Learning temporal context for activity recognition

Claudio Coppola and **Tomáš Krajník** and **Tom Duckett** and **Nicola Bellotto** [1]

Abstract. We present a method that allows to improve activity recognition using temporal and spatial context. We investigate how incremental learning of long-term human activity patterns improves the accuracy of activity classification over time. Two datasets collected over several months containing hand-annotated activity in residential and office environments were chosen to evaluate the approach. Several types of spatial and temporal models were evaluated for each of these datasets and the efficiency of each method was assessed by the way it improved activity classification. The results indicate that incremental learning of daily routines allows to dramatically improve activity classification. For example, a weak classifier deployed in a single-inhabited apartment for a period of three weeks was enhanced with a temporal model that increased its accuracy from 20% to 60%.

Figure 1. Witham dataset - ceiling camera view.

1 Introduction

Automated recognition human activities has recently become a hot topic of research. It enables a wide range of applications such as security, retail or healthcare, but recently a huge focus has been given to the recognition of the Activities of Daily Living (ADL) due to its potential application in Ambient Assisted Living (AAL). This technology could help to overcome the predicted need of health workers and improve the quality of life of the increasing elderly population in the near future, by assisting people in their daily tasks and identifying potential problems. Furthermore, it could be used also in security applications to detect anomalous situations which could endanger people or property. The introduction of new technologies has made this problem easier to address. In particular, RGB-D sensors together with the pose estimation software and the smart sensors for the Internet of Things have enabled the possibility of acquiring data for such applications, giving birth to many related datasets [3, 14, 34, 1]. The development of Activity Recognition is furthermore supported by novel techniques to manage huge quantities of data (Big Data) and the increased computational power of modern computers, enabling real-time implementations.

The main focus of the recognition models has been the recognition of patterns derived from the data acquired from the sensors. The features used for pattern recognition typically relate to the body movement and the surrounding context, in the case of RGB-D sensors, or by the sensor events in a smart environment. By contrast, in this work we aim to exploit the long-term patterns of recurring activities to improve the performances of activity classification. Prior work [16] showed that the patterns of the spatio-temporal dynamics of the environment can be exploited to improve the indoor localization of a mobile robot.

[1] Lincoln Centre for Autonomous Systems, University of Lincoln, UK email: ccoppola@lincoln.ac.uk

In a similar way this work proposes an approach to calculate prior probabilities of an activity happening at a certain time, which improves the error rate of a given classification algorithm. We analyse several possible techniques, including a novel approach based on Adaptive Interval Based Models, which delivers continuous improvement to the recognition performance on-the-fly, by incrementally performing naive Bayesian learning. We evaluate our methods on the Aruba Dataset [3], based on the ADL activities and the Witham Dataset, manually annotated from camera recording (Figure 1) in an office environment.

There are two main contributions in this paper: (i) The introduction of a probabilistic formulation to incrementally model temporal and spatial context to improve Activity Recognition performance of a given classifier with a spatial or temporal model. (ii) The introduction of novel probabilistic models of temporal and spatial context. (iii) Comparison of different temporal models in order to understand which ones can better represent the temporal structure of daily activities.

The remainder of this paper is organized as follows. Section 2 will give an overview on the state-of-the-art for Activity Recognition performed with smart sensors and RGB-D cameras and on the use of temporal and spatial models for Activity Recognition. Section 3 will provide a formulation of the Activity Recognition problem. Section 4 introduces the models we are based on. Section 5 explains our method of evaluation for the temporal models. Section 6 will comment the results of our experiment, and finally Section 7 presents the conclusion and future work.

2 Related work

Human activity recognition aims to recognize the actions and goals of human agents using a sequence of observations on the agents' actions and the environmental conditions. Tracking and understanding

human behaviour through videos is a very important and challenging problem with various useful applications. Activity Recognition has originally been performed on RGB video streams with a wide spectra of solutions [13, 27].

The development of cheap RGB-D cameras has contributed to the increased focus on this problem, since they allow to reduce the computational requirement for estimating the pose of human body and the contextual patterns in the scene in real-time. In [9, 10] a probabilistic ensemble of classifiers called Dynamic Bayesian Mixture Model (DBMM) is proposed to combine different posterior probabilities from a set of classifiers for activity recognition. Wang et al. [35] show a deep structured model built with layered convolutional neural networks. A biologically inspired approach adopting an artificial neural network to combine pose and motion features for action perception is proposed by [25]. In [5], a simple way to apply qualitative trajectory calculus to model 3D movements of the tracked human body using hidden Markov models (HMMs) is presented. Sung et al. in [28] and [29] perform activity recognition in unstructured environments such as homes and offices with an RGBD camera. The movement is modelled by transforming the rotation matrix of each joint to the body torso and inferring the activities and sub-activities with a 2-layered Maximum Entropy Markov Model (MEMM). A three-level hierarchical discriminative approach is presented in [20]. The activities are decomposed into a lower level representing the pose data, an intermediate level where the poses are combined into simple human actions, and a high level where the actions are spatially and temporally combined into complex human activities. The approach presented in [26] uses HMMs combined with Gaussian Mixture Models (GMM) to model the combination of continuous joint positions over time for activity recognition. In [33], the authors use random occupancy patterns to model activities using context from depth data.

Smart environments allow to mine though the sensor events to classify which activity has happened. [12] presents a dataset with smart sensors for ADL recognition, where the classification is performed using Support Vector Machines (SVM). A mining technique to find the association rules between the activities and their frequent patterns in smart environments is presented in [36]. In [8], the authors use the Back-Propagation algorithm to train a feed-forward Neural Network with features extracted from the motion sensor events. In [7], a method for evaluating the confidence of classification is presented. It is performed with SVM for a certain activity to reduce false positives so that samples with low confidence can be further investigated by a human operator. In [4] an activity discovery algorithm is presented which identifies patterns in sensor data with a greedy approach. It searches for a sequence pattern that best compresses the input data; the data is scanned to create initial patterns of length one, which are extended in every loop while minimizing the description of the data.

In [24] analysis of human activities in an office environment is performed using a Layered Hidden Markov Model (LHMM) architecture based on real-time streams of evidence from video, acoustic, and computer interactions. Similarly, a multi-level HMM is presented in [37] for recognising office activities and tracking the users across the rooms. In [23] a solution for office activity recognition is proposed, which handles multiple-user, multiple-area situations, based on an ontological approach, using low-cost, binary and wireless sensors. The idea of exploiting long-term analysis has been presented already by Van Laerhoven et al. [32], using wrist-worn sensors to collect daily activity data to create rhythmic models of the activities. These models are created off-line using a frequentist approach, accumulating the amount of times an annotated activity starts and stops within a certain time interval, which is represented as a bin. In [21] a long-term annotated dataset using many different sensors is introduced. The classification is performed using a binary classifier for each learned activity, collecting features from the sensor data in particular time windows. Daily routines are recognized in [2] from features extracted with a sliding window approach. These are clustered with k-means to calculate their occurrence statistics and store them in a histogram which is classified using a Joint Boosting technique. The authors in [30] introduce a wellness determination process to help healthcare providers to assess the performance of the elderly in their daily activities. It verifies the behaviour of elderly people at three different stages (usage of appliances, activity recognition and forecast levels) in a smart home monitoring environment integrating the spatial and temporal information.

In [6] a model is introduced for long-term monitoring of activities in a smart home. The classification is performed with a Probabilistic Neural Network (PNN), and the daily schedules of activities are then clustered with K-means. The clusters with highest inter-variation ae considered as normal and the others as their deviations. [22] presents a way of predicting future activity occurrences, with a recurrent predictor, based on the structure of the temporal sequence of the activities. Long-term modelling of indoor environments has been exploited also in other cases. In [17], the authors argue that part of the environment variations exhibit periodicities and represent the environment states by their frequency spectra. The concept of Frequency-based Map Enhancement (FreMEn) was applied to occupancy grids in [19] to achieve compression of the observed environment variations and to landmark-based maps in order to increase robustness of mobile robot localization [15]. In this paper, we proposed a method that can be applied to existing classification algorithms for activity recognition, learning the temporal structure of the classified activities in order to incrementally improve the classification results on-line. Furthermore we investigate several possible models which can be used to model the (prior) occurrence probability of the learned activities.

3 Problem formulation

We formulate the activity classification problem simply as a Bayesian decision making problem. Let us assume that at time t, a person is performing an (unknown) activity from the set of possible activities $\mathcal{A}$ while being observed by a set of sensors. Let some algorithm C process the sensory readings and classify that the activity being performed is $o \in \mathcal{A}$. Let us assume that we have experimentally established the performance of C on some representative dataset and thus, we know C's confusion matrix, i.e. we can characterise the performance of C as a conditional probability $p(o|a)$. Thus, every time the algorithm C provides us with an observation o, we can establish the posterior distribution $p(a|o,t)$ over the possible activities at time t as:

$$p(a|o,t) = \frac{p(o|a)\,p(a,t)}{\sum_{b\in A} p(o|b)p(b,t)}. \tag{1}$$

In our case, we will use a separate spatial/temporal model per each activity. To emphasize that the models are calculated separately, we rewrite the Equation (1) for a single activity a as

$$p_a(o,t) = \frac{p(o|a)\,p_a(t)}{p(o|a)p_a(t) + p(o|\neg a)(1-p_a(t))}, \tag{2}$$

where $p_a(t)$ represents the probability of the activity a being performed at time t, i.e. the temporal prior of a. The expression $p_a(t)$ was chosen to emphasize that the temporal models are built independently - it corresponds to $p(a,t)$ in Equation (1).

While most of the research in activity recognition is aimed at the performance of the activity recognition algorithm C, which increases the chance of correct activity classification by improving $p(o|a)$ in Equation (2), our work is not concerned with the actual method that is used to determine the activity from the sensory readings. Instead, we focus on the term $p_a(t)$ in (2, which effectively represents the temporal context of a given activity. We hypothesize that since people tend to perform certain activities on a regular basis, $p_a(t)$ is a (pseudo-)periodic function that can be learned over time and that better knowledge of $p_a(t)$ would positively impact the performance of the classification system represented by Equation (2).

To learn $p_a(t)$, we apply Equation (2) iteratively. Initially, we start with all $p_a(t) = 1/|\mathcal{A}|$, i.e. we assume that the activities occur with the same probability regardless of the time. Whenever an activity is classified by (2), we use the output of (2) to update $p_a(t)$ and use the updated $p_a(t)$ in the following classification step.

The key questions that our paper addresses are:

1. Which model should be used to represent the temporal activity context (or prior) $p_a(t)$?
2. How much does the temporal context impact the performance of state-of-the-art classifiers ?
3. Can we learn the temporal context even with a weak classifier ?

To answer these questions, we tested three different temporal models on two datasets, which contain human activities labelled minute-by-minute over several weeks.

4 Temporal models

In our work, a temporal model of activity a is a function $p_a(t)$, which represents the probability of the activity a occurring at time t. We consider four types of temporal models: Frequency Map Enhancement (FreMEn), which represents cyclic processes by their frequency spectra, Gaussian Mixtures, which are well established in several domains, and naïve and adaptive versions of interval-based models.

4.1 Frequency map enhancement

Frequency Map Enhancement (FreMEn) is an emerging technique that improves the efficiency of mobile robots that operate autonomously for long periods of time [15, 11]. The method assumes that states of the robots' operational environments are affected by pseudo-periodic processes, whose influence and periodicity can be obtained through the Fourier transform. Thus, the uncertainty of a given state $s(t)$ is represented as a probabilistic function of time that is a combination of harmonic functions:

$$p(t) = \alpha_0 + \sum_{i=1}^{n} \alpha_i cos(\omega_i t + \varphi_i), \tag{3}$$

where the amplitude α_i, phase shift φ_i and frequency ω_i correspond to the most prominent spectral components of the observations of the original state $s(t)$.

In our case, the state $s(t)$ of the FreMEn model is a binary function of time $o_a(t)$ which indicates if the activity a was observed at time t and $p_a(t)$ will be our probabilistic function $p(t)$. To build the FreMEn model, we simply take the results of the past classifications and form a sequence $o_a(t)$ for each activity $a \in \mathcal{A}$. Then, we calculate the Fourier spectrum of each sequence $o_a(t)$, select n of its most prominent (i.e. with highest amplitudes) spectral components and use their amplitudes, periodicities and phase shifts as (α_i, ω_i and φ_i) parameters of the predictive FreMEn model in Equation (3), which is used as a prior for classification in Equation (2. Since the performance of the FreMEn model is affected by the choice of the model order n, we run our experiments with n ranging from 0 to 9 and chose the best performing setting, which was $n = 3$. To speed up calculations, we used the version of FreMEn introduced in [18], which allows for incremental updates.

The main advantage of the FreMEn is that it naturally represents multiple periodicities that are inferred automatically from the data. However, it poorly presents periodic, but short duration activities, such as teeth brushing or tea making.

4.2 Gaussian Mixture Models

Gaussian Mixture Models, which approximate multi-dimensional functions as weighted sums of Gaussian component densities, are a well-established method that find their applications in numerous fields from Psychology to Astrophysics [31]. A Gaussian Mixture Model of a function $f(t)$ is a weighted sum of m Gaussian functions:

$$f(t) = \frac{1}{\sqrt{2\pi}} \sum_{j=1}^{m} \frac{w_j}{\sigma_j} e^{-\frac{(t-\mu_j)^2}{2\sigma_j^2}}. \tag{4}$$

The parameters of the GMM components, i.e. the means μ_j, variances σ_j and weights w_k, are typically calculated from the training data by an iterative Expectation Maximization (EM) or Maximum Aposteriori (MAP) algorithms. Since the classic GMMs are not meant to represent periodic functions, we simply assume that people perform most of their activities on a daily basis and limit the time domain of GMM-based models to one day. While this assumption is not entirely correct (as activities of weekdays differ from the weekend ones), such a temporal model might still perform better than a 'static' one, where the probability of a given activity is constant in time.

To build the GMM model of $p_a(t)$, we first create a temporal sequence of observations $o_a(t)$ for each activity in the same way as in the FreMEn case. Then, we calculate an initial prior as follows:

$$p'_a(t) = \frac{k}{\tau} \sum_{i=1}^{\lfloor k/\tau \rfloor} o_a(t + (i-1)\tau), \tag{5}$$

where τ is the assumed period (in our case $\tau = 86400$ s) and k is the $s(t)$ sequence length. After calculating $p'_a(t)$, we employ the Expectation Maximization algorithm to find the means μ_i, standard deviations σ_i and weights w_i of its Gaussian Mixture approximation:

$$p_a(t) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{n} \frac{w_i}{\sigma_i} e^{-\frac{((t \bmod \tau)-\mu_i)^2}{2\sigma_i^2}}, \tag{6}$$

where τ is the apriori known period of the function $p_a(t)$ and $\bmod$ is a modulo operator.

The weaknesses of periodic GMMs (PerGaM) are complementary to the advantages of the FreMEn. Periodic GMMs can approximate short-duration activities, but they can represent only one period that has to be known a priori. Similarly to FreMEn, the performance of GMMs depends on the choice of n, which represents the number of Gaussians used in the mixture model. Again, we run our experiments with n ranging from 0 to 9 and chose the best performing setting, which was $n = 3$.

4.3 Interval-based Models

Another temporal model that has been considered partitions the time in disjoint intervals, each with a different prior probability $p_a(t)$. Similarly to the GMM-based models, the partitioning requires that the periodicity τ and model order n (the number of intervals) are chosen a priori. In our interval-based model, $p_a(t)$ is represented by n values $p'_a(k)$ that denote prior probabilities of a given activity occurring between $\tau m + \tau\frac{k}{n}$ and $\tau m + \tau\frac{k+1}{n}$, where $m \in \mathbb{N}$ and $k \in \{0, 1 \dots k-1\}$. In the following text, we will refer to the time interval τ/n as the "interval width". To update or retrieve $p_a(t)$, one has to simply determine the index k of the relevant interval:

$$p_a(t) = p'_a(k) = p'_a(\lfloor (t \bmod \tau)\frac{n}{\tau} \rfloor), \tag{7}$$

where $\lfloor x \rfloor$ is a floor operator, that returns the integer part of x.

Unlike in FreMEn and GMM models, the interval-based model is updated according to Bayes rule in Equation (2). Thus, when a classification is performed at time t, we first calculate k by Equation (7) and then perform the model update as follows:

$$p'_a(k) \leftarrow \frac{p(o|a)\, p'_a(k)}{\sum_{a \in A} p(o|a) p'_a(k)}, \tag{8}$$

Again, a crucial question here is model granularity (i.e. the interval width that is determined by the number of the represented intervals n). Models with wide intervals cannot represent short-duration activities, whereas models with short intervals require larger amounts of data for training, therefore their learning rate is slow.

4.4 Adaptive Interval Models

To deal with the aforementioned problem, we can store the number of updates performed for each interval $u(k)$ and calculate $p_a(t)$ by aggregating the probabilistic values of neighbouring intervals, so that $p_a(t)$ is based at least on l updates. While the model update remains the same as in the previous case (see Equation (8)) and the only difference it that the value of $u(k)$ is increased by 1, calculating $p_a(t)$ differs. To determine $p_a(t)$, we first calculate the index of the relevant interval k as $\lfloor (t \bmod \tau)\frac{n}{\tau} \rfloor$ (see Equation (7)). We check if the number of updates performed to calculate $p'_a(k)$ is at least l and if not, we include the neighbouring intervals and calculate $p(t)$ as the weighted (by the number of updates) average. This is repeated until the number of measurements used to determine $p_a(t)$ exceeds l. See Algorithm 1 for more details.

Algorithm 1 Adaptive interval prior calculation

function CALCULATEPRIOR($t, \tau, n, \boldsymbol{u}, \boldsymbol{p'_a}, l$)
  $k \leftarrow \lfloor (t \bmod \tau)\frac{n}{\tau} \rfloor$ ▷ determine interval index
  $m \leftarrow u(k)$ ▷ initialize total number of measurements
  $p \leftarrow m\, p'_a(k)$ ▷ initialize prior probability
  while $m < l$ **do** ▷ num. of measurements must be at least l
    $p \leftarrow p + p'_a(k+1)u(k+1)$ ▷ add neighbour prior
    $p \leftarrow p + p'_a(k-1)u(k-1)$ ▷ add neighbour prior
    $m \leftarrow m + u(k+1) + u(k-1)$ ▷ update meas.num.
  end while
  $p_a(t) \leftarrow p/m$ ▷ the resulting prior is a weighted average
end function

This "adaptive interval" method calculates $p_a(t)$ over several intervals in case there is not enough data available, which is equivalent to adjusting the interval width to the number of data gathered. However, one still has to choose the minimal interval width (we selected 60 s), the periodicity (we choose $\tau = 1$ day) and l, which is the minimal number of measurements required to calculate $p_a(t)$. The optimal number of measurements l is subject to investigation in the following sections.

4.5 Modelling the spatial context

We also evaluated the use of spatial context without temporal information in activity recognition. The use of spatial context context is motivated by the fact that certain activities are tied to certain locations, e.g. cooking typically occurs in a kitchen. Similarly to temporal models, we formalise a spatial model of activity a as a function $p_a(l)$, which represents the probability of the activity a performed by a person when at location l. The process of using and building a spatial context model is similar to the interval temporal models:

$$p_a(l) \leftarrow \frac{p(o|a)\, p_a(l)}{\sum_{a \in A} p(o|a) p_a(l)}, \tag{9}$$

The only difference is that the location l is not calculated based on the time, but on the position of the person. Combination of spatial and temporal context is considered for an extended version of this paper.

4.6 Model overview and evaluation

Each of the aforementioned models has advantages and drawbacks. A comparison of the PerGaM and FreMEn models applied to a "reading" activity in an office is shown in Figure 2. We assume that the interval-based models do not require an illustrative example. The

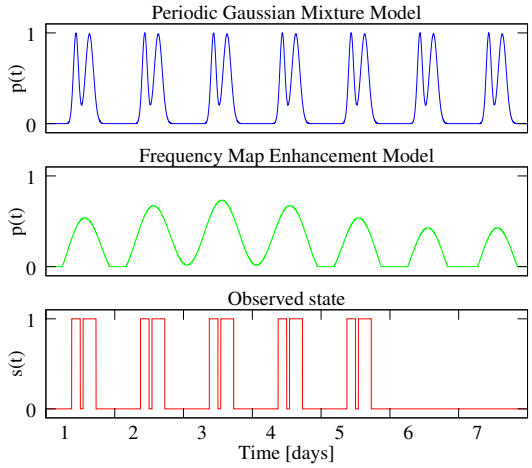

Figure 2. PerGaM and FreMEn models examples comparison.

main aim of this work is to investigate how these models perform when being used as priors for activity recognition. We abstract from the actual algorithm that is used for classification - we simply assume that the classifier can use the priors provided by our spatial and temporal models to estimate which activity is being performed. We assume that if the priors are not provided, the performance of a given classifier depends on its confusion matrix, which represents the conditional probability $p(o|a)$. The primary metric to be investigated is the overall activity recognition error, i.e. the probability that $o \neq a$.

Table 1. Activities of the Aruba and Witham experiments.

Aruba dataset	Witham dataset
Bed to Toilet	Outside
Eating	Reading
Enter Home	Writing
Housekeeping	Watching a video
Leave Home	Cooking
Meal Preparation	Talking
Relax	Sleeping
Resperate	Phonecall
Sleeping	Go to toilett
Wash Dishes	Other
Work	

5 Experiments

To evaluate the usefulness of the individual models for activity recognition, we performed their comparison on two datasets that cover several weeks of human activity at home and at work.

The first, 'Aruba' dataset was collected by the Center for Advanced Studies in Adaptive Systems (CASAS) to support their research concerning smart environments [3]. The Aruba dataset contains ground-truthed activities (Table 1) of a home-bound person in a small apartment for 16 weeks. The second 'Witham' dataset was gathered at the Lincoln Centre for Autonomous System (L-CAS) as part of the large-scale EU-funded STRANDS project, which aims to enable long-term autonomous operation of intelligent robots in human-populated environments. The Witham dataset, which was gathered for four weeks, contains activities (Table 1) of one of the L-CAS researchers.

5.1 Aruba dataset

The Aruba dataset [3] consists of measurements collected by 50 different sensors distributed over a 10×12 m^2, seven-room apartment over a period of 16 weeks.

During data collection, the apartment was occupied by a single person who was occasionally visited by other people. While the starting and finishing times of activities are provided with the CASAS dataset, the location of the person is not. Thus, we partitioned the apartment into nine different locations, seven of which represent different rooms and two correspond to corridors, and estimated the person location from the events of the apartment's motion detectors. Thus, the Aruba dataset contains a minute-by-minute timeline of 12 different activities performed at 9 different locations over the course of 16 weeks.

5.2 Witham Wharf dataset

The Witham dataset was collected in an open-plan office of the Lincoln Centre for Autonomous Systems (L-CAS). The office consists of a kitchenette, resting area, lounge and 20 working places that are occupied by students and postdoctoral researchers. We installed a ceiling camera that took a snapshot of the office every 10 seconds for 3 weeks, see Figure 1, and we hand-annotated activities and locations of one of the researchers over time.

The 'Witham' dataset contains a minute-by-minute timeline of 10 different activities performed at 10 different locations over the course of 3 weeks.

Figure 3. Aruba dataset - reconstructed layout of the apartment [3].

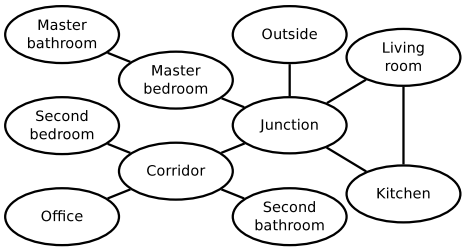

Figure 4. Aruba dataset - topological structure of the apartment.

5.3 Evaluation

As mentioned before, we abstract from the internal working of the classifier itself and we simply assume that it can take into account the priors that we provide by our spatial and temporal models. Thus, we base our evaluation on the fact that we know the conditional probabilities $p(o|a)$ which are represented by the confusion matrix of the evaluated classifier.

The evaluation starts with the prior models being invariant to time (and location) and equal to each other, i.e.

$$p_a(t) = \frac{1}{|\mathcal{A}|}, \quad \forall a \in \mathcal{A}, \quad \forall t \in \mathbb{R}. \tag{10}$$

Then, we retrieve the activity performed at time $t = 0$ from the given dataset and, using the priors initialised by Equation (10) and known $p(o|a)$, we calculate the posterior probabilities $p_a(t|o)$ with the Bayes Equation (2). After that, we simulate the stochastic nature of the activity classification process by running a Monte-Carlo scheme over the probabilities $p_a(t|o)$ and we obtain the simulated

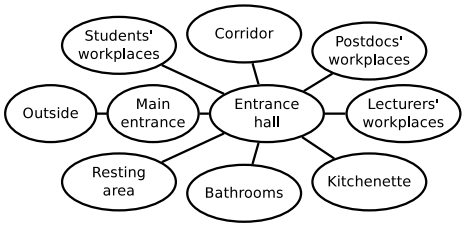

Figure 5. Witham dataset - topological structure of the apartment.

classification result $o(t) \in \mathcal{A}$. Then, we update the binary sequences $o_a(t)$ of each activity as follows:

$$\begin{aligned} o_a(t) = 1 &\iff o(t) = a, \\ o_a(t) = 0 &\iff o(t) \neq a. \end{aligned} \tag{11}$$

These sequences are then processed by the models. Then, we increment the time by 60 s and repeat the procedure again. After 1440 iterations, which represent the activity recognition results minute-by-minute for a full day, we compare the ground truth to the results of the simulated activity recognition $o(t)$ and calculate the activity classification error for that particular day. This error is calculated for every day of the available datasets.

5.3.1 *Classifiers considered*

In our experiments, we evaluate the spatial and temporal models with three different classifiers represented by different distributions $p(o|a)$. The first "weak" classifier has only a 20% probability of correct recognition, i.e. its confusion matrix has 0.2 on the diagonal and the other elements are equal. This corresponds to a 80% classification error (Figure 6a). The second "good" classifier has a 20% (Figure 6b) classification error - its confusion matrix diagonal elements are equal to 0.8 and the non-diagonal elements are identical.

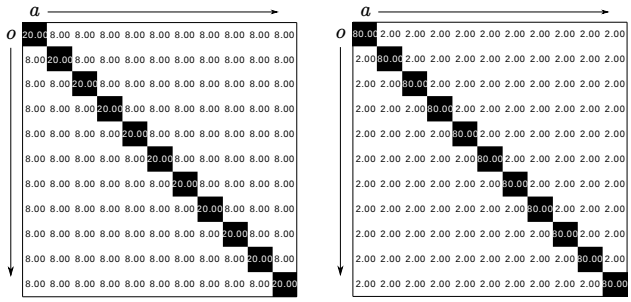

Figure 6. Confusion matrices of the "weak"(a) and the "good"(b) classifiers which characterize their $p(o|a)$ with the size of the Aruba dataset.

Finally, we consider a "real" classifier that was evaluated on the Aruba dataset in [7]. Here, the authors evaluate the performance of a classifier that can indicate lack of evidence to perform an actual classification. This is represented by a special type of observation, called "Irregular", which constitute an additional column in their classifier's confusion matrix. To obtain a square confusion matrix required by our method, the conditional probabilities represented by this additional column are uniformly redistributed across the matrix. The average value of the diagonal elements of the "real" classifier's confusion matrix is 85.14% (Figure 7a).

On the Witham dataset, instead, there are no classifiers existing from previous works. To represent the $p(o|a)$ of a "real" classifier for the Witham dataset, we used a 10×10 submatrix of the "real" classifier used with the Aruba dataset (Figure 7b).

6 Experimental results

Each of the models mentioned in Section 4 depend on a parameter as summarised in Figure 8. Here we discuss the sensitivity of these models to the parameter values and how well the models perform on the aforementioned datasets.

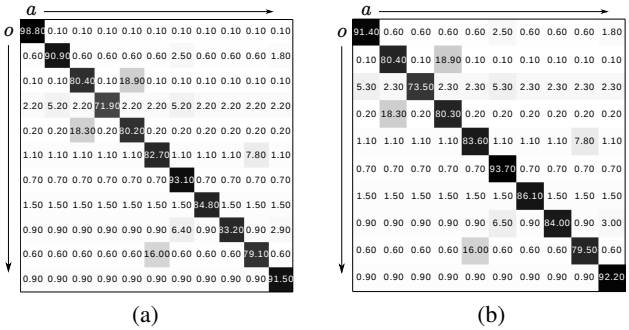

Figure 7. Confusion matrices which charaterize the $p(o|a)$ of the 'real' classifiers for the Aruba(a) and Witham(b) datasets.

Temporal model	Parameter type	Units	Used value
GMM	num. of Gaussians	-	3
Fremen	num. of periodics	-	3
Interval-based	interval width	minutes	60
Adaptive interv.	num. of samples	-	1000

Figure 8. The list parameters for each temporal model which improve the results the most on the datasets.

6.1 Model Parameters

The Fremen results seem to be very stable to the order of the model. Increasing the order does actually increase the classification performance but not significantly, as shown in Figures 9 and 10. The only exception is the Static component in the Aruba dataset, since in case of a weak base-classifier the performance increase does not reach the same magnitude of the higher orders. This suggests that using a FreMEn model of order 3 is sufficient to obtain a good reduction of the error rate.

A similar result was observed using Gaussian Mixture Model based priors. Indeed, as can be seen in Figures 11 and 12, the results are fairly stable with respect to the order of the model, although a model of order 5 seems to overfit the data in case of a real classifier, increasing the error rate accordingly.

For the Interval-based Models, the choice of the interval width can be very crucial, as shown in Figures 13 and 14. In case of a weak base classifier, a bin width of one hour produced the best results. Furthermore, this choice is the only one improving the same classifier on the Aruba dataset. In all the other cases the sensitivity of the error rate is not very strong.

The Adaptive intervals adapt the interval width according to the available quantity of evidence, so the smaller the number of samples the closer the behaviour will be to the atomic unit (1 minute in our case). As shown in Figures 15 and 16, the adaptive interval with a single sample has the same behaviour of the static interval with 1 minute width. In case of weak classifiers, the number of samples for the adaptation of the intervals does not influence the classification performance, and the same happens with a real base-classifier. In case of a good classifier (20% error rate) instead, the increase of the samples highly modify the performances of the model, lowering the error rate.

According to our experiments, the models which are the least sen-

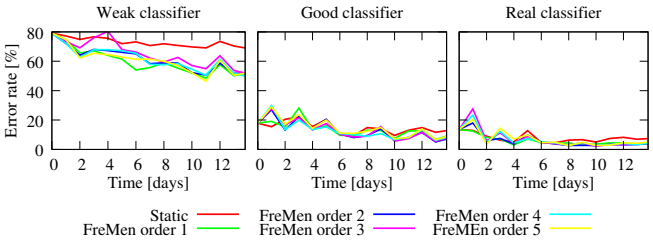

Figure 9. Impact of the number of modelled periodical processes on the FreMEn model - Aruba dataset.

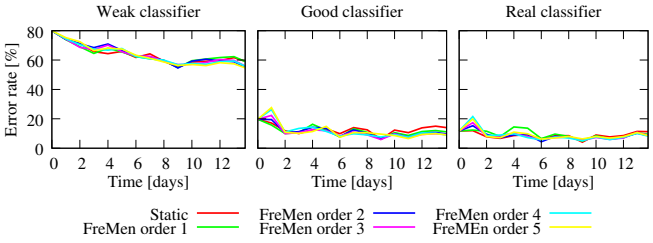

Figure 10. Impact of the number of modelled periodical processes on the FreMEn model - Witham dataset.

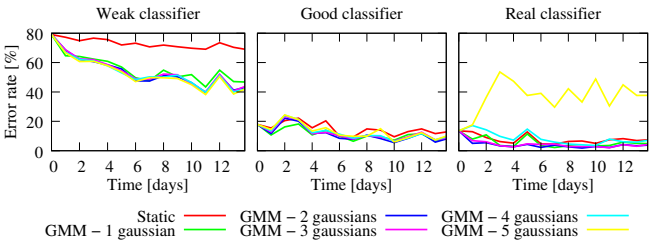

Figure 11. Impact of the number of Gaussians included on the performance of the Gaussian Mixtures - Aruba dataset.

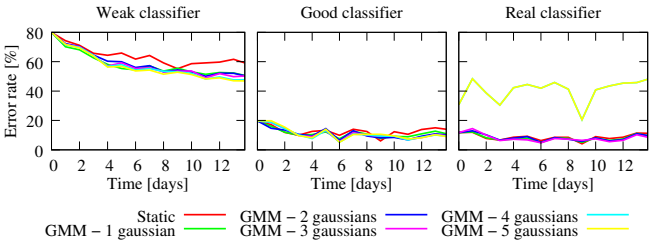

Figure 12. Impact of the number of Gaussians included on the performance of the Gaussian Mixtures - Witham dataset.

sitive to the variation of classifier and to the parameter choice are the Fremen and the GMM models.

Following these results, we will use the best performing cases to compare the models. The parameters used are the ones shown in Table 8.

6.2 Model Comparison

Our experiments showed that the use of incrementally learnt models for spatial and temporal context can actually improve the performances of an Activity Recognition system. In Figure 17, it can be seen that all the temporal models could improve the classification results, without much difference in the results reducing the error rate to the half. It is interesting to notice how the Location-based model on the Aruba dataset reduced them, while on the Witham dataset it outperformed all the temporal models. This might depend on the fact that the association activity-location has a higher correlation in an office-like environment rather than in an home-like one, requiring lower accuracy for the base classifier to learn the context of the activities. Furthermore, we can observe that the Static component of FreMEn is improving, but only slightly compared to the other models, showing the need of having higher frequencies in weak base-classifiers. Figure 18 shows how the Interval Models tend to fail in adapting to the temporal contest, especially without the adaptive intervals, being unable to improve the results in this case. As in the previous case, the Location-based model works only on the Witham dataset. The remaining models are able to keep the error rate down to the half again. Finally, Figure 19 shows how a realistic base-classifier would benefit by the contextual prior probabilities learning. With the only exception of the static Interval-based models, which only worsen the performances in this case.

The results show that, using the right model and parameters, the error rate can be halved in less than two weeks, as can be seen in Figures 17, 18 and 19.

The models that produced the most reliable results were the GMM and Fremen, which had similar performances in reduction of the error rates and stability to the choice of the parameter. The only real difference lies in the fact that the GMM starts to reduce the errors right from the beginning, while Fremen tends to increase the errors, creating pronounced spikes in the error rate during the early days of execution. The Interval-based Models can actually perform an improvement comparable to the aforementioned models, in case of a weak classifier (Figure 17), while they appear to worsen performance if the classifier is a strong one (Figures 18, 19). This might be caused by the lack of sufficient evidence during the estimation of the probability priors when the confidence of the classifier is high. The latter can be demonstrated by the fact that the adaptation of the intervals according to the actual evidence does actually benefit the model behaviour, reaching performances similar to the GMM and Fremen in most cases.

The Location-Based probability priors had discordant results on the two datasets. In the Aruba dataset, it had a worsening effect on the error rate of the classification, although it improves when a strong classifier is used. This could mean that the model requires high base accuracy in complex indoor environments, in which the activities do not have a direct association to the place where they happen. On the Witham dataset instead, it did not only improve performance, but also outperformed all the other temporal prior models. This depends directly on the high association of the activities performed with places in office environments; for example, the activity of writing on the keyboard will always be performed close to the workplace.

7 Conclusion

This paper presented a novel approach to activity recognition for indoor environments based on incremental modelling of long-term spatial and temporal context. The presented approach allows to integrate several observations of the same environment in spatial and temporal models that captures the periodic behaviour of the activity occurrences and uses this knowledge to construct time and location dependent probability priors to improve the recognition of the activities. In other words, given the assumption of spatial and temporal structure of the activities, we have tried to learn those patterns to improve the performance of a base classifier with different models. Among those, the novel Interval-based model with Adaptive intervals has been in-

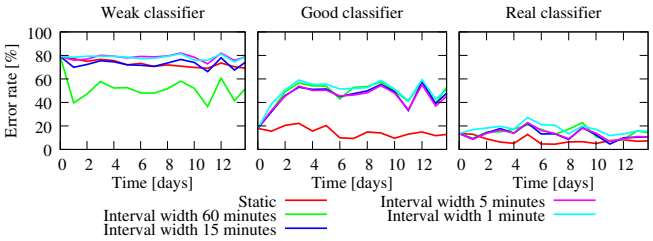

Figure 13. Impact of the interval width on the performance of the Interval-based models - Aruba dataset.

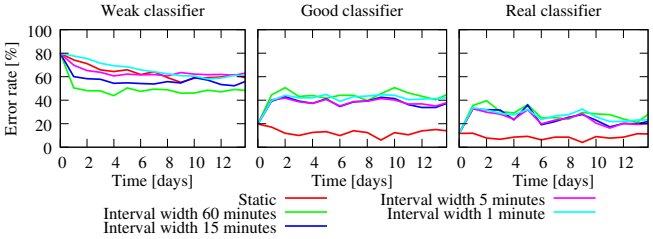

Figure 14. Impact of the interval width on the performance of the Interval-based models - Witham dataset.

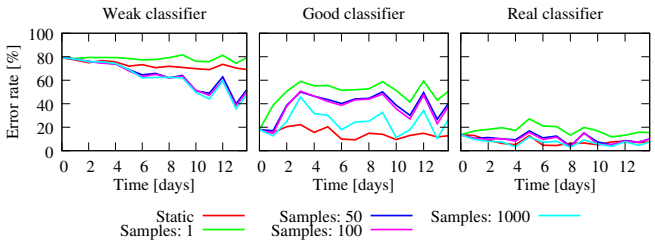

Figure 15. Impact of the number of samples used for prior estimation on the performance of the Adaptive-interval models - Aruba dataset.

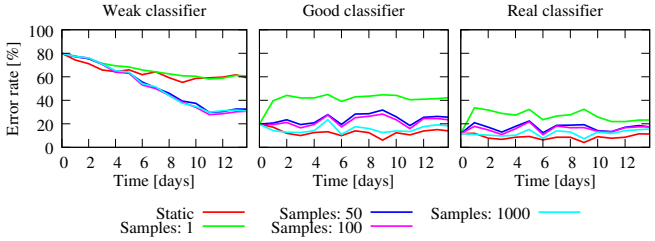

Figure 16. Impact of the number of samples used for prior estimation on the performance of the Adaptive-interval models - Witham dataset.

troduced, giving encouraging results. All of the models were benchmarked on two datasets, representing home and office environments, to show which models perform better in learning the temporal context of the activities, reducing in the best cases the error rate in time by half. Furthermore, an example of a location-based model has been introduced. This achieved an improvement in performances, but only in the office environment, due to the high correlation of typical office activities with their location. All of the models have been shown to be able to learn the recurrent patterns of the activities, even in cases of very weak base classification systems. Possible future works will include the merging of spatial and temporal models. A possibility could be applying a different temporal model in each spatial element of the environment. This could also require more data for a complete modelling of the contextual information, which could be overcome with an adaptive behaviour between the spatial units, like the one applied in the Adaptive intervals. The results are encouraging, although they are still applied to data delivered densely every minute. Future works will need to deal with data sparsity so that the model can be built on a mobile robot, which is not able to collect the activity data densely.

Acknowledgments

The work was supported by the EU ICT project 600623 'STRANDS' and by the European (H2020-PHC) project ENRICHME

REFERENCES

[1] H. Alerndar, H. Ertan, O.D. Incel, and C. Ersoy, 'Aras human activity datasets in multiple homes with multiple residents', in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*, pp. 232–235, (May 2013).

[2] Ulf Blanke and Bernt Schiele, 'Daily routine recognition through activity spotting', in *Location and Context Awareness*, 192–206, Springer, (2009).

[3] Diane J Cook, 'Learning setting-generalized activity models for smart spaces', *IEEE Intelligent Systems*, **2010**(99), 1, (2010).

[4] D.J. Cook, N.C. Krishnan, and P. Rashidi, 'Activity discovery and activity recognition: A new partnership', *Cybernetics, IEEE Transactions on*, **43**(3), 820–828, (June 2013).

[5] Claudio Coppola, Oscar Martinez Mozos, Nicola Bellotto, et al., 'Applying a 3d qualitative trajectory calculus to human action recognition using depth cameras', in *Proceedings of Intelligent Robots and Systems Workshops (IROSW 2015), 2015 IEEE/RSJ International Conference on*. IEEE, (2015).

[6] Labiba Gillani Fahad, Arshad Ali, and Muttukrishnan Rajarajan, 'Long term analysis of daily activities in smart home', in *Proc. of the European Symp. on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 419–424, (2013).

[7] Labiba Gillani Fahad, Asifullah Khan, and Muttukrishnan Rajarajan, 'Activity recognition in smart homes with self verification of assignments', *Neurocomputing*, **149**, 1286–1298, (2015).

[8] Hongqing Fang, Lei He, Hao Si, Peng Liu, and Xiaolei Xie, 'Human activity recognition based on feature selection in smart home using back-propagation algorithm', *ISA transactions*, **53**(5), 1629–1638, (2014).

[9] Diego R. Faria, Cristiano Premebida, and Urbano Nunes, 'A probalistic approach for human everyday activities recognition using body motion from RGB-D images', in *IEEE RO-MAN'14*, (2014).

[10] Diego R. Faria, Mario Vieira, Cristiano Premebida, and Urbano Nunes, 'Probabilistic human daily activity recognition towards robot-assisted living', in *IEEE RO-MAN'15: IEEE Int. Symposium on Robot and Human Interactive Communication. Kobe, Japan.*, (2015).

[11] Jaime Pulido Fentanes, Bruno Lacerda, Tomáš Krajník, Nick Hawes, and Marc Hanheide, 'Now or later? predicting and maximising success of navigation actions from long-term experience', in *International Conference on Robotics and Automation (ICRA)*, (2015).

[12] A. Fleury, N. Noury, and M. Vacher, 'Introducing knowledge in the process of supervised classification of activities of daily living in health smart homes', in *e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference on*, pp. 322–329, (July 2010).

[13] Shian-Ru Ke, Hoang Le Uyen Thuc, Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, and Kyoung-Ho Choi, 'A review on video-based human activity recognition', *Computers*, **2**(2), 88–131, (2013).

[14] Hema S Koppula, Rudhir Gupta, and Ashutosh Saxena, 'Learning human activities and object affordances from RGB-D videos', in *IJRR journal*, (2012).

[15] Tomáš Krajník et al., 'Long-term topological localization for service robots in dynamic environments using spectral maps', in *International Conference on Intelligent Robots and Systems (IROS)*, (2014).

[16] Tomas Krajnik, Jaime P Fentanes, Oscar Martinez Mozos, Tom Duckett, Johan Ekekrantz, and Marc Hanheide, 'Long-term topological localisation for service robots in dynamic environments using spectral maps', in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pp. 4537–4542. IEEE, (2014).

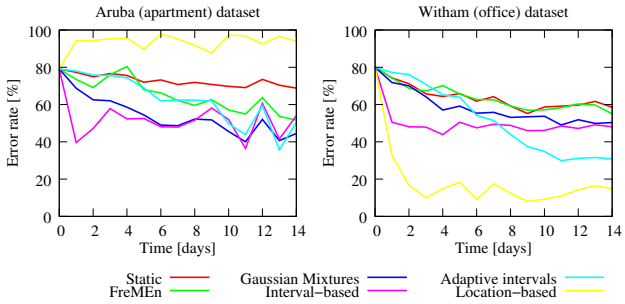

Figure 17. The impact of various spatial and temporal priors on the activity recognition error over time - weak classifier with 80% classification error.

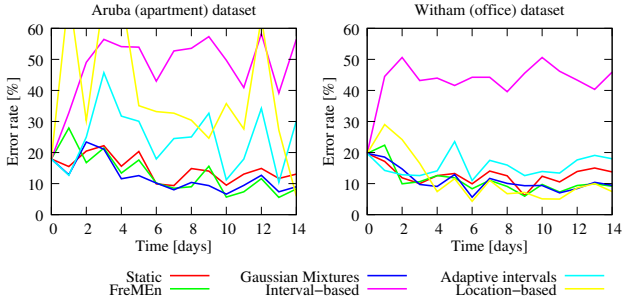

Figure 18. The impact of various spatial and temporal priors on the activity recognition error over time - good classifier with 20% classification error.

[17] Tomáš Krajník, Jaime Pulido Fentanes, Grzegorz Cielniak, Christian Dondrup, and Tom Duckett, 'Spectral analysis for long-term robotic mapping', in *International Conference on Robotics and Automation (ICRA)*, (2014).

[18] Tomáš Krajník, Joao Santos, and Tom Duckett, 'Life-long spatio-temporal exploration of dynamic environments', in *ECMR*, (2015).

[19] Tomáš Krajník, João Santos, Bianca Seemann, and Tom Duckett, 'FrOctomap: An efficient spatio-temporal environment representation', in *Proceedings of Towards Autonomous Robotic Systems (TAROS)*, (2014).

[20] Ivan Lillo, Alvaro Soto, and Juan Niebles, 'Discriminative hierarchical modeling of spatio-temporally composable human activities', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 812–819, (2014).

[21] Beth Logan, Jennifer Healey, Matthai Philipose, Emmanuel Munguia Tapia, and Stephen Intille, *A long-term evaluation of sensing modalities for activity recognition*, Springer, 2007.

[22] Bryan Minor, Janardhan Rao Doppa, and Diane J Cook, 'Data-driven activity prediction: Algorithms, evaluation methodology, and applications', in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 805–814. ACM, (2015).

[23] Tuan Anh Nguyen, Andrea Raspitzu, and Marco Aiello, 'Ontology-based office activity recognition with applications for energy savings', *Journal of Ambient Intelligence and Humanized Computing*, **5**(5), 667–681, (2014).

[24] Nuria Oliver, Eric Horvitz, and Ashutosh Garg, 'Layered representations for human activity recognition', in *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on*, pp. 3–8. IEEE, (2002).

[25] GI Parisi, C Weber, and S Wermter, 'Self-organizing neural integration of pose-motion features for human action recognition', *Name: Frontiers in Neurorobotics*, **9**(3), (2015).

[26] Lasitha Piyathilaka and Sarath Kodagoda, 'Human activity recognition for domestic robots', in *Field and Service Robotics*. Springer, (2015).

[27] Ronald Poppe, 'A survey on vision-based human action recognition', *Image and vision computing*, **28**(6), 976–990, (2010).

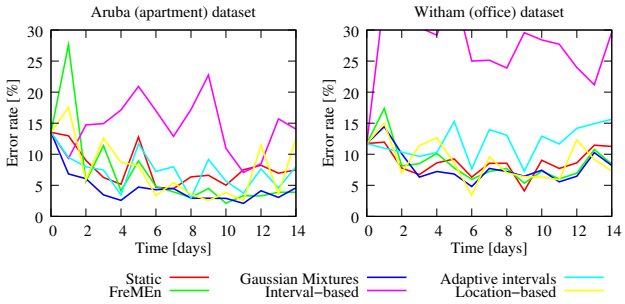

Figure 19. The impact of various spatial and temporal priors on the activity recognition error over time - real classifiers with ∼10% classification error.

[28] Jaeyong Sung, Colin Ponce, Bart Selman, and Ashutosh Saxena, 'Human activity detection from rgbd images.', *plan, activity, and intent recognition*, **64**, (2011).

[29] Jaeyong Sung, Colin Ponce, Bart Selman, and Ashutosh Saxena, 'Unstructured human activity detection from rgbd images', in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pp. 842–849. IEEE, (2012).

[30] NK Suryadevara, Subhas C Mukhopadhyay, R Wang, and RK Rayudu, 'Forecasting the behavior of an elderly using wireless sensors data in a smart home', *Engineering Applications of Artificial Intelligence*, **26**(10), 2641–2652, (2013).

[31] D Michael Titterington, Adrian FM Smith, Udi E Makov, et al., *Statistical analysis of finite mixture distributions*, volume 7, Wiley New York, 1985.

[32] Kristof Van Laerhoven, David Kilian, and Bernt Schiele, 'Using rhythm awareness in long-term activity recognition', in *Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on*, pp. 63–66. IEEE, (2008).

[33] Jiang Wang, Zicheng Liu, Jan Chorowski, Zhuoyuan Chen, and Ying Wu, 'Robust 3D action recognition with random occupancy patterns', in *European Conference on Computer Vision (ECCV)*, (2012).

[34] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan, 'Learning actionlet ensemble for 3D human action recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**(5), 914–927, (2014).

[35] Keze Wang, Xiaolong Wang, Liang Lin, Meng Wang, and Wangmeng Zuo, '3d human activity recognition with reconfigurable convolutional neural networks', in *Proceedings of the ACM International Conference on Multimedia*, pp. 97–106. ACM, (2014).

[36] Jiahui Wen, Mingyang Zhong, and Zhiying Wang, 'Activity recognition with weighted frequent patterns mining in smart environments', *Expert Systems with Applications*, **42**(17), 6423–6432, (2015).

[37] Christian Wojek, Kai Nickel, and Rainer Stiefelhagen, 'Activity recognition and room-level tracking in an office environment', in *Multisensor Fusion and Integration for Intelligent Systems, 2006 IEEE International Conference on*, pp. 25–30. IEEE, (2006).