Accelerating the teaching of industrial robots by re-using semantic knowledge from various domains

Karinne Ramirez-Amaro and Gordon Cheng

Abstract—It is envisioned that the next generation of robots will work in heterogeneous production lines by efficiently interacting and collaborating with human co-workers. To enable a natural collaboration between robots and operators, robots should also understand and recognize the actions of the operators. For this purpose, an automated process to program industrial robots needs to be developed. In this paper we present a teaching by demonstration method based on semantic representations that enables a standard industrial robots to be flexible, modular and adaptable to different production requirements. The proposed semantic-based method is able to re-use the knowledge obtained from household demonstrations to accelerate the teaching of unknown tasks in industrial settings such as packing oranges. Furthermore, the proposed method is able to automatically understand the actions of the operator during their interaction. This new learning method enables non-experts operators to intuitively program robots new tasks.

I. INTRODUCTION

The successful automation of adaptable production processes demands flexible, usable and acceptable robotic solutions. Flexibility implicates that robotic systems have to be quickly deployable with short installation times, easy to move to different production sites or stations and to allow quick and easy adjustments to cope with current production demands. Usability and acceptability imply simple and intuitive programming (teaching) methods, enabling non-experts and untrained personnel to effortlessly reconfigure the robot system, thus, providing technology for a broad spectrum of users, regardless the age or background skills. Combining all these solutions will enable affordable and effective Human-Robot Collaborations since the deployment of this new generation of robots will produce minimal changes (disruptions) in the production line. These robots will be able to generate Human-Robot Collaborations just as if they were Human-Human Collaborations, considering that they will have ideally the same set of skills and requirements as a human co-worker in the context of a specific production process or operation.

An interesting method to extend the flexibility and capabilities of a robot is to integrate it in close interaction with human co-workers. The fusion of the high adaptability of the human and the accuracy of a robot system can facilitate the automation of industrial processes. In this case, safety in physical human–robot interaction [2] is a fundamental aspect on developing robot technologies. Especially for the new way of teaching robots sequences using programming by demonstration methods which requires physical interactions with the robot [1]. These new programming by demonstration method is needed to advance the current robot systems and it is the main focus of this paper which is being developed as part of the project Factory-in-a-Day1. This project aimed at improving the competitiveness of European manufacturing SMEs by optimizing the robot installation time and installation cost. For this, we have developed novel reasoning and knowledge-based methods [3] to allow a natural way to teach industrial robots new tasks by physically interacting with them. A semantic-based reasoning approach was proposed to integrate different input sensors [1], such as the joint encoders of a robot, skin sensors (tactile and proximity) and visual information from the cameras embedded in the robot (first view perspective) [4], [5]. Additionally, we developed a state-of-the-art knowledge-based representation to incrementally learn new representations from the demonstrations to accelerate the teaching of the unknown tasks. We show that our presented framework enables a standard industrial robot to be flexible, modular and adaptable to different production requirements.

II. RELATED WORK

In the context of industrial scenarios, programming robots in an easy and intuitive manner is an important requirement [6]. To fulfill this requirement, a new promising way for robot programming seems to be the Programming by

1http://www.factory-in-a-day.eu/
Demonstration (PbD) [7], [8], which allows the operator
to teach tasks to the robot in an easy and natural way,
thus requiring no experience in robot programming. PbD
methods can enable fast and flexible modifications on robot
behaviors to perform a wide variety of tasks [9]. For example,
[10] proposed a three-stage PbD method based on Dynamic
Motion Primitives (DMPs) to Kinesthetically teach industrial
robots, this method learned low-level profiles such as force
and pose trajectories. However, generalizing the learned
models to different domains is not straight-forward.

Allowing robots to recognize activities through different
sensors and re-using its previous experiences is a prominent
way to program robots. For this, a recognition method
needs to be proposed such that it is transferable toward different
domains such as household or industrial domains. One
key component for such generalization is the definition of
common representations. Ramirez-Amaro et al. [3] presented
a flexible system to extract symbolic representations of the
perceived scenario which adapts to different sensors, such as
cameras [11], multi-modal skin [5], robot joint data [4],
and virtual environments [12], [13], among other sources.
Thus, it has been demonstrated that the method is sensor
agnostic since it can adapt to the available sensors. Activity
recognition is a broad topic and includes the ability to extract
semantic representations, recognize new actions when there
is no pre-existing knowledge (on-line learning), and predict
ahead of time human behaviours. Aksoy et al [14] proposed
an action learning system based on the physical relationships
between objects during manipulation. Summers-Stay et al
[15] used a simpler detection and segmentation method
using a tree structure. Recent work, on one- and zero-shot
learning techniques are used, which seek to minimize both
the volume of and reliance on training datasets, with the
extreme case being systems capable of classifying previously
unseen actions without any prior training [16].

III. SEMANTIC REASONING LEARNER

In this paper, we summarize the proposed hierarchical ap-
proach to extract the meaning of kinesthetic demonstrations
by means of symbolic and semantic representations. This
means that the movements from the operator are tracked,
segmented and recognized on-line by robots while kine-
sthetically demonstrating a new process, e.g. packing oranges.
The lowest level of our hierarchical method finds the relevant
information from the demonstrations from multiple sensors.
This obtained information represents the input to the highest
level, which infers the demonstrated activities using the
automatically extracted semantic representations. The pre-
sented hierarchical approach works in two different spaces,
the Problem Space and the Execution Space as depicted in
Fig. 2. The Problem Space provides semantic descriptions
which represents robot-agnostic knowledge, therefore it can
be transferred to different domains. The Execution Space is
the specific information and routines that depends on the
current robot, then only this information needs to be changed
when using a different robot.

A. Workflow hierarchical structure

The following vocabulary has been used to recognize the
robot demonstrations at different levels of abstraction [5].
The highest level is the Process, which is defined as the
combination of sequential Tasks. Tasks are the combination
of ordered Activities. Activities are semantic descriptions of
Skills, and finally Skills (lowest level) represent the primitives
that robots need to execute, see Fig. 2.

For example, the Process “Pack good oranges into boxes”
is composed of two Tasks “Pick an orange” and “Place
orange in box”. The first task contains three activities namely
“Idle”, “Reach” and “Take”, while the second task is defined
by the activities “Put” and “Release”. Each of these Activities
is connected to a Skill. For example, “ Reach” is linked to
the “Reach Skill” primitive.

Process, Tasks and Activities are described in the Problem
Space, and they are considered robot agnostic descriptions.
They represent what the robot should perform, and not how
it should be done. On the other hand, the Skills are defined in
the Execution Space, and they explicitly define how the robot
should execute an Activity. They represent specific routines
or robot programs to execute a given Activity.

The main advantage of this hierarchical architecture is the
re-usability and generalization of the acquired knowledge.
Thus allowing the transference of knowledge generated in
the Problem Space to different domains, see Fig. 2.

B. Automatic recognition of human interactions

In order to automatically interpret the kinesthetic demon-
strations, our learning system transforms the continuous
signals obtained from the demonstrations to symbolic rep-
resentations [11]. For example, the motions (m) of
the robot’s end-effector (ef) are interpreted as either Move
or Not Move symbols. Where Move: the end-effector is moving,
i.e. \( \dot{x} > \varepsilon \) and Not Move: the end-effector stops its motion,
i.e. \( \dot{x} \to 0 \), where \( \dot{x} \) is the end-effector velocity and \( \varepsilon \) is a
heuristically defined threshold. In addition, the information
about the perceived environment is also transformed into
symbolic representations. For the demonstration scenario
described in Section IV-A, the robot TOMM can perceive
its environment through the following sensors: robot skin,
RGB-D camera, and joint sensors. From these sensors the
following abstract properties can be defined: a) ObjectActe-
dOn\(^2\) (o\(_h\)): the end-effector is moving towards an object,
\( d(x_{ef}, x_{o_h}) \to 0 \); b) ObjectInHand (o\(_h\)): the object is in

\(^2\)The information from the object can be obtained either from the vision
system or the proximity sensor of the skin. The same is valid for the property
ObjectInHand.
the end-effector, i.e. \( d(x_{ef}, x_{o}) \approx 0 \), where \( d(\cdot, \cdot) \) is the Euclidean distance between the end-effector \( x_{ef} \) and the detected object \( x_{o} \); c) GripperState \( g_s \): the current state of the gripper (open/closed).

In order to extract semantic rules, we randomly select one participant that demonstrates a sandwich making from a kitchen data set\(^3\). Then, we obtained a decision tree using the information of the ground-truth\(^4\) data of the analyzed subject. We split the training and testing data as follows: the first 60% of the trails are used for training and the rest 40% for testing, similar to [1]. Then, we obtained the tree \( T_{\text{sandwich}} \), see Fig. 3.

![Tree obtained from the sandwich making scenario (\( T_{\text{sandwich}} \))](image)

**Fig. 3.** Tree obtained from the sandwich making scenario \( (T_{\text{sandwich}}) \) [1].

### IV. ROBOT KINESTHETIC TEACHING RESULTS

The next challenge is to test our obtained semantic rules in a completely new environment for a new industrial task of packing oranges. Our proposed demonstration has been successfully implemented in our robotic platform Tactile Omni-directional Mobile Manipulator (TOMM), see Fig. 1. TOMM is composed of two industrial robot arms (UR-5) covered with artificial skin, two Allegro hands from SimLab also covered with our artificial skin and 2 cameras on its fixed head used to obtain the 3D position of target objects [5].

#### A. Industrial scenario

We consider the task of packing oranges. With this scenario, we can highlight the benefits of using the tactile and proximity sensors on the robot skin to sense the quality of the fruits\(^5\). The user teaches the robot the activities and the intermediate tasks required to sort oranges: a) Good oranges (with stiff texture) will be placed in a box, and b) Bad oranges (with soft texture) will be thrown into the trash bin. The texture of the oranges is evaluated using the force sensors from the robot skin placed in the external finger patches of the grippers. The stiffness threshold to discriminate the texture of the fruits is defined during the demonstration.

Then, we further tested the robustness of the obtained model \( T_{\text{sandwich}} \) for the new industrial scenario. Therefore,

\(^3\)The used data set is publically available at the following link [http://www.ics.ei.tum.de/ics-data-sets/cooking-data-set](http://www.ics.ei.tum.de/ics-data-sets/cooking-data-set)

\(^4\)The ground-truth was manually labeled by a person considered as an expert since this person received a training session.

\(^5\)This scenario was inspired by the standard process of orange sorting where humans use their tactile sensation to discriminate good and bad oranges.

for this experiment there is no training phase, thus allowing the robot to re-use the inferences that it learn from previous experiences, e.g. sandwich-making as shown in Fig. 1. Then, we expect that the inference module automatically segments and infers the demonstrated activities on-the-fly by re-using the learned semantic models. To quantitatively validate the robustness and generalization of our system, we tested our semantic models \( T_{\text{sandwich}} \) with different variations on the Kinesthetic demonstrations for the packing oranges scenario performed by two different participants\(^6\). A total of four demonstrations\(^7\) are considered and our system is able to infer the Kinesthetically demonstrated activities on-the-fly. For these four demonstrations, the position of the oranges in all the experiments is randomly selected. The overall results\(^8\) of the segmentation and recognition of the demonstrated activities is around 83.15% of recognition accuracy [1].

The presented semantic-based method, not only segments and recognizes human demonstration, but also allow the user to generate task plans to define new processes. The tasks generated by the user are also stored in the knowledge-based and a new process can be generated. This process is generated with un-bounded variables, which will be instantiated during running time, taking in consideration the information obtained from the multi-modal perception system (tactile skin, vision, robot state, etc.). In this case, the user creates the process of packing oranges which consists on the following sequential tasks:

\( T_1 \{ \text{Pick Fruit} \} = [1) \text{Reach(object)}, 2) \text{Take(object)} \}, T_2 \{ \text{Identify Good Fruit} \} = [3) \text{Put(object, place)}, 4) \text{Release(object)}, 5) \text{Squeeze(object), ..., 6) Take(object)} \}

where \( object = \text{orange} \) and \( place = \text{squeezable area} \). If the stiffness of the orange is high, then it is considered as “good-orange” and the following task is executed: \( T_3 \{ \text{Place Fruit Box} \} = [a1) \text{Put(object, place)}, a2) \text{Release(object)} \}, \) where \( object = \text{orange} \) and \( place = \text{box} \). On the other hand, if the orange is soft, then it is considered as “bad orange”, then the following task is executed: \( T_4 \{ \text{Place Fruit Trash} \} = [b1) \text{Put(object, place), b2) Release(object)} \} \) in this case \( object = \text{orange} \) and \( place = \text{trash} \). After the new process has been defined, the user only needs to indicate the robot to start executing the new process according to a stopping criteria, also defined by the user, for example, a maximum number of oranges to be packed. Then, the robot will start executing the new learned process until it reaches the stopping criteria or there are no more oranges to be packed.

### V. CONCLUSIONS

This paper summarizes the results of a novel semantic-based method to obtain general recognition models. The obtained semantic representations are robust and invariant to

\(^6\)One participant was a robotic expert and the other non-expert. We are planning to extend this study to a larger group of participants.

\(^7\)Note that the data from the robot Kinesthetic demonstrations was not used to improve in any way the semantic models \( T_{\text{sandwich}} \).

\(^8\)The ground-truth is obtained from visual information of the robot cameras, used by each participant to segment and label the taught activities.
different demonstration styles of the same activity. Additionally, the obtained semantic representations are able to re-use the acquired knowledge to infer different types of activities from household to industrial scenarios. We presented an approach that automatically extracts the meaning of the demonstrated activities by means of semantic representations. This new learning by demonstration approach enables non-expert operators to teach new task to industrial robots.

ACKNOWLEDGMENTS

The research reported in this paper has been (partially) supported by the German Research Foundation DFG, as part of Collaborative Research Center (Sonderforschungsbereich) 1320 EASE - Everyday Activity Science and Engineering, University of Bremen (http://www.ease-crc.org/). The research was conducted in subproject R01, “NEEM-based embodied knowledge framework”.

REFERENCES


