Analysis of Shopping Behavior based on Surveillance System

Mirela Popa, Leon Rothkrantz, Zhenke Yang, and Pascal Wiggers
MMI Department, TU Delft
Delft, the Netherlands
{m.c.popa; l.j.m.rothkrantz; z.yang;p.wiggers}@tudelft.nl

Leon Rothkrantz and Zhenke Yang
Sensor Technology, SEWACO, Netherlands Defence Academy
Den Helder, the Netherlands
{ljm.rothkrantz; z.yang}@nlda.nl

Mirela Popa, Ralph Braspenning, and Caifeng Shan
Video and Image Processing, Philips Research
Eindhoven, the Netherlands
{mirela.popa; ralph.braspenning; caifeng.shan}@philips.com

***Abstract*—Closed Circuit Television systems in shopping malls could be used to monitor the shopping behavior of people. From the tracked path, features can be extracted such as the relation with the shopping area, the orientation of the head, speed of walking and direction, pauses which are supposed to be related to the interest of the shopper. Once the interest has been detected the next step is to assess the shopper's positive or negative appreciation to the focused products by analyzing the (non-)verbal behavior of the shopper. Ultimately the system goal is to assess the opportunities for selling, by detecting if a customer needs support. In this paper we present our methodology towards developing such a system consisting of participating observation, designing shopping behavioral models, assessing the associated features and analyzing the underlying technology. In order to validate our observations we made recordings in our shop lab. Next we describe the used tracking technology and the results from experiments.**

***Keywords*—Model, Bayesian Networks, Surveillance, Tracking, Shopping behavior.**

I. INTRODUCTION

Surveillance in public places by means of Closed Circuit TeleVision (CCTV) systems is currently widely used to monitor locations [4] and the behavior of the people in those areas. Since events like the terrorist attack in Madrid and London, there has been a further increasing demand for video sensor network systems to guarantee the safety of people in public areas. But also events like football games, music concerts and large venues like shopping malls where many people gather, have a need for video surveillance systems to guarantee safety. In this paper we propose to use the existing surveillance system to investigate the shopping behavior of people. In a shop, products are displayed in such a way to optimize the buying behavior of shoppers. Using the surveillance systems, the ideas about placement of products can be validated and improved.

However, the greater the number of cameras, the greater the number of operators and supervisors needed to monitor the video streams. A fully automated surveillance system for shopping behavior analysis is currently not commercially available. Some software packages do exist [19], but they mostly record video streams and provide little further analysis.

Motion detection and human tracking, as well as behavior analysis methods are widely researched topics. Understanding the way in which customers interact with products and developing an automatic system for recognizing their behavior and assessing possible business opportunities can be of a great benefit for shops, leading to improved marketing strategies and helping them building a better relationship with their customers.

A multimodal surveillance system would use audio and video data from the feed and analyze them to determine the behavior that is present in the currently observed scene. To analyze a situation, the scene must first be interpreted.

Separate systems are proposed to analyze both the video and the audio stream.

The goal of our research consists of designing empirical based models of shopping behavior. Next, we aim at building a system for recognizing and analyzing the customers' behavior in relation with products. Finally we plan to test our system on real life, spontaneous data. This work will focus on methods of analyzing and processing the video data. The goal is to extract as much relevant features as possible from the footage and to find a way to interpret this data in a meaningful manner given the context.

The outline of the paper is as follows. In the next section related work is reported, then we present our proposed methodology consisting of participating observation based on which we define shopping behavior models, followed by the assessment of the relevant features and the analysis of the underlying technology. Next tracking algorithms are discussed followed by the description of the experimental results. Finally we formulate our conclusions and directions for future work.

II. RELATED WORK

A. Shopping behavior analysis

Shopping behavior represents the decision processes and acts of people involved in buying and using products.

978-1-4244-6587-3/10/$25.00 É2010 IEEE

Monitoring shopping behavior [7] presents interest for both the academic society and also for the private sector. The ViCoMo project [22] focuses on the recognition of people behavior in general, having applicability in security and surveillance but also in the consumer market. There were also earlier attempts towards observing and analyzing shopping behavior made by Wells and Sciuto [25] based on participating observation. Regarding companies, Shopping Behavior Xplained [19] investigates and analyses the customers' behavior, in order to find the conscious and sub-conscious motivations that influence the decision making process of purchasing a certain item, by using both video recordings of shopping sessions and also interviews with the customers.

The purpose of automated surveillance systems concerns mainly security issues, but they can be also useful in assessing customers' interactions with products, as their main capabilities consist in localizing, tracking people, and analyzing their behavior.

B. Enabling Technology

There are currently a number of computer vision techniques available, especially tracking, face detection and motion interpretation [18]. Next we will discuss the possibilities of each method and the feasibility of their application to our work.

1) Motion detection

Since most of the behavior we wish to detect is associated with some kind of body movements, being able to detect motion in a scene is the first and most important basic step towards understanding behavior. Motion detection aims at detecting the non-static parts of a scene by comparing for example two scene captures. Techniques used for motion detection could be divided in several categories: background subtraction [20], temporal differencing [11], or optical flow estimation [12].

Most of the following operations we wish to perform, such as person tracking and behavior interpretation are highly dependent on motion detection. One important aspect involved in motion detection algorithms regards background modeling.

2) Background modeling

Background modeling is very important for motion detection since it provides a description of the scene which can help in interpreting the observed motion data. In our case a shop will be split up in walking areas, product areas, and areas for shop assistants. Background modeling can greatly reduce the cost of computation and help eliminate false positives. One of the challenges is being able to model the background pixels under varying lighting conditions. A simple approach is based on using a previously acquired image of the scene without any objects or persons in the scene, possibly under varying lighting conditions, so that it can later be used for background subtraction. More advanced methods exist, some including Gaussian pixel models [20], others using Kalman filters to reduce the variance in illumination, e.g. Hu et al. [8].

3) Object detection

Object-class detection aims to detect all objects in a scene belonging to a certain class, such as vehicles, bags, but also human bodies or animals. Mikolajczyk [14] combined AdaBoost with local orientation histograms and built a detector using object parts. Another approach was proposed by Tuzel et al. [21], in which human detection is achieved by using classification on Riemannian manifolds. The object descriptors they use are covariance matrices of image features computed over image regions.

4) Face detection

Face detection can be regarded as a special case of object-class detection. While most face detection methods try to detect frontal views of faces, newer algorithms attempt to detect faces from multiple angles, or multi-view face detection [24]. A wide variety of techniques exist, ranging from simple edge-based algorithms, to complex high-level approaches using pattern recognition methods. The method used by Albiol et al. [1] detects faces by first detecting skin pixels and then applies a segmentation algorithm to find skin regions. A watershed segmentation algorithm is used to find clusters, after which the most face-like blobs are selected as the faces.

Viola and Jones [23] introduced a new approach for visual object detection using the principle of a boosted cascade of classifiers. Their detector can be trained for face detection, and is capable of processing images extremely rapidly while achieving high detection rates.

5) Object tracking

A common tracking method is to use a filtering mechanism to predict each movement of the recognized object. The most commonly used filter in surveillance systems is the Kalman filter [15]. Condensation [10] is a very well-known tracking algorithm, having the advantage that it can be used more or less independent of the object representation; still it is less efficient at tracking more than one object at the same time. A less known alternative for Condensation is the Mean Shift algorithm [5] which was recently proposed as an efficient tool to handle partial occlusions and significant clutters [3].

6) Behaviour analysis

In the view of an automatic surveillance system, object detection and tracking needs to be followed by object behavior analysis and recognition. Hu et al. [8] present behavior understanding as a classification of motion patterns produced by the object tracking module. Other methods which proved their efficacy in generating behavior models use maximum entropy and Markov mixture models [13]. Dynamic time warping (DTW) is a time-varying technique widely used in speech recognition, image patterns and recently in human movement patterns [16]. In [17] the recognition of behaviors and activities is done using a declarative model to represent scenarios and a logic-based approach to recognize predefined scenario based models.

III. SHOPPING BEHAVIOR METHODOLOGY

To design and implement a system for automatic assessment of users' appreciation of products and opportunities for selling we used the following methodology composed of several steps. First we did participating observation of the customers' behavior while shopping, which will be presented in Section III.A. Based on these observations we build behavioral models of shopping behavior, which will

be presented in details in Section III.B. We observed the features characteristic for each type of behavior and we grouped them depending on the addressed modality (people detection, face detection, gesture recognition, or voice analysis). We employ a Bayesian Network to model the relationships between the sensors, the observed features and the associated shopping behavior. Next we investigated the underlying technologies needed to assess the proposed features and we present the system architecture in Section IV.B. In order to validate and test our models we made recordings of shopping behavior in our shop lab. Finally we present the obtained experimental results and comment upon them.

A. Partcipating Observation

There is a continuous interest in building and testing consumer behavior models [9]. One methodology which is useful for defining user models is participating observation. In the shops, the researchers observed in an unobtrusive manner the shopping behavior of people.

One researcher was standing near the entrance of the shop, while another one was following the customer at a reasonable distance such that he did not interfere in the shopping activity.

Both researchers wrote notes of their observations. In total we collected 20 hours of observations. We present next an example of an observation session:

"A lady enters the shop. She goes directly (1) to the elevator, at a high speed (2). She chooses the 2nd floor and she goes to the make-up rayon. She looks around (4) for a moment and selects a product. Then she raises her hand (7), asking (8) for a shop assistant. She asks (8) for special offers from the advertising magazine. She looks happy (6) as she discovers that there is a new offer for the product she wants. She follows (1) the assistant to the pay desk and goes away (1)."

In the brackets, the relevant features which characterize shopping behavior are indicated. They are presented together with the associated sensing devices in Table 1 from Section V.A.

B. Behavior Models

Our behavior models are based on the observations made in real shops. Not surprisingly, there are many individual differences in shopping behavior of people. The ultimate goal of shopping is to buy a required product. In order to realize that goal a shopper has to perform some actions. In case a shopper knows what he wants, he has to find the location of the product and to put the product in his basket. Next we have the helpless people, who cannot find the product and are actively looking for support. Finally we have the "fun"-shoppers, who have no idea what to buy and first look around for interesting products or just enjoy being in a shop. All kinds of shoppers show different but characteristic behavior and our surveillance system should be able to extract relevant features for the recognition of the specific type of shopper. We considered the following types of shoppers: goal oriented, disoriented shopper, looking for support shopper, fun-shopper and duo shopper.

A *goal oriented shopper* has a shopping list, knows the location of the product and walks directly to that place at a high speed, without looking left or right. An example of this behavior can be noticed in Fig. 1 below.

The *disoriented shopper* has no specific idea about what he wants, doesn't know if the product is available or where to find it. He walks at a low speed, is frequently looking left and right, and walks without any apparent plan.

Figure 1: Goal oriented behavior

Some shoppers are *looking for the shop assistant* or waiving their hand, asking information about the location of the products, alternatives, and characteristics of the product (see Fig. 2).

Figure 2: Looking for help behavior

A *fun-shopper* wants to be where the action is, he joins the crowded area; he has interest for expositions, demonstrations, or products promotions.

Some people like to shop in a group, especially the fun-shoppers but also partners or friends. They talk a lot to each other and comment the choices and appreciation of products to their partner. An example of this type of behavior is depicted in Fig. 3.

A special case of the *duo-shopper* is represented by a parent accompanied by his child/children. In this case we can expect a lot of walking around, as the child is running away, but this behavior is not related to products. This type of behavior should be interpreted and recognized correctly.

Figure 3: Commenting of products

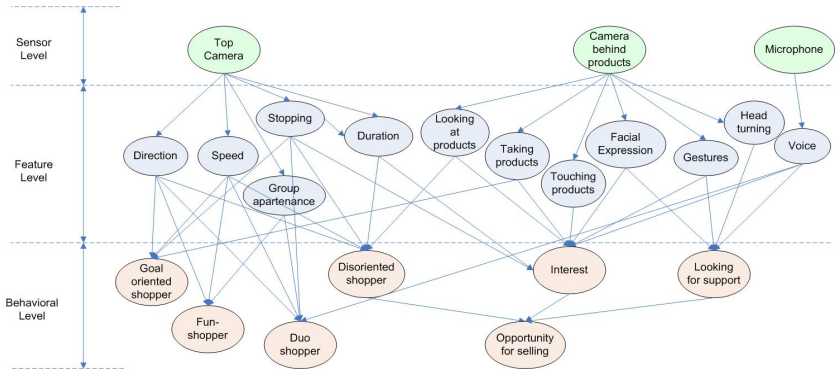

Figure 4: Bayesian Network Model of shopping behavior

C. *Analysis of Shopping Behavior*

By observing shopping people we extracted the following general steps in the shopping behavior: orientation, selection, appreciation, and decision making.

- *Orientation.* During orientation the shopper has to decide what he wants and has to make a plan for how to find the product(s). If the shopper is not familiar with the shop, he will inspect the shop information lists and the routing schema, he will just walk, or he will ask for help.
- *Selection.* Once the product has been discovered a shopper will look for characteristics and alternatives. A shopper is inspecting the product, reading the text, comparing products, and comparing prices, or looking for promotions.
- *Appreciation.* During the selection process, positive and/or negative emotions will surface and result in a final positive or negative appreciation. From facial expressions or other non-verbal behavior the appreciation can be assessed.
- *Decision making.* The last step of the shopping process is to decide whether to buy the product or not. In the case of duo-shopper, the partner will be consulted. Shoppers take a last close look at the product and the facial expression can be one of puzzlement. Once the decision has been made shoppers show a positive or negative facial expression.

Each of the previously described steps can be further divided into sub-steps specific for each type of shopping behavior.

IV. PROPOSED SYSTEM

A. *Reasoning*

Probabilistic methods, such as Bayesian Networks offer a number of advantages for the representation and processing of knowledge and uncertainty, being able to cope with missing or incomplete information.

We employed a Bayesian Network to condense information and represent the relationships between the observed features and the corresponding shopping behavior.

The proposed model is presented in Fig. 4. Our model is organized on three levels; on the first level we display the used sensors (cameras, microphone). The next level contains the observed features (e.g. walking in relation with speed and direction, or activities related to products: looking, taking, touching), we regard all these features as observables. The third level corresponds to the shopping behavior, giving an indication regarding the interest of shoppers or the opportunity for selling. This level is regarded as a high semanthical level, where based on the observables, hypotheses can be formulated. The relevant features are fused, each having associated a weight. For example, in order to detect the "Goal Oriented" behavior, the direction and the speed of walking are considered and observed by the top camera. When the customer stops for a period of time, the system changes the focus to the camera behind the products and the customer's actions are assessed. We are interested if he/she is looking at products, if he/she takes one, touches it and then puts it back or puts it in his basket. Furthermore we extract relevant information from his facial expression, gestures, and voice (both the tonality and the semantics). Finally we conclude if he is interested in that product and also the type of interest (positive or negative).

In order to learn the model parameters from data we would need a lot of training examples, which are not currently available. We will tackle this issue by asking experts in this field to specify or to correct the Conditional Probabilities Tables (CPT) for each parameter.

B. *System Architecture*

In this section, the design of our system for multimodal assessment of users' appreciation of products is presented. We propose a modular approach and we describe next the functionality of each module. A diagram of the proposed system is shown in Fig. 5.

The system is composed of two basic modules for video and audio data analysis. First the video file is processed in order to extract the image sequences and the audio file. The video data analysis module consists of motion detection, human detection, facial expression recognition, and also hand detection sub-modules. Motion based analysis regards motion detection and motion recognition tasks. Motion energy classification is important as it can give indication regarding the general direction or the amount of movement. The human detection part aims at recognizing people and tracking them accordingly. Besides playing an important role in human

tracking, face detection and tracking is also used for assessing user's appreciation of products, by enabling the analysis of facial expressions. The process of recognizing the user's facial expression consists of applying the Active Appearance Model (AAM) [6] to the facial region, extracting the relevant features, and then using a classification method such as Hidden Markov Models (HMMs).

From the video data we also extract information regarding gestures, a separate module being designated for this task.

Skin color detection together with pose estimation and edge detection contribute to hand detection and tracking.

The audio analysis module consists of feature extraction (pitch, energy, MFCC, jitter, etc.), task followed by classification in two phases. First we detect interest or non-interest. Next, in case interest was detected we classify it into positive or negative.

The output of the system is obtained by fusing the intermediary results of each module (motion detection, human behavior detection, facial expression recognition, gesture recognition, and voice analysis). In order to take into account the relationships and the importance of each module, a Dynamic Bayesian Network (DBN) is employed (see Fig. 4). One of the strengths of the presented system consists in its adaptability. If one modality is not available or the quality of the result is below a certain threshold, then it will be discarded in the fusion process. Furthermore each modality has associated a certain weight which reflects both its importance and reliability. We have implemented modules for motion energy classification, human detection and tracking, facial expression analysis, and voice analysis.

Still all these modalities need to be integrated and fused in a common framework, task on which we are working at this moment. Next we present in more details the human tracking module.

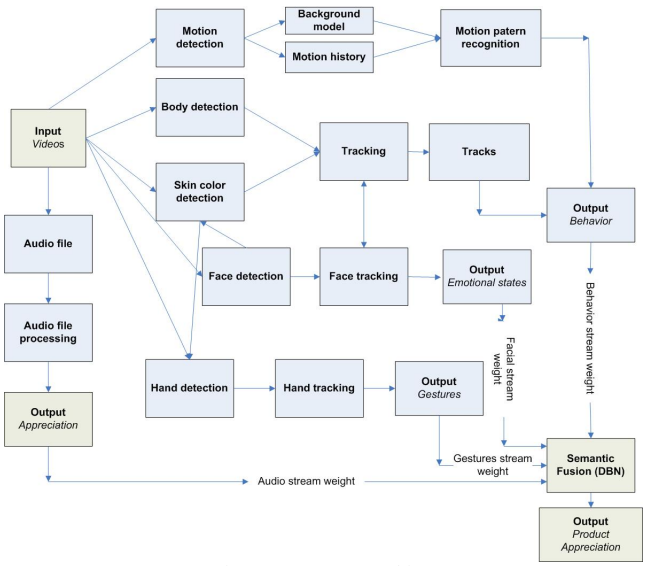

Figure 5: System Architecture

C. *Human Tracking*

As it was stated in Section II.B.5 the Mean Shift algorithm [5] represents an efficient method for human tracking, reason why we chose to use it in our approach. The 'mean shift' represents the estimated direction and distance in which the target moves, and these parameters are computed without using a dynamic model, but only by comparing a candidate target with the model. Regarding object representation, in the Mean Shift approach every object is represented by an ellipse.

In the initialization phase, for every object that has to be tracked, the model color histogram of the ellipse is computed in the Region of Interest (ROI). To increase the robustness, to every histogram a convex and monotonic *kernel mask* is added: pixels in the center of the ellipse get a higher value than the ones on the border. The weight decreases with squared distance from center. Next the histogram is normalized. All the histograms used are in three dimensions (one dimension for each color). To compute the distances between the histograms, the Bhattacharyya [5] distance is used.

During every processed frame, the target moves toward its most probable position in multiple iterations. In order to accomplish this task we used an algorithm which maximizes the Bhattacharyya coefficient (higher coefficient means a higher similarity and thus a shorter distance).

1) First for candidate location y_0, the current candidate histogram $p(y_0)$ will be computed, together with the kernel mask. After that, the Bhattacharyya coefficient between this histogram and the model histogram q is computed.

2) For every pixel, compute a weight as defined by:

$$w_i = \sum_{u=1}^{m} \delta[b(x_i) - u] \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}} \quad (1)$$

where $b(x_i)$ is the bin for the color of pixel x_i, u is the current bin and δ is the Kronecker delta function. This means that every weight is the square root of the value of the model bin of the pixel color, divided by the value of the candidate bin of the pixel color.

3) Compute the mean shift, which represents the new estimated location y_1:

$$\hat{y}_1 = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} \quad (2)$$

Then compute again the Bhattacharyya coefficient, between the new candidate histogram $p(y_1)$ and the model histogram q.

4) As long as the coefficient between $p(y_0)$ and q is larger than the one between $p(y_1)$ and q, the target has not yet been reached, so the location of y_1 must be updated: $y_1 \leftarrow \frac{1}{2}(y_0 + y_1)$. Repeat this step until the target has been reached.

5) If $\| y_1 - y_0 \| < \varepsilon$, stop the iterations, and continue with the next frame. Otherwise, start a new iteration at step 1 with the new candidate ellipse: $y_0 \leftarrow y_1$.

In the following section we present our recordings acquisition process which provided the data on which we could test our algorithms.

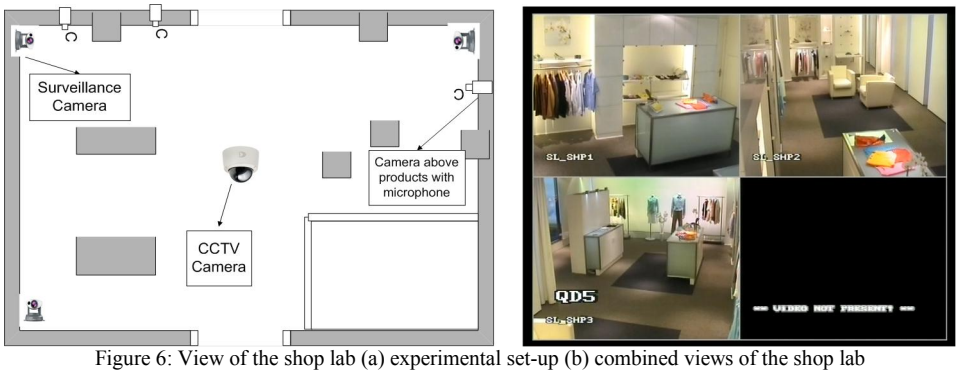

Figure 6: View of the shop lab (a) experimental set-up (b) combined views of the shop lab

V. EXPERIMENTS

A. Data Collection

In order to test our system we need recordings of shopping behavior. Our aim is to have realistic data consisting of spontaneous shopping behavior of people. But before we can obtain this kind of data we have to solve first some ethical problems regarding privacy. Therefore we test our system for the time being on recordings made in our shop lab (see Fig. 6b). Cameras and microphones were installed at different points, at the locations depicted in Fig. 6a. The multimodal devices used in our lab consist of a camera attached to the ceiling, a camera with a microphone behind the products, and three surveillance cameras in the corners of the room. A combined view of the shop lab as captured by the surveillance cameras is shown in Fig. 6b. The purpose of each camera is different, each being designated to capture a certain type of reaction/behavior. In Table 1 we present the extracted features and the associated capturing devices.

TABLE 1. FEATURES AND SENSING DEVICES

Behavior	Sensing Devices	Features
(1) Walking directions	Top camera/ Surveillance camera	Trajectory analysis Motion estimation
(2) Speed of walking	Top camera/ Surveillance camera	Relative distance of positions-points at regular time intervals
(3) Stop	Top camera/ Surveillance camera	Cluster of points
(4) Looking at products	All cameras	Head position/Gaze estimation
(5) Touching products	All cameras	Hand movements
(6) Facial expressions	Camera behind/above products	Positive/negative appreciation
(7) Gestures	All cameras	Hand raising, waiving
(8) Speech	Microphone	Utterances asking for help

We asked ten students and researchers to show the shopping behavior of the five shopping models presented above and we made recordings of the main steps characteristic for each behavior. Next we present the experimental results obtained using the recorded data.

B. Experimental Results

In this section we discuss the first experimental results in tracking and analyzing shoppers' behavior. We tested first the motion detection module.

Bobick and Davis [2] propose a view-based approach, which uses motion history images (MHI) and motion energy images (MEI) to interpret human behavior. This method is based on the assumption that different actions have different motion history patterns which can be used to detect and classify human actions. We computed the amount of motion by comparing the movement of pixels in successive frames resulting in different energy distributions (see Fig. 7a, b). Fig. 7a presents a shopper standing in front of a product display, searching for a certain item and then starting to walk around, with a low speed and looking around. These cues are similar to the ones displayed by the "disoriented shopper" and represent a first indication of this type of behavior. In Fig. 7b is presented the motion energy for a shopper, standing in front of a product taking a closer look at it or taking it. This motion pattern represents an indication that the customer is interested in that product.

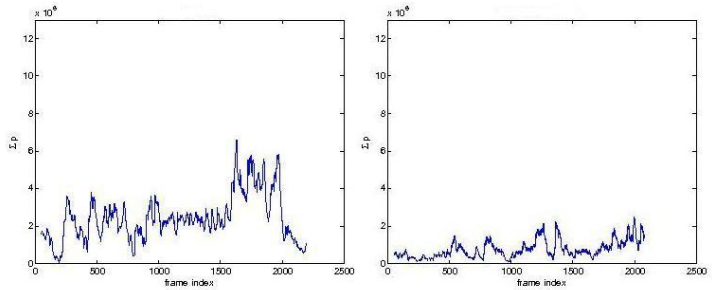

Figure 7: Energy graphs of: (a) shopper searching and walking (b) shopper standing in front of a product

An analysis of the presented energy graphs indicates that higher peaks can be correlated with large movements (e.g. walking), the medium peaks can be associated with movements with a smaller amplitude such as hands movements for example, while the motion pattern depicted in graph 7b suggests low movement or no motion at all. The information derived from the motion energy graphs refers to the amount of motion. To be relevant for behavior classification, this information needs to be used in

combination with other features: with a low or high speed, constant or variable way of walking, certain types of activities or facial expressions, in order to infer if the motion pattern corresponds to a certain type of behavior or not.

The motion detection module provides input for the next module: human tracking, by indicating which parts of the image are interesting to follow. Finding the connected components is based on the foreground pixels. The components which have an area larger than a predefined threshold will be used as tracked objects in the Mean Shift algorithm (see the example in Fig. 8).

Figure 8: Tracking customers

In Fig. 9 we show 5 tracks. At fixed time intervals we plot the position of the tracked blob. In this way we get a trajectory with dots along it. The distance between the dots is an indication of the speed of walking. The red track corresponds to a shopper walking at a high speed at a straight line but his walking speed slows down at the end, probably close to the products of his interest. The blue track depicts a similar walking pattern as the red track. The yellow and the green track correspond to shoppers wandering around.

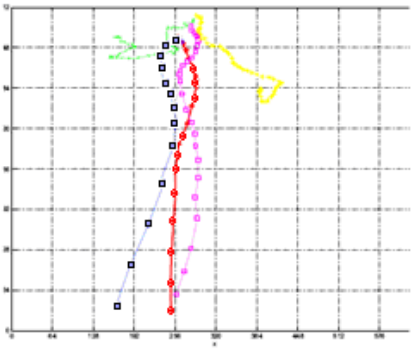
Figure 9: Trajectories of shoppers walking with different speed along different tracks (in colors)

Given the outline of the shopping area and the position of the products, we are able to detect who is interested in what.

Trajectory analysis consisting of walking pattern and speed provides us with a first indication of the type of behavior of the customers. The red, blue and purple tracks can be associated with a goal oriented shopper, while the green and the yellow ones represent a first indication of a disoriented shopper.

Still multiple people tracking is not always performing in a reliable way, therefore we combined it with the face detection algorithm of Viola and Jones [23]. This is a very robust algorithm but nevertheless results in lot of false positive and false negatives. To cope with the high amount of misses, we try to recover a face in the following frame(s) as follows.

Since we have 25 frames in a second, there is a lot of redundant information available. Once a face was detected of size $n \times n$, it can be expected that in one of the next frames the same face is also visible. We consider a search window $(n+m)\times(n+m)$ around the face-coordinates and look in the following p frames for a face which is assumed to be the expected face. We set p=10 and m=5. Table 2 presents the results obtained after applying different optimizations. The first column shows the raw face detection rate. Column 2 shows the results after ignoring detection misses due to turned heads (e.g. profile views). The false positive rate is reduced by considering masking areas (column 4) and analysing a detected face in relation to a blob.

TABLE 2. FACE DETECTION RESULTS OF A VIDEO SEQUENCE OF 1420 FRAMES (118 sec).

Detection rate	Frontal correction	False positive rate	FP rate after masking	FP rate after blob alignment
62%	86%	20%	11%	4%

A noticeable decreasing of the false positive rate was achieved by applying a mask (dividing the shop into areas designated for products, customers and shop assistants), as a lot of faces were detected previously in the products area due to the similarity with skin color. By combining the human tracking and the face tracking modules we achieved a lower false positive rate as showed in Table 2. The face tracking module has also other advantages, being useful at detecting a customer looking left and right, feature which is characteristic for a certain type of behavior.

Next, in order to validate our assumptions related to shopping behavior interpretation, we annotated segments of the recorded videos as either goal oriented or disoriented type of behavior and we extracted a set of features (min, max, standard deviation of position, speed, and acceleration on x and y axes).

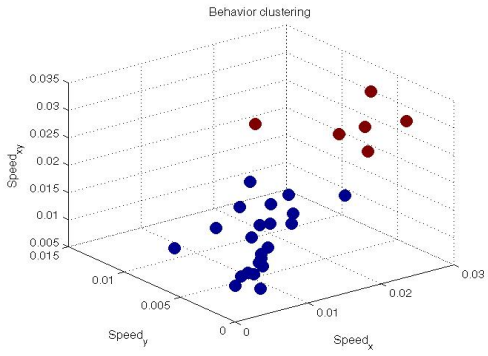

Figure 10: Behavior clustering based on customers' speed (in colors)

As expected, speed proved to be a discriminative feature for the two considered types of behavior and the obtained results are shown in Fig. 10 (blue points correspond to disoriented shoppers, while the brown points depict the goal oriented ones).

The presented results are preliminary and we plan to continue our recordings and to test our prototype system on real life shopping behavior data from a shopping mall. Next we present our conclusion and give indication regarding future work.

VI. CONCLUSIONS

In this paper we report about a surveillance system in a shop lab, analyzing the shopping behavior. By participating observation in real shops we were able to model shoppers with different shopping behavior. We designed and implemented a first running prototype and tested the modules: motion detection, trajectory analysis based on human tracking, and face localization and tracking for different shoppers. The time annotated tracks show characteristic features which enable us to classify the shopping activities. Furthermore, by analyzing the customers' speed of walking we were able to make a first classification of their shopping behavior into 'goal oriented' or 'disoriented' type. As future work we plan to test the other implemented modules especially the modules related to sound recording and analysis. The next step consists of fusing the data from cameras at different location and view angles and the data from different modalities. In the last years we developed software systems to fuse data from different modalities and we plan to use them in the shopping behavior surveillance context. Further on we plan to make real life recordings of shopping people, which can enable the validation of our proposed models. Finally our system should serve as an automatic assessment tool of users' appreciation of products and opportunities for selling. Knowing the layout of the shopping area an intelligent surveillance system can also infer the type of products a customer is interested in and might be used to propose personalized advertisements.

ACKNOWLEDGMENT

This work was supported by the Netherlands Organization for Scientific Research (NWO) under Grant 018.003.017.

REFERENCES

[1] A. Albiol, L. Torres, and E. Delp, "An unsupervised color image segmentation algorithm for face detection applications", In IEEE International Conference on Image Processing, pp. 681-684, Oct. 2001.

[2] A. Bobick and J. Davis, "Real-time recognition of activity using temporal templates", In WACV '96: Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision (WACV '96), pp. 39, Washington, DC, USA.

[3] Y. Cai, N. de Freitas, and J. Little, "Robust visual tracking for multiple targets," In 9th European Conference on Computer Vision, 2006, pp. 107–118.

[4] R. Collins, A. Lipton, and T. Kanade, "A system for video surveillance and monitoring," In American Nuclear Society 8th Internal Topical Meeting on Robotics and Remote Systems, 1999.

[5] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects using Mean Shift," IEEE Conf. on Computer Vision and Pattern Rec., vol. 2, pp. 142-149, 2000.

[6] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active Appearance Models," In H. Burkhardt and B. Neumann, editors, 5th European Conference on Computer Vision 1998, Vol. 2, pp. 484-498, Springer, Berlin.

[7] I. Haritaoglu and M. Flickner, "Attentive billboards: Towards to video based customer behavior," In Proc. IEEE Workshop on Applications of Computer Vision, pp. 127-131, Orlando, FL, USA, Dec. 2002.

[8] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," System, Man and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 34(3):334-352, Aug. 2004.

[9] S. K. Hui, P. S. Fader, and E. Bradlow, "Path Data in Marketing: An Integrative Framework and Prospectus for Model Building," Marketing Science, March 1, 2009; 28(2): 320 - 335.

[10] M. Isard and A. Blake, "Condensation - Conditional Density Propagation for Visual Tracking," International Journal of Computer Vision, 29(1):5-28, 1998.

[11] G. Jing, C. E. Siong, and D. Rajan, "Foreground motion detection by difference-based spatial temporal entropy image," In Proc. IEEE Region 10 Conference (TENCON), vol. A, pp. 379-382, Chiang Mai, Thailand, Nov. 2004.

[12] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision (darpa)," In Proc. of the 1981 DARPA Image Understanding Workshop, pp. 121-130.

[13] E. Manavoglu, D. Pavlov, and C. L. Giles, "Probabilistic User Behavior Models," Data Mining, ICDM 2003, Third IEEE International Conference on Data Mining, pp. 203-210.

[14] K. Mikolajczyk, C. Schmid, and A. Zisserman, "Human detection based on a probabilistic assembly of robust part detectors," In Proc. European Conference on Computer Vision, volume 3021 of Lecture Notes in Computer Science, pp. 69–81, Prague, Czech Republic, May 2004.

[15] N. T. Nguyen, S. Venkatesh, G. West, and H. H. Bui, "Multiple camera coordination in a surveillance system," Acta Automatica Sinica, 2003, 29, (3), pp. 408–421.

[16] T. Oates, M. D. Schmill and P.R. Cohen, "A method for clustering the experiences of a mobile robot with human judgements," Proc. of the 17th National Conference on Artificial Intelligence and Twelth Conf. on Innovative Applic. of Artificial Intelligence, AAAI 2000, pp. 846-851.

[17] N. Rota and M. Thonnat, "Video sequence interpretation for visual surveillance," 3rd IEEE Int. Workshop on Visual Surveillance, Dublin, 2000, pp. 59-68.

[18] A. W. Senior, L. Brown, A. Hampapur, C. F. Shu, Y. Zhai, R. S. Feris, Y.L. Tian, S. Borger, and C. Carlson, "Video analytics for retail," In Proc. IEEE Conference on Advanced Video and Signal-based Surveillance, pp. 423-428, London, UK, Sep. 2007.

[19] Shopping Behavior Xplained (SBLX), 2009. "Axis cameras watch shopper's behavior," http://www.axis.com/files/success_stories/ss_ret_sbxl_36113_en_0907_lo.pdf

[20] Z. Tang and Z. Miao, "Fast background subtraction and shadow elimination using improved Gaussian mixture model," In Proc. IEEE International Workshop on Haptic Audio Visual Environments and their Applications, pp. 38-41, Ottawa, Canada, Oct. 2007.

[21] O. Tuzel, F. Porikli, and P.Meer, "Human detection via classification on riemannian manifolds," In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, Minneapolis, MN, USA, June 2007.

[22] ViCoMo, "Visual Context Modelling," September 2009, http://www.itea2.org/public/project_leaflets/VICOMO_profile_oct-09.pdf

[23] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Vol. 1, pp. I–511– I–518.

[24] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A literature survey," ACM Comput. Survey, 35(4):399–458, 2003.

[25] W. D. Wells, L. A. Lo Sciuto, "Direct Observation of Purchasing Behavior," Journal of Marketing Research, Vol. 3, No. 3, pp. 227-233, Aug. 1966.