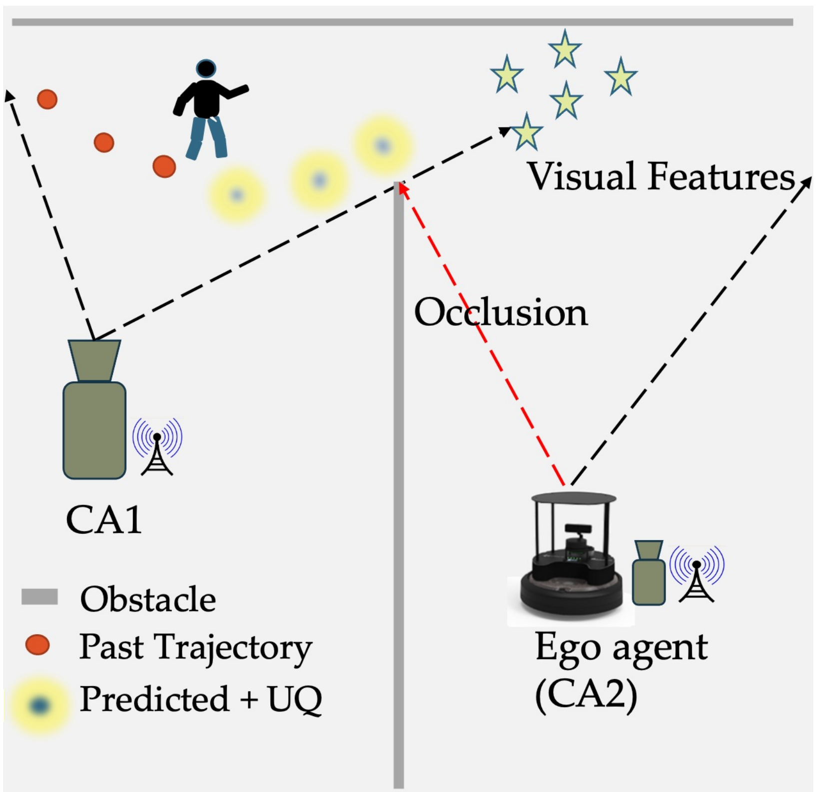
Cooperative Probabilistic Trajectory Forecasting under occlusion

Anshul Nayak
Azim Eskandarian

Motivation



- Estimation and prediction of dynamic occluded objects is essential for planning.
- GPS denied environment such as indoor navigation.
- Communication of rich sensor information computationally expensive.

Design an end-to-end network that estimates relative pose and forecasts future trajectory of occluded object.

Methods : Pose Recovery

❖ Pose Recovery between two cameras has been used to obtain rigid body transformation of occluded pedestrian.

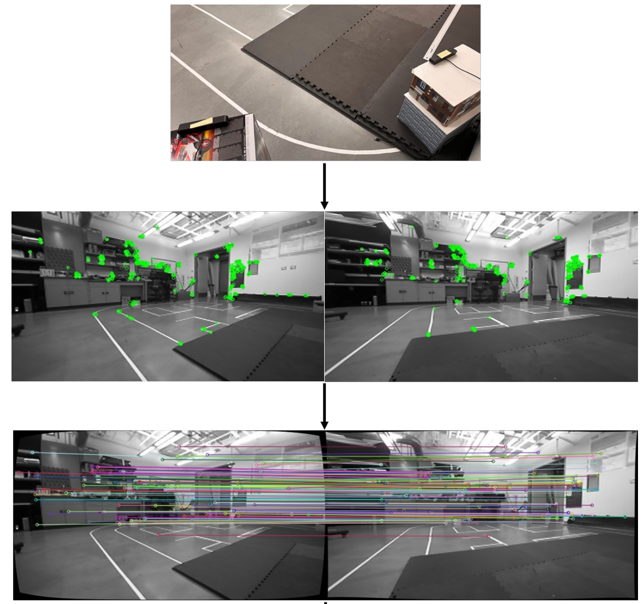
Step 1: Relative Orientation: Cameras take perspective images from two orientation.

Step 2: Feature detection and description:
- Difference of Gaussian
- SIFT, ORB, SURF

Step 3: Feature Matching:
- FLANN or KNN matcher
- RANSAC to eliminate outliers

Step 4: Relative Pose Recovery:
- Fundamental matrix : x’Fx = 0
- Camera intrinsic: K1, K2
- Essential matrix decomposition: SVD(E)

Visual Odometry can be pivotal for cheap pose recovery between two moving frames if the initial relative pose is known

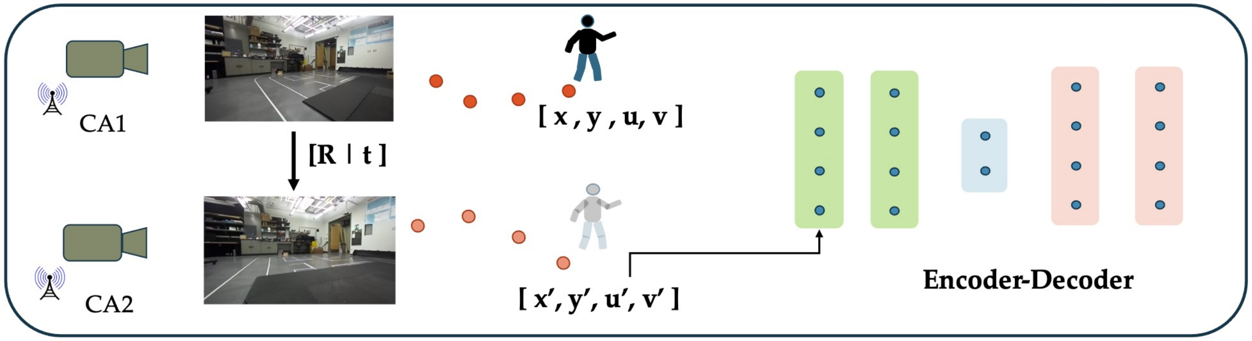


[R | t]

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

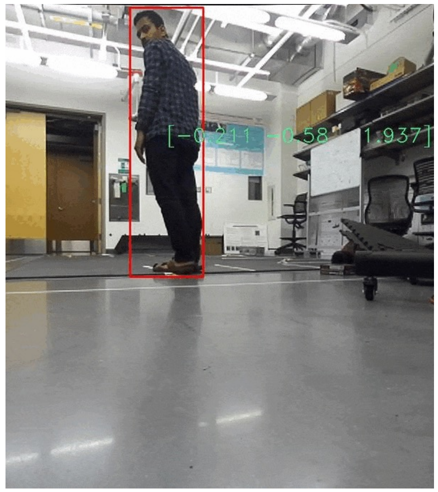
How accurate is pose recovery?

Methods : Uncertainty –aware Prediction



- ❖ Pose recovery followed by rigid body transformation to estimate occluded object's transformed states.
- ❖ Uncertainty-aware prediction using Bayesian inference of neural network models:
  - Monte Carlo Dropout
  - Deep Ensembles

Experiments: Pose Recovery



Ground Truth	[1.31, -1.767, 19.12]
Estimate	[1.44, -3.018, 21.878]
Feature points	1290
Good Matches	128

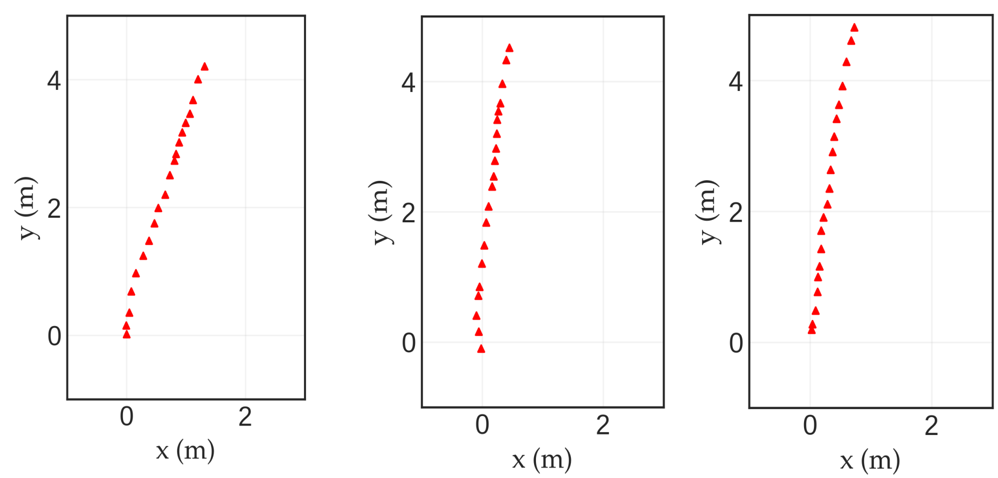
How reliably can pose recovery be used to transform pedestrian coordinates from camera 1 to camera 2 frame?

Rigid body Transformation:

- $[x', y', z'] = R^T([x, y, z] - t)$

Average Displacement error:

- $\frac{1}{T}\Sigma_{t=t0}^{tf} \|\hat{Y} - Y\|$

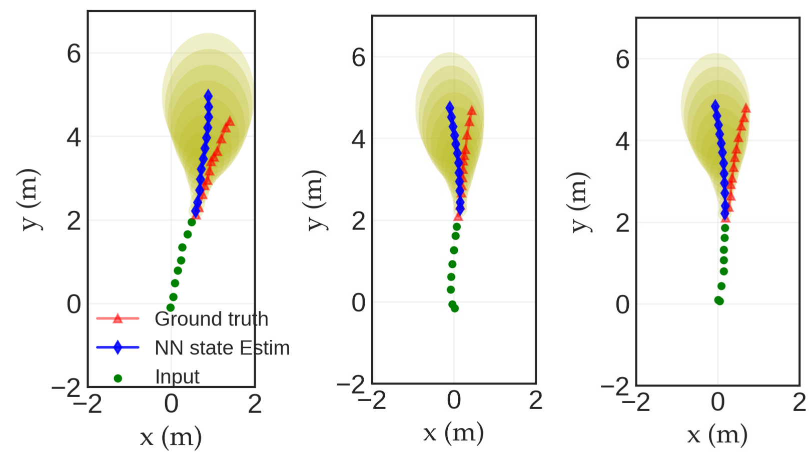


(a) Camera 1 (b) Camera 1 Transf. (c) Camera 2

Results

	ETH	HOTEL	ZARA1	ZARA2	UNIV	AVERAGE
S-LSTM [12]	1.09/2.35	0.79/1.76	0.47/1.00	0.56/1.17	0.67/1.40	0.72/1.54
SGAN [32]	0.87/1.62	0.67/1.37	0.35/0.68	0.42/0.84	0.76/1.52	0.61/1.21
Sophie [33]	0.70/1.43	0.76/1.67	0.30/0.63	0.38/0.78	0.54/1.24	0.54/1.15
Social-BiGAT [34]	0.69/1.29	0.49/1.01	**0.30/0.62**	**0.36/0.75**	0.55/1.32	0.48/1.00
LSTM	0.54/0.94	0.33/0.46	0.51/0.96	0.53/0.96	0.75/0.93	0.53/0.85
1D CNN	0.71/0.90	0.71/1.04	0.75/1.02	0.86/1.16	0.95/1.24	0.79/1.07
CNN-LSTM	0.68/1.11	0.98/1.29	0.73/0.99	0.95/1.27	0.87/1.11	0.84/1.15
LSTM + MC	0.55/0.94	**0.32/0.45**	0.51/0.96	0.54/0.96	0.59/0.84	**0.50/0.83**
1D CNN + MC	0.69/0.84	0.58/0.79	0.73/0.99	0.85/1.15	0.71/0.85	0.71/0.92
CNN-LSTM + MC	**0.48/0.82**	**0.3/0.48**	0.50/0.83	0.77/1.12	**0.53/0.86**	**0.51/0.82**

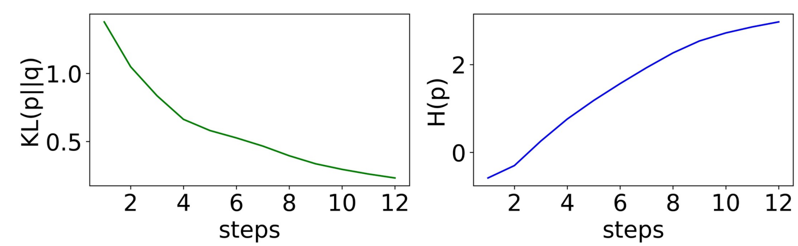
* Our model CNN-LSTM with dropout showed improvement in ADE and FDE for mean path on pedestrian dataset.



(a) Cam 1 (b) Cam 1 Transf. (c) Cam 2

Experimental results:

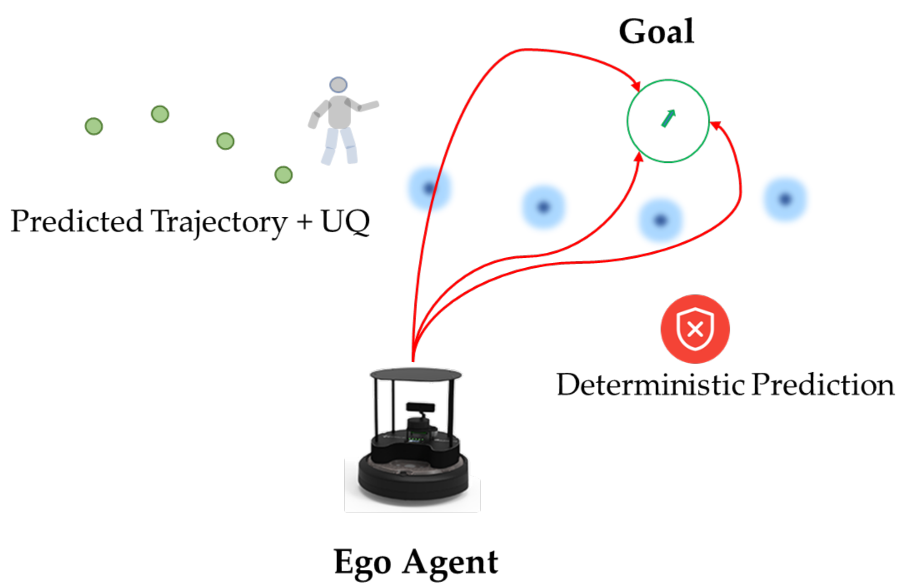
- Probabilistic predicted states for Camera 1 transformed (Fig. b) matches camera 2 (Fig. c).
- Ground truth lies within the 2σ predicted distribution.



KL div. between predicted distribution of transformed and ground truth is minimum.

Conclusion

- End-to-end cooperative trajectory prediction with safety guarantees under occlusion.
- Extended to dynamic agents using visual odometry if initial pose is known.



How uncertain human-robot interaction can be modeled to handle occlusion-aware planning?

Prediction uncertainty-aware robust planning for dynamic objects under occlusion.