UNIVERSITY OF BIELEFELD

BACHELOR THESIS

Efficient Target Identification during Haptic Search in a Three-dimensional Environment

Author:
Julian Nowainski

Supervisor:

A thesis submitted in fulfillment of the requirements
for the degree of Bachelor of Science

in the

Neuroinformatics Group
CITEC

September 17, 2017

Declaration of Authorship

I, Julian Nowainski, declare that this thesis titled, "Efficient Target Identification during Haptic Search in a Three-dimensional Environment" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

University of Bielefeld

Abstract

Faculty of Technology
CITEC

Bachelor of Science

Efficient Target Identification during Haptic Search in a Three-dimensional Environment

by Julian Nowainski

In this work the goal was to analyze the human way of efficient target identification during a haptic search in a three-dimensional environment. Therefore an experiment was proposed in which blindfolded participants were asked to localize a target object on a modular haptic stimulus board (MHSB) among different items. Both the target and distractor objects were wooden bricks of five different tactile shapes whereof multiple objects of each shape were embedded in a configurable wooden board. The participants had to perform this task in different scenarios each with its own distribution of objects and a different target to search for. During this experiment multimodal data was recorded with a glove capable of capturing both haptic data and joint angles between the fingers.
By performing multiple classification tasks it should be investigated how the data of same stimuli correlates to their role as a target or distractor object in the scenarios. It was found that from the data models could be build to distinguish between these roles and it was also possible to show that these models performed better when they were trained only on the data from targets. This means that the efficiency of searching a target is based on a set of salient features that is sufficient enough to differentiate between the target or a distractor role, but not necessarily between the objects itself.

Contents

Declaration of Authorship iii

Abstract v

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Goals . . . 1

2 Haptic Search Experiment 3
- 2.1 Haptic Search Experiment . . . 3
  - 2.1.1 Experimental Setup . . . 3
  - 2.1.2 Execution . . . 3
- 2.2 Hardware . . . 4
  - 2.2.1 Glove . . . 4
  - 2.2.2 Vicon . . . 4
  - 2.2.3 Setting . . . 5

3 Data Generation and Analysis 7
- 3.1 Data Structure and Requirements . . . 7
- 3.2 Recording . . . 7
- 3.3 Postprocessing Vicon Data . . . 8
- 3.4 Generating Labels . . . 8
  - 3.4.1 Synchronizing Glove and Vicon Data . . . 8
  - 3.4.2 Representing Glove and Objects . . . 9
  - 3.4.3 Finding Labels . . . 10
- 3.5 Analyzing the Data . . . 11

4 Model and Training 13
- 4.1 Model . . . 13
- 4.2 Preprocessing . . . 13
- 4.3 Training . . . 13

5 Evaluation 15

6 Discussion 17

Chapter 1

Introduction

Hier eine Introduction mit der zugrunde liegenden Motivation und den daraus abgeleiteten Zielen für meine Bachelorarbeit

1.1 Motivation

1.2 Goals

Chapter 2

Haptic Search Experiment

2.1 Haptic Search Experiment

!!!!!!!!!Include small introduction to haptic search here!!!!!!!!

2.1.1 Experimental Setup

The Modular Haptic Stimuli Board (MHSB) makes up the core part of the experiment. It is a setting with two wooden frames that hold stimuli objects. These objects are 3 x 3 cm big wooden blocks, which have a primitive three-dimensional shape on top of it or are just plane. The whole set consists of 360 blocks with 55 different shapes.
The first wooden frame can fit 25 objects and is used for learning a target object whereas the second frame has a capacity of 100 objects and is used for searching target objects. The stimuli are statically installed in the frames and not manipulable to allow a focus on just the search task itself [See figure of setting with both frames].

For this experiment, a subset of stimuli was chosen, consisting of 5 different shapes and plane ones [See fig. XXX -> picture of the shapes]. The target consists of one object and is placed central in the small frame with the rest of the space consisting of plane stimuli. The big frame contains the rest of this subset, where each shape exist 4 to 5 times, including the target. The objects were distributed mostly equally and kept the same rotation throughout the experiment. Only the distribution and the target were changed with each trial.

2.1.2 Execution

For this experiment, 7 participants were invited and asked to solve a haptic search task while being blindfolded. The participants were 23 to 28 years old and included both genders. All participants were right-handed and have never seen the stimuli objects, so that during the task they never knew how the set of objects looked like and their perception was purely based on the haptic features.
Each participant performed on maximal 5 trials, where after each trial, the target was exchanged and the distribution of the stimuli on the big frame was changed. Before the beginning, there were 2 rehearsals to accustom the subjects to the setting. No participant had the same target twice or more and the task was done with just the right hand, while wearing a glove to record relevant data [see section Hardware].

For the procedure, each participant was given a description of the task [see appendix]. The task consisted of two parts.
The first task was to explore the target object on the small frame and remembering it

just by its haptic features. When collected enough information about the target stimulus, the subject should proceed to the big frame and search for the learned target. The only goal in this part was to remember the approximate position of the target and not saying that it was found or pointing at it, so the recorded data would not contain pauses or pointing postures.
It was not necessary to find every target shape in the big frame, just as many as one could. The time was limited to 30 seconds to guarantee that the focus lies only on the salient features. An acoustic signal by the examiner determined the start- and endpoint of the experiment.
The second part of the experiment was to figure out if the subjects found the target object between the non-target objects, called distractors, and how well they could remember the approximate position on the frame. Again an acoustic signal determined start and end of the trial. For the second part, the subjects had just 10 seconds left to find the targets and point on them. The short period of time was set to prevent the subjects from exploring too much of the frame and focusing only on the smallest set of haptic features that were sufficient enough to differentiate between target and distractor.

2.2 Hardware

In this section the used hardware will be described as well as the overall setting that was used to record the data for the experiment.
There will be first a brief description of the glove that was used to capture tactile relevant and hand posture data followed by a description of the Vicon system to capture position data in a three-dimensional environment. At the end the implementation of the hardware into the experimental setting is explained.

2.2.1 Glove

A detailed explanation of the underlying technical properties and its implementation into the glove can be found in the work of Bianchi [Refernce to the journal paper].

To record data for this experiment, a device was needed that would be able to capture the most relevant patterns underlying a haptic search task. These so called exploratory features (EPs) describe the behavior of the hand during the exploration [refernce to EP paper]. Furthermore a device for recording the tactile properties was needed.
The multi-modal sensing glove combines both of these features. On the bottom side of the glove 64 tactile cells are mounted, covering hand palm and fingers. These fabric-based sensors record local pressures with a frequency of 150 Hz. The top side consists of 18 bending sensors, used to capture the joint angles representing the hand pose with a frequency of 50 Hz [See fig. of glove].

2.2.2 Vicon

For capturing the position of the hand and the MHSB, the Vicon system was used. It records motion data with a frequency of 200 Hz, using retroreflective markers that are tracked by infrared cameras [reference to vicon].
Also included is a Basler camera, generating a top-down view for the experiment.

2.2.3 Setting

To record motion data from the subjects hand, 17 reflective markers has been placed on an extra glove that the participant wears atop of the multi-modal one. The markers were placed in a position as seen in figure XYZ to guarantee a good reconstruction of the finger and hand movements.
The most time-consuming part was to find a setting of the Vicon cameras that would capture the reflective markers continuously, making sure to minimize the occurrence of gaps. The result is seen in figure xyz. There were 14 Vicon cameras placed in a semicircle around the MHSB. The Basler camera was placed directly above the frames. As an addition, there were 2 cameras placed on the left and right side to record also the side-view of the experiment.
The glove was connected via USB and serial-port to a nearby computer. A second computer controlled the Vicon system. To simultanetly start the recording, a synchronising tool called MSS was used [see chapter 3.2 recording].

Chapter 3

Data Generation and Analysis

This chapter addresses the methods and efforts to tackle the task of data generation as well as labeling the huge amount of data that was recorded.
It was the most time-consuming part of this work, since it involved a lot of post-processing and data cleansing work that was necessary due to the multi-modality of the recording devices and capturing data with different frequencies with various data formats which also were partly unsynchronized. Furthermore methods are explained that were used to label the generated data mostly automatically, based on position data of the hand and the MHSB, as well as a representation of the distribution of the stimuli objects on the frames.

3.1 Data Structure and Requirements

The data in this experiment was recorded with multiple devices, including the Vicon system, the 2 parts of the glove as well as 3 cameras, generating side- and top-views. To be able to train a classifier with supervised learning, there were a number of requirements to the data:

1. Simultaneous data acquisition
    - Capturing all devices at the same time will facilitate upcoming processing steps
2. Postprocessing raw data
    - To be able to work with the data, raw data needs to be processed and all files need to be in the same format
3. Synchronizing the time-series
    - Delays in the data acquisition and different device frequencies make this step necessary
4. Generating the labels
    - For supervised learning, the whole dataset needs to be labeled

3.2 Recording

To record the data of all devices preferably at the same time and with giving just one start signal, a tool called Multiple Start Synchronizer (MSS) was used. MSS sends a trigger signal to all registered devices which makes them start and stop capturing data.

The Vicon and Basler camera data were captured directly within the Vicon Nexus program. For the glove, data was recorded as rosbag consisting of two topics for each part of the glove. Side-view camera images were captured directly as image files.

Despite using MSS, there were still delays among the different devices that had to be synchronized separately.

3.3 Postprocessing Vicon Data

The first step in the pipeline was to postprocess the Vicon data. In this procedure, a three-dimensional hand model with marker positions was fitted to an image of the subjects hand, see figure xyz(fitting and reconstruct). This model was then used to reconstruct the hand movement during the experiment to approximate marker positions that occurred during gaps in the recording when no camera captured a marker see right side of figure xyz.

The resulting file contains a time-series of the x-,y- and z-position of each marker. Furthermore a file with the joint-angles was generated.

3.4 Generating Labels

This section will describe the methodology that was used to generate labels for the recorded data. The challenge was to write a program, that will do most of the work automatically and handle the huge amount of data generated by this experiment. With 7 subjects participating in up to 5 trials each, and a time series containing between 5000 and 7000 data points for each trial, a manual labeling of the data would be too time-consuming. Also having to cope with unsynchronized data due to delays between the modalities would make this task hard to tackle without proper preprocessing. The solution was a program that used the trajectories of the Vicon data to extract objects that were explored during the search experiment and to label them appropriately. The exact procedure is described in the subsections below. It can be summarized to three mandatory steps:

1. Synchronizing data from Vicon and glove to allocate positions to tactile data
2. Building a representation of the experimental setting to recreate it in sense of the hand trajectories and object distribution
3. Generating labels by replaying these trajectories and constructing a vector containing the explored objects at each timestep

3.4.1 Synchronizing Glove and Vicon Data

A problem that occurred during the acquisition was the delay between starting the Vicon system and the glove recording. Although sending a trigger signal to both systems at the same time, the glove started capturing data approximately 3 to 5 seconds later. Additionally the beginning of the Vicon data had to be cut by 100 to 1000 frames for postprocessing reasons. Fitting the three-dimensional model was only successful if the markers of the first frames had a nearly perfect plane position. As a consequence, an offset had to be defined pointing to the beginning of the Vicon

time series because the data only contains a timestamp describing the beginning of the recording. On the other hand the recorded rosbags from the glove came with a timestamp for each sample.
Since the frequency of the tactile glove with 150 Hz is lower than the frequency of Vicon with 200 Hz, the trajectory data should be reduced to the length of the tactile glove.

Consider we have two time series $V = \{v_t \mid t \in T_V\}$ and $G = \{g_t \mid t \in T_G\}$ describing the set for the Vicon data and glove data. The set of timestamps T_G was given for the tactile data and consisting of unix time values. For T_V the timestamps had to be calculated for each sample from the initial timestamp, the offset and the frequency.
To synchronize, a new time series $V' \subset V$ was defined with

$$V' = \{v_t \mid \forall g_{t_g} \in G \exists v_{t_v} \in V : t_v \geq t_g \wedge t_v < t_{g+1}, t_v \in T_V, t_g \in T_G\}$$

This new time series has now equal length to G and each time value from V' matches exactly one time value from the time series G.

3.4.2 Representing Glove and Objects

The core idea behind this program was to use the hands trajectories and approximated object positions to detect which object was covered by the hand during which time. Having the trajectory data given, the only thing that had to be done manually was the object distribution. For this, a representation of the board was generated in form of a matrix $B \in \mathbb{R}^{10\times10}$ for each trial. In this representation, $b_{11} \in \Omega = \{0, .., 5\}$ would be the top left object and from there rows and columns were generated accordingly where Ω is the set of labels. To decrease the number of false-positives, only explored objects were considered in this representation.(see fig. with rep next to picture of MHSB)
The next level was to represent this information in a coordinate system by generating polygons for each object in B to embody the MHSB. Since only the top corners of the boards were assembled with markers, the positions for respective corners of the polygons had to be calculated based on this. First for each object $b_{ij} \in B$ a polygon $P_{ij} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ was created where each element describes the x- and y-position of a corner. In the second step, this polygon was represented as its center position $z = \begin{pmatrix} z_x \\ z_y \end{pmatrix}$. The result is a matrix $B' = \begin{pmatrix} z_{11} & \cdots & z_{1n} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nn} \end{pmatrix}$ that represents every object in B through a position.

The remaining problem was that we used the top left position of the bigger frame as a base to build our polygons using a step size corresponding to the edge length of the stimuli. As a result, the represented board was placed parallel to the x- and y-axis into the coordinate system. Because this did not match the real setting, the matrix B' had to be rotated. For this the actual angle α between the vector $\vec{t} = \mu_{tr} - \mu_{tl}$, where μ_{tl} is the mean position of the top left corner and μ_{tr} the mean position of the top right corner, and the x-axis had to be calculated to build a rotation matrix R_α. The angle α describes the angle with whom we need to rotate the representation so that the orientation of B' matches the one in the real setting. The final matrix for the stimuli representation is $B'' = \begin{pmatrix} R_\alpha z_{11} & \dots & R_\alpha z_{1n} \\ \vdots & \ddots & \vdots \\ R_\alpha z_{n1} & \dots & R_\alpha z_{nn} \end{pmatrix}$.

With B'' a matrix was now given that could be used for assigning objects, or in this case their labels, to hand positions. But in order to do this, a representation for the hand had to be thought of. For now, the hand consisted of 17 trajectories t_i, one for each marker i.
A first approach was to use a convex hull H_c of the hand as seen in figure (figure with shapes representing hand) and check for each time step in V' if H_c contained points $p \in B''$. This idea was discarded quickly because the seize of the convex hull was to large, resulting in multiple possible objects p for each time step. Furthermore it was computationally costly since for every step the whole matrix B'' had to be checked for contained points. This led to an extension of B'' to save its elements in a k-d-tree.
The second approach simplified the representation by only using finger markers. Instead of the convex hull for the whole hand, the trajectories t_i were averaged for each finger, resulting in only 5 positions and a representation H'_c. Also there were no checks for points $p \in B''$ that were contained in the convex hull anymore, but rather finding the point p_i with minimum distance to the center of H'_c. These distances could be looked up now efficiently in the k-d-tree and there were no multiple possible objects for each time step but only one. This approach improved the performance greatly so that only a few points were still labeled falsely due to the size of H'_c.
The last approach was fine tuned by simplifying even more. Now only three fingers were used, the index-, middle- and ring finger. Observations showed, that the little finger wasn't used often during exploration. Moreover the thumb played a redundant role, since it never touched a stimuli alone but instead together with index-, or middle finger to apply pressure. Removing the trajectories from these fingers resulted in a pyramid-like polygon(see fig) that was precise enough to exclude false labels almost completely.

3.4.3 Finding Labels

Having now a representation of the glove and the objects in the MHSB, the remaining task was to bring it all together to find the labels for tactile data. In this subsection, the algorithm is explained that was written to label data almost automatically as well as further cleaning steps that were mandatory.

The algorithm(1) requires synchronized data, the representations of glove, objects and the target label. These were also part of the program, but were treated separately in the previous subsection since the focus here is the procedure on how to find labels.
The algorithm starts by iterating over all time steps in V'. For each iteration, first the hand representation is calculated by the current positions as explained previously. The center position of this polygon is then passed on to query the nearest object in the k-d-tree. Furthermore a mean position μ_z is calculated for the hands z-position. This will serve as validation condition to see if $\mu_z <= \delta$ with δ describing a threshold for the minimum height of the hand. It is approximately a bit above the

boards height in three-dimensional space, since the representations are just in two-dimensional space (see fig of viz) and there is no information about the z-axis given. Additionally it validates whether the polygon center is inside one of the boards. If both conditions are true, the label is assigned for this time step.
For debugging purpose, the program also includes a simple visualization tool to follow the process that shows the representations of the objects and for each iteration the representation of the glove as well as the assigned label(see fig viz).

Algorithm 1 Finding and assigning labels to a time series

```
Require: time series V' containing marker positions, k-d-tree T
begin
    Initialize l = {} and threshold δ
    for every time step t in V' do
        p ← generatePolygon(V'(t))
        label ← T.query(p.center)
        mean_z ← getMeanZ(V'(t))

        if mean_z ≤ δ and IsInside(p.center) then
            append(l, label)
        else
            append(l, 0)  //means no relevant object explored or hand outside
        end if
    end for
    return l
end
```

A small addition was made after the first few observations. As a result of using only the center of the polygon, small noise in the position led to false labels when the center appeared to be closer to a neighbor object. To fix this problem, and additional parameter γ was added describing an attraction variable. If the same label occurs consecutively, meaning an object is explored for some duration, γ increases. When then the label changes, but the old one is still near, the algorithm will stick to the previous one while decreasing γ. This ensures to avoid gaps of false labels.

After generating the label vectors for the trials, almost no additional work had to be done manually. However if the data was too noisy, few gaps had to filled by hand with the help of the visualization tool or good guessing.

3.5 Analyzing the Data

Chapter 4

Model and Training

Hier muss ich mir noch Gedanken über das Model machen, auch was das Preprocessing angeht. In diesem Kapitel wird wahrscheinlich noch einiges umgebaut

4.1 Model

4.2 Preprocessing

4.3 Training

Chapter 5

Evaluation

Evaluation der Ergebnisse vom Training mit Visualisiserung

Chapter 6

Discussion

Weiterführende Diskusion und Fazit über die Studie