Visual Customer Behavior Analysis at the Point of Sale

Johannes Kröckel, Freimut Bodendorf
Department of Information Systems
University of Erlangen-Nuremberg
Lange Gasse 20, Nuremberg, Germany
{johannes.kroeckel, freimut.bodendorf}@wiso.uni-erlangen.de

***Abstract*— In times of low-cost broadband and mobile internet flat rates advantages such as continuous availability and a large range of products enable online retailers to mature. Since competition grows, web shop owners have to adapt to their customers to remain competitive. Therefore, they record and analyze data about all kinds of customer activities on the web shop to derive optimization potentials and to offer customized services. Stationary retailers, which are regularly in direct competition to web shops lack these possibilities. They are limited to data derived from enterprise resource planning systems, checkout systems or loyalty cards. Hence, the behavior of customers at the point of sale cannot be considered, yet. To address this lack of information an approach is presented that applies video and infrared cameras to record and analyze customer movements and activities at the point of sale. The approach aims at extracting valuable information for the management and sales staff of stationary retailers. Based on the information, services are developed to support decisions regarding the store structure, product range and customer approach.**

Customer tracking and tracing, Customer behavior, Retailer Support.

I. INTRODUCTION

Low prices, short delivery periods and 24/7 availability are only three reasons that help web shops to extend their customer base at the expense of stationary retailers. Besides, there are few reasons left for customers to buy goods like books or music in stationary stores since a physical experience is not required. Consequently, stationary shop operators have to come up with sophisticated, individual approaches for attracting and retaining customers. For this, knowledge about the customers as well as their on-site buying behavior is required [1].

While click paths, bounce rates and page impressions as well as time spent on websites are common key figures for internet shops stationary retailers lack any sources of information about their customers' behavior [2][3][4]. Data recorded from electronic checkout counters or merchandise planning and control systems are not sufficient to reveal individual customer behavior.

Companies like Envirosell or Shopperception try to overcome this information shortage by data gathered from manually conducted observations [5]. However, these strategies are designed for a limited time only. Continuous observation over a longer period of time like weeks or months would be very expensive due to the required human resources. Besides that, an objective documentation of results by the observing persons cannot be guaranteed. To enable quick reaction to contemporary customer behavior a continuous automated monitoring similar to the one for web shops is needed.

Movements and activities can be considered as real world equivalents of clicks. Movements describe how customers walk through the shopping environment and provide useful data on their speed, regions of interest and behavior towards other people in the surrounding area. Activities in terms of interactions with products enable to gain information about viewed or purchased products and therefore about customers' buying behavior.

The extraction of movement and activity data using various kinds of sensors is strongly discussed in fields like computer vision and data mining (see Section II). However, data is rarely used for gaining information for retail managers and sales personnel. Besides that, tracking of movements and activities is mainly conducted with expensive high tech equipment, which enables scientific applications but rarely allows feasible solutions for real world applications. Although companies like Visapix and Vitracom offer tracking systems for sensing and analyzing position data their software solutions require specific hardware components. The additional hardware expenses often exceed retailer's budgets. Moreover, these software solutions provide limited data analysis. Therefore, a cost-efficient and practical approach is required.

For the proposed concept existing algorithms are combined and work with low-cost sensors. The extracted data is analyzed in a way to allow retailers to improve their retail environment. Thus, they can manage better to retain existing and attract new customers.

To gain this information, first, raw data about the movements and activities need to be recorded. Therefore, Section III.A describes the extraction of walking paths using surveillance cameras and methods from the field of video mining. Then, Section III.B gives an overview of capturing and extracting activities using the Microsoft Kinect Controller and data mining methods.

Analyses of the derived raw data are described in the succeeding sections IV and V. In Section IV data are processed to gain an overview of the comprehensive behavior of customers at the point of sale. Subsequently, in Section V the behavior of individual customers is considered.

2012, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

II. Related Work

In recent years, a variety of sensor-based solutions has been developed to record context information [6][7][8]. Especially the extraction and analysis of location data is a frequently discussed topic. The approaches mainly use gathered location data in the field of ubiquitous and mobile computing [9][10][11]. Therefore, the authors apply cell phone compatible positioning technologies like GSM or GPS for outdoor location tracking. The presented approach requires position determination inside a store. That means, it has to be more accurate than GSM and, in contrast to GPS, available indoors [12]. Indoor position localization is among others achieved using technologies such as RFID, Bluetooth and WIFI [13]. However, these technologies lack of accurate results in indoor environments (minimum deviation ~1.0-2.0m). For example, that makes it impossible to capture product group related behavior. Beyond that, transmitters or receivers need to be carried around by the persons to be tracked, which might influence their behavior. Therefore, the presented approach applies surveillance cameras, infrared cameras and algorithms from the fields of video mining to extract movement and activity data in a retail environment without bothering customers.

The extraction of position data by using image sensors is among others described by Wang et al. [11], Gavrila [14] and Perl [15]. However, none of the mentioned approaches considers the conditions at the point of sale. Besides, the use of the gathered data is discussed very little.

The analysis of movement data is among others described by Andrienko et al. [16] as well as Ashbrook and Starner [9]. In their works waypoints are assigned to well-known points of interest such as buildings or places. Based on the aggregation further movements are predicted. Gutjahr [10] extended this approach to include other sources for position data. Because the structure of a shopping environment changes continuously, it makes little sense to highlight static objects as characteristic points of interest. Rather, useful information for retailers is desired that can be extracted without further knowledge of the environment.

Shortly after its introduction the Microsoft Kinect, which was originally conceived as a controller for the Microsoft Xbox 360 game console has been used in various fields of application and especially for research purposes. The easy to use gesture recognition was applied for different purposes such as innovative user interfaces or robot steering [17][18][19]. However, using the Kinect for point of sale data collection is not considered yet.

Furthermore, the combined analysis of both movement and activity data to reveal customer behavior information for retailers has not been discussed yet.

III. Data Collection

The section comprises a brief overview of the approaches being used to extract movement and activity data from raw footage. Section III.A presents two methods using image data captured from an aerial and a lateral perspective. Subsequently, the methods are weighed against each other. Section III.B addresses the extraction of activity data based on infrared sensors implemented by the Microsoft Kinect Controller.

A. Movement Data

The overall concept presented in this work is inspired by Fillbrandt [20]. In his doctoral thesis he introduces an approach for a modular single or multi camera system tracking human movements in a well-known environment. Therefore, persons are detected on images by a set of computer vision algorithms. Afterwards, location estimation is executed. Finally, the single location data of a person are connected resulting in a trajectory.

In this work two approaches are presented to capture movement data. The lateral approach tracks customers by using cameras with a lateral point of view. The aerial approach applies cameras mounted on the ceiling.

1) Lateral Tracking

The lateral approach for customer tracking uses cameras mounted at the upper end of a corridor, which enables the observation of an entire corridor area. Footage is recorded by network cameras that enable real-time applications as well as subsequent analyses.

For person detection the histogram of oriented gradients algorithm proposed by Dalal and Triggs [21] is applied. The approach is suitable for the detection of people on images. First, the images are converted into gray scale. After that, they are transformed into gradient maps. Then, small pixel areas are analyzed regarding their one-dimensional gradient direction. The gradient maps of a variety of images containing and lacking persons are used to train a support vector machine to extract distinctive features [22] (see Figure 1a).

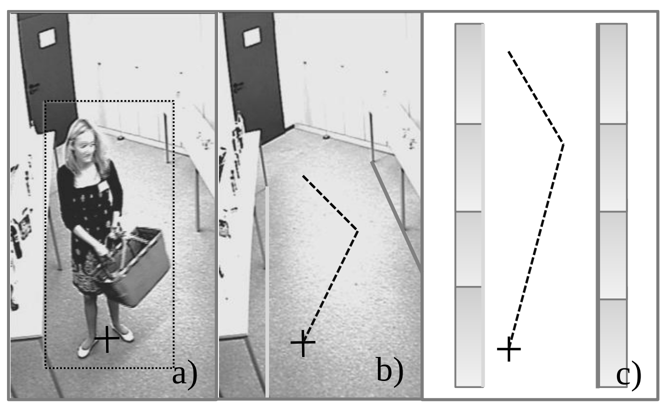

Figure 1. Lateral person tracking

Due to the wide angle distortion effect of the camera's lens especially the locations of persons being further away from the camera are perspectively distorted. That means they cannot be used for true to scale calculations yet. In consequence a perspective transformation by calculating a 3x3 warp matrix based on four source and four destination points is executed. The points have to mark equal positions on the image and on a true to scale map to calculate the factor of distortion.

2012, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

The approach is evaluated using a test environment comprising one corridor delimited by two shelves. The corridor has a width of 2.0m and a length of 4.0m, which compares with a corridor of a typical small store. Raw data (e.g., 20,000 frames) are recorded for 30 different people viewing and buying products from the shelves while traversing the corridor. A maximum of three people is staying in the corridor at the same time. The results reveal issues that are caused by occultation and minor contrast between the people and the background. As a result the approach has a sensitivity of ~61%. The standard deviation between actual and measured position is 0.17m.

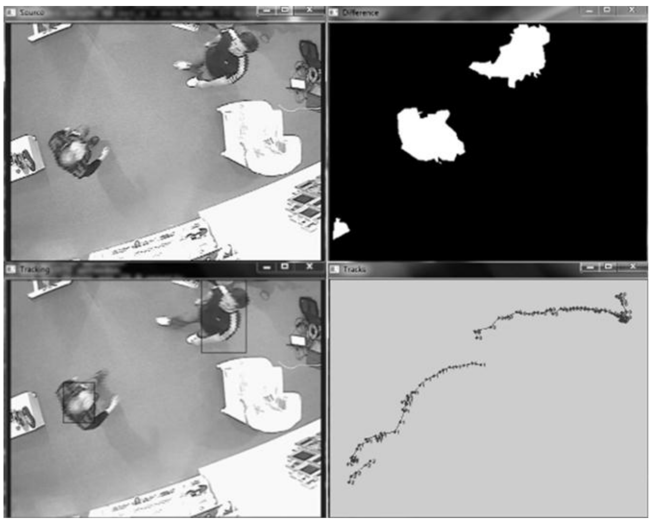

Figure 2. Aerial person tracking

2) Aerial Tracking

By using aerial mounted cameras the size of the observed area depends on the camera's focal distance and altitude. For the detection and tracking of persons within a dedicated area the following set of algorithms is applied. First, background differencing among others described by [23] and Yoshida [24] is used for object detection in single frames being captured by a camera. For this, a reference image is needed, which shows the captured areas without any objects. Comparing the reference image with the actual considered frame excludes all similarities between the two pictures. Differences are highlighted (see Figure 2. , upper right area). After that, image noise is reduced. Eliminating objects that are smaller than the average human shape reveals all objects that could possibly be considered as persons. This step is mostly accomplished by using a template or contour matching algorithm as described by Hu [25] or Zhang and Burkhardt [26]. However, this is not feasible for the presented approach. Contour or template matching algorithms are not able to detect human shapes with high reliability as a result of the varying distance and view of the camera. Besides that, people carrying bags or driving shopping carts as well as disabled people using wheel chairs would not be recognized correctly by the algorithm. Therefore, the detected shapes are filtered by a minimum surface threshold. This leads to significantly better results, i.e., persons can be recognized correctly in most cases.

Subsequently, the continuously adaptive mean shift (camshift) algorithm presented by Bradski [27] is applied for tracking detected persons. The algorithm is based on the mean-shift algorithm originally introduced by Fukunaga and Hostetler [28] and was originally invented for face tracking. Thenceforth, it has been applied for a great variety of tracking purposes.

The mean-shift algorithm is used to track motions of objects by iteratively computing the center of mass of the HUV (hue, saturation, value) vectors within a defined window [29]. For every frame of a video stream the centers of mass are calculated and then defined as new centers of the corresponding windows (see Figure 3). By connecting subsequently occurring centers of windows a trajectory of the movement is obtained. Defining windows as smallest rectangle areas covering shapes of persons extracted by the background differencing approach enables to apply this concept for person tracking purposes.

While the mean-shift algorithm considers windows of static size, the camshift implementation adapts the window size dynamically. This is of great importance for the presented application because persons moving away from or to the center of the observed area occur in different sizes. Using the mean-shift algorithm would lead to an increasing amount of vectors from areas around the considered person. If the amount of these vectors becomes too high, the scope on the person will be lost and errors occur.

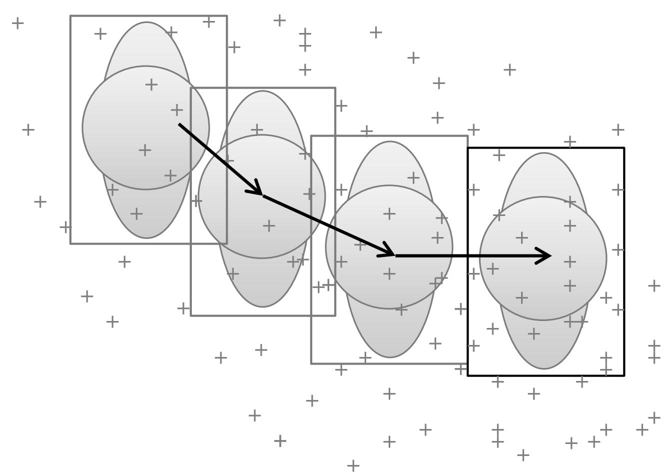

Figure 3. Mean-shift: window shifts

To achieve better results, especially for crowded places the good features to track algorithm by Shi and Tomasi [30] and the optical flow algorithm by Lucas and Kanade [31] are applied as a backup strategy. The good features to track algorithm uses corner detection to find pixels, which differ from those in their surrounding area. Subsequently, the optical flow algorithm tries to find these pixels in the

2012, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

following frame within the surrounding area of their original location on the image.

The combination with a color constancy algorithm (e.g., Barrera et al. [32]) or spatio-temporal rules enables to track persons across several cameras and therefore several corridors. This implies that persons leaving one camera area to another one have to be handed over while crossing an overlapping area (see Figure 4).

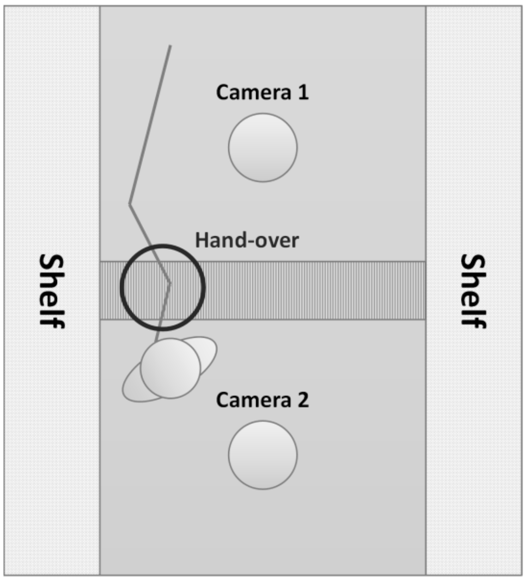

Figure 4. Camera hand-over

An evaluation study was performed for the same test environment that was used for the lateral approach. The camera is placed 3.0m aboveground. The sample footage consists of 16,000 frames showing 30 different people with a maximum of three people walking through the observed area at the same time. Lossless tracking is obtained for ~82% of the observed walks. The average deviation between the real and the automatically determined position is 0.11m.

3) Discussion

The evaluation results reveal that the aerial approach is clearly more accurate and robust. While the lateral approach struggles to overcome contrast and occultation issues the aerial approach doesn't show such problems. Apart from that, a careful evaluation of parameters for the aerial approach is mandatory to gain the described results. This is not necessary for the lateral approach since they are mainly chosen by the algorithms themselves.

Although the delineated tests included a maximum of three people at the same time both approaches are able to handle more people at the same time. Later tests in a grocery showed similar results for transition areas with up to eight persons.

When it comes to expenditures the lateral approach might be preferred by retail managers since the approach is able to use existing surveillance cameras without large-scale alternations. Therefore, expenses for new cameras are reduced to a minimum. Beyond that, the lateral approach regularly requires more cameras that are solely mounted for tracking purposes.

Nevertheless, since the aerial approach provides a better overall reliability its results will be used for the further analyses. Besides, since most other methods like WIFI and RFID lack an inch-perfect accuracy as well as a high reliability especially the aerial approach is considered as a reasonable alternative [13].

B. Activity Data

For the extraction of customer activities the Microsoft Kinect Controller is applied. The controller unifies low costs and high reliability and is therefore widely spread. It allows the tracking of parts of the body and limbs by using an infrared emitter for projecting a distributed grid of 10,000 to 20,000 individual infrared light spots into the physical space. Based on the distortions between the field captured by the infrared camera and a field in empty space the underlying system is able to extract objects. Since the camera features a resolution of 640x480 pixels, the result is an interpolated three-dimensional depth map. By using the Microsoft Software Development Kit positions of limbs of individual persons can be identified.

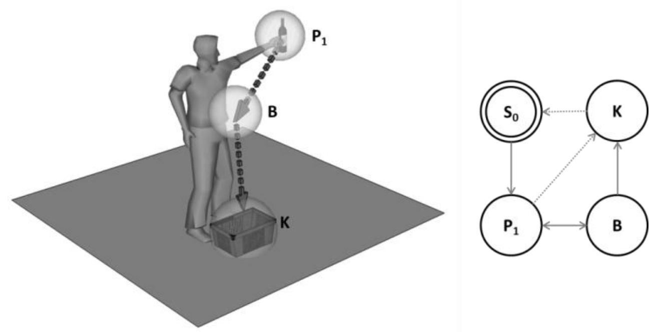

Figure 5. Activity recording

Prior to detecting activities the respective activities have to be determined. To cover the most common interactions of customers with products at the point of sale, this approach is limited to the activities “beholding a product” and “putting a product in the shopping cart”. This reduces the amount of limbs, which have to be considered, to a customer’s arms and the position of his body related to the arms. Since detecting gestures based on temporal and spatial movements is disproportionately complex, following Hong et al. [33] the detection task is reduced to limb movements between predefined states, e.g., three-dimensional areas. The states are represented as a finite state machine.

Figure 5 shows an example of a typical motion sequence. The customer moves an arm from a neutral position to a product. The gesture is recorded as "customer reaches for product 1" and the state change from S_0 to P_1. Then, he moves his hand in a pre-defined area in front of his body. The system recognizes a transition from state P_1 to B and, therefore, the gesture "customer moves hand to body". The

2012, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

sequence of both gestures is assigned to the activity "customer beholds product P1".

IV. Customer Stream Analysis

Two groups of stakeholders are differentiated that require information about customers at the point of sale. Services derived from the customer stream analysis are dedicated to the retail management. They provide a more abstract overview of the situation at the point of sale, like information about the quality of a store's structure or the range of products. However, single customer analyses are conducted to provide customer related information to the sales stuff. They require detailed information about individual customers or customer groups, e.g., to coordinate an optimized customer approach.

To analyze customers' stream behavior the DBSCAN algorithm is used to extract regions of interest, i.e., areas that are most interesting to customers. The method was chosen because of its comparably small consumption of resources and its ability to accurately distinguish between high and low density areas, e.g., in contrast to the popular regions algorithm by Giannotti et al. [34]. Movements between areas are modeled as a Markov chain showing transition probabilities and therefore the most likely paths between the regions of interest.

A. DBSCAN

The 'density-based spatial clustering of applications with noise' algorithm originally proposed by Ester et al. [35] was developed to distinguish between clusters and noise in spatial databases.

Figure 6. DBSCAN cluster center search

Clusters are defined as areas with a considerable higher density than outside of the cluster. To distinguish clusters from noise the following steps have to be accomplished. First, an arbitrary point p is selected. Then, all points that can be reached from p are retrieved. If p turns out to be a core point of a cluster a new cluster is formed. Limitations are made regarding the minimum points (minPts) to be reached by p as well as the maximum distance ε between p and the considered neighboring points (see Figure 6). If one of the constraints is not met no new cluster is formed and another randomly chosen point is considered.

The overall datasets of all trajectories extracted by the movement tracking approach are analyzed by the DBSCAN algorithm using a minimum threshold (minPts) of five points and a maximum real world distance (ε) of 0.02 m. The analysis reveals an amount of 551 clusters.

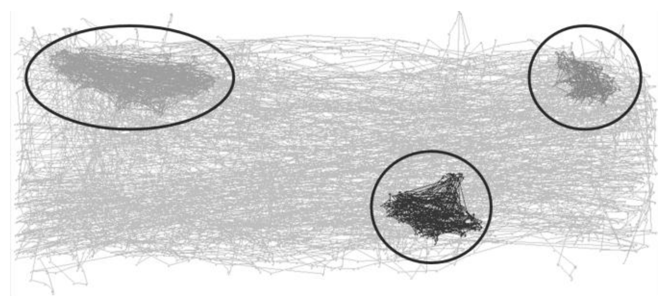

Figure 7. Clusters revealed by DBSCAN algorithm

For further processing all clusters including points from less than 60% of the trajectories are eliminated since they cover positions of too less customers. This step reveals three clusters comprising points of between 60.0% and 90.5% of the trajectories within the test environment (see Figure 7). The areas covered by the clusters are considered as hotspots that are significantly higher visited than other areas of the test retail environment.

B. Markov Chains

A Markov chain comprises states of a system as well as transition probabilities between them [36]. A transition probability is defined as the probability of a system's change from one state to another one. For a first order Markov chain it is only based on the current state. In the presented approach probabilities describe the chances of movements between two clusters. Recursive transitions are neglected because for the presented approach only the succession of movements between different states, i.e., clusters is relevant. That means a transition between two hotspot clusters exists when two temporally succeeding points of a customer trajectory belong to two different clusters. The points do not have to be temporally adjacent points in the database but all of the intermediate points must not be part of another hotspot.

Regarding the movements between clusters, the datasets resulting from the computer vision algorithms described in Section III.A.2 are taken into account. Points that are not part of one of the three considered clusters are ignored.

2012, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

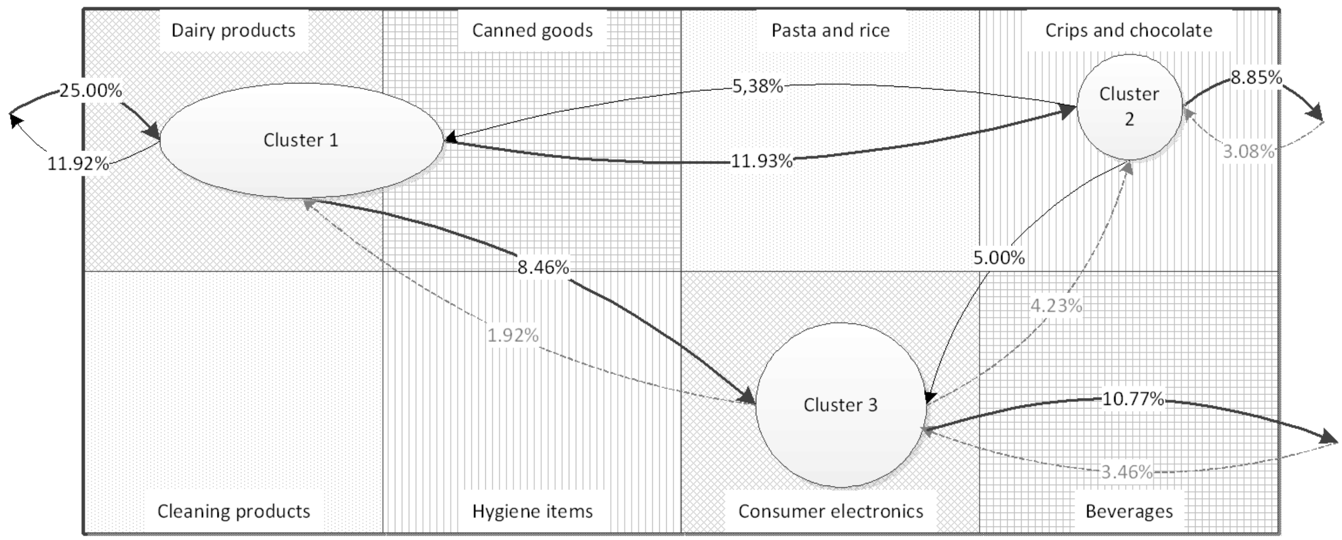

Figure 8. Prototypical retail environment – Scenario 1

C. Results

Considering the clusters and the movements between them allows a closer look on how customers act within a retail store. As an use case two shopping scenarios within the prototypical retail environment are analyzed. The test environment comprises eight product categories (see Figure 8).

The first scenario describes a regular product setup without any advertisements and signs and therefore is used to observe the regular behavior of customers. The second one is based on the findings of the first scenario and includes advertisements for selected products (see Figure 9). Both of the scenarios are compared eventually.

For the first scenario three clusters exceeding the 60% threshold are found. One of them covers the area with shelves containing dairy products. The second, smaller one is located near shelves with crisps ad chocolate. The third one covers the area in front of the shelves with consumer electronics. Figure 8 shows these three clusters as well as the transitions between them. The percentage indicates the proportion of transitions to the total number of transitions being extracted from the customer movements. Transition paths below the limit of a 5% share are greyed out to achieve a clear visualization.

For the given scenario the majority of customers enter the corridor from the left side heading to the first hotspot (dairy products). Afterwards they are more likely moving on to the second one (crisps and chocolate). Then, either they go back to the area of cluster 1 (dairy products) or go on to cluster 3 (consumer electronics).

After that, the customers are most likely leaving the observed area. Besides showing hotspots within the retail environment the graph of Figure 8 also reveals typical paths customers use to move through the store. Looking at visited products it is apparent that products located on the lower left (cleaning products and hygiene items) are less considered. Therefore, advertisements in frequently attended areas are used to call attention for these products.

This idea is seized for the second scenario (see Figure 9). The prototypical retail store is extended by two promotional signs for cleaning and hygiene products. This leads to notable changes of the customers' behavior. While the first scenario leads to three hotspots the second one includes four hotspots. An additional hotspot covers the area between cleaning and hygiene products.

Considering the transitions customers still most likely enter the observed retail environment from the left side attending the area near dairy products first. Afterwards, they are moving on either to crisps and chocolate or to the area in front of the shelves containing consumer electronics.

While most of the consumers move from crisps and chocolate back to the area of dairy products there is also a notable percentage of customers walking to an area in front of cleaning and hygiene products. This could mean that the promotion campaign was successful.

The visualization enables retail managers to get an overview of the movements at the point of sale. Thus, the behavior of customers is monitored and changes are revealed. Besides adding new advertisements existing ones can be evaluated regarding their effectiveness. The same applies to the structure of the environment itself. If the structure is adapted to change the customer flow the movement behavior can be evaluated eventually. If the results don't correspond to the expectation the store might be adjusted iteratively.

2012, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

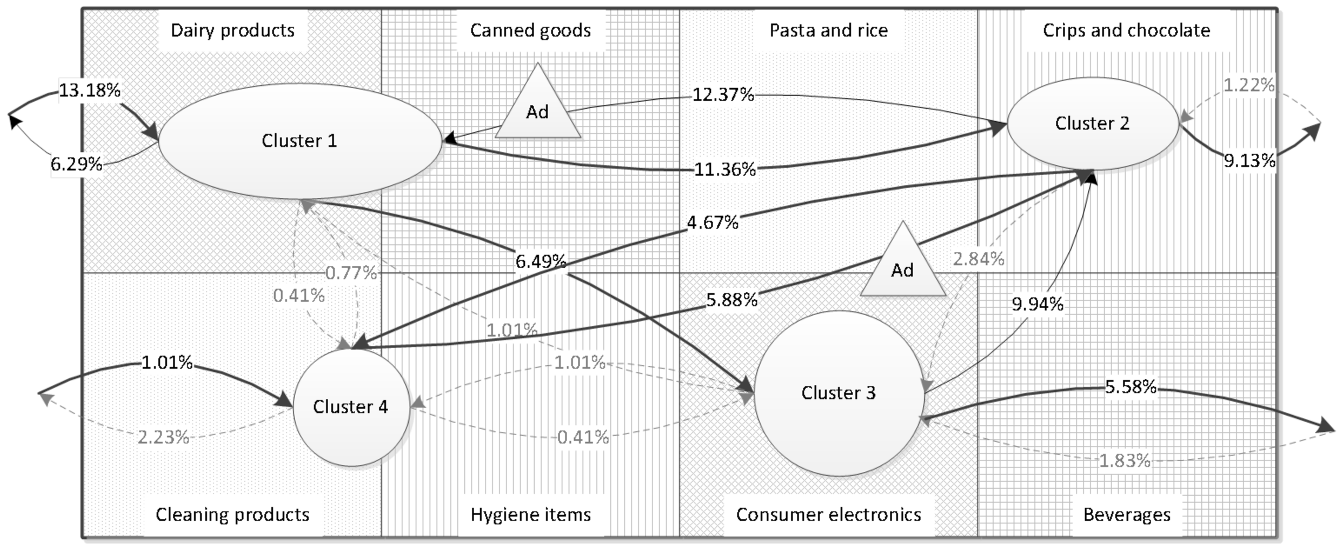

Figure 9. Prototypical retail environment – Scenario 2

V. Single Customer Analysis

Single customer analyses are used to provide information to sales people assisting them to address customers. Therefore, a cockpit is presented that provides relevant customer related information to the sale personnel in real-time. Key figures are used to deliver information about common characteristics of customers. Companionships provide an overview of which customers belong together and estimated behavior and next steps are used to interpret customers' behavior.

A. Key Figures

Key figures are discrete values that provide insight into single dimensions of individual customer behavior. Here, key figures are duration of stay and velocity as well as the number of stops, direction changes and visited sections.

a) Duration of Stay

The duration of stay describes the overall time a customer spent in a section or the entire environment. Therefore, it is an indicator for the interest of customers in certain sections. Customers that are more interested in products of a certain section will spend more time there. In contrast, less interested customers will leave the section faster. Measuring the duration of stay in real-time enables to estimate the time left in the store based on the average time spent.

b) Velocity

A customer's average velocity while walking through the retail environment is derived from the total distance between the entire recorded positions and the total time spent at the store or a section. Since stops distort the average velocity they are considered separately and are excluded from the velocity calculation. The average velocity helps to distinguish between hurrying and slowly traversing customers and is an indicator for the interest of customers in certain areas or the entire store. Besides, it helps to estimate a customer's interest in consultation.

c) Stops

Stops of customers describe a sequence of recorded positions that are close nearby each other or in the same place for a defined amount of time. Considering the average stops per section reveals information about customers' interest for the products exposed in this section. In addition, an above average amount of stops is an indicator for customers that are searching for specific products or comparing them and therefore might be used to address customers different.

d) Direction Changes

Direction changes are another feature of customer movements and indicate how well customers know the retail environment or sections of it. A high number of direction changes in one section indicate that a customer is searching for or comparing a specific product. A high number of changes within the entire environment might be evidence that the customer entered the store for the first time.

Direction changes are calculated as the smaller angle between preceding and succeeding points connected through an apex. If the angle undercuts 60° it is considered as an intentional change in a customer's movement direction.

e) Visited Sections

To calculate the number of visited sections the retail environment has to be separated in clearly delimited areas, e.g., "cereals" or "hygiene items". Then, counting the number of distinct visited sections enables to draw conclusions, which areas of a retail store are visited more or less often by customers. Besides, considering sequences of sections reveal typical sequence patterns. These patterns help to understand in which way customers traverse the retail environment and therefore, which sections are commonly visited in a row.

2012, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

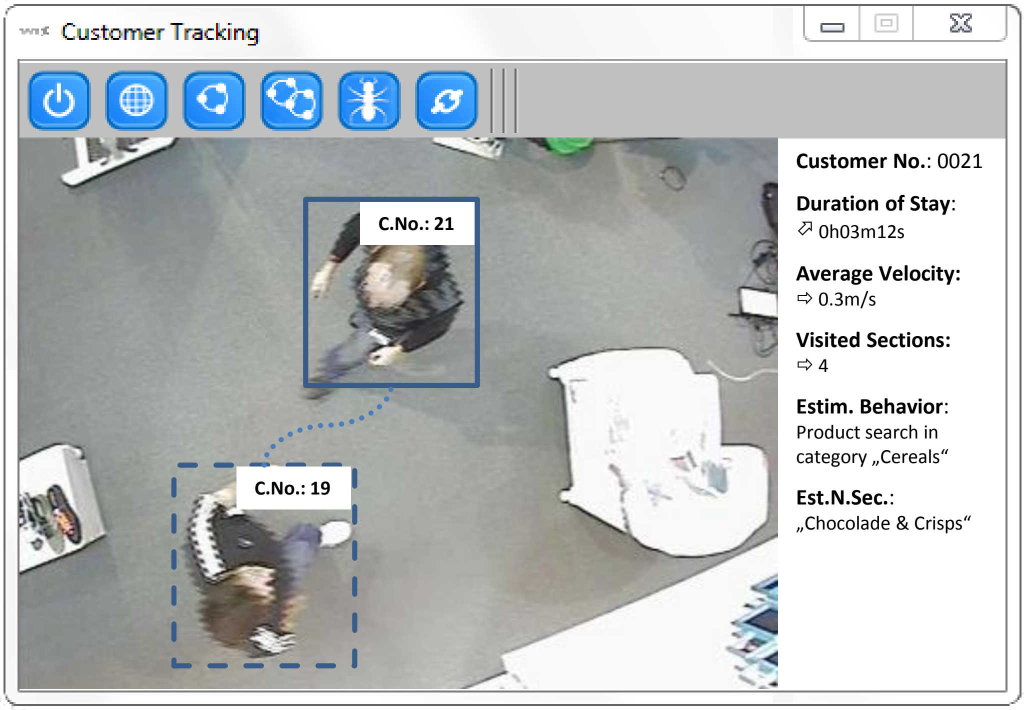

Figure 10. Cockpit from single customer analysis

f) Activities

In contrast to the other key figures activities are recorded through the approach described in Section III.B. Using activity data for product categories reveals information about how customers behave before they buy a specific product. E.g., if a customer views several different products of a product category before he/she puts it into the cart, the activities indicate that the customer is comparing product of a product category to find the most suitable product according to his requirements. In contrast, when a product gains less attention before it is bought it most probably is an item that is purchased habitually.

B. Companions

Companions are people shopping together. To determine companionships all people in the retail environment are continuously observed regarding the distance between each other. If the customers spend most of their time nearby each other the two persons are considered as related.

Figure 10 shows a prototypical cockpit visualizing the companionship of two customers by a dotted line.

C. Estimated Behavior

Estimated behavior describes the behavior that customers show based on their current movements and activities. For that, customer related information such as the number of stops, changes of direction, the average velocity and the activities, e.g., viewed products are determined. A pattern recognition system then uses this information to estimate a predefined class of behavior.

E.g., when a customer shows a significantly higher number of stops, direction changes and recently viewed a lot of products by taking them of the shelves without putting them in his cart his behavior is considered as a product search behavior.

D. Estimated Next Steps

Estimated next steps means the movement behavior that customers will most likely show based on their previous movements. For this, the movement history of a customer is compared to a set of rules being derived from previously analyzed trajectories. The system then estimates the most likely next steps. The approach is based on Markov models to represent transition probabilities between certain areas of a retail environment [36][37].

Figure 11 shows the prediction of the further path, based on a grid, which is superimposed over the shopping environment. The pathway starts at field 5-0 following the grey colored fields. The last performed step is at field 1-F. Based on the underlying model, the system estimates 2-F and 3-F as the most likely next steps.

E. Results

For the described test data the average duration of stay is 3.12 minutes. In that time customers move with an average velocity of 0.48m/s. Compared to the velocity including stopping times the adjusted velocity provides a greater variance (0.024m/s) and therefore enables a better differentiation between different customers. In average, customers perform 7.6 stops and 4.1 significant direction changes. It is striking that the stops and direction changes are mainly located at the beginning of the shopping getting

2012, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

less until the end of it. It looks like customers are losing interest the longer shopping takes. Besides, stops and direction changes are mainly found nearby more expensive products such as consumer electronics.

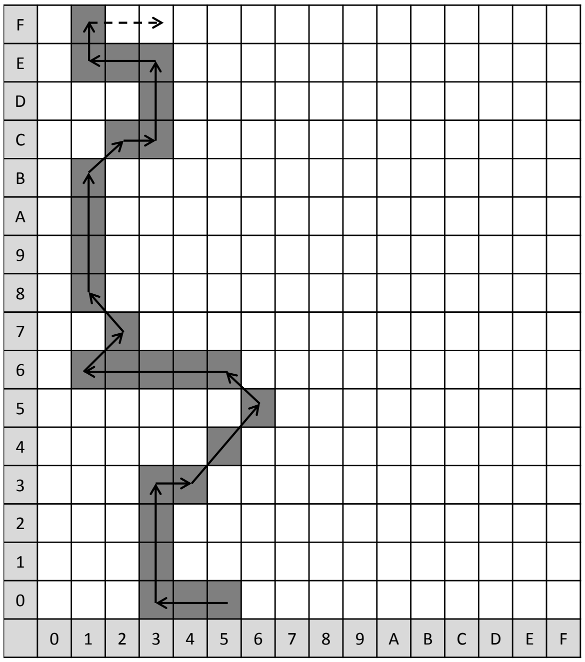

Figure 11. Movement estimation

Since the testing environment comprises only a small area, the average number of visited sections (7.3 of 8 sections) is comparatively high and therefore has less validity. Activities are mainly performed in sections like “crisps and chocolate” as well as “consumer electronics” and therefore corroborate the hypothesis that everyday life products are bought faster.

Using a k-Means clustering algorithm for detecting stereotypic behavior of customers works well for estimating customer’s behavior. However, results show solid accuracy for less than 4 behavior classes.

Initial tests of the companionship extraction show that the use of movement data provides reliable results for couples or friends. In contrast, families are harder to recognize since children are moving around more inconsistent. Therefore, distances measured for people staying in the same section are rated higher. As a result the rate of correct assignments is increased to a maximum of 86.1%.

For the estimated next steps Markov chains are tested from first to third order for the pre-defined sections shown in Figure 8 and 9. Besides, an evaluation of Markov chains with uniform grids is performed. The testing results reveal that second order Markov chains using data from grid fields are the best tradeoff between performance and accuracy.

VI. Conclusion and Future Work

The paper introduces an approach for recording and analyzing customer behavior data based on image and infrared sensors.

Surveillance cameras produce large amounts of video data. Intelligent methods for processing these data are crucial in order to gain customer insight especially over a longer period of time. Therefore, methods are introduced for capturing and extracting movements using network cameras and algorithms from the field of computer vision.

The same applies to customer activity data that are rarely considered yet but reveal valuable information. They are extracted by infrared sensors being implemented in the Microsoft Kinect Controller. Although it is difficult to cover an entire retail environment by these controllers the infrared sensor technology turned out to be a reliable way to record activity data.

Both the movement and the activity data are used for customer behavior analysis. Analyses are conducted using sets of pattern recognition algorithms. As a result information for retail managers and sale staff is gained and visualized through cockpits.

Considering longer periods of time might reveal different customer behavior not only for different setups but also for different times of a day, days of the week or seasons. In addition to that, comparative studies of different stores of the same chain are possible. Information gained from these analyses is the basis for planning dynamic product placements or seasonal offers. Besides, the knowledge about customers is considered as an additional source for management information systems. It helps to identify rarely visited areas or products and therefore enables retail store managers to analyze and optimize their shopping environment. New settings can be evaluated by considering the ex-ante and the ex-post change status.

Single customer related information put sales staff in the position of knowing their customers before they address them. Therefore, they are able to react more specific to customer requests and can provide them a better consultation.

Extending the described information by further information such as bought products and socio-demographic factors like gender or age might increase the validity of the information and therefore reveal new possibilities for management and sales force related services. For instance, an extended database allows setting up customer typologies considering different periods of time or different shops.

References

[1] J. Kröckel and F. Bodendorf, “Intelligent Processing of Video Streams for Visual Customer Behavior Analysis,” in *ICONS 2012, The Seventh International Conference on Systems*, 2012, pp. 163–168.

[2] H.-F. Li, S.-Y. Lee, and M.-K. Shan, “DSM-TKP: Mining Top-K Path Traversal Patterns over Web Click-Streams,” *The 2005 IEEEWICACM International Conference on Web Intelligence WI05*, pp. 326–329, 2005.

[3] I. Nagy and C. Gaspar-Papanek, "User Behaviour Analysis Based on Time Spent on Web Pages," in *Web Mining Applications in Ecommerce and Eservices*, Springer Berlin / Heidelberg, 2009, pp. 117–136.

[4] X. Zhang, W. Gong, and Y. Kawamura, "Customer Behavior Pattern Discovering," in *Advanced Web Technologies and Applications*, J. Yu, X. Lin, H. Lu, and Y. Zhang, Eds. Berlin/Heidelberg: Springer, 2004, pp. 844–853.

[5] P. Underhill, *Why we buy: The Science of Shopping*. Textere, 2000.

[6] B. Rao and L. Minakakis, "Evolution of mobile location-based services," *Communications of the ACM*, vol. 46, no. 12, pp. 61–65, 2003.

[7] J. Hightower, B. Brumitt, and G. Borriello, "The location stack: a layered model for location in ubiquitous computing," *Proceedings Fourth IEEE Workshop on Mobile Computing Systems and Applications*, pp. 22–28, 2002.

[8] G. D. Abowd and E. D. Mynatt, "Charting past, present, and future research in ubiquitous computing," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 1, pp. 29–58, 2000.

[9] D. Ashbrook and T. Starner, "Using GPS to learn significant locations and predict movement across multiple users," *Personal and Ubiquitous Computing*, vol. 7, no. 5, pp. 275–286, 2003.

[10] A. Gutjahr, "Bewegungsprofile und -vorhersage," 2008.

[11] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, no. 3, pp. 585–601, 2003.

[12] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classication and tracking from real-time video," in *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*, 1998, pp. 8–14.

[13] H. Koyuncu and S. H. Yang, "A survey of indoor positioning and object locating systems," *Journal of Computer Science and Network*, vol. 10, no. 5, pp. 121–128, 2010.

[14] D. M. Gavrila, "The Visual Analysis of Human Movement: A Survey," *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82–98, 1999.

[15] J. Perl, "A neural network approach to movement pattern analysis.," *Human Movement Science*, vol. 23, no. 5, pp. 605–620, 2004.

[16] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, and D. Pedreschi, "*A Visual Analytics Toolkit for Cluster-Based Classification of Mobility Data*," pp. 432–435, 2009.

[17] A. Bleiweiss, D. Eshar, G. Kutliroff, A. Lerner, Y. Oshrat, and Y. Yanai, "Enhanced interactive gaming by blending full-body tracking and gesture animation," *ACM SIGGRAPH ASIA 2010 Sketches on - SA '10*, pp. 1–2, 2010.

[18] K. Lai, J. Konrad, and P. Ishwar, "A gesture-driven computer interface using Kinect," *2012 IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 185–188, 2012.

[19] Z. Ren and J. Meng, "Robust hand gesture recognition with kinect sensor," *MM '11 Proceedings of the 19th ACM international conference on Multimedia*, pp. 759–760, 2011.

[20] H. Fillbrandt, "Videobasiertes Multi-Personentracking in komplexen Innenräumen," Rheinisch-Westfälische Technische Hochschule Aachen, 2008.

[21] N. Dalal and W. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05*, vol. 1, no. 3, pp. 886–893, 2004.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[23] M. Piccardi, "Background subtraction techniques: a review," *2004 IEEE International Conference on Systems Man and Cybernetics IEEE Cat No04CH37583*, vol. 4, no. C, pp. 3099–3104, 2004.

[24] T. Yoshida, "Background differencing technique for image segmentation based on the status of reference pixels," *2004 International Conference on Image Processing, 2004. ICIP '04.*, vol. 1, no. 1, pp. 3487–3490, 2004.

[25] M.-K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[26] G. Zhao, N. Zhang, and Z. Liu, "A case investigation on the scaling behaviors in web browsing," *Web Intelligence and Intelligent*, no. 70971089, pp. 160–163, 2010.

[27] G. R. Bradski, "Computer Vision Face Tracking For Use in a Perceptual User Interface," *Interface*, vol. 2, no. 2, pp. 12–21, 1998.

[28] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[29] D. Comaniciu and P. Meer, "Mean shift analysis and applications," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, no. 2, pp. 1197–1203 vol.2, 1999.

[30] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994, vol. 94, no. June, pp. 593–600.

[31] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *International Joint Conference on Artificial Intelligence*, vol. 3, pp. 674–679, 1981.

[32] P. Barrera, J. M. Canas, and V. Matellán, "Visual object tracking in 3D with color based particle filter," *Int Journal of Information Technology*, vol. 2, no. 1, pp. 61–65, 2005.

[33] P. Hong, M. Turk, and T. Huang, "Gesture modeling and recognition using finite state machines," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, 2000, pp. 410–415.

[34] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 330–339.

[35] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining*, 1996, vol. 1996, pp. 226–231.

[36] A. A. Markov, "Extension of the limit theorems of probability theory to a sum of variables connected in a chain (Reprint in Appendix B)," John Wiley and Sons, 1971.

[37] I. Nižetic, F. Krešimir, and K. Damir, "A prototype for the short-term prediction of moving object's movement using Markov chains," *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces*, pp. 559–564, Jun. 2009.

2012, © Copyright by authors, Published under agreement with IARIA - www.iaria.org