Lifestyle understanding through the analysis of egocentric photo-streams

Talavera Martínez, Estefanía

DOI:
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):

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 1

Introduction

How can we improve and contribute to the people’s quality of life? The personal development process is described as the assessment of people’s qualities and behaviours. By tracking people’s daily behaviours we can help them draw a picture of their lifestyle. The obtained information can be used to later improve their personal development (Ryff, 1995). However, the self-awareness and personal development process are not trivial. They include the enhancement of the following activities: self-knowledge, health, strengths, aspirations, social relations, enhancing lifestyle, quality of life and time-management, among others (Ryff, 1995). For instance, the quantification of their daily activities helps to define goals for future changes and/or advances in their personal needs and ambitions.

This thesis addresses the development of automatic computer vision tools for the study of people’s behaviours. To this end, we rely on the analysis of egocentric photo-streams recorded by a wearable camera. These pictures show an egocentric view of the camera wearers’ experiences, allowing an objective description of their days (Bolaños et al., 2017). They describe the users’ daily activities, including people they meet, time spent working on their computers, outdoor activities, sports, eating, or shopping. The first-view perspective shown by the images describes how the lives of the camera wearers look like. We believe that this data is a powerful source of information since it is a raw description of the behaviours of people in society. Our goal is to demonstrate that egocentric images can help us draw a picture of the days of the camera wearer, that can be used to improve the healthy living of individuals.

1.1 Scope

This thesis aims to develop and introduce automatic computer vision tools that allow the study and characterization of the lifestyle of people. To do so, we rely on egocentric images recorded by wearable cameras, see Fig. 1.2. An egocentric photo-stream or egocentric photo-sequence is defined as a collection of temporal consecutive images. Fig. 1.1 illustrates a collection of photo-streams recorded by a camera.
1. Introduction

Figure 1.1: Illustration of recorded days in the form of egocentric photo-streams. These images were collected by the Narrative Clip wearable camera and describe the life of the camera wearer. The information that we can obtain from the recorded photo-streams is broad because of the wide range of applications that can be addressed. More specifically, in this work, we focus on the analysis of the following behavioural traits:

- **Temporal segmentation**: Days are composed of moments when the camera wearer spends time at certain environments. To find such moments, we look for sequences of similar images. Given an egocentric photo-sequence, our model decides the temporal boundaries that divided the photo-stream into moments based on the global and semantic features of the images.

- **Routine discovery**: Implement an automatic tool for the discovery of Routine-related days among days recorded by different users. To this end, we evaluate the role of semantics extracted from the egocentric photo-streams.

- **Recognition of food-related scenes**: Identify food-related environments where the user spends time to describe food-related activity routines.

- **Sentiment retrieval**: Given images describing recorded scenes by the user the aim is to determine their sentiment associated based on the extraction of either visual features or semantic concepts with sentiment associated, or their combination.

- **Social pattern characterization**: Provide an automated description of patterns of the experienced social interactions, according to the detection of people and the occurrence of their appearance throughout the recorded photo-streams.
1.1. Scope

Egocentric images describe from an egocentric point of view the wearer’s life. The extracted information allows us to get insight into the lifestyle of the camera users, for the later improvement of their health. Moreover, wearable cameras are lightweight, financially affordable and with potential for other applications to assist or improve the quality of life of people.

1.1.1 Societal impact

Nowadays, describing people’s lives has become a hot topic in several disciplines. In psychology, this topic is addressed aiming to help ordinary people, and especially people with some kind of need (Martin et al., 1986; de Haan et al., 1997; Yesavage, 1983), where an automatic evaluation of lifestyle would be of much help for the practitioners.

Healthy ageing is of relevance due to the ever-increasing number of elderly people in the population. These collections of digital data can serve as cues to trigger autobiographical memory about past events and can be used as an important tool for prevention or hindrance of cognitive and functional decline in elderly people (Doherty et al., 2013), and memory enhancement (Lee and Dey, 2008). In (Sellen et al., 2007), it was discussed that if memory cues are provided to people suffering from Mild Cognitive Impairment (MCI), they would be helped to mentally ‘re-live’ specific past life experiences. Studies have shown how different cues, such as: time, place, people, and events, trigger autobiographical memories, suggesting that place, events, and people are the stronger ones. A collaboration with neuropsychologist from the Hospital of Terrassa, Spain, shows the good acceptability of older adults of wearable devices, where the potential benefits for memory outweigh concerns related to privacy (Gelonch et al., 2019). Our novel proposed system will contribute to healthy ageing by improving peace of mind of elderly people. The developed models that we applied in such situations have shown promising outcomes.

In the last few years, there has been an exponential increase in the use of self-monitoring devices (Trickler, 2013) by ordinary people who want to get to know themselves better. These devices offer information about daily habits, by logging daily data of the user, such as: how many steps the user walks (Cadmus-Bertram et al., 2015), how and how long smartphones and apps are used (Wei et al., 2011), heart-rate with the use of smart bracelets or watches (Reeder and David, 2016), to name some. People want to increase their self-knowledge automatically, expecting that it will lead to psychological well-being and the improvement of their lifestyle (Ryff, 1995). Self-knowledge is a psychology term that describes a person’s answers to the question “How am I like?” (Neisser, 1988). To answer this, there is often need of external information mainly because of two causes. On one side, it is a difficult
1. Introduction

The task to describe our behavioural patterns. On the other side, we tend to alter and not be accurate when describing what it is like (Silvia and Gendolla, 2001).

From another point of view, big companies started looking for information about their employees and clients with the aim of improving productivity and customer acquisition (Chin et al., 2011; Sanlier and Seren Karakus, 2010; Spiliopoulou et al., 1999). Furthermore, behavioural psychologists from the University of Otago, New Zealand, already shown their interest in this tool since they are working on the characterization of the lifestyle of students. The identification of which whom students tend to interact and the duration of such interactions is of high importance when aiming to understand their daily habits, and ultimately improve them.

1.1 Privacy issues

Personal data relates to any information that can be obtained from the living of an individual. The use of wearable devices to track our lifestyle can be seen as intrusive, but can help to promote life-enhancing. Following the General Data Protection Regulation (EU) 2016/679 (GDPR), we consider data protection and ensure personal data privacy from different perspectives:

- **Researchers**: People working on the analysis of the collected data were asked to sign a consent form confirming that they will use the data for research purposes, respecting the privacy of the participants.

- **Participants**: Camera wearers were asked to give their written consent for the later use of their collected data. The collected data is then linked to an identifier that ensures the anonymization of the camera user. In the case of models were detected faces are needed for the analysis, we do not blur the identity of the persons with whom the participant interacts, but we do ask for their consent to be part of the dataset. The participants have the right to revoke their consent at any time.

1.2 Background

In this section, we describe the main concepts that we refer to throughout this thesis, such as lifelogging, wearable cameras, egocentric vision, and egocentric photo-streams. Moreover, we briefly introduce the framework of the different applications that we later describe and address in the following chapters of this thesis.

Before the emergence of static and wearable sensors, people’s daily habits were manually recorded. For instance, Activities of Daily Living (ADL) were manually
1.2. Background

annotated by either individual users and/or specialists, as in (Andersen et al., 2004; Wood et al., 2002). In (Andersen et al., 2004), manually recorded information about the ability of someone’s ADL performance was examined to classify the patients’ dependence, as either dependent or independent.

**Lifelogging** Nowadays, the development of new wearable technologies allows to automatically record data from our daily living. Lifelogging appeared in the 1960s as the process of recording and tracking personal activity data generated by the daily behaviour of a person. Through the analysis of recorded visual data, information about the lifestyle of the camera wearer can be obtained and retrieved. By recording people’s own view of the world, lifelogging opens new questions and goes a step forward to the desired and personalized analysis of the lifestyle of individuals. The objective perspective offered by the recorded data of what happened during different moments of the day, represents a robust tool for the analysis of the lifestyle of people.

![Figure 1.2: Wearable camera - Narrative Clip.](image)

**Cameras** Among the advances in wearable technology during the last few years, wearable cameras specifically have gained more popularity (Bolaños et al., 2017). In Fig. 1.3 we present some examples of wearable cameras that are available on the market. These cameras are used for different purposes and have different specifications (see Table 1.1). All the mentioned devices allow capturing high-quality images in a hands-free fashion from the first-person point of view.

Wearable video cameras, such as GoPro and Looxcie, which have a relatively high frame rate, ranging from 25 to 60 fps, are mostly used for recording the user activities for a few hours. Instead, wearable photo cameras, such as the Narrative
Clip and SenseCam, capture only 2 or 3 fpm and are therefore mostly used for image acquisition during longer periods of time (e.g. a whole day). By using wearable cameras with a low temporal resolution, the camera wearer captures each day up to 1000 egocentric photo-streams.

Figure 1.3: Some of the available wearable cameras that can be found on the market. While the a) torso mounted cameras are commonly used for visual diary creation and security, the (b) glass mounted wearable cameras are often used for augmented reality [Google Glasses and Spectacles]. Finally, the (c) head mounted cameras are used for recording sports and leisure activities [GoPro and Polaroid Cube].

<table>
<thead>
<tr>
<th>Camera</th>
<th>Main use</th>
<th>Temporal Resolution (FPS/FPM)</th>
<th>Worn on</th>
<th>Size (mm)</th>
<th>Weight (gr.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoPro Hero5</td>
<td>Entertainment</td>
<td>High (60fps)</td>
<td>Head and Torso</td>
<td>38x38</td>
<td>73</td>
</tr>
<tr>
<td>Google Glasses</td>
<td>Augmented Reality</td>
<td>High (60fps)</td>
<td>Head</td>
<td>up to 133.3x203</td>
<td>36</td>
</tr>
<tr>
<td>Spectacles</td>
<td>Social Networks</td>
<td>High (60 fps)</td>
<td>Head</td>
<td>33x149</td>
<td>48</td>
</tr>
<tr>
<td>Axon Body 2</td>
<td>Security</td>
<td>High (30 fps)</td>
<td>Torso</td>
<td>70x87</td>
<td>141</td>
</tr>
<tr>
<td>Narrative Clip 2</td>
<td>Lifelogging</td>
<td>Low (2-3fps)</td>
<td>Torso</td>
<td>36x36</td>
<td>19</td>
</tr>
<tr>
<td>SenseCam</td>
<td>Lifelogging</td>
<td>Low (2 fps)</td>
<td>Torso</td>
<td>74x50</td>
<td>90</td>
</tr>
<tr>
<td>Autographer</td>
<td>Lifelogging</td>
<td>Low (2-3fps)</td>
<td>Torso</td>
<td>90x36</td>
<td>58</td>
</tr>
</tbody>
</table>

Egocentric Photo-streams The recorded photo-streams offer a first-person view of the world (see Fig. [1.1]). The big advantage of image-based lifelogging is that it gives rich information able to generate explanations and visualize the circumstances of the person’s activities, scenes, state, environment and social context that influence his/her way of life, as it defines the contextual information. Through the analysis of images collected by continuously recording the user’s life, information about daily routines, eating habits, or positive memories can be obtained and retrieved.

1.2.1 Temporal Segmentation

Egocentric photo-streams generally appear in the form of long unstructured sequences of images, often with a high degree of redundancy and abrupt appearance changes even in temporally adjacent frames, that harden the extraction of semantically meaningful content. Temporal segmentation, the process of organizing un-
structured data into homogeneous chapters, provides a large potential for extracting semantic information. Video segmentation aims to temporally divide the video into different groups of consecutive images called events or scenes that describe the performance of an activity or a specific environment where the user is spending time (see Figure 1.4). Many segmentation techniques have been proposed in the literature in an attempt to deal with this problem, such as video summarization based on clustering methods or object detection. The work described in Goldman et al. (2006) was a first approach where the user selected the frames considered important as key-frames (considered as the frame that best represents the scene), generating the storyboard that reported object’s trajectory. Other studies incorporate audio or linguistic information (Nam and Tewfik, 1999; Smith, 1997) to the segmentation approach looking for the semantic meaning of the video.

![Figure 1.4: Example of temporal segmentation of an egocentric sequence based on what the camera wearer sees. In addition to the segmentation, our method provides a set of semantic attributes that characterize each segment.]

We believe that the division of the photo-stream into a set of homogeneous and manageable segments is important for the better characterization of the collection of images. Each segment can be represented by a small number of key-frames and indexed by semantic features. This division provides a basis for understanding the semantic structure of the event. Hence, in this work, we aim to study and discuss the following related research lines: Can we obtain a good enough division of the recorded photo-streams into events? Which are the features that help us achieve the best temporal segmentation? Is the manually temporal segmentation process robust on its own? These questions will be developed in Chapter 2 of this thesis.

1.2.2 Routine Discovery

Human behaviour analysis is of high interest in our society and a recent research area in computer vision. Routine-related days have common patterns that describe situations of the daily life of a person. More specifically, routine was described as
regularity in the activity in Sevtsuk and Ratti (2010). Fig 1.5 is an illustration of what can be considered as the routine of a person. Social psychologists exposed in Society for Personality and Social Psychology (2014) that each day 40% of people’s daily activities are performed in similar situations. However, Routine has no concrete definition, since it varies depending on the lifestyle of the individual under study. Therefore, supervised approaches are not useful due to the need for prior information in the form of annotated data or predefined categories. For the discovery of routine-related days, unsupervised methods are necessary to enable an analysis of the dataset with minimal prior knowledge. Moreover, we need to apply automatic methods that can extract and group the days of an individual using correlated daily elements. We address the discovery of routine-related days following two different approaches:

- On one side, we evaluate outlier detection methods for the discovery of clusters corresponding to routine-related days, i.e. outliers to non-routine related days. In this approach, days are described as the aggregation of the images’ global features.

- On the other side, we propose a novel automatic unsupervised pipeline for the identification and characterization of routine-related days from egocentric photo-streams. We perform an ablation study at different levels of the proposed architecture for the characterization and comparison of days.

Figure 1.5: The routine of the camera wearer is described by his or her performed activities throughout the days. We aim to discover the daily habits of people to get a better understanding of their behaviour.

Together with the proposed models, we introduce EgoRoutine, a new egocentric dataset composed of a total of 100,000 images, from 104 days. Further description of the proposed methodology and experiments can be found in Chapter 3.
1.2.3 Food Related scene classification

From another perspective, nutritional habits are of importance for the understanding of the lifestyle of a person. Recent studies in nutrition argue that it is not only important what people eat but also how/where people eat (Laska et al., 2015). We propose the analysis of collected egocentric photo-streams for the automatic characterization and monitoring of the health habits of the camera wearer. To this end, we focus on the classification of 15 different food-related scenes. Scenes recorded by an egocentric perspective and related to food consumption, acquisition or preparation share visual information, which makes difficult to distinguish them. Therefore, we propose a hierarchical classification model that organizes the classes based on their semantic relation. We illustrate the three main food-related activities and some of the scenes in Fig. 1.6. The intermediate probabilities help to improve the final classification by re-enforcing the predictions of the classifiers. There are no previous works on this field, and therefore, the proposed model represents the baseline for food-related scenes classification.

Moreover, we propose and make publicly available EgoFoodPlaces, a dataset composed of more than 33,000 images representing food-related scenes. We describe EgoFoodPlaces, the proposed model and the performed experiments in Chapter 4 of this thesis.

Figure 1.6: Daily health habits related to food consumption, acquisition or preparation can be studied by the examination of recorded egocentric photo-streams. The analysis of food-related scenes and activities can help us understand the lifestyle of the camera wearer for the improvement of his or her nutritional behaviour.
1.2.4 Inferring associated sentiment to images

Understanding emotions plays an important role in personal growth and development, and gives insight into how human intelligence works. Moreover, selected memories can be used as a tool for mental imagery, which is described as the process in which the feeling of an experience is imagined by a person in the absence of external stimuli. The process of reliving previous experiences is illustrated in Fig. 1.7. Therapists assumed it is directly related to emotions (Holmes and et al., 2006), opening some questions when images describing past moments of our lives are available: Can an image facilitate the process of mental imagery? or Can specific images help us to retrieve or imply feelings and moods? Semantic concepts extracted from the collection of egocentric images help us describing the emotions related to memories that the photo-streams capture.

Part of the recorded egocentric images are redundant, non-informative or routine and thus without special value for the wearer to be preserved. Usually, users are interested in keeping special moments, images with sentiments that will allow them in the future to re-live the personal moments captured by the camera. An automatic tool for sentiment analysis of egocentric images is of high interest to make possible the processing of the big collection of lifelogging data and keeping out just the images of interest i.e. of high charge of positive sentiments. To the best of our knowledge, no previous works had addressed this topic from egocentric photo-streams in the literature. In Chapter 5 we study how egocentric images can be analyzed to discover events that would invoke positive, neutral or negative feelings to the user.

Figure 1.7: Illustration of a camera user reviewing his or her collected events, being affected by their associated sentiment.
1.2. Background

1.2.5 Social pattern analysis

Human social behaviour involves how people influence and interact with others, and how they are affected by others. This behaviour varies depending on the person and is influenced by ethics, attitudes, or culture (Allport 1985). Understanding the behaviour of an individual is of high interest in social psychology. In (House et al., 1988), the authors addressed the problem of how social relationships affect health and demonstrated that social isolation leads to major risk factors for mortality. Moreover, in (Yang et al., 2016), the authors observed that lack of social connections is associated with health risks in specific life stages, such as the risk of inflammation in adolescence, or hypertension in old age. Also, as in (Kawachi and Berkman, 2001) it was highlighted that social ties have a beneficial effect when maintaining psychological well-being.

Considering the importance of the matter, automatic discovery and understanding of the social interactions are of high importance to the scientists, as they remove the need for manual labour. On the other hand, egocentric cameras are useful tools as they offer the opportunity to obtain images of the daily activities of users from their own perspective. Therefore, providing a tool for automatic detection and characterization of social interactions through these recorded visual data can lead to personalized social pattern discoveries, see Fig. 1.8. We discuss the proposed model and findings in Chapter 6.

![Figure 1.8: Example of social profile given a set of collected photo-streams associated with one person. First, we detect appearing faces in the photo-streams. Later, we apply the OpenFace tool to convert the faces into feature vectors. We propose to define the re-identification problem as a clustering problem with a later analysis of the grouped faces occurrence.](image-url)
1.3 Objectives

The main goal of this dissertation is to give appropriate tools for the analysis and interpretation of egocentric photo-streams for the understanding of the behavioural patterns of the camera wearer. Given the previous general lines that represent the ground of this thesis, we defined the following particular objectives:

- To temporally segment egocentric photo-streams into moments within the day for their later analysis according to global and semantic features extracted from the images.
- To provide an automatic tool for routine discovery through the recognition of days with similar patterns within the egocentric photo-streams collection.
- To automatically classify egocentric photo-streams into food-related scenes to get an understanding of the user’s eating habits.
- To define a simple social pattern analysis framework to compare different user’s social behavioural patterns.
- To identify the sentiment that a retrieved moment would provoke the users when reviewing it.

1.4 Research Contributions

This thesis argues that behavioural patterns can be analysed in the domain of egocentric photo-streams since they represent a first-person perspective of the life experiences of the camera user. The analysis of egocentric photo-streams allows us to extract information which gives us insight into the lifestyle of the camera wearer. Our contributions aim to improve a person’s lifestyle. The presented models can be easily adapted for personalized behavioural patterns analysis from images recorded from a first-person view.

Specifically, the contributions of this thesis can be summarized as follows:

1. Due to the free movement of the camera and its low frame rate, abrupt changes are visible even among temporally adjacent images (see Fig. 2.1 and Fig. 2.8). Under these conditions motion and low-level features such as colour or image layout are prone to fail for event representation, hence urges the need to incorporate higher-level semantic information. Instead of representing each image by its contextual global features, which capture the basic environment appearance, we detect segments as a set of temporally adjacent images with the same
1.4. Research Contributions

contextual representation in terms of semantic visual concepts. Nonetheless, not all the semantic concepts in an image are equally discriminant for environment classification: objects like trees and buildings can be more discriminant than objects like dogs or mobile phones, since the former characterizes a specific environment such as forest or street, whereas the latter can be found in many different environments. In this paper, we propose a method called Semantic Regularized Clustering (SR-Clustering), which takes into account semantic concepts in the image together with the global image context for event representation. These are the contributions within this line of research:

- Methodology for the description of egocentric photo-streams based on semantic information.
- Set of evaluation metrics applied to ground truth consistency estimation.
- Evaluation on an extensive number of datasets, including our own, which was published with this work.
- Exhaustive evaluation on a broader number of methods to compare with.

The proposed model for temporal segmentation was published as [Talavera et al., 2015] and [Dimiccoli et al., 2017].

2. We address for the first time the discovery of routine-related days from egocentric photo-streams. With this aim, we propose two different approaches. On one hand, we propose an unsupervised and automatic model for the discovery of routine days following a novelty detection approach. This model is based on the analysis of the aggregation of descriptors of the images within the photo-stream. We tested the proposed model over a home-made collected egocentric dataset. This dataset describes the daily life of the camera wearers. It is composed of a total of 73,000 images, from 72 recorded days by 5 different users. We name this dataset EgoRoutine. This work was published in a conference as [Talavera et al., 2019]. On the other hand, we introduce a novel automatic unsupervised pipeline for the identification and characterization of Routine-related days from egocentric photo-streams. In our proposed model we first extract semantic features from the egocentric images in terms of detected concepts. Later, we translate them to documents following the temporal distribution of the labels. To do so, the detected words in images that were recorded during pre-defined time-slots define a document. Then, we apply the topic modelling technique to the created documents to find abstract topics related to the person’s behaviour and his/her daily habits. We prove that topic modelling is a powerful tool for pattern discovery when addressing Bag-of-Words representation of photo-streams. Later, Dynamic Time Warping
1. Introduction

(DTW) and Spectral Clustering are applied for the unsupervised routine discovery. We prove that using DTW and Distance-based clustering is a robust technique to detect the cluster of routine days being tolerant to small temporal differences in the daily events. The proposed pipeline is evaluated over an extension of the previous EgoRoutine dataset, which is composed of more than 100,000 images, from 104 days collected by 7 different users. This work was submitted and is currently under review.

3. A novel model for food-related scenes classification is introduced in Chapter 4. Food-related scenes that commonly appear in the collected egocentric photo-streams tend to be semantically related. There exists a high intra-class variance in addition to not a high inter-class similarity, leading to a challenging classification task. To face this classification problem the contributions of the chapter are three-fold. On one side, we define a taxonomy with the relation of the studied classes, where food-related environments are organized in a fine-grained way that take into account the main food-related activities (eating, cooking, buying, etc). On the other side, we propose a hierarchical model composed of different layers of deep neural networks. The model is adapted to the defined taxonomy for food-related scenes classification in egocentric photo-streams. Our hierarchical model can classify at the different levels of the taxonomy. Finally, we introduce a new egocentric dataset of more than 33,000 images describing 15 food-related environments. We call it FoodEgoPlaces and along with its ground-truth are publicly available in http://www.ub.edu/cvub/dataset/. This work is published as (Talavera et al., 2014).

4. We present innovative models for emotion classification in egocentric photo-streams setting, see Chapter 5. In this chapter, we present two models: one is based on the analysis of semantic concepts extracted from images that belong to the same event, while the other analyses the combination of semantic concepts and general visual features of such images. In our proposed analysis, we evaluate the role of considered semantic concepts in terms of Adjective-Noun-Pairs (ANPs), given that they have sentiment values associated (Borth et al., 2013), and their combination with general visual features extracted with a CNN (Krizhevsky, Sulskever and Hinton, 2012). With this work, we prove the importance of such a combination in the invoked sentiment detection. Moreover, we test our method on a new egocentric dataset of 12,088 pictures with ternary sentiment values acquired from 3 users in a total of 20 days. Our contribution is an analytic tool for positive emotion retrieval seeking events that best represent a pleasant moment to be invoked within the whole set of a
day photo-stream. We focus on the event’s sentiment description from an objective point of view of the moment under analysis. The results given in this chapter are published in two conferences (Talavera, Radeva and Petkov 2017; Talavera, Strisciuglio, Petkov and Radeva 2017).

5. We propose a method that enables us to automatically analyse and answer to questions such as *Do I socialize throughout my days?* or *With how many people do I interact daily?*. To do so, we rely on the analysis of egocentric photo-streams. Given sets of captured days by camera wearers, our proposed model employs a person re-identification model to achieve social pattern descriptions. First, a Haar-like feature-based cascade classifier is applied (Viola et al., 2001) to detect the appearing faces in the photo-streams. Detected faces in this step are converted into feature descriptors by applying the OpenFace tool (Amos et al., 2016). Finally, we propose to define the person re-identification problem as a clustering problem. The clustering is applied over the pile of photo-streams recorded by the users along the days to find the recurrent faces within photo-streams. Shaping an idea about the social behaviour of the users becomes possible through referring to the time and day when the recurrences were appearing. The proposed work was presented in a conference as (Talavera et al., 2018).

### 1.5 Thesis Organization

The remaining chapters of this thesis are organised as follows: Chapter 2 describes our proposed temporal segmentation method, which divides egocentric sequences into sequential similar images, that we call *events*. In Chapter 3, we present an automatic model for the discovery of routine-related days from the photo-stream collection of a user. Following, in Chapter 4, we introduce a hierarchical network for the classification of images into food-related scenes. Later, in Chapter 5, we address the recognition of what an image would invoke to the camera wearer. In Chapter 6, we focus on the analysis of social interactions of the user to then infer a social pattern that describes his or her social daily behaviour. Finally, Chapter 7 provides a summary of the thesis and gives an outlook of how the proposed techniques can be developed further and applied in different computer vision applications.