Pick from Here! - An Interactive Mobile Cart using In-Situ Projection for Order Picking

Markus Funk[1], Alireza Sahami Shirazi[2*], Sven Mayer[1], Lars Lischke[1], Albrecht Schmidt[1]

[1]University of Stuttgart (Pfaffenwaldring 5a, 70569 Stuttgart, Germany)
[2]Yahoo Inc. (701 1st Ave, Sunnyvale, CA 94089, USA)
[1]firstname.lastname@vis.uni-stuttgart.de
[2]alireza@yahoo-inc.com

ABSTRACT
Order Picking is not only one of the most important but also most mentally demanding and error-prone tasks in the industry. Both stationary and wearable systems have been introduced to facilitate this task. Existing stationary systems are not scalable because of the high cost and wearable systems have issues being accepted by the workers. In this paper, we introduce a mobile camera-projector cart called OrderPickAR, which combines the benefits of both stationary and mobile systems to support order picking through Augmented Reality. Our system dynamically projects in-situ picking information into the storage system and automatically detects when a picking task is done. In a lab study, we compare our system to existing approaches, i.e, Pick-by-Paper, Pick-by-Voice, and Pick-by-Vision. The results show that using the proposed system, order picking is almost twice as fast as other approaches, the error rate is decreased up to 9 times, and mental demands are reduced up to 50%.

Author Keywords
Industrial Augmented Reality; Order Picking; Projected Displays;

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Humans tend to organize and store items according to the task they are performing [16]. In the industry, organizing, sorting, and providing items is a task called order picking. Thereby, workers need to pick an exact number of parts from a shelf in a warehouse and put them in designated boxes. As this task is very mentally demanding and requires high concentration, various approaches have been proposed to support this task.

*The majority of the work has been conducted while he was a researcher at the University of Stuttgart.

UbiComp '15, September 7–11, 2015, Osaka, Japan.

http://dx.doi.org/10.1145/2750858.2804268

Figure 1. A participant is picking items from a shelf. The projector cart highlights the correct box and provides in-situ information about how many items to pick.

Early systems used Pick-by-Paper (PbP), where the worker can process a paper list step by step. The paper lists the name, the quantity, location, and destination of items which should be picked.

Also Pick-by-Vision (PbVi) systems [25] using head-mounted displays (HMDs) have been proposed to support order picking tasks [26]. Providing different visual feedback on HMDs such as overlaying a shelf with an attention funnel [22] or displaying the layout of the shelf and highlighting the position [10] has been investigated. However, it is reported that users wearing HMDs complain about headaches and having problems to focus [25]. It also limits the worker's field of view [22]. Further, such systems cannot automatically detect whether the correct item is picked. Barcode scanners are used in state-of-the-art systems to solve this problem, e.g., ReadyToShip[1]. Thereby, each picked item has to be scanned explicitly in each step. However, holding the scanner can interfere with the user's picking task and scanning can increase the task completion time (TCT). Fully automated approaches, e.g., Amazon's Kiva system[2] have been developed for automatically storing and picking items. But, such fully automated warehouse systems are very expensive. Another alternative to support order picking is the Pick-by-Light system (e.g. ULMA Pick-to-Light[3]). Thereby, the supporting system is directly

[1]`www.readytoshippicking.com` (last access 03-02-2015)
[2]`www.kivasystems.com` (last access 03-02-2015)
[3]`www.ulmahandling.com/en/picking` (last access 03-02-2015)

integrated into warehouse and does not require the worker to wear any technology. In this approach, each position of a bin in a warehouse is augmented with an LED. The LED indicates where the worker needs to pick the next item from and how many items should be picked from the bin. After items have been picked, the worker has to confirm the pick directly at the bin. However, integrating hardware into every bin in a warehouse is not scalable and is therefore only used in smaller settings. Overall, state of the art approaches have recognized the need for step-aware in-situ support, but require stationary hardware and advancing the steps manually.

In this paper we introduce a mobile order picking system, called *OrderPickAR*, that uses cart-mounted projectors for providing in-situ feedback during an order picking task (see Figure 1). Furthermore, the system is step-aware by using cart-mounted depth cameras for monitoring order picking steps during the task. To the best of the authors' knowledge, the system is the first hands-free order picking system that can automatically and implicitly detect picking steps.

The contribution of this paper is twofold: (1) We present an interactive order picking system that provides projected in-situ feedback for order picking tasks and automatically monitors steps during the tasks, (2) through a user study and comparing the proposed system with Pick-by-Paper, Pick-by-Voice, and Pick-by-Vision, we discuss that the proposed system reduces the mental load and the TCT compared to all other interactive systems. Further our system reduces the error rate (ER) compared to the Pick-by-Voice (PbVo) and the PbVi approaches. Finally, we derive design guidelines for an interactive system for supporting order picking based on our participants qualitative statements.

RELATED WORK

Augmenting the real world with digital instructions using interactive systems has been the subject of various research. In the following we provide an overview in relevant research areas for presenting interactive projected information and supporting workers during order picking.

Projected User Interfaces

Augmenting the physical world with projected content has been around for some years. In 2001, Pinhanez [20] proposed to augment physical objects with digital content by projecting on them using a stationary camera-projector pair. As distortion was a problem, he suggested to correct the projection using a camera to enable a distortion free projection on curved surfaces. Linder et al. [18] are installing a camera-projector pair into an anglepoise lamp. In addition to displaying projected content, their LuminAR prototype is able to identify objects that are placed inside the projection area and augment the objects with information. With the proliferation of Kinect depth-cameras in 2010, sensing touch on projected interfaces became easily possible on arbitrary surfaces [28]. Their algorithm was improved by Hardy et al. [11] by using KD-trees to handle multitouch with 30 frames per seconds.

While the previous projects mainly are designed for interacting with stationary projection, Beardsley et. al [3] use a mobile camera-projector system. They are using their system to align the projection next to unique points in the environment. Thereby, the projection can stay at a defined position even when moving the projector. Further, they were able to use the movement of the hand-held projector as a digital cursor for interaction. Willis et al. [27] scaled using a projector as an input device up to using multiple mobile projections that are interacting with each other. They use an IR channel to communicate between multiple mobile camera-projector pairs. On the other hand, Raskar et al. [21] created a geometrically aware camera-projector system by adding a tilt sensor and creating a 3D-mesh from the camera input. Projected images are then transformed according to the 3D-mesh and corrected to be viewed distortion free even on non-planar surfaces. They further used their system to project onto picking bins, however their focus was mainly on distortion free viewing and combining multiple projectors to enable a projection in 3D space. Schwerdtfeger et al. [24] are using head-mounted and environment-mounted laser projectors to display information in a welding context. Their findings comprise that head-mounted projectors are too heavy to use in long-term tasks e.g. at the workplace. Löchtefeld et al. [19] use a hand-held camera-projector system that can be directed at a shelf to categorize products according to a user's personal profile. They are using an RGB-camera to visually identify objects in the shelf and augment them with information. Harrison et. al [12] targeted a mobile scenario by mounting a camera-projector pair on a user's shoulder. Recently, Winkler et. al [29] proposed a backpack-mounted solution for both projector and depth-sensing camera in their AMP-D project. Especially for interacting with mobile projectors, many areas of application have been suggested. Rukzio et al. [23] provide a comprehensive overview.

Augmented Order Picking

Systems for supporting workers during order picking and finding objects have been the topic of various projects. A strand of work has embedded cameras and projectors to the environment to track objects and provide visual feedback. Butz et al. [5] use a stationary camera-projector system, which is firmly mounted at a room's ceiling. Their system can detect books that are equipped with visual markers automatically and later highlight their position using the projector. Furthermore, Crasto et al. [6] use a foreground detection algorithm to sense changes in a bookshelf. On the other hand, Bannat et al. [1] use a similar camera-projector setup to automatically detect when a worker picks an item from a box during an assembly task. While their work is mainly focusing on a manual assembly workplace, their camera-based detection of a box's position gives the worker great flexibility. Li et al. [17] are using a stationary Kinect together with computer vision algorithms to identify picked objects based on their shape and visual appearance. In their approach, the worker has to explicitly place the object in front of the camera which results in an extra working step and might increase in the TCT.

Another strand of work has used HMDs to deliver feedback directly to users. Reif et al. [22] and Schwerdtfeger et al. [25] use a modification of the attention funnel, proposed by Biocca et al. [4], on HMDs to guide the worker to the next shelving unit. They use an optical tracking system to display the

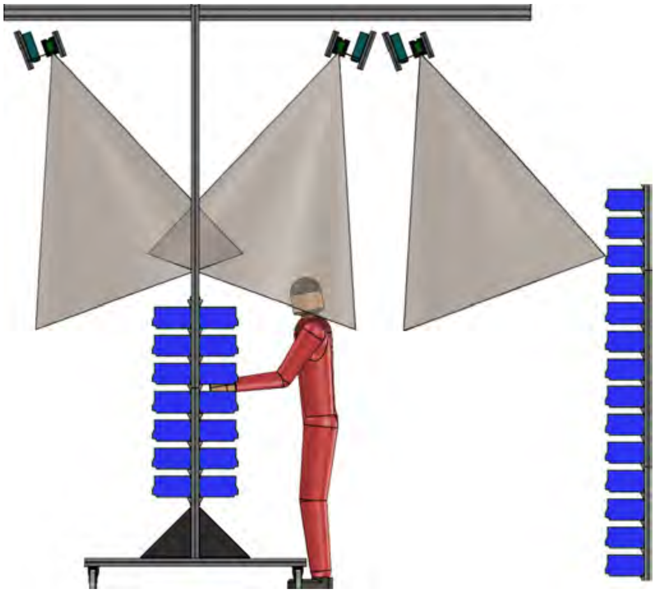

Figure 2. A sketch of our projector cart prototype. Two Kinect-projector pairs are facing the wagon while one Kinect-projector pair is facing the shelves. The projector cart can be moved freely in the isle.

funnel correctly on a HMD. It is reported that the ER can be reduced using the attention funnel. However, the authors also report that the weight of the HMD can disturb the worker, the visual clutter produced by the funnel limits the field of view, and the precise position of the HMD is crucial for the visualization. Furthermore, content that is displayed on a HMD could potentially block the safety-critical real-world view [14]. This makes it difficult to deploy in industrial settings. To overcome the need for a precise location of the HMD, Weaver et al. [26] suggest to display a 2D model of the shelf on the HMD and highlight the box to pick from in the model. Guo et al. [10] compare this approach to a cart-mounted display (CMD), which is displaying the 2D graphical representation and the order picking tasks. In their results, the 2D representation that is displayed in the HMD is significantly faster than the traditional PbP. Furthermore, their analysis also shows that schematic representation of the warehouse on the CMD is also favored compared to a PbP approach regarding ER, TCT, and cognitive load.

Overall, previous work shows that assistance systems can improve the picking performance of users but HMDs have problems being accepted during longtime usage. Instead of the user, we propose augmenting a picking cart with a projector and a depth camera to project in-situ picking instructions and automatically detect when a picking step is performed.

OrderPickAR: AN INTERACTIVE MOBILE CART

We design an interactive mobile cart for order picking tasks, called *OrderPickAR*. The system is a regular order picking cart for a classical man-to-goods system that is extended with a top-mounted beam holding three pairs of camera-projector (see Figure 2). Two pairs are facing the boxes holding the processed orders located at the two sides of the cart. One pair is facing the shelves in the warehouse containing the items

Figure 3. Arrows that are projected when the target compartment is not in the projector's field of view. The arrows are moving together with the cart and showing the shortest path to the target compartment.

that can be picked. The cameras mounted on the cart are depth cameras that monitor the boxes mounted on the cart and the shelves in the warehouse. The projectors are used for providing in-situ feedback by highlighting shelves to pick items from and boxes to store the processed orders.

The field-of-view (FoV) of the cart-facing camera-projector pairs cover all boxes that are mounted on the order picking cart. As they are moved together with the boxes mounted on the cart, the layout of the boxes can be predefined. In the current version, the cart holds 49 boxes on each side (a 7×7 grid) and can store up to 98 orders at the same time. Furthermore, the height of the cart can be adjusted to the warehouse, as we constructed the frame holding the beam to be height-adjustable. In our configuration, the height of the cart was set to 3.38m to perfectly cover the boxes and the height of the shelves.

For building the cart we used aluminum profiles which are typically used in industry. The projectors facing the cart are Acer K335 LED-projectors with 1000 ANSI Lumen. The projector facing the shelves is an Optoma EW610ST DLP projector with 3100 ANSI Lumen. The depth cameras are Kinect for Windows running with a 640×480 resolution. At the bottom of the cart, we installed a PC that runs the pick detection and calculates the projection on both sides of the cart and in the environment. The system is powered through a ceiling-mounted electric cable.

Making the System Step-Aware

OrderPickAR can detect when the worker is performing a pick and can detect when the worker is storing an item from an order in the cart using the top-mounted depth cameras. To achieve this, the system requires knowledge about the 3D model of the warehouse. The 3D model can be specified in the system using a graphical editor (see Figure 4). The graphical editor is developed using Unity3D[4]. Using the editor, the user can position interactive zones, so called trigger spheres, to overlay the compartments of the shelves in the model at the position

[4] `www.unity3d.com` (last access 03-02-2015)

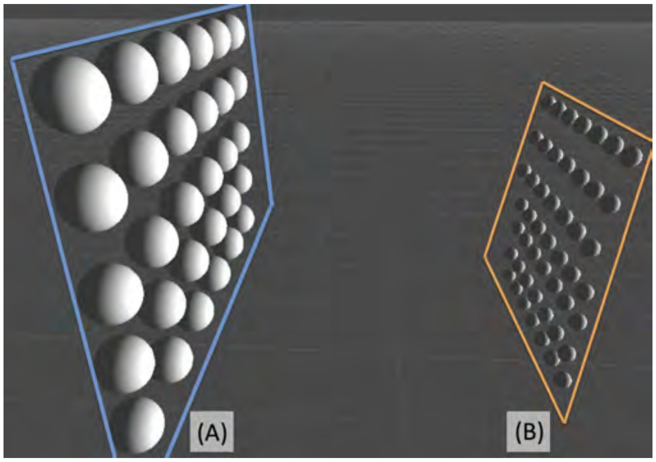

Figure 4. A visual representation of the spheres that trigger a pick or place event. (A) The stationary spheres from the shelves in the warehouse. (B) The mobile spheres belonging to one side of the cart triggering when an item is placed into an order in a box.

Figure 5. The layout of the warehouse that was used in the study. Each compartment is labeled with a compartment number. The warehouse consists of 30 different compartments.

where the compartments are in the physical world. Figure 4 (A) shows the stationary trigger spheres of the environment. The trigger spheres are used to identify the compartments and sense if the user picked an item from the correct compartment. We use OptiTrack[5] motion capturing system to track the position and orientation of the cart in the warehouse. We equipped the warehouse with 17 OptiTrack Flex3 cameras and positioned a marker at the upper frame of the cart. The cameras were positioned throughout the warehouse in a way that for every possible position of the cart, at least 4 cameras were able to track the cart's marker. According to the specification, this allows the system to track the cart's position and orientation within an accuracy of millimeters. The OptiTrack system is connected to a stationary PC that streams the position data to the cart-mounted PC via WiFi. For detecting the actual pick, the top-mounted depth camera facing the warehouse observes if there was a movement in the a trigger sphere. If the change in depth data is beyond a threshold, the sphere triggers that the user picked from the associated compartment. An informal experiment suggested using a threshold of 61% concerning the changed depth pixels for reliably triggering the interaction for our compartments. This value has to be adjusted to the size of the compartment, as the interactive area changes according to the compartment's size.

For detecting when the user places an item in one of the boxes mounted on the cart, the system uses a similar approach. Both cart-facing depth cameras are mounted at the cart's beam in a 90° angle. They are used like a light barrier and can detect when a user is putting an object or a hand into one of the cart's boxes. Figure 4 (B) shows the mobile trigger spheres that belong to the cart. As the position of the trigger spheres move according to the position of the cart, their position inside the 3D space can change. But, the distance and angle between the the depth camera and the cart-mounted boxes always stay the same, as they are firmly mounted on the cart. The algorithm determines spheres triggered and provides the feedback at the corresponding compartment.

Calibration

Before using the system for the first time, the trigger spheres for the environment have to be defined in the underlying 3D model of the warehouse to match the position of the compartments. In our prototype, this is done using Unity3D. This calibration step has to be performed once when deploying the system in a new environment. After this step, the calibration is stored and the system can be used in the environment. The trigger spheres have a unique ID, which are used to identify the picked items. In the current version of the system, we are using a textfile-based approach to load orders to pick. However, this system can be easily integrated into an enterprise resource planning system, which could send and display orders in the moment they are issued by the customer. The trigger spheres representing the boxes in the cart (see Figure 4 B), also have to be defined only once when adjusting the height of the cart to the warehouse. As the camera-projector pair is firmly mounted on the cart's beam, the distance between the pair and the boxes stays the same when moving the cart. Finally, the projector and the camera have to be calibrated. We are doing this using a simple 4 point calibration. Here, the projector displays four targets which are recorded by the camera. The user has to click at the target inside the camera's recorded image. Thereby, the system has a mapping between projector and camera spaces. This procedure has to be done once for each of the three camera-projector pairs.

Displaying Visual Feedback

We deliberately chose the size of the warehouse to be larger than the field of view of the camera-projector pair facing the shelves. When the target compartment is inside the pair's field of view, the target is highlighted with a green light that is directly projected into the shelf (see Figure 1). Inspired by our previous research [9], we designed the visual feedback in a way that the compartment is highlighted by using a simple color-based visualization. To communicate the quantity of

[5] `www.naturalpoint.com/optitrack` (last access 03-02-2015)

(a) (b) (c) (d)

Figure 6. The picking methods that were used in our study. (a) A participant using the Pick-by-Paper list. (b) The Pick-by-Voice approach. (c) Perspective of the participant using Pick-by-Vision with an attention funnel visualization. (d) The projector cart highlights the box to pick from.

items to be picked, the system also projects the information directly into the compartment. In case the target is outside the field of view, the system projects an off-screen visualization (inspired by [2]), which displays green arrows that are pointing towards the target compartment (see Figure 3). When the target compartment is inside the cart's projection range, the arrows slowly disappear and the compartment is illuminated.

The visual in-situ feedback is calculated according to the process order. The order contains the ID of the target compartment's trigger sphere. This information is used to calculate the position of the next target and to show directions towards it. When the cart is in front of the shelf, the system uses the shelf-facing projector to highlight the shelf using a green light. As the depth camera and the projector are calibrated, the depth camera is used to detect when the user picks an item from the shelves. In case the user picks an item from a wrong box, the box is highlighted in red. Using this approach, the system is aware of the user's current working step. After the user picked the order from the shelf, the system uses the projectors facing the cart to highlight the box where the user should put the previously picked item using a green light (Figure 6d). Again, the depth camera is used to check if the user has put the item in the correct box. If the user puts an item into a wrong box, the box is highlighted in red. It should be mentioned that the current version of the system can only detect whether an item was picked from a compartment without knowing the quantity. Orders that require the worker to walk from the compartment to the cart several times can be challenging. To support the worker when an order consists of too many items to carry from the source compartment to the target box in one run, OrderPickAR implements a function that highlights the last order's compartment and target using a yellow light. This is confusing: Thereby, the user can easier find the shelves and finish the order using the yellow light although the feedback was already advanced. Afterwards the user can continue with processing the next order by following the green light. The yellow light is advanced when the following order is completed or when the cart is moved to another location.

USER STUDY

We conducted a user study to evaluate our system. In the following we describe the setting of the study and other picking methods that we derived from previous research. Further, we describe the design, procedure, participants, and the picking task.

Warehouse Layout

For conducting our user study, we designed a warehouse in our research lab consisting of three shelves. Figure 5 shows the layout of the warehouse. The shelves are aligned to form a grid consisting of 5 rows $\times$ 6 columns. Each compartment is labeled with its identifying number. In total, the warehouse is 2.07m high and 3.62m wide. We designed the warehouse in a way that the distance to travel between the picks is minimal. However, the warehouse is approx. twice the size of the shelf-facing camera-projector pair's field of view depending on the distance of the cart to the shelves. The position of the compartments in the warehouse is not ordered according to features of the stored items. Each compartment contains 10 items of the same type. We did not use boxes inside the compartments and stored the items directly in the shelves. As items to pick, we are using 30 different Lego bricks in different shapes and colors. This warehouse layout is a designed to represent a classical picker-to-parts low-level warehouse [7].

Picking Methods

In the following, we describe the picking methods that we used in our evaluation. In addition to our OrderPickAR projector cart system, we considered three other existing approaches: PbP, PbVi, and PbVo.

Pick by Paper (PbP)

The PbP approach (Figure 6a), where a worker gets a paper list containing the picking information, is still used in many warehouses in the industry. As other research [10, 26, 25] uses PbP to compare their system, we include PbP as a baseline in our study. We are using a paper list containing the following information about each picking task: article's description, article's number, quantity to pick, source compartment, and destination box. The user has to find the source compartment in the warehouse, pick the correct quantity, and place the items into the destination box.

Pick by Voice (PbVo)

For the PbVo approach, we recorded audio instructions for all picking tasks. The user can control the audio instruction with the following commands: *next*, *back*, and *replay*. We designed the commands to reflect state of the art systems[6]. The command *next* jumps to the next (or first) instruction

[6] **`http://www.dematic.com/en/Supply-Chain-Solutions/By-Technology/Voice-and-Light-Systems/Pick-to-Voice`** (last access 03-02-2015)

and plays it back. The *back* command jumps to the previous instruction and plays it back. The *replay* command replays the current instruction again without proceeding the instruction. The instructions were played back in a headset (see Figure 6b). As we did not implement a voice input, a wizard of oz issued the playback of the correct picking instruction according to the participants command. The audio instructions contained the information needed for the current picking task, e.g., *"grab 3 items from shelf 02-10 and put them into box 39"*. To make the system more comparable to the other systems used in the study, we decided not to include a *ready* command to confirm the pick.

Pick by Vision (PbVi)

To further compare our projector-cart approach to existing approaches, we implemented a PbVi system using the attention funnel visualization [4] on a Epson Moverio BT-200[7] HMD. This approach is similar to Schwerdtfeger et al. [25]. The attention funnel visualization displays circles towards the compartment to pick from (see Figure 6c). Further it displays the quantity of the items that have to be picked in the bottom left corner of the HMD. For tracking the position and orientation of the HMD, we equipped it with OptiTrack markers at each side. The position and orientation information is calculated at a desktop computer and transmitted to the HMD via WiFi. Further, the HMD is running a 3D visualization of the attention funnel using the Unity3D engine. The program then uses the position and orientation information of the HMD and the position of the target compartment to adjust the visualization. The position and orientation information is received with 100 frames per second. However, due to the limited processing power of the Moverio BT-200, our system was only able to display 9-10 frames per second. The system is implemented in a way that it only renders the most recent position information. Position frames that cannot be processed in time will be dropped.

Method

We designed the study using a 4-level repeated measures design with the guidance system used as the only independent variable. The guidance systems used were: PbVi, PbVo, PbP, and the OrderPickAR projector cart. As dependent variables we measured the error rate (ER), task completion time (TCT), and the NASA-Task Load Index (NASA-TLX) [13] score.

After welcoming the participant and explaining the course of the study, a general introduction about order picking was given. Before each condition, the participant was allowed to perform 3 picking tasks to get familiar with the current guidance system and moving the picking cart. The picking task we are using in the study is the so-called *discrete picking* [8]. This means that the participant takes parts from one compartment and puts it into a single box on the cart afterwards. After placing the picked items the user picks from another compartment.

We considered four picking tasks each consisting of 10 different steps and 30 items to pick. The items were located in different compartments of the previously described warehouse. The participant was told to always move the cart in front of the shelf they need to pick from next. The movement of the cart was needed to simulate a regular sized warehouse. Furthermore, we chose to only use the shelf-facing side of the cart as it provided enough boxes. We designed all picking tasks to consist of the same distance that needs to be walked between the shelves and the cart. Further, the distance that the cart has to be moved is the same in each task. We counterbalanced the order of the conditions and tasks using Balanced Latin Square. The participants were also instructed to carry all items belonging to a step in one single walk from the shelf to the cart and not to split the picking task into multiple walks. As this study focuses on the different types of feedback, we designed the study that all feedback is proceeded by a wizard-of-oz during all conditions. Further, the facilitator counted the ER. After completing the tasks using one condition, the participant was asked to fill in a NASA-TLX [13] questionnaire. We repeated the procedure for all conditions. At the end, we collected additional qualitative feedback.

We instructed all participants to focus on not making any errors during the picking tasks, which is considered the primary goal. Further, we told the participants that a fast picking of the orders is only considered the secondary goal, nevertheless the time to pick the orders is measured during the study.

We recruited 16 participants (4 female, 12 male) via our university's mailing list. The participants were aged from 20 to 43 years ($M = 24.81$, $SD = 5.39$) and were students with various majors and a secretary. None of the participants had experience with order picking. All participants were not familiar with our system or the picking tasks. The study took approximately 60 minutes per participant. The participants were compensated with 5€.

Results

We statistically compared TCT, ER, and NASA-TLX between the guidance systems using a one-way ANOVA test. Mauchly's test showed that the sphericity assumption was violated for TCT($\chi^2(5) = 37.70$, $p < .001$) and ER($\chi^2(5) = 18.16$, $p < .003$). Therefore, we used the Greenhouse-Geisser correction to adjust the degrees of freedom ($\epsilon = .42$ for TCT and $\epsilon = .57$ for ER). Otherwise stated, the Bonferroni correction was used for all the post-hoc tests.

A repeated measures ANOVA showed a statistically significant difference in TCT between the approaches $F(1.287, 19.305) = 93.99$, $p < .001$. The post-hoc tests depicted that the difference between all approaches are significant (all $p < .05$) except between PbVo and PbP ($p = n.s.$). Figure 7a shows the average TCT of all approaches. TCT was fastest using the OrderPickAR system ($M = 3.55$ minutes, $SD = 0.38$) followed by PbVo ($M = 6.81$ minutes, $SD = 1.12$), PbP ($M = 7.11$ minutes, $SD = 1.40$), and PbVi ($M = 15.31$ minutes, $SD = 3.89$).

The statistical analysis also revealed a significant difference in the ER between the approaches $F(1.711, 25.667) = 22.49$, $p < .001$. The post hoc test only showed a significant difference (all $p < .05$) between PbVi and all other approaches. The OrderPickAR projector cart ($M = 1$, $SD = 3.24$) and PbP ($M = 1$, $SD = .96$) had the lowest ER, followed by PbVo

[7] `http://www.epson.com/moverio` (last access 03-02-2015)

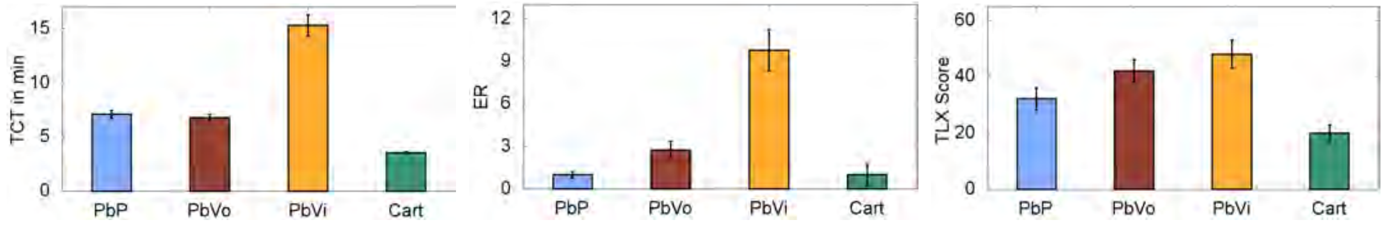

Figure 7. The results of our study: (a) task completion time in minutes, (b) error rate, and (c) the mental load indicated by the NASA-TLX score.

($M = 2.75$, $SD = 2.56$) and PbVi ($M = 9.75$, $SD = 6.07$). The results are shown in Figure 7b.

The analysis reveals a significant effect on NASA-TLX scores between the approaches $F(3, 45) = 11.06$, $p < .001$. The post-hoc test only revealed a significant difference (all $p < .05$) between OrderPickAR and all other approaches (Figure 7c). OrderPickAR had the lowest score ($M = 20.25$, $SD = 12.64$) following by PbP ($M = 32.25$, $SD = 17.48$), PbVo ($M = 42.44$, $SD = 17.30$), and PbVi ($M = 48.81$, $SD = 21.00$).

The qualitative feedback indicated that participants did not find the visualization provided on the HMD helpful. They mentioned that the glasses slightly moved when performing a pick, which caused the funnel to become inaccurate (P10, P14). It was also mentioned *"[I] would not like the head-mounted display when working together with co-workers"* (P11). The participants found the in-situ feedback OrderPickAR provided fast and easy to use (P10). But they sometimes occluded the projection when picking from lower boxes (P7). Overall, participants liked that they have both hands free for performing the picking task in the PbVi, PbVo, and projector cart conditions. They disliked that they have to carry the picking list at all times in the PbP condition.

DISCUSSION

The results suggest that using our system has several advantages. First, TCT is almost 2 times faster than using the PbVo and PbP approaches and even more than 4 times faster than using the PbVi approach. Second, the ER is significantly lower up to 9 times compared to the HMD approach. The difference in number of errors in comparison to the classical paper-based approach was not significant. Interestingly, all errors that were made with the OrderPickAR approach in the user study was made by one single user, who did not pay attention to the displayed number of items to pick. Third, the mental demand during the order picking is more than two times lower than using other interactive approaches. The qualitative feedback also conveys that users find the in-situ feedback projected directly on the shelves and boxes better than using a HMD. The feedback on HMDs may hinder users to communicate with each other. As the HMD slightly moves during a picking task, the visualization on it introduces an offset.

When comparing our results to previous work using the same PbVi representation, our PbVi approach performs worse compared to the respective PbP approach. Schwerdtfeger et al. [25] and Reif et al. [22] report a similar TCT comparing PbP and PbVi, while our PbVi approach differs significantly from our PbP approach. Concerning ER, previous work reported a slightly higher rate when using the PbVi approach compared to PbP [25]. However, our PbVi approach performs significantly worse compared to our PbP approach. This difference might be caused by the used HMD, the Epson Moverio BT-200. Especially, as the qualitative feedback of the participants revealed that although fitting the HMD to each user in a calibration step, the viewing accuracy in the PbVi system was prone to fast head movements. Therefore, the HMD needed to be corrected before continuing. The participants mentioned that the HMD is relatively heavy compared to normal glasses. Also the HMD's rubber parts behind the participants' ears and on the participants' noses sometimes caused the HMD to slip to the bottom of the participants' noses. We think that using a different HMD for the study, the results of the PbVi system would be different.

Limitations

Despite the fact that our system could support multiple users working on a picking task, we limited the task to only support a single user as this is favored by the industry. A multi-user scenario could e.g. use color coding to assign targets to users or use the top-mounted Kinects to track different users. Further, the task used in the evaluation of the system was limited to 10 steps. Tasks with other number of steps may reveal other results. In the study, we were using a Wizard of Oz approach to advance the feedback in case the pick was not registered correctly. E.g. if the user is occluding the Kinect's field of view with the head, the user has to manually advance the feedback in an industry setting. Additionally, the OrderPickAR approach is only able to detect that the user picked from the correct compartment, however it is not able to detect how many items were picked from it. This problem is solved in industry settings by adding a scale into the process. Especially with larger quantity, the order's weight is checked as an additional step. For determining the indoor position of the cart in our proof-of-concept system, we use the OptiTrack motion capturing system. Using this technology is not feasible in larger warehouses because the cost to cover larger areas is too high. To scale this up to larger warehouses and to reduce the costs of the system, an approach using visual markers for determining the position and orientation could be used in future versions of the system. E.g. Kim et al. [15] use an optical marker-based solution for tracking the indoor position of the user. A similar approach could also be used for tracking the picking cart in a large warehouse.

Design Implications

In the following, we provide design guidelines for building interactive assistance systems for order picking based on the qualitative statements of the participants.

Place information directly in the environment. The PbVo and PbP systems require the user to transfer the given information into the physical setting themselves. On the other hand the PbVi and projector-cart systems reduce the cognitive effort by overlaying physical objects with the picking information. Qualitative feedback indicated that the users liked the idea of not having to search for the appropriate box. Therefore, we argue to design order picking systems in a way that the information is directly visible.

Automatically proceed visual feedback. Participants reported that the step-aware processing of the feedback that was proceeded in the PbVi and projector cart condition based on the workers actions was well perceived. On the other hand, participants disliked that they had to issue commands manually in the PbVo approach. Thus, assistance systems should take the context into account and proceed feedback automatically.

Design assistance systems for hand-free usage. Several participants suggested that the PbP approach would interfere with their picking task in case they need both hands to handle the items. Therefore, we argue to design assistance systems for order picking in a way that both hands can be used freely.

Add motivating quantified-self information. During our study, participants occasionally asked how many items were left in the current task and how fast they were. Some participants suggested to have this information always present during the working task.

CONCLUSION

In this paper we presented the design, implementation, and evaluation of a mobile step-aware projector cart for supporting order picking tasks in warehouses. We compared our approach to existing approaches, i.e., Pick-by-Vision, Pick-by-Voice, and Pick-by-Paper. The evaluation shows that the OrderPickAR projector cart is faster than the other approaches and significantly reduces the user's mental load. Additionally, we found that our approach reduces the number of errors compared to a Pick-by-Vision approach. The users find the in-situ feedback provided directly on the warehouse and shelves more helpful than feedback on HMDs or on a picking list. We further provided design guidelines for designing interactive order picking systems based on user's qualitative opinions. As future work, we want to combine our system with visual Augmented Reality markers instead of using the motion capturing system, test our system in larger warehouses, and conduct a long term study. Furthermore, we are planning to assess the possibilities of OrderPickAR in supporting impaired workers during order picking tasks.

ACKNOWLEDGMENTS

This work is funded by the German Federal Ministry for Economic Affairs and Energy in the project motionEAP, grant no. 01MT12021E. We thank Sebastian Pickl and Johannes Herter for their hard work in implementing the software and conducting the user study.

REFERENCES

1. Bannat, A., Gast, J., Rigoll, G., and Wallhoff, F. Event analysis and interpretation of human activity for augmented reality-based assistant systems. In *Intelligent Computer Communication and Processing, 2008. ICCP 2008. 4th International Conference on*, IEEE (2008), 1–8.
2. Baudisch, P., and Rosenholtz, R. Halo: a technique for visualizing off-screen objects. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2003), 481–488.
3. Beardsley, P., Van Baar, J., Raskar, R., and Forlines, C. Interaction using a handheld projector. *Computer Graphics and Applications 25*, 1 (2005), 39–43.
4. Biocca, F., Tang, A., Owen, C., and Xiao, F. Attention funnel: omnidirectional 3d cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM (2006), 1115–1122.
5. Butz, A., Schneider, M., and Spassova, M. Searchlight–a lightweight search function for pervasive environments. In *Pervasive Computing*. Springer, 2004, 351–356.
6. Crasto, D., Kale, A., and Jaynes, C. The smart bookshelf: A study of camera projector scene augmentation of an everyday environment. In *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, vol. 1, IEEE (2005), 218–225.
7. De Koster, R., Le-Duc, T., and Roodbergen, K. J. Design and control of warehouse order picking: A literature review. *European Journal of Operational Research 182*, 2 (2007), 481–501.
8. Eisenstein, D. D. Analysis and optimal design of discrete order picking technologies along a line. *Naval Research Logistics (NRL) 55*, 4 (2008), 350–362.
9. Funk, M., Baechler, A., Baechler, L., Korn, O., Krieger, C., Heidenreich, T., and Schmidt, A. Comparing projected in-situ feedback at the manual assembly workplace with impaired workers. In *Proceedings of the 8th International Conference on PErvasive Technologies Related to Assistive Environments*, ACM (2015).
10. Guo, A., Raghu, S., Xie, X., Ismail, S., Luo, X., Simoneau, J., Gilliland, S., Baumann, H., Southern, C., and Starner, T. 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*, ACM (2014), 71–78.
11. Hardy, J., and Alexander, J. Toolkit support for interactive projected displays. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, ACM (2012), 42.

12. Harrison, C., Benko, H., and Wilson, A. D. Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, ACM (2011), 441–450.
13. Hart, S. G., and Staveland, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology 52* (1988), 139–183.
14. Hua, H., and Gao, C. Design of a bright polarized head-mounted projection display. *Applied optics 46*, 14 (2007), 2600–2610.
15. Kim, J., and Jun, H. Vision-based location positioning using augmented reality for indoor navigation. *Consumer Electronics, IEEE Transactions on 54*, 3 (2008), 954–962.
16. Kirsh, D. The intelligent use of space. *Artificial intelligence 73*, 1 (1995), 31–68.
17. Li, X., Chen, I. Y.-H., Thomas, S., and MacDonald, B. A. Using kinect for monitoring warehouse order picking operations. In *Proceedings of Australasian Conference on Robotics and Automation* (2012), 15.
18. Linder, N., and Maes, P. Luminar: portable robotic augmented reality interface design and prototype. In *Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology*, ACM (2010), 395–396.
19. Löchtefeld, M., Gehring, S., Schöning, J., and Krüger, A. Shelftorchlight: Augmenting a shelf using a camera projector unit. In *Adjunct Proceedings of the Eighth International Conference on Pervasive Computing*, Citeseer (2010), 1–4.
20. Pinhanez, C. The everywhere displays projector: A device to create ubiquitous graphical interfaces. In *Ubicomp 2001: Ubiquitous Computing*, Springer (2001), 315–331.
21. Raskar, R., Van Baar, J., Beardsley, P., Willwacher, T., Rao, S., and Forlines, C. ilamps: geometrically aware and self-configuring projectors. In *ACM SIGGRAPH 2006 Courses*, ACM (2006), 7.
22. Reif, R., Günthner, W. A., Schwerdtfeger, B., and Klinker, G. 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*, ACM (2009), 23–31.
23. Rukzio, E., Holleis, P., and Gellersen, H. Personal projectors for pervasive computing. *IEEE Pervasive Computing 11*, 2 (2012), 30–37.
24. Schwerdtfeger, B., Pustka, D., Hofhauser, A., and Klinker, G. Using laser projectors for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, ACM (2008), 134–137.
25. Schwerdtfeger, B., Reif, R., Gunthner, W. A., Klinker, G., Hamacher, D., Schega, L., Bockelmann, I., Doil, F., and Tumler, J. Pick-by-vision: A first stress test. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, IEEE (2009), 115–124.
26. Weaver, K. A., Baumann, H., Starner, T., Iben, H., and Lawo, M. An empirical task analysis of warehouse order picking using head-mounted displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2010), 1695–1704.
27. Willis, K. D., Poupyrev, I., Hudson, S. E., and Mahler, M. Sidebyside: ad-hoc multi-user interaction with handheld projectors. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, ACM (2011), 431–440.
28. Wilson, A. D. Using a depth camera as a touch sensor. In *ACM international conference on interactive tabletops and surfaces*, ACM (2010), 69–72.
29. Winkler, C., Seifert, J., Dobbelstein, D., and Rukzio, E. Pervasive information through constant personal projection: the ambient mobile pervasive display (amp-d). In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 4117–4126.