Information about the Human Morning Routine Dataset

Michael Karg   KARGM@IN.TUM.DE

Institute for Advanced Study, Technische Universität München, Lichtenbergstrasse 2a, D-85748 Garching, Germany, http://hcai.in.tum.de

Alexandra Kirsch   ALEXANDRA.KIRSCH@UNI-TUEBINGEN.DE

Department of Computer Science, University of Tübingen, Sand 14, D-72076 Tübingen, Germany, http://www.hci.uni-tuebingen.de

Abstract

This document provides an overview about the Morning Routine Dataset recorded at TUM München by members of the Human Centered Artificial Intelligence Group. The dataset was designed to capture different types of information about a typical human morning routine and make those information available to the scientific community to test and evaluate different approaches in the domain of domestic service robots.

1. Introduction

The dataset described in this document has been created with the aim of capturing a typical morning routine of a single person in an apartment. In contrast to other datasets like the TUM kitchen dataset (Tenorth, Bandouch, and Beetz, 2009) which consists of several persons performing the same task, the focus for this dataset was to capture several activities of daily living performed by the same person at different days during an extended period of time. For this dataset, the same scenario as in the TUM kitchen dataset was used, but we specifically focused on a low cost sensor setting using only two Kinect sensors which cost below 250 USD in contrast to relying on complex motion tracking systems that are very costly. We decided to go for such a sensor setting since we think that a commercial service robot that will be deployed on the consumer market will use rather low cost components in order to be affordable for a wide range of people. Thus our dataset reflects a realistic impression about how such a robot would perceive routine activities of its user when it is deployed in a human apartment.

2. Description of the Dataset

2.1 Locations, Times and durations

We investigated a period of 14 workdays of a voluntary male test person that did not know about our system and was living in a rental apartment by himself. We asked the test person to note the activities that he performed before going to work as well as the locations at which he stands still

Figure 1. One day of the notes the participant took showing the order of locations he visited during his morning routine. The left picture shows a schematic top view of the apartment while at the right you can see the annotated activities that the participant performed at the corresponding locations.

while performing those activities over 3 weeks. The exemplary notes of one day are shown in Figure 1.

> The notes that the participant took can be found as a jpeg file for each day in the folder:
> **./apartment_locations**

The participant was furthermore asked to write down the approximate times and durations at which he performed each activity, keeping the naming of re-occurring activities consistent.

> The notes about the times and durations of each activity can be found as a csv-file at:
> **./activities_durations.csv**

2.2 Motion tracking data

To obtain real sensor data of human motion tracking as well as detected objects, we asked our participant to reenact his morning routine (according to his notes) in a sensor equipped kitchen environment. Therefore, we equipped an experimental kitchen-environment with two Kinect sensors. We use one of the Kinects to perform motion tracking of the human using the OpenNI tracker [1] and the second Kinect for object detections based on visual markers using the ar_pose [2] wrapper for ARToolKit. An overview of the kitchen scenario as well as a visualization of the corresponding semantic map is shown in Figure 2.

---

1. http://ros.org/wiki/openni_tracker
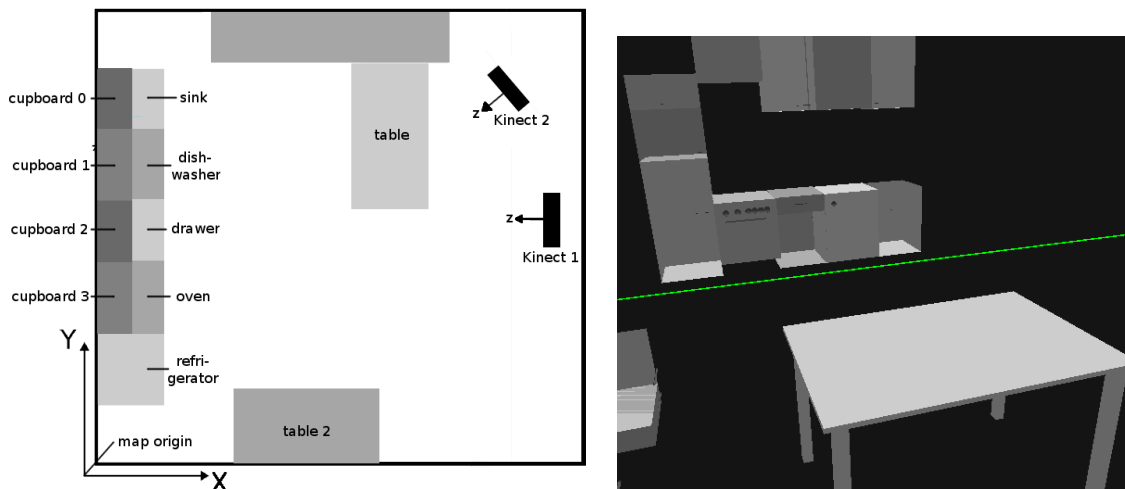2. http://www.ros.org/wiki/ar_pose

Figure 2. Left: Structural overview of the kitchen scenario (not true to scale!). Right: Visualization of the semantic map in the KnowRob framework.

The exact locations of the Kinects are posted as coordinate transforms "camera_1_link" (Kinect 1, used for motion tracking) and "camera_link" (Kinect 2, used for object detection) via ROS TF frames [3]. The approximate positions of the Kinects are:

- Kinect 1: (<x,y,z><x,y,z,w>): <4,3.02,1.11><0,0,1,0>
- Kinect 2: (<x,y,z><x,y,z,w>): <2.986, 4.414,0.870><-0.271,-0.653,0.653,0.271>.

Where the first three values are x,y and z coordinates in meters according to the origin of the map and the latter four values represent a quaternion according to the definitions of ROS Quaternions.

The exact locations of the furniture objects can be obtained by querying the semantic map "ias_semantic_map.owl" from KnowRob. The "ias_semantic_map.owl" is part of the standard installation of KnowRob. Please consult the KnowRob Website[4] for installation instructions. You can use the Semantic Map Editor[5] to visualize and edit the semantic map or query the KnowRob knowledge base directly or via ROS. If you are not into using KnowRob, you can also look up the exact locations of the furniture objects in the simulated scenario of the MORSE simulator. Furthermore the dataset contains a 2D gridmap of the environment that can be used with the ROS map server [6] for visualization.

The simulated scenario can be found as a Blender file at: **./garching_kitchen.blend**.

A 2d gridmap of the scenario can be found at: **./2d_gridmap**.

3. http://wiki.ros.org/tf

4. http://www.knowrob.org/

5. http://www.knowrob.org/doc/semantic_map_editor
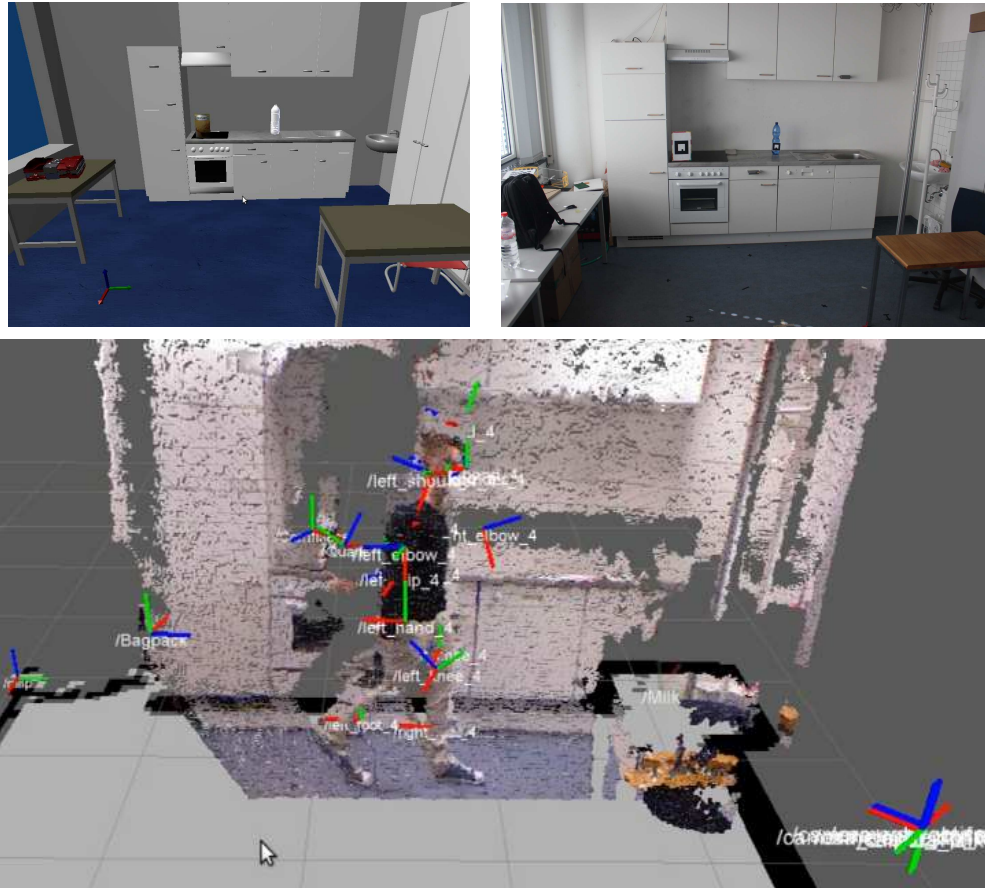
6. http://wiki.ros.org/map_server

Figure 3. The simulated environment in the MORSE simulator is shown on the upper left picture, while the upper right picture shows the real experimental kitchen environment. In the lower picture the sensor input of the real scenario is shown. The motion tracking returns coordinate transformations for each joint of the human while the visual marker detection returns a coordinate transform for each detected object. The map and the Pointcloud data are only shown for visualization.

Since the size of the visual markers was of critical importance for the detection, we could only use markers on objects that were of a certain size. We equipped the following objects with markers: Cornflakes, Milk, Backpack, Curd cheese, Bottle. The ar_pose object detection returns the approximate pose of all detected objects in reference to the Kinect as coordinate transformations in the ROS TF framework [7].

We also set up a simulated environment of the same kitchen using the MORSE simulator[8] (Echeverria et al., 2011). This simulator provides us with a human avatar that can be controlled by a human and used to perform pick- and place tasks in the simulated environment like in 3D computer games. Figure 3 depicts the experimental kitchen in its simulated version as well as the real kitchen

---

7. http://www.ros.org/wiki/tf
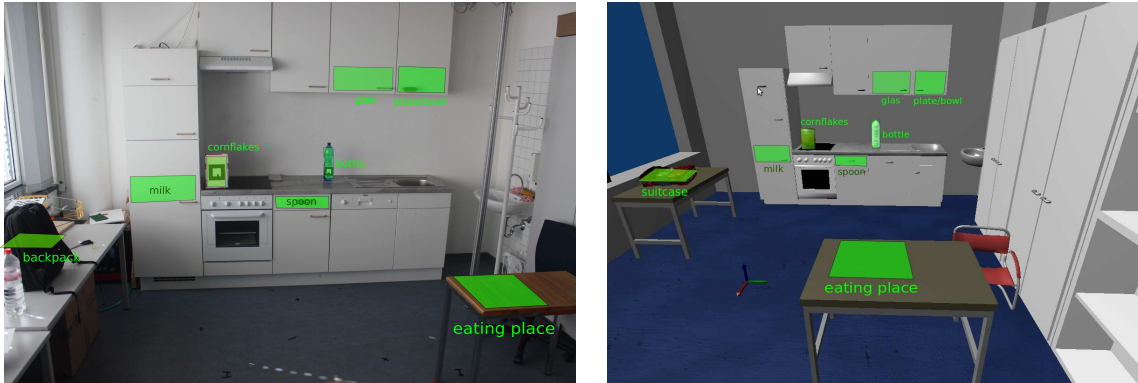
8. http://morse.openrobots.org

Figure 4. Approximate storage locations of objects in the real (left image) and simulated (right image) scenarios.
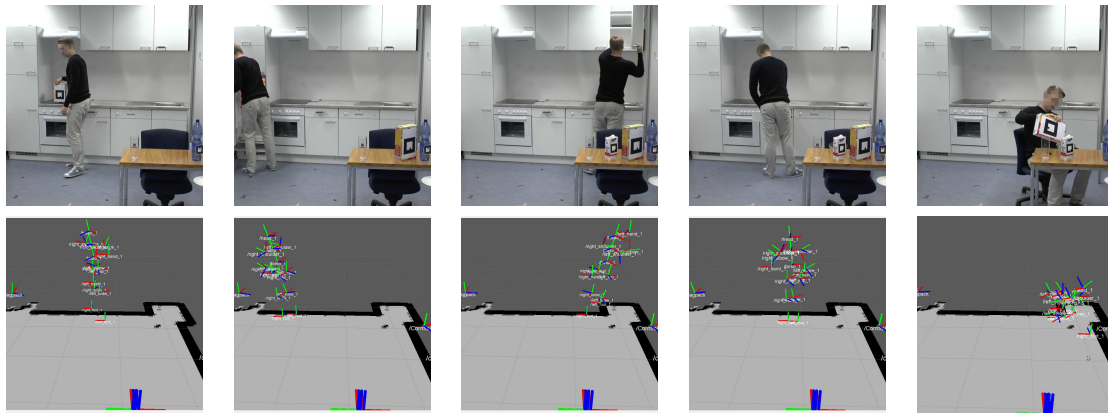

Figure 5. A typical table-setting activity of the participant performed as part of his morning routine in the experimental kitchen environment. The upper images show video data of different parts of the preparation for having cereals as breakfast. The lower pictures show a visualization of the motion tracking data and object detections that were recorded using two Kinect sensors. For visualization, a 2D gridmap of the kitchen environment is shown as well. In the rightmost picture of the lower images, parts of the body of the participant were occluded by the table, resulting in inaccurate and even wrong tracking for many joints of the tracked human model.

environment. Furthermore, the picture shows a visualization of the motion tracking data and object detections that are obtained by the two Kinect sensors. The objects were placed at the same typical locations in both scenarios (butter in the refrigerator, cutlery in a drawer etc.). The approximate locations of the objects are shown in figure 4 for the simulated as well as the real scenario.

The participant was then asked to perform his typical morning routine in simulation as well as in the real kitchen in the same way as he did in the 14 days that he took notes of. A visualization of video data of parts of one activity along with the corresponding motion tracking data is shown in Figure 5.

Although the motion tracker worked quite well while the test person was walking, it sometimes lost track of the person while sitting at the table or getting objects from the refrigerator. Especially when parts of the body are occluded the motion tracking data is very inaccurate, which results in a "jumping" of some joints of the tracked human. The performance of the marker-based object detection was varying in its quality quite a lot. While in many cases we had correct object detections, it often happened that objects were detected at wrong places or some detections of objects During task performance in the real scenario, we recorded the following information on the corresponding ROS topic into ROS bag-files [9]:

- **/tf** - Full tf transformations of human motion tracking data and object detections in reference to the map origin
- **/object_detections** - Detected objects output as a string message which consists of CommonLISP Code in reference to Kinect 2. The list is generated for each object in the following form: (object.name object.type object.position.x object.position.y object.position.z object.orientation.x object.orientation.y object.orientation.z object.orientation.w), where the orientation is generated as a Quaternion according to ROS definitions.

The bag-files of the real scenario can be found in ./real_scenario/bag.

In the simulated scenario, the test person used the human avatar of the MORSE simulator to open and close drawers, and move objects around. We tracked the position of all objects using a simulated object tracker and the human using a simulated person tracker. Instead of eating or drinking, the person was asked to take each object (glass, spoon) that is part if the corresponding task with the simulated human avatar for a short time and wait for a while (1 minute for drinking, 3 minutes for eating).

For the simulated scenario, we recorded the following information in the corresponding topics:

- **/Human/Pose** - The position of the human in reference to the map origin as ROS nav_msgs/Odometry [10]
- **/b21/object_tracker** - Detected objects output as a string message which consists of CommonLISP Code in reference to Kinect 2. The list is generated for each object in the following form: (object.name object.type object.position.x object.position.y object.position.z object.orientation.x object.orientation.y object.orientation.z object.orientation.w), where the orientation is generated as a Quaternion according to ROS definitions.

The bag-files of the simulated scenario can be found in ./simulated_scenario/bag

2.3 Ground Truth Data

Each experiment has been manually labelled to provide a ground for the evaluation of experiments. The ground-truth files are in the CSV-file format and provide information about which location the

---

9. http://wiki.ros.org/rosbag

person visited for how long and which task he currently executed. The ground truth CSV files use the following format:

`id,start_time,end_time,location,plan`

- **id**: unique ID of the location
- **start_time:** Time when the person entered the location
- **end_time:** Time when the person left the location
- **location:** Name of the location visited plus plan abbreviation name as prefix
- **plan:** Name of the plan the human is currently executing

The "location" tag use an abbreviation of the corresponding plan as a prefix. The abbreviations are according to the following scheme:

- **DW:** Drink Water
- **STC:** Set table for cereals
- **STQ:** Set table for curd cheese
- **STB:** Set table for bread
- **CTC:** Clean table after having cereals
- **CTQ:** Clean table after having curd cheese
- **CTB:** Clean table after having bread

> The ground truth files of the real scenario can be found in: **./real_scenario/ground_truth**.
> The ground truth files of the simulated scenario can be found in:
> **./simulated_scenario/ground_truth**.

3. Conclusion

This document was intended as an extended Readme to provide researchers with an overview about the recorded data and provide data for the research community. We hope that this dataset will free some researchers from the tedious task of recording datasets about human habitual activities. In case, you have further questions, don't hesitate to contact the authors using the mail addresses provided in the document header.

References

Echeverria, G.; Lassabe, N.; Degroote, A.; and Lemaignan, S. 2011. Modular openrobots simulation engine: Morse. In *Proceedings of the IEEE ICRA*.

Tenorth, M.; Bandouch, J.; and Beetz, M. 2009. The TUM Kitchen Data Set of Everyday Manipulation Activities for Motion Tracking and Action Recognition. In *IEEE Int. Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS). In conjunction with ICCV2009*.