Tracking Pointing Gestures to Support Sales Conversations

Markus Löchtefeld
German Research Center for Artificial Intelligence
Saarbrücken, Germany
markus.loechtefeld @dfki.de

Florian Daiber
German Research Center for Artificial Intelligence
Saarbrücken, Germany
florian.daiber@dfki.de

Sven Gehring
German Research Center for Artificial Intelligence
Saarbrücken, Germany
sven.gehring@dfki.de

Antonio Krüger
German Research Center for Artificial Intelligence
Saarbrücken, Germany
krueger@dfki.de

Johannes Schöning
German Research Center for Artificial Intelligence
Saarbrücken, Germany
schoening@dfki.de

Copyright is held by the author/owner(s).

Abstract
Meat and cheese counters are among the most important spots in a supermarket where employees interact with customers. Currently a lot of communication problems occur at these counters due to the distance between the customer and the service staff and the separation by the counter. Often it is ambiguous about what 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 the interactions and communications at a meat counter and introduce a prototype to ease the communication between customer and employees to higher customer satisfaction and decrease service times using a Microsoft Kinect.

Introduction
Customers are able to buy more and more fresh food in supermarkets. Today's efficient logistic chains make it possible to sell more than 500 different sorts of meat, cheese and fish in one single market. These products are available at special meat and cheese counters. Due to hygiene requirements, meat and cheese counters are served by employees to make sure that customers do not touch meat or cheese. The products are presented in counters that have a glass front and are accessible from the back by the service personal. In a sales conversation, pointing gestures of employee and

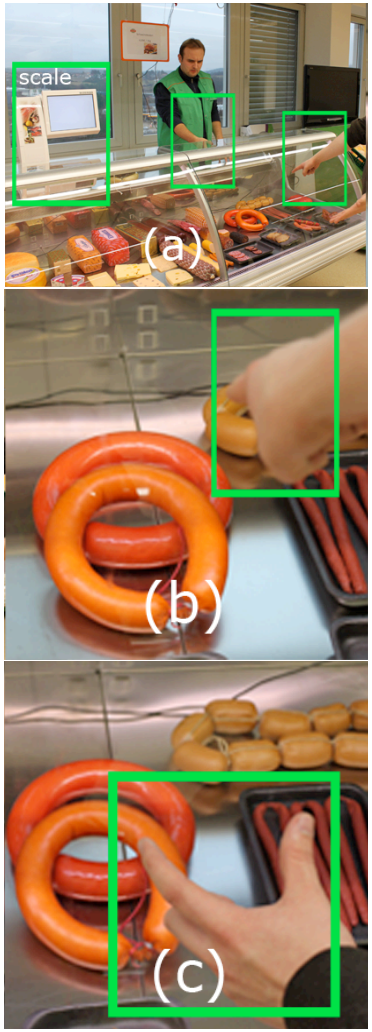

Figure 1: Pointing gestures during a sales conversation at a meat and cheese counter. The counter and the scale can be seen in (a). In (b) and (c), a customer is performing gestures above a product in the counter. In (c) a customer indicates the desired size of a certain product is shown.

customer generally support the negotiation on the desired items and the associated amounts. This negotiation can be problematic due to the physical nature of the counter that serves as a spatial separation between the customer and the products. From the employee's perspective, it remains often unclear to which products the customer is pointing. The different perspectives from each side of the counter, the pointing distance and the high density of items in the counter make pointing gestures ambiguous and therefore hard to understand. Common sales conversations on a product therefore need several iterations of pointing and confirming. This leads to longer service times and respectively to longer waiting times for customers and can lead to customer frustration. In this paper a prototype is presented to support the sales conversation between customers employees. A camera above the counter provides information about the products the customers are pointing at. This information is transferred to the employee using a display at a scale that is standing on the counter to weigh out and to label the products.

Informal Field Study

In order to identify the current interaction protocol and communication problems at meat and cheese counters we observed the meat counter of a supermarket with approximately 12000 clients per day for one day in Saarbrücken, Germany. Approximately 350 customers visit the counter and bought 3.6 products in average. We focused on how customers express their desires 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 customer and employee negotiated on product and quantity (e.g. how many iterations the negotiation process takes). From these 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 and indicates 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 1 (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 user.

In addition to this basic protocol, when the costumer points to the product, we also made a second observation: 5.) 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

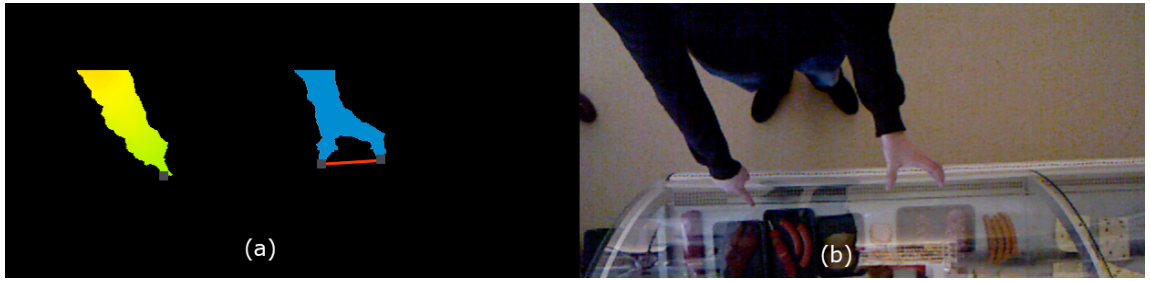

Figure 2: The filtered depth image (a) with detection of *touch* and *size* gestures and the original image (b).

asking what the advertised product is or stating that they want to buy the displayed product. To support and simplify this described interactions and to reduce the service times and the iterations we propose an application concept.

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 customers interacts with the human employee. More recently, Holzapfel et al. [2] again focus on supporting the natural interaction and communication with a human robot using gesture tracking. Oviatt and Dumas et al. provide a good overview on work in this area [4]. In the retail domain, Wasinger et al. describe in [5] a user interface that integrates multimodal input on a handheld device with external gestures performed with products in a shelf. Newcomb et al. [6] presented 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. A detail review of literature on how gestures are used in communications is provided by Kendon [8].

Our work differs from the related work, because it is to our knowledge the first approach to use gesture recognition to improve sales conversation - human to human communication - in the retail domain.

Application Setup

We implemented an application that recognizes the pointing gestures of customer and shopping assistant by using a Microsoft Kinect[1] depth camera. It supports customers during the sales conversation by recognizing the pointing gesture and providing feedback. The scale on the counter consists of a scale unit and two displays. One display 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 (e.g. advertisements). An example of the setup is shown in Figure 1. 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 with point 5). 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 as described above in the observation section in point 3.2 tracks the gestures of the customer. This amount is again roughly highlighted on scale display on the employee's side.

Implementation

The setup of the scenario can bee seen in Figure 3. As can be seen in Figure 3, the Kinect is mounted above the counter. As proposed by Wilson [8], we use the depth image that we obtain from the camera with the freenect[2] library to track the users arm. This can be seen in Figure 2. With the depth information of the part

[1] http://www.xbox.com/kinect

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

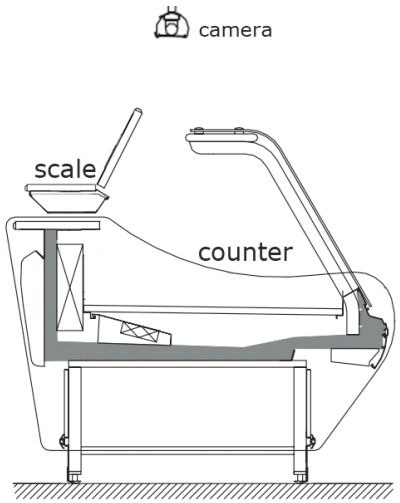

Figure 3: The figure shows the illustrated application setup with the counter, the scale with display for employee and customer and the Kinect mounted above the counter.

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 users pointing arm and the pointing finger - so to say the top of the pointing gesture - we calculated the declination of the arm. By calculating the angle between the users arm and the alignment of the image, we can then estimate the position in the counter where the user is pointing. 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.

Conclusion & Future Work

In this paper, we described a first concept on how to integrate HCI methods to ease the sales conversations at meat and cheese counters in a retail store by utilizing a depth camera and a display attached to a scale to provide further information to both customer and employee. This work in progress has the following contribution. 1) 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. 2) In addition, we proposed a prototypical system to solve these communication problems with pointing gesture recognition.

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 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. *Proceedings ofSIGGRAPH'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. *Proceedings of 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] Newcomb, E., Pashley, T. and Stasko, J. Mobile Computing in the Retail Area. *Proceedings of CHI'03*, ACM (2003).

[7] Wilson, A. Using a Depth Camera as a Touch Sensor. *Proceedings of ITS'10*, ACM (2010).

[8] Kendon, A. Do Gestures Communicate? In *Research on Language & Social Interaction*, 1532-7973, Volume 27, Issue 3, Pages 175 – 200 (1994)