Research on Intelligent Manufacturing

Jing Xu · Hao Su ·
Rui Chen · Zhimin Hou

Robotic Intelligent Assembly

華中科技大學出版社
Huazhong University of Science and Technology Press

Springer

Research on Intelligent Manufacturing

Research on Intelligent Manufacturing (RIM) publishes the latest developments and applications of research in intelligent manufacturing—rapidly, informally and in high quality. It combines theory and practice to analyse related cases in fields including but not limited to:

Intelligent design theory and technologies
Intelligent manufacturing equipment and technologies
Intelligent sensing and control technologies
Intelligent manufacturing systems and services

This book series aims to address hot technological spots and solve challenging problems in the field of intelligent manufacturing. It brings together scientists and engineers working in all related branches from both East and West, under the support of national strategies like Industry 4.0 and Made in China 2025. With its wide coverage in all related branches, such as Industrial Internet of Things (IoT), Cloud Computing, 3D Printing and Virtual Reality Technology, we hope this book series can provide the researchers with a scientific platform to exchange and share the latest findings, ideas, and advances, and to chart the frontiers of intelligent manufacturing.

The series' scope includes monographs, professional books and graduate textbooks, edited volumes, and reference works intended to support education in related areas at the graduate and post-graduate levels.

Jing Xu · Hao Su · Rui Chen · Zhimin Hou

Robotic Intelligent Assembly

華中科技大學出版社
Huazhong University of Science and Technology Press

Springer

Jing Xu
Department of Mechanical Engineering
Tsinghua University
Beijing, China

Hao Su
Department of Computer Science
and Engineering
University of California
San Diego, CA, USA

Rui Chen
Department of Mechanical Engineering
Tsinghua University
Beijing, China

Zhimin Hou
Mechanical and Aerospace Engineering
North Carolina State University
Raleigh, NC, USA

ISSN 2523-3386 ISSN 2523-3394 (electronic)
Research on Intelligent Manufacturing
ISBN 978-981-96-2656-4 ISBN 978-981-96-2657-1 (eBook)
https://doi.org/10.1007/978-981-96-2657-1

Jointly published with Huazhong University of Science and Technology Press
The print edition is not for sale in China (Mainland). Customers from China (Mainland) please order the print book from: Huazhong University of Science and Technology Press.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

Robotic assembly is a fundamental technique in a wide range of manufacturing processes. While the classical rule-based robotic assembly methods have reached an elevated level of maturity, they have limitations when it comes to handling complex tasks with a large variety of components, which significantly restricts their applications.

Fortunately, in the past decade, deep learning and artificial intelligence have undergone transformative breakthroughs, showcasing remarkable success across various domains such as computer vision and natural language processing. These data-driven approaches have demonstrated great generalizability and have paved the way for significant advancements in robotic assembly by improving the robot capability of perception, planning and control to meet the requirement of increasing complexity and variety of assembly tasks.

This book is devoted to presenting the latest advances in robotic intelligent assembly strategies, with a focus on pig-in-hole tasks. The book covers a range of perspectives related to robotic PiH assembly strategies, including perception, model-based control, and learning-based control. The authors have put great effort into organizing previous research works in a systematic manner, offering readers a comprehensive understanding of the field.

The first part of the book introduces the background and drawbacks of existing approaches. The second part delves into improving the accuracy and computational efficiency of learning-based multi-view stereo and narrowing the simulation-to-reality domain gap for depth sensors. These techniques improve the robot's ability to perceive the geometry information of the assembly environment. The third part describes the model-based strategies. And the fourth part presents state-of-the-art learning-based strategies for general PiH tasks. These approaches enhance the generalizability of robot assembly by utilizing learning-based algorithms. By leveraging the power of deep learning, robots can adapt to new assembly scenarios and handle a wider range of components efficiently. The authors also explore techniques for accelerating the training process of existing learning algorithms, enabling faster deployment and improved performance.

The outcome of this book is a comprehensive treatise that establishes significant advances in the theoretical formulation and experimental validation of robotic assembly techniques. By leveraging the power of deep learning and artificial intelligence, the authors have made substantial contributions to the field. This book serves as a vivid signature, showcasing the remarkable progress in robotic intelligent assembly and inspiring future advancements in the field.

Hong Kong, China Ning Xi

Preface

Over a decade ago, we embarked on tackling one of the most fundamental yet challenging tasks in manufacturing—the robot assembly at a time when reinforcement learning (RL) methods were not yet well-developed or widely adopted.

Our initial efforts focused on developing models to accurately represent the intricate contact status between the peg and the hole during insertion. We reasoned that by precisely capturing and recognizing this critical contact status, we could adjust the peg's movements accordingly and accomplish the assembly task. The model-based approaches offered the advantages of efficiency and explainability.

However, as our research progressed, we gradually uncovered the limitations of this methodology. Complex contact scenarios involving multiple points of interaction and substantial variability in real-world conditions posed significant challenges that our models struggled to accommodate. It became evident that a paradigm shift was necessary to overcome these bottlenecks and achieve robust, adaptive assembly capabilities.

Consequently, we pivoted our focus towards learning-based strategies, embracing the power of reinforcement learning to derive optimal policies from data and experience. Unlike the vast body of work on RL in simulation environments or video games, two critical factors emerged as paramount for the successful adoption of RL in PiH tasks on physical robots: safety and sample efficiency. Any learning algorithm deployed in a real manufacturing setting must ensure the safety of the equipment and components, while also minimizing the need for costly trial-and-error during the training phase.

To address these challenges, we delved into the study of impedance-conditioned actions and hierarchical RL techniques. By incorporating compliant control and hierarchies of low-level policies, our methods demonstrated the ability to learn intricate insertion strategies that could successfully handle complex contact situations. Experimental validations on real PiH tasks verified the efficacy of our approaches, accomplishing these intricate assembly tasks with remarkable efficiency and reliability.

Despite these promising results, we recognize that there remain opportunities for further advancement. Aspects such as the autonomous discovery and definition of low-level policies, seamless integration with higher-level task planners, and generalization to diverse manufacturing scenarios warrant continued investigation. Our team remains committed to exploring these fundamental scientific problems, driven by the ambition to push the boundaries of intelligent robotic systems.

Beijing, China | Jing Xu
June 2024 | Rui Chen

Contents

Acronyms

ABS	Acrylonitrile Butadiene Styren
AI	Artificial Intelligence
ARIE	Attractive Region in Environment
BC	Behavior Cloning
BSDF	Bidirectional Scattering Distribution Function
CPS	Contextual Policy Search
CSCT	Center-Symmetric Census Transform
DDPG	Deep Deterministic Policy Gradient
DMP	Dynamic Movement Primitive
DRL	Deep Reinforcement Learning
EM	Expectation Maximization
F/T	Force-Torque
FC	Fuzzy Classifier
FEM	Finite Element Method
FLDVTSP	Fuzzy Logic-Driven Variable Time-Scale Prediction
FLS	Fuzzy Logic System
FOV	Field of View
FPS	Frame Per Second
GAN	Generative Adversarial Network
GMM	Gaussian Mixture Model
GMR	Gaussian Mixture Regression
GPR	Gaussian Process Regression
GPS	Guided Policy Search
GS	Gravitational Search
GVF	General Value Function
HC-CPS	Hierarchical Compliance-based Contextual Policy Search
HMM	Hidden Markov Model
HRL	Hierarchical Reinforcement Learning
iLQG	Iterative Linear-Quadratic-Gaussian
IPC	Incremental Potential Contact
IR	Infrared

kNN	k Nearest Neighbor
LFD	Learning From Demonstrations
LFE	Learning From Environments
MDDPG	Model-driven Deep Deterministic Policy Gradient
MDP	Markov Decision Process
MSD	Mean Standard Deviation
MVS	Multi-View Stereo
OC	Option Critic
PBR	Physically Based Rendering
PiH	Peg-in-Hole
PTFE	Polytetrafluoroethylene
REPS	Relative Entropy Policy Search
RF	Random Forest
RL	Reinforcement Learning
RMSE	Root Mean Square Error
ROI	Region of Interest
SAC	Soft Actor-Critic
SGBM	Semi-Global Block Matching
SGM	Semi-Global Matchin
SPP	Samples Per Pixel
SVM	Support Vector Machine
TD	Temporal Difference
TD3	Twin-Delayed-DDPG
ToF	Time-of-Flight
VAE	Variational Auto-Encoder

Chapter 1
Introduction

1.1 Background

ASSEMBLY is a capstone stage in the manufacturing process, which directly affects the quality of the final product [154]. Statistical evidence indicates that the time spent on assembly typically accounts for 20–50% of the total production time, while the associated costs constitute approximately 20–30% of the overall cost of a given product [160]. With the development and advancement of automation technologies, robots have been widely used in assembly tasks to overcome the inherent challenges associated with manual assembly, such as low efficiency, high cost, and increased accident rate [166]. Nevertheless, the escalating complexity and specialization of assembly tasks necessitate the deployment of an increasing number of assembly robots within unstructured and dynamic environments, thereby presenting substantial challenges such as inadequate adaptability to assembly processes, limited sensing capabilities, and demanding assembly environment prerequisites [61, 62]. Consequently, robotic assembly necessitates a transition towards enhanced intelligence to address these challenges.

The rapid advancement of Artificial Intelligence (AI) technology has promoted the advancement of intelligent robotic assembly system. When combined with advanced sensing system and control strategy, AI can significantly augment the capabilities of assembly robots in all the phases, including perception, decision-making, and control [20].

Peg-in-Hole (PiH) assembly is a fundamental and widely employed technique in industrial manufacturing [99, 154]. It accounts for approximately 40% of the overall assembly workload [61]. PiH has been extensively investigated across various domains, ranging from large-scale aviation component assembly [149] to small-scale component assembly in automotive production lines [133] and mold casting manufacturing [95]. Furthermore, it extends to the assembly of electronic components [60] and even micro-product assembly [83]. Notably, there has been a recent surge in research on PiH assembly for household tasks in human environments [6],

J. Xu et al., *Robotic Intelligent Assembly*, Research on Intelligent Manufacturing,
https://doi.org/10.1007/978-981-96-2657-1_1

e.g., chair assembly [136]. Due to its prominent status in assembly processes, the study of PiH assembly is of importance for the advancement of robotic intelligent assembly technology. Consequently, this monograph focuses explicitly on robotic PiH assembly.

1.2 Robotic Peg-in-Hole Assembly System

We first introduce a typical robotic PiH system. As shown in Fig. 1.1a, the holes are fixed, while manipulators grasp pegs to complete part mating based on feedback from the sensing system. Therefore, the sensing system is crucial for acquiring the feedback for successful assembly.

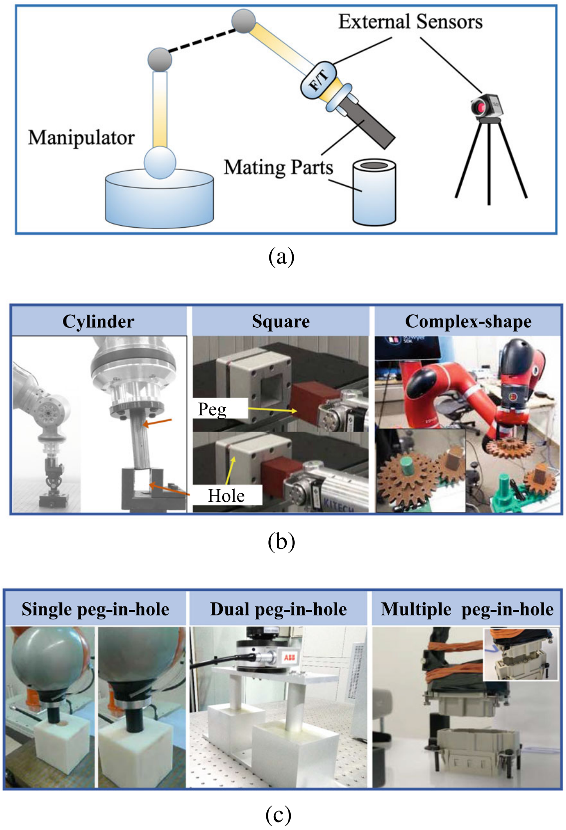

Fig. 1.1 Robotic PiH assembly system. **a** Framework. **b** Cylinder [115], square [67], and complex-shaped cases [131]. **c** Single [57], dual [166], and multiple PiH assemblies [120]

1.2.1 Manipulators

Manipulators have evolved significantly to meet the diverse requirements of assembly tasks. Industrial robots excel in large-scale and heavy component assembly and possess advanced position and velocity control for high-precision tasks. In contrast, collaborative robots find extensive application in the assembly of small products and deformable parts due to their ability to achieve compliant behaviors through precise force and torque control.

1.2.2 Mating Parts

Mating parts consist of a peg and a corresponding hole (Fig. 1.1a). While cylindrical PiH assemblies serve as the foundation, other geometries, including square pegs [67, 107], pegs with key slots, and complex shapes like gear assemblies [85], are employed in special cases. Besides single PiH [57], dual PiH [166], multiple PiH [120, 148] are also necessary for certain industrial applications, like electronics industry (see Fig. 1.1b).

The scale of mating parts varies significantly depending on the application. It ranges from macro-assembly, involving large aviation parts [149], to micro-assembly, focused on electronic components in circuit boards [83]. Certain assembly tasks require high precision, with clearances smaller than the resolution and accuracy of robots, often ranging from 0.02 mm to 0.2 mm [49]. Additionally, in addition to common rigid parts with high stiffness properties, other tasks necessitate flexible components made of materials such as plastics and wood [57, 85]. The clearance and surface characteristics of mating parts contribute to the complexity of assembly tasks, resulting in varying degrees of difficulty. While this monograph aims to provide an in-depth analysis of robotic PiH systems, the examination of clearance and surface characteristics falls outside its scope.

1.2.3 Sensing

Similar to human tactile and vision sensing, the sensing in PiH includes force and vision.

Force feedback is commonly used to monitor contact force and recognize environmental uncertainties. As shown in Fig. 1.1a, an external Force-Torque (F/T) sensor is amounted to the industrial robot wrist to obtain wrench signals (forces and torques) [158]. Wrench signals can also be estimated based on joint currents [108]. For compliant robots, joint torque sensors have been utilized to calculate wrench signals at the end-effector [74, 115]. Some researchers have embedded force sensors in grippers or dexterous hands [9, 14, 109]. A table force sensor has been applied to guide PiH insertion [101, 161].

Vision is commonly used to capture the pose (position and orientation) of holes and pegs (see Fig. 1.1a). 2D cameras are widely utilized for coarse localization of holes by extracting their boundaries from images [92]. Marker points captured by 2D cameras are used to calculate the pose of pegs [149]. Image-based visual servo systems have been designed to track the relationship between pegs and holes [82, 110]. Position-based visual servo systems have been improved to enable high-speed microscale PiH assembly [48].

In contrast to 2D cameras, depth sensors [2, 108, 151] have been applied to capture the 3D geometry information. The obtained depth information is advantageous for estimating the spatial relationship between mating parts. In recent years, the cost reduction of 3D measurement systems and the rapid development of 3D vision and deep learning have significantly improved the efficiency and robustness of robotic assembly based on 3D vision, making it a new mainstream solution. In this monograph, we mainly focus on **multi-view stereo (MVS)** in general environment.

1.2.4 Assembly Strategies

Assembly trajectories are generally generated by a high-level planning module, which takes into account task specifications and assembly knowledge. Subsequently, low-level controllers are employed to track the planned trajectories and compensate for uncertainties arising from the environment or the robot. Since these low-level controllers have matured in implementing accurate motion and force tracking. The assembly strategies for PiH tasks can be categorized based on the high-level planning module.

With advancements in robotic techniques and task specifications, the strategies for robotic PiH assembly can be broadly classified into three types: **model-based strategies**, **learning from demonstrations (LFD)** [119], and **learning from environments (LFE)** [70]. It should be noted that the term "model" within the context of **model-based strategies** pertains to the process of modeling the contact status between the peg and the hole. In the subsequent chapters of this monograph, we will explore these assembly strategies in detail, focusing on their application to various PiH tasks.

1.3 Literature Review

In this section, we present a literature overview of MVS, tactile sensor simulation, model-based strategies, LFD and LFE. We organize the review in the aforementioned order to provide a coherent and structured presentation of these topics.

1.3.1 Multi-view Stereo

The geometry information of the peg and the hole is critical for assembly strategy. MVS aims to reconstruct the dense geometry of the 3D object from a sequence of images and corresponding camera poses and intrinsic parameters. MVS is a classical problem that had been extensively studied before the rise of deep learning. A number of 3D representations were adopted, including volumes [23, 44, 45], deformation models [25, 127, 165], and patches [8, 31, 39], which were iteratively updated through multi-view photo-consistency and regularization optimization.

Learning-Based Multi-view Stereo

Inspired by the success of deep learning in image recognition tasks, researchers began to apply learning techniques to stereo reconstruction tasks for better patch representation and matching [37, 69, 122]. Although these methods, in which only 2D networks were used, had made a great improvement on stereo tasks, it was difficult to extend them to MVS tasks, and their performance was limited in challenging scenes due to the lack of contextual geometry knowledge. Concurrently, 3D cost volume regularization approaches were proposed [59, 64, 65], where a 3D cost volume was built either in the camera frustum or the scene. Next, the multi-view 2D image features were warped into the cost volume, so that 3D CNNs could be applied to it. The key advantage of 3D cost volume is that the 3D geometry of the scene can be captured by the network explicitly, and the photo-metric matching can be performed in 3D space, alleviating the influence of image distortion caused by perspective transformation, which makes these methods achieve better results than 2D learning-based methods.

More recently, Luo et al. [86] proposed to use a learnable patchwise aggregation function and apply isotropic and anisotropic 3D convolutions on the 3D cost volume to improve the matching accuracy and robustness. Xue et al. [163] proposed Multi-view StereoCRF, where Multi-Scale Conditional Random Fields are adopted to constraint the smoothness of depth prediction explicitly.

High-Resolution and Hierarchical Multi-View Stereo

High-resolution MVS is critical to PiH assembly. Traditional methods [31, 77, 104] generated dense 3D patches by expanding from confident matching key points repeatedly, which was potentially time-consuming. These methods were also sensitive to noise and change of viewpoint owing to the usage of hand-crafted features. Recent learning methods tried to ease memory consumption by advanced space partitioning [116, 142, 150] or by replacing 3D CNNs with RNN [164]. However, most of these methods constructed a fixed cost volume representation for the whole scene, lacking flexibility.

Hierarchical MVS generates high-resolution depth maps in a coarse-to-fine manner, which reduces unnecessary computation and leads to improved efficiency. For classic methods, hierarchical Mutual Information computation was utilized to initialize and refine disparity maps [41, 117]. Learning-based methods were proposed to predict the residual of the depth map from warped images [105] or by constructing cascade narrow cost volume [33, 169].

1.3.2 Tactile Sensor Simulation

The simulation process of tactile sensors is typically divided into two phases. The first phase involves the simulation of the sensor's deformation caused by contact. The second phase involves the simulation of the transduction of physical quantities from the deformation. For marker-based visuotactile sensors, the first phase is a fundamental step for the second phase. This subsection primarily summarizes methods related to the first phase. Tactile sensors that operate on other principles, such as BioTac [30], share a similar first phase as marker-based visuotactile sensors and are also included in this subsection.

In [32, 168], a Gazebo-based GelSight simulator was proposed, where the deformation was simulated from the contact geometry using Gaussian filtering followed by difference of Gaussian. In [18, 81, 152], PyBullet was used to simulate the contact between the object and the sensor, and the deformation was approximated by calculating the penetration of the object into the sensor. A similar method is also reported in [3]. The focus of these works is the simulation of sensors' optical properties. Despite their computational efficiency, these geometry-based methods cannot model the sensor's tangential deformation, which is crucial for many robot applications such as slip detection and peg-in-hole insertion. Xu et al. [159] proposed a penalty-based tactile model upon rigid body dynamics, which is able to simulate both normal and shear tactile force fields at high speed. While the penalty-based simulation can approximate the sensor deformation in manipulation, it cannot simulate the elastic behavior of the elastomer accurately, especially the contact force caused by the tangential deformation. Therefore, its Sim2Real transferability is limited, which has been demonstrated by our experimental results.

Compared with rigid-body-based simulation, Finite Element Methods (FEM) can model the deformation of the sensor's elastomer more accurately. Bi et al. [5] developed a FEM-based tactile sensor simulator and achieved zero-shot Sim2Real transfer of RL policies for aggressive swing-up manipulation. However, they utilized the cylindrical geometry of the poles to simplify the simulation, and constrained the motion of the pole in the *x*-*y* plane, rendering it inapplicable to objects with various geometries. Si and Yuan [124] proposed a superposition method to approximate the FEM dynamics and successfully simulate the sensor's tangential deformation, but no manipulation tasks were demonstrated. Narang et al. built a linear-FEM-based tactile simulator for BioTac with Isaac Gym [96], achieving faster speeds than the commercial FEM software (ANSYS) [97]. Recently, Luu et al. employed SOFA [28] to build a simulator for large-scale marker-cum-vision-based-tactile sensor [87]. However, both [87, 96] primarily used their simulators to collect supervised datasets for interpreting tactile signals, leaving the potential of using simulation to train manipulation policies unexplored.

1.3.3 Model-Based Strategies

Model-based strategies employ contact state recognition to generate assembly motion trajectories. Hence, constructing an accurate contact state model becomes crucial. In the 1990s, several analytical models were proposed for contact state recognition. However, with the growing amount of data and task complexity, statistical approaches have been researched to identify contact states from assembly data.

Analytical Contact Model

Contact states can be recognized relying on the analysis of the geometrical and environmental constraints [153]. It can be implemented via two stages: contact state modeling and contact state determining [157].

Firstly, contact states can be modeled according to the mating features. Commonly used features include geometrical information and topology information [121]. Desai et al. [22] classified the contact states as the set of Elemental Contacts with same topology information. Xiao et al. [138, 156] introduced the concept of Principal Contacts for topological representation of contact states and generated the graph to represent the transition of contact states. Bruyninckx et al. [11, 12] built the kinematic models based on the constraints analysis for compliant motion in the presence of uncertainties [89]. Ohwovoriole et al. [103] defined three screw pairs to represent the contact constraints based on the screw theory. Hirai and Asada [40] extended the theory of polyhedral convex cones to solve manipulation problems governed by unidirectional constraints.

Secondly, the contact states can be determined or verified depending on the force/torque information [4]. McCarragher and Asada [91] developed a neural network-based qualitative recognition method to identify the transition of contact states quickly, where the dynamics effects were analyzed with the Petri Net. In an edge mating task, the contact states were detected via a singular value decomposition technique depending on analyzing the force/moment signals, which has shown better robustness to noise [68]. Geometric interpretation based on polyhedral convex cones was utilized to determine the likelihood of each feasible contact formation even without accurate contact force [27]. To reduce the redundancies to build the static equilibrium equations, a part knowledge-based system was developed to identify the contact states [144].

In summary, the contact constraints can be analyzed at a certain level [22]. However, the environmental uncertainties (e.g., part poses, rigidity or elasticity of the assembly components, and friction model) always result in challenges to recognize the contact state exactly. Additionally, the analytical analysis is sensitive to uncertainties, and it is impossible for any perfect model to account for all uncertainties. Moreover, the analytical analysis relies solely on past contact states, which limits its ability to generalize to new situations.

Statistical Contact Model

Statistical analysis methods often model the assembly trajectory as several segmented phases with different contact states [134]. Contact state recognition via statistical analysis is often formulated as a classification problem given possible contact states [93]. Statistical strategies can recognize the contact states from the pattern of collected samples without any assumption [51]. Commonly used statistical techniques for contact states consist of non-parametric classification methods, e.g., Fuzzy Classifier (FC) and Random Forest (RF); and parametric classification methods, e.g., Neural Network, Support Vector Machine (SVM), Gaussian Mixture Model (GMM), and Hidden Markov Model (HMM).

FC was applied in recognizing contact states via accommodating the uncertainties based on task knowledge [111, 128]. To enhance the robustness of the fuzzy system, the Gravitational Search (GS) was used to adjust the fuzzy rules [55]. GS-FC could solve the simple classification with little computing time through giving more accurate logic rules. Additionally, RF [16] and binary Stochastic Gradient Boosting [15] with strong classifier diversity were explored for multiple classification problems.

NN was researched to map the nonlinear relationship between the force information and contact states [10]. Two NN-based efficient classifiers combining with FC had achieved better sample efficiency performance in terms of the environmental noise [128, 130]. SVMs could reduce the risk and controlling the confidence interval for correct classification. It was suitable and applicable to design a practical contact states recognition method or real-world tasks [51, 53]. A Fuzzy Inference Mechanism with an adaptive classifier boundary was proposed to improve the robustness of the contact states classification [52].

GMM was employed to encode the states of PiH assembly, and Expectation Maximization (EM) had been demonstrated to optimize the parameters of GMMs. Bayesian classification had been incorporated to estimate a binary classification of the given GMMs [56, 58]. Jasim et al. [58] utilized the Distribution Similarity Measure to determine the optimal number of GMM components. HMMs could recognize both the contact states and state transitions via considering the temporal information [38]. It was suitable for dynamic assembly with sensor noise because the parameters of HMMs can be learned from empirical data [46]. A practical contact states identification system based on HMMs was implemented via incorporating with the prior spatial relationships of contact formations [21, 72].