Adaptable Human Intention and Trajectory Prediction for Human-Robot Collaboration

Abulikemu Abuduweili — Peking University
Siyan Li — Georgia Institute of Technology
Changliu Liu — Carnegie Mellon University

Introduction

- Must generate high-fidelity trajectory and intention predictions to ensure safe and efficient HRC.
- Previous works
  - Separated intention and trajectory prediction
  - Rarely used adaptation
- We propose a framework that enables:
  (1) Fast HRC integration
  (2) **Simultaneous** trajectory and intention prediction
  (3) Online model **adaptation** to address heterogeneity

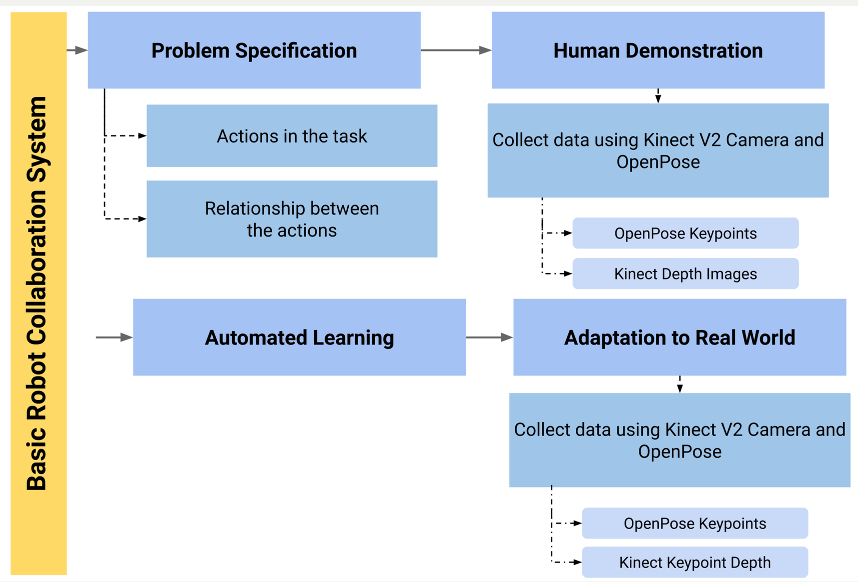

Fig 1. The structure of our pipeline

Framework Structure

Problem Definition

- Simple card-making task containing 11 actions, we defined action 12 to be the retrieval action in every "taking" action for convenience in intention labeling
- Can be represented using an and-or-graph
- This step must be performed manually at the current stage

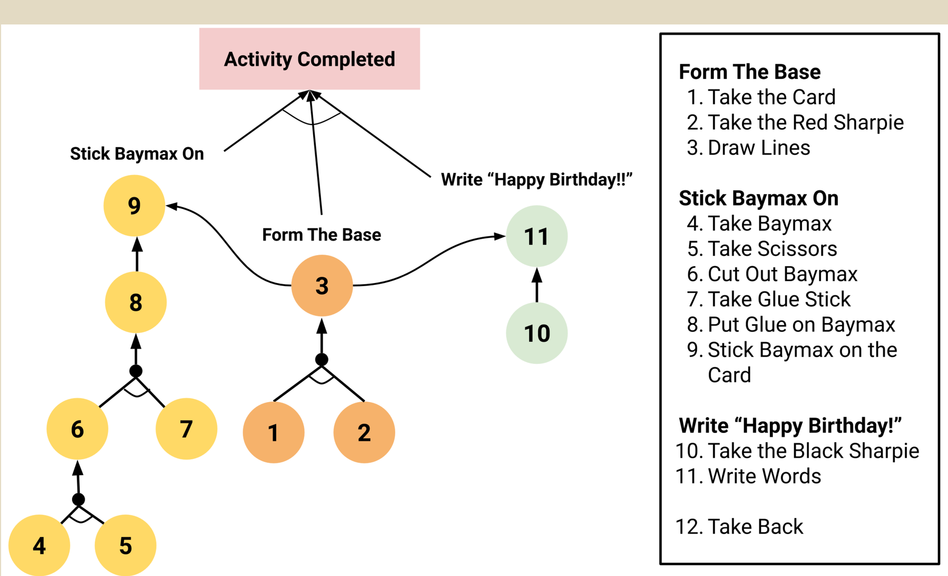

Fig 2. The and-or graph representation of our task. Each arrow pointing from node A to B indicates that A must occur before B.

Human Demonstration

- All of the atomic actions must be repeated by a human so that the neural network in the pipeline can then learn the different trajectories and their corresponding intentions
- To collect data, we utilized a Kinect V2 camera along with OpenPose (Cao et al. 2018), a framework that can generate pixel positions of joint keypoints

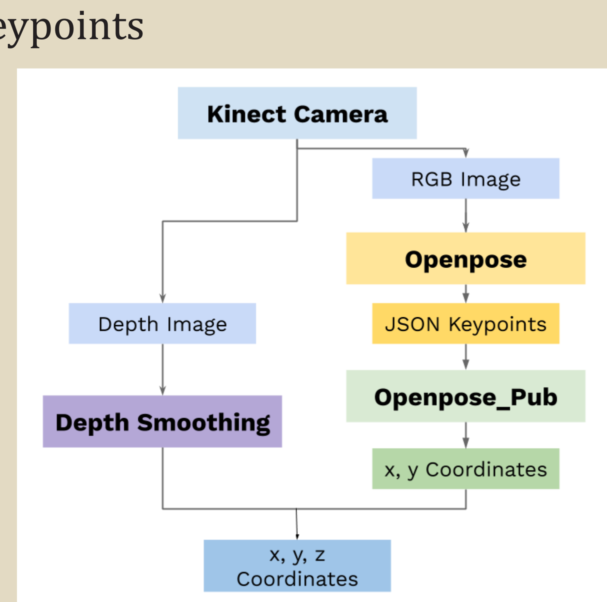

Fig 3. Our process for data collection and pre-processsing

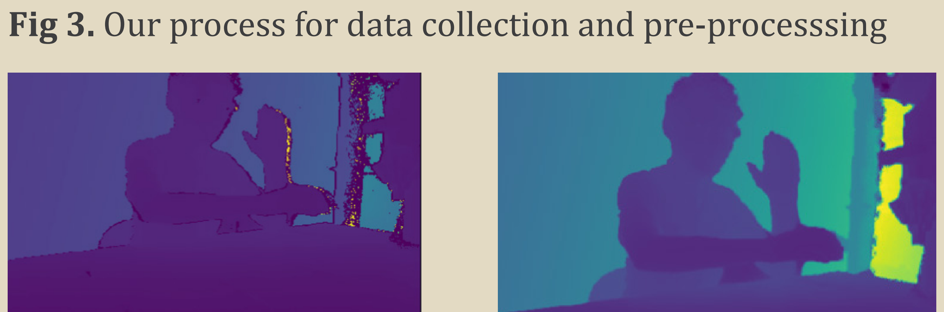

Fig 4. Noisy original Kinect image (top) and smoothened image using the averaging method (bottom)

Framework Structure (Continued)

Learning Trajectories and Intentions

We propose a multitask model, with a regression section for trajectory prediction and a classification section for intention prediction. The training loss is a weighted sum of MSE for regression and cross-entropy for classification.

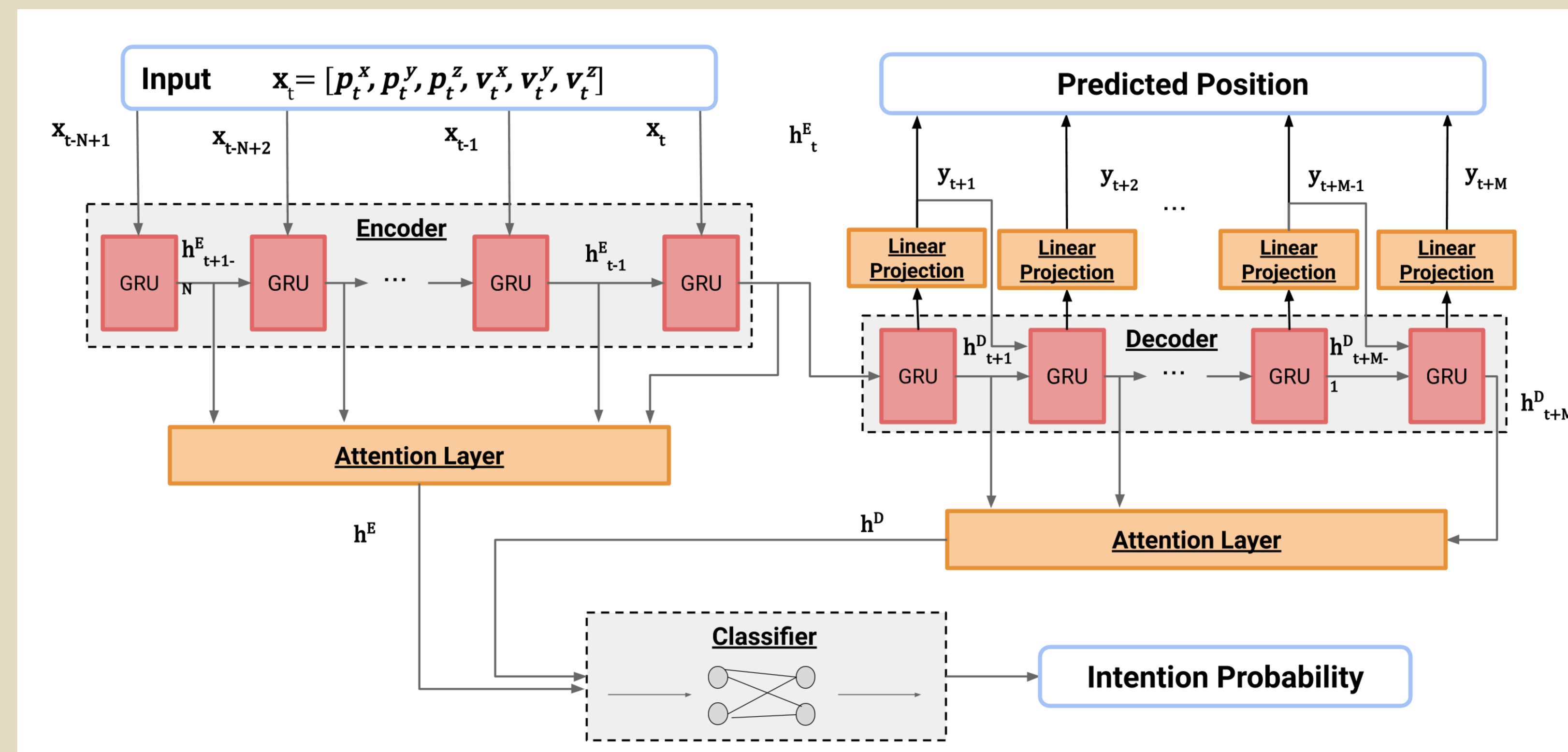

Fig 5. The neural network structure of our multi-task model.

Adapting to Real-World Tasks

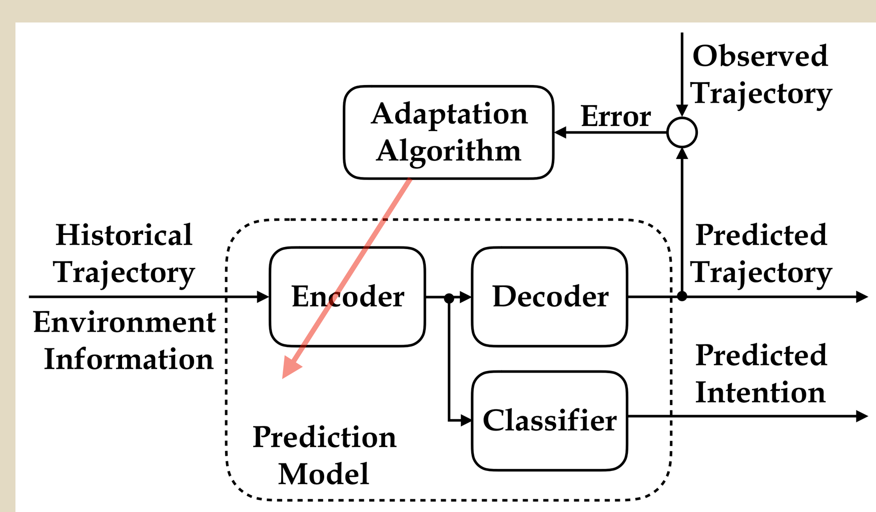

Fig 6. The general framework of the online adaptation system. At every time step, the ground truth is received and the error calculated. The encoder is then updated using the error.

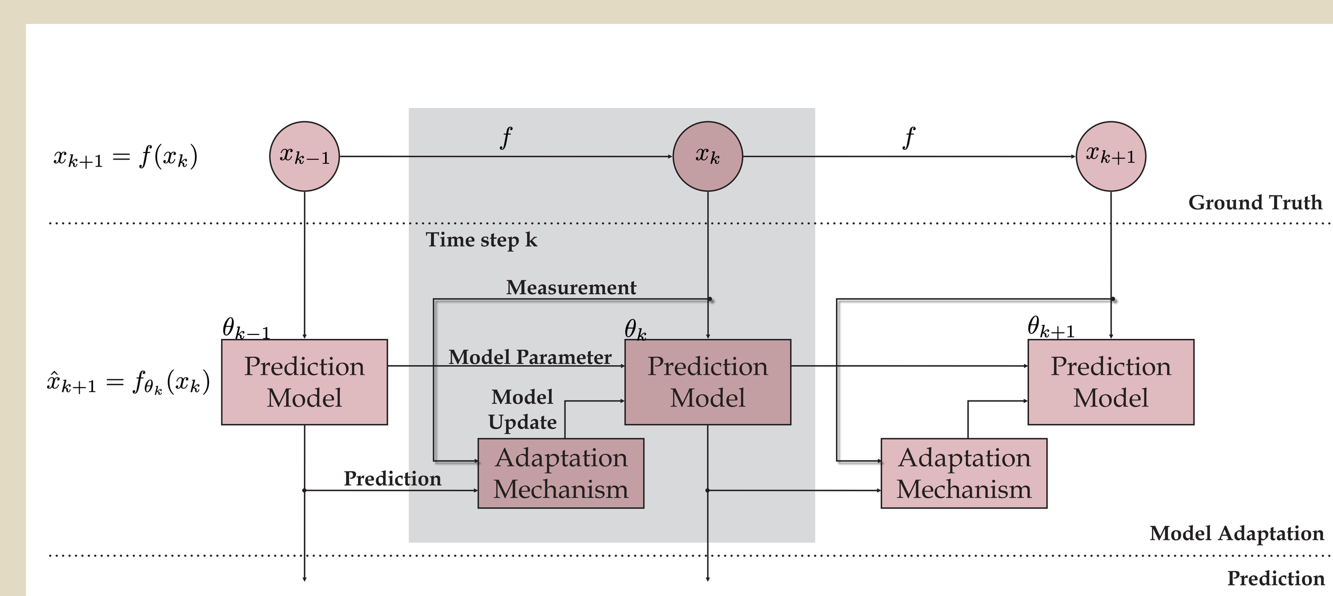

Fig 7. The detailed procedure for online adaptation.

Experiments

- Collected data for the 12 pre-defined actions, Actor A performed each action 50 times (80% offline training, 20% offline validation), Actor B 10 times (100% online testing).
- Implemented 1-step, 2-step, and 5-step adaptation
- Compared performances of the single-task and the multi-task models

	Accuracy	MSE (cm^2)
Without Adaptation	0.930	5.508
1-step Adaptation	0.938	4.919
2-step Adaptation	0.938	4.488
5-step Adaptation	**0.946**	**3.964**

Table 1. The results from using (1) no adaptation, (2) 1-step adaptation, (3) 2-step adaptation, and (4) 5-step adaptation. We also learned that there is a time-accuracy trade-off; increasing the number of adaptation steps boosts accuracy but also adds to running time.

	Accuracy	MSE (cm^2)
(Single) Intention Prediction	0.899	-
(Single) Trajectory Prediction	-	5.909
(Multi) Our Model	**0.930**	**5.508**

Table 2. The results from comparing performances of the single-task and the multi-task models. Our multi-task model out-performs the single-task models.

Conclusion and Future Work

- Our simultaneous trajectory and intention prediction model outperforms its single-task counterparts
- Online adaptation improves online prediction accuracy by 28%
- The more adaptation steps it takes, the more accurate the predictions are, but prediction time also increases

- Future work
  - Integrate the prediction model into a robot collaboration system
  - Further experimentation
  - Enable automatic learning of task specifications
  - Enable full-arm prediction

References

- [Cao et al. 2018] Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.-E.; and Sheikh, Y. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In arXiv preprint arXiv:1812.08008.