Smart Home Simulation using Avatar Control and Probabilistic Sampling

J. Lundström *, J. Synnott †, E. Järpe ‡, C.D.Nugent †

*Intelligent Systems Laboratory, ‡Department MPE Halmstad University, Box 823, S 301 18 Halmstad Sweden
{jenlun, erja}@hh.se

†School of Computing and Mathematics, University of Ulster, Jordanstown, Shore Road, Newtownabbey, Co. Antrim BT37 0QB
{j.synnott, cd.nugent}@ulster.ac.uk

Abstract—Development, testing and validation of algorithms for smart home applications are often complex, expensive and tedious processes. Research on simulation of resident activity patterns in Smart Homes is an active research area and facilitates development of algorithms of smart home applications. However, the simulation of passive infrared (PIR) sensors is often used in a static fashion by generating equidistant events while an intended occupant is within sensor proximity. This paper suggests the combination of avatar-based control and probabilistic sampling in order to increase realism of the simulated data. The number of PIR events during a time interval is assumed to be Poisson distributed and this assumption is used in the simulation of Smart Home data. Results suggest that the proposed approach increase realism of simulated data, however results also indicate that improvements could be achieved using the geometric distribution as a model for the number of PIR events during a time interval.

I. INTRODUCTION

There is currently a global trend towards population ageing. This is as a result of several factors including an increase in life expectancy combined with a reduction in fertility levels. The population of older adults aged 60 or over is predicted to increase from the 600 million recorded in 2000 to 2 billion by 2050. The fastest growing portion of the older adult population is those aged 80 and over, who accounted for 1 in 7 of older adults in 2009 and are predicted to increase to account for 1 in 5 older adults by 2050 [1].

An increase in life expectancy is a positive situation, however the increased proportion of older adults poses a challenge for the sustainability of adequate healthcare provision. As the prevalence of such conditions increases, so will the demand on resources. Intelligent Environments, and in particular Smart Homes, are seen as a potential solution to ease the predicted burden on healthcare resources by facilitating ageing in place, ambient assisted living and the provision of objective healthcare metrics.

The development, testing and validation of algorithms for human activity modelling and deviation detection in smart homes are often complex, expensive and tedious processes. There are many factors to consider, including sensor accuracy and placement, in addition to the complexity of residents' behaviour. Performances of activities of daily living (ADLs) may consist of many different valid combinations of steps. A recent study with simulator *IE Sim* demonstrated a constrained ADL graph limiting the process of preparing a cup of tea to

8 possible combinations of steps, although in reality there are far more [2].

Access to comprehensive sensor datasets is required in order to facilitate the development, testing and validation of new models and algorithms. However, access to such datasets is limited for several reasons. As with any physical environment, Intelligent Environments and Smart Homes are expensive to construct, requiring a significant investment of resources in terms of time for planning, space, and cost for construction.

The use of Intelligent Environment simulation software can ease these limitations, allowing researchers to create vast, diverse datasets. The use of simulators enables researchers and developers to test software elements early in the process, e.g. to test a deviation detection algorithm's ability to detect a person falling in a smart home. For such simulations to be useful the output data should show a high correlation to data collected in a real-life setting. The use of simulation software during the design phase of a novel approach is likely to result in more robust and inclusive designs [3].

It is not common for current simulators to specify details (i.e. arms and legs movements) on how the human moves with respect to the PIR sensors on a level of activity or time, or both. However, simulators such as *IE Sim* [2], let the software user control an avatar in a 2D environment. Moreover, a 2D perspective of an activity where the avatar is eating breakfast is visualised as an avatar with a fixed position and renders no PIR events (or simulates events at a fixed rate). This could be a problem when such motion data is crucial for the tested algorithms, e.g. for anomaly detection algorithms. Therefore, in this paper the avatar controlled approach of *IE Sim* is combined with estimation of human motion in rooms, in the goal to produce more realistic data sets. The approach is evaluated using a small study where similarities between simulated and real data (from a specific scenario) are analysed.

II. BACKGROUND AND RELATED WORK

Systems for detecting deviating activity patterns is one example of smart environment applications, which is in particular limited to *smart homes*. The methods of such applications are based on *ubiquitous computing* which is a concept of integrating computers into everyday life in order to facilitate the modern life of humans, a concept coined by Weiser nearly three decades ago [4]. Ubiquitous computing also involve

computers able to adapt to the environment by sensing various features, e.g. by sensing *location*. Realisations of the concept can be found in the research area of *Intelligent environments* which includes examples of smart homes [5], class rooms [6], and cities [7].

An example from academia is the large scale intelligent environment at Ulster University, which encompasses a home-like environment including a sensor equipped kitchen, living room and a meeting room. The environment, which is used to facilitate research and is used in synergy with developed simulator tools, is equipped with a range of sensor technologies including X10 and Tynetec PIR, contact, chair pressure and floor pressure sensors [8].

Another example is the technology demonstrator developed at Halmstad University, which is a portable smart home of 10 m^2 size that let researchers test sensors, sensor placement, algorithms and scenarios. The space is especially used for testing anomaly detection algorithms and is used as a bridge between simulations and real-life measurements. Moreover, currently six elderly residents that live in their own home in Halmstad, Sweden, have sensors and data logging equipment installed in order to log their normal activity patterns. The long-term goal is to develop algorithms which could detect potentially dangerous situations. Despite the ongoing data collection, realistic simulation data is required to facilitate the research process (e.g. by focusing on specific deviating event patterns).

There are a number of existing related works exploring the use of virtual environments to generate simulated datasets. Buchmayr et al. (2011) [9] created a simulator involving the use of floor plans that facilitated sensor interaction through mouse clicks. Poland et al. (2009) [10] and McGinn et al (2010) [11] created simulation tools which facilied inter-action with 3-dimensional environments through the use of an avatar. Synnott et al. created IE Sim [2], which was created to expand upon previous works in the area by facilitating rapid creation of 2D floor plans, with the ability to create new objects and sensors for use within an environment. This approach, which generates data in the HomeML format, is the target of expansion in this paper.

III. PROPOSED APPROACH

To increase realism it is proposed to combine the interactive avatar-based approach with probabilistic modelling of PIR sensors. This section describes each of the components and the proposed combination thereof.

A. Modelling number of PIR events by the Poisson distribution

The events generated by an PIR sensor could be grouped in bins over different times of the day. Consider bin b_i which is defined by the start and stop times t_i, t_{i+l} where l denotes the width of the bin, in time. Moreover, the sum of all events which occur in b_i is denoted as x_i. Given multiple observations of x_i (e.g. the number of events occurring between seven and eight AM during 30 days) we assume that x_i is the realisation of the random variable X_i which is distributed according to some arbitrary distribution. It is reasonable to assume that X_i is Poisson distributed due to some properties of x_i: being non-negative due to the *count* of number of events.

The Poisson distribution is governed by a single parameter, λ (expected number of events during a time period), which controls both expected value and variance, as seen in the Poisson probability density function:

$$P(X = x) = \frac{\lambda^x \exp^{-\lambda}}{x!} \qquad (1)$$

For a certain time period t_i we estimate λ_i by taking the sample mean of all events x_i which occur in bins during time period i. Hence, it is possible to generate realisations of the Poisson distributed PIR events using the estimated λ values. Other authors have suggested to model Smart Home data by similar assumptions based on the Poisson distribution [12].

To generate PIR events, the inter-arrival time between two successive events is randomly drawn from the the exponential distribution, which is a distribution explaining the elapsed time between events and is also governed by the λ parameter.

B. Modelling interaction using IE Sim

IE Sim [2] was developed in line with Buchmayr et al.s recommendations for the functionality of a simulator designed for the generation and visualisation of sensor data [9]. These recommendations include several essential requirements including environment creation, sensor creation, error simula-tion, real-time data synthesis and real-time visualisation. IE Sim facilitates the creation of simulated environments through the use of an object toolbox. This object toolbox provides a range of structural objects (including rooms and walls), household objects (including furniture and appliances such as sofas, chairs, televisions, kettles, microwaves etc.) and sensors (including PIR, door contact, humidity, light etc.). Users can select objects required in the environment and drag them into an environment workspace to create an interactive floor plan of the environment. The properties of any objects and sensors can be adjusted, including position, orientation, size, name and ID. For interactive objects such as doors and appliances, properties such as the interaction range can be adjusted. Additionally, properties of the data generated by sensors may be adjusted, such as the data values generated, the detection range and the frequency of data generation. IE Sim promotes flexibility by supporting the creation of new objects or sensors using an intuitive user interface. User-created environments can be saved in XML format for distribution.

Interaction with environments is facilitated through the use of an avatar-based approach. The software interface is easy to use and is designed for non-technical users through an intuitive user interface and intuitive avatar controls. This approach involves the use of the arrow keys to navigate an avatar throughout the environment. The implementation of realistic collision detection ensures the user must navigate the avatar around objects such as furniture, and interact with objects such as doors in order to access various areas of the environment. Interaction with the environment is performed either passively or actively. Passive interaction includes the firing of PIR sensors when the avatar is within the PIR sensors field of view, or the firing of a pressure sensor when the avatar makes contact with it. Active interaction involves the use of a context menu which updates to provide interactions options (e.g. Open door, Close door) once the avatar is within range

of an interactive object. IE Sim supports the adjustment of simulation time, facilitating the generation of datasets spanning large amounts of time. For example, activity recording can be completed in real-time when simulating detailed ADL performance, and can then be fast-forwarded during longer periods of low activity, such as sleeping.

This interactive avatar-based approach has the advantage of facilitating the recording of detailed, fine-grained activity performances. The user has complete control over when and where activities occur, in addition to the exact order of events in ADL performance. For example, IE Sim has been shown to facilitate the recording of the steps involved in the making a cup of tea activity, including valid and invalid activity performances. The approach also facilitates the recording of activity interruptions, such as a user completing some of the steps of an activity before leaving the room to answer the telephone and then returning to complete the activity. The ability to record such detailed activity performances while generating associated sensor data can facilitate the generation of detailed sensor datasets for use within a range of applications.

C. Combining IE Sim and probabilistic sampling

Previous versions of IE Sim produced PIR data generated at fixed intervals, as determined by the *Frequency* parameter. Our proposed approach was to combine the Poisson distribution modelling approach with IE Sim for the production of more realistic datasets. This augmentation was implemented by adding a parameter file containing estimates of λ during various time periods to the virtual PIR sensor class.

A PIR sensor becomes *active* when the avatar enters the field of view (FOV) of the sensor. While a sensor is active, the *TotalTimeActivated* property of that PIR sensor updates to track the duration of time that the avatar is within the sensor's FOV. Once the avatar leaves the FOV of the sensor, or if the simulation is ended, the data for the active period i is retrospectively generated by sampling inter-arrival times from the exponential distribution using λ_i. This data generation process involves iterating through the active time and comparing the simulated time to the associated lambda values.

IV. EXPERIMENTAL SETUP

A virtual representation of a sensor equipped home of an elderly resident in Halmstad, Sweden, was created using the IE Sim software. This was based on an architectural floor plan of the environment, which facilitated the precise replication of room layout and dimensions, in addition to the placement of objects and sensors within the environment. The real-world setup contains six PIR sensors, eight door switch sensors and one bed sensor each thoroughly tested when installing to get the desired output. However, for the sake of simplicity only a subset of the sensors (nine sensors) is contained in this experiment. The virtual environment is shown in Figure 1. In this example, the avatar (D) is located in the bedroom. The bedroom and hallway PIR (C) sensors are both generating PIR events as they are detecting the avatar's presence, indicated by a red glow. Inactive PIR sensors (A) are represented as a white colour.

A scenario was selected as it represented common activities recorded in the Halmstad dataset, and thereby manually picked

from the collected data. The scenario was based on typical late night and early morning routines.

Estimations of λ for five time periods were computed from 72 days of data, dataset D_1, collected in the home of the elderly resident. Only occasions when at least one PIR event was measured were used in the lambda parameter estimation, hence the estimation is based only on when the resident was in the sensor proximity. The scenario was selected from additional 74 days of data, denoted D_2. The project and data collection was granted an ethical approval by the regional ethical review board and the resident signed a written consent. Further, events generated by the visiting night patrol were automatically removed from the data before estimating λ using sensors installed at the front door.

Bias was eliminated by ensuring the recording of the scenario was performed by a user who had no familiarity with the Halmstad dataset, the behaviour of the inhabitants, or the environment other than the floor plan. The scenario was presented to a single user as a series of timestamped instructions and locations. The following example contains the instructions for the scenario:

1) Stay in bed until 07:45:00, thereafter go towards bathroom (07:47:30).
2) Spend time in bathroom (with door open) until 07:49:30.
3) Go towards living room starting 07:49:30.
4) Spend time in living room until 07:51:10, thereafter go to bathroom.
5) Spend time in bathroom until 08:11 (with door open).
6) At 08:11 go back to the kitchen and spend time at the table until 08:14:00.
7) Use trashcan and spend time by the trashcan until 08:17:00.
8) Use trashcan.
9) Go back to the table and spend time there until 08:34.
10) Use trashcan (shortly after 08:34).
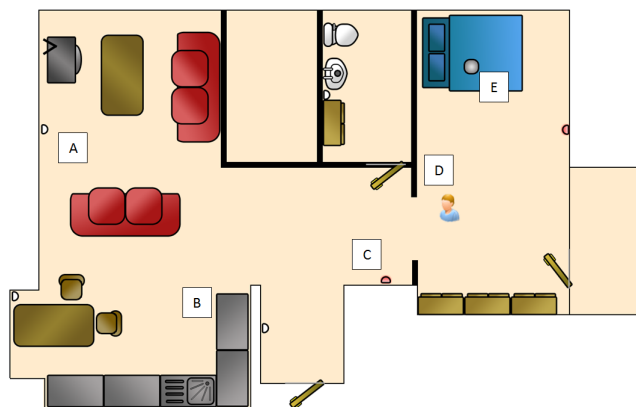11) Go back to the table and spend time there until 08:45:00.

Fig. 1. The virtual environment. (A) An inactive PIR sensor, (B) the refrigerator with a door contact sensor (as well as a door contact sensor installed at the trashcan door), (C) an active PIR sensor, (D) the avatar, (E) an inactive bed pressure sensor

A. Evaluation of simulated data

The estimated values of λ using D_1 and D_2 are compared in short intervals of two minutes in order to investigate parameter stability for 146 days. Further, simulation and collected real-world data of the scenario is investigated by visual inspection and by computing a similarity score using Dynamic Time Warping (DTW) [13]. To let DTW compute the similarity score, the simulated sequence and real-world sequence were converted to a one-dimensional discrete signal where each level represents a unique sensor activation. Moreover, DTW requires a distance function which defines the distance between two observations. Therefore a distance function designed to allow for physical proximity, or overlap of FOV, of sensors was used (e.g. the PIR sensors in kitchen and living room could be triggered by the same human movement).

V. RESULTS

To study how the rate of PIR events change during the day for D_1 and D_2 the sample mean of two minutes intervals is plotted in Figure 2 and Figure 3. Solid and dashed lines illustrate D_1 and D_2 respectively. It can be seen that movement in the bedroom (in green) increases after wake-up time (approximately at 8 am) which is followed by activity in kitchen (in black). Moreover, the two datasets D_1 and D_2 separated in time closely relates to each other which shows the consistent behaviour patterns over the two time periods. This is also shown for the PIR sensors in the living room (in blue) and bathroom (in red), see Figure 3.
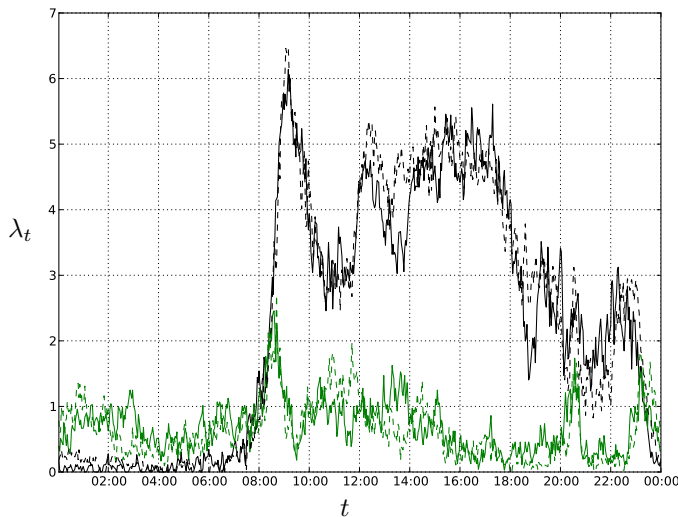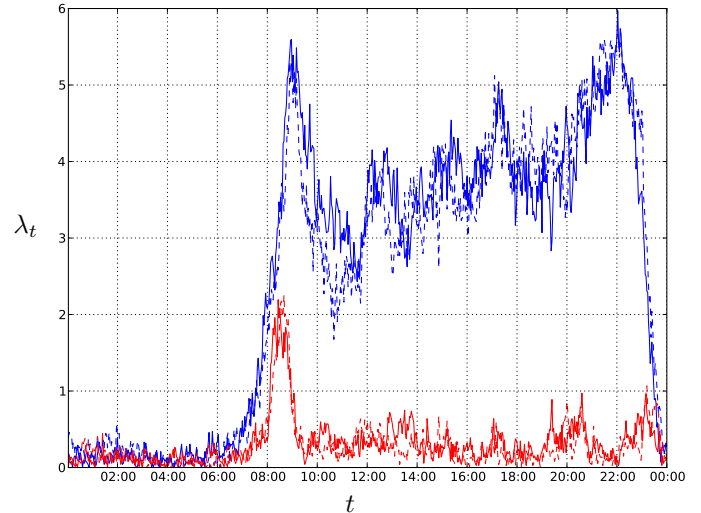

Fig. 3. Estimated λ over intervals of two minutes for livingroom (blue) and bathroom (red) for D_1 (solid line) and D_2 (dashed line) over time t.

intensities of λ the correlation is better than for high rates which shows an underestimation. The Spearman correlation coefficient and p-value for λ_s and λ_r was $r = 0.72$ and $p < 0.01$.

Fig. 2. Estimated λ over intervals of two minutes for bedroom (green) and kitchen (black) for D_1 (solid line) and D_2 (dashed line) over time t.

The consistent but daily-varying rate of PIR event, λ, during the day also suggest that different time periods during the day should be simulated using different number of PIR events.

The bedroom PIR sensor is in particular interesting for healthcare applications and the simulation capability is further analyzed in Figure 4. The plot shows the simulated (λ_s) and real-world (λ_r) sample mean of the number events during scenario, calculated using a 15 minute intervals. For low
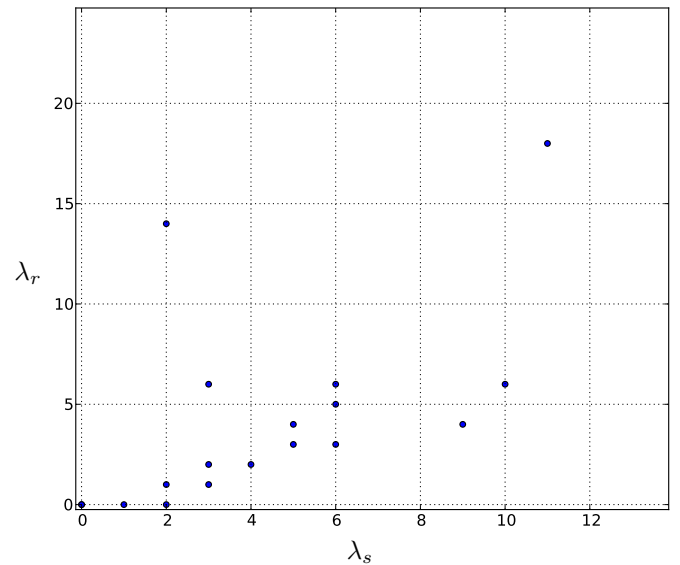

Fig. 4. Sample mean of the bedroom PIR event rate during the scenario for simulated (λ_s) and real data (λ_r).

The last two hours of simulation data and real-world collected data for the scenario is shown in Figure 5 and Figure 6 respectively. Rows indicate a specific type of sensor event (*bar* is a PIR event, *arrow down/up* is door opening/closing event, and *square* is a bed exit event) for the different sensors used.

During time period from 06:35 to 07:45 the resident is still in bed. Clearly for the latter part of this time period there is an underestimation in the number of bedroom PIR events, S_5, (just before the resident leaves the bed). This
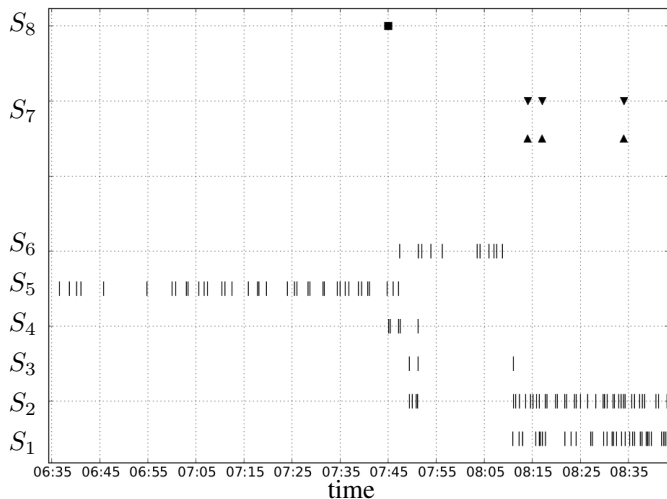
From Figures 5 and 6 an underestimation of λ for S_1 and S_2 during the time period 08:15-08:35, can be seen. This is also captured by the DTW similarity measure, Figure 7.

Fig. 5. Simulated events generated for a scenario (during a single day) where (*bar* is a PIR event, *arrow down/up* is door opening/closing event, and *square* is a bed exit event for the sensors PIR in kitchen (S_1), livingroom (S_2), hallway (S_3), corridor (S_4), bedroom (S_5), bathroom (S_6). Door sensor at trashcan in kitchen (S_7) and pressure sensor in bed (S_8).

increase of intensity is most likely context specific and is therefore something that sensors did not capture. Hallway and corridor sensors did capture transitions between the larger rooms of bedroom and living room. However, the period spent in bathroom during 07:55 and 08:05 is mainly captured by the corridor sensor outside (see S_4 in Figure 6) which is not captured by the simulator.
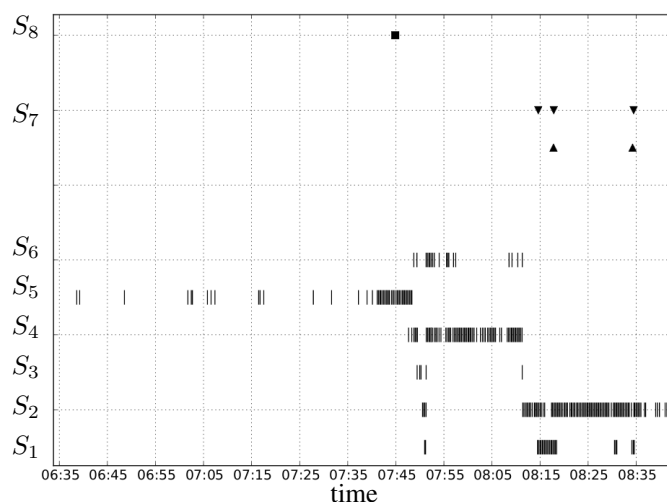

Fig. 6. Events from sensors installed in a real home for the specific scenario (during a single day) where (*bar* is a PIR event, *arrow down/up* is door opening/closing event, and *square* is a bed exit event for the sensors PIR in kitchen (S_1), livingroom (S_2), hallway (S_3), corridor (S_4), bedroom (S_5), bathroom (S_6). Door sensor at trashcan in kitchen (S_7) and pressure sensor in bed (S_8).
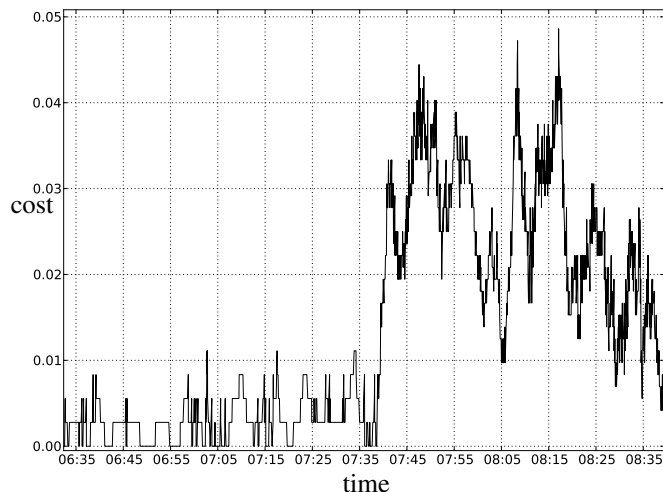

Fig. 7. Computed cost (using the DTW algorithm) of the similarity for two-minute intervals (sliding window) plotted from 06:35 to 08:35.

The apparent underestimation suggests that higher values of λ is not well aligned with the Poisson distribution. Therefore an alternative was considered, the geometric distribution. This distribution was tested for a goodness of fit using D_1 by adopting a series of χ^2-tests. The PIR data for kitchen, living room, bedroom and bathroom was used, see Table I.

TABLE I. GOODNESS OF FIT FOR GEOMETRIC DISTRIBUTION USING EXPECTED VALUE ESTIMATION FROM THREE MONTHS OF DATA, D_1. TIME PERIODS ARE t_1 (00:00-07:00), t_2 (07:00-07:30), t_3 (07:30-08:00), t_4 (08:00-08:30) AND t_5 (08:30-09:00).

χ^2/p-value	t_1	t_2	t_3	t_4	t_5
Kitchen	1.02/0.58	0.67/0.70	0.31/0.90	0.45/0.81	**18.61/0.0005**
Living room	0.91/0.65	0.10/0.96	3.14/0.21	2.32/0.29	**45.4/0.0005**
Bedroom	1.63/0.46	2.00/0.36	1.19/0.57	4.24/0.12	0.73/0.71
Bathroom	3.03/0.25	3.59/0.17	2.07/0.35	2.54/0.26	2.64/0.28

Majority of the computed values of χ^2, were below the critical-value (5.99), and almost all p-values above a five percent significance level. This means that the null hypothesis, suggesting that the data is consistent with the geometric distribution, can not be rejected for most of the goodness of fit tests.

VI. CONCLUSIONS AND DISCUSSION

This paper presents a novel approach to simulate smart home data using a combination of avatar-based control and probabilistic sampling. The proposed approach was tested by simulating data from a scenario acquired from real-world data, and then comparing similarities of simulated and real data.

It can be concluded that the average PIR rate of events do follow a regular pattern of the two datasets studied for 146 days. This is encouraging as the proposed approach is based on λ estimates of different PIR sensors.

For low PIR event rates there is reason to believe that the approach is successful. However, for higher rates the probabilistic sampling is underestimating the PIR event rate. This could probably be due to the thin tails of the Poisson distribution, which seems reasonable due to the high goodness of fit with the geometric distribution (which has a thicker tail).

Few conclusions on the results of DTW can be taken. This is due to fact that inactivity (no sensor activity) is overrepresented in the data and therefore makes the result difficult to interpret. However, the average cost level can provide information about the general similarity.

VII.   Future Work

Future work includes testing the proposed approach using more scenarios. Moreover, further work on using a sampling function which relates to the inter-arrival times of the geometric distribution is also considered. A study on how the simulated data affects activity recognition and anomaly detection algorithms is also planned as future work.

Acknowledgment

References

[1] UnitedNations, "World population ageing 2009," United Nations, Economics and Social Affairs, Tech. Rep., 2009.

[2] J. Synnott, L. Chen, C. Nugent, and G. Moore, "The creation of simulated activity datasets using a graphical intelligent environment simulation tool," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, Aug 2014, pp. 4143–4146.

[3] S. Helal, J. W. Lee, S. Hossain, E. Kim, H. Hagras, and D. Cook, "Persim - simulator for human activities in pervasive spaces," in *Intelligent Environments (IE), 2011 7th International Conference on*, July 2011, pp. 192–199.

[4] M. Weiser, "The computer for the 21st century," *Scientific american*, vol. 265, no. 3, pp. 94–104, 1991.

[5] X. Hong, C. Nugent, M. Mulvenna, S. McClean, B. Scotney, and S. Devlin, "Evidential fusion of sensor data for activity recognition in smart homes," *Pervasive and Mobile Computing*, vol. 5, no. 3, pp. 236–252, 2009.

[6] S. S. Yau, S. K. Gupta, F. Karim, S. I. Ahamed, Y. Wang, and B. Wang, "Smart classroom: Enhancing collaborative learning using pervasive computing technology," *II American Society of Engineering Education (ASEE)*, 2003.

[7] H. Schaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson, and A. Oliveira, "Smart cities and the future internet: towards cooperation frameworks for open innovation," in *The future internet*.   Springer, 2011, pp. 431–446.

[8] X. H. S. D. C.D. Nugent, M.D. Mulvenna, "Experiences in the development of a smart lab," *International Journal of Biomedical Engineering and Technology*, vol. 2, no. 4, pp. 319–331, 2009.

[9] M. Buchmayr, W. Kurschl, and J. Kng, "A simulator for generating and visualizing sensor data for ambient intelligence environments," *Procedia Computer Science*, vol. 5, no. 0, pp. 90 – 97, 2011, the 2nd International Conference on Ambient Systems, Networks and Technologies (ANT-2011) / The 8th International Conference on Mobile Web Information Systems (MobiWIS 2011). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050911003395

[10] M. P. Poland, C. D. Nugent, H. Wang, and L. Chen, "Development of a smart home simulator for use as a heuristic tool for management of sensor distribution," *Technology and Health Care*, vol. 17, no. 3, pp. 171–182, 2009.

[11] K. McGlinn, E. O'Neill, A. Gibney, D. O'Sullivan, and D. Lewis, "Simcon: A tool to support rapid evaluation of smart building application design using context simulation and virtual reality." *J. UCS*, vol. 16, no. 15, pp. 1992–2018, 2010.

[12] A. Mendez-Vazquez, A. Helal, and D. Cook, "Simulating events to generate synthetic data for pervasive spaces," in *Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research*.   Citeseer, 2009.

[13] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 26, no. 1, pp. 43–49, 1978.