Chapter 3
Routine Discovery from Egocentric Images

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 these works, 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.
3. Routine Discovery from Egocentric Images

3.1 Introduction

The characterization of people’s life has become an active area of research with the increasing availability of wearable sensors (Doherty et al., 2013). Lifelogging is the process of collecting data about the life of people; this data can describe their activities, emotions and interactions along the day. It 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, a.k.a. egocentric point of view of the camera wearer’s. Egocentric images are a valuable source of information in many domains due to the similarity to human perception and memory. However, egocentric collections used 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 these photo-streams is an important step for several applications, such as: self-awareness, how does my daily life look like?; monitoring patients or health-care and assistance of elderly people, it is essential to know the person’s common behaviour and Routine; 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 (Oliveira-Barra et al., 2017).

![Figure 3.1: Example of images recorded by one of the camera wearers.](image.png)

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, un-
supervised 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:

- In Section 3.3, we propose a personalized and automatic tool for the discovery of routine related days within recorded photo-streams by a camera wearer. We hypothesize that discovering routine related days can be addressed as a clustering problem where methods such as $k$-means with, for instance, $k = 2$ could potentially classify the days in terms of the behaviour they represent. However, some days present abnormal behaviour. These days correspond to non-routine related days. Most of the time they are not related to each other, which can be interpreted as outliers within the user’s recorded photo-streams. Experience has shown that it is difficult to describe what non-routine related days are for a given photo-stream collection. In the context of outlier detection, samples considered as outliers do not form the cluster with higher density when representing the days in a feature space. We propose an unsupervised classification method that assumes that outliers are situated in low-density areas. Outlier detection methods are commonly used in data mining to indicate variability in measurements, errors or novel samples (Ding and Fei, 2013; Hodge and Austin, 2004). Among their applications are fraud detection (Ghosh and Reilly, 1994) and satellite image analysis (Alvera-Azcárate et al., 2012). However, up to our knowledge for first-time routine detection is defined through an outlier detection approach. Within the available outlier detection algorithms, we propose Isolation Forest algorithm (Liu et al., 2008). This method has shown a good performance when detection outliers in multi-dimensional space, not seeking normal data points but identifying anomalies. Our model is unsupervised because routine differs per person and our aim is to propose a generic model able to discover routine of unknown users. However, since we have the labels of the recorded photo-streams that compose our dataset, we use them to validate if we are able to discover their routine related days.

- In Section 3.4, we apply Topic Modelling (TM) technique to help us detect correlated elements of the individual’s day (e.g. objects that use to appear together 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 recorded days allows the understanding of the different situations
3. Routine Discovery from Egocentric Images

where the user spends time: working, shopping, walking outside, etc. These elements define the context of the person’s lifestyle. Our goal is to address the routine discovery by analyzing the appearance of these patterns in the life of a person. Our goal is to address the routine discovery by analyzing the appearance of these patterns in the life of a person. This pattern gives 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. We select this technique because it has demonstrated to be a powerful tool for the discovery of abstract topics appearing in collections of documents. 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 used 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 chapter are the following:

- We address for the first time the problem of routine extraction from egocentric data.
- We propose an unsupervised and automatic model for the analysis of routine days following an anomaly detection approach. This model is based on the aggregation of the descriptors of the images within the photo-stream.
- 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/](http://www.ub.edu/cvub/dataset/).

This chapter is organized as follows: in Section 3.2, we highlight relevant work related to the routine discovery. In Section 3.3 and 3.4 we describe the approaches proposed for Routine. Within the approach section we also described the proposed dataset, outline the experiments performed and the results obtained, and discuss the achieved results. Finally, in Sections 3.5 and 3.6 we globally discuss our findings and present our conclusions, respectively.

### 3.2 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.

#### 3.2.1 Routine from manual annotation

The manual annotation of daily habits tend to be common practise for its later analysis by either the own person [Andersen et al., 2004] or physicians [Wood et al., 2002]. In [Andersen et al., 2004], manually recorded information about the ability of someone performing ADL was examined to classify the patients' dependence, as either dependent or independent. Also, in [Wood et al., 2002] the authors studied diaries from 70 undergraduate students, who rated the assiduity of activity during the previous month through a questionnaire.

#### 3.2.2 Routine from sensors

With the increasing availability of wearable sensors, the aim for automatic data collecting and understanding the behaviour of people 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 [Eagle and Pentland, 2006]. The model relied on information
3. Routine Discovery from Egocentric Images

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 (Li et al., 2015).

One of the seminal works on routine discovery was presented in (Seiter et al., 2015) 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 (Farrahi and Gatica, 2011), 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 (Biagioni and Krumm, 2013), 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) (Keogh and Pazzani, 2001) 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 (Yürüten et al., 2014) 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.

3.2.3 Routine from conventional images

In (Xu and Damen, 2018), the authors addressed the problem of recognition of routine changes from short-term video sequences. Short-term refers to shortly defined time-slots while long-term tends to define the continuous collection throughout the
day. The dataset 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. Although 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.

3.2.4 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 (Furnari et al., 2017, 2015, 2016), social interactions characterization (Aghaei et al., 2017; Alletto et al., 2015; Talavera et al., 2018), food-scenes classification (Sarker et al., 2018; Talavera et al., 2014), photo-stream segmentation (Dimiccoli et al., 2017) and summarization (Bolanos et al., 2015), and sentiment analysis (Talavera, Strisciuglio, Petkov and Radeva, 2017). 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 (Ermes et al., 2008) and (Furnari et al., 2016) 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. In contrast with the mentioned above, this chapter goes some steps further by automatically discovering routines as well as visualizing and describing behavioural patterns of the camera wearer from his or her collected photo-streams.
3.3 Unsupervised routine discovery following an outlier detection approach

In this section, we propose an innovative and unsupervised routine discovery method. Its application scheme is given in Fig. 3.2.

Figure 3.2: The pipeline of the proposed model. Given a set of recorded days, a) they are translated to a set of global or semantic features. Later, b) days are considered as routine or non-routine based on their resemblance.

Our proposed method is based on an outlier detection algorithm. For outlier detection models, an outlier sample is known as a sample outside the ‘boundary’ of the known classes. In our case, these samples relate to non-routine related days. Hence, we assume that routine related days define a class, of which the samples are close to each other within the feature space. The proposed model indicates routine of the person by detecting the sample days that can be clustered together. In the following subsections, we describe the steps in the proposed pipeline as shown in Fig. 3.2.

a) From days to feature vectors

As mentioned above, a day is described by a collection of images and takes the form of photo-stream. We address the day classification by translating the recorded photo-streams into feature vectors for their later analysis and comparison.

Based on the high accuracy recently achieved for the classification of daily activities in egocentric images in [Cartas et al., 2017], we use their proposed network for the characterization of the recorded days. Given an image, this network classifies it into 21 Activities of Daily Living. A day of the user is represented by $Day = \frac{\sum_{image}^{N}}{N}$, where $N$ is the number of images within a day, and $image$ represents the feature vector of the recorded images.

We consider the following descriptors obtained from the collected photo-streams:

1. Activity occurrence within the day: We consider the occurrence of activities throughout the day for the characterization of routine, i.e. bag-of-activities.
This feature vector gives an overview of the activities the user performs in a day. However, it does not include temporal information.

2. Global descriptors: We use the ResNet CNN model [He et al., 2016] to extract global descriptors from the images. We use the activation over the entire image given by the last fully connected layer. Given an image, we obtain a 2048 features vector.

3. We concatenate the mentioned features in 1) and 2).

b) Routine related days recognition

More specifically, we rely on the unsupervised outlier detection Isolation Forest [Liu et al., 2008] algorithm, and use its available implementation in Scikit-learn [Pedregosa et al., 2011]. It is a tree ensemble method that analyses the density of the space to ‘isolate’ outliers. The algorithm works as follows:

First, it randomly selects a feature. Then, for the selected feature, it randomly selects a split value between its maximum and minimum value. By recursive partitioning, it can be represented by a tree structure. As the number of trees increases, the algorithm reaches the convergence. The length of the path from the root to the end node can be considered as the number of splittings needed to isolate a sample. By randomly partitioning the data, the paths for anomalies become shorter. Therefore, samples with shorter path lengths are likely to be anomalies. Later, the anomaly score is calculated per sample based on the averaged and normalized distance of the path. Finally, samples considered as outliers have an anomaly score of 1, while samples with values close to 0 are considered as regular.

The Isolation Forest algorithm, given a set of \( n \) samples and an observation \( x \), computes the anomaly score \( s(x) \) as follows:

\[
s(x, n) = 2 \left( 1 - \frac{E(h(x))}{c(n)} \right),
\]

where \( h(x) \) is the path length of a point \( x \) measured by the number of edges that the point traverses from the root node until the last external node. \( E(h(x)) \) corresponds to the average of \( h(x) \) from a collection of isolation trees. \( c(n) \) is the average path length, and it is defined as follows:

\[
c(n) = 2H(n - 1) - (2(n - 1)/n),
\]

where \( H(i) \) is the harmonic number and it can be estimated by \( \ln(i) + 0.5772156649 \) (Euler’s constant).
3. Routine Discovery from Egocentric Images

<table>
<thead>
<tr>
<th>User ID</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num Days</td>
<td>14</td>
<td>10</td>
<td>16</td>
<td>19</td>
<td>13</td>
<td>72</td>
</tr>
<tr>
<td>Images per day</td>
<td>20k</td>
<td>8k</td>
<td>21k</td>
<td>13k</td>
<td>11k</td>
<td>73k</td>
</tr>
</tbody>
</table>

Table 3.1: Description of the collected Egoroutine dataset by 5 users.

To summarize, given a collection of photo-streams recorded by a camera wearer, our proposed personalized and automatic tool will detect the non-routine related days by computing the density within the feature space. The proposed Isolation Forest algorithm considers as routine related days if their samples are in a dense region of samples. In contrast, samples that represent non-routine related days correspond to points in a low-density area. This will have as an output the distinction among days, giving insight into the daily habits and lifestyle of the person.

3.3.1 Experiments

In this section, we describe the experimental setup, the metrics used to evaluate the analysis, and the obtained results.

Dataset

We collected data from 5 different subjects who were asked to record their daily life during at least a week. To this end, the users wore the Narrative Clip camera fixed to their chest, with a resolution of 2 fpm. The introduced dataset consists of 100k images, from a total of 72 recorded days, see Table 3.1. They captured information about their daily routine, such as the people with whom they interacted, the activities they performed or how often they walked outside. Since there is no training involved in this approach, the whole dataset is analysed by our proposed model. Moreover, in order to show the variance among collected days, Fig. 3.3 shows the average number of images per day. We can observe how the amount of images differs per day and user.

Process of creating the Ground-truth

The annotators got the following definition of “Life routine; a sequence of actions which are followed regularly, or at specific intervals of time, daily or weekly”. Next, they were shown mosaics of images representing days of the user. They were asked to first have a look at all the mosaics to get an impression of how routine looks like for that specific user. Later, they gave a binary label: routine or non-routine related.

In Table 3.2, 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

1[^1](http://getnarrative.com/)
3.3. Unsupervised routine discovery following an outlier detection approach

Figure 3.3: Average number of images per recorded egocentric photo-stream. We give the number of collected days per user between parenthesis.

<table>
<thead>
<tr>
<th>Class</th>
<th>Six Agree</th>
<th>Five Agree</th>
<th>At Least Four Agree</th>
<th>At Least Three Agree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>34</td>
<td>21</td>
<td>11</td>
<td>6</td>
<td>72</td>
</tr>
<tr>
<td>routine</td>
<td>28</td>
<td>16</td>
<td>7</td>
<td>0</td>
<td>51</td>
</tr>
<tr>
<td>non-routine</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of the labelling results for the Egoroutine dataset.

non-routine is not an easy task. Routine can be easily described in general terms, but it becomes challenging when sequences of images describing a long time period are classified. We can observe how in the majority of cases, the annotators agreed when it comes to label days as routine. However, the non-routine related days are more difficult to perceive leading to disagreement among the annotators. Finally, we have considered as routine related days when >4 of the labels agreed. In case of a draw, the day is labelled as non-routine related. Therefore, from a total of 72 recorded days, 51 days are routine related, and 21 are non-routine related. If we extrapolate to a common life scenario, 72 days correspond to almost 15 recorded weeks. If the users followed what could be considered common routine (a week has 5 working days and 2 weekend days or holiday), in 10 weeks we have 20 weekend days and 50 working days.
Validation

We evaluate the performance of the proposed model and compare it with the baseline models by computing the Accuracy (Acc), Recall (R), Precision (P), and F-Score metrics, where: 
\[ F - Score = 2 \cdot \frac{P \cdot R}{P + R} \]

Precision computes the ratio between True Positive (TP) samples and False Positive (FP) samples following: 
\[ TP / (TP + FP) \]

Recall evaluates the ratio of TP and False Negative (FN), showing the ability of the model to find the positive samples, the formula is 
\[ TP / (TP + FN) \]

Due to the unbalanced dataset we calculate and compare their ‘macro’ and ‘weighted’ mean. The ‘weighted’ mean evaluates the true classification per label, while ‘macro’ calculates the unweighted mean per label. The weighted measures provide the strength of the classifier when applied to unbalanced data.

Experimental setup

To the best of our knowledge, no previous works have addressed the recognition of routine discovery from egocentric photo-streams. Therefore, we evaluate the performance of the proposed model and compare it with what we introduced as baseline methods. We select several outlier detection algorithms namely: Robust Covariance, and One-class SVM. Moreover, we propose to apply unsupervised clustering techniques that allow the identification of outliers or isolation of samples outside the high-density space. These methods allow the recognition of non-similar samples or with non-convex boundaries within the sample collection. Specifically, we evaluate the performance of DBSCAN and Spectral clustering.

Here we give a brief explanation of how these baseline methods work:

- **Robust Covariance** \[ \text{Rousseeuw and Driesen} \] \[ \text{1999} \], also called elliptic envelope, assumes that the data follow Gaussian distribution and learns an ellipse. Its drawback is that it degrades when the data is not uni-modal.

- **One-class SVM** \[ \text{Platt et al.} \] \[ \text{1999} \] is an unsupervised algorithm that estimates the support of the dimensional distribution.

- **DBSCAN** \[ \text{Ester et al.} \] \[ \text{1996} \], short for Density-Based Spatial Clustering of Applications with Noise, finds samples with high density and defines them as the centre of a cluster. From the center, it expands the cluster. Its \( \epsilon \) parameter determines the maximum distance between samples to be considered as in the same cluster. Outliers are samples that lie alone in low-density regions.

- **Spectral Clustering** \[ \text{Yu and Shi} \] \[ \text{2003} \] works on the similarity graph between samples. It computes the first \( k \) eigenvectors of its Laplacian matrix and defines a feature vector per sample. Later, \( k \)-Means is applied to these feature
3.3. Unsupervised routine discovery following an outlier detection approach

<table>
<thead>
<tr>
<th>Methods</th>
<th>Feature Vector</th>
<th>All Users</th>
<th>All Users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Acc</td>
<td>F-Score</td>
</tr>
<tr>
<td>Robust covariance</td>
<td>Activity Occurrence (Act)</td>
<td>0.61</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Global Features (Glo)</td>
<td>0.71</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Act - Glo</td>
<td>0.54</td>
<td>0.39</td>
</tr>
<tr>
<td>One-Class SVM</td>
<td>Activity Occurrence (Act)</td>
<td>0.72</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Global Features (Glo)</td>
<td>0.67</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Act - Glo</td>
<td>0.65</td>
<td>0.58</td>
</tr>
<tr>
<td>DBSCAN</td>
<td>Activity Occurrence (Act)</td>
<td>0.61</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Global Features (Glo)</td>
<td>0.69</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>Act - Glo</td>
<td>0.63</td>
<td>0.56</td>
</tr>
<tr>
<td>SpectralClustering</td>
<td>Activity Occurrence (Act)</td>
<td>0.66</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Global Features (Glo)</td>
<td>0.66</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Act - Glo</td>
<td>0.62</td>
<td>0.46</td>
</tr>
<tr>
<td>Isolation Forest</td>
<td>Activity Occurrence (Act)</td>
<td>0.76</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>Global Features (Glo)</td>
<td>0.76</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Table 3.3: Performance of the different methods implemented for the discovery of routine and non-routine days.

vectors to separate them into \( k \) classes. In our case, we set \( k = 2 \), so we evaluate its performance when addressing routine vs non-routine classification.

For the last two proposed unsupervised model, DBSCAN and Spectral Clustering, the closeness among the recorded days is computed based on their shared similarities, following an all-vs-all strategy. To do so, we use the well-known Euclidean metric. The computed similarity matrix is fed to the unsupervised classifier algorithm for the detection of outliers within the set samples. The outlier detection methods are fed with the feature matrix describing the samples.

Results

We present the obtained classification accuracy at day level for the performed experiments in Table 3.3. The proposed model, based on the Isolation Forest algorithm and with global features as descriptors of the recorded days, achieved the best performance with respect to the rest of the tested baseline methods. Our model achieves an average of 76% Accuracy and 68% Weighted F-Score for all the users, outperforming the rest of the tested methods. The highest performance is when analysing global features, which cover most of the possible present activities.

Moreover, in Fig. 3.5, we visualize the days as points in the feature space drawn by the first two principal components of the dataset. We can see the Ground-truth indicated with the boundaries of the circles and the prediction of the model, for both cases red corresponds to routine related days and blue to non-routine related. As it can be observed, our model is the one that obtains the best results.
3. Routine Discovery from Egocentric Images

Figure 3.4: Histograms showing the occurrence of activities throughout the days of 3 of the 5 users that worn the camera. As we can appreciate, some activities are more related to non-routine related days, while ‘working’ and ‘walking indoor’ characterizes routine related days.

In Fig. 3.4 we can observe the occurrence of activities per day in the form of a histogram. This representation allows us to better infer and understand how routine (orange) and non-routine (blue) related days vary for the different camera wearers. From this representation we can confirm our initial assumptions: i) the set of ac-
3.3. Unsupervised routine discovery following an outlier detection approach

Figure 3.5: Visualization of the obtained classification results based on the analysis of the histogram of activities occurring throughout the day for User1, User2 and User5. We show the classification per user and per studied method. Each dot in the graph corresponds to one day recorded by the user. Each of the 4 subplots shows the classification into routine or non-routine by the baseline methods. The colour of the boundaries of the dots represents the given Ground-truth and the filling the classification label; Red routine and Blue non-routine.

Activities performed as routine and non-routine related days differs per person, ii) a subset of activities is commonly shared when it comes to routine, such as ‘working’, which is mostly described by a laptop/pc as central object in the scene, or ‘using mobile’. In contrast, some activities are specific per user: The routine of User 5 is characterized by ‘cooking’, ‘reading’, and ‘meeting’. In contrast, for User 2 ‘walking outdoor’, ‘shopping’, and ‘mobile’ are the more representative activities.

Limitations: The presented analysis can be improved in several directions as by augmenting the number of subjects and the amount of collected data. We believe this is a good starting point for this new field of unsupervised routine analysis from a first-person perspective. Moreover, and even though in this work we consider that there exists one routine per person, future lines will address the discovery of several routines. However, for that, it is needed a bigger amount of data.
3.4 Unsupervised routine discovery relying on topic models

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. 3.6 illustrates the main steps that our model follows given a set of collected long-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 (Chollet, 2017; Redmon, 2018), places (Zhou et al., 2017), and activities (Cartas et al., 2017).

Let us consider that for each image \( I \) the CNNs return, \( L \) 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 \( L_r, r = 1 \ldots R \). 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 \((i,j)\) corresponds to day \( i = 1, \ldots |I_u| \), and \( j = 1, \ldots J \times R \). 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)(Blei et al., 2003), 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 (Blei et al., 2003):

(a) Choose \( \theta_i \sim \text{Dirichlet}(\alpha) \), where \( i \in \{1, ..., D\} \).

(b) For each of the \( N_i \) words \( w_{ij} \) in document \( i \):
   i. choose a topic \( z_{ij} \sim \text{Multinomial}(\theta_i) \)
   ii. choose a word \( w_{ij} \) from \( P(w_{ij}|z_{ij}, \beta) \sim \text{Multinomial probability on} \)
       the topic \( z_{ij} \).

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

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

where the parameters \( \alpha \) and \( \beta \) are sampled only once in the process of generating the corpus, while the variables \( \theta_i \) 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 (Blei et al., 2003):

- 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 $\theta_i$.

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. 3.7 showing a day represented by the most important topics (with the highest probability) and the relations between topics and words.

Figure 3.7: 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.

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. 3.7). In order to find similar days, we need a
3.4. Unsupervised routine discovery relying on topic models

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 (Keogh and Pazzani, 2001) for computing the similarity of topics representation among days. DTW is an algorithm that computes the optimal alignment between two sequences, 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 \((w_1, w_2, ..., w_Q)\) 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:

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

In our proposed model, we employ the fastDTW algorithm (Salvador and Chan, 2007), 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 similarity. 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 (Yu and Shi, 2003) 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 \(G = (V_n, E, W_e)\), being \(V_n\) the set of nodes, \(E\) the set of edges and \(W_e\) 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 \(k_c\) of clusters to be defined, which without loss of generality, we set to 2 for the discovery of Routine and Non-Routine related days.
3. Routine Discovery from Egocentric Images

3.4.1 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.

EgoRoutine - An egocentric dataset for behaviour analysis

In this work, we propose and make publicly available the EgoRoutine dataset\(^2\). This dataset is composed of recorded days by 7 individuals who wore the Narrative Clip camera\(^3\) 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 3.4 and Fig. 3.8 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.

<table>
<thead>
<tr>
<th>User ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num Days</td>
<td>14</td>
<td>10</td>
<td>16</td>
<td>20</td>
<td>13</td>
<td>18</td>
<td>13</td>
<td>104</td>
</tr>
<tr>
<td>Images per day</td>
<td>20521</td>
<td>9583</td>
<td>21606</td>
<td>19152</td>
<td>17046</td>
<td>16592</td>
<td>10957</td>
<td>115430</td>
</tr>
</tbody>
</table>

Table 3.4: Total number of recorded days and collected images per user.

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. 3.9, 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 3.5, 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 classify sequences of images describing a long period of time. We observed that in the majority of 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.

\(^2\)http://www.ub.edu/cvub/dataset/
\(^3\)http://getnarrative.com/
3.4. Unsupervised routine discovery relying on topic models

Figure 3.8: 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.

Figure 3.9: 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.)
Routine Discovery from Egocentric Images

related. Therefore, from a total of 104 recorded days, 65 days are Routine related, and 39 are Non-Routine related. In Fig. 3.10 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.

<table>
<thead>
<tr>
<th>Class</th>
<th>Six Agree</th>
<th>Five Agree</th>
<th>At Least Four Agree</th>
<th>At Least Three Agree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>47</td>
<td>29</td>
<td>18</td>
<td>10</td>
<td>104</td>
</tr>
<tr>
<td>Routine</td>
<td>35</td>
<td>22</td>
<td>8</td>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td>Non-Routine</td>
<td>13</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 3.5: Summary of the agreement among the 6 individuals that labelled the collected photo-streams into Routine or Non-Routine related days.

![Figure 3.10: Number of Routine and Non-Routine days for each user (U) in the EgoRoutine dataset.](image)

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:

$$F_1 = \frac{2P \cdot R}{P + R}, P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.3)$$

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.

**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 [Redmon, 2018] and Xception [Chollet, 2017]. These models were trained on the COCO [Lin et al., 2014] and ImageNet dataset [Deng et al., 2009], 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 [Zhou et al., 2017].

3. **Activities recognition**: We use the activity labels given by the CNN proposed in [Cartas et al., 2017], which was trained for the recognition of 21 different daily activities. We select the activity label with the highest probability per image.

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.
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 [Farrahi and Gatica, 2011]: 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 [Rokach and Maimon, 2005] for the discrimination among days.

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 3.6, 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
3.4. Unsupervised routine discovery relying on topic models

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. In Table 3.8, we show the concepts that are detected by the different evaluated CNNs in a given photo-stream. Overall, the detected places given 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

<table>
<thead>
<tr>
<th>Concept</th>
<th>Xception</th>
<th>Yolo Redmon</th>
<th>Activities Cartas et al.</th>
<th>Places Zhou et al.</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>P</td>
<td>R</td>
<td>Acc</td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>SpClus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HierClus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TimeSlot</th>
<th>SpClus</th>
<th>HierClus</th>
<th>Generic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.6: 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 (Farrahi and Gatica, 2011). 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).
3. Routine Discovery from Egocentric Images

<table>
<thead>
<tr>
<th>User</th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
<th>User 4</th>
<th>User 5</th>
<th>User 6</th>
<th>User 7</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>0.79</td>
<td>0.74</td>
<td>0.75</td>
<td>0.90</td>
<td>0.92</td>
<td>0.56</td>
<td>0.92</td>
<td>0.80</td>
</tr>
<tr>
<td>F1</td>
<td>0.75</td>
<td>0.70</td>
<td>0.71</td>
<td>0.89</td>
<td>0.92</td>
<td>0.50</td>
<td>0.92</td>
<td>0.77</td>
</tr>
<tr>
<td>P</td>
<td>0.75</td>
<td>0.75</td>
<td>0.70</td>
<td>0.89</td>
<td>0.93</td>
<td>0.56</td>
<td>0.94</td>
<td>0.79</td>
</tr>
<tr>
<td>R</td>
<td>0.86</td>
<td>0.79</td>
<td>0.75</td>
<td>0.89</td>
<td>0.93</td>
<td>0.60</td>
<td>0.92</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 3.7: 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.

is due to the fact that the CNN was trained with 80 different categories corresponding to Common Objects in Context (COCO [Lin et al., 2014]). 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 [Deng et al., 2009]). 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 the network used 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 miss-interpreted as the lights in the cabin of an airplane.

<table>
<thead>
<tr>
<th>Time-slot (h)</th>
<th>9-11</th>
<th>11-14</th>
<th>14-17</th>
<th>17-19</th>
<th>19-21</th>
<th>21-24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xception</td>
<td>office</td>
<td>23</td>
<td>screen</td>
<td>203</td>
<td>desktop pc</td>
<td>75</td>
</tr>
<tr>
<td>Places</td>
<td>airplane cabin</td>
<td>167</td>
<td>conference room</td>
<td>49</td>
<td>office</td>
<td>131</td>
</tr>
<tr>
<td>Cartas et al. [2017]</td>
<td>shopping</td>
<td>40</td>
<td>mobile</td>
<td>29</td>
<td>mobile</td>
<td>40</td>
</tr>
<tr>
<td>Activity</td>
<td>walking</td>
<td>36</td>
<td>walking</td>
<td>36</td>
<td>meeting</td>
<td>46</td>
</tr>
<tr>
<td>Yolo</td>
<td>person</td>
<td>146</td>
<td>tv monitor</td>
<td>383</td>
<td>person</td>
<td>202</td>
</tr>
<tr>
<td></td>
<td>chair</td>
<td>38</td>
<td>laptop</td>
<td>534</td>
<td>chair</td>
<td>108</td>
</tr>
</tbody>
</table>

Table 3.8: 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.

- **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 3.6, we can observe that the model better performs when the day is described by analyzing the time division proposed in [Farrahi and Gatica](#).
3.4. Unsupervised routine discovery relying on topic models

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 wearer. 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 activities 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 80% accuracy and 77% $F_1$ (see Table 3.6).

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 [Farrahi and Gatica, 2011]. 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 [Rokach and Maimon, 2005] (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
3. Routine Discovery from Egocentric Images

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 3.6 that the Spectral Clustering classifier leads to 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 3.7 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 the difference 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. 3.9, 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 until late in the evening. In contrast, on Day 1, User 1 spent much less hours at work and left the office much earlier. This could be a cause of miss-classification by the model. Non-Routine related days contain events where the user works for short periods and spends longer time interacting with colleagues or friends. Day 7 is an example where User 1 is going 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 to be a rich tool for behaviour visualization. In Fig. 3.11 we present how the found topics could be analysed by the wearer or an expert. As an example for 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 environments. We understand that activ-
3.4. Unsupervised routine discovery relying on topic models

Figure 3.11: 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 (Farrahi and Gatica, 2011), 8 topics and personalized approach.

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. 3.12 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’ correspond 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. 3.12 between similar days and given labels.
Finally, in Table 3.9 we compare the obtained results for routine discovery to the routine discovery in (Talavera et al., 2019). As one can see the method in (Talavera et al., 2019) 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 (Talavera et al., 2019) 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 modeling, 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 (Talavera et al., 2019), 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 modeling DTW and spectral clustering are applied on the collection of documents composed of detected semantic concepts.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Users</th>
<th>Acc</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine discovery (Talavera et al., 2019)</td>
<td>5</td>
<td>0.76</td>
<td>0.69</td>
</tr>
<tr>
<td>Routine discovery propose here</td>
<td>7</td>
<td>0.82</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Table 3.9: Comparison between our previous work introduced in (Talavera et al., 2019) and the model here proposed for routine discovery from egocentric photo-streams.
3.5 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\footnote{\url{http://www.ub.edu/cvub/dataset/}} 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.

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