Visual Customer Behavior Analysis Based on Customer Movements

Johannes Kröckel, Freimut Bodendorf

Abstract—Today’s retail stores are facing new challenges. While a few years ago competitive advantages could be achieved by well-stocked product portfolios and bargain prices both domains have been lost to internet shops. To preserve and improve customer retention in stationary retail environments sophisticated customer approaches such as customized services or optimized retail environments are crucial. Therefore, comprehensive knowledge of the customers has to be available. In order to gain this knowledge without using vague customer surveys or short-time observations an automated solution is desirable.

An approach for capturing and processing customer movement data by using network cameras and algorithms from the field of computer vision is introduced. Two image processing methods are presented and evaluated. The extracted information is used for analyzing and visualizing customer behavior by the popular regions algorithm and Markov chains. The approach is exemplarily applied to a testing environment showing considerable differences in customer behavior for two settings.

Index Terms—Customer Behavior Analysis, Customer Tracking, Markov Models, Regions of Interest

I. INTRODUCTION

Compared to stationary retailers cheaper storage costs and smaller personnel expenditures enable internet shop operators to retain customers with cheap prices, a great variety of products and 24/7 availability. Especially when it comes to highly exchangeable products, benefits like physical product experience and personal product advice mostly cannot compensate the advantages of internet shopping sufficiently. Therefore, stationary retailers need new approaches to address customers by their actual context and needs.

While click paths, bounce and click rates as well as page impressions and time spent on websites are common key figures for internet shops (see [1]-[4]) stationary retailers lack sources of information about their customers and their individual context. Solely information like sales figures and product combinations are recorded at the checkout. Hence, addressing customers at the point of sale based on their behavior is not possible. Strategies like the one described by Underhill [5] try to compensate this knowledge shortage by manually conducted observations. However, these strategies also imply limitations. Among others, only small periods of time can be considered due to the need for great amounts of human resources. This means, continuous observation of variances over a longer period of time is not possible. Furthermore, an objective documentation of results by the observing people cannot be ensured.

The approach presented in this paper aims at detecting customer behavior patterns comprising customer movements in retail environments. Therefore, three steps have to be performed. First, customer tracking is performed by cameras recording raw data and computer vision algorithms extracting customer movements (section III). Second, the discrete movement data are analyzed regarding regions of interest and the customers’ movements between them (section IV). Third, the regions of interest and movements between them are used for customer behavior analysis and visualization (section V).

For visual tracking of customers two different methods are discussed. One of them considers video streams from lateral mounted cameras, the other one streams recorded from a bird’s eye view. The methods for tracking customers come from the field of computer vision. As a part of artificial intelligence the field of computer vision deals with the extraction of information from pictures and video streams. Therefore, knowledge from electrical engineering, mathematics and informatics is combined to imitate human understanding of images. An important area of application of computer vision algorithms is the detection of persons and their movements (“trajectories”) based on footage. The presented approach utilizes these algorithms for customer tracking and tracing. The discrete movement data derived from the computer vision analyses comprise information like coordinates and timestamps. Based on these datasets hotspots are identified by detection of regions of interest (RoI). The detected areas and the movements of customers between them are used as a basis for analyzing and visualizing typical movement patterns for different settings. These patterns are represented as Markov chains. The gained knowledge is used to achieve a better understanding of the behavior of customers for different product and environmental settings at the point of sale. The gathered knowledge is a valuable source of information for retail managers. Customer behavior can be observed on an abstract basis. This means that typical customer stereotypes may be identified by considering typical movement patterns between regions of interest.

J. Kröckel is with the Institute of Information Systems, University of Erlangen-Nuremberg, Nuremberg, Germany (e-mail: johannes.kroeckel@wiso.uni-erlangen.de)

Prof. Dr. F. Bodendorf is head of the Institute of Information Systems, University of Erlangen-Nuremberg, Nuremberg, Germany (e-mail: freimut.bodendorf@wiso.uni-erlangen.de)

ISBN: 978-0-9803267-4-1
II. RELATED WORK

Gathering and analyzing persons’ locations are strongly discussed topics in the field of ubiquitous computing (see [6]-[8]). The exemplarily named approaches use technologies integrated in mobile devices like GSM or GPS for position recording in outdoor environments. The collected data are mainly used for location based services without regarding historic data or future locations. In this work a different approach is chosen that enables the position determination in indoor environments. Besides, historic customer movements are analyzed in order to gain valuable insights into customer behavior and to predict future movements.

Movement prediction and analysis using GPS data is considered by Andrienko et al. [9] or Ashbrock and Starner [10]. They present an approach to predict peoples’ movements by location data collected from mobile devices. Highly frequented places (regions of interest) are extracted by a clustering algorithm and matched with well-known points of interest in their surrounding area. Prediction is based on Markov models. A similar approach is described by Gutjahr [11] extending the work of Ashbrock and Starner [10] by a greater variety of methods for location capturing. These authors are considering static points of interest like buildings or squares. However, due to continuing structural changes (e.g., bargain bins, seasonal products) such approaches are not applicable to retail environments. Points of interest can only be recognized during a limited period of time. In addition, they do not consider behavioral information that can be revealed from movement patterns.

The extraction of human movements from video streams is discussed by Wang et al. [12], Gavrila [13] and Perl [14]. However, these works are solely focusing on the extraction and comparison of person trajectories from footage and do not analyze the persons’ behavior.

III. CUSTOMER TRACKING

Approaches for person detection and tracking are strongly discussed topics in the field of computer vision. Algorithms are for example described by Muñoz-Salinas et al. [15], Arens [16] and Bischof [17]. While most concepts address applications like surveillance of places and facilities customer behavior within retail environments is less considered yet. The overall concept presented in this work is inspired by Fillbrandt [18]. 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 locations of a person are connected resulting in a trajectory.

A. Lateral Approach

For the lateral approach cameras are mounted at the upper end of a corridor. This facilitates the observation of an entire corridor area. Raw data are recorded by network cameras that enable real-time applications as well as subsequent analyses.

For the detection of persons entering the corridor the algorithm histogram of oriented gradients proposed by Dalal & Triggs [19] is applied. The approach is very suitable for the detection of people on images. First, the images are converted into gray scale. After that, the images are transformed into gradient maps. Therefore, small pixel areas are analyzed regarding their one-dimensional gradient direction by consideration of the areas’ histograms.

The gradient maps of a variety of images containing and not containing persons are used as a training set for a support vector machine to extract distinctive features that are typical for persons on images (Cortes and Vapnik [20]) (see Fig. 1a).

![Fig. 1. Lateral person tracking](image-url)
shopping at the same time reveals great occultation issues which cannot be compensated by optical flow. Besides, people with shopping carts or disabled people in wheel chairs are rarely recognized by the histogram of oriented gradients approach. The result is that tracking is unsuccessful for about 39% of the people. The average deviation between actual and automatically determined location is 0.17 m.

B. Aerial Approach

The aerial approach uses cameras mounted on the ceiling observing an area from above. The size of the area depends on the camera’s focal distance and altitude. In contrast to the lateral approach no explicit detection of persons is conducted. However, persons are determined by background differencing (among others described by Piccardi [24]) and a minimum expected shape size based on testing. Therefore, in a first step, a reference image is needed. Comparing the reference image with the actual considered frame excludes all similarities between the two pictures. Differences are highlighted (Fig. 2a and Fig. 2b). Consequently, shapes not similar to those of humans have to be eliminated. For this, in most cases a template or contour matching algorithm as described by Hu [25] or Zhang and Burckhardt [26] would be suitable. However, for the considered application these approaches are too error-prone because of the continuously changing shapes of persons moving to or away from the center of an observed area. Besides, people carrying bags or driving shopping carts as well as disabled people using wheel chairs would not be recognizable. Therefore, the detected shapes are solely filtered by a minimum size.

Fig. 2. Aerial person tracking

Subsequently, the continuously adaptive mean shift (camshift) algorithm presented by Bradski [27] is applied for person tracking. The algorithm is a further development of the mean-shift algorithm originally presented by Fukunaga and Hostetler [28]. Applied to computer vision 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 consequently defined as new centers of the corresponding windows (see Fig. 3).

By connecting subsequently occurring centers of windows a trajectory of the movements of the window is obtained. By defining windows as smallest rectangle areas covering shapes of persons extracted by the background differencing approach this concept can be applied for person tracking purposes.

While the mean-shift algorithm considers windows of static size, the camshift implementation adapts the window size dynamically. This is especially important for the presented application because persons moving away from or to the center of the observed area will occur in different sizes and shapes. 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.

To ensure better results, especially for crowded places, movements are verified by the good features to track and optical flow algorithms (see section III.A, [21], [22]). Coordinates of persons are derived from the centers of windows of the camshift algorithm. Similar to the lateral approach perspective distortion caused by the fish eye effect of the lens is compensated by 3x3 matrix transformation (see Fig. 2c).

Evaluation was performed for the same test environment as in the lateral approach. The sample video set contained 16,000 frames also showing 30 different people with a maximum of three people at the same time. Tracking is lost for only ~18% of the observed people. The average deviation between the real and the automatically determined location is 0.11 m.

C. Discussion

As described in III.A and III.B both approaches were tested for sets of test data. The lateral approach might gain a higher acceptance due to the fact that most surveillance cameras in retail stores are adjusted laterally to observe larger areas. However, mainly because of occultation and contrast issues reliability is lower compared to the aerial approach. Especially when it comes to crowded situations within corridors reliability decreases. The same applies to the accuracy of extracted waypoints for large areas.

The aerial approach provides a better overall reliability to the expense of a higher number of required cameras. Besides that,
the accuracy of the extracted location is more precise than achieved by the lateral approach. Occultation and therefore another source of error can be avoided. Further analyses are done by using data derived from the aerial approach because of the mentioned advantages.

IV. CUSTOMER MOVEMENT ANALYSIS

A. Region of Interest

The term “regions of interest” (RoI) describes significant subsets of one- or more-dimensional datasets. These datasets are mainly extracted by algorithms from the field of pattern recognition. Two-dimensional detection of significant areas is mainly used in medicine e.g. for tumor detection or search for anomalies on images (Evans et al. [30], Huesman et al. [31], Gibbs et al. [32]).

The presented approach uses region of interest detection as part of customer behavior observation. The regions of interest considered here are significantly more frequent areas in a shopping environment. The basic algorithm for RoI detection is derived from Gianotti et al. [33] and is adapted for the described application.

Before applying the popular regions algorithm data have to be adjusted. The data derived from the methods described in section III consist of discrete coordinates, timestamps and IDs connecting datasets of the same trajectory (i.e. person). This data structure does not meet the requirements of the popular regions algorithm. It has to be aggregated regarding time and place. This means, the area observed by network cameras is separated into several commensurate fields by superimposing a defined grid. Periods of time are calculated by considering the time being spent by people in the corresponding fields (see Fig. 4).

Summing up all periods of time people spent standing in a field results in a field’s overall time. However, only periods of time exceeding a defined threshold are included. The pre-calculated datasets now consist of time periods assigned to specific grid fields and can, for example, be depicted as a heat map (see Fig. 5).

Fig. 4. Data transformation for popular regions algorithm

Fig. 5. Example for a heat map depicting grid values

Succeeding to pre-calculation, popular regions algorithm is applied to the data. At first, an overall time threshold for RoIs, called density threshold, has to be defined. The subarea with the highest density (e.g., overall time) is determined as first RoI if it exceeds the defined threshold. Subsequently, all fields with time values beyond this threshold are considered, too. If they are not part of an already existing region of interest they are determined as a new one. Because these areas can consist of more than one grid field, the adjacent fields have to be considered, too. Therefore the area with the highest density in the surrounding of a region of interest is included to it. If the average density of the overall region also containing the new area is beyond the threshold, the subarea will be added as a part of the RoI. Otherwise, this RoI is considered as complete neglecting the recently added field.

B. Markov Chains

A simple Markov model consists of states of a system as well as transition probabilities between them [34]. Transition probability is defined as the probability that a system changes from one state to another one. In the presented approach the probabilities describe the chances that a customer moves from one region of interest to another one. Recursive transitions are neglected because for the presented approach only the succession of movements between different states (i.e. regions of interest) is relevant.

Regarding the movements between RoIs, the datasets derived from the computer vision algorithms described in section III.B are used. To compensate deviation between true and automatically derived coordinates adjacent subareas are added. Visited fields that are not part of a region of interest are ignored.

V. CUSTOMER BEHAVIOR ANALYSIS

Using the regions of interest and the movements between them enables a closer look on how customers behave within a retail environment. Therefore, two analyztes are conducted within a test shopping environment encompassing eight product categories (see Fig. 6).

The first analysis considers the regular behavior of customers within the observed area. The second one comprises a test setting based on the findings of the first one.
At first, a set of datasets being recorded without any modifications of the environment is analyzed. For this, data set three regions of interest covering areas of the product categories “dairy products”, “crisps and chocolate” as well as “consumer electronics” are identified by the popular regions algorithm. Considering the movements between the single RoIs reveals the transitions depicted in Fig. 6.

For the examples transitions with a percentage less than 6% of the entire transitions are not considered to get a better overview on the main transitions. The example depicted in Fig. 7 shows that the majority of customers enters the observed area from the left side moving to RoI 1 (“dairy products”). Only a much smaller number of customers move in from the right mainly first attending the product category “crisps and chocolate”. Furthermore, it can be recognized that the majority of people moves to “crisps and chocolate” (RoI 2) after having visited one of the two other RoIs. After staying at RoI 2 people are used to leave the observed area. Especially the section on the lower left covering “cleaning products” and “hygiene items” of seems to be visited rarely. For example, the detected regions of interest and transitions suggest an advertisement campaign for the cleaning products located in the lower left area. This may increase the interest in these products and, therefore, may also increase sales. Based on the results from the first analysis the test environment was adapted by setting up promotion signs for “cleaning products” on the paths between RoI 1 and RoI 2 as well as the one between RoI 1 and RoI 3. By testing the new setup the analysis results changed as depicted in Fig. 7. As in the above illustration only overall percentage rates greater than 6% are considered for a better overview. Still the majority of customers enter the observed area from the left side moving to “dairy products”. But, opposite to the first analysis a fourth region of interest within the area of “cleaning products” was detected. It proves that the promotion signs worked. A second indication of the success of the marketing campaign is the fact that customers mostly move from the second and third but less from the first region of interest to the new, fourth one.

VI. CONCLUSION

The presented approach enables analysis of customer shopping behavior on the basis of customer movements. Since surveillance cameras produce large amounts of video data intelligent methods for its processing are crucial in order to gain customer insight. Therefore, two methods for capturing and extracting movements using network cameras and algorithms from the field of computer vision are discussed. While both imply advantages and disadvantages the aerial approach is chosen due to better reliability and accuracy of the test results.
Subsequently, the derived data are used for customer behavior analysis. The data are aggregated and analyzed by the popular regions algorithms. Subsequently, the extracted regions of interest and the movements of customers between them are represented as Markov chains. Both, the regions of interest and the transitions are used to analyze the quality of the retail environment structure. Besides, the transitions reveal possibilities for placing marketing campaigns near highly frequented paths. One the one hand, this enables retailers to advertise products with locations away from the main paths like described in section V. On the other hand, they can advertise additional products to the ones located in the areas of previously attended regions.

Considering data covering longer periods of time might reveal different customer behavior during different times of day, on different days of the week or different seasons. Information gained from these analyses provides the basis for dynamic product placements or seasonal offers. The knowledge about customers can also be used as a valuable 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. Moreover, new settings can be tested and evaluated by considering the ex-ante and the post change status in a very short time.

The presented approach comprises an overall customer behavior observation. Regarding single customer paths and combinations of transitions allows identifying customer types. Therefore, the datasets will be extended by further information about movement (e.g. speed) or socio-demographic factors like gender or age.

Besides extending the data analysis the customer tracking approaches will be improved by a greater set of algorithms for improvement of reliability and accuracy.

REFERENCES