Intelligent Processing of Video Streams for Visual Customer Behavior Analysis

Johannes Kröckel, Freimut Bodendorf
Institute of Information Systems
University of Erlangen-Nuremberg
Nuremberg, Germany
johannes.kroeckel@wiso.uni-erlangen.de, bodendorf@wiso.uni-erlangen.de

***Abstract* - In today's society purchasing goods through web shops has become habitual. Some years ago only a few products like books, computer games and music CDs were intensely sold by online retailers. Today's internet shops are offering almost every imaginable product and service. This also leads to an increasing competition for traditional retailers offering products in stationary retail stores. Losing more and more customers stationary retailers need to think of new approaches for customer retention. Since customer retention is based on knowledge about the customers and their behavior store managers have to come up with new concepts for gaining and using customer knowledge to compete with bargain prices and 24/7 availability. In order to gain this knowledge without using vague customer surveys or short-time observations an automated solution is desirable. In this paper an approach is introduced, which allows to track and analyze customer movements through the store. Person tracking is accomplished by using aerial mounted cameras and a set of computer vision algorithms. Based on the captured movement data customer behavior analysis is performed by applying the dbscan algorithm and Markov models. The approach is illustrated by a test environment showing considerable differences in customer behavior for two settings.**

Keywords - Customer tracking; video analysis; behavior analysis; retail; point of sale.

I. INTRODUCTION

Low prices, short delivery periods and 24/7 availability are only three of the greatest advantages of web shops over stationary retailers. Moreover, highly exchangeable products like books need no physical experience and therefore lack reasons for buying them in a stationary shop. Consequently, stationary shop operators need to come up with sophisticated, individual approaches for attracting and retaining customers. This requires knowledge about the customers, their actual context as well as their on-site buying behavior.

Stationary retailers lack sources of information about their customers and their individual context, while click paths, bounce rates and page impressions as well as time spent on websites are common key figures for internet shops [1-3]. Sales figures and product combinations recorded by electronic checkout counters don't reveal individual customer behavior. Approaches like the one described by Underhill [4] try to overcome this knowledge shortage by manually conducted observations. However, these strategies are limited. Continuous observation over a longer period of time like weeks or months require too expensive human resources. Besides that, an objective documentation of results by the observing persons cannot be guaranteed.

The approach presented in this paper aims at detecting customer behavior patterns by analyzing movements of customers within retail environments. Therefore, customer movement data is extracted by cameras recording raw data and computer vision algorithms extracting movement information. Subsequently, the discrete movement datasets are analyzed regarding frequently attended areas and the customers' movements between them. Finally, the hotspots and movements between them are used for customer behavior analysis.

II. RELATED WORK

Extraction and analysis of location data is strongly discussed in the field of ubiquitous computing (see [5-7]). The approaches described in those papers apply cell phone compatible technologies like GSM or GPS for location tracking in outdoor environments. 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 indoor.

Data recorded by GSM and GPS are mainly used for location based services which are a surplus for the service users but not for companies. The approach presented here uses historic customer movements in order to gain valuable insights into customer behavior and to predict future actions. Thus, it is especially interesting for retail managers.

ISBN: 978-1-61208-184-7

The prediction and analysis of movement data derived from GPS is among others addressed by Stauffer and Grimson [8], Andrienko et al. [9] or Ashbrock and Starner [10]. The presented approach is used to predict peoples' movements by location data collected from mobile devices. Highly frequented places are extracted and matched with well-known points of interest in their surrounding area. Prediction is based on first-order Markov models. Gutjahr [11] extends the work of Ashbrock and Starner [10] by a greater variety of methods for location capturing. These authors consider static points of interest like buildings or squares. However, due to continuing structural changes (e.g., bargain bins, seasonal products) such approaches cannot be applied to retail environments. Points of interest are only recognized during a limited period of time. In addition, there is no consideration of behavioral information that can be revealed from movement patterns.

III. MOVEMENT TRACKING

Person detection and tracking approaches are part of the field of computer vision. Algorithms are among others proposed by Bischop [12], Wang et al. [13] and Perl [14]. Fields of application are mainly the surveillance of places and facilities. Little attention is being paid to customer behavior within retail environments yet. The overall concept presented in this work is inspired by Fillbrandt [15]. In his doctoral thesis he introduces an approach for a modular single or multi camera system tracking human movements in a well-known environment. Fillbrandt's approach is applied for tracking people in airplanes. Persons are detected on images by a set of computer vision algorithms and position estimation is executed.

Camera based person tracking can be accomplished by two similar settings. The lateral approach uses cameras being mounted edgewise [16] whereas the aerial approach utilizes cameras mounted on the ceiling. While lateral mounted cameras enable the observation of larger areas, approaches using aerial mounted cameras reveal more accurate results. Therefore, an aerial approach is chosen.

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 a set of algorithms is applied. First, background differencing (among others described by Piccardi [17] and Yoshida [18]) is used for object detection in single frames being captured by a camera.

In a first step, 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 Fig. 1, upper right area). After that, image noise is reduced. Eliminating objects that are smaller than the average human shape reveals all objects that could be considered as persons. This step is mostly accomplished by using a template or contour matching algorithm as described by Hu [19] or Zhang and Burckhardt [20]. 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 as humans by the algorithm. Therefore, the detected shapes are filtered by a minimum area.

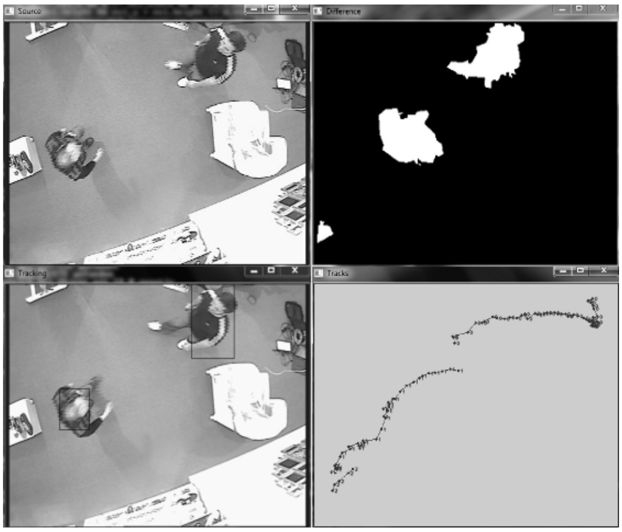

Figure 1. Aerial person tracking

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 [21] is applied for tracking the detected persons. The algorithm is based on the mean-shift algorithm originally introduced by Fukunaga and Hostetler [22]. The approach 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 [23]. 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. 2). 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 especially important 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.

To come to better results, especially for crowded places the good features to track algorithm by Shi and Tomasi [24] and the optical flow algorithm by Lucas and Kanade [25] are applied. The good features to track algorithm uses corner detection to find pixels, which differ from those in their surrounding area.

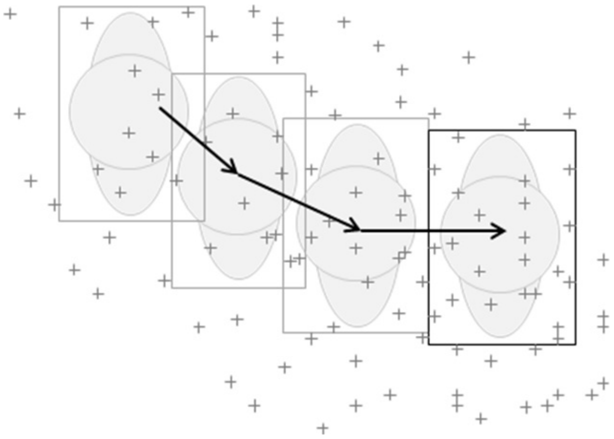

Figure 2. Mean-shift: window shifts

The optical flow algorithm compares sequential images regarding noticeable changes. So, prominent pixels are obtained from the good features to track algorithm. Subsequently, the algorithm tries to find these pixels in the 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. [26]) 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 Fig. 3).

Due to the fish eye 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. An evaluation study was performed for a test environment including one corridor and two shelves. The corridor between these two shelves has a width of 2.0 m and a length of 4.0 m. The width compares with the typical scales of a small grocery store. The camera is placed 3.0 m 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 0.11 m.

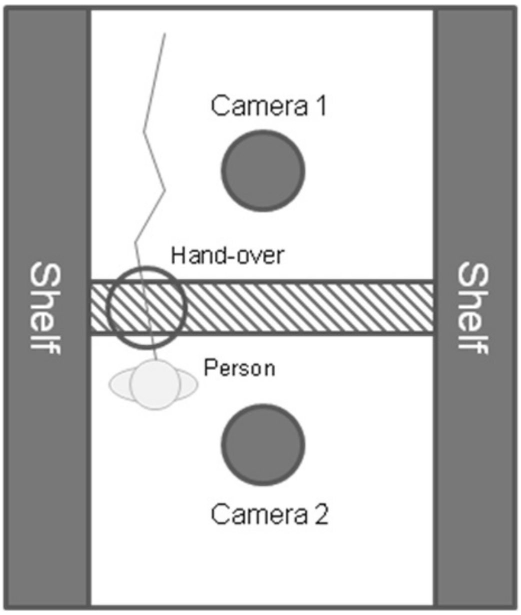

Figure 3. Camera hand-over

IV. MOVEMENT ANALYSIS

A. DBScan

The algorithm called 'density-based spatial clustering of applications with noise' originally proposed by Ether et al. [27] was developed to distinguish between clusters and noise in spatial databases. 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. 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 distance between p and the considered neighboring points. If one of the constraints is not met no new cluster is formed and another 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 5 points and a maximum real world distance of 0.02 m. The analysis reveals an amount of 551 clusters.

For further processing all clusters including points from less than 60% of the trajectories are eliminated. This reveals 3 clusters comprising points of between 60% and 90.5% of the trajectories within the test environment (see Fig. 4). The areas covered by the clusters are considered as hotspots that are significantly higher visited than other areas of the test retail environment.

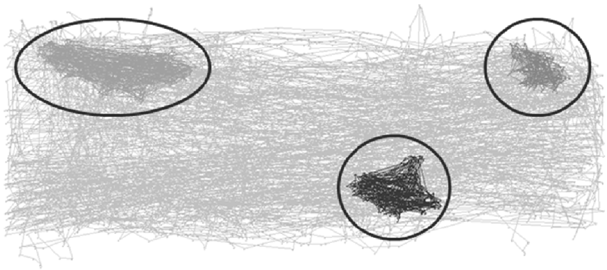

Figure 4. Clusters revealed by DBScan algorithm

B. Markov Chains

A first-order Markov model includes states of a system as well as transition probabilities between them [28]. A transition probability is defined as the probability of a system change from one state to another one. In the presented approach probabilities describe the chances of moves 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 succeeding 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 are taken into account. Points that are not part of one of the three considered clusters are ignored.

V. BEHAVIOR ANALYSIS

Considering the clusters and the movements between them allows a closer look on how customers act within a retail store. As use cases two shopping scenarios within a prototypical retail environment have been analyzed. The test environment comprises eight product categories (see Fig. 5).

The first scenario describes a regular product setup without any advertisements and signs and therefore observes the regular behavior of customers. The second one is based on the findings of the first scenario and includes advertisements for selected products. 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.

Fig. 5 shows these three clusters as well as the transitions between them. The percentage values describe the percentage of transitions been made between two clusters compared to the total transitions. Transition paths below the limit of a 5% share are ignored.

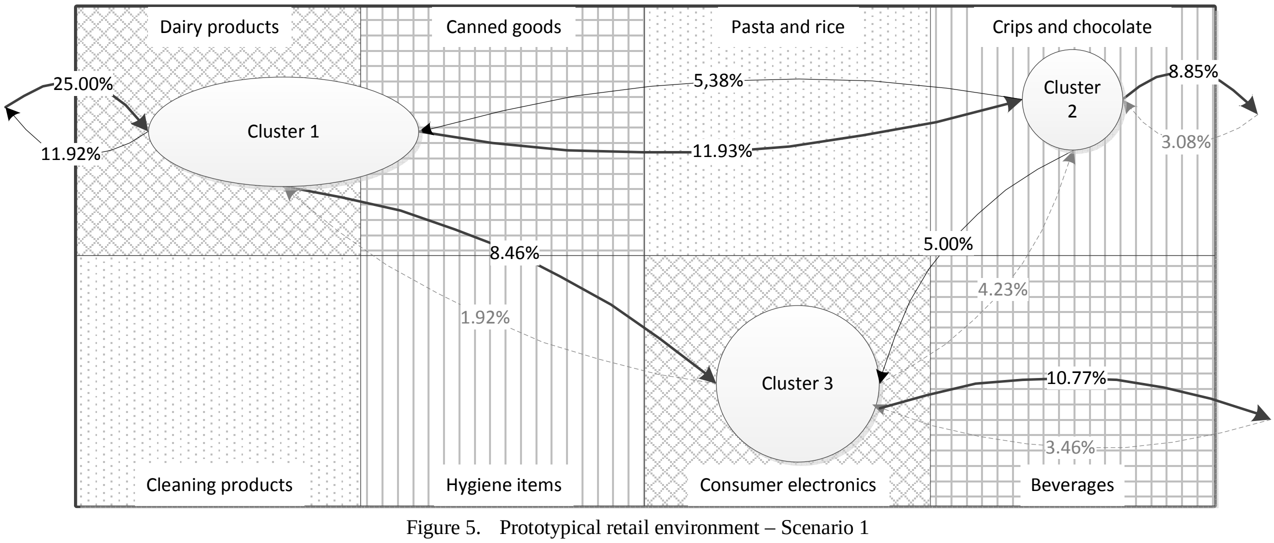

Figure 5. Prototypical retail environment – Scenario 1

For the given scenario the majority of customers enter the corridor from the left side heading for 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 the one of 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 Fig. 5 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 of advertisements near frequently used paths to call attention for these products.

This idea is seized for the second scenario (see Fig. 6.). 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 enter the observed retail environment most likely from the left side attending the area near dairy products first. Consequently, they are most likely moving on either to crisps and chocolate or 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. But this approach cannot only be used for advertisement evaluation. Furthermore, it is possible to evaluate the entire structure of a retail environment regarding product placement or shelf structure.

VI. CONCLUSION

This paper introduces an approach for the analysis of customer behavior on the basis of video recordings of customer movements within a real-world shopping environment. Surveillance cameras produce large amounts of video data. Intelligent methods for processing this data are crucial in order to gain customer insight especially over a longer period of time. Therefore, a method for capturing and extracting movements using network cameras and algorithms from the field of computer vision is provided. The resulting data is used for customer behavior analysis. Analysis is done using the dbscan algorithm. Finally, the extracted clusters and the movements of customers between them are represented as Markov chains. Both, the cluster areas and the transitions are used to assess the quality of the retail environment. The transitions reveal possibilities for marketing campaigns considering highly frequented paths. On the one hand, this enables retailers to advertise products with locations away from the main paths as described in the preceding section. On the other hand, retailers can promote additional products located near previously visited hotspots. Moreover, based on the tracked trajectories next customer movements can be predicted and individualized messages posted on advertisement screens in real-time.

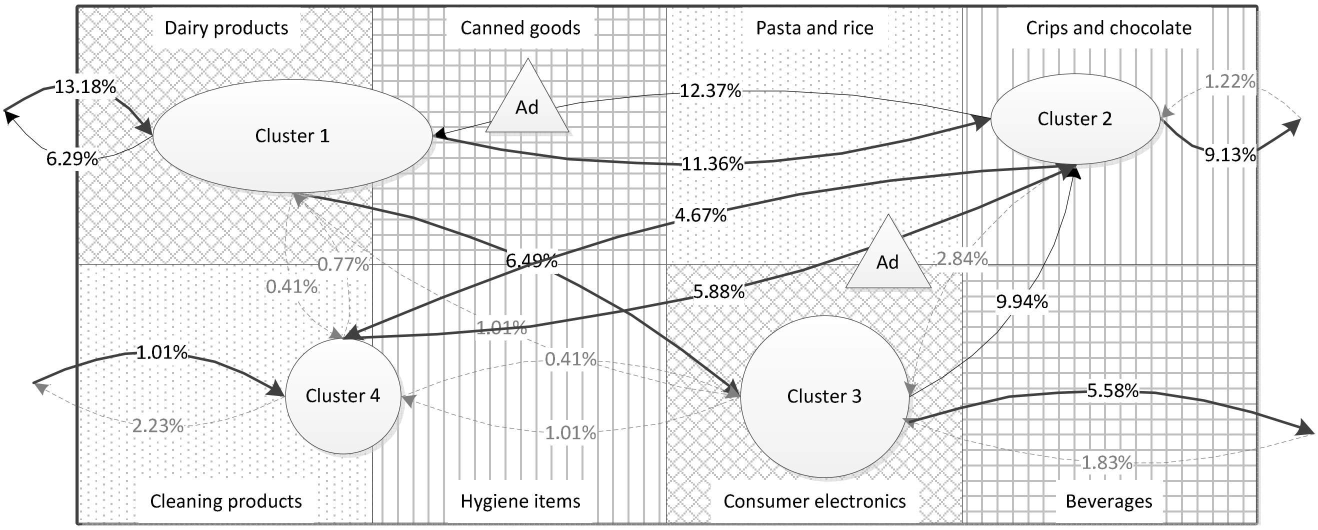

Figure 6. Prototypical retail environment – Scenario 2

Considering longer periods of time might reveal different customer behavior not only for different setups but also for different times of a day, for different days of the week or for different 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. This knowledge about customers is also 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 post change status.

REFERENCES

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

[2] I. Nagy and C. Gaspar-Papanek, “User Behavior Analysis Based on Time Spent on Web Pages,” in Web Mining Applications in Ecommerce and Eservices, Springer Berlin / Heidelberg, 2009, pp. 117-136.

[3] A. Goldfarb, “Analyzing Website Choice Using Clickstream Data,” Advances in Applied Microeconomics A Research Annual, vol. 11, 2001, pp. 26.

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

[5] C. Feature, “Systems for Ubiquitous Computing,” Computer, vol. 34, August 2001, pp. 57-66, 2001.

[6] C. Becker and F. Dürr, “On location models for ubiquitous computing,” Personal and Ubiquitous Computing, vol. 9, no. 1, 2004, pp. 20-31.

[7] J. Hightower and G. Borriello, “Location systems for ubiquitous computing,” Computer, vol. 34, no. 8, 2001, pp. 57-66.

[8] C. Stauffer and W. E. L. Grimson, “Learning patterns of activity using real-time tracking,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, 2000, pp. 747-757.

[9] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti, “Interactive visual clustering of large collections of trajectories,” Proc. IEEE Symposium on Visual Analytics Science and Technology, no. ii, 2009, pp. 3-10.

[10] 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, 2003, pp. 275-286.

[11] A. Gutjahr, “Bewegungsprofile und -vorhersage,” 2008.

[12] H. Bischof, “Robust Person Detection for Surveillance Using Online Learning,” Proc. of the Ninth International Workshop on Image Analysis for Multimedia Interactive Services, 2008, p. 1.

[13] L. Wang, W. Hu, and T. Tan, “Recent developments in human motion analysis,” Pattern Recognition, vol. 36, no. 3, 2003, pp. 585-601.

[14] J. Perl, “A neural network approach to movement pattern analysis,” Human Movement Science, vol. 23, no. 5, 2004, pp. 605-620.

[15] H. Fillbrandt, “Videobasiertes Multi-Personentracking in komplexen Innenräumen,” Rheinisch-Westfälische Technische Hochschule Aachen, 2008.

[16] J. Kröckel and F. Bodendorf, “Extraction and Application of Person Trajectories in Retail Environments,” Proc. of the IADIS International Conference - Intelligent Systems and Agents, 2010, pp. 109-113.

[17] M. Piccardi, “Background subtraction techniques: a review,” Proc. IEEE International Conference on Systems Man and Cybernetics IEEE, vol. 4, no. C, 2004, pp. 3099-3104.

[18] T. Yoshida, “Background differencing technique for image segmentation based on the status of reference pixels,” Proc. International Conference on Image Processing, ICIP ’04., vol. 1, no. 1, 2004, pp. 3487-3490.

[19] M.-K. Hu, “Visual pattern recognition by moment invariants,” IEEE Trans Information Theory, vol. 8, no. 2, 1962, pp. 179-187.

[20] W. Zhang, C. K. Chang, H.-i Yang, and H.-yi Jiang, “A Hybrid Approach to Data Clustering Analysis with K-means and Enhanced Ant-based Template Mechanism”, 2010, vol. 1.

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

[22] 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, 1975, pp. 32-40.

[23] D. Comaniciu and P. Meer, “Mean shift analysis and applications,” Proc. of the Seventh IEEE International Conference on Computer Vision, vol. 2, no. 2, 1999, pp. 1197-1203.

[24] J. Shi and C. Tomasi, “Good features to track,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, vol. 94, 1994, pp. 593-600.

[25] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” Proc. International Joint Conference on Artificial Intelligence, vol. 3, 1981, pp. 674-679.

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

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

[28] 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.