Pattern Recognition Letters 34 (2013) 1879–1889

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Shopping behavior recognition using a language modeling analogy

M.C. Popa [a,b,*], L.J.M. Rothkrantz [a,c], P. Wiggers [a], C. Shan [b]

[a] Section of Interactive Intelligence, Department of Intelligence Systems, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
[b] Video and Image Processing Department, Philips Research, HTC 36, 5656 AE Eindhoven, The Netherlands
[c] Sensor Technology, SEWACO Department, Netherlands Defence Academy, Nieuwe Diep 8, 1781 AC Den Helder, The Netherlands

ARTICLE INFO

Article history:
Available online 7 December 2012

Keywords:
Shopping behavior
Semantic analysis
Language model
Action recognition
Hidden Markov Models

ABSTRACT

Automatic understanding and recognition of human shopping behavior has many potential applications, attracting an increasing interest in the marketing domain. The reliability and performance of the automatic recognition system is highly influenced by the adopted theoretical model of behavior. In this work, we address the analogy between human shopping behavior and a natural language. The adopted methodology associates low-level information extracted from video data with semantic information using the proposed behavior language model. Our contribution on the action recognition level consists of proposing a new feature set which fuses Histograms of Optical Flow (HOF) with directional features. On the behavior level we propose combining smoothed bi-grams with the maximum dependency in a chain of conditional probabilities. The experiments are performed on both laboratory and real-life datasets. The introduced behavior language model achieves an accuracy of 87% on the laboratory data and 76% on the real-life dataset, an improvement of 11% and 8% respectively over the baseline model, by incorporating semantic knowledge and capturing correlations between the basic actions.

1. Introduction

Creating machines that can understand human behavior has been a goal of Artificial Intelligence (AI) since its beginning. However, even today's most successful algorithms and systems have only a partial knowledge of the meaning of a person's behavior. Advances in technology have known a tremendous progress in the past years, enabling tracking, motion detection, action recognition, and to some extent behavior understanding (Candamo et al., 2010). The area of application domains to which awareness of the users' behavior could contribute, ranges from affective computing, gaming industry, surveillance, to elderly care or marketing. In the marketing domain it is of great interest to build a satisfactory relation with the customer, by assessing his/her emotional state (Tsai and Huang, 2002) and intentions. The shopping experience could be enhanced by facilitating easy access to the products for which the customer shows interest or by offering timely assistance whenever a customer needs help in finding or selecting a product. Furthermore, understanding customers' behavior has many other advantages, such as gathering statistics regarding shoppers' preferences in terms of products, moments of the day to visit the shop, or areas in the shop, which could lead to an optimized marketing strategy both for the company and the customers (Karolefski, 2012). Automatic recognition and assessment of human shopping behavior is definitely preferred over the traditional methods represented by questionnaires, interviews, or human observations. Still this task is a challenging one and needs to be modeled accordingly.

Therefore, we need first to introduce the adopted terminology for human behavior, which was inspired from Candamo et al. (2010). In this paper, the authors define human behavior as a composition of multiple events, where an event represents a single spatio-temporal entity which cannot be further decomposed (e.g. a person standing or walking). Given that we aim at placing human behavior in a specific context, namely shopping environment, we continue by identifying the relevant types of interactions. Human shopping behavior can be subdivided in several groups depending on the type of interaction and the number of involved participants: (1) single person or no interaction; (2) person–products interactions; (3) person–shopping basket/cart interactions; (4) person–location interactions; (5) multiple–person interactions.

We propose in this paper a representation of the shopping behavior analogous to a language processing model (Chatera and Manning, 2006). We mention that we use probabilistic language processing in the same manner as in speech recognition or information retrieval, where the probability of a sequence of words is computed based on the statistics of words frequencies in a corpora of real language (Hiemstra and de Jong, 1999). Following a top-down approach, behavior can be decomposed into activities, activities into basic actions, and finally basic actions into human

* Corresponding author at: Section of Interactive Intelligence, Department of Intelligence Systems, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands. Tel.: +31 (0) 15 2781537; fax: +31 (0) 15 2787141.
E-mail addresses: m.c.popa@tudelft.nl, mirela.popa@philips.com (M.C. Popa).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.11.015

postures. In the same manner, spoken language is composed of sentences, sentences are formed of words and words at their turn can be split into phonemes. Still it is well known that not every combination of phonemes forms a word (e.g. "ghklf") and that some sequences are more likely than others. Furthermore, regarding words, not any combination of words makes sense and can form a sentence (e.g. "man world try give home"). Similar rules apply also for sequences of basic actions, some combinations being very improbable to happen (e.g. "pick an item, take it off, check how it looks on you"), while others are very likely because they are performed in a meaningful order ("pick an item, check its characteristics, try it on, check how it looks in the mirror, and take it off"). The proposed analogy refers both to the hierarchical composition of behavior/language and to their structure which follows certain rules. Human behavior can be expressed according to a grammar, which sets rules to the basic actions order from a semantic point of view. Chomsky (1955, 1957) introduced in his theory about transformational grammar, claiming that sequences of words have a syntax characterized by a formal grammar, in particular a context free grammar extended with transformation rules. A string of basic symbols (letters from an alphabet) is considered as grammatically correct if it is composed from basic symbols using composition rules, or from the other side can be decomposed into basic symbols using decomposition rules. His theory was in line with the AI view around that time favouring a rule based approach. According to Chomsky, probabilistic models give no insight in the basic problems of syntactic structure. We agree with that, but in our work we use a computational approach, inspired by the work of Jelinek (1976) in the area of speech recognition. We apply statistical models for the recognition of shopping behavior. Given a video sequence we compute the probabilities of basic shopping actions represented in that interval. Next, we use HMMs to compute the most probable sequence of basic shopping actions. The recognition of the behavior pattern is improved by applying a language model, which computes based on statistics the most likely sequence of basic shopping actions from a semantic point of view. Furthermore, in speech recognition, context plays an important role, contributing to the overall recognition accuracy (Kimberly et al., 2008). Visual scene understanding and human behavior recognition is also affected by contextual knowledge and could benefit a lot from this type of information.

Therefore, we propose in this paper an approach towards shopping behavior modeling inspired from a probabilistic language processing model, which combines in an efficient manner the characteristics of the vision model with semantic information as depicted in Fig. 1. In speech recognition systems, a language model is used to complement the results of the acoustic model which models the relation between words (or parts of words called phonemes) and the acoustic signal.

The proposed approach is a bottom-up one, consisting of recognizing basic actions related to shopping on the low-level, detecting sequences of basic actions on the intermediary level which are translated into activities, and finally recognizing different types of shopping behavior as combinations of activities, on the high-semantic level. Previously, we investigated different methods towards basic action recognition (Popa et al., 2011a,b), while the focus of this work is on shopping behavior modeling. The outline of the paper is as follows. In Section 2 we give an overview of related work. Next, we provide in Section 3 the theoretical foundations of the proposed shopping behavior model. We introduce in Section 4 the behavior recognition model, consisting of the feature extraction methods and classification technique. Next, we provide a description of the used datasets and the experimental results in Section 5. Finally, we formulate our conclusions and give directions for future work.

2. Related work

Automatic analysis of human behavior has been attracting a lot of attention lately due to its potential applications in a variety of domains and posed scientific challenges. To achieve such a challenging task, several research fields focus on modeling human behavior and several methods for modeling behavior have been investigated. Probabilistic graphical models represent one possible alternative due to their properties, such as the ability to capture uncertainty and of generalising to complex activities. Common graphical models for interpreting behavior include static (Intille and Bobick, 1999) and dynamic Bayesian networks (Du et al., 2006), latent Dirichlet allocation model (Mehran et al., 2009), context free-grammars (Brand, 1996), and stochastic context-free grammars (Ivanov and Bobick, 2000). Besides probabilistic graphical models, there are also rule-based methods (Tran and Davis,

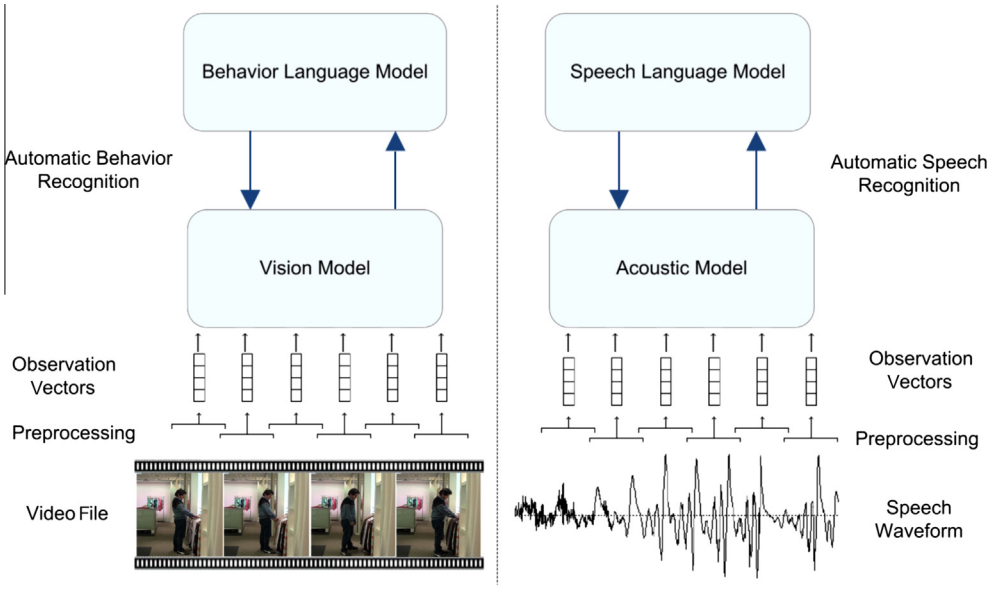

Fig. 1. Analogy between behavior recognition and speech recognition.