Using Intelligent Natural User Interfaces to Support Sales Conversations

Sven Gehring, Markus Löchtefeld, Florian Daiber, Matthias Böhmer, Antonio Krüger
Deutsches Forschungszentrum für Künstliche Intelligenz
Saarland University, Campus D3 2, Saarbrücken, Germany
{sven.gehring, markus.loechtefeld, florian.daiber, matthias.boehmer, krueger}@dfki.de

ABSTRACT

During sales conversations, gestures and mimics are of high importance to communicate information about a product. One prominent example for such sales gestures is the meat and cheese counter, which is one of the remaining spots in supermarkets where sales persons interact with customers. Interactions at such counters in supermarkets normally follow a simple protocol. The customer points at an item of choice. The employee takes out the item and, in most of the cases the product needs to be cut to fit the amount the customer wants to buy. Often it is ambiguous about what specific product the customer and the employees are talking about. Up to now, there are just a few efforts in HCI research to enrich communication at the point of sale. In this paper we report and analyze one scenario in which an intelligent natural user interface can support communication between customer and employee in a sales conversation. Furthermore, we report on our prototype that is able to track pointing gestures by using a depth camera and to display information about items pointed at.

Author Keywords
Gestural Interaction, Sales Conversation, Pointing.

ACM Classification Keywords
H.5.2 [Information Interfaces And Presentation]: User Interfaces - Input devices and strategies.

General Terms
Human Factors; Design

INTRODUCTION

The most crucial factor for today's retailers is to deliver true customer satisfaction and achieve superior competitive advantages despite the small margins that they have to face. For most customers the experience of shopping has become more important and the demand for exact product information has drastically increased e.g. due to food scandals. Today's customers tend to be enlightened customers. In addition, in some cases the intention for shopping is simply different: People not only go shopping when they need something; from time to time they consider shopping as a popular leisure time activity (e.g., shopping with the whole family on weekends). During sales conversations, gestures and mimics are of high importance to communicate information about a product. This especially becomes clear when considering fresh food counters. These counters are served by employees and can contain up to 500 different products. Customers often have an increased trust towards the employees compared to technology. Therefore it is of high importance to deliver exact and interesting information at such counters. The sales process at such a counter is generally a repetitive negotiation on product and amount that is supported by pointing gestures of employee and customer. Due to the physical nature of the counter that serves as a spatial separation between the customer and the products, the pointing gestures are often ambiguous. From the employee's perspective, it often remains unclear which product the customer is pointing at. The high density of items in the counter even intensifies the ambiguity. Hence, sales conversations need several iterations of pointing and confirming to disambiguate the meant product. This leads to longer service times and to a frustration of the customer. In this paper, we present an intelligent natural user interface to support the negotiation on a product during a sales conversation at a fresh food counter. A depth camera above the counter infers which products the customers are pointing at. The product is displayed to the employee on the display of the digital scale that is usually only used to weigh out and

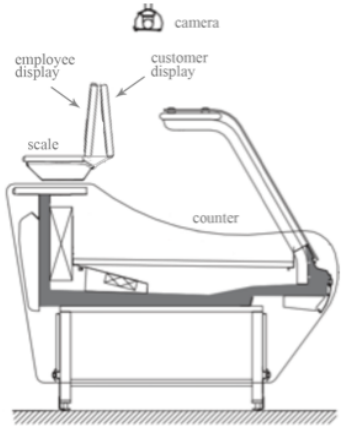

Figure 1: The application setup with the counter, a scale with two displays and a depth camera above the counter.

IUI'12, February 14–17, 2012, Lisbon, Portugal.

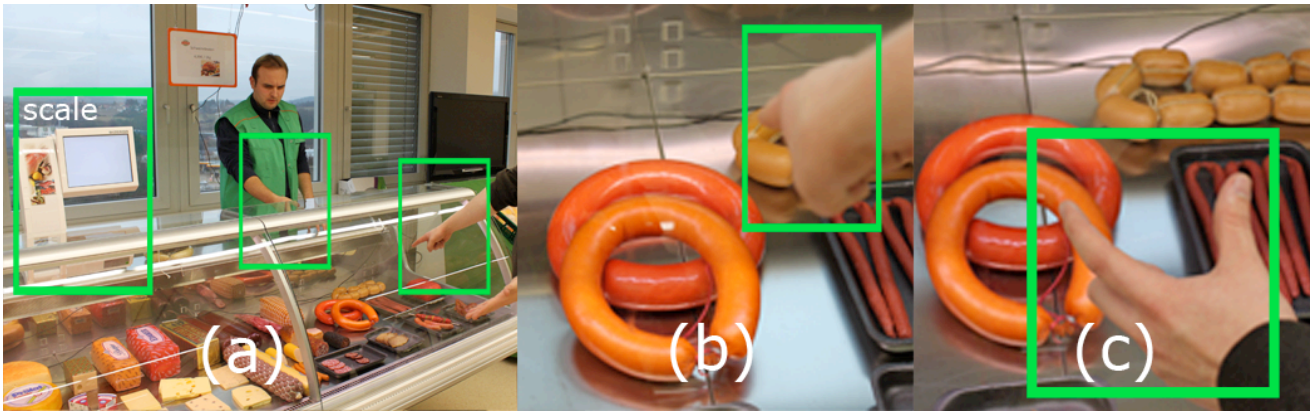

Figure 2: Pointing gestures during a sales conversation at a meat and cheese counter. The counter and the scale of our prototype implementation can be seen in (a). In (b) and (c), a customer is performing gestures above a product in the counter. In (c) a customer indicates the desired size of a certain product is shown.

label the products. Furthermore, additional information (e.g., about the origin of the product) can easily be given to even untrained employees to optimize the sales talk.

This paper is structured as follows: First, we report on common interaction problems at fresh food counters. Second, we briefly describe related work in this area. Based on this, we propose an interaction concept. Afterwards, we depict the implementation of our system. The paper concludes with an outlook and ideas for future work and use cases of such systems.

INFORMAL FIELD STUDY

In order to identify the current interaction protocol and communication problems at meat and cheese counters, we observed the meat counter in a supermarket with approximately 12,000 clients per day for one day. We continuously queued at the counter and observed regular customers during sales conversations. We focused on how customers express their desired products and how they negotiate on the desired quantities with the employees. The observers were standing in front of the meat counter and monitoring the sales conversations between customers and employees. They recorded how the customers expressed their preferences and how customers and employees negotiated on product and quantity (e.g. how many iterations the negotiation process takes) as well as how employees gave additional information on the products. During the day, approximately 350 customers visited the counter and bought 3.6 products in average (derived from the sales data of the whole day). From the observations, we obtained the following basic interaction protocol for a sale conversation at a meat and cheese counter:

1) When a customer is at his turn, he normally points at a product in the counter to indicate his preference to the employee behind the counter. 2) The employee points at the desired product and asks the customers if this certain product is the preferred one. Due to the spatial separation between customer, product and employee, this initial step often needs several iterations until customer and employee agree on a product. 3) After customer and employee agreed on the product, the employee normally asks how much the customer wants of the product. 3.1) The customer either tells the employee an amount or 3.2) he often performs a gesture to indicate the amount he wants. This can be seen in Figure 2 (c). This gesture is than mimicked by the employee to confirm the amount of the product. For example the employee points with a knife on the product to indicate how much he will cut off. The customer either agrees or corrects the employee by stating that he wants some more or less of the product. 4) After customer and employee agreed on the amount, the desired amount is cut off, packed and handed to the customer.

In addition to this basic protocol, when the costumer points to the product, we also made a second observation: On the wall behind meat and cheese counters, there are often advertisements for certain products in the counter. Customers not always point at the product in the counter but also on an advertisement asking what the advertised product is or stating that they want to buy the displayed product. To support and simplify the described interactions and to reduce the service times and the iterations we propose an application concept based on the interaction protocol we found.

RELATED WORK

The general concept of this work is inspired by the “Put-that-there” prototype by Bolt in 1980 [1]. Instead of interacting with a computer system in our case the user or customer interacts with the human employee. We track the gestures to support the discussion. More recently, Holzapfel et al. [2] focus on supporting the natural interaction and communication with a human robot using gesture tracking. In addition, they present architecture for the fusion of multimodal input streams for natural interaction. Oviatt et al. [4] provide a good overview on work in this area. In the retail domain, Wasinger et al. [5] describe a user interface that integrates multimodal input on a handheld device with

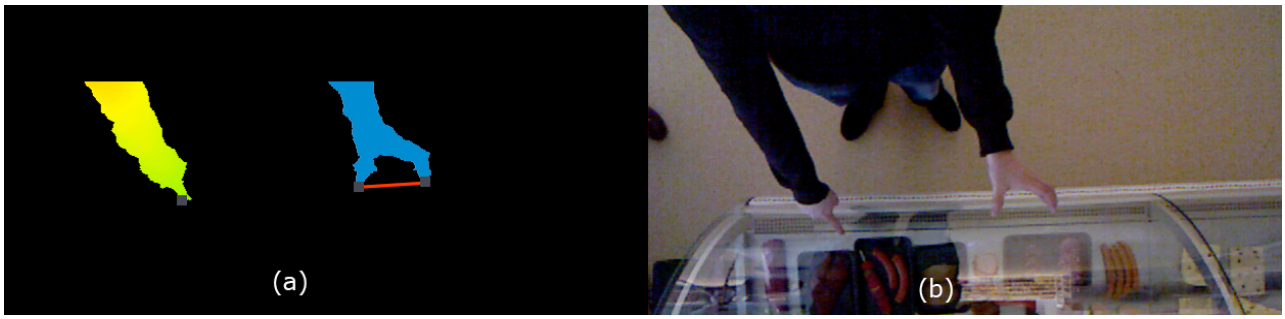

Figure 3: The filtered depth image (a) with detection of *pointing* and *size* gestures and the original image (b).

external gestures performed with products in a shelf. Spassova et al. [6] further evolve this approach. They introduced the concept of product-associated displays as a way of providing visual feedback to users interacting with physical objects in an instrumented environment. Similarly, Newcomb et al. [7] pre-sent design guidelines for a PDA based shopping assistant in a grocery store. They also state that one important aspect that has to be regarded during the design process is that shoppers often use their hands to touch and or point to the products – something we have tried to incorporate into our system and that is also a key aspect in the work of Wasinger [5]. A detailed review on literature on how gestures are used in communications is provided by Kendon [9].

Our work differs from the related work, because to the best of our knowledge, it is the first approach to use gesture recognition to improve sales conversations (i.e. human to human communication) in the retail domain.

APPLICATION SETUP

Our observations revealed that a vast amount of gesturing is already happening during sales conversations and that the protocols are already established. Therefore, instead of introducing new gestures that people would have had to learn and to not interrupt these protocols, we make use of the natural gestures that already exist. So, having the gestures already well established, we try to reduce misunderstanding and misinterpretation of the gestures to ease the sales conversation.

We implemented an application that recognizes the pointing gestures of customers and shopping assistants by using a Microsoft Kinect depth camera. It can support customers during the sales conversation by recognizing the pointing gesture and providing feedback. As can be seen in Figure 1, the scale on the counter consists of a scale unit and two displays. One display of this dual-display scale is faced towards the employee to enable weighting of the product and attaching of price labels. The second display is used to display additional information to the costumer. The layout of the displays is the same in our use-case but they are used to display additional information about the product the customer is pointing at. We provide the customer with information on origin and ingredients of the product, e.g. the fat content of a cheese. On the display facing the employee, even more information such as wine recommendations, recipes or details of the production process can be revealed. This allows the employee to give further advice and negotiation about the product. We believe that this can enhance sales conversations and with that increase customer satisfaction. An example of the setup is shown in Figure 2.

In addition, the system also allows the customer to point at an advertisement on the wall behind the counter as described in the observation section. The employee takes out the product from the presentation area of the counter and places it on a dedicated spot on the counter for cutting and preparing the products. Again the depth camera tracks the gestures of the customer, as described above in the observation section in point 3.2. This amount is again roughly highlighted on scale display on the employee's side.

IMPLEMENTATION

To test the technical feasibility of this application concept we implemented a prototype using a depth camera, a real meat and cheese counter and a dual-display scale. The setup of the scenario can bee seen in Figure 1. As proposed by Wilson [8], we use the depth image that we obtain from the camera with the OpenKinect[1] library to track the users arm. After a background subtraction only the pointing part of the user arm is left, this can be seen in Figure 3. With the depth information of the part of the image displaying the users arm, we determine the height of the pointing. By calculating the difference in the depth image between the beginning of the user's pointing arm and the pointing finger – i.e. so to say the top of the pointing gesture – we calculated the declination of the arm. By calculating the angle between the user's arm and the alignment of the image, we can then estimate the position in the counter the user is pointing at. The system automatically looks up in a database which product belongs to the estimated position. If the user points at an advertisement instead of the counter, the procedure is analogous. This pointing information is transferred via a network connection to the scale. The main restriction of our

[1] http://openkinect.org/wiki/Main_Page

system is the resolution of the Kinect. At a distance of two meters the resolution is approximately 2cm per pixel, this restrains the maximum amount of items in the shelf drastically. Furthermore we have no possibility to determine if the customer thinks that the pointing gesture is a projection of his arm or the ray casted by his hand and his eyes. This can cause severe inaccuracies and confuse the customer.

DISCUSSION

Besides the basic requirements that we extracted from the observations of customers' current interaction at fresh-food counters, we have also presented our system to various users and experts (e.g. on retail trade fairs like the EuroShop 2011[2]) and collected preliminary feedback to inform the further design of the system. We got positive feedback from customers supporting our initial observations. They agreed, that our system supports the problem of disambiguating the meant item and the desired amount. Further, they saw a benefit in the possibility of "browsing through the cheese counter while waiting in a cue" when no sales person is available. However, first tests with customers revealed that people point differently (e.g. with index finger or whole hand). So far, our system is designed for the use with stretched index finger for pointing. One solution for that might be gesture recognition that is based on a user model. A small projected curser as a visual cue might be another solution. Such a curser can easily guide the user while pointing at objects. Further, this will also support the sales person, since he will have the information what the customer is pointing at directly projected on the piece of cheese or meat. Our users also came up with the point, that our system could be useful when they buy food in foreign countries, where they do not speak the language.

We also presented our prototype implementation to experts, i.e. sales personal and retail managers. They liked the idea of unobtrusively supporting the sales talk at cheese and meat counters. Further, they suggested implementing a similar system for self-service frozen food where no sales people are available, e.g. convenience food. However, they also raised two problems that might arise when deploying such a system in the wild. On the one hand, they concerned the density of items, i.e. pieces of different sorts of cheese or meat might be a problem that might lead to inaccuracy. However, we think that we can improve the accuracy of the system by using a more precise depth camera. On the other hand, they mentioned that different pieces of cheese and meat are "traveling" during the day, i.e. when sales persons take out pieces of the counter and put it back, the size of the item changes and it might be placed a little bit apart from its previous position. However, we discussed with these experts and came to the concept of using the technique also as an input device: since the sales people know the type of cheese and meat they can point at items and update the information in the database.

[2] http://www.euroshop.de/

CONCLUSION

In this paper, we described a first concept on how to integrate intelligent natural user interfaces to ease the sales conversations at meat and cheese counters in a retail store by utilizing a depth camera and two displays attached to a scale to provide further information to both customer and employee. Our work makes the following contributions: First we identified the general communication problems that occur in such sales conversations and report on an informal field study we did in a supermarket. Second, we proposed a prototypical system to solve these communication problems with pointing gesture recognition. In addition, we reported on preliminary feedback of customers and sales people to inform a next iteration of the system.

We plan to carry out a user study in a retail store to evaluate our ideas and see how our prototype can improve the sale conversation. We want to compare the time a sales conversation takes with and without the system. Furthermore, we want to measure the effect of the proposed system on the retail sales figures at meat and cheese counters. For the future, we want to further refine the system based on the customers' and experts' feedback and investigate how we can integrate further HCI methods to support sales conversations between customer and shop assistant in retail stores.

REFERENCES

1. Bolt, R. A. Put-that-there: Voice and gesture at the graphics interface. Proc. SIGGRAPH'80, ACM (1980).
2. Holzapfel, H., Nickel, K. and Stiefelhagen, R. Implementation and evaluation of a constrained-based multimodal fusion systems for speech and 3D pointing gestures. Proc. ICMI'04, ACM (2004).
3. Oviatt, S. L. Multimodal Interfaces. In The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Lawrence Erlbaum Associates (2003).
4. Lalanne, D., Kohlas, J., Dumas, B. and Oviatt, S. Multimodal Interfaces: A Survey of Principles, Models and Frameworks. In Human Machine Interaction, Springer Lecture Notes in Computer Science (2009).
5. Wasinger, R., Krüger, A. and Jacobs, O. Integrating Intra and Extra Gestures into a Mobile and Multimodal Shopping Assistant. In Pervasive Computing, Springer Lecture Notes in Computer Science (2005).
6. Spassova, L., Wasinger, R., Baus, J., Krüger, A. Product Associated Displays in a Shopping Scenario. Proc. ISMAR'05. IEEE (2005).
7. Newcomb, E., Pashley, T. and Stasko, J. Mobile Computing in the Retail Area. Proc. CHI'03, ACM (2003).
8. Wilson, A. Using a Depth Camera as a Touch Sensor. Proc. ITS'10, ACM (2010).
9. Kendon, A. Do Gestures Communicate? In Research on Language & Social Interaction, 1532-7973, Volume 27, Issue 3, Pages 175 – 200 (1994).