Report Title

- Subtitle -

Project Report

Group Name/Number

Aalborg University
Electronics and IT

Here you can write something about which tools and software you have used for typesetting the document, running simulations and creating figures. If you do not know what to write, either leave this page blank or have a look at the colophon in some of your books.

Electronics and IT
Aalborg University
http://www.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Project Title

Theme:
Collaborative Robotics

Project Period:
Spring Semester 2022

Project Group:
863

Participant(s):
Bjarke Gjerlev Stück
Marco António de O. Q. F. Alemão
Marek Rashka
Robin Leo Emil Cotman
Reshad Zadran

Supervisor(s):
Strahinja Dosen
Miguel Nobre Castro

Copies: 1

Page Numbers: 25

Date of Completion:
March 12, 2022

Abstract:

Here is the abstract

agreement with the author.

Contents

Todo list

Preface

Here is the preface. You should put your signatures at the end of the preface.

Aalborg University, March 12, 2022

Author 1
<username1@XX.aau.dk>

Author 2
<username2@XX.aau.dk>

Author 3
<username3@XX.aau.dk>

Chapter 1

Introduction

Social interaction between people is an important and common element of life. The importance of social interaction lies in the impact it has on the quality of life [1] and health of a person[2]. Among others, the life of amputees is impacted by social interaction. Amputees are able to perceive human interactions like a handshake, but are not able to interact due to their amputation.

Nowadays prosthesis are used for different application. They are commonly used to replace the missing limbs of an amputee. Michelangelo hand is an upper-limb transradial prosthesis capable of different grasp types. The social interaction problem persists with the prosthesis because it does not function as a complete replacement for the lost limb. The current market for prostheses is growing and products with various functionality exist. On the other hand, software for controlling a prostheses is lacking behind the hardware and a desire for greater control of prosthesis with software exists.

is anyone complaining about it? who? why? source...

The focus of this project is to develop a control solution for improving social interaction between an upper-limb transradial prothesis user and others. Where the focus lies on grasping an object with a prosthesis, from the hand of another person. One of the challenges occurring with grasping are occlusion, by the hand holding the object and another one is grasping strategy for different kind of objects, held in hand of a person.

The report is structured such that in chapter one the problem analysis defined...(continue)

Chapter 2

Problem Analysis

This chapter starts by formulating an initial problem in a concise manner, then it describes the problem in greater depth. Afterwards, it continues by reviewing existing solutions, to certain aspects of the problem. Next, possible methodology is explained. Finally, the chapter is summarized and a final problem formulation will be derived.

2.1 Initial Problem Formulation

How can current semi-autonomous trans-radial prosthesis solutions be built upon to improve social interactions by perceiving human intention?

2.2 Problem Description

According to a study performed in 2005, there exists over 1.2 million people in the US alone suffering from a loss of limb. With approximately an average of 185,000 amputations performed yearly. Among those amputee patients, 541,000 of them suffer from upper extremity amputations. Limb loss being a consequence of traumatic injuries, vascular diseases, malignancy and more. [6]
Not being able to perform simple tasks such as grasping and manipulating objects, as well as the limited social interactions, impair a patient not only physically but also psychologically. The loss of autonomy in a patients body can result in psychological conditions such as clinical depression. [4] [1]
Figure 2.1 showcases the different levels of upper limb amputations, depending on the placement on the arm. For simplicity sake, this project focus on prosthesis hands with wrist rotation, only transradial amputations will be taken into account.

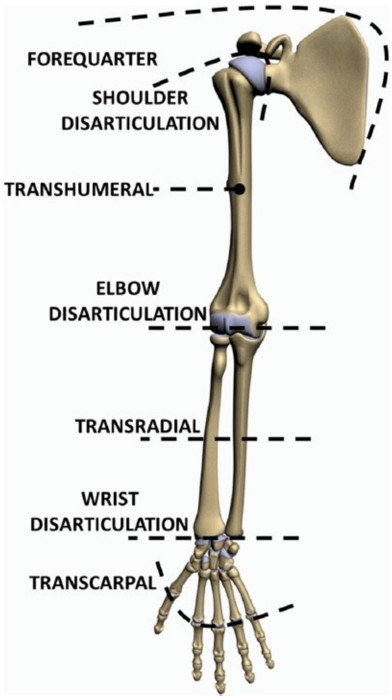

Figure 2.1: Different levels of upper limb amputations, with transradial being between the elbow and the wrist.[4]

The focus of this project lies on improving the social interaction capababilities of patients suffering from a transradial amputation, by developing better prosthetic solutions.

2.3 State of the art

This section will present three papers that proposes a solution for recognition of objects for grasping the object. Then three other papers that proposes different solutions for recognising hand gestures, will be presented. The methods used, the results and the drawbacks will be presented for each paper, to be used when creating the requirements for this project.

2.3.1 Deep learning-based artificial vision for grasp classification in myoelectric hands

Ghazal Ghazaei et al.[7] proposes to use a deep neural network (DNN) to augment the grasping functionality of a two-channel myoelectric prosthetic hand. The authors trained a convolutional neural network (CNN) with images of graspable objects. The CNN categorised the objects into four different grasp classes, pinch, tripod, palmar wrist neutral and palmar wrist pronated. Their approach was tested with two trans-radial prosthesis hands with a mounted webcam. After training, the subjects successfully picked up and moved the target objects, with an overall success rate of up to 88%.

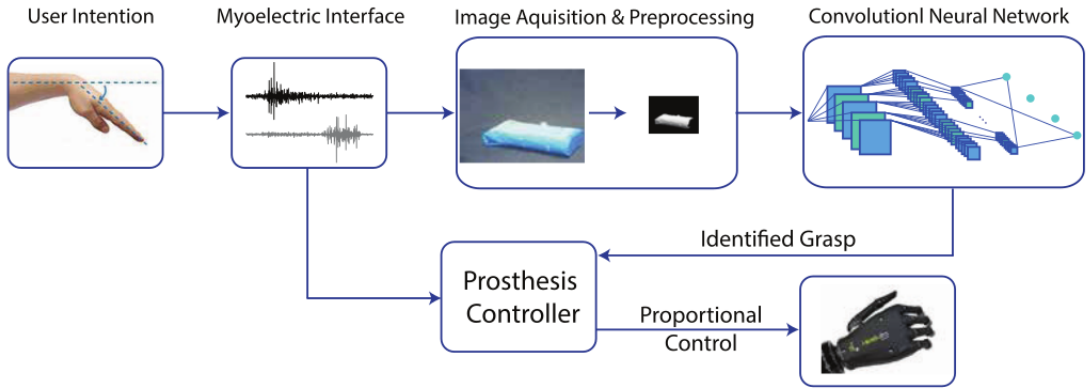

Figure 2.2: write text here.[7]

2.3.2 Cognitive vision system for control of dexterous prosthetic hands: experimental evaluation

This paper proposes a method of semi-autonomously selecting a grasp type for a dexterous prosthetic hand for transradial amputees.[15] The concept, named Cognitive Vision System (CVS), consists of an architecture which uses artificial vision to extract depth information from objects, and uses a decision tree to determine the appropriate grasp type, as well as the size of the grasp, for the selected object. The CVS can select between four different grasp types. Palmar, lateral, 3-digit and 2-digit pinch grasps, with palmar and lateral having three different variants depending on the object size. The decision tree consisting of IF-THEN rules, which contains thresholds regarding specific dimensions of the object. An illustration of said procedure can be found in Figure 2.3. The procedure requires a user to aim a laser onto the target object and capture an image using a myoelectric signal to activate the camera. Afterwards the appropriate grasp type is selected.

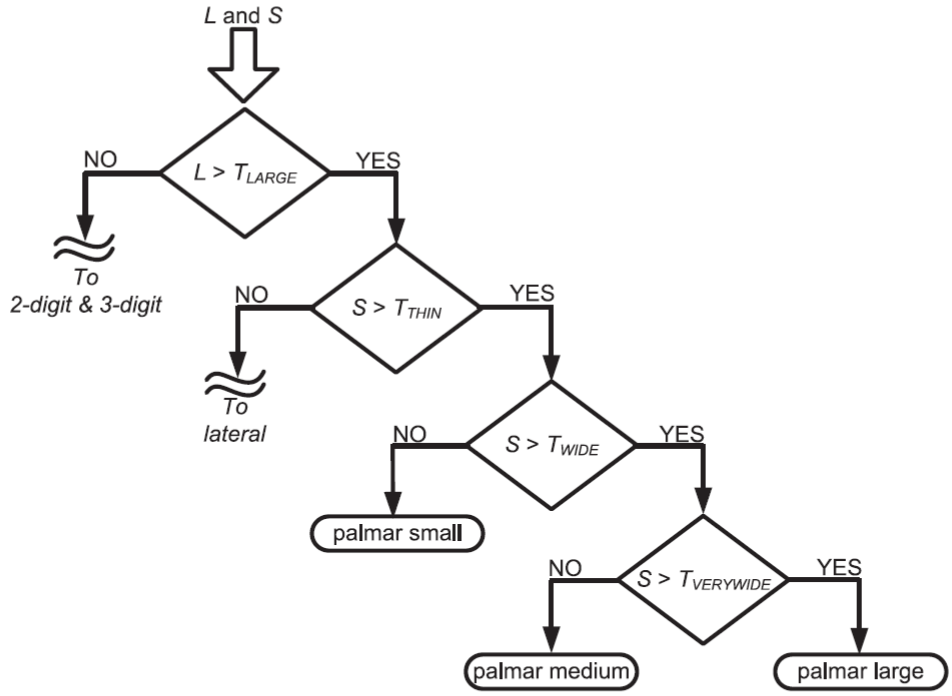

Figure 2.3: Illustration of the decision tree, asking IF-THEN questions when determining a grasp type. [15]

The CVS had a success rate on 84% in total of the 612 grasp trials performed between 13 able bodied subjects. Downsides to the CVS, were that the user would be required to highlight an object by aiming the laser which could become a problem in social gatherings, and when picking up translucent objects. Another downside is the limited amount of possible grasp types and grasp sizes. In total, the CVS provides a total of 9 different options, with 4 different grasp types, limiting the number of objects the user can get a solid grip when picking up the object.

2.3.3 Stereovision and augmented reality for closed-loop control of grasping in hand prostheses [9]

Grasp restoring prostheses are available, but their control is far from effective or effortless, which come from failure to integrate easy way to feedforward, in different words, command and feedback information such as sensation of any kind. This studies objective is to solve both problems with augmented reality glasses, in this paper argue could be both satisfied with one device, such as Google glasses connected to the semi-autonomous controller which would get feedforward from myoelectric sensors in the remainder of patients arm for triggering actions or making corrections for the controller.

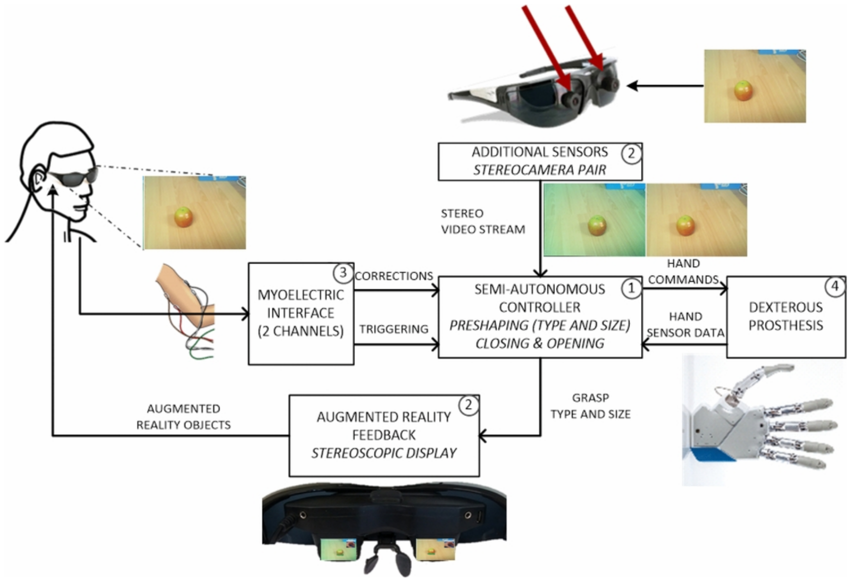

Figure 2.4: System architecture.[9]

Their proposed system architecture functions as follows. Semi-autonomous controller is focused on pre-shaping hand for grasping different sized and shaped objects, as well as closing and opening the hand itself, for that purpose it is connected to stereo camera that is sensing the depth of the object in question and sends it in AR feedback to person which has the myoelectric interface connected to his hand, in this studies case left hand, that triggers suggested actions or corrects them, such as different grip types and how much the hand actually needs to close before the object can be considered hand held by the prosthesis, which is the last part of the system, which does the commands of the controller after they were chosen by the person. This prosthesis also gives feedback in form of hand sensory data.

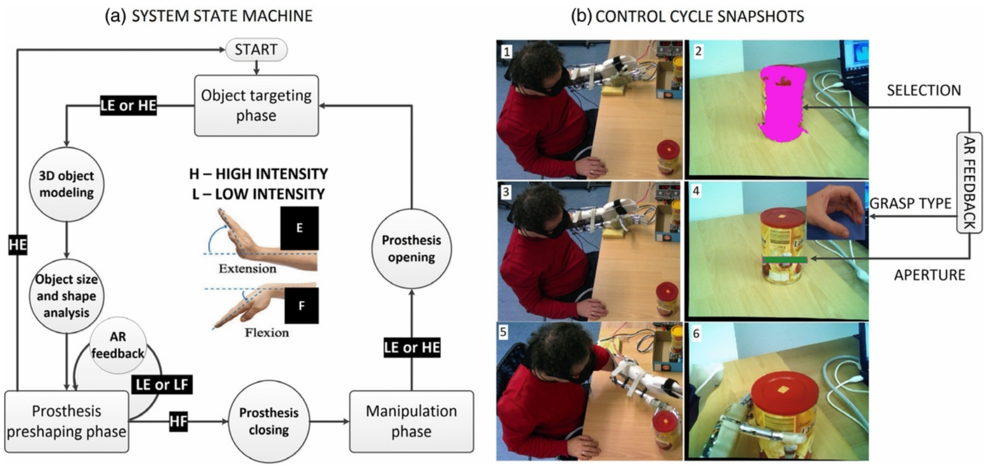

Figure 2.5: State machine design and implementation.[9]

Control loop is operated as a state machine with the state transition simply triggered by two-channel interface (flexion/extension with high/low intensity of action). It can be seen on right side of the picture in testing environment. This consists of object targeting, followed by pre-shaping with AR feedback on grasp type and width of object and final closing phase.

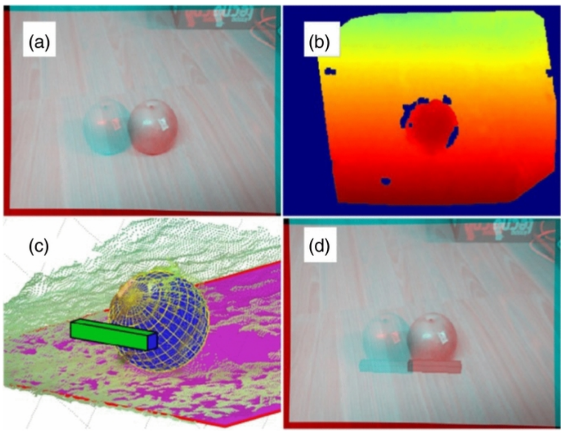

Figure 2.6: Computer vision processing pipeline.[9]

In this figure we can see how our system actually sees the objects before they are selected, in this case an apple. This image is then processed as depth image, then a geometrical model best representing the object is fitted through the cloud of points representing the object in AR feedback inserted back into 3D image that is displayed by the stereoscopic display in AR glasses.

This study achieved very high success rate with or without added errors for correction by the test subjects, which in this case is over 90% ,but it should be kept in mind that this control group was very homogeneous as everyone was between 25-33 years old, everyone was using their left hand only. The whole operation of commanding arm to accomplish task was on average 2,77 seconds without any human interference, which was a lot less successful in task completion as opposed to 3,47 with human interference with no added errors. With errors added, some for shape and some for width, it was between 5,2 and 5,9 seconds respectively. The stereo camera which was recognised as more powerful and robust tool compared to single camera, also had some problems such as pixel correspondence or segmenting the desired object without problems such as "spillover" from targeted object onto its neighboring objects/background. This system in the end should be taken as illustrative example how AR could help in prosthesis control.

2.3.4 Deep attention network for joint hand gesture localization and recognition using static RGB-D images

This paper proposes a way to recognise different hand gestures in RGB-D images, using a CNN.[17] The pipeline of the proposal can be seen in Figure 2.7.

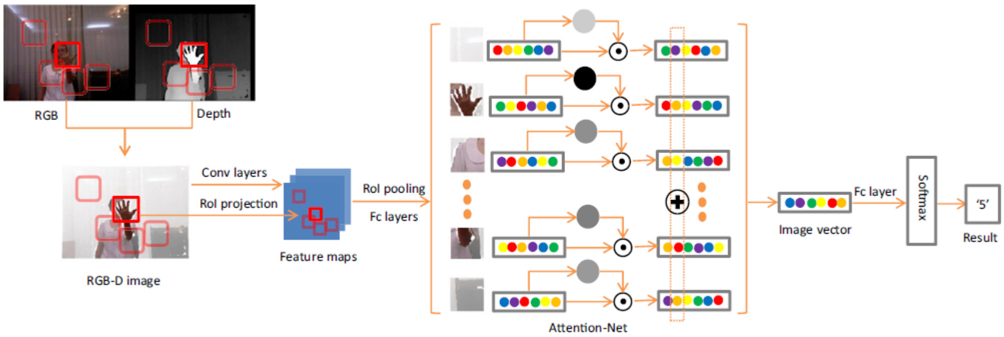

Figure 2.7: The pipeline for the proposed solution. [17]

By utilizing the depth information together with a sliding window, the area of focus is found. Afterwards the features are extracted using a Region of Interest (RoI) pooling layer. Next is the attention layer, which will determine the classification for the hand gesture, based on the feature weights that has been determined by the two datasets Nanyang Technological University Hand Digits (NTU-HD) containing 1000 images with 10 hand gestures and Huazhong University of Science and Technology (HUST) American Sign Language (HUST-ASL), containing 5440 images of 34 different hand gestures. Then, by finding the softmax loss of the feature vector, a class can be assigned to the hand gesture. The proposed method was compared to 6 other methods with the HUST-ASL dataset, and outperformed all six with an accuracy on 73.4% but was however the third fastest with a test time on an average of 0.726 seconds. By removing the segmentation element of the

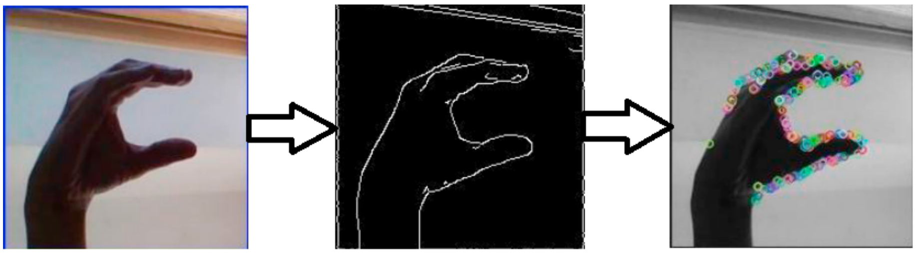

Figure 2.8: Image processing and feature extraction technique pipeline for feature detection. [14]

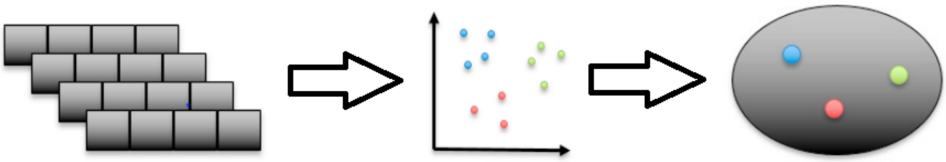

Figure 2.9: Image processing and feature extraction technique pipeline for classifiers. [14]

classification, the subjects hand can be in a more cluttered image, but will however be required to have background objects in order to differentiate the hand from the environment. Altough the proposed method surpased the other methods in accuracy, it had a longer test time than half of the other methods. This could be due to the sliding window segmentation process taking longer when compared to other solutions. Another downside is since the method relies on a CNN library, it has a limited amount of hand gesture classes.

2.3.5 Hand Gesture Recognition using Image Processing and Feature Extraction Techniques

This paper focuses on finding the best feature extraction technique for a gesture recognition model, featuring HOG, PCA, LBP and their own novel technique, using ORB features.

The image is pre-processed before ORB is used, in order to find the most important features. Which are then added to the visual vocabulary, which consists of K-mean clusters. Each cluster is used to classify the different gestures.
Similar techniques were used, such as Random Forest, Naive Bayes, Logistic Regression, K Nearest Neighbor, Support Vector Machine and Multi-Layer Perceptron. This techniques were run on a testing set, containing 17400 images. The proposed method outperforms all others in Naive Bayes, Logic Regression and KNN, but lacks behind PCA in the rest: MLP, Random Forest and SVM. The proposed method has the best average accuracy for recognition of gestures between

all used feature extractors, it was only tested on static images.
The workflow can be seen in Figures 2.8 and 2.9.

Clustered Dynamic Graph CNN for Biometric 3D Hand Shape Recognition

This paper presents a method for 3D hand shape recognition of point cloud data, obtained from RGB-D images, using geometric deep learning techniques. The architecture of the proposed method can be seen on Figure 2.10.

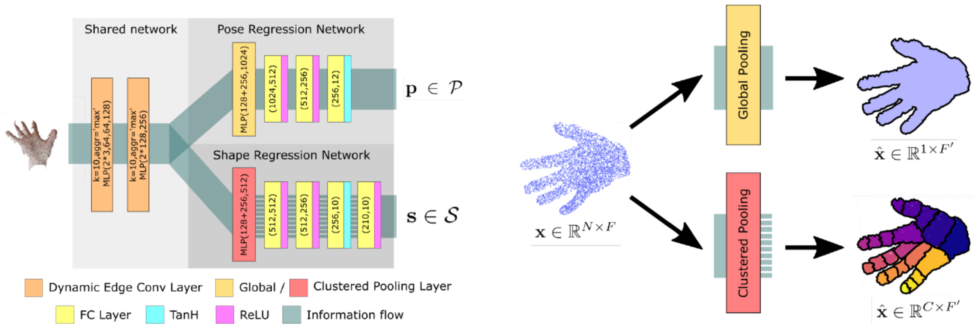

Figure 2.10: Image processing and feature extraction technique pipeline for classifiers. [8]

Based on Dynamic Graph CNN, by using a custom clustered pooling layer, instead of exclusively using a global pooling layer, it allows for a more piece-wise description of the hand, sorting the point cloud into segments. The algorithm was trained using a synthetic dataset of hand point clouds, obtained from recordings of subjects moving their hands into different hand shapes. The proposed methods showcased a good performance when compared to other methods such as PointNet++ and DGCNN when recognizing hand shapes.

Summary This section went over different proposals to solve similar problems in regards to recognising hand gestures and objects, along with displaying the methods used, and the advantages and disadvantages of said methods.

2.4 Methods

This section describes some of the algorithms that can be used to isolate objects in an image, and presents a way to spatially represent 3-dimensional objects. Subsequently, a method was presented which simplify and treats the 3-dimentional data. The section concludes with a summary.

2.4.1 Hand shape variations

The human hand possesses a large amount of Degrees of Freedom (DoF), which allows humans to grasp and manipulate a variety of objects with a stable grip. A study was conducted on this subject, which concluded that the human hand is capable of posing in 33 different grasp types in order to pick up objects. [16] An illustration of all 33 grasp types can be found in Figure 2.11.

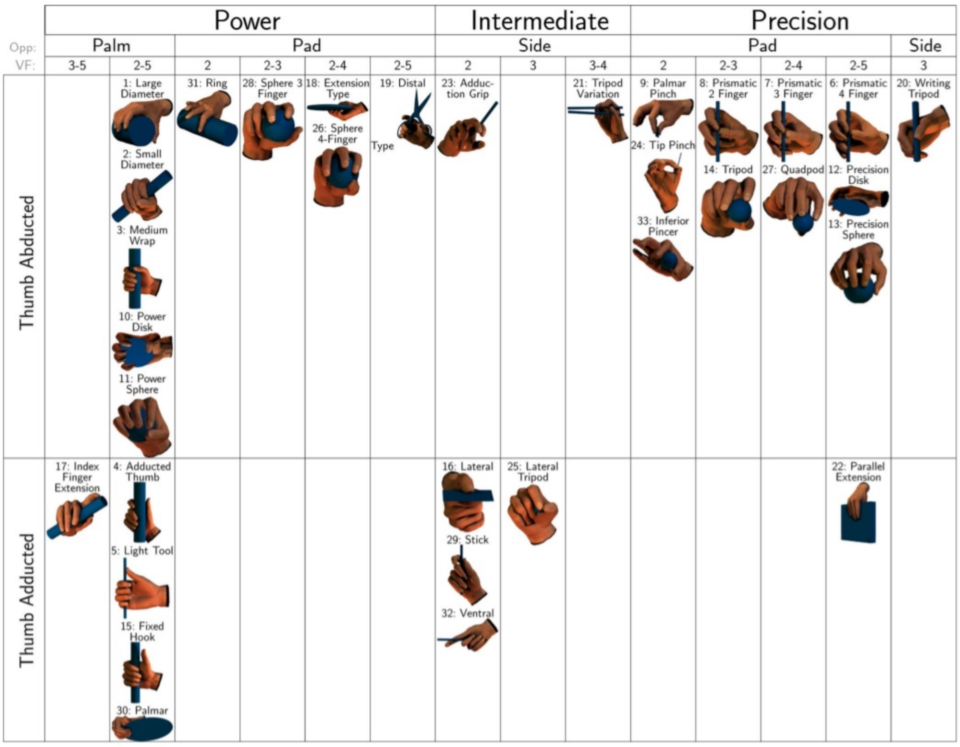

Figure 2.11: This figure showcases the 33 different grasp shapes for human hands.[16]

These 33 grasp types can be divided in five different main categories, depending on the dimensions and shape of the objects.[16] The first category being power palm, for objects with a large diameter or width. Furthermore power pad, which is similar to palm, but applied for smaller objects. Another category, intermediate side, which has the object held by one or more fingers, not including the thumb. precision pad which is when an object is held with the finger tips. With the last being precision side, which is when the person is holding a pen or something similar. For these five categories, sub categories are applied, being thumb abducted and thumb abducted, meaning whether or not the thumb applies force that oppose the other fingers.

Out of these 33 different grasp types, the eight most frequent grasps are the medium wrap, prismatic 2 finger, lateral tripod, lateral, tripod, power sphere, precision disk and palmar pinch. These grasp types can be seen illustrated in Figure 2.11.

Medium wrap, a power grip, which is meant for grasping cylinder shapes thin enough for the fingers to wrap around it. Almost all the fingers wrap around one side of the objects which is vertically placed, while the thumb applies force to the opposite side.

- Prismatic 2 finger, which is a precision grip, and has the hand hold a vertically long object with a small diameter or width/length, with two fingers, including the thumb.
- Lateral tripod, an intermediate grip for objects with very small dimensions. It holds the objects with two fingers, including the thumb.
- Lateral, is another intermediate grip for very thin objects. It involves the thin object resting horizontally folded together fist, on the index finger, while being hold in place by the thumb.
- Tripod, is a precision grasp, and is a way to grasp small spherical objects using the index finger, long finger and thumb.
- Power sphere, a power grasp which involves holding a small spherical objects similar to tripod, but instead using all the fingers, wrapped around the sphere.
- Palmar pinch, which is a precision grasp meant for very small objects that has the hand hold the object with the index finger and the thumb.
- Lastly, precision disk, is another precision grasp which, similar to palmar pinch, involves holding an object with the index finger and thumb. However, this grasp type is meant for disks consisting of a higher diameter, like a CD.

2.4.2 Regions of Interest

It is often the case that data has to be processed, before it can be used. By isolating Regions of Interest (RoI), one can remove part of the noise, or simply to regularize the data in a way that the next part of the process is simplified.
In an RGB image, a RoI can be represented by a mask, in other words, a binary image where each pixel has a Boolean value, corresponding to whether or not it belongs to the RoI. This concept is illustrated by Figure 2.12.

Figure 2.12: RGB image of a mug and mask of the same mug

The concept of RoI is not limited to 2 dimensional images. In fact, the binary condition previously described may be passed from pixels to points in a point clouds, there is a direct correlation between them. An example of this can be found, by using an RGB-D camera, as it may allow a mapping between the RGB data and the point cloud data.

Image Segmentation

If a person delivers an object to another person, the second is expected to grab the object with as little contact as possible. Ideally, the exchange is proceeded without any skin to skin contact between the two parties, according to the unspoken social norms.

source

With this in mind, one may draw a spacial representation of the object to be transferred, and discretize it in regards to everything else. By doing so, one may attempt to calculate a hand pose, capable of grabbing the object without touching the hand, or anything else.
This representation can be made by combining a point cloud with a mask resultant from a correlated RGB image.

Semantic Segmentation Semantic segmentation focuses on materials with undefined shapes, which possess characteristic patterns or homogeneous colours. Figure 2.13 highlights the background, using semantic segmentation.[3]
Semantic segmentation is often used to classify the background, but it can also be used to define objects that are closer to the camera, such as occluding objects. When an object is too close to the camera, and its shape can not be fully understood.

Figure 2.13: 2019 leader-board of COCO Stuff Segmentation Task. [3]

In 2019 COCO held the **Stuff Segmentation Task** competition, which distinguished the best algorithms for semantic segmentation, according to the metrics represented in Figure 2.15. This competition may be a good indicator of the state of the art algorithms in this category. The top 5 best algorithms are presented in Figure.[3]

	mIoU	fIoU	mAcc	pAcc	mIoU^S	fIoU^S	mAcc^S	pAcc^S	date
ResNeXt-FPN	0.287	0.560	0.423	0.694	0.563	0.690	0.706	0.807	2017-11-01
G-RMI	0.264	0.522	0.403	0.656	0.525	0.650	0.681	0.776	2017-11-01
Oxford Active Vision Lab	0.232	0.505	0.339	0.653	0.494	0.633	0.627	0.771	2017-11-01
Deeplab VGG-16	0.200	0.479	0.280	0.649	0.461	0.605	0.574	0.753	2017-11-01
Vllab	0.124	0.394	0.176	0.581	0.352	0.512	0.444	0.684	2017-11-01

Figure 2.14: 2019 leader-board of COCO Stuff Segmentation Task. [3]

```
Leaf category metrics:
  mIoU              % Mean IoU (primary challenge metric)
  fIoU              % Frequency Weighted IoU
  mAcc              % Mean Accuracy
  pAcc              % Pixel Accuracy
Supercategory metrics:
  mIoU^S            % Mean IoU
  fIoU^S            % Frequency Weighted IoU
  mAcc^S            % Mean Accuracy
  pAcc^S            % Pixel Accuracy
```

Figure 2.15: 2019 leader-board of COCO Stuff Segmentation Task. [3]

Panoptic Segmentation Panoptic segmentation combines semantic segmentation with instance segmentation. This includes extra information, but contains the type of object that a certain pixel corresponds to, such is the case in semantic segmentation. Therefore, this category is also relevant for this projects case of study.
Figure 2.16 shows an example of panoptic segmentation. There a mask for each instance of a certain type of object can be observed.

Figure 2.16: Panoptic segmentated images. [3]

In 2020, instead of the Stuff Segmentation Task competition, COCO held the **Panoptic Segmentation Task** competition. Figure 2.17 presents the top 3 algorithms that won the competition. From those,

	PQ	SQ	RQ	PQ^{Th}	SQ^{Th}	RQ^{Th}	PQ^{St}	SQ^{St}	RQ^{St}	date
Megvii [2019 winner]	0.546	0.836	0.643	0.642	0.860	0.744	0.402	0.800	0.491	2019-10-05
ELEME	0.519	0.820	0.622	0.618	0.840	0.732	0.379	0.790	0.457	2020-08-23
AIC-OpenVision	0.501	0.820	0.599	0.576	0.841	0.680	0.388	0.790	0.477	2020-08-23

Figure 2.17: 2020 leader-board of COCO Panoptic Segmentation Task. [3]

Average Panoptic Metrics:
PQ % Panoptic Quality **(primary challenge metric)**
SQ % Segmentation Quality component of PQ
RQ % Recognition Quality component of PQ
Panoptic Metrics for Things Categories:
PQ^{Th} % PQ for things categories only
SQ^{Th} % SQ for things categories only
RQ^{Th} % RQ for over things categories only
Panoptic Metrics for Stuff Categories:
PQ^{St} % PQ for stuff categories only
SQ^{St} % SQ for stuff categories only
RQ^{St} % RQ for stuff categories only

Figure 2.18: Panoptic Segmentation Task metrics. [3]

Point Cloud

When targeting an object, one may consider two questions. What constitutes a potential target is a question that depends on the specific case, but what information is necessary to interact with the targeted is something that one can start discussing, for instance, commenting on the use of point clouds and images.

When a potential target is selected, it is necessary to identify the shape, in order to plan a grasping strategy.
A shape can be represented by a set of vectors in a 3D space, this kind of structure is called **point cloud**. [12]
A teapot, for instance can be represented by a point cloud and its points can be rendered such as in Figure2.19.

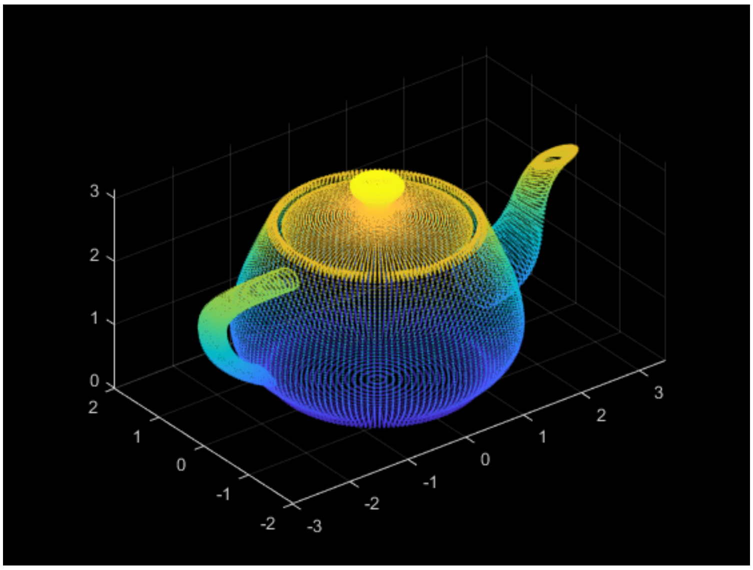

Figure 2.19: Visualization of point cloud. [12]

There are multiple libraries in Matlab designed for point cloud processing. Examples of this are *pcread* which allows point .ply and .pcd file reading. This are two type of files commonly used to store point clouds.

There are also python libraries dedicated to point cloud processing, for instance open3d [10], and pcl [11].

Summary It is possible to both isolate objects in an image and compute spatial representation of a 3D environment. When working with devices that correlate pixels and points of a point cloud, one can combine information from both types of data.

2.4.3 Object shape recognition

The way objects are grasped depends a great deal on how these objects are shaped. Most machine and human made objects can be simplified to represent geometric primitives such as spheres or cylinders or a combination of these primitives. Based on the fact that different objects can be represented by their geometric primitives it is possible to group different objects based on this, making the choice of selecting how to grasp an object more manageable. The following subsection will look into different methods that can be used to define the shape of an object.

RANdom SAmple Consensus(RANSAC)

RANSAC, can be used when processing images as a predictive modelling tool. This algorithm can effectively be used on point cloud data to find the a sought after model. RANSAC is effective at dealing with noise captured in the point cloud data, and will find a model which best fits the model. The way RANSAC works is by randomly selecting a certain amount of points in the data and a minimum amount of points needed to represent the model. After this, every point in the data set is tested against this random model to see which are inliers and which are outliers to the model. The previous step is then repeated for a set amount of time, where the model with the most inliers will be picked or until a goal amount of inliers has been reached. An example of RANSAC on a point cloud dataset can be seen in figure 2.20

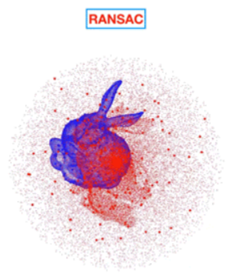

Figure 2.20: Example of RANSAC trying to fit a 3D model to a point cloud.[2]

Schnabel et al. [13], presents a method based on RANSAC to detect primitive geometric shapes in a cluttered point cloud.

N Adjacent Points SAmple(NAPSAC) Since the initial publication of RANSAC it has already been widely used in computer vision and has become a standard tool. Several RANSAC based algorithms have been developed to increase the robustness against outliers, the speed and the accuracy. NAPSAC is an algorithm introduced by Myatt et al.[5]. NAPSAC, in comparison to RANSAC which selects the initial points for the model randomly, chooses these points based on locality.

What they try to achieve with this is to reduce the number of iterations the algorithm has to run, which can prove advantageous in regards to speed. Myatt et al. show that, when selecting the data points for model instantiation using proximity, the probability of finding a set containing inliers with fewer iterations is increased. The concept comes from logical intuition that, the probability of initial points representing the surface of a model is larger, when the points are chosen based on their locality, rather than be picked completely at random. NAPSAC has been proven to be more robust, when it comes to higher dimensional data and data with more noise, in comparison to the original RANSAC algorithm.

2.5 Conclusion

This section presented an analysis of the problems that an upper arm amputees faces when interacting socially. Furthermore, various papers consisting of proposed solutions to a similar issues that amputees face.

2.5.1 Final Problem Formulation

Chapter 3

Conclusion

Bibliography

[1] Ravi Balasubramanian and Veronica J. Santos. *The Human Hand as an Inspiration for Robot Hand Development*. 1. ed. Springer International, 2014.

[2] Jennifer Chu. *Algorithm quickly finds hidden objects in dense point clouds*. https://techxplore.com/news/2019-06-algorithm-quickly-hidden-dense-clouds.html. 2019.

[3] *COCO-Common Objects in Context*. https://cocodataset.org.

[4] Francesca Cordella et al. *Literature Review on Needs of Upper Limb Prosthesis Users*. 2016.

[5] D.R. Myatt P.H.S. Torr S.J. Nasuto J.M. Bishop R. Craddock. *NAPSAC: HIGH NOISE, HIGH DIMENSIONAL ROBUST ESTIMATION - IT'S IN THE BAG*. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=ED412F3CFDE5CEFE493CD063DDE5C5 doi=10.1.1.62.7082&rep=rep1&type=pdf.

[6] Beth D. Darnall et al. *Depressive Symptoms and Mental Health Service Utilization Among Persons With Limb Loss: Results of a National Survey*. https://www.archives-pmr.org/action/showPdf?pii=S0003-9993%2804%2901316-4. 2005.

[7] Ghazal Ghazaei et al. "Deep learning-based artificial vision for grasp classification in myoelectric hands". In: *Journal of neural engineering* 14.3 (2017), p. 036025.

[8] Davide Boscaini Jonathan Masci Michael Bronstein Jan Svoboda Pietro Astolfi. *Clustered Dynamic Graph CNN for Biometric 3D Hand Shape Recognition*. 2020.

[9] Marko Markovic et al. "Stereovision and augmented reality for closed-loop control of grasping in hand prostheses". In: *Journal of Neural Engineering* 11.4 (2014), p. 046001. DOI: 10.1088/1741-2560/11/4/046001. URL: https://doi.org/10.1088/1741-2560/11/4/046001.

[10] *Open3D*. http://www.open3d.org/.

[11] *pcl*. https://github.com/strawlab/python-pcl/.

[12] *Point Cloud Processing*. https://se.mathworks.com/help/vision/point-cloud-processing.html.

[13] Roland Wahl Reinhard Klein Ruwen Schnabel Raoul Wessel. *Shape Recognition in 3D Point-Clouds*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.477.8247&rep=rep1&type=pdf.

[14] Ashish Sharma et al. *Hand gesture recognition using image processing and feature extraction techniques*. 2020. URL: https://www.sciencedirect.com/science/article/pii/S187705092031526X.

[15] Miloš Kostić Marco Controzzi Maria C Carrozza Dejan B Popović Strahinja Došen Christian Cipriani. *Cognitive vision system for control of dexterous prosthetic hands: Experimental evaluation*. 2010.

[16] Heinz-Bodo Schmiedmayer Aaron M. Dollar Thomas Feix Javier Romero and Danica Kragic. *The GRASP Taxonomy of Human Grasp Types*. IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, VOL. 46, NO. 1. February 2016.

[17] Wenyu Liu Bin Feng Yuan Li Xinggang Wang. *Deep attention network for joint hand gesture localization and recognition using static RGB-D images*. Elsevier. 2018.

Appendix A

Appendix A name

Here is the first appendix