Chapter 1
Introduction to Robot Introspection

Abstract In this chapter, we mainly introduce the definition, background, significance, and the start-of-the-art methods of collaborative robot multimodal introspection. The current issues of robot introspection are also introduced, which including the complex task representation, anomaly monitoring, diagnoses and recovery by assessing the quality of multimodal sensory data during robot manipulation. The overall content of this book is presented at the end.

1.1 What is Robot Introspection?

As humans, doubt and understanding our own limitations, failures and shortcomings is a key for improvement and development, such knowledge alters our behaviors e.g. to execute tasks in a more cautious way [1, 2]. Due to the humans can learn and recognize the environment for a long time through the five senses (visual, acoustic, gustatory, osphretic, and tactile) and generate corresponding internal models in the daily execution of operational tasks [3–7]. Correspondingly, equipping robots with a set of skills that allows them to assess the quality of their sensory data, internal models, used methods etc. is correspondingly believed to greatly improve the autonomous operability and safety performance of an autonomous system, e.g. presented in literature [8, 9].

In recent years, the ability of robots to learn intrinsic introspective models like humans has been favored by robotics researchers [10, 11]. With respect to humans, introspection is the process of examining one’s internal state by evaluating the potential patterns of multi-modal sensing data [12, 13, 15], as shown in Fig. 1.1. Its application lies in giving robots three capabilities: “What’s it doing?”, “How to do?”, and “How’s it doing?”. Among them, “What’s doing?” is to solve the problem of continuous state estimation during the robot execution; “How’s it doing?” is to monitor the real-time state during robot execution, such as normal or abnormal; “How to do?” including the robot task planning and decision making on future motion. The development and application of robot introspection can effectively
(1) estimate, monitor and prevent abnormal events; (2) evaluate the internal state of the robot; (3) strengthen the ability to recovery abnormalities; (4) optimize control and motion decisions.

However, robots do not have thoughts neither feelings, only data, hardware and algorithms. Therefore, robots can only assess the quality of sensor data, internal models, representations, information, perception input etc. Such knowledge can later lead to a modification of robots behavior by including the assessed quality score in the planning process. Consequently, robot introspection relates to safety, active perception, mapping and many topics. These topics have a direct impact on a variety of research areas, such as long term autonomy, search and rescue, and many others. Long term autonomy can benefit from autonomous failure recovery and active learning. For search and rescue, estimation of the confidence of the sensor input and used maps is essential for overall risk assessment. Moreover, for a large variety of tasks assessing the quality of sensor data, internal models, representation, information will directly affect mission success. The ability to reason and solve its own failures, and proactively enrich own knowledge is a direct way to improve autonomous behaviours of a robot.

### 1.2 Background and Significance

Traditional industrial robots have played a significant role in standardization and automated production by isolating from human, which is difficult to meet unstructured, personalized, and flexible non-standard complex task requirements in a dynamic environment such as the human robot interactive scenarios that shown in Fig. 1.2. Due to they lack of multi-functional and intelligent capabilities that per-
1.2 Background and Significance

Fig. 1.2 Robots are applied in diverse, individualized and flexible non-standard tasks in unstructured dynamic environments

Robots are applied in diverse, individualized and flexible non-standard tasks in unstructured dynamic environments. In addition, with the widespread application and development of collaborative robots [16, 17], robots will gradually move from a traditional closed manufacturing environment to a shared space that coexists and interacts with human [18], from semi-automatic operation tasks to autonomously perform, which inevitably leads to various unpredictable anomalies, such as objects slipping, collisions between end-effector and the environment, human collisions, and system abnormalities. Therefore, in order to give the robot a longer-term autonomy and a safer human-machine collaborative environment, the robot should perform real-time multi-modal fusion modeling to achieve accurate introspection of its own movement behavior (identification of movement behaviors, abnormal monitoring, and abnormal diagnoses), and anomaly recovery are essential in the next generation of intelligent collaborative robots.

Generally speaking, the robot follows the control framework of the Sense-Plan-Act (SPA) in a structured environment. First, the robot observes the surrounding environment and builds an internal model [14]. Then, it formulates a task execution plan, and finally executes this plan. However, the control framework is difficult to meet the increasingly complex robot operation tasks in unstructured dynamic environments [20, 23]. In the recent decades, increasing the robot multi-modal introspection and anomaly recovery policy after the robot execution would be an effective way to address the aforementioned issues. To this end, a novel Sense-Plan-Act-Introspect-Recover (SPAIR) control framework is proposed in this book, which can provide a theoretical framework and solutions for autonomous robot operation, human-computer interaction, and human-machine integration in the future.

1.3 Issues to be Addressed in Robot Introspection

This book intends to provide the reader with a comprehensive overview of the newly developed methods for improving robot introspection. In recent years, with the rapid development of collaborative robots, scholars in the field of robotics and artificial intelligence have carried out explorations on learning from human demonstration, integration with humans, multimodal sensing, and learning methods and theories.
of robots. Although lots of research results and research ideas within those topics are presented, the difficulties and key issues of multi-modal modeling and its application of robots are not fully considered, and effective theoretical frameworks and systematic integration have not yet been formed. The application research of robot multi-modal sensing makes the research in this book more valuable. Therefore, this book develops a theoretical study of robot multi-modal introspection and learning based on non-parametric Bayesian models, aiming to endow robots a longer-term autonomy and a safer human-robot collaboration environment. With respect to the state-of-the-art status of robot introspection, the following five issues need to be addressed urgently:

- **(1) How to assess the quality of internal models, methods and sensor data, used by robots and how to alter their behavior upon this information?**

Humans can accurately perceive the state of the outside world or the objects they need to operate through the five senses based on past experience and internal models, and can self-recognize the advantages and disadvantages to learn and adjust existing experiences and models. However, robots can only sensing the world through sensory data such that how to model and analyze of multi-modal time series is a difficult problem to realize robot introspection.

- **(2) How to represent the complex robot task and generalize to the environmental changes during unstructured scenarios?**

It’s difficult to complete tasks from the beginning to the end with fixed and pre-programmed movements in the case of an unstructured dynamic human-robot collaboration environment. Resulting in encountering fault or abnormality of the task because of the unstructured environment, the changing pose of the operating object, and the unpredictable robot state. Therefore, how to learn and generalize the robot complex tasks from human demonstration for adapting to the changes in the environment is an valuable question.

- **(3) How to implement the anomaly monitoring and diagnoses of multimodal time series during robot execution?**

The robot anomalies usually derived from joint encoders, the environmental changes, and human interference in the human robot interaction scenarios. Assuming that the robot can monitor and classify these anomalies, it will reduce or prevent the robot from being potentially injured and improve the safety of human-robot collaboration. Due to the diversity and complexity of the types of anomalies, it isn’t possible to directly enumerate and model all anomalies, resulting the anomaly monitoring cannot be achieved through supervised learning methods. Generally, the robot anomaly monitoring were implemented by learning the outstanding difference between the normal ones. In other words, anomalies are identified from models of normal robot behavior. Making the robot anomaly monitoring and diagnoses should be critical factor for realizing the robot longer-term autonomy.
• (4) How to learn the recovery policy from human demonstration for specific anomalous event?

Robot anomaly recovery cannot be performed by the traditional robot’s own motion planning algorithm under the human robot interaction scenarios. Whereas human expectations for robot motion should be taken into consideration. The main challenge is how to incrementally learn the robot recovery policy when encounter various abnormal events from unstructured human demonstrations.

• (5) How to develop a stable and extendable software framework for integrating the introspective abilities by modeling the multimodal sensory data?

How to develop a software for robot multimodal introspection that including multi-functional modules for complex task representation, multi-modal fusion, anomaly monitoring, anomaly diagnoses, and anomaly recovery. It’s another common issue to be solved for realizing long-term autonomy and safe human robot collaborative environment.

1.4 System Framework for Robot Multimodal Introspection

1.4.1 Introduction

With the rapid development of collaborative robots, robots are moving from the traditional structured environment of industrial manufacturing to unstructured dynamic shared workspaces that coexist with, collaborate with, and integrate with people. Although the robot has a more robust and stable control algorithm, the algorithm cannot precisely model the unstructured environment, and unexpected events will occur. In order to endow the robot with a longer-term autonomy and a safer human-robot collaborative environment, the robot must implement self-monitoring and recovery capabilities by evaluating multi-modal sensory sensor in real time.

This book aim to propose a theoretical framework and extendable system platform for robot multimodal introspection and learning based on non-parametric Bayesian models. It mainly includes the following four aspects:

• Modeling of multimodal time series, details in Chap. 2;
• Learning and representation of robot complex task, details in Chap. 3;
• Multimodal anomaly monitoring, details in Chap. 4;
• Multimodal anomaly diagnose, details in Chap. 5;
• Heuristic incremental learning robot recovery policies, details in Chap. 6.

The proposed framework divides the complex tasks of the robot into the directed graph, and learns the motion primitive model and realizes the recognition of the motion behavior for the motion behavior between the nodes in the directed graph. The non-parametric Bayesian time series model learns models for anomaly monitoring
and anomaly diagnoses models, and compares and analyzes the performance and results of anomaly monitoring under three different anomaly threshold conditions. A robot anomaly recovery policy is designed at the top of the framework to handle anomalies encountered by external disturbances based on the representation of robot complex task and introspection. We assume the considered recovery policy should be learned incrementally from the context of task. Therefore, the proposed Re-enactment and Adaptation recovery policies correspond to accidental anomalies and persistent anomalies, respectively. Among them, the reenact policy uses a probability model of a polynomial distribution to count the humans decision-making at different anomaly situations for the accidental anomalies. Additionally, the robot learns and stores its recovery policy, while the motion adaptation uses human intuition to teach the recovery of persistent anomalies.

The anomalous event of robot is unpredictable, so that the recovery policy of the robot under different abnormal conditions cannot be planned offline. Therefore, learning the recovery policy incrementally is the key to achieving human-robot safety collaboration. The framework has fast and robust anomaly monitoring and diagnoses capabilities for both initially planned motion primitives and newly learned recovery motion primitives. In addition, by using non-parametric Bayesian models to incrementally learn models from multi-modal sensing data, the stability, reliability and fault tolerance of the framework are well improved.

1.4.2 Modules of Introspective System

To endow robots a longer-term autonomy and a safer human-robot collaboration environment, the robot’s Introspection phase and the Recovery phase were added to the traditional robot control framework SPA. The system framework of the multimodal sensing and learning is shown in Fig. 1.3 and named SPAIR. The framework mainly includes four modules: (1) directed graph representation of complex robot tasks; (2) generalization and identification of robot motion behavior; (3) real-time abnormal monitoring and diagnoses during robot execution; (4) learning recovery policy for recovering from robot abnormal events.

In recent years, the directed graph representation of complex tasks of robots has been investigated by a large number of scholars, such as Kappler et al. [19], Fox et al. [12], Kroemer et al. [21], and Niekum et al. [22]. To address the problem of robot task representation, many scholars tried to represent the complex task by a series of movement primitives, and realized the contact state estimation and task representation of the robot [13, 24–26]. This book assumes that the robot’s tasks are represented by a directed manipulation graph that include a set of movement primitives, and each primitive is composed of nodes: the start node and the target node, and along with the transition between different movement primitives is expressed too. In order to accurately represent the robot complex tasks and the subsequent recovery behavior, the movement primitive is formulated as follows: (1) The behavior of the robot is used to represent a complete movement of the robot, as shown in Fig. 1.3.
1.4 System Framework for Robot Multimodal Introspection

The complete process from the “Home” node to the “Pre-pick” node; (2) The robot task is usually composed of one or more movement primitives, as shown in Fig. 1.3. The “Pick-and-Place” task of the robot consists of four motions: The behavior is composed of five action nodes, among which the movement primitives are “Home → Pre-pick”, “Pre-pick → Pick”, “Pick → Pre-place” and “Pre-place → Place”.

In particular, the directed manipulation graph will be updated when the robot encounter an abnormal event, which including: $N_{ij}$ indicates a recovery node between nodes $N_i$ and $N_j$ and node $N_{ijk}$ denotes a recovery node between nodes $N_{ij}$ and $N_k$. By analogy, nodes for recovery will be generated in the original graph structure, and the connection between the nodes (including the recovery node and the node of the original task) will generate new motion behavior. From the recovery module at the top of Fig. 1.3, it can be known that the manipulation graph $G$ of the robot tasks will gradually robust and stable with the increasing recovery behaviors.
The nodes in the SPAIR framework will include multiple functions performed in parallel, including: the robot movement properties $\mathcal{M}$, the robot introspective properties $\mathcal{I}$, and the robot sensing external environment properties $\mathcal{V}$. The $\mathcal{M}$ in this book refer to the robot movement primitives; $\mathcal{I}$ includes the robot’s movement identification, anomaly monitoring, and anomaly diagnoses; $\mathcal{V}$ represent the visual estimation of the object’s pose in the environment. Among them, the robot’s introspective attributes are mainly responsible for real-time monitoring of robot anomalies. Once anomalies are detected, the system will classify the anomalies, and the system will recovery according to a limited recovery strategy for different types of anomalies. If the recovery strategy does not exist, The system will prompt the need for human recovery demonstration, and learn and update its recovery behavior. According to the diagnoses result, if it is a accidental anomaly, the robot will perform the motion redo recovery behavior, as shown in “anomaly $− i$” and “anomaly $− j$” shown in Fig. 1.3; if it is a persistent anomaly, the robot will execute the motion adaptation recovery behavior, as shown in “anomaly $− k$” and “anomaly $− m$” in Fig. 1.3.

The proposed framework is built on the ROS-Indigo\(^1\) platform, allowing signals of different modalities to communicate in the form of topics, and the sampling frequency of different sensors is fixed to a specific value by synchronous sampling (this book uses 10Hz), and combines multiple open source software and code, such as the description of the robotic task by the finite state machine SMACH\(^2\) and the ar_track_alvar\(^3\) for pose estimation of the object. In addition, the system will automatically collect multi-modal sensor signal data for each motion primitive and the subsequent proposed recovery behaviors during the robot’s movement. So that the robot can learn experience and generalize movement from the generated data. To this end, the relevant code, videos, pictures and usage documents of this book will be all open source,\(^4\) so that others researchers can improve and share it.

1.5 Summary

In this chapter, we mainly introduce the definition and significance of robot introspection in the background of human-robot collaboration. We inspired from human that doubt and understanding our own limitations, failures and shortcomings is a key for improvement and development, such knowledge alters our behaviors e.g.to execute tasks in a more cautious way. To endow robots a longer-term autonomy and a safer human-robot collaboration environment, we referred to the state-of-the-art status of robot introspection, five issues are needed to be addressed urgently. To this end, we proposed an introspective system framework with four modules, named Sense-Plan-

\(^1\)http://wiki.ros.org/indigo.
\(^2\)http://wiki.ros.org/smach.
\(^3\)http://wiki.ros.org/ar_track_alvar.
\(^4\)https://github.com/birlrobotics/smach_based_introspection_framework.
Act-Introspect-Recover (SPAIR) for address the problems including robot complex task graphical representation and execution (skill) identification from unstructured demonstrations, multimodel time series modelling for robot anomaly monitoring and diagnose as well as the recovery policies learning for different anomalies.

References

1. LaViers A. Make robot motions natural. Nature. 2019;565:422–4.
2. Les Z, Les M. Human perception. Machine understanding. Cham: Springer; 2020. p. 1–8.
3. Teixeira T, Dublon G, Savvides A. A survey of human-sensing: methods for detecting presence, count, location, track, and identity. ACM Comput Surv. 2010;5(1):59–69.
4. Sharma R, Pavlović VI, Huang TS. Toward multimodal human–computer interface. Advances in image processing and understanding: a festschrift for Thomas S Huang. Singapore: World Scientific; 2002. p. 349–365.
5. Pan B, Zheng Y, Wilkie D, et al. Crowd sensing of traffic anomalies based on human mobility and social media. In: Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems. ACM; 2013. p. 344–353.
6. Fong T, Nourbakhsh I, Dautenhahn K. A survey of socially interactive robots. Robot Auton Syst. 2003;42(3–4):143–66.
7. Jaimes A, Sebe N. Multimodal human-computer interaction: a survey. Comput Vis Image Underst. 2007;108(1–2):116–34.
8. Cifuentes CA, Rodriguez C, Frizeraneto A, et al. Multimodal human-robot interaction for walker-assisted gait; 2017 10(3):933–943.
9. de Gea Fernàndez J, Mronga D, Günther M, et al. Multimodal sensor-based whole-body control for human-robot collaboration in industrial settings. Robot Auton Syst. 2017; 94:102–119.
10. Do HM, Pham M, Sheng W et al. RiSH: a robot-integrated smart home for elderly care. Robot Auton Syst. 2017; 101:74–92.
11. Gao ZD, Su JB, Zhou W. Robot autonomous perception model for internet-based intelligent robotic system. In: IEEE proceedings of international conference on machine learning and cybernetics; 2005.
12. Fox M, Ghallab M, Infantes G, et al. Robot introspection through learned hidden markov models. Artif Intell. 2006;170(2):59–113.
13. Rojas J, Luo S, Zhu D, et al. Online robot introspection via wrench-based action grammars. 2017 IEEE-RAS international conference on robotics and system. IEEE; 2017. p. 113–119.
14. Xu Z, Li S, Zhou X, Wu Y, Cheng T, Huang D. Dynamic neural networks based kinematic control for redundant manipulators with model uncertainties. Neurocomputing. 2019;329(1):255–66.
15. Wu H, Lin H, Guan Y, et al. Robot introspection with Bayesian nonparametric vector autoregressive hidden Markov models. 2017 IEEE-RAS 17th international conference on humanoid robotics (Humanoids). IEEE; 2017. p. 882–888.
16. Braganza S, Costa E, Castellucci I, Arezes PM. A brief overview of the use of collaborative robots in industry 4.0: human role and safety. In: Occupational and environmental safety and health. Cham: Springer; 2019. p. 641–650
17. Goldberg K. Robots and the return to collaborative intelligence. Nat Mach Intell. 2019; 1(1):2.
18. Cherubini A, Passama R, Crosnier A, Lasnier A, Fraisse P. Collaborative manufacturing with physical human–robot interaction. Robot Comput-Integr Manuf. 2016; 1(40):1–3.
19. Kappler D, Pastor P, Kalakrishnan M, et al. Data-driven online decision making for autonomous manipulation. In: Robotics: science and systems; 2015.
20. Xu Z, Li S, Zhou X, Cheng T. Dynamic neural networks based adaptive admittance control for redundant manipulators with model uncertainties. Neurocomputing. 2019;357(1):271–81.
21. Kroemer O, Daniel C, Neumann G, et al. Towards learning hierarchical skills for multi-phase manipulation tasks. In: 2015 IEEE international conference on robotics and automation (ICRA). IEEE, 2015. p. 1503–1510.

22. Scott N, Sarah O, George K, et al. Learning grounded finite state representations from unstructured demonstrations. Int J Robot Res. 2015; 34(2):131–57.

23. Xu Z, Li S, Zhou X, Yan W, Cheng T, Dan H. Dynamic neural networks for motion-force control of redundant manipulators: an optimization perspective. IEEE Trans Ind Electron. Early access, 2020. https://doi.org/10.1109/TIE.2020.2970635.

24. Rojas J, Huang Z, Harada K. Robot contact task state estimation via position-based action grammars. In: IEEE-RAS 16th international conference on humanoid robots; 2016. p. 249–255.

25. Rojas J, Newman W, Jie H, et al. A steering wheel manipulation scheme by an anthropomorphic humanoid robot in a constrained vehicle environment. In: IEEE international conference on robotics and biomimetics; 2015. p. 1566–71.

26. Rojas J, Harada K, Onda H, et al. Contextualized early failure characterization of cantilever snap assemblies. In: The 14th IEEE-RAS international conference on humanoid robots; 2014. p. 380–387.