Pattern Recognition 104 (2020) 107330

Contents lists available at ScienceDirect

Pattern Recognition

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

Topic modelling for routine discovery from egocentric photo-streams

Estefania Talavera[a,b,*], Carolin Wuerich[b], Nicolai Petkov[a], Petia Radeva[b]

[a] University of Groningen, Johann Bernoulli Institue, Nijenborgh 9, 9747 AG Groningen, Netherlands
[b] University of Barcelona, Department Mathematics and Computer Science and Computer Vision Center, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain

ARTICLE INFO

Article history:
Received 17 September 2019
Revised 28 February 2020
Accepted 12 March 2020
Available online 19 March 2020

Keywords:
Routine
Egocentric vision
Lifestyle
Behaviour analysis
Topic modelling

ABSTRACT

Developing tools to understand and visualize lifestyle is of high interest when addressing the improvement of habits and well-being of people. Routine, defined as the usual things that a person does daily, helps describe the individuals' lifestyle. With this paper, we are the first ones to address the development of novel tools for automatic discovery of routine days of an individual from his/her egocentric images. In the proposed model, sequences of images are firstly characterized by semantic labels detected by pre-trained CNNs. Then, these features are organized in temporal-semantic documents to later be embedded into a topic models space. Finally, Dynamic-Time-Warping and Spectral-Clustering methods are used for final day routine/non-routine discrimination. Moreover, we introduce a new *EgoRoutine*-dataset, a collection of 104 egocentric days with more than 100.000 images recorded by 7 users. Results show that routine can be discovered and behavioural patterns can be observed.

1. Introduction

With the dynamization of the day-by-day of our century, many people need to improve the quality of their life, and the first step is to get a better understanding of it. A characterization of the behaviour of a person can help us draw a picture of his or her lifestyle. In [29], the authors claimed that the definition of patterns of behaviour allows people to reach goals by creating associations between actions that are repeated in a stable context and their responses. In our study, we relate patterns to the combination of elements that describe the context of the days of someone, such as: environment, objects around the person, and his/her activities. Patterns of behaviour were also described as *ordered sequences of activities* [16] and are important elements when describing the *Routine* of a person. At the same time, *Routine* describes habits and sequences of activities of someone's days, and tends to be unique. More specifically, *Routine* has been described as *regularity in the activity* [35]. The ability to perform Activities of Daily Living (ADL) directly affects someone's quality of life. Health problems can be detected when certain activities are not performed due to some issues, such as isolation or depression. Therefore, the discovery of the routine of a person is of importance for its later analysis in order to assure the healthy living of individuals.

The characterization of people's life has become an active area of research with the increasing availability of wearable sensors [25]. Lifelogging is the process of collecting data about the life of people; this data can describe their activities, emotions and interactions throughout the day. In Fig. 1, we show a set of photo-streams collected by a camera wearer. This collection offers a rich source of information that allows understanding of the lifestyle of a person. More specifically, by using wearable cameras, images can be automatically collected from a first-person [12], a.k.a. egocentric point of view of the camera wearer. Egocentric images are a valuable source of information in many domains due to the similarity to human perception and memory. However, egocentric collections use to be large (of order of thousands of pictures per day), which makes difficult its analysis. In this work, we rely on long temporal resolution (2fpm) egocentric images for the discovery and study of Routine-related days of people since they allow to monitor and visualize most of their day. The discovery of *Routine* and *Non-Routine* days from egocentric photo-streams is an important step for several applications, such as: self-awareness i.e. how does my daily life look like?; monitoring patients or assistance of elderly people (it is essential to know the person's common behaviour and *Routine*) [9]; or, for memory enhancement and rehabilitation, which benefits from structuring the photo-stream into *Routine* and *Non-Routine* to easily find important events used in memory reminiscence therapy and interventions [28].

* Corresponding author.
E-mail addresses: e.talavera.martinez@rug.nl (E. Talavera), carolin@wuerich.eu (C. Wuerich), n.petkov@rug.nl (N. Petkov), petia.ivanova@ub.edu (P. Radeva).

https://doi.org/10.1016/j.patcog.2020.107330

Fig. 1. Example of images recorded by one of the camera wearers.

Routine-related days have common patterns that describe situations of the daily life of the person. 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. In this paper, we propose to apply Topic Modelling (TM) technique [5] to detect correlated elements of the individual's day (e.g. objects that appear together often in the environment of the wearer). We use TM as an unsupervised approach for the analysis of behavioural habits with the final goal of detecting *Routine* from egocentric images and thus, to describe and understand the daily patterns of conduct of the camera wearer. The analysis of the appearing topics throughout the recorded days allows the understanding of the different environments where the user spends time: working, shopping, walking outside, etc. These elements define the context of the lifestyle of the person. Our goal is to address the routine discovery by analyzing the appearance of these patterns in the life of a person. This pattern give us the opportunity to compare and evaluate days. They also allow us to describe what Routine represents for a person given a collection of his or her days.

In this work, we propose to apply TM to our problem by translating collected egocentric photo-streams into documents, as we describe in Section 3. We select this technique because it has demonstrated to be a powerful tool for the discovery of abstract topics appearing in collections of documents, audio, and images [15,17,21,22]. The input images are translated to a Bag-of-Word (BoW) representation, where an image is described by the objects around the wearer, activities of the wearer and the scene the image depicts. Next, the BoW is converted to a new representation of the day in terms of a set of discovered probabilistic topics. Then, the following step is to discover similar days. Routine can present daily small variations thus, the similarity measure use to compare performed activities during the day by the camera wearer should be tolerant to small differences. For instance, having breakfast at 6am and going to work from 7am to 5pm exhibits the same *Routine* as having breakfast at 7am and working from 9am to 7pm. We argue that this allows flexibility in the occurrence of performed activities during the day while temporal order among day elements is maintained. Therefore, in our model, we define similarities among days by evaluating distances between time-slots of a certain duration. To discover similar days we use Dynamic Time Warping for the computation of similarities/distances among the collected photo-streams, allowing that daily habits are tolerant to small differences in starting time and duration.

The contributions of this work are the following:

- We introduce an automatic unsupervised pipeline for the identification and characterization of Routine-related days from egocentric photo-streams. This pipeline can be adapted to different characterizations of days. Our model is based on the topics that describe the day-by-day from egocentric photo-streams for their classification into *Routine* and *Non-Routine* days.
- We present a new egocentric dataset describing the daily life of the camera wearers. It is composed of a total of 100.000 images, from 104 days recorded by 7 different users. We call it *EgoRoutine* and together with its ground-truth are publicly available in http://www.ub.edu/cvub/dataset/.

This paper is organized as follows: in Section 2, we highlight relevant work related to the routine discovery. In Section 3, we describe the approach proposed for *Routine* discovery. In Section 4, we introduce our EgoRoutine dataset, outline the experiments performed and the results obtained, and discuss the achieved results. Finally, in Sections 5 and 6, we discuss our findings and present our conclusions, respectively.

2. Related works

In this section, we describe how the routine behaviour of people was studied before the raise of wearable devices and what has been studied since then.

2.1. Routines from manually annotated data

The manual annotation of daily habits tend to be common practise for its later analysis by either the own person [3] or physicians [39]. In [3], manually recorded information about the ability of someone performing ADL was examined to classify the patients' dependence, as either dependent or independent. Also, in [39] the authors studied diaries from 70 undergraduate students, who rated the assiduity of activity during the previous month through a questionnaire.

2.2. Automatic routine discovery from sensors data

With the increasing availability of wearable sensors, the aim for automatic data collecting and understanding the behaviour of peo-

ple have become active areas of research. These sensors allow the automatic collection of big amount of data describing the life of the person who uses them. One of the first works on analyzing regularities in human behaviour from a large scale dataset in an unsupervised manner was presented in [13]. The model relied on information from mobile phones, such as locations, Bluetooth device proximity, application usage, and phone status. Other works relied on data collected by sensors placed in smart homes, such as the one in [26].

One of the seminal works on routine discovery was presented in [34] that applied a Latent Dirichlet Allocation (LDA) model for detecting activities and a subsequent assessment of the similarity of a person's days. There, topic modelling was employed to discover daily life activities related to rehabilitation patients from wearable sensors. Specific activity groups were applied to define the user's routine. The main 6 categories are eating/leisure (social interactions, eating, playing games), cognitive training (using pc, puzzles), medical fitness, kitchen work (household activities), motor training, and rest. In [15], the authors focused on *Routine* discovery by analyzing the localization patterns in a phone location dataset collected by 97 people over one year. Their proposed model is based on LDA and word analyses that are built based on location sequences. Sequences of words are defined by translating the pre-defined locations 'home', 'work', 'others' and 'no reception' to H, W, O, and N, respectively. Combining a fine-grain (30 minutes) and coarse-grain (several hours) consideration, they construct a bag representation of location sequences. Every location sequence consists of three consecutive location labels for the fine-grain intervals, followed by a number indicating the coarse-grain time-slot. This approach identifies *Routines* which dominate the entire group's behaviour such as 'going to work late' or 'working non-stop'. Furthermore, they characterize or classify individuals by those *Routines*. From another perspective, in [4], the behaviour information comes from phone GPS location and is used to assess the similarity of a person's day. The authors applied a modified version of Dynamic Time Warping (DTW) [24] method to sequences of GPS points sampled at an interval of 10 seconds. Thereafter, a spectral clustering algorithm is employed to cluster similar days and find anomalous behaviours. The authors in [44] proposed a model for the discovery of clusters of daily activity routines based on accelerometer data, which describes the expenditure data and steps. The model applies a low rank and sparse decomposition of the data signal to later isolate routine and deviations as two different sets of clusters. DTW and hierarchical clustering are used for the computation of pairwise distances and final classification, respectively.

2.3. Routine from conventional images

In [40], the authors addressed the problem of recognition of routine changes from short-term video sequences. Note that short-term refers to shortly defined time-slots (e.g. 3–4 hours as it is the case of a GoPro data) while long-term tends to define the continuous collection throughout the day. The dataset in [40] was recorded by a static camera at the entrance of a kitchen and for periods of time in 6 consecutive days, in 3 different years. In their approach, they first proposed to define a model per year. This model represents the structure of the sequential activities performed by the individual during that week and makes use of Dynamic Bayesian Network to estimate the similarity among sliding windows of the collected video sequences against the evaluated model. By evaluating the differences between each time frame and the model, their algorithm detects the changes between years in the performed activities when the person is in the kitchen. Despite the excellent results of this work, this method is applied on strongly controlled environments under the field of view of the static camera and so are not applicable to detect routine days of individuals.

The analysis of the behaviour of people has been previously addressed for personalized applications such as route planning [7], travel [31,41] or point-of-interest Recommendations [23,43], among others. In [41], the authors use dynamic topic modelling to mine the visited places described by intentionally collected photos by individuals. Based on the discovered topics, other similar locations are recommended. However, none of the above-mentioned approaches could describe and deal with the analysis of collections of photo-streams, which are what we believe can help to better understand the behaviour of the user.

2.4. Routine from egocentric images

The availability of wearable cameras allows to collect large amount of egocentric photo-streams, showing a first-view perspective of the performed activities by the camera wearer. Since the egocentric vision field emerged, several works have addressed the analysis of such collections of data from different perspectives: activity recognition [18–20], social interactions characterization [1,2], food-scenes classification [36], photo-stream segmentation [11], and sentiment analysis [38]. Especially difficult is the problem of analysis of long-term egocentric photo-streams (e.g. activity recognition), as they are recorded with a lower frame rate (2 fpm) and therefore provide sparser contextual information. Other related works mainly focus on the analysis of ADL. For instance, the works presented in [14] and [20] analyze egocentric images, focusing on recognizing the activities the camera wearer was performing. These studies do not go deeper into the analysis of how regularly the recognized activities or environment appear in the recorded photo-streams. Such pattern of appearance is what we believe will allow us to discover *Routine-related* days.

Whereas most of the long-term *Routine* analysis approaches rely on mobile phone locations or sensor data, our approach models patterns of behaviour based on visual data from egocentric images. This source of data allows us to understand the surrounding world and to give a visual explanation to our findings. To the best of our knowledge, the only work addressing routine behavioural analysis from egocentric images is [37], being a very preliminary work and a proof of concept of our proposal here. There, we addressed the classification of egocentric photo-streams into *Routine* or *Non-Routine* related days as an Anomaly Detection problem. That model achieved an average of 76% Accuracy and 69% F-Score. However, there we addressed the problem of *Routine* discovery from egocentric photo-streams following a very basic and straightforward solution. The proposed model was based on the Isolation Forest algorithm that partitions the data based on an anomaly score. A day was described as the average of the obtained global features for its sequence of images. The method evaluates the given feature vector as the descriptor for a day. Moreover, there we did not describe patterns of behaviour of people since days were represented by the aggregation of the global features of all images that composed the photo-stream. In contrast with the mentioned above, this article goes one step further by automatically discovering routines as well as visualizing and describing behavioural patterns of the camera wearer from his or her collected photo-streams.

3. Discovery of routine-related days from egocentric photo-streams

In this section, we describe our proposed model for the characterization of egocentric photo-streams for their later classification into *Routine* and *Non-Routine* related days. Fig. 2 illustrates the main steps that our model follows given a set of collected long-

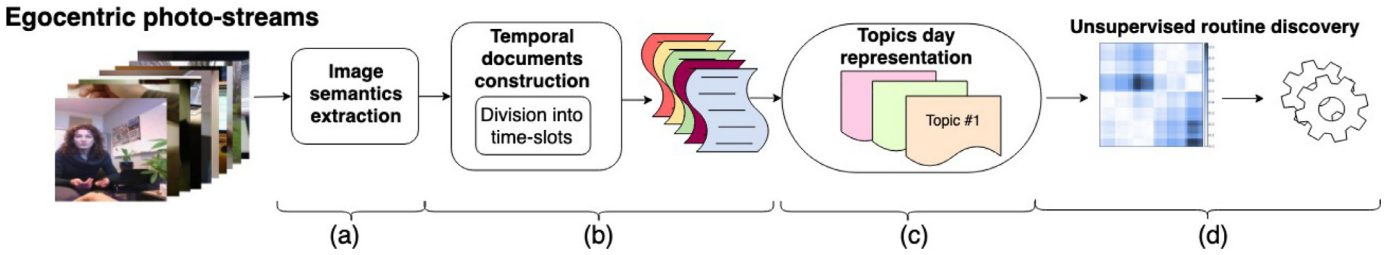

Fig. 2. Illustration of the proposed pipeline for the discovery of routine from sets of egocentric photo-streams collected by a user. The model proceeds as follows: (a) image semantics extraction, (b) temporal documents construction, (c) topics day representation, and finally, (d) unsupervised routine discovery.

term temporal resolution photo-streams. Below, we describe in detail how they are implemented.

a) *Image semantics extraction*
Describing sequences of photo-streams is not a trivial task due to the unknown visual content. In this work, we propose to describe our daily recorded images through detected concepts by an already pre-trained CNN. For a broad analysis of the scene depicted on a given image, we make use of CNNs pre-trained for the recognition of objects [8,30], places [45], and activities [6].
Let us consider that for each image I the CNNs return, L_r labels related to a total of R concepts found in the images; objects, scene, and activities of the wearer. Thus, each image is represented by a Bag-of-Words composed of these detected semantic concepts (CNN labels).

b) *Temporal documents construction*
To model the patterns of behaviour of the camera wearer, we embed the detected semantic labels extracted from the egocentric images into a temporal document. The detected concepts by the CNNs represent the words that describe the day i.e. that form the document.
In order to maintain the temporal information about the appearance of the extracted semantics, we define J time intervals within the day (e.g. from 7-9h, 9-11h, etc.). For each time-interval we estimate the frequency of appearing of each concept ($L_r, r = 1 \dots R$). For the time-intervals in which no images are taken, we create a dummy variable. Hence, each day is represented by a vector of $J \times R$ dimension.
Given a set I_u of egocentric photo-streams (days) for user u, a matrix $M_{i,j}$ is constructed where each of its elements (ij) corresponds to day $i = 1, \dots, |I_u|$, and $j = 1, \dots, J \times R$. This temporal document is composed of the concepts detected in the images recorded at a specific range of time. Thus, the proposed model translates a recorded day that is composed of a sequence of egocentric images, to a temporal document represented by the matrix M_{ij} defined in terms of the frequency of the detected concepts (words) in the photo-stream.

c) *Topics day representation*
Topic modelling allows the transformation of the dataset by factorisation of a set D of documents. A document is composed of a vector of words frequencies, and at the same time, it is assumed that it defines a certain number, K, of topics. In this work, we rely on Latent Dirichlet Allocation (LDA) [5], a topic modelling approach that is a generative probabilistic model applied to explain multinomial observations using unsupervised learning. The LDA method follows a generative process described as follows [5]:

(a) Choose $\theta_i \sim Dirichlet(\alpha)$, where $i \in \{1, \dots, D\}$.
(b) For each of the N_i words w_{ij} in document i:
i. choose a topic $z_{ij} \sim$ Multinomial (θ_i)
ii. choose a word w_{ij} from $P(w_{ij}|z_{ij}, \beta) \sim$ Multinomial probability on the topic z_{ij}.

where the parameters of the multinomials for topics in a document θ_i and words in a topic z_{ij} have Dirichlet priors, $Dir(\alpha)$ and $Dir(\beta)$ respectively. The probability of a corpus with D documents is defined as follows:

$$P(D|\alpha, \beta) = \prod_{i=1}^{|D|} \int P(\theta_i|\alpha) \left(\prod_{j=1}^{N_i} \sum_{z_{ij}} P(w_{ij}|z_{ij}, \beta) P(z_{ij}|\theta_i)\right) d\theta_i$$

where the parameters α and β are sampled only once in the process of generating the corpus, while the variables θ_i are sampled once per document. Lastly, the variables z_{ij} and w_{ij} are word-level variables which are sampled once per word j in each document i.
As a result, given a corpus (set) of D documents and K topics to be discovered, LDA gives [5]:

- the structure or combination of words that best fits the number of topics, by giving a *topic-word matrix* $P(w_{ij}|z_{ij}, \beta)$ where each element of it defines the probability of assigning word w_{ij} to topic z_{ij}.
- a *document-topic matrix* $P(z_{ij}|\theta_i)$ so that each element of it defines the probability of a topic z_{ij} for given a document θ_i.

In our case, we apply the LDA to decompose the elements $M_{i,j}$ of the temporal documents M corresponding to day i and time-slot j. LDA returns a document-topic matrix $P(z_{ij}|M_{ij})$ with the probabilities of all K topics associated with each element M_{ij} and the topic-words matrix $P(w_{ij}|z_{ij})$ that defines the relations between topics and words. This is illustrated in Fig. 3 showing a day represented by the most important topics (with the highest probability) and the relations between topics and words.

d) *Unsupervised routine discovery*
Once we have the representation of each day in terms of the most relevant topics with their probabilities, we need to find similarities among days for their later classification as *Routine* or *Non-Routine* days. For example, we expect that days that used to repeat (e.g. defined by topics related to *breakfast, metro, work, lunch, work, metro, and dinner*), appear frequently and thus correspond to a user's routine days.
At this point, a day is represented as a J-dimensional vector, where each element is a K-dimensional vector composed of the probabilities of the detected topics describing it (see Fig. 3). In order to find similar days, we need a metric to compare topics representation. However, it should be tolerant to small temporal differences, since events during the days can begin and last differently. To this purpose, we propose to apply DTW [24] for computing the similarity of topics representation among days. DTW is an algorithm that computes the optimal alignment between two sequences,

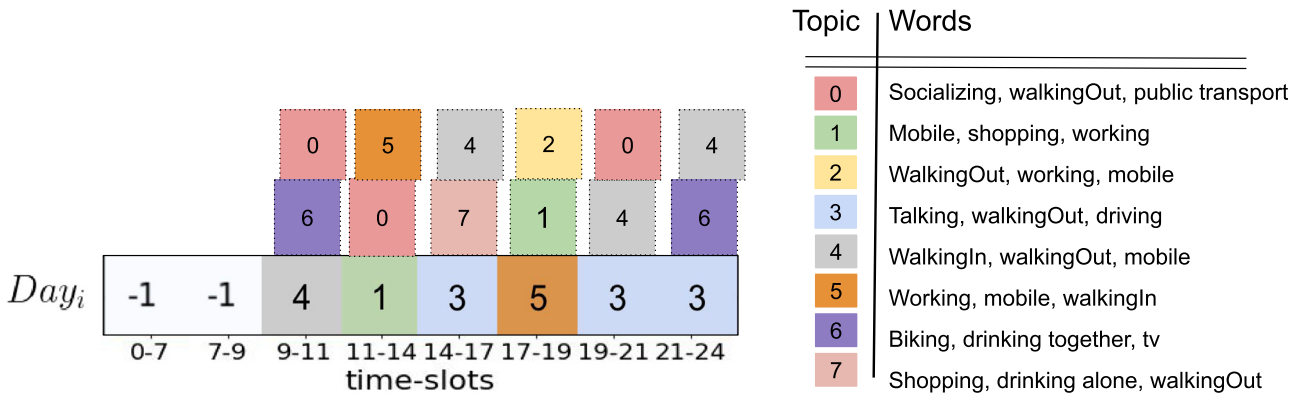

Fig. 3. Illustration of how a photo-stream/document (Day_i) is described by different proportions of topics throughout the day. We present the winning topic for each time-slot, together with the following N= 2 topics with the higher representation.

Table 1
Total number of recorded days and collected images per user.

User ID	1	2	3	4	5	6	7	Total
Num Days	14	10	16	20	13	18	13	104
Images per day	20,521	9583	21,606	19,152	17,046	16,592	10,957	115,430

where one of them might be stretched or shrunken non-linearly along the time axis. Given two sequences (or vectors) corresponding to two day representations, a warp path $(w_1, w_2, \ldots, w_Q)$ is constructed, where Q is the length of the path and every element w_q is a pair $(w_q[1], w_q[2])$ that indicates the mapping of element $w_q[1]$ in the first sequence s' to element $w_q[2]$ in the second one s''. Further, $w_q[1]$ and $w_q[2]$ have to monotonically increase. The optimal warp path defines the best correspondence of elements of both sequences represented by the path with minimal distance and is computed as follows:

$$dist_{DTW}(s', s'') = \sum_{r=1}^{Q} dist(s'_{w_q[1]}, s''_{w_q[2]}).$$

In our proposed model, we employ the fastDTW algorithm [33], which is an accurate approximation of the DTW method, but has a linear time and space complexity. In contrast to the standard DTW, the fastDTW algorithm shrinks a time series into smaller ones with fewer data points trying to preserve as much information about the original curve as possible. Given two sequences describing two days, the fastDTW algorithm computes the distance among them and gives as output the cost of aligning two days, i.e. their dissimilarity. To compare the topics representation of each time-slot, we apply Euclidean distance.

DTW only gives the distance between pairs of days. Next, we need to discover clusters of similar days. For that purpose, we cannot rely on the days topics representation but on the computed distances among pairs. We apply the *Spectral clustering* algorithm [42] over the computed affinity matrix of the distances between the days. This method does not make assumptions about the global structure of the data, but bases its decision on local evidence of how likely two elements (days) might belong to the same cluster. From the affinity matrix, the algorithm constructs a weighted graph $G = (Vn, E, We)$, being Vn the set of nodes, E the set of edges and We the weights of the edges. The global optimum is then computed by eigen-decomposition. This clustering method relies on k-Means for the final classification and thus, needs a number kc of clusters to be defined, which without loss of generality, we set to 2 for the discovery of *Routine* and *Non-Routine* related days.

4. Experimental framework and results

In this section, we detail a newly introduced EgoRoutine dataset. Then, we describe the metrics used for the evaluation of the performed experiments. Next, we depict the experimental setup with the proposed baseline approaches. Finally, we analyze the obtained results at different stages of the proposed pipeline.

4.1. Egoroutine - An egocentric dataset for behaviour analysis

In this work, we propose and make publicly available the *EgoRoutine* dataset[1]. This dataset is composed of recorded days by 7 individuals who wore the Narrative Clip camera[2] fixed to their chest and were asked to record their daily life. *EgoRoutine* consists of 115.430 images, from a total of 104 recorded days. In Table 1 and Fig. 4, we indicate the number of days and images collected per user. The camera wearers captured information about their daily *Routine*, taking pictures of the activities they performed and their occurrence as well as the people with whom they interacted.

GT evaluation: The collected dataset was labelled by 6 annotators who were asked to classify days into *Routine* or *Non-Routine* related. The annotators got the following definition "Life *Routine* is a sequence of actions which are followed regularly, or at specific intervals of time, daily or weekly". Days were shown to them in the form of a mosaics.

In Fig. 5, we present a representation of some of the collected photo-streams of User 1 with their final routine (R) or Non-Routine (NR) labels given on the right. In Table 2, we present the summary of the labels given by the different annotators. From the labelling results we can deduce that defining what is *Routine* and *Non-Routine* is not an easy task. *Routine* can be easily verbally described, but it becomes challenging when we want to discover

[1] http://www.ub.edu/cvub/dataset/.
[2] http://getnarrative.com/.

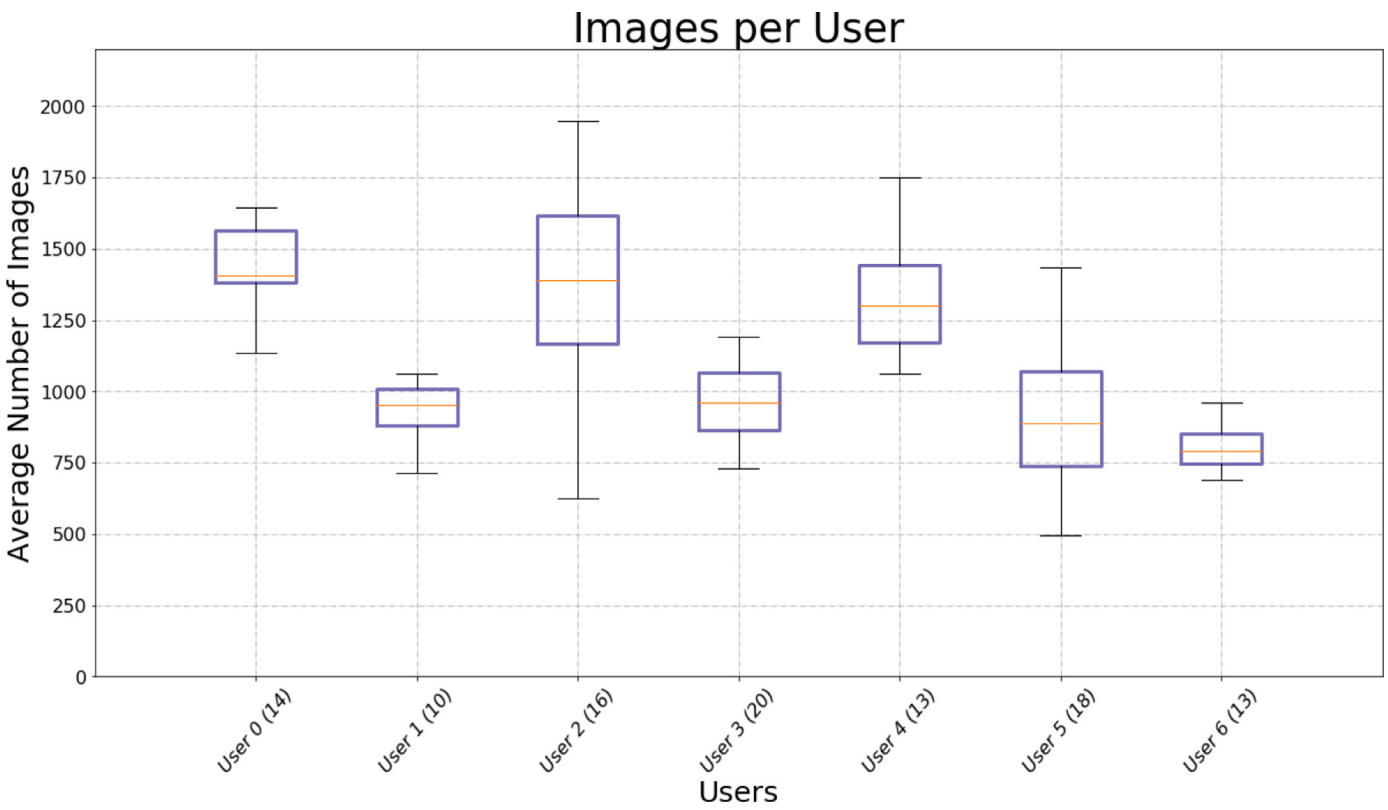

Fig. 4. Average number and variance of egocentric images per recorded photo-stream for the 7 users. Between parenthesis, we show the number of recorded days per user.

Table 2
Summary of the agreement among the 6 individuals that labelled the collected photo-streams into Routine or Non-Routine related days.

Class	Six Agree	Five Agree	At Least Four Agree	At Least Three Agree	Total
All	47	29	18	10	104
Routine	35	22	8	0	65
Non-Routine	13	7	9	10	39

it through the analysis of sequences of images describing a long period of time. We observed that in most cases, the annotators agreed when labelling days related to *Routine*. However, the *Non-Routine* related days were more difficult to perceive leading to disagreement among the annotators. For the final distinction, we have considered as *Routine* related days when more than 4 annotators agreed on the label. In case of a draw, the day is labelled as *Non-Routine* related. Therefore, from a total of 104 recorded days, 65 days are *Routine* related, and 39 are *Non-Routine* related. In Fig. 6 we present the number of labelled days per user into *Routine* and *Non-Routine*. If we extrapolate to a common life scenario, then 104 days correspond to almost 15 recorded weeks. If the users followed what could be considered as common *Routine*, where a week has 5 working days and 2 weekend days, in 15 weeks we have 30 weekend days and 75 working days. This could be an explanation of the resulted labels since it is proportional to the working days reported by the camera wearers.

4.2. Evaluation

In this section, we describe the metrics that we use to evaluate our proposed model for the discovery of *Routine* and *Non-Routine* related days.

The discovery of routine behaviour is an unsupervised problem with non-trivial evaluation. We evaluate the results in terms of *Accuracy (A), Precision (P)* and *Recall (R)* and F_1 score in terms of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), when classifying days into *Routine* or *Non-Routine*, defined as follows:

$$F_1 = \frac{2P \cdot R}{P+R}, P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}, Acc = \frac{TP+TN}{TP+TN+FP+FN}$$

Moreover, since the proposed pipeline for the discovery of routine behavioural patterns is composed of several steps, we also present qualitative results of the intermediate steps of our proposal.

4.3. Implementation setting

Regarding the concepts detected in the egocentric images, we perform an ablation study using the following different CNNs:

1. *Objects detection:* Detected objects by Yolo [30] and Xception [8]. These models were trained on the COCO [27] and ImageNet dataset [10], respectively.
2. *Scene recognition:* We represent an image by the top-1 probability scene label obtained by the VGG16, a pre-trained network previously trained on the Places365 dataset [45].
3. *Activities recognition:* We use the activity labels given by the CNN proposed in [6], which was trained for the recognition of 21 different daily activities. We select the activity label with the highest probability per image.

Fig. 5. Example of selected images throughout some of the recorded photo-streams of User1. On the right, we can see the given ground-truth (R for routine and NR for non-routine) and the predicted binary label by the best combination of parameters (1 for Non-routine and 0 for Routine days).

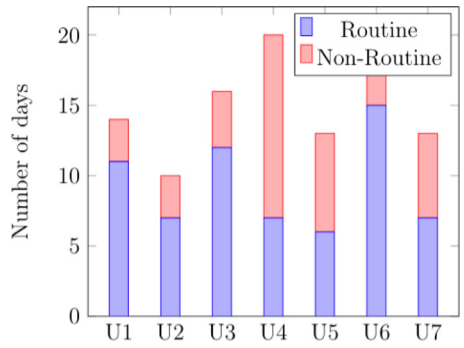

Fig. 6. Number of *Routine* and *Non-Routine* days for each user (U) in the *EgoRoutine* dataset.

Concerning DTW, we use the *Euclidean* metric to compute the distance among samples. Finally, with respect to the Spectral clustering, we set k equal to 2 to discover Routine and Non-Routine related days.

4.4. Experimental setup

We evaluate the performance of the different steps of our approach:

- **Image semantics extraction** in terms of the detected concepts in the egocentric images by the pre-trained CNNs as descriptors of the egocentric photo-streams.
- **Temporal documents construction** by the conversion of photo-streams concepts to documents. To evaluate the effect of this, we test the following:
  1. *Long duration time-slots:* We define J number of time-slots following the ones proposed in [15]: 0am-7am, 7am-9am, 9am-11am, 11am-2pm, 2pm-5pm, 5pm-7pm, 7pm-9pm, 9pm-12pm.
  2. *Short duration time-slots:* Of one hour each, 00:00-01:00, 01:00-02:00, 02:00-03:00, etc, with a result of 24 time-slots.
- **Topics day representation**, we evaluate the importance and the robustness of the proposal on the number of topics. Moreover, we study the need of individual vs. generic topic models in order to explore if the information about the routine of other users improve the final classification. Given multiple camera users, the LDA model can be computed either using the images of all users (generic) or considering the set of documents collected by each person separately (personalized).
- **Unsupervised routine discovery** of photo-streams. We assess the goodness of the proposed clustering method for the discovery of routine-related days, comparing it to the one achieved when using the *Agglomerative Hierarchical Clustering* [32] for the discrimination among days.

4.5. Results and discussions

Next, we present quantitative and qualitative results of the performance on the different stages of our approach for routine discovery validated on our *EgoRoutine* dataset.

- **Image semantics extraction performance:** in terms of the detected concepts: objects, activities and scenes. Within an ablation study we evaluate the performance of the different concept descriptors when they are considered separately or as a combination. In Table 3, we depict the performance of the experiments obtained. As it can be observed, the combination of labels of detected objects, activity and places better describes the data leading to the best results when addressing routine discovery, with $Acc = 80\%$ and $F_1 = 77\%$. This makes sense since a richer description of the image helps to better draw the description of the behaviour of people. Depending on the final goal and application, it could be that independently studying information about activities, objects and/or places helps describe better the routine of people.

Table 3
Results of the proposed pipeline and baseline models. We report results when evaluating different lengths of the time-slots in which we divide the photo-streams: per hour or the ones introduced in [15]. We also quantify the performance when evaluating 2, 4, 6, 8 and 10 topics. Moreover, we present the obtained results when applying Hierarchical (HierClus) and Spectral Clustering (SpClus). Finally, we show the output of the model when evaluating collected days by the user (Personalized) or by the whole set of user (Generic topics).

				Xception [8]				Yolo [30]				Activities [6]				Places [45]				Combination			
	TimeSlot	Clustering	#Topics	Acc	F_1	P	R	Acc	F_1	P	R	Acc	F_1	P	R	Acc	F_1	P	R	Acc	F_1	P	R
Personalize	Per Hour	SpClus	2	0.72	0.68	0.70	0.71	0.71	0.68	0.73	0.75	0.72	0.70	0.72	0.73	0.68	0.65	0.69	0.70	0.72	0.69	0.70	0.72
			4	0.75	0.73	0.74	0.77	0.72	0.71	0.74	0.77	0.72	0.69	0.70	0.71	0.78	0.76	0.77	0.81	0.75	0.72	0.74	0.75
			6	0.72	0.70	0.73	0.76	0.76	0.73	0.74	0.76	0.76	0.73	0.75	0.77	0.74	0.72	0.75	0.78	0.76	0.72	0.74	0.76
			8	0.78	0.75	0.76	0.79	0.76	0.73	0.75	0.78	0.77	0.75	0.78	0.81	0.71	0.70	0.75	0.76	0.77	0.73	0.76	0.80
			10	0.73	0.72	0.75	0.78	0.73	0.70	0.72	0.74	0.69	0.66	0.69	0.71	0.72	0.69	0.72	0.74	0.74	0.71	0.74	0.75
		HierClus	2	0.68	0.64	0.71	0.71	0.66	0.64	0.73	0.74	0.71	0.69	0.74	0.76	0.71	0.69	0.73	0.74	0.71	0.68	0.76	0.74
			4	0.75	0.72	0.77	0.77	0.76	0.74	0.76	0.78	0.71	0.67	0.72	0.72	0.75	0.72	0.76	0.77	0.73	0.69	0.72	0.74
			6	0.66	0.60	0.66	0.67	0.76	0.73	0.77	0.79	0.71	0.65	0.71	0.69	0.75	0.71	0.78	0.75	0.70	0.68	0.71	0.74
			8	0.79	0.75	0.83	0.79	0.72	0.68	0.71	0.71	0.72	0.66	0.73	0.72	0.77	0.75	0.81	0.82	0.75	0.72	0.78	0.77
			10	0.72	0.64	0.69	0.68	0.71	0.63	0.67	0.71	0.67	0.61	0.67	0.69	0.76	0.71	0.71	0.75	0.73	0.66	0.74	0.73
	As in [15]	SpClus	2	0.69	0.66	0.69	0.71	0.66	0.63	0.67	0.68	0.68	0.66	0.71	0.72	0.68	0.67	0.70	0.72	0.69	0.68	0.71	0.73
			4	0.72	0.71	0.74	0.77	0.75	0.72	0.75	0.77	0.74	0.72	0.74	0.77	0.75	0.73	0.77	0.79	0.77	0.75	0.77	0.80
			6	0.77	0.75	0.77	0.80	0.71	0.68	0.72	0.74	0.72	0.68	0.70	0.72	0.74	0.71	0.74	0.76	**0.80**	**0.77**	**0.79**	**0.82**
			8	0.70	0.67	0.70	0.72	0.66	0.63	0.70	0.70	0.76	0.72	0.73	0.74	0.76	0.73	0.74	0.77	0.72	0.69	0.72	0.74
			10	0.76	0.73	0.74	0.76	0.70	0.66	0.72	0.72	0.75	0.73	0.74	0.76	0.77	0.75	0.77	0.80	0.77	0.75	0.76	0.79
		HierClus	2	0.73	0.70	0.72	0.73	0.69	0.67	0.72	0.72	0.69	0.63	0.65	0.67	0.64	0.60	0.67	0.66	0.72	0.63	0.64	0.68
			4	0.70	0.68	0.72	0.74	0.70	0.68	0.71	0.74	0.69	0.68	0.72	0.74	0.68	0.65	0.69	0.71	0.74	0.73	0.75	0.77
			6	0.73	0.72	0.76	0.79	0.63	0.57	0.64	0.65	0.65	0.56	0.60	0.63	0.71	0.69	0.72	0.74	0.75	0.72	0.75	0.75
			8	0.66	0.62	0.70	0.69	0.67	0.62	0.68	0.69	0.71	0.66	0.69	0.70	0.71	0.66	0.70	0.71	0.75	0.70	0.71	0.73
			10	0.67	0.59	0.61	0.66	0.72	0.64	0.69	0.69	0.67	0.60	0.68	0.68	0.71	0.69	0.72	0.75	0.73	0.66	0.71	0.71
Generic	Per Hour	SpClus	2	0.74	0.69	0.70	0.71	0.76	0.74	0.76	0.79	0.79	0.75	0.75	0.77	0.72	0.69	0.70	0.72	0.76	0.72	0.73	0.75
			4	0.74	0.70	0.73	0.75	0.78	0.74	0.75	0.78	0.77	0.75	0.78	0.80	0.74	0.72	0.75	0.78	0.77	0.74	0.76	0.77
			6	0.76	0.72	0.74	0.76	0.75	0.71	0.73	0.76	0.74	0.73	0.76	0.79	0.76	0.74	0.75	0.78	0.75	0.71	0.73	0.75
			8	0.72	0.69	0.72	0.74	0.74	0.71	0.73	0.75	0.73	0.71	0.74	0.76	0.76	0.74	0.76	0.78	0.76	0.72	0.74	0.76
			10	0.76	0.72	0.74	0.76	0.75	0.72	0.74	0.76	0.73	0.71	0.72	0.75	0.75	0.73	0.76	0.79	0.74	0.71	0.74	0.75
		HierClus	2	0.69	0.65	0.69	0.71	0.67	0.59	0.65	0.65	0.68	0.65	0.71	0.72	0.68	0.65	0.72	0.72	0.67	0.63	0.70	0.70
			4	0.75	0.71	0.78	0.76	0.74	0.68	0.70	0.73	0.75	0.72	0.77	0.76	0.67	0.63	0.70	0.69	0.74	0.70	0.72	0.74
			6	0.72	0.66	0.67	0.71	0.67	0.63	0.71	0.71	0.73	0.68	0.72	0.75	0.79	0.75	0.81	0.76	0.73	0.70	0.75	0.77
			8	0.67	0.63	0.77	0.72	0.69	0.65	0.75	0.73	0.73	0.64	0.65	0.70	0.75	0.70	0.76	0.74	0.76	0.73	0.75	0.78
			10	0.68	0.66	0.73	0.75	0.74	0.67	0.70	0.70	0.70	0.63	0.71	0.70	0.73	0.69	0.76	0.73	0.76	0.70	0.77	0.74
	As in [15]	SpClus	2	0.70	0.68	0.71	0.73	0.71	0.69	0.73	0.74	0.67	0.66	0.68	0.71	0.69	0.66	0.70	0.71	0.69	0.67	0.72	0.73
			4	0.69	0.66	0.70	0.72	0.71	0.68	0.73	0.74	0.70	0.67	0.68	0.70	0.73	0.71	0.75	0.77	0.78	0.76	0.78	0.81
			6	0.75	0.72	0.74	0.77	0.73	0.71	0.73	0.76	0.69	0.65	0.67	0.68	0.74	0.70	0.72	0.73	0.78	0.76	0.77	0.80
			8	0.74	0.71	0.72	0.75	0.69	0.64	0.67	0.68	0.72	0.68	0.70	0.73	0.72	0.70	0.73	0.75	0.75	0.72	0.74	0.76
			10	0.72	0.69	0.71	0.74	0.73	0.70	0.74	0.76	0.73	0.70	0.72	0.74	0.76	0.74	0.76	0.79	0.76	0.74	0.76	0.78
		HierClus	2	0.73	0.68	0.71	0.73	0.67	0.65	0.70	0.71	0.73	0.70	0.71	0.73	0.70	0.64	0.69	0.70	0.65	0.63	0.70	0.70
			4	0.68	0.65	0.68	0.70	0.66	0.64	0.71	0.71	0.64	0.58	0.62	0.63	0.60	0.54	0.64	0.63	0.64	0.59	0.65	0.67
			6	0.74	0.67	0.68	0.72	0.69	0.64	0.69	0.70	0.70	0.65	0.73	0.70	0.69	0.63	0.75	0.69	0.72	0.67	0.68	0.73
			8	0.69	0.64	0.69	0.70	0.67	0.61	0.64	0.64	0.74	0.70	0.74	0.75	0.69	0.61	0.67	0.65	0.70	0.68	0.75	0.75
			10	0.75	0.68	0.73	0.73	0.72	0.66	0.70	0.72	0.71	0.67	0.70	0.70	0.75	0.71	0.77	0.75	0.67	0.61	0.67	0.69

Table 4
Results of the proposed pipeline for the best setting of the parameters: analysing the set of collected photo-streams of User1, seeking for 6 topics to describe the data, with time-slots of long duration, and with spectral clustering as the final classifier.

	User 1	User 2	User 3	User 4	User 5	User 6	User 7	Avg
Acc	0.79	0.74	0.75	0.90	0.92	0.56	0.92	0.80
F_1	0.75	0.70	0.71	0.89	0.92	0.50	0.92	0.77
P	0.75	0.75	0.70	0.89	0.93	0.56	0.94	0.79
R	0.86	0.79	0.75	0.89	0.93	0.60	0.92	0.82

In Table 5, we show concepts that are detected by the different evaluated CNNs in a given photo-stream. Overall, the detected places by the network get close enough to reality and therefore are evaluated. In the case of *activity* recognition, and since the network was trained with egocentric images, the results are more consistent. For the detection of objects, *YOLO* seems more consistent when detecting objects of the daily living. We understand that this is due to the fact that the CNN was trained with 80 different categories corresponding to Common Objects in Context (COCO [27]). In contrast, *Xception* might be able to recognize uncommon objects since it was trained over a bigger dataset composed of 1000 different categories (the ImageNet [10]). We can observe some inconsistencies in the classes given by the network trained over *Places365*, such as finding the 'airplane cabin' label early in the morning. We explain it by the fact that this network was not trained with egocentric pictures. The change of perspective modifies how scenes are understood, and lights in the ceiling of an office or corridor can be missinterpreted as the lights in the cabin of an airplane.

- **Evaluation of the temporal documents construction:** We study the effect on the discovered topics for the final classification when analyzing time-slots of different duration. Time-slots of longer duration might affect the result by smoothing activities happening during a short time. In contrast, fine-grained time-slots might lead to noise in the final classification. From the results shown in Table 3, we can observe that the model better performs when the day is described by analyzing the time division proposed in [15]. We deduce that time-slots with a longer duration smooth the activities performed during short periods of time when comparing days. A fine-grained time-slots with an hour duration might include noise to the description of a day.
- **Evaluation of the topics day representation performance:** Topic models discover abstract topics within given documents. A natural question that may arise is the data used for the discovery of topics: should they be discovered from the set involving all users or they should be extracted for each user individually?. A hypothesis is that if more documents are given (joining all data), more robust topics will be discovered, and thus, better they will be able to describe the behavioural patterns of the camera wearers. Thus, when learning the topic-word distribution following the generic approach, we could take advantage of a bigger dataset. A negative aspect of seeking generalization is that user-specific activities can be missed, since they would become not relevant to be detected. In contrast, we assume that individually learned topics might find more personalized representations of every specific activity of the user, since the places of their daily life, e.g. the office desk or living room of different people, might be described differently. Therefore, we evaluate the performance of the model when obtaining the topics just based on the collected photo-streams by the user under study (personalized approach), or when analyzing all the collected photo-streams that compose the EgoRoutine dataset (generic approach). From the results and for the goal of routine discovery, the personalized approach allows the model to better distinguish Routine-related days with a 80% accuracy and 77% F_1 (see Table 3).
  The goodness of the model when varying the number of topics is also tested. We present results when discovering 2, 4, 6, 8 and 10 topics. As it can be observed, the performance of the classifier is highest when discovering 6 and addressing the time-division proposed in [15]. However, it could be that for a more detailed analysis of what is happening at a specific time, a higher number of fine-grained time-slots might describe in more detail, in terms of objects, activities and places.
- **Evaluation of the Unsupervised routine discovery performance**: We compare the performance of the proposed Spectral Clustering algorithm with the results obtained by the *Agglomerative Hierarchical Clustering* [32] (HC) when classifying into *Routine* or *Non-Routine* related days. HC method follows a bottom-up approach where each data point starts as a single cluster, and pairs of samples are recursively merged following the path that minimally increases the given linkage distance. The process continues as samples are clustered moving up in the similarity hierarchy. We select the HC since we need to compare against methods that are able to analyse pre-computed distance matrices.
  We can observe in Table 3 that the Spectral Clustering classifier leads to a more accurate discovery of the Routine-related days, outperforming the classification by the HC. We believe this is due to the ability of the Spectral clustering to adapt to complex shapes of the data in the data space.
  For a more detailed understanding of the performance at user level, in Table 4 we show results of the best performing model. We can observe that for some of the users the classification into *Routine* and *Non-Routine* related days is rather clear, such as for User 5 or User 7, while for User 6 the classification is close to random. This is due to differences between the lifestyle of the users. Some of them have a clear distribution of routine (e.g. work) and non-routine (e.g. non-work) related activities, while others recorded days for periods when their activities were not following an established routine pattern.
  In Fig. 5, we present some collected days of User 1 and the predicted label by the best combination of parameters (personalize analysis of documents, combination of labels as images descriptors, 6 topics, and Spectral clustering). Days predicted as Non-Routine related are assigned label '1' and Routine-related days label '0'. Day 1 is miss-classified as Non-Routine related. From observing the data, we can guess that this user tends to start working at noon and until late in the evening. In contrast, on Day 1, User 1 spent much fewer hours at work and left the office much earlier. This could be a cause of miss-classification by the model. Non-Routine related days contained events where the user worked for short periods and spent longer time interacting with colleagues or friends. Day 7 is an example where User 1 went for dinner to a restaurant right after working for a short time.
- **Final routine characterization and visualization for behaviour modelling**: The characterization of days based on detected concepts and the later inferred topics have demonstrated

Table 5
Example of detected concepts in a given recorded day by User 1. This table aims to give an idea of how documents develop throughout the day of the person. Each column represents a time slot of a specific duration. Rows present the top-3 concepts detected by the pre-trained networks referenced in the left of the table. The presented numbers describe the numbers of times a concept was present in that time-slot.

	Time-slot (h)											
	9–11		11–14		14–17		17–19		19–21		21–24	
Xception [8]	screen	29	desktop pc	266	desktop pc	85	desktop pc	83	radio	16	photocopier	80
	menu	23	screen	265	desktop pc	75	screen	80	CD player	16	desk	59
	monitor	19	monitor	254	screen	51	monitor	74	slot	16	projector	42
Places [45]	airplane cabin	90	airplane cabin	167	conference room	49	office	41	airplane cabin	31	reception	28
	atrium/public	8	office	113	office	43	airplane cabin	26	bowling alley	14	airplane cabin	26
	office cubicles	8	office cubicles	42	reception	37	computer room	23	airport terminal	10	hotel room	14
Activity [6]	WalkingIn	50	Mobile	227	Mobile	60	Working	78	Mobile	30	Talking	50
	Shopping	40	Shopping	94	Talking	46	Mobile	39	Driving	25	WalkingOut	37
	WalkingOut	36	Working	75	meeting	46	WalkingOut	32	WalkingOut	16	Mobile	27
Yolo [30]	person	146	tvmonitor	383	person	202	person	132	person	107	person	198
	laptop	38	cup	354	laptop	112	tvmonitor	122	chair	32	chair	155
	chair	38	laptop	334	chair	108	keyboard	73	cell phone	23	diningtable	53

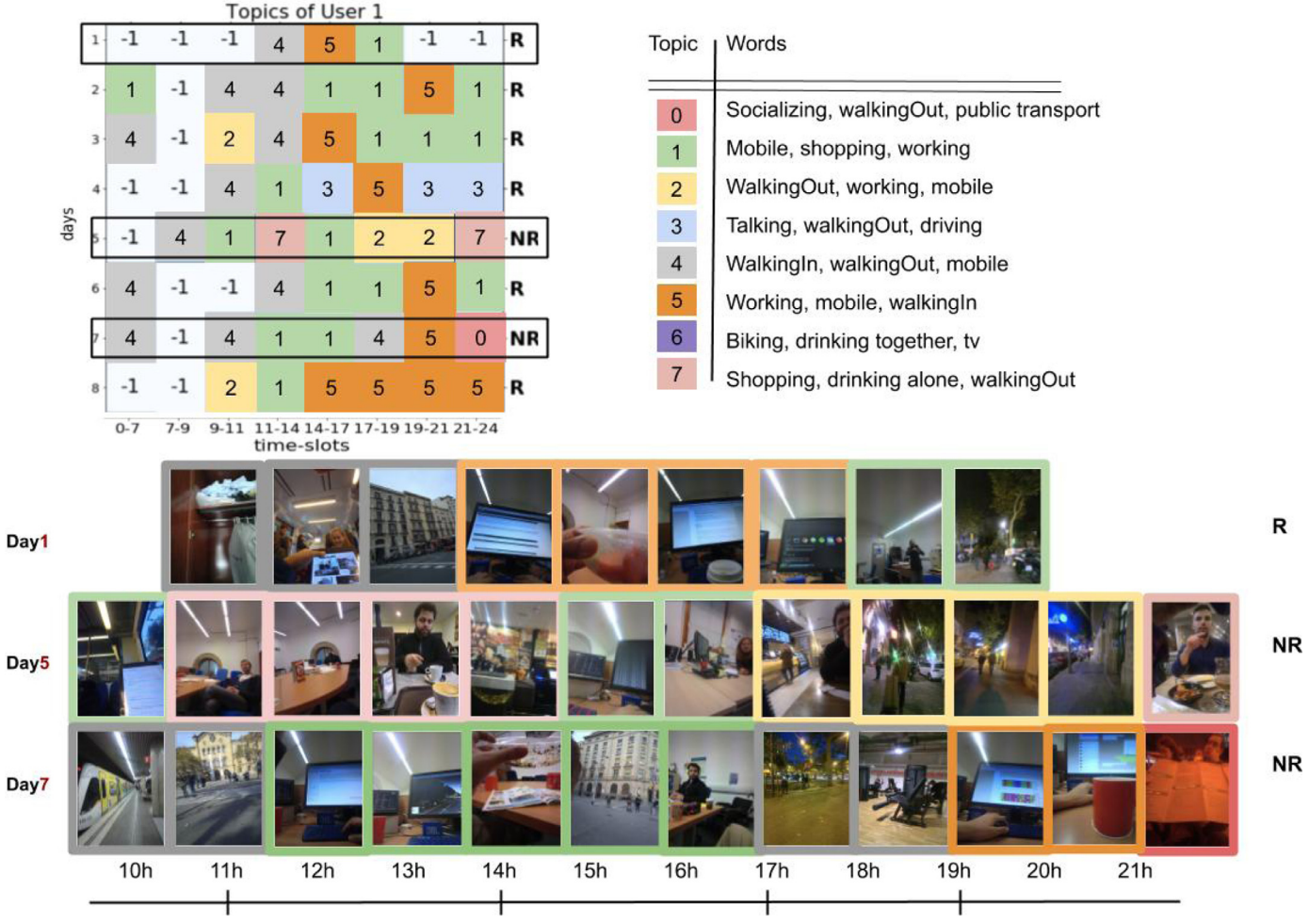

Fig. 7. Example of given photo-streams, sample images at several time-slots, their representative topics, and the concepts that compose them. We present results with the following combination of the parameters of our model: activity labels, time-slots as in [15], 8 topics and personalized approach.

to be a rich tool for behaviour visualization. In Fig. 7 we present how the found topics could be analysed by the wearer or an expert. As an example of visualization, results are shown following a personalized analysis of the data collected by User 1 described with activity labels, and discovering 8 topics. As we can observe, Non-routine related days differ from the Routine-related days as the first one presents Topic 0 and Topic 7, which are composed of activity labels describing social interaction in food-related environments. Routine-related days are mainly described by Topic 1, 3, 4, and 5, which describe working environ-

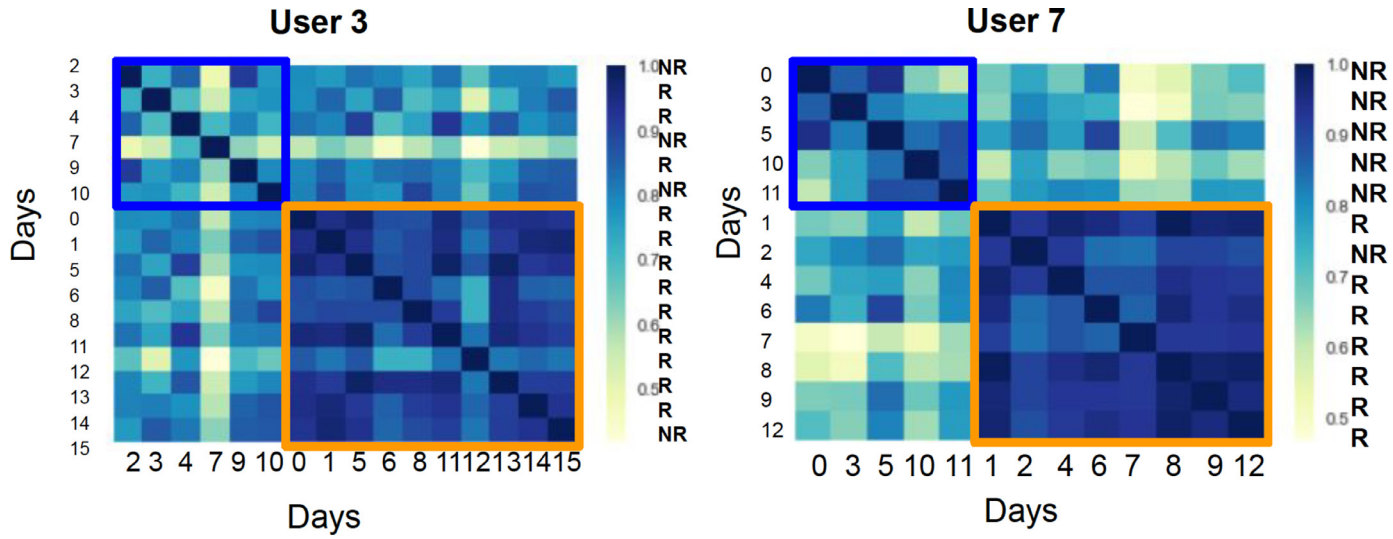

Fig. 8. Affinity matrix obtained from the distances computed by DTW for the later discrimination as Routine or Non-Routine related days by Spectral Clustering of collected days by users 3 and 7. Days are divided with orange and blue boxes as the two final clusters. On the right, we indicate the ground-truth labels per day. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

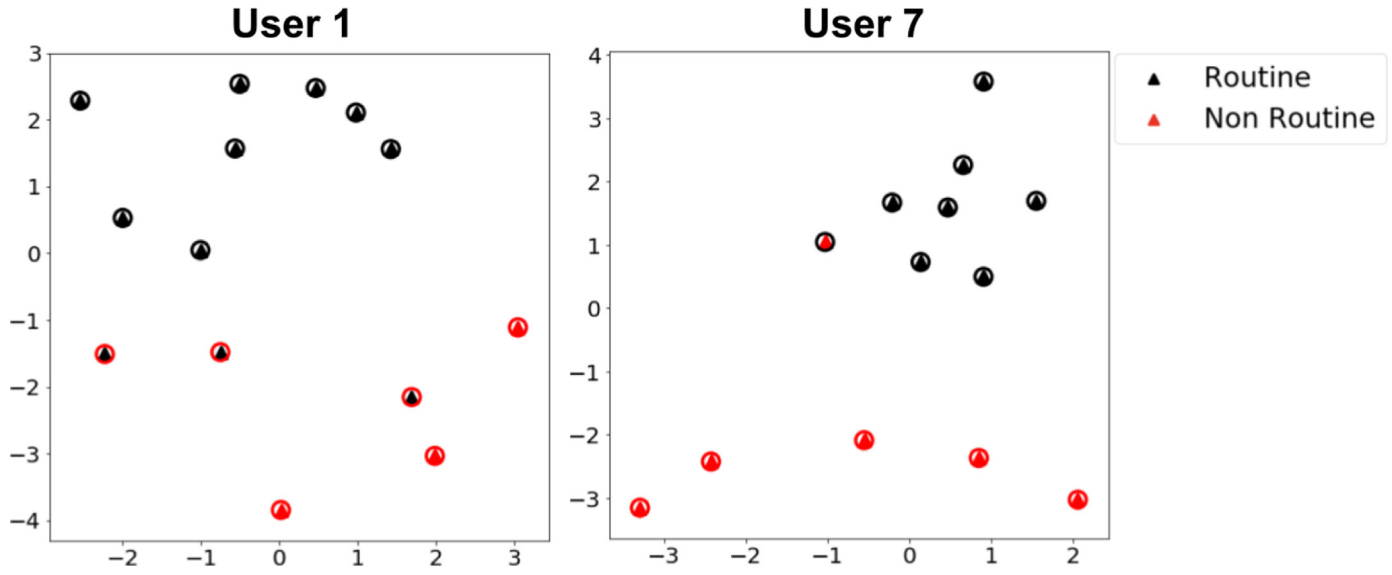

Fig. 9. Visualization using multi-dimensional scaling (MDS) of the distribution of samples for users 1 and 7. Each dot corresponds to a collected day by the user. We use two colors to distinguish between the two classes. The inside color of the dots is the given ground-truth and the colour of the boundaries of the dots represents the classification label ('R' black and 'NR' red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ments. We understand that activity labels such as *mobile, talking*, and *walking Indoor/Outdoor* can be understood as screen, meeting, and commuting, respectively.

To get insight at the classification level, we present in Fig. 8 the affinity matrix that the Spectral Clustering uses for the discrimination among the collected days by User 3 and User 7. The given labels for the collected days are indicated in the figure on the right of the matrix, where 'R' corresponds to Routine-related and 'NR' to Non-Routine related. In the presented affinity matrix, we highlight the two final clusters with orange and blue. We can observe how in the case of these users clear R-related clusters are defined, while NR-related clusters are scattered. The accuracy for User 3 and User 7 is of 75% and 92%, respectively, which agree with the visual association in Fig. 8 between similar days and given labels.

Furthermore, in Fig. 9 we visually illustrate the produced results of our model for users 1 and 7. We applied Multi-Dimensional Scaling (MDS) for this visualization since it allows to visualize spatial distribution of data from their similarity matrix instead of explicit coordinates/representations. We use it to display the mutual spatial distribution of the user days representations expressed by the obtained similarity matrix when applying the DTW. We can see the ground-truth indicated as the inside color of the sample and the classification label as the boundaries of the circles. In both cases black corresponds to Routine-related days and red to Non-routine related days. This visualization allows us to better explore classification results.

Moreover, in Fig. 10 we can observe the computed silhouette scores for the obtained final routine and non-routine related clusters. Note that the silhouette score can take a value between 1 and -1. Values close to 1 indicate differentiable clus-

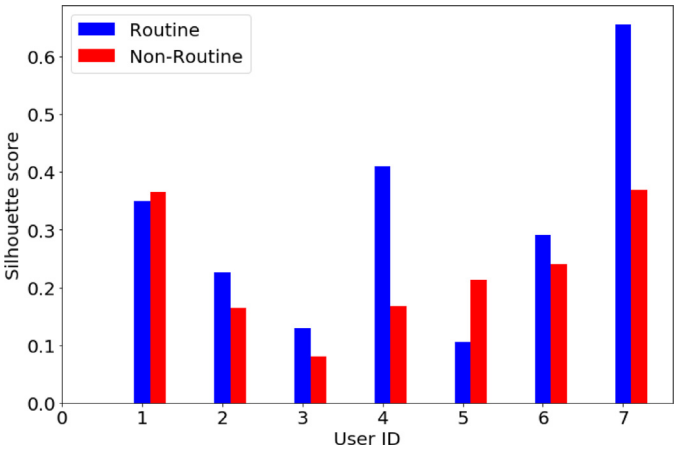

Fig. 10. Silhouette score per user for the two discovered clusters, Routine and Non-routine related days.

Table 6
Comparison between our previous work introduced in [37] and the model here proposed for routine discovery from egocentric photo-streams.

Method	Number of Users	Acc	F_1
Routine discovery [37]	5	0.76	0.69
Routine discovery propose here		**0.82**	**0.79**
Routine discovery propose here	7	**0.81**	**0.80**

ters, 0 overlapping of clusters, and negative samples represent the wrong classification of samples. We can see how for the majority of the users, routine-related days share a higher score than the non-routine related days clusters. This reinforces our hypothesis that routine related days correspond to more compact clusters, while non-routine related days form a more sparse cluster. In two of the cases, the silhouette score for the routine related days is lower than for non-routine related ones. Looking closer at the data we observed that in these cases there were more than one routine groups of days. However, the problem of discovering the optimal number of clusters and thus the routines is out of the scope of this paper.

Finally, in Table 6 we compare the obtained results for routine discovery to the routine discovery in [37]. As one can see the method in [37] run on 5 users achieved 0.76 of accuracy and 0.69 of F_1 score while the method proposed here achieved 0.81 of accuracy and 0.80 of F_1 score. A possible explanation is that the work proposed in [37] relied on the aggregation of global features of all the images composing a day for its description. In contrast, the model proposed here relies on semantic concepts combined with topic modelling, DTW and spectral clustering, which results also allow understanding of what is happening in the life of the camera user. We also present the results of our method for the subset of five users that were analyzed in [37], with a performance of $Acc = 0.82$ and $F_1 = 0.79$. As we can observe, the results are quite similar: moreover, higher classification performance is achieved when topics modelling DTW and spectral clustering are applied to the collection of documents composed of detected semantic concepts.

5. Discussions

In this work, we presented a new method for the analysis of routine behavioural patterns from collected egocentric visual data. We demonstrated that these images are a rich source of information and that detected concepts from the images can help us draw a picture of the lifestyle of the camera wearer.

One of the important advantages of this work is the unsupervised discovery of routine and non-routine related days. Given a new user, we can discriminate routine days and characterize their collected photo-streams. In particular, given a collection of photo-streams, our model can discover routine-related days by relying on the found topics when considering detected concepts as image descriptors. The input is a Bag-of-Word representation of the images, where an image is described by the objects and the scene it depicts. This is treated as a document for the discovery of abstract topics describing the themes of the lifestyle of the individual under study. Documents are fed to an LDA model that organizes semantic labels into topics computing a topic-word distribution and a document-topic distribution, thus, obtaining topics distribution for each given document. Moreover, we show that using temporal documents based on time-slots into which days are divided, allows flexibility when comparing the behaviour at different times of the day. The distances between the days can be computed using DTW to finally cluster days and assign them into *Routine* and *Non-Routine* ones by applying Spectral clustering.

Moreover, we introduced a new *EgoRoutine* dataset, on which we tested and validated our proposed model. The dataset is composed of a total of 104 days, recorded by 7 users, and we make it publicly available[3] for the future development of this line of research. The analysis of the model could be improved by the augmentation of the dataset. For further steps in this direction, we need richer data. However, this is not a trivial task and we are working on it. Moreover, more accurate detected concepts would be of help when describing the collected days. For this, we would need trained networks on egocentric images.

We hypothesize that Routine-related days will share similar traits and thus, will represent a cluster. Commonly, Non-routine related days, tend to be the ones non-work related. These days share their own routine-patterns, i.e. there can be more than one routine in the life of people; cleaning, cooking, or going out with friends could describe one of them. A limitation of our work is that Non-Routine related days might not define a cluster. In future works, we plan to evaluate if the combination of outlier detection with topic modelling allows a better understanding of the lifestyle of the camera wearer.

We hope that our proposed dataset and the shown results will be a call for other researchers who aim to study people's behaviour for its understanding and providing tools for lifestyle improvement.

6. Conclusions

In this work, we conclude that behavioural analysis from visual data is possible. Moreover, topic models proved to be a powerful tool for the discovery of patterns when addressing Bag-of-Words representation of photo-streams. From the obtained results, we observed that discovered topic models following a personalized approach improve the classification of days. This provides a more detailed explanation of wearer daily behaviour. However, a generic or personalized approach can be applied depending on if the goal is to detect general information or peculiarities of the life of a person. One of the important advantages of this work is the unsupervised discovery of routine and non-routine related days. Given a new user, we can discriminate routine days and characterize their collected photo-streams.

Further works will explore the inclusion of outlier detection techniques and the discovery of specific behaviours, such as: social interactions and nutritional behaviour by studying the appearance of people in certain situations and food-related scenes, re-

[3] http://www.ub.edu/cvub/dataset/

spectively. Furthermore, we are interested in studying how topic modelling and CNNs can be interconnected.

We hope that our proposed dataset and the shown results will be a call for other researchers who aim to study people's behaviour for its understanding and providing tools for lifestyle improvement.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was partially founded by projects TIN2015-66951-C2, RTI2018-095232-B-C2, SGR 1742, CERCA, Nestore Horizon2020 SC1-PM-15-2017 (n 769643), Validithi EIT Health Program and *ICREA Academia 2014*. The founders had no role in the study design, data collection, analysis, and preparation of the manuscript. The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of several Titan Xp GPU used for this research. The collected data as part of the study and given labels is publicly available from the website of our research group: http://www.ub.edu/cvub/dataset/

References

[1] M. Aghaei, M. Dimiccoli, C.C. Ferrer, P. Radeva, Towards social pattern characterization in egocentric photo-streams, Comput. Vision Image Understanding (2018) 104–117.
[2] S. Alletto, G. Serra, S. Calderara, R. Cucchiara, Understanding social relationships in egocentric vision, Pattern Recognit. 48 (12) (2015) 4082–4096.
[3] C.K. Andersen, K.U. Wittrup-Jensen, A. Lolk, K. Andersen, P. Kragh-Sørensen, Ability to perform activities of daily living is the main factor affecting quality of life in patients with dementia, Health Qual. Life Outcomes 2 (1) (2004) 52.
[4] J. Biagioni, J. Krumm, Days of our lives: assessing day similarity from location traces, International Conference on User Modeling, Adaptation, and Personalization (2013) 89–101.
[5] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, Journal of machine Learning research 3 (Jan) (2003) 993–1022.
[6] A. Cartas, J. Marín, P. Radeva, M. Dimiccoli, Batch-based activity recognition from egocentric photo-streams revisited, Pattern Analysis and Applications 21 (4) (2018) 953–965.
[7] D. Chen, D. Kim, L. Xie, M. Shin, A.K. Menon, C.S. Ong, I. Avazpour, J. Grundy, Pathrec: Visual analysis of travel route recommendations, in: Proceedings of the Eleventh ACM Conference on Recommender Systems, 2017, pp. 364–365.
[8] F. Chollet, Xception: deep learning with depthwise separable convolutions, IEEE Conference on Computer Vision and Pattern Recognition (2017) 1800–1807.
[9] P.-C. Chung, C.-D. Liu, A daily behavior enabled hidden markov model for human behavior understanding, Pattern Recognit 41 (5) (2008) 1572–1580.
[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, IEEE Conference on Computer Vision and Pattern Recognition (2009) 248–255.
[11] M. Dimiccoli, M. Bolaños, E. Talavera, M. Aghaei, S.G. Nikolov, P. Radeva, Sr–clustering: semantic regularized clustering for egocentric photo streams segmentation, Comput. Vision Image Understanding 155 (2017) 55–69.
[12] A.R. Doherty, S.E. Hodges, A.C. King, A.F. Smeaton, E. Berry, C.J. Moulin, S. Lindley, P. Kelly, C. Foster, Wearable cameras in health: the state of the art and future possibilities, Am. J. Prev. Med. 44 (3) (2013) 320–323.
[13] N. Eagle, A. Pentland, Reality mining: sensing complex social systems, Personal Ubiquitous Comput. 10 (4) (2006) 255–268.
[14] M. Ermes, J. PÄrkkÄ, J. MÄntyjÄrvi, I. Korhonen, Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions, IEEE Trans. Inf. Technol. Biomed. 12 (1) (2008) 20–26.
[15] K. Farrahi, D. Gatica-Perez, Discovering routines from large-scale human locations using probabilistic topic models, ACM Trans. Intell. Syst. Technol. 2 (1) (2011) 3.
[16] I. Fatima, M. Fahim, Y.-K. Lee, S. Lee, A unified framework for activity recognition-based behavior analysis and action prediction in smart homes, Sensors 13 (2) (2013) 2682–2699.
[17] R. Fernandez-Beltran, F. Pla, Latent topics-based relevance feedback for video retrieval, Pattern Recognit. 51 (2016) 72–84.
[18] A. Furnari, G. Farinella, S. Battiato, Recognizing personal locations from egocentric videos, IEEE Trans Hum Mach Syst 47 (1) (2017) 1–13.
[19] A. Furnari, G.M. Farinella, S. Battiato, Recognizing personal contexts from egocentric images, IEEE International Conference on Computer Vision Workshop (2015) 393–401.
[20] A. Furnari, G.M. Farinella, S. Battiato, Temporal segmentation of egocentric videos to highlight personal locations of interest, European Conference on Computer Vision (2016) 474–489.
[21] S. Hou, L. Chen, D. Tao, S. Zhou, W. Liu, Y. Zheng, Multi-layer multi-view topic model for classifying advertising video, Pattern Recognit. 68 (2017) 66–81.
[22] P. Hu, W. Liu, W. Jiang, Z. Yang, Latent topic model for audio retrieval, Pattern Recognit. 47 (3) (2014) 1138–1143.
[23] S. Jiang, X. Qian, J. Shen, Y. Fu, T. Mei, Author topic model-based collaborative filtering for personalized poi recommendations, IEEE Trans. Multimedia 17 (6) (2015) 907–918.
[24] E.J. Keogh, M.J. Pazzani, Derivative dynamic time warping, SIAM international conference on data mining (2001) 1–11.
[25] O.D. Lara, M.A. Labrador, A survey on human activity recognition using wearable sensors, IEEE Communications Surveys & Tutorials 15 (3) (2012) 1192–1209.
[26] C. Li, W.K. Cheung, J. Liu, Elderly mobility and daily routine analysis based on behavior-aware flow graph modeling, International Conference on Healthcare Informatics (2015) 427–436.
[27] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: common objects in context, European Conference on Computer Vision (2014) 740–755.
[28] G. Oliveira-Barra, M. Bolaños, E. Talavera, A. Dueñas, O. Gelonch, M. Garolera, Serious games application for memory training using egocentric images, International Conference on Image Analysis and Processing (2017) 120–130.
[29] Society for Personality, Social Psychology, How we form habits, change existing ones, ScienceDaily (2014).
[30] J. Redmon, A. Farhadi, Yolov3: an incremental improvement, arXiv (2018).
[31] S. Renjith, A. Sreekumar, M. Jathavedan, An extensive study on the evolution of context-aware personalized travel recommender systems, Inf. Process. Manag. 57 (1) (2020) 102078.
[32] L. Rokach, O. Maimon, Clustering methods, Data mining and knowledge discovery handbook (2005) 321–352.
[33] S. Salvador, P. Chan, Toward accurate dynamic time warping in linear time and space, Intell. Data Analalysis 11 (5) (2007) 561–580.
[34] J. Seiter, A. Derungs, C. Schuster-Amft, O. Amft, G. Tröster, Daily life activity routine discovery in hemiparetic rehabilitation patients using topic models, Methods Inf. Med. 54 (3) (2015) 248–255.
[35] A. Sevtsuk, C. Ratti, Does urban mobility have a daily routine? learning from the aggregate data of mobile networks, J. Urban Technol. 1 (17) (2010) 41–60.
[36] E. Talavera, M. Leyva-Vallina, M. Sarker, D. Puig, N. Petkov, P. Radeva, Hierarchical approach to classify food scenes in egocentric photo-streams, J. Biomed. and Health Informatics (2019).
[37] E. Talavera, N. Petkov, P. Radeva, Unsupervised routine discovery in egocentric photo-streams, 18th Conference on Computer Analysis of Images and Patterns (2019).
[38] E. Talavera, N. Strisciuglio, N. Petkov, P. Radeva, Sentiment recognition in egocentric photostreams, Iberian Conference on Pattern Recognition and Image Analysis (2017) 471–479.
[39] W. Wood, J. Quinn, D. Kashy, Habits in everyday life: thought, emotion, and action, J. Pers. Soc. Psychol. 83 (6) (2002) 1281–1297.
[40] Y. Xu, D. Damen, Human routine change detection using bayesian modelling, International Conference on Pattern Recognition (2018) 1833–1838.
[41] Z. Xu, L. Chen, Y. Dai, G. Chen, A dynamic topic model and matrix factorization-based travel recommendation method exploiting ubiquitous data, IEEE Trans. Multimedia 19 (8) (2017) 1933–1945.
[42] S.X. Yu, J. Shi, Multiclass spectral clustering, IEEE International Conference on Computer Vision 2 (2003).
[43] Z. Yu, H. Xu, Z. Yang, B. Guo, Personalized travel package with multi–point-of-interest recommendation based on crowdsourced user footprints, IEEE Trans Hum Mach Syst 46 (1) (2015) 151–158.
[44] O. Yürüten, J. Zhang, P. Pu, Decomposing activities of daily living to discover routine clusters, Conference on Artificial Intelligence (2014).
[45] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba, Places: a 10 million image database for scene recognition, IEEE Trans. Pattern Anal. Mach. Intell. (2017).

Estefania Talavera received her BSc degree in Electronic engineering from Balearic Islands University in 2012 and her MSc degree in Biomedical Engineering from Polytechnic University of Catalonia in 2014. She is currently a PhD student at the University of Barcelona and University of Groningen. Her research interests are lifelogging and health applications.

Carolin Wuerich received her BEng degree in Electrical Engineering from the Baden-Wuerttemberg Cooperative State University Stuttgart (Germany) in 2017 and her MSc degree in Artificial Intelligence from the Polytechnic University of Catalonia in 2019.

Prof. Nicolai Petkov received the Dr.Sc.Techn. degree in Computer Engineering from the Dresden University of Technology, Germany. He is a Full Professor and Head of the Intelligent Systems group of the Bernoulli Institute of the University of Groningen, the Netherlands. His current research is in image processing, computer vision and pattern recognition.

Prof. Petia Radeva is a Senior researcher and a Full professor at the University of Barcelona (UB). She is Head of Computer Vision at the UB group and the MiLab of Computer Vision Center. Her present research interests are in the development of learning-based approaches for computer vision, egocentric vision and medical imaging.