

# **Multimodal Assessment of Shopping Behavior**

Mirela Carmia Popa

*Intelligent and Interactive Systems, University of Innsbruck, Technikerstr. 21a, Innsbruck, Austria*

*Thesis Advisors: Prof. dr. drs. L. J. M. Rothkrantz, Prof. dr. C. M. Jonker*

*Date and location of PhD thesis defense: 8th January 2013, University of Delft, Delft, the Netherlands*

Received 2nd February 2015; accepted 27th July 2015

---

## **Abstract**

An increasing number of public places (e.g. cities, schools, transit districts, and public buildings) are deploying CCTV surveillance systems to monitor and protect the people in those areas. Since events like the terrorist attacks 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 and monitor the behavior of people in these places.

Currently, the existing video surveillance systems in public places are used by human operators for monitoring the situation, analysing the data, and detecting abnormal or unwanted human behaviour such as theft or aggression or for later retrieval in case an unwanted event was reported. Human monitoring has benefits such as intelligent reasoning about the situation, but also limitations such as fatigue or loss of concentration, especially when nothing happens for a long period of time, or difficulty to cope with the overwhelming number of cameras and to watch them continuously. Therefore, a supporting alternative is represented by the development of automatic systems designated at monitoring the video streams and signalling the human operators only in the case of unusual or unwanted events.

Another applicability of an automatic surveillance system consists of detecting and analysing human behavior, being useful in several domains such as patients monitoring, supporting elderly people, detecting anti-social behavior of people in public places, or aggression detection in trains.

The dissertation [7] presents a study about video surveillance and behavior analysis applied to shopping. The existing video surveillance infrastructure in shopping malls, designated usually at serving security purposes such as aggression or theft detection, could be extended to other purposes, such as investigating the shopping behaviour of customers. For improving marketing strategies and offering a better service to the customers it is important to understand customers' shopping behavior, their relation with products, what catches their attention and what remains unobserved.

In order to meet this goal different types of information are employed, which are relevant at understanding and analyzing human behavior. Furthermore, different video cameras are synchronized and used in a collaborative manner, serving different goals. Fish eye cameras mounted on the ceiling are useful at detecting people and tracking them through the environment, high-definition cameras are installed in the relevant regions of interest,

---

Correspondence to: <Mirela.Popa@uibk.ac.at>

Recommended for acceptance by Jorge Bernal

DOI <http://dx.doi.org/10.5565/rev/elcvia.703>

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

to facilitate the recognition of the interaction patterns with objects in the environment, while for a more refined analysis of the people's reaction to objects (e.g. products), another type of video cameras is employed, such as high quality web cameras, devoted at recording the people's frontal facial expressions.

We investigated which information cues are relevant for behavior recognition and how to assess them automatically.

A first behavior cue regards to the human movement patterns, being useful at detecting when a person is disoriented and seems not to find what he/she is looking for, case in which offering support would be useful. People detection and tracking modules provide peoples' tracks, which are described using trajectory features. The customers' walking patterns are discriminated using Hidden Markov Models (HMMs) [3].

For obtaining a better overview of what is happening inside an environment, context information is used, which is based on the segmentation of the area into Regions of Interest (ROIs) (e.g. products, passing areas, pay desk, or resting areas). Features related to the ROIs, such as the time spent in each ROI together with the transitions between different ROIs contribute to a better recognition of the behavior, as an action can have different meanings in different ROIs [5].

More information regarding behavior can be extracted, by analyzing the interaction patterns with objects in the environment (e.g. products, shopping baskets, or shopping carts) [4]. In order to assess this type of information, relevant features are extracted, regarding both the appearance of a person and his/her movements. Next, we investigated different classification methods, both spatial (e.g. SVM, k-NN, Adaboost, Fisher) and spatial-temporal ones (HMMs and DBNs) [1], striving for finding the best suited one for discriminating between the different types of actions. The analysis of interaction patterns is important, as different sequences of actions can have different semantic meanings, contributing to the recognition of the behavioral models.

Another informative behavior cue is represented by facial expressions, which can be used to assess a person's reaction to an object or in our case study to a product [2]. Even though a huge amount of research was devoted to the recognition of the six basic emotions (happiness, sadness, surprise, anger, disgust, and fear), there is little known about product related emotions, reason why they need to be investigated. After identifying relevant product related emotional classes, the next step regards to automatic facial expressions recognition. This task consists of face detection, identifying facial landmarks (e.g. mouth, nose, eyes, and eyebrows), extracting discriminative features, and finally using a classification method in order to distinguish between different facial expressions.

Each intermediary analysis stream (trajectory analysis, action recognition, ROI detection module, and facial expression analysis), provides an input to the reasoning model, which based on the observables formulates a hypothesis regarding the most likely behavioral model. Two types of models are used for fusing the different data streams, both deterministic and probabilistic ones.

Based on expert knowledge, a rule-based system is defined, which takes into account the intermediary outputs of the trajectory analysis and action recognition modules, associates them with context information regarding visited ROIs, and draws a conclusion about the behavioural type [5].

The second proposed behavioural model is a probabilistic one, inspired by language models developed for speech recognition, as depicted in Fig. 1. We developed a behavioral model which combines a bi-gram model with the maximum dependency in a chain of conditional probabilities, satisfying the ordering and the dependency requirements. It uses a visual grammar to impose a structure on the basic actions. By adding semantics to the computational approach we were able to filter out unlikely combinations of basic actions and to improve the recognition accuracy of the behavioral types. A characteristic of the behavioral model is the ability of dealing with continuous data, meaning combinations of actions, and not being restricted to the discrete case in which only predefined patterns would be considered [6].

In order to validate the proposed approach, we constructed a proof-of-concept prototype for shopping. In particular, we identified that human movement patterns, their interaction with products, and their facial expressions represent relevant behavioral cues. Next, we introduced in the thesis [7] a behavioral model inspired by language models. Furthermore, we integrated the different types of information on the semantic level, using a multi-level framework. Finally, we evaluated this system in the ShopLab, in a real supermarket, and the

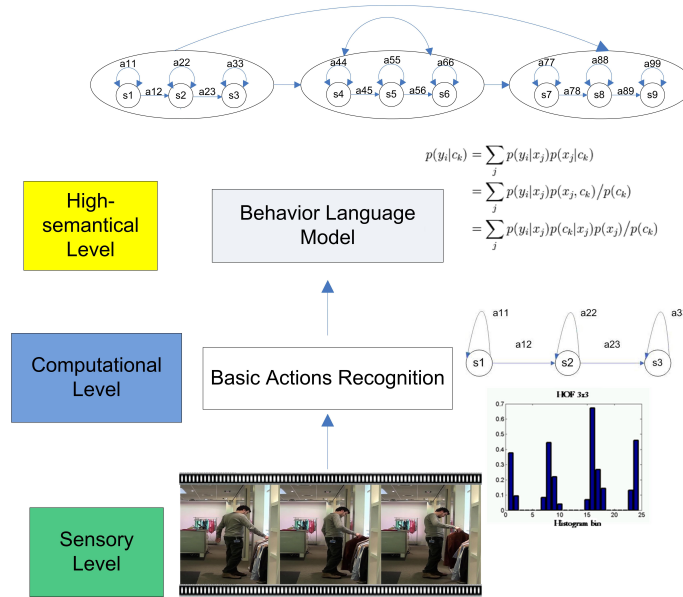


Figure 1: Computational approach from a hierarchical perspective.

product appreciation in a laboratory setting. The results show the feasibility of the approach in the recognition of trajectories (93%), shopping actions (91.6%), action units (93%), facial expressions (84%), and the most important behavioral types (87%).

## References

- [1] M. C. Popa, L. J. M. Rothkrantz, D. Datcu, P. Wiggers, R. Braspenning, and C. Shan, "A comparative study of HMMs and DBNs applied to Facial Action Units Recognition", *Neural Network World*, vol. 20, no. 6, pp. 737-760, 2010.
- [2] M. C. Popa, L. J. M. Rothkrantz, and P. Wiggers, "Products appreciation by facial expression analysis", *11th International Conference on Computer Systems and Technologies (CompSysTech'10)*, pp. 293-298, 2010.
- [3] M. C. Popa, T. Gritti, L. J. M. Rothkrantz, C. Shan, and P. Wiggers, "Detecting Customers' Buying Events on a Real-life Database", *Computer Analysis of Images and Patterns, 14th International Conference (CAIP 2011)*, Springer Heidelberg, ISBN 978-3-642-23672-3, Seville, Spain, pp. 17-25, 2011.
- [4] M. C. Popa, L. J. M. Rothkrantz, C. Shan, and P. Wiggers, "Assessment of Customers Level of Interest", *IEEE Int. Conference on Image Processing (ICIP)*, Florida, U.S.A, 2012.
- [5] M. C. Popa, L. J. M. Rothkrantz, C. Shan, T. Gritti, and P. Wiggers, "Semantic Assessment of Shopping Behavior Using Trajectories, Shopping Related Actions, and Context Information", *Pattern Recognition Letters*, 34(7), 809-819, 2013.
- [6] M. C. Popa, L. J. M. Rothkrantz, P. Wiggers, and C. Shan, "Shopping behavior recognition using a language modeling analogy", *Pattern Recognition Letters*, 34(15), 1879-1889, 2013.
- [7] M. C. Popa, "Assessment of Shopping Behavior: Automatic System for Behavioral Cues Analysis", *PhD Thesis*, 2013, (DOI:10.4233/uuid:570088d7-e134-4da1-b952-c70302ee77e9).