University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Xia, Yu

Title:
Robotic Haptic Object Identification through Grasping

Date:
2023-05

Persistent Link:
https://hdl.handle.net/11343/337747

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

Robotic Haptic Object Identification through Grasping

by

Yu Xia

ORCID: 0000-0003-1120-3164

Doctor of Philosophy

October 2023

in the

Department of Mechanical Engineering

Faculty of Engineering and Information Technology

THE UNIVERSITY OF MELBOURNE

A thesis submitted in total fulfillment for the degree of Doctor of Philosophy

Abstract

Robotic haptic object identification is the process of identifying objects out of a given object set using a robotic hand equipped with tactile and finger-joint displacement sensors. Efficiency and accuracy, two essential evaluation metrics in haptic object identification, are the focus of this thesis. In terms of identification efficiency, when the robotic hand is smaller than an object, multiple grasps are required to capture the whole information of the object. However, from the practical consideration for robotic haptic object identification, it is always preferred to have the least number of grasps to identify an object. Regarding identification accuracy, when taking measurements by grasping the object, the uncertainties in the pose of the object relative to the hand will affect the identification accuracy. Each tactile sensor can capture contact within their specific local areas, thus, any change in the positions where objects make contact in relation to the robotic hand will significantly affect the tactile measurements.

This thesis, therefore, aims to address the issues proposed above. The contributions of the thesis are:

- An information gain-based method is proposed to improve the efficiency of haptic object identification by determining where to grasp to obtain the most distinguishing information about the object, thereby minimising the number of grasps needed;
- A statistical method based on the Beta mixture model is proposed to improve the accuracy of haptic object identification by systematically characterising the uncertainties in the haptic measurements caused by the deviation in the relative pose between the object and the end-effector.

Declaration of Authorship

I, YU XIA, declare that this thesis titled, 'Robotic Haptic Object Identification through Grasping' and the work presented in it are my own. I confirm that:

- The thesis comprises only my original work towards the Doctor of Philosophy except where indicated in the preface;
- due acknowledgement has been made in the text to all other material used; and
- the thesis is fewer than the maximum word limit in length, exclusive of tables, maps, bibliographies and appendices as approved by the Research Higher Degrees Committee.

Signed: _____

Date: _____

Preface

The work presented in Chapter 3 was published by **Frontiers in Robotics and AI** in July 2021.

The work presented in Chapter 5 was published by **IEEE Transactions on Mechatronics** in August 2022. This work was presented at the 2022 ASME International Conference on Advanced Intelligent Mechatronics (AIM).

Details of two published journal papers are as follows:

1. Y. Xia, A. Mohammadi, Y. Tan, B. Chen, P. Choong, and D. Oetomo, "On the efficiency of haptic based object identification: Determining where to grasp to get the most distinguishing information", Front. Robot. AI, vol. 8, p. 231, 2021. ([Chapter 3])
2. Y. Xia, A. Mohammadi, L. Peng, Y. Tan, B. Chen, P. Choong, and D. Oetomo, "Beta Mixture Model for the Uncertainties in Robotic Haptic Object Identification", IEEE/ASME Trans. Mechatron., 2022. ([Chapter 5])

In the first paper, the estimated contribution of the first author is 60%. The authors' contributions are as follows:

- YX, AM, YT, and DO: literature, experiment, data analysis, and paper drafting.
- BC and PC: experiment design, and paper review.

In the second paper, the estimated contribution of the first author is 60%. The authors' contributions are as follows:

- YX, AM, YT, and DO: literature, experiment, data analysis, and paper drafting.
- BC and PC: experiment design, and paper review.
- LP: data analysis, and paper review.

Acknowledgements

My PhD study commenced in 2019, presenting both difficulties and thrilling moments. Throughout my PhD study, I have received invaluable assistance and support from a multitude of individuals.

First and foremost, I would like to express my sincere appreciation to my supervisors, including Prof. Denny Oetomo, Prof. Ying Tan, Dr. Alireza Mohammadi, and A/Prof. Bernard Chen. Their constant encouragement and guidance have been a tremendous source of inspiration to me, especially during the period when I changed my research topic. I feel very fortunate to have such an approachable group of supervisors during the past four years.

I would also like to extend my gratitude to my lab colleagues, with whom I have shared many ideas, discussions, and activities such as sports and travelling. In early 2020, the COVID-19 pandemic caused Melbourne to enter into lockdown, compelling us to work from home. It was not until 2022 that things began to slowly return to normal. Despite the challenges brought about by the pandemic, we were able to support each other and maintain our motivation even through virtual meetings. I hope we will have more opportunities to collaborate in the future.

Additionally, I would like to thank Valma Angliss Trust for the financial support, which has been instrumental in allowing me to complete my studies. Moreover, I am grateful to all the individuals who have contributed to my project and provided me with valuable suggestions throughout my PhD study.

My family has played an essential role in helping me overcome difficulties during this journey. They have given me the confidence and emotional support necessary to persevere through challenges. I would like to express my special thanks to my wife, who has been my partner for the past ten years and has been my constant source of support. Lastly, I want to express my gratitude to my 9-month-old baby, who has been a precious gift to our family. I hope he grows up to be a kind and compassionate person.

Contents

Abstract ii

Declaration of Authorship iii

Preface iv

Acknowledgements v

List of Figures ix

List of Tables xi

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Problem Statement . . . 2
- 1.3 Contributions and Structure of the Thesis . . . 3

2 Literature Review 6
- 2.1 Haptic Exploratory Procedures . . . 6
  - 2.1.1 Exploratory procedures . . . 7
  - 2.1.2 Comparison of different exploratory procedures . . . 8
- 2.2 Haptic Measurement Collection Hardware . . . 9
  - 2.2.1 Information collection hardware . . . 10
  - 2.2.2 Comparison of haptic information collection hardware . . . 12
- 2.3 Haptic Object Identification Algorithms . . . 13
  - 2.3.1 Object identification approaches . . . 13
    - 2.3.1.1 Local information based identification . . . 13
    - 2.3.1.2 Global information based method . . . 16
  - 2.3.2 Efficiency of haptic object identification with grasping . . . 17
  - 2.3.3 Accuracy of haptic object identification with grasping . . . 18
- 2.4 Research Objectives . . . 21

3 Efficiency of Haptic Object Identification 23
- 3.1 Preliminaries . . . 24
  - 3.1.1 Min-max normalisation . . . 24
  - 3.1.2 K-Means based clustering technique . . . 24

3.1.2.1 Introduction on K-Means . . . 25
3.1.2.2 The elbow method . . . 27
3.1.3 Information gain . . . 27
3.1.4 Procedure of haptic object identification . . . 29
3.1.4.1 Object dataset construction . . . 30
3.1.4.2 Object identification . . . 30
3.2 Proposed Methodology . . . 30
3.2.1 Object dataset construction . . . 31
3.2.1.1 Object data acquisition and normalisation . . . 31
3.2.1.2 Categorising the grasp clusters . . . 32
3.2.1.3 Constructing the object description . . . 33
3.2.2 Object identification . . . 34
3.2.2.1 Information gain-based approach (the proposed approach) 35
3.2.2.2 Baseline approach . . . 40
3.3 Experimental Evaluation . . . 41
3.3.1 Experimental setup . . . 41
3.3.1.1 Object set . . . 41
3.3.1.2 Experimental platform . . . 45
3.3.2 Object dataset construction . . . 47
3.3.3 Object identification . . . 48
3.4 Results and Discussion . . . 51
3.4.1 Object dataset construction . . . 51
3.4.2 Object identification . . . 53
3.5 Summary . . . 59

4 Efficiency of Simultaneous Haptic Identification and Localisation 60
4.1 Preliminary: Available Transformation . . . 60
4.2 Proposed Methodology . . . 61
4.2.1 Object dataset construction . . . 62
4.2.2 Object identification and localisation . . . 63
4.2.2.1 The proposed approach . . . 63
4.2.2.2 Baseline approach . . . 71
4.3 Experimental Evaluation of the Object Identification and Localisation . . . 72
4.4 Results and Discussion of the Object Identification and Localisation . . . 75
4.4.1 Scenario 1 (The grasp is carried out without uncertainty) . . . 75
4.4.2 Scenario 2 (The grasp is carried out with uncertainty) . . . 77
4.5 Summary . . . 79

5 Accuracy of Haptic Object Identification 81
5.1 Preliminaries . . . 82
5.1.1 Beta mixture model . . . 82
5.1.1.1 Beta model . . . 82
5.1.1.2 Beta mixture model (BMM) . . . 82
5.1.1.3 Maximum likelihood estimation (MLE) . . . 84
5.1.2 Bayes' theorem . . . 85
5.1.3 Cramer-von Mises criterion . . . 86
5.1.4 Hypothesis testing . . . 86

5.1.4.1 Paired t-test . . . 87
5.1.4.2 Two-sample proportion test . . . 88
5.2 Proposed Methodology . . . 89
5.2.1 Object dataset construction . . . 89
5.2.1.1 Re-organising data into segments . . . 91
5.2.1.2 Determining the parameters in Beta mixture model . . . 91
5.2.2 Object identification . . . 92
5.2.2.1 Updating the posterior probability the object o_i . . . 92
5.2.2.2 Stopping rule . . . 93
5.3 Experimental Evaluation . . . 95
5.3.1 Experimental setup . . . 95
5.3.1.1 The object set . . . 95
5.3.1.2 Experimental platform . . . 98
5.3.2 Object dataset construction . . . 98
5.3.2.1 Data acquisition and normalisation . . . 98
5.3.2.2 Estimation of parameters of Beta mixture model . . . 99
5.3.3 Object identification . . . 99
5.4 Results and Discussion . . . 100
5.4.1 Object dataset construction . . . 100
5.4.2 Object identification . . . 101
5.5 Summary . . . 106

6 Conclusion **107**
6.1 Contributions . . . 107
6.2 Limitations and Future Work . . . 108

Bibliography **111**

List of Figures

2.1 Illustration of exploratory procedures [1]. 8
2.2 Illustrative example showing the tactile information captured by different types of tactile sensors when grasping a mug (The finger CAD model is referenced from [2]). 10
2.3 Illustrative example showing the impact of uncertain contact positions in the measurements (The finger CAD model is referenced from [2]). 19

3.1 Illustrative example of the elbow method. 28
3.2 The characteristics of an object are fully captured by grasping the object at N_g different poses. The N_g poses are spread, if possible uniformly, across the surface of the object. 32
3.3 An illustration of a 2D projection of the clusters formed by the M-dimensional normalised measurements (represented by each point on the plot) for all N_g poses of all ℓ objects in the dataset. The points are categorised into N_C grasp clusters. 34
3.4 The description of an arbitrary object in the object set. 34
3.5 Consists of 23 objects in a variety of shapes, sizes and stiffness. 44
3.6 A diagram of the finger of the Reflex Takktile hand showing the location of the tactile sensors. 45
3.7 A TakkTile sensor strip consists of 5 sensors end-to-end spaced at $8mm$ spacing [3]. 46
3.8 Experimental platform. **(a)** The experimental setup. **(b)** The schematic of the experimental setup (side view). 47
3.9 The number of grasps required to correctly identify each of 23 objects using the proposed approach and the baseline approach: Scenario 1 (without uncertainty). 54
3.10 The accuracy of identifying each of 23 objects through the proposed approach: Scenario 2 (with uncertainty). 57
3.11 The number of grasps required to correctly identify each of 23 objects using the proposed approach: Scenario 2 (with uncertainty). 58

4.1 The characteristics of an object are captured by grasping the object at eight uniformly distributed poses. 62
4.2 An illustrative example demonstrates the information that can be obtained for each option after every available displacement. 66
4.3 The number of grasps required to correctly identify each of 23 objects using the proposed approach and the baseline approach: Scenario 1 (without uncertainty). 76
4.4 The accuracy of identifying each of 23 objects using the proposed approach. 78

4.5 The number of grasps required to correctly identify each of 23 objects using the proposed approach and the baseline approach: Scenario 2 (with uncertainty). 79

5.1 Examples of the beta distribution with different values of shape parameters 83
5.2 The characteristics of an object are regarded to be captured by grasping the object at different poses (only three poses are shown). In this illustrative example, one segment contains three poses. 90
5.3 The list of the object set with 20 objects in a variety of shapes, sizes and stiffness. 97
5.4 The fitting result of the BMM ($K = 2$) and the Gaussian model. The histogram shows the readings from Tactile sensor 3 of Finger 1 in the first segment of Object 4 (a bottle of handwash). 100
5.5 The averaged identification accuracy of all objects in four methods and decreasing number of segments. 104

List of Tables

2.1 Haptic exploratory procedure 7
2.2 The summary of haptic information collection hardware 10
2.3 The summary of previously studied local information based haptic identification 14
2.4 The summary of previously studied global information based haptic identification 16

3.1 The object set 43
3.2 The list of the constructed labels for the object dataset used in the experiment 52
3.3 IG value for each object at each pose: first iteration 55

5.1 The object set 96
5.2 Experimental results: the identification accuracy for all 6 cases and 4 methods[1]. 102
5.3 Two-sample proportion test result comparing Gaussian and BMM on identification accuracy (one-tailed p-value)[1]. 104

Chapter 1

Introduction

1.1 Background

Object identification, which refers to the ability to identify a previously experienced object [4], is a common but essential skill in human daily life. As the surroundings are always composed of a variety of objects, without the ability to identify objects, the surrounding environments cannot be well perceived and interacted with [5]. In light of recent technological advancements, object identification has become progressively essential in numerous robotics applications, such as in-hand object manipulations [6, 7], industrial assemblies [8, 9], service robots [10] and simultaneous localisation and mapping (SLAMs) [11]. Motivated by its significance and broad applicability, object identification has become a popular research topic.

Identifying objects through haptic sensation is one of the major identification pathways [12, 13]. In this thesis, haptics refers to the study of touch sensing, specifically via kinesthetic and cutaneous receptors, associated with perception and manipulation [14, 15]. It involves not only the tactile information at the contact points/surface but also the proprioceptive information such as the pose of the fingers upon grasping the objects. Through haptic sensations, a person is able to perceive the physical and surface properties of objects, such as the volume, stiffness and texture of the object [16, 17]. Therefore, haptic object identification plays a crucial role when the identification process needs to rely on the information provided only through the physical interaction between

the end-effector and the objects, for example, identifying the objects with the same visual input but different object weight, surface texture or stiffness. Haptic identification is also important when the identification process cannot be conveniently achieved by other means (such as vision) [18], for instance, finding a specific object in a bag or a pocket. Although haptic sensing has been widely used in conjunction with other sensing modality/modalities in identifying objects (e.g., with vision) [19, 20], it is still very important to study haptic object identification in isolation to understand the extent of its capabilities fundamentally.

1.2 Problem Statement

This thesis focuses on how to improve the performance of object identification through haptic information obtained using a robotic hand. Accuracy is the foremost important performance requirement for object identification. For haptic object identification, efficiency is also important. Specifically, each touch or grasp may only provide partial information about the object. If the distinguishing information necessary to identify the object is not obtained from a single grasp, multiple grasps are required to extract sufficient information for identification [21]. This situation is often observed when the object is larger than the robotic hand, but it is also applicable when the hand is larger than the object. From the practical consideration for robotic haptic object identification, it is always preferred to have the fewest possible number of grasps to identify an object. Understanding which pose can provide the most distinctive information for object identification is crucial for determining where to grasp. This leads to the research question of how to determine where to grasp to improve haptic identification efficiency.

The second research question is related to the accuracy of haptic object identification. More specifically, it will answer the question of "how to improve the accuracy of haptic object identification?". In haptic object identification, the haptic information collected by grasping/touching objects has the problem that for the fixed locations of tactile sensors on the robotic hand/grasper, the measurements are dependent upon the contact pose of the object relative to the hand at grasping. In other words, the measurements of each sensor (embedded on specific locations the robotic hand/grasper used to perform the grasping and touching) will be susceptible to position deviation, which represents the difference in relative pose between the object and the end-effector. Given the high

likelihood of uncertainties in this relative pose, it introduces a significant source of uncertainties in the problem, potentially reducing the identification accuracy. Therefore, another research question is how to account for such uncertainties to improve haptic identification accuracy.

Based on the discussion above, two research questions have been proposed in the thesis:

- **Research question 1**: Is it feasible to design a haptic identification algorithm to improve the identification efficiency by determining where to grasp to get the most distinguishing information?
- **Research question 2**: Is it feasible to design a haptic identification algorithm to improve the identification accuracy by accounting for the uncertainties in the haptic measurements?

1.3 Contributions and Structure of the Thesis

This thesis, therefore, aims to address the questions proposed in the previous section. The contributions of the thesis are:

- **Contribution 1**: Improving the efficiency of haptic object identification by determining where to grasp to obtain the most distinguishing information of the object to minimise the number of grasps needed [Chapter 3 and Chapter 4];
- **Contribution 2**: Improving the accuracy of haptic object identification by systematically characterising the uncertainties in the haptic measurements caused by the deviation in the relative pose between the object and the end-effector [Chapter 5].

The results of this thesis are published in the following two journal papers:

1. Y. Xia, A. Mohammadi, Y. Tan, B. Chen, P. Choong, and D. Oetomo, "On the efficiency of haptic based object identification: Determining where to grasp to get the most distinguishing information", Front. Robot. AI, vol. 8, p. 231, 2021. ([Chapter 3])

2. Y. Xia, A. Mohammadi, L. Peng, Y. Tan, B. Chen, P. Choong, and D. Oetomo, "Beta Mixture Model for the Uncertainties in Robotic Haptic Object Identification", IEEE/ASME Trans. Mechatron., 2022. ([Chapter 5])

The structure of the thesis is constructed as follows:

In **Chapter** 2, the state-of-the-art techniques in haptic object identification are reviewed. This chapter starts with an overview of the haptic exploratory procedure. The haptic measurement collection setup is then summarised based on the type of information that can be obtained. At last, as the main focus of this thesis, the identification algorithms are categorised and compared in different dimensions. Moreover, this chapter identifies the challenges of existing haptic identification algorithms.

Chapter 3, which corresponds to Contribution 1, presents the methodology for the efficiency analysis of haptic identification. An object descriptor construction method to preserve the spatial haptic information of the object and an information gain method to calculate the amount of information at each pose that can be obtained to distinguish the object are proposed. Each subsequent grasp in the identification process will be carried out at the pose with the most distinguishing information, therefore, the number of grasps needed to identify the objects will be minimised. The result is evaluated by comparing the performance of the proposed method with a baseline method.

Chapter 4 also corresponds to Contribution 1. It focuses on solving the efficiency of haptic object identification when the pose position is unknown. When determining where to grasp, the proposed algorithm will output a spatial position of the next grasp relative to the current position (rather than directly determining which specific pose should be grasped). This work removes the assumption that is generally not valid in a practical scenario, thus this chapter improves the practical applicability of the resulting algorithm. The work in this chapter can be considered as the extension of the work in Chapter 3.

Chapter 5, which corresponds to Contribution 2, introduces the methodology for improving the accuracy of haptic object identification in the presence of uncertainties coming from the relative pose between the object and the end-effector. A statistical modelling method is utilised to account for the uncertainties in the haptic measurements, leading to a novel haptic object identification algorithm which is robust to these

uncertainties. The experimental results on identification accuracy are then presented and discussed.

Chapter 6 summarises the main outcomes of this thesis and introduces the possible future research directions.

Chapter 2

Literature Review

This chapter summarises the state-of-the-art in haptic object identification. Three topics about haptic object identification are discussed in detail: exploratory procedures, measurement collection hardware, and object identification algorithms. In the exploratory procedures section, different hand movement patterns are investigated. The measurement collection hardware section then reviews different hardware for obtaining the haptic information through tactile and joint sensors. With different exploratory procedures and data collection hardware, the types of haptic information that can be collected are different, resulting in the development of a wide range of haptic identification algorithms. Therefore, reviewing the first two topics can provide guidance in choosing the appropriate exploratory procedure and measurement collection hardware to be used in this thesis. The haptic object identification algorithms, which are the main focus of this thesis, are then reviewed and subsequently discussed based on the evaluation metrics of efficiency and accuracy. By identifying the challenges of the existing identification algorithms, the research objectives are summarised in the end.

2.1 Haptic Exploratory Procedures

Since haptic information can only be obtained by contacting the objects, it is essential to consider the exploratory procedure (EP) of haptic identification [19].

2.1.1 Exploratory procedures

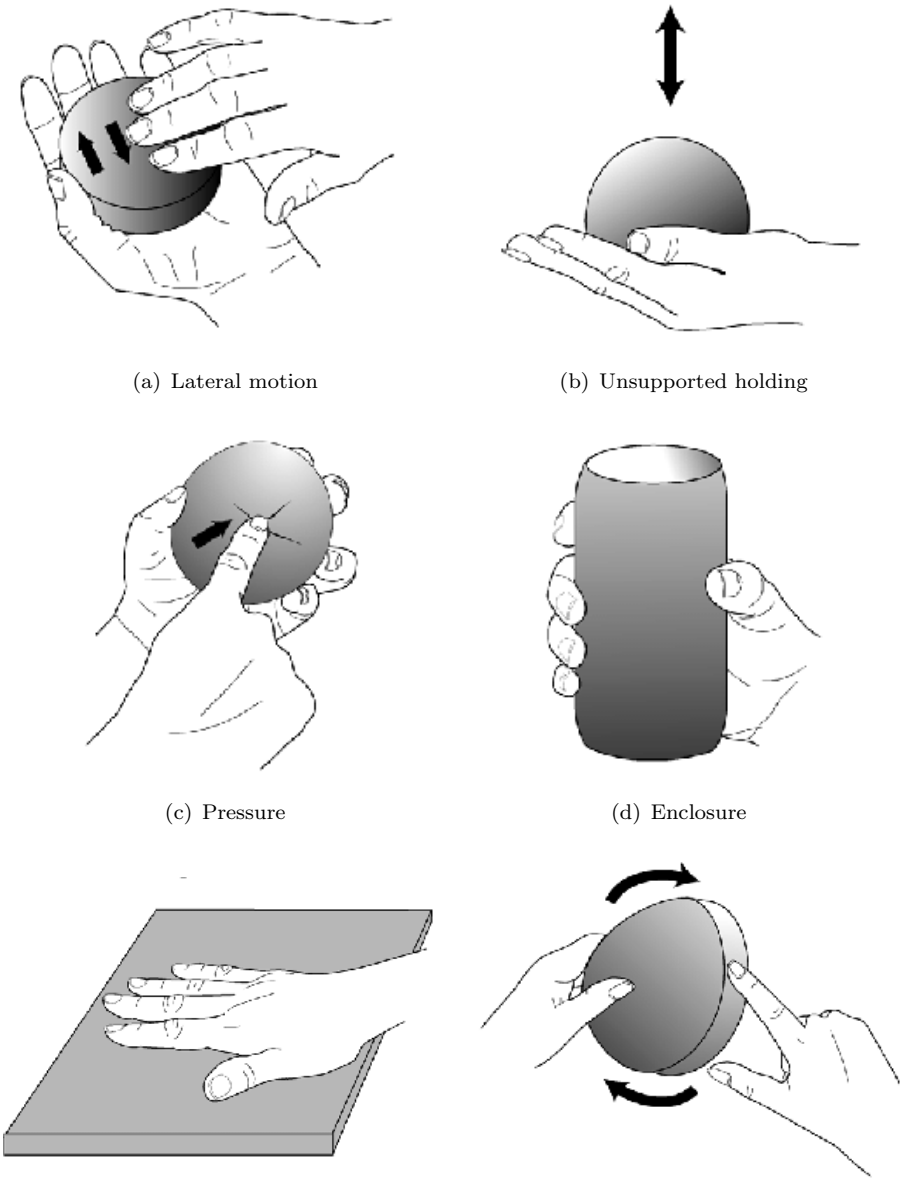
The EPs and the corresponding properties of the object are summarised in Table 2.1, and the graphic illustration of each EP is shown in Fig. 2.1. In general, the movement patterns can be categorised into six EPs [1], which are listed as follows:

1. In the lateral motion EP (Fig. 2.1(a)), the fingers of the hand will slide on the surface of the object to perceive the texture information.
2. Unsupported holding (Fig. 2.1(b)). This EP can be achieved by lifting the object up from the supporting surface. The weight of the object therefore can be obtained.
3. The pressure EP (Fig. 2.1(c)) represents the procedure of squeezing or pressing the object. By applying the forces or torque to an object, the stiffness of the object can be determined.
4. Enclosure (Fig. 2.1(d)). The hand will envelop the object to determine the global shape and volume of the object.
5. The static contact EP (Fig. 2.1(e)) can perceive the thermal conductivity or the temperature of the object.
6. Contour following (Fig. 2.1(f)). The hand will maintain contact with the contour of the object. Therefore, the exact/local shape can be obtained.

TABLE 2.1: Haptic exploratory procedure

Exploratory Procedure	Sub-properties of the object	Properties of the object
Enclosure	Global shape/Volume	Shape
Contour following	Exact shape	
Lateral motion	Texture	Material
Pressure	Stiffness	
Static contact	Temperature/Thermal conductivity	
Unsupported holding	Weight	

In the literature on robotic haptic object identification, different procedures are adopted. Specifically, [22–26] did the enclosure exploratory procedure (such as grasping or touching) with the objects, [27–32] performed the lateral motion on the surface of the objects to obtain the surface information, [30, 33, 34] involve pressing the objects to identify the stiffness related features of them, [28] and [30] also utilise the static contact and unsupported holding to collect the corresponding haptic information.



(a) Lateral motion (b) Unsupported holding

(c) Pressure (d) Enclosure

(e) Static contact (f) Contour following

FIGURE 2.1: Illustration of exploratory procedures [1].

2.1.2 Comparison of different exploratory procedures

When selecting the EP/EPs for haptic identification, the following factors might need to be considered:

1. The amount of information that can be obtained. The amount of distinguishing information that each EP can provide to identify an object varies depending on the scenario. However, in most cases, the enclosure exploratory procedure can provide

some rich information for object identification [35]. Besides, lateral motion can be informative when material information is required to differentiate between objects.

2. Implementability. The selection of the EPs requires the consideration of what manipulations the end-effector should do and if these manipulations can be achieved by the manipulator. The implementation of EPs such as the enclosure, the static contact, or the unsupported holding is relatively straightforward. For example, executing the closure exploratory procedure simply requires the end-effector to grasp or touch the object, which is only a fundamental requirement for a manipulator. In contrast, performing lateral motion or contour following EPs needs to consider the path planning for finger movement, and force control issues in identifying objects with different shapes.

Based on the above discussions, among the EPs, the enclosure is preferred as it is informative and relatively straightforward to execute [35]. Therefore, it will be adopted for this thesis where the object will be identified through grasping using a robotic hand, for haptic object identification.

2.2 Haptic Measurement Collection Hardware

In robotic haptic sensing, what haptic information (especially the tactile information) can be collected and how the haptic identification performs will be determined by the end-effectors and the types and numbers of tactile sensors used. Therefore, this section will introduce the haptic information collection hardware based on the tactile information that can be obtained. The concept of categorising hardware based on tactile sensor resolution, as introduced by [36], was employed. A categorisation criterion is established based on the extent of detailed local information that a robotic hand can obtain while grasping an object. An illustrative example can be seen in Fig. 2.2. The summary of the information collection hardware and the corresponding schematic illustrations are shown in Table 2.2.

TABLE 2.2: The summary of haptic information collection hardware

Tactile information	proprioceptive information	Reference
Single point contact	with	[37]
	without	[30], [38], [39]
Low resolution tactile imprint	with	[22], [23], [33], [40], [26]
	without	[41], [28], [42], [43], [44], [45], [46], [47], [48]
High resolution tactile image	with	[49], [50], [51]
	without	[52], [25], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63]

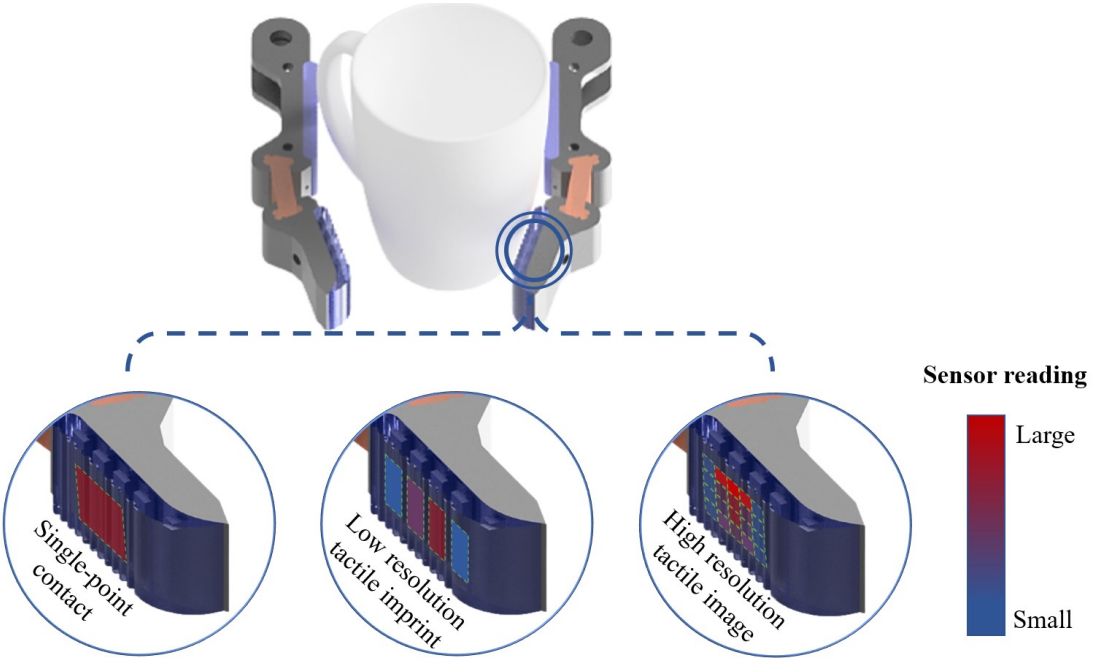


FIGURE 2.2: Illustrative example showing the tactile information captured by different types of tactile sensors when grasping a mug (The finger CAD model is referenced from [2]).

2.2.1 Information collection hardware

Hardware for collecting single point contact data

The single-point contact tactile sensor can determine the contact between the object and the end-effector and the contact force/pressure at the contact point. Several studies about haptic identification have been achieved by using this type of tactile sensor. Specifically, [30] adopted the OptoForce sensor (Berkeley Heights, NJ, USA), which can measure the 3-dimensional force at the contact point, to identify the textures, stiffness

and centre of mass of the objects. ATI Nano 17 sensor (Apex, NC, USA), a single point 6-axis force-torque sensor, is used by [37] to extract the surface properties of the object. Besides the mentioned sensors, [38, 39] also designed their own tactile sensors to measure the contact forces/pressures, and the experimental results showed that the data collected by them are able to distinguish the objects with different properties.

Hardware for collecting low-resolution tactile imprint data

The low-resolution imprint (the partial local information of the object at the contact area) can be obtained by the low spatial resolution tactile sensor array. Compared with the single-point contact tactile sensors, it can provide more detailed haptic information. Meanwhile, the low-resolution sensor array has a good response time, therefore, it gains the attraction of researchers. In the literature on haptic object identification, one commonly used sensor would be the BioTac sensor (Montrose, CA, USA) [28, 41–45]. It is capable to perceive the thermal conductivity and the signal vibration information between the object to be touched and itself, which is a benefit to the material identification. In addition, with 17 impedance sensing electrodes on board, the pressure at the contact points will be measured by the BioTac sensor. The uSkin sensors (Tokyo, Japan) is another low-resolution tactile sensor array designed for the Allegro hand [22, 23, 46]. A four-by-four sensor array was attached to the finger phalanx and each cell can provide the 3-dimensional force (both the normal and shear forces). The Takktile sensor array (Somerville, MA, USA), which is based on the MEMS barometers, can sense the pressure on each independent sensor with a spatial resolution of $8mm$. The Takktile sensor arrays are generally embedded on the Reflex Takktile hand, while a one-by-five array is on the proximal phalanx and a one-by-four array is on the distal phalanx [26, 33, 40]. Other low-resolution sensor arrays such as the PPS sensors and NeuTouch sensors are also used in their studies, which can be found in [47, 48].

Hardware for collecting high-resolution tactile image data

The high spatial resolution tactile sensor array can capture the majority of the local information of the object at the contact area. Weiss sensor (Ludwigsburg, Germany) comprises a grid of 14×6 sensing cells, with each cell's size being $3.4mm \times 3.4mm$ [25]. It is regarded as a high-resolution tactile array in this thesis because of its ability to capture local information about an object, despite having a lower resolution compared to other listed high-resolution tactile sensors. Weiss sensor has already been adopted by [25, 49, 50, 52] for haptic identification. Another popular commercial sensor is the TekScan array

model 6077 (South Boston, MA, USA) which contains 1400 tactels (also called taxels o sensels) with $1.3mm \times 1.3mm$ spatial resolution each [51, 53, 54]. Recently, the camera-based sensor has come to researchers' attention due to its outstanding performance. The camera-based sensor can capture the contact and the force based on the deformation of the membranes. Example prototypes are GelSlim [55], GelSight [59] and TacTip [56, 57]. Besides the mentioned commercial sensors, different high spatial resolution sensors can also be found in [58, 60–63]. For the high spatial resolution sensors, although the spatial resolution might vary, the information collected by these sensors can be treated as an image and hence can be directly processed using computer vision techniques. Details of the haptic object identification algorithm will be discussed in Section 2.3.

2.2.2 Comparison of haptic information collection hardware

What haptic information collection hardware to use is one of the factors that can affect the performance of haptic object identification. Generally, the following criteria need to be taken into consideration:

1. Spatial resolution of the sensor. With more detailed haptic information can be obtained, accurate identification is more likely. Therefore, high spatial resolution sensors are preferable.

2. Frequency response. It is important to consider the frequency response of the tactile sensor when identifying objects in real-time or when capturing the surface information of the objects by doing the lateral motion. In general, the single point contact tactile sensors have the best response performance as their output is of low dimension. In contrast, the high spatial resolution sensors will have a relatively low update rate. For instance, the frequency response of the ATI Nano 17 sensor and the TekScan sensor (high spatial resolution sensor) is around $7200Hz$ [64] and $200Hz$ [36], respectively.

3. Wiring and size of the sensor. The wiring of sensors should not affect the workspace of the robotic hand. Integrating a higher number of sensors into the end-effector might be challenging due to the wiring constraints [65]. Additionally, the size of the tactile sensor plays an important role and should be carefully considered. For

instance, camera-based sensors typically have larger sizes, making it challenging to integrate them into the fingertips of certain robotic hands.

In this thesis, by considering the issues mentioned above, the low spatial resolution array (Takktile sensor) is selected as the tactile sensor due to its comprehensive performance. Details of the hardware utilised in this thesis will be introduced in the next chapter.

2.3 Haptic Object Identification Algorithms

The purpose of haptic object identification algorithms is to use the haptic measurements collected by the information collection hardware as the input to predict what the object being identified is. In this section, different identification approaches proposed in the literature are listed and discussed.

2.3.1 Object identification approaches

To identify the objects, different algorithms are proposed in the literature. Generally, the haptic identification of the object can be achieved on two scales: local and global. The former method relies more on the local distinguishing information to identify the objects in the set, while the latter focuses on the global characteristics of the object. Take a cup as an example, the cup's global characteristics could include its volume, the distribution of its colour, or the overall shape, while the local features could refer to the shape of its handle or the texture in specific regions. Details will be discussed in the following subsections.

2.3.1.1 Local information based identification

The identification by this method is usually performed after one single grasp at a local pose of the object. Methods for local information based identification can be classified into four categories: raw haptic readings based method, statistical modelling method, handcrafted feature based method and self-organising based method (as shown in Table 2.3).

TABLE 2.3: The summary of previously studied local information based haptic identification

Local information based method	Reference
Raw haptic readings based method	[33], [66]
Statistical modelling method	[28], [27]
Handcrafted feature based method	[67], [68], [62], [30], [37], [69], [41]
Self-organising based method	[51], [55], [22], [23], [53], [58], [56], [38], [61], [24], [48], [39], [57]

Raw haptic readings based method

This method directly utilises the raw haptic readings to represent the characteristics of the objects. In other words, each haptic reading will be considered as one feature of the object. This method is able to be applied to any type of haptic measurement, especially the low dimensional sparse haptic measurements (measurements collected by the single point contact tactile sensors or the low resolution tactile sensor arrays).

Statistical modelling method

Another method is to use statistical modelling techniques to perform haptic object identification from measured data [27, 28]. Given a set of collected haptic measurements, the statistical model can characterise the probability density distribution of measurements offline. This information can be utilised to recognise the patterns within this dataset. Consequently, this statistical model is used to interpret the likelihood that a testing measurement belongs to an object based on the prior knowledge of the dataset [70]. In the literature, the Gaussian model has been used to account for the tactile sensor noise [27, 28]. However, it should be noted that as the uncertainties in the haptic measurement are not always Gaussian distributed, therefore, how to select the model of measurements would be the major challenge of the statistical approach.

Handcrafted feature based method

In this method, the object will be presented with hand-crafted features predefined by the researchers. Different features were proposed in the literature. The first is features adapted from computer vision. This type of feature treats the readings from tactile sensors as an image. [67] utilised Speed-up Robust Features (SURF) to extract the image features. These features extract the key points that are insensitive to the scale,

translation or rotation of the image, therefore, the result of the haptic identification will be robust to the pose variance. Besides, [62, 68] calculated the image moment (such as Zernike moments and Hu's moment) as the features of the object. It is noted that since these types of features are adopted from computer vision, the haptic information collection hardware should be equipped with high resolution tactile sensor array or camera-based sensors. Features with physical meanings are another type of feature used in the literature [30, 37]. In this approach, the physical properties of the object, such as the texture, stiffness or friction coefficients, will be extracted to represent the object. Besides the mentioned features, [50] proposed the Iterative Closest Labelled Point (iCLAP) method for haptic object identification. This approach includes local tactile features, which are defined as distinct label numbers categorised through the use of the K-Means clustering technique, alongside kinesthetic features, represented by the 3D position of the hand. Other hand-crafted features such as the minimum or maximum readings of the tactile sensors [46, 69] or Fourier transform based features [41] were also adopted in the literature.

Self-organising based method

The self-organising based method is the method that can give the identification results by analysing the haptic measurements itself and does not require human involvement or any a priori knowledge of the data. Generally, this type of method refers to the deep learning methods (deep neural networks), which use multiple layers to progressively extract higher-level features from the raw haptic measurements. [24] proposed the multi-layers artificial neural network (ANN) with dropout to avoid the overfitting problem in neural networks. [22, 23, 38, 51, 53, 55, 56, 58, 61] considered tactile measurements as tactile images and employed convolutional neural network (CNN) based methods with various network structures and measurement feeding patterns for identification purposes. As tactile measurements were treated as images, high spatial resolution tactile sensors were always utilised to capture intricate details. [22] adopted recurrent neural network (RNN) to process time series haptic measurements and compared the performance with CNN. [48] utilised a spiking graph neural network (SGN) to process the haptic data. Compared to conventional neural networks, this model enables better exploitation of local taxel structures, which may exhibit high irregularity [48]. Besides, in the literature, unsupervised neural network learning techniques, such as self-organising maps (SOM) [39] and auto-encoder [57], are also utilised to decrease the dimensionality

of the measurements employed in the identification process.

2.3.1.2 Global information based method

To identify the object using this method, multiple grasps at different poses will always need to be carried out to collect the global characteristics of the object. There are two widely used global descriptors in the literature, which are the point cloud based method and the histogram based method (as shown in Table 2.4).

TABLE 2.4: The summary of previously studied global information based haptic identification

Global information based method	Reference
Point cloud based method	[71], [72]
Histogram based method	[49], [73], [25], [52], [26], [74]

Point cloud based method

The point cloud based method utilises the information of the object contour and treats each object as a set of data points in the 3-dimensional space. To identify the object, the end-effector will touch or grasp the object at different locations to get the contact locations in the space. By using the Iterative Closest Point (ICP) based algorithm, the obtained contact points will be compared with the point cloud of each object in the set, and the most similar object (with the minimum matching error) in the set will be predicted as the testing object [71, 72]. As the ICP algorithm is invariant to the translation and rotation of the object, the point cloud based method will be robust to the pose variance of the object. However, in cases where the end-effector is smaller than the object and the spatial resolution of the tactile sensors is relatively low, constructing the contact locations of the whole object will be time-consuming.

Histogram based method

Histogram is another commonly used global information based method which aims to utilise the overall distribution of the characteristics of the object at different contact locations to represent that object. Given a histogram, its horizontal axis contains all features used to represent an object, while the vertical axis is the frequency of occurrence of each feature counted at different local poses. To define the features of each object, different methods were proposed in the literature. One popular approach is the Bag of Words (BoW) which is initially an approach for natural language processing.

By collecting the information on the objects at all contact locations, the BoW model will categorise all the measurements or features and generate the "codebook" of haptic features. An object therefore can be represented by the different frequency occurrences of each haptic feature in the "codebook" (a feature occurrence vector). The BoW model has been used by [25, 49, 52, 73] for haptic identification. Besides the BoW model, [26, 74] used the distance between the contact points as the features of the histogram. It can be noted that since the histogram of the object is constructed based on multiple local information, therefore, features used in the local information based method can also be applied to the histogram based method. As the global information of the object is considered by the histogram based method, it has an appropriate degree of robustness for the identification. However, the histogram based approach does not preserve the association of the spatial information on the objects and the haptic information they yield, which affects the upper limit of its identification performance.

2.3.2 Efficiency of haptic object identification with grasping

Efficiency is one of the important evaluation metrics in object identification in general as well as in haptic object identification. As mentioned, in haptic object identification, multiple grasps of the object are often required to identify it from the given set of objects. Specifically, when characteristics of an object are fully captured by grasping the object at different poses, where poses are spread across the surface of the object, each grasp performed at a different local pose relative to the object will only obtain piece-wise partial information. Therefore, if the distinguishing information required to identify an object is not obtained by a single grasp, multiple grasps are required. However, it might be possible to identify an object without grasping all local poses exhaustively. This is also naturally observed in human efforts of object identification when relying on handling the object without vision [21, 50]. The efficiency of the identification process is therefore to do with identifying the object in as few numbers of grasps as possible.

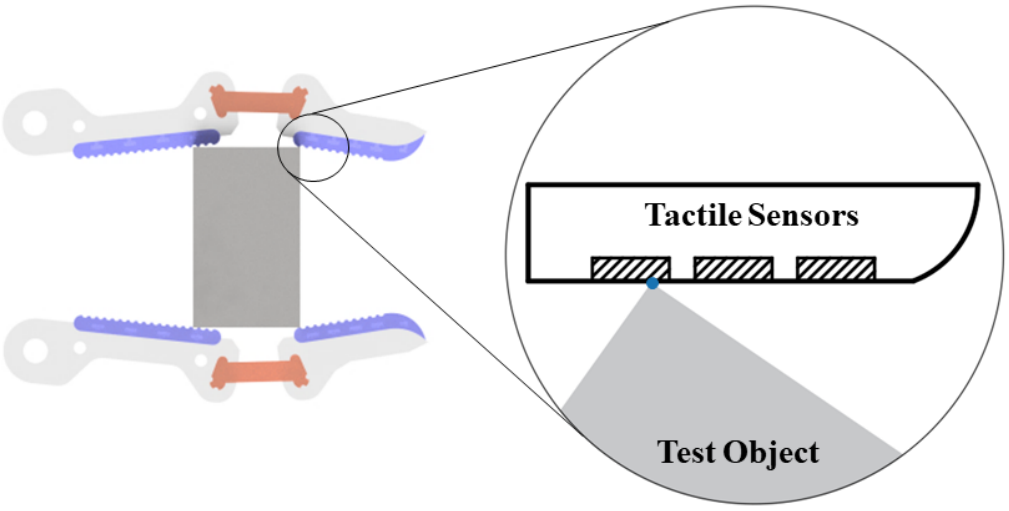
The identification efficiency of the global information based method is relatively low. As it focuses on the global characteristics of the objects, multiple actions are always needed to construct the corresponding descriptor. For example, multiple grasps are required to obtain the contact positions for the point cloud based method and to count the frequency of occurrence of each feature for the histogram based method.

The local information based method is potentially more efficient than the global information based method. However, relative to the identification accuracy, the efficiency of local information based haptic identification techniques has been less investigated [28, 30]. In [28], an efficient exploratory algorithm was presented specifically for texture identification. In [30], a method was presented to improve the efficiency of the learning process of object physical properties for the purpose of constructing object descriptions, not object identification. In addition to the above-mentioned works, other studies were prone to carry out the grasp on a random pose of the object. However, if the grasp is carried out on a pose without distinguishing information (e.g., it is not able to distinguish the mallet and the hammer with the same handle if the grasp is entirely on the handle), the efficiency of the identification will be reduced and the misidentification may occur.

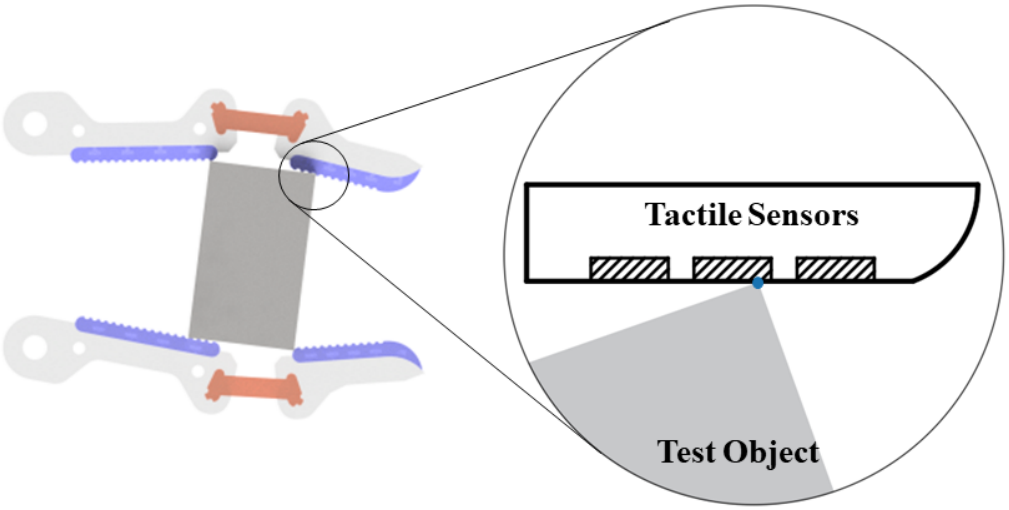
2.3.3 Accuracy of haptic object identification with grasping

Following the idea proposed in the last section, when selecting a pose for grasping, it is essential to accurately extract the information associated with that pose. However, for haptic object identification, the information collected by grasping/touching objects has the problem that for the fixed locations of tactile sensors on the robotic hand, the measurements are dependent upon the contact position of the object relative to the hand at grasping, such as shown in Fig. 2.3. In other words, the measurements of each sensor (embedded in specific locations where the robotic hand/grasper is used to perform the grasping and touching) will be susceptible to the deviation in the relative pose between the object and the grasper. Given the high likelihood of uncertainties in this relative pose, it introduces a significant source of uncertainties in the problem.

In terms of the global information based method, as it focuses on the global characteristics of the object rather than the specific local information, after multiple grasps, it can have the ability to resist uncertainties to some extent. For the point cloud based method, as long as the contact positions in the 3-dimensional space vary in a certain range, the overall distribution of an object point cloud would not be significantly affected. The robustness of the histogram based method to uncertainties depends on the type of features selected to represent the object.



(a) In the first grasp, the left tactile sensor is activated.



(b) In the second grasp, the middle tactile sensor is activated while the left tactile sensor is not activated.

FIGURE 2.3: Illustrative example showing the impact of uncertain contact positions in the measurements (The finger CAD model is referenced from [2]).

Different local information based methods have different characteristics. The raw haptic reading based method has the advantage of being applicable to all types of haptic measurements. However, the corresponding trade-off is that it is sensitive to uncertainties, which reduces the upper limit of identification accuracy of this method.

The statistical modelling method can characterise the distribution of the haptic measurement. It provides interpretable results from haptic measurements [70]. Besides,

this technique utilises all information measured from sensors and further reduces the subjective human involvement in the selection of appropriate features, making it more systematic. As mentioned, the challenge of the statistical approach would be the selection of the model of measurements. In the literature, the Gaussian model has been used to account for the tactile sensor noise [27, 28]. However, it is questionable whether a Gaussian model is able to cope with uncertainties. For objects other than perfect cylinders or spheres in shape, when grasped at a posture with uncertainties, tactile sensors on a given location of the robotic hand may or may not come into contact with the object. This is because multiple sensors are used to measure contact, placed adjacent to one another. If there is no contact pressure on a tactile sensor, then the readings of it would be zero. With a slight amount of uncertainties between the hand and the object, the resulting probability density function (*pdf*) of the measurement of a tactile sensor will peak around the "no contact" case and the "firm contact" case, resulting in two peaks. The Gaussian model, which is typically used in many existing algorithms, is not a good candidate to characterise the *pdf* of tactile sensors as it is symmetrically distributed and only works when most data are around the central peak.

For the handcrafted feature based method, with suitable features being constructed, it usually has good resistance to the uncertainties in the haptic measurements, in other words, good identification accuracy can be obtained. However, with different end-effectors and different types or numbers of embedded tactile sensors, the features that can be extracted are different. Under this situation (especially for the case where the acquired tactile information is very sparse), finding features, which keep the majority of information from sensors and are less sensitive to uncertainties, might be challenging sometimes [36].

Based on the powerful performance of deep learning, the identification performance of the self-organising method presented in the literature is always competitive. On the other hand, the self-organising method also has disadvantages that need to be taken into consideration:

- Parameters tuning. The performance of deep learning techniques relies on the selected structure and parameters, which usually are selected by trial and error. Using inappropriate parameters or structures might lead to poor identification performance as indicated in [36, 75].

- Interpretability. As a typical black-box identification technique, the results lack interpretability, which might limit the scope of the methods in some applications, such as medical, in which interpretability plays an important role in diagnosis as pointed out in [76].

- Dimensionality and quantity of the input data. According to the literature mentioned in Section 2.3.1, such as [22, 51, 53, 55, 58], the amount of data in the haptic dataset is much smaller than the visual one as the process of collecting data by the robotic hand is more complex and time-consuming. The deep learning method generally performs better in high-dimensional data with a huge amount of data. It might be questionable how well the neural networks perform on low-dimensional data (from the point based tactile sensor or low-resolution tactile sensor array) with less amount of measurements.

2.4 Research Objectives

Based on the discussion above, each grasp can provide only partial information about an object. If the distinguishing information of the object is not extracted by a single grasp, then multiple grasps are required for haptic object identification. Therefore, the main objective of this thesis is to develop an algorithm that can efficiently and accurately identify objects through grasping. For the global information based method, there is a trade-off between efficiency and accuracy. The global information based method can be robust to the uncertainties in the measurement once the global descriptor of the object to be identified can be fully constructed. However, it normally will take many numbers to grasp, which contradicts the requirement of efficiency. In contrast, the local information based identification method has the potential to achieve both efficiency and accuracy. However, its efficiency has not been thoroughly investigated, and therefore an algorithm for identification efficiency should be proposed. For accurate identification, to account for the uncertainties in the haptic measurements, this thesis will utilise the statistical modelling method as it has fewer limitations, e.g., it allows for interpretable results from physical measurements and reduces the need for subjective human involvement, making the identification more systematic. Therefore, the main objectives of this thesis are:

- Constructing the local information based descriptor of the object which preserves the spatial information and developing an identification algorithm to improve the efficiency (to minimise the number of grasps needed) of the haptic object identification.
- Proposing a local information based statistical modelling identification method that improves identification accuracy by accounting for the uncertainties in the measurements.

In this thesis, Chapters 3 and Chapter 4 will focus on achieving the first research objective, while Chapter 5 will be dedicated to addressing the second research objective.

Chapter 3

Efficiency of Haptic Object Identification

This chapter presents a systematic method for analysing and quantifying the efficiency of haptic object identification (determining where to grasp to minimise the number of grasps needed for identification). The process affects two stages of identification: the object dataset construction process and the haptic object identification process based on the information gain technique. An approach for object dataset construction is proposed to preserve the spatial information on the objects and the haptic information they yield. The clustering technique is employed to reduce the dimensionality and complexity of the object descriptions, such as in [73]. Based on the obtained spatial information (haptic object description), the haptic information associated with each pose on the object is available to the algorithm.

With the prior knowledge of the pose and the associated haptic information for each object obtained from the constructed object dataset, the proposed information gain-based technique calculates the amount of information at each pose that can be obtained to distinguish the object from the rest of the objects in the set. Efficiency in haptic identification is pursued in this algorithm by selecting the pose that contains the largest amount of information. It should be noted that in this chapter, during the object identification, we assume the pose of the object is known, and this assumption will be relaxed in Chapter 4.

3.1 Preliminaries

This section provides the preliminaries used in this chapter. In particular, the min-max normalisation method, the K-Means clustering technique (which is used to label the measurements), the information gain technique (which is used to calculate the distinguishing information of each pose), and the procedure of haptic object identification are introduced.

3.1.1 Min-max normalisation

Min-max normalisation will perform the linear transformation on the original data and scale them to the range $[0, 1]$. In this chapter, the min-max normalisation was utilised to scale the readings obtained by each specific sensor contacting different objects in the given set. Compared with other normalisation methods, such as Z-score normalisation and log transformation normalisation, it has the advantage that it can preserve the relationships and relative differences between data points [77]. For the n^{th} sensor, the formula of min-max normalisation is given as:

$$V_n' = \frac{V_n - \min(V_n)}{\max(V_n) - \min(V_n)}, \tag{3.1}$$

where:

- V_n is a set of original values from the n^{th} sensor obtained by grasping all objects;
- V_n' is the normalised values from the n^{th} sensor obtained by grasping all objects;
- $\min(V_n)$ is the minimum value in V_n;
- $\max(V_n)$ is the maximum value in V_n.

3.1.2 K-Means based clustering technique

Assuming that X is a set consisting of N normalised data points collected. Each each point $\mathbf{x}_i, i = 1, \ldots, N$ is obtained from the measurements of M sensors, i.e., $\mathbf{x}_i \in \mathcal{R}^M$, where $\mathcal{R}$ is the set of real numbers. We define $\mathbf{x}_i = \begin{bmatrix} x_{i,1} & x_{i,2} & \cdots & x_{i,n} & \cdots & x_{i,M} \end{bmatrix}^T$, where $x_{i,n}$ represents the reading of the n^{th} sensor of data $\mathbf{x}_i$. In this chapter, as

each entry of the measurements will be normalised with $[0, 1]$, it has $|\mathbf{x}_i|_\infty \leq 1$, where $|\mathbf{x}|_\infty = \max_{n=1,\ldots,M} |x_n|$ for any $\mathbf{x} \in \mathcal{R}^M$.

In haptic object identification, different unsupervised learning techniques have been proposed to label the measurement data collected from grasps to reduce the effort of labelling. Among them, the K-Means clustering technique is one of the most widely used [78].

3.1.2.1 Introduction on K-Means

The k-Means clustering technique is used to form clusters from the measured data. The data set X can form N_C clusters with the centre $\mathbf{x}_k^c, \forall k = 1, \ldots, N_C$. For simplicity of notation, we define an operator of K-Means as:

$$\mathcal{K}_{means} : \{\mathbf{x}_i\}_{i=1,\ldots,N} \rightarrow \mathbf{x}_k^c, k = 1, \ldots, N_C. \tag{3.2}$$

K-Means clustering technique usually minimises the within-cluster sum of the squared error (E) between each data point and its corresponding centroid $\mathbf{x}_k^c$, which is defined as follows:

$$E = \sum_{k=1}^{N_C} \sum_{i=1}^{N} u_{i,k} \cdot d(\mathbf{x}_i, \mathbf{x}_k^c)^2, \tag{3.3}$$

where the nonlinear mapping $d(\cdot, \cdot)$ is the distance measure. Usually, the Euclidean distance is used in K-Means, i.e.,

$$d(\mathbf{x}_i, \mathbf{x}_k^c) = \sqrt{\sum_{n=1}^{M} (x_{i,n} - x_{k,n}^c)^2}. \tag{3.4}$$

Some modifications of K-Means use different "distance" measures. For example, for the well-known weighted K-Means, the weighted distance is used, i.e.,

$$d^w(\mathbf{x}_i, \mathbf{x}_k^c) = \sqrt{\sum_{n=1}^{M} w_n \cdot (x_{i,n} - x_{k,n}^c)^2}, \tag{3.5}$$

where w_n is the weight assigned to the n^{th} feature or sensor (where $0 \leq w_n \leq 1$).

The variable $u_{i,k}$ is used to represent whether the entity $\mathbf{x}_i$ belongs to the k^{th} cluster, which is defined as:

$$u_{i,k} = \begin{cases} 1, \text{if } \mathbf{x}_i \text{ belongs to cluster } k, \\ 0, \text{if } \mathbf{x}_i \text{ does not belong to cluster } k. \end{cases} \tag{3.6}$$

The minimisation of E can be achieved by some iterative processes. It usually has three steps:

Assuming that there are N_C clusters. The first step randomly selects N_C data points from the dataset X as the initial centroids. Many different algorithms have been proposed on how to select the number of clusters [79].

The second step assigns each data point $\mathbf{x}_i$ to the closest centroid (hence to the closest cluster) based on the Euclidean or weighted distance. This can be represented as:

$$\lambda_i = \underset{k=1,\ldots,N_C}{\operatorname{argmin}} \; d(\mathbf{x}_i, \mathbf{x}_k^c)^2, \tag{3.7}$$

where $d(\cdot,\cdot)$ is defined in Eq. (3.4) or Eq. (3.5).

The third step updates the centroid of each cluster by using the following criteria:

$$\mathbf{x}_k^c = \text{mean}(\{\mathbf{x}_i \in X | \lambda_i = k\}), \tag{3.8}$$

which provides the average of the data in the cluster.

The second and third steps will be repeated until the change of E is less than a predefined threshold. The pseudo-code of the K-Means algorithm is shown below.

Algorithm 1: K-Means Clustering Algorithm

Input: Dataset $X = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$; The number of clusters N_C.;

Randomly selects N_C data points from the dataset X as the initial centroids;

Calculate the initial E using Eq. (3.3).

while *the change of* $E \geq$ *predefined threshold :* **do**

  Assign each data point $\mathbf{x}_i$ to the closest centroid using Eq. (3.7);

  Update the centroid of each cluster using Eq. (3.8);

  Update E using Eq. (3.3).

end

Onput: Clustered data and the centroids of each cluster;

3.1.2.2 The elbow method

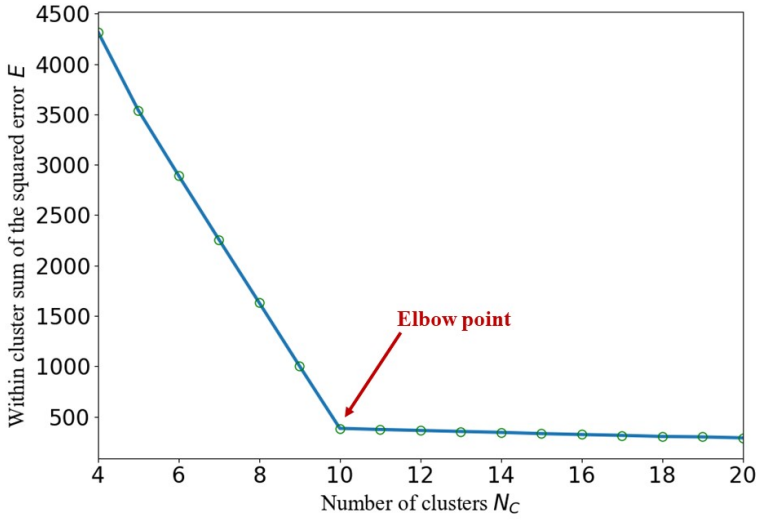
The elbow method is one of the commonly used methods to determine the optimal number of clusters N_C in K-Means [79]. It involves plotting the values of a metric (such as the within-cluster sum of the squared error) against a range of different numbers of clusters, and then identifying the point of diminishing returns, which is referred to as the "elbow point". This point represents the optimal number of clusters to use in a clustering analysis. Mathematically, the following four steps can be used to illustrate the elbow method:

1. Initialise K-Means algorithm for a range of N_C (number of clusters) values;
2. For each K-Means algorithm, once it converges, record the within-cluster sum of the squared error E (see Eq. (3.3));
3. Calculate the second order derivative $\ddot{E}$ with respect to N_C;
4. The N_C with the maximum $\ddot{E}$ will be selected as the optimal number of clusters.

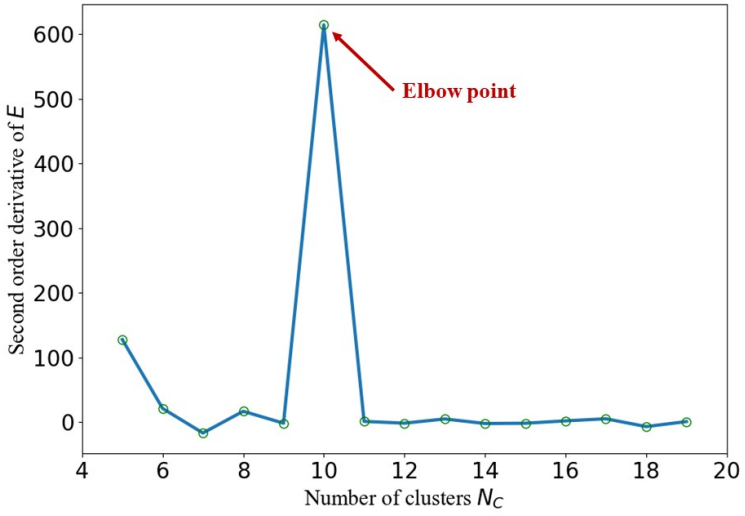
The illustrative example with the optimal number of clusters of ten is shown in Fig. 3.1.

3.1.3 Information gain

Information gain is a commonly used concept to represent the amount of information that is gained by knowing the value of a feature (variable). Specifically, IG assigns the most distinguishable feature with the highest information value. Currently, IG has been



(a) Within cluster sum of the squared error E versus number of clusters N_C.



(b) Second order derivative of E versus number of clusters N_C.

FIGURE 3.1: Illustrative example of the elbow method.

developed and applied to decision trees, simultaneous localisation and mapping (SLAM), feature selection techniques [80, 81] and in object shape re-construction by improving the exploration efficiency [82]. In this chapter, IG is utilised to measure the amount of distinguishing information contained in each pose.

For a random variable Y with N_A possible outcomes, y_a, $a = 1, \ldots, N_A$, assume that each outcome has a probability of $p(y_a)$. The *information entropy* $H(Y)$ is defined as:

$$H(Y) = -\sum_{a=1}^{N_A} p(y_a)\log_2(p(y_a)). \tag{3.9}$$

Conditional entropy [83] is then used to quantify the information needed to describe the outcome of a random variable Y given the value of another random variable Q. Assume that there are N_B possible outcomes of the random variable Q. Each outcome q_b of the random variable Q has a probability of $p(q_b)$, $b = 1, \ldots, N_B$. Let $p(y_a|q_b)$ be the conditional probability of y_a given q_b. $H(Y|Q)$ represents the information entropy of the variable Y conditioned upon the random variable Q, which can be computed as:

$$\begin{aligned} H(Y|Q) &= \sum_{b=1}^{N_B} p(q_b)H(Y|q_b) \\ &= -\sum_{b=1}^{N_B} p(q_b) \sum_{a=1}^{N_A} p(y_a|q_b)\log_2\left(p(y_a|q_b)\right) . \end{aligned} \tag{3.10}$$

Information Gain [83] represents the degree to which uncertainty in the information of Y is reduced when the random variable Q happens, which is defined as:

$$\begin{aligned} G(Y,Q) &= H(Y) - H(Y|Q) \\ &= -\sum_{a=1}^{N_A} p(y_a)\log_2(p(y_a)) \\ &\quad + \sum_{b=1}^{N_B} p(q_b) \sum_{a=1}^{N_A} p(y_a|q_b)\log_2\left(p(y_a|q_b)\right) . \end{aligned} \tag{3.11}$$

3.1.4 Procedure of haptic object identification

For a given set of objects O, defined as $O = \{o_1, ..., o_\ell\}$, $\ell \in \mathcal{N}$ where $\mathcal{N}$ is the set of integers, the aim of haptic object identification is to identify an object among possible objects in the object set. It should be highlighted that this thesis works on object identification [84], which involves identifying a specific object from a given object set, rather than the classification [85], which focuses on categorising a new object not contained in the object set into a specific class. This is currently the case as haptic is relatively less advanced than computer vision research and the information content in haptic is more

sparse. The procedure involved in haptic object identification includes object dataset construction and object identification.

3.1.4.1 Object dataset construction

To construct the dataset, each object in the set will be grasped in different poses multiple times depending on the tasks, and the corresponding haptic measurements of each object at different poses will be recorded in the dataset. The collected measurements then will be used for further processing to represent the characteristics of each object.

3.1.4.2 Object identification

In this step, the identification algorithm is presented with an object randomly selected from the object set O. By grasping the object, the haptic measurements will be acquired and subsequently processed to compare with the haptic characteristics of each object in the constructed dataset. The identification algorithm will output the predicted label of the tested object, and a binary identification result (correctly identified or misidentified) will be obtained by comparing the predicted label of the object with its ground truth label. The performance of the haptic identification will be evaluated by running the process multiple times and taking the average of the number of times the algorithm correctly identifies an object out of the total number of attempts.

Each object identification attempt can be an iterative process. During each iteration, once the haptic measurements are obtained, the identification algorithm will calculate the similarity between the object under test and each object o_i in the object set O. When the identification procedure satisfies the stopping rule, the algorithm will stop, and the label of the object in the set O with the maximum probability of similarity will output as the predicted label.

3.2 Proposed Methodology

The objective of the proposed methodology is to identify an object presented to the algorithm with the fewest number of grasps. The possible grasp poses (the location and

orientation that a robotic hand is used to grasp the object, relative to the object) are common for all objects.

The process includes the construction of the object dataset (that incorporates the grasp pose information to the corresponding haptic information) and the object identification. Once the object dataset is established according to the proposed method, the problem is to identify which object (out of the object set) is being presented to the algorithm.

3.2.1 Object dataset construction

The procedure can be presented in three steps: object data acquisition and normalisation, categorising the grasp clusters and constructing the object description. The details of each step are presented below.

3.2.1.1 Object data acquisition and normalisation

Haptic measurements for all objects are taken at N_g predefined poses Figure 3.2 shows an object with coordinate frames 1 to N_g representing the finite number of predefined poses where a robotic grasper with relevant tactile and proprioceptive sensors is placed to grasp the objects to gather measurements. These poses are specified only as the angle of approach to the object at each increment in object height (the example in Figure 3.2 shows 3 levels of object heights and the poses defined at every $\pi/2$ *rad* for each height). Therefore, after grasping at N_g poses, all parts of the object, at all heights and angular displacements measured from a reference frame attached to the object, have been encapsulated by the grasp of the robotic hand. In other words, the robotic hand would collect all the haptic information necessary to describe the object. Each grasp on the object will yield an M dimensional measurement $\hat{\mathbf{z}} = \left[\hat{z}^1, ..., \hat{z}^j, ..., \hat{z}^M\right]^T$, corresponding to the number of sensors on the robotic hand.

To account for the expected amount of uncertainty associated with the grasping process, in each pose, the grasp will be carried out N_T times, and each time the grasp orientation relative to the object will be randomly determined. These N_T data will capture the data distribution, enabling systematic analysis of the uncertainties present in the measurements. Due to the different dimensions associated with the different types of physical variables involved (joint displacement and tactile pressure), a normalisation method is

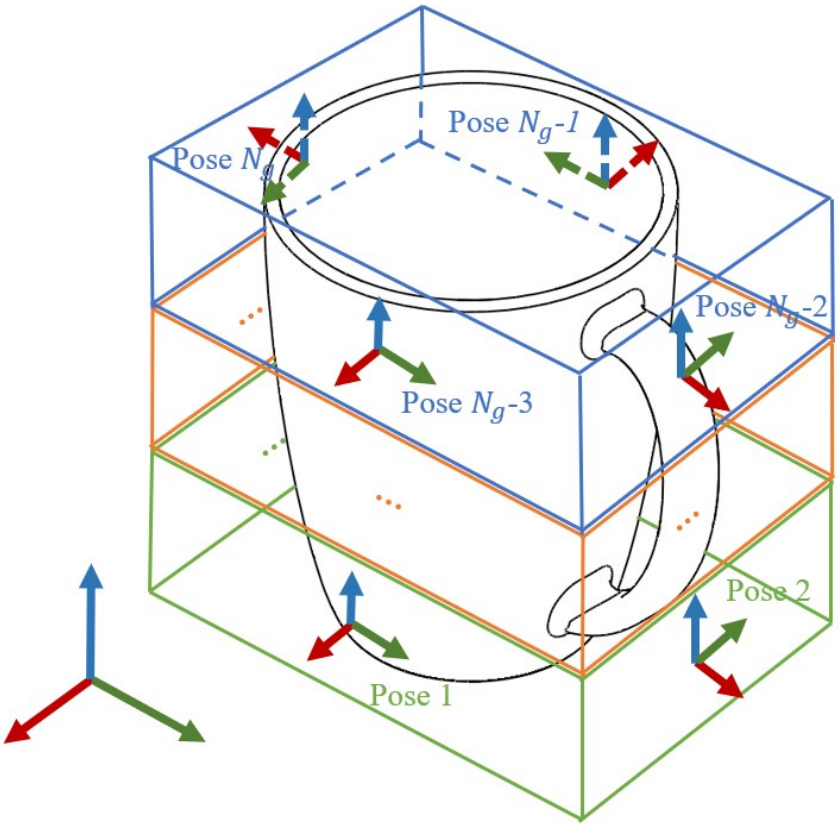


FIGURE 3.2: The characteristics of an object are fully captured by grasping the object at N_g different poses. The N_g poses are spread, if possible uniformly, across the surface of the object.

needed to remove the influence of the dimension and the unit. In this thesis, the min-max normalisation method (as introduced in Section 3.1.1), which scales the values of all variables to the range of zero and one, is employed.

In summary, for ℓ objects in the object set, the dataset consists of $\ell \times N_g \times N_T$ total measurements. Each measurement is a M-dimensional vector, representing M sensors used. The data can then be further processed with different algorithms to represent the characteristics of the objects according to the different task requirements.

3.2.1.2 Categorising the grasp clusters

An unsupervised learning technique is utilised to cluster all the M dimensional normalised measurements for all N_g poses in an object for all objects in the dataset into N_C grasp clusters. In this thesis, the well-established K-Means approach is adopted due

to its high-speed performance. To determine the number of clusters, the elbow method is used (as introduced in Section 3.1.2.2).

As an illustrative example, a 2-dimensional plot representing the clustering results is shown in Fig. 3.3. Each point on Fig. 3.3 represents one M dimensional normalised measurement (for ease of reading, the M dimension plot is conceptually represented in a 2-dimensional plot). Therefore, in order to focus on the efficiency of the algorithm, it is assumed that the object dataset is well-designed with distinguishable objects and noise/uncertainty in the haptic measurements are bounded in a range that is clearly smaller than the variance in the distinguishing information between distinct grasps. Therefore, grasping an object at a given pose repeated multiple times will result in measurements that should all be categorised into the same cluster, which guarantees the identification accuracy can be decoupled from the efficiency. Furthermore, for each cluster, the spread of measurements will be completely bounded (represented by the red circles in Fig. 3.3). Besides, the minimum distance between centroids of two clusters (clusters i and j), denoted as $d_{i,j}$ (inter-cluster distance), should be larger than the sum of two radii representing the data spread within each cluster, denoted as $\sigma_i + \sigma_j$ (intra-cluster distance). Mathematically, $\frac{d_{i,j}}{\sigma_i+\sigma_j} \geq 1$. It should be highlighted that this assumption is only used for object dataset construction. During the object identification step, practical uncertainties are included in the test, where the effect of such uncertainties is evaluated in the experiment.

3.2.1.3 Constructing the object description

After categorisation, an object can be labelled uniquely by the string (of length N_g) for the N_g poses across the object. Each pose $p_1, ..., p_{N_g}$ is assigned a cluster number, which is the cluster that the haptic measurements at that pose for that object have been categorised into. An example of the label of an object in the set is represented in Fig. 3.4. In this example, when grasped at Pose 1, the object yields haptic measurements across the M sensors on the robotic grasper that has been categorised into grasp cluster number 7 through the clustering process. The resulting object description, therefore, preserves the information of which pose yields which grasp measurement, which is now represented as the grasp cluster number.

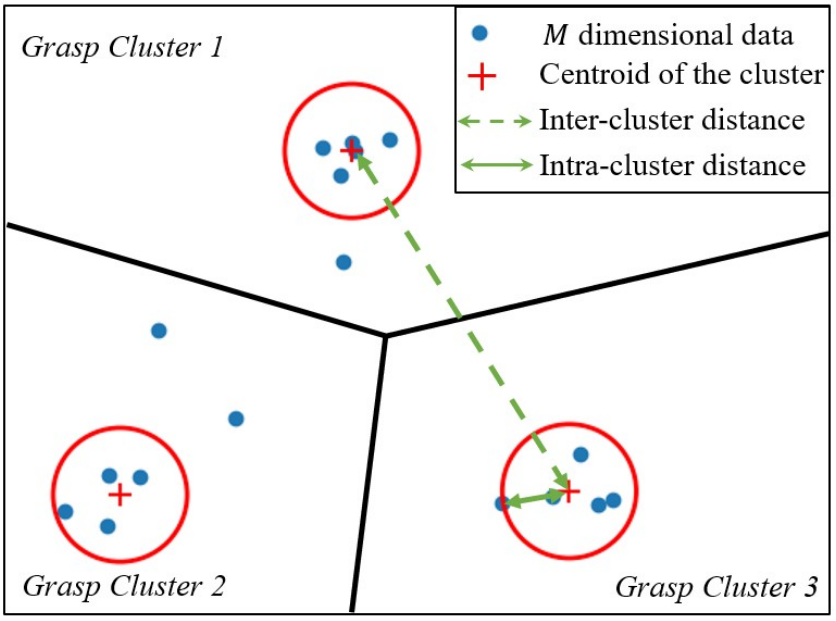


FIGURE 3.3: An illustration of a 2D projection of the clusters formed by the M-dimensional normalised measurements (represented by each point on the plot) for all N_g poses of all ℓ objects in the dataset. The points are categorised into N_C grasp clusters.

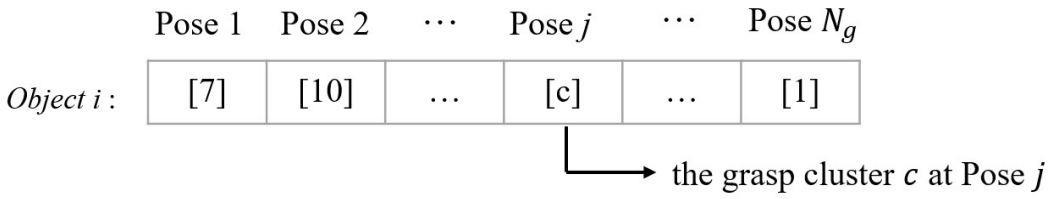


FIGURE 3.4: The description of an arbitrary object in the object set.

3.2.2 Object identification

The proposed object identification approach utilises the information gain-based (IG-based) technique. In this thesis, it will be compared to a baseline approach, where the information preserved in each pose of objects is not utilised. It is noted that any objects to be identified should belong to the object set. If a new object (which does not belong to the object set) is presented, it should first be added to the object data set, by undergoing the object dataset construction step.

3.2.2.1 Information gain-based approach (the proposed approach)

An IG-based approach is proposed here that estimates the degree to which the information uncertainty is reduced by each choice of the pose in the haptic identification applied to a set of ℓ candidate objects.

After each grasp, the algorithm will calculate the information gain for each pose for every remaining object in the set. Firstly, the information entropy is calculated for each object. Considering that at the t^{th} iteration/grasp, there are N_o^t remaining objects in the set O^t. Note that when the first iteration (first grasp) starts, $N_o^1 = \ell$. When the object to be identified is assumed to be the l^{th} element in the set ($O^t[l]$), all other remaining objects in the set will then be named $\neg O^t[l]$ (*i.e.*, not $O^t[l]$). Therefore, the object set for $O^t[l]$ can be represented as two possible outcomes: $O_l^t = \left\{O^t[l], \neg O^t[l]\right\}$, representing the random variable. Assuming the probability that the object being identified is any of the objects in the remaining set is equal (uniformly distributed), the probability of each possible outcome can then be calculated as:

$$p(O^t[l]) = \tfrac{1}{N_o^t}; \tag{3.12}$$

$$p(\neg O^t[l]) = \tfrac{N_o^t - 1}{N_o^t}. \tag{3.13}$$

Therefore, the information entropy for the object $O^t[l]$ can be calculated as:

$$H(O^t[l]) = -p(O^t[l])\log_2(p(O^t[l])) - p(\neg O^t[l])\log_2(p(\neg O^t[l])). \tag{3.14}$$

A second variable considered here is $C_j^t = \left\{{c_{j,1}}^t, ..., c_{j,N_h^t}^t\right\}$ containing non-repeating grasp cluster numbers at pose p_j for all remaining objects in O_l^t, while N_h^t is the number of non-repeating grasp cluster numbers at the t^{th} iteration. At pose p_j, each possible outcome $c_{j,v}^t \in C_j^t$ (where $v \leq N_h^t,\ v \in \mathcal{N}$) has a probability $p(c_{j,v}^t)$. Defining a set $F_j = \left\{{f_{j,1}}^t, ..., f_{j,N_h^t}^t\right\}$ to represent the frequency of occurrence of each non-repeating grasp cluster number $c_{j,v}^t$ at pose p_j (for example, at pose p_j, the occurrence frequency of $c_{j,1}^t$ is ${f_{j,1}}^t$), the probability of the outcome $c_{j,v}^t$ can be calculated as:

$$p(c_{j,v}^t) = \frac{f_{j,v}^t}{N_o^t}. \tag{3.15}$$

Therefore, the conditional entropy for $O^t[l]$ at pose p_j can be computed as:

$$\begin{aligned} H(O^t[l]|p_j) &= \sum_{v=1}^{N_h^t} p(c_{j,v}^t) H(O_l^t|c_{j,v}^t) \\ &= -\sum_{v=1}^{N_h^t} p(c_{j,v}^t) \Bigg[p(O^t[l]|c_{j,v}^t) \log_2 \left(p(O^t[l]|c_{j,v}^t) \right) \\ &\quad + p(\neg O^t[l]|c_{j,v}^t) \log_2 \left(p(\neg O^t[l]|c_{j,v}^t) \right) \Bigg]. \end{aligned} \tag{3.16}$$

The $p(O^t[l]|c_{j,v}^t)$ and $p(\neg O^t[l]|c_{j,v}^t)$ are the conditional probability of $O^t[l]$ given $c_{j,v}^t$ and $\neg O^t[l]$ given $c_{j,v}^t$ at the t^{th} iteration, respectively. Specifically, if the grasp cluster number of $O^t[l]$ at p_j is $c_{j,v}^t$, the conditional probability can be calculated as:

$$p(O^t[l]|c_{j,v}^t) = \tfrac{1}{f_{j,v}^t}; \tag{3.17}$$

$$p(\neg O^t[l]|c_{j,v}^t) = \tfrac{f_{j,v}^t - 1}{f_{j,v}^t}, \tag{3.18}$$

otherwise, if the grasp cluster number of $O^t[l]$ at p_j is not $c_{j,v}^t$, then:

$$p(O^t[l]|c_{j,v}^t) = 0; \tag{3.19}$$

$$p(\neg O^t[l]|c_{j,v}^t) = 1. \tag{3.20}$$

The information gain (IG) for $O^t[l]$ at pose p_j can be calculated as:

$$\begin{aligned} G(O^t[l], p_j) &= H(O^t[l]) - H(O^t[l]|p_j) \\ &= -p(O^t[l]) \log_2(p(O^t[l])) - p(\neg O^t[l]) \log_2(p(\neg O^t[l])) \\ &\quad + \sum_{v=1}^{N_h^t} p(c_{j,v}^t) \Bigg[p(O^t[l]|c_{j,v}^t) \log_2 \left(p(O^t[l]|c_{j,v}^t) \right) \\ &\quad + p(\neg O^t[l]|c_{j,v}^t) \log_2 \left(p(\neg O^t[l]|c_{j,v}^t) \right) \Bigg]. \end{aligned} \tag{3.21}$$

The information gain can be calculated for all remaining objects for each pose. The algorithm then selects the pose that contains the highest IG. In other words, the algorithm will grasp the object at the pose that would provide the most information gain.

For each pose, each object has its IG value, representing the degree of distinction of this object from all other remaining objects in the set at this pose. At a given pose, the greater the IG value of an object, the greater the degree of distinction between this object and other remaining objects at this pose. There may be a few ways to quantify the measure of the "highest information gain". In this proposed method, the algorithm first selects the pose with the largest value of minimum information gain (MIG). When the largest value of MIG is shared by more than one pose, the algorithm then chooses the pose with the highest value of average IG (AIG) out of these poses. If the highest value of AIG is shared by more than one pose, any of these poses can be chosen since they all have the same amount of information for object identification purposes. If such a situation arises, the pose selection will follow the order of poses.

The summary of the proposed method is shown in Algorithm 2. Note that for a given object dataset, when the position of the object pose is known, the first pose in the identification process is always the same, as the algorithm will go first for the pose with the largest value of minimum IG for the entire object dataset. After the first grasp, depending on the measurements encountered on the object to the identified, the set of candidate objects is eliminated from consideration and a "reduced object set" is generated, containing the remaining possible options (of objects) that could not be eliminated.

The proposed algorithm will be repeated until the number of objects in the set is less than or equal to one. The outcome of the IG-based approach includes the predicted label of the object and the number of grasps used for its identification. The prediction result can be one of the following: 1) the object is correctly identified, or 2) the object is misidentified (as another in the object set). It is also possible in the case of misidentification that the resulting object description does not match any in the object dataset. This case (Case 3) represents that the object is unidentified. Note that this does not imply that the new object is not contained in the object dataset. It simply means that the uncertainties in the measurements resulted in an object description that cannot be matched to the dataset. Regarding the number of grasps required for the identification ('Grasps $\leftarrow t-1$'), each iteration corresponds to one grasp, with t representing the current iteration. As the algorithm increments t by one before the next iteration begins, it is necessary to subtract one from the output value to account for this adjustment. Hence, when the algorithm stops, the number of grasps used for identification will be $t-1$.

Algorithm 2: Information Gain Based Approach

Input: The object to be identified; The descriptor of each object;

Iteration $t \leftarrow 1$;

while *size(O^t)>1:* **do**

  for *each* $p_j \in [p_1, p_n]$ **do**

    for *each* $O^t[l] \in O^t$ **do**

      Calculate the information gain $G(O^t[l], p_j)$ using Eq. (3.21);

    end

  end

  for *each* $p_j \in [p_1, p_{N_g}]$ **do**

    Calculate MIG and AIG at each pose:

$$G_{MIG,p_j} \leftarrow \min_{O^t[l] \in O^t} (G(O^t[l], p_j))$$

$$G_{AIG,p_j} \leftarrow \operatorname{avg}_{O^t[l] \in O^t} (G(O^t[l], p_j))$$

  end

  Find the subset of P in which the poses have the same largest MIG:

$$P_{max,MIG} \leftarrow \operatorname*{argmax}_{p_j \in [p_1, p_{N_g}]} (G_{MIG,p_j})$$

  if *size($P_{max,MIG}$)=1:* **then**

    Grasp at the only pose in $P_{max,MIG}$;

  else

    if *size($P_{max,MIG}$)>1:* **then**

      Find subset of $P_{max,MIG}$ in which the poses have the same largest AIG:

$$P_{max,AIG} \leftarrow \operatorname*{argmax}_{p_j \in P_{max,MIG}} (G_{AIG,p_j})$$

      if *size($P_{max,AIG}$)=1:* **then**

        Grasp at the only pose in $P_{max,AIG}$;

      else

        In $P_{max,AIG}$, choose a pose based on the order;

        Grasp at the selected pose;

      end

    end

  end

  Obtain the grasp cluster number;

  $t \leftarrow t + 1$;

  Eliminate inadmissible candidates from set O^t;

end

if *size(O^t)=1:* **then**

  Output: the predicted label;

  Grasps $\leftarrow t-1$;

else

  Output: *"Null"*(unidentified);

  Grasps $\leftarrow t-1$;

end

3.2.2.2 Baseline approach

In this chapter, the baseline approaches are constructed to compare with the proposed IG-based approaches, and the summaries of it are shown in Algorithm 3. In this chapter, the baseline approach indicates that the decision on where to grasp next in the haptic object identification process does not consider the information content of the object dataset. The procedure of the baseline approach is similar to the IG-based approach (it still eliminates inadmissible candidates following the grasps performed), but each grasp position is selected at random. The grasp poses where measurements are already taken are removed from the set P, which means this grasp position can not be chosen in the following identification iterations. The baseline approach has also been utilised in [26, 62, 86] where efficiency is not the focus of their study.

Algorithm 3: Baseline Approach

Input: The object to be identified; The descriptor of each object;

Iteration $t \leftarrow 1$;

while *size(O^t)>1:* **do**

  Randomly choose a pose from the pose set P ;

  Grasp at the selected pose;

  Obtain the corresponding grasp cluster number;

  Eliminate inadmissible candidate objects from the object set O^t;

  Eliminate the selected pose from the pose set P ;

  $t \leftarrow t + 1$;

end

if *size(O^t)=1:* **then**

  Output: the predicted label;

  Grasps $\leftarrow t - 1$;

else

  Output: *"Null"*(unidentified);

  Grasps $\leftarrow t - 1$;

end

3.3 Experimental Evaluation

3.3.1 Experimental setup

To evaluate the performance of the proposed object identification algorithm, the following experimental setup is used in this chapter:

3.3.1.1 Object set

The descriptions of the objects will be presented in the format of [explanation of the object - material]. To represent a wide range of objects with different values of stiffness and shapes, 23 objects are selected (as shown in Table 3.1 and Fig. 5.3). Objects used are: 1) A soft drink bottle - glass, 2) A can of beetroot - metal, 3) A peppercorn dispenser - hard plastic, 4) A drinking mug - ceramic, 5) A cleaning spray bottle of non-axisymmetric shape - medium stiffness plastic, 6) A can of meat - metal, 7) A 3D printed cylinder - PLA plastic, 8) An arbitrarily shaped 3D-printed object - PLA plastic, 9) A soft drink bottle - soft plastic, and 10) A soft drink bottle otherwise identical to Object 9) but with a bump on the bottle cap - soft plastic, 11) A tennis ball - uniform felt-covered rubber compound, 12) A silicone ball - hard silicone, 13) A bottle of mouthwash - hard stiffness plastic, 14) Another bottle of mouthwash - medium stiffness plastic, 15) A sports drink bottle - soft plastic, 16) A sport drink bottle of a different shape (full) - soft plastic , 17) A sport drink bottle of a different shape (empty) - soft plastic, 18) Assembled Lego blocks - hard plastic, 19) Assembled Lego blocks with a specific part at Height 1 - hard plastic, 20) Assembled Lego blocks with a specific part at Height 2 - hard plastic, 21) Assembled Lego blocks with a specific part at Height 3 - hard plastic, 22) Assembled Lego block with a specific part at Height 4 - hard plastic, 23) Assembled Lego blocks with a shorter height. All selected objects satisfy the assumption of distinguishability of the objects.

The object set is constructed based on the Yale-CMU-Berkeley (YCB) model set [87]. Specifically, Objects 1, 2, 6, 9, 10, 15, 16, and 17 belong to the 'Food items' category in the YCB set, and Objects 3, 4, 5, 13, and 14 are in the 'Kitchen items' category, Objects 7, 8, 11, 12 are included in the 'Shape items' category, and Objects 18-23 belongs to

'Assembly objects' category. Furthermore, to analyse the efficiency of object identification, objects in the set are intentionally selected to have a high level of similarity, such as Objects 9 and 10 which are identical except for the bump on the bottle cap on Object 10. Objects 11 and 12 have the same shape as their main body, with a stiffness difference overall. Objects 13 and 14 have similar sizes and shapes but with different stiffness at Pose 1. Therefore, when grasping these two, the joint angle will be similar, but the tactile map will be different. Objects 15 and 16 have similar object convex envelopes, such that they result in similar finger displacements when grasped. However, they differ in that one has clearly defined edges while the other shows a rounded shape. The tactile measurements upon grasping these two objects are significantly different, while the finger displacements of the grasps are similar. Objects 15 and 17 are the same object (identical plastic bottles), but one is filled full of water while the other is empty. Objects 18-23 have similar shapes on the majority of the portions of the objects, except for one specific part on each object.

Table 3.1: The object set

Object Number	Object Name	Material
1	A soft drink bottle	Glass
2	A can of beetroot	Metal
3	A peppercorn dispenser	Hard plastic
4	A drinking mug	Ceramic
5	A cleaning spray bottle of non-axisymmetric shape	Medium stiffness plastic
6	A can of meat	Metal
7	A 3D printed cylinder	PLA plastic
8	An arbitrarily shaped 3D-printed object	PLA plastic
9	A soft drink bottle	Soft plastic
10	A soft drink bottle with a bump on the bottle cap	Soft plastic
11	A tennis ball	Uniform felt-covered rubber compound
12	A silicone ball	Hard silicone
13	A bottle of mouthwash	Hard stiffness plastic
14	Another bottle of mouthwash	Medium stiffness plastic
15	A sports drink bottle	Soft plastic
16	A sports drink bottle of a different shape (full)	Soft plastic
17	A sports drink bottle of a different shape (empty)	Soft plastic
18	Assembled Lego blocks	Hard plastic
19	Assembled Lego blocks with a specific part at Height 1	Hard plastic
20	Assembled Lego blocks with a specific part at Height 2	Hard plastic
21	Assembled Lego blocks with a specific part at Height 3	Hard plastic
22	Assembled Lego blocks with a specific part at Height 4	Hard plastic
23	Assembled Lego blocks with a shorter height	Hard plastic



FIGURE 3.5: Consists of 23 objects in a variety of shapes, sizes and stiffness.

3.3.1.2 Experimental platform

Robotic hand

The ReFlex TakkTile hand (Right Hand Robotics, US), which consists of three under-actuated fingers, is used in this chapter. The ReFlex hand has 4 Degrees of Freedom (DoFs): the flexion/extension of each finger and 1 coupled rotation between the orientation of finger $No.1$ and $No.2$). In this project, only the 3 DoFs of finger flexion/extension are used. Each under-actuated finger is controlled by a Dynamixel $MX-28$ actuator that can drive the tendon spanning both the proximal and distal joints. The proximal joint connects the proximal link to the base, and the distal joint connects the distal link to the proximal link. For each finger, there is one proximal joint encoder, one tendon spool encoder, and nine embedded Takktile pressure sensors arranged along the finger (as shown in Fig. 3.6). The reading of the tendon spool encoder represents the angular displacement of the whole finger (the sum of the angular displacement of the proximal joint and distal joint), representing the proprioceptive information. Therefore, the degree of flexion of the distal joint can be calculated from the difference between the tendon spool decoder and the proximal joint encoder.

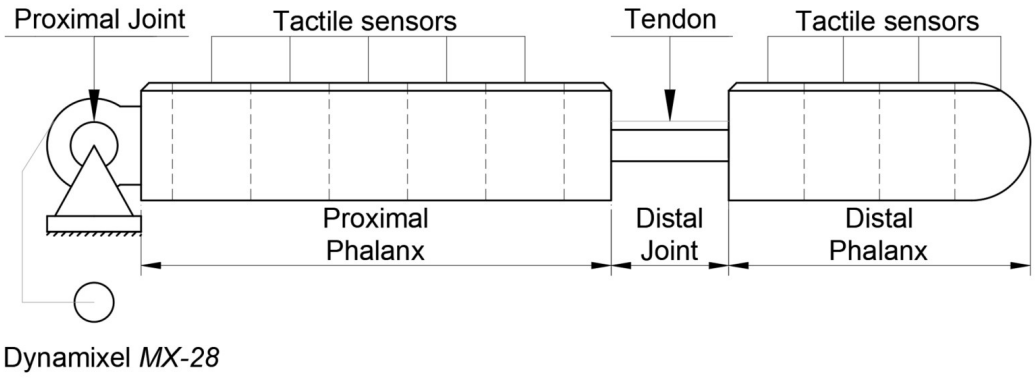


FIGURE 3.6: A diagram of the finger of the Reflex Takktile hand showing the location of the tactile sensors.

Tactile Sensors

The TakkTile pressure sensors [88], which are designed based on the urethane-encased micro-electro-mechanical system (MEMS) barometers, are embedded in the compliant pads of each finger as the tactile sensors (as shown in Fig. 3.7). There are five tactile sensors on the proximal phalanx and four on the distal phalanx of each finger (as shown in Fig. 3.6). For each phalanx of the finger, the sensors are mounted on a printed circuit board with an $8mm$ separation. The Takktile sensors have the excellent linearity ($< 1\%$

typical), low noise levels ($< 0.01N$) and bandwidth ($> 100Hz$). Each tactile sensor can measure the normal contact pressure.

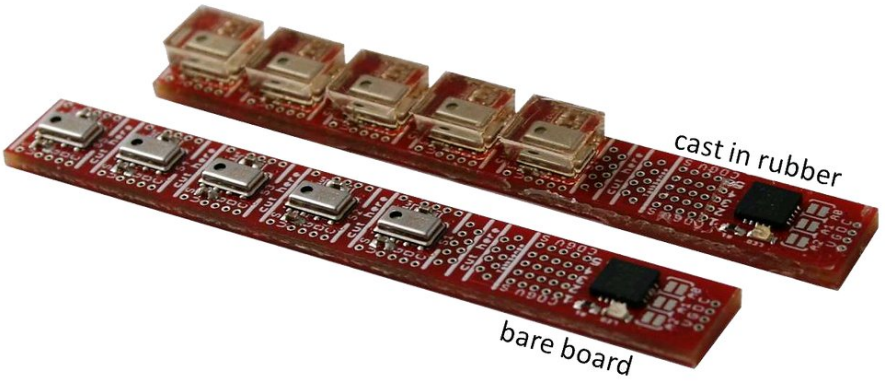
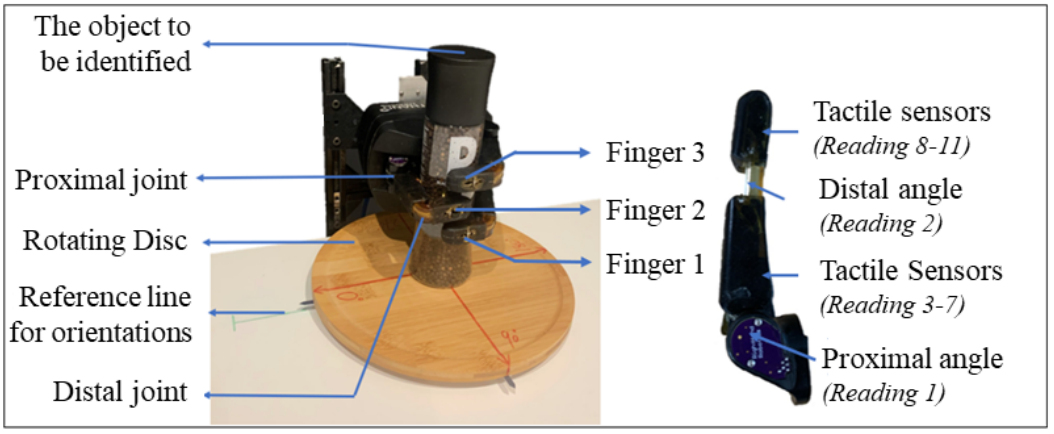


FIGURE 3.7: A TakkTile sensor strip consists of 5 sensors end-to-end spaced at $8mm$ spacing [3].

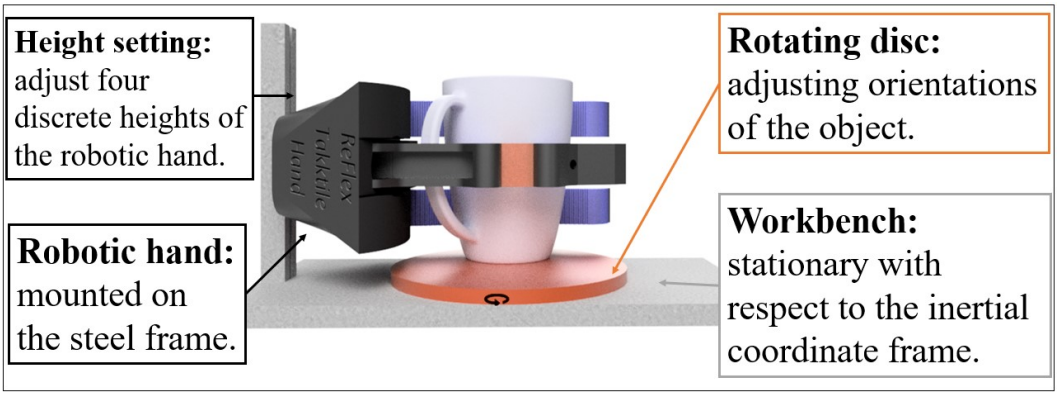
Software

The robotic hand operation and data collection are all performed via Robot Operating System (ROS), version 'ROS-indigo', under Linux Ubuntu 14.04 computer system. The information (joint angle and pressure value information) is recorded via the 'rosbag' command and is saved in the 'comma-separated values'('.csv') file. All recorded data are then processed using Python3.

The overall experimental platform is shown in Figure 3.8. In the thesis, instead of mounting the ReFlex hand on a robot arm, it is mounted on a steel frame, with its three fingers positioned parallel to the horizontal direction. The height of the robotic hand can be determined by adjusting the vertical position of the steel frame. The rotating disc can adjust the orientations of the object. Therefore, the haptic measurements of each object at different poses can be collected. Such a setup simplifies the data collection procedure and prevents the setup from becoming complex. It should be noted that when collecting the data, the object is not stationary and it can be moved during the grasping procedure. All the data is collected at $40Hz$ sampling. During each grasp, all fingers are commanded to move 3 *rad* over 6 seconds. A torque threshold is also set for each motor. For each finger, once this threshold is reached, this finger will stop moving and keep that pose as the final pose for measurements.



(a)



(b)

FIGURE 3.8: Experimental platform. **(a)** The experimental setup. **(b)** The schematic of the experimental setup (side view).

3.3.2 Object dataset construction

Each object $o_i \in O, i = 1, \ldots, 23$ is grasped at 16 different poses, $N_g = 16$, divided across 4 grasp heights (with $55mm$ increments). The object is grasped in 4 different orientations about the vertical axis at each height setting (with 90^o increments). For each object, the grasp is repeated 10 times at each pose. The collected data are then normalised using the min-max normalisation method. To categorise the data, fifty-three grasp clusters are identified through the elbow method. Since each object is grasped at $N_g = 16$ different poses, the label for each object is made up of a series of $N_g = 16$ grasp cluster numbers.

3.3.3 Object identification

Using the established data set, the object identification procedure with the IG-based algorithm for efficiency was validated for two scenarios, where the measurements taken for the object identification process were done:

- **Scenario 1**: The grasp is carried out without uncertainty. This represents a sanity check for the theoretical-best-possible outcome, evaluating only the ability of the IG-based algorithm to utilise the information of the objects in the dataset to make decisions on where to make the subsequent measurement grasp. Note that the sensor measurement errors (of the robotic grasper encoders and tactile sensors) are still present in this case but were not significant to cause any false outcomes in identifying the grasp types.
- **Scenario 2**: The grasp is carried out with an amount of uncertainties approximated to a typical practical process. The amount of practical uncertainties was realised as the positioning deviation of the grasper relative to the object, bounded within $\pm$ 20 degrees of grasping orientation uncertainty (with respect to the object). It can be highlighted that with the uncertainty involved, this scenario is able to evaluate the performance in a practical situation, where measurement and grasp positioning uncertainty is present.

Two tests are conducted in the two scenarios: the accuracy evaluation and the efficiency evaluation. For both evaluations, 10 trials are conducted, meaning each object will be identified 10 times. In the *accuracy evaluation*, the identification result (correctly identified, misidentified or unidentified) from each trial will be recorded. The procedure is shown in Algorithm 4. In the *efficiency evaluation*, the number of grasps used to correctly identify the object in each trial will be counted. The procedure of the efficiency evaluation is summarised in Algorithm 5. The measure of evaluation is defined as the number of grasps used to correctly identify the presented object. Specifically, when uncertainties are involved in the identification, there will be a possibility of the algorithm misidentifying the grasp type when measuring one of the poses. The definition hence means that in each (of the ten attempts per object presented) to identify an object correctly, the number of grasps made on the object when a correct identification is not achieved are also counted towards calculating the efficiency. Each time the algorithm

misidentifies the presented object, the algorithm will restart again.

Algorithm 4: Procedure of the accuracy evaluation

Input: The object to be identified, its true label, the number of trials, and the identification approaches;

trial ← 0;

correctly_identified ← 0;

mis_identified ← 0;

un_identified ← 0;

while *trial* ≤ *number of trials:* **do**

  if *identification approach = IG based:* **then**

    prediction, grasps ← **Call** Algorithm 2;

  else if *identification approach = baseline:* **then**

    prediction, grasps ← **Call** Algorithm 3;

  end

  if *prediction = object's true label:* **then**

    correctly_identified ← correctly_identified +1;

  else if *prediction = "Null":* **then**

    un_identified ← un_identified +1;

  else

    mis_identified ← mis_identified +1;

  end

  trial ← trial+1;

end

correctly_identified ← correctly_identified / number of trials;

mis_identified ← mis_identified / number of trials;

un_identified ← un_identified / number of trials;

Output: correctly_identified, mis_identified, un_identified;

Algorithm 5: Procedure of the efficiency evaluation

Input: The object to be identified, its true label, the number of trials, and the identification approaches;

```
trial ← 0;
while trial ≤ the number of trials: do
    Identification ← false;
    Num_of_grasps ← 0;
    while Identification ! = true: do
        if identification approach = IG based: then
            prediction, grasps ← Call Algorithm 2;
        else
            if identification approach = baseline: then
                prediction, grasps ← Call Algorithm 3;
            end
        end
        Num_of_grasps ← Num_of_grasps + grasps;
        if prediction = object's true label: then
            Identification ← true;
        end
    end
    trial ← trial+1;
    Output: Num_of_grasps;
end
```

3.4 Results and Discussion

3.4.1 Object dataset construction

The result of the object dataset construction used in the experimental evaluation is represented in Table 3.2. Each of the 23 objects is described as a 16-number long string, representing the 16 clusters for that specific object from the measurements taken at the 16 grasping poses performed on the object. For example, according to Table 3.2, a cluster [1] can be obtained when grasping Object 1 (o_1) at Pose 5 (p_5).

TABLE 3.2: The list of the constructed labels for the object dataset used in the experiment

	Pose 1	Pose 2	Pose 3	Pose 4	Pose 5	Pose 6	Pose 7	Pose 8	Pose 9	Pose 10	Pose 11	Pose 12	Pose 13	Pose 14	Pose 15	Pose 16
Object 1	[0]	[0]	[0]	[0]	[1]	[1]	[1]	[1]	[2]	[2]	[2]	[2]	[3]	[3]	[3]	[3]
Object 2	[4]	[4]	[4]	[4]	[5]	[5]	[5]	[5]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]
Object 3	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[7]	[7]	[7]	[7]
Object 4	[8]	[8]	[8]	[9]	[10]	[11]	[12]	[13]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]
Object 5	[14]	[9]	[14]	[9]	[14]	[9]	[14]	[9]	[14]	[15]	[14]	[15]	[16]	[16]	[16]	[16]
Object 6	[17]	[9]	[17]	[9]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]
Object 7	[18]	[18]	[18]	[18]	[19]	[19]	[19]	[19]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]
Object 8	[18]	[18]	[18]	[18]	[20]	[20]	[20]	[20]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]
Object 9	[21]	[21]	[21]	[21]	[22]	[22]	[22]	[22]	[23]	[23]	[23]	[23]	[24]	[24]	[24]	[24]
Object 10	[21]	[21]	[21]	[21]	[22]	[22]	[22]	[22]	[23]	[23]	[23]	[23]	[25]	[24]	[24]	[24]
Object 11	[26]	[26]	[26]	[26]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]
Object 12	[27]	[27]	[27]	[27]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]
Object 13	[28]	[29]	[28]	[29]	[30]	[31]	[30]	[31]	[32]	[33]	[32]	[33]	[34]	[34]	[34]	[34]
Object 14	[35]	[29]	[35]	[29]	[30]	[31]	[30]	[31]	[32]	[33]	[32]	[33]	[36]	[36]	[36]	[36]
Object 15	[37]	[37]	[37]	[37]	[38]	[38]	[38]	[38]	[39]	[39]	[39]	[39]	[40]	[40]	[40]	[40]
Object 16	[39]	[41]	[39]	[41]	[42]	[43]	[42]	[43]	[44]	[44]	[44]	[44]	[45]	[45]	[45]	[45]
Object 17	[39]	[41]	[39]	[41]	[46]	[43]	[46]	[43]	[44]	[44]	[44]	[44]	[45]	[45]	[45]	[45]
Object 18	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]
Object 19	[49]	[50]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]
Object 20	[47]	[48]	[47]	[48]	[49]	[50]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]
Object 21	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[49]	[50]	[47]	[48]	[47]	[48]	[47]	[48]
Object 22	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[49]	[50]	[47]	[48]
Object 23	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[47]	[48]	[51]	[52]	[51]	[52]

3.4.2 Object identification

Scenario 1 (The grasp is carried out without uncertainties)

Accuracy Evaluation

Ten sets of identification trials for each object are conducted. In this scenario where no uncertainties were included, it was confirmed that all trials yielded a 100% accuracy for both the IG-based approach and the baseline approach, as expected.

Efficiency Evaluation

The number of grasps required to correctly identify the presented object, using the baseline approach and the proposed IG-based approach, is shown in Fig. 3.9. Using the constructed object dataset, the IG value for each object at each pose can be calculated for the first iteration. The IG values for all the objects for all possible poses are shown in Table 3.3.

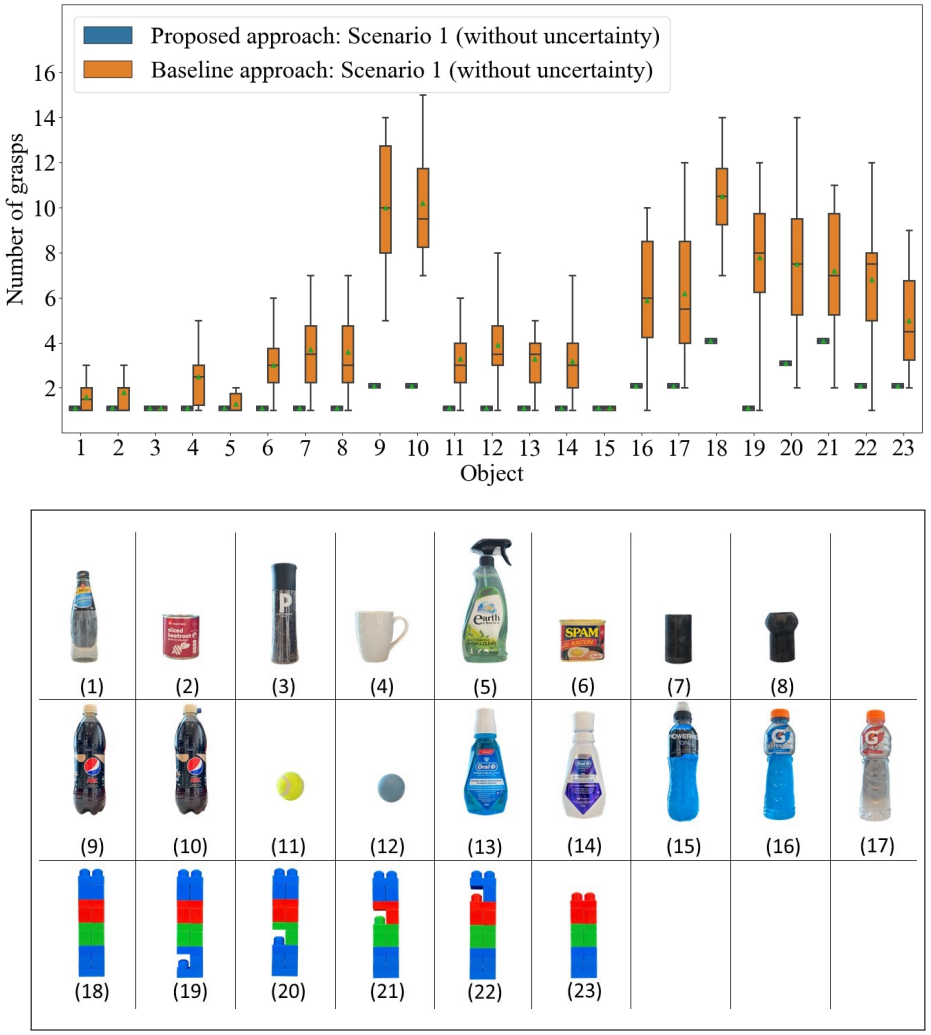


FIGURE 3.9: The number of grasps required to correctly identify each of 23 objects using the proposed approach and the baseline approach: Scenario 1 (without uncertainty).

TABLE 3.3: IG value for each object at each pose: first iteration

	Pose 1	Pose 2	Pose 3	Pose 4	Pose 5	Pose 6	Pose 7	Pose 8	Pose 9	Pose 10	Pose 11	Pose 12	Pose 13	Pose 14	Pose 15	Pose 16
Object 1	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.07	0.07	0.07	0.07
Object 2	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07
Object 3	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
Object 4	0.26	0.26	0.26	0.14	0.26	0.26	0.26	0.26	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07
Object 5	0.26	0.17	0.26	0.14	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
Object 6	0.26	0.17	0.26	0.14	0.14	0.14	0.14	0.14	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07
Object 7	0.17	0.17	0.17	0.17	0.26	0.26	0.26	0.26	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07
Object 8	0.17	0.17	0.17	0.17	0.26	0.26	0.26	0.26	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07
Object 9	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.26	0.17	0.17	0.17
Object 10	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.26	0.17	0.17	0.17
Object 11	0.26	0.26	0.26	0.26	0.14	0.14	0.14	0.14	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07
Object 12	0.26	0.26	0.26	0.26	0.14	0.14	0.14	0.14	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07
Object 13	0.26	0.17	0.26	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.26	0.26	0.26	0.26
Object 14	0.26	0.17	0.26	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.26	0.26	0.26	0.26
Object 15	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
Object 16	0.17	0.17	0.17	0.17	0.26	0.17	0.26	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
Object 17	0.17	0.17	0.17	0.17	0.26	0.17	0.26	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
Object 18	0.10	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.12	0.12	0.10	0.10
Object 19	0.26	0.26	0.09	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.12	0.12	0.10	0.10
Object 20	0.10	0.10	0.09	0.09	0.26	0.26	0.09	0.09	0.10	0.10	0.09	0.09	0.12	0.12	0.10	0.10
Object 21	0.10	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.26	0.26	0.09	0.09	0.12	0.12	0.10	0.10
Object 22	0.10	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.26	0.26	0.10	0.10
Object 23	0.10	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.26	0.26	0.26	0.26
AIG	0.20	0.19	0.19	0.17	0.19	0.19	0.18	0.18	0.15	0.15	0.14	0.14	0.16	0.15	0.14	0.14
MIG	0.10	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07

To decide on the pose for the first grasp, the algorithm looks at the largest minimum IG (MIG) values for all the poses in Table 3.3. In our example, a MIG value of 0.10 can be found for Pose 1 and 2, and Pose 5 and 6 (see the last row of Table 3.3). The algorithm then selects out of these poses, the one with the highest averaged IG (AIG) value, which is 0.20 for Pose 1. Therefore, for the first grasp, Pose 1 is selected. Note that once the first grasp is made, depending on the grasp type identified, inadmissible candidates are removed and the IG values for the subsequent grasp will need to be calculated with the set of the remaining objects.

The experiment results of the number of grasps required to correctly identify the object through the IG-based approach are shown in Fig. 3.9. It is observed that the IG-based approach outperforms the baseline approach in general, even if the difference in performance differs from object to object. The average number of grasps needed for the baseline approach is 4.81 while the IG-based approach was observed to correctly identify the object on average in 1.65 grasps. As expected, the number of grasps required by the baseline approach to identify an object varies widely between different attempts due to the random selection in deciding the subsequent grasps. The IG-based approach is repeatable for all objects in the same set in the absence of practical uncertainties. This demonstrates the theoretically achievable performance of the proposed approach.

The advantage of the proposed approach is highlighted in the more challenging cases, where objects are more similar to others in the object set. Objects 9 and 10 are identical except for a dent in one of the two bottle caps - and the baseline approach struggled to identify the object without exploiting the knowledge of the objects in the dataset. Similar observations can be made to Objects 18 to 23.

Scenario 2 (The grasp is carried out with uncertainty)

This section explores how the proposed object identification approach performs when the measurement includes a level of uncertainty typical to that seen in practical conditions. The identification procedure remains the same as shown in Section 3.2.2.1 where the first grasp is carried out at Pose 1.

Accuracy evaluation

The resulting percentage accuracy of the proposed approach in the presence of practical uncertainties is shown in Fig. 3.10. The percentage of misidentified (when an object is identified as another) and unidentified (when the resulting descriptor does not match

any object in the dataset due to the uncertainties involved) results are also shown. It can be seen that some objects are more susceptible to uncertainties. In general, cylindrical objects are less susceptible, while objects with sharp, angular features are more prone to misidentification. This is due to the nature of the tactile sensors on the robotic grasper used in our experiment, where individual sensors are embedded in the fingers. A change in the grasping pose will cause a sharp angular feature on an object to press on the adjacent tactile sensor, activating a completely different sensor in the resulting reading.

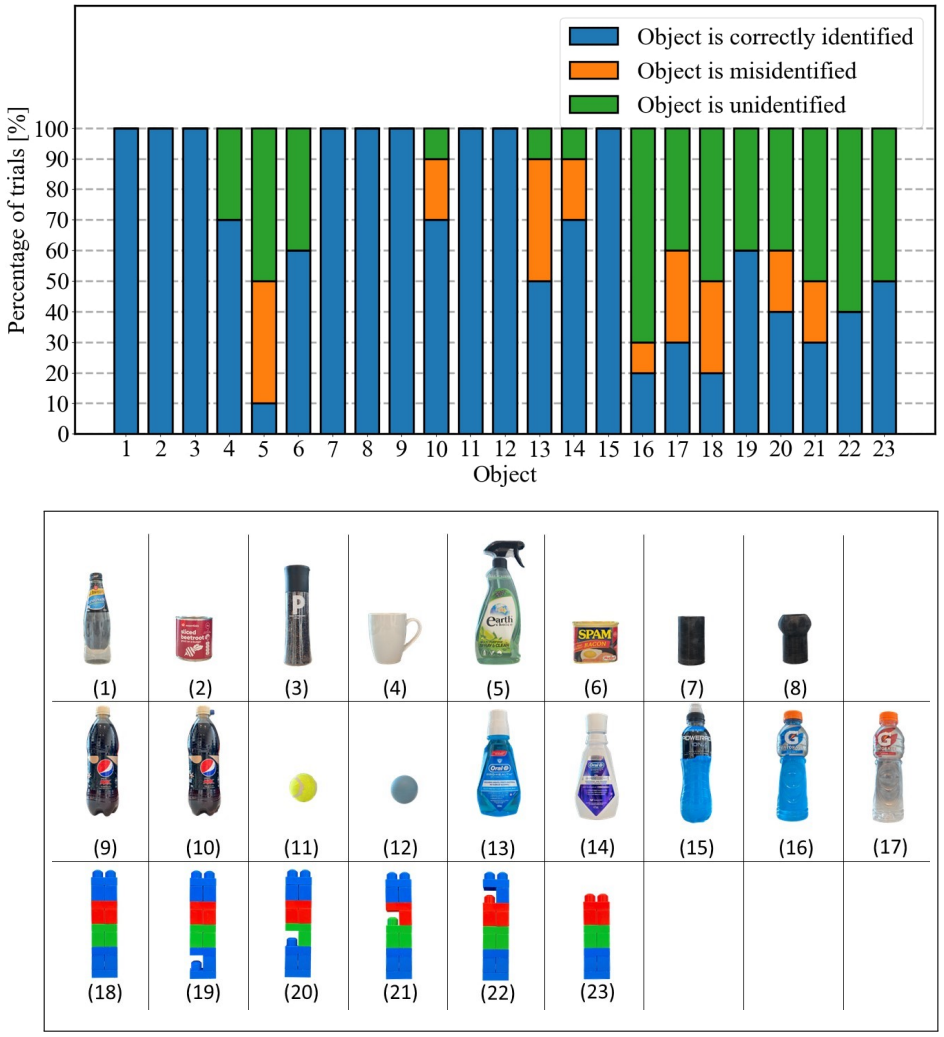


FIGURE 3.10: The accuracy of identifying each of 23 objects through the proposed approach: Scenario 2 (with uncertainty).

Efficiency evaluation

The effect of the inclusion of the practical uncertainties in the performance of the proposed approach is shown in Fig. 3.11, where it is compared to the performance without

practical uncertainties. The average number of grasps required for correct identification increases from 1.65 in Scenario 1 to 4.08 in this scenario.

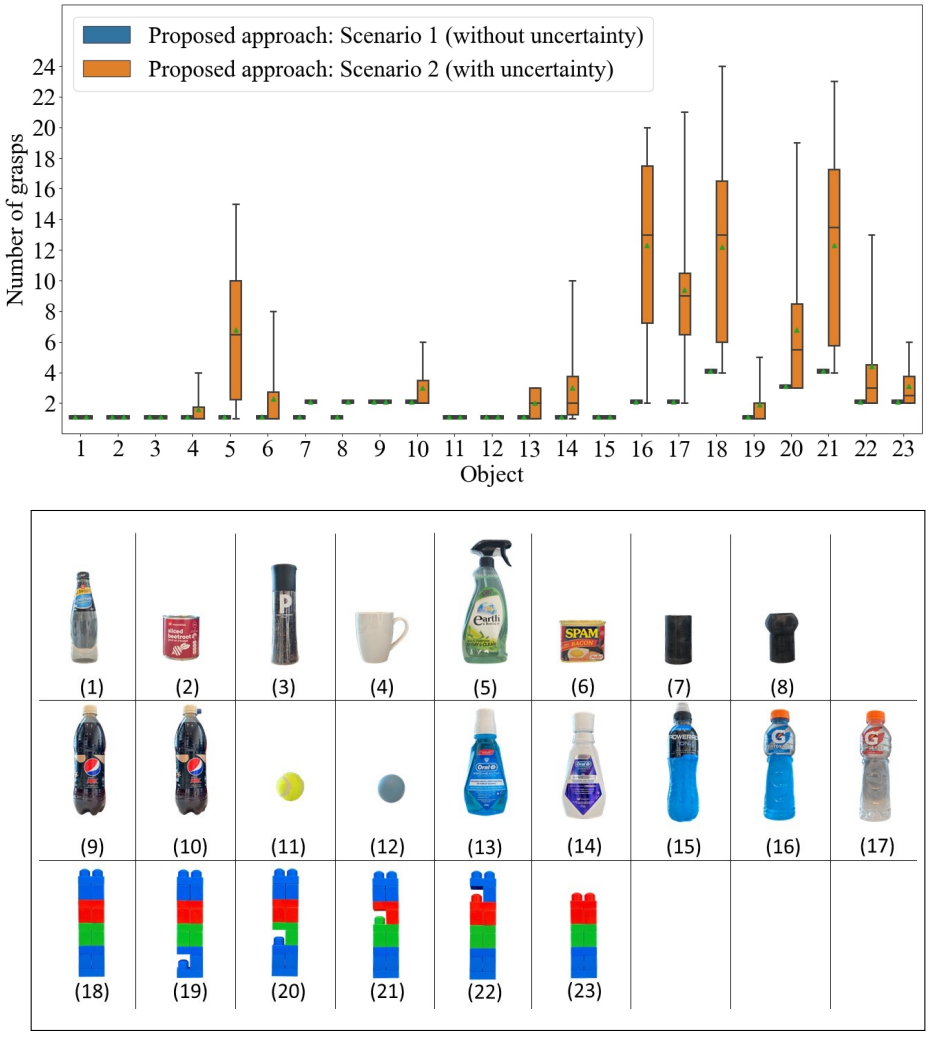


FIGURE 3.11: The number of grasps required to correctly identify each of 23 objects using the proposed approach: Scenario 2 (with uncertainty).

As seen in the outcome of the accuracy evaluation, the identification efficiency of rectangular-shaped objects - objects with sharp angular features on their surface (such as Object 5, 6, 13, 14, and 18) are the most affected, due to its poor accuracy, thus requiring, on average, more attempts (thus more grasps) to make a correct identification. It is also demonstrated that the constructed measure of efficiency is successful in capturing/accounting for the accuracy of the identification approach in reflecting the efficiency performance of the approach.

3.5 Summary

In this chapter, we proposed an approach to utilise the information in the object dataset to improve the efficiency of haptic object identification. It is shown that compared to an equivalent haptic-based object identification approach that does not exploit such information (representative of the conventional approaches), the number of grasps required to correctly identify the presented object was significantly reduced. In the experimental evaluation with 23 objects in the object set, the proposed approach required an average of 1.65 grasps to 4.81 grasps in the baseline approach. The presence of uncertainties in any practical applications, such as in the positioning inaccuracies of the pose of the robotic grasper upon the object or the sensor noise in haptic sensors onboard the robotic grasper, affects the performance of the approach. This is because inaccuracies in the identification approach require the algorithms more attempts to achieve the same number of correct identification of the objects. It was observed that the shapes of the objects and their interaction with the robotic grasper used in the task also determined the susceptibility of the object to these uncertainties.

Chapter 4

Efficiency of Simultaneous Haptic Identification and Localisation

As an extension of Chapter 3, where it is assumed that the object location is known a *priori* throughout the identification process, this chapter aims to provide an efficient solution for haptic object identification when the location of the object during identification is unknown. As the location has to be estimated when simultaneously identifying the object, this problem can be treated as simultaneous haptic object identification and localisation. Specifically, the output of the proposed method will be not only the predicted label but also the pose of the object relative to the grasping hand. The information gain techniques proposed in Chapter 3 will be extended to decide which pose to grasp to obtain efficient identification when localisation is also required.

4.1 Preliminary: Available Transformation

Given a set of objects $O = \{o_1, ..., o_\ell\}$, $\ell \in \mathcal{N}$, where $\mathcal{N}$ is the set of integers, and a set of poses $P = \{p_1, ..., p_{N_g}\}$, $N_g \in \mathcal{N}$, it is assumed that grasping each object $o_i \in O$ (where $i \leq \ell, i \in \mathcal{N}$) at all N_g poses is sufficient to capture the characteristics that would allow it to be distinguished from others in the object set. These poses relative to the object (i.e., relative to a coordinate frame attached on the object) are predetermined, so the transformation matrix ${}^{l}_{k}T$, which represents the relative displacement from pose p_l to any other pose p_k (where $j \neq k$), can be known. In general, there will be $N_g - 1$

transformation matrices for each pose relative to any other predetermined poses. Define $A = \{{}^{l}_{k}T|$ for distinct $l, k \in [1, N_g]\}$ as the set containing all available transformation matrices for all N_g poses.

All distinct transformation matrices in the set A will be then recorded in the set R, which is denoted as $R = \{T_m | T_m, T_n \in A, T_m \neq T_n$ for distinct $m, n \in [1, N_r]\}$, where N_r is the number of available distinct matrices.

In terms of the specific setup introduced in Chapter 3 where all N_g poses relative to the object are uniformly distributed (including N_h levels of heights and N_s uniformly distributed segments at each height), the number of available distinct matrices will be $N_r = (2N_h - 1)N_s - 1 = 2N_g - N_s - 1$. To illustrate this, the following example is used.

Example 4.1. *As the example setting shown in Fig. 4.1, the object has* 2 *levels of height and* 4 *uniformly distributed segments (*90^o *increment) at each height. In this example, each pose would have* 7 *transformation matrices, and there would be* 11 *distinct transformation matrices in total (*$N_r = 11$*). The distinct transformation matrices in this example are: 1)* T_1*−upward translation; 2)* T_2*−upward translation and* 90^o *angular displacement; 3)* T_3*−upward translation and* 180^o *angular displacement; 4)* T_4*−upward translation and* 270^o *angular displacement; 5)* T_5*−*90^o *angular displacement; 6)* T_6*−*180^o *angular displacement; 7)* T_7*−*270^o *angular displacement; 8)* T_8*−downward translation; 9)* T_9*−downward translation and* 90^o *angular displacement; 10)* T_{10}*−downward translation and* 180^o *angular displacement; and 11)* T_{11}*−downward translation and* 270^o *angular displacement.*

In the following section, we will explain how to effectively identify and localise an object from a given object set by utilising the information of the transformation matrices, as well as the information gain method previously introduced in Chapter 3.

4.2 Proposed Methodology

The objective of this chapter is to generate algorithms to improve the efficiency of haptic identification with an appropriate localisation. Specifically, the proposed algorithm aims to identify and localise the object using the minimum number of grasps. Similar to the

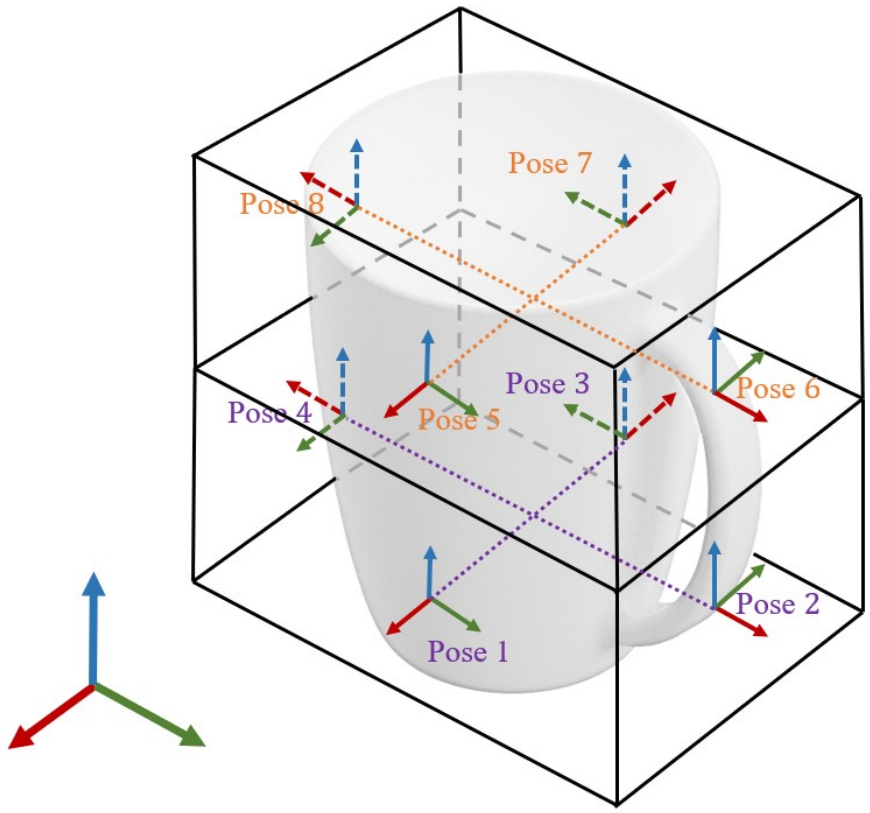


FIGURE 4.1: The characteristics of an object are captured by grasping the object at eight uniformly distributed poses.

discussion in Chapter 3, the process includes object dataset construction object identification and localisation. It should be noted that during the object dataset construction, knowledge of both the pose and the associated haptic information for each object at each pose are considered known as a *priori*. In contrast, this knowledge becomes inaccessible during the identification process.

4.2.1 Object dataset construction

The procedure of object dataset construction is illustrated in Section 3.2.1. This procedure consists of three steps: 1) object data acquisition and normalisation, 2) categorising the grasp clusters, 3) and constructing the object description. Each object in the dataset will be split into N_g poses (as shown in Fig. 3.2).

In the first step, it is assumed that we have sufficient predefined poses so that each object can be uniquely represented by these predefined poses. The robotic hand will then grasp at each pose $(p_1, ..., p_{N_g})$ of each object o_i, $i = 1, \ldots, \ell$, N_T times to account

for the uncertainties associated with the grasping process. During the data acquisition, the min-max normalisation method (introduced in Section 3.1.1) is used for normalising the collected measurements to remove the influence of the dimension and the unit.

The normalised haptic measurements will then be categorised into N_C clusters using unsupervised clustering techniques. In particular, the K-Means clustering technique, which was introduced in Section 3.1.2), is utilised. The number of clusters N_C is determined by the elbow method (as introduced in Section 3.1.2.2). It is assumed that the object dataset is well designed with distinguishable objects, and noise/uncertainty in the haptic measurements are bounded in a range that is clearly smaller than the variance in the distinguishing information between distinct grasps.

After categorisation, the object descriptor will be constructed as a string of N_g cluster numbers. The cluster number at each pose $p_1, ..., p_{N_g}$ represents the cluster that the haptic measurements at that pose for that object have been categorised into. The illustrative example of an object descriptor is shown in Fig. 3.4.

4.2.2 Object identification and localisation

The proposed object identification and localisation approach utilises the information gain-based technique. This extends the algorithm in Chapter 3 towards a more practical scenario of utilising haptic sensors to touch and identify objects, where the actual pose of the object relative to the grasping hand is initially unknown. The performance of the proposed method is compared with a baseline approach where the knowledge of the characteristics in the object dataset is not utilised. It means that the baseline approach would perform identification and localisation where the decision of where to grasp next to obtain more information to identify the object (and to localise itself relative to the object) is done at random. Details of the proposed IG approach and the baseline approach were discussed in Section 3.2.2.1 and Section 3.2.2.2, respectively.

4.2.2.1 The proposed approach

In the context of haptic object identification and localisation, the information gain (IG) algorithm has no prior knowledge of the object's pose relative to the grasping robotic hand. Thus, when determining where to grasp next, instead of selecting a specific

predefined pose for grasping, the output of the IG algorithm is an intended grasp pose relative to the object, which is an available transformation/displacement introduced in Section 4.1. Define the option as (o_i, p_k) for $i \in \{1, 2, ..., \ell\}$ and $k \in \{1, 2, ..., N_g\}$, representing the possible object being identified is and the possible pose relative to the object that the first grasp is carried out. The general procedure of the proposed IG algorithm can be summarised as follows:

1 The first grasp to an unknown object will be randomly carried out. It will grasp the object with no prior information about the object's location relative to the grasping hand.

2 According to all available options, IG based algorithm will determine the displacement relative to the pose of the first grasp that the next grasp should be executed.

3 Inadmissible options are removed after each iteration/grasp.

4 Repeat the process until the number of options in the set is less than or equal to one.

5 If only one option remains, it would be the output of the algorithm. If there are no options remaining (i.e., zero options), it means that the IG algorithm failed to identify or localise the object, hence the object is considered unidentified.

To record the available options of each iteration, in this chapter, we define S^t as the set containing all possible options at the t^{th} iteration (before doing the t^{th} grasp), where $t \in \mathcal{N}_{>0}$. At the first iteration, the option set would be $S^1 = [(o_1, p_1), (o_1, p_2), ..., (o_1, p_{N_g}), (o_2, p_1), ..., (o_2, p_{N_g}), ..., (o_\ell, p_1), ..., (o_\ell, p_{N_g})]$ containing all the combinations of all objects and all predefined poses. After the t^{th} grasps, based on the obtained haptic measurements $\mathbf{z}_t$, the option set for the next iteration S^{t+1} will be formed by removing the inadmissible options from the set S^t, for example, after the first grasp/iteration, the set could be $S^2 = [(o_1, p_2), (o_2, p_5), (o_5, p_4), ..., (o_\ell, p_{N_g-1})]$. In other words, the option set at the iteration $t + 1$ will be a subset of the last iteration t, represented as $S^{t+1} \subseteq S^t$.

To determine where to grasp, for each remaining option $(o_i, p_k) \in S^2$, the information/ cluster that can be obtained after each available displacement (relative to p_k) should be known. It should be noted that, for simplicity, we use the transformation matrix $R[j]$

(also expressed as T_j), $\forall j = 1, ..., N_r$, to represent the pose reached by the displacement represented by this matrix. Two possible scenarios may arise when acquiring the information contained in each $R[j]$:

- The information can be obtained from the descriptor of o_i (constructed in Section 4.2.1). In this case, the pose reached by the displacement represented by $R[j]$ is one of the predefined poses. Then, the cluster number assigned to $R[j]$ would be the same as the one assigned to the predefined pose for which the relative displacement from p_k to it can be represented by $R[j]$.
- The information can NOT be obtained from the descriptor of o_i. In other words, the pose reached by the displacement represented by $R[j]$ is NOT any of the predefined poses. In this scenario, the algorithm will infer the information/cluster for that pose based on the corresponding circumstances. For example, if the first grasp is assumed to be performed at the pose on the lowest height level, then reaching the pose by a downward translation will cause the robotic hand to collide with the workbench (when assuming the object is placed on the workbench), resulting in [null] being assigned to this pose. Similarly, if the first grasp is assumed at the pose on the highest height level, moving the robotic hand upward for the next grasp will lead to a blank grasp, hence this pose (reached by moving upwards) will be filled as the grasp cluster number that represents the blank.

An illustrative example of acquiring the information contained in each $R[j]$ is shown as follows.

Example 4.2. *This example utilises the setting from* ***Example 1****, in which the object contains* 2 *height levels and* 4 *uniformly distributed segments (*90^o *increment) at each height. This leads to* 11 *distinct transformation matrices in total (*$N_r = 11$*). The example is shown in Fig. 4.2. There are* ℓ *objects in the object set, and each object is labelled as* 8 *cluster numbers, representing the cluster that the haptic measurements will be categorised into. For instance, a cluster [1] can be obtained when grasping Object 1 (*o_1*) at Pose 1 (*p_1*). When identifying the object, the first grasp is carried out randomly. Assuming the obtained measurement is categorised into cluster* [3]*, therefore, the available options in the set would be* (o_1, p_2) *and* (o_2, p_1)*. For option* (o_1, p_2)*,* T_1*,* T_2*,* T_3 *and* T_4 *represent the relative displacement from* p_2 *to* p_6*,* p_7*,* p_8 *and* p_5*, respectively. Therefore,*

the cluster number [5], [2], [1] *and* [1] *will be assigned to* T_1 *to* T_4 *for* (o_1, p_2). *Besides,* T_8, T_9, T_{10}, *and* T_{11} *will be assigned a* [*null*] *as moving to a lower height will cause the robotic hand to collide with the workbench. For option* (o_2, p_1), *the same approach would be applied.* T_1 *to* T_4 *represent the relative displacement from* p_1 *to* p_5, p_6, p_7 *and* p_8, *respectively. Therefore, the cluster number* [4], [2], [5] *and* [2] *will be assigned to* T_1 *to* T_4 *for* (o_2, p_1).

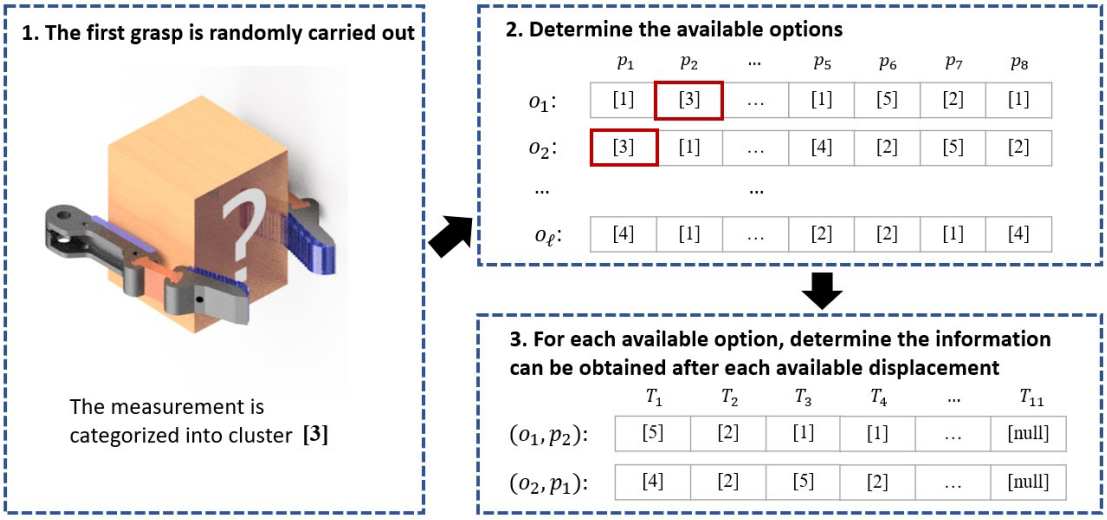


FIGURE 4.2: An illustrative example demonstrates the information that can be obtained for each option after every available displacement.

The proposed IG-based algorithm will then be used to determine where to grasp next. Specifically, at each iteration/grasp, the minimum information gain (MIG) and average information gain (AIG) value will be calculated given all options in S^t, and the pose with the largest value of IG will be selected for the next grasp.

For the l^{th} option in the set S^t, assuming that there are N_p^t remaining options, define $S_l^t = \{S^t[l], \neg S^t[l]\}$ as the set containing two possible outcomes. The information entropy for the option $S^t[l]$ can be calculated as:

$$H(S^t[l]) = -p(S^t[l])\log_2(p(S^t[l])) - p(\neg S^t[l])\log_2(p(\neg S^t[l])), \tag{4.1}$$

where $p(S^t[l])$ and $p(\neg S^t[l])$ are the probability of each outcome in S_l^t respectively.

Assuming the probability that the option being identified is any of the options in the remaining set is equal (uniformly distributed), $p(S^t[l])$ and $p(\neg S^t[l])$ can be calculated

as:

$$p(S^t[l]) = \frac{1}{N_p^t}; \tag{4.2}$$

$$p(\neg S^t[l]) = \frac{N_p^t - 1}{N_p^t}. \tag{4.3}$$

For the conditional entropy calculation, two sets $C_j^t = \left\{c_{j,1}^t, ..., c_{j,v}^t, ..., c_{j,N_h^t}^t\right\}$ and $F_j^t = \left\{f_{j,1}^t, ..., f_{j,v}^t, ..., f_{j,N_h^t}^t\right\}$ are defined to contain non-repeating grasp cluster numbers and the corresponding frequency of occurrence at $R[j]$, respectively, for all remaining options in S_l^t. N_h^t is the number of different cluster numbers at $R[j]$. The probability $c_{j,v}^t$ can be calculated as (considering uniformly distributed):

$$p(c_{j,v}) = \frac{f_{j,v}^t}{N_p^t}. \tag{4.4}$$

Therefore, the conditional entropy for $S^t[l]$ at $R[j]$ can be computed as follows:

$$\begin{aligned} H(S^t[l]|R[j]) &= \sum_{v=1}^{N_h^t} p(c_{j,v}^t) H(S_l^t|c_{j,v}^t) \\ &= -\sum_{v=1}^{N_h^t} p(c_{j,v}^t) \Bigg[p(S^t[l]|c_{j,v}^t) \log_2 \left(p(S^t[l]|c_{j,v}^t)\right) \\ &\quad + p(\neg S^t[l]|c_{j,v}^t) \log_2 \left(p(\neg S^t[l]|c_{j,v}^t)\right) \Bigg]. \end{aligned} \tag{4.5}$$

The $p(S^t[l]|c_{j,v}^t)$ and $p(\neg S^t[l]|c_{j,v}^t)$ are the conditional probability of $S^t[l]$ given $c_{j,v}^t$ and $\neg S^t[l]$ given $c_{j,v}^t$, respectively. To calculate these conditional probabilities, if the grasp cluster number of $S^t[l]$ at $R[j]$ is not $c_{j,v}^t$, then:

$$p(S^t[l]|c_{j,v}^t) = 0; \tag{4.6}$$

$$p(\neg S^t[l]|c_{j,v}^t) = 1. \tag{4.7}$$

otherwise:

$$p(S^t[l]|c_{j,v}^t) = \frac{1}{f_{j,v}^t}; \tag{4.8}$$

$$p(\neg S^t[l]|c_{j,v}^t) = \frac{f_{j,v}^t - 1}{f_{j,v}^t}. \tag{4.9}$$

The information gain for $S^t[l]$ at $R[j]$ then can be calculated as:

$$\begin{aligned} G(S^t[l], R[j]) &= H(S^t[l]) - H(S^t[l]|R[j]) \\ &= -p(S^t[l])\log_2(p(S^t[l])) - p(\neg S^t[l])\log_2(p(\neg S^t[l])) \\ &\quad + \sum_{v=1}^{N_h^t} p(c_{j,v}^t) \Bigg[p(S^t[l]|c_{j,v}^t)\log_2\left(p(S^t[l]|c_{j,v}^t)\right) \\ &\quad + p(\neg S^t[l]|c_{j,v}^t)\log_2\left(p(\neg S^t[l]|c_{j,v}^t)\right) \Bigg]. \end{aligned} \tag{4.10}$$

In summary, when the object location is unknown, the first grasp will be carried out randomly as there is no prior information about the object pose that can be utilised for the identification algorithm. According to each option (o_i, p_k), the proposed IG-based algorithm will be applied to determine the displacement relative to p_k for the next grasp. It should be noted that if there is more than one option coming from the same object and having the same grasp cluster numbers at all poses, then these options will be counted as one option in the set S^t. The proposed algorithm will be repeated until the number of inadmissible options in the set is less than or equal to one. The algorithm is summarised as follows:

Algorithm 6: IG Based Approach: Unknown object pose

Input: The object to be identified; The descriptor of each object;

Randomly grasp the object at one of the predefined poses;

Iteration $t \leftarrow 1$;

Construct the option set S^t;

while *size(S^t)>1:* **do**

  Construct the descriptor of all options based on transformation matrices set R;

  for *each* $R[j] \in R$ **do**

    for *each* $S^t[l] \in S^t$ **do**

      Calculate the information gain $G(S^t[l], R[j])$ using Eq. (4.10);

    end

  end

  for *each* $R[j] \in R$ **do**

    Calculate MIG and AIG at each pose:

$$G_{MIG,R[j]} \leftarrow \min_{S^t[l] \in S^t} (G(S^t[l], R[j]))$$

$$G_{AIG,R[j]} \leftarrow \operatorname{avg}_{S^t[l] \in S^t} (G(S^t[l], R[j]))$$

  end

  Find the subset of R in which poses have the same largest MIG:

$$R_{max,MIG} \leftarrow \operatorname*{argmax}_{R[j] \in R} (G_{MIG,R[j]})$$

  if *size($R_{max,MIG}$)=1:* **then**

    Grasp at the only pose in $R_{max,MIG}$;

  else

    if *size($R_{max,MIG}$)>1:* **then**

      Find subset of $R_{max,MIG}$ in which poses have same largest AIG:

$$R_{max,AIG} \leftarrow \operatorname*{argmax}_{R[j] \in R_{max,AIG}} (G_{AIG,R[j]})$$

      if *size($R_{max,AIG}$)=1:* **then**

        Grasp at the only pose in $R_{max,AIG}$;

      else

        In $R_{max,AIG}$, choose a pose based on the order;

        Grasp at the selected pose;

      end

    end

  end

  Obtain the corresponding grasp cluster number;

  $t \leftarrow t + 1$;

  Update option set S^t;

end

---

if *size(S^t)=1:* **then**

| ***Output:*** the option set S^t;
| Grasps $\leftarrow t-1$;

else

| ***Output:*** *"Null"*(unidentified);
| Grasps $\leftarrow t-1$;

end

4.2.2.2 Baseline approach

In order to evaluate the performance of the proposed method, a baseline method is used to compare the performance with the proposed method. Here, the baseline approach determines the subsequent grasp pose randomly without utilising the information content of the object dataset. Details of the baseline approach are summarised in Algorithm 7.

Algorithm 7: Baseline Approach: Unknown object pose

Input: The object to be identified; The descriptor of each object;

Randomly grasp the object at one of the predefined poses;

Iteration $t \leftarrow 1$;

Construct the option set S^t;

while *size(S^t)>1:* **do**

- Construct the descriptor of all options based on transformation matrices set R;
- Randomly choose a pose from the set R;
- Grasp at the selected pose;
- Obtain the corresponding grasp cluster number;
- Update option set S^t;
- Eliminate the selected pose from the set R;
- $t \leftarrow t + 1$;

end

if *size(S^t)=1:* **then**

- ***Output:*** the option set S^t;
  Grasps $\leftarrow t - 1$;

else

- ***Output:*** *"Null"*(unidentified);
  Grasps $\leftarrow t - 1$;

end

4.3 Experimental Evaluation of the Object Identification and Localisation

The experimental setup and the set of objects used in this chapter are the same as in Section 3.3.1.1 and Section 3.3.1.2, respectively.

In terms of the evaluation of the object identification and localisation, two scenarios (as introduced in Section 3.3.3) are used for validating the performance of the proposed method:

- **Scenario 1:** The grasp is carried out without uncertainty. This scenario evaluates the theoretical-best-possible outcome in the noise-free condition.
- **Scenario 2:** The grasp is carried out with ±20 degrees of grasping orientation uncertainty. In this scenario, a practical situation where the grasping uncertainty in the haptic measurement is evaluated. It should be highlighted that this uncertainty is only applied in the validation (object identification) step, meaning the grasps for object dataset identification are performed accurately at the predetermined grasp poses.

In each scenario, both the accuracy and efficiency evaluations are conducted. Details of the evaluation method are summarised in Algorithm 8 and Algorithm 9, respectively. It can be noted that the efficiency evaluation is defined as the number of grasps used to correctly identify and localise the presented object. Specifically, when there are uncertainties in the measurements, there is a possibility that the object will be misidentified or unidentified. Under such a situation, each time the algorithm does not correctly identify or localise the presented object, the algorithm will continue to run (restarted again for each failure) until the correct object is identified. The total number of grasps performed throughout, for all attempts until the object is correctly identified, are tallied as the number of grasps required, as a measure of performance of efficiency. The average over multiple attempts will reflect the typical number of grasps required by the algorithm (containing uncertainties) to correctly identify the objects.

Algorithm 8: Procedure of the accuracy evaluation

Input: The object to be identified, its true label, the number of trials, and the identification approach;

trial ← 0;

correctly_identified ← 0;

mis_identified ← 0;

un_identified ← 0;

while *trial* ≤ *number of trials:* **do**

  if *identification approach = IG with unknown object pose:* **then**

    prediction, grasps ← **Call** Algorithm 6;

  else if *identification approach = baseline with unknown object pose:* **then**

    prediction, grasps ← **Call** Algorithm 7;

  end

  if *prediction = object's true label:* **then**

    correctly_identified ← correctly_identified +1;

  else if *prediction = "Null":* **then**

    un_identified ← un_identified +1;

  else

    mis_identified ← mis_identified +1;

  end

  trial ← trial+1;

end

correctly_identified ← correctly_identified / number of trials;

mis_identified ← mis_identified / number of trials;

un_identified ← un_identified / number of trials;

Output: correctly_identified, mis_identified, un_identified;

Algorithm 9: Procedure of the efficiency evaluation

Input: The object to be identified, its true label, the number of trials, and the identification approach;

trial ← 0;

while *trial* ≤ *the number of trials:* **do**

  Identification ← $false$;

  Num_of_grasps ← 0;

  while *Identification* ! = *true:* **do**

    if *identification approach = IG with unknown object pose:* **then**

      prediction, Grasps ← **Call** Algorithm 6;

    else

      if *identification approach = baseline with unknown object pose:* **then**

        prediction, Grasps ← **Call** Algorithm 7;

      end

    end

    Num_of_grasps ← Num_of_grasps + Grasps;

    if *prediction = object's true label:* **then**

      Identification ← *true*;

    end

  end

  trial ← trial+1;

  Output: Num_of_grasps;

end

4.4 Results and Discussion of the Object Identification and Localisation

The result of the object dataset construction is shown in Table. 3.1. The descriptor of each of the 23 objects is a string for all 16 poses, preserving the spatial haptic information of that object at the corresponding poses. The accuracy and efficiency results will be presented for two scenarios. It is important to note that the identification of the object's pose is also included in the results below. For example, if the object's label is correctly predicted but the pose is identified incorrectly, then this output is still considered incorrectly identified (misidentified). However, in the figures presented, we primarily show the efficiency and accuracy results based on the object labels, which allows for a comparison of the identification performance among different objects and within a single object.

4.4.1 Scenario 1 (The grasp is carried out without uncertainty)

Accuracy Evaluation

Ten sets of identification trials for each object are conducted. In this scenario where no uncertainties were included, it was confirmed that all trials yielded a 100% accuracy for both the IG-based approach and the baseline approach, as expected.

Efficiency Evaluation

The experiment results of both the IG-based approach and the random approach are shown in Fig. 4.3. The results turn out that the proposed method improves the identification efficiency significantly compared with the random approach. The average number of grasps needed for the random approach is 7.39 while the IG-based approach was observed to correctly identify the object on average in 2.93 grasps. As the first grasp is carried out at a random pose, the number of grasps needed is less consistent compared with the efficiency result of IG based algorithm shown in Fig. 3.9 (when the object location is known). However, compared with the result of the random approach, the spread of the number of grasps required is still much narrower.

An outstanding improvement in the proposed approach compared to randomly grasping the object can be seen in the case of Objects 9 and 18. Randomly grasping these objects require on average 20.7 and 19.9 attempts respectively. The IG-based approach exploits

the information in the dataset resulting in accurate identification in only 4.70 and 8.90 grasps respectively.

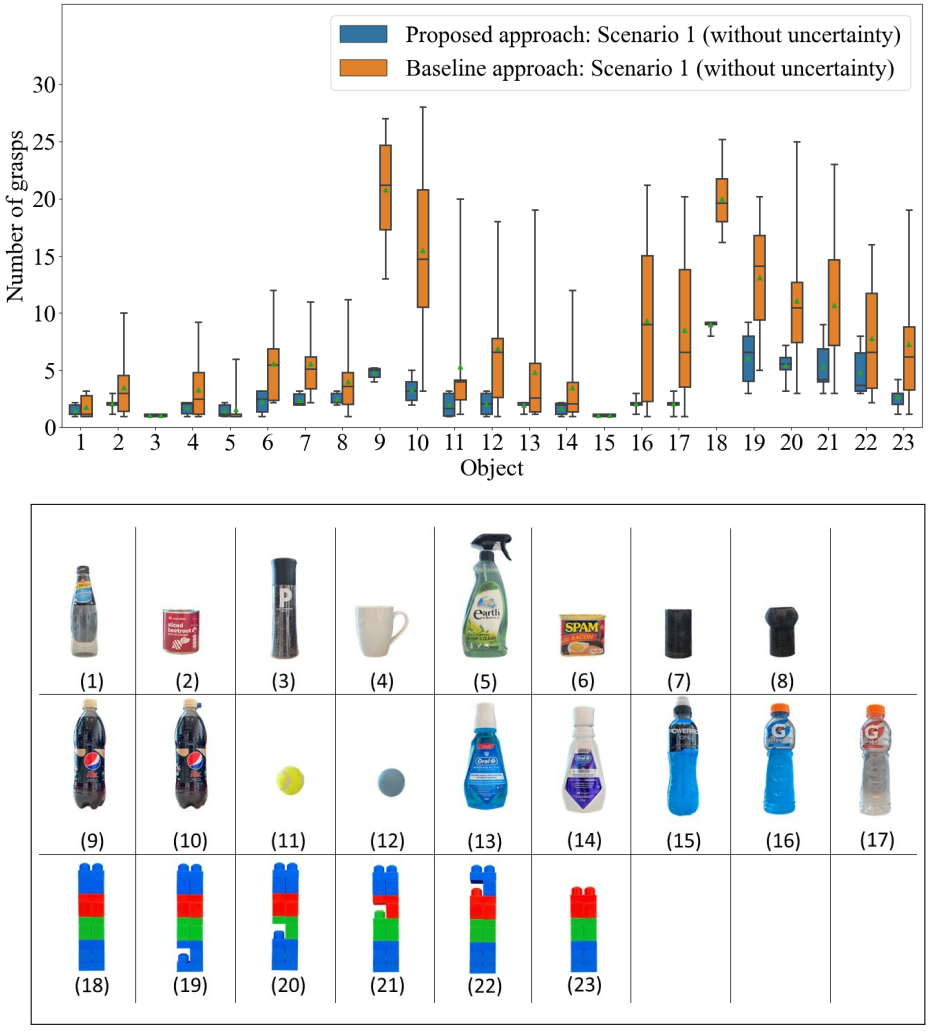


FIGURE 4.3: The number of grasps required to correctly identify each of 23 objects using the proposed approach and the baseline approach: Scenario 1 (without uncertainty).

4.4.2 Scenario 2 (The grasp is carried out with uncertainty)

Accuracy Evaluation

The identification accuracy of the proposed approach (the percentage of each object being correctly identified, misidentified and unidentified) in the presence of uncertainties is shown in Fig. 4.4. It can be seen that the cylindrical objects are not affected by the grasping orientation uncertainty and the accuracy of identifying them is all 100%. In contrast, objects with sharp edges are more prone to get the wrong identification result. This is due to the fact that the measurements of each sensor (embedded in specific locations the robotic hand/grasper used to perform the grasping and touching) will be susceptible to the deviation in the pose between the non-cylindrical object and the robotic hand. Therefore, the grasp on objects with sharp edges can result in different sensors being activated, leading to incorrect identification.

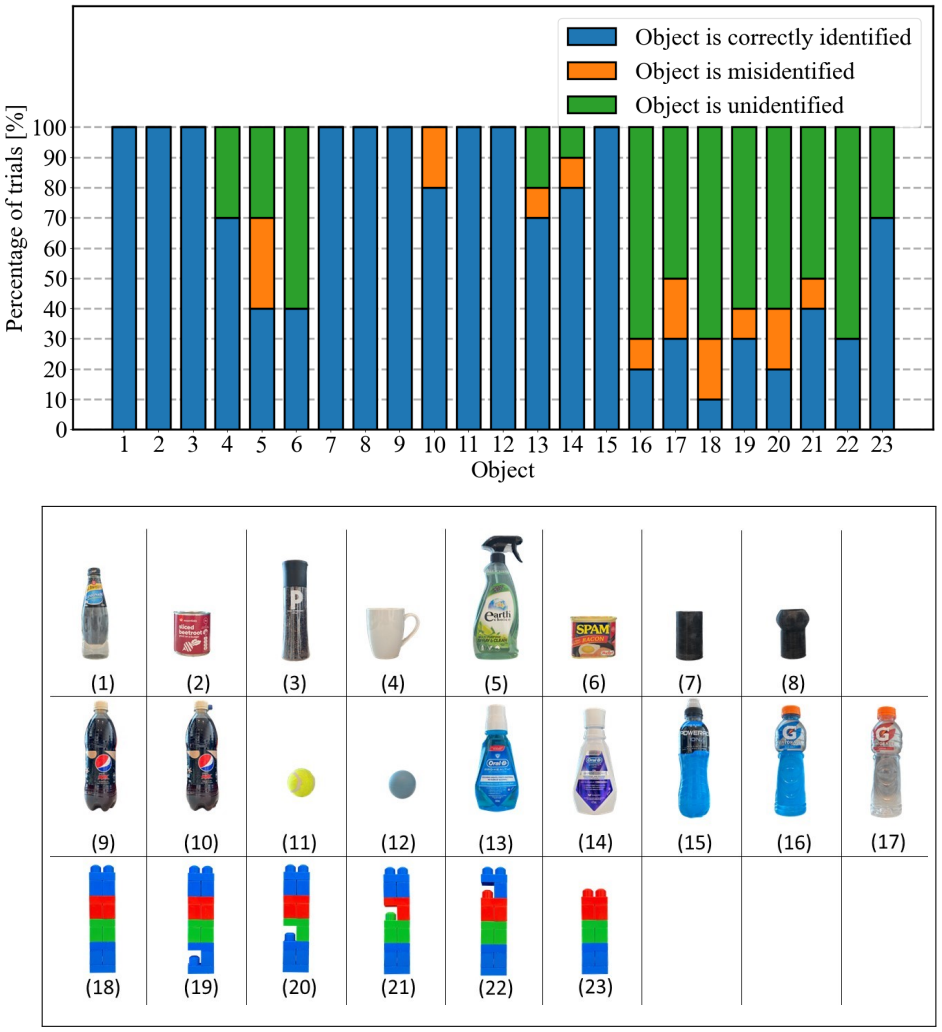


FIGURE 4.4: The accuracy of identifying each of 23 objects using the proposed approach.

Efficiency Evaluation

The efficiency results of the proposed approach and the random approach are shown in Fig. 4.5. It can be seen that the efficiency performance of the proposed approach outperforms the baseline approach. The average number of grasps needed for the IG-based approach to correctly identify the object is 4.37 while the baseline approach requires 10.33 grasps for the correct identification.

Compared to the efficiency result of Scenario 1 and Scenario 2 (as the result shown in Fig. 4.3 and Fig. 4.5), both approaches require more grasps to identify an object in Scenario 2 compared to Scenario 1. This is because, in Scenario 2, uncertainties in the measurements are more likely to lead to incorrect identification, therefore, more

attempts (thus more grasps) might be required to make a correct identification.

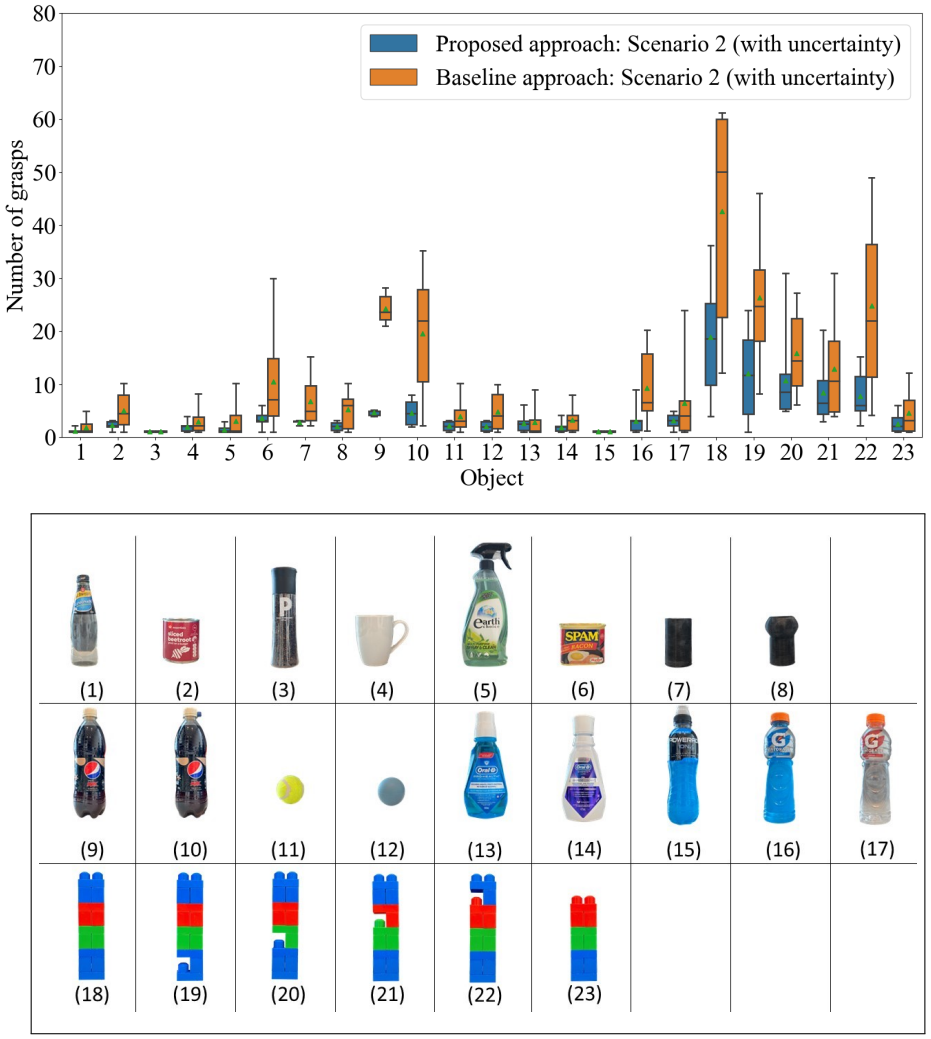


FIGURE 4.5: The number of grasps required to correctly identify each of 23 objects using the proposed approach and the baseline approach: Scenario 2 (with uncertainty).

4.5 Summary

This chapter presented an algorithm for improving the efficiency of haptic object identification with localisation. With the information on transformation matrices, the efficiency of the proposed method was achieved by using IG based algorithm to determine the displacement for the next grasp. The results showed that the number of grasps needed for correctly identifying the objects was significantly reduced compared to the baseline approach. In the experimental evaluation with 23 objects, the proposed approach required

on average 2.97 grasps to 7.14 grasps in the baseline approach when the grasp is carried out without uncertainty and 4.37 grasps to 10.33 grasps when the grasp is carried out with uncertainty. Besides, due to the inaccuracies in the identification, the number of grasps required to correctly identify an object in the presence of uncertainty will be higher compared to the scenario without uncertainty. The results show the effectiveness of the proposed method.

Chapter 5

Accuracy of Haptic Object Identification

This chapter presents a statistical modelling method which accounts for the uncertainties of the relative pose between the object to be identified and the robotic grasping hand to improve the accuracy of haptic object identification. Based on the statistical properties observed from the measurements of haptic sensors, the Beta mixture model (BMM), which is a statistical model that can accommodate probability density function with multiple peaks, has been adopted for haptic identification. Compared with deep learning-based methods, the proposed approach can overcome the limitations of deep learning, such as the potential for poor identification performance resulting from inappropriate parameters or structures. Besides, such a statistical model method provides interpretable results from physical measurements, a factor often ignored in deep learning-based methods. Furthermore, it reduces the subjective human involvement in the selection of appropriate features, making haptic identification more systematic.

In the haptic object identification process, once the dataset of the objects is established, the parameters of the BMM of each object can be determined, and the identified parameters can then be used as the prior information to compute the posterior probability of a grasp belonging to an object [89]. The knowledge of the identified BMM can improve *prior* information, leading to better identification accuracy.

5.1 Preliminaries

This section provides preliminaries, including the introduction of the Beta mixture model and Bayes' theorem.

5.1.1 Beta mixture model

5.1.1.1 Beta model

The Beta model is a family of continuous probability distributions defined for a normalised random variable parameterised by two positive shape parameters. For $x \in [0, 1]$, the probability density function (*pdf*) of a Beta model [90] is defined as:

$$\begin{aligned} f^{\boldsymbol{\theta}}_{\text{beta}}(x) &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathbf{B}(\alpha,\beta)} \\ &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du}, \end{aligned} \tag{5.1}$$

where:

- α and β are positive shape parameters of *pdf*;
- $\mathbf{B}(\alpha, \beta)$ is the function that ensures the total probability is one;
- $\boldsymbol{\theta} = [\alpha, \beta]$ represents the parameter vector.

With different parameters used, the *pdf* of the beta distribution will show different shapes (examples are shown in Fig. 5.1).

5.1.1.2 Beta mixture model (BMM)

The BMM is a combination of K Beta models, where $K \in \mathcal{N}$ is called the number of components of the BMM. Here, $\mathcal{N}_{>0}$ is the set of positive integers. The *pdf* of the BMM

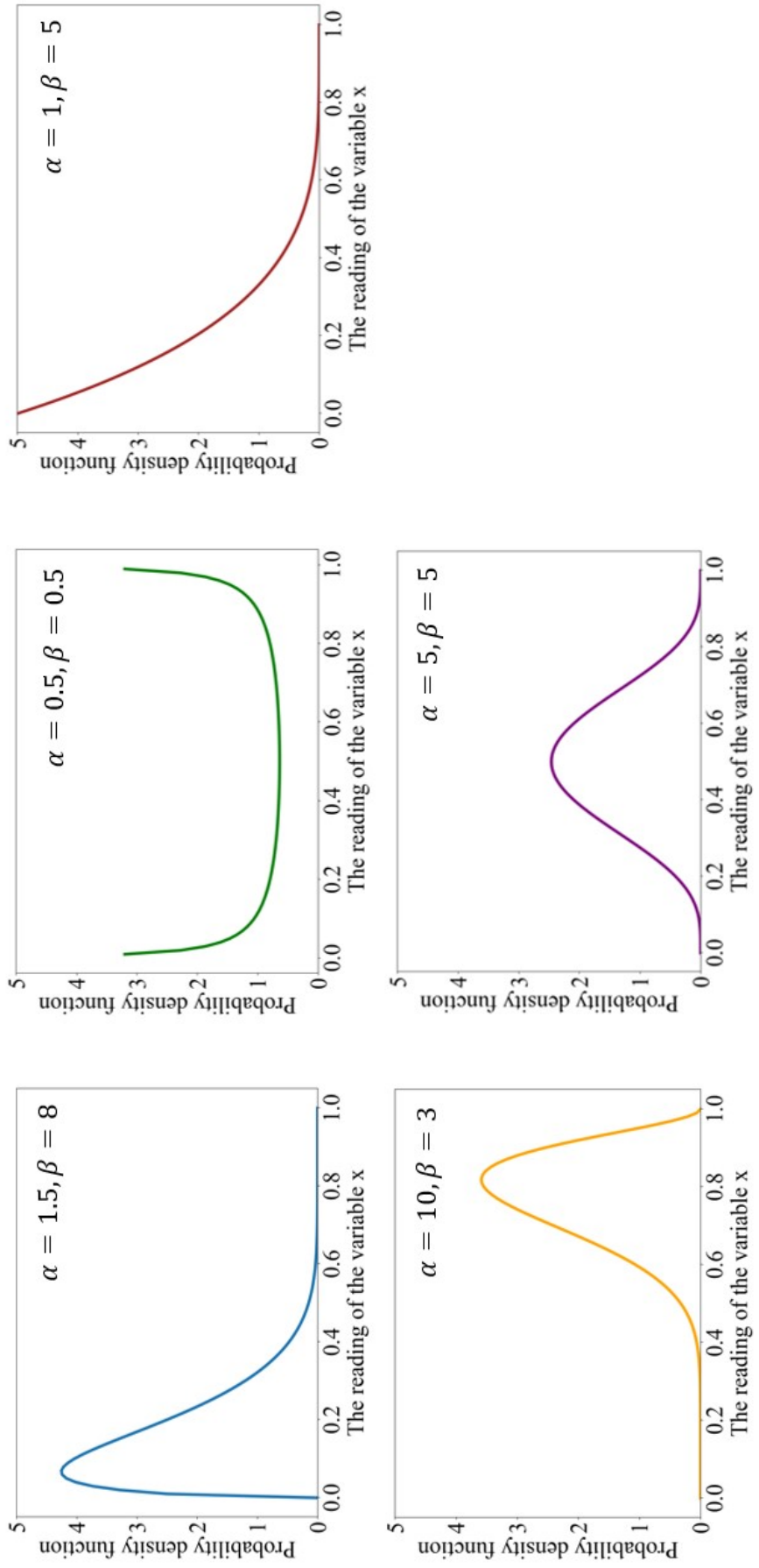


FIGURE 5.1: Examples of the beta distribution with different values of shape parameters

can be represented as:

$$\begin{aligned} f_{mix}^{\boldsymbol{\theta}_{mix}}(x) &= \sum_{m=1}^{K} \pi_m f_{\text{beta}}^{\boldsymbol{\theta}_m}(x) \\ &= \sum_{m=1}^{K} \pi_m \left[\frac{x^{\alpha_m - 1}(1-m)^{\beta_m - 1}}{\int_0^1 u^{\alpha_m - 1}(1-u)^{\beta_m - 1} du} \right], \end{aligned} \tag{5.2}$$

where:

- $\boldsymbol{\theta}_m = [\alpha_m, \beta_m]^T$ is the parameter vector of the m^{th} component of *pdf*;
- π_m is the mixture coefficient of the m^{th} component which satisfies $\pi_m \geq 0$ and $\sum_{m=1}^{K} \pi_m = 1$;
- $\boldsymbol{\theta}_{mix} = \left[\pi_1, \boldsymbol{\theta}_1^T, \pi_2, \boldsymbol{\theta}_2^T, \ldots, \pi_K, \boldsymbol{\theta}_K^T\right]^T \in LP^{3K}$ in Eq. (5.2) is the lumped parameter vector, where $LP^{3K} = \left\{ \left[\pi_1, \boldsymbol{\theta}_1^T, \pi_2, \boldsymbol{\theta}_2^T, \ldots, \pi_K, \boldsymbol{\theta}_K^T\right]^T : \sum_{m=1}^{K} \pi_m = 1, \pi_m \geq 0, \right.$ $\left. \alpha_m > 0, \beta_m > 0 \text{ for } m = 1, \ldots, K \right\}$ is the parameter space.

5.1.1.3 Maximum likelihood estimation (MLE)

MLE is a statistical method that can estimate the values for the parameters of an assumed statistical model. The parameter values are determined by maximising the likelihood function that the assumed probability distribution results in the observed data [91].

In this chapter, for a set of measurements $(x_1, ..., x_N)$ (or the dataset D), it is assumed that the measurements satisfy the BMM with the unknown $\boldsymbol{\theta}_{mix}$. The MLE method [91] has the following cost function:

$$\underset{\boldsymbol{\theta}_{mix} \in LP^{3K}}{\text{maximise}} \ \mathcal{L}(\boldsymbol{\theta}_{mix}), \tag{5.3}$$

where:

$$\mathcal{L}(\boldsymbol{\theta}_{mix}) = \sum_{i=1}^{N} \ln f_{mix}^{\boldsymbol{\theta}_{mix}}(x_i). \tag{5.4}$$

It is noted that the higher the dimension of $\boldsymbol{\theta}_{mix}$, the more data are needed to accurately estimate the parameters using the MLE. If the dataset is not rich enough, there will be a non-negligible bias in the parameter estimation.

5.1.2 Bayes' theorem

The Bayesian framework [89] can be used to calculate the probability of a random event H_i (the i^{th} event in the event space) given the prior knowledge of it and the observed evidence e. The equation can be represented as:

$$p(H_i|e) = \frac{p(e|H_i)p(H_i)}{p(e)}, \tag{5.5}$$

where:

- $p(H_i|e)$ is the posterior probability of the random event H_i given the random observation e;
- $p(e|H_i)$ is the probability likelihood of observing e given H_i;
- $p(H_i)$ is the prior probability of the i^{th} event;
- $p(e)$ represents the marginal likelihood which can be considered as a normalisation term, ensuring the posterior probability will sum to one among all events.

Based on the law of total probability [92], $p(e)$ can be calculated as:

$$p(e) = \sum_{i=1} p(e|H_i)p(H_i). \tag{5.6}$$

It is noted that Bayes' theorem can be used iteratively. For the t^{th} iteration, by observing the new evidence e_t and getting new prior probability $P(H|e_{1:t-1})$, the posterior probability $P(H|e_{1:t})$ can be continuously calculated (as shown in Eq.(5.7)) until a threshold is reached.

$$p(H|e_{1:t}) = \frac{p(e_t|H)p(H|e_{1:t-1})}{p(e_t)}. \tag{5.7}$$

5.1.3 Cramer-von Mises criterion

The Cramer-von Mises criterion can be used to judge the goodness of fit. In this chapter, it was utilised to evaluate the haptic measurements fitting performance of each sensor for each object at each segment (the concept of the segment will be introduced in Section 5.2). Generally, given the observed values $x_1, ..., x_i, ..., x_{N_m}$ arranged in increasing order, the test statistic T_{cramer} can be calculated as follows:

$$T_{cramer} = \frac{1}{12N_m} + \sum_{i=1}^{N_m} \left[\frac{2i-1}{2N_m} - F(x_i) \right]^2, \tag{5.8}$$

where:

- N_m is the number of samples;
- $F(\cdot)$ is the specified cumulative distribution function.

The smaller the value T_{cramer}, the better the fit. In this chapter, the number of samples N_m refers to the number of haptic measurements in each segment of each object. The observed value x_i is the i^{th} reading of a certain sensor in a segment of an object. $F(\cdot)$ is the cumulative distribution function of the BMMs or the Gaussian model. For each sensor, each segment of each object should have one test statistic T_{cramer}, representing the fitting performance of the Gaussian model or the BMMs at that segment. To evaluate the difference in the overall fitting performance of the Gaussian model and BMMs, the paired t-test will then be utilised where the Cramer-von Mises test statistic of each sensor at each segment of each object will be the input. Details of paired t-test will be introduced in Section 5.1.4.1.

5.1.4 Hypothesis testing

Hypothesis testing is a form of statistical inference used to evaluate the plausibility of a hypothesis based on the obtained data. In this chapter, two hypothesis testing approaches (paired t-test and two-sample proportion test) are adopted for testing the performance difference between the Gaussian model (the baseline approach) and the BMMs. For any hypothesis testing approach, the process can be divided into four steps:

1. Specify the null hypothesis and alternative hypothesis (research hypothesis).
   The null hypothesis states that there is no difference between the two groups of data, while the alternative hypothesis is the opposite of the null hypothesis. Generally, the alternative hypothesis can be one-tailed (only provides one direction, e.g., one group of data is greater or less than the other group of data) or two-tailed (provides two directions, e.g., there is a difference between two groups of data).

2. Set the significance level.
   The significance level represents the probability of accepting the alternative hypothesis when the null hypothesis is true. It is generally set as 0.01, 0.05 or 0.1.

3. Calculate the test statistic and corresponding p-value.
   In this chapter, the paired t-test and the two-sample proportion test are utilised to evaluate the distribution fitting performance difference and identification accuracy difference between the Gaussian model and the BMM. Details of the testing approaches are introduced in Sections 5.1.4.1 and 5.1.4.2. Once the test statistic is calculated, the corresponding p-value can be found by comparing the test statistic against the p-value table.

4. Analyse the results and decide whether to accept the alternative hypothesis.
   When the p-value is smaller than the significance level, the alternative hypothesis will be accepted, therefore, we can conclude that the difference between two groups of data is statistically significant (two-tailed test), or one group of data is statistically greater/less than the other group (one-tailed test).

5.1.4.1 Paired t-test

The paired t-test can be used to compare the mean of two matched groups of data [93]. In this chapter, the paired t-test is utilised to evaluate the difference in the overall distribution fitting performance. The test statistic of the paired t-test can be calculated as:

$$T_{paired} = \frac{\bar{d}}{\hat{\sigma}/\sqrt{N_z}}, \tag{5.9}$$

where:

- $\bar{d}$ is the sample mean of differences;
- $\hat{\sigma}$ is the sample standard deviation of the differences;
- N_z is the sample size.

The sample mean of the differences $\bar{d}$ can be calculated as:

$$\bar{d} = \frac{\sum_{i=1}^{N_z} d_i}{N_z}, \tag{5.10}$$

where d_i is the difference between the i^{th} paired data.
The sample standard deviation of the differences $\hat{\sigma}$ can be calculated as:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N_z}(d_i - \bar{d})^2}{N_z - 1}}. \tag{5.11}$$

In this chapter, d_i is the difference in the output of "Cramer-von Mises criterion" (as introduced in Section 5.1.3) between the Gaussian model and BMMs at the i^{th} sample and N_z should equal the total number of segments for all objects (This is because, for each sensor, each segment of each object should have one BMM/Gaussian distribution).

5.1.4.2 Two-sample proportion test

The two-sample proportion test can be used to conduct the hypothesis test about the difference between the proportions of two populations/samples [94]. In this chapter, the two-sample proportion test is used to evaluate the difference in the identification accuracy of the algorithm with the BMMs against the Gaussian model. The test statistic can be calculated as:

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}, \tag{5.12}$$

where:

- $\hat{p_1}$ and $\hat{p_2}$ are the observed proportions from sample one and sample two;
- n_1 and n_2 are the size of sample one and sample two;

- $\hat{p}$ is the pooled sample proportion, which can be calculated as:

$$p = \frac{\hat{p_1} n_1 - \hat{p_2} n_2}{n_1 + n_2}. \tag{5.13}$$

In this chapter, n_1 and n_2 refer to the number of identifications of all objects by the Gaussian model and the BMM, respectively. $\hat{p_1}$ and $\hat{p_2}$ are the number of times that all objects were correctly identified by the Gaussian model and the BMM, respectively.

5.2 Proposed Methodology

The objective of the proposed methodology is to address the challenges introduced by the uncertainties from sensors. Specifically, when a robotic hand is grasping an object, the probability density function (*pdf*) of tactile sensor measurements generally has more than one peak. With a slight amount of uncertainties between the hand and the object, the resulting *pdf* of the measurement of a tactile sensor will peak around the "no contact" case and the "firm contact" case. In this chapter, the BMM is utilised to account for such uncertainties. It should be noted that the BMM can fit the data with multiple peaks, not limited to just "no contact" and "firm contact" cases. When the "insufficient contact" occurs, generating another peak in the data, the BMM can also capture it. This capability enhances its performance in haptic object identification. The process involved in this method includes both the object dataset construction and the object identification.

5.2.1 Object dataset construction

The data of all objects are collected and normalised using the method mentioned in Section 3.2.1.1. In summary, each object will be split into N_g uniform distributed orientation intervals along the vertical axis. After grasping at N_g intervals, all parts of the object have been encapsulated by the grasp of the robotic hand. Each grasp on the object will yield an M dimensional measurement $\hat{\mathbf{z}} = \left[\hat{z}^1, ..., \hat{z}^j, ..., \hat{z}^M\right]^T$, corresponding to the number of sensors on the robotic hand. To account for the expected amount of uncertainty associated with the grasping process, in each interval, the grasp will be

carried out N_T times, and each time the grasp orientation relative to the object will be randomly determined.

Once the data for all ℓ objects are collected, the BMM will then be utilised to mathematically account for the uncertainties. It should be noted that as the focus is the identification accuracy, in this method, the grasps are performed on a single height of the object for simplicity of the experimental procedure, such as practised in [33, 95], where all objects are distinguishable from the haptic information at this height. In general, the grasp poses of objects can be extended to different heights as per the information requirement of the dataset [40]. The details of the procedure are presented below.

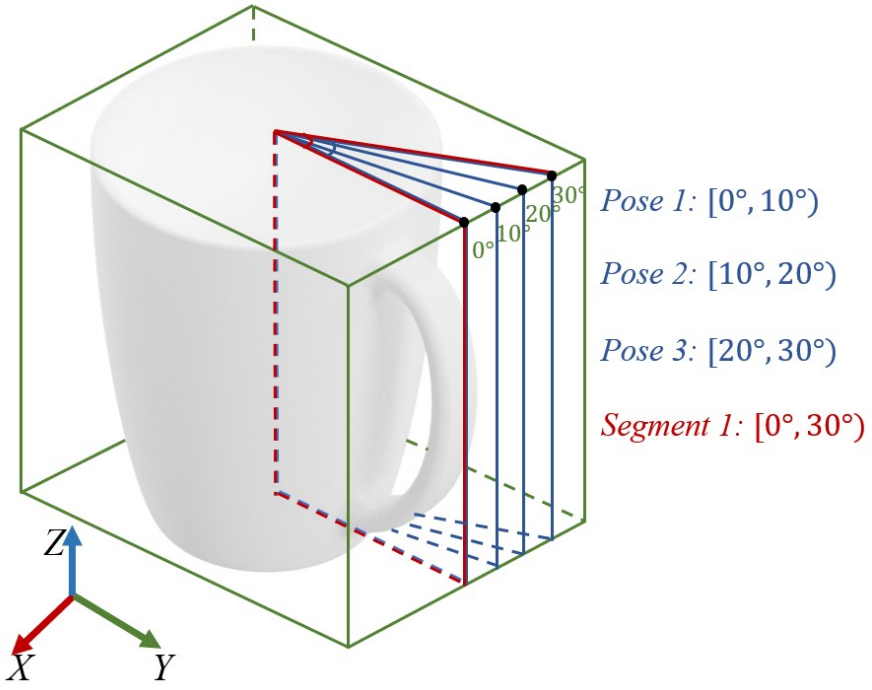


FIGURE 5.2: The characteristics of an object are regarded to be captured by grasping the object at different poses (only three poses are shown). In this illustrative example, one segment contains three poses.

For each pose of the object, the measurements of each sensor should have a statistical model to capture the uncertainties for N_T repetitions. Therefore, $\ell \times N_g \times M$ BMMs are required in total. In this proposed method, we fix N_T to simplify the data collection. Therefore, to have enough measurements to estimate the parameters of each BMM, the concept of the segment is introduced. A segment combines the measurements collected at a few poses to estimate the parameters of one BMM. The number of segments is defined as $N_s = \frac{N_g}{N_I}$ for some $N_I \in \mathcal{N}$, where N_I is the number of poses contained in

each segment N_s. Fig. 5.2 shows an example of a segment containing 3 poses (N_I=3). Details of the procedure of object dataset construction are presented below.

5.2.1.1 Re-organising data into segments

The measurements in the dataset are re-organised in terms of N_s segments. For each object, each segment will contain all measurements from N_I poses (as shown in Fig. 5.2), leading to N_M measurements for the estimation of the parameters, where $N_M = N_T \times N_I$. For the j^{th} segment s_j of object o_i, a dataset $D_{i,j} := \{\hat{\mathbf{z}}_{i,j}^1, ..., \hat{\mathbf{z}}_{i,j}^c, ..., \hat{\mathbf{z}}_{i,j}^{N_M}\}$, where $\hat{\mathbf{z}} \in \mathcal{R}^M$, represents the readings of M sensors in the collected dataset. As the subset of $D_{i,j}$, $\bar{D}_{i,j}^n$, which represents the measurements of the n^{th} sensor, will be utilised to identify the parameters of the BMM of the n^{th} sensor.

5.2.1.2 Determining the parameters in Beta mixture model

Given the dataset $\bar{D}_{i,j}^n$, parameters $\boldsymbol{\theta}_{mix}^{n,i,j}$ of the BMM can be estimated using Eq. (5.3). More precisely, the optimal parameters are select to fit $\bar{D}_{i,j}^n$ to maximise the following cost:

$$\underset{\boldsymbol{\theta}_{mix}^{n,i,j} \in LP^{3K}}{\text{maximise}} \quad \mathcal{L}(\boldsymbol{\theta}_{mix}^{n,i,j}), \tag{5.14}$$

where $\mathcal{L}$ is defined in (5.3) for the dataset $\bar{D}_{i,j}^n$.

In this proposed method, the MLE procedure is achieved by using the Constrained Optimisation BY Linear Approximation (COBYLA) algorithm [96] which is integrated into Python's *"scipy.optimize.minimize"* package. It is noted that this Python package is only for minimising a cost, therefore, to maximise the cost (as shown in Eq. (5.14)), a simple mathematical adjustment will be made, and Eq. (5.14) will then become:

$$\underset{\boldsymbol{\theta}_{mix}^{n,i,j} \in LP^{3K}}{\text{minimise}} \quad -\mathcal{L}(\boldsymbol{\theta}_{mix}^{n,i,j}). \tag{5.15}$$

It is noted that for a fixed N_T, if the number of the segments is small, fewer BMMs will be used. This will reduce the complexity of the identification algorithm, at the cost of lower identification accuracy. On the other hand, each dataset $\bar{D}_{i,j}^n$ will have sufficient

measurements to properly estimate the parameters of the BMM in LP^{3K}, leading to better estimation.

5.2.2 Object identification

In this section, the Bayesian-based method is adopted as the identification approach for testing the effectiveness of the proposed BMM.

Given the object to be identified, a testing measurement $\mathbf{z} = \left[z^1, ..., z^n, \ldots, z^M\right]^T$ can be obtained when a random grasp is performed. We can form a sequence of grasps $\mathbf{z}_{1:t} = \{\mathbf{z}_1, ..., \mathbf{z}_t\}$ in the iterative identification procedure.

5.2.2.1 Updating the posterior probability the object o_i

Given the t^{th} grasp in the identification stage, by applying the Bayesian equation in an iterative process, the probability of object o_i is computed as:

$$\begin{aligned} p(o_i|\mathbf{z}_{1:t}) &= \max_{s_j, j=1,...,N_s} \left[p(o_i, s_j|\mathbf{z}_{1:t})\right] \\ &= \max_{s_j, j=1,...,N_s} \left[\frac{p(\mathbf{z}_t|o_i, s_j)p(o_i, s_j|\mathbf{z}_{1:t-1})}{p(\mathbf{z}_t)}\right], \end{aligned} \tag{5.16}$$

where:

- $p(o_i, s_j|\mathbf{z}_{1:t})$ is the posterior probability of the object o_i at segment s_j given $\mathbf{z}_{1:t}$;
- $p(\mathbf{z}_t|o_i, s_j)$ is the likelihood of observing $\mathbf{z}_t$ given o_i and s_j;
- $p(o_i, s_j|\mathbf{z}_{1:t-1})$ is the prior probability;
- $p(\mathbf{z}_t)$ represents the marginal likelihood.

The likelihood $p(\mathbf{z}_t|o_i, s_j)$ is calculated using the *pdf* of the BMM. Assuming each dimension of the measurement is independent, according to the conditional independence

[97], $p(\mathbf{z}_t|o_i, s_j)$ can be calculated by:

$$\begin{aligned} p(\mathbf{z}_t|o_i, s_j) &= p(z_t^1, ..., z_t^n, ..., z_t^M|o_i, s_j) \\ &= \prod_{n=1}^{M} p(z_t^n|o_i, s_j) \\ &= \prod_{n=1}^{M} f_{mix}^{\boldsymbol{\theta}_{mix}^{n,i,j}}(z_t^n), \end{aligned} \tag{5.17}$$

where $f_{mix}^{\boldsymbol{\theta}_{mix}^{n,i,j}}(z^n)$ is calculated using Eq. (5.2). Parameters $\boldsymbol{\theta}_{mix}^{n,i,j}$ are identified from the constructed dataset using the MLE.

The prior probability of the current iteration t can be considered as the posterior probability of the last iteration $t-1$. For $t=1$, the prior probability of each object at each segment will be considered as a uniform distribution. Therefore, the prior probability of the first iteration is set as $p(o_i, s_j|\mathbf{z}_0) = \frac{1}{\ell N_s}$.

The marginal likelihood $p(\mathbf{z}_t)$ is calculated as:

$$p(\mathbf{z}_t) = \sum_{i=1}^{\ell} \sum_{j=1}^{N_s} p(\mathbf{z}_t|o_i, s_j) p(o_i, s_j|\mathbf{z}_{1:t-1}). \tag{5.18}$$

5.2.2.2 Stopping rule

As the Bayesian haptic object identification is an iterative process, an appropriate stopping rule is thus needed. There are two stopping rules discussed in the literature: 1) The posterior probability of the object $p(o_i|\mathbf{z}_{1:t})$ (which is calculated using Eq. (5.16)) reaches a pre-defined threshold τ_b [28, 30]; or 2) The number of grasps reaches the threshold τ_g [27, 62]. These thresholds are subjectively defined according to different needs or tasks. Both rules can be used to stop the process.

The algorithm of Bayesian haptic object identification is summarised in Algorithm 10.

Algorithm 10: Bayesian haptic object identification (presented in Section 5.2.2)

Input: The constructed object dataset (in Section 5.2.2);

The estimated parameters (in Section 5.2.2);

Initialise: The iteration t: $t \leftarrow 0$;

The stopping rule: Rule $\leftarrow \tau_g$ or τ_b;

The condition of stopping rule: Stop $\leftarrow false$;

while *Stop* $!= true$*:* **do**

  $t \leftarrow t + 1$;

  Do the t^{th} grasp and get the haptic measurements $\mathbf{z}_t$

  for *each* $o_i \in [o_1, o_\ell]$ **do**

    for *each* $s_j \in [s_1, s_{N_s}]$ **do**

      Calculate likelihood $p(\mathbf{z}_t|o_i, s_j)$ using Eq. (5.17);

      Update prior probability $p(o_i, s_j|\mathbf{z}_{1:t-1})$;

    end

  end

  Calculate marginal likelihood $p(\mathbf{z}_t)$ using Eq. (5.18);

  for *each* $o_i \in [o_1, o_\ell]$ **do**

    Calculate posterior probability $p(o_i|\mathbf{z}_{1:t})$ using Eq. (5.16);

  end

  if *Rule* $== \tau_b$ *and* $p(o_i|\mathbf{z}_t) \geq \tau_b$*:* **then**

    Stop $\leftarrow true$;

  else

    if *Rule* $== \tau_g$ *and* $t \geq \tau_g$*:* **then**

      Stop $\leftarrow true$;

    end

  end

end

predicted_object_label $\leftarrow \underset{o_i \in [o_1, o_\ell]}{\operatorname{argmax}} p(o_i|\mathbf{z}_{1:t})$;

Output: predicted_object_label;

5.3 Experimental Evaluation

To evaluate the performance of the proposed haptic object identification algorithm based on the BMM, the Reflex TakkTile robotic hand with integrated joint displacement and tactile sensors (as introduced in section 3.3.1.2) is used to perform the object identification for an object set consisting of 20 objects.

This section will present the experimental setup, followed by detailed steps of the haptic object identification algorithm for the different number of segments N_s, and the different number of components of BMMs (K in Eq. (5.2)). The performance of the proposed method with the different number of segments and different components is compared with the Bayesian identification based on the Gaussian model.

5.3.1 Experimental setup

5.3.1.1 The object set

To represent a wide range of objects with different stiffnesses, sizes, and shapes, 20 objects are selected (as shown in Table 5.1 and Fig. 5.3). Objects $1-3$, $7-10$, and 12 are the cylindrical/spherical objects, while Objects $4-6$, 11, and $13-20$ are with well defined edges. The objects are selected with reference to the Yale-CMU-Berkeley (YCB) Object and Model Set [87].

TABLE 5.1: The object set

Object Number	Object Name	Material
1	A soft drink bottle	Glass
2	A can of beetroot	Metal
3	A popcorn dispenser	Hard plastic
4	A A bottle of handwash	Soft plastic
5	A drinking mug	Ceramic
6	Another drinking mug	Ceramic
7	An arbitrarily shaped 3D-printed object	PLA plastic
8	A second soft drink bottle	Soft plastic
9	A tennis ball	Uniform felt-covered rubber compound
10	A silicone ball	Hard silicone
11	A bottle of mouthwash	Medium stiffness plastic
12	A sports drink bottle	Soft plastic
13	A Lego blocks	Hard plastic
14	Another drink bottle of a different shape	Soft plastic
15	A milk bottle	Soft plastic
16	A 3D-printed apple	PLA plastic
17	A stain remover spray bottle of non-axisymmetric shape	Medium stiffness plastic
18	A cleaning spray bottle of non-axisymmetric shape	Medium stiffness plastic
19	A can of meat	Metal
20	Another Lego block	Hard plastic

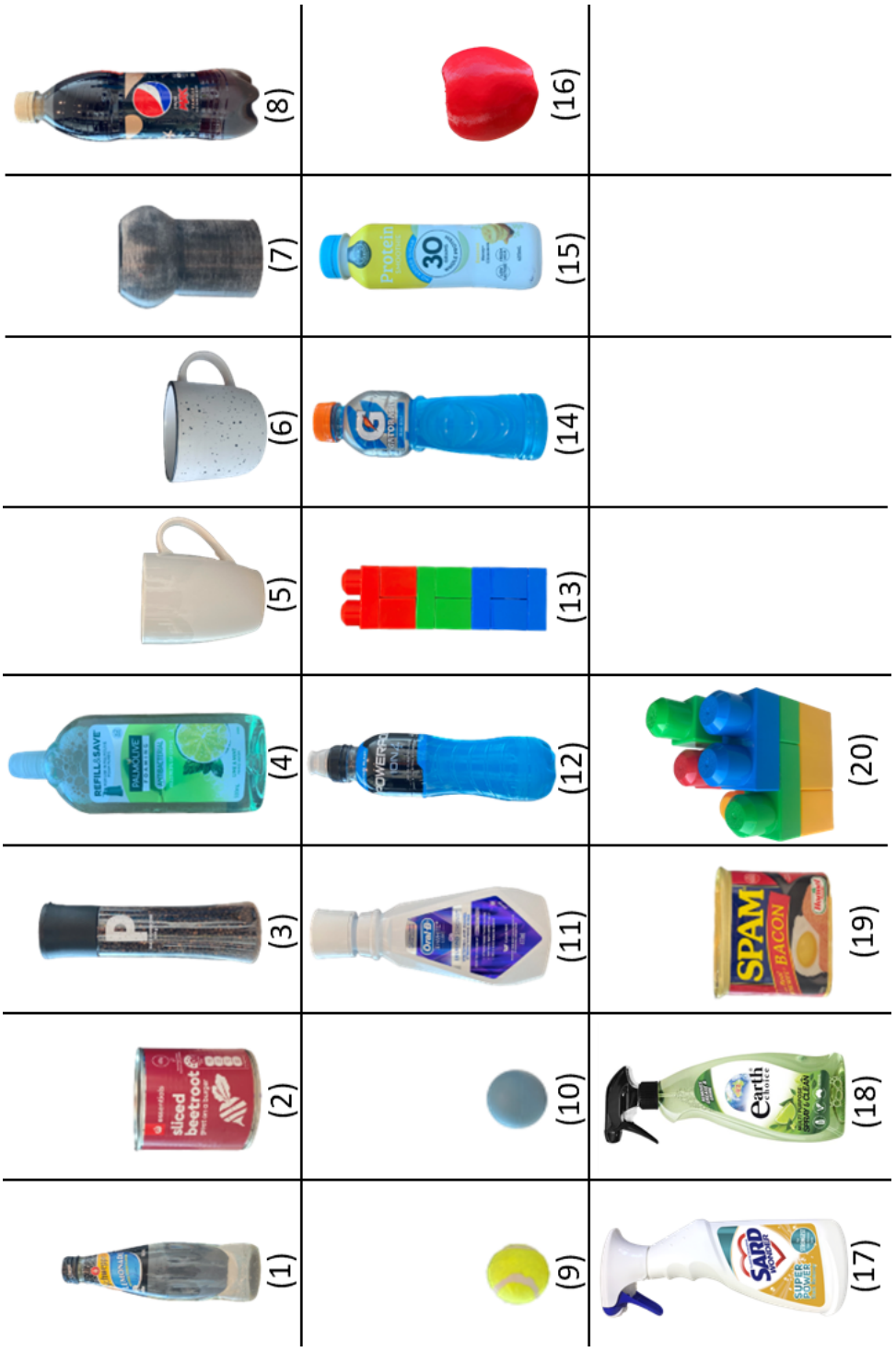


FIGURE 5.3: The list of the object set with 20 objects in a variety of shapes, sizes and stiffness.

The descriptions of the objects will be presented in the format of [explanation of the object - the material]. The material description is included to provide the physical property of the objects, while the shape is evident from Fig. 5.3. Objects used are 1) A soft drink bottle - glass, 2) A can of beetroot - metal, 3) A peppercorn dispenser - hard plastic, 4) A bottle of handwash - soft plastic, 5) A drinking mug - ceramic, 6) Another drinking mug - ceramic, 7) An arbitrarily shaped 3D-printed object - PLA plastic, 8) A second soft drink bottle - soft plastic, 9) A tennis ball - uniform felt-covered rubber compound, 10) A silicone ball - hard silicone, 11) A bottle of mouthwash - medium stiffness plastic, 12) A sports drink bottle - soft plastic, 13) A Lego block - hard plastic, 14) Another drink bottle of a different shape - soft plastic, 15) A milk bottle - soft plastic, 16) A 3D-printed apple - PLA plastic, 17) A stain remover spray bottle of non-axisymmetric shape - medium stiffness plastic, 18) A cleaning spray bottle of non-axisymmetric shape - medium stiffness plastic, 19) A can of meat - metal, 20) Another Lego block - hard plastic.

5.3.1.2 Experimental platform

The experimental platform (robotic hand, tactile sensors and software) used in this chapter is described in Section 3.3.1.2.

5.3.2 Object dataset construction

5.3.2.1 Data acquisition and normalisation

Each object $o_i \in O, i = 1, \ldots, 20$ in the object set is grasped in 36 different poses about the vertical axis (from 0 to 360^o with the resolution of 10^o). Hence, the number of poses to cover the surface is $N_g = 36$. It is noted that the reason for choosing $N_g = 36$ is that collecting the haptic measurement with 10 degrees pose of the grasping pose of the robotic hand around the object allows comparing the performance of the algorithm at the different number of segments (as 36 is divisible by a good few numbers: 36, 18, 9, 6, 4 and 2). At each pose, the grasp is repeated 3 times ($N_T = 3$), and during each grasping, the object will be rotated to make the pose face the robotic hand. The collected data are then normalised using the min-max normalisation method (as introduced in Section 3.1.1).

5.3.2.2 Estimation of parameters of Beta mixture model

As discussed in Section 5.2.1, the parameter estimation step starts by re-organising the measurements into segments. In the experiments, 6 different segment arrangements are evaluated:

- **Case 1**: 36 segments ($N_I = 1$) with 3 measurements in each segment ($N_M = 3$).
- **Case 2**: 18 segments ($N_I = 2$) with 6 measurements in each segment ($N_M = 6$).
- **Case 3**: 12 segments ($N_I = 3$) with 9 measurements in each segment ($N_M = 9$).
- **Case 4**: 6 segments ($N_I = 6$) with 18 measurements in each segment ($N_M = 18$).
- **Case 5**: 4 segments ($N_I = 9$) with 27 measurements in each segment ($N_M = 27$).
- **Case 6**: 2 segments ($N_I = 18$) with 54 measurements in each segment ($N_M = 54$).

In all 6 cases, 4 methods for statistical modelling are used:

- **Method 1**: Gaussian model (the baseline method). The *pdf* of the Gaussian model is defined as:

$$f_{norm}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{5.19}$$

  where μ and σ^2, denoting the mean and variance of the Gaussian model, are parameters that need to be identified.

- **Method 2**: BMM ($K = 1$). 2 parameters to identify.
- **Method 3**: BMM ($K = 2$). 5 parameters to identify.
- **Method 4**: BMM ($K = 3$). 8 parameters to identify.

5.3.3 Object identification

Methods $1 - 4$ are used to identify the object. Each object will be identified 6 times, therefore, 6 trials of identification for each object are conducted for the identification methods described. The accuracy measure is obtained by taking the average of the

number of times the algorithm correctly identifies the object out of the total 6 attempts. As the focus is to evaluate the performance of the BMM for identifying the object, in each identification trial, the threshold τ_g of the stopping rule is set to 1, to simplify the evaluation procedure without loss of generality.

5.4 Results and Discussion

5.4.1 Object dataset construction

To evaluate the haptic measurements fitting performance of each sensor for each object at each segment, the "Cramer–von Mises criterion" is adopted. The "paired t-test" is then used to statistically compare the overall performance of the BMMs and the Gaussian model for each segment. Since there are 33 sensors measuring over the different number of segments on 20 objects, as an illustrative result, only the performance of different methods at 6 segments (Case 4) is represented. For all three BMM methods, according to the p-value calculated using the "paired t-test", the BMMs demonstrated a closer fit than that of Gaussian at a significant level of 0.001. A representative distribution of tactile measurements is given in Fig. 5.4, which specifically shows the tactile measurements of the third tactile sensor of the first finger when grasping Object 4. Two peaks are observed, corresponding to the "no contact" and the "firm contact" cases, respectively. This is widely observed in the measurements of tactile sensors in this experiment. The resulting distribution modelled by a BMM (in this case, at $K = 2$) is shown in the figure, as well as a comparison to that with a Gaussian model.

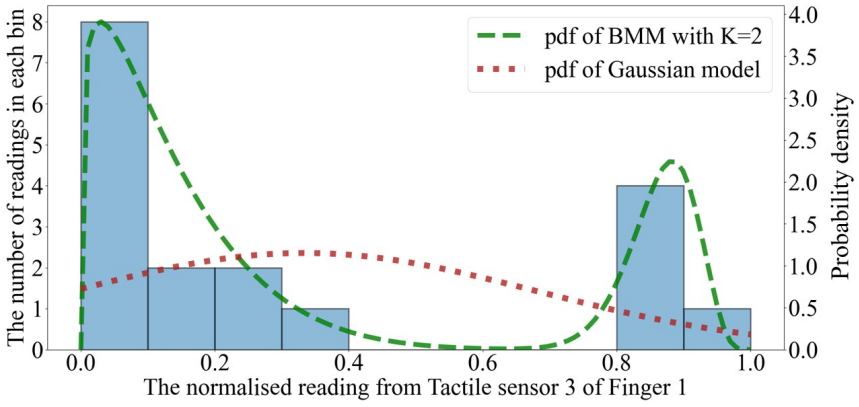


FIGURE 5.4: The fitting result of the BMM ($K = 2$) and the Gaussian model. The histogram shows the readings from Tactile sensor 3 of Finger 1 in the first segment of Object 4 (a bottle of handwash).

5.4.2 Object identification

The complete results of the experiment, comparing the accuracy of the haptic object identification algorithm using the Gaussian model and BMMs for all cases, are shown in Table 5.2. The averaged accuracy is presented in Fig. 5.5 for ease of comparison. The p-values of the identification accuracy performance of the algorithm with the proposed BMMs against the Gaussian model, based on the "two-sample proportion test", are presented in Table 5.3 to indicate the statistical significance of the difference in performance.

TABLE 5.2: Experimental results: the identification accuracy for all 6 cases and 4 methods[1].

	Method 1: Gaussian Model						**Method 2: BMM with $K = 1$**					
Object	Case 1 36 segments	Case 2 18 segments	Case 3 12 segments	Case 4 6 segments	Case 5 4 segments	Case 6 2 segments	Case 1 36 segments	Case 2 18 segments	Case 3 12 segments	Case 4 6 segments	Case 5 4 segments	Case 6 2 segments
4	66.7%	83.3%	66.7%	66.7%	100.0%	33.3%	100.0%	100.0%	100.0%	100.0%	100.0%	50.0%
5	83.3%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
6	66.7%	66.7%	50.0%	100.0%	100.0%	100.0%	83.3%	100.0%	100.0%	100.0%	83.3%	100.0%
11	33.3%	66.7%	66.7%	66.7%	33.3%	16.7%	83.3%	50.0%	33.3%	33.3%	66.7%	16.7%
13	66.7%	66.7%	83.3%	66.7%	66.7%	50.0%	100.0%	83.3%	83.3%	83.3%	83.3%	83.3%
14	66.7%	83.3%	83.3%	83.3%	66.7%	100.0%	83.3%	100.0%	83.3%	83.3%	100.0%	100.0%
15	33.3%	83.3%	50.0%	50.0%	50.0%	33.3%	66.7%	100.0%	100.0%	100.0%	100.0%	100.0%
16	83.3%	83.3%	100.0%	100.0%	83.3%	83.3%	100.0%	83.3%	100.0%	100.0%	100.0%	100.0%
17	33.3%	33.3%	33.3%	33.3%	33.3%	16.7%	66.7%	83.3%	83.3%	66.7%	50.0%	33.3%
18	66.7%	66.7%	83.3%	83.3%	83.3%	33.3%	83.3%	100.0%	83.3%	50.0%	66.7%	16.7%
19	83.3%	83.3%	83.3%	66.7%	50.0%	66.7%	83.3%	88.3%	50.0%	50.0%	50.0%	66.7%
20	66.7%	66.7%	83.3%	50.0%	50.0%	33.3%	100.0%	100.0%	100.0%	100.0%	100.0%	83.3%
Avg_nc [2]	62.5%	**73.6%**	**73.6%**	72.2%	66.7%	55.6%	87.5%	**90.3%**	84.7%	80.6%	83.3%	70.8%
Std_nc [2]	0.174	0.153	0.190	0.200	0.253	0.335	0.115	0.138	0.201	0.225	0.185	0.307
Avg_all [3]	77.5%	**84.2%**	**84.2%**	83.3%	80.0%	73.4%	92.5%**	**94.2%****	90.8%*	88.4%	90.0%**	82.5%**
Std_all [3]	0.231	0.179	0.200	0.211	0.261	0.347	0.111	0.121	0.179	0.205	0.170	0.286

[1] For all 6 cases and 4 methods, the identification accuracy of all cylindrical and spherical objects (Objects $1-3$, $7-10$, and 12) is 100%. Hence only the results for non-cylindrical objects (Objects $4-6$, 11, and $13-20$) are shown.

[2] Avg_nc and Std_nc denote the average and standard deviation of the identification accuracy of non-cylindrical/spherical objects, respectively.

[3] Avg_all and Std_all denote the average and standard deviation of the identification accuracy of all objects, respectively.

Note: The values with '**' and '*' represent the identification accuracy of that method is statistically higher than that of the Gaussian model (Method 1) in the corresponding case at significance level 0.05 and 0.1 respectively.

Object	Method 3: BMM with $K = 2$						Method 4: BMM with $K = 3$					
	Case 1 36 segments	Case 2 18 segments	Case 3 12 segments	Case 4 6 segments	Case 5 4 segments	Case 6 2 segments	Case 1 36 segments	Case 2 18 segments	Case 3 12 segments	Case 4 6 segments	Case 5 4 segments	Case 6 2 segments
4	–	83.3%	83.3%	100.0%	66.7%	66.7%	–	–	66.7%	100.0%	83.3%	33.3%
5	–	100.0%	100.0%	100.0%	100.0%	100.0%	–	–	100.0%	100.0%	100.0%	83.3%
6	–	83.3%	83.3%	66.7%	50.0%	33.3%	–	–	100.0%	100.0%	100.0%	83.3%
11	–	66.7%	50.0%	50.0%	50.0%	33.3%	–	–	33.3%	50.0%	66.7%	33.3%
13	–	83.3%	100.0%	100.0%	83.3%	66.7%	–	–	83.3%	83.3%	66.7%	16.7%
14	–	100.0%	100.0%	100.0%	100.0%	100.0%	–	–	100.0%	100.0%	100.0%	100.0%
15	–	100.0%	100.0%	100.0%	100.0%	100.0%	–	–	100.0%	100.0%	100.0%	100.0%
16	–	100.0%	100.0%	100.0%	100.0%	66.7%	–	–	100.0%	100.0%	100.0%	100.0%
17	–	83.3%	100.0%	100.0%	100.0%	100.0%	–	–	100.0%	66.7%	33.3%	50.0%
18	–	83.3%	83.3%	100.0%	100.0%	83.3%	–	–	83.3%	83.3%	83.3%	83.3%
19	–	83.3%	50.0%	50.0%	50.0%	50.0%	–	–	83.3%	83.3%	66.7%	50.0%
20	–	83.3%	83.3%	66.7%	100.0%	83.3%	–	–	100.0%	100.0%	50.0%	50.0%
Avg_nc [2]	–	**87.5%**	86.1%	84.7%	83.3%	72.2%	–	–	87.5%	**88.9%**	79.1%	65.2%
Std_nc [2]	–	0.095	0.171	0.190	0.207	0.220	–	–	0.186	0.151	0.208	0.273
Avg_all [3]	–	**92.5%****	91.7%**	90.8%**	90.0%**	83.3%**	–	–	92.5%**	**93.3%****	87.5%*	83.3%**
Std_all [3]	–	0.098	0.154	0.171	0.186	0.223	–	–	0.162	0.133	0.196	0.278

[1] For all 6 cases and 4 methods, the identification accuracy of all cylindrical and spherical objects (Objects $1-3$, $7-10$, and 12) is 100%. Hence only the results for non-cylindrical objects (Objects $4-6$, 11, and $13-20$) are shown.

[2] Avg_nc and Std_nc denote the average and standard deviation of the identification accuracy of non-cylindrical/spherical objects, respectively.

[3] Avg_all and Std_all denote the average and standard deviation of the identification accuracy of all objects, respectively.

Note: The values with '**' and '*' represent the identification accuracy of that method is statistically higher than that of the Gaussian model (Method 1) in the corresponding case at significance level 0.05 and 0.1 respectively.

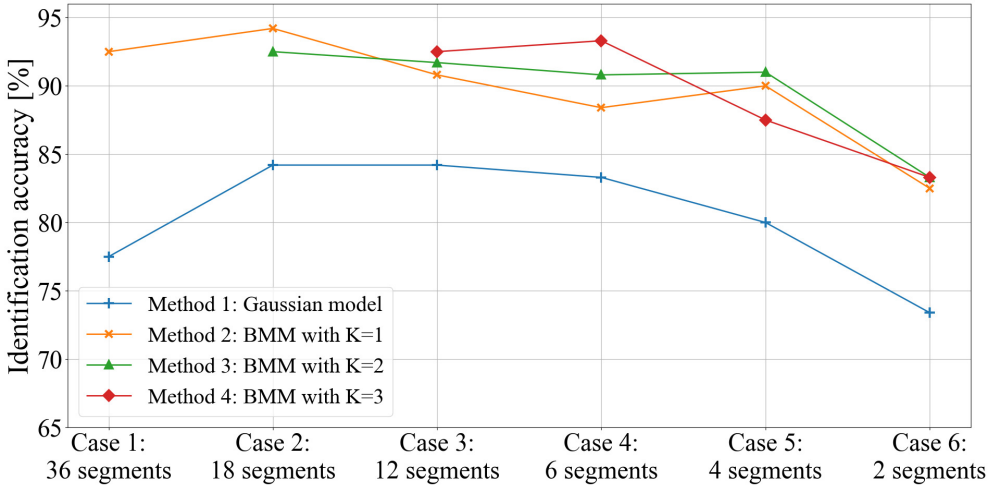


FIGURE 5.5: The averaged identification accuracy of all objects in four methods and decreasing number of segments.

TABLE 5.3: Two-sample proportion test result comparing Gaussian and BMM on identification accuracy (one-tailed p-value)[1].

	Method 1 and 2	Method 1 and 3	Method 1 and 4
p-value	5.05×10^{-7}	2.34×10^{-6}	3.36×10^{-5}

[1] The p-value is calculated using all 20 objects in the object set and the identification result of all cases.

From Fig. 5.5, it can be observed that the Gaussian model (Method 1) has the lowest identification accuracy on average and BMMs (Methods 2-4) outperform the Gaussian model in all cases. The statistical significance of the difference in performance for Methods 2-4, compared to Method 1, is validated by the p-values presented in Table 5.3. This supports our hypothesis that the BMM more accurately captures the distribution of the uncertainties typical in object grasping than the Gaussian model. From Fig. 5.5, the identification performance can be observed to drop with the lower number of segments for all methods. This is because a lower number of segments (to represent each object) means that each Gaussian or BMM would need to capture a wider range of variations in the object, while a higher number of segments per object allows each model to capture a localised area on an object more accurately (thus resulting in more accurate identification performance). It should be noted that a higher number of segments also means a higher number of measurements needed, as there would be a fixed number of parameters that need to be identified per segment. In our experiment, a fixed number of overall measurements was collected, thus they are divided among the different number of segments for different cases. The cases with a higher number of segments therefore will have

fewer measurements assigned per segment. In this case, it can be observed that for the highest number of segments and higher number of components of the BMMs (Methods 3 and 4, corresponding to $K = 2, 3$, respectively), there were insufficient measurements to produce the estimation of the parameters of the BMMs. Thus in Fig. 5.5 and Table 5.2, it can be observed that these results were not available.

According to Table 5.2, for all 4 methods (the BMMs and the Gaussian model), the identification accuracy of cylindrical/spherical objects (Objects $1-3$, $7-10$, and 12) was 100%. As expected, due to their cylindrical symmetric feature, the measurements were not affected by uncertainties. The primary source of noise in these measurements is sensor noise, which is bounded within a range that is much smaller than the variance in the distinguishing information between objects. In contrast, the better performance of the BMMs relative to the Gaussian model can be found in identifying objects with sharp, clear, or well-defined edges (Objects $4-6$, 11, and $13-20$). Due to the nature of the tactile sensors used in our experiment, individual sensors are embedded in the fingers and only measure local forces applied near their location (this is a typical nature of tactile sensors, observed in many robotic hands [98, 99]), a change in the grasping position will cause a sharp angular feature on an object to press on the adjacent tactile sensor, activating a completely different sensor in the resulting reading. Therefore, the measurement distribution of each tactile sensor will contain multiple peaks, and BMMs are able to fit such distribution better than the Gaussian model.

It should be clarified that this chapter only focuses on the identification problem (identifying the specific object from a given object set) rather than the classification (categorising the object into a certain class). However, if the object to be identified is not in the set, the following scenarios might occur: 1) If the object to be identified is not similar to any objects in the set, the posterior probability of being any objects in the set will be small. When using the stopping rule τ_b and τ_g, if the posterior probability cannot reach the threshold τ_b after τ_g grasps, then the algorithm can conclude that the object is not the one in the constructed object set. The object then can be added to the object set, hence it will be able to be identified next time; 2) If the object to be identified is similar to one object in the set, then this object would be predicted as the most similar object in the given object set.

5.5 Summary

In this chapter, we utilised the Beta mixture model (BMM) to characterise the uncertainties typical to haptic object identification using robotic hand grasps. The BMMs were shown to be effective in representing the distribution of the tactile measurements due to the pose uncertainties relative to the grasping hand. The advantage of the BMM over the Gaussian model is clearly shown in the identification of objects with well-defined edges. The higher number of segments used to represent an object (N_s) was found to improve the precision of the model of the probability density function and this in turn results in higher accuracy in the haptic object identification. It was noted that as the model of the probability density function for each segment requires a set number of parameters to be estimated, the number of measurements (the number of times the object is grasped and measurements taken) required to construct the model for an object increases with the number of segments used to represent it. The study did not find any significant difference in the resulting identification performance as a function of the number of components (K) in the BMM.

Chapter 6

Conclusion

This thesis investigated haptic object identification from the following two topics: a) Identifying the object from a given set using the minimum number of grasps (the efficiency of haptic object identification); b) Accounting for the uncertainties characteristic to the haptic measurements of robot hand mounted tactile sensors in improving the accuracy of the object identification.

6.1 Contributions

This thesis answered two proposed research questions. Regarding the efficiency of haptic object identification, an information gain-based approach is proposed to determine the grasp pose to minimise the number of grasps needed. For identification accuracy, a statistical method based on the Beta mixture model is proposed to characterise the uncertainties in the haptic measurements. Specifically, regarding the first research question, to improve the efficiency of haptic identification, an object descriptor construction method to preserve the spatial haptic information of the object was proposed, and an information gain-based identification method was developed to determine where to grasp to get the most distinguishing information. Two scenarios were considered: one assumes that the object poses are known and the other relaxes this assumption (this assumption, thus requires an implementation of simultaneously object identification and localisation). The results showed that the number of grasps needed for identification can be significantly reduced for both scenarios compared to an equivalent haptic object identification

approach without such information. It was also observed that the uncertainties, which are the positioning deviation of the grasper relative to the object, increase the number of attempts to achieve the same number of correct identification.

Secondly, to address the problem caused by the uncertainties of the relative pose between the object to be identified and the robotic grasping hand (to improve the identification accuracy), the Beta mixture model is utilised to account for the uncertainties in the haptic measurements. The results showed that the Beta mixture model clearly outperforms the Gaussian model in identifying objects with well-defined edges. Besides, increasing the number of segments into which an object is divided, i.e., implementing finer segmentation, would result in greater identification accuracy because there would be fewer uncertainties in each segment.

Enhancing efficiency and accuracy in haptic identification makes this research important for applications such as automation and prosthetics. The methods proposed in the thesis offer innovative solutions to the challenges of haptic object identification, laying the foundation for future research in this area.

6.2 Limitations and Future Work

An assumption used throughout this thesis is that the orientation of the object is assumed to be known as *priori*. Specifically, all the objects in the set will always be positioned upright (the z-axis of an object will always be perpendicular to the support surface). For example, as the object (mug) shown in Fig. 3.2, it is assumed to remain standing on the table and will not fall down during identification. This assumption helps to simplify the problem of object identification and to avoid the effect of other variables on the core research objectives (the efficiency and accuracy of haptic object identification).

Although this assumption is commonly used in the literature, such as [33, 100], it limits the applicability to some well-controlled problems, such as where objects are well placed on a surface, such as a shelf or a table. It will face challenges in scenarios where the objects are placed in an arbitrary orientation, for example, identifying a book or a phone in the bag. The reason is that the constructed object dataset only has the data or features at that specific orientation. Therefore, the online identification measurements

collected at other orientations of the object cannot be processed to match the data or the features recorded in the offline dataset, hence the identification is very likely to be inaccurate.

Future works can focus on relaxing this assumption to make the haptic identification algorithms applicable. There might be three potential solutions that can be explored for this problem:

- Expanding the offline object dataset. When constructing the object dataset, the measurements of each object can be systemically collected at different orientations. Therefore, the corresponding features can be matched in the dataset whenever the object is identified at which orientation. However, this approach might need a significant amount of data, which would be time-consuming. A potential method to solve this problem would be "sim-to-real" [101, 102], which constructs the offline dataset using the simulation model. The performance of this method depends on how similar the simulated model is to the real robotic hand.

- Finding the reference orientation during online identification. This approach aims to determine the orientation of the object first during the online identification step. Once the orientation is determined, an object is able to be identified by contacting the poses whose measurements are included in the offline dataset. However, the possible problem with this approach is that as the haptic information is sparse, determining the orientation of the object will require many contacts, which is contrary to the evaluation criteria of efficiency.

- Designing the features of objects that are invariant to the orientation. This approach characterises the object with features that have distinguishable information and do not change with orientation. However, based on the sparse haptic measurement obtained from one contact, it is hard to find features that have similar values for the same object at different orientations and very different values for different objects.

The robustness analysis of the unsupervised clustering techniques in haptic object identification could be another future work that can be focused on. The data-driven unsupervised clustering techniques play an important role in haptic identification for labelling

[25, 49]. For example, in Chapter 3, the unsupervised clustering technique K-Means was used to categorise the offline collected measurements. However, it was noticed that when doing the identification, the online measurements have uncertainties from inaccurate grasp positions or sensor noise. When the offline clustering results are sensitive to noise/uncertainties coming from the sensors, the online measurement is very likely to be categorised into the wrong cluster, hence the accuracy of identification will be reduced.

A potential solution for this problem would be assigning different weights to different sensors or features to account for the measurement uncertainties/noise. The assignment of weights depends on two metrics: the importance of the feature (weight of the importance of each feature) and the ability of the uncertainties/noise to affect the clustering results (weight of uncertainty). The weight of the importance might be able to be achieved by utilising the existing unsupervised feature selection method, such as the Laplacian score method [103]. The weight of uncertainty could be determined by the ratio between the variance of the measurement uncertainties and the uncertainties threshold identified in the unsupervised clustering technique. The combination of these two weights will determine the overall weight of each sensor/feature used in the process of offline data collection and online object identification.

Bibliography

[1] S. J. Lederman and R. L. Klatzky, "Hand movements: A window into haptic object recognition," *Cognitive Psychology*, vol. 19, no. 3, pp. 342–368, 1987.

[2] R. R. Ma, J. T. Belter, and A. M. Dollar, "Hybrid deposition manufacturing: design strategies for multimaterial mechanisms via three-dimensional printing and material deposition," *Journal of Mechanisms and Robotics*, vol. 7, no. 2, p. 021002, 2015.

[3] R. LABS, "Technical specification of the reflex takktile hand." https://www.labs.righthandrobotics.com/reflexhand.

[4] C. Weiss and J. F. Disterhoft, "Cognitive neuroscience," in *Encyclopedia of Social Measurement*, pp. 341–349, New York: Elsevier, 2005.

[5] K. Segaert, G. E. Nygård, and J. Wagemans, "Identification of everyday objects on the basis of kinetic contours," *Vision Research*, vol. 49, no. 4, pp. 417–428, 2009.

[6] A. A. Shahid, J. S. V. Sesin, D. Pecioski, F. Braghin, D. Piga, and L. Roveda, "Decentralized multi-agent control of a manipulator in continuous task learning," *Applied Sciences*, vol. 11, no. 21, p. 10227, 2021.

[7] H. Van Hoof, T. Hermans, G. Neumann, and J. Peters, "Learning robot in-hand manipulation with tactile features," *In Proceedings of International Conference on Humanoid Robots*, pp. 121–127, 2015.

[8] F. Vicentini *et al.*, "Piros: Cooperative, safe and reconfigurable robotic companion for cnc pallets load/unload stations," in *Bringing Innovative Robotic Technologies from Research Labs to Industrial End-users*, pp. 57–96, Springer, 2020.

[9] F. Chen, F. Cannella, C. Canali, T. Hauptman, G. Sofia, and D. Caldwell, "In-hand precise twisting and positioning by a novel dexterous robotic gripper for industrial high-speed assembly," *In Proceedings of International Conference on Robotics and Automation*, pp. 270–275, 2014.

[10] L. Acosta, E. González, J. N. Rodríguez, A. F. Hamilton, *et al.*, "Design and implementation of a service robot for a restaurant," *International Journal of Robotics and Automation*, vol. 21, no. 4, p. 273, 2006.

[11] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," *In Proceedings of International Conference on Computer Vision*, vol. 3, pp. 1403–1403, 2003.

[12] K. Overvliet, J. B. Smeets, and E. Brenner, "The use of proprioception and tactile information in haptic search," *Acta Psychologica*, vol. 129, no. 1, pp. 83–90, 2008.

[13] M. Grunwald, *Human haptic perception: Basics and applications*. Springer Science & Business Media, 2008.

[14] O. Azulay, I. Ben-David, and A. Sintov, "Learning haptic-based object pose estimation for in-hand manipulation control with underactuated robotic hands," *IEEE Transactions on Haptics*, vol. 16, no. 1, pp. 73–85, 2022.

[15] B. Hannaford and A. M. Okamura, "Haptics," in *Springer Handbook of Robotics*, pp. 1063–1084, Springer, 2016.

[16] W. Shaw Cortez, D. Oetomo, C. Manzie, and P. Choong, "Robust object manipulation for tactile-based blind grasping," *Control Engineering Practice*, vol. 92, p. 104136, 2019.

[17] L. Scimeca, P. Maiolino, E. Bray, and F. Iida, "Structuring of tactile sensory information for category formation in robotics palpation," *Autonomous Robots*, pp. 1–17, 2020.

[18] J. Dargahi and S. Najarian, "Human tactile perception as a standard for artificial tactile sensing—a review," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 1, no. 1, pp. 23–35, 2004.

[19] H. Liu, Y. Wu, F. Sun, and G. Di, "Recent progress on tactile object recognition," *International Journal of Advanced Robotic Systems*, vol. 14, no. 4, pp. 1–12, 2017.

[20] A. Faragasso, J. Bimbo, A. Stilli, H. A. Wurdemann, K. Althoefer, and H. Asama, "Real-time vision-based stiffness mapping," *Sensors*, vol. 18, no. 5, p. 1347, 2018.

[21] H. Gu, S. Fan, H. Zong, M. Jin, and H. Liu, "Haptic perception of unknown object by robot hand: Exploration strategy and recognition approach," *International Journal of Humanoid Robotics*, vol. 13, no. 03, p. 1650008, 2016.

[22] S. Funabashi, S. Morikuni, A. Geier, A. Schmitz, S. Ogasa, T. P. Torno, S. Somlor, and S. Sugano, "Object recognition through active sensing using a multi-fingered robot hand with 3D tactile sensors," *In Proceedings of International Conference on Intelligent Robots and Systems*, pp. 2589–2595, 2018.

[23] S. Funabashi, G. Yan, A. Geier, A. Schmitz, T. Ogata, and S. Sugano, "Morphology-specific convolutional neural networks for tactile object recognition with a multi-fingered hand," *In Proceedings of International Conference on Robotics and Automation*, pp. 57–63, 2019.

[24] A. Schmitz, Y. Bansho, K. Noda, H. Iwata, T. Ogata, and S. Sugano, "Tactile object recognition using deep learning and dropout," *In Proceedings of International Conference on Humanoid Robots*, pp. 1044–1050, 2015.

[25] S. Luo, W. Mou, K. Althoefer, and H. Liu, "Novel tactile-SIFT descriptor for object shape recognition," *IEEE Sens. J.*, vol. 15, no. 9, pp. 5001–5009, 2015.

[26] M. M. Zhang, M. D. Kennedy, M. A. Hsieh, and K. Daniilidis, "A triangle histogram for object classification by tactile sensing," *In Proceedings of International Conference on Intelligent Robots and Systems*, pp. 4931–4938, 2016.

[27] J. A. Fishel and G. E. Loeb, "Bayesian exploration for intelligent identification of textures," *Frontiers in Neurorobotics*, vol. 6, p. 4, 2012.

[28] D. Xu, G. E. Loeb, and J. A. Fishel, "Tactile identification of objects using bayesian exploration," *In Proceedings of International Conference on Robotics and Automation*, pp. 3056–3061, 2013.

[29] P. Dallaire, P. Giguère, D. Émond, and B. Chaib-Draa, "Autonomous tactile perception: A combined improved sensing and bayesian nonparametric approach," *Robotics and Autonomous Systems*, vol. 62, no. 4, pp. 422–435, 2014.

[30] M. Kaboli, K. Yao, D. Feng, and G. Cheng, "Tactile-based active object discrimination and target object search in an unknown workspace," *Autonomous Robots*, vol. 43, no. 1, pp. 123–152, 2019.

[31] J. M. Romano and K. J. Kuchenbecker, "Methods for robotic tool-mediated haptic surface recognition," *In Proceedings of Haptics Symposium*, pp. 49–56, 2014.

[32] N. Jamali and C. Sammut, "Majority voting: Material classification by tactile sensing using surface texture," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 508–521, 2011.

[33] A. J. Spiers, M. V. Liarokapis, B. Calli, and A. M. Dollar, "Single-grasp object classification and feature extraction with simple robot hands and tactile sensors," *IEEE Transactions on Haptics*, vol. 9, no. 2, pp. 207–220, 2016.

[34] S. Chitta, J. Sturm, M. Piccoli, and W. Burgard, "Tactile sensing for mobile manipulation," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 558–568, 2011.

[35] S. J. Lederman and R. L. Klatzky, "Extracting object properties through haptic exploration," *Acta Psychologica*, vol. 84, no. 1, pp. 29–40, 1993.

[36] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, pp. 54–67, 2017.

[37] T. Sun, J. Back, and H. Liu, "Combining contact forces and geometry to recognize objects during surface haptic exploration," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2509–2514, 2018.

[38] G. Li, S. Liu, L. Wang, and R. Zhu, "Skin-inspired quadruple tactile sensors integrated on a robot hand enable object recognition," *Science Robotics*, vol. 5, no. 49, p. eabc8134, 2020.

[39] L. Jamone, G. Metta, F. Nori, and G. Sandini, "James: A humanoid robot acting over an unstructured world," *In Proceedings of International Conference on Humanoid Robots*, pp. 143–150, 2006.

[40] Y. Xia, A. Mohammadi, Y. Tan, B. Chen, P. Choong, and D. Oetomo, "On the efficiency of haptic based object identification: Determining where to grasp to get the most distinguishing information," *Frontiers in Robotics and AI*, vol. 8, p. 231, 2021.

[41] M. Kaboli and G. Cheng, "Robust tactile descriptors for discriminating objects from textural properties via artificial robotic skin," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 985–1003, 2018.

[42] H. Liu and F. Sun, "Material identification using tactile perception: A semantics-regularized dictionary learning method," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 3, pp. 1050–1058, 2017.

[43] D. Tanaka, T. Matsubara, K. Ichien, and K. Sugimoto, "Object manifold learning with action features for active tactile object recognition," *In Proceedings of International Conference on Intelligent Robots and Systems*, pp. 608–614, 2014.

[44] E. Kerr, T. M. McGinnity, and S. Coleman, "Material recognition using tactile sensing," *Expert Systems with Applications*, vol. 94, pp. 94–111, 2018.

[45] J. Hoelscher, J. Peters, and T. Hermans, "Evaluation of tactile feature extraction for interactive object recognition," *In Proceedings of International Conference on Humanoid Robots*, pp. 310–317, 2015.

[46] E. Kirby, R. Zenha, and L. Jamone, "Comparing single touch to dynamic exploratory procedures for robotic tactile object recognition," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4252–4258, 2022.

[47] H. Liu and F. Sun, "Material Identification Using Tactile Perception: A Semantics-Regularized Dictionary Learning Method," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 3, pp. 1050–1058, 2018.

[48] F. Gu, W. Sng, T. Taunyazov, and H. Soh, "Tactilesgnet: A spiking graph neural network for event-based tactile object recognition," *In Proceedings of International Conference on Intelligent Robots and Systems*, pp. 9876–9882, 2020.

[49] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard, "Object identification with tactile sensors using bag-of-features," *In Proceedings of IEEE International Conference on Intelligent Robots and Systems*, pp. 243–248, 2009.

[50] S. Luo, W. Mou, K. Althoefer, and H. Liu, "iclap: Shape recognition by combining proprioception and touch sensing," *Autonomous Robots*, vol. 43, no. 4, pp. 993–1004, 2019.

[51] F. Pastor, J. M. Gandarias, A. J. García-Cerezo, and J. M. Gómez-de Gabriel, "Using 3d convolutional neural networks for tactile object recognition with robotic palpation," *Sensors*, vol. 19, no. 24, p. 5356, 2019.

[52] M. Madry, L. Bo, D. Kragic, and D. Fox, "ST-HMP: Unsupervised spatio-temporal feature learning for tactile data," *In Proceedings of International Conference on Robotics and Automation*, pp. 2262–2269, 2014.

[53] J. M. Gandarias, J. M. Gómez-de Gabriel, and A. J. García-Cerezo, "Enhancing perception with tactile object recognition in adaptive grippers for human–robot interaction," *Sensors*, vol. 18, no. 3, p. 692, 2018.

[54] B. Hollis, S. Patterson, and J. Trinkle, "Compressed learning for tactile object recognition," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1616–1623, 2018.

[55] M. Bauza, O. Canal, and A. Rodriguez, "Tactile mapping and localization from high-resolution tactile imprints," *In Proceedings of International Conference on Robotics and Automation*, pp. 3811–3817, 2019.

[56] J. W. James, A. Church, L. Cramphorn, and N. F. Lepora, "Tactile model o: Fabrication and testing of a 3d-printed, three-fingered tactile robot hand," *Soft Robotics*, vol. 8, no. 5, pp. 594–610, 2021.

[57] M. Polic, I. Krajacic, N. Lepora, and M. Orsag, "Convolutional autoencoder for feature extraction in tactile sensing," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3671–3678, 2019.

[58] P. Sarakon, H. Kawano, K. Shimonomura, and S. Serikawa, "Efficient and small network using multi-trim network structure for tactile object recognition on embedded systems," *IEEE Access*, vol. 8, pp. 144277–144291, 2020.

[59] W. Yuan, Y. Mo, S. Wang, and E. H. Adelson, "Active clothing material perception using tactile sensing and deep learning," *In Proceedings of International Conference on Robotics and Automation*, pp. 4842–4849, 2018.

[60] M. Schöpfer, M. Pardowitz, R. Haschke, and H. Ritter, "Identifying relevant tactile features for object identification," in *Towards Service Robots for Everyday Environments*, pp. 417–430, Springer, 2012.

[61] L. Cao, R. Kotagiri, F. Sun, H. Li, W. Huang, and Z. M. M. Aye, "Efficient spatio-temporal tactile object recognition with randomized tiling convolutional networks in a hierarchical fusion strategy," *In Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.

[62] T. Corradi, P. Hall, and P. Iravani, "Bayesian tactile object recognition: Learning and recognising objects using a new inexpensive tactile sensor," *In Proceedings of International Conference on Robotics and Automation*, pp. 3909–3914, 2015.

[63] X. Qian, E. Li, J. Zhang, S.-N. Zhao, Q.-E. Wu, H. Zhang, W. Wang, and Y. Wu, "Hardness recognition of robotic forearm based on semi-supervised generative adversarial networks," *Frontiers in Neurorobotics*, p. 73, 2019.

[64] ATI Industrial Automation, "F/t sensor: nano17." Available at http://www.ati-ia.com/products/ft/ft_models.aspx?id=nano17.

[65] Z. Kappassov, J.-A. Corrales, and V. Perdereau, "Tactile sensing in dexterous robot hands," *Robotics and Autonomous Systems*, vol. 74, pp. 195–220, 2015.

[66] H. Dang, J. Weisz, and P. K. Allen, "Blind grasping: Stable robotic grasping using tactile feedback and hand kinematics," *In Proceedings of International Conference on Robotics and Automation*, pp. 5917–5922, 2011.

[67] J. M. Gandarias, J. M. Gómez-de Gabriel, and A. García-Cerezo, "Human and object recognition with a high-resolution tactile sensor," *In Proceedings of IEEE Sensors*, pp. 1–3, 2017.

[68] R. A. Russell, "Object recognition by a 'smart' tactile sensor," in *Proceedings of the Australian Conference on Robotics and Automation*, pp. 93–8, 2000.

[69] Z. Flintoff, B. Johnston, and M. Liarokapis, "Single-grasp, model-free object classification using a hyper-adaptive hand, google soli, and tactile sensors," *In Proceedings of International Conference on Intelligent Robots and Systems*, pp. 1943–1950, 2018.

[70] A. J. Dobson, *Introduction to statistical modelling*. Springer, 2013.

[71] A. Aggarwal and F. Kirchner, "Object recognition and localization: The role of tactile sensors," *Sensors*, vol. 14, no. 2, pp. 3227–3266, 2014.

[72] P. Falco, S. Lu, A. Cirillo, C. Natale, S. Pirozzi, and D. Lee, "Cross-modal visuo-tactile object recognition using robotic active exploration," *In Proceedings of International Conference on Robotics and Automation*, pp. 5273–5280, 2017.

[73] Z. Pezzementi, E. Plaku, C. Reyda, and G. D. Hager, "Tactile-object recognition from appearance information," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 473–487, 2011.

[74] M. M. Zhang, N. Atanasov, and K. Daniilidis, "Active end-effector pose selection for tactile object recognition through Monte Carlo tree search," *IEEE International Conference on Intelligent Robots and Systems*, pp. 3258–3265, 2017.

[75] W. De Mulder, S. Bethard, and M.-F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," *Computer Speech and Language*, vol. 30, no. 1, pp. 61–98, 2015.

[76] A. Vellido, "The importance of interpretability and visualization in machine learning for applications in medicine and health care," *Neural Computing and Applications*, vol. 32, no. 24, pp. 18069–18083, 2020.

[77] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[78] M. Parsian, *Data algorithms: Recipes for scaling up with hadoop and spark*. O'Reilly Media, Inc., 2015.

[79] P. Bholowalia and A. Kumar, "Ebk-means: A clustering technique based on elbow method and k-means in wsn," *International Journal of Computer Applications*, vol. 105, no. 9, 2014.

[80] J. T. Kent, "Information gain and a general measure of correlation," *Biometrika*, vol. 70, no. 1, pp. 163–173, 1983.

[81] C. Lee and G. G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization," *Information Processing and Management*, vol. 42, no. 1, pp. 155–165, 2006.

[82] S. Ottenhaus, L. Kaul, N. Vahrenkamp, and T. Asfour, "Active tactile exploration based on cost-aware information gain maximization," *International Journal of Humanoid Robotics*, vol. 15, no. 01, p. 1850015, 2018.

[83] R. M. Gray, *Entropy and information theory*. Springer Science & Business Media, 2011.

[84] M. Bennamoun and G. J. Mamic, *Object recognition: fundamentals and case studies*. Springer Science & Business Media, 2001.

[85] R. M. Cormack, "A review of classification," *Journal of the Royal Statistical Society: Series A (General)*, vol. 134, no. 3, pp. 321–353, 1971.

[86] G. Vezzani, N. Jamali, U. Pattacini, G. Battistelli, L. Chisci, and L. Natale, "A novel bayesian filtering approach to tactile object recognition," *In Proceedings of International Conference on Humanoid Robots*, pp. 256–263, 2016.

[87] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in manipulation research: Using the Yale-CMU-Berkeley object and model set," *IEEE Robotics and Automation Magazine*, vol. 22, no. 3, pp. 36–52, 2015.

[88] Y. Tenzer, L. P. Jentoft, and R. D. Howe, "The feel of MEMS barometers: Inexpensive and easily customized tactile array sensors," *IEEE Robotics and Automation Magazine*, vol. 21, no. 3, pp. 89–95, 2014.

[89] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[90] C. Walck *et al.*, *Hand-book on statistical distributions for experimentalists*. University of Stockholm, 2007.

[91] S. R. Eliason, *Maximum likelihood estimation: Logic and practice*. No. 96, Sage, 1993.

[92] R. V. Hogg, J. McKean, and A. T. Craig, *Introduction to mathematical statistics*. Pearson Education, 2005.

[93] A. Ross and V. L. Willson, "Paired samples t-test," in *Basic and advanced statistical tests*, pp. 17–19, Springer, 2017.

[94] R. A. Johnson, I. Miller, and J. E. Freund, "Probability and statistics for engineers," 2000.

[95] H. Liu, D. Guo, and F. Sun, "Object recognition using tactile measurements: Kernel sparse coding methods," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 656–665, 2016.

[96] M. J. D. Powell, *A direct search optimization method that models the objective and constraint functions by linear interpolation*, pp. 51–67. Springer, 1994.

[97] A. P. Dawid, "Conditional independence in statistical theory," *Journal of the Royal Statistical Society, Series B Methodology*, vol. 41, no. 1, pp. 1–15, 1979.

[98] A. Mohammadi, Y. Tan, P. Choong, and D. Oetomo, "Flexible mechanical metamaterials enabling soft tactile sensors with multiple sensitivities at multiple force sensing ranges," *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021.

[99] A. Mohammadi, Y. Xu, Y. Tan, P. Choong, and D. Oetomo, "Magnetic-based soft tactile sensors with deformable continuous force transfer medium for resolving contact locations in robotic grasping and manipulation," *Sensors*, vol. 19, no. 22, p. 4925, 2019.

[100] N. F. Lepora, U. Martinez-Hernandez, and T. J. Prescott, "Active bayesian perception for simultaneous object localization and identification.," *In Proceedings of Robotics: Science and Systems*, 2013.

[101] D. F. Gomes, P. Paoletti, and S. Luo, "Generation of gelsight tactile images for sim2real learning," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 4177–4184, 2021.

[102] R. Gao, Z. Si, Y.-Y. Chang, S. Clarke, J. Bohg, L. Fei-Fei, W. Yuan, and J. Wu, "Objectfolder 2.0: A multisensory object dataset for sim2real transfer," *In Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 10598–10608, 2022.

[103] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," *Advances in neural information processing systems*, vol. 18, pp. 507–514, 2005.