Institute for Visualization and Interactive Systems

University of Stuttgart
Pfaffenwaldring 5a
D–70569 Stuttgart

Bachelor Thesis Nr. 131

# Augmented Reality supported Order Picking using Projected User Interfaces

Johannes Herter

**Course of Study:** Softwaretechnik

**Examiner:** Prof. Dr. Albrecht Schmidt

**Supervisor:** Dipl.-Inf. Markus Funk,
Dipl.-Inf. Lars Lischke,
Dipl.-Inf. Sven Mayer

**Commenced:** April 24, 2014

**Completed:** October 24, 2014

**CR-Classification:** H.5.2

## Acknowledgement

At this point I would like to thank all people involved in this thesis. Thank you for helping and supporting me and giving me the chance to produce this thesis.

Special thanks go to my main supervisor Markus Funk for the interesting and exciting topic as well as the help provided during the development, the study and the printing of the thesis.

I also like to thank my other supervisors Lars Lischke and Sven Mayer for giving me hints when I was stuck and also helping me out whenever I had any questions.

Additionally I want to thank the motionEAP project for the chance to contribute in such an interesting field.

Furthermore I want to acknowledge the work of Schnaithmann Maschinenbau GmbH, especially Bernhard Bader, for building the order picking cart.

I would also like to thank Sebastian Pickl for contributing to OrderPickAR. The help in the development of the pick detection as well as the position streaming is highly appreciated.

My thanks also go to the participants of the study for sacrificing their time and providing me with additional input on what to improve on OrderPickAR.

Last but not least I want to thank all proof readers of my thesis, for correcting errors and giving me hints on how to improve the read quality.

# Abstract

Order Picking is one of the most important tasks in modern warehouses. Since most work is still done manually, new methods to improve efficiency of the task are being researched. While the currently most used approaches Pick-by-Paper and Pick-by-Light are either prone to error or only scalable with high costs, other methods are considered. These methods include Pick-by-Vision systems based on Augmented Reality although these systems mostly rely on head-mounted displays. In order to evaluate a new method, we developed OrderPickAR which uses an order picking cart as well as projected user interfaces. OrderPickAR is part of the motionEAP project of the University of Stuttgart and relies on in-situ projection as well as motion recognition to guide the user and present feedback. The intuitive feedback provided by the in-situ projection as well as the motion recognition gives OrderPickAR the chance to effectivly eliminate errors while lowering the task completion time. With the use of a mobile workstation we also address the scalability of OrderPickAR. Since the developement is not sufficiant, we also conducted a study in which we compared OrderPickAR to currently used approaches. In addition we included a Pick-by-Vision approach developed in a related project by Sebastian Pickl. We analysed and compared different error types as well as the task completion time.

# Kurzfassung

Kommissionierung ist einer der wichtigsten Aufgaben in modernen Lagerhäusern. Da die meiste Arbeit noch immer manuell getätigt wird, werden neue Methoden zur Steigerung der Effizienz untersucht. Während die aktuell am meisten genutzten Ansätze, Pick-by-Paper und Pick-by-Light, entweder fehleranfällig oder nur unter hohen Kosten skalierbar sind, werden neue Methoden in Betracht gezogen. Diese Methoden schließen Pick-by-Vision Systeme basierend auf Augmended Reality ein, welche aber hauptsächlich auf den Nutzen von Head-Mounted Displays setzen. Um eine neue Methode zu untersuchen, haben wir OrderPickAR entwickelt, welches einen Kommissionierwagen und projizierte User Interfaces nutzt. OrderPickAR ist Teil des motionEAP Projekts der Universität Stuttgart und nutzt in-situ Projektion sowie Bewegungserkennung um den Nutzern zu leiten und Feedback zu präsentieren. Das intuitive Feedback der in-situ Projektion und die Bewegungserkennung geben OrderPickAR die Chance, Fehler auszumerzen, während gleichzeitig die Bearbeitungszeit einer Aufgabe reduziert wird. Durch die Nutzung einer mobilen Arbeitsstation berücksichtigen wir außerdem die Skalierbarkeit von OrderPickAR. Da die Entwicklung alleine nicht ausreicht, haben wir zusätzlich eine Studie durchgeführt, in welcher wir OrderPickAR mit aktuell genutzten Methoden verglichen haben. Außerdem haben wir einen Pick-by-Vision Ansatz, der von Sebastin Pickl in einem verwandten Projekt entwickelt wurde, in die Studie eingebunden. Wir haben unterschiedliche Fehlerarten und die Bearbeitungszeit untersucht und verglichen.

# Contents

# List of Figures

# List of Acronyms

**TCT** Task Completion Time

**PbV** Pick-by-Voice

**PbVi** Pick-by-Vision

**PbP** Pick-by-Paper

**PbL** Pick-by-Light

**TCP** Transmission Control Protocol

**UDP** User Datagram Protocol

**SDK** Software Development Kit

**NASA-TLX** NASA Task Load Index

**WPF** Windows Presentation Foundation

# 1. Introduction

The gathering of goods out of a warehouse following a specific order is called order picking. This task accounts for the most costs in modern warehouses [BH08] and thus raising efficiency is essential. While the most commonly used method is called Pick-by-Paper (PbP) [GRX+14], where the worker has a paper written list which states the items to pick, newer technology slowly enters the field. Pick-by-Light (PbL) is another popular method, where lights mounted at the shelves indicate which item to pick [Bau13]. Due to these methods being prone to errors (PbP) [GRX+14] or only scalable with high costs (PbL) still newer approaches are investigated. With the possible intuitive presentation of information, Augmented Reality is currently researched for the use in order picking environments. Although most researchers are focusing their work on the use of head-mounted displays, effectively creating a Pick-by-Vision (PbVi) system, Augmented Reality can also be used with projections. At the University of Stuttgart a project called motionEAP [1] aims to improve assembly tasks with the use of projection and motion recognition in the context of Augmented Reality.

MotionEAP is an assistance system in production processes, developed in cooperation with the University of Stuttgart, the Hochschule Esslingen as well as several companies (Audi AG, BESSEY Tool GmbH & Co. KG, GWW - Gemeinnützige Werkstätten und Wohnstätten GmbH, Schnaithmann Maschinenbau GmbH, Robert Bosch GmbH, KORION Simulation & Assistive Technology GmbH). With the use of motion recognition and in-situ projection, the system aims to improve efficiency of these processes. The system provides context sensitive help with in-situ projection to guide through tasks [FKS14a] and hint at errors. With changing and adapting the provided user feedback, motionEAP is also an ideal support for impaired persons.

In the current state, the main application of the motionEAP project is the assembly table construction. This construction can be placed on any horizontal surface and is designed to hold a projector as well as a depth sensor [FKS14b]. In addition, the construction provides several mounts for boxes. With a Kinect mounted above the table, it is possible to detect movement and objects [FKS14b] in the space above the table or on top of the table. It is also possible to detect interaction with objects or recognize gestures. The projector, mounted next to the Kinect, provides the possibility to project information directly into the workspace of the user [FKS14a]. These in-situ projection provide the user with feedback and can either guide the user through working tasks or show additional information on the current task. In order

[1] http://www.motioneap.de

to allow an accurate and permanent calibration the projector and Kinect are mounted (and movable) separately.

In total the motionEAP project includes four scenarios. The first scenario, the training, has several possible use-cases. It includes the instruction of a worker, the addition of new workflows to the system as well as the correction of errors to create an unmistaken process. Another scenario is the assembly of work pieces using the assembly table construction. As described before, task guidance or manuals as well as hints can be projected and in addition errors can be recognized. Thinking about a production or assembly line, the next scenario extends the previous with the use of multiple assembly tables. By connecting the tables to form an assembly line, the execution of complex and extensive workflows becomes possible. This scenario provides high flexibility but comes with extensive training and introduction costs as well as a high error rate. The last scenario is the order picking. With a combination of projection and motion recognition a guidance system can be developed. Additionally this system can control picks and also monitor the deposition.

As part of the motionEAP project, this thesis focuses on the order picking scenario. With the use of in-situ projection and motion recognition an Augmented Reality system called OrderPickAR was developed and implemented. In order to proof efficiency and utility of OrderPickAR a study was conducted, comparing OrderPickAR to other approaches, including PbP, Pick-by-Voice (PbV) as well as a PbVi approach developed in a related project by Sebastian Pickl.

The thesis is structured as follows:

**Chapter 2** - Related Work: An overview on previous work, related to the topic and project.

**Chapter 3** - Apparatus: Hardware and devices we used as part of OrderPickAR.

**Chapter 4** - OrderPickAR: Software we developed to show in-situ feedback and guide a user through the order picking task.

**Chapter 5** - Study: Comparative study with four different approaches to order picking.

**Chapter 6** - Future Work: Explaining important steps to take in order to improve OrderPickAR.

**Chapter 7** - Conclusion: An overview of the work done and the results.

# 2. Related Work

To get a good understanding of the necessary fields for our project studying related work is essential. Keeping our project and our goals in mind, previously developed and currently used approaches of order picking have to be investigated. In addition the utility and usage of Augmented Reality has to be considered. On the back of this, projected user interfaces have to be examined. We start with an overview of Augmented Reality, including definitions and use cases. Then we explain projected user interfaces, problems and interaction capabilities. At the end we introduce the importance of order picking as well as different approaches and new technical methods.

## 2.1. Augmented Reality

Being called "variation" [$A^+97$] or section ("Teilbereich" [Gün09]) of Virtual Reality, Augmented Reality is a growing technology finding its way in many fields. Following Günthner et al. [Gün09] Augmented Reality is a visual replenishment or also a superimposition for the real world. Furthermore Günthner et al. state that, depending on the visual context, information, generated by a computer, is added to the environment. Similar statements are found by Azuma [$A^+97$], where the contrast between Virtual Reality and Augmented Reality is emphasized. Describing Virtual Reality as immersion without sight on the real world around the user, Augmented Reality is explained to let the user see the real world "[...] with virtual objects superimposed upon or composited with the real world." [$A^+97$]. Due to the requirements to head-mounted displays, sometimes seen in literature, Azuma introduced a definition of Augmented Reality without limitation to a specific technology. The three characteristics defined by Azuma [$A^+97$] include the combination of virtual and real world, the interaction in real time as well as the 3D-registration of virtual content in the real world. Krevelen and Poelman [VKP10] also mention that the definition by Azuma does not limit Augmented Reality to only visual feedback nor does it restrict Augmented Reality to only the addition of objects.

As noted before, the possible and actual use of Augmented Reality is spread over several areas of application. Azuma [$A^+97$] mentions the possible use in a medical context, using Augmented Reality as training aid or visualization support during surgeries. Azuma especially explains the potential improvement for minimally-invasive surgery in which the treating physician has little chance of seeing much. The possibility of guiding in precision tasks is also mentioned by Azuma, giving the example of drilling a hole into a skull for a brain surgery. Yuen et al.

[YYJ11] present similar approaches for the use of Augmented Reality in a medical context. They mention the support for surgeons "[...] with navigation and orientation [...]" [YYJ11] as well as improvement of efficiency and safety by enhancing "[...] medical surgical and clinical procedures [...]" [YYJ11].

Furthermore Azuma [A+97] notes the use for manufacturing and repair. He explains the ease of understanding manuals if the instructions are "[...] superimposed upon the actual equipment [...]" [A+97] or 3-D drawings are shown. He also mentions the possible use of Augmented Reality as step-by-step guidance for consecutive tasks.

Yuen et al. [YYJ11] describe the use for construction tasks. By giving "[...] designers, workers, customers and employers [...]" [YYJ11] the chance to "[...] visualize and experience a virtual facility [...]" [YYJ11], possible architectural flaws or cosmetic misconceptions can be discovered. In a similar enumeration Azuma [A+97] additionally notes the possible preview of skylines if a new skyscraper is built. The company Metaio, which is specified on Augmented Reality applications, developed an application which also gives you a preview, not of a skyline but rather of your direct surroundings with additional furniture.

An application of Augmented Reality, which is common in the broadcasting of sport events, but is not realized as Augmented Reality, is the live overlay of current results. Krevelen and Poelman [VKP10] mention the indication of the current ranking in racing events or the first down line shown in American Football matches.

## 2.2. Projected User Interfaces

Projectors, in contrast to other screening devices (e.g. LCD, LED), do not have the restriction of a fixed screen but rather "[...] different surfaces become available to be used as displays." [Pin01]. Rukzio et al. [RHG12] mention the decreasing size of projectors as reason for non static use as handheld, wearable or stand-alone mobile devices.

With the mobility of projectors in mind, interaction with the projected information becomes essential. The user interface provides the connection between a human and a technical device. While the best known user interface, the graphical user interface, has a wide range of usage, other types of interfaces are developed and already in use. This includes speech recognition, touchscreens and tangible user interfaces which all focus on the type of interaction.

As said before and also stated by Huber [Hub14], projectors are becoming smaller, finally leading to pico projectors. Following Huber these devices fit in the palm of a hand and open up "[...] interesting oppertunities for novel user interfaces [...]" [Hub14]. Huber also mentions the overlay of information on a physical object as "[...] key application scenario [...]" [Hub14] but also alludes to the first of several problems in the use of projected user interfaces.

The problem mentioned by Huber is the necessity of complex algorithms for object recognition and tracking as well as projection mapping and alignment. Still referring to the use of "[...] mixed reality interfaces [...]" [Hub14] there are even more challenging implementations required to ensure a decent quality of interfaces. For example the real-world registration of objects, nonplanar projection surfaces and sensor fusion are mentioned by Huber [Hub14].

Another problem that comes in mind is brightness and contrast of projections. In typical surroundings the sunlight is cumbersome and sun-blinds need to be shut to prevent the image from looking blurry or even be barely visible. But with closed sun-blinds the lighting in the room can still be a problem and depending on the conditions a strong projector can be necessary. Relating to Pinhanez [Pin01] a projector with 1200 lumen suffices for a office room with the lights on. Bimber and Raskar [BR04] mention a brightness of 1100 lumen for "[...] a normally lit environment"[BR04] but don't specify a normally lit environment. Pinhanez also alludes another projector with 3000 lumen specifically for the projection on "[...] horizontal surfaces such as tables or desk [...]"[Pin01]. This special need is justified by a higher brightness of reflected light because those surfaces are oriented "[...] orthogonal to the sources of ambient light [...]"[Pin01]. Butz et al. [BSS04] mention 3000 lumen necessary for sufficient contrast and brightness in daylight.

Oblique projections are another relevant issue. If the projector is not positioned correctly in front of the surface, the projected image and shapes will get deformed. But not only a wrong positioned projector leads to oblique projections. Also the shape of the surface to which should be projected is important. For example the projection of a square on a globular shaped surface will show a distorted image of the square. To avert such oblique projections, " [...] the image to be projected must be inversely distorted prior to projection [...]" [Pin01]. Pinhanez [Pin01] explains an approach using a 3D computer simulation to transform the image so the projection does not look oblique.

If the requirements for a clearly visible projection are met, there is still another problem. A person can stand in front of the projector and obstruct the projection. To reduce the frequency and area of obstructions the placement of the projector is vital. Pinhanez [Pin01] states that placing the projector in the upper corner of a room is more effective than to place it in the middle of the ceiling. He names the size of the projection cone as well as the cone being closer to the wall as essential reasons.

Due to an interface not only communicating from computer to the human, but also the other way around, an interaction channel for the human has to be integrated. This channel can be based on different modalities, including external devices, direct interaction with the projection or manipulating the projection surface [Hub14, RHG12]. With newer technology in mind, the interaction via mouse, keyboard or a remote control seems rather old fashioned. Using additional sensors, like a depth sensor or an infrared camera, direct interaction with the projection as well as the use of gestures can be implemented. Referring to Huber [Hub14] this approach is preferably used when the projector is fixed in its position, projecting onto a fixed surface. Hardy and Alexander [HA12] describe their implementation of a touch surface using depth cameras and also refer to previous work like the dSensingNI framework [KNF12] or the work by Wilson [Wil10] which focused on such an implementation.

Referring to Rukzio et al. [RHG12] and Huber [Hub14] there are other possibilities of interaction for the human. In both workings the movement of the projector itself as well as manipulating the projection surface are presented as possibility, although in the context of mobile projection. Interacting by moving the projector implies tracking of the motion, either by the device itself or by an external tracking system [RHG12]. As presented by Huber [Hub14] the interaction with the projection surface is a tangible interaction. By moving "[...] physical objects that are projected onto [...]" [Hub14] in physical space the user can control the interface.

## 2.3. Order Picking

One of the most relevant operating sequence in modern warehouses is the order picking. Referring to Bartholdi et al. [BH08] it accounts for about 55% to 65% of costs in modern warehouses. Order picking is the process of collecting items from the stock by navigating between the shelves and picking up articles to complete an order. According to Tompkins [Tom10] a coworker spends 50% of the time traveling, 20% searching, 15% picking, 10% in setup and 5% doing other tasks while collecting an order. Most current work concentrates on efficient path planning [Bau13] because that is the biggest part of the order picking task. Following Koster et al. [DKLDR07] many other papers concentrate on automatic systems for order picking, although in most practical environments the task is still done manually.

The current state of the art is somewhat unclear. There are many different approaches using more or less technical aids. As said previously, in most warehouses the order picking task is still performed with a paper list called PbP [GRX+14]. This means a person uses a paper written list of the items to pick and then walks from one shelf to another to pick them up. In some areas this approach is supported by newer technical devices for example “mobile data entry devices” [RG09] to simplify the task for the coworker. Other approaches like PbL where lights at the shelves indicate which item to pick, and PbV, where a computer voice explains which item to pick, are rarely used. Current work focuses on PbVi, an approach mostly using head-mounted displays to guide the coworker.

There are two major strategies for the task of order picking. The first to mention is called “picker-to-parts” [DKLDR07, Par12] where “[...] the picker walks or rides to the picking location to retrieve items [...]” [Par12]. So to get a part of the order, the coworker has to navigate through the storage room and find the shelf in which the part is stored. This can either happen by foot or with a motorized cart. The other strategy is called “parts-to-picker” [DKLDR07, Par12] where necessary parts are transported to the picker. In contrast to picker-to-parts the costs of such a system are very high and the flexibility very low.

For picker-to-parts Koster et al. [DKLDR07] differ even further in low-level and high-level picker-to-parts. The level stands mostly for the height of the shelves the coworker has to pick from. So in case of low-level picker-to-parts the coworker can collect orders by foot or with the help of a low lift elevating platform truck. There are no further technical aids necessary because the shelves are small enough for the user to reach everything and there are no pallets to pick up. However for high-level picker-to-parks bigger technical vehicles are required. The shelves can be several levels high and the user can not reach any of the higher levels. So a fork-lift truck or a similar vehicle is needed, that can even pick up a whole pallet of items. Park [Par12] differs in three sections called "pallet pick", "case pick" and "broken-case pick". "Depending on the types of retrieval units [. . . ]" [Par12] this classification is scaled. So in case of a pallet pick several full pallet loads are collected thus a fork lift truck is needed, comparable to the high-level description by Koster et al. [DKLDR07]. Park calls it a case pick when less than a full pallet load but still several cases are collected. A broken-case pick is described as "[. . . ] an order pick where the picking quantity is less than a full case or in pieces." [Par12]. Park [Par12] states that the efficiency of order picking highly depends on storage and retrieval equipment. Koster et al. [DKLDR07] mention the estimation that about 80% of warehouses in Western Europe use low-level picker-to-parts.

In most warehouses the coworker has not only one order to pick up but up to 14 at once. In these cases the process is called batch picking [Bau13]. Following Baumann [Bau13] the efficiency of batch picking is higher and can differ depending on the sorting technique. If the coworker sorts the items while he is collecting them (called sort-while-picking) the efficiency in general is higher than if all parts are selected and sorted afterwards (called sort-after-picking).

There is still another strategy called zone picking. As the name suggests the storage is divided in different zones. Following Park [Par12] every picker gets a zone assigned and is responsible for the parts in the zone. For zone picking there are two different approaches. In "pick-and-pass" [Bau13, Par12] one order is handed from zone to zone and thus the list of needed items is completed. In "wave-picking" [Bau13, Par12] the order is handed to all zones at once and reaches a collection point where the items are merged. For zone picking the "[...] balance of workload between the zones [...]" [Par12] is essential to hold efficiency high.

As previously stated there are several methods the order picking task can be accomplished with. The most commonly used method is PbP [GRX$^+$14]. For this method a paper written list of the necessary items stating the location as well as the amount is handed to the coworker. The coworker than has to navigate to the given locations and pick up the needed amount of items. Although this approach shows a high flexibility and low costs [Bau13], it still has a high error rate [GRX$^+$14].

Another approach is called PbV where the coworker gets the information for the items broadcast with the help of a mobile computer and an ear plug. Following Starner [Sta02] wearable computers are used in most cases to free both hands for the coworker to pick parts. By using speech recognition on the device the coworker can interact with it by giving commands. If a pick is completed he can go ahead by saying “next pick” [WBS+10] or use one of the other commands (“repeat”, “back”, “empty” [WBS+10]) to interact with the device. Unfortunately this approach can be inapplicable because the noise level in the surrounding environment is too high.

A further method is PbL where the coworker is guided by lights which are mounted “[...] under or over each pick location [...]” [Bau13]. Activated lights indicate for the coworker which item to pick and can also be used to tell the coworker where to put this item afterwards. The advantages of this method include a high picking speed and a low error rate. Unfortunately the installation of a PbL system is very expensive.

The newest method in this field is called PbVi where the coworker gets supported by an Augmented Reality system [SRGK11] while executing the order picking task. Referring to Reif et al. [RGSK09] all necessary information is shown in the field of vision of the coworker. Most work focuses on using head-mounted displays to support the process. Schwerdtfeger et al. [SRGK11] mention several different possibilities for the visualization in a head-mounted display. This includes the navigation using arrows as well as an attention funnel or simple box highlighting. Schwerdtfeger et al. [SRGK11] also mention the coworker not having to look away from the shelf as the biggest advantage of PbVi over PbP.

To measure accuracy of a pick there are three types of errors to detect. Following Weaver et al. [WBS+10] and Baumann [Bau13] these three error types are substitutions, insertions and deletions. An error is called substitution if the coworker picks a wrong part respectively “[..] one part was swapped for another part [...]” [Bau13]. Insertions are when an additional item was picked or “[...] an unrequested part was put in an order bin [...]“ [WBS+10]. A deletion is the oblivion of an item when no replacement occurred or following Weaver et al. “[...] deletions are when a part was forgotten and not replaced by another object [...]”[WBS+10].

There are several causes for the errors mentioned before. Following Günthner and Rammelmeier [GR12] these causes can be ranged in four categories. While human factors are grouped together, errors attributed to the used method are another category as well as factors referring to the environment. The fourth category includes all factors which are not in the other three and is called others. Referring to Günthner and Rammelmeier [GR12] the primary source of errors is the human himself. Günthner and Rammelmeier [GR12] mention carelessness as one of the most common causes for errors. Depending on the used method carelessness leads to misjudging and miscounting as well as a false recognition of the number of picked items. This effect gets reinforced by the pressure of time often set in modern working environments. Günthner and Rammelmeier [GR12] also mention the wrong handling of order picking systems as another common cause for errors. They state that marking one or more steps as finished before executing these steps often leads to errors.

There are several approaches to reduce the number of errors. Referring to Günthner and Rammelmeier [GR12] using newer approaches for order picking like PbL or PbV already reduce the error rate in contrast to PbP. Weaver et al. [WBS+10] also state a lower error rate in newer approaches compared to PbP. Using additional technical devices like barcode scanners and scales help the coworker to avoid errors while cameras and depth sensors can be used to detect errors. Following Günthner and Rammelmeier [GR12] an intuitive presentation of the data can lead to reduced error rates and therefore improve productivity.

## 2.4. Summary and Discussion

So far, we have introduced a definition of Augmented Reality as presented by Azuma [$A^+97$] also supported by other authors [Gün09, VKP10], which does not limit Augmented Reality to specific technologies. This definition includes the combination of virtual and real world, the interaction in real time as well as the 3D-registration of virtual content in the real world. We also explained the interaction possibilities with projected user interfaces, including direct interaction with the projection or moving the projector itself. In addition we showed problems which come along with the use of projections. These problems include occlusion, lighting of the environment as well as oblique projections. Last but not least, we pointed out different approaches to order picking. While only PbP and PbL are commonly used, newer approaches focus on the use of head-mounted displays. These newer approaches try to overcome issues of the commonly used approaches, especially the high error rate of PbP and the low scalability of PbL.

Projections are seldom used in newer approaches, although they come with the possible fusion of advantages of PbP and PbL. A low error rate with a low Task Completion Time (TCT) included in a highly scalable system. Additionally currently tested pick validations aim to identify the picked items and their count by either using a scale or an algorithm for image recognition. The validation of a pick in the correct box is not introduced. These are exactly the ideas we want to address with our system and explain in the following pages.

# 3. Apparatus

To ensure functionality and utility of OrderPickAR, we had several hardware parts to design and mount. While we relied on depth sensors, projectors and a location tracking system for the registration, visualization and pick detection, we also designed an order picking cart. We equipped this cart with the depth sensors and projectors to provide a mobile workstation. We first present the cart itself, including design and utility. Afterwards we take a brief look at the used projectors. At the end we introduce the used depth camera as well as the location tracking system.

## 3.1. Cart

We designed an order picking cart (see Figure 3.1) which was not only suitable to hold our equipment but also reduced the possibility of occlusion by the user. To ensure competent manufacturing of the cart we engaged Schnaithmann Maschinenbau GmbH, a partner of the motionEAP project, to build and construct the cart.

The cart holds 49 boxes on either side, seven boxes in each of the seven rows. Mounts for Kinects (see 3.3) and projectors were attached to the upper construction, which is variable in its height, while the boxes remain at their position. The variable height was designed by us to ensure utility of the cart in different warehouses and additionally giving the chance to reduce occlusion. The mechanism supports a height of up to 3.38 meters. On the upper construction a total of three Kinects and three projectors were installed. While one pair of Kinect and projector was arranged to face the surrounding environment especially the shelf, the other two pairs were mounted to face either side of the cart. This adjustment allowed us to address every box on the cart as well as every box in the shelf even though movement of the cart is necessary for the latter. Having two of four wheels steerable, maneuvering of the cart is quiet simple and similar to the behavior of a car. To handle the pictures provided by the Kinects as well as calculating the needed visualization we also mounted a computer on the cart. This computer runs the main part of our software, not only analyzing the pictures of the Kinects to recognize potential picks, but also calculating the visualization to show on the projectors. In addition we added a router to the devices mounted on the cart in order to provide our own network and thus eliminate additional overhead from other users.

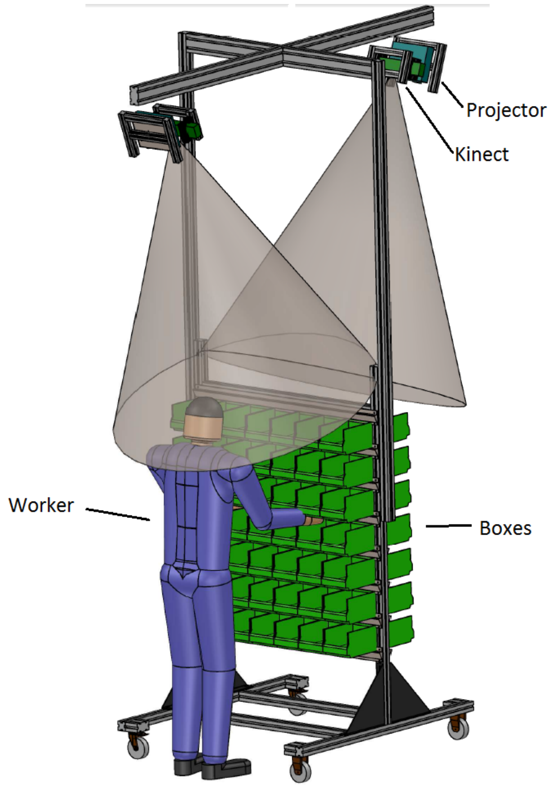



**Figure 3.1.:** Design drawing of our cart, showing the mounted devices and boxes as well as the projection cones and the worker.

## 3.2. Projectors

In order to address every box on the cart a projector with a normal projection field proofed to be sufficient. For this cause we used the Acer K335 [1] on either side of the cart. Depending on the distance between cart and shelf a similar projector could not reach the boxes in every altitude of the shelf. Therefore we used a short throw projector by Optoma, the EW610ST [2], with a bigger projection field on short distances to highlight boxes in the shelf. This allowed us not only to project in every altitude of the shelf but also in a broader field seen horizontally.

[1]http://www.acer.de/ac/de/DE/content/professional-model/MR.JG711.002
[2]www.optoma.de/uploads/brochures/EW610ST-B-de.pdf

## 3.3. Depth Camera

Although mostly known for their software solutions on personal computers like Windows or Office, Microsoft also developed a motion capture device called Kinect [3]. This device was primarily developed for the use with an Xbox 360 to allow the user interaction via gestures or body movement. The combination of a depth sensor, a RGB camera and several microphones allows the Kinect to track and locate players in open space in front of the device. With the development of the Kinect Software Development Kit (SDK) [4], users were allowed to use the built in sensors for their own projects and ideas. The big advantage of the Kinect is the marker free tracking and possible manipulation of the provided frames. Although there are several better depth sensors on the market, the Kinect provides sufficient accuracy with an adequate frame rate while still being affordable. We also decided to use the Kinect because it was already an established part of the motionEAP project.

## 3.4. Location Tracking System

NaturalPoint [5] is a company specialized on motion capturing and real-time motion tracking. OptiTrack is the name of NaturalPoint's optical motion capture technology which is used in films, games and other fields. Their website states their motion capture cameras are industry's best-performing tracking cameras. The motion recognition is based on infrared tracking of reflective markers which are small spheres. An OptiTrack system features high accuracy, with low response time and high frame rate if the setup is calibrated well. Furthermore the cameras come with great software support and provide excellent results. The reason we are using OptiTrack is the easy determination of position and orientation of multiple objects in a large scale.

In our setup 5.3 we used a total of 17 cameras which were mounted and oriented to cover the area in front of the shelf. For the best recognition of the cart we placed a total of seven markers on the cart, allowing us to determine the position and orientation even if not all seven markers were in the field of view of the cameras. In addition we used NaturalPoints Motive to collect and manage the pictures provided by the cameras. Motive allows us to calibrate the cameras, view live footage and also to create rigid bodies from multiple tracked markers.

[3] http://www.microsoft.com/en-us/kinectforwindows/
[4] http://www.microsoft.com/en-us/kinectforwindows/develop/default.aspx
[5] http://www.naturalpoint.com

# 4. OrderPickAR

We developed a system called OrderPickAR to calculate the visualization as well as handling the pick detection and orders. This software includes a component to handle the input by the Kinects and detect picks, as well as a component to calculate the images for the projectors. An additional component handles orders and the course of an order. We start by explaining the implemented Kinect Control. Afterwards we take a detailed look at the visualization for cart and shelf separately. Thereafter we present how the orders are managed and also the course of an order. At the end we take a brief look at the position streaming as well as the network infrastructure.

## 4.1. Pick Detection

To overcome possible problems like occlusion or a rather complex implementation of the pick detection we decided to mount the Kinects, responsible for the cart, vertically above the boxes. With this approach we eliminated occlusion given normal usage of our system. We also had the chance of a fairly simple implementation of the pick detection. With the Kinect looking down vertically, we developed an algorithm that only watches a field of about two centimeter directly in front of the boxes. In order to put a part into a box the user has to reach through this field and we can determine by depth changes in this field which row of the boxes the user tried to hit. By partitioning the rows in up to seven sections, we were able to create virtual 3D boxes in front of every real box on the cart. These virtual boxes allow us to verify the exact spot and real box the user tries to reach. To eliminate false pick detections occurring in areas where the border of two boxes meet, we added spaces in between the virtual boxes effectively reducing false picks. The resulting virtual 3D boxes covered the opening of the real boxes and helped us achieve an almost flawless pick detection on the cart.

In order to get more precise results we did not make the creation of the pick detection dynamically but rather implemented a manual configuration for the boxes. To accomplish best outcomes every row has to be configured individually due to the fact that boxes further away are seen smaller by the Kinect. With the help of a special calibration device and the graphical view of the image seen by the Kinect, the process to configure the boxes was kept simple. The calibration device had the length of a row of seven boxes and if laid on top of a row would reach into the area we were watching with the Kinect. To configure the boxes of a row the calibration device has to be placed on top of this row and a user has to click on the right and

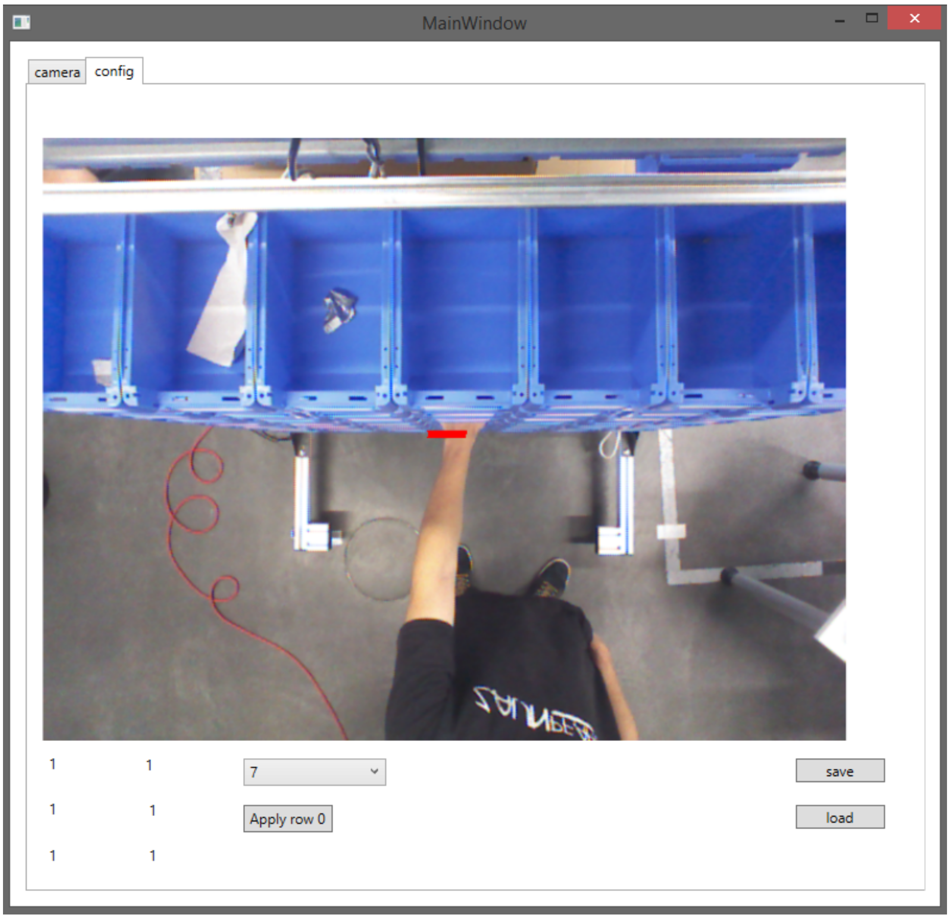



**Figure 4.1.:** The implemented control of the Kinect, giving visual feedback and allowing us to configure the pick detection.

left end of the device seen in the graphical view. Our algorithm then creates seven boxes in between these two clicks considering distances between boxes as well as the desired height of each box. By mapping the depth frame of the Kinect to its color image frame we can determine the altitude of the calibration device at both clicks. We then calculate the mean of those two values, add an offset and set this as positional elevation of the upper boundary of the row.

Finally in a running state, the software (see Figure 4.1) checks every third frame provided by the Kinect to avoid performance issues. In these frames the software inspects every pixel, which is part of our detection area and our boxes, for its depth value. If more than five depth values are inside the range of a box, we declared this box as triggered. In several trials with different quantity in depth values, five proofed to be a fairly good amount. It showed not only to eliminate false triggers caused by flickering in the Kinect image, but also detected most picks where the pieces were simply tossed inside the box without reaching inside. If a pick was detected ultimately, the software sends the pick under use of the Transmission Control Protocol (TCP) to notify the visualization of the detected pick. The TCP stream is used for communicating the detected picks without loss.

To overcome the disadvantage of the calibration, we implemented functions to save and load calibrations. This enabled us to configure the Kinect for different heights of the cart and made reconfiguring of the software easier than having to calibrate with every change of the height.

Note that our pick detection does not verify actual picks of items. It rather checks if the user reaches into the intended box.

## 4.2. Visualization

As most important part of our software the visualization not only provides the user with feedback but also functions as guidance for every task. In order to build an efficient system that shows and renders the visualization properly, we used two 3D models, one of the cart and one of the shelf, built in a Viewport3D in C#. With this approach we avoided the manual computation of complex matrices because it is already integrated in the viewport system of C#. We created two different Windows Presentation Foundation (WPF) windows to hold and manage the viewports, each showing on the relating projector. The models were created as 3D mesh put together with four points in open space and a material. The material determines the behavior of the surface if it is hit by light and additionally holds a brush. The brush is responsible for the color or respectively the look of the object. We used two different brushes. The SolidColorBrush class provided by C# allowed us to cover the object in one color, while the ImageBrush class gave us the opportunity to show a predefined image on the surface of the object. To make the models visible, cameras have to be added to a viewport. While the viewport itself represents the 3D world, added models represent objects and the camera is the point of view from which the visualization is rendered. We did not find it necessary to create the complete boxes, but rather focused on the (open) front side of the boxes. To get the required visualization, the settings of the cameras had to fit the specification of the used projector, especially in concern of the field of view. By changing position, look direction and also up direction of the camera in the viewport, we were able to adjust the cameras to the projectors on the cart.

Handling color or surface changes received from events, the visualization first compares the received data with the current data of the relating box. If a red material is sent, and the box is currently black (indicating an error occurred), it will be redrawn in red. If a red material is sent while the box is currently shown in red (correction of an error), it will be changed back to black. Finally, if a red material is sent and the box is currently yellow, nothing will be changed. In any other case the material will be changed without checking for current settings.

### 4.2.1. Color Conventions

We had three conventions applied for the visualization focusing on the used colors of the user feedback. These conventions apply for visualization of both cart and shelf, still considering the difference between picking and putting.

- green: indicating box to pick from or to put into
- red: indicating an error made by reaching into the wrong box
- yellow: indicating the boxes of the previous pick

For the user to complete a task, he or she simply has to follow the green highlighted boxes. Since our system does not verify the number of items picked and we did not want the user to pick every single item separately from one box, we introduced the yellow feedback. This gives the user the chance to pick another item if he or she picked too few or even put items back if the reception of too many items is recognized.

### 4.2.2. Cart

The visualization on the boxes held by the cart has one big advantage. The position and angle of the projector in relation to the boxes does not change when the cart is moving. So without implementing a dynamic visualization, a perfectly accurate view for the boxes is possible. We decided to highlight only the labels of the boxes (see Figure 4.2) changing the colors between the values presented with our conventions. Considering this, we only added the labels to our model in the viewport. In our model we used the fact that the labels are arranged on a vertical plane and made an exact copy of this alignment. The adjustment of the position and the angle of the camera then yielded the correct visualization for the boxes.


**Figure 4.2.:** A participant of our study, focusing the green highlighted box with number 49 to put items into it.

To adjust the camera position and angle we implemented a simple controller using the arrow keys as well as several other keys. With this implementation we were able to configure the camera while seeing instant feedback on the projected image. Since we do not want to adjust the camera every time we start the software, we also support saving and loading of calibrations. Anyway calibration is still necessary when the angle or position of the projector is changed.

### 4.2.3. Shelf

Although the visualization for the shelf is similar to the cart, there are some major differences. As mentioned previously, we used a short-throw projector for the visualization of the shelf resulting in a larger field fo view for the camera in the model. In contrast to the boxes on the cart, we did not use the labels of the boxes in the shelf but rather illuminated the entire box (see Figure 4.3). To achieve a complete illumination of a box, modeling the (open) side oriented towards the cart suffices and shows great results. Similar to the model of the labels on the cart, all boxes were aligned on a vertical plane. We decided to illuminate not only the inner part of a box but additionally the frame around a box. To avert graphical bugs caused by interleaving of (modeled) boxes on the particular frames, we added small variations of one pixel in the respective depth values. The resulting boxes were not aligned on a exact vertical plane but arranged within four pixels in front or behind it.

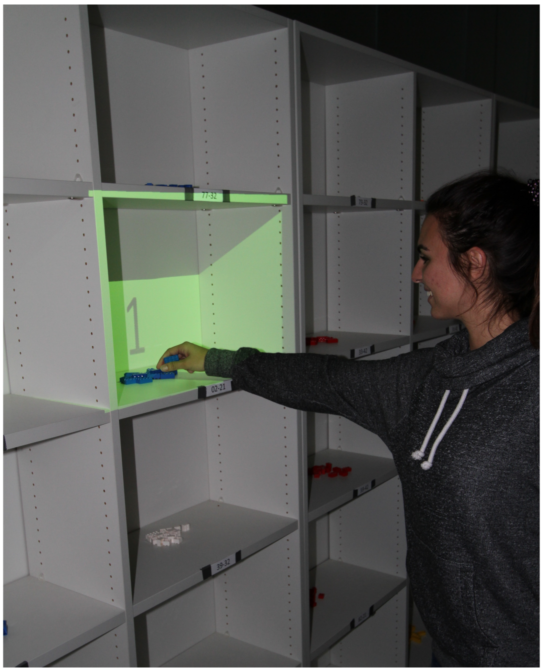


**Figure 4.3.:** A participant of our study, reaching into the highlighted box to pick up one item as indicated by the projection.

Another difference to the visualization on the cart is the dynamic positioning of the projector in relation to the shelf. By moving the cart, position and angle of the projector can change and thus the camera in our model has to be adjusted. To achieve preferably precise results, we used the data collected by the OptiTrack cameras to adapt the position and orientation of the camera in the model. During the implementation we encountered several problems with the orientation and adaption and decided to restrict the movement of the cart to a straight line. Other reasons for this constraint attracted our attention and are explained in the chapter Limitations and Problems.

In order to ensure a smooth course of tasks, we added two features to our visualization on the shelf. First we added numbers to the boxes representing the amount of items to pick from this box. Other possibilities like confining the user to pick only one item and flashing the highlighted box for multiple picks seemed to have greater disadvantages resulting in loss of time. The second feature we added are arrows shown and pointing to the right or left. These arrows are displayed if the next box to pick from cannot be highlighted by the projector because of the projection cone not reaching this box. The arrows indicate a necessary movement of the cart in the direction they are pointing.

## 4.3. Order Manager

Visualization and guiding tasks require a management to hold and administrate the necessary steps in order to execute and finish tasks. We developed a system containing four levels (see Figure 4.4). Starting from the lowest layer, a step is defined by a position and two indices. While the position differs in "C" for cart and "S" for shelf, the indices indicate row and column in the respective arrangement. The next layer includes tasks which are defined by exactly two steps, picking and putting, as well as the amount of items to pick. The first step holds position and indices for the shelf which represent the picking of items. In contrast the second step holds position and indices for the cart representing the destination of the items. The layer above the tasks is described by orders. In general an order holds all tasks necessary to fulfill an appointment which basically represents a list of items required to complete an assembly. On the topmost level the previously mentioned management is settled. Containing not only a list of orders, but saving the current and previous order, task and step separately, the order manager handles and validates picks, detected by the Kinects.

When receiving a detected pick, the order manager compares the values with the stored current step (see Figure 4.5). If the values match, several actions will take place. At first, the order manager creates a material with a SolidColorBrush containing the color yellow. Afterwards an event is sent to the visualization to change the material of the relating box to show a yellow surface. In addition the previous step will be colored in black, if it is not null, and the current step will be stored as previous step. Thereafter the current task is polled for the next step. If the request provides a valid step, it gets set as current step and a material containing a

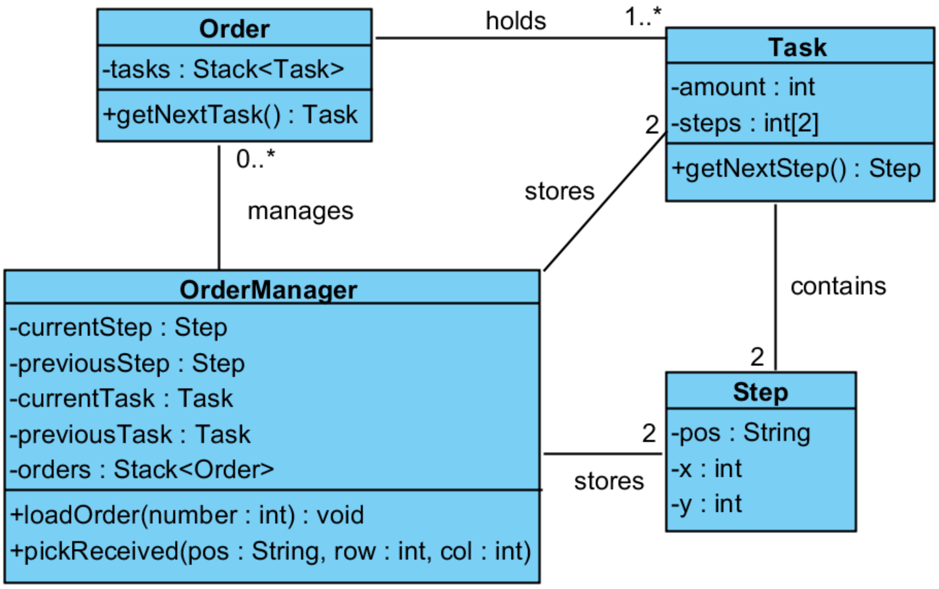



**Figure 4.4.:** A class diagram showing connections between Order Manager, orders, tasks and steps.

green SolidColorBrush is sent to the visualization. However if the requests provides null, the current task is finished and the next task is loaded. If the next task happens to be null, the current order is finished but otherwise a material containing an ImageBrush will be sent to the visualization. This ImageBrush loads a predefined image showing a black number on a green background. Depending on the amount of items specified in the task the image containing the appropriate number is loaded.

If the values do not match upon receiving a detected pick, a material with a SolidColorBrush containing the color red will be sent to the visualization.

## 4.4. Position Streaming

For the cause of calculating and processing the positions tracked by the OptiTrack cameras, a software developed by NaturalPoint is used. The software with the name Motive [1] provides a rich interface, visualizing the data of the cameras as well as calibrating the cameras and also streaming position of markers or rigid bodies. A rigid body is a combination of up to seven markers and allows the software to recognize the arrangement of these markers, even after leaving and reentering the covered area. By adding markers to the cart and creating a rigid

[1] https://www.naturalpoint.com/optitrack/products/motive/

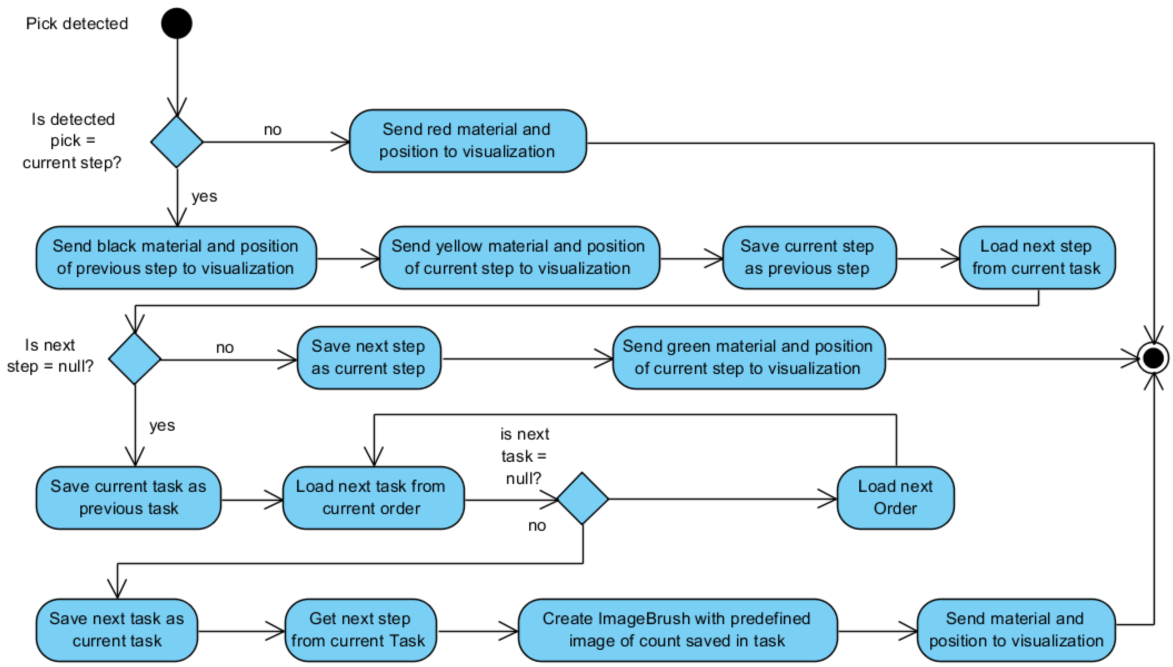



**Figure 4.5.:** The activity diagram shows the steps which are executed if a detected pick is sent to the Order Manager.

body of these markers, we are able to determine the position of the cart. With the supported feature for multicast streaming of Motive, we send the position of the cart into our network. To receive and process the stream, we developed a C# application, using the NatNet SDK [2] by Naturalpoint. Due to the associated project by Sebastian Pickl using an Android device, we first intercept the multicast stream on the same computer and transform it into a User Datagram Protocol (UDP) stream. This is vital because the used Android device cannot receive a multicast stream and a TCP stream simultaneously. The created UDP stream is then streamed to the computer on the cart to update the visualization.

## 4.5. Network Infrastructure

To ensure connectivity between the different parts of our system, we built our own network which also reduced traffic caused by other users. We had one computer which had the OptiTrack cameras connected via USB and was responsible for the input provided by the cameras. The other task of this computer was to stream the data collected by the cameras to the computer

[2] https://www.naturalpoint.com/optitrack/products/natnet-sdk/

on the cart via UDP. Serving as our central processing unit, the computer on the cart not only organized the connected projectors but also managed the Kinects plugged in via USB. The last part of our infrastructure was another computer acting as Wizard of Oz input for the pick detection in the shelf. This computer sent the picks via TCP to our computer on the cart.

# 5. Study

To compare our system to other methods and proof the utility, we conducted a study in which we measured the performance of several participants, each performing in every method. Then we compared the results of the different methods and also discuss these results. First we present the study design, with which we introduce the tasks we created as well as the measurements we noted. Afterwards we take a brief look at the participants of our study. Thereafter we introduce the environment in which the study took place, the concluding questionnaire we created and other equipment we used during the study. This is followed by the introduction of the conditions which include PbP, PbV and PbVi. Following the conditions, the procedure of the study is explained. At the end of the chapter the results as well as the discussion are presented.

## 5.1. Study Design

We introduced a study, using the "repeated measure" design in which we compared four different conditions on order picking. These conditions were PbP, PbV, PbVi and OrderPickAR. Every participant had to complete an order with all conditions.

Since we wanted to provide balanced task loads and comparable orders, we decided on several conventions for each order. We designed an order to be a collection of 16 picks, while a pick was the process of moving the cart (if necessary), picking the items from the shelf and putting the collected items into a box on the cart. In each order the participant had to collect a total of 44 items divided in three picks of one item, four picks of two items, five picks of three items, two picks of four items and two picks of five items. Due to the fact that small people could not be able to see or reach into the upper row in the shelf, we did not address this row in any of our orders. For the remaining four rows, we added four picks from each row to an order. Furthermore we added movement of the cart to the tasks, because we wanted to represent a big warehouse in our comparatively small room. Since we also wanted the cart movement to be balanced, we specified three positions in which the cart had to be placed and added the same traveled distance in the same increments to every order.

In order to compare the different methods, we observed the TCT as well as errors made during the execution. Although we did not measure the time needed for a single pick, we still took the time needed to complete all picks of one condition.

To analyze if the methods are prone to error we introduced several types to the participants. Three types of errors were controlled during the process and are called initial errors. One of our team members followed the participants and wrote down if they reached into a wrong box on either the shelf or the cart. Due to the implementation of our pick detection we decided to count reaching into a wrong box, on either the shelf or the cart, as an error. In this case the participant must not necessarily pick an item. By recording these initial errors, we wanted to show how often such errors result in an actual error. An actual error was wrong parts or wrong amount of parts in a box at the end of the task. In addition to the previous mentioned errors we counted wrong placement of the cart as error. If the participant placed the cart too far from the marked spots an error has been noted.

After completing a condition, we investigated each box on the cart for the containing items. If we found a wrong amount of items in a box we have been noting an error. This included boxes which should be empty but contained one or multiple items as well as empty boxes which should contain a certain amount of items. Besides, a box containing more or less items than it should, counted as error.

## 5.2. Participants

A total of sixteen participants attended our study. Most of the participants were regular students of the university, but we also welcomed two Ph.D. students, one research assistant and one secretary. While all four female participants were between 23 and 43 years (M=29.25, SD=6.9) old, the twelve male participants ranged from 20 to 31 years (M=23.3, SD=2.2).

All participants were volunteers. Since every participant had to complete each condition, which took between 50 and 60 minutes, we decided to pay them five Euros for attending our study.

To exclude accustoming as possible factor of the outcome of our study, we shuffled the order in which each participant executed the conditions. Since we still wanted a balanced positioning for each method, we used a balanced Latin square to determine the order for each participant.

## 5.3. Apparatus

In order to get information about the subjective opinion of the participant on the used methods, we instructed them to fill out a raw NASA Task Load Index (NASA-TLX) [HS88](NASA Task Load Index, see Appendix A.3) after completing a condition.

To give the participants a chance of contributing to our system, we introduced a concluding questionnaire (see Appendix A.1) about the executed conditions. In the first question the participants were asked to rate every condition on a Likert scale ranging from "dislike a lot" (1) to "like a lot" (5). The second question was about which method they would like to use if they

**Figure 5.1.:** The order picking cart holding several Kinects and projectors, and standing in front of the shelves.

worked in a fulltime job as order picker. The next question was about possible improvements of the methods or respectively what the participant would wish for. In the last question we asked the participant what they did not like about the processed conditions. With these questions we want to find out, which condition was the best in the eyes of the participants. In addition we want to know which improvements could be made for any of the presented conditions.

To provide an environment in which we could perform the study, we built a small warehouse(see Figure 5.1) in a room appropriated by the University of Stuttgart. Our warehouse includes three shelves, each comprising ten boxes (see Figure 5.2), to hold our picking items. The resulting thirty boxes were arranged in five rows and six columns. Despite the few boxes the layout proofed sufficient for our intentions. The items we placed in the shelves were Lego stones in different sizes and colors. In addition to the shelves we used the OptiTrack system to monitor our warehouse. Last but not least we placed our order picking cart, which could be moved unrestricted, in the warehouse.

## 5.4. Conditions

In order to monitor the efficiency and utility of our system, we compared a total of four methods. These methods include one wide spread approach, PbP, as well as a rather seldom


**Figure 5.2.:** The box with number 02-21 in a shelf, holding blue Lego stones.

used PbV approach. Furthermore we introduced an approach also based on Augmented Reality, but relying on head-mounted displays, called PbVi. Our system completed the set of the methods.

### 5.4.1. Pick-by-Paper

As described previously in 2.3, PbP is based on a list of datasets for every order. This list contains different information for every pick which has to be executed to finish an order. Every line provides the location of the items in the shelf (from), the location of the box to put the picked items into (to), the amount of items to pick (count), as well as additional information such as pricing or article number (see Appendix A.2). We instructed the user to follow the order scheduled on the list because of the balancing of the different orders we defined previously. So to complete an order, the participant had to search for the necessary information on the list, move the cart, pick the items and put them into the predefined box.

### 5.4.2. Pick-by-Voice

The PbV method was also introduced in 2.3. This method provides the participant with verbal input, (see Figure 5.3) stating the location of items in the shelf, the amount of items to pick as well as the location of the box in which the items are to be put.


**Figure 5.3.:** A user reaching into the shelf, equipped with a device for PbV commands.

We did not implement a prototype of this method but rather used a Wizard of Oz approach. On of our team members lectured the instructions and also reacted to commands given by the participants. These commands had to be vocalized by the participants and included "wiederholen" (repeat), "weiter" (next) and "zurück" (back). An example of an instruction is "Nimm fünf aus Fach 28-32 und lege sie in Fach 47" (Pick five from 28-32 and put them into 47).

### 5.4.3. Pick-by-Vision

The last condition we would like to introduce is a method based on Augmented Reality using a head-mounted display (see Figure 5.4). The system, developed by Sebastian Pickl in collaboration with our project, availed the OptiTrack system as well as the developed pick detection. Relying on the Epson Moverio BT-200 [1], smart glasses based on Android, the system showed a tunnel leading the participants to the destined location. Spheres in different colors were shown at the destination to help the participants with identifying the correct box. The colors of the spheres followed the same conventions as OrderPickAR, introducing green as next pick, yellow as previous pick and red as error. By equipping the device with OptiTrack markers, tracking of movement as well as head position and angle could be determined.

[1] http://www.epson.de/de/de/viewcon/corporatesite/products/mainunits/overview/12411

**Figure 5.4.:** A participant reaching into the shelf while following the instructions given by the head-mounted display.

## 5.5. Procedure

Since we attended every participant individually, we started with an introduction to order picking for every participant. We explained our warehouse setup as well as the measurements we took during the course of the study. In addition we showed them which actions would result in errors. After the introduction the participants had to fill an initial questionnaire about their gender, age and employment. Furthermore they had to fill a consent form, with which they allowed us to use the collected data of the study and could disown us from taking pictures of them.

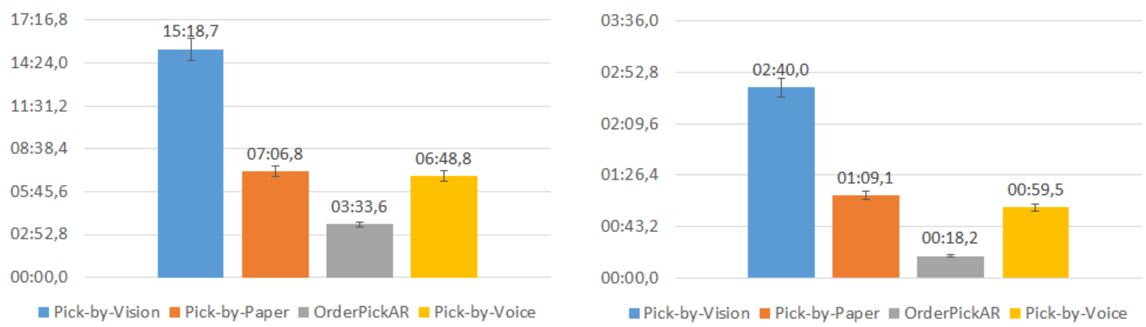
Since every participant had to complete all conditions, the following steps were identical for all conditions. We began with an introduction of the condition, what it is based on and what the participants had to do. Then we gave the participants the chance to accustom to the condition by executing three tasks. After placing the cart in the start position, we commenced the time measurement and the condition with a signal given by the facilitator. During the execution the facilitator accompanied the participants to note initial errors. When the participants put the last item in the cart, we stopped the time. Then the participants had to fill out a raw NASA-TLX about the condition they just completed. In the same time the cart was inspected for actual errors and the picked parts were sorted back.

After completing the last condition and filling out the relating NASA-TLX, the participants had to answer the concluding questionnaire. Thereafter the participants were given the reward for attending the study.

## 5.6. Results

During the evaluation of the results, we identified considerable differences in the average TCT for each condition (see Figure 5.5). While OrderPickAR proofed to be the fastest with an average TCT of 3:33.6 minutes, the PbVi approach was the slowest with an average TCT of 15:18.7 minutes. The mean TCT of PbP (7:06.8 minutes) and PbV (6:48.8 minutes) were close to each other, with a small advantage of PbV. The SPSS [2] analysis showed us the significance of some of these differences. Due to the test with assumed sphericity not delivering expressive results, we had to use the Greenhouse-Geisser correction. This showed us a significant difference between the TCT of PbVi compared to all other conditions. It also showed no significant difference between PbV and PbP. However a significant difference between OrderPickAR and all other conditions was stated.

In addition, the mean derivation to the average time had the same succession as the TCT (see Figure 5.5). With a mean derivation of 0:18.2 minutes, OrderPickAR showed the smallest difference. The PbVi approach yet again produced the highest value with a mean derivation of 2:40.0 minutes. PbV, with a mean derivation of 0:59.5 minutes, indicated a small advantage over PbP with a mean derivation of 1:09.1 minutes.




(a) (b)

**Figure 5.5.:** Representation of the (a) mean task completion time and the (b) mean derivation of task completion time per condition.

The errors we recorded showed some interesting results (see Figure 5.6). While we encountered no initial errors during the use of OrderPickAR (wrong cart placement not included), it showed 0.6875 actual errors per participant. This result was caused by one participant who picked a wrong amount of items eleven times. With an average of 0.4375 initial errors per participant, PbP recorded the second lowest results. In addition it also showed the second lowest results in actual errors with 0.5625 errors per participants. In PbV an average of 0.9375 initial errors

[2] http://www-01.ibm.com/software/de/analytics/spss/

per participant resulted in 1.25 actual errors. We encountered the highest initial error rate of 9.625 per participant during the PbVi approach. However, these initial errors only resulted in 0.125 actual errors. Yet again the SPSS analysis showed us the significant differences of the error rates. While we had to use Greenhouse-Geisser again, a significant difference between PbVi and all other conditions was stated. However, the difference between PbV, PbP and OrderPickAR did not account for a significant result.

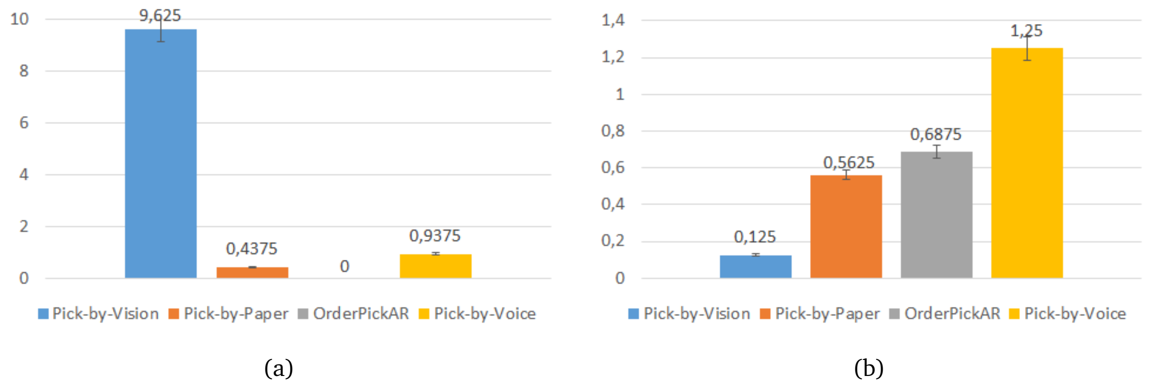



**Figure 5.6.:** Representation of the (a) initial and (b) resulting errors, not including wrong placement of the cart.

Reviewing the SPSS analysis of the NASA-TLX, we could rely on the spherical test (see Figure 5.7). The results showed a significant difference between OrderPickAR and all other conditions. The other conditions showed no significant difference between each other.

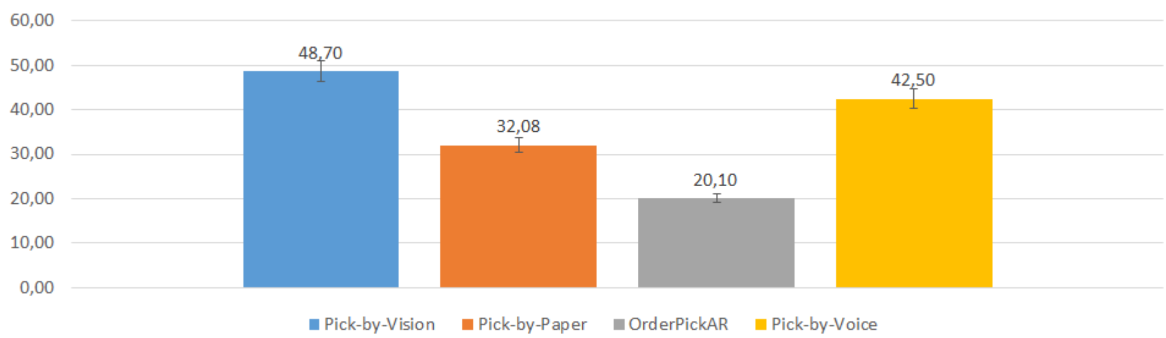



**Figure 5.7.:** Representation of the results of the NASA-TLX.

On the questionnaire, the participants had to fill out at the end of the study, a Likert scale for each method was given. While thirteen of the participants answered "like a lot"(5) for OrderPickAR, the most common value for the PbP was neutral (3). Rather unpopular were PbV, with seven values on "dislike" (2) and five on "dislike a lot" (1), as well as PbVi with eight values on "dislike a lot" (1) (see Figure 5.8).

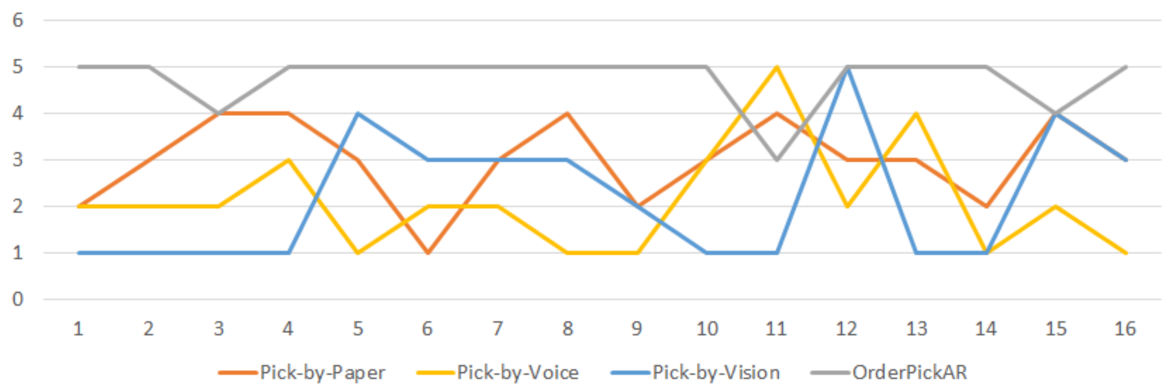



**Figure 5.8.:** Representation of the results of the Likert scale.

Fourteen participants would like to use OrderPickAR, if they were employed as order picker. Only one participant would like to use PbP, while another participant did not want to use any of the conditions. Most of the comments to possible improvements were focused on the PbVi system, mentioning things like faster presentation with less delay, more accurate registration and faster feedback. One participant mentioned the improvement of the color codes as well as the visualization of the numbers for OrderPickAR. When the participants were asked what they did not like, most of them answered with PbV. These answers were justified with annoyance of the system and high mental demand. Multiple participants described PbVi as uncomfortable and slow. One participant mentioned the occlusion of the lower boxes during the use of OrderPickAR.

## 5.7. Discussion

A clear advantage of OrderPickAR over the other conditions was shown by our study. Not only the TCT, but also the acceptance by the users demonstrated the benefit of OrderPickAR. However, the resulting errors did not show the same advantage but still proofed the utility and competitive ability of OrderPickAR. Only one participant made errors while using OrderPickAR, while all other participants finished the condition flawless. In contrast, the errors of the other conditions were spread over all participants. Our study also showed the similarity between PbP and PbV, not only TCT, but also error rate. The PbVi approach showed a low resulting error rate, but with an unacceptable slow TCT.

The results of our study do not coincide with previous studies. While Schwerdtfeger et al. [SRG+09, SRGK11] and Günthner [Gün09] showed a slightly faster completion time of PbVi compared to PbP or PbV, our study did not affirm these results. However, we also discovered a lower error rate of the PbVi approach but this was due to the used pick detection. The participants started to use trial and error to find the correct box to pick from and to put into.

We executed the study with several limitations, which do not affect the confidence in our findings. For instance, we did not implement the pick detection for the shelf. Due to the high complexity of this implementation and the occlusion by the user, hindering us from using the bottom rows of the shelf, we used a Wizard of Oz to detect the picks. We also limited the movement of the cart to a straight line. While our prototype did not consider other movements, we also decided that movement in all directions is rather unusual in warehouses with shelves standing close together. Future work has to focus on the implementation of the pick detection for the shelves, keeping the occlusion in mind.

### 5.7.1. Limitations

During the study we encountered two limitations to our system. These limitations include the lighting of the surrounding environment as well as the occlusion by the user.

The efficient use of projections often depends on the surrounding environment. With modern warehouses being well-lighted most projections won't be clearly visible. Depending on the used projector or respectively the lumen provided by its bulb, images can be blurry or close to invisible even on a close range. This is not only a problem our system faces but rather all similar projects using projections will encounter this problem. In the chapter about related work (on page 18) we already presented the problems that occur with projections, including the lighting of the surrounding environment.

The biggest problem we encountered during the development and the study was the obstruction by the user. Standing in front of the projector causes the visualization to be only partial visible on the shelf, or not visible at all. This can lead to confusion of the user and a waste of time. The occlusion of the Kinects was also a problem. Due to the mounting above the boxes on the cart, the pick detection for the cart had almost no problems with occlusion. However, the Kinect for the shelf suffered heavily from occlusion. While investigating possible solutions we realized the complete occlusion of the two bottom rows, if the user crouched to pick an item.

# 6. Future Work

Although our system proofed to be efficient and beneficial for the order picking task, it is still a prototype and needs further improvement.

In order to refine our prototype, a different method to determine the position of the cart has to be used. Even though the OptiTrack system proofed to be fast and reliable, it is still not applicable for a bigger warehouse. Therefore the usage of other technologies has to be inspected. Using one of the Kinects to determine the position of the cart may be adaptable. Regarding the approach of Butz [BSS04], placing additional markers in the environment or on the shelves would be an approach to start with.

Another improvement of our prototype concerns our visualization. While it is obvious that the movement of the cart in all directions has to be implemented, also the used colors need to be investigated. During the course of our study, participants complained about the yellow highlighting and some were even confused by it.

Due to the fact that the cart is big and unhandy, the design of the cart has to be reconsidered. Changing the layout of the whole cart as well as the arrangement of the boxes and devices has to be thought over. A cart without variable height might be acceptable, because changing the height of our cart comes with necessary adjustments of the mounted devices.

While the pick detection for the cart works almost flawless, the pick detection for the shelves is not yet implemented. While tracking the hand of the user with the Kinect might be an applicable approach, developing a detection, similar to our working system for the cart, could be a better implementations. In order to detect all errors, a verification of the picked amount of items has to be implemented. Previous work already introduced systems based on picture comparison [LCTM12].

Last of all, future work has to address the problems of the electricity supply for the cart. Long cables would block paths in big warehouses and also hinder the movement of the cart. Using an accumulator might be the best approach, but it is unclear if the provided energy would suffice for one workday or if the technology simply takes to much space.

# 7. Conclusion

We designed and developed a system using Augmented Reality and projected user interfaces to improve and enhance the order picking task. The name of the system is OrderPickAR. During the development we relied on hardware such as the built order picking cart, the Kinects, the projectors and the OptiTrack system. We implemented the software of OrderPickAR around the used hardware. With in-situ projection directly on the shelves an intuitive user interface is granted. Adding an implicit pick detection using Kinects, errors can be detected and also be revoked. In order to proof the utility and benefit of OrderPickAR we conducted a study, in which we compared different approaches to order picking. These approaches included PbP, PbV and PbVi. The results of the study underline the potential of OrderPickAR. Although still a prototype, it outnumbered the other conditions in TCT and acceptance by the users. Furthermore the errors made during the study showed that OrderPickAR is in no way inferior to commonly used methods in the industry. However the course of the study also showed and accentuated work, which has to be put in development and further improvement of OrderPickAR.

# A. Appendix

Bewerten Sie die angewandten Methoden auf der folgenden Skala:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Paperpicking** | *gefällt mir nicht* | o | o | o | o | o | *gefällt mir sehr* |
| **Pick-By-Voice** | *gefällt mir nicht* | o | o | o | o | o | *gefällt mir sehr* |
| **Head Mounted Display** | *gefällt mir nicht* | o | o | o | o | o | *gefällt mir sehr* |
| **Projektionen** | *gefällt mir nicht* | o | o | o | o | o | *gefällt mir sehr* |

Welche Methode würden Sie 8 Stunden am Tag (also ein Arbeitstag) benutzen wollen?

Was könnte man verbessern bzw. was würden Sie sich wünschen?

Was hat Ihnen weniger gefallen?

**Figure A.1.:** Questionnaire the participants had to fill out at the end of the study.

| Pos | Article Number | Count | From | To | Price | Checked |
|---|---|---|---|---|---|---|
| 1 | 757-567-346 | 3 | 16-22 | 43 | | Gelbes 2x2 |
| 2 | 999-234-313 | 1 | 91-22 | 30 | | Gelbes 2x1 |
| 3 | 640-324-346 | 1 | 33-42 | 46 | | Rotes 4x2 |
| 4 | 443-563-564 | 2 | 82-21 | 38 | | Rotes 2x2 |
| 5 | 172-391-134 | 4 | 44-22 | 42 | | Grünes 2x1 |
| 6 | 743-849-173 | 5 | 13-22 | 36 | | Weißes 3x2 |
| 7 | 920-876-776 | 2 | 39-32 | 9 | | Weißes 2x2 |
| 8 | 774-143-332 | 3 | 23-32 | 3 | | Hellblaues 4x2 |
| 9 | 867-174-651 | 3 | 59-42 | 10 | | Rotes 3x2 |
| 10 | 593-739-096 | 2 | 31-42 | 28 | | Orange 2x2 |
| 11 | 632-649-756 | 3 | 16-42 | 18 | | Grünes 3x2 |
| 12 | 113-413-073 | 5 | 24-22 | 32 | | Schwarzes 2x1 |
| 13 | 467-432-097 | 4 | 25-21 | 34 | | HellGrün 4x2 |
| 14 | 967-312-532 | 3 | 00-42 | 37 | | Grünes 4x2 |
| 15 | 777-654-333 | 1 | 02-21 | 11 | | Blaues 4x2 |
| 16 | 423-563-109 | 2 | 13-42 | 20 | | Weißes 4x2 |

**Figure A.2.:** The complete paperpicklist we handed to a participant.

Proband ID: ______________ Bedingung: _______________ Datum: ___________________

Mentaler Aufwand Wie geistig anspruchsvoll war die Aufgabe?

Sehr niedrig Sehr hoch

Körperlicher Aufwand Wie anstrengend war die Aufgabe?

Sehr niedrig Sehr hoch

Zeitlicher Aufwand Wie hastig oder gehetzt war das gesetzte Tempo der Aufgabe?

Sehr niedrig Sehr hoch

Performance Wie erfolgreich waren Sie im Lösen der Aufgabe?

Perfekt Gescheitert

Aufwand Wie sehr mussten sie sich anstrengen, um Ihre Leistung zu erreichen?

Sehr niedrig Sehr hoch

Frustration Wie unsicher, entmutigt, irritiert, gestresst, und verärgert waren Sie?

Sehr niedrig Sehr hoch

**Figure A.3.:** The NASA-TLX participants had to fill out after completing a condition.

# Bibliography


[A$^{+}$97] R. T. Azuma, et al. A survey of augmented reality. *Presence*, 6(4):355–385, 1997. (Cited on pages 15, 16 and 21)

[Bau13] H. Baumann. *Order picking supported by mobile computing*. Ph.D. thesis, University of Bremen, 2013. (Cited on pages 13, 18, 19 and 20)

[BH08] J. J. Bartholdi, S. T. Hackman. *Warehouse & Distribution Science: Release 0.89*. Supply Chain and Logistics Institute, 2008. (Cited on pages 13 and 18)

[BR04] O. Bimber, R. Raskar. Spatial Augmented Reality. In *ISMAR*, p. 306. 2004. (Cited on page 17)

[BSS04] A. Butz, M. Schneider, M. Spassova. SearchLight–a lightweight search function for pervasive environments. In *Pervasive Computing*, pp. 351–356. Springer, 2004. (Cited on pages 17 and 47)

[DKLDR07] R. De Koster, T. Le-Duc, K. J. Roodbergen. Design and control of warehouse order picking: A literature review. *European Journal of Operational Research*, 182(2):481–501, 2007. (Cited on pages 18 and 19)

[FKS14a] M. Funk, O. Korn, A. Schmidt. Assistive Augmentation at the Manual Assembly Workplace using In-Situ Projection. 2014. (Cited on page 13)

[FKS14b] M. Funk, O. Korn, A. Schmidt. An augmented workplace for enabling user-defined tangibles. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pp. 1285–1290. ACM, 2014. (Cited on page 13)

[GR12] W. A. Günthner, T. Rammelmeier. Auf dem Weg zur Null-Fehler-Kommissionierung. 2012. (Cited on pages 20 and 21)

[GRX$^{+}$14] A. Guo, S. Raghu, X. Xie, S. Ismail, X. Luo, J. Simoneau, S. Gilliland, H. Baumann, C. Southern, T. Starner. A comparison of order picking assisted by head-up display (HUD), cart-mounted display (CMD), light, and paper pick list. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, pp. 71–78. ACM, 2014. (Cited on pages 13, 18 and 19)

[Gün09] W. A. Günthner. *Pick-by-Vision: augmented reality unterstützte Kommissionierung*. Lehrstuhl für Fördertechnik Materialfluss Logistik, 2009. (Cited on pages 15, 21 and 45)

[HA12] J. Hardy, J. Alexander. Toolkit support for interactive projected displays. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, p. 42. ACM, 2012. (Cited on page 17)

[HS88] S. G. Hart, L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988. (Cited on page 38)

[Hub14] J. Huber. A Research Overview of Mobile Projected User Interfaces. *Informatik-Spektrum*, pp. 1–10, 2014. (Cited on pages 16, 17 and 18)

[KNF12] F. Klompmaker, K. Nebe, A. Fast. dSensingNI: a framework for advanced tangible interaction using a depth camera. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, pp. 217–224. ACM, 2012. (Cited on page 18)

[LCTM12] X. Li, I. Y.-H. Chen, S. Thomas, B. A. MacDonald. Using Kinect for monitoring warehouse order picking operations. 2012. (Cited on page 47)

[Par12] B. C. Park. Order Picking: Issues, Systems and Models. In *Warehousing in the Global Supply Chain*, pp. 1–30. Springer, 2012. (Cited on pages 18 and 19)

[Pin01] C. Pinhanez. The everywhere displays projector: A device to create ubiquitous graphical interfaces. In *Ubicomp 2001: Ubiquitous Computing*, pp. 315–331. Springer, 2001. (Cited on pages 16 and 17)

[RG09] R. Reif, W. A. Günthner. Pick-by-vision: augmented reality supported order picking. *The Visual Computer*, 25(5-7):461–467, 2009. (Cited on page 18)

[RGSK09] R. Reif, W. A. Günthner, B. Schwerdtfeger, G. Klinker. Pick-by-Vision comes on age: evaluation of an augmented reality supported picking system in a real storage environment. In *Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, pp. 23–31. ACM, 2009. (Cited on page 20)

[RHG12] E. Rukzio, P. Holleis, H. Gellersen. Personal projectors for pervasive computing. *IEEE Pervasive Computing*, 11(2):30–37, 2012. (Cited on pages 16, 17 and 18)

[SRG+09] B. Schwerdtfeger, R. Reif, W. A. Gunthner, G. Klinker, D. Hamacher, L. Schega, I. Bockelmann, F. Doil, J. Tumler. Pick-by-Vision: A first stress test. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pp. 115–124. IEEE, 2009. (Cited on page 45)

[SRGK11] B. Schwerdtfeger, R. Reif, W. A. Günthner, G. Klinker. Pick-by-vision: there is something to pick at the end of the augmented tunnel. *Virtual reality*, 15(2-3):213–223, 2011. (Cited on pages 20 and 45)

[Sta02] T. E. Starner. Wearable computers: No longer science fiction. *Pervasive Computing, IEEE*, 1(1):86–88, 2002. (Cited on page 19)

[Tom10] J. A. Tompkins. *Facilities planning*. John Wiley & Sons, 2010. (Cited on page 18)

[VKP10] D. Van Krevelen, R. Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2):1, 2010. (Cited on pages 15, 16 and 21)

[WBS+10] K. A. Weaver, H. Baumann, T. Starner, H. Iben, M. Lawo. An empirical task analysis of warehouse order picking using head-mounted displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1695–1704. ACM, 2010. (Cited on page 20)

[Wil10] A. D. Wilson. Using a depth camera as a touch sensor. In *ACM international conference on interactive tabletops and surfaces*, pp. 69–72. ACM, 2010. (Cited on page 18)

[YYJ11] S. Yuen, G. Yaoyuneyong, E. Johnson. Augmented reality: An overview and five directions for AR in education. *Journal of Educational Technology Development and Exchange*, 4(1):119–140, 2011. (Cited on page 16)

All links were last followed on September 27, 2014.

**Declaration**

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

place, date, signature