

# Discovering activity patterns in office environment using a network of low-resolution visual sensors

Mohamed Eldib<sup>1</sup> · Francis Deboeverie<sup>1</sup> · Wilfried Philips<sup>1</sup> · Hamid Aghajan<sup>1,2</sup>

Received: 30 August 2016 / Accepted: 21 May 2017 / Published online: 21 June 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** Understanding activity patterns in office environments is important in order to increase workers' comfort and productivity. This paper proposes an automated system for discovering activity patterns of multiple persons in a work environment using a network of cheap low-resolution visual sensors (900 pixels). Firstly, the users' locations are obtained from a robust people tracker based on recursive maximum likelihood principles. Secondly, based on the users' mobility tracks, the high density positions are found using a bivariate kernel density estimation. Then, the hotspots are detected using a confidence region estimation. Thirdly, we analyze the individual's tracks to find the starting and ending hotspots. The starting and ending hotspots form an observation sequence, where the user's presence and absence are detected using three powerful Probabilistic Graphical Models (PGMs). We describe two approaches to identify the user's status: a single model approach and a two-model mining approach. We evaluate both approaches on video sequences captured in a real work environment, where the persons' daily routines are recorded over 5 months. We show how the second approach achieves a better performance than the first approach. Routines dominating the entire group's activities are identified with a methodology based on the Latent Dirichlet Allocation topic model. We also detect routines which are characteristic of persons. More specifically, we perform various analysis to

determine regions with high variations, which may correspond to specific events.

**Keywords** Visual sensor network · Supervised learning · Probabilistic graphical models · Topic models · Sequence mining

## List of symbols

$(x', y')$	Tracked position
$f'$	Probability density function of tracked positions
$B$	Kernel function in $f'$
$H$	Bandwidth matrix in $f'$
$(x'^{(m)}, y'^{(m)})$	Position in cluster $m$
$m$	Cluster index
$L$	Number of clusters
$K^{(m)}$	Number of positions in cluster $m$
$U^{(m)}$	Positions matrix
$C^{(m)}$	Covariance matrix of clustered positions
$\lambda_1^{(m)}, \lambda_2^{(m)}$	Eigenvalues of the covariance matrix $C^{(m)}$
$\rho$	Correlation coefficient
$\sigma_{x'^{(m)}}$	Standard deviation of clustered $x'^{(m)}$ position
$\sigma_{y'^{(m)}}$	Standard deviation of clustered $y'^{(m)}$ position
$A$	Ellipse confidence level
$V^{(m)}$	Eigenvectors matrix
$D^{(m)}$	Diagonal matrix
$\theta^{(m)}$	Rotation angel of the ellipse
$S$	Start of hotspot
$E$	End of hotspot
$T$	Track with start hotspot $S$ and end hotspot $E$
$I$	Number of positions in track $T$
$(g_x^{(m)}, g_y^{(m)})$	Center of hotspot (cluster) $m$
$F$	Hotspot threshold
$x_t$	Feature vector at time instant $t$
$x_{1:T}$	Sequence of observations

✉ Mohamed Eldib  
mohamed.eldib@ugent.be

<sup>1</sup> Image Processing and Interpretation, TELIN, Ghent University/imec, Sint Pietersnieuwstraat 41, 9000 Gent, Belgium

<sup>2</sup> Ambient Intelligence Research Lab, David Packard Building, Stanford, CA 94305-9515, USA

$\mathbf{y}_{1:T}$	Sequence of states
$T$	Number of time steps
$\hat{N}$	Number of persons
$Q$	Number of states
$p(\mathbf{x}_{1:T}, y)$	Joint distribution of $\mathbf{x}_{1:T}$ and $y$
$\gamma, \delta$	Potential functions
$\epsilon$	Actual potential
$f$	Feature function
$p(\mathbf{y}   \mathbf{x})$	Conditional probability of $\mathbf{y}$ given $\mathbf{x}$
$Z_x$	Normalization function
$N$	Pattern length
$e$	Word in a document (presence status)
$d$	Document of words (presence status sequence)
$z$	Topic (activity pattern)
$K$	Number of topics
$G$	Number of constructed words
$M$	Number of documents
$V$	Vocabulary of words
$\theta$	Matrix of day-specific mixture weights
$\Phi$	Matrix of word-specific mixture weights
$\alpha$	Per-document activity pattern distributions hyperparameter
$\beta$	Per-activity word distribution hyperparameter
MAE	Mean absolute error
$v_r$	Estimated presence duration for hour $r$
$v'_r$	Actual presence duration for hour $r$
$\hat{H}$	Number of hours
RAE	Relative absolute error
$\xi$	Training corpus model
$G_m$	Length of the document $m$

## 1 Introduction

The productivity of a person in a work environment is associated with several factors such as workloads, social support and time pressures. These factors can contribute to increase or decrease the stress levels in the workplace. Stress is undesirable, because it is the second cause of Europe's health problems (EU-OSHA 2013a). It costs the European Union 20 billion Euro (Cosemans et al. 2014). In 2005, 22% of Europe's workers suffered from it (Milczarek et al. 2009), 51% of Europe's workers confess stress is common in their workplace (EU-OSHA 2013b), and 50–60% lost working days in Europe are due to stress (EU-OSHA 2013a).

Stress can cause long-term health and economic consequences. Workers may suffer from big long-term physical and mental problems (Bickford 2005) such as depression, anxiety, heart disease, chronic fatigue syndrome, diabetes and osteoporosis. These health problems lead to economic consequences to organizations such as absenteeism, staff

turnover and tardiness increase (Milczarek et al. 2009) which decrease the organization's production. Also, workers may present in the workplace, but they do not work with their full capacity and this is known as "presenteeism". A recent study (Cosemans et al. 2014) showed that presenteeism and absenteeism cost the organizations an annual loss of 242 billion Euro in terms of decreased productivity.

It is of a significant importance to detect changes over time in the psychological patterns and activity patterns to ensure a less stressful work environment and a more productive worker. If unhealthy or inefficient activity patterns are detected, then change toward more healthy or more efficient habits can be recommended. Finally, understanding activity patterns benefit individual well-being and personal productivity. The analysis of the psychological changes are hard to detect directly. This requires the worker to fill self-report questionnaires such as Stress Self Rating Scale (SSRS) or being interviewed by a psychologist. The psychological analysis can be taken from time to time, but may not be suitable for detecting the subtle changes which could lead to an early sign of a major problem. Also, the psychological analysis is only conducted when the worker asks for the analysis or the people around him notice that the severity of the situation increased. Sometimes, people may not be able to assess themselves in problems.

A work environment equipped with appropriate sensor devices and actuators is referred to an "Intelligent Office". Understanding activity patterns of persons in an intelligent office can be used to optimize the productivity and the comfort of the workers'. The sensory signal outputs from an office monitoring system can be used to recognize several activity patterns such as "arriving to work late", "leaving the office early", "working non-stop" and so on. By learning and detecting activity patterns for long-term, the environment becomes aware of each person's preferences in order to increase work productivity and decrease stress. For example, a person who works continuously for longer hours than usual without a break, the environment can recommend him to have a coffee break. In another situation, when the environment notices a change in a person's behavior by arriving and leaving the office late, the environment can notify him how such a change in his habit can make him less social interactive. Based on observations and learned models, the environment compares how the observations deviate from previous activity patterns, in order to suggest healthier habits.

Humans perform activities based on habits, so inferring patterns which describe the past and present activities is important in order to define future activities as well. In that sense, an environment can proactively activate and deactivate some devices based on learnt patterns (e.g. switching off the computer automatically when a person leaves his office). Apart from automating actions or devices, patterns

can also be used to understand a person's activity behavior (Oliver et al. 2004) and act in accordance with it (e.g. issuing meeting reminders). Also, making the environment more efficient in terms of saving energy (Cheng and Lee 2014; Salamone et al. 2016) (e.g. switching off the lights when a person has gone to lunch or a meeting) or increasing safety (Mrazovac et al. 2011) (e.g. locking office door when a person is not present). Having such a system installed in an environment could help to improve work productivity and encourage people to manage stress.

In this paper, we utilize low-resolution visual sensors to build an office monitoring system. The system is installed in an office environment where multiple persons are working. The system has been operational for 5 months. The computer vision algorithms used in this paper are based on vision algorithms developed in the research project "Little Sister: Low-cost monitoring for care and retail (iMinds 2013) which focuses on creating a sensor-based monitoring system that can match, in terms of performance, a combination of the body-worn devices and the high-resolution cameras at a much reduced cost. They are also one of the core components of the Ambient Assisted Living Joint Programme project "SONOPA: Social Networks for Older adults to Promote an Active life" (Docobo 2013). In SONOPA, the aim is to combine a social network with activity recognition in a smart home environment to stimulate and support activities and daily life tasks. SONOPA suggests suitable activities and social connections to the senior citizen automatically, proactively and at the optimal time, while providing a simple bridge to the social network of the senior citizen. SONOPA achieves this by analyzing both physical and online activities of senior citizen users in their smart homes. This paper extends and improves the work of SONOPA and Little Sister with probabilistic graphical models, sequence mining techniques and topic models.

Our focus is the automatic discovery of activities from persons' trajectories collected by low-resolution visual sensors over the course of 5 months. We define activities to be temporal regularities in people's lives. An activity often involves patterns of being present or absent in the office over time (e.g. being in the office or going to lunch), possibly over varying time scales and for different time intervals. Automatic activity classification and discovery face several challenges and obstacles as people's habits often vary from day to day and from individual to individual, and sensors can deliver incomplete and noisy data. A supervised learning approach to activity recognition would require data to be labeled with the actual activities (the "ground truth" labels) (Kim et al. 2010). In contrast, an unsupervised learning approach can automatically discover meaningful patterns in the emerging activities of people without requiring training data. Activity discovery enables the

possibility of sifting through large amounts of noisy data. Furthermore, the data can be clustered (i.e. people or days) corresponding to the most common activities (those of several people) and discover the dataset structure with minimal prior knowledge.

In this work, we develop a framework built on several components to discover activity patterns. The contributions of this work are the following:

1. We install a network of low-resolution visual sensors in an office environment, in order to discover several activity patterns such as arriving to the office early or late, leaving the office early or late, going to lunch outside the office, eating lunch inside the office and attend meetings. The activity patterns span 5 months of real-life data in an office environment of multiple persons. In contrast to earlier research (Oliver and Horvitz 2005), we monitor real-life office activities without resorting to simulations. Simulated data are obtained by people acting office life-style may risk not being representative. Moreover, they are by necessity short, making it difficult to study the analysis of long-term trends.
2. We propose a methodology to estimate the users hotspots. Firstly, the persons' positions are extracted using a recursive maximum likelihood tracker (Bo et al. 2014). Then, the underlying distribution of the mobility tracks is examined using a bivariate kernel density estimation in order to extract the high estimated density of the persons' positions. Finally, the confidences ellipses of the high density positions are computed to define the persons' hotspots.
3. We introduce two approaches to estimate the presence or absence of users in the office. We use supervised learning methods to train the models in our two proposed approaches. Both approaches use three powerful Probabilistic Graphical Models (PGM), namely Naïve Bayes (NB), Hidden Markov Model (HMM) and Linear-Chain Conditional Random Field (LC-CRF). The first approach is based on a single model, while the second approach employs sequence mining technique with two models. We compare both approaches against collected ground truth for 12 days using three persons. In this step, the parameters of the models are trained using 2 days of data.
4. We present a methodology for the automatic discovery of daily activity patterns with Latent Dirichlet Allocation (LDA) (Blei et al. 2003), where we discover activity characteristics of all days in the dataset.
5. We analyze the model outputs to recommend more healthy and more efficient activity patterns. Our analysis includes finding activities which dominate on certain kinds of days; finding days which are well repre-

sented by few or many topics; finding a given person's dominating daily patterns; finding low-entropy and high-entropy activity characteristic days; determining when a large variation occurs for a given person's activity over time; and discovering groups of persons that follow certain trends.

Our overall objective is to determine what individual and group routines are contained in the low-resolution video dataset. The discovered routines could help us to understand how we can optimize the work environment by providing recommendations in case of unhealthy habits, issuing reminders in case of meetings or social events, and making the environment more efficient in terms of saving energy. The remainder of the paper is organized as follows. Related work in literature is listed in the next section. Section 3 gives an overview of the work environment set-up. Then, we discuss the hotspot detection method in Sect. 4, followed by explaining the proposed architectures for person status identification in Sect. 5. Section 6 introduces topic model for discovering activity patterns. We present and discuss the experimental results in Sect. 7. Finally, Sect. 8 draws conclusions.

## 2 Related work

The sensors used in office environments can be divided into two main categories: (1) wearable sensors and (2) ambient sensors. In the first category (Cinaz et al. 2013; Okada et al. 2013; Healey and Picard 2005), various wearable sensors, such as accelerometers, gyroscopes, proximity sensors, and e-textile sensors are attached to the subject to monitor physiological signals such as electrocardiogram (ECG), electroencephalogram (EEG), electromyogram (EMG), blood pressure, and respiration. Wearable sensors face a few disadvantages, such as limited battery life, high cost, missing data when the user forgets to wear the device, and the need to attach them to specific body parts to provide reliable measurements. In the second category, ambient sensors are installed in office environment by mounting them on the wall or the ceiling and/or embedding them in furniture and appliances. The advantage of using ambient sensors to measure activity patterns is that unlike wearable sensors, they can normally be done in a totally unobtrusive manner, and without the need of expensive extra equipment. The common ways to study the activity patterns of individuals are Keystroke (Zimmermann et al. 2003), mouse dynamics (Liao et al. 2005), computer exposure (Eijkelhof et al. 2014), and intelligent environments (Azitiria 2010). On the other hand, the most popular ambient sensors in research are Passive Infrared Motion (PIR) sensors, visual sensors (including special technologies such as depth cameras) and Radio Frequency Identification (RFID).

Tables 1 and 2 summarize the different capabilities and properties of three sensors: PIR, Kinect and visual sensors. In Table 1, four capabilities such as location, presence, shape and tracking of the three technologies are compared. PIR sensors have limited capabilities when they are compared to Kinect and visual sensors. PIR sensors can provide good presence detection accuracy, but they can not provide very accurate information about the exact location (e.g.  $x$  and  $y$  positions). Also, PIR sensors can not track multiple persons at the same time or do shape detection. On the

**Table 1** Comparison between the different capabilities of PIR, visual and Kinect sensors

Technology	Location detection	Presence detection	Shape detection	Tracking
PIR sensors	L	M	Not possible	Single person
Visual sensors	H	H	L	Multiperson
Kinect sensors	H	H	H	Multiperson

*H, M and L* stand for high, medium and low values, respectively

**Table 2** Comparison between the different properties of PIR, visual and Kinect sensors

Properties	PIR sensors	Visual sensors	Kinect sensors
Network density	H	M	M
Resolution	Single pixel (on/off)	30 × 30 pixels	IR depth sensor: 640 × 480 pixels Color camera: 1280 × 1024 pixels
Space occupancy	L	L	H
Cost	L	M	H
Privacy concern	L	L	H
Operation (lighting)	No	Yes	IR depth sensor: no; Color camera: yes
Applicability	Indoors	Indoors	Indoors/Outdoors
Battery life	H	L	Not possible

*H, M and L* stand for high, medium and low values, respectively

contrary, Kinect and visual sensors have highly accurate location and presence detections, and both technologies can track multiple persons. Shape detection and skeleton extraction can be done more accurately using Kinect than visual sensors.

Table 2 shows several properties of PIR, Kinect and visual sensors:

- *Network density* The number of sensors required to be installed in an area to provide some specific service. In (Teixeira et al. 2010), the authors quantified the network density (ND) using the order of magnitude (in base 2) of the number of sensors. For instance, if a single camera can detect a person within area  $A$ , then the density of the camera solution is  $\log_2(1) = 0$ . PIR sensors require a high network density to provide accurate locations ( $ND = 4$ ). A high ND requires a complex infrastructure, cumbersome to install and manage.
- *Resolution* PIR sensors return a state “on” if human presence is detected within a certain sensing area, otherwise a state “off” is returned. Kinect has an Infrared depth sensor with an image resolution of  $640 \times 480$  pixels and a color camera sensor with an image resolution of  $1280 \times 1024$  pixels. Visual sensors provide an image resolution of  $30 \times 30$  pixels.
- *Space occupancy*: The dimensions of Kinect, visual, and PIR sensors are ( $w \times d \times h$ ):  $37 \times 15 \times 12$ ,  $6.2 \times 4.1 \times 2$  cm<sup>3</sup> (Camilli and Kleihorst 2011),  $3.2 \times 2.5 \times 2.8$  cm<sup>3</sup>, respectively. The Kinect sensor clearly occupies more space than PIR and visual sensors.
- *Cost* The Kinect sensor has advanced hardware components. This increases the price per unit (above 100 Euros), while the bill material of the visual sensor is under 25 Euros (Camilli and Kleihorst 2011). The PIR sensor is the cheapest solution.
- *Privacy concern* User studies in the projects Little Sister and SONOPA indicated that the users attach high priority to privacy, and they agreed to install low-resolution cameras (e.g. visual sensors) or PIR sensors, but not high-resolution cameras (e.g. Kinect) which often raises privacy concerns. Visual sensors pose very little privacy issues since they are not capable of gathering detailed information.
- *Operation* PIR sensors and the infrared depth sensor in Kinect do not require lighting conditions to operate, while visual sensors and the color camera in Kinect require sufficient lighting conditions to operate.
- *Applicability* PIR and visual sensors can only be used in indoors scenarios (e.g. behavior analysis), while Kinect sensors can be used indoors and outdoors (e.g. car tracking).
- *Battery life* PIR sensors have a longer battery life than Kinect and visual sensors, because PIR sensors con-

sume less processing power. Kinect and visual sensors are installed in a wired setup and powered by mains electricity. Given the low power consumption of the visual sensors, it is possible to operate them on battery over prolonged periods of time.

From the detailed comparison in Tables 1 and 2, Kinect and visual sensors have similar and more powerful capabilities than PIR sensors. Furthermore, the properties of the visual sensors are more suitable than Kinect for office monitoring systems, because of the affordable price, and the preservation of privacy (Ziefle et al. 2011). The images produced by the visual sensors are  $30 \times 30$  pixels. In these images privacy is maintained, thus it is for instance hard to recognize faces. However, they are very useful in our office monitoring system to recognize activity patterns. Examples of activity patterns are arriving to the office, leaving the office. An example of a behavioral change is increased or decreased mobility measured from speed or walked distance (Bo et al. 2014).

A single PIR sensor records the worker’s activities with only a binary state indicating whether there is a motion detected within its detection range. Thus, datasets recorded using PIR sensors are in fact a time series of sensor activation events, which contain very limited information that can be used to identify the corresponding individual. While, a single camera can capture rich information of different levels of granularities, from the gross movements of subjects similar to that provided by simple motion detection sensors to richer information about posture, body motion, head and body orientation, fidgeting, and so on. In most cases, multiple PIR sensors and cameras are used in office environments.

In the activity analysis field, researchers have developed and applied several machine learning methods to recognize human activities (e.g. sitting, standing, or walking) from various types of sensor data. The machine learning methods are divided into supervised learning and unsupervised learning approaches. In the supervised approach, the task of recognizing activities can be easily formatted into a classification problem where the model relies on labeled data for training the desired activities. Tao et al. (2011) introduced a system of 43 PIR sensors which were attached to the ceiling of a research room. The system used person localization algorithm for providing various personalized services. The algorithm assumes every person wants to go back to their desk after a certain task. The system achieved an accuracy of 84% using support vector machine. Jaramillo and Amft (2013) studied the energy efficiency by controlling desk appliances such as computer screens. They used PIR sensors and screen-attached ultrasound sensors to recognize desk activities (ScreenWork, DeskWork, Away) through classification.



Then, the classifier output is mapped into on/off switching states for the screen power controller.

Moreover, probabilistic graphical models, such as HMM, dynamic Bayesian network, and Conditional Random Fields (CRFs), have been used to model the activity transition sequence for activity recognition purposes. In Oliver and Horvitz (2005), the authors compared Layered HMMs (LHMMs) (Oliver et al. 2002, 2004) and dynamic Bayesian networks for identifying office activities from multi-modal sensors such as video, audio and the user's interaction with the computer. Then, dynamic Bayesian networks are only included at higher levels of the LHMMs, where the results of previous layers (inferential layers using HMMs) are used. 90 minutes of activity data were used to test the performance of both models. In Milenkovic and Amft (2013), the authors used LHMMs and Finite State Machines (FSMs) to recognize office worker activities that are relevant for energy-related control of appliances using PIR sensors. They evaluated their approach in a living-lab office, including three private and multi-person office rooms for 5 days. Wojek et al. (2006) proposed a multi-level HMM framework for multi-person activity recognition (meeting, paperwork, discussion, etc) with simultaneous tracking of users in the room using audio and video cues. Chen et al. (2011a, b, c) studied the problem of discovering the social interactions in office environments using a network of high-resolution cameras and RFID. The head poses and the locations of people are tracked using Chamfer matching. Then, a classifier is used to estimate the head orientation based on the location, relative distance and head orientation of people, a probabilistic model is used to infer the use of space by individuals and their interactive behavioral patterns.

Even though the majority of the proposed activity recognition approaches are supervised methods, most of them share the same limitation that the accurate activity labels for PIR sensor datasets and cameras are very difficult to get. For almost all of the current testbeds with PIR sensors and cameras, the data collection and data labeling are two separate processes for which the activity labeling for the collected data is extremely time consuming and laborious because it is usually based on direct video coding and manually labeling. Clearly, this limitation prevents the supervised approaches from being easily generalized to the real-world situation where activity labels are usually not available for a huge amount of sensor data. Therefore, many unsupervised approaches have been proposed to handle the problem that activity labels are not available. In Chen et al. (2011a), a system consisting of a visual processing and a learning module are proposed to discover accurate patterns that represent the user's frequent behaviors in office by associating the semantic locations of the user to activities. Hamid et al. (2009) proposed the idea that global structural information of human activities can be encoded

using a subset of their local event sequences. They regarded discovering structure patterns of activity as a feature selection process. Si et al. (2011) studied the daily activities of students in office from videos, by automatically learning event grammar under the information projection and minimum description length principles in a coherent probabilistic framework, without manual supervision about what events happen and when they happen.

Topic models (Blei et al. 2003) have gained an increasing attention in recent years as an unsupervised learning approach for activity discovery. The topic models are designed for text mining and discovering the main themes that pervade a large corpus of documents. In topic models, the documents are represented as mixtures of topics, learned in a latent space, and they offer ways to organize documents, words and other entities through clustering and ranking. They have the ability to characterize discrete data represented by bags. These models are advantageous to capture which words are important to a relevant topic as well as the prevalence of those topics within a document, resulting in a rank measure. In (Farrahi and Gatica-Perez 2011), the authors studied the routines of 97 subjects using mobile phone sensor data over one year. They applied the probabilistic topic models to automatically discover routines, such as "being at work" or "going home from work". They replaced words with bag of location sequences, documents with days and topics with routines. Huynh et al. (2008) used topic models to discover routines, such as "lunch" and "office work" from recognized activity primitives. The authors used on-body sensor data from one subject over 16 days. They tested their approach on short-term scenarios using 7 days. One of the limitations in Huynh et al. (2008), their approach requires a higher level information regarding a person's activities. Kim et al. (2010) proposed a topic model approach based on pairing activity recognition and activity discovery. In Castanedo et al. (2014), the authors discovered patterns in sensor data for long-term using topic models. Their analysis provided insights on the ability to discover routines that represent the common activities gathered from the sensor network. They tested their model on two real datasets with more than 100 sensors and over 50 weeks of data. Varadarajan et al. (2013) identified recurrent activity sequences from motion patterns in traffic videos using topic models.

In our approach of discovering activity patterns, we do not use supervised learning as in Tao et al. (2011) and we do not analyze the power use of office equipments as in Milenkovic and Amft (2013). The authors in Chen and Aghajan (2011); Wojek et al. (2006); Oliver et al. (2002)), used high-resolution cameras which offer access to details of office activities, but are regarded with caution in terms of coping with user privacy concerns and increasing the cost of the sensor network. Additionally, they used data of

simulated office activities for relatively short-time periods (several days). We took a different approach by employing a network of low-resolution visual sensors ( $30 \times 30$  pixels) (Camilli and Kleihorst 2011). The low-resolution nature of the visual sensors maintains the user’s privacy. Our activity pattern study includes multiple persons and spans a long-term period of 5 months using real-data recordings. On the other hand, topic models have been used with PIR sensors (Castanedo et al. 2014), mobile phones (Farrahi and Gatica-Perez 2011) and wearable sensors (Huynh et al. 2008) data. There is relatively little work on topic models using visual sensors network, their use has been limited to motion patterns in traffic videos (Varadarajan et al. 2013), but to our knowledge, their use for real-life activity discovery in

office environment from a multi-camera system is novel. The proposed low-resolution visual sensor network has shown promising results in the application of ambient assisted living (Eldib et al. 2015a; Xie et al. 2014; Eldib et al. 2016a, 2014b, 2015b, 2016c), absenteeism detection (Eldib et al. 2016b), and for person tracking (Eldib et al. 2014a; Bo et al. 2014).

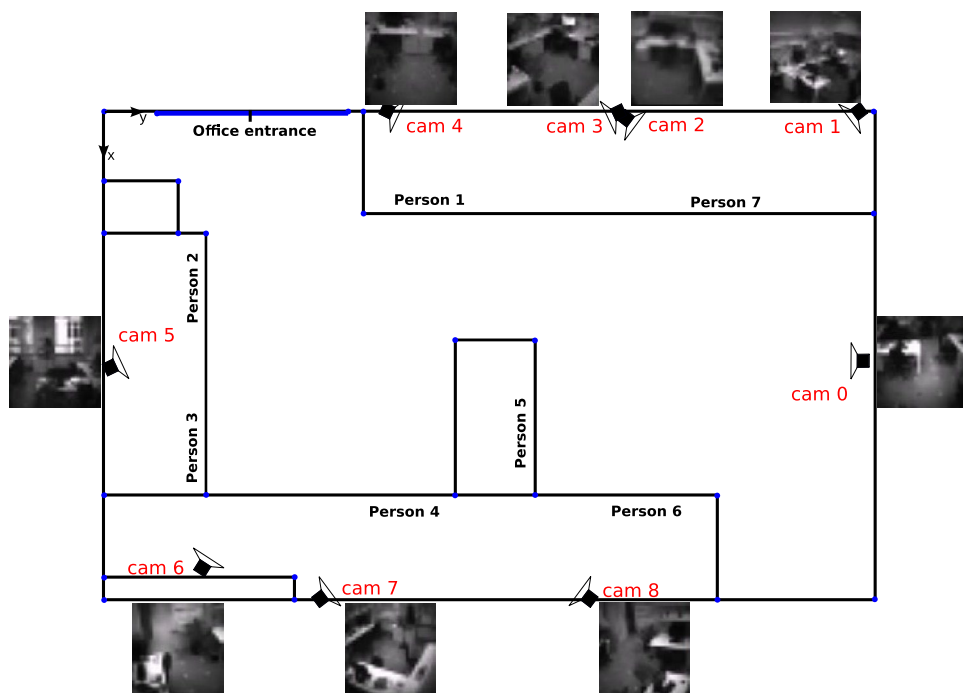
### 3 Office environment setup

The office environment is equipped with a network of nine visual sensors covering an area of  $8 \times 5 \text{ m}^2$ . Each visual sensor has a pair of image sensors ( $30 \times 30$  pixels resolution sensors used in computer mice) as shown in Fig. 1. An overview of the location of the visual sensors in the office environment is shown in Fig. 2. The visual sensor images often suffer from artifacts due to read-out problems such as electrical interference, and it does not have built-in processing capabilities, such as lens shading correction resulting in a reduction of the image’s brightness. The used lens in the low-resolution visual sensors need to focus the light properly on the imaging sensor, in order to produce a sharp image of the outside world. This typically causes an effect known as “vignetting”: the amount of light energy projected by the lens onto the sensor which create a pattern of concentric circle. This problem can be solved by correcting the peripheral shading which is known as “devignetting” on the digital signal controller.



**Fig. 1** The camera consists of stereo pair of image sensors controlled by a digital signal controller. Each image sensor delivers an image with a resolution of  $30 \times 30$  pixels

**Fig. 2** Office environment layout showing the configuration of nine visual sensors covering an area of  $8 \times 5 \text{ m}^2$

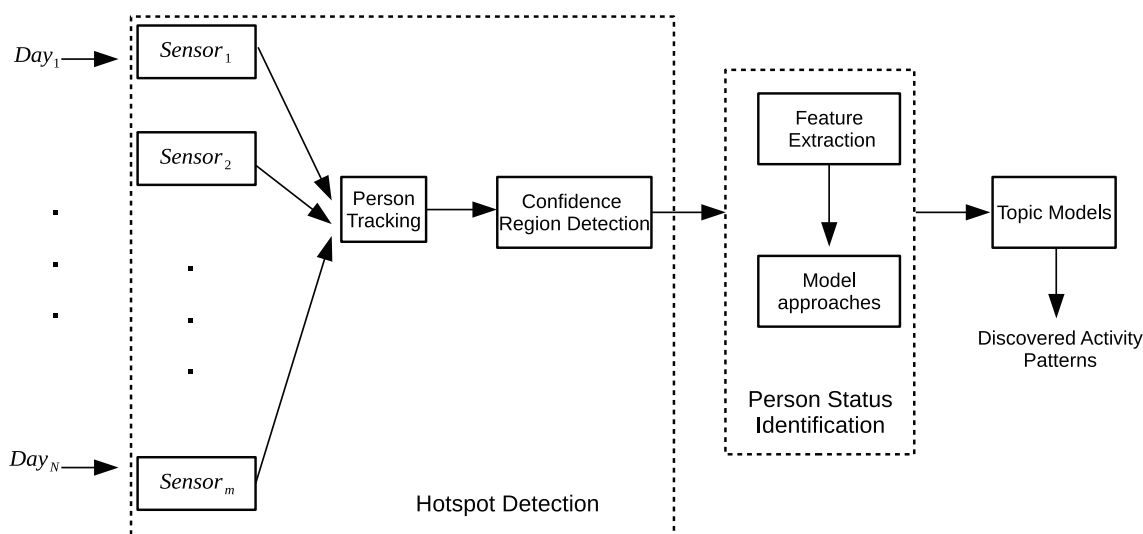


The cameras consist of two Agilent ADNS-3060 high-performance optical mouse sensors. These sensors are used in gaming applications. Camilli et al. (2011) used this sensor with small adaptation to produce video of  $30 \times 30$  pixels at 100 frames per second. The sensors connect over a Serial Peripheral Interface bus directly to the internal memory of the DSP which performs the video processing. In our work, each microcontroller in each sensor performs preprocessing, including devignetting (correcting for lower brightness at the periphery of the image), automatic gain control, and noise reduction.

Learning and understanding the activity patterns of each person in the current setup is challenging due to the following:

- More than six persons work in the same office room.
- Different activity patterns for each person (meetings, lunch time, arrival time, leaving time, etc).
- Regular visits from other colleagues to the office room.
- Real-life office activities without resorting to simulations.

Figure 3 shows a block diagram of our framework. First, the images are captured by different visual sensors. Then, the mobility patterns of several persons are extracted using a recursive likelihood tracker (Bo et al. 2014). From the persons' positions, the desk locations (hotspots) are found by examining the underlying distribution of the mobility tracks by employing a bivariate kernel density estimation. Using the start and end hotspots as a feature vector, we predict the people's presence inside the office by exploring two approaches. Based on the people's presence and the time of day, topic models are utilized for activity discovery.



**Fig. 3** A block diagram of the proposed framework

## 4 Hotspot detection

### 4.1 Tracking

In this component, the visual sensor video capturing and pre-processing are done as in Bo et al. (2014). We operate the visual sensor to produce images of  $30 \times 30$  pixels and an image depth of 6 bit per pixel. In the pre-processing stage, a de-noising step is applied by averaging the gray values of each pixel over time. The second pre-processing step is to produce a sharp image of the outside world by applying devignetting and also by correcting any pixel-dependent dark stream current in the visual sensors.

The images captured by the visual sensors suffer from noisiness and poor and quickly changing lighting conditions which are quite prominent indoors. In a previous study (Bo et al. 2014), several foreground/background algorithms have been tested to handle this effect. The correlation method has shown sufficient robust to illumination changes. In this paper, we opted to use the correlation method, as shown in Fig. 4. The correlation method parameters have been tuned to produce the best visualization results and to work with the minimum lighting conditions. As a future work, we plan to study different parameter settings. Table 3 summarizes the tuned parameters.

In previous studies (Eldib et al. 2014a; Bo et al. 2014), the Recursive Maximum Likelihood (RML) tracker has shown promising results for person tracking using low-resolution visual sensors. In this work, we use the RML tracker to extract the users' positions. After each visual sensor captures a new frame, the RML tracker analyzes the frame to separate moving objects from the static background using a correlation-base foreground detection method. This produces





Fig. 4 Foreground detection by the correlation method

Table 3 Tuned parameters of the correlation method

Parameters	Values
Size of $\omega$	3
$\rho_{min}$	0.985
$\alpha$	0.005

a number of blobs, some of which correspond to noise or non-interesting moving objects such as chairs. Each blob is then checked if it is overlapping with the bounding box of the tracked persons in the previous frame. Only non-overlapping blobs are matched across all camera views using homography and well-matched blobs are initialized as a new person for tracking. Next, in each camera view the likelihood that a person is in a particular position in the room is calculated using the known position in the previous frame as prior knowledge. The fusion center computes joint-likelihood based on the likelihood computed by each camera, and estimates the most likely new position of the person. Finally, jointly estimated positions are sent back to all camera views as a prior for likelihood computation in the next frame.

### 4.2 Confidence region detection

A hotspot is defined as a region or multiple regions where most of the persons' positions occur or where most of the time is spent. There are seven desk locations and one door entrance location. In order to obtain an occupancy map with the users' hotspots, we need to define the confidence region of the desk locations where each person spends most of the time. For this purpose, we use 1 week of observed data samples to estimate the underlying probability density function  $f'$ . Let  $\mathbf{s} = (x', y')$  be the output of the RML tracker which represents the person's position on ground plane in world coordinates. Let  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$  be a sample data of the persons' positions drawn from unknown density function  $f'$ . Then, the kernel density estimation function for bivariate data (Simonoff 1996) is defined as follows:

$$f'(\mathbf{w}; \mathbf{H}) = \frac{1}{n} \sum_{i=1}^n B_{\mathbf{H}}(\mathbf{w} - \mathbf{s}_i), \tag{1}$$

Table 4 Bandwidth selectors for kernel density estimation

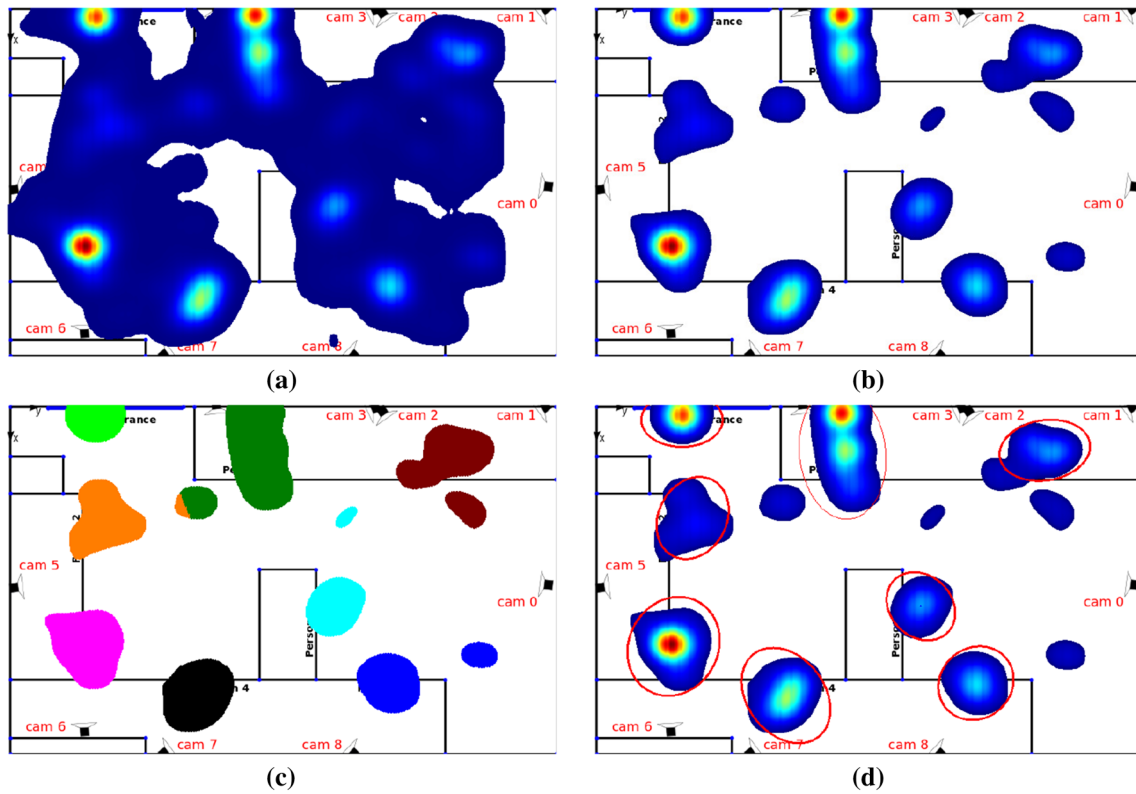
Bandwidth selector	$h_1^2$	$h_2^2$
Plug-in	54.044	61.850
Smoothed cross validation	58.505	66.956
Rule of thumb	54.426	62.287

where  $\mathbf{w} = (w_1', w_2')^T$ ,  $\mathbf{s}_i = (x_i', y_i')^T$  and  $i = 1, 2, \dots, n$ . Here  $B(\mathbf{w})$  is the kernel which is a symmetric probability density function.  $\mathbf{H}$  is the bandwidth matrix which is symmetric and positive-definite:

$$\mathbf{H} = \begin{bmatrix} h_1^2 & 0 \\ 0 & h_2^2 \end{bmatrix}, \tag{2}$$

where  $B_{\mathbf{H}}(\mathbf{w}) = |\mathbf{H}|^{-1/2} B(\mathbf{H}^{-1/2} \mathbf{w})$ . The choice of the kernel function  $B$  is not crucial. There are many kernel functions but the most popular are uniform, Epanechnikov and Gaussian kernels. We chose to use the standard normal throughout due to its convenient mathematical properties:  $B(\mathbf{w}) = (2\pi)^{-1} \exp(-\frac{1}{2} \mathbf{w}^T \mathbf{w})$ . In contrast, the choice of  $\mathbf{H}$  is important in evaluating the performance of  $f'$ . There are several approaches to select the optimal bandwidth matrix  $\mathbf{H}$  automatically such as plug-in (Sheather and Jones 1991), smoothed cross validation (Duong and Hazelton 2005) and rule of thumb (Silverman 1986). The three approaches generate similar bandwidth matrix  $\mathbf{H}$  for our data. Table 4 shows the output of  $\mathbf{H}$  using the three approaches. We compute the average results of the three approaches to get the final result of  $\mathbf{H}$ .

Figure 5a shows the bivariate kernel density estimation of the users' positions. Figure 5b shows the users' positions after only considering positions with high estimated density. We use the k-means clustering to detect and highlight the desk and door entrance locations from the users' positions (hotspots). We chose the number of clusters to be eight since there are seven desk locations and one door entrance location. Figure 5c shows the hotspots after applying the k-means clustering. Each hotspot (cluster) represents a distinct location (*Person 1, Person 2*, etc). Finally, we calculate the confidence



**Fig. 5** The steps to estimate the confidence regions for 1 week of observed data: **a** the kernel density estimation of the users’ positions; **b** the high estimated density of the users’ positions; **c** k-means clusters; **d** confidence ellipses

ellipse of each hotspot to define the region that contains most of the samples that can be drawn from the underlying distribution. Let  $\mathbf{x}^{(m)} = (x_1^{(m)}, x_2^{(m)}, \dots, x_{K'}^{(m)})$  and  $\mathbf{y}^{(m)} = (y_1^{(m)}, y_2^{(m)}, \dots, y_{K'}^{(m)})$  be the  $x'$  and  $y'$  positions that belong to cluster  $m$ , where  $m = 1, \dots, L$ . Let  $\mathbf{U}^{(m)} = \begin{bmatrix} \mathbf{x}^{(m)} \\ \mathbf{y}^{(m)} \end{bmatrix}$  be a matrix that holds  $\mathbf{x}^{(m)}$  and  $\mathbf{y}^{(m)}$  positions in  $m$ . Let  $\mathbf{C}^{(m)}$  be the covariance matrix of  $\mathbf{U}^{(m)}$  which is given by the equation:

$$\mathbf{C}^{(m)} = \frac{1}{K^{(m)} - 1} \mathbf{U}^{(m)} \mathbf{U}^{(m)T} \tag{3}$$

A confidence region with an ellipse shape can be defined as follows:

$$\left( \frac{\mathbf{x}^{(m)}}{\sigma_{x^{(m)}}} \right)^2 + \left( \frac{\mathbf{y}^{(m)}}{\sigma_{y^{(m)}}} \right)^2 = A, \tag{4}$$

where  $\sigma_{x^{(m)}}$  and  $\sigma_{y^{(m)}}$  are the standard deviations and  $A$  defines the scale of the ellipse. The choice of  $A$  represents a chosen confidence level. our data is sampled from a distribution with a Gaussian kernel. This implies that  $\mathbf{x}^{(m)}$  and  $\mathbf{y}^{(m)}$  are normally distributed. In probability theory, a sum of the squares of independent normally distributed

data samples is known to be distributed according to chi-squared distribution with  $j$  degrees of freedom (Lancaster and Seneta 1969). In our case there are two unknowns, and therefore  $j = 2$ . To find the value of  $A$ , Table 5 gives the cumulative chi-square distribution (Lancaster and Seneta 1969) for 2-degrees of freedom and the probability values of different confidence intervals. For example,  $A$  is 5.99 when the confidence interval is 95% ( $p' = 1 - 0.95$ ). There are two cases need to be considered to find the confidence ellipse:

- If  $\mathbf{C}^{(m)}$  is a diagonal matrix, which happens when  $\mathbf{x}^{(m)}$  and  $\mathbf{y}^{(m)}$  are uncorrelated, and the ellipse axis are aligned with the frame axis (e.g.  $p = 0$ ).
- If  $\mathbf{C}^{(m)}$  is a non-diagonal matrix, the ellipse axis are not aligned with the frame axis (e.g.  $p \neq 0$ ).

**Table 5** The Chi-squared distribution table for 2-degrees of freedoms and confidence intervals of 90, 95 and 99%

$p'$ -value (probability)	$A$
0.10	4.60
0.05	5.99
0.01	9.21

In both cases, the length of the ellipse axis is related with the eigenvalues of covariance matrix  $\mathbf{C}^{(m)}$  given by:

$$\lambda_1^{(m)} = \frac{1}{2} \left( \sigma_{x^{(m)}}^2 + \sigma_{y^{(m)}}^2 + \sqrt{(\sigma_{x^{(m)}}^2 - \sigma_{y^{(m)}}^2) + 4\sigma_{x^{(m)}}^2 \sigma_{y^{(m)}}^2 p^2} \right) \tag{5}$$

$$\lambda_2^{(m)} = \frac{1}{2} \left( \sigma_{x^{(m)}}^2 + \sigma_{y^{(m)}}^2 - \sqrt{(\sigma_{x^{(m)}}^2 - \sigma_{y^{(m)}}^2) + 4\sigma_{x^{(m)}}^2 \sigma_{y^{(m)}}^2 p^2} \right) \tag{6}$$

In the first case, when  $p = 0$ , then the eigenvalues particularize to  $\lambda_1^{(m)} = \sigma_{x^{(m)}}^2$  and  $\lambda_2^{(m)} = \sigma_{y^{(m)}}^2$ . The confidence ellipse is aligned parallel to the frame axis with a major axis length equals to  $2\sigma_{x^{(m)}}\sqrt{A}$  and a minor axis length equals to  $2\sigma_{y^{(m)}}\sqrt{A}$ .

In the second case, when  $p \neq 0$ , the confidence ellipse is not axis aligned. In the sequel we evaluate the angle between the ellipse axis and those of the coordinate frame. The corresponding eigenvectors are orthogonal when  $\sigma_{x^{(m)}} \neq \sigma_{y^{(m)}}$ . Then, the relation between the linear transformation  $\mathbf{V}^{(m)}$  and  $\mathbf{C}^{(m)}$  can be expressed as follows:

$$\mathbf{C}^{(m)} = \mathbf{V}^{(m)} \mathbf{D}^{(m)} \mathbf{V}^{(m)-1}, \tag{7}$$

where  $\mathbf{V}^{(m)}$  contains the eigenvectors of  $\mathbf{C}^{(m)}$  and  $\mathbf{D}^{(m)}$  is the diagonal matrix whose non-zero elements are the corresponding eigenvalues. In this particular case the ellipse under analysis may be written as:

$$\mathbf{U}^{(m)T} \mathbf{C}^{(m)-1} \mathbf{U}^{(m)} = A \tag{8}$$

Replacing Eq. 7 in Eq. 8:

$$\mathbf{U}^{(m)T} \mathbf{V}^{(m)} \mathbf{D}^{(m)-1} \mathbf{V}^{(m)-1} \mathbf{U}^{(m)} = A \tag{9}$$

Let  $\mathbf{Q}^{(m)} = \mathbf{V}^{(m)-1} \mathbf{U}^{(m)}$  and given that  $\mathbf{V}^{(m)}$  is an orthogonal matrix,  $\mathbf{V}^{(m)-1} = \mathbf{V}^{(m)T}$ . Then, Eq. 9 can be expressed as follows:

$$\mathbf{Q}^{(m)T} \mathbf{D}^{(m)-1} \mathbf{Q}^{(m)} = A \tag{10}$$

The confidence ellipse is aligned to the new coordinate system  $\mathbf{Q}^{(m)}$  with a major axis length equals to  $2\sqrt{A\lambda_1^{(m)}}$  and a minor axis length equals to  $2\sqrt{A\lambda_2^{(m)}}$ . Finally, the rotation angel  $\theta$  is computed to obtain the orientation of the confidence ellipse:

$$\theta^{(m)} = \frac{1}{2} \tan^{-1} \left( \frac{2p\sigma_{x^{(m)}}\sigma_{y^{(m)}}}{\sigma_{x^{(m)}}^2 - \sigma_{y^{(m)}}^2} \right), \quad -\frac{\pi}{4} \leq \theta^{(m)} \leq \frac{\pi}{4}, \sigma_{x^{(m)}} \neq \sigma_{y^{(m)}} \tag{11}$$

Figure 5d shows the 95% confidence ellipse of each hotspot in the office. The confidence ellipses are used to represent the hotspots. In the following section, we will use the confidence ellipse to find the start and the end of tracks. This forms a simple feature vector from which it will be

used to build models to identify the persons' statuses in the office.

## 5 Person status identification

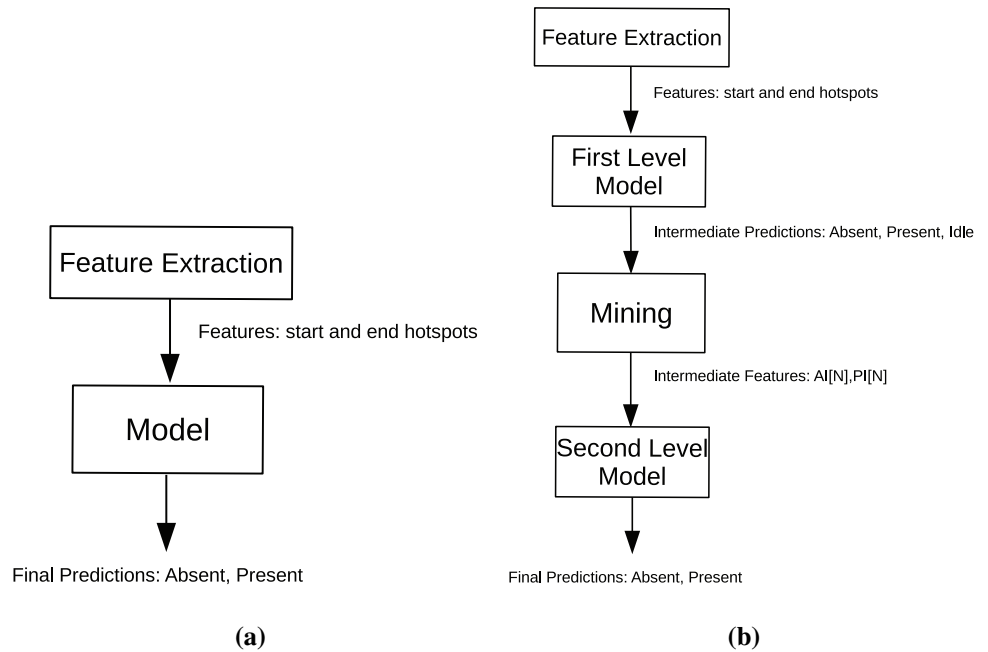
In order to determine the people's presence inside the office, we propose two approaches: (1) a single model approach and (2) a two-model mining approach. In the first approach, we simply train a model using the start and end hotspots as a feature vector to predict the person's presence. For this purpose, we compare and evaluate three probabilistic graphical models: Naïve Bayes (NB), Hidden Markov Model (HMM) and Linear Chain-Conditional Random Field (LC-CRF), where the role of each model is to predict the person's status sequence (*Absent* or *Present*). This approach did not yield a good representation of the people's status due to tracking loss and the inability of the tracker to track multiple persons accurately in certain situations such as group lunch. Figure 6a shows the use of the single model approach.

In the second approach, we introduce the second approach to solve these problems as shown in Fig. 6b. We use a first model level, where we increase the number of variables from two to three by including an additional *Idle* variable. The model is trained to predict the person's status sequence (*Absent*, *Present* and *Idle*) Then, a mining step is performed to extract two sequences:  $\mathbf{AI}[N]$  and  $\mathbf{PI}[N]$ , where  $N$  is the sequence length. Finally, a second level model is trained to predict the final person's status (*Absent* and *Present*) based on the sequence length of  $\mathbf{AI}[N]$  and  $\mathbf{PI}[N]$ . Similar to the single model approach, we compare and evaluate three PGMs, where the same model is used in the first and the second levels.

### 5.1 Feature extraction

The extraction of the start and the end hotspots of tracks is common between the single model approach and the two-model mining approach. Detecting the start and the end of tracks play an important role to identify the status of the persons in the office. Each person has an estimated confidence ellipse which defines the person's hotspot. A track starts from the door and ends to one of the person's hotspots shows a person's presence. Similarly, a track starts from one of the person's hotspots and ends to the door shows a person's absence. We propose to use the start and the end hotspots to form a feature vector from which we will estimate the persons' statuses. Let  $x'_i$  and  $y'_i$  be the positions associated with a given track  $T$ , where  $i = 1, \dots, I$ . Let  $g_x^{(m)}$  and  $g_y^{(m)}$  be the hotspots centres. Let  $a = (x'_i - g_x^{(m)}) \cos(\theta^{(m)}) + (y'_i - g_y^{(m)}) \sin(\theta^{(m)})$  and

**Fig. 6** The block diagram of the two approaches for people’s status sequence prediction: **a** the single model approach; **b** the two-model mining approach



$b = (x'_i - g_x^{(m)}) \sin(\theta^{(m)}) - (y'_i - g_y^{(m)}) \cos(\theta^{(m)})$ . The start hotspot  $S$  of track  $T$  can be found as follows:

$$S = m, \quad \frac{a^2}{A\lambda_1^{(m)}} + \frac{b^2}{A\lambda_2^{(m)}} \leq 1, \quad i \leq F, \quad i = 1 \dots I, \quad m = 1 \dots L, \tag{12}$$

where the positions should be inside the hotspot and only the first  $F$  positions are evaluated to find the start hotspot. Similarly, the end hotspot  $E$  of track  $T$  can be found as follows:

$$E = m, \quad \frac{a^2}{A\lambda_1^{(m)}} + \frac{b^2}{A\lambda_2^{(m)}} \leq 1, \quad i \geq I - F, \quad i = 1 \dots I, \quad m = 1 \dots L, \tag{13}$$

where only the last  $I - F$  positions are evaluated to find the end hotspot. Finally, let  $\mathbf{x}_t = (S, E)$  forms a feature vector to represent the start and the end hotspots at time instant  $t$ . Our objective is to recognize the presence or absence of persons from their tracks in the office. We typically have a sequence of observations  $\mathbf{x}_{1:T} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$  and we wish to infer the matching sequence of states  $\mathbf{y}_{1:T} = (y_1, y_2, \dots, y_T)$ . In order to work with different models, we divide our time series data in time slices of constant length. We denote the duration of a time slice with  $\Delta t$ ; we will state the chosen value for  $\Delta t$  in the experiments section. We will denote the start and the end hotspots for time  $t$  as  $\mathbf{x}_t^i$ , indicating whether person  $i$  initiated a track with a start hotspot  $S_t^i$  and an end hotspot  $E_t^i$  at least once between  $t$  and  $t + \Delta t$ , with  $\mathbf{x}_t^i = (S_t^i, E_t^i)$ . The person’s status at time slice  $t$  is denoted with  $y_t^i$ . In an office with  $\hat{N}$

persons, our task is to find a mapping between a sequence of observations  $\mathbf{x}^i = (\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_T^i)$  and a sequence of states  $\mathbf{y}^i = (y_1^i, y_2^i, \dots, y_T^i)$  for a total  $T$  time steps, where  $i = 1, \dots, \hat{N}$  and  $y_t^i$  can assume one of  $Q$  possible states  $1 \dots, Q$ .

### 5.2 Models description

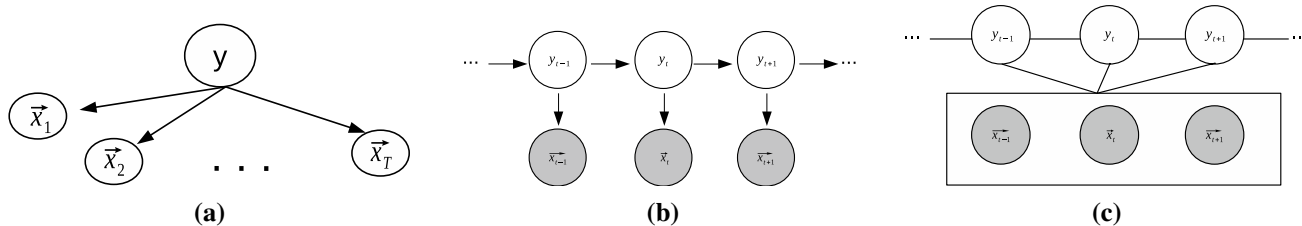
#### 5.2.1 Naïve Bayes model

This model utilizes the assumption that data attributes are conditionally independent given the class value (person’s status label). Let  $y$  denotes the class label. Our Naïve Bayes model (Rish 2001) assumes that the observation variable  $\mathbf{x}_t$  is only dependent on  $y$  as depicted in Fig. 7a. The likelihood can thus be computed as the product of the probability estimates for each particular observation value given the class label:

$$p(\mathbf{x}_{1:T}, y) = p(y) \prod_{t=1}^T p(\mathbf{x}_t | y) \tag{14}$$

#### 5.2.2 Hidden Markov model

An HMM is a generative model consisting of a hidden variable  $y_t$  and an observable variable  $\mathbf{x}_t$ . In this paper, the HMM is used as a supervised learning method to classify the people’s status sequence  $y_t$  from the feature vector  $\mathbf{x}_t$ . These variables change with time  $t$ . Our HMM assumes



**Fig. 7** The graphical representation of **a** Naïve Bayes where  $y$  denotes the class label,  $x_t$  denotes the feature vector of the start and end hotspots; **b** HMM and **c** LC-CRF. The *dark nodes* represent observable variables, whereas the *white nodes* represent hidden variables

that only two dependencies exist, represented by directed arrows in Fig. 7b. First, the hidden variable  $y_t$  at time  $t$  statistically depends only on the previous hidden variable  $y_{t-1}$  (first order Markov assumption). Second, the observable variable  $x_t$  at time  $t$  depends only on the hidden variable  $y_t$  at the same time instant. We can, therefore, specify the HMM using three probability distributions:

- The probability of the initial states,  $p(y_1)$  representing the probability that a person’s status  $y$  occurs at the beginning of the state sequence.
- The probability of the state transition,  $p(y_t | y_{t-1})$  representing the probability of switching from one state  $y_{t-1} = i$  (e.g. present) at time  $t - 1$  to another state  $y_t = j$  (e.g. absent) at the next time step,  $t$ . This represents the probability of transitions between person’s statuses in the office.
- The probability of the observation,  $p(x_t | y_t)$ , indicating the probability that state  $y_t$  (e.g. present) would generate observation  $x_t$ . This represents the probability of a particular person’s status generating a specific associated start and end hotspots.

Learning the parameters of these distributions corresponds to maximizing the joint probability of a sequence of states  $y$  and corresponding observations  $x$ . The joint probability of all observations and hidden states is:

$$p(\mathbf{y}, \mathbf{x}) = \prod_{t=1}^T p(\mathbf{x}_t | y_t) p(y_t | y_{t-1}). \tag{15}$$

The inference problem consists of finding the single best state sequence (path) that maximizes  $p(\mathbf{y}, \mathbf{x})$ . Although the number of possible paths grows exponentially with the length of the sequence, the best state sequence can be found efficiently using the Viterbi algorithm (Rabiner 1989). Using dynamic programming, we can discard a number of paths at each time step. This results in a computational complexity of  $O(TQ^2)$  for the entire sequence. Our HMM is fully-connected, where all the transitions are allowed. Finally, the HMM model is trained based on the

Baum-Welch parameter estimation algorithm (Baggenstoss 2001).

### 5.2.3 Linear chain-conditional random field

A LC-CRF (Lafferty et al. 2001) is a discriminative model that is used for segmenting and labeling sequence data. This model examines the “context” of the neighboring samples while classifying a sample. The LC-CRF still consists of a hidden variable (the person’s status)  $y_t$  and an observable variable (start and end hotspots)  $x_t$  at each time step  $t$  as shown in Fig. 7c. In contrast to the HMM model illustrated in Fig. 7b, the arrows on the edges have disappeared in the LC-CRF, making this an undirected model. This denotes that two connected nodes no longer represent a conditional distribution, as an alternative we refer to the potential between two connected nodes. Unlike probability functions, potentials (also referred as feature functions) are not limited to a value between 0 and 1.

The potential functions that specify the LC-CRF are  $\gamma(y_t, y_{t-1})$  and  $\delta(y_t, \mathbf{x}_t)$ . The  $\gamma$  function captures the relationship between the person’s status at the current time step and the person’s status at the preceding time step, while the  $\delta$  function captures the relationship between the person’s status and the observed variables at the current time step. Let  $f(y_t, y_{t-1}, \mathbf{x}_t)$  represents both  $\gamma(y_t, y_{t-1})$  and  $\delta(y_t, \mathbf{x}_t)$ . The first potential function is defined as follows:  $\gamma(y_t = i, y_{t-1} = j) = \epsilon_{ij} f_{ijl}(y_t, y_{t-1}, \mathbf{x}_t)$  in which the  $\epsilon_{ijl}$  is the actual potential and  $f_{ijl}(y_t, y_{t-1}, \mathbf{x}_t)$  is a feature function that in the simplest case returns 1 when  $y_t = i$  and  $y_{t-1} = j$  and 0 otherwise. Similarly, the second potential function is defined:  $\gamma(y_t = i, \mathbf{x}_t = \mathbf{x}_t) = \epsilon_{ijl} f_{ijl}(y_t, y_{t-1}, \mathbf{x}_t)$ , where  $\epsilon_{ijl}$  is the feature potential and the feature function now returns 1 when  $y_t = i$  and  $\mathbf{x}_t = \mathbf{x}_t$  and 0 otherwise. In order to easily represent the summation over all the different potential functions (Sutton and McCallum 2012), the index  $ijl$  is typically replaced by a one-dimensional index.

In LC-CRF, we learn the parameters by maximizing the conditional probability  $p(\mathbf{y}|\mathbf{x})$  which belongs to the family of exponential distributions as (Sutton and McCallum 2012):



$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z_x} \exp \left\{ \sum_{l=1}^L \epsilon_l f_l(y_t, y_{t-1}, \mathbf{x}_t) \right\} \quad (16)$$

where  $Z_x$  is an instance-specific normalization function, which guarantees the outcome as a probability:

$$Z_x = \sum_{\mathbf{y}} \exp \left\{ \sum_{l=1}^L \epsilon_l f_l(y_t, y_{t-1}, \mathbf{x}_t) \right\} \quad (17)$$

The feature function  $f_l(y_t, y_{t-1}, \mathbf{x}_t)$  will return a 0 or 1 depending on the values of the input variables and therefore defines whether a potential should be included in the computation. Since LC-CRF is a discriminative model, we can only use LC-CRF to perform inference (and not to generate data as in HMM). While learning the parameters of the model, we avoid modeling the distribution of the observations  $p(x)$ . Finally, an iterative gradient algorithm can learn the model parameters,  $\epsilon_l$ . Some particularly successful methods include quasi-Newton methods such as BFGS (Liu and Nocedal 1989), because they take into account the curvature of the likelihood function. The Viterbi algorithm (Rabiner 1989) can be used to generate person's status labels that correspond to an input sequence of observed start and end hotspots given a learned LC-CRF model.

There are modeling similarities between LC-CRF and HMM, note that the HMM's transition probability  $p(y_t|y_{t-1})$  and emission probability  $p(\mathbf{x}_t|y_t)$  have been replaced by the potentials  $\gamma$  and  $\delta$ , respectively. The essential difference lies in the way the model parameters are learned. Given a sequence of observations  $\mathbf{x}$  and corresponding sequence states  $\mathbf{y}$ , the HMM learns the parameters by maximizing the joint probability distribution  $p(\mathbf{x}, \mathbf{y})$ . By contrast, the LC-CRF learns the parameters by maximizing the conditional probability distribution  $p(\mathbf{y}|\mathbf{x})$ .

### 5.3 Single model approach

In this approach, a single model is built using one of the three PGMs, where the start and end hotspots are used as a feature vector to train the model to predict the person's status. The person's statuses are: *Present* (**P**) or *Absent* (**A**). For each time slice  $t$ , an observation sequence  $\mathbf{x}_t^i$  is generated for person  $i$ . When person  $i$  does not produce an observation sequence for the next time slice  $t + 1$ . Then, the last observation sequence from the previous time slice  $t$  is used for the next time slices, until person  $i$  generates a new observation sequence. The single model approach did not produce an accurate representation of the person's status, when each PGM is used for status prediction. Figure 8 shows a comparison between the output of the single model approach and the ground truth. The ground truth in Fig. 8a

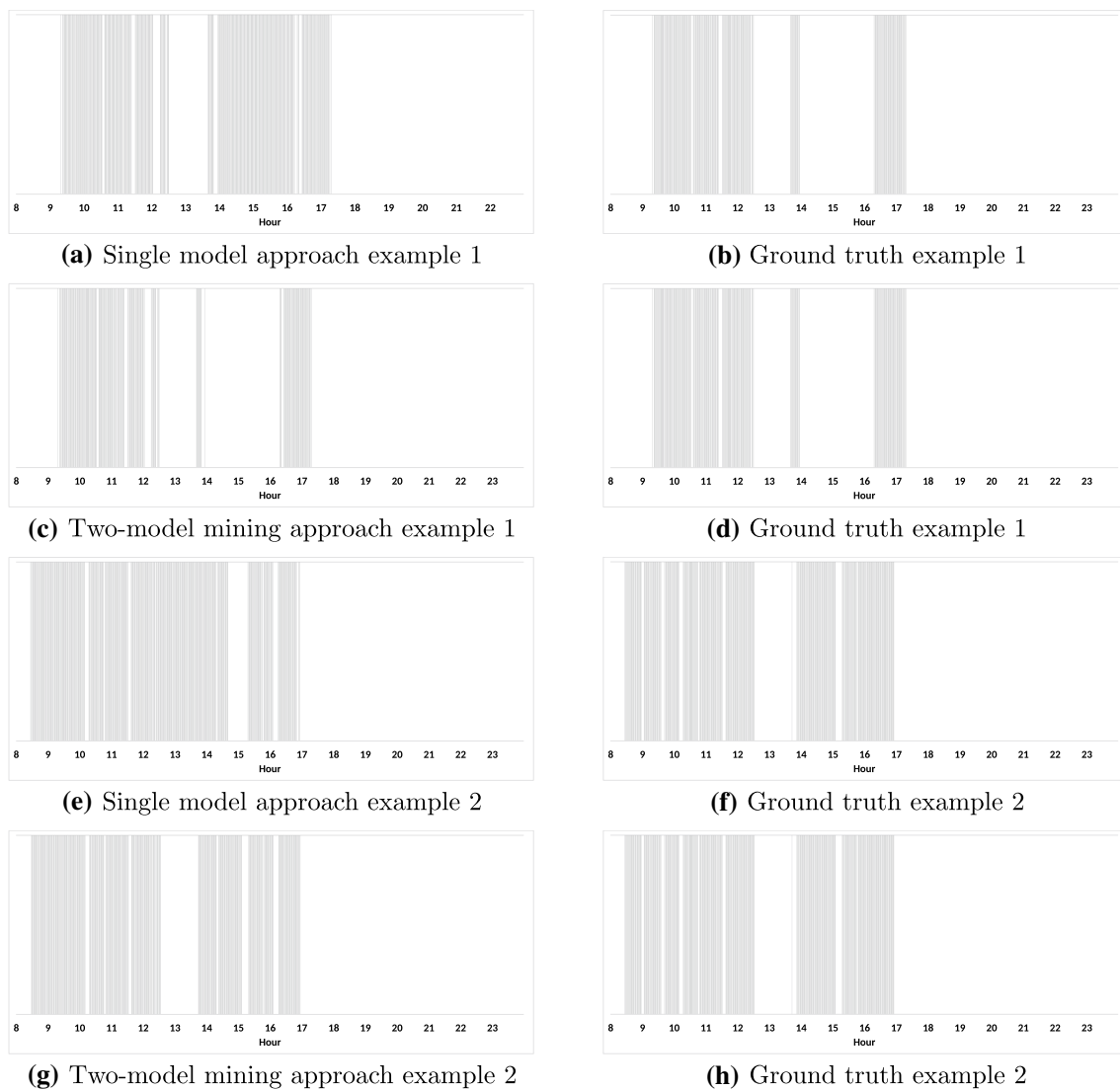
shows that a person has left the office for more than 2 h from 14:00 to 16:30. While, the single model approach output in Fig. 8b shows the person still in the office in the same period. Similarly, Fig. 8e shows a person has left between 12:30–14:00, while Fig. 8f shows the person still in the office. This inaccuracy happens because the RML tracker sometimes fails to produce accurate tracks for the person who leaves his desk location towards the door entrance. So, the status of the person remains present although he is absent. In the results section, the reported accuracy of each model is presented against the ground truth.

### 5.4 Two-model mining approach

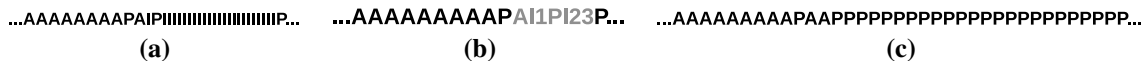
To overcome the inaccuracy of the single model approach, an obvious initial approach to discovering person's status patterns is to mine sample sequence states data from models for common, or frequent, recurring sequence patterns. Sequential pattern mining is commonly used to identify common progressions of purchasing patterns and searches for recurring patterns. One criterion in sequence mining is frequency, or the number of times the sequence pattern appears in the sample data.

In the single model approach, there were two states, namely, *Present* (**P**) or *Absent* (**A**). But in this approach, we increase the number of states from two to three by introducing a new state *Idle* (**I**). The model generates the *Idle* state, when there is no observation sequence  $\mathbf{x}_t^i$  at time slice  $t$  produced by person  $i$ . This forms the first model level. Figure 9a shows the state sequence output of the first model level. Then, a sequence mining algorithm performs a search through the space of candidate sequences to identify interesting patterns. A pattern here consists of a sequence definition and all of its occurrences in the data. Each candidate sequence pattern is evaluated according to a predefined criterion. We apply regular expressions as a sequence mining technique.

Regular expressions are simple, natural syntax for the succinct specification of families of sequential patterns. It includes a wide interesting pattern constraints. The sequence in Fig. 9a has two types of repeated sequence patterns: **AI** pattern and **PI** pattern. We use the following regular expressions: "**P(I+)**" and "**A(I+)**" to find these two sequence patterns. The quantifier character "+" matches the preceding element one or more times, while the parentheses define a marked subexpression. After applying the regular expression patterns in each iteration, the input sequence length is decreased to be in the following reduced form **PI[N]** or **AI[N]**, where  $N$  is the pattern length. Figure 9b shows the sequence mining output. We are interested to know if **AI[N]** and **PI[M]** sequence patterns are **P** or **A** patterns. We use k-means clustering algorithm to cluster **PI** and **AI** sequence patterns based on the pattern



**Fig. 8** A comparison between the output of the single model approach and the two-model mining approach against ground truth: **a** ground truth example 1; **b** single model approach example 1; **c** two-model mining approach example 1; **d** ground truth example 1; **e** ground truth example 2; **f** single model approach example 2; **g** two-model mining approach example 2; **h** ground truth example 2



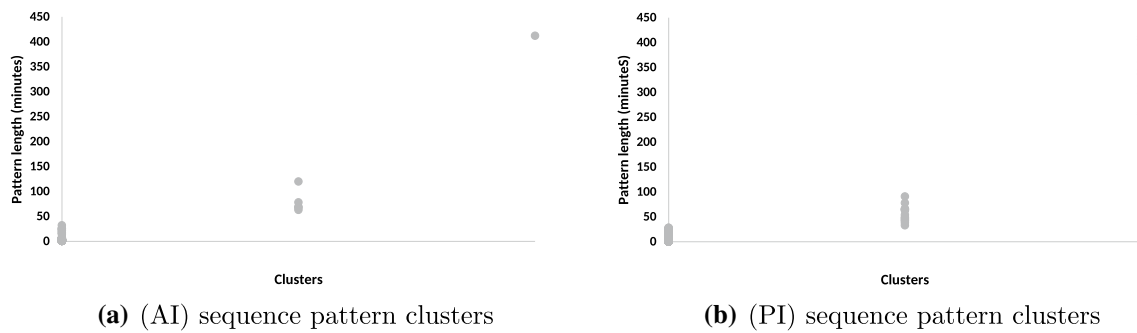
**Fig. 9** The output sequence of each model from the two-model mining approach: **a** first model level output, there are two interesting sequence patterns: **PI** and **AI**; **b** sequence mining output, the interest-

ing sequence patterns **PI**[*N*] and **AI**[*N*] are highlighted in *gray color*; **c** second model level output

length *N*. Figure 10a shows the **PI** patterns clustered into three groups based on the length of the pattern. The first cluster contains short length **PI** sequence patterns which are possible indications of **P** pattern. While, the other two clusters contain medium and long length **PI** sequence patterns which are possible indications of **A** pattern. Similarly, the **AI** patterns are clustered into three groups as shown in

Fig. 10b. The **AI** sequence patterns are assumed to be only indications of **A** pattern, regardless of the pattern length.

The objective of the second model level is to map the output sequence from the regular expression to the corresponding **P** and **A** state sequence. In the second model level, the **AI**[*N*] and **PI**[*N*] act as observation variables and the hidden variables are **P** and **A**. Figure 9c shows the output of the second model level after processing the sequence



**Fig. 10** Clustering the sequence mining output using k-means into short, medium and long clusters based on the pattern: **a** AI sequence pattern clusters; **b** PI sequence pattern clusters

mining output. In the single model approach, there is inconsistency between the estimated results and ground truth in some periods as shown in Fig. 8. This inconsistency does not exist between the two-model mining approach output and the ground truth as shown in Fig. 8d, g. In Fig. 8d, the person has left the office from 14:00 to 16:30. This is the same as the ground truth result in Fig. 8c. Similarly, the estimated result and the ground truth agree that the person has left from 12:30 to 14:00 as shown in Fig. 8g, h. More analysis and comparisons are shown in the results section.

The two-model mining approach is better than the single model approach due to the new introduced *Idle* (I) state in the first model level and the use of mining step. In the single model approach, when a person does not produce an observation for a given time slice  $t$ , then the last observation from the previous time slice  $t - 1$  is used until the person produces a new observation. If the used observation is false due to tracking loss or a group activity. Then, this false observation will propagate in the next time slices, leading to false states. This problem is addressed by making the first model level to generate an *Idle* state, when there is no observation produced by the person. Then, the regular expression technique looks for short, medium and long length patterns to provide a meaningful observation sequence to the second model level. Based on the pattern length and the pattern sequence, the final status of the person is determined by the second model level.

## 6 Activity patterns discovery

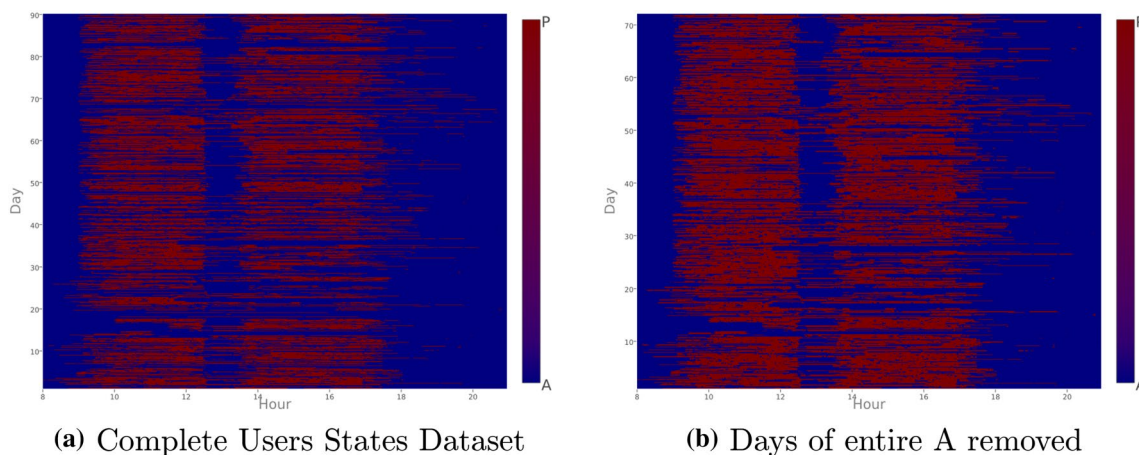
A semantic label is assigned to the user's status of *Present* (P) or *Absent* (A) provided by the previous component. At this point, we can represent a day in the life of an office worker in terms of user's status labels. For visualization and description purposes, the users' status patterns are visualized as a function of time of day, as in Fig. 11a, b. Each row in the figures is a day of a person's life in terms of his status, where the x-axis is the time of day and the

two colors represent the two user status labels. Figure 11a shows our entire dataset for the seven users and their 5 months of activities, many of which contain absence the entire day. The input dataset used is shown in Fig. 11b after removing days containing entirely absence labels. Looking at Fig. 11b, there is immense quantity of data and complex mixture of activities. Moreover, it is not clear how to detect dominating group activities and how to characterize individuals in terms of the groups' activities. These are a few of the points we address by using topic models.

The user's status sequences are not suitable for topic models in their original time sequence form since words in the topic model should be interchangeable. Table 6 shows used terms and their definitions in the context of natural language processing and activity discovery problem. We construct a *bag of user's status sequences* which can be viewed as analogous to words for text mining. Overall, we make an analogy between the bag of user's status sequences (or words) for activity discovery and a bag of words for text documents, where a user's status sequence is analogous to a text word, a day in the life of a user is analogous to a document, and a user is analogous to the author of a document. Finally, we use the Latent Dirichlet Allocation (LDA) topic model to discover activities, in which the input is the bag of user's status sequences, and the output is a set of probability distributions over words and latent topics, capturing the dominating underlying activities in the dataset.

### 6.1 Building the corpus

In order to generate the artificial words to construct the bag of user's status sequences, we follow a similar approach as in (Farrahi and Gatica-Perez 2011; Castanedo et al. 2014). We chose to divide a day into 15-min time intervals, resulting in 52 time blocks per day. A 15-min slot is used to ensure no vocabulary size explosion, and to remove some of the potential noise due to minor time differences between daily activities. For example, if a user arrives to the office at 09:04 am as opposed to 09:10 am, we want



**Fig. 11** Visualizations of the users states data for **a** all the users and the entire set of days and **b** all the users and days excluding days which contain entirely absence data. The *x-axis* corresponds to the time of day (in hours). The *y-axis* corresponds to days

**Table 6** Definitions of the natural language processing terms used in the context of activity discovery problem

Term	Definition
Word	Person’s status
Bag of words	Person’s status sequence
Topic	Activity pattern (e.g. lunch)
Document	Day with status sequences (e.g. 21-09-2015)
Author	A corresponding person in the office
Corpus	Dataset of daily status sequences of each person

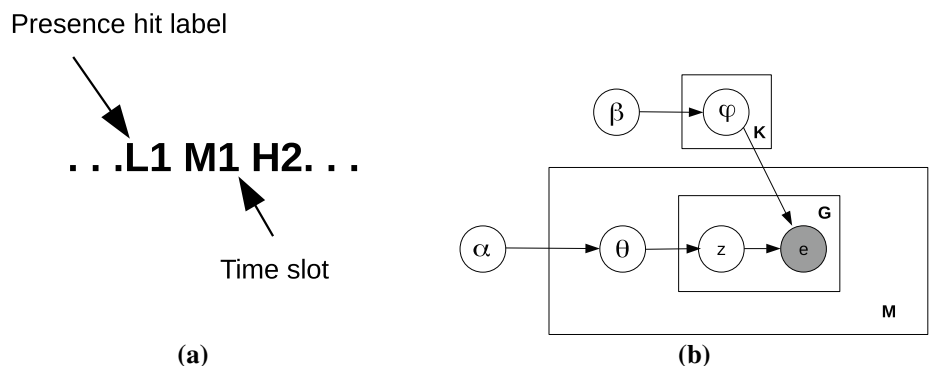
to capture the important feature of “arriving to the office early in the morning” and not the minor time difference of this activity between days. The choice of the timeslots is also guided by common sense about daily activities (e.g. typical lunch times, meeting times, leaving times). For each block of time, we compute the number of times the user’s status is present. Then, we map the presence hit to one of three discrete labels: *Low (L)*, *Medium (M)* and *High (H)* presence. We divide a day into the timeslots as follows:

(1) from 08:00 to 10:00, (2) from 10:00 to 12:00, (3) from 12:00 to 14:00, (4) from 14:00 to 16:00, (5) from 16:00 to 18:00 and (6) from 18:00 to 21:00. Finally, the last step in building the bag of user’s status sequences is the word construction. Each word will contain a presence hit label, followed by one of the 6 timeslots in which it occurred. Figure 12a shows an example of a user’s status sequence.

### 6.2 Latent Dirichlet allocation

Latent Dirichlet allocation (LDA) is a probabilistic generative model, introduced by Blei et al. (2003), in which every document is modeled as a multinomial distribution of topics and every topic is modeled as a multinomial distribution of words. LDA can be extended to include other collections of discrete data. LDA allows to infer the inherent activity patterns from our dataset. For a particular day *d*, it picks a set of activity patterns with different emphasis. Thus, we model the mixture of activity patterns as multinomial probability distribution  $p(z|d)$  over activity pattern *z*. Similarly, the importance of each

**Fig. 12** The two steps required for activity pattern discovery; **a** an example of user’s status sequence construction to build the corpus; **b** graphical model of Latent Dirichlet allocation (LDA)



constructed word  $e$  for each activity pattern  $z$  is also modeled as a multinomial probability distribution  $p(e|z)$  over words  $e$  of a vocabulary. Given these two distributions, we can compute the probability of a constructed word  $e$  occurring in day  $d$ :

$$p(e|d) = \sum_{z=1}^K p(e|z)p(z|d), \quad (18)$$

assuming that there are  $K$  activities. Having many days in the corpus, we observe a data matrix of observed  $p(eld)$  as a result of a matrix product of the word relevance for each activity pattern  $p(e|z)$  and a mixture of activity patterns  $p(z|d)$  for each day. Thereby, recovering the characteristic words for each activity pattern and the mixture of activity patterns for each day. Using the LDA model, each day in the corpus is modeled as a finite mixture over an underlying set of  $K$  activity patterns. The activity pattern mixture is drawn from a Dirichlet prior to the entire corpus. In a corpus of  $M$  days, the generative process begins by specifying a distribution over activity patterns  $\mathbf{z} = (z_{1:K})$  for a given day  $d$ , where  $K$  is the number of activity patterns. Given a distribution of activity patterns for a day, words are generated by sampling activity patterns from this distribution. The result is a vector of  $G$  constructed words  $\mathbf{e} = (e_{1:G})$ . LDA places a Dirichlet prior distribution on the activity pattern mixture parameters  $\theta$  and  $\Phi$ , to provide a complete generative model for days.  $\theta$  is an  $M \times K$  matrix of day-specific mixture weights for the  $K$  activity patterns, each drawn from a Dirichlet prior, with hyperparameter  $\alpha$ .  $\Phi$  is a  $V \times K$  matrix of word-specific mixture weights over  $V$  vocabulary items for the  $K$  activity patterns, each drawn from a Dirichlet prior, with hyperparameter  $\beta$ .

A graphical representation of the LDA topic model is shown in Fig. 12b. The inner plate over  $z$  and  $e$  shows the repeated sampling of activity patterns  $\mathbf{z}$  as a distribution over  $G$  words  $\mathbf{e}$ . The plate surrounding  $\theta$  shows the sampling of a distribution over activity patterns for a total of  $M$  days in the corpus. The plate surrounding  $\Phi$  shows the repeated sampling of word distributions for each activity pattern until  $K$  activity patterns have been generated. Words are further dependent on a Dirichlet distribution ( $\beta$ ), from which they are drawn. While, the mixture weights  $\theta$  that describe each day as a distribution over activity patterns are again assumed to be Dirichlet distributed ( $\alpha$ ). The main objectives of LDA inference are to find the probability of a constructed word given each activity pattern  $k$ :  $p(e = t|z = k) = \phi_k^t$ , and to find the probability of an activity pattern given each day  $m$ :  $p(z = k|d = m) = \theta_m^k$ . Several approximation techniques have been developed for inference and learning in the LDA model (Blei et al. 2003; Griffiths and Steyvers

2004). In this work we adopt the Gibbs sampling approach (Griffiths and Steyvers 2004).

## 7 Results and discussion

### 7.1 Dataset

For validating the performance of our proposed approach, we collected 5 months of real-life recordings using a network of nine low-resolution visual sensors producing synchronized images of  $30 \times 30$  pixels at a frame rate of 50 fps. Each day of data corresponds to a 13 h period starting from 08:00 to 21:00. The recording period started in November 2014 and lasted till March 2015. The minimum number of running visual sensors is 4 and the maximum is 9 in our dataset. The dataset includes 90% of 9 running visual sensors (82 days) and 10% of 4–5 running visual sensors (9 days). The low number of running visual sensors is due to reaching the maximum storage capacity of the hard disk while recording. The resulting dataset is massive, amounting to 637 days, and over 8200 h of video recording data for seven persons.

The low-resolution visual sensor data is stored in a platform that is used by the consortium to store all the project work. This platform offers server service that stocks the data safely and controls the access to the data files; only registered and appointed users (username and password) have access to the data files. With this platform the different data captured from the various sensors can be stored in the same place and easily combined for further analysis.

We performed a visual inspection of the videos in order to collect ground truth about the persons' statuses. We selected three persons out of seven for the evaluation. For each person, 10% of the dataset, which corresponds to 12 days, was selected for the evaluation. In our experiments, we chose to have  $\Delta t = 60$  seconds. This time slice duration is long enough to be discriminative and short enough to provide high accuracy labeling result. Each minute in the ground truth is annotated with **A** and **P** tags, yielding to 780 labels per day. To compare the performance of the three PGMs in the single model approach and the two-model mining approach against ground truth, the original data is split into a test and training set, 2 days were used for training the models, and 10 days were used for testing the models in each approach.

### 7.2 Person status identification analysis

A first step at evaluating the performance of the two approaches against ground truth, we compute the accuracy. This measure can be calculated using the confusion



matrix shown in Table 7. The accuracy can be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

Table 8 shows the accuracy values for the three persons. In the single model approach, the NB and LC-CRF have similar and less accurate results (an average accuracy of 51.50%). By checking the outputs of the NB and LC-CRF against the ground truth, we found that when a person leaves the office, the NB and LC-CRF models generate **P** state, while the ground truth state is **A**. This inaccuracy can be attributed to several reasons: (1) the adapted random walk model in the RML tracker only imposes some weak constraints on the temporal continuity of the tracks. This property causes tracking loss. (2) In multiple-person activities such as group lunch, or meetings, the RML tracker can not track multiple persons who are leaving or entering the office accurately. This leads the tracker to generate false observation, and as a result, wrong state sequence. (3) The very low resolution of the cameras and the associated limitations in image processing and calibration. The HMM has a higher accuracy (an average accuracy of 86.90%).

In the two-model mining approach, the accuracy of the NB has an average increase of 17.82%, while the LC-CRF has an average accuracy increase of 21.23%. The high accuracy increase in the NB and LC-CRF models is due to the new introduced *Idle* (**I**) state in the first model and the use of regular expression sequence mining technique. Finally, the HMM has an average accuracy increase of 8.90%. The HMM produces the best accuracy for the three persons in both approaches, because the HMM is able to deal with temporal patterns.

We analyze the trade-off between true positive rate (TPR) and false positive rate (FPR) of both approaches in the form of Receiver Operating Characteristic curve

(ROC). The true positive and false positive rates can be calculated as follows:

$$TRP = \frac{TP}{TP + FN} \tag{20}$$

$$FPR = \frac{FP}{FP + FN} \tag{21}$$

The ROC curve is a two-dimensional graph with the false positive rate on the x-axis and the true positive rate on the y-axis. Figure 13 shows the ROC plots of the single model and the two-model mining approaches for three persons. For each person, the visualization of the single model approach ROC curve, entitled “X\_a”, while the two-model mining approach, entitled “X\_b”. A model is considered as superior to another if its point is closer to the (0,1) coordinate (the upper left corner) than the other. It clearly that “HMM\_a” and “HMM\_b” has better ROC curves than the others. While, “LCR\_b” scores the second best ROC curve. The rest of the ROC curves indicate poor performance of the models in both approaches.

To further analyze the two-model mining approach, we compute the time the person spent being present per hour. Then, we compute the mean absolute error (MAE):

$$MAE = \sum_{r=1}^N \frac{|v_r - v'_r|}{\hat{H}}, \tag{22}$$

where  $v_r$  is the estimated presence duration for hour  $r$ ,  $v'_r$  is the actual presence duration for hour  $r$ , and  $\hat{H}$  is the number of hours. The relative absolute error (RAE) is computed to measure the error percentage:

$$RAE = \frac{\sum_{r=1}^N \frac{|v_r - v'_r|}{v'_r}}{\hat{H}} \times 100 \tag{23}$$

Additionally, we measure the Spearman’s rank correlation coefficient ( $\rho$ ) to assess the relationship between the estimated presence duration and the ground truth. The MAE, the Spearman’s correlation coefficient and the RAE results of three PGMs are shown in Table 9 for three persons. Clearly, HMM outperforms the LC-CRF and NB in MAE, RAE and  $\rho$  measures. The LC-CRF has

**Table 7** Confusion matrix showing the true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) for each class

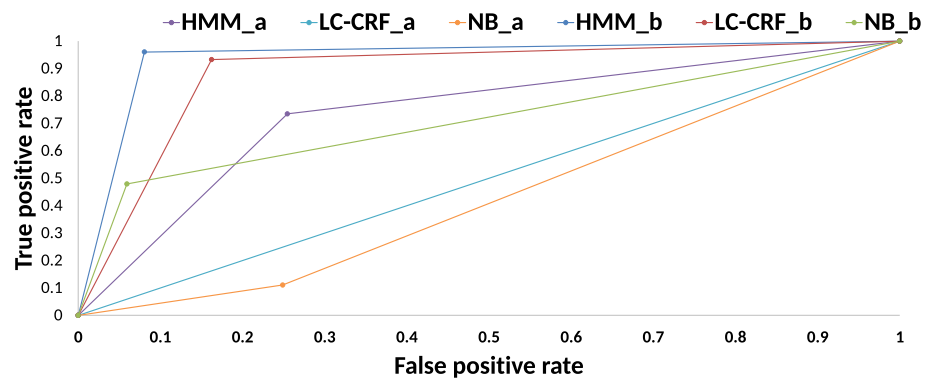
		Prediction	
		P	A
Truth	P	TP	FP
	A	FN	TN

**Table 8** Results for single model and two-model mining approaches

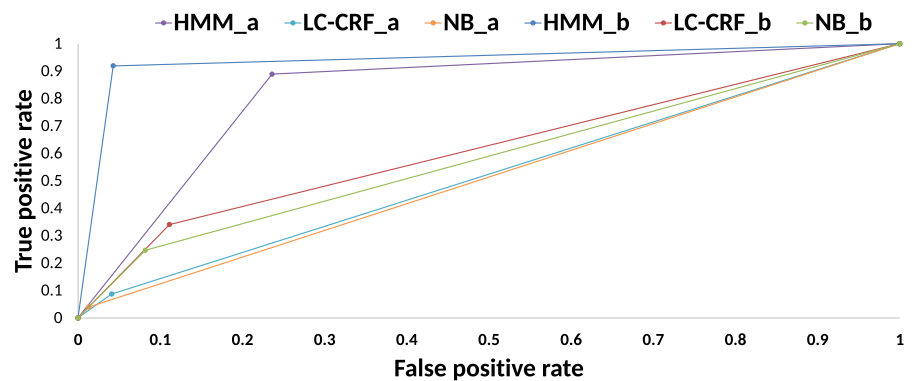
	Single model approach			Two-model mining approach		
	HMM (%)	LC-CRF (%)	NB (%)	HMM (%)	LC-CRF (%)	NB (%)
Person 1	90.20	53.28	44.31	95.26	86.91	71.72
Person 3	85.82	56.83	56.50	95.79	63.61	61.32
Person 5	84.68	48.69	49.44	96.37	71.99	70.68

The table shows the accuracy of NB, LC-CRF and HMM for Person 1, 3 and 5

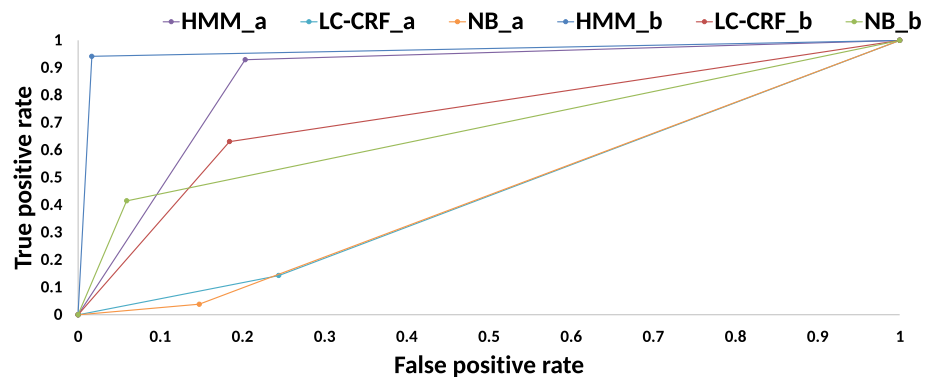
**Fig. 13** ROC curves for single model and two-model mining approaches: **a** Person 1; **b** Person 3; **c** Person 5. “X\_a” represents models in the single model approach, while “X\_b” represents models in the two-model mining approach



(a) Person 1



(b) Person 3



(c) Person 5

a slight performance increase than the NB, but not spectacularly so. As the HMM produces the best result for the three persons, we only consider the results produced by it in this analysis. Finally, we compare between the average presence duration per hour produced by HMM from the two-model mining approach against the average presence duration per hour produced by the ground truth as shown in Fig. 14. The vertical error bars show the overestimates and the underestimates of presence durations. About 30% overestimates for *Person 1* between 12:00 and 13:00.

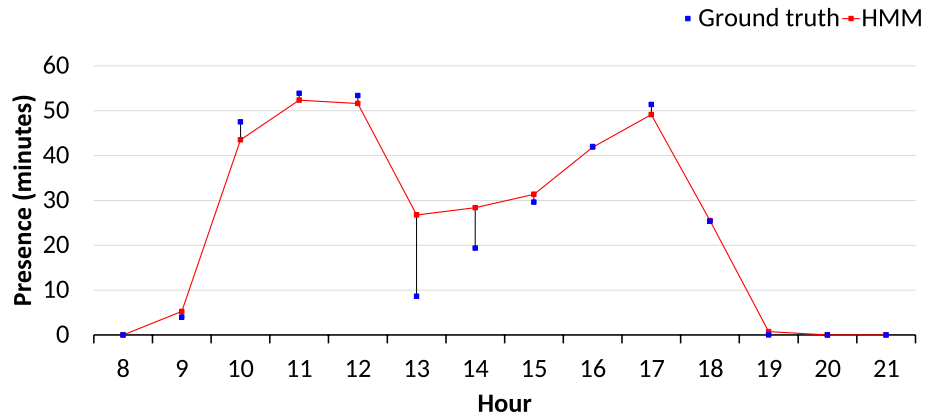
From the visual inspection, when *Person 1* goes to lunch between 12:00 and 13:00, our approach shows *Person 1* is present, although he is absent. This is attributed to the very close distance between *Person 1*'s desk location and the door entrance as shown in Fig. 2, so visitors who tend to stand next to the door entrance or close to *Person 1*'s desk location generate indications of presence for *Person 1*. In other circumstances, when *Person 1* leaves the office, the RML tracker fails to generate a trajectory from *Person 1*'s desk location to the door entrance due to

**Table 9** Results for two-model mining approach

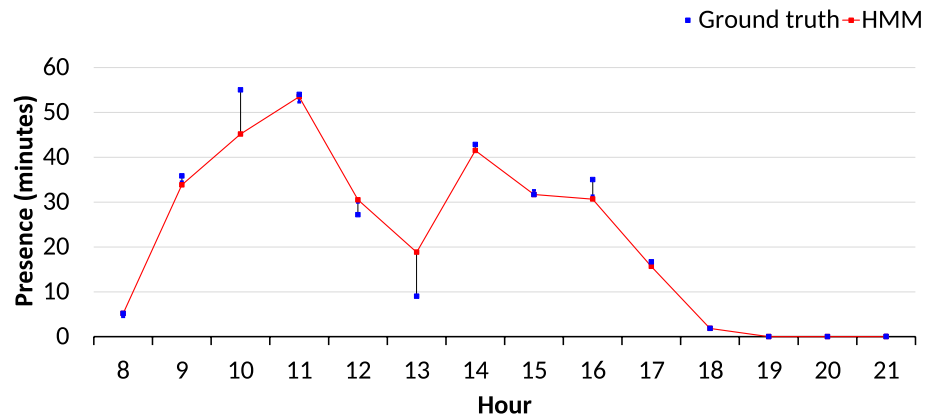
	HMM			LC-CRF			NB		
	MAE	RAE (%)	$\rho$	MAE	RAE(%)	$\rho$	MAE	RAE(%)	$\rho$
Person 1	3.14	29.26	0.94	5.58	39.37	0.90	11.63	33.25	0.87
Person 3	2.03	9.86	0.95	12.16	40.09	0.70	12.72	41.08	0.71
Person 5	1.39	7.10	0.79	12.65	27.5	0.53	13.19	27.16	0.54

This table shows the MAE, the Spearman’s rank correlation coefficient ( $\rho$ ), and the RAE for the time spent being present per hour for Person 1, 3 and 5

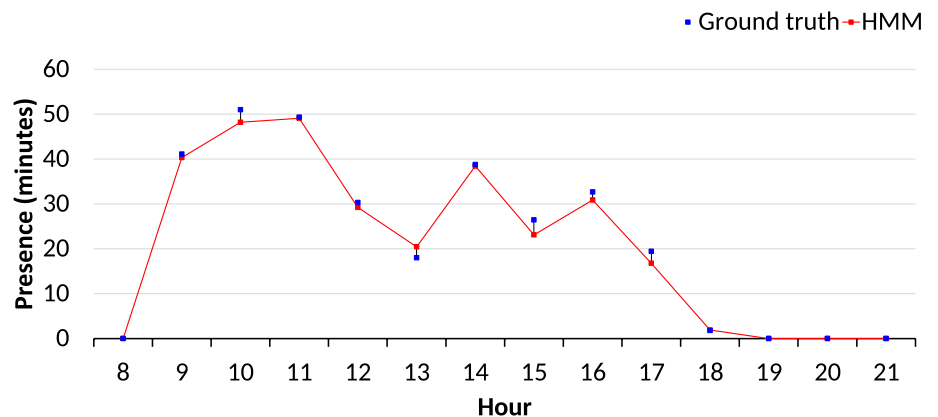
**Fig. 14** A comparison between presence duration estimates and ground truth: **a** Person 1; **b** person 3; **c** person 5



(a) Person 1



(b) Person 3



(c) Person 5

the very close distance. About 15% overestimates for *Person 1* and *Person 3* between 13:00 and 14:00. The visual inspection indicated that, some visitors tend to occupy their desks while they are absent. *Person 5* has overestimates and underestimates of less than 2%. Our approach of estimating the presence duration provides promising results close to the ground truth. The accuracy can be increased by using RFID or computer usage logs.

### 7.3 LDA model selection

LDA and other topic models are frequently evaluated in terms of their ability to generalize to unseen data. A common performance measure for this purpose is perplexity. In the context of topic modeling, perplexity measures how well the topic model learned from a training corpus generalizes to a set of unseen documents in a test corpus. The lower the perplexity of a model, the better its predictive power. Perplexity is defined as the reciprocal geometric mean of the likelihood of a test corpus given a model  $\xi$ :

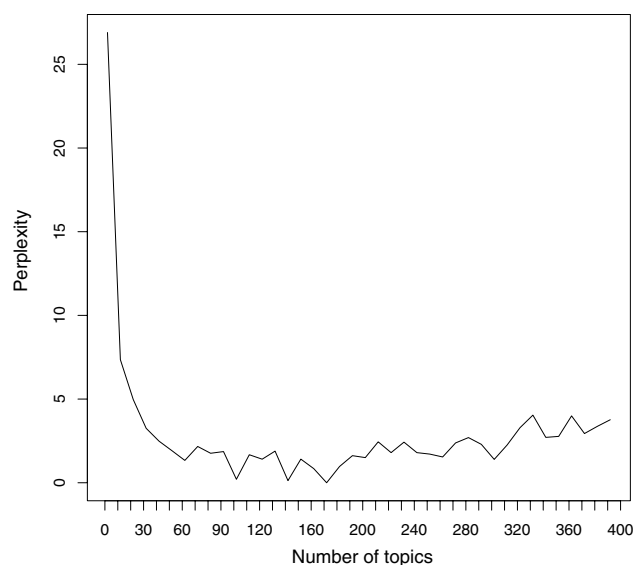
$$\text{Perplexity} = \exp \left[ - \frac{\sum_{m=1}^M \log p(e_m | \xi)}{\sum_{m=1}^M G_m} \right], \quad (24)$$

where  $G_m$  is the length of the document  $m$  and  $e_m$  is the set of unseen words in document  $m$ . We use perplexity as an indicator to choose the optimal number of latent topics,  $K$ . Establishing the number of topics (or activity patterns) that the model must learn is one important decision when training a topic model. In this work, we performed several analysis by increasing the number of topics and evaluated the obtained scores with the aim of choosing a good model. First, we randomly chose proportions of 90% training and 10% test documents. Then, we computed perplexity for LDA using  $K$  values from 2 to 400 with increments of 10. For all values of  $K$ , initialization was followed by 1000 iterations of the Gibbs sampling algorithm. We used  $\beta = 0.1$  and  $\alpha = 50/K$  as suggested in Griffiths and Steyvers (2004).

Figure 15 reports the perplexity results against the number of topics. A lower perplexity value indicates a better prediction over data. It can be shown that perplexity values decrease while we increase the number of topics till 80, after which the perplexity stabilizes. We choose  $K = 80$  as the number of latent topics for the remaining experiments.

### 7.4 Group activity discovery analysis

The LDA model successfully found topics over all persons and days, and contain the dominating



**Fig. 15** Perplexity plot as a function of the number of topics,  $K$ . At  $K = 80$ , the perplexity mostly stabilizes to a low value

activity patterns. The unsupervised clustering of presence/absence routines showed different types of activity patterns, allocating intervals of days which follow characteristic trends to different topics with a probability measure. To illustrate the discovered activity patterns, for each group of topics we rank the 5 most probable words, ranked by  $p(e|z)$ , and show them in tables. For group of topics, we also rank the most probable days, ranked by  $p(z|d)$ , and visualize them in plots. In Table 10, the topics 11, 16 and 52 capture “attend a meeting” activity pattern where the most probable word is L4, which indicates low presence in timeslot 4 (14:00–16:00). While, topic 39 captures “leaving the office late” activity pattern where the two most probable words are H6 and H5, which indicate a high presence in timeslots 5 (16:00–18:00) and 6 (18:00–21:00). Figure 16a, c visualize the days for topics 11, 16, 52 and 39, and can see that topics 11, 16 and 52 identify 65 days as “attend a meeting” activity pattern, wherein topic 39 identifies 20 days as “leaving the office late” activity pattern. Note that in all these topics, the top words account for over 90% of the probability mass, which suggests that the topics are discriminant of very characteristic patterns.

Other activity patterns discovered are visualized in Fig. 16 with their corresponding labels as the title:

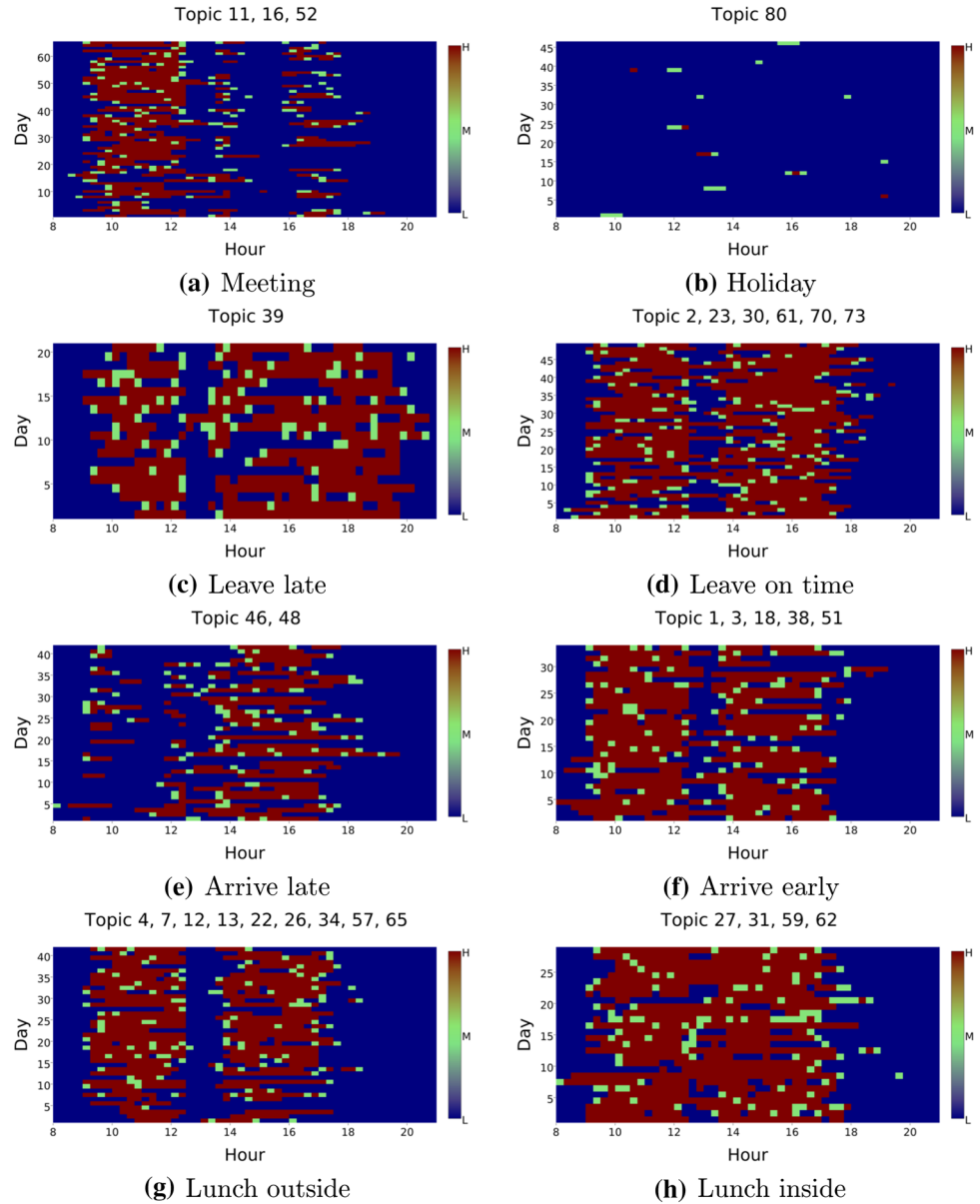
- Topic 80 captures holidays activity pattern. It is clear that all the timeslots have low presence.
- Topics 2, 23, 30, 61, 70 and 73 capture leave on time activity pattern which correspond to low presence in timeslots 5 (16:00–18:00) and 6 (18:00–21:00).

**Table 10** The table lists the five most probable user’s status sequence ranked by  $p(eltz)$  for topics 11, 16, 52 and 39

Topic 11		Topic 16		Topic 52		Topic 39	
Word	$p(eltz)$	Word	$p(eltz)$	Word	$p(eltz)$	Word	$p(eltz)$
L4	0.9709	L4	0.9847	L4	0.9856	H6	0.5337
M3	0.0251	M5	0.0048	H2	0.0045	H5	0.3240
H5	0.0002	H1	0.0025	H3	0.0024	M6	0.0958
H4	0.0002	L5	0.0025	L3	0.0024	M2	0.0162
H1	0.0002	L2	0.0025	M1	0.0024	M1	0.0135

The topics 11, 16 and 52 capture “attend a meeting” or “leaving the office early” activity patterns. While, topic 39 captures “leaving the office late” activity pattern

**Fig. 16** The discovery activity patterns visualized for several days for group of topics. The corresponding activity pattern name is displayed below the discovered topics ranked by  $p(zld)$



- Topics 46 and 48 capture arrive late activity pattern which correspond to low presence in timeslots 1 (08:00–10:00) and 2 (10:00–12:00).
- Topics 1, 3, 18, 38 and 51 capture arrive early activity pattern. This is indicated by a high presence in timeslot 1 (08:00–10:00).



- Topics 4, 7, 12, 13, 22, 26, 34, 57 and 65 capture lunch break activity pattern outside the office where timeslot 3 (12:00–14:00) has low presence.
- Topics 27, 31, 59 and 62 capture lunch break activity pattern inside the office, with high presence in timeslot 3 (12:00–14:00).

On a weekly level, some trends characteristic of weekends appeared with the activity patterns discovered by LDA. Topics 29 and 44 captured the activity pattern of working on the weekends. The discovered topics show only 3 days which belong to *Person 3* and *Person 7*. The visual inspection of the weekends has confirmed the LDA results. Figure 17 shows the visualization of both topics and their corresponding heatmaps. The heatmaps show the hotspots of each person and the active paths between their desk locations and the door entrance. Both persons have tracks that lead to *Person 5*'s desk location, because there is a wall clothes hanger in this area. Some topics such as 80 and 33 demonstrate holidays and days off activity patterns as shown in Fig. 16b.

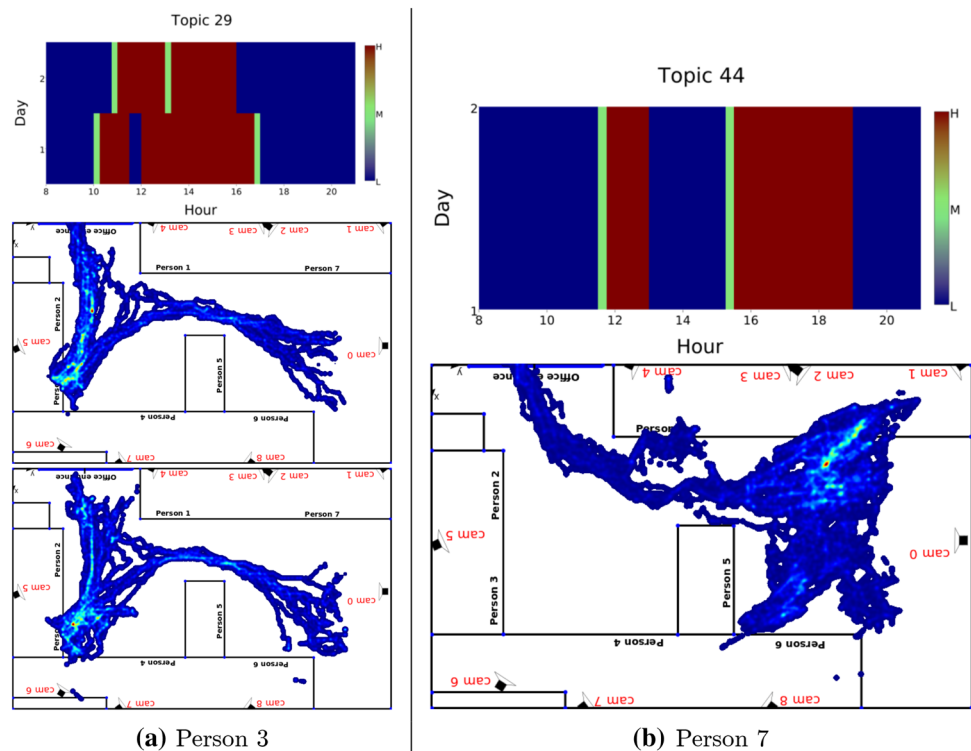
Finally, we are interested in finding how evident is the “mixture of topic” assumption in our data. Are days about one topic or several topics? Our LDA methodology also allows us to find days which vary over many topics, and days which are best represented by a few topics. In Fig. 18a, we show a histogram of the number of “dominating” topics per day. We compute the number of topics composing at

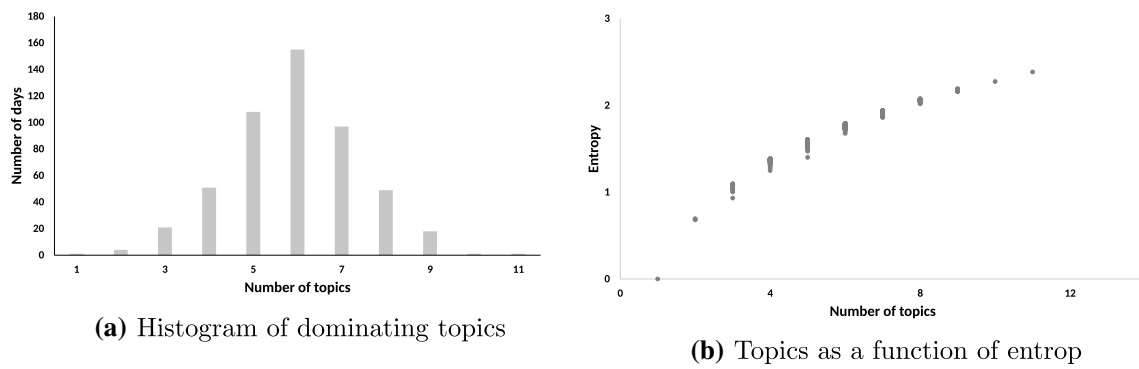
least 50% of the probability mass of each day in the study, and plot a histogram of the results. In general, all days are well described by fewer than 11 topics. Thus, at most 13.75% (11/80) of the topics can describe the probability mass of any day in the dataset. On the lower end of the histogram, very few days are described by less than three topics (21 days, or 3.29% of the days in the dataset). The same can be observed for high number of topics, very few days require 9 or more topics to be well defined (18 days, or 2.82% of the days in the dataset). The average number of topics in the study is 6 topics. In Fig. 18b we plot the entropy for each day, computed on the topic distribution, as a function of the number of dominating topics. Each data point represents a day. We can see that the number of topics as a function of entropy is about linear, proposing that number of dominating topics is indeed a good measure of day entropy and variation in daily activities.

### 7.5 Individual activity discovery analysis

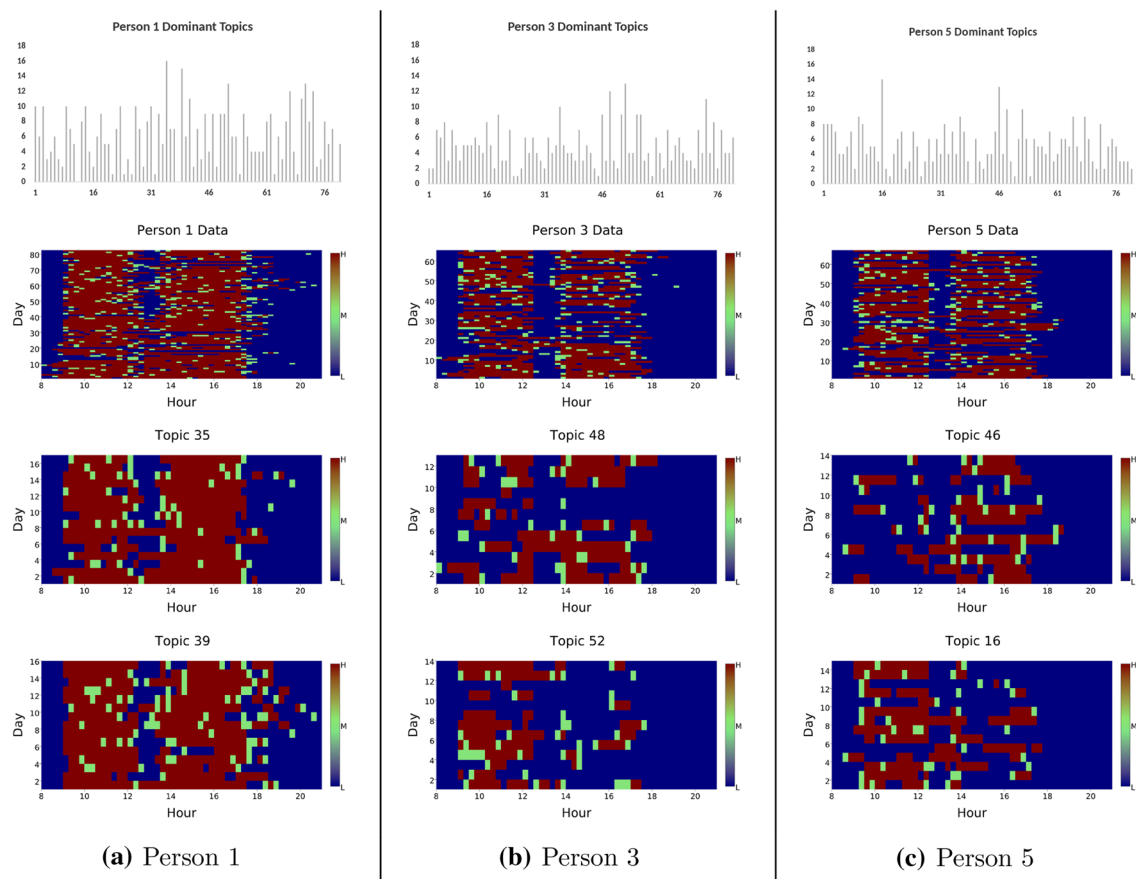
After having discovered the activity patterns of all persons in the office, we can also examine the topic distributions over individuals with LDA. For each individual  $i$ 's day  $d_i$ , we count the topics for which the ranked probability of the topics given the day,  $p(z|d_i)$  is greater than  $T$  (set here to 0.03), aggregate for all the individual's days and illustrate them in the histogram entitled “Person  $i$  Dominant Topics” in Fig. 19. Some persons' days

**Fig. 17** Topics discovered by LDA for weekend activity patterns. Topics 29 and 44 show two persons are working on the weekends: **a** Person 3 has worked 2 days on the weekends. The heatmaps show the hotspot and a high active path between the desk location and the door entrance; **b** Person 7 has worked only 1 day on the weekend. The heatmap displays the hotspot and a low active path between the desk location and the door entrance





**Fig. 18** LDA results. **a** Histogram of number of “dominating” topics per day for the LDA model. **b** Number of topics plot as a function of entropy for each day, showing an approximate linear relationship between the two measures



**Fig. 19** Individual person analysis. The histograms “Person x Dominant Topics” demonstrate dominant topics for persons x. Plots “Person x Data” corresponds to the raw input user’s status data of person x. The two topics below are the two dominating activity patterns for person x

are expressed well by a few topics, other persons have a rich set of varying activity which are expressed as a mixture of many topics. For example, noting the varying y-axis scales, *Person 1* has 15 topics, whereas *Person 3* and *Person 5* have 4 topics, respectively, in which 10 or more documents are assigned to each topic. It can be noted that, *Person 3* and *Person 5* have a very high

probability of a few topics for most days, while *Person 1* days are expressed as a mixture over many topics. We plot the persons’ status data in the plots entitled “Person x Data”. Each person has a different number of days (y-axis), since they have varying number of days after removing fully *absence* days. Beneath the persons’ days are the two topics which dominate the given persons’

daily activities. For instance, the two topics dominating *Person 1* daily activities are topics 35 and 39. *Person 1* dominating activities are “office work for the whole day with regular lunch breaks”, as well as “being at work late in the evening”. Looking at “Person 1 Data”, we can confirm that *Person 1* does work a lot, especially in the afternoon. *Person 1* daily activities are thus a mixture over several topics, as can be seen by the histogram “Person 1 Dominant Topics”. *Person 3* most common activities are “arriving to work before 11:00”, and “attending meetings in the afternoon”. Looking at *Person 3* status data, we can see this person arrives to work early in the morning, then goes to lunch, except for some days when he arrives late, after that he attends meetings or leaves the office early in the afternoon. *Person 5* mostly arrives to the office late in the morning, as seen by the dominant topic 46 dominating most of his daily activities. *Person 5* is mostly out in the afternoon attending meetings as captured by topic 16. Looking at *Person 3* and *Person 5* lunch breaks, this suggests that both persons go to lunch together. Finally, *Person 3* and *Person 5* dominant topics are less of a mixture over several topics than *Person 1*.

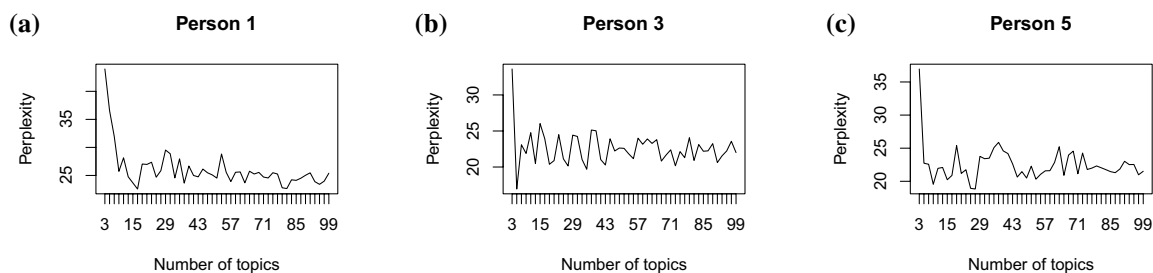
Most persons’ daily activities are described well by a few topics, others require more. We focus on analyzing and comparing the topics activation for 1 day of several individuals against ground truth. We use the days which only belong to each individual to build the LDA model. We computed perplexity for LDA using  $K$  values from 3 to 100 with increments of 2, because the dataset of each individual is small, amounting to 91 days. Figure 20 shows the perplexity results against the number of topics  $K$  for three persons, the lowest perplexity value varies slightly for each individual, we choose  $K$  to be 6. The perplexity does not stabilize due to the fact that each individual dataset can not converge (with the 1000 max iterations established) when so many topics are used. For each individual, the LDA estimation was performed on the whole dataset except 1 day, where the inference is done on the remaining day.

Figure 21a shows the topic activations on the day that was left out during training for *Person 5*, the topics were estimated from 90 days of data. For each topic  $z$  we list all

user’s status labels  $e$  with  $p(e|z) \geq 0.01$ . Figure 21b shows the ground truth activities. The first important observation which can be made from the results shown in Fig. 21 is that there are topics that clearly correlate with the daily activities of the person’s day. This can be seen by comparing the topic activations to the daily ground truth activities. Topics 1 and 2 are active during morning office work. The lunch activity is represented by topic 3. As the typical lunch activity is composed of a visit to the cafeteria or a visit to the restaurant. In the afternoon, topics 4 and 5 are active during afternoon office work so that their joint or individual activation is a good indication of office work. The remaining daily activity, commuting, is not directly correlated with a single topic but rather with a combination of topics. Both in the evening and in the morning, the co-activation of various topics including topics 1, 4 and 6 allow to identify this activity.

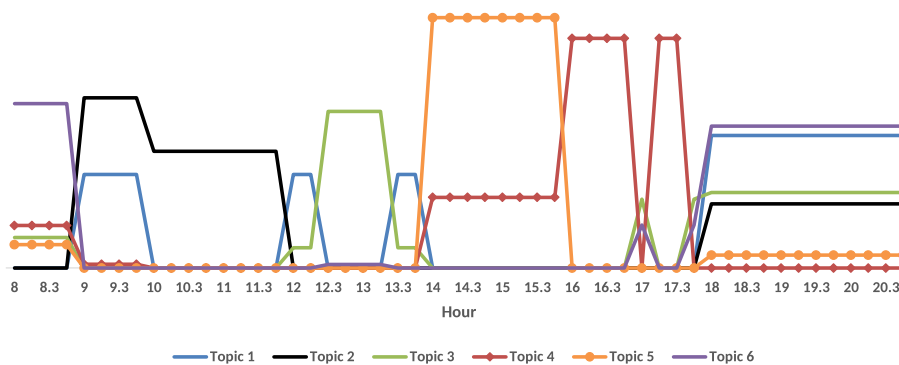
Table 11 shows the contents of the topics. The content often represents a meaningful set of user’s status labels to discover activity patterns. For lunch activity, the prominent words in topic 3 are L3, L6, L5, L1 and H3. Topics 4 and 5 have words H5 and H4 which represent afternoon office work. Similarly, topics 1 and 2 are a mixture of H1 and H2 words which represent morning office work.

Finally, Table 12 shows the relation between the ground truth activities and the average number of the activated topics for all persons. We manually calculated the average number of activated topics for each persons, and we selected the common and different topics between individuals. The average number of activated topics are high for leaving and arriving activities. This is attributed to different activity patterns of each individual, where some persons prefer to arrive early to work and others prefer to arrive late, the same observation applies to leaving activity. Also, each topic has different generated list of words, which reflects a variety in people’s preferences. The office activity has the highest average number of activated topics because each person has different working habit, where some persons may stay in the office for long periods without any breaks and others may take a coffee break or leave the office for certain amount of time. This generates

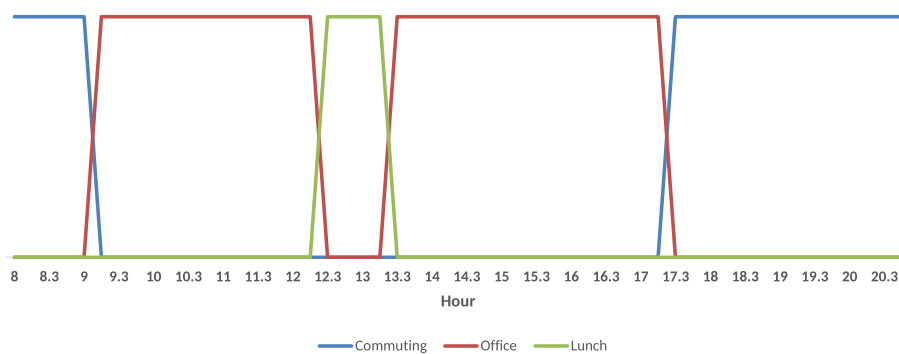


**Fig. 20** Different perplexity plots for three individuals: **a** person 1; **b** person 3 and **c** person 5

**Fig. 21** **a** The inferred topic activations for the day that was left out during training; and **b** ground truth for 1 day



**(a)** Topic activations



**(b)** Ground truth

**Table 11** Topics and its activities

Topic	Content of the topics
1	L6 (0.37), H1 (0.26), H3 (0.26), M1 (0.07) and M2 (0.02)
2	H1 (0.47), H2 (0.32), L6 (0.18), M4 (0.02)
3	L3 (0.43), L6 (0.21), L5 (0.19), L1 (0.08) and H3 (0.03)
4	H5 (0.64), H4 (0.20), L1 (0.12), M1 (0.03) and H1 (0.01)
5	H4 (0.69), M3 (0.14), L1 (0.06), M5 (0.04) and L6 (0.04)
6	L1 (0.46), L6 (0.39), L5 (0.12), M5 (0.02) and L3 (0.01)

The numbers in brackets indicate  $p(e|z)$ , i.e. the probability of the user's status label  $e$  given the current topic  $z$  (labels  $e$  with  $p(e|z) < 0.01$  are not shown). The distributions were estimated from 90 days of data

various list of words of each topic. In case of group activities such as meeting, lunch and holiday, we have noticed that there are three groups with different lunch activity. One group prefers to eat lunch from 12:00 to 13:00 outside the office, and another group prefers to eat lunch from 12:30 to 13:30 outside the office and the last group prefers to eat lunch inside the office from 12:30 to 13:30. These different lunch activities have been captured by three common activated topics. From the ground truth data, there are two group meetings which are taking place on two different

**Table 12** The relation between the ground truth activities and the average number of the activated topics

Ground truth activities	Average number of activated topics
Leaving	9
Arriving	6
Office (work)	16
Meeting	3
Lunch	2
Holiday	1

days. These meetings happen bi-weekly. This can be shown by two common topics between all individuals. All persons share the same activated topic for holiday activity.

### 7.6 Activity pattern variation analysis

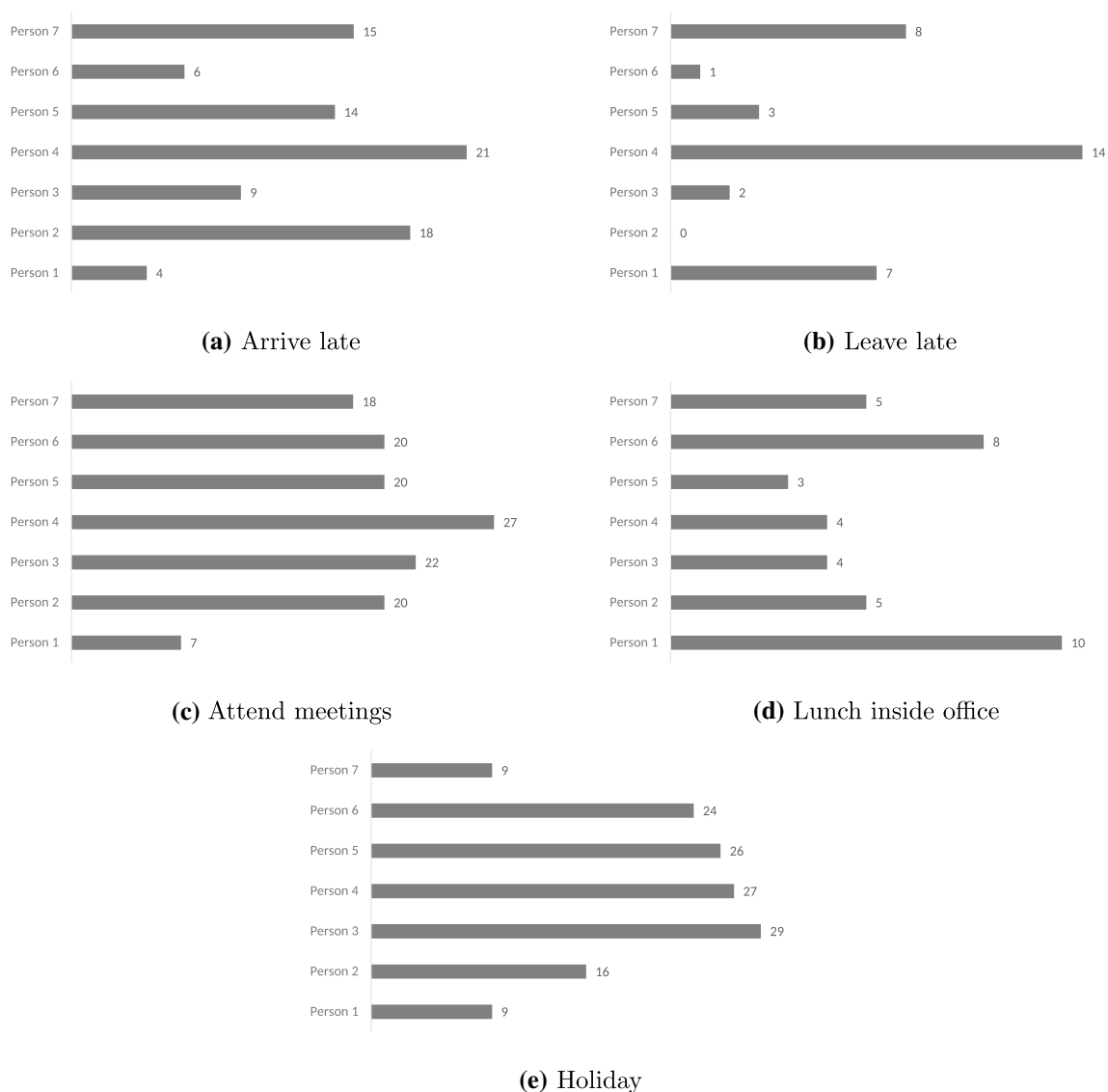
Previously, we have shown in Fig. 16 different activity patterns for group of topics. Some persons follow very regular, non-varying lifestyles, and others have more highly varying lifestyles such as working late in the evening, arriving to work late in the morning and having lunch breaks inside the

office. These variations may correspond to specific events. By analyzing how often a person works late in the evening or how often he attends meetings, we can recommend more healthy and efficient habits. We find topics that display certain activities we wish to inspect such as “leaving the office late”. We use LDA to rank days for these activities, and then count the number of times each person performs this activity pattern. Figure 22 compares five activity patterns between persons in the office. In Fig. 22a, *Person 2*, *Person 4*, *Person 5* and *Person 7* prefer to arrive to the office late in the morning, while the rest prefer to arrive early. Looking at Fig. 22b, *Person 1*, *Person 4* and *Person 7* work until late hour. All persons were attending meetings regularly as shown in Fig. 22c, except *Person 1*, because he had family emergencies. According to Fig. 22d *Person 1* and *Person*

6 prefer to eat lunch inside the office sometimes, while the others have high preference to eat the lunch in the cafeteria. Finally, Fig. 22e shows how often the persons in the office take holidays. *Person 1* and *Person 7* are used to come to the office more often than taking holidays. While, the rest of the office members show preferences of taking holidays.

## 8 Conclusions

We have installed a network of low-resolution visual sensors in an office environment of multiple persons for activity discovery. The low-resolution visual sensors ensure cheap and privacy preserved monitoring solution. Using a long-term and a real-life video dataset over a period of



**Fig. 22** A comparison between five activity patterns which represent special events for office workers: **a** arrive late; **b** leave late; **c** attend meetings; **d** lunch inside office; **e** holiday



5 months, we have presented a framework to discover the activity patterns by analyzing the users' positions. The analysis started by detecting the users' hotspots. Then, we have proposed two architectures to identify the persons' presence and absence using probabilistic graphical models and sequence mining technique. The detailed analysis and comparisons have showed how accurate the two-model mining approach than the single model approach.

Based on the persons' statuses, we have successfully discovered routines characteristic of days and persons in the study in an unsupervised manner using LDA topic model. The resulting distributions of words for latent topics, as well as topics given days, and topics given persons, reveal hidden structure of activity patterns which we use to perform varying tasks, including finding persons or groups of persons that display given activity patterns, and determining times when certain events or changes in events occur.

The PIR sensors may not raise privacy concerns as the low-resolution visual sensors, since there are no images captured of the users. There are two ways to address privacy concerns using low-resolution visual sensor. One way is to decrease the quantity and quality of the image data captured to the point where it no longer provides any visual information about the users. However, this will also decrease the accuracy of discovering activities. The number of visual sensors and their locations and resolutions are three important data dimensions that significantly impact both visual information and activity discovery accuracy. In this work, we used visual sensors with an image resolution of  $30 \times 30$  pixels for activity discovery. We showed that it is visible to discover office activities in low-resolution constraints. As a future work, we plan to study the limits to which we can reduce these data dimensions more (less than  $30 \times 30$  pixels) without significantly impacting activity discovery accuracy. Another way to address privacy concerns is to use post-processing algorithms to modify the original image, concealing different details using a level-based visualisation scheme, and at the same time, the usefulness of the information is retained.

While we have shown that many insights about activity patterns can be obtained with our approach, One of the major limitations in our work, it is the way we select the number of topics. For LDA, perplexity measure is used as a way to evaluate the performance of the model on unseen data. However, perplexity is not a "perfect" evaluation criteria for model selection, since topics with similar results are not considered in the perplexity computation. In practice, choosing smaller values of K would have yielded to less duplication of topics but also the topics become more general. Overall, perplexity measure is not a perfect way to select model, though other ways of determining model parameters do not give better results, and the problem of

model selection for topic models is an active problem (Blei et al. 2003).

Currently, the reported results are based on activity patterns being discovered using video data captured in the office environment. As a future work, we are interested to look at models which could account for varying activity pattern time intervals, specifically analyzing activity patterns on varying timescales, such as hourly, daily and weekly. Furthermore, we are planning to study the fusion of different heterogeneous sensor information such as the interaction activity with the computer, RFID and PIR sensors, along with the visual sensors. The study of sensor fusion helps to find the best combination of sensor information, and to build rich dataset for activity discovery.

**Acknowledgements** This research has been financed by the Belgian National Fund for Scientific Research (FWO Flanders) and Ghent University through the FWO project G.0.398.11.N.10. The evaluation was performed in the context of the projects "LittleSister" and the European AAL project "SONOPA," financed by the agency for Innovation by Science and Technology (IWT), imec and the EU Ambient Assisted Living programme.

## References

- Aztiria A (2010) Learning frequent behaviours of the users in intelligent environments. *J Ambient Intell Smart Environ* 2(4):435–436. doi:[10.3233/AIS-2010-0084](https://doi.org/10.3233/AIS-2010-0084)
- Baggenstoss PM (2001) A modified baum-welch algorithm for hidden markov models with multiple observation spaces. *IEEE Trans Speech Audio Process* 9(4):411–416. doi:[10.1109/89.917686](https://doi.org/10.1109/89.917686)
- Bickford M (2005) Stress in the workplace: a general overview of the causes, the effects, and the solutions. *Canadian Mental Health Association Newfoundland and Labrador Division*, pp 1–3
- Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3:993–1022
- Bo NB, Deboeverie F, Eldib M, Guan J, Xie X, Niño J, Van Haerenborgh D, Slembrouck M, Van de Velde S, Steendam H et al (2014) Human mobility monitoring in very low resolution visual sensor network. *Sensors* 14(11):20,800–20,824. doi:[10.3390/s141120800](https://doi.org/10.3390/s141120800)
- Camilli M, Kleihorst R (2011) Demo: Mouse sensor networks, the smart camera. In: *Fifth ACM/IEEE International Conference on Distributed Smart Cameras*, pp 1–3. doi:[10.1109/ICDSC.2011.6042944](https://doi.org/10.1109/ICDSC.2011.6042944)
- Castanedo F, de Ipia DL, Aghajan HK, Kleihorst R (2014) Learning routines over long-term sensor data using topic models. *Expert Syst* 31(4):365–377. doi:[10.1111/exsy.12033](https://doi.org/10.1111/exsy.12033)
- Chen CW, Aghajan H (2011) Multiview social behavior analysis in work environments. In: *Fifth ACM/IEEE International Conference on Distributed Smart Cameras*, pp 1–6. doi:[10.1109/ICDSC.2011.6042910](https://doi.org/10.1109/ICDSC.2011.6042910)
- Chen CW, Aztiria A, Aghajan H (2011a) Learning human behaviour patterns in work environments. In: *CVPR 2011 WORKSHOPS*, pp 47–52. doi:[10.1109/CVPRW.2011.5981696](https://doi.org/10.1109/CVPRW.2011.5981696)
- Chen CW, Aztiria A, Allouch SB, Aghajan H (2011b) Understanding the influence of social interactions on individuals behavior pattern in a work environment. In: *International Workshop on Human Behavior Understanding*, Springer, pp 146–157. doi:[10.1007/978-3-642-25446-8\\_16](https://doi.org/10.1007/978-3-642-25446-8_16)

- Chen CW, Ugarte RC, Wu C, Aghajan H (2011) Discovering social interactions in real work environments. *Face Gesture* 2011:933–938. doi:[10.1109/FG.2011.5771376](https://doi.org/10.1109/FG.2011.5771376)
- Cheng CC, Lee D (2014) Smart sensors enable smart air conditioning control. *Sensors* 14(6):11,179–11,203. doi:[10.3390/s140611179](https://doi.org/10.3390/s140611179)
- Cinaz B, Arnrich B, Marca R, Tröster G (2013) Monitoring of mental workload levels during an everyday life office-work scenario. *Person Ubiquitous Comput* 17(2):229–239. doi:[10.1007/s00779-011-0466-1](https://doi.org/10.1007/s00779-011-0466-1)
- Cosemans B, Cosmar M, Gründler R, Flemming D, Van den Broek K (2014) Calculating the cost of work-related stress and psychosocial risks. In: Tech. rep., European Agency for Safety and Health at Work, Luxembourg. doi:[10.2802/20493](https://doi.org/10.2802/20493)
- Docobo (2013) Sonopa:social networks for older adults to promote an active life. <http://www.sonopa.eu> (Online). Accessed 12 May 2016
- Duong T, Hazelton ML (2005) Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scand J Stat* 32(3):485–506. doi:[10.1111/j.1467-9469.2005.00445.x](https://doi.org/10.1111/j.1467-9469.2005.00445.x)
- Eijkelhof BH, Huysmans MA, Blatter BM, Leider PC, Johnson PW, van Dien JH, Dennerlein JT, van der Beek AJ (2014) Office workers' computer use patterns are associated with workplace stressors. *Appl Ergon* 45(6):1660–1667. doi:[10.1016/j.apergo.2014.05.013](https://doi.org/10.1016/j.apergo.2014.05.013)
- Eldib M, Bo NB, Deboeverie F, Nino J, Guan J, Van de Velde S, Steendam H, Aghajan H, Philips W (2014a) A low resolution multi-camera system for person tracking. In: *IEEE International Conference on Image Processing (ICIP)*, IEEE, pp 378–382. doi:[10.1109/ICIP.2014.7025075](https://doi.org/10.1109/ICIP.2014.7025075)
- Eldib M, Bo NB, Deboeverie F, Xie X, Philips W, Aghajan H (2014b) Behavior analysis for aging-in-place using similarity heatmaps. In: *Proceedings of the International Conference on Distributed Smart Cameras*, ACM, ICDSC '14, vol 6, pp 1–34. doi:[10.1145/2659021.2659038](https://doi.org/10.1145/2659021.2659038)
- Eldib M, Deboeverie F, Haerenborgh DV, Philips W, Aghajan H (2015a) Detection of visitors in elderly care using a low-resolution visual sensor network. In: *Proceedings of the 9th International Conference on Distributed Smart Cameras*, ACM, pp 56–61. doi:[10.1145/2789116.2789137](https://doi.org/10.1145/2789116.2789137)
- Eldib M, Deboeverie F, Philips W, Aghajan H (2015b) Sleep analysis for elderly care using a low-resolution visual sensor network. In: *Human Behavior Understanding*, Springer, pp 26–38. doi:[10.1007/978-3-319-24195-1\\_3](https://doi.org/10.1007/978-3-319-24195-1_3)
- Eldib M, Deboeverie F, Philips W, Aghajan H (2016a) Behavior analysis for elderly care using a network of low-resolution visual sensors. *J Electron Imaging* 25(4):041,003–041,003. doi:[10.1117/1.JEI.25.4.041003](https://doi.org/10.1117/1.JEI.25.4.041003)
- Eldib M, Deboeverie F, Philips W, Aghajan H (2016b) Towards more efficient use of office space. In: *Proceedings of the 10th International Conference on Distributed Smart Camera*, ACM, pp 37–43. doi:[10.1145/2967413.2967424](https://doi.org/10.1145/2967413.2967424)
- Eldib M, Zhang T, Deboeverie F, Philips W, Aghajan H (2016c) A data fusion approach for identifying lifestyle patterns in elderly care. In: *Active and Assisted Living: Technologies and Applications*, Healthcare Technologies, Institution of Engineering and Technology, pp 81–102. doi:[10.1049/PBHE006E\\_ch5](https://doi.org/10.1049/PBHE006E_ch5)
- EU-OSHA (2013a) Campaign guide managing stress and psychosocial risks at work. <https://www.healthy-workplaces.eu/en/campaign-materials/guide> (Online). Accessed 12 May 2016
- EU-OSHA (2013b) European opinion poll on occupational safety and health. In: Tech. rep. doi:[10.2802/55505](https://doi.org/10.2802/55505)
- Farrahi K, Gatica-Perez D (2011) Discovering routines from large-scale human locations using probabilistic topic models. *ACM Trans Intell Syst Technol* 2(1):3:1–3:27. doi:[10.1145/1889681.1889684](https://doi.org/10.1145/1889681.1889684)
- Griffiths TL, Steyvers M (2004) Finding scientific topics. *Proc Natl Acad Sci* 101(suppl 1):5228–5235. doi:[10.1073/pnas.0307752101](https://doi.org/10.1073/pnas.0307752101)
- Hamid R, Maddi S, Johnson A, Bobick A, Essa I, Isbell C (2009) A novel sequence representation for unsupervised analysis of human activities. *Artif Intell* 173(14):1221–1244. doi:[10.1016/j.artint.2009.05.002](https://doi.org/10.1016/j.artint.2009.05.002)
- Healey JA, Picard RW (2005) Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans Intell Transport Syst* 6(2):156–166. doi:[10.1109/TITS.2005.848368](https://doi.org/10.1109/TITS.2005.848368)
- Huynh T, Fritz M, Schiele B (2008) Discovery of activity patterns using topic models. In: *Proceedings of the 10th International Conference on Ubiquitous Computing, ACM, UbiComp '08*, pp 10–19. doi:[10.1145/1409635.1409638](https://doi.org/10.1145/1409635.1409638)
- iMinds (2013) Little sister: low-cost monitoring for care and retail. <https://www.iminds.be/en/projects/littlesister> (Online). Accessed 12 May 2016
- Jaramillo P, Amft O (2013) Improving energy efficiency through activity-aware control of office appliances using proximity sensing—a real-life study. In: *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, IEEE, pp 664–669. doi:[10.1109/PerComW.2013.6529576](https://doi.org/10.1109/PerComW.2013.6529576)
- Kim E, Helal S, Cook D (2010) Human activity recognition and pattern discovery. *IEEE Pervasive Comput* 9(1):48–53. doi:[10.1109/MPRV.2010.7](https://doi.org/10.1109/MPRV.2010.7)
- Lafferty J, McCallum A, Pereira F et al (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the eighteenth international conference on machine learning, ICML*. Morgan Kaufmann Publishers Inc., vol 1, pp 282–289
- Lancaster HO, Seneta E (1969) Chi-square distribution. *Wiley Online Library*. doi:[10.1002/0470011815.b2a15018](https://doi.org/10.1002/0470011815.b2a15018)
- Liao W, Zhang W, Zhu Z, Ji Q (2005) A real-time human stress monitoring system using dynamic bayesian network. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)—Workshops*, pp 70–70. doi:[10.1109/CVPR.2005.394](https://doi.org/10.1109/CVPR.2005.394)
- Liu DC, Nocedal J (1989) On the limited memory bfgs method for large scale optimization. *Math Prog* 45(1):503–528. doi:[10.1007/BF01589116](https://doi.org/10.1007/BF01589116)
- Milczarek M, Rial-González E, Schneider E (2009) OSH [Occupational safety and health] in figures: stress at work-facts and figures. Office for Official Publications of the European Communities
- Milenkovic M, Amft O (2013) An opportunistic activity-sensing approach to save energy in office buildings. In: *Proceedings of the fourth international conference on Future energy systems*, ACM, pp 247–258. doi:[10.1145/2487166.2487194](https://doi.org/10.1145/2487166.2487194)
- Mrazovac B, Bjelica MZ, Teslic N, Papp I (2011) Towards ubiquitous smart outlets for safety and energetic efficiency of home electric appliances. In: *2011 IEEE International Conference on Consumer Electronics—Berlin (ICCE-Berlin)*, pp 322–326. doi:[10.1109/ICCE-Berlin.2011.6031795](https://doi.org/10.1109/ICCE-Berlin.2011.6031795)
- Okada Y, Yoto TY, Suzuki T, Sakuragawa S, Sugiura T (2013) Wearable ecg recorder with acceleration sensors for monitoring daily stress: Office work simulation study. In: *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp 4718–4721. doi:[10.1109/EMBC.2013.6610601](https://doi.org/10.1109/EMBC.2013.6610601)
- Oliver N, Horvitz E (2005) A comparison of hmms and dynamic bayesian networks for recognizing office activities. In: *Proceedings of the 10th International Conference on User Modeling, UM'05*, pp 199–209. doi:[10.1007/11527886\\_26](https://doi.org/10.1007/11527886_26)
- Oliver N, Horvitz E, Garg A (2002) Layered representations for human activity recognition. In: *Multimodal Interfaces*, 2002.

- Proceedings. Fourth IEEE International Conference on, IEEE, pp 3–8. doi:[10.1109/ICMI.2002.1166960](https://doi.org/10.1109/ICMI.2002.1166960)
- Oliver N, Garg A, Horvitz E (2004) Layered representations for learning and inferring office activity from multiple sensory channels. *Comput Vis Image Underst* 96(2):163–180. doi:[10.1016/j.cviu.2004.02.004](https://doi.org/10.1016/j.cviu.2004.02.004)
- Rabiner L (1989) A tutorial on hidden markov models and selected applications in speech recognition. *Proc IEEE* 77(2):257–286. doi:[10.1109/5.18626](https://doi.org/10.1109/5.18626)
- Rish I (2001) An empirical study of the naive bayes classifier. In: *IJCAI 2001 workshop on empirical methods in artificial intelligence*. IBM New York, vol 3, pp 41–46
- Salamone F, Belussi L, Danza L, Ghellere M, Meroni I (2016) An open source smart lamp for the optimization of plant systems and thermal comfort of offices. *Sensors* 16(3):338. doi:[10.3390/s16030338](https://doi.org/10.3390/s16030338)
- Sheather SJ, Jones MC (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J R Stat Soc Ser B (Methodol)*. doi:[10.2307/2345597](https://doi.org/10.2307/2345597)
- Si Z, Pei M, Yao B, Zhu SC (2011) Unsupervised learning of event and-or grammar and semantics from video. In: *International Conference on Computer Vision, IEEE*, pp 41–48. doi:[10.1109/ICCV.2011.6126223](https://doi.org/10.1109/ICCV.2011.6126223)
- Silverman BW (1986) *Density estimation for statistics and data analysis*, vol 26. CRC Press. doi:[10.1007/978-1-4899-3324-9](https://doi.org/10.1007/978-1-4899-3324-9)
- Simonoff J (1996) *Smoothing methods in statistics*. Springer, Berlin
- Sutton C, McCallum A (2012) An introduction to conditional random fields. *Found Trends Mach Learn* 4(4):267373. doi:[10.1561/22000000013](https://doi.org/10.1561/22000000013)
- Tao S, Kudo M, Nonaka H, Toyama J (2011) Person authentication and activities analysis in an office environment using a sensor network. In: *International Joint Conference on Ambient Intelligence*, Springer, pp 119–127. doi:[10.1007/978-3-642-31479-7\\_19](https://doi.org/10.1007/978-3-642-31479-7_19)
- Teixeira T, Dublon G, Savvides A (2010) A survey of human-sensing: methods for detecting presence, count, location, track, and identity. ENALAB technical report
- Varadarajan J, Emonet R, Odobez JM (2013) A sequential topic model for mining recurrent activities from long term video logs. *Int J Comp Vis* 103(1):100–126. doi:[10.1007/s11263-012-0596-6](https://doi.org/10.1007/s11263-012-0596-6)
- Wojek C, Nickel K, Stiefelwagen R (2006) Activity recognition and room-level tracking in an office environment. In: *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp 25–30. doi:[10.1109/MFI.2006.265608](https://doi.org/10.1109/MFI.2006.265608)
- Xie X, Deboeverie F, Eldib M, Philips W, Aghajan H (2014) Phd forum: Analyzing behaviors patterns of the elderly from low-precision trajectories. In: *Proceedings of the International Conference on Distributed Smart Cameras, ACM, ICDSC '14*, vol 2, pp 1–47. doi:[10.1145/2659021.2675057](https://doi.org/10.1145/2659021.2675057)
- Ziefle M, Rocker C, Holzinger A (2011) Medical technology in smart homes: exploring the user's perspective on privacy, intimacy and trust. In: *IEEE 35th Annual Computer Software and Applications Conference Workshops*, pp 410–415. doi:[10.1109/COMPSACW.2011.75](https://doi.org/10.1109/COMPSACW.2011.75)
- Zimmermann P, Guttormsen S, Danuser B, Gomez P (2003) Affective computing a rationale for measuring mood with mouse and keyboard. *Int J Occup Saf Ergon* 9(4):539–551. doi:[10.1080/10803548.2003.11076589](https://doi.org/10.1080/10803548.2003.11076589)