Activity Recognition in Sensor Data Streams for Active and Assisted Living Environments

Fadi Al Machot, Ahmad Haj Mosa, Mouhannad Ali, and Kyandoghere Kyamakya

***Abstract*— In active and assisted living environments, a major service that can be provided is the automated assessment of elderly people's well-being. Therefore, activity recognition is required to detect what types of help disabled persons need to support them in their daily life activities. Unfortunately, it is still a difficult task to estimate the size of the required window for online sensor data streams to recognize a specific activity, especially when new sensor events are recorded. This paper proposes a windowing algorithm, which presents promising results to recognize complex human activities for multi-resident homes. The approach is based on the analysis of the sensor data to identify the best fitting sensors that should be considered in a specified window. Moreover, the second part of this paper proposes a set of different statistical spatio-temporal features to recognize human activities. In order to check the overall performance, this approach is tested using the CASAS data set and artificially generated laboratory data using our HBMS simulator. The results show high performance based on different evaluation metrics compared to other approaches. We believe that the proposed windowing approach provides a good approximation of the required window size in order to robustly detect human activities in comparison to other windowing approaches.**

***Index Terms*— Classification, support vector machines (SVMs), activity recognition, windowing, active and assisted living environments (AAL), smart homes.**

I. Introduction

In RECENT years, advances in machine learning and pervasive computing technologies have opened the doors for researchers to explore human behaviors in a more advanced manner. Smart home environments are one of such application domains where human activities are monitored by modern technologies to assist the residents when required. Hensel *et al.* [1] define smart home as a residence which is equipped with technology to provide safety of residents and their well-being. Modern technologies have provided state of the art solutions to enhance the quality of independent living for old or disabled people. Especially the smart homes do enable this by providing a platform to deploy ambient assisted living systems.

Generally, smart home environments are equipped with sensors which are localized inside a resident home. Thus, the behavior of residents and their interaction with the environment is observed by those sensors. An action is considered as a sequence of movements (simple events) performed by a human agent during the performance of a task. Consequently, activity recognition is a process of identifying those movements (simple events) and combining them using reasoning or machine learning models. The main purpose of activity recognition is to identify the most suitable activity label (class), even when such an activity is performed by a different agent under different circumstances regardless of the environment where the activity has been performed.

There are many stages involved in the process of recognizing human activities. In general, this process starts from collecting activity information based on raw sensor data and ends with the recognized activity. Initially, the collected sensor data pass through a data cleaning and preparation phase to remove the noise and unnecessary data from the raw sensor stream. Then, the preprocessed data pass through a windowing phase to extract features out of the window. Then, the feature extraction phase is followed by a dimensionality reduction phase to increase the accuracy of the extracted features and reduce the computational effort needed for the classification. Finally, the selected feature vector is used as a classification model input for classification and recognition purposes.

Designing such an activity recognition system is challenging due to the spontaneous nature of human activities and because of the problem of estimating the correct required dataset window, in particular, when new sensor events are recorded in real life scenarios. Additionally, capturing every element of the human behavior is important. Therefore, various types of sensors are used to extract different types of low level features from the environment.

In this work, we consider the CASAS project dataset [2] and the HBMS dataset [3]. The HBMS dataset is an artificially created dataset using HBMS simulation tool and the CASAS dataset comes from real life scenarios to support residents in a smart home environment to receive the status of the inhabitants. Our aim is to support old persons in a given AAL scenario by a robust windowing and activity recognition approach which performs well even for multi-resident homes compared to other state of the art approaches.

The paper is organized as follows: Section 2 gives an overview of the state-of-the-art approaches. Section 3 defines the terminology and the considered problem. Section 4 presents the overall architecture of the proposed windowing and activity recognition system. Section 5 shows the results and the overall performance evaluation. Section 6 discusses

Manuscript received January 13, 2017; revised June 18, 2017 and July 28, 2017; accepted October 6, 2017. Date of publication October 20, 2017; date of current version October 24, 2018. This paper was recommended by Associate Editor F. Al Machot. *(Corresponding author: Fadi Al Machot.)*

The authors are with Alpen-Adria-Universität, 9020 Klagenfurt, Austria (e-mail: fadi.almachot@aau.at; ahmad.hajmosa@aau.at; mouhannad.ali@aau.at; kyandoghere.kyamakya@aau.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2017.2764868

the obtained results and compares them to other approaches. The paper ends in Section 7 with a conclusion.

II. Related Works

During the last decades, different approaches for human activity recognition have been reported. They can be classified into 3 major categories along their underlying model types: logic-based context models, graphical models, and syntactic models. In this section, we review the state-of-the-art on activity recognition systems and windowing techniques to recognize activities in sensor data streams.

A. Knowledge-Based Approaches

In logic-based context models, a context is defined as using facts (context properties) with expressions and rules to describe and define relationships and constraints [4]. There are different logic-based frameworks for behavior recognition depending on sensors which coordinate the concurrent execution of the available movements (simple events) so as to mimic the target behavior [5].

Logic-based methods for activity recognition have been proposed to represent domain knowledge and model each event. In these approaches, an event is generally specified as a set of logical rules that enables them to be recognized by using logical inference techniques such as resolution or abduction [6]. In particular, Shed *et al.* [7] proposed a framework that combines computer vision algorithms with logic programming to represent and recognize activities in a parking lot in the domain of video surveillance.

Other event activity recognition systems are presented by [5] and [8], both systems are based on a logic programming implementation of an Event Calculus [9]. The Event Calculus is a set of first-order predicate calculus including temporal formalism for representing and reasoning events and their effects. These approaches do not consider the problems of noise or missing observations that always exist in real world applications. To cope with these issues, some extensions to logic approaches have been presented. Moreover, a domain knowledge system is described by [10] as first-order logic production rules with associated weights to indicate their confidence.

Additionally, Ontology Web Language (OWL), Semantic Web Rule Language (SWRL) and SPARQL Inferencing Notation (SPIN) have been proposed by the World Wide Web Consortium (W3C) as language standards for representing ontologies and rules, respectively. SPARQL Protocol and RDF Query Language (SPARQL) has been approved by W3C recommendation as query language for the Semantic Web technologies. An overview of such languages is presented in [11]. These languages enable autonomic agents to reason about complex events.

Activity recognition based on the Ontology Web Language (OWL 2) is still a major research area [12]–[14] and fortunately, OWL 2 ontologies are supported by a fuzzy logic-based reasoner to handle uncertainty.

Furthermore, a well-known logic-based approach is Answer Set Programming (ASP) which supports a number of arithmetic functions that are evaluated during grounding. Unfortunately, little research has been done within the frame of using Answer Set Programming (ASP) in AAL applications [15]–[17].

B. Graphical Models

Graphical models are used to model complex scenes or multimodal sensors data because of the characteristics of the inherent structure and semantics of complex activities that require higher level representation and reasoning methods such as Bayesian propagation networks (BNs) [18], [19], Dynamic Bayesian Networks [20], Hidden Markov Models [21], [22], Dempster-Shafer [23], [24], Conditional Random Fields (CRFs) [25] and Gaussian Mixture Models (GMM) [26].

An extensive research has been done within the frame of Bayesian networks to detect simple and complex events. For example, BNs are used to detect events and analysis pattern of soccer games [27]. Dynamic Bayesian Network (DBN) is used to characterize the spatio-temporal nature of the semantic objects [28]. Moreover, many papers have been published that consider Bayesian Networks to detect abnormal events in smart homes [29].

A hybrid approach which combines Bayesian Networks and Hidden Markov Models also proposed recently to model temporal inter-actions for complex event detection [21]. Additionally, the Gaussian Mixture Models (GMM) that are considered as special cases of Bayesian Networks are used to recognize static postures and non-temporal event patterns, but they do not perform in real time if the dimensionality of the problem is too high. Therefore, they are not suitable for spatio-temporal reasoning [30].

Furthermore, Conditional Random Fields (CRFs) are a probabilistic framework for labeling and segmenting structured data. It defines a conditional probability distribution over specific sequences given rather than a joint distribution over both label and observation sequences. Many researchers have used CRFs for activity recognition in smart homes. They used conditional random fields (CRFs) to model the spatio-temporal relations for human activity recognition using CASAS dataset [2], [31].

Moreover, Hierarchical HMMs have also been used for activity recognition by [32] and [33]. Originally, the hierarchical HMMs were restricted to a tree structure, meaning that two different states cannot have the same child and share common substructures. In contrast to that, dynamically multi-linked HMMs perform a better model group activity in noisy outdoor scenes [34]. In decomposed HMMs, the production layer consists of several independent state channels with several abstract layers to represent long-term interacting activities.

C. Syntactic Approaches

Syntactic approaches are used to express the structure of a process by using a set of production rules to describe the real world events, e.g., rough set theory [35] and fuzzy logic [36]. Fuzzy Logic offers cheap computation time to solve very complicated problems in AAL [37], e.g., [51] proposed a behavior-modulation technique using a fuzzy discrete event system (FDES) for behavior-based robotic control. The method exploits the different features of fuzzy logic and event-driven properties of a Discrete Event System (DES) to detect human behavior through the use of fuzzy state vectors [38].

Furthermore, Subramanian and Suresh [39] proposed a neuro fuzzy approach to recognize human actions in videos. Their approach is basically based on McFIS classifier and its sequential learning algorithm using the principle of self-regulation observed in human meta-cognition data.

In order to perform event detection successfully, in case of fusing sensor data that does not require preliminary or additional information such as data distribution or membership function, rough set theory is suitable [40]. Unfortunately, only very little research has been done using rough set to detect complex events for human activity recognition.

D. Deep Learning

Considering human activity recognition approaches that are based on deep learning, different approaches have appeared recently which consist of the following architecture: (a) data pre- processing, (b) convolutional architecture for activity recognition, and (c) fusion strategies for the final classification [41]–[43].

Additionally, deep learning models have been applied using smartphones for human activity recognition; e.g., a deep convolutional neural network (convnet) has been proposed by exploiting the inherent characteristics of activities and 1D time-series signals [44].

Other works like in [45] proposed a technique based on a deep learning methodology to offer real-time activity recognition to be applied on limited power devices. The proposed approach considers the invariance against changes in sensor orientation, sensor placement, and sensor acquisition rates. Moreover, a human activity recognition model based on RFID and deep convolutional neural network is introduced by [46]. They feed the RFID data into a deep convolutional neural network for activity recognition instead of selecting features.

Furthermore, in [47], a 3D human activity recognition model has been proposed based on Convolutional Neural Networks using 3D depth sensors. The proposed approach aims to find out the temporal structures for every activity by representing each activity as an ensemble of cubic-like video segments. Another interesting work is ADORE (Adaptive Holons Representation) [48] which combines the advantages of local and global cues for human pose estimation. The proposed approach first infers the location of the joints on the global level using the ILPNs concept (Independent Losses Pose Nets), and then applies Convolutional Local Detectors (CLDs) to detect the joints in the potential regions generated by ILPN adaptively.

E. Windowing Approaches

In order to provide robust activity related information for real-life applications, researchers suggested different approaches to recognize human activities in smart homes. In literature, there are three major approaches that deal with the windowing of streaming data. This can be summarized as follows:

Time periods of equal length: This approach offers a simpler method to learn the activity models during the training phase. On the other hand, some activities might spread over more than one time slice [49], [50].

Chunks of equal number of sensors: This approach has the advantage that the resulting windows have a varying duration. However then, there might be periods in which not enough sensors are activated [51].

Probabilistic dynamic windowing: This approach uses a probabilistic approach which maximizes the probability of the most likely window size for a specific activity. The idea is to incorporate the time decay and mutual information using weightings of sensor events within a window [52]. A limitation of such approaches is its inefficiency in modeling complex activities; where many similar sensors are shared.

Other researchers proposed an offline phase to preprocess sensor events before the classification phase. For example in [53], an approach to monitor the change in daily routine of a person is proposed by analyzing residents' behavior over a long term. They identified specific activities as a group that can be performed during a day and monitor such behavior over a period. The goal was to separate the normal routine from unusual and suspicious routines.

Moreover, in [54] they considered the activity as a set of active sensors related data, which are involved in the performance of a certain task in a smart home environment. In the proposed approach, they have carried out a kernel fusion method for accurate activity recognition. Then, they further classified the significant sequential behaviors from the recognized activities. As a result, the behavior patterns are further used to predict the future actions from past activities.

In addition, in [55], they defined a sliding window which encodes in the comparison of the actual observation the mean and standard deviation computed in the sliding window.

III. Terminology and Problem Definition

This paper does address the problem of human activity recognition with an online windowing approach whenever new sensor events are recorded in the AAL environment. The human activities we discuss here are the well-known activities that can be performed at home. For example, preparing a meal or a drink, washing dishes, cleaning, going to bed, etc.

The whole scenario consists of two phases. The first phase is the offline phase where the annotated training data are considered to select the sets of best fitting sensors for each activity. The second phase is the online phase which aims at building the observation window by collecting sensor data in a window until all sensors of a set of the best fitting sensors are activated. After the previous steps in the online phase, the features extraction step is applied on the observation window which is followed by the multiclass classification model that provides the class of the performed activity. Figure 1 shows the overall architecture of the proposed windowing and activity recognition approach.

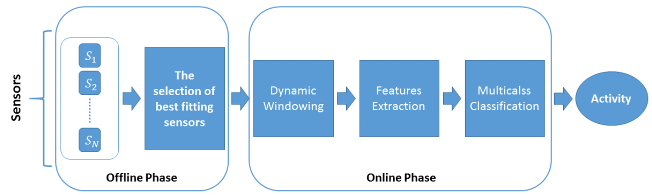


Fig. 1. The overall architecture of the proposed system.

The best fitting sensors for each activity will be determined in the offline phase, and they will be used in the online phase as reference sensors to build the estimated window when the person is acting in his/her smart environment.

A. Terminology

1) Annotated Training Dataset: A training set is the data set used to train the classification model. In this paper, there are two kinds of training sets: the initial training to determine the best fitting sensors and the classification training set which is used to train the multiclass classifier.

2) Training Sensors: Training sensors are the sensors that belong to the annotated training dataset of a specific activity. They are used to select the best fitting sensors for each activity out of them (offline). We denote them by S_{tr_i} where $i = 1, 2, 3, \ldots Num$ and Num is the number of sensors of the observed activity.

3) Observation Window: A window where sensors' events are added into it after using the proposed windowing algorithm (see Algorithm 1). We denote it by w. For example, if we say the observation window length is 9, it means the data window contains 9 data samples. We denote an observation window by $S = S_1, S_2, \ldots, S_N$ where $i = 1, 2, 3, \ldots N$ and N is the observation window length.

4) Sensor Event: A sensor event consists of the sensors' readings at a single time. It is the unit data that is sent to the activity recognition system to be added into the observation window when the user acts in the AAL environment (online). We denote it by S.

5) Best Fitting Sensors: They are the sets of the best fitting sensors which provide the highest Information Gain for each activity [56]. This is performed in the offline phase. We represent these sets as a matrix which is denoted by BestSensors. The BestSensors matrix has two dimension: its first dimension represents the list of activities (IDs) and the second dimension represent the list of sensors (IDs).

6) Best Fitting Sensors Reference: It is the reference matrix which is used to determine the final observation window in the online phase. We denote it by tmpBestSensors. This matrix is a zeros matrix which has the same dimensions as BestSensors. It is a reference window which lets the system decides where to stop adding new sensors' events into the observation window.

7) Feature Vector: A feature vector is a vector that describes the characteristics of each observation window.

8) Occurrences Histogram: The occurrences histogram consists of two dimensions. The first dimension is the sensors' IDs, and its second dimension consists of the number of activations of each sensor within the observation window w.

B. Problem Definition

The recognition of human activity is very complex due to the following facts [57]: (a) the selection of the attributes, (b) the design of feature extraction and inference methods, (c) challenges concerning the context awareness that involve knowledge discovery, since the raw sensor records deliver a lot of useless information, and (d) a multi-resident activity recognition where different persons might have different patterns to perform the same activity. Therefore, the offline phase is proposed to overcome the challenges (a) and (b), as it selects the best fitting sensors that help to provide meaningful features out of them whenever they might be found in the observed window, which keeps the extracted features related to the correct context. Additionally, regarding points (c) and (d), the offline phase aims at finding the best fitting sensors with respect to different residents based on the "Info Gain" which represents how far a given attribute of the features' vectors is not dependent of different residents or subjects. In this way, the universality of the model is increased as this is confirmed in Section 5.

Furthermore, the online phase aims to estimate the most accurate window when the inference starts. The problem here is that neither a time-based windowing nor considering a specific set of sensors (chunks) can determine the best window for subject-independent activity recognition. Some persons might behave slower than others or perform the same activity differently [52]. Thus, considering the context of the activity while considering only a single person leads to a better estimation. In contrast to this, a probabilistic or statistic-based window estimation may work well to recognize the activities of a single resident but could fail to estimate the correct model parameter patterns for multi-residents.

IV. The Proposed Human Activity Recognition Approach

A. Offline Phase

In order to determine the best fitting sensors matrix, we extracted several features of the given annotated training dataset.[1] The impact of these features is analyzed by using the Information Gain (IG) attribute evaluation [56]. The results showed that the number of activations for each sensor is the most relevant attribute. The overall general steps can be summarized in Figure 2. The output would be the matrix of best fitting sensors for each activity $BestSensors$.

The overall general steps are summarized as follows: for the given annotated training dataset, the most relevant feature should be selected which can be a spatial, temporal or a causal feature. Then, sensors should be sorted based on the selected attribute in a descending order. After that, for every activity, the first $1..N$ sensors will be selected iteratively whereby after each iteration the features quality should be checked, e.g.,

[1] Number of activations of each sensor, activation duration of each sensor, number of activated sensors for each activity, and the location of the sensor

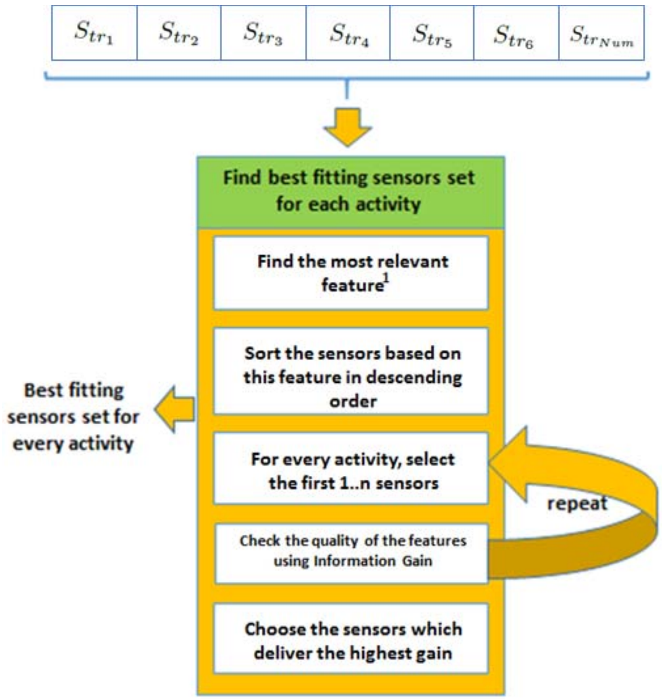


Fig. 2. The steps to find the best fitting sensors in the (offline phase).

Information Gain (IG) Attribute Evaluation. In this step, the training sensors are used (see III-A). Finally, sensors which deliver the highest gain should be selected. The information gain of an attribute measures the difference in information between the cases where the value of an attribute is known and where it is unknown.

The output will be the list of the best sensors for every activity. This list will be stored in the $BestSensors(Activities, Sensors)$, where $Activities$ is the number of activities (IDs) and $Sensors$ is the number of sensors (IDs).

B. Online Phase

To dynamically determine the observation from a sensor data stream, we use Algorithm 1. In this algorithm, the first input is the $BestSensors$ matrix. The second input is the $tmpBestSensors$ matrix. The algorithm starts with getting the next activated sensor from the sensor events stream ($SensorsStream.HasNext()$) by calling the function $GetNextActivatedSensor()$. The obtained sensor is added to the observation Window w which is initialized as an empty vector. Then, it iterates over the BestSensors matrix to check whether the activated best sensor is ON/active, and then the $tmpBestSensors$ is accumulated by 1. This process is repeated until all the best fitting sensors of a specific activity are activated in the environment, which means that the row of a specific activity in $tmpBestSensors$ does not contain zeros. The output is then the observation window w. This algorithm is applied on the given dataset by using the proposed features in Section IV-C.

Algorithm 1 shows how to create the final window where the sensor data should be collected until all the optimal sensors of an activity are activated. As soon as the observation window w has been determined, we create a two-dimensional occurrences histogram (see Figure 3).

Algorithm 1 Online Dynamic Algorithm - in the Online Phase

Input: $BestSensors, tmpBestSensors$
Output: $The\ estimated\ window\ w$

```
initialize w as an empty vector
while SensorsStream.HasNext() do
    s = GetNextActivatedSensor()
    add sensor s to the window w
    for each i in Activities do
        for each j in Sensors do
            if BestSensors(i, j) == s then
                tmpBestSensors(i, j) = 1
            end if
        end for
    end for
    for each p in Activities do
        if tmpBestSensors(p,AllSensors) has no zeros then
            return w
        end if
    end for
end while
```

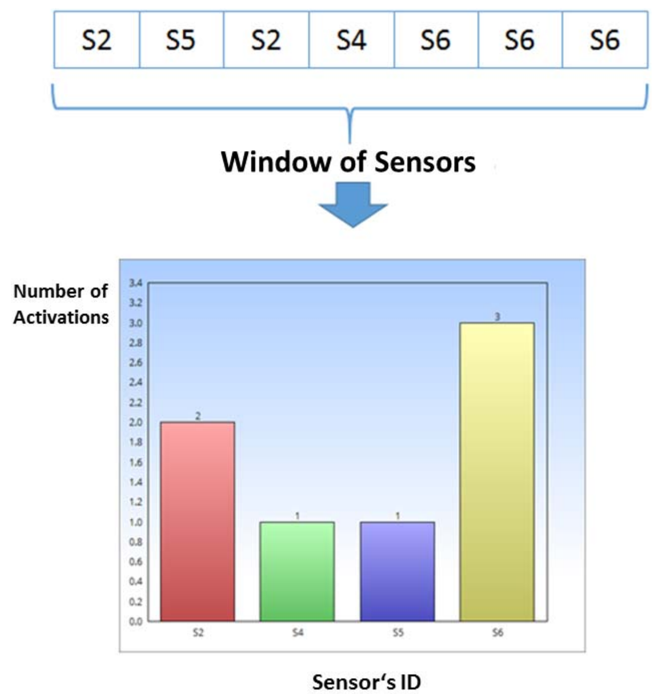


Fig. 3. An occurrences histogram example which shows that first dimension is the sensors' Ids, and its second dimension consists of the number of activations of each sensor.

C. Features Extraction

In this section, we introduce the set of specific spatio-temporal features that we use in our approach for calculation, evaluation and the final features vector composition.

1) Temporal Features: Human activity recognition systems consider two types of temporal features [58]:

- **Absolute temporal features** which are independent of other activities. They represent the duration of each activity individually,
- **Relative temporal features** that relate to other activities,

In Eq. 1, absTime is defined to be the sum of the durations of all activities observed where N is the total number of activities.

$$absTime(a_i) = \sum_{i=1}^{N} startTime(a_i) - endTime(a_i) \quad (1)$$

Relative temporal features are supposed to represent some useful information for classification [22]. In other words, they can be used for identifying relation patterns between activities using reasoning mechanisms. For example, a person might not have a routine time for taking a shower but might brush teeth after each meal. Eq. (2) defines the relative time (relTime) for an activity a_p where again $N = |V|$ is the total number of activities and V is the vector of activities. g is a function which gives 1 if two specific activities are performed after each other and 0 otherwise.

$$relTime(a_p, a_q) = \sum_{i=1}^{N} g_i(a_p, a_q) \quad (2)$$

$$g(a_p, a_q) = \begin{cases} 1, & \text{if } if\ \ V(i) = a_p,\ \ V(i+1) = a_q \\ 0, & \text{if } otherwise \end{cases}$$

2) Spatial Features: Using the occurrence histogram as defined in section IV-B (see Figure 3), we extracted the following arithmetic statistics [59]: the mean, the median, the range, the standard deviation, the skewness, the kurtosis and the empirical isotropic semi-variogram. The occurrence histogram o is the number of activations for each sensor s_p (see Eq. 3).

$$O(S_p) = \sum_{i=1}^{S} h_j(S_p) \quad (3)$$

$$h(S_p) = \begin{cases} 1, & \text{if } if\ \ S_p\ is\ On \\ 0, & \text{if } if\ \ S_p\ is\ Off \end{cases}$$

h is a function which gives 1 if the sensor s_p is on and 0 otherwise.

a) Mean: The arithmetic mean μ is the average value of the data, i.e., the sum of all occurrence histogram fields o_i divided by the total number of sensors S (i.e., the vector length) (see Eq. 4).

$$\mu = \frac{1}{S} \sum_{i=1}^{S} o_i \quad (4)$$

b) Median: The median is the middle value in a sorted list of data, in our case in the middle value of o (see Eq. 5).

$$median = \begin{cases} o_{(\frac{S+1}{2})}, & \text{if } if\ \ o\ is\ odd \\ \frac{1}{2}(o_{(\frac{S}{2})} + o_{(\frac{S}{2}+1)}), & \text{if } if\ \ o\ is\ even \end{cases} \quad (5)$$

c) Range: The range is the difference between the largest value and the smallest value of a dataset. In our case, the difference between the largest value and the smallest value of the given occurrences histogram (see Eq. 6).

$$o_{range} = o_{max} - o_{min} \quad (6)$$

d) Standard deviation: The standard deviation is the square root of the variance. The variance is the sum of all squared differences from each occurrence value to the mean divided by the number of sensors S minus, (see Eq. 7).

$$\sigma^2 = \frac{1}{S-1} \sum_{i=1}^{S} (O_i - \mu)^2 \quad (7)$$

e) Skewness: The skewness λ is defined as the quotient of the third central moment m_3 of a dataset, and the cubed standard deviation, (see Eq. 8).

$$\lambda = \frac{m_3}{\sigma^3} = \frac{\frac{1}{S}\sum_{i=1}^{S}(O_i - \mu)^3}{(\sqrt{\frac{1}{S}\sum_{i=1}^{S}((O_i - \mu))})^3} \quad (8)$$

f) Kurtosis: The kurtosis $\lambda'$is defined as the quotient of the fourth central moment of a dataset m_4, and the standard deviation σ power 4, (see Eq. 9).

$$\lambda' = \frac{m_4}{\sigma^4} = \frac{\frac{1}{S}\sum_{i=1}^{S}(O_i - \mu)^4}{(\sqrt{\frac{1}{S}\sum_{i=1}^{S}((O_i - \mu))})^4} \quad (9)$$

g) Empirical isotropic semi-variogram: Semi-variograms describe the roughness of spatial data. An elementary characteristic of spatial data is spatial autocorrelation. Spatial autocorrelation presumes that spatial data values nearby similar compared to spatial data values having a greater distance. Such a spatial autocorrelation is expressed by a semi-variogram function. It contains one half of the variance of differences between data pair values.

In our smart home datasets, we calculate the empirical isotropic semi-variogram which is given by Eq. 10), where var is the variance. If two occurrences values, s_i and s_j, are close to each other in terms of the euclidean distance measure of $d(si, sj)$, we expect them to be similar. Thus, the difference in their values, $o(s_i){-}o(s_j)$, will be small and larger otherwise.

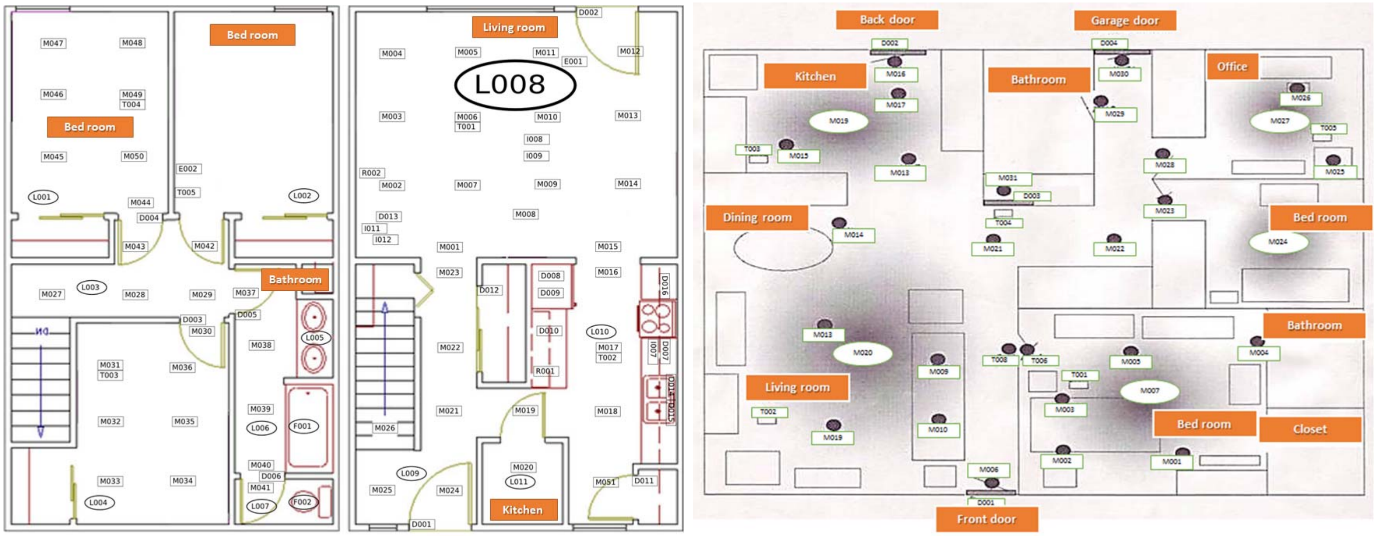
$$\lambda(S_i, S_j) = \frac{1}{2} var(o(S_i) - o(S_j)) \quad (10)$$

The final feature vector consists of the occurrences histogram and the previous mentioned features.

D. Multiclass Classification

Typically, AAL datasets have multiple activities (classes). Thus, we have a multiclass classification problem which is a major problem in machine learning. The most common approach to multiclass classification is the one-versus-all approach which makes direct use of standard binary classifiers to encode and train the output labels. It assumes that for each class there exists a single separator between that class and all the other classes.

Additionally, another common approach is all-versus-all approach which assumes the existence of a separator between



(a) (b)

Fig. 4. The overall environmental settings (smart homes) that are used for our evaluation, where M means motion sensors, I item sensor (kitchen items), D means door sensor, T means temperature sensor, P means electricity usage and AD1-A, AD1-B, AD1-C mean burner, cold water and hot water sensors respectively. (a) The smart home of Towor CASAS dataset. (b) The smart home of Aruba CASAS dataset.

any two classes. In this paper, we choose the one-versus-all approach, advantage of which is that each class is represented by only one classifier. Thus, it is possible to obtain knowledge about a specific class by checking out its corresponding classifier.

Other authors also proposed methods that consider all classes at once, but thereby the corresponding complexity of the model increases and a better optimization approach is required [60]. However, as it is recommended in [61] one-versus-all still shows promising results compared to other approaches which tries to consider all classes based on Support Vector Machine (SVM).

The classification model used is SVM [62]–[64] to multi-class problems. Support Vector Machines have applications basically in pattern recognition, regression, prediction and feature selection. Therefore, they have become the method of choice to solve complex classification problems. This is due to the fact that (a) SVMs offer solutions for classification problems even in case of high-dimensional data and a small set of training samples, and b) they can efficiently classify non-linearly separable data using the "kernel trick". The two most commonly used kernel functions for real valued data are the polynomial and the Gaussian kernel.

The choice of a kernel depends on the problem at hand because it depends on what we are trying to model. A polynomial kernel, for example, allows us to model feature conjunctions up to the order of the polynomial. Radial basis functions allow to pick out circles (or hyperspheres) in contrast with the Linear kernel, which allows only to pick out lines (or hyperplanes). In this paper, we consider a polynomial kernel of degree 3 which delivers the highest performance.

For imbalanced datasets, it is typical to change the misclassification penalty per class. This is called class-weighted SVM [65]. The misclassification penalty for the minority class is chosen to be larger than that of the majority class. This has been considered during our one versus all classification with respect to the proportion between the minority and majority classes. Additionally, this extension has no real impact on the complexity of the problem since only the bounding box constraints does change thereby.

V. RESULTS OBTAINED

For activity recognition, an evaluation of features is applied to check their effectiveness in the intended classification task. For this purpose, we checked both the Information Gain (IG) Attribute Evaluation [56] and the Gain Ration (GR) Attribute Evaluation [66].

Moreover, we considered three smart home datasets: the Aruba and Towor CASAS dataset [2] and the HBMS[2] laboratory dataset. Figure 4 shows the smart homes that are used for our evaluation.

Additionally, for multiclass classification, SVMs is used based on the $10-fold$ cross-validation approach. Therefore, the dataset is randomly divided into ten equal-sized slices. Each slice is used as a test set and a training set. The test results are then averaged over the ten cases. For evaluation, the WEKA library is used, which is a machine learning library based on JAVA [67].

Furthermore, the following evaluation metrics have been used for the overall evaluation.

- TP is true positives,
- FP is false positives,
- Pr is the precision,
- Re is the recall,
- F-M is the F-Measure, and
- ROC is the Receiver-Operating-Characteristic Curve.

A. Aruba CASAS Dataset

Aruba CASAS dataset has been collected in a house which has a single bedroom, a kitchen, a bathroom, a dining room,

[2]hbms-ainf.aau.at/de

TABLE I
THE CLASSIFICATION PERFORMANCE USING ARUBA CASAS DATASET

Class	TP	FP	Pr	Re	F-M	ROC
C1	0.94	0.003	0.96	0.94	0.95	0.96
C2	0.54	0.001	0.71	0.54	0.61	0.77
C3	0.95	0.15	0.77	0.95	0.85	0.90
C4	0.93	0.05	0.93	0.93	0.93	0.94
C5	0.20	0.005	0.54	0.20	0.29	0.59
C6	0.48	0.002	0.85	0.48	0.62	0.74
C7	0.09	0.002	0.20	0.009	0.01	0.50
C8	0.80	0	0.66	0.80	0.72	0.90
C9	0.56	0	0.90	0.56	0.69	0.78
C10	0.85	0.01	0.83	0.85	0.84	0.92
Weighted AVR	0.85	0.07	0.82	0.85	0.82	0.88

TABLE II
THE CONFUSION MATRIX USING ARUBA CASAS DATASET

Class	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	896	11	2	39	3	1	0	0	0	0
C2	22	27	0	1	0	0	0	0	0	0
C3	3	0	3816	141	14	11	15	0	0	0
C4	6	0	270	4918	33	2	5	0	1	12
C5	4	0	117	140	68	3	4	0	0	0
C6	0	0	93	26	4	117	0	0	0	0
C7	0	0	655	11	2	3	6	0	0	0
C8	0	0	0	1	0	0	0	4	0	0
C9	0	0	4	1	0	0	0	2	9	0
C10	0	0	1	9	0	0	2	0	0	59

and an office. The home Aruba is "smart" as 34 sensors are installed to collect information such as door closure, motion and temperature. All activities are collected by a single inhabitant within the period from 2010-11-04 to 2011-06-11. The following activities are annotated in the dataset: Sleeping, Bed to Toilet, Meal Preparation, Dish Washing, Relaxing, House Keeping, Eating, Leave Home, Enter Home, and Work.

The quality of the spatial features using Info Gain and Gain Ratio for this dataset varied between 0.13 and 0.68. The closer the quality is to one, the better it is.

In Tables I and II the classes are:

- C1: Sleeping,
- C2: Go to Bed,
- C3: Meal Preparation,
- C4: Relaxing,
- C5: House Keeping,
- C6: Eating,
- C7: Wash Dishes,
- C8: Leave Home,
- C9: Enter Home, and
- C10: Work

Table I shows that the overall F-Measure value is 0.82 although there are bad cases, first worst class is C7 "Wash dishes" with value 0.01 and second worst class is C5 "House Keeping" with value 0.29. This can be explained by Table II which shows the confusion matrix for all ten classes. For example, the misclassification for C5 occurred due to the dominance of class C3 "Meal Preparation" and C4 "Relaxing." Moreover, the detection performance of C7 was very low due to the low number of recorded cases for this activity in the given dataset.

TABLE III
THE CLASSIFICATION PERFORMANCE USING TOWOR CASAS DATASET

Class	TP	FP	Pr	Re	F-M	ROC
C1	0.40	0.06	0.62	0.40	0.49	0.70
C2	0.95	0.04	0.91	0.95	0.93	0.95
C3	0.76	0.07	0.73	0.76	0.74	0.84
C4	0.98	0.001	0.96	0.98	0.97	0.99
C5	0.50	0.008	0.80	0.50	0.62	0.74
C6	0.71	0.06	0.73	0.71	0.72	0.82
C7	0.82	0.04	0.75	0.82	0.78	0.89
Weighted AVR	0.81	0.04	0.80	0.81	0.80	0.88

B. Towor CASAS Dataset

This data set represents time series information collected from sensors recording ADL of two residents R1 and R2 during the summer 2009. 51 motion sensors, five temperature sensors, fifteen door sensors, a burner sensor, hot and cold water sensor and an electric sensor were used. It consists of the following annotated activities: Cleaning, Meal-Preparation, R1-Bed-to-Toilet, R1-Personal-Hygiene, R1-Sleep, R1-Work, R2 Bed to Toilet, R2-Personal-Hygiene, R2-Sleep, R2-Work, Study, Wash-Bathtub and Watch-TV, where $R1$ refers to first resident and $R2$ refers to the second resident. The classes "Study" and "Wash-Bathtub" were excluded due to their low number of occurrences which make the training phase insufficient.

The quality of the spatial features using Info Gain and Gain Ratio for this dataset varied between 0.17 and 0.92. The closer the quality is to one, the better it is. For this dataset, we excluded the temperature sensors.

The classes in Tables III and IV are:

- C1: Cleaning,
- C2: Meal Preparation,
- C3: Bed to Toilet,
- C4: Sleeping,
- C5: Work,
- C6: Personal-Hygiene, and
- C7: Watch a TV.

Table III shows the overall F-Measure value is 0.80 although there are bad cases, the first worst is class C1 "Cleaning" with value 0.49; the second worst is class C5 "Work" with value 0.62. This can be explained by Table IV which shows the confusion matrix (see Table IV) for all seven classes. For example, the missclassifications occurred due to the dominance of class C2 "Meal Preparation" where most of other activities are recognized to be C2. Additionally, some missclassifications are

TABLE IV
THE CLASSIFICATION PERFORMANCE USING TOWOR CASAS DATASET

Class	C1	C2	C3	C4	C5	C6	C7
C1	15	9	1	0	0	0	12
C2	7	476	5	0	3	4	8
C3	0	0	229	1	0	71	0
C4	0	0	0	57	1	0	0
C5	1	11	0	1	44	0	30
C6	0	4	77	0	0	208	1
C7	1	21	1	0	10	0	157

TABLE V
THE CLASSIFICATION PERFORMANCE USING HBMS LABORATORY DATASET

Class	TP	FP	Pr	Re	F-M	ROC
C1	1	0.02	0.95	1	0.97	1
C2	0.74	0.04	0.74	0.74	0.74	0.86
C3	0.94	0	1	0.94	0.97	1
C4	1	0	1	1	1	1
C5	0.82	0.03	0.80	0.82	0.81	0.86
C6	0.93	0.004	0.96	0.93	0.95	0.99
Weighted AVR	0.92	0.01	0.91	0.92	0.91	0.95

occurred because of the shared sensors between C6 "Personal-Hygiene", C3 "Bed to Toilet" and C4 "Sleeping".

C. HBMS Laboratory Dataset

This dataset is created based on 22 sensors (switches and motion sensors). Each sensor generates a binary output only, 1 if it is activated, 0 otherwise. The dataset is annotated with 6 activities such as Cleaning, Going for Shopping, Checking blood pressure, Getting a drink, Meal Preparation and Washing Dishes. None of these activities occur at the same time. Due to the binary nature of the sensors, context values for these sensors provide simple events for example if dishes are taken or fridge is opened. The lab had three virtual rooms (a living room, a kitchen, and a bed room plus a bath).

Activities were simulated using the simulation tool in the HBMS lab. We have simulated 5 residents and evaluated the overall performance for all of them. The quality of the spatial features assessed using Info Gain and Gain Ratio for this dataset varied between 0.19 and 0.94. The closer the quality is to one, the better it is.

In Tables V and VI, the classes are:

- C1: Cleaning
- C2: Meal Preparation,
- C3: Going for Shopping,
- C4: Checking blood pressure,
- C5: Getting a drink, and
- C6: Washing Dishes.

Table V shows that the overall F-Measure value is 0.91 although there are bad cases, first worst is class C2 "Meal Preparation" with value 0.74, second worst is class C5 "Getting a drink" with value 0.81. This can be explained by the fact that to different sensors are shared between these activities. Table V shows the confusion matrix for all six classes.

TABLE VI
THE CONFUSION MATRIX USING HBMS LABORATORY DATASET

Class	C1	C2	C3	C4	C5	C6
C1	94	0	0	0	0	0
C2	1	29	0	0	8	1
C3	1	0	33	0	0	0
C4	0	0	0	45	0	0
C5	1	6	0	0	33	0
C6	0	2	0	0	0	29

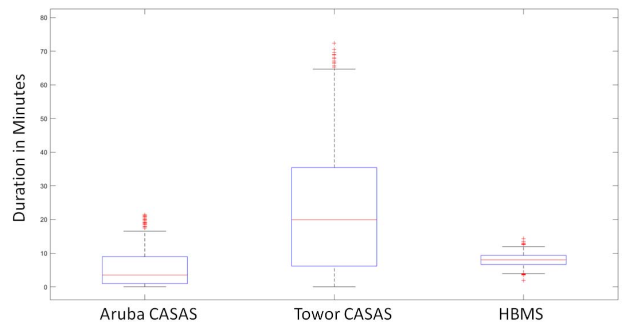


Fig. 5. The duration of the selected windows for Aruba, Towor and HBMS datasets.

Concerning the duration of the selected windows, Figure 5 shows the Boxplot chart for Aruba, Towor and HBMS datasets respectively.

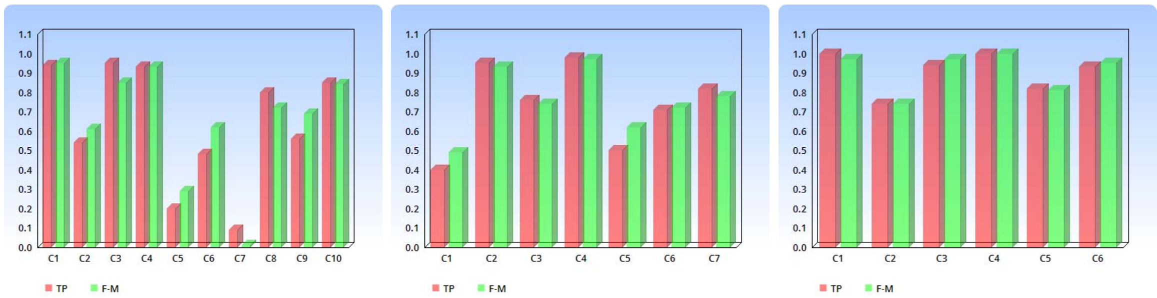
Figure 5 shows that the upper and lower quartile for Aruba dataset varies between 1 and 9 minutes. However, it varies between 6 and 34 minutes for Towor and between 6 and 9 minutes for HBMS. It clearly shows the dynamic length of the proposed windowing approach.

VI. DISCUSSION OF THE OBTAINED RESULTS

This paper has presented the classification of human activities in AAL environments based on Support Vector Machines (SVMs). We have considered both real smart homes by using the CASAS dataset and the artificially generated laboratory data using our HBMS simulator.

Moreover, we have focused mostly on motion sensors and door sensors that are used to present the selection and extraction of spatial features. Our statistical features were chosen because of their representative correct differentiation between each other. Additionally, our experiments considered the classification between the activities both for a single resident and for multiple residents. An extensive feature set is analysed using several statistical measurements. Furthermore, in order to correlate between sensor activation, the semi-variogram is designed to obtain a robust and efficient activity recognition system.

Different classification models have been tested as shown in Tables VII and VIII. Table VII shows the F-measure of each dataset and Table VIII shows the accuracy of each dataset using different classification models. The classification models have been applied on the dataset after windowing. The best performance has been obtained by support vector



(a) (b) (c)

Fig. 6. The F-measure (F-M) and the true positives (TP) measurements for all datasets. (a) The F-measure (F-M) and the true positives (TP) measurements for Aruba CASAS dataset. (b) The F-measure (F-M) and the true positives (TP) measurements for Towor CASAS dataset.

TABLE VII
F-MEASURE USING DIFFERENT CLASSIFIERS

Class	Towor	Aruba	HBMS
P-SVM	**0.80**	**0.82**	**0.91**
RBF-SVM	0.76	0.82	0.89
Naive Bayes	0.64	0.54	0.89
logistic regression	0.77	0.82	0.90
RNN	0.76	0.84	0.88

TABLE VIII
ACCURACY USING DIFFERENT CLASSIFIERS

Class	Towor	Aruba	HBMS
P-SVM	**0.81**	**0.85**	**0.92**
RBF-SVM	0.76	0.85	0.90
Naive-Bayes	0.63	0.51	0.89
logistic regression	0.78	0.85	0.91
RNN	0.71	0.85	0.90

machines (SVMs) using a polynomial kernel as the rates of accuracy and F-measure are over 0.90 for laboratory data and above 0.80 with real life activities. In contrast to that, the Naive Bayes classifier does not perform well compared to other models, but it performs well using the laboratory data. This can be explained by the overlapping features in the CASAS dataset. Based on the high performance of the polynomial SVM for all datasets, it makes it a proper candidate to be used as a multiclass classification model for human activity recognition.

Additionally, based on our observation, we have realized that human activities generate sequences, which contain sensor data records that are dependent. Therefore, a Recurrent Neural Network could be a proper model because of the connections between neurons that form a directed cycle (i.e. a loop). Consequently, each neuron can use these loops as an internal memory to maintain information about the previous input to recognize human activity in smart environments. However, the windowing approach is still needed to train the RNN model. Thus, to compare the obtained results with deep learning models, a Recurrent Neural Network (RNN) has been created which consists of 8 hidden layers and a softmax layer as a decision layer [68]. The input of the RNN model are the proposed observation windows. Due to the small dimension of the observation windows, a very deep neural model here is not required. The obtained accuracy is 0.85, 0.71 and 0.90, and the obtained F-measure is 0.84, 0.76 and 0.88 respectively for Aruba CASAS, Towor CASAS and HBMS. It is noticed that the F-measure of the accuracy using RNN for Aruba dataset performs better than the polynomial SVM which might be explained by the high number of training samples in Aruba compared to other datasets.

Furthermore, we have measured the performance of our system not only in terms of classification accuracy, but also we have considered different types of evaluation metrics which can provide a significant evaluation of the proposed approach. Also, the extracted features perform well for most activities (see Figure 6) where the figure shows the F-measure (F-M) and the true positive (TP) measurements for all datasets. Although these figures show that the results for real life scenarios are slightly poorer compared to the laboratory data generated using our simulation tool.

This is fully reasonable because, as might be expected, the nature of the activities monitored in real life scenarios contain an extensive set of different ways of acting and many different a priori unrelated actions. Consequently, the features that should be extracted and selected have to deal with a widespread range of situations. This was the goal of using the proposed statistical moments and semi-variogram features.

The proposed approach offers: (a) an effective online activity recognition system which can run on smart devices without processing constraints. (b) has the flexibility to support different residents, which is highly required for AAL. (c) the proposed dynamic windowing approach does not provide a fixed window size; in contrast to that, it varies depending on residents' actions, and (d) the proposed approach shows promising results when compared to other approaches.

Concerning the comparison to other works, in [69], they used the Towor dataset to classify human activities based on Naive Bayes, C4.5 Decision Tree, and Support Vector Machine while considering the annotated dataset. The overall accuracy was 0.77, 0.76 and 0.76, respectively. In [54], they classified the activities based on Hidden Markov Models (HMM) and Conditional Random Fields (CRF) using the kernel fusion method and a sequential behavior analysis. The overall F-measure was 0.73 and 0.79, respectively.

With respect to the Aruba dataset, in [70], a Probabilistic Neural Network (PNN) was used, and in [53], an Activity Recognition approach by a Clustering based Classification (AR-CbC) were proposed and compared to the Evidence Theoretic K-Nearest Neighbors (ET-KNN) model. The F-measure for PNN, AR-CbC and ET-KNN was 0.69, 0.75 and 0.71 respectively.

The advantage of previous works arises from the fact that they can recognize specific activities even better than our proposed approach with respect to the accuracy measurement. However, our approach performs better with respect to F-measure.

Nevertheless, as these ADL support systems are designed to be used by people, such systems should not be intrusive due to the fact that wearing such intrusive sensors by an old or disabled person is still a controversial issue. According to [71], the acceptance does not only rely on people's capabilities and limitations, it also depends on the personal, socioeconomic and cultural contexts. To avoid such hinders, choosing non-intrusive sensors is still the best choice. Moreover, cameras and microphones are non-intrusive but are still a problem for both governments and patients due to privacy issues. Thus, they are not preferable in practical daily life situations, and this is what we have considered in this work by using sensors that do not affect users' privacy.

VII. Conclusion and Outlook

This paper has proposed an activity recognition system which provides high performance based on a windowing technique using the most suitable features approach. Given the disadvantages of previous mentioned windowing techniques, this situation might lead to low performance compared to our proposed approach. This paper focuses on the development of an effective activity recognition approach. It has discussed the consideration of non-intrusive sensors using different statistical features based on both real life and artificially simulated data. Moreover, we have analyzed diverse state-of-the-art approaches for activity recognition, the advantages and disadvantages of each technique, and a comprehensive evaluation of the performance of the proposed system using different evaluation metrics has been conducted.

In our future work, we will test the proposed approach using other datasets, increase the number of activities, implement the proposed system on non-intrusive devices and think about the best way to interact with disabled or aged users for their help and daily support.

References

[1] B. K. Hensel, G. Demiris, and K. L. Courtney, "Defining obtrusiveness in home telehealth technologies: A conceptual framework," *J. Amer. Med. Informat. Assoc.*, vol. 13, no. 4, pp. 428–431, 2006.

[2] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "CASAS: A smart home in a box," *Computer*, vol. 46, no. 7, pp. 62–69, 2013.

[3] J. Michael, A. Grießer, T. Strobl, and H. C. Mayr, "Cognitive modeling and support for ambient assistance," in *Information Systems: Methods, Models, and Applications*. Yalta, Ukraine: Springer, 2012, pp. 96–107.

[4] H. Kress-Gazit, G. E. Fainekos, and G. J. Pappas, "Temporal-logic-based reactive mission and motion planning," *IEEE Trans. Robot.*, vol. 25, no. 6, pp. 1370–1381, Dec. 2009.

[5] A. Artikis, M. Sergot, and G. Paliouras, "A logic programming approach to activity recognition," in *Proc. 2nd ACM Int. Workshop Events Multimedia*, 2010, pp. 3–8.

[6] C. Dousson and P. Le Maigat, "Chronicle recognition improvement using temporal focusing and hierarchization," in *Proc. IJCAI*, vol. 7. 2007, pp. 324–329.

[7] V. D. Shet, D. Harwood, and L. S. Davis, "VidMAP: Video monitoring of activity with prolog," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill. (AVSS)*, Sep. 2005, pp. 224–229.

[8] A. Paschke and M. Bichler, "Knowledge representation concepts for automated sla management," *Decision Support Syst.*, vol. 46, no. 1, pp. 187–205, 2008.

[9] R. Kowalski and M. Sergot, "A logic-based calculus of events," in *Foundations of Knowledge Base Management*. Springer, 1989, pp. 23–55.

[10] S. D. Tran and L. S. Davis, "Event modeling and recognition using Markov logic networks," in *Computer Vision*. Marseille, France: Springer, 2008, pp. 610–623.

[11] J. A. Miller and G. Baramidze, "Simulation and the semantic Web," in *Proc. 37th Conf. Winter Simulation*, 2005, pp. 2371–2377.

[12] G. Meditskos, S. Dasiopoulou, and I. Kompatsiaris, "MetaQ: A knowledge-driven framework for context-aware activity recognition combining SPARQL and OWL 2 activity patterns," *Pervasive Mobile Comput.*, vol. 25, pp. 104–124, Jan. 2015.

[13] G. Okeyo, L. Chen, H. Wang, and R. Sterritt, "Dynamic sensor data segmentation for real-time knowledge-driven activity recognition," *Pervasive Mobile Comput.*, vol. 10, pp. 155–172, Feb. 2014.

[14] S. Zolfaghari, R. Zall, and M. R. Keyvanpour, "SOnAr: Smart ontology activity recognition framework to fulfill semantic Web in smart homes," in *Proc. 2nd Int. Conf. Web Res. (ICWR)*, Apr. 2016, pp. 139–144.

[15] F. Al Machot, K. Kyamakya, B. Dieber, and B. Rinner, "Real time complex event detection for resource-limited multimedia sensor networks," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (AVSS)*, Aug. 2011, pp. 468–473.

[16] F. Al Machot *et al.*, "Smart resource-aware multimedia sensor network for automatic detection of complex events," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (AVSS)*, Aug./Sep. 2011, pp. 402–407.

[17] F. Al Machot, H. C. Mayr, and J. Michael, "Behavior modeling and reasoning for ambient support: HCM-L modeler," in *Proc. Int. Conf. Ind., Eng. Other Appl. Appl. Intell. Syst.*, 2014, pp. 388–397.

[18] Y. Zhang, Y. Zhang, E. Swears, N. Larios, Z. Wang, and Q. Ji, "Modeling temporal interactions with interval temporal Bayesian networks for complex activity recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2468–2483, Oct. 2013.

[19] A. Veenendaal, E. Jones, Z. Gang, E. Daly, S. Vartak, and R. Patwardhan. (2016). "Dynamic probabilistic network based human action recognition." [Online]. Available: https://arxiv.org/abs/1610.06395

[20] S. Ribaric and T. Hrkac, "A model of fuzzy spatio-temporal knowledge representation and reasoning based on high-level Petri nets," *Inf. Syst.*, vol. 37, no. 3, pp. 238–256, 2012.

[21] F. J. Ordonez, G. Englebienne, P. de Toledo, T. van Kasteren, A. Sanchis, and B. Kröse, "In-home activity recognition: Bayesian inference for hidden Markov models," *IEEE Pervasive Comput.*, vol. 13, no. 3, pp. 67–75, Jul./Sep. 2014.

[22] D. Trabelsi, S. Mohammed, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "An unsupervised approach for automatic activity recognition based on hidden Markov model regression," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 829–835, Jul. 2013.

[23] F. Sebbak, F. Benhammadi, A. Chibani, Y. Amirat, and A. Mokhtari, "Dempster–Shafer theory-based human activity recognition in smart home environments," *Ann. Telecommun.*, vol. 69, nos. 3–4, pp. 171–184, 2014.

[24] E. Javadi, B. Moshiri, and H. S. Yazdi, "Activity recognition in smart home using weighted Dempster–Shafer theory," *Int. J. Smart Home*, vol. 7, no. 6, pp. 23–34, 2013.

[25] K. Zhan, S. Faux, and F. Ramos, "Multi-scale conditional random fields for first-person activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2014, pp. 51–59.

[26] X. Zhuang, J. Huang, G. Potamianos, and M. Hasegawa-Johnson, "Acoustic fall detection using Gaussian mixture models and GMM supervectors," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2009, pp. 69–72.

[27] C.-L. Huang, H.-C. Shih, and C.-Y. Chao, "Semantic analysis of soccer video using dynamic Bayesian network," *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 749–760, Aug. 2006.

[28] Y. Luo, T.-D. Wu, and J.-N. Hwang, "Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks," *Comput. Vis. Image Understand.*, vol. 92, no. 2, pp. 196–216, 2003.

[29] Y.-L. Hsueh, N.-H. Lin, C.-C. Chang, O. T.-C. Chen, and W.-N. Lie, "Abnormal event detection using Bayesian networks at a smart home," in *Proc. 8th Int. Conf. Ubi-Media Comput. (UMEDIA)*, Aug. 2015, pp. 273–277.

[30] D. Du, H. Qi, Q. Huang, W. Zeng, and C. Zhang, "Abnormal event detection in crowded scenes based on structural multi-scale motion interrelated patterns," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2013, pp. 1–6.

[31] Y. Zhu, N. M. Nayak, and A. K. Roy-Chowdhury, "Context-aware activity modeling using hierarchical conditional random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1360–1372, Jul. 2015.

[32] D. Kawanaka, T. Okatani, and K. Deguchi, "HHMM based recognition of human activity," *IEICE Trans. Inf. Syst.*, vol. E89-D, no. 7, pp. 2180–2185, 2006.

[33] M. H. Kolekar and D. P. Dash, "Hidden Markov model based human activity recognition using shape and optical flow based features," in *Proc. IEEE Region 10 Conf. (TENCON)*, Nov. 2016, pp. 393–397.

[34] W. Zhang, F. Chen, W. Xu, and Z. Cao, "Decomposition in hidden Markov models for activity recognition," in *Multimedia Content Analysis and Mining*. Weihai, China: Springer, 2007, pp. 232–241.

[35] B. Walczak and D. L. Massart, "Rough sets theory," *Chemometrics Intell. Lab. Syst.*, vol. 47, no. 1, pp. 1–16, 1999.

[36] S. Chen, Y. Deng, and J. Wu, "Fuzzy sensor fusion based on evidence theory and its application," *Appl. Artif. Intell.*, vol. 27, no. 3, pp. 235–248, 2013.

[37] J.-L. Reyes-Ortiz, L. Oneto, A. Sama, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, Jan. 2016.

[38] R. Huq, G. K. I. Mann, and R. G. Gosine, "Behavior-modulation technique in mobile robotics using fuzzy discrete event system," *IEEE Trans. Robot.*, vol. 22, no. 5, pp. 903–916, Oct. 2006.

[39] K. Subramanian and S. Suresh, "Human action recognition using metacognitive neuro-fuzzy inference system," *Int. J. Neural Syst.*, vol. 22, no. 6, p. 1250028, 2012.

[40] C. Liang-Zhou, S. Wen-Kang, D. Yong, and Z. Zhen-Fu, "A new fusion approach based on distance of evidences," *J. Zhejiang Univ. Sci. A*, vol. 6, no. 5, pp. 476–482, 2005.

[41] R. Granada, J. Monteiro, R. C. Barros, and F. Meneguzzi, "A deep neural architecture for kitchen activity recognition," in *Proc. Assoc. Adv. Artif. Intell.*, 2017, pp. 1–6.

[42] Z. Luo, B. Peng, D.-A. Huang, A. Alahi, and L. Fei-Fei. (2017). "Unsupervised learning of long-term motion dynamics for videos." [Online]. Available: https://arxiv.org/abs/1701.01821

[43] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal, and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features," *Pattern Recognit.*, vol. 61, pp. 295–308, Jan. 2017.

[44] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016.

[45] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," in *Proc. IEEE 13th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Jun. 2016, pp. 71–76.

[46] X. Li, Y. Zhang, I. Marsic, A. Sarcevic, and R. S. Burd, "Deep learning for RFID-based activity recognition," in *Proc. 14th ACM Conf. Embedded Netw. Sensor Syst. CD-ROM*, 2016, pp. 164–175.

[47] K. Wang, X. Wang, L. Lin, M. Wang, and W. Zuo, "3D human activity recognition with reconfigurable convolutional neural networks," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 97–106.

[48] L. Dong, X. Chen, R. Wang, Q. Zhang, and E. Izquierdo, "ADORE: An adaptive Holons representation framework for human pose estimation," *IEEE Trans. Circuits Syst. Video Technol.* [Online]. Available: http://ieeexplore.ieee.org/document/7932915/

[49] T. L. M. Van Kasteren, G. Englebienne, and B. J. Kröse, "An activity monitoring system for elderly care using generative and discriminative models," *Pers. Ubiquitous Comput.*, vol. 14, no. 6, pp. 489–498, 2010.

[50] N. C. Krishnan and S. Panchanathan, "Analysis of low resolution accelerometer data for continuous human activity recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2008, pp. 3337–3340.

[51] D. J. Cook and N. C. Krishnan, *Activity Learning: Discovering, Recognizing, and Predicting Human Behavior From Sensor Data*. Hoboken, NJ, USA: Wiley, 2015.

[52] N. C. Krishnan and D. J. Cook, "Activity recognition on streaming sensor data," *Pervasive Mobile Comput.*, vol. 10, pp. 138–154, Feb. 2014.

[53] L. G. Fahad, A. Ali, and M. Rajarajan, "Long term analysis of daily activities in smart home," in *Proc. ESANN*, 2013, pp. 1–6.

[54] I. Fatima, M. Fahim, Y.-K. Lee, and S. Lee, "A unified framework for activity recognition-based behavior analysis and action prediction in smart homes," *Sensors*, vol. 13, no. 2, pp. 2682–2699, 2013.

[55] N. Bicocchi, M. Mamei, and F. Zambonelli, "Detecting activities from body-worn accelerometers via instance-based algorithms," *Pervasive Mobile Comput.*, vol. 6, no. 4, pp. 482–495, Aug. 2010.

[56] J. Novakovic, "Using information gain attribute evaluation to classify sonar targets," in *Proc. 17th Telecommun. Forum TELFOR*, 2009, pp. 24–26.

[57] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.

[58] J. Ye, A. K. Clear, L. Coyle, and S. Dobson, "On using temporal features to create more accurate human-activity classifiers," in *Artificial Intelligence and Cognitive Science*. Dublin, Ireland: Springer, 2009, pp. 273–282.

[59] J. Rice, *Mathematical Statistics and Data Analysis*. Scarborough, ON, USA: Nelson Education, 2006.

[60] K. Crammer and Y. Singer, "On the learnability and design of output codes for multiclass problems," *Mach. Learn.*, vol. 47, no. 2, pp. 201–233, 2002.

[61] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[62] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 2013.

[63] V. N. Vapnik and V. Vapnik, *Statistical Learning Theory*, vol. 1. New York, NY, USA: Wiley, 1998.

[64] V. Vapnik, "The support vector method of function estimation," in *Nonlinear Modeling: Advanced Black-Box Techniques*. Boston, MA, USA: Kluwer, 1998, p. 86.

[65] E. E. Osuna, "Support vector machines: Training and applications," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 1998.

[66] M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," in *Proc. 21st Austral. Comput. Sci. Conf. ACSC*, 1998.

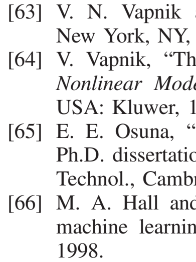
[67] I. H. Witten, E. Frank, L. Trigg, M. Hall, G. Holmes, and S. J. Cunningham, "Weka: Practical machine learning tools and techniques with java implementations," Dept. Comput. Sci., Univ. Waikato, Hamilton, New Zealand, Tech. Rep., 1999.

[68] H. Jaeger, "Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the 'Echo State Network' approach," GMD Res. Center Inf. Technol., German Nat. Res. Inst. Comput. Sci., Tech. Rep. 159, 2002, vol. 5.

[69] C. Phua, K. Sim, and J. Biswas, "Multiple people activity recognition using simple sensors," in *Proc. PECCS*, 2011, pp. 224–231.

[70] L. G. Fahad, S. F. Tahir, and M. Rajarajan, "Activity recognition in smart homes using clustering based classification," in *Proc. 22nd Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2014, pp. 1348–1353.

[71] Q. Ni, A. B. García Hernando, and I. P. de la Cruz, "The elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development," *Sensors*, vol. 15, no. 5, pp. 11312–11362, 2015.



Fadi Al Machot received the degree in computer science from the University of Potsdam in 2010, and the Diploma degree in the field of data mining and computer vision and artificial intelligence and the Ph.D. degree from Alpen-Adria University Klagenfurt in 2013. He was a research member in different European projects. As a researcher, he developed different algorithms in the areas of complex event detection in video/audio surveillance systems, advanced driver assistance systems, and human cognitive reasoning. His work has been patented and published in different international conferences and journals. He is currently a Lecturer and a Senior Researcher with Klagenfurt University and a Senior Data Scientist with Research Center Borstel–Leibniz Center for Medicine and Biosciences.

Ahmad Haj Mosa received the B.S. degree in mechatronics engineering from Tishreen University, Lattakia, Syria, in 2008, the M.S. degree in health care IT from the Carinthia University of Applied Science, Klagenfurt, Austria, and the Ph.D. degree from Klagenfurt University in 2016. He is currently a Lecturer with Klagenfurt University. He is also a Manager in the field of artificial intelligence for tax technology with PwC Austria. His research of interests includes machine vision, machine learning, applied mathematics, and neurocomputing. He developed a variety of methods in the scope of human-machine interaction and pattern recognition. He has participated in several projects the field of transportation informatics, and the outcomes were patented.

Mouhannad Ali received the B.Sc. degree in information technology from Virtual Syrian University, Syria, in 2008, and the Diploma degree in health care IT from Carinthia University of Applied Science of Klagenfurt, Austrian, in 2012. He is currently pursuing the Ph.D. degree with the Institute for Smart System Technology, Alpen-Adria-Universität Klagenfurt, Austria. His research interests include Signal Processing, Machine Learning and Data Mining.

Kyandoghere Kyamakya received the Ingénieur Civil degree in electrical engineering from the University of Kinshasa, Congo, in 1990, and the Ph.D. degree in electrical engineering from the University of Hagen, Germany, in 1999. He was experienced three years as a Post-Doctoral Researcher in the field of mobility management in wireless networks with the Leibniz University of Hannover. From 2002 to 2005, he was a Junior Professor of Positioning and Location-Based Services at Leibniz University of Hannover. Since 2005, he has been a Full Professor of transportation informatics and the Director of the Institute for Smart Systems Technologies, Alpen-Adria-Universität Klagenfurt, Austria. He is actively conducting research involving modeling, simulation, and test-bed evaluations for a series of concepts in the frame of the application of information and communication technology in transportation. Three main transportation systems are addressed, such as intelligent transportation systems, intelligent vehicles and mobile robotics systems, and intelligent logistics and supply chains. In the research addressing these systems, a series of fundamental and theoretical tools from the fields of applied mathematics, electronics, and computer science are either extensively exploited or are source of inspiration for innovative solutions and concepts: nonlinear dynamics and synchronization, cellular neural networks, nonlinear image processing, cellular-neural-networks-based analog computing, systems science, and computational intelligence.