ABSTRACT
Order picking is one of the most important process steps in logistics. Because of their flexibility human beings cannot be replaced by machines. But if workers in order picking systems are equipped with a head-mounted display, Augmented Reality can improve the information visualization.

In this paper the development of such a system – called Pick-by-Vision - is presented. The system is evaluated in a user study performed in a real storage environment. Important logistics figures as well as subjective figures were measured. The results show that Pick-by-Vision can improve considerably order picking processes.

Keywords
Augmented Reality, order picking, user studies, 3D geometries.

1 INTRODUCTION
Over the last decade globalization has led to an increasing division of labor along the value creation chain. Companies focus on their core competences and the trend is moving towards outsourcing processes and tasks. Because of this, the domain of logistics and most of all order picking as one of its core functions are becoming more and more important. Order picking is the gathering of goods out of a prepared range of items following some customer orders [VDI94]. As such it is the last process step before the goods are delivered to the customers. Mistakes have a strong influence on the quality of delivery and the relationship between clients and suppliers. Thus, zero defect picking is one important goal. However, this won’t be achieved, no matter which technologies are used [Gud07]. One way to minimize errors is complete process automation. Machines usually cannot replace the human being with his flexibility and fine motor skills [Gud07]. Flexibility is needed because the product range and thus the variety of items increases while, in contrast the size of orders decreases. Human beings are often the best solution for picking in storages. Accordingly, the aim is to support optimally workers by technical devices during their task fulfillment. In this paper we focus on the provision of information for order pickers.

The Department for Materials Handling, Material Flow, Logistics (fml) of the Technische Universität München (TUM) is participating in a research project on visual information assistance – the Augmented Reality (AR) supported order picking system Pick-by-Vision. The order picker wears a head-mounted display (HMD) which visualizes all the required data directly in his field of view. Thus, he does not have to move his head, which leads to a decrease in dead times caused by looking e. g. at a mobile data terminal (MDT). In combination with a voice system for data input, the application is hands-free and the worker can use both hands for his real task.

First this paper gives an overview of the state of the art. After selecting the important hardware components, the test bed at the Department fml for an evaluation of the Pick-by-Vision system is presented. The experimental setup, the execution of the test and its results are presented in the next section. This evaluation is the basis for some optimizations and for the second system which includes a tracking device and which is displayed next. The final section gives an outlook on future fields of application of the Pick-by-Vision technology.
2 STATE OF THE ART
This section introduces the state of the art in order picking and industrial AR.

2.1 Order Picking
The basic conditions in the field of logistics have changed rapidly over recent years. The market demands customized products. Thus, production and logistics systems as well as the workers within these systems have to become more flexible to fulfill the customers’ needs. A lot of different techniques exist for order picking in warehouses [Gud07]. Conventionally, workers execute their orders with paper lists which are intuitive for human beings but laborious to handle. Modern systems go without paper work. They include mobile data entry devices still having a high handling effort but which are usually connected online to the warehouse management system (WMS) processing the data.

Figure 1: Order picking technologies: scanner (1), Pick-by-Light (2) and Pick-by-Voice (3)

In modern warehouses, worker support based on the usual paper lists is often replaced by MDT, Pick-by-Voice (PbV) or Pick-by-Light (PbL) systems (see Figure 1). All these technologies have specific advantages as well as disadvantages. PbV supports the worker by giving him all instructions through the computer’s speech output. Unfortunately, these systems face difficulties in noisy industrial environments. Furthermore, it is questionable whether the warehouseman, as the user of such a system, likes it when he is bossed by a monotone voice the whole day. Compared to voice support systems, PbL offers the worker visual aid by installing small lamps on each storage compartment. PbL systems have the problem that the displays have to be elaborately integrated into the shelf construction and are thus very expensive and inflexible towards rebuilding. PbL is suitable for order picking stations with a high throughput because the display addresses the human optical sense, the favoured sense for the provision of information.

Brynzér and Johannsson evaluated several picking applications in some case studies [Bry95]. Their results are that more logical information systems and more logical storage strategies can save a lot of time. Thus, new technologies of information provision like Pick-by-Vision can improve this logistics process.

2.2 Augmented Reality
Augmented Reality is a technology which can support the human visual sense. Following the definition of Azuma we define AR as a combination of the real and virtual world with 3D registration and interaction in real time [Azu97]. This definition requires a tracking system for positioning virtual objects. For order picking, the support with static data like text information via an HMD could be enough.

AR has many possible fields of application in industrial environments. In the ARVIKA research project, AR applications for development, production, and service were implemented [Fri04]. The first industrial application was the wire bundle assembly project carried out by Boeing in the 1990s [Miz01]. The use of AR for maintenance of a printer was introduced by Feiner et al. [Fei93]. Production planning is another field where AR is used productively in industrial applications [Doi03]. In this paper we focus on supporting the operative staff. One application is an intelligent welding gun for experimental vehicle construction [Ech04]. Tang et al. showed in a user study that AR can improve manual assembly tasks [Tan04]. The subjects were faster and made fewer errors. Training of the operative staff is another field [Wal07].

The biggest potential of AR is the parallelization of information gathering with secondary employment. Thus, dead times can be minimized and the time for information search can be reduced when the data is displayed in the user’s field of view. Dangelmaier et al. compared an AR system with a video see-through (VST) HMD to a the usual paper list [Dan06]. The view of the order picker’s real environment was superimposed with text information, with a map of the storage and the storing compartment was highlighted with an optical frame. With the HMD, a clear learning curve was seen and all the implemented picking errors were found. But some users faced orientation and equilibrium problems due to the VST HMD with its small field of view. In the research project ForLog, an evaluation for information visualization in storing environments was performed [Kli06]. Among other things, an AR system with an optical see-through (OST) HMD was tested. The small field of view and the bad depth
perception of this non-stereoscopic HMD caused problems. Most of the subjects could not clearly identify the real spatial position of the 3D arrows pointing on the storage compartment. The Department fml also evaluated a first prototype of a Pick-by-Vision system [Rei07]. The results and the consequences are mentioned later in this paper in chapter 5.

3 HARDWARE COMPONENTS
An AR system consists of some typical hardware devices: the visualization, the interaction and the tracking system. In this paper only mobile systems with HMDs as visualization medium are considered.

3.1 Head-Mounted Display
The HMD is the most important hardware for Pick-by-Vision because it is the interface between the human and the technical system. Its task is to display the necessary information to the order picker. The visualization of the data is one aspect; the other aspect is ergonomics and the physiological harmlessness of the device. Furthermore, aspects concerning the use in industrial environments should be considered. The most significant requirement is that the worker has to wear the HMD over a shift of eight hours. Because of that, the HMD should be light and ergonomically designed, but also rugged and with an eight-hour battery operation. Another critical point is that the field of view must not be limited due to reasons of labor safety.

In this project more than 40 HMDs where considered and evaluated in terms of suitability but only ten have the potential to be used in storages. VST was a knock-out criterion because a power failure leaves the worker completely blind. Other problems were a too small field of view or the weight. Virtual Retinal Displays (VRD) like the Nomad from Microvision are best suited for order picking applications [Tid95]. They don’t limit the field of view because of their construction based on a semi-permeable mirror. The mirror is used to project the image with a laser beam directly into the eye. The HMD is the decisive factor towards user acceptance of Pick-by-Vision. The Nomad HMD was presented to some order pickers employed by some industrial partners in this project. The feedback was predominantly positive, but the workers didn’t wear the HMD for more than 15 minutes. Because if this, the time dependent effects did not have any influence on their opinion. Most of them can imagine working with an HMD for one day or even longer for further evaluation. But there are also workers who generally dislike the HMD.

3.2 Interaction Device
The second important piece of hardware is the interaction device. Order picking processes vary slightly from one company to the next. But the acknowledgement of the pick, the input of the zero crossing (if there are not enough items in the storage compartment) or the input of errors are necessary for every order picking system. These interactions are implemented in the fml Pick-by-Vision system. Different input devices were evaluated in terms of their suitability. The interaction device should be robust and should not limit the worker’s freedom of movement. Two devices were chosen for Pick-by-Vision. An adjustment knob and speech input can be used best for this application. The degrees of freedom of the adjustment knob (turning left or right and pushing) can be transferred easily to the user interface. Speech input is the most intuitive form of interaction for humans and it is the only technology which allows hands-free interaction.

3.3 Tracking System
Besides the HMD, the tracking system is the most problematic hardware component of a mobile AR application, especially in industrial environments. A lot of different factors, like degrees of freedom, accuracy, resolution, update rate and range characterize tracking systems [Rol01]. If these technical issues are suitable for an industrial application there is still another important factor: the price. On the one hand, an AR system must work robustly and safely in practical operation. On the other hand, it must have better performance than the system used before so that a short return on investments can be achieved.

There are many different functional principles for tracking systems like electromagnetic, inertial, mechanical, optical, radio-based or ultrasonic systems [Rol01]. They all have their specific advantages and disadvantages. Electromagnetic, ultrasonic or radio-based systems have problems in storages because of the high proportion of metallic structures. Magic Map is a WLAN-based technology which is used in storages to locate devices and loading aids [Iba05]. The position can be measured well but this system gives no orientation. Ubesense developed a tracking system based on the ultra-wideband technology which works in metallic environments [Ube08]. In most publications, optical tracking systems are seen as the best choice for use in industrial environments. After an evaluation optical tracking systems were also chosen for Pick-by-Vision. Three different variants are possible. The first one is an inside-out system with a video camera and paper markers mounted in the storage. Paper markers are always crucial if they are used in industrial applications because they can become dirty and perform more poorly. The second idea is an inside-out system with infrared sensors and active LEDs on the ceiling like the Hi-Ball system [Wel01].
But in our application it wasn’t possible to implement LEDs on the ceiling. Thus, we decided to use the third variant, an outside-in system with infrared cameras and with spherical reflective markers mounted on the HMD.

It is questionable if a tracking system is needed for a Pick-by-Vision system. In the evaluation of Tang et al. the tracking system leads to a better performance especially to less errors [Tan04]. However, is the performance in order picking systems with tracking sufficiently better than without so that the additional costs are amortized in an acceptable time? This is one question which should be answered in this project. Therefore, the first Pick-by-Vision system was installed without a tracking system. This system will be explained and evaluated in the next chapters. The system with tracking is introduced in chapter 6.

4 PICK-BY-VISION SYSTEM

The first Pick-by-Vision system was implemented in close collaboration with our industrial partner CIM GmbH which is the developer of the WMS used in this application. The information of the WMS which is normally shown on a fork lift terminal is displayed on the HMD. Thus, the AR application is embedded in the WMS. As mentioned above, the Microvision Nomad VRD is used (see Figure 2).

Figure 2: System used in this set-up and a possible visualization (small picture)

One of the most important things about this system is the Graphical User Interface (GUI), because the virtual information must be displayed at the right time and at the right position. A GUI was implemented following special AR guidelines [Bow05], [Fri04]. For his daily work the order picker needs essential text information about the orders, e.g. storage locations, article number or required quantities. The data input, e.g. the acknowledgement, is done by the order picker with a rudimentary speech input system based on the software from MediaInterface Dresden GmbH. Only a small vocabulary is needed for order picking. In this case the system can be operated with ca. 20 words. The system must function properly independently of the speaker. This is very important for an industrial application with changing workers or for an evaluation with several test persons. Only then every user can work with the system without a special familiarisation.

The workflow of this Pick-by-Vision system includes all important tasks found in the general order picking process. In real order picking applications there are more company specific tasks but for our system the following workflow is sufficient. First, there is a short login dialog. After choosing the next order the user has to take a picking trolley where he sets down the picking container. Then, the system shows him the next storage compartment. During his way to this storage compartment he can already read the picking information (article number and amount). After acknowledging the pick the next order line is displayed. When the last order line is completed, he is told to go to the delivery station to finish the order. Every command is confirmed with the same simple speech input – except for the case where the picking amount also has to be acknowledged by repeating the amount. Because of the small number of speech commands the user becomes familiar with the system in a little while.

5 EVALUATION

A first simple Pick-by-Vision system was evaluated in 2007 [Rei07]. The users’ acceptance was high but the measured logistic operating figures were worse when compared to the paper list. Several factors were responsible for this. The speech input system was not well implemented, the GUI could use improvement and the test storage at the Department fiml was too small. After improving the system a new evaluation was carried out in a larger storage.

5.1 Experimental Setup

In this experimental series the Pick-by-Vision system was compared to a usual paper list in a compartment shelving system hosted by our industrial partner Kühne + Nagel (AG & Co.) KG in their distribution centre in Langenbach. The storage consists of eight shelves with four aisles with more than 600 stock locations (see Figure 3).

Figure 3: The storage for the experimental setup

Each subject had to finish 14 orders using both techniques. The orders had between two and six
order lines with one to six items each. Altogether there were 52 order lines with 125 items. The items were boxes in different sizes and with different weight. Other items were booklets or sweets which were also the reward for the test persons. The orders were picked in the same sequence with each technique. This means that each subject started with order 1 and finished with order 14. For every order the WMS optimized the route through the storage.

16 subjects took part in this test series. Most of them were male (13), the average age was 27.6 years (between 20 and 52 years, standard deviation 8.13). Among them were students or researchers, but also non academic people like skilled workers. Six had experience with 3D visualization, e.g. from computer games. Five were familiar with order picking processes. The data for this analysis was collected in personal questionnaires.

5.2 Null Hypotheses

During the test series two very important logistic operating figures were measured: order picking time and errors. Order picking time is important for the throughput time of the orders in the storage. It is a part of the reaction time between the order of the costumer and the delivery to him. This time becomes shorter and shorter and is an essential factor of success for a company. In addition, it can be used to calculate the order picking performance which is the average number of picked order lines per hour. Picking errors can have a big effect when they are not recognized before shipping. They can result in high contract penalties. The picking errors are translated into an error rate. This represents the amount of errors within all picked order lines.

For these figures some null hypotheses can be introduced comparing Pick-by-Vision to a paper list based system. The first null hypothesis is that the picking time $t$ for both techniques is equal:

$$ H_{0.1}: t_{\text{Pick-by-Vision}} = t_{\text{paper list}} \quad (1) $$

For Pick-by-Vision little training is needed. So the order picking times for Pick-by-Vision can be different between the subjects who started with Pick-by-Vision and who started with the paper list. Thus, the second null hypothesis is that the times $t_{\text{start}}$ are equal and do not depend on the starting technology:

$$ H_{0.2}: t_{\text{start Pick-by-Vision}} = t_{\text{start paper list}} \quad (2) $$

For the error rates $f$ between both techniques it is expected that they are equal. This is the third null hypothesis:

$$ H_{0.3}: f_{\text{Pick-by-Vision}} = f_{\text{paper list}} \quad (3) $$

5.3 Analysis of the test series

The measurement of the logistic operating figures is one side, their interpretation is another. Identifying a difference between Pick-by-Vision and the paper list for only one certain value in this test series does not necessarily imply a universal validity of this result. First, descriptive values like the mean value or the standard derivation are calculated. Based on these results the null hypotheses are proven. For all analysis in this paper the level of significance is 5%.

1.1.1 Picking Time

![Figure 4: mean values, maximum and minimum of the order picking times over all 16 subjects](image)

There is only a small difference between the mean values of the order picking times. With Pick-by-Vision the subjects were about one minute (4%) faster than with the paper list (see Figure 4). In both tests the Grubbs’ Test for outliers shows no outliers with a confidence level of 99%. Both samples are very homogenous and are normally distributed (Kolmogorov-Smirnov-Test). This is the condition for the significance test. A t-Test for paired dependent samples is used because the number of values for both samples is equal but the variance is different (Pick-by-Vision: 4.97, paper list 3.08). The difference between the order picking times is not significant and the null hypothesis $H_{0.1}$ cannot be discarded. Remarkable is that the statistical spread is bigger with Pick-by-Vision. The difference between the slowest and the fastest subject is 18 minutes (ca. 47%) whereas it is only 12 minutes (35%) with the paper list. With a special test the significance can be proven [Bor05]. The difference is significant. Thus,
the null hypothesis $H_{0,1}$ can be discarded. Order picking with Pick-by-Vision is significantly faster.

1.1.2 Learning Effects
There are always some learning effects, e.g. concerning the layout of the storage, the workflow or the look of the articles. To minimise these effects the technique with which the subjects had to start was randomized. Hence, eight subjects started with Pick-by-Vision and eight with the paper list. We can notice an interesting effect. When the picking times with Pick-by-Vision are compared, there is a difference of six minutes (ca. 19%). So the subjects were noticeably faster with Pick-by-Vision when they had picked with the paper list before (see Figure 5). This effect is checked for significance. Both samples are normally distributed (Kolmogorov-Smirnov-Test). A t-Test for paired dependent samples is used again because the number of values for both samples is equal but the variance is different (start Pick-by-Vision: 3.90, start paper list 2.23). The test shows a significant difference between the order picking times. Thus, $H_{0,2}$ is discarded. The explanation for this effect is that the subjects work more confidently with Pick-by-Vision if they know the storage and the workflow. So they can concentrate on the new AR technology.

![Figure 5: Boxplot of the order picking time with Pick-by-Vision](image)

1.1.3 Picking Errors
Among other things the error rate depends on the order picking technology. For a paper list it is normally 0.35%, for Pick-by-Light 0.40% or Pick-by-Voice 0.08% [Ten04]. This means e.g. for an error rate of 0.40% that four order items within 1,000 are faulty. There are different types of errors, e.g. a wrong item was picked or the amount is incorrect. Even one error within 1,000 is usually not acceptable to the customers, because each mistake can lead to halting the production line.

In this test series the error rate for the paper list is seven times higher than for Pick-by-Vision (see Figure 6). With Pick-by-Vision only one error was made for each 1,904 picked order lines. Significance is questionable because of such a small number of errors in the whole evaluation. Both samples are not normally distributed. Error rates are subject to the Poisson distribution because the error count can only take integer, non-negative values and picking errors are a rare event. Thus, the significance test was made with the Mann-Whitney-U-Test. The test shows no significant difference although the difference seems to be big enough. The null hypothesis $H_{0,3}$ cannot be discarded. The reasons are that there were only 16 subjects and they made too few errors.

![Figure 6: Mean values, maximum and minimum of the order picking times over all 16 subjects](image)

1.1.4 Questionnaires
Besides the logistic operating figures the psychological factors motivation, usability, impression and the cognitive load were measured with questionnaires. The subjects accepted the system very well and the subjective load is lower than with the paper list. The distinct usability and the low cognitive load lead to a very high motivation to work with Pick-by-Vision. The motivation is the major difference when compared to the paper list whereas the other examined factors showed no clear differences. There were also free questions where the subjects could state their opinion in their own words. On the one hand, some had problems with the HMD because they could read worse due to changing brightness or the shift of the Nomad VRD mounted on a usual baseball cap. Most disliked the monotone speech input and that they had no overview of the size of an order (foremost the amount of order lines). On the other hand, the subjects liked that they could work hands-free and that the information was displayed clear and in their field of view. Another advantage is the acknowledgment of the pick with the right amount. This is the main reason for the low error rate. Altogether the result is very positive but the subjects wore the HMD only between 30 and 45
minutes. Therefore, it is hard to make predictions concerning full-time use.

6 ADDING A TRACKING SYSTEM

The second functional model of the Pick-by-Vision system is developed independently from the first one together with the Fachgebiet für Augmented Reality (FAR) of the TUM. For this system, the same HMD is used but the GUI is different. Besides text information, pictures of the articles or 3D geometries can also be visualized. One important point is to find the right amount of displayed information because it should not occlude too much of the real environment due to labor safety.

![Figure 7: Pick-by-Vision including an infrared tracking system with targets fixed on the HMD](image)

The first step for this system was selecting the tracking system. As mentioned in chapter 3.3 an infrared tracking system was chosen. In the AR Lab of the FAR four ARTrack cameras from the A.R.T. GmbH are installed and cover an area with two shelves of one meter in length. This setup was used to develop the visualization. For the small storage at the Department fml with two aisles of four meters in length at least eight of these cameras were needed. Because of a low budget we decided to use the prototype ARLiveCam from our industrial partner metaio GmbH in this application. The two infrared cameras have a wider aperture angle. Thus, hanging above the shelves they cover the whole storage (see Figure 7). With the tracking system it is possible to display 3D information in correct spatial position. These geometries are used for wayfinding and for marking the storage compartment. Different geometries like arrows, boxes or tunnels were tested. The combination of a tunnel with a box seems to be the best solution. This is a further development of the attention funnel [Bio06]. The tunnel shows the user in which direction he should look. If his view is towards the right shelf the tunnel becomes transparent and at its end the user can see a frame around the storing compartment (see Figure 8). In this system an adjustment knob is used for interaction because the speech recognition in the other functional model does not work satisfactorily.

![Figure 8: Frame and tunnel becoming transparent if the right storage compartment](image)

A first evaluation was made [Sch08]. Several visualizations were tested at the FAR’s AR Lab considering the error rate. The result was that the subject made no errors regardless of the chosen visualization. The next step is that the system must run stably with the new tracking system at the storage of the Department fml. Afterwards, it will be compared to a paper list or a PbV system. One of the most interesting questions will be if the tracking system brings any further advantages for the performance of order picking processes.

7 CONCLUSION

Our results underline the potentials of Pick-by-Vision. Our evaluations show that the users are faster and make fewer errors. But not only logistics operating figures were considered. The user acceptance is high, resulting in a steep training curve. But there are still some problems. The biggest obstacle for porting such systems from the research stage into practical applications is the hardware components, especially the HMD and the tracking system. However, there is a continuous further development of these components because the gaming industry is slowly discovering AR and HMDs will soon be a part of everyday life within mobile multimedia applications. Therefore, it seems probably that HMDs will be used in industrial applications within the next five years.

8 ACKNOWLEDGEMENTS

The Department fml would like to thank their industry partners for their support with the development and the evaluation of the above-mentioned demonstrators; especially CIM GmbH, metaio GmbH and Kühne + Nagel (AG & Co) KG. A special thanks to the FAR of the TUM, especially to Björn Schwerdtfeger for his work on the 3D
visualization. Many students were involved in the project. Two should be mentioned here: Xueming Pan and Thomas Schmidt. Finally, the Department thanks all participants who took part in the different kinds of experiments and gave us feedback for optimizing the demonstrators and the German Federal Ministry of Economics and Technology which supports this project.

9 REFERENCES


