

University of Groningen

Lifestyle understanding through the analysis of egocentric photo-streams

Talavera Martínez, Estefanía

DOI:
[10.33612/diss.112971105](https://doi.org/10.33612/diss.112971105)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Talavera Martínez, E. (2020). *Lifestyle understanding through the analysis of egocentric photo-streams*. Rijksuniversiteit Groningen. <https://doi.org/10.33612/diss.112971105>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 7

Summary and Outlook

7.1 Work Summary

This dissertation provides with several solutions for the understanding of the lifestyle and behavioural patterns from egocentric photo-streams collected with wearable cameras. Five main applications have guided this work: the temporal segmentation of egocentric photo-streams; the discovery of Routine and Non-routine related days; the classification of food-scenes in egocentric images; the recognition of induced sentiment when reviewing own collected photo-streams; and the characterization of social interactions.

Deep learning has had a huge impact in the computer vision community for the classification and description of images. In particular, in this thesis, we have relied on the use of these techniques for the above-mentioned applications. Due to the limited amount of collected data, we use *transfer learning* theory in the different classification problems that we have addressed. Moreover, we made use of detected objects, places, and faces for the semantic description of the images. This obtained information allowed us to built on top of pre-trained models for the understanding of the lifestyle and behavioural patterns of the camera wearer.

Egocentric photo-streams describe a first-person view of the life of the camera wearer. These images are collected with a wearable camera that usually has a low frame-rate and describe where users spend time by following their body movements. This fact leads to consecutive highly visual different images when the user walks, as well as to similar ones when the user stays static doing some activity, such as *watching tv*. For the summarization and analysis of specific activities or time frames, temporal segmentation of egocentric photo-streams is a useful tool. In this work, we introduced a novel temporal segmentation model based on the hierarchical clustering method which computed the relation among images based on global and semantic features extracted from them. The obtained segmentation was coherent among several manual segmentations provided by different people, showing its robustness for further applications within the egocentric vision field.

The analysis of behavioural patterns relates to the description of the routine. We proposed a model for the discovery of behavioural patterns relying on topic modelling for the automatic finding of abstract topics within the recorded days. Topic modelling is a method for natural language analysis. Therefore, we translate the collected days into documents, which are composed of the detected objects in the images that constitute the photo-stream. Discovered abstract topics are evaluated when analyzing the documents collected by all users, or just at a personal level. We also evaluate the performance of the found topics for the task of classifying the collected days into Routine or Non-Routine related. We also test the performance of the model when evaluating time slots of different duration. The performed experiments establish a robust baseline showing that it is feasible to get insight into the behavioural patterns of people.

The places where people spend their time describe their lifestyle. More specifically, the food-related scenes showed to have an impact on their health. Therefore, the identification of food-related scenes in the collected photo-streams is critical for a better understanding of the lifestyle of a person. We proposed a hierarchical model for the classification of egocentric images into 15 different food-related scenes. Food-related scenes are visually and semantically related. Therefore, we have introduced a taxonomy describing the relationship between the studied classes. This taxonomy allows the analysis of the collected photo-streams through activity and location label. The proposed model adapts to the proposed taxonomy. The first stage of classification is between three food-related activities: eating, preparing and acquiring food. The final classification differentiates among the 15 proposed classes: bakery, bar, beer hall, cafeteria, coffee shop, dining room, food court, ice cream parlour, kitchen, market indoor, market outdoor, picnic area, pub indoor, restaurant, and supermarket. In order to give a robust classification output, the final classification probabilities for a given image are computed as the multiplication of prior probabilities of the given classification tree. The proposed model has shown to be a powerful tool for the later characterization of the nutritional habits of the camera wearer.

The analysis of inferred sentiment by reviewing past experiences through the collected photo-streams started with a collaboration with psychologists who worked with people suffering from depression. We focus on the classification of the images into three main classes: positive, negative, or neutral. The proposed model bases its classification on the semantic features obtained from the images. These semantic features were obtained from an existent model that detected objects with an associated sentiment value, such as: *beautiful view*, *lonely chair*, or *damaged building* with associated values of 1.69, -0.44, and -1.42, respectively. Our hypothesis was that the detected semantics would allow us to describe the feeling the images irradiate to the owner in the reviewing process. However, the available tool was not able to cor-

rectly detect such concepts in our images. Moreover, the final classification ended into: negative images as the non-informative ones; neutral images as the ones that describe work-related scenes; and the positive images were those with scenes related to social interactions, eating, or walking outside. This research was the starting point of the analysis of the routine of a person.

Appearing faces in the collected photo-streams show the social interactions of the camera wearers throughout their days. The detection and analysis of these faces in images is possible with existent and publicly available tools, such as OpenCV and OpenFace. In this work, we made use of such tools for the identification of social interactions and build a model to characterize them. Our proposed model performs person re-identification throughout the sequences for the identification of familiar people. Due to the lack of baseline works in this specific field of research, we proposed several metrics related to the occurrence and duration of the detected social interactions for their quantification.

Benchmarking: Together with the approaches described, we have introduced two novel and home-made datasets: *EgoFoodPlaces* and *EgoRoutine*. The first one is composed of more than 33,000 images and describes 15 different food-related scenes. It is further described in Chapter 4. The latest was collected by 7 different users for periods of at least two weeks. It is composed of 103 days, with a total of more than 100,000 images. This dataset is described in Chapter 3. Both datasets are available on the website of our research group: <http://www.ub.edu/cvub/dataset/>. This will encourage and allow other researchers to evaluate their algorithms with ours. If feasible, it would represent a significant step forward for the automatic and personalized characterization of a person's social life. One could argue that the final usability and applicability of this technology is not ensured. However, in (Gelonch et al., 2019), psychologists discuss that older adults have a high level of acceptance, concluding that the benefits for memory overcome previous privacy concerns.

7.2 Outlook

In this section, we describe ideas and open directions on how the work presented in this dissertation can be extended for future studies.

Fig. 7.1 illustrates the future lines of research that will further develop the proposed applications pipelines in this manuscript for the understanding of human behaviour from collected egocentric photo-streams.

First, we discuss the classification of food-related scenes in collected photo-streams. The improvement of the performance of the proposed classifiers will allow a better characterization of the nutritional routine of people leading to an improvement of

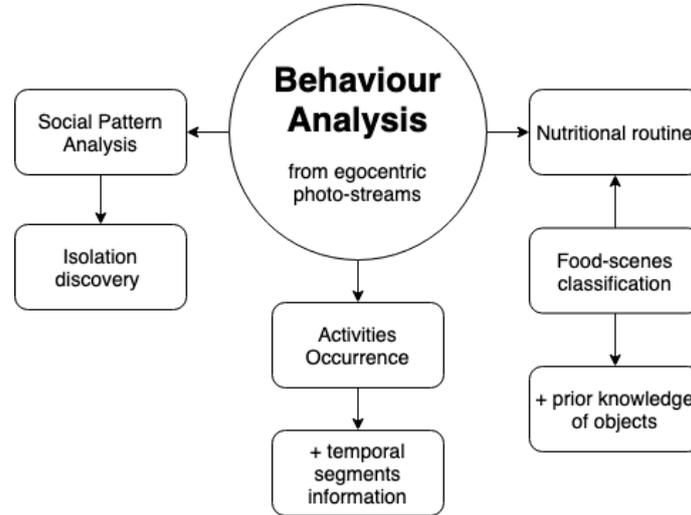


Figure 7.1: An overview of directions for further development for the understanding of behavioural patterns from egocentric photo-streams. These directions include various aspects: the characterization of the social interactions of the camera wearer, with the possibility of detecting isolation; the analysis of the nutritional routine relying on the analysis of occurrence of food-related scenes; and the inclusion of prior knowledge about appearing objects in certain scenes for the improvement of food-scenes recognition; the inclusion of information about temporal boundaries of scenes within the collected days for the analysis of activities in the frame of behavioural analysis.

their healthy lifestyle. Further work will analyze the classification of food-related scenes based on detected concepts in the images. We plan to study how the inclusion of priors of appearing objects in the target scenes affect the final classification. As an example, the final classification of a given image will be modified if it is labelled by the proposed model as *market outdoor* and at the same time, objects like *television* are found with a relatively high probability. We believe that this will help us improve the classification accuracy by avoiding “non-common sense” associations.

The process of identifying behavioural patterns through the discovery of abstract topics showed its potential for the characterization of the lifestyle of the camera wearer. Our proposed model relies on the application of the statistical process of topic modelling to the semantic features obtained from the image. Following this up, we believe that further research should go on the lines of nutritional behaviour and social patterns analysis. The classification of images into food-related scenes is the first step for the analysis of nutritional habits of the camera wearer. A more

personalized advice from specialists can be given if details about the occurrence of food-related scenes in the life of a person are automatically and objectively obtained. This shows the importance of developing tools for the analysis of the nutritional routine of a person. Future work will address the analysis of how days are related based on the food-related activities: eating, preparing and acquiring food. We believe this will help in describing the lifestyle of people for the later improvement of their health.

Second, we propose several research lines following the work done on analysis of behavioural patterns. The field of behaviour analysis is a wide area since different studies can focus on different and specific aspects of the lifestyle of a person. In Chapter 3, we concluded that activity patterns give relevant information that allows us to better perform a distinction among similar days. However, that was a general approach for the classification of days based on detected objects that gives an overview of the behaviour of a person. Following that line, we foresee that the characterization of a person's behavioural patterns can be addressed by the quantification of the performed activities throughout his or her collected days in the form of photo-streams.

In Chapter 6, we addressed the characterization of social patterns of behaviour by quantifying the social relations of the camera wearer. We will explore in future works the analysis of social habits and social activities occurrence through the collected days. We believe that this analysis will allow a better understanding of the habits of the person, helping to automatically detect isolation or certain behaviours related to specific disorders such as a depression.

Finally, dividing days into time-slots has helped us to characterize them more accurately and to when classify them into Routine and Non-routine compare them by maintaining temporal information. Future work will evaluate the performance of the proposed model for behaviour analysis when including information about temporal boundaries of events happening through the day. This information might be useful for the definition of time-slots to be compared, making the comparison more flexible with respect to activities duration. Methods such as the one proposed in Chapter 2 (Dimiccoli et al., 2017) can be used for the detection of boundaries within the collected days.

