Vision-Based Adaptive Impedance Control for Soft Material Manipulation

by

Jeffrey Lee

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Mechanical and Mechatronics Engineering

Waterloo, Ontario, Canada, 2025

© Jeffrey Lee 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The manipulation of soft materials presents a longstanding challenge in robotics due to their nonlinear, variable, and often unpredictable physical properties. Traditional control strategies either require complex physical modeling or extensive data collection, limiting their applicability in dynamic, real-world settings. This thesis proposes a model-free, vision-based adaptive impedance control framework that enables collaborative robots to regulate contact forces during manipulation using only real-time visual feedback. By dynamically adjusting stiffness based on vision measurements of object deformation, the framework allows for compliant and adaptive interaction with a wide range of soft materials without prior knowledge of their physical characteristics.

The effectiveness of the framework is demonstrated through two representative studies. The first, VAIRO (Vision-based Adaptive Impedance-control Robotic framework), addresses the challenge of rolling croissant dough in a craft bakery setting. This study involved using a standalone stereo vision camera to monitor the rolling process of puff pastry into croissants with a robotic manipulator. Employing RGB-D data and real-time image processing, VAIRO detects and segments croissant layer thickness and inter-layer gaps, which are used to adapt the stiffness of a Cartesian impedance controller. The result of the study showed VAIRO creating croissants closer to artisinal quality compared to a constant stiffness approach, where VAIRO produced croissants with up to 6 times less variance in heights and mean error of heights less than 0.51mm.

The second study, VAISI (Vision-based Adaptive Impedance-control for Surgical Incisions), applies similar methodology to perform depth-controlled incisions in soft biological tissues using point-clouds captured by a stereo vision-enabled scalpel end-tool on the robotic manipulator. Using geometric measurements on the point cloud, the scalpel's depth and skin tissue deformation is tracked in real-time to adapt both the incision depth and force through visual servoing and stiffness adaptation of the Cartesian impedance controller. VAISI demonstrated the capability to create incisions with sub-millimeter accuracy and a maximum variance in incision depths of 1.23mm.

The vision-based adaptive impedance control framework presented in this thesis offers a foundation for future research in flexible automation, medical robotics, and other domains requiring interaction with soft materials. This contributes towards the goals of Industry 4.0—namely, adaptable and robust robotic systems.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Professor Yue Hu and Professor Alexander Wong, for their invaluable guidance and support throughout the development of this thesis. I also thank Professor Stewart McLachlin and Teresa Marotta for their collaboration and contributions to the work presented in Chapter 4. Special thanks to Duo Patisserie for their consultative input in shaping the application explored in Chapter 3. Thank you to Andrea Herold and Simon Hellwig from the KUKA Innovation Award for the KUKA equipment support.

This thesis was produced with funding from University of Waterloo's Faculty of Engineering's Engineering Excellence Master's Fellowship.

Dedication

This is dedicated to all my friends and family that have supported me through life and my academic journey, leading me to this point.

Table of Contents

Author's Declaration ii

Abstract iii

Acknowledgements iv

Dedication v

List of Figures ix

List of Tables xi

List of Abbreviations xii

1 Introduction 1

1.1 Thesis Contribution and Outline 3

2 Background 5

2.1 Adaptive Cartesian Impedance Control 5

2.1.1 Cartesian Impedance Control 5

2.1.2 Adaptive Cartesian Impedance Control 8

2.2 Stereo Vision 8

2.3 Image Processing 12

2.3.1 Object Detection . . . 13
2.3.2 Object Instance Segmentation . . . 14
2.3.3 Homogeneous Transformations . . . 14
2.4 Soft Material Manipulation . . . 16

3 Vision Based Adaptive Impedance Control Applied to Croissant Rolling 19
3.1 Introduction . . . 19
3.1.1 Background . . . 22
3.1.2 Contributions . . . 22
3.2 Methodology . . . 23
3.2.1 Experimental Setup . . . 23
3.2.2 End Effector Design . . . 24
3.2.3 Cartesian Impedance Control Implementation . . . 24
3.2.4 Croissant Parts Instance Segmentation . . . 25
3.2.5 Vision-Based Stiffness Adaptation . . . 27
3.3 Results . . . 30
3.3.1 Nominal Stiffness Identification . . . 30
3.3.2 Croissant Instance Segmentation . . . 31
3.3.3 Adaptive Impedance Dough Rolling . . . 31
3.4 Conclusions and future work . . . 32

4 Vision based Adaptive Impedance Control Applied to Skin Incisions 36
4.1 Introduction . . . 36
4.1.1 Background . . . 38
4.1.2 Contributions . . . 39
4.2 Methodology . . . 40
4.2.1 Design of the Stereo Vision Scalpel End-Tool . . . 40
4.2.2 Incision Feature Estimation using Stereo Vision . . . 41

4.2.3 Vision-Based Adaptive Cartesian Impedance Control 43
4.3 Experiments and Results 45
4.3.1 Evaluation of the Stereo Vision Scalpel End-Tool 45
4.3.2 Experimental Setup 47
4.3.3 Results 47
4.4 Conclusion and Future Work 50

5 Conclusion and Future Work **53**

References **56**

List of Figures

2.1 The robots used in this thesis: The Kinova Gen3 (left) and the KUKA LBR Med (right) . . . 6

2.2 Point projection to the image plane [13] . . . 9

2.3 The Intel RealSense D405 camera used in this thesis . . . 10

2.4 Example of finding the disparity of a point between the left and right sensor images from a parallel-axis stereo camera [14] . . . 11

2.5 Example of an object detection output of people and animals ©2016 IEEE [57] 13

2.6 Example of an object instance segmentation output of animals, people, and signs ©2023 IEEE [43] . . . 15

2.7 Example of cloth manipulation ©2024 IEEE [47] . . . 17

2.8 Deformable object manipulation of foam ©2024 IEEE [22] . . . 18

3.1 Small-scale pastry production kitchen at Duo Patisserie . . . 20

3.2 Figs. 3.2a and 3.2b illustrate the rolling technique by a pastry chef at Duo Patisserie. Figs. 3.2c and 3.2d is the technique reproduced using a robotic arm (with bounding boxes and segmentation). See the video attachment for the movements. . . . 21

3.3 Croissant production with a Collaborative Robot (CoBot) demonstration for the KUKA Innovation Award 2024 at the Hannover Messe 2024 by Jeffrey Lee, Cheng Tang, and Yue Hu ©KUKA Group . . . 21

3.4 Experimental Setup with the Kinova Gen3 robot arm and the RealSense D405 camera. . . . 24

3.5 Croissant Rolling Tool for a two-finger gripper. . . . 24

3.6 Visual perception pipeline for croissant top layer and gap thickness estimation within VAIRO . . . 25
3.7 Croissant segmentation result with the developed vision model. The black box is the bounding box of the *rolling_croissant_dough* object, the red outline shows the *croissant_top_layer* mask, green shows the *croissant_gap* mask, and blue shows the *croissant_core* mask. . . . 27
3.8 Z-stiffness causing 1% height loss in rolled croissants. Box plots show 10 trials per material with means labeled. . . . 30
3.9 Box plot of rolled croissant heights for each material over 10 trials using VAIRO vs. constant z-stiffness (100 N/m). The ideal height is shown by the green dashed line. . . . 32
3.10 Z-stiffness adaptation using VAIRO during rolling. . . . 33

4.1 Scalpel end-tool with the RealSense D405 close-range stereo vision camera 37
4.2 Surgical CoBot showcase with KUKA at the 2025 42nd IEEE-RAS International Conference of Robotics and Automation (ICRA) by Jeffrey Lee, Teresa Marotta, Stewart McLachlin, and Yue Hu . . . 39
4.3 (a) depicts the RGB image captured by the end-tool camera during an incision and (b) is the visualization of the processed point-cloud using Alg.2, where green represent points used for *scalpel_tip_z*, red represent points calculated for *indent_r*, and blue represents points used for *nominal_skin_z* 42
4.4 VAISI Framework Overview . . . 44
4.5 Incision depth measurement on a piece of foam from the LBR Med vs. the Realsense D405 on the end-tool at four incision velocities. . . . 46
4.6 Experimental setup for performing incisions on porcine skin using the LBR Med and scalpel camera end-tool . . . 47
4.7 Incision depth and force using constant stiffness of 200 N/m on porcine belly skin . . . 48
4.8 Incision depth and force using constant stiffness of 200 N/m on porcine hock skin . . . 48
4.9 Vision measured incision depth, skin indentation, and adapted impedance controller z-stiffness during the incision process of each test condition on porcine. . . . 51

List of Tables

3.1 Mean error from ideal height (ME), height variance of sample (σ^2), and sample result of rolled croissant of each material over 10 trials using VAIRO vs. a constant z-stiffness of 100 N/m. 34

4.1 Bias and noise of incision depth measurements using the RealSense D405 on the end-tool at four desired incision velocities, using the z-displacement measured from the LBR Med as the ground truth. 46

4.2 Performance at the start, middle, and end of incisions over 10 trials of each test condition 52

List of Abbreviations

CoBot Collaborative Robot ix, x, 1, 2, 6, 7, 19–22, 35, 36, 38–40, 44, 50, 53–55

DRL Deep reinforcement learning 2

PID proportional-integral-derivative 44, 45, 49

R-CNN Region-based Convolutional Neural Network 13, 14

RGB-D Red,Green,Blue (Colour) - Depth data 12, 23

YOLO You Only Look Once object detection algorithm 14, 25, 27, 31, 53

Chapter 1

Introduction

As Industry 4.0 progresses, CoBots are being increasingly adopted in various industries with their smaller footprint compared to common industrial robots and flexibility in performing tasks [66]. Their design to work alongside humans provide the added benefit for integration in smaller workspaces, allowing for a greater range of industries and environments, such as surgery, therapeutics, and manufacturing, to leverage this technology [66]. The role of robotics has shifted towards their integration with intelligent systems, being flexible and adapt to work efficiently to complete their tasks [27]. This involves using sensors and data for robots to determine their actions. As society continues to adopt CoBots to complete a greater variety of tasks, the adoption of intelligent systems and greater research is needed on control methodology to handle complex materials, such as soft objects.

Robots have been widely used in industrial settings, designed for automation in highly controlled environments [76], and for repetitive tasks [41]. In manufacturing spaces, manipulated objects' positions are often precisely controlled in automated manufacturing lines, whether that be with robotic manipulators or conveyor systems. Environmental conditions such as temperature and humidity are also controlled such that the manipulated materials are always in the same state. Further, most industrial manufacturing is targeted at the manipulation of rigid objects [87]. The great amount of control that is invested into making sure objects are in the same state and position allows robotic controls to be simplified to the point of simple position control, yet still remain robust due to the consistency of the conditions. However, this is very costly. Most small-medium sized enterprises could not afford this kind of control. In addition, as we look into integrating robotics further into society, this high level of environmental control is simply not feasible to implement everywhere in the real-world. Robots must be flexible and easy to integrate into a work environment to align with the Industry 4.0 paradigm [41]. Therefore, the ability to adapt

to uncontrolled environments and situations must be accounted for in robotics in order for robots to be integrated into society.

CoBots, or robotic manipulators in general, are not often used today to interact with soft material. The problem of soft material manipulation has historically been less investigated compared to rigid object manipulation due to the complexity of the problem [85]. Soft objects pose a significant technical challenge as their deformation must be sensed or modeled for the robot controls to adapt to their changes. To compound this problem, many soft objects have different properties from variations in their composition and in various environmental conditions. Consequently, applications that require handling of soft materials often rely on highly specialized and inflexible equipment designed for specific materials and conditions, contributing further to system complexity and cost, a contradiction to the principles of Industry 4.0, which advocates for adaptable and cost-effective automation solutions.

In the domain of robotic manipulation and learning, most of the recent progress primarily focuses on rigid objects. For example, Wolf et al. [83] review diffusion models for robotic manipulation, highlighting advances in trajectory generation, grasping, and task planning—none of which specifically address soft material manipulation. Similarly, Sapkota et al. [64] survey vision-language action models where manipulation tasks are predominantly limited to object handover, pick-and-place operations, and grasping rigid bodies. Deep reinforcement learning (DRL) approaches also face significant challenges when applied to soft materials; as reviewed by Han et al. [29], DRL techniques typically require either highly accurate simulation models or substantial real-world training data. However, the inherent high-dimensional and nonlinear nature of soft materials complicates the development of accurate simulation environments and impedes the feasibility of large-scale real-world data collection. Although initiatives such as SoftGym [51] and Nvidia, Disney Research, and Google DeepMinds's collaboration on Newton[1] have begun efforts to simulate soft material manipulation for DRL, their focus remains limited to ropes, fabrics, and fluids—materials that exhibit relatively low variability in physical properties and environmental sensitivity. In contrast, many soft materials encountered in real-world applications demonstrate substantial variability in mechanical properties depending on composition and environmental conditions. As a result, robust generalization across varying material properties would necessitate extensive real-world demonstrations or sophisticated adaptive models, further complicating the learning process. The complexity of soft material dynamics, combined with the challenges in modeling and data acquisition, has significantly hindered progress in this field, leaving soft material manipulation as an open and pressing research problem.

[1] https://github.com/newton-physics/newton

Nevertheless, interaction with soft materials is pervasive across a multitude of everyday and industrial activities, encompassing applications in food handling, textile management, and medical procedures involving biological tissues. The prevalence of soft material interactions, coupled with the comparatively limited research attention in this domain, underscores a significant research frontier. In the domain of soft material manipulation, a large majority of the research that has been done are model-based approaches [28, 82, 85], requiring the target material dynamics to be modeled for feedback. However, modeling soft materials is not desirable as many soft materials exhibit highly nonlinear and variable dynamics that are difficult to capture accurately. Such models often fail to generalize across different materials or environmental conditions, limiting their effectiveness in real-world applications.

Building upon these insights, the research presented in this thesis proposes a model-free, vision-based approach to soft material manipulation. By leveraging visual sensing, this work directly addresses the challenges posed by the high-dimensional, nonlinear, and variable physical properties of deformable soft materials. Vision systems enable rich, real-time perception of object deformations without the need for complex analytical models, thereby providing the flexibility required to handle diverse material behaviors and environmental variations. Specifically, this work employs visual feedback to adaptively regulate contact forces through the real-time adjustment of Cartesian impedance control parameters. The proposed vision-based adaptive impedance control framework capitalizes on recent advances in sensing technologies to enhance robotic adaptability and robustness, introducing a novel pathway toward achieving reliable and generalizable soft material manipulation.

1.1 Thesis Contribution and Outline

This thesis contributes to the emerging field of soft material manipulation by introducing a novel vision-based adaptive impedance control framework. The proposed approach is model-free, eliminating the need for extensive datasets or analytical physics models to estimate the deformations of soft objects. Furthermore, this work aims to demonstrate the framework's generalizability by achieving successful manipulation across materials with varying compositions and properties. The effectiveness and versatility of the proposed framework are demonstrated through two distinct investigations: one in the context of artisinal food production and another in surgery, illustrating its applicability across domains. The key contributions of this thesis are outlined as follows:

- **Chapter 1** introduces the motivation and problem statement that frames the context for this research.
- **Chapter 2** provides background on adaptive impedance control, stereo vision, image processing, and soft material manipulation, forming the theoretical foundation for the proposed framework.
- **Chapter 3** presents the first investigation of this work, *VAIRO* (**V**ision-based **A**daptive **I**mpedance-control **RO**botic framework), which applies the proposed method to the task of rolling puff pastry for croissant preparation. This work was accepted to the 2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).
- **Chapter 4** presents the second investigation, *VAISI* (**V**ision-based **A**daptive **I**mpedance-control for **S**urgical **I**ncisions), focusing on achieving depth-precise incisions in soft biological tissues. The tool design involved in this chapter was accepted as *See, Cut, Control: A Stereo Vision Scalpel for Surgical Robotics* in the 1st International Workshop on the Evolving Landscape of Surgical Robotics (ELSR 2025), held at the 2025 42nd IEEE-RAS International Conference of Robotics and Automation (ICRA). The full work was accepted to the 2025 24th IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS).
- **Chapter 5** concludes the thesis by summarizing the contributions and discussing potential directions for future research in vision-based adaptive impedance control for soft material manipulation.

Chapter 2

Background

2.1 Adaptive Cartesian Impedance Control

The aim of this work is to develop a control framework capable of modulating contact forces while executing a desired motion trajectory. Among various compliant control strategies, impedance control is particularly well-suited for this objective, as it accepts motion input and produces a corresponding force response [35]. In contrast, alternative compliant control strategies such as admittance control, which requires force input and yields motion output [35], or direct force control, which tracks a predefined force trajectory, are less compatible with the goals of this work. These approaches do not inherently support simultaneous motion execution with adaptive force modulation, making impedance control the most appropriate choice for the envisioned tasks.

This section begins by introducing the fundamental implementation of Cartesian impedance control. The following builds on this foundation by describing how impedance control is made adaptive for dynamic contact force regulation.

2.1.1 Cartesian Impedance Control

Cartesian Impedance control was developed as a compliant control strategy for robotic manipulators to dynamically interact with its environment [35]. Compared to other control strategies such as position or torque control, impedance control distinct itself as it attempts to implement dynamic relation between both manipulator end-point position and force rather than just control these variables alone [35].



Figure 2.1: The robots used in this thesis: The Kinova Gen3 (left) and the KUKA LBR Med (right)

Impedance control establishes a relationship between displacement and velocity with force through stiffness and damping, respectively. The stiffness relation, $K \in \mathbb{R}^{6\times6}$, is expressed by

$$F_K = K(x_0 - x) \tag{2.1}$$

where $F_K \in \mathbb{R}^6$ is the Cartesian force exerted by the end-effector from stiffness, $x_0 \in \mathbb{R}^6$ is the desired end-effector position, and $x \in \mathbb{R}^6$ is the measured end-effector position [35]. For robotic manipulators with joint-torque actuation and sensing capabilities, such as the CoBots used in this thesis: the Kinova Gen3[1] and the KUKA LBR Med[2], see Fig.2.1, this relation can be extended to the joint-space to solve for the joint torques, $\tau_K \in \mathbb{R}^7$, needed for the robot to exhibit stiffness behaviour using the robot's Jacobian calculated at the end-effector, $J(q) \in \mathbb{R}^{6\times7}$. The CoBots used in this work all have 7 joints, thus the joint-space is in $\mathbb{R}^7$, and its joint angles are represented as $q \in \mathbb{R}^7$.

$$\tau_K = J^T(q)F_K \tag{2.2}$$

Using this relation in eq.2.2, the joint torques required for the CoBot to exhibit Cartesian

[1]https://www.kinovarobotics.com/product/gen3-robots

[2]https://www.kuka.com/en-ca/industries/robots-used-in-medicine/kuka-solutions-for-medical-robots/lbr-med-medical-robot

impedance behaviour from end-effector displacement can be described by

$$\tau_K = J^T(q)K(x_0 - x) \tag{2.3}$$

Furthermore, damping, $D \in \mathbb{R}^{6\times6}$, is expressed by

$$F_D = Dv \tag{2.4}$$

where $F_D \in \mathbb{R}^6$ is the damping force, $v \in \mathbb{R}^6$ is the measured end-effector velocities of the CoBot. Again, using the Jacobian relation, the end-effector velocities can be found from the measured joint velocities of the CoBot using

$$v = J(q)\dot{q} \tag{2.5}$$

where $\dot{q} \in \mathbb{R}^7$ is the measured joint velocities of the CoBot. Using the relations of eq.2.2 and eq.2.5, the joint torques, $\tau_D \in \mathbb{R}^7$, required for a CoBot to express Cartesian damping behaviour can be described by

$$\tau_D = J^T(q)D(J(q)\dot{q}) \tag{2.6}$$

Summing the stiffness and damping torques from eq.2.1 and eq.2.6, respectively, the torque required for a CoBot to achieve Cartesian impedance behaviour, τ_{ci}, is expressed as

$$\tau_{ci} = J^T(q)(K(x_0 - x) + D(J(q)\dot{q})) \tag{2.7}$$

In order to implement Cartesian impedance control, however, robot dynamics must also be accounted for. The dynamics of a robotic manipulator can be captured by the following dynamic equation of motion [71]

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + g(q) + J^T(q)F_{ext} = \tau \tag{2.8}$$

where $M(q) \in \mathbb{R}^{7\times7}$ is the inertia matrix, $C(q,\dot{q}) \in \mathbb{R}^{7\times7}$ captures the effects of Coriolis and centripetal forces, $g \in \mathbb{R}^7$ represents the effect of gravity, and $F_{ext} \in \mathbb{R}^6$ is the external force acting on the end-effector. Therefore, the control torque, $\tau_c \in \mathbb{R}^7$, must compensate for the inertial, Coriolis, centripetal, and gravity forces acting on the robotic manipulator, while τ_{ci}, as expressed in eq.2.7, reacts to the external joint-torques $J^T(q)F_{ext}$ to exert Cartesian impedance behaviour. Therefore, the control torque, $\tau_c \in \mathbb{R}^7$, for Cartesian impedance control becomes as follows

$$\tau_c = J^T(q)(K(x_0 - x) + D(J(q)\dot{q})) + M(q)\ddot{q} + C(q,\dot{q})\dot{q} + g(q) \tag{2.9}$$

2.1.2 Adaptive Cartesian Impedance Control

In principle, stiffness, K, and damping, D, parameters can be kept constant if fixed interaction forces are desired for a given task. However, adaptive impedance control, also known as variable impedance control, has been a popular and effective strategy for enabling robots to interact with dynamic environments [1]. Recent studies have predominantly adopted learning-based approaches, both offline and online, to modulate impedance behavior in contact-rich tasks [1]. Among these, online learning methods have demonstrated effectiveness for tasks involving interactions with unknown objects and environments.

Adaptive impedance control works by modulating the stiffness, K, and damping, D, parameters of eq.2.7. As the term suggests, stiffness governs the interaction forces based on the deviation of the robot's end-effector from a desired position. Consequently, a higher stiffness results in stronger interaction forces, while a lower stiffness leads to more compliant behavior. This enables a direct mapping between task force requirements and stiffness: increasing stiffness when greater force is needed, and decreasing it when a softer interaction is desired.

On the other hand, damping modulates interaction forces based on end-effector velocity. However, in contact-based tasks—such as those investigated in this work—the robot is expected to remain in continuous contact with the object, leading to minimal end-effector velocity. Thus, the effect of damping on force modulation is relatively limited in such scenarios. For this reason, the damping parameter is fixed in this work to achieve critical damping, using the standard relationship:

$$D = 2\zeta\sqrt{K} \tag{2.10}$$

where the damping factor, ζ, is set to 1. Furthermore, setting the controller to critical damping enables more isolated and consistent analysis of force modulation through stiffness alone in this work.

2.2 Stereo Vision

Stereo vision has become a more accessible technology in recent years thanks to the wider range of off the shelf stereo camera hardware, software, and algorithms becoming available. This propelled the use of stereo vision in robotics as it allows 3-dimensional scene representation of a robot's environment for feedback, which is especially important as robots are continuously being deployed in new environments.

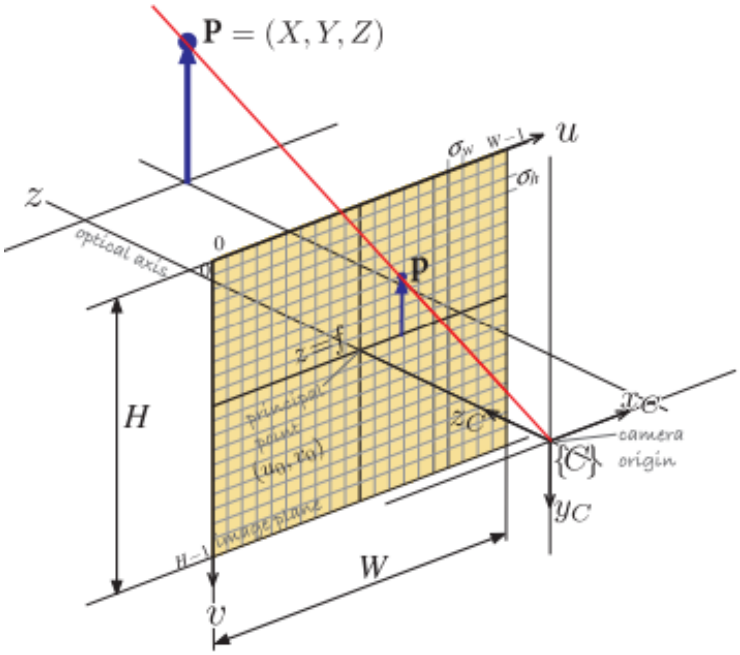


Figure 2.2: Point projection to the image plane[3] [13]

To understand how stereo vision captures a 3-dimensional (3D) scene, it is first essential to examine how a 2-dimensional (2D) image is formed using a single camera. According to the *Image Formation* chapter in *Robotic Vision* by Peter Corke [13], image formation is governed by a projection model. A real-world point $P = (X, Y, Z)$ in 3D space is projected onto the image plane as a point $p = (x, y, z)$, and subsequently represented in a digital image as a 2D pixel coordinate (u, v), as illustrated in Fig. 2.2. By applying the principle of similar triangles, the projection from 3D to the image plane is defined as:

$$x = z\frac{X}{Z}, \quad y = z\frac{Y}{Z} \tag{2.11}$$

The transformation from image plane coordinates to pixel coordinates involves the camera intrinsics, which encapsulate the hardware configuration of the camera. Key intrinsic parameters include:

- The focal length, f, representing the distance from the camera origin to the image plane along the optical axis,

[3] *Reproduced with permission from Springer Nature*



Figure 2.3: The Intel RealSense D405 camera used in this thesis

- The principal point, (u_0, v_0), where the optical axis intersects the image sensor,
- The physical pixel dimensions, ρ_u and ρ_v, representing the real-world width and height of each pixel, respectively.

Using these parameters, the pixel coordinates (u, v) are related to the image plane coordinates by:

$$u = \frac{x}{\rho_u} + u_0, \quad v = \frac{y}{\rho_v} + v_0 \tag{2.12}$$

Substituting Eq. 2.11 into Eq. 2.12, the projection from real-world coordinates to image pixels becomes:

$$u = \frac{fX}{Z\rho_u} + u_0, \quad v = \frac{fY}{Z\rho_v} + v_0 \tag{2.13}$$

Although the camera intrinsic parameters are preserved during image capture, the depth term Z is lost, making it impossible to recover the full 3D coordinates (X, Y, Z) from a single image alone. This fundamental limitation necessitates the use of stereo vision, which employs two spatially separated cameras to recover depth information by triangulating corresponding image features across both views, thereby reconstructing the full (X, Y, Z) position of points in the scene.

The stereo camera used in this work is the Intel RealSense D405[4], see Fig.2.3. This camera has parallel-axis camera geometry (i.e. the two camera sensors are aligned horizontally) and provides dense stereo matching, where the real-world (X, Y, Z) coordinates are

[4]https://www.intelrealsense.com/depth-camera-d405/

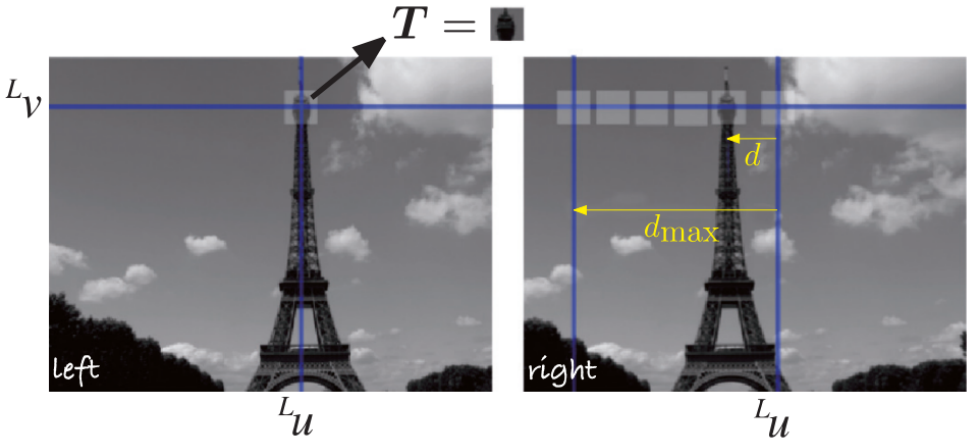


Figure 2.4: Example of finding the disparity of a point between the left and right sensor images from a parallel-axis stereo camera[5] [14]

computed for every pixel in the image. According to the *Stereo Vision* chapter in *Robotic Vision* by Peter Corke [14], stereo vision estimates the (X, Y, Z) positions of pixels by using the disparity of real-world points in an image taken from two different viewpoints. If the pixel coordinates corresponding to a point are (u^L, v^L) from the left image sensor and (u^R, v^R) from the right image sensor, then $v^L = v^R$ because of the camera's parallel-axis geometry, and the disparity of the point in the two images is $d = u^L - u^R$. For a pixel (u^L, v^L) in the left camera sensor, the same pixel is at some coordinate $(u^L - d, v^L)$ in the right camera sensor. To compute the disparity, a square template window T of size $N \times N$ is extracted around the pixel u_L in the left image. This window is then slid horizontally along the corresponding row in the right image. The most similar matching window is identified, and the disparity d is computed. This concept is illustrated by an example depicted in Fig.2.4.

With the additional information provided by the *disparity* d, defined as the horizontal pixel difference between corresponding points in the left and right images, and the known *baseline* b, which is the physical distance between the two camera sensors, the real-world coordinates of image points can be recovered.

To compute the X coordinate of a point, consider the horizontal displacement in pixel units from the principal point, u_0, in the left camera image, denoted u^L. By geometric similarity, the pixel ratio $\frac{u^L - u_0}{d}$ is equivalent to the physical ratio $\frac{X}{b}$. Solving for X yields:

[5] *Reproduced with permission from Springer Nature*

$$X = \frac{b(u^L - u_0)}{d} \tag{2.14}$$

Similarly, the vertical displacement from the principal point v_0 can be used to solve for the Y coordinate:

$$Y = \frac{b(v^L - v_0)}{d} \tag{2.15}$$

Using the concept of similar triangles again, the following relation must hold

$$\frac{Z}{X} = \frac{f}{\rho_u(u^L - u_0)} \tag{2.16}$$

and substituting in Eq.2.14, Z can be found:

$$Z = \frac{fb}{d\rho_u} \tag{2.17}$$

Using Eqs.2.14, 2.15, and 2.17, the physical (X, Y, Z) location of a point can be found, with respect to the left sensor on a stereo camera.

With this foundation, stereo vision data can be utilized in two primary ways: as Red,Green,Blue (Colour) - Depth data (RGB-D) or as point clouds. In the RGB-D representation, images are processed similarly to conventional 2D RGB images, with the depth information incorporated as a fourth channel alongside the red, green, and blue colour channels. This format enables the integration of depth values directly into 2D image-based processing pipelines and is used in the work presented in Chapter 3. In contrast, the point cloud representation captures the 3D geometry of a scene by mapping each pixel to its corresponding spatial coordinates, (X, Y, Z). This enables direct geometric reasoning in 3D space and is employed in the work described in Chapter 4.

2.3 Image Processing

Computer vision is widely used in today's robotic applications. Just like with humans, vision enables robots to understand its environment by identifying objects and knowing where they are. This is especially important for manipulation tasks as this allows for the automation of finding an object and visually tracking an object's state for feedback. This section provides background on the image processing techniques used in this thesis to enable vision-based control and feedback.

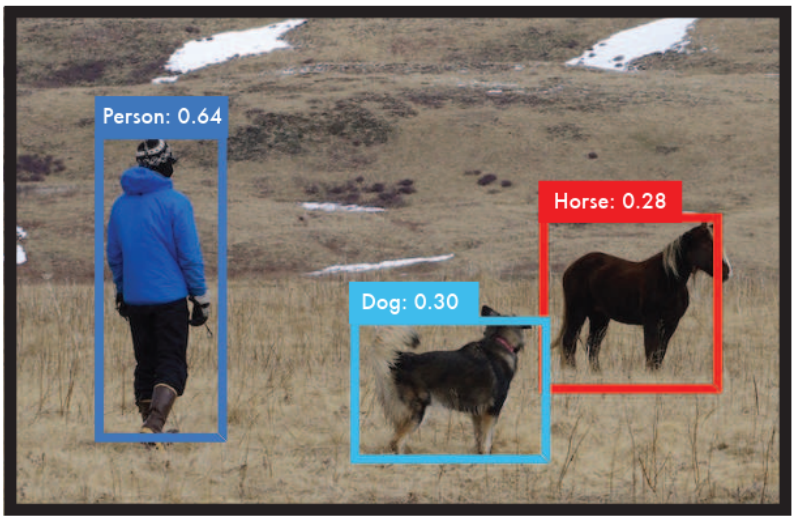


Figure 2.5: Example of an object detection output of people and animals ©2016 IEEE [57]

2.3.1 Object Detection

Object detection in image processing refers to the task of recognizing whether objects are in an image and if they belong to a particular class. An example can be seen in Fig.2.5. Modern approaches primarily use learning-based techniques, where object features are learned from a dataset of objects annotated with bounding boxes in the image. These learned features are then used to identify objects in new, unseen images through similarity matching [75]. Current research in learning-based object detection is largely dominated by two families of methods: region-based techniques and single-stage networks [75].

Region-based techniques performs object detection in a two-stage process. First, a region proposal algorithm identifies candidate rectangular regions of objects. Second, a classification algorithm assigns object classes to each region [75]. The combined regions and class labels are then the output of the object detection algorithm. The most well-known examples of this approach are the Region-based Convolutional Neural Network (R-CNN) family, including R-CNN [25], Fast R-CNN [24], and Faster R-CNN [58]. While successive iterations have significantly improved inference speed, the two-stage nature of R-CNN inherently leads to higher computational latency. Slow detection limits the performance of robotic applications where real-time feedback is needed. Delays in feedback could result in failure of a task due to the slow reaction in adapting the controls. As the works presented in here rely on real-time vision feedback for robotic manipulators to adapt, this is not desirable and thus single-stage networks are used instead.

In contrast, single-stage networks use a single neural network architecture to directly output object detections across the image in a single pass [75]. Among these, the You Only Look Once object detection algorithm (YOLO) is employed in this work due to its accurate performance and real-time capabilities [79], ease of integration via the Ultralytics API[6], and compatibility with training tools like Roboflow[7]. YOLO reformulates object detection as a single regression problem, mapping image pixels to bounding boxes and class labels [57]. By framing the problem in this way, object detection can be performed using a single neural network, eliminating the need for separate region proposal and classification stages. To accomplish this, YOLO simultaneously predicts the class of an object and it's bounding box's position and dimensions, achieving all object detection necessities at once [57]. By utilizing a single pipeline of convolutional, pooling, and fully connected layers, YOLO performs object detection significantly faster compared to two-stage approaches, while achieving similar performance. As such, YOLO is well-suited for the real-time object detection demands of the work presented in Chapter 3.

2.3.2 Object Instance Segmentation

Object instance segmentation extends on object detection by providing more precise pixel-level boundaries, known as masks, around each object, rather than simple bounding boxes. An example can be seen in Fig.2.6. As a result, training such networks requires more detailed annotations of object masks. Similar to object detection, modern instance segmentation approaches also use learning based approaches with convolutional neural networks, and often stem from architectures developed for object detection [75]. For example, Mask R-CNN extends on the Faster R-CNN architecture with an additional branch for the task of generating a mask for the object [30]. Similarly, Ultralyics has extended on YOLO's single stage architecture to output object masks[8]. As with the rationale for object detection, YOLO is employed in the work of Chapter 3 for its detection speed.

2.3.3 Homogeneous Transformations

To make sensor measurements useful for robotic actions, they should be made relative to a robot's frame of reference. This can be done through homogeneous transformations,

[6] https://docs.ultralytics.com
[7] https://roboflow.com
[8] https://docs.ultralytics.com/tasks/segment



Figure 2.6: Example of an object instance segmentation output of animals, people, and signs ©2023 IEEE [43]

where sensor data is both rotated and translated from the sensor's frame of reference to the robot's frame of reference.

A common method to represent rotation in 3-dimensional space between two reference frames is through roll (rotation about the z axis), θ, pitch (rotation about the y axis), ϕ, and yaw (rotation about the x axis), ψ, angles. When conducting rotations of ψ, ϕ, θ, in that order, a rotation of the sensor frame with respect to the robot frame can be formed as the following [72]:

$$R_S^R = R_\theta R_\phi R_\psi \tag{2.18}$$

$$= \begin{bmatrix} c_\phi & -s_\phi & 0 \\ s_\phi & c_\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_\theta & 0 & s_\theta \\ 0 & 1 & 0 \\ -s_\theta & 0 & c_\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_\psi & -s_\psi \\ 0 & s_\psi & c_\psi \end{bmatrix} \tag{2.19}$$

$$= \begin{bmatrix} c_\phi c_\theta & -s_\phi c_\psi + c_\phi s_\theta s_\psi & s_\phi s_\psi + c_\phi s_\theta c_\psi \\ s_\phi c_\theta & c_\phi c_\psi + s_\phi s_\theta s_\psi & -c_\phi s_\psi + s_\phi s_\theta c_\psi \\ -s_\theta & c_\theta s_\psi & c_\theta c_\psi \end{bmatrix} \tag{2.20}$$

$$\tag{2.21}$$

The next part of a transformation that needs to be represented in the translation between the two frames. In a 3-dimensional space, the x, y, z translation of the sensor frame with respect to the robot's frame can be represented as [72]:

$$p_S^R = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{2.22}$$

Altogether, a homogeneous transformation matrix H_S^R can be formed to capture both the rotation and translation needed to transform measurements from a sensor frame to the robot's frame of reference [72]:

$$H_S^R = \begin{bmatrix} R_S^R & p_S^R \\ 0 & 1 \end{bmatrix} \tag{2.23}$$

Now, any non-homogeneous point, $p_S = \begin{bmatrix} x_S \\ y_S \\ z_S \end{bmatrix}$, such as points in a point cloud can be transformed to a non-homogeneous point, $p_R = \begin{bmatrix} x_R \\ y_R \\ z_R \end{bmatrix}$, relative to a robot's frame of reference for robot control by using H_S^R:

$$\begin{bmatrix} p_R \\ 1 \end{bmatrix} = H_S^R \begin{bmatrix} p_S \\ 1 \end{bmatrix} \tag{2.24}$$

2.4 Soft Material Manipulation

Deformable and soft materials presents a significant challenge in robotic manipulation, particularly in tracking an object's shape and pose for effective feedback control. Historically, research in this field has primarily adopted model-based approaches, wherein object dynamics are analytically estimated through physical modeling [28, 82, 85]. These models are used to guide objects into desired configurations, with performance typically evaluated by comparing the positions of key points against predefined targets [80, 81]. Such methods have largely focused on structured deformable objects—such as ropes and cloth—that exhibit consistent, recoverable shapes and relatively predictable mechanical properties [80, 81, 85].

However, the effectiveness of using model-based techniques to capture the target object's dynamics is inherently limited by the accuracy and adaptability of the physical models used. While these methods may perform well in controlled laboratory settings, they often degrade in real-world scenarios involving materials with high variability or dynamic

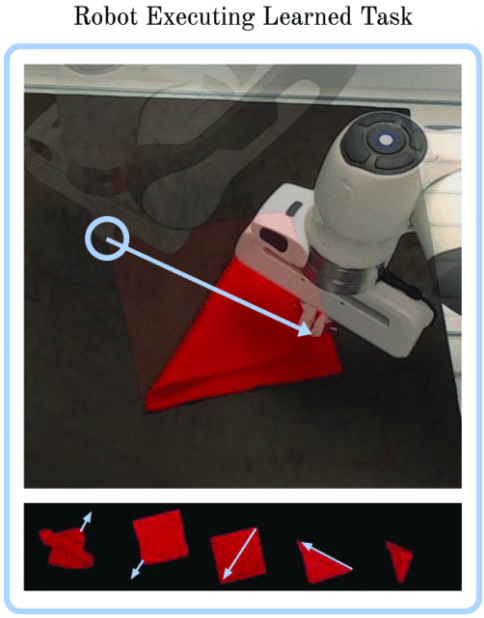


Figure 2.7: Example of cloth manipulation ©2024 IEEE [47]

properties. To address these limitations, there has been a growing shift toward learning- and data-driven approaches, enabled by recent advances in machine learning (ML) and the increasing availability of high-resolution sensor technologies [28,82]. Notably, surveys in the field highlight sensing as the most critical enabler for advancing soft material manipulation, emphasizing the importance of robust, real-time perception systems [87].

In response to these challenges, using model-free approaches to capture the target object's dynamics have been developed to leverage machine learning to directly map sensor measurements to control policies. Several recent studies have demonstrated the use of vision-based systems to extract masks, meshes, or key point features from real-time point cloud or RGB-D data of deformable objects during manipulation [22, 37, 38, 62, 77]. These features are compared against desired configurations to compute the pose error, which is then used to generate reference velocities for visual servoing of the robotic manipulator. In some cases, vision has also been used in learning-by-demonstration frameworks to visually observe and replicate how a human manipulates a deformable object [47, 63]. This model-free paradigm bypasses the need for explicit physical modeling and offers greater adaptability in handling complex, deformable objects under variable conditions.

However, the works of [22, 37, 38, 62, 77], all focus on the manipulation of foam, as seen in Fig.2.8, a structured deformable object which exhibit elasticity and can return to

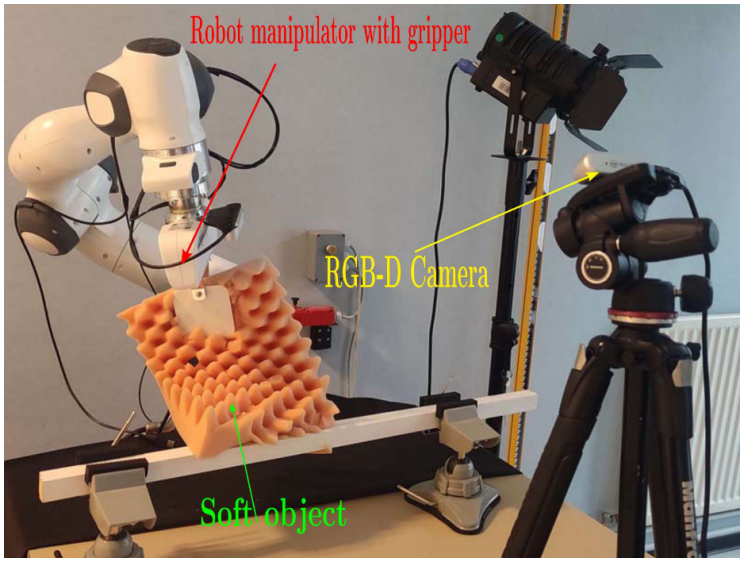


Figure 2.8: Deformable object manipulation of foam ©2024 IEEE [22]

their original shape. In such cases, interaction forces are of lesser concern, as deformation is reversible and visual servoing alone is sufficient to guide the object toward a target configuration. In contrast, unstructured soft materials, which deform irreversibly and lack inherent elasticity, require not only pose tracking, but also explicit regulation of interaction forces to avoid damage or irreversible deformation.

This work aims to extend the current progress in vision-based soft material manipulation by incorporating force regulation into the manipulation loop, enabling robust interaction with unstructured, non-recoverable soft materials.

Chapter 3

Vision Based Adaptive Impedance Control Applied to Croissant Rolling

3.1 Introduction

While robotic automation can offer precise solutions to repetitive tasks, leading to time and cost savings in production lines, such precise repetition is not always desired. This is especially the case with artisanal craft goods, where consumers seek the artisanal touch, believed to be of higher quality [7]. An example is baked goods, where machine-made pastries are cheaper and perceived as lower-grade with respect to more expensive but artisanal hand-made ones [17].

Despite this, there is interest from the small-scale craft industry in adopting recent technologies such as robotics [20], especially in a human-robot collaboration fashion. Current industrial machinery for high-volume batch production of goods tends to be bulky and costly, limiting its flexibility and making it less suitable for small-scale or customized production in environments such as the one seen in Fig.3.1. CoBots represent a small-scale versatile solution that can be adopted in a human-centered environment such as current craft laboratories and/or kitchens. Large robot manufacturers also showed interest in this domain, with the KUKA 2024 Innovation Award [33] calling for solutions applied to craft, tailored to small-medium enterprises. In this context, we investigated the problem of artisanal baking with local pastry chefs at Duo Patisserie in Toronto, Canada. The chefs have expressed the need to automate several highly repetitive baking procedures with versatile machines such as CoBots, with the requirement that the CoBots must reproduce the artisanal quality of the baked goods. Especially, they wished for the robot to be able to work



Figure 3.1: Small-scale pastry production kitchen at Duo Patisserie

the high-volume production croissants, highlighting that the rolling action is the one that defines the difference between artisanal and machine-made.

A common way of rolling croissants[1] is by creating an initial lip of the dough (puff pastry) and pushing one's hand over the puff pastry such that static friction pulls the flat puff pastry over itself into a croissant shape as shown in Figs. 3.2a and 3.2b. During this action, the downward pressure applied to the dough is adjusted to eliminate all gaps between layers, while maintaining consistent thickness throughout all layers of the croissant. This requirement is the most crucial task in preserving the artisanal quality. An initial solution tailored to the identified requirements was presented at the Kuka Innovation Award final at the Hannover Messe 2024, [33], where the CoBot-based concept sparked the interest of many experts in the crafts industry, see Fig.3.3. In this chapter, we present **VAIRO**: a **V**ision-Based **A**daptive **I**mpedance-control **RO**botic framework for soft material manipulation capable of achieving croissant dough rolling, with evaluations on different materials.

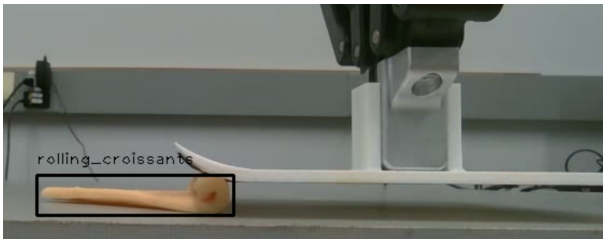
[1]This information was obtained from site visits at Duo Patisserie, interviews, and exchanges with their chefs.



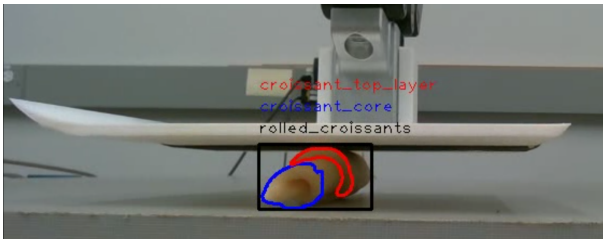
(a) Hand rolling - start



(b) Hand rolling - end



(c) Robot rolling - start



(d) Robot rolling - end

Figure 3.2: Figs. 3.2a and 3.2b illustrate the rolling technique by a pastry chef at Duo Patisserie. Figs. 3.2c and 3.2d is the technique reproduced using a robotic arm (with bounding boxes and segmentation). See the video attachment for the movements.



Figure 3.3: Croissant production with a CoBot demonstration for the KUKA Innovation Award 2024 at the Hannover Messe 2024 by Jeffrey Lee, Cheng Tang, and Yue Hu ©KUKA Group

3.1.1 Background

To tackle the manipulation of puff pastry, it is necessary to first understand its rheology, such as its yield stress, viscosity, Young's Modulus, etc. Dough itself has non-linear rheological behaviors, with many factors such as the type of flour, dough ingredients, temperature, water composition, and air incorporation all affecting its rheological properties [53]. For puff pastry specifically, [69] found that the dough resistance continuously increased while dough extensibility varied with increasing levels of salt and [59] observed a variation of rheological properties from just the variation in fat that is used in the puff pastry and a general softening behavior with increasing sheeting steps used for puff pastry. This variability of physical properties poses a difficult problem for manipulation as estimations on material properties cannot be confidently relied upon for force regulation. Therefore, the problem of dough manipulation is restricted to using only sensor measurements for the state estimation of the dough.

With the great variability of puff pastry rheology, our vision-based feedback solution allows us to steer away from the estimation of the physical properties of the dough. To the best of our knowledge, works prior to this have not used vision feedback for force regulation with impedance control, which is essential to create a craft bakery robot helper that can automate baked goods production mimicking the techniques of pastry chefs. Furthermore, in contrast to existing automated croissant production systems, which rely on large-scale, high-cost machinery operating in highly controlled environments to produce uniform, synthetic-quality products, our method of using CoBots enables small-scale manufacturing of croissants with artisanal quality, aligning more closely with traditional pastry-making practices.

3.1.2 Contributions

In this work, we introduce a low-cost, small-scale solution using a robotic manipulator with vision-based real-time force regulation for craft bakeries, with a focus on rolling croissants into their iconic shape, while maintaining artisanal qualities. To achieve this, a small-size collaborative manipulator will be used with an adaptive Cartesian impedance controller to regulate the rolling force to form the dough into the croissant shape. The stiffness parameters of the Cartesian impedance controller are adapted in real time to maintain consistent layer thickness and eliminate gaps through vision measurements. Specifically, this work presents the VAIRO framework, which consists of the following:

1. A computer vision process to effectively segment and estimate the size of croissant

parts with RGB-D data;

2. A vision feedback loop to adjust for an adaptive Cartesian impedance controller's stiffness parameters for force regulation.

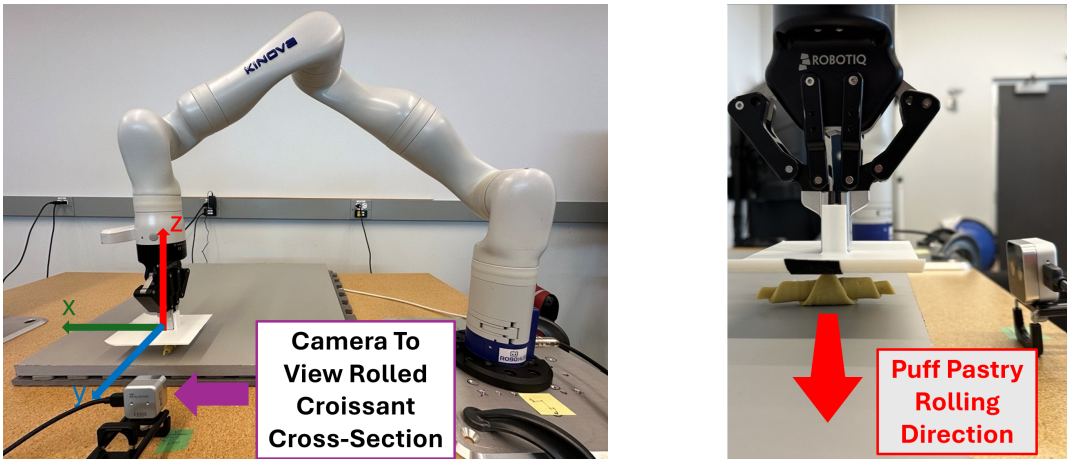
The framework is experimentally verified with the Kinova Gen3 using a custom end-effector designed to mimic a human's hand for puff pastry dough rolling (as shown in Figs. 3.2c and 3.2d), with four different types of dough. It is worth noting that while the VAIRO framework was motivated and centered around the shaping of croissants, this paper contributes a novel framework to the manipulation of diverse soft materials whose physical properties are difficult or impossible to estimate in real time, which is the case of many real-world applications.

3.2 Methodology

The proposed framework uses vision to measure both the thickness of the rolled croissant layers and the gaps between layers during the croissant rolling process as feedback to adjust the stiffness and damping parameters of a Cartesian impedance controller in real-time. To achieve this, only a robotic manipulator and a depth camera are needed. The approach consists of four parts: end-effector design, adaptive impedance control, computer vision model for segmentation, and integration of visual measurements to adjust the stiffness and damping parameters of the impedance controller. In this section, we present all the components along with the experimental setup and implementation details.

3.2.1 Experimental Setup

The setup used for experimentally validating VAIRO consists of the Kinova Gen3 for the robotic manipulator (and not a KUKA arm as the framework was extended and results were collected after the Innovation Award 2024, furthermore, the Gen3 is a more compact and smaller size robot) and the Intel RealSense D405 camera for the close-range vision. Fig. 4.6 shows the setup, where the camera faces perpendicularly to the rolling direction of the robotic manipulator to capture the cross-section of the puff-pastry dough during the rolling process. Additionally, the camera is placed 7 [cm] away from the dough as this is the minimum depth measurement for the D405, and is centered at the same height as the dough to achieve the best measurement accuracy as possible. This setup allows physical isolation of the camera from the robot to reduce measurement noise from the robot's movement.



(a) Front View (b) Side View

Figure 3.4: Experimental Setup with the Kinova Gen3 robot arm and the RealSense D405 camera.

3.2.2 End Effector Design

A custom end effector was designed to mimic the shape of the fingertips and palm of a flat human hand while rolling dough, as seen in Fig.3.2, but also with a uniform width. This results in a tool design with a large flat base as the palm and a rounded tip as the fingertips, as shown in Fig. 4.1.

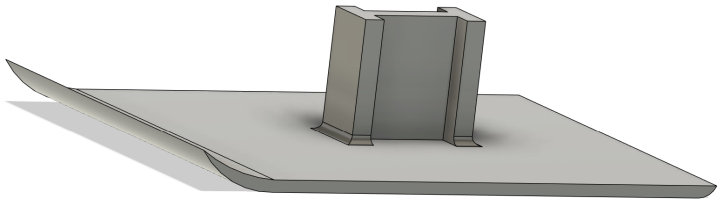


Figure 3.5: Croissant Rolling Tool for a two-finger gripper.

3.2.3 Cartesian Impedance Control Implementation

As introduced in Sec.2.1, the development of the adaptive impedance controller relies on Eq. 2.9, where the K and D stiffness and damping parameters are adapted in real-time. For the purposes of dough rolling, only the z-directional (vertical axis perpendicular to the

working surface, as in Fig. 4.6) stiffness needs to be adapted for force regulation. As a result, all other axes (x, y) and orientations (roll pitch, yaw) should use high stiffnesses for improved tracking performances.

To achieve the desired Cartesian impedance behavior on the Kinova Gen3, the Pinocchio rigid body dynamics library [12] is used to compute the $M(q)$, $C(q, \dot{q})$, $g(q)$, and $J(q)$ paramters of Eq.2.9 using the dynamic model of the robot. To implement the controller on the Kinova Gen3 as described by Eq. 2.9, Kinova's Kortex API[2] is used for real-time communication with the Gen3 in torque control mode, and is also used to retrieve the joint positions (q), velocities ($\dot{q}$), and torques (τ).

3.2.4 Croissant Parts Instance Segmentation

Although this application involves the manipulation of a 3D object, the problem can be reduced to a 2D one through tool design. The designed end-effector allows for the same force to be applied throughout the width of the croissant during the rolling process with its symmetric design (see Sec. 3.2.2). Leveraging on this symmetric manipulation, more robust, open-source computer vision tools designed for 2D images, such as Roboflow[3] and YOLO[4], can be used.

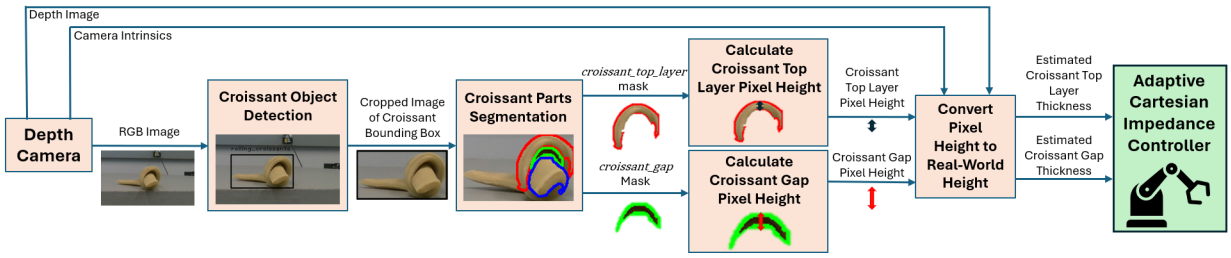


Figure 3.6: Visual perception pipeline for croissant top layer and gap thickness estimation within VAIRO

The overview of the entire computer vision process is outlined in Fig. 3.6. To determine the croissant layer thickness and gaps, a two-stage object detection and segmentation technique is employed. First, object detection is used to detect a bounding box for the manipulated croissant dough. Second, using the cropped image of this bounding box,

[2]https://github.com/Kinovarobotics/Kinova-kortex2
[3]https://roboflow.com/
[4]https://github.com/ultralytics/ultralytics

image segmentation is used to detect different parts of the croissant. Specifically, instance segmentation is used to detect the croissant's core, top layer, and any gaps in between the rolled layers. This two-stage detection architecture significantly improves instance segmentation accuracy by constraining the segmentation area to the croissant's bounding box. This reduces the number of pixels that need classification and directs training toward positive detection, as the model is certain it is analyzing only a croissant. For example, performing image segmentation only on the croissant allows dark spots to be classified as gaps between its layers, whereas segmenting the whole scene would require additional training to reduce false positive detection of shadows as croissant gaps.

To estimate the croissant's top layer (t_{layer}) and gap thickness (t_{gap}) during rolling, Alg.1 is used on both *croissant_top_layer* and *croissant_gap* masks respectively:

Algorithm 1 Estimate Layer or Gap Thickness

Input: *croissant_top_layer* or *croissant_gap*

Output: t_{layer} or t_{gap}

1 **Function** `computeRealMaskHeight(`*mask*`):`
2 Find $p_{ctr} = (x_{ctr}, y_{ctr})$ where $y_{ctr} > y, \forall p = (x, y) \in mask$.
3 Initialize height_list = []
4 **for** x_i *in range(*$x_{ctr} - N$, $x_{ctr} + N$*)* **do**
5 Find y_{top} where $y_{top} \geq y, \forall p = (x_i, y) \in mask$
6 Find y_{btm} where $y_{btm} \leq y, \forall p = (x_i, y) \in mask$
7 Convert $p_{top} = (x_i, y_{top})$ and $p_{btm} = (x_i, y_{btm})$ to real-world coordinates r_{top} and r_{btm} using camera intrinsics and depth data.
8 Compute height $h_i = r_{top_z} - r_{btm_z}$
9 Append h_i to height_list
10 Compute estimated_height = mean(height_list)
11 **return** estimated_height

Alg.1 estimates the thickness of the croissant parts by measuring the height of the masks at its center. In line 2, the center pixel p_{ctr} of the mask is defined as the pixel in the mask with the highest y-value because this is the point of contact between the end-tool and the dough, making this the most critical area to measure for real-time feedback. However, relying solely on the center pixel's height introduces measurement noise. To mitigate this, in lines 4-9, a padding of N pixels surrounding p_{ctr} is included, and the mean height at these pixels is computed. The real-world heights of the mask in the interval $[x_{ctr} - N, x_{ctr} + N]$, where $N = 10$ in this work, are found by taking the difference in real-world z-values between the top most and bottom most pixels at each x_i in this interval.

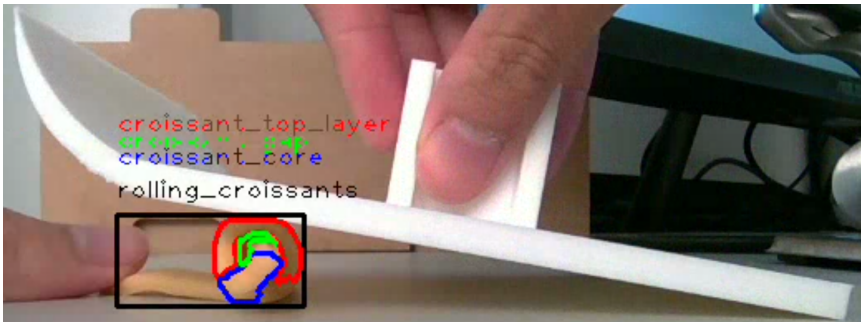


Figure 3.7: Croissant segmentation result with the developed vision model. The black box is the bounding box of the *rolling_croissant_dough* object, the red outline shows the *croissant_top_layer* mask, green shows the *croissant_gap* mask, and blue shows the *croissant_core* mask.

Lastly, in line 10, the mean of the heights in this interval is used as the estimated height of the croissant parts used for adaptive feedback.

In this work, the Intel RealSense D405 camera is used at 848x480 resolution and 60 fps. This setting was found to provide consistent feedback while still maintaining submillimeter measurements with the available computing hardware. The croissant segmentation architecture is implemented by fine-tuning YOLOv8's detect and segment models[5] on 238 croissant rolling images that were collected using the RealSense D405 and labeled using Roboflow. Fig. 3.7 shows a cropped still-shot of successful detection and segmentation of all relevant parts of the croissant during a live video capture of manual croissant rolling with the end-effector. The measurements from this vision-pipeline is then used as feedback in an asynchronous manner by the Kinova Gen3, which operates at 1000 Hz, introduced in the next section.

3.2.5 Vision-Based Stiffness Adaptation

To maintain consistent layer thickness without gaps when rolling croissants, the t_{gap} and t_{layer} thickness measurement from the vision module will be used to adjust the z-stiffness parameter with critical damping of the Cartesian impedance controller while rolling in real time.

Without knowing material properties, it is necessary to first determine a nominal stiffness for the controller to use to start rolling such that the desired manipulation can be

[5]https://github.com/ultralytics/ultralytics

performed without compressing the dough before the vision system can take any measurements. In this study, the nominal stiffness for rolling a croissant is approximated by placing the robotic arm's end-tool over a croissant and gradually increasing the stiffness value of the Cartesian impedance controller. The reference point is set at the table surface ($z = 0$), and stiffness is increased until a deviation of 1% in the croissant's height is detected, indicated by a change in the measured z value of the manipulator's end tool. This process is illustrated for different materials in Sec. 3.3.1. Next, a reference trajectory for the controller to roll the dough is required. A linear trajectory following the top of the rolling dough is calculated using the roll length formula:

$$l = \frac{\pi h^2}{4t} \tag{3.1}$$

where l is the length of the rolled dough, t is the nominal thickness of the dough, and h is the height of the croissant. Creating a linear trajectory in the x direction (Fig. 4.6), Eq. 3.1 can be used to calculate the needed z to maintain contact with the croissant by substituting l with x and h with z:

$$z = \sqrt{\frac{4xt}{\pi}} \tag{3.2}$$

From this trajectory, a manipulator should be able to roll the croissant by using position control with slight compression. However, any slip in the contact between the end-effector and the surface of the dough will cause rolling to fail, as the manipulator cannot recover the contact. Therefore, an adaptive force-based control is necessary.

At the start of the rolling process, the rolling tool is positioned at $z = 0$, in contact with the tip of the croissant dough. As the tool follows the aforementioned linear trajectory to roll the dough, it should maintain consistent contact to ensure proper croissant shaping. However, due to slipping and high static friction needed for rolling, maintaining this contact requires adjusting the tool's position to a z-value 10% lower (determined experimentally) than the calculated trajectory. To maintain accurate trajectory tracking, the impedance controller applies a high stiffness in all directions except the adaptive z direction. Additionally, the high orientation stiffness helps maintain a uniform rolling force across the width of the dough. In the z direction, the nominal stiffness is used until defects in the rolled dough are detected with the vision module. The following equation is used for determining the z stiffness:

$$K_p = K_{p\_nom} - K_{p\_layer} + K_{p\_gap} \tag{3.3}$$

where K_{p_nom} is the nominal stiffness, K_{p_layer} is the stiffness adjustment based on the top layer's thickness deviation and K_{p_gap} is the stiffness adjustment based on the gaps

in between the croissant layers. K_{p_layer} is calculated as the nominal stiffness multiplied by the croissant's top layer thickness percentage deviation from the nominal dough layer thickness. This is calculated using:

$$K_{p\_layer} = \begin{cases} 0 & \text{if } t_{layer} \leq t_{nom} \\ K_{p\_nom} \times \left(1 - \frac{t_{nom} - t_{layer}}{t_{nom}}\right) & \text{if } t_{layer} > t_{nom} \end{cases} \tag{3.4}$$

where t_{nom} is the nominal dough layer thickness and t_{layer} is the estimated croissant's top layer thickness during the rolling process from Sec.3.2.4. The rationale behind this calculation is that detecting compression in the croissant's top layer indicates excessive rolling pressure, necessitating a reduction in the applied force. Consequently, this stiffness adjustment functions solely as a reduction from the nominal value in Eq. 3.3. Moreover, this decrease in stiffness is only applied at the instantaneous moment that the croissant layer is detected to be smaller than nominal until the next measurement is received.

To get a relative size reference of the measured gap, a fractional comparison of the gap compared to the nominal layer thickness is used. This relative value is then used for the K_{p_gap} stiffness gain calculation as a product with the nominal stiffness as follows:

$$K_{p\_gap} \mathrel{+}= \frac{K_{p\_nom}}{100} \times \frac{t_{gap}}{t_{nom}} \tag{3.5}$$

where t_{gap} is the estimated height of the gap from Sec.3.2.4. The K_{p_gap} stiffness adjustment is applied to increase the stiffness of the controller, as shown in Eq. 3.3. The detection of a gap indicates that the dough is stiffer than the nominal value, necessitating an increase in downward rolling pressure to eliminate the gap and ensure proper dough formation. Furthermore, the gap stiffness adjustment division by 100 in Eq. 3.5 was determined experimentally to moderate the rate of stiffness adaptation. Experimental observations revealed that without this scaling, the gain would increase too rapidly and compress the croissants before the vision feedback can take measurements to correct for this error.

In contrast to the instantaneous K_{p_layer} gains, the K_{p_gap} gain here is cumulative, i.e. stiffness will continue to increase until gaps are eliminated and its value is then kept for the rest of the rolling process. Through experimentation with instantaneous K_{p_gap} gains, it was observed that rolling higher stiffness materials resulted in periodic stiffness increases in an oscillatory pattern. This formed many gaps in the croissants when the stiffness decreased. This behavior indicated that stiffness should be kept higher when rolling stiffer material, thus this cumulative increase in stiffness scheme was used instead. Additionally, with this cumulative approach, the K_{p_layer} adjustment can still provide instantaneous pressure relaxation during rolling to correct for overcompensation in stiffness gain or variations in the puff pastry dough's stiffness.

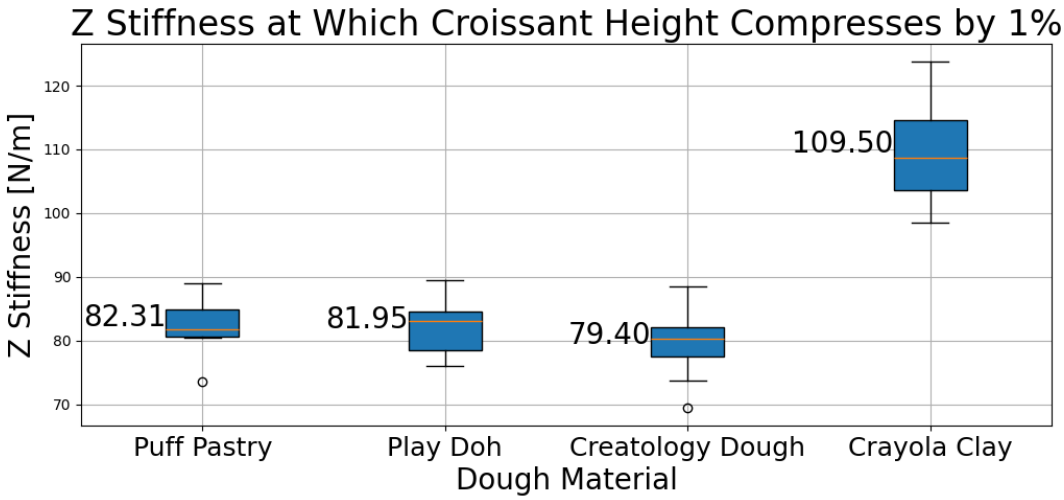


Figure 3.8: Z-stiffness causing 1% height loss in rolled croissants. Box plots show 10 trials per material with means labeled.

3.3 Results

To validate the effectiveness of VAIRO, four different materials were used as "dough" to test the effectiveness of the framework [67]. The materials are puff pastry (thawed from frozen ones), Play-Doh, Creatology sensory dough, and Crayola modeling clay. These materials were chosen to have a good degree of variation in material stiffness. To create consistent test pieces between these materials, they are all processed through a mechanical sheeter to create sheets of 3mm thickness and then cut into equilateral triangles with 140mm side-lengths using a croissant roller cutter.

3.3.1 Nominal Stiffness Identification

To find the nominal stiffness, the stiffness test as described in Sec. 3.2.5 was performed on all the four test materials. Fig. 3.8 shows the result of this stiffness test performed over 10 trials for each material using the Kinova Gen3. These results show that the materials, ranked in order of increasing stiffness, are: Creatology sensory dough, Play-Doh, puff pastry, and finally Crayola modeling clay, which exhibits significantly higher stiffness than the others. From this, a nominal z-stiffness of 100 N/m was chosen to show the effectiveness of VAIRO.

3.3.2 Croissant Instance Segmentation

The trained croissant detection model has a mean average precision calculated at an intersection over union threshold of 0.50 (mAP50) score of 0.994 for three classes: $flat\_croissant\_dough$, $rolling\_croissant\_dough$, and $rolled\_croissants$. The croissant segmentation model has a mAP50 score of 0.889 for three classes: the $croissant\_top\_layer$, the $croissant\_core$, and the $croissant\_gap$. These high accuracy results along with the fast predictions of YOLOv8 allowed for accurate real-time croissant segmentation at 60 Hz.

3.3.3 Adaptive Impedance Dough Rolling

To test the performance of our proposed framework, 10 trials of croissant rolling were performed using both VAIRO and a constant stiffness of 100 N/m on each test "dough" material. Table 3.1 contains the empirical results of the data samples from this test, and Fig. 3.9 shows the box plots of the resultant heights of rolled croissant using VAIRO (blue) vs. using a constant stiffness of 100 N/m (red) on the test "dough" materials over 10 trials. Annotated by the green dashed line is the ideal croissant height of 17.60mm calculated using eq. 3.2 with $x = 81mm$ and $t = 3mm$ dimensions of the test pieces. These results clearly indicate that VAIRO produced croissants closer to the ideal height across various materials compared to using a constant stiffness. For softer materials with measured stiffness below 100 N/m, constant stiffness rolling at 100 N/m resulted in overly compressed croissants, whereas VAIRO achieved heights much closer to the ideal one. Figures 3.10a, b, and c demonstrate VAIRO's ability to detect layer compression and adjust the rolling force accordingly. Notably, Figures 3.10a and b show VAIRO reducing stiffness to approximately 80–90 N/m, aligning with the measured stiffness of puff pastry and Play-Doh in Figure 3.8. The Creatology Sensory Dough tests yielded particularly insightful results, with VAIRO dynamically decreasing and increasing stiffness throughout the rolling process (Figure 3.10c), showcasing its capacity to adapt to both layer compressions and gaps. For the stiffest material tested in this study, Crayola modeling clay, Figure 3.10d illustrates VAIRO increasing rolling stiffness, up to 4 times that of every other material, to close gaps caused by the higher forces required for the stiffer material of Crayola modeling clay. The standard deviations reported in Figure 3.10, ranging from around 30 N/m to 400 N/m, suggests that stiffness adaptations are also necessary to successfully roll the same type of dough between trials and that material properties within the same material also fluctuates.

Furthermore, these results demonstrate that VAIRO enhances the precision of rolled croissants. In the test cases of Play-Doh, Crayola clay, and especially puff pastry, there

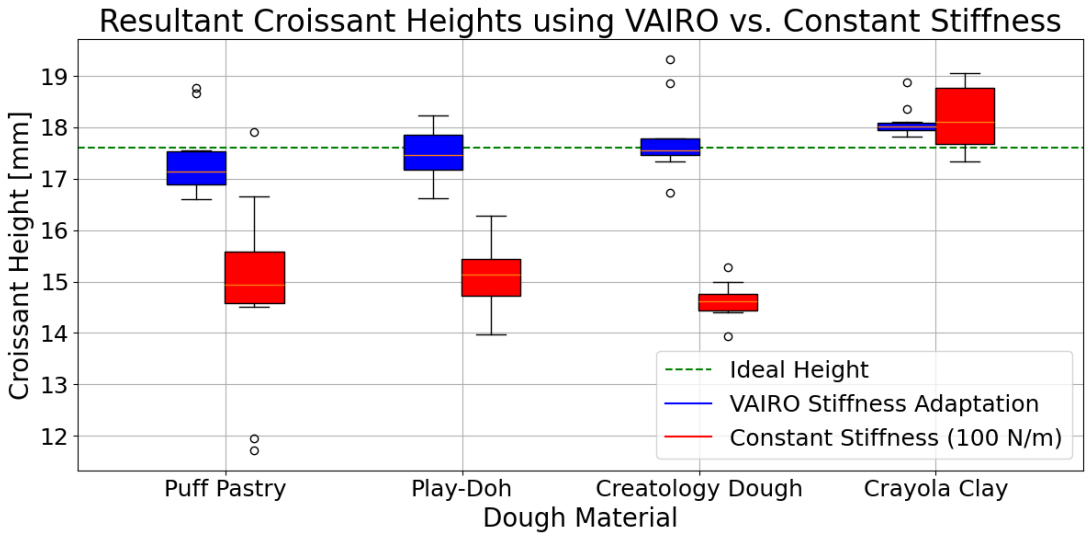
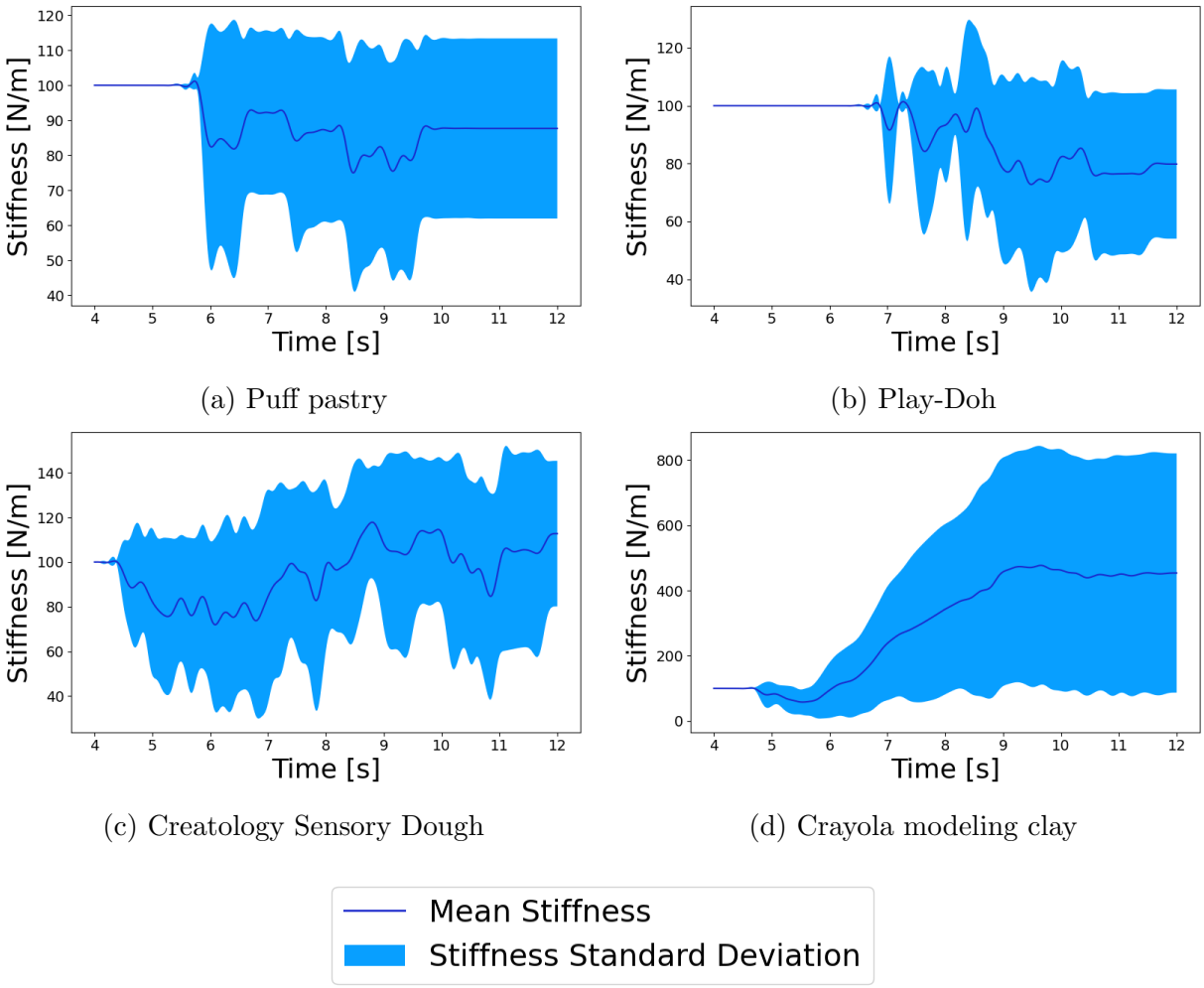


Figure 3.9: Box plot of rolled croissant heights for each material over 10 trials using VAIRO vs. constant z-stiffness (100 N/m). The ideal height is shown by the green dashed line.

is a decrease in the variance of the heights of the rolled croissants when using VAIRO as opposed to constant stiffness. Dough properties can fluctuate within the same material due to factors such as temperature, moisture, composition, and manufacturing defects, leading to inconsistent results when applying a fixed force profile. This effect is especially apparent in the puff pastry tests, where the dough must be kept frozen before use. As the high moisture content altered the dough's rheology during thawing, the pastry softened and became stickier within minutes between each rolling test, causing greater variability than in other test materials. Rolling puff pastry with constant stiffness produced results with 6–7 times higher variance than other trials, including outliers where croissant heights were compressed to around 12 mm. In contrast, using VAIRO under identical conditions yielded more consistent results, with a variance of only 0.58 and a mean height error of -0.21 mm from the target, emphasizing VAIRO's effectiveness in forming soft materials into uniform shapes.

3.4 Conclusions and future work

VAIRO, a Vision-based Adaptive Impedance-control RObotic framework to estimate soft material measurements for manipulation was presented. The development of the framework



(a) Puff pastry

(b) Play-Doh

(c) Creatology Sensory Dough

(d) Crayola modeling clay

Figure 3.10: Z-stiffness adaptation using VAIRO during rolling.

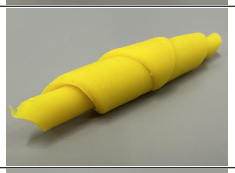
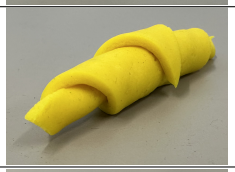
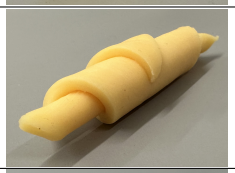
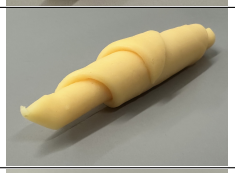
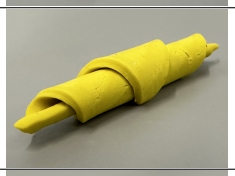
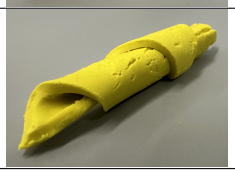
Dough Material	Control	ME [mm]	σ^2	Sample Result
Puff pastry	VAIRO	-0.21	0.58	
	Constant	-2.76	3.55	
Play-Doh	VAIRO	-0.10	0.24	
	Constant	-2.52	0.45	
Creatology sensory dough	VAIRO	0.18	0.57	
	Constant	-2.97	0.13	
Crayola modeling clay	VAIRO	+0.51	0.09	
	Constant	+0.57	0.43	

Table 3.1: Mean error from ideal height (ME), height variance of sample (σ^2), and sample result of rolled croissant of each material over 10 trials using VAIRO vs. a constant z-stiffness of 100 N/m.

was centered around dough rolling for the manufacturing of artisanal grade croissants using CoBots as a proof-of-concept test case. A two-stage computer vision approach, consisting of a first stage dough detection and a second stage croissant parts segmentation were presented to extract a croissant's top layer and gap masks to estimate their thicknesses. These measurements were then used to adjust stiffness parameters for force regulation using a Cartesian impedance controller, which was able to increase its stiffness based on measured gaps and decrease its stiffness based on layer compressions to roll dough into croissant shapes. Experiments with a physical CoBot on four different types of "dough": puff-pastry, Play-Doh, Creatology sensory dough, and Crayola modeling clay, show the effectiveness of VAIRO, highlighting how the adaptation of the controller stiffness in real-time using visual measurements aids in maintaining layer thickness and eliminating gaps, leading to the final tightly rolled croissant shape without flattening.

The friction model of the Kinova Gen3 was not available and, therefore, was not implemented in this work. Due to the high joint friction, the resultant force regulation was not as precise as desired and with latency. This resulted in the manipulator exerting undesired forces that do not correlate to the desired stiffness, inhibiting proper analysis of the force regulation. Future work to investigate this adaptive Cartesian impedance controller includes the development of a more reactive and precise controller such that stiffness modulation schemes can be investigated to determine better measurement to stiffness adjustment mappings for soft material manipulation. VAIRO implemented on manipulators with higher precision force sensing or friction compensation on the Kinova Gen3 may also lead to improved outcomes.

Croissant parts segmentation could also be improved upon in future work with the integration of object tracking techniques such as optical flow. The current vision framework performs croissant detection on each frame independently, so failed detection in certain frames results in inconsistent feedback to the controller. Using object tracking can assist in providing consistent estimation throughout the entire rolling motion and provide more reactive feedback to the controller.

The VAIRO framework can be easily translated to the manipulation of other soft materials where rheological properties are unpredictable or a general purpose solution is desired, such as surgical operations on soft tissue. The only translation needed to use this adaptive impedance control scheme is to define visual indications and metrics of the soft material's stiffness.

Chapter 4

Vision based Adaptive Impedance Control Applied to Skin Incisions

4.1 Introduction

Robotic-assisted and autonomous robotic surgery can allow for safer, more efficient, and more optimized surgical procedures [65] [56]. Many current robotic-assisted procedures leverage robots for improved access to the underlying anatomy (minimally invasive), integrate preoperative planning within the surgical workflow, and allow for more targeted, precise interventions [56]. Soft tissue incisions are the critical first step in many surgical procedures and key for minimally-invasive procedures, where smaller incisions facilitate healing and reduce the formation of scar tissue. Teleoperated robotic systems, such as the da Vinci robot, are commonly used for these types of surgical procedures as robots allow for increased precision, dexterity, and haptic feedback [55]. Leveraging CoBots for minimally-invasive procedures, though, could offer surgeons additional flexibility during procedures, where CoBots could be adapted to perform a variety of tasks using different end-effectors, acting as surgical assistants, working alongside surgeons in operating rooms.

Soft tissue incisions are challenging for CoBots to perform due to the nonlinear deformation of tissue during manipulation [23]. Effective robotic control must adapt to this dynamic behavior to successfully execute surgical tasks. This is further challenged by the variability in mechanical properties of skin, with thickness varying by body region [86], elastic modulus varying with age and skin layers (epidermis, dermis, and hypodermis) [45], and environmental conditions (temperature and humidity) affecting mechanical properties [32]. As a result, skin models are complex to use and develop for deformation tracking,

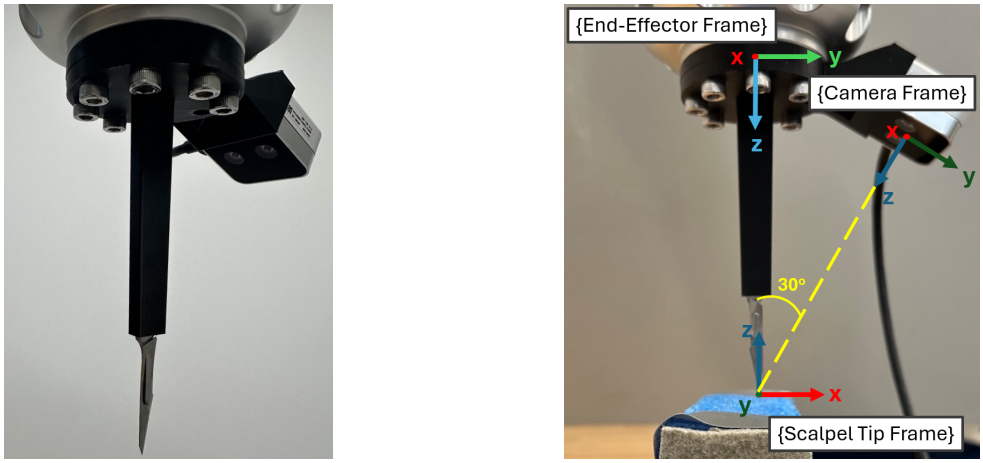


Figure 4.1: Scalpel end-tool with the RealSense D405 close-range stereo vision camera

and do not generalize well to different cases. To overcome this issue, vision feedback can be leveraged to allow for model-free approaches to track dynamic tissue movement and enable robots to perform soft tissue manipulation.

While incisions must be optimized for surgical success, post-operative implications of surgical incisions are also key to consider. Scarring and scar tissue formation are common outcomes post-surgical incisions, which can potentially lead to psychological and functional side-effects due to physical impairment (stiffer, less elastic tissue [16]) and societal stigma [6] [10]. From a procedural standpoint, it has been recognized that tension forces across a wound affect scar formations [6]. Studies have found that excessive mechanical stress exerted in the wound environment induces scarring via activation of mechanotransduction pathways [6] [5] (conversion of mechanical stimuli into biochemical signals that result in intracellular changes [31]). Therefore, minimizing mechanical forces during incision is crucial to reduce post-operative scarring and its associated complications.

Autonomous soft tissue incisions and manipulation is a growing area of medical robotics research, with a key focus on developing different methods to address the unique mechanical behaviour of soft tissues [68]. Leonard *et al.* developed a robotic system capable of performing suturing [48] named Smart Tissue Anastomosis Robot (STAR). STAR is a robot arm equipped with a monocular colour camera and an actuated suturing tool [48]; it allows surgeons to specify the location of each stitch or allows for the automatic placement of stitches [48]. Li *et al.* proposed a surgical perception framework that continuously collects 3D geometric data to map deformable surgical fields and tracking instruments [49]. Their framework proposed using visual perception from endoscopic image data along with

a robotic control loop to allow for accurate soft tissue manipulation [49]. Li *et al.* implemented the framework for evaluation on the da Vinci Surgical System and demonstrated that the framework allowed for reliable tracking during significant tissue deformations [49]. Liang *et al.* developed a system capable of performing autonomous incision of the cornea by leveraging a monocular camera [50], which is used to determine the position of the scalpel and to capture the surgical scene [50]. Experiments performed with the robotic system on ex vivo porcine eyes showed that the system was capable of performing precise cuts [50]. All these works demonstrated precise results; however, they all also require specialized robotic systems or cameras to perform their tasks, limiting the scalability of their work to more general-purpose robots such as CoBots or humanoids. More generalized surgical solutions were recently investigated using a humanoid for assistance in surgical practices [3], but still rely on teleoperation of the robot with motion mapping, where incision procedures were not optimized. In addition, given that the system requires a medical professional to teleoperate the robot, it does not address the issue of the increasing demand for healthcare workers.

With the Industry 4.0 paradigm, the integration of robotics into the medical field has accelerated. Large robotics manufacturers are driving this effort with the KUKA Innovation Award 2025 focusing on medical robotics for a second time [34], and Kinova breaking into medical CoBot production[1]. Furthermore, the integration of CoBots into the surgical field has shown to have a great interest with the concept of a CoBot surgical assistant in anterior cruciate ligament (ACL) surgery, including a soft-tissue incision task, drawing attention from robotics leaders in both industry and academia at the 2025 42nd IEEE-RAS International Conference of Robotics and Automation (ICRA)[2], see Fig.4.2.

4.1.1 Background

Soft tissues are unmineralized, flexible tissues that comprise of a variety of tissues within the body, including skin, tendons, ligaments, blood vessels, and more [36]. They display a nonlinear stress-strain relationship, are capable of large deformations, and have viscoelastic properties [23] [44]. Additionally, the mechanics of soft tissues are highly variable, depending on tissue type, age, and sex [44].

To model the behaviour of soft tissues, many mathematical models have been proposed [21]. In particular, different models can be used to describe specific behaviours of tissue, including: hyperelasticity (nonlinear or linear), viscoelasticity (nonlinear or linear),

[1]https://www.kinovarobotics.com/sector/medical-robotics
[2]https://2025.ieee-icra.org/



Figure 4.2: Surgical CoBot showcase with KUKA at the 2025 42nd IEEE-RAS International Conference of Robotics and Automation (ICRA) by Jeffrey Lee, Teresa Marotta, Stewart McLachlin, and Yue Hu

anisotropy, and heterogeneity [21]. Additionally, soft tissue behaviour changes over time and during damage, a critical behaviour present during surgical procedures [21]. Since many mathematical models exist for soft tissue behaviour, selecting the appropriate model for use is highly dependent on the application and intended tissue for manipulation or incisions [21]. As a result, developing a model-based generalized solution for incision automation would require significant development to create the vast number of models needed for various tissues.

Consistent with the prior work of Chapter 3, the inherent complexity and variability of the mechanics of the targeted material, soft tissue in this case, motivates the use of model-free approaches. The vision-based adaptive impedance control framework proposed in this work aligns well with this method and is employed in the context of soft tissue manipulation.

4.1.2 Contributions

In this work, we introduce a generalized autonomous solution to performing surgical incisions using CoBots and stereo vision feedback. This solution is designed to be adaptable

across various procedures and robotic platforms. Our main contributions are as follows:

1. A compact end-tool scalpel with an integrated stereo vision camera (Fig. 4.1), capable of capturing fine-grained soft tissue deformations for extensible and vision-based feedback control during incisions.
2. The VAISI framework: a model-free, vision-driven adaptive impedance control system for CoBots, designed to autonomously perform skin incisions while minimizing tissue trauma through:
   (a) Millimeter-scale incision depth precision;
   (b) Adaptively minimizing incision force.

This work is experimentally verified using the KUKA LBR Med CoBot (Kuka Robotics, Germany), with incisions performed on porcine belly and hock skin.

4.2 Methodology

4.2.1 Design of the Stereo Vision Scalpel End-Tool

The development of the compact scalpel end-tool with a stereo vision sensor presented in this work consists of two main components: selection of a stereo camera for accurate measurements suitable for surgical applications and the mechanical design of the scalpel camera assembly.

Vision Module

The stereo vision module selected for integration with the end-effector is the Intel RealSense D405. This sensor offers key advantages for robotic applications thanks to the broad compatibility of the Intel RealSense SDK with widely used robotics platforms (C++, Python, and both Robotics Operating System - ROS 1 and 2). Additionally, the D405 is optimized for close-range depth sensing and delivers measurement accuracies of up to 0.1 mm at a distance of 7 cm [15]. This accuracy is key for the surgical environment, where high precision is critical. Furthermore, the compact form factor of the D405 supports the development of space-constrained end-effector designs, making it well-suited for minimally invasive or collaborative surgical environments.

Stereo Vision Scalpel End-Tool Design

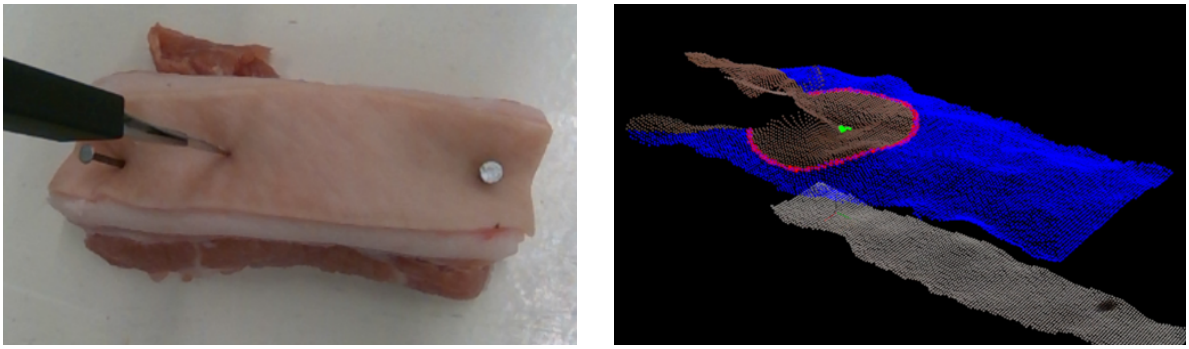
The scalpel end-tool with a stereo vision module designed for this work is depicted in Fig. 4.1. To optimize the balance between minimizing the end-tool's footprint and enhancing the accuracy of the camera's measurements of the manipulated soft tissue, the camera is positioned 7 cm from the heel of the blade. The camera's face is oriented at a 30 deg angle relative to the scalpel and is aligned parallel to the blade (see Fig. 4.1), thereby reducing the occlusion of the tissue by the scalpel. The deformability of soft tissues often results in surfaces with significant curvature, which can complicate accurate imaging. This issue is especially prominent during incisions, where protrusions and cavities often form from this manipulation. To mitigate this issue, it is crucial for the vision system to be positioned and angled at a sufficient distance from the tissue's surface, which is achieved with this design. Furthermore, the 30 deg angle offers the additional advantage of utilizing the tool's vertical height to achieve the necessary camera-to-tissue distance, thereby reducing the tool's overall cross-sectional footprint. This configuration allows for greater visibility and monitoring of incisions in a surgical environment.

4.2.2 Incision Feature Estimation using Stereo Vision

To accomplish the task of incisions in soft tissue in robotics, two critical pieces of feedback must be captured: the location of the tissue and its deformation while it is being manipulated. With the goal of depth precision in mind, the vision algorithm developed for VAISI measures:

1. The depth of the tip of the scalpel blade from the surface of the soft tissue.
2. The indentation deformation on the soft tissue from the incision interaction of the scalpel blade.

Using the point-cloud captured by the stereo-vision camera, Alg. 2 was developed using the Point Cloud Library (PCL) [61] to capture these measurements as *incision_depth* and *indent_depth*, respectively. Fig. 4.3 shows a visualization of a point-cloud processed by Alg. 2 during an incision.



(a) (b)

Figure 4.3: (a) depicts the RGB image captured by the end-tool camera during an incision and (b) is the visualization of the processed point-cloud using Alg.2, where green represent points used for *scalpel_tip_z*, red represent points calculated for *indent_r*, and blue represents points used for *nominal_skin_z*

.

Algorithm 2 Point cloud processing to calculate incision depth and indentation deformation

Input: Point cloud, pc, captured by the stereo vision camera on the scalpel end-tool

Output: Distance between the scalpel tip and skin surface, and skin indentation

Function `computeSkinFeatures(`pc`):`

- Transform pc to {Scalpel Tip Frame}.
- Crop pc
- Remove outliers in pc
- $scalpel\_tip\_z = \frac{1}{10}\sum_{i=1}^{10} z_i$, where$\{(x_i, y_i, z_i)\}_{i=1}^{10} = \text{argmin}\left(\sqrt{x^2+y^2}\right) \in pc$
- $indent\_r = \min\{indent\_r = \sqrt{x_i^2+y_i^2} \mid (x_i, y_i, z_i) \in pc, \nabla z_i = 0\}$
- $indent\_z = \frac{1}{N}\sum_{i=1}^{N} z_i \quad \forall \quad \sqrt{x_i^2+y_i^2} = indent\_r, (x_i, y_i, z_i) \in pc$
- $indent\_depth = indent\_z$ - $scalpel\_tip\_z$
- $nominal\_skin\_z = \frac{1}{N}\sum_{i=1}^{N} z_i$, $\forall\sqrt{x_i^2+y_i^2} > indent\_r,\ (x_i, y_i, z_i) \in point\_cloud$
- $incision\_depth = nominal\_skin\_z$ - $scalpel\_tip\_z$
- **return** *incision_depth* and *indent_depth*

Firstly, in line 2, the point cloud is transformed from the Camera Frame to the tip of the scalpel, defined at the origin of the Scalpel Tip Frame, as shown in Fig. 4.1. In line 3, the point cloud is cropped to a 2cm radius bounding box around the origin of the Scalpel Tip Frame to reduce the computation for the following steps while still capturing local features. In line 4, statistical outliers are removed by calculating the mean distance to each point's 25 nearest neighbors and eliminating points with mean distances greater than one standard deviation above the global mean [61]. In line 5, the depth of the scalpel tip at the incision point, defined as $scalpel\_tip\_z$, is calculated as the mean z of the 10 points closest to the x-y origin of the Scalpel Tip Frame. In line 6, skin indentation deformation is identified. Assuming that the skin indentation forms in the shape of a 3D Gaussian, an indentation is identified by finding the local maxima radius closest to the origin of the Scalpel Tip Frame. This radius is defined as $indent\_r$ and, in line 7, the average z-value of the points lying at this radius is defined as $indent\_z$. In line 8, $indent\_depth$ is computed as the difference between $indent\_z$ and $scalpel\_tip\_z$. In line 9, the average z-value of the points lying outside of $indent\_r$ is defined as $nominal\_skin\_z$ to represent the height of the skin surface at its nominal (relaxed and uncompressed) state. Finally, in line 10, under the assumption that the desired incision depth is relative to the surface of the soft tissue at its nominal state, $incision\_depth$ is calculated as the difference between the $scalpel\_tip\_z$ and $nominal\_skin\_z$.

4.2.3 Vision-Based Adaptive Cartesian Impedance Control

The desired end-tool position, x_0, and stiffness, K, of the impedance controller as introduced in Sec. 2.1, are adapted from the vision feedback. $incision\_depth$ provides feedback on the position of the soft tissue, and therefore provides the z-position required for the scalpel end-tool to reach a desired incision depth. $indent\_depth$ provides feedback on skin deformation caused by forces insufficient to penetrate the tissue layers, resulting in an indentation through compressing the skin surface rather than incision. As such, VAISI simultaneously uses the $incision\_depth$ feedback for visual servoing to the incision depth with x_0, and the $indent\_depth$ feedback to control the incision force through adaptive stiffness, K, and critical damping, D, to create precise-depth incisions. The implementation of this combined control is outlined in the following, supplemented by Alg. 3 and Fig. 4.4.

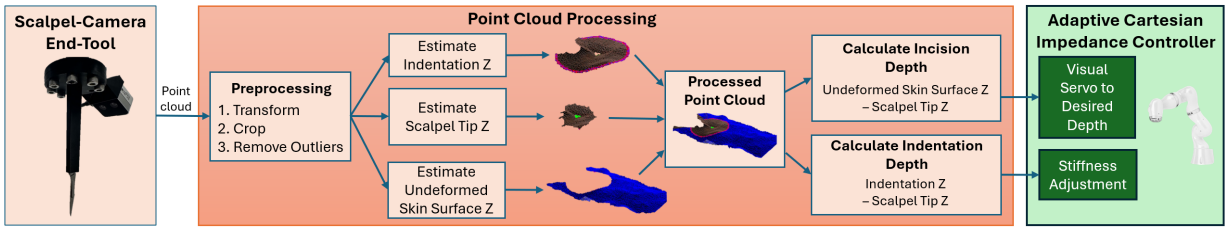


Figure 4.4: VAISI Framework Overview

Algorithm 3 Incision Control

Input: *incision_depth* and *indent_depth*

Output: CoBot end-effector position p_{ctrl}

1 **Function** `incisionControl`(*incision_depth*, *indent_depth*)**:**
2 **while** *incision_depth* $<$ *desired_depth* **do**
3 $p_{des}.z = \min(\text{PID}_z(incision\_depth), desired\_depth \times 1.5)$
4 **if** *indent_depth* $\geq$ *previous_indent_depth* **then**
5 K_z += 5
6 **else**
7 K_z -= 5
8 *previous_indent_depth* = *indent_depth*
9 **for** *number of x trajectory steps* **do**
10 $p_{des}.z = \min(\text{PID}_z(incision\_depth), desired\_depth \times 1.5)$
11 $p_{des}.x+ = x\_step$
12 **return** p_{ctrl}

Visual Servoing via Incision Depth Feedback

To control the incision depth on the deformable soft-tissue, the control point, x_0, for the Cartesian impedance controller is adaptively set using the vision feedback *incision_depth*. The incision is first performed by a linear downward trajectory using proportional-integral-derivative (PID) control in the negative z-direction of the scalpel tip frame, with *desired_depth* as the input and *incision_depth* as the feedback signal (line 3). In addition, a maximum

50% overshoot for the control points is also imposed to prevent damage of underlying anatomy and excessive cuts (line 3). The framework waits until the desired depth is reached (line 2), then a linear x-trajectory is used to create an incision while the depth-PID control remains active (lines 9-11).

Adaptive Stiffness via Skin Indentation Feedback

Surgeons perform an incision by modulating downward pressure until the scalpel penetrates the skin and then continues to create a smooth cut [84]. Similarly, VAISI modulates the downward stiffness of the Cartesian impedance controller to penetrate the skin by using the *indent_depth* feedback. With the notion of indents indicating greater incision forces needed to split the tissue, as *indent_depth* stays the same or increases during the incisions, VAISI will adapt by increasing z-stiffness and thereby the incision force (lines 4-5). If *indent_depth* decreases, z-stiffness decreases (lines 6-7). Once the desired incision depth is reached, the adjusted stiffness is then maintained to create a smooth cut.

4.3 Experiments and Results

4.3.1 Evaluation of the Stereo Vision Scalpel End-Tool

To verify the performance of the stereo vision scalpel end-tool introduced in Sec. 4.2.1, the tool is mounted on a 7 DoF KUKA LBR Med, with a #11 scalpel blade. Foam was selected as the test specimen as it is penetrable with minimal deformation, enabling evaluation of vision measurements independent of material deformation. The end-tool is positioned perpendicularly on top of foam so that a z-trajectory results in a straight stab incision, as seen in Fig. 4.1. Using this setup, desired incision trajectories at 0.5 mm/s, 1 mm/s, 1.5 mm/s, and 2 mm/s were used to create incisions of 10mm in the foam using position control on the robot. The z-displacement measured from the robot is recorded as the ground truth of the incision depth to be compared to the *incision_depth* measurements from Alg.2.

Measurements obtained from the experiments can be seen in Fig. 4.5, where the dotted lines represent the robot measured displacement and the solid lines are the vision measured displacement. The bias (compared to the displacement measured by the robot) and noise (measured using the standard deviation of residuals) are reported in Table 4.1. The experiment reveals that using the RealSense D405 camera for the scalpel end-tool achieves

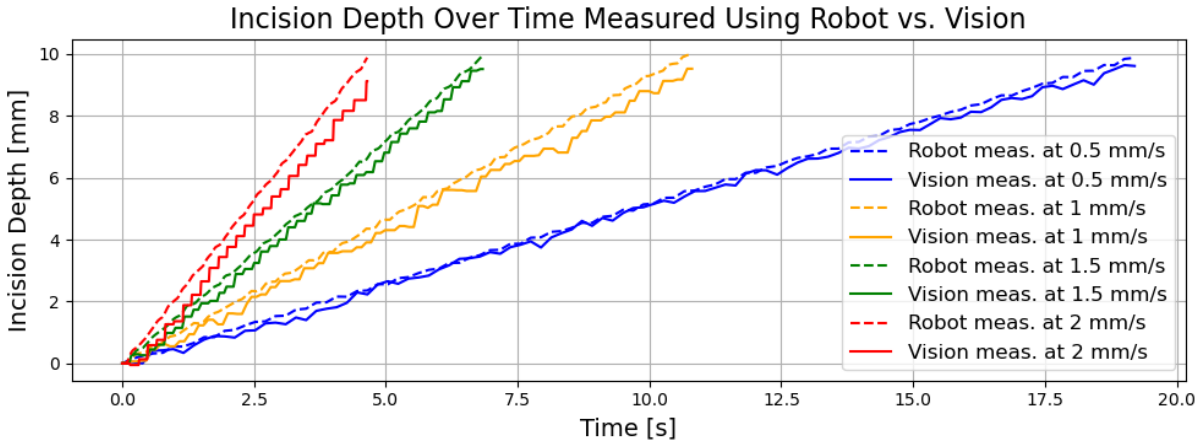
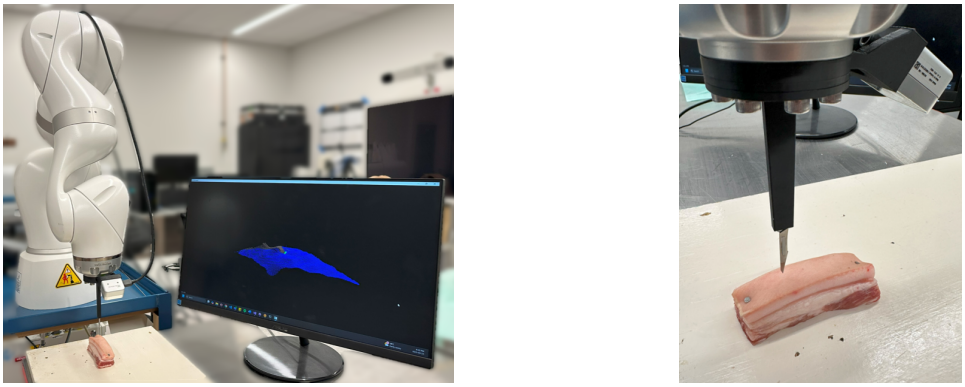


Figure 4.5: Incision depth measurement on a piece of foam from the LBR Med vs. the Realsense D405 on the end-tool at four incision velocities.

Desired Incision Velocity [mm/s]	Bias [mm]	Noise (Standard Deviation of Residuals) [mm]
0.5	-0.144	0.144
1.0	-0.372	0.202
1.5	-0.321	0.130
2.0	-0.717	0.225

Table 4.1: Bias and noise of incision depth measurements using the RealSense D405 on the end-tool at four desired incision velocities, using the z-displacement measured from the LBR Med as the ground truth.

sub-millimeter accuracy at the reported velocities. However, a decrease in both accuracy and precision is seen as the velocity of the incision increases. This degradation in performance can be attributed to two main factors. Firstly, there is high latency from point cloud processing. Using PCL with the available hardware, an AMD Ryzen™ 7 7745HX processor, for this experiment produced vision measurements between 4-8 Hz, resulting in an increase of measurement errors at higher speeds. Secondly, it was observed that there was an increase in vibrations from increasing the robot's operating speed, resulting in vibrations on the camera sensor. This likely attributed to the increase in measurement noise observed in this test. From this result, the incision performed in the next sections to evaluate VAISI was limited to a maximum of 0.5 mm/s to optimize the accuracy of the vision measurements.



(a) Overall setup (b) Close-up view

Figure 4.6: Experimental setup for performing incisions on porcine skin using the LBR Med and scalpel camera end-tool

4.3.2 Experimental Setup

The same hardware setup as the previous section is used to conduct robotic skin incisions using the VAISI framework. Two types of tissues were used as the specimens in this work: porcine belly skin and hock skin. Porcine models were chosen as they provide mechanical forces across their skin most similar to those found in humans [6]. These specimens were secured to a hardwood surface to keep the tissue in place during the incision process (as in a real surgical procedure, the incision location would likely be immobile). Fig. 4.6 depicts this setup with porcine belly as the test tissue.

4.3.3 Results

To validate the effectiveness of the VAISI framework on creating incisions with minimal force, incisions using no stiffness adaptation and incisions using VAISI were both performed. To further demonstrate the VAISI framework's effectiveness, incisions of different desired depths were also performed.

Without Stiffness Adaptation

Due to the unknown mechanical properties of the skin tissue and underlying anatomical structures, it is not feasible to define a constant force required to achieve incision. Nevertheless, with the objective of minimizing incision force, incisions were conducted using a

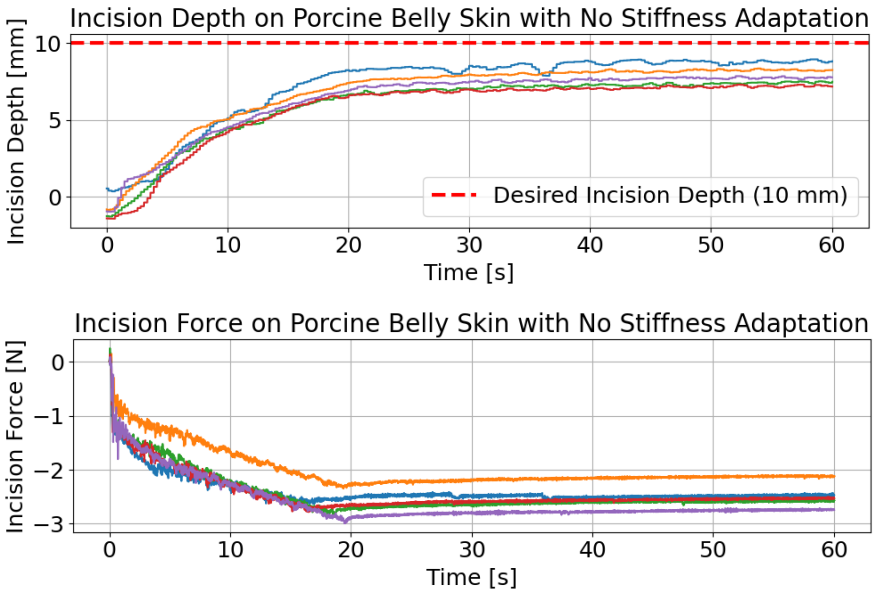


Figure 4.7: Incision depth and force using constant stiffness of 200 N/m on porcine belly skin

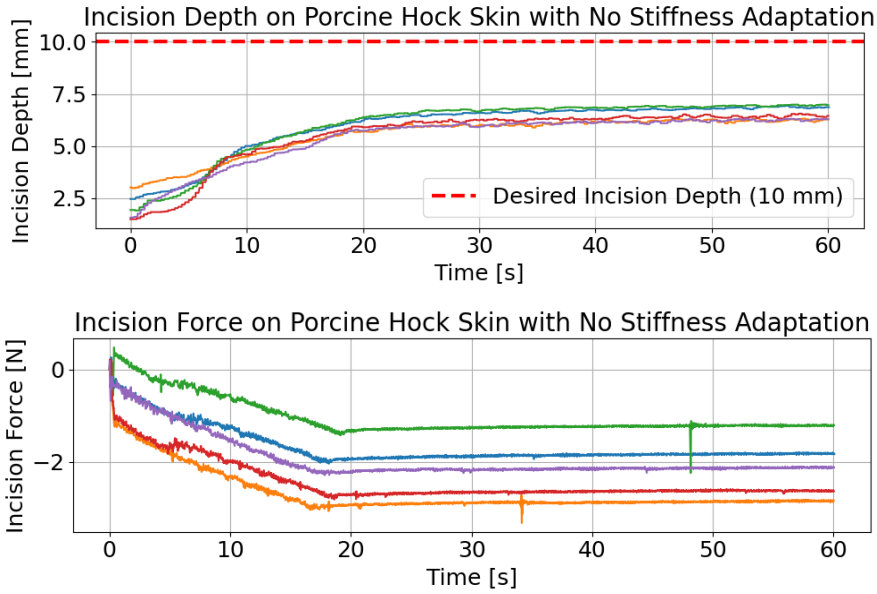


Figure 4.8: Incision depth and force using constant stiffness of 200 N/m on porcine hock skin

low constant stiffness of 200 N/m was arbitrarily chosen to assess whether incisions can be achieved through the application of minimal force alone. Figs. 4.7 and 4.8 show the results on porcine belly and hock skin, respectively, with a target incision depth of 10 mm over five trials. These results demonstrate that a constant stiffness of 200 N/m was insufficient to reach the target depth. Incision force plots, measured as the z-direction force on the end-effector, showed convergence to a steady state across all trials, while incision depth plateaued below the desired value. This indicates that automated incisions to a specified depth cannot be achieved using a constant force or stiffness that fails to overcome tissue mechanics. The failure of these trials highlights the necessity of stiffness adaptation to modulate applied forces and achieve precise incisions in tissues with unknown mechanical properties.

VAISI Framework

To validate the performance of the VAISI framework, incisions on porcine hock skin and belly skin were performed with desired depths of 5mm and 10mm. 10 trials of each test case were conducted. PID values for Alg. 3 were 0.75, 0.25, and 0.05 respectively. The vision measurements of *incision_depth* and *indent_depth* along with the adapted impedance controller stiffness is reported in Fig. 4.9. The incisions were then measured at the start, middle, and end using a caliper to verify the depth of the incisions. The mean error, overshoot, undershoot, and standard deviation from the desired depth of these measurements are reported in Table 4.2.

It is worth noting that the specimens used in this study exhibited variation in the thickness of the skin, fat, and muscle layers, even within the same anatomical region. This resulted in variations in indent deformation measurements and therefore stiffness adaptations. Especially, specimens with a large fat layer underneath the tough skin allowed for greater deformation under the incision force, instead of remaining taut for the blade to slice through the skin. This resulted in an especially tough manipulation challenge for a shallow 5mm incision on porcine belly, which had a large fat layer underneath a skin layer close to 5mm, resulting in incisions too shallow and the worse performing case.

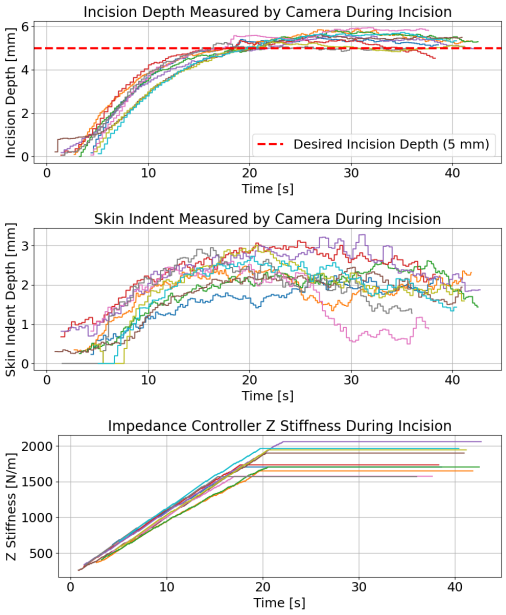
Furthermore, skin indentation increased during the entire incision process to the desired depth for all cases. This is likely due to the triangular geometry of the #11 scalpel blade preventing the tissue to return to its uncompressed state around the incision point. This indicates that the stiffness reduction in the VAISI framework was not necessary in this case, but may be useful for cases where the incision instrument has a uniform shape, such as a needle.

Overall, VAISI demonstrated highly promising performance, achieving sub-millimeter mean depth accuracy and maintaining maximum errors below 2 mm across all trials on porcine hock skin and 10 mm incisions on porcine belly skin. Notably, the low standard deviation observed across experiments highlights the system's excellent precision and repeatability. These results underscore VAISI's potential as a robust and reliable tool for precise incision control in soft tissue, paving the way for future applications in both preclinical research and clinical settings where high accuracy and consistency are critical.

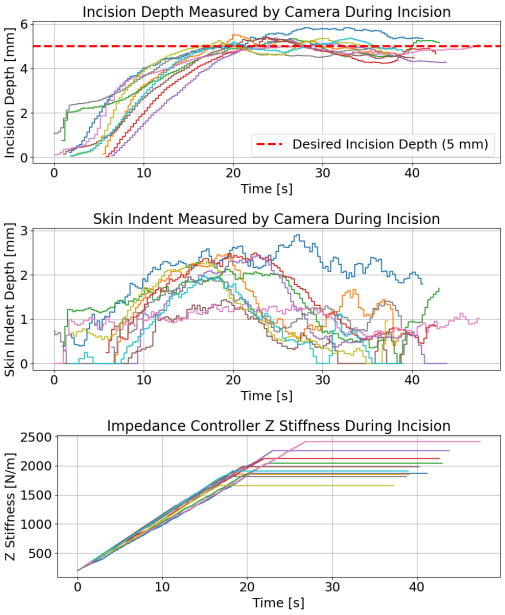
4.4 Conclusion and Future Work

VAISI, a **V**ision-based **A**daptive Impedance-control framework for **S**urgical **I**ncisions along with a compact stereo-vision scalpel end-tool was presented. VAISI is a model-free framework designed with the goal of depth precision and minimizing tissue damage by controlling both a CoBot's position and applied force through adaptive impedance control. Utilizing point-cloud data of a target soft-tissue, its position relative to the CoBot and its deformation during manipulation is captured. These two visual feedbacks are then used to visual servo the CoBot to a desired incision depth and control the applied force by adapting the stiffness of the impedance controller, respectively. Furthermore, a compact stereo vision scalpel end-tool design was developed for CoBots that can serve as surgical assistants in operating rooms. The overall system was experimentally validated with incisions performed on porcine hock skin and belly skin with desired depths of 5mm and 10mm. The experiments revealed promising results with sub-millimeter mean errors for three of the four test cases, and precision with a maximum of 1.23mm standard deviation in incision depths reported between all cases.

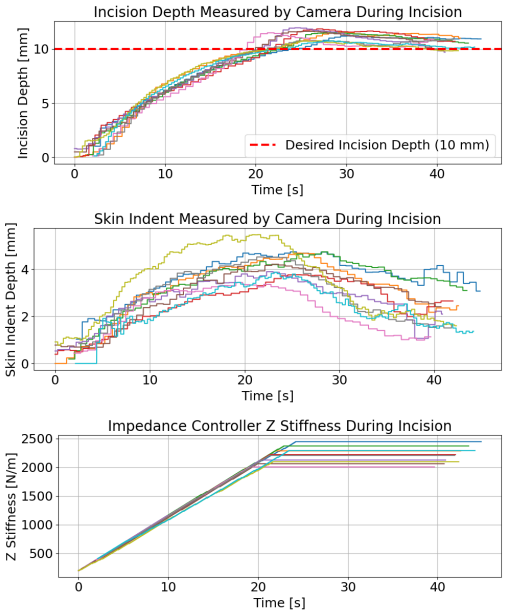
The work presented serves as a step towards fully autonomous surgery by tackling a common procedure in surgery: skin incisions. Currently, VAISI has shown promising results for automating small stab incisions with #11 scalpel blades. Investigations on using different surgical scalpel blades and larger, different incisions could be extended on the VAISI framework. Multi-modal force adaptation could also be investigated by measuring the reaction forces on the scalpel as feedback. Ultimately, VAISI, in conjunction with the stereo vision scalpel end-tool, offers a generalized, model-free, vision-based solution for robotic incision tasks in soft materials such as soft-tissue. This approach has the potential to be deployed across various robotic platforms, including CoBots, mobile manipulators, and humanoids.



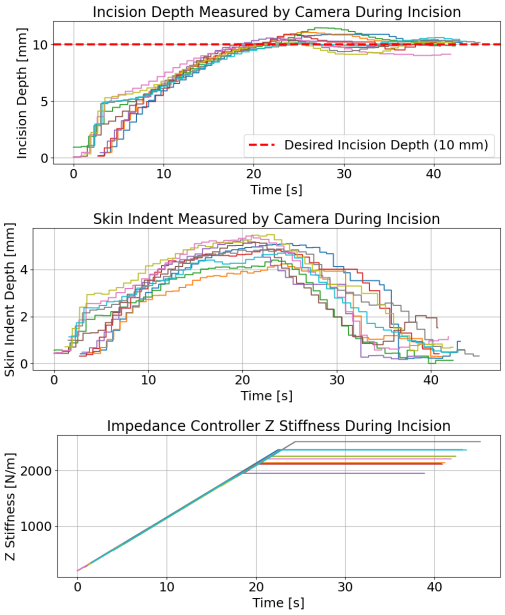
(a) Type: belly skin; Desired depth: 5mm



(b) Type: hock skin; Desired depth: 5mm



(c) Type: belly skin; Desired depth: 10mm.



(d) Type: hock skin; Desired depth: 10mm

Figure 4.9: Vision measured incision depth, skin indentation, and adapted impedance controller z-stiffness during the incision process of each test condition on porcine.

Point of Incision	Start	Middle	End
Mean Error from Desired Depth [mm]	-2.30	-1.87	-1.70
Maximum Overshoot from Desired Depth [mm]	No Overshoot		
Maximum Undershoot from Desired Depth [mm]	-2.56	-2.59	-2.6
Standard Deviation of Incision Depth [mm]	0.48	0.87	1.06

(a) Porcine belly skin with 5mm desired incision depth

Point of Incision	Start	Middle	End
Mean Error from Desired Depth [mm]	-0.46	0.03	-0.24
Maximum Overshoot from Desired Depth [mm]	1.56	1.72	1.64
Maximum Undershoot from Desired Depth [mm]	-1.85	-1.46	-1.70
Standard Deviation of Incision Depth [mm]	1.23	1.23	1.19

(b) Porcine hock skin with 5mm desired incision depth

Point of Incision	Start	Middle	End
Mean Error from Desired Depth [mm]	-0.18	-0.70	-0.74
Maximum Overshoot from Desired Depth [mm]	0.77	0.39	0.36
Maximum Undershoot from Desired Depth [mm]	-1.7	-1.81	-1.85
Standard Deviation of Incision Depth [mm]	0.86	0.71	0.80

(c) Porcine belly skin with 10mm desired incision depth

Point of Incision	Start	Middle	End
Mean Error from Desired Depth [mm]	0.56	0.57	0.47
Maximum Overshoot from Desired Depth [mm]	1.69	1.61	1.94
Maximum Undershoot from Desired Depth [mm]	-0.94	-1.06	-1.22
Standard Deviation of Incision Depth [mm]	0.82	0.90	1.00

(d) Porcine hock skin with 10mm desired incision depth

Table 4.2: Performance at the start, middle, and end of incisions over 10 trials of each test condition

Chapter 5

Conclusion and Future Work

This thesis presented a model-free, vision-based adaptive impedance control framework to address the challenges of robotic manipulation of soft materials with variable and unmodeled properties. Unlike conventional methods that rely heavily on predefined physical models or large-scale training data, the proposed framework dynamically adjusts contact forces using real-time visual feedback. This approach enables CoBots to adaptively interact with deformable materials. The framework was adapted and validated in two distinct domains, demonstrating both its generalizability and effectiveness across vastly different tasks involving soft material manipulation.

In Chapter 3, the proposed framework was first applied to the domain of craft food preparation, where a CoBot was tasked with rolling croissant dough—a task traditionally reliant on human skill to ensure artisanal quality. An end-effector was developed to mimic the shape of a human hand during the rolling action, while a vision module was trained using YOLOv8-based detection and segmentation to extract croissant layer and gap features in real time. These measurements were used to modulate the stiffness of the adaptive Cartesian impedance controller, ensuring consistent layer thickness and gap eliminations despite variations in dough stiffness across different materials. Experimental evaluations using four types of dough with different stiffness showed that the framework created croissants with greater consistency, outperforming fixed-stiffness impedance controllers. The VAIRO system thus demonstrated how real-time visual feedback can maintain product consistency and quality in soft material manipulation tasks without the need for manual tuning or rigid assumptions about material behaviour.

Chapter 4 extended the application of the proposed framework into a medical robotics context, where precision and force control are critical. A stereo vision-enabled scalpel end-

tool was developed and integrated with a CoBot to perform controlled incisions on soft biological tissues. Through point cloud processing and incision feature estimation algorithms, the system measured parameters such as incision depth and skin deformation in real time. This information was used to adapt the stiffness of the Cartesian impedance controller to maintain desired incision depth while avoiding excessive force. Experiments showed that the VAISI framework achieved greater depth consistency compared to constant stiffness control, under varying tissue properties between porcine belly and hock skin. These results highlight the system's ability to complete manipulation tasks on highly variable soft materials like soft-tissue surgery.

All together, the work in this thesis has been presented in three papers:

1. *VAIRO* (**V**ision-based **A**daptive **I**mpedance-control **RO**botic framework), was accepted to the 2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).

2. *VAISI* (**V**ision-based **A**daptive **I**mpedance-control for **S**urgical **I**ncisions) was submitted and under review for the 2025 24th IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS).

3. See, Cut, Control: A Stereo Vision Scalpel for Surgical Robotics, was accepted to the 1st International Workshop on the Evolving Landscape of Surgical Robotics (ELSR 2025) at the 2025 42nd IEEE-RAS International Conference of Robotics and Automation (ICRA).

and was part of larger projects for two international competitions:

1. The Collaborative Deliciousness project for the KUKA Innovation Award 2024: Robots for the People

2. The Surgical Environment Robotic Assistant (SERA) project for the KUKA Innovation Award 2025: Medical Robotics Challenge 2.0

This thesis demonstrated the versatility of the vision-based adaptive impedance control framework for tasks where traditional modeling is either infeasible or impractical. The framework's modularity also allows it to be ported to other domains involving soft materials, including textile handling, deformable object assembly, or cleaning with compliant tools. However, challenges remain, particularly in improving the responsiveness and accuracy of force regulation. In both case studies, limitations such as sensing latency reduced

the system's ability to perfectly match intended stiffness profiles. Future work should investigate friction compensation methods, higher-bandwidth torque control loops, and more robust vision pipelines to mitigate these limitations.

Future work will explore the application of vision-based adaptive impedance control in other soft material manipulation tasks where precise interaction force regulation is critical. These applications span a wide range of domains, including but not limited to:

- Other medical robotics tasks on soft tissue, such as suturing. Visual measurements could once again use skin deformation to minimize the suturing force.
- Assistive care, where a CoBot could assist in bathing individuals. Vision-based measurements could use both skin deformation and cleaning tool deformation, such as with sponges, to modulate the force used to clean.
- Artisanal crafts, such as clay modeling, where vision could be used to modulate forces to shape clay into desired geometries matching a point cloud.

Comparisons on using various model-based and model-free approaches in tracking the soft material could also be done on the applications explored in this thesis and in these suggested applications.

In addition to visual feedback, future investigations could also explore the integration of multi-modal sensing into the adaptive impedance control framework. For instance, incorporating a force sensor or load cell at the end-effector would enable more precise force estimation, facilitating finer control over force modulation. Furthermore, environmental sensors, such as temperature and humidity sensors, could enhance performance in tasks where material properties are sensitive to external conditions. Overall, augmenting the vision-based framework with additional sensing modalities could enable more dynamic, accurate, and responsive modulation of stiffness, leading to improved precision in complex soft material interaction tasks.

References

[1] Fares J. Abu-Dakka and Matteo Saveriano. Variable impedance control and learning—a review. *Frontiers in Robotics and AI*, 7:590681, 2020.

[2] A. Albu-Schaffer and G. Hirzinger. Cartesian impedance control techniques for torque controlled light-weight robots. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, volume 1, pages 657–663 vol.1, 2002.

[3] Soofiyan Atar, Xiao Liang, Calvin Joyce, Florian Richter, Wood Ricardo, Charles Goldberg, Preetham Suresh, and Michael Yip. Humanoids in hospitals: A technical study of humanoid robot surrogates for dexterous medical interventions, 2025.

[4] J. Banks and P. Corke. Quantitative evaluation of matching methods and validity measures for stereo vision. *The International Journal of Robotics Research*, 20(7):512–532, 2001.

[5] Leandra A. Barnes, Clement D. Marshall, Tripp Leavitt, Michael S. Hu, Alessandra L. Moore, Jennifer G. Gonzalez, Michael T. Longaker, and Geoffrey C. Gurtner. Mechanical forces in cutaneous wound healing: Emerging therapies to minimize scar formation. *Advances in Wound Care*, 7(2):47–56, 2018.

[6] Charlotte E. Berry, Mauricio Downer Jr, Annah G. Morgan, Michelle Griffin, Norah E. Liang, Lionel Kameni, Jennifer B. Laufey Parker, Jason Guo, and Michael T. Longaker. The effects of mechanical force on fibroblast behavior in cutaneous injury. *Frontiers in Surgery*, 10:1167067, 2023.

[7] Gargi Bhaduri and Nancy Stanforth. To (or not to) label products as artisanal: effect of fashion involvement on customer perceived value. *Journal of Product & Brand Management*, 26(2):177–189, 2017.

[8] Miroslav Bogdanovic, Majid Khadiv, and Ludovic Righetti. Learning variable impedance control for contact sensitive tasks. *IEEE Robotics and Automation Letters*, 5(4):6129–6136, 2020.

[9] Rakesh P. Borase, D. K. Maghade, S. Y. Sondkar, and S. N. Pawar. A review of pid control, tuning methods and applications. *International Journal of Dynamics and Control*, 9:818–827, 2021.

[10] B.C. Brown, S.P. McKenna, K. Siddhi, D.A. McGrouther, and A. Bayat. The hidden cost of skin scars: quality of life after skin scarring. *Journal of Plastic, Reconstructive Aesthetic Surgery*, 61(9):1049–1058, 2008.

[11] Justin Carpentier, Nicolas Mansard, Florian Valenza, Joseph Mirabel, Guilhem Saurel, and Rohan Budhiraja. Pinocchio - efficient and versatile rigid body dynamics algorithms, jul. 2021.

[12] Justin Carpentier, Guilhem Saurel, Gabriele Buondonno, Joseph Mirabel, Florent Lamiraux, Olivier Stasse, and Nicolas Mansard. The pinocchio c++ library : A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives. In *2019 IEEE/SICE International Symposium on System Integration*, pages 614–619, 2019.

[13] Peter Corke. *Image Formation*, pages 207–251. Springer International Publishing, Cham, 2022.

[14] Peter Corke. *Using Multiple Images*, pages 253–326. Springer International Publishing, Cham, 2022.

[15] Intel Corporation. Intel realsense depth camera d405, 2024. Accessed: 2024-10-04.

[16] David T Corr and David A Hart. Biomechanics of scar tissue and uninjured skin. *Advances in wound care*, 2(2):37–43, 2013.

[17] Angela Dettori, Michela Floris, and Cinzia Dessì. Customer-perceived quality, innovation and tradition: some empirical evidence. *The TQM Journal*, 32(6):1467–1486, 2020.

[18] Innotech Digital. How to calculate the length of part rolls of material, 2023. Accessed: 2024-09-05.

[19] B. Dwyer, J. Nelson, T. Hansen, et al. Roboflow (version 1.0), 2024. Computer vision.

[20] Ron Eglash, Lionel Robert, Audrey Bennett, Kwame Porter Robinson, Michael Lachney, and William Babbitt. Automation for the artisanal economy: enhancing the economic and environmental sustainability of crafting professions with human–machine collaboration. *Ai & Society*, 35(3):595–609, 2020.

[21] Nele Famaey and Jos Vander Sloten. Soft tissue modelling for applications in virtual surgery and surgical robotics. *Computer methods in biomechanics and biomedical engineering*, 11(4):351–366, 2008.

[22] Mandela Ouafo Fonkoua, François Chaumette, and Alexandre Krupa. Deformation control of a 3d soft object using rgb-d visual servoing and fem-based dynamic model. *IEEE Robotics and Automation Letters*, 9(8):6943–6950, 2024.

[23] Yuan-cheng Fung. *Biomechanics: mechanical properties of living tissues*. Springer Science & Business Media, 2013.

[24] Ross Girshick. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[25] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[26] KUKA Deutschland GmbH. Kuka sunrise.os med 2.7 v1, 2023. Accessed: 2024-10-04.

[27] Ruchi Goel and Pooja Gupta. *Robotics and Industry 4.0*, pages 157–169. Springer International Publishing, Cham, 2020.

[28] Feida Gu, Yanmin Zhou, Zhipeng Wang, Shuo Jiang, and Bin He. A survey on robotic manipulation of deformable objects: Recent advances, open challenges and new frontiers, 2023.

[29] Dong Han, Beni Mulyana, Vladimir Stankovic, and Samuel Cheng. A survey on deep reinforcement learning algorithms for robotic manipulation. *Sensors*, 23(7), 2023.

[30] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[31] Penny L. Head. Rehabilitation considerations in regenerative medicine. *Physical Medicine and Rehabilitation Clinics of North America*, 27(4):1043–1054, 2016. Regenerative Medicine.

[32] F.M. Hendriks, D. Brokken, C.W.J. Oomens, D.L. Bader, and F.P.T. Baaijens. The relative contributions of different skin layers to the mechanical behavior of human skin in vivo using suction experiments. *Medical Engineering Physics*, 28(3):259–266, 2006.

[33] Andrea Herold. KUKA Innovation Award 2024, 2024.

[34] Andrea Herold. KUKA Innovation Award 2025, 2025.

[35] Neville Hogan. Impedance control: An approach to manipulation. In *1984 American Control Conference*, pages 304–313, 1984.

[36] Gerhard A Holzapfel et al. Biomechanics of soft tissue. *The handbook of materials behavior models*, 3(1):1049–1063, 2001.

[37] Zhe Hu, Tao Han, Peigen Sun, Jia Pan, and Dinesh Manocha. 3-d deformable object manipulation using deep neural networks. *IEEE Robotics and Automation Letters*, 4(4):4255–4261, 2019.

[38] Zhe Hu, Peigen Sun, and Jia Pan. Three-dimensional deformable object manipulation using fast online gaussian process regression. *IEEE Robotics and Automation Letters*, 3(2):979–986, 2018.

[39] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO, January 2023.

[40] John Keller and Sebastian Scherer. A stereo algorithm for thin obstacles and reflective objects, 2019.

[41] Ali Keshvarparast, Daria Battini, Olga Battaïa, and Amir Pirayesh. Collaborative robots in manufacturing and assembly systems: literature review and future research agenda. *Journal of Intelligent Manufacturing*, 35, 05 2023.

[42] Min Jun Kim, Fabian Beck, Christian Ott, and Alin Albu-Schäffer. Model-free friction observers for flexible joint robots with torque measurements. *IEEE Transactions on Robotics*, 35(6):1508–1515, 2019.

[43] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3992–4003, Oct 2023.

[44] Sina Nia Kosari, Srikrishnan Ramadurai, Howard Jay Chizeck, and Blake Hannaford. Robotic compression of soft tissue. In *IEEE International Conference on Robotics and Automation*, pages 4654–4659, 2012.

[45] Ilja L. Kruglikov and Philipp E. Scherer. Skin aging as a mechanical phenomenon: The main weak links. *Nutrition and Healthy Aging*, 4(4):291–307, 2018.

[46] Emprise Lab. Gen3 compliant controllers, 2024. Accessed: 2024-08-01.

[47] Robert Lee, Jad Abou-Chakra, Fangyi Zhang, and Peter Corke. Learning fabric manipulation in the real world with human videos. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3124–3130, 2024.

[48] Simon Leonard, Kyle L. Wu, Yonjae Kim, Axel Krieger, and Peter C.W. Kim. Smart tissue anastomosis robot (star): A vision-guided robotics system for laparoscopic suturing. *IEEE Transactions on Biomedical Engineering*, 61(4):1305–1317, 2014.

[49] Yang Li, Florian Richter, Jingpei Lu, Emily K. Funk, Ryan K. Orosco, Jianke Zhu, and Michael C. Yip. Super: A surgical perception framework for endoscopic tissue manipulation with surgical robotics. *IEEE Robotics and Automation Letters*, 5(2):2294–2301, 2020.

[50] Hongli Liang, Ting Wang, Jun Xia, M. Ali Nasseri, Haotian Lin, and Kai Huang. Autonomous clear corneal incision guided by force–vision fusion. *IEEE Transactions on Industrial Electronics*, 71(8):9319–9327, 2024.

[51] Xingyu Lin, Yufei Wang, Jake Olkin, and David Held. Softgym: Benchmarking deep reinforcement learning for deformable object manipulation. *CoRR*, abs/2011.07215, 2020.

[52] Matthias Mayr and Julian M. Salt-Ducaju. A c++ implementation of a cartesian impedance controller for robotic manipulators. *Journal of Open Source Software*, 9(93):5194, 2024.

[53] Hossein Mirsaeedghazi, Zahra Emam-Djomeh, Sayed Mohammad Ali Mousavi, et al. Rheometric measurement of dough rheological characteristics and factors affecting it. *International Journal of Agriculture and Biology*, 10(1):112–119, 2008.

[54] Manjula P. Mony, Kelly A. Harmon, Ryan Hess, Amir H. Dorafshar, and Sasha H. Shafikhani. An updated review of hypertrophic scarring. *Cells*, 12(5):678, 2023.

[55] Patrick Probst. A review of the role of robotics in surgery: To davinci and beyond! *Missouri medicine*, 120(5):389, 2023.

[56] Kavyanjali Reddy, Pankaj Gharde, Harshal Tayade, Mihir Patil, Lucky Srivani Reddy, Dheeraj Surya, and Lucky srivani Reddy. Advancements in robotic surgery: a comprehensive overview of current utilizations and upcoming frontiers. *Cureus*, 15(12), 2023.

[57] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.

[58] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[59] Stefano Renzetti, Roderik de Harder, and Albert Jurgens. Puff pastry with low saturated fat contents: The role of fat and dough physical interactions in the development of a layered structure. *Journal of Food Engineering*, 170:24–32, 2016.

[60] Kinova Robotics. Kinova-kortex2_gen3_g3l, 2024. https://github.com/Kinovarobotics/Kinova-kortex2_Gen3_G3L, Accessed: 2025-03-11.

[61] Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation*, Shanghai, China, May 9-13 2011.

[62] Célia Saghour, Mathieu Célérier, Philippe Fraisse, and Andrea Cherubini. Visual servoing for dual arm shaping of soft objects in 3d. In *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, pages 1–7, 2023.

[63] Gautam Salhotra, I-Chun Arthur Liu, Marcus Dominguez-Kuhne, and Gaurav S. Sukhatme. Learning deformable object manipulation from expert demonstrations. *IEEE Robotics and Automation Letters*, 7(4):8775–8782, 2022.

[64] Ranjan Sapkota, Yang Cao, Konstantinos I. Roumeliotis, and Manoj Karkee. Vision-language-action models: Concepts, progress, applications and challenges, 2025.

[65] Azad Shademan, Ryan S Decker, Justin D Opfermann, Simon Leonard, Axel Krieger, and Peter CW Kim. Supervised autonomous robotic soft tissue surgery. *Science translational medicine*, 8(337):337ra64–337ra64, 2016.

[66] F. Sherwani, Muhammad Mujtaba Asad, and B.S.K.K. Ibrahim. Collaborative robots and industrial revolution 4.0 (ir 4.0). In *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, pages 1–5, 2020.

[67] Haochen Shi, Huazhe Xu, Samuel Clarke, Yunzhu Li, and Jiajun Wu. Robocook: Long-horizon elasto-plastic object manipulation with diverse tools. In *Conference on Robot Learning*, pages 642–660. PMLR, 2023.

[68] Peixuan Shi, Bin Cheng, Ye Zong, Yi Gong, Hao Yan, Yida Hu, and *et al.* A soft tissue scalpel cutting robotic system with sucker fixation. In *IEEE 14th International Conference on Control and Automation*, pages 1162–1167, 2018.

[69] Christoph Silow, Emanuele Zannini, Claudia Axel, Kieran M. Lynch, and Elke K. Arendt. Effect of salt reduction on wheat-dough properties and quality characteristics of puff pastry with full and reduced fat content. *Food Research International*, 89:330–337, 2016.

[70] Hutchinson Seth Vidyasagar M. Spong, Mark W. 8.5 PID Control, 2020.

[71] Mark W. Spong, Seth Hutchinson, and M. Vidyasagar. Force Control, 2005.

[72] Mark W. Spong, Seth Hutchinson, and M. Vidyasagar. Rigid Motions and Homogeneous Transformations, 2005.

[73] Mark W. Spong, Seth Hutchinson, and M. Vidyasagar. 10.3.1 Impedance Control, 2020.

[74] Mark W. Spong, Seth Hutchinson, and M. Vidyasagar. 11.2.1 The Geometry of Image Formation, 2020.

[75] Richard Szeliski. Recognition, 2022.

[76] Khalid Hasan Tantawi, Alexandr Sokolov, and Omar Tantawi. Advances in industrial robotics: From industry 3.0 automation to industry 4.0 collaboration. In *2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, pages 1–4, 2019.

[77] Bao Thach, Brian Y. Cho, Alan Kuntz, and Tucker Hermans. Learning visual shape control of novel 3d deformable objects from partial-view point clouds. In *International Conference on Robotics and Automation*, pages 8274–8281, 2022.

[78] Xujing Tian, Mengyao Lv, Jiazheng Sun, Hongzheng Zhao, Ziyuan Jiang, Jinshuo Han, Wei Gu, and Gang Cheng. An adaptive impedance control method for polishing system of an optical mirror processing robot. *Robotica*, 42(1):21–39, 2024.

[79] Rejin Varghese and Sambath M. Yolov8: A novel object detection algorithm with enhanced performance and robustness. In *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, pages 1–6, 2024.

[80] T. Wada, S. Hirai, and S. Kawamura. Indirect simultaneous positioning operations of extensionally deformable objects. In *Proceedings. IEEE/RSJ International Conference on Intelligent Robots and Systems. Innovations in Theory, Practice and Applications (Cat. No.98CH36190)*, volume 2, pages 1333–1338 vol.2, 1998.

[81] T. Wada, S. Hirai, S. Kawamura, and N. Kamiji. Robust manipulation of deformable objects by a simple pid feedback. In *IEEE International Conference on Robotics and Automation*, volume 1, pages 85–90 vol.1, 2001.

[82] Liman Wang and Jihong Zhu. Deformable object manipulation in caregiving scenarios: A review. *Machines*, 11(11), 2023.

[83] Rosa Wolf, Yitian Shi, Sheng Liu, and Rania Rayyes. Diffusion models for robotic manipulation: A survey, 2025.

[84] Andrew S. Wright, Aaron Jensen, Sara Kim, and Karen Horvath. Instrument handling: Scalpels, 2024. Accessed: 2024-10-04.

[85] Hang Yin, Anastasia Varava, and Danica Kragic. Modeling, learning, perception, and control methods for deformable object manipulation. *Science Robotics*, 6(54):eabd8803, 2021.

[86] Hani Yousef, Mandy Alhajj, Adegbenro O. Fakoya, and Sandeep Sharma. *Anatomy, Skin (Integument), Epidermis*. StatPearls Publishing, Treasure Island (FL), 2025. Updated June 8, 2024.

[87] Jihong Zhu, Andrea Cherubini, Claire Dune, David Navarro-Alarcon, Farshid Alambeigi, Dmitry Berenson, Fanny Ficuciello, Kensuke Harada, Jens Kober, Xiang Li, Jia Pan, Wenzhen Yuan, and Michael Gienger. Challenges and outlook in robotic manipulation of deformable objects. *IEEE Robotics and Automation Magazine*, 29(3):67–77, 2022.