

Multimodal Integration Processes in Plan-Based Service Robot Control*

Dominik Off, Jianwei Zhang**

TAMS, Department of Informatics, University of Hamburg, 22527 Hamburg, Germany

Abstract: Cross-modal integration processes are essential for service robots to reliably perceive relevant parts of the partially known unstructured environment. We demonstrate how multimodal integration on different abstraction levels leads to reasonable behavior that would be difficult to achieve with unimodal approaches. Sensing and acting modalities are composed to multimodal robot skills via a fuzzy multisensor fusion approach. Single modalities constitute basic robot skills that can dynamically be composed to appropriate behavior by symbolic planning. Furthermore, multimodal integration is exploited to answer relevant queries about the partially known environment. All these approaches are successfully implemented and tested on our mobile service robot platform TASER.

Key words: service robotic; multimodal integration; plan-based robot control

Introduction

The plan-based robot control paradigm is the most promising approach to adapt robot behavior to a dynamically changing environment. Nevertheless, generating and executing plans that perform high-level tasks is difficult for robots – like our service robot TASER (see Fig. 1) – that act in dynamic unstructured environments, for several reasons. One reason is that often not all necessary knowledge is available in order to generate a plan for a given task. Especially dynamic facts, like the state of doors or the location of moveable objects (e.g., cups, plates, buckets, etc.), should not be stored for a longer period of time in the robot's memory. We developed the planning system for artificial cognitive systems (ACogPlan) as the deliberative planning component of the overall plan-based robot control architecture. ACogPlan is able to deal with a partially known environment by reasoning about

necessary unknown information, generating queries about this information, and submitting these queries to external knowledge sources. One major knowledge



Fig. 1 TASER executing a manipulation task

Received: 2010-11-29

* Supported by the German Research Foundation (No. GRK 1247/1)

** To whom correspondence should be addressed.

E-mail: zhang@informatik.uni-hamburg.de

source for service robots – that are usually equipped with several sensor systems – is perception. However, reliably perceiving the external world is still a challenging task for robot platforms.

The integrative processing of different sensing and acting modalities is an essential approach to provide reliable information about complex dynamic environments^[1]. In order to perform high-level tasks in a plan-based robot control architecture, we propose to integrate sensing and acting modalities in the following – not exclusive – two ways:

(1) Acting and sensing modalities are composed to multimodal robot skills; and

(2) Each sensing and acting modality is a unimodal robot skill and integrated by symbolic planning to reasonable multimodal robot behavior.

“An atomic robot skill is the generic term for pairs of primitive operators and control programs. It describes robot actions in both symbolic representation and numerical implementation”^[2]. In this sense robot skills connect the symbolic planning layer with the robot control programs. Figure 2 illustrates the two proposed multimodal integration approaches. Unimodal skills are statically integrated into multimodal skills and dynamically composed into multimodal plans by symbolic planning. Uni- and multimodal robot skills then serve as the basic building blocks for the generation of plans that perform high-level tasks.

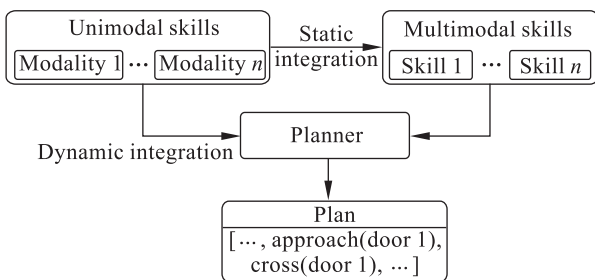


Fig. 2 Static and dynamic multimodal integration

1 State-of-the-Art in Multimodal Plan-Based Robot Control

1.1 Open world robot planning

Conformant, contingent, or probabilistic planning approaches can be used to generate plans in situations where insufficient information is available at planning time^[3-5]. These approaches generate conditional plans – or policies – for all possible contingencies. Unfortunately, these approaches are computationally hard, scale

badly in dynamic unstructured domains, and are only applicable if it is possible to foresee all possible outcomes of a knowledge acquisition process. Therefore, we agree with Brenner and Nebel^[6] that these approaches can hardly be applied in real world robotics.

The recently published continual planning system^[6] deals – like our approach – with the challenge of generating a plan without initially having sufficient knowledge, but focuses on a different question. While Brenner and Nebel^[6] described why and when agents should switch between planning and acting, ACogPlan is able to generate knowledge acquisition queries for indeterminable preconditions in order to find necessary information.

The approach of Dornhege et al.^[7] also integrates external components – so-called semantic attachments – into the planning process. However, semantic attachments do not directly consider multimodal perception as a source of external information, but mainly aims at the integration of high-level general purpose symbolic planning and domain specific lower-level reasoning (e.g., path or motion planning). Furthermore, integration is not done autonomously (i.e., by reasoning on the need to acquire information from external sources), but predefined in the domain description.

1.2 Multimodal integration in robotics

Luo and Kay^[1] classified the combination of different sensor modalities into multisensor integration, which is the synergistic use of information provided by multiple sensors to assist in the accomplishment of a task, and multisensor fusion, which refers to an actual combination of different sources of sensory information into one representational format.

The integration of multiple sensing and acting modalities occurs in almost every robot action, because most of the actions are guided – or at least monitored – by sensors and thus at least two modalities are integrated. Typical service robot manipulation tasks like opening a door^[8] or pushing a wheel chair^[9] usually involve even more modalities. For multimodal perception laser range data and camera images are often fused due to their complementary properties while being easily directed at a common workspace^[10,11]. A technique to fuse visual and depth information acquired by a camera and laser range scanner is proposed by Scheibe and Scheel^[12].

2 Multimodal Robot Skills

Mobile service robots are usually equipped with many different sensor systems. The integration of multiple sensors can result in more informative percepts of the external world. In this way multimodal integration can significantly improve the reliable perception of relevant parts (e.g., type of objects, state of doors, etc.) of the environment.

2.1 Multisensor fusion

Sensor data may always contain errors and thus may lead to incorrect and incomplete internal representations of the external world. Several probabilistic approaches exist that are expressive enough to represent these uncertainties in a reasonable manner. Unfortunately, probabilistic representations lead to an enormous rise in computational complexity because probability values have to be assigned to all possible locations and configurations for all known objects.

In order to avoid these computational problems we propose – as published in Weser et al.^[13] – the practical approach to adopt a fuzzy representation of certainty about the existence and identity of objects and their defining features. An object/feature can be existent, doubtful or impossible. Objects are defined by sets of features. Features are defined as subsets of the sensory data stream that can be differentiated. The approach is extendible to new algorithms that may provide features that are not detectable yet. Some detectors provide features that overestimate occurrences of the requested object, others provide features underestimating them. Others, however, gain from previous assumptions (i.e., they confirm/dismiss object evidences derived from other features).

Unimodal features are integrated to multimodal perceived objects using membership to fuzzy sets. The underlying fuzzy rules can be visualized in Fig. 3. The presented rules define how unimodal detectors are combined to multimodal percepts. If the integration

Laser Camera	Impossible	Doubtful	Assured
Impossible	Impossible	Impossible	Doubtful
Doubtful	Impossible	Doubtful	Assured
Assured	Impossible	Assured	Assured

Fig. 3 Fuzzy rules that describe how unimodal detectors are combined to multimodal percepts

results in the doubtful detection of an object, then it is tried to use additional modalities or other knowledge sources (e.g., human-robot interaction) in order to determine the sufficiently exact state of the external world. The fuzzy multisensory fusion approach is developed for goal driven scenarios, where only objects of interests are perceived. Thus, it is not necessary to match detected features from the bottom up to corresponding objects.

The fuzzy multisensor fusion approach is used on our service robot platform TASER (see Fig. 1) to reliably detect relevant parts of its natural office environment (e.g., the state of doors, detection of tables, etc.). See Weser et al.^[13] for more details.

2.2 Encapsulate multimodal robot skills

Multimodal robot skills are described on a symbolic level in the form of planning actions – sometimes also called planning operators^[14]. Planning actions are defined by a precondition and an effect (i.e., a set of facts that are removed from the world model and a set of effects that will be added to the world model after the execution) and are the basic building blocks for symbolic planning processes^[4].

We define the availability of necessary sensing and acting modalities as part of the precondition of the planning action for each multimodal robot skill. This approach makes it possible to find plans even in the case of unavailable modalities via dynamically composing robot actions that do not depend on unavailable modalities. If, for example, the main visual perception system is not available (e.g., due to malfunction or insufficient lighting conditions) the planner tries to compose the remaining modalities to a plan that performs the currently desired task.

Figure 4 demonstrates how we encapsulate the multimodal robot skill `check_latch` – which is able to determine whether the door latch is in the notch – by a robot action.

```

action(
  % task
  check_latch(Door),
  % precondition
  (available(arm), available(hand),
   available(force), near(robot,Door)),
  % removed facts
  [],
  % added facts
  [determinable latched(Door)]
).

```

Fig. 4 Symbolic description of multimodal robot skill `check_latch`

3 Integrated Architecture

The integrated planning and control architecture is sketched in Fig. 4. Facts, axioms, and taxonomic knowledge are stored in the robot's memory system. State-of-the-art planning systems – like SHOP2^[15] – usually store the planning methods and actions in separate domain files (i.e., the world is *a priori* partitioned into fine-grained domains). In this manner the performance can be increased, because only relevant methods and actions have to be considered by the planning system. This approach is usually inflexible in the context of embodied service robots. Furthermore, it is impossible to partition an unstructured environment into fine-grained subdomains. Thus, we store all planning related information in the same memory system. Information that is known to be irrelevant for a given task can be temporally hidden from the planning system in order to improve the performance by not considering irrelevant information.

Actions – as already mentioned – describe primitive executable robot actions which are directly dependent on the available robot devices. In contrast, HTN methods represent domain knowledge that defines how to decompose tasks into subtasks.

The human-robot interaction (HRI) module features the instruction of tasks and the acquisition of knowledge. A dedicated predicate-based language is currently used to communicate with the robot platform. The controller is responsible for the execution and monitoring of generated plans. If the controller detects that the existing plan has become invalid, it initializes the replanning process.

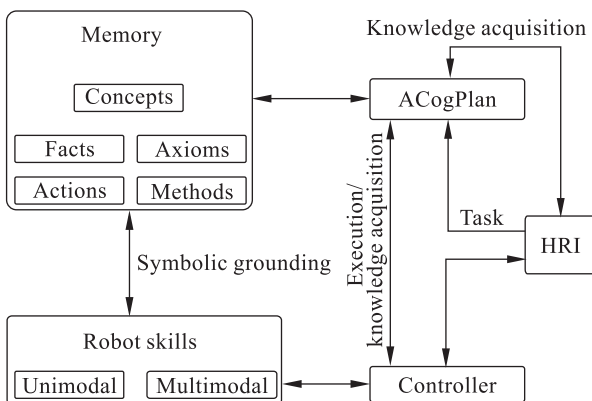


Fig. 5 Overall system architecture

4 High-Level Multimodal Integration via Symbolic Planning

The sensing and planning modalities of robot platform constitutes its basic skills. All available modalities of our mobile manipulator TASER are symbolically denoted by corresponding planning actions. This approach enables the planning system to dynamically compose available modalities to reasonable multimodal behavior. While the integration of different modalities to multimodal robot skills (as described in Section 2) is suited for low-level sensor fusion, the integration of unimodal skills by deliberative planning enables the flexible high-level integration of sensing and acting modalities.

5 Multimodal Perception Planning

As already mentioned above, generating plans for mobile service robots that act in unstructured environments is difficult, because often not all necessary information is available at planning time. Our planning system ACOGPlan is based on an open world model. Thus, in a given world model w a logical statement, st , is called provable if w entails that st is true, disprovable if w entails that st is false, or indeterminable otherwise. ACOGPlan is able to (1) reason about relevant unknown information, (2) generate queries about this information, (3) execute knowledge acquisition actions in order to submit these queries to external knowledge sources, and (4) reasonably integrate the responses into the planning process. We distinguish between ground queries and new instance queries. Let us assume that the task of the robot is to grasp a red cup which is located in the lab and thus it tries to find a true ground instance of the statement: $cup(X)$, $red(X)$, and $in_room(X, lab)$, where 'X' denotes an existentially quantified variable. Furthermore, let us assume that the robot only knows that $cup1$ is located in the lab. Then first the ground query $red(cup1)?$ would be generated. If this query can not be determined to be true the query new instance($cup(X)$, [$cup(cup1)$], [$red(X)$, $in_room(X, lab)$])? would be generated. New instance queries consist of the unground literal ($cup(X)$), the known ground instance ($[cup(cup1)]$), and the context ($[red(X)$, $in_room(X, lab)]$).

Since perception is a major source of information

for mobile service robots we propose to exploit available sensing modalities in order to answer these abstract queries. Special HTN methods encapsulate domain knowledge that describes how to compose the basic acting and sensing capabilities for the purpose of answering the desired query.

6 Results

All the described multimodal integration approaches are implemented on our service robot platform TASER. During the experiments TASER executed many different pick-up and delivery tasks like “bring me a cup of coffee from the kitchen” in our office environment. The current state of the dynamic office environment is usually only partially known by TASER. Especially, dynamic facts (e.g., the state of doors or the location of moveable objects) are not stored in the robots’ long-term memory and have to be acquired on demand. TASER was able to reason about relevant unknown information (i.e., find out what additional information may help to solve a currently unsolvable planning problem) and acquire this information by reasonably composing multi- and unimodal perception skills. For example TASER autonomously detects unknown doors during a pick-up and delivery task, as already published in Weser et al.^[13] Figures 6-8 demonstrate how the multimodal integration of an omnidirectional camera and a laser range finder are used to detect doors. Simple features in laser range scans and omnidirectional images are used to estimate door candidates. The detection method in laser range scans searches for line segments that show a sufficiently long gap that is not caused by occlusion. In omnidirectional images, pairs of sufficiently long horizontal edges are considered as potential door candidates. As shown in Figs. 6 and 7 the separate results of each feature detector are not reliable. The overall result of a fuzzy multimodal combination, however, renders higher elaborated methods unnecessary. Figure 8 shows the combined result using fuzzy rules. The combined multimodal result shows only doors belonging to assured.

We demonstrated how service robots can benefit from multimodal integration processes on different abstraction levels. The fuzzy sensor fusion approach enabled us to compose sensing and acting modalities to multimodal robot skills without the need to use computational hard probabilistic approaches. Additionally,



Fig. 6 Feature detector using omnidirectional image



Fig. 7 Feature detector using laser range scan, the results are transformed to image coordinates for better visualization.



Fig. 8 Integrated multimodal result

we defined each single modality as a basic robot skill. In this way modalities can be dynamically integrated by symbolic planning techniques to reasonable robot behavior. The plan-based robot control paradigm is the most promising approach to endowing robots with autonomy and intelligence. Unfortunately, reasoning about external information and the acquisition of relevant knowledge has not been sufficiently considered in existing planning approaches and is seen as an important direction of further growth^[16]. Our open world planning system ACogPlan takes a step in this direction. Abstract queries about relevant unknown information are generated and then – if possible – available sensing modalities are composed into multimodal perception skills that are able to answer these queries.

Directions of future work include the integration of incremental natural language processing and the world wide web as additional knowledge sources.

References

- [1] Luo R C, Kay M G. *Multisensor Integration and Fusion for Intelligent Machines and Systems*. New York, USA: Ablex Publish Corporation, 1995.
- [2] Weser M. *Hierarchical memory organization of multimodal robot skills for plan-based robot control* [Dissertation]. University of Hamburg, Germany, 2009.
- [3] Russel S J, Norvig P. *Artificial Intelligence: A Modern*

- Approach. Prentice Hall, 2009.
- [4] Ghallab M, Nau S, Traverso P. *Automated Planning: Theory and Practice*. Elsevier Science, 2004.
- [5] Kuter U, Nau D, Reisner E, et al. Conditionalization: Adapting forward-chaining planners to partially observable environments. In: *Proceedings of ICAPS-07 Workshop on Planning and Plan Execution for Real-World Systems – Principles and Practice for Planning in Execution*. Providence, USA, 2007.
- [6] Brenner M, Nebel B. Continual planning and acting in dynamic multiagent environments. *Autonomous Agents and Multiagent Systems*, 2009, **19**(3): 297-331.
- [7] Dornhege C, Eyerich P, Keller T, et al. Semantic attachments for domain independent planning systems. In: *Proceedings of the 19th International Conference on Planning and Scheduling (ICAPS)*. Thessaloniki, Greece: AAAI Press, 2009: 114-121.
- [8] Kim D, Kang J, Hwang C, et al. Mobile robot for door opening in a house. In: *Lecture Notes in Computer Science*. Springer Berlin, 2004.
- [9] Rubrecht S, Bouzid R, Mebal N, et al. HAMMI: Intelligent robot of assistance to people with motor disabilities. In: *10th International Conference on Control Application*. Quebec City, Canada, 2008.
- [10] Weser M, Westhoff D, Hüser M, et al. Multimodal people tracking and trajectory prediction based on learned generalized motion patterns. In: *International Conference on Multisensor Fusion and Integration (MFI)*. Heidelberg, Germany, 2006: 541-546.
- [11] Weser M, Westhoff D, Hüser M, et al. Real-time fusion of multimodal tracking data and generalization of motion patterns for trajectory prediction. In: *International Conference on Information Acquisition (ICIA)*. Weihai, China, 2006: 786-791.
- [12] Scheibe K, Scheel M. Multi-sensor panorama fusion and visualization. *Communication and information technology research*. Technical report 141. University of Auckland, New Zealand, 2004.
- [13] Weser M, Jockel S, Zhang J. Fuzzy multisensory fusion for autonomous proactive robot perception. In: *Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ) at IEEE World Congress on Computational Intelligence (WCCI)*. Hong Kong, China, 2008: 2262-2267.
- [14] Weser M, Off D, Zhang J. HTN robot planning in partially observable dynamic environments. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. Anchorage, AK, USA, 2010: 1505-1510.
- [15] Nau D, Au T, Ilghamy O, et al. SHOP2: An HTN planning system. *Journal on Artificial Intelligence Research*, 2003, **20**: 379-404.
- [16] Nau D. Current trends in automated planning. *AI Magazine*, 2007, **28**(4): 43-58.