HHS Public Access
Author manuscript
J Ambient Intell Humaniz Comput. Author manuscript; available in PMC 2016 April 01.

Published in final edited form as:
J Ambient Intell Humaniz Comput. 2016 February ; 7(1): 1–19. doi:10.1007/s12652-015-0294-7.

Ambient and smartphone sensor assisted ADL recognition in multi-inhabitant smart environments

Nirmalya Roy[1], **Archan Misra**[2], and **Diane Cook**[3]

Nirmalya Roy: nroy@umbc.edu; Archan Misra: archanm@smu.edu.sg; Diane Cook: cook@eecs.wsu.edu

[1]Department of Information Systems, University of Maryland, Baltimore County, Baltimore, USA

[2]School of Information Systems, Singapore Management University, Singapore, Singapore

[3]School of Electrical Engineering and Computer Science, Washington State University, Pullman, USA

Abstract

Activity recognition in smart environments is an evolving research problem due to the advancement and proliferation of sensing, monitoring and actuation technologies to make it possible for large scale and real deployment. While activities in smart home are interleaved, complex and volatile; the number of inhabitants in the environment is also dynamic. A key challenge in designing robust smart home activity recognition approaches is to exploit the users' spatiotemporal behavior and location, focus on the availability of multitude of devices capable of providing different dimensions of information and fulfill the underpinning needs for scaling the system beyond a single user or a home environment. In this paper, we propose a hybrid approach for recognizing complex activities of daily living (ADL), that lie in between the two extremes of intensive use of body-worn sensors and the use of ambient sensors. Our approach harnesses the power of simple ambient sensors (e.g., motion sensors) to provide additional 'hidden' context (e.g., room-level location) of an individual, and then combines this context with smartphone-based sensing of micro-level postural/locomotive states. The major novelty is our focus on multi-inhabitant environments, where we show how the use of spatiotemporal constraints along with multitude of data sources can be used to significantly improve the accuracy and computational overhead of traditional activity recognition based approaches such as coupled-hidden Markov models. Experimental results on two separate smart home datasets demonstrate that this approach improves the accuracy of complex ADL classification by over 30 %, compared to pure smartphone-based solutions.

1 Introduction

Smart environment has the potential to revolutionize the way people can live and age gracefully in their own environment. The growing number of aging baby boomers and increasing healthcare costs accelerate the need for smart home technologies for healthy independent living. Wireless sensor networks, be it ambient, wearable, object, or smart phone sensors open up an avenue of smart home services, if implemented successfully, that

Correspondence to: Nirmalya Roy, nroy@umbc.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

help older adults to live in their own environment for a longer period of time. The wearable sensors could collect the biometric data, and help update the patient's electronic health records or monitor the activities, behavior or location of the inhabitants over space and time to help design the novel activity recognition algorithms in complex smart home situations.

Activity recognition can be investigated to explore healthy living, societal interaction, environmental sustainability and many other human centric applications. Simple activity recognition, while proven to be useful, need to be scaled to encompass fine-grained exploration on microscopic activities over the space, time, number of people and data sources to help design more robust and novel activity recognition techniques. Scaling the activity recognition approaches beyond a single user, home or an uni-modal data source brings up innovative research challenges. Human activities are interleaved. For example, cooking activities may concurrently happen while an individual is also watching television and may continue even after the watching TV activity ends. Similarly many activities of daily living (ADLs) are complex. For example, the high-level cooking activity is composed of low-level activities, like standing and walking in the kitchen and perhaps sitting in the living room. Multiple residents can be present at a given time with underlying spatiotemporal constraints in a smart environment and make it obviously hard to infer who is doing what (Roy et al. 2013)?

Activity recognition research in smart environments (e.g., homes or assisted-living facilities) traditionally falls into two extremes:

- *Body-worn* In the wearable computing paradigm, multiple body-worn sensors (such as accelerometers, sound, gyro sensors) are placed on an individual's body to help track their locomotive and postural movements at a very fine-granularity (e.g., Wang et al. 2011).
- *Ambient* In this alternate model, the environment itself is augmented with a variety of sensors, such as RF readers, object tags, video cameras, or motion sensors mounted in different rooms.

Unfortunately, the evidence of the last decade of research suggests that these two extremes both face steep operational and human acceptability challenges. In particular, individuals [even elderly patients (Bergmann and McGregor 2011)] appear reluctant to continually wear multiple sensors on the body. In addition, such sensors are often susceptible to placement-related artifacts. On the other hand, embedding sensors on myriad objects of daily living, such as microwaves and kitchen cabinets (Intille et al. 2006) or mounting them on the ceiling has challenging operational costs and battery-life issues. Video sensors are often viewed as too intrusive to be acceptable in assisted living homes due to privacy concerns.

Driven by these observations, we ask a basic question: *does there exist a middle ground for sensing in smart environments, especially one that can combine an everyday personal device (the smartphone) with low-cost, coarsegrained ambient sensors?* If so, what advances in activity recognition and learning algorithms do we need to jointly harness the power of these diverse sources of sensor data? Our research is motivated by the emergence of the smartphone as a de-facto pervasive and *personal* device, and its demonstrated use for detecting basic low-level activities (such as sitting, walking etc.) through simple feature-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

based classification of smartphone-embedded accelerometers (e.g., Gyorbiro et al. 2008; Kwapisz et al. 2010). Likewise, simple infrared based occupancy or motion sensors are now widely deployed, and accepted by consumers, in many indoor environments (often to automate simple tasks such as lighting control).

While this idea of combining body-worn and infrastructural sensing certainly is not new, our unique differentiator lies in the fact that we explicitly consider *multi-inhabitant settings*, where multiple individuals simultaneously occupy the smart environment and engage in individual and collective *ADLs*. In this case, the key challenge is to effectively disambiguate the association between the infrastructure sensor observations and each individual, especially when the infrastructure sensors measure *ambient conditions that are inherently non-person specific*. For example, when individual phone-mounted accelerometers suggest that both persons A and B are walking around, and occupancy sensors indicate that both the kitchen and living room are occupied, how do we map individuals to specific locations—i.e., decide if A is located in the kitchen, and B is in the living room, or vice versa? Resolving such location context, as an exemplar, in a multi-inhabitant environment, is key to more accurately profiling and classifying the activities of each individual, for various applications, such as wellness monitoring, timely in-situ reminders (e.g., medication reminder when sitting down for dinner) and lifestyle recommendations (Bergmann and McGregor 2011).

In this paper, we consider the challenge of discerning such ‘hidden’ or ‘ambiguous’ individual context, by appropriately combining both low-level person-specific individual context and person-independent ambient context. At a high-level, we model each individual's activity context as a *multi-dimensional* set of attributes, some of which are observable from the smartphone (e.g., whether the individual is walking, standing or sitting) and some of which are ‘hidden’ (e.g., is the person in the kitchen vs. living room, is she alone or with other occupants?). The *temporal evolution* of each person's activity is jointly modeled as a coupled hidden Markov model (CHMM); our unique innovation lies in the specification of a set of constraints to this model, arising from the presence of a combination of mobile and ambient sensing data. The constraints are both *intra-personal* (an individual is more or less likely to follow a certain activity pattern) and *interpersonal* (the ‘hidden context’ of different individuals is often likely to possess some mutual exclusionary properties). We then build such a CHMM through appropriate modifications to the standard expectation maximization algorithm, and use a modified Viterbi algorithm during the testing phase to determine the most likely *temporal evolution* of each person's activity.

Our investigations in this paper address several key research questions. First, given the reality of an indoor multi-inhabitant environment with cheap ambient sensors, what sort of constraints, both inter-personal and intra-personal, arise due to the combination of mobile sensing and ambient environmental data? Second, how can we combine such constraints *across multiple users*, across both time and space, to infer the ‘hidden context attributes’ of each individual, in a computationally efficient fashion? Finally, how much quantitative improvement do we observe in our ability to infer complex ADLs via such ‘hidden context’, as compared to alternatives that rely solely on the mobile sensing or the ambient observations?

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We believe that our innovations and results provide strong preliminary evidence that such a hybrid model, where mobile sensing is augmented with ambient context from cheap everyday sensors (and, in the medium-term future, sensing via wearable devices), can prove to be an attractive and practically viable alternative. Specifically, we show how the set of viable 'hidden context states' is associated with a set of possible spatial and temporal constraints, generated as a consequence of the available *combination* of mobile and ambient sensing. Besides a generic formulation, we specifically combine smartphone-based activity recognition with motion/occupancy sensor-based ambient monitoring to help identify the *indoor location or space* inhabited by different users. Such location context is crucial to correctly classifying ADLs, and this overcomes a challenge of indoor localization in smart homes (as opposed to commercial spaces blanketed by Wi-Fi APs). In addition, we develop a modified coupled HMM to express the temporal evolution of the context of multiple individuals subject to such constraints, and then present a computationally-efficient, modified Viterbi algorithm to determine the most likely temporal evolution of each individual's context. We provide results that show that this approach can be viable at least for multi-inhabitant environments, such as assisted living facilities, where the number of individuals is relatively small (e.g., below 5). Finally, we use test data, generated by appropriately synthesizing *real-life activity traces*, to quantify the performance of our algorithms and show that the intelligent fusion of such mobile plus ambient context data can improve the accuracy of 'hidden' context estimation by over 70 %, and the accuracy of ADL classification by ≈30%.

2 Related work

We cover the work on activity recognition that is closest to our focus in this paper.

Activity recognition and mobile sensing

Most of the existing work on multi-user activity recognition used video data only. HMMs and CHMMs for modeling and classifying interactions between multiple users are addressed in Oliver et al. (2000) and Wang et al. (2011), while Gong and Xiang (2003) has developed a dynamically multi-linked HMM model to interpret group activities based on video observations. Activity recognition in smart environments using unsupervised clustering of data collected by a rich set of wearable sensors has been explored in Clarkson et al. (2000). The recent proliferation of sensor-equipped smartphones suggests that a vast amount of individual-specific data can be collected via the phone's microphone, accelerometer, gyro, and magnetometer (Gyorbiro et al. 2008; Kwapisz et al. 2010; Khan et al. 2015b). A zero-configuration infrastructure-less occupancy detection techniques based on smartphone's accelerometer, microphone and magnetometer sensor have been proposed in Khan et al. (2015a). Microphone senor based acoustic noise detection has been used to count number of people in a gathering or meeting place whereas accelerometer sensor based locomotive context detection has been augmented in absence of conversational data. Roy and Kindle (2014) has investigated simple classification algorithms for remotely monitoring patient recovery using wireless physiotherapy devices while Roy and Julien (2014) has articulated the challenges for smart living environments and inclusive communities for immersive physiotherapy applications.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Classifying ADLs

Sensor-based activity recognition strategies are typically probabilistic and can be categorized into static and temporal categories (Chen et al. 2012). Naive Bayes (Logan et al. 2007), decision trees (Logan et al. 2007), K-nearest neighbors (Huynh et al. 2007) and SVM (Huynh et al. 2007) have been used extensively as static classifiers; temporal classification approaches infer the values of hidden context states using approaches such as HMMs (Lester et al. 2006), dynamic Bayesian network (Philipose et al. 2004), conditional random fields (Kasteren et al. 2008) and CHMM (Wang et al. 2011). SAMMPLE (Yan et al. 2012) is a recent attempt at classifying ADLs using only accelerometer-data via a layered (two-tier) approach, where the lower layer first classifies low-level micro-activities, whereas the higher level uses micro-activity based features to classify complex ADLs. We believe our approach is distinct from these approaches, in its judicious combination of available smartphone sensors and minimal usage of ambient sensors. A dynamic Bayesian networks (extended to a CHMM) based multi-resident activity model has been proposed in Yi-Ting et al. (2010). While this work categorized the sensor observations based on data association and semantic information, our work exploits the underlying microscopic features of the activities and the spatiotemporal nature of the state spaces with an ambient augmented mobile sensing based methodology to model the multi-residents activity patterns (Roy et al. 2006). An active learning based scalable sleep monitoring framework has been proposed in Hossain et al. (2015). A factorial hidden Markov based model has been proposed for acoustic based appliance state identifications for fine grained energy analytics in building environment (Pathak et al. 2015; Khan et al. 2015c).

Combining body-worn and ambient sensor data

The notion of using simple, ambient sensors (such as motion sensors) to infer individualized context in a multi-inhabitant smart environment was first studied in Wilson and Atkeson (2005), which uses a particle filtering approach to infer the evolution of coupled HMMs, based on events generated by multiple infrastructure-embedded sensors. Unlike Wilson and Atkeson (2005), we exploit the pervasiveness of body-worn smartphone sensors to infer some amount of person-specific context; additionally, while Wilson and Atkeson (2005) focuses only on inferring whether an individual is in movement or stationary, our focus is on inferring complex ADLs. We explore the technical feasibility of a vision where the sensing capabilities of ambient sensors are combined with the smartphone sensing of user finer movements to provide significantly greater insight into the microscopic activities of daily activities of individuals in smart home environment. While the empirical investigations carried out in this paper utilize smartphones (that may or may not be always carried around inside a home), an eventual embodiment will likely rely on wearable devices [e.g., smartwatches (Android Wear: Information 2015); smart-bracelets (Huawei Smart Bracelet 2015)] that are now gaining wider market acceptance and that a user will likely wear almost-continuously (Intel Make it Wearable 2014). A smartphone and iBeacon sensor based real time activity recognition framework has been proposed in Alam et al. (2015). A bagging ensemble learning and packaged naive bayes classification algorithm have been proposed for high level activity recognition on smartphone. In Roy et al. (2015), smartphone sensing based activity recognition approaches help reduce the set of appliances usage at a time which then combined with powerline sensing in green building environment for fine-grained

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

appliance usage and energy monitoring. Alam and Roy (2014) proposed a gestural activity recognition model for predicting behavioral health based on a smart earring.

Techniques and technologies for activity recognition

Activity recognition is an interdisciplinary research problem spanning across multiple research domains, such as internet of things (Sheng et al. 2014; Yan et al. 2014), cloud computing (Rahimi et al. 2012, 2014), wireless sensor networks (Hayajneh et al. 2014), machine learning, context-awareness and modeling (Zhou et al. 2010) etc. Ambient and artificial intelligence based methodologies have been investigated for discovery, recognition, and learning activity models (Acampora et al. 2013). Body-area sensor networks based physiological signal detection and their augmentation with the ADL to help older adults live independently has been proposed in Chen et al. (2011) while the data collected using the wearable devices around the users have been encrypted in Zhang et al. (2012). Integrating a multitude of ambient and wearable devices and making them interoperable are key research tasks for building wireless sensor networks based health and activity recognition model and tool kit (Lin et al. 2015b). While body-area sensor networks help gather meaningful data based on user movements, activities, and contexts, but storing and processing data on the cloud have become inevitable components in activity recognition pipeline (Fortino et al. 2014; Almashaqbeh et al. 2014). Accessing data on real time (Zheng et al. 2013) and labeling data through crowd sourcing (Feng et al. 2014) are of great importance for realizing scalable activity recognition model across multiple users and premises while providing just-intime intervention and proactive healthcare decision to the target population. Meeting the quality of service for inferring the activity of the users and trading the delicate balance between cost and accuracy in presence of multiple healthcare activity recognition applications have also been investigated in Roy et al. (2009, 2011, 2012), Lin et al. (2015a) and Lee et al. (2013).

3 The constrained multi-user activity model

We first mathematically describe the evolution of the context state of an individual, and then consider the various spatiotemporal constraints associated with the *combination* of smartphone-based and ambient sensing observations. We also outline how these ‘micro-context’ observations and inferences can then be used to derive the higher-layer ADLs, using a variant of the two-tier SAMMPLE approach (Yan et al. 2012).

Consider a smart environment (such as an assisted living facility) with N distinct individuals. The ith individual's *micro-context*, at a given time instant t, is captured by a M-dimensional tuple $\mathrm{Context}^i(t)=\langle c_1^i(t), c_2^i(t), \ldots, c_M^i(t)\rangle$, where each of the M elements of the tuple corresponds to a specific type of context attribute. In the canonical case considered in this paper, context is viewed as a $\langle$*microactivity*, *location*$\rangle$ tuple, where microactivity refers to an individual's postural state (such as {*walking*, *sitting*, *standing*, …,}) and location can assume values such as {*bedroom*, *bathroom*, *kitchen*,…}. In general, assuming time to be discretely slotted, an individual i's activity pattern may be represented by a micro-context stream, i.e., $Context^i(t)$, $Context^i(t+1)$, … An important characteristic of our model is that a subset of the M elements are ‘observable’. They may be inferred (with varying levels of estimation

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

error) using solely the sensors embedded within individual's body-worn and personal mobile device. For example, the determination of postural microactivity can be made using the 3-axis accelerometer (Gyorbiro et al. 2008; Kwapisz et al. 2010) universally available in modern smartphones. The remaining elements of each tuple are, however 'hidden'. The user's location is not directly revealed by the smartphone accelerometer data. The key goal of our research is to propose a technique to infer these hidden attributes.

Our smart environment is also assumed to possess J different *types* of inexpensive ambient sensors. Assume that the environment has a total of K such sensors, each of which is deployed at a well-known location. The kth : $k = 1, \ldots, K$ sensors, located at an a-priori known location $Loc(k)$, is assumed to provide some measure of ambient context, denoted by $ConAmbient(k)$ for the ambience. For example, as a canonical exemplar, the environment consists of $K = 10$ different motion sensors ($J = 1$), each of which is placed in a location such as {*bedroom*, *bathroom*, *kitchen*, …}.

3.1 Two-tier inferencing for individual/multiple inhabitants

Given our formulation above, the evolution of the micro-activities of the ith user can be represented by a state transition matrix over $Context^i(t)$. More specifically, we assume that the evolution of the state is Markovian (Rabiner 1989) with order 1 (higher order Markovian models are conceptually similar, but mathematically more elaborate), so that the $P(Context^i(t)|; Context^i(t-1))$ denotes the likelihood of the current context state, given the past context state.

Our context extraction process is illustrated in Fig. 1 and consists of two tiers [similar to the conceptual stages of the SAMMPLE approach (Yan et al. 2012)]. The first goal of our research (illustrated in the "lower tier" of Fig. 1) is to infer the 'hidden states' (specifically *location* in our experiments), given the observable (or directly inferrable) values of *postural activity*. In Fig. 1, the smartphone sensor data of an individual are first transformed into corresponding low-level 'observable' context (e.g., using the accelerometer data to infer the postural states). Note that this transformation is not the focus of this paper: we simply assume the use of well-known feature based classification techniques to perform this basic inferencing. The core contribution of the paper lies in the next step: *inferring the hidden states of an individual's low-level context, based on the combination of phone-generated and ambient sensor data*. As shown in Fig. 1, this lower-tier's challenge is to infer the 'hidden states' of multiple individuals *concurrently*, utilizing both their observable low-level individual context and the non-personal ambient context.

After inferring these hidden states, we now have a complete set of $Context^i(t)$ observations for each individual. In the next step of the two-tier process (the "higher tier" in Fig. 1), the entire set of an individual's context stream is then used to classify his/her 'higher level' (or so-called 'complex') ADLs. More specifically, based on the inferencing performed in the lower-tier, the joint (*postural activity*, *location*) stream is used to identify each individual's complex activity. The interesting question that we experimentally answer is: *how much improvement in the accuracy of complex activity classification do we obtain as a result of this additional availability of the hitherto 'unobservable' location context?*

3.2 Capturing spatial and temporal constraints

Our process for performing the 'lower-tier' of context recognition is driven by a key observation: in a multi-inhabitant environment, the context attributes of different individuals are often mutually coupled, and related to the environmental context sensed by the ambient sensors. In particular, we observe that the 'unobserved' components of each individual's micro-level context are subject (probabilistically) to both *temporal* and *spatial* constraints. As specific examples, consider the case of two users occupying a smart environment. We can see the following constraints (also shown in Fig. 3):

a. *Intra-user temporal constraints* For a specific user i, if $Context^i(t-1)$ = (*sitting*, *livingroom*), $Context^i(t)$ cannot equal (*sitting*, *bathroom*); i.e., the user cannot simply change rooms while remaining in a 'sitting' state!

b. *Inter-user spatial constraints* Given two users i and j, both $Context^i(t)$ and $Context^j(t)$ cannot be (*sitting*, *bathroom*); i.e., both the users are very unlikely to be sitting in the bathroom concurrently.

3.3 Coupled HMM for multiple inhabitants

We investigate the CHMM (Brand 1996) which has been used for inferring the users' activities in a multi-resident environment. A basic block diagram of CHMM is shown in Fig. 2 which represents the temporal interaction between the two users given their respective observation sequences. Given our assumption of Markovian evolution of each individual's context, and the demonstrated constraints or 'coupling' that arise between the various 'hidden' contextual attributes of different individuals, we can then model the evolution of each individual's 'low-level context' [i.e., $Context^i(t)$] as a CHMM (Brand 1996). To define this HMM, let $O(t)$ denote the "observable stream" (in our canonical example, this consists of the accelerometer readings on the smartphone and the motion readings reported by the occupancy sensors).

If the environment was inhabited by only a single user i, the most probable context sequence, $Context^i(t)$, given an observed sequence, is that which maximizes the joint probability $P(O^i|Context^i)$ as shown by:

$$P(O^i|\text{Context}^i)=\prod_{t=1}^{T}P(\text{Context}^i(t)|\text{Context}^i(t-1))\times P(O^i(t)|\text{Context}^i(t)). \quad (1)$$

In our case, there are multiple users inhabiting the same environment with various spatiotemporal constraints expressed across their combined context. In this case, assuming N users, we have an N-chain coupled HMMs, where each chain is associated with a distinct user as shown below:

$$P(\text{Context}^{(n)}|O)=\prod_{(n)}^{N}\left(\pi_{s_1^{(n)}}P_{s_1^{(n)}}(o_1^{(n)})\prod_{t=2}^{T}\left(P_{s_t^{(n)}}(o_t^{(n)})\prod_{(d)}^{N}P_{s_t^{(n)}|s_{t-1}^{(n)}}\right)\right)/P(O) \quad (2)$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where a different user is indexed by the superscript. $P_{s_t^{(n)}}(o_t^{(n)})$ is the emission probability given a state in chain n, $P_{s_t^{(n)}|s_{t-1}^{(d)}}$ is the transition probability of a state in chain n given a previous state in chain d and $\pi_{s_1^{(n)}}$ is the initial state probability.

Simplifying the N chain couplings as shown in Eq. 2 by considering two users, the posterior of the CHMM for any user can be represented as follows.[1]

$$P(\text{Context}|O)=\frac{\pi_{s_1}P_{s_1}(o_1)\pi_{s_1'}P_{s_1'}(o_1')}{P(O)}\prod_{t=2}^{T}P_{s_t|s_{t-1}}P_{s_t'|s_{t-1}'}P_{s_t'|s_{t-1}}P_{s_t|s_{t-1}'}P_{s_t}(o_t)P_{s_t'}(o_t') \quad (3)$$

where π_{s_1} and $\pi_{s_1'}$ are initial state probabilities; $P_{s_t|s_{t-1}}$ and $P_{s_t'|s_{t-1}'}$ are intra-user state transition probabilities; $P_{s_t|s_{t-1}'}$ and $P_{s_t'|s_{t-1}}$ are inter-user state transition probabilities; $P_{s_t}(o_t)$ and $P_{s_t'}(o_t')$ are the emission probabilities of the states respectively for User i and User j. Incorporating the spatial constraints across users as shown in Fig. 3, we modify the posterior of the state sequence for two users by:

$$P(\text{Context}|O)=\frac{\pi_{s_1}P_{s_1}(o_1)\pi_{s_1'}P_{s_1'}(o_1')}{P(O)}\times\prod_{t=2}^{T}P_{s_t|s_{t-1}}P_{s_t'|s_{t-1}'}P_{s_t'|s_{t-1}}P_{s_t|s_{t-1}'}P_{s_t|s_t'}P_{s_t'|s_t}P_{s_t}(o_t)P_{s_t'}(o_t') \quad (4)$$

where $P_{s_t|s_t'}$ and $P_{s_t'|s_t}$ denote the inter-user spatial state transition probabilities (constraints can be modeled with zero or low probability values) at the same time instant.

4 Solving the coupled activity model

Having defined the CHMM, we now discuss how we can solve this model to infer the 'hidden' context variables for multiple occupants simultaneously. Unlike prior work (-Brand 1996) which only considers the conditional probabilities in time (i.e., the likelihood of an individual to exhibit a specific context value at time t, *given* the context value at time $t-1$), we consider both the spatial effect on conditional probabilities (coupled across users) as well as the additional *constraints* imposed by the joint observation of smartphone and ambient sensor data. We first show (using the case of two simultaneous occupants as a canonical example) how to prune the possible state-space based on the spatiotemporal constraints. We then propose an efficient dynamic programming algorithm for multiple users, based on forward–backward analysis (Rabiner 1989) to train a model during the training phase and subsequently describe a modified Viterbi algorithm to infer context during the regular testing phase.

[1]We interchangeably use *Context* as a state s in our HMM model. For brevity we denote $\text{Context}^i(t)=s_t$ and $\text{Context}^j(t)=s_t'$ in equations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

4.1 State space filtering from spatial/temporal constraints

In this section we introduce a pruning technique for accelerating the evaluation of HMMs from multiple users. By using the spatiotemporal constraints between the micro-activities (of different users) across multiple HMMs, we can limit the viable state space for the micro-activities of each individual, and thereby significantly reduce the computational complexity. Unlike existing approaches (e.g., Plotz and Flink 2004) where such pruning is performed only during the runtime estimation of states, we perform our pruning during the offline building of the CHMM as well.

To illustrate our approach, consider the state-trellis for two users, User-1 and User-2, illustrated in Fig. 4. In this figure, User-1 is assumed (for illustration purposes) to have 3 possible values for its context tuple [i.e., (postural activity, location)] at each time instant, whereas User-2 is assumed to have four such values for her context tuple; each such context tuple is denoted by a node q_i in the trellis diagram. Assume that User-1's postural activity (inferred from the smartphone accelerometer) at time $t-2$ is 'sitting', while User-2's postural activity equals 'standing'. Furthermore, we observe that the *living room* infrastructure sensor was activated at time stamp $t-2$, indicating that the *living room* was occupied at $t-2$. In this case, of the three possible values: {(*sitting*, *living room*), (*sitting*, *bathroom*), (*sitting*, *kitchen*)} in the trellis for User-1, only the (*sitting*, *living room*) state is possible at time $t-2$. Likewise, of the four possible values: {(*standing*, *living room*), (*standing*, *bathroom*), (*standing*, *kitchen*), (*walking*, *corridor*)} for User-2, only the (*standing*, *living room*) state is possible. Clearly, in this case, the ambient context has enabled us to prune the state space for each user unambiguously.

Continuing the example, imagine now that two infrastructure sensors, say *kitchen* and *living room*, are observed to be triggered at time stamp $t-1$, while User-1's postural activity remains 'sitting', while User-2's activity is now 'walking'. In this case, while an individual HMM may allow ($2 \times 2 =$) 4 possible state pairs (the Cartesian product of {(*sitting*, *kitchen*), (*sitting*, *livingroom*)} for User-1 and {(*walking*, *kitchen*), (*walking*, *livingroom*)} for User-2), our coupled HMM spatially permits the concurrent occurrence of only some of these context states (namely, the ones where both User-1 and User-2 inhabit different rooms). In effect, this reduces the possible set of concurrent context states (for the two users) from 4 to 2. Furthermore, now considering the *temporal* constraint, we note that User-1 cannot have the state (*sitting*, *kitchen*) at time $t-1$, as she cannot have changed location while remaining in the 'sitting' state across ($t-2$, $t-1$). As a consequence, the only legitimate choice of states at time $t-1$ is (*sitting*, *living room*) for User-1, and (*walking*, *kitchen*) for User-2.

Mathematically, this filtering approach can be expressed more generically as a form of *constraint reasoning*. In general, we can limit the temporal constraint propagation to K successive instants. If each of the N individuals in the smart environment have M possible choices for their context state at any instant, this constraint filtering approach effectively involves the creation of a *K-dimensional* binary array, with length $M \times N$ in each dimension, and then applying the reasoning process to mark each cell of this array as either 'permitted' or 'prohibited'. In practice, this process of exhaustively evaluating all possible $(M \times N)^K$ choices can be significantly curtailed by both (a) starting with those time instants where the

Author Manuscript

context is deterministic (in our example, the $t-2$ choices are unambiguous as shown in Fig. 4) and (b) keeping the dimension K small (for our experimental studies, $K = 2$ provided good-enough results).

4.2 Model likelihood estimation

To intuitively understand the algorithm, consider the case where we have a sequence of T observations (T consecutive time instants), with $\bar{M}$ underlying states (reduced from the $M \times N$ original states by the pruning process) at each step. As shown in Fig. 4, this reduced trellis can be viewed as a matrix of *Context* tuple, where $\alpha[i, t]$ is the probability of being in *Context* tuple i while seeing the observation at t. In case of our coupled activity model, to calculate the model likelihood $P(O|\lambda)$, where λ = (transition, emission probabilities), two state paths have to be followed over time considering the temporal coupling, one path keep track of the head, probable *Context* tuple of User 1 in one chain (represented with subscript h) and the other path keep track of the sidekick, *Context* tuple of User 2 with respect to this head in another chain (represented with subscript k) as shown in Fig. 4. First, for each observation O_t, we compute the full posterior probability $\alpha^*[i, t]$ for all context streams i considering all the previous trellis $\alpha^*[j, t-1]$ in User 1 and inter-chain transition probabilities of sidekick trellis for User 2 (line 14 in Fig. 5).

In each step of the forward analysis we calculate the *maximum a posterior* (MAP) for $\{Context^i(t), Context^{j'}(t-1) = head, sidekick\}$ pairs given all antecedent paths. Here there are multiple trellises for a specific user. We use i, j for User 1 and i', j' for User 2, where h_i, h_j and $k_i, k_j \in Context^i, Context^j$ and $k_{i'}, k_{j'}$ and $h_{i'}, h_{j'} \in Context^{i'}, Context^{j'}$. Every $Context^i(t)$ tuple for User 1 sums over the same set of antecedent paths, and thus share the same $Context^{j'}(t-1)$ tuple as a sidekick from User 2. We choose the $Context^{j'}(t-1)$ tuple in User 2 that has maximum marginal posterior given all antecedent paths as a sidekick (line 10 in Fig. 5). In each chain, we choose the MAP state given all antecedent paths. This is again taken as a sidekick to heads in other chains. We calculate a new path posterior given antecedent paths and sidekicks for each head. We marginalize the sidekicks to calculate the forward variable $\alpha[i, t]$ associated with each head (line 18 in Fig. 5). This forward analysis algorithm pseudocode is articulated in Fig. 5 and explained with a pictorial diagram in Fig. 4 where $h_{i,t}$ and $k_{i,t}$ represents the heads and sidekicks indices at each time stamp t, $\alpha^*[i, t]$ is the probability mass associated with each head and $pp[i, t]$ is the partial posterior probability of a state given all $\alpha^*[j, t-1]$.

4.3 Determination of most-likely activity sequence

Subsequent to state pruning and model likelihood determination through forward analysis, the inference of the hidden context states can be computed by the Viterbi algorithm, which determines the most likely path (sequence of states) through the trellis. Given the model constructed as described above, we then use the Viterbi algorithm to find the *most likely path* among all unpruned state paths through the trellis. For our coupled activity model, we calculate the MAP value given all antecedent paths. Given our coupled model, for each head at time t, the Viterbi algorithm must also choose an antecedent path in $t-1$ for a single HMM, as well as a sidekick in t. This can be achieved in two steps: (1) select MAP sidekicks in t for each antecedent path in $t-1$ and (2) select the antecedent path and associated sidekick that maximizes the new head's posterior for each head in *t*. Figure 6 presents the pseudocode for our modified Viterbi algorithm, developed for multi-inhabitant environments.

5 Implementation and results

In this section, we report on our experiments that investigate the benefit of this proposed approach for recognizing complex ADLs using a combination of smartphone and simple ambient testing. Our experiments are conducted using ten participants at the WSU CASAS smart home.

5.1 Data collection

To validate our approach, we collected data from ten subjects (a.k.a PUCK dataset), each of whom carried an Android 2.1 OS based Samsung Captivate smart phone (containing a tri-axial accelerometer and a gyroscope) (Dernbach et al. 2012). Each subject carried the phone while performing different ADL. The location and orientation of the phone was not standardized and was left to the convenience of the subject though most of the users were encouraged to keep it in their pant pockets. However, orientation information was taken into consideration while extracting the different data features. We utilized a custom application on the phone to collect the corresponding accelerometer sensor data; while the accelerometer sampling rate could be varied if required, our studies are conducted based on a sampling frequency of 80 Hz. In tandem, we also collected data from ceiling-mounted infrared motion sensors (embedded as part of the SHIMMER platform), providing us a combination of concurrent smartphone and ambient sensor data streams. Using a smartphone-based application, subjects could stop and start the sensor data that was being collected, as well as manually input the activity they were about to perform. As each individual performed these tasks separately from the others, the multi-user sensor stream (for the ambient sensors) was then obtained by synthetically combining (for each time slot) the readings from all the simultaneously activated ambient sensors. We superimposed the data-traces of two randomly chosen users to generate the multi-user sensor data streams.

5.2 Enumeration of activities

Consistent with our proposed two-tier architecture, the activities we monitored consist of two types: (1) *low-level* (or *micro*): these consist of the postural or motion activities that can be classified by a phone-mounted accelerometer. For our study, the micro-activity set consisted of six labels: {*sitting*, *standing*, *walking*, *running*, *lying*, *climbing stairs*}. (2) *High-level* (or *complex*): these consisted of semantically meaningful ADLs, and included six labels:

- *Cleaning* Subject wiped down the kitchen counter top and sink.
- *Cooking* Subject simulated cooking by heating a bowl of water in the microwave and pouring a glass of water from a pitcher in the fridge.
- *Medication* Subject retrieved pills from the cupboard and sorted out a week's worth of doses.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- *Sweeping* Subject swept the kitchen area.
- *Washing hands* Subject washed hands using the soap in the bathroom.
- *Watering plants* Subject filled a watering can and watered three plants in living room.

Note that each instance of the ADL had definite (start, end) times, manually annotated by each subject. Thus, in this paper, we assume that we have a priori knowledge of the exact mapping between an instance of a complex activity and the underlying set of micro-activities. The subjects repeated execution of these complex activities four times.

5.3 Micro-activity classification

Our goal is to apply feature-based classification techniques for the micro-activities, and then apply the micro-activity stream in a two-tier manner to understand the impact on complex activity classification. To classify the micro-activities, the 3-axis accelerometer streams and the 3-axis gyroscope data were broken up into successive *frames* (we experimented with frame lengths of {1, 2, 4, 8, 12} s and report results here for the representative case of 4 s), and a 30-dimensional feature vector (see Table 1) was computed over each frame. The ground-truth annotated training set (aggregated across all ten users) was then fed into the Weka toolkit (Witten and Frank 1999) and used to train six classifiers: multi-layer perceptron, naive Bayes, Bayesian network, decision table, best-first tree, and K-star. The accuracy of the classifiers was tested using tenfold cross-validation. Figure 7 plots the average classification accuracy for the micro-activities: we see that, except for naive Bayes, all the other classifiers had similar classification accuracy of above 90 %. Our experimental results confirm that the smartphone-mounted sensors indeed provide accurate recognition of the low-level micro-activities. For subsequent results, we utilize the best-first tree classifier (as this provides the best results for the Naive-Mobile approach described in Sect. 5.6).

5.4 Location classification

As explained previously, the subject's indoor location is the 'hidden context' state in our studies. Accordingly, we fed the combination of individual-specific micro-activity streams features (not accelerometer sensor features as shown in Table 1 but micro activity features as explained as Naive-SAMMPLE in Sect. 5.6) and the infrastructure (motion sensor) specific location feature into our activity recognition (ar version 1.2) code (Activity Recognition Code 2014) on our multi-user datasets to train each individual HMM model. Our Viterbi algorithm then operates on the test data to infer each subject's most likely location trajectory. Figure 8 reports on the accuracy of the location estimate of 4 individuals randomly chosen. The location accuracy for each individual user has been calculated by taking the average of all pair-wise combinations of that specific user. The standard deviation of the average location accuracy of all the users remains within 2–4% as shown in Fig. 8. We see that our use of additional intra-person and inter-person constraints results in an overall accuracy of room-level location inference of approx. 72 % on average. In contrast, given the presence of multiple occupants, a naive strategy would be able to declare the location unambiguously for only those instants where either (a) only one inhabitant was present in the smart home, or (b) all the occupants were located in the same room. We found this to be the case for only ≈5–

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

6 % of our collected data set, implying that our constrained coupled-HMM technique is able to achieve over a 12-fold increase in the ability to meaningfully infer individual-specific location.

5.5 Micro activity feature assisted macro activity classification

The goal of this study is to take the data frames gathered from users performing complex "macro" activities and reclassify them as simple "micro" activities (as referred as Naive-Mobile approach described in Sect. 5.6). We count the occurrences of micro activities within a certain window size, and use those counts as attributes in a new data frame. We perform the micro-classification step because micro classifications tend to be more accurate than trying to directly classify macro activities. In other words, we are attempting to identify the micro activities that make up a more complex activity and use these to correctly classify the complex activity being performed. The first set of experiments was performed on the PUCK data set, collected at WSU as referred before. The composition of the PUCK set activity frames are shown in Tables 2 and 3.

We created an application which reads a file containing data for a single complex activity, averages the frames within a window size of 5, prints the averaged frames to a new condensed file as unlabeled frames, reclassifies the condensed, unlabeled frames as micro activities using a classifier trained by our micro training set, counts the occurrences of micro activities within a second window size of 4, 8, or 12, and prints out these numbers of occurrences as attributes in a new frame with the original complex activity as a label. We perform this process on each individual complex activity and then combine the resulting files into one file. We then use this file as a training set for multiple classifiers and test the classifiers on the training set.

Figure 9 shows results from the PUCK study data. We notice that there is a strong bias in the model towards classifying complex frames as the micro activity '*climbing*', and so we design a second experiment in which we remove '*Climbing*' from the training set to observe if there is an improvement. Figure 10 represents that the classification accuracy has been improved from 40 to 50 % by excluding the micro activity '*Climbing*'. While the confusion in the classifier is albeit reduced by removing the activity '*climbing*', but a new bias towards the micro activity '*running*' has become apparent. Removing the attribute '*running*' resulted in a loss of accuracy, suggesting a certain minimum number of micro-activities are necessary to model the macro-activities.

We conclude that the micro activities included in our training set do not correctly represent the complex activities which we are attempting to model. Many of our macro activities involve a large amount of hand motion whereas our micro activities are predominately foot or lower body motions. We confirm this by isolating a few of the more distinctive macro activities which could possibly be modeled with the micro activities available. Of course, the improved accuracies must stem in part from the reduced number of classification options. However, from these results it appears that the micro activities we have do not correctly represent the micro activities that make up the complex activity. With the goal of correcting this, we repeat our experiments on a second data set, Activity Recognition Challenge-Opportunity dataset (Activity Recognition Challenge 2013).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

5.5.1 Opportunity dataset—We tried our approach on the Opportunity data set (Activity Recognition Challenge 2013). The opportunity dataset comprises the readings of motion sensors recorded while users executed typical daily activities at the micro and macro-level as shown in Tables 4 and 5. The following sensors and recordings have been used to benchmark human activity recognition algorithms (Chavarriaga et al. 2013).

- Body-worn sensors: seven inertial measurement units, 12 3D acceleration sensors, four 3D localization information.
- Object sensors: 12 objects with 3D acceleration and 2D rate of turn.
- Ambient sensors: 13 switches and eight 3D acceleration sensors.
- Recordings: four users, six runs per users.

We used a single sensor from the Opportunity set to simulate the data which we would be able to collect using a phone—specifically the Inertial Measurement Unit. The raw data in the Opportunity set was preprocessed by selecting a time window in which to calculate features used as attributes in our arff frames. These include the mean, maximum, minimum, standard deviation, and correlation of the acceleration and orientation measures on the x, y, and z axes over the chosen time frame. The preprocessed data frames from the Opportunity set has been shown in Tables 4 and 5.

Figure 11 represents a significant improvement (almost a twofold increment) in micro-feature assisted macro activity classification. We believe here the micro activities were more closely related to the macro activities and thus help produced the improved results. However there still exists confusion between similar activities like *coffee making* and *sandwich making*. It is possible that difference in the micro activities which make up these similar complex activities are primarily small variations in the hand motions which are not captured by phone sensors.

We conclude that in order to achieve reliable accuracy levels, our approach must ensure that our micro activities accurately model the macro activities and that the macro activities are distinct enough for phone sensors to pick up differences in the data. The improvement in accuracies with the Opportunity data set over the PUCK data set suggests that a broader range of micro activities including distinctive hand motions would be helpful to accurately classifying the complex activities in question.

5.6 Macro/complex activity classification

Finally, we investigate the issue of whether this infrastructure-assisted activity recognition approach really helps to improve the accuracy of complex activity recognition. In particular, we experimented with four different strategies, which differ in their use of the additional infrastructure assistance (the motion sensor readings) and the adoption of a one-tier or two-tier classification strategy:

1. *Naive-Mobile (NM)* In this approach, we use only the mobile sensor data (i.e., accelerometer and gyroscope-based features) to classify the complex activities. More specifically, this approach is similar to the step of micro-activity classification in that the classifier is trained with features computed over individual frames, with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the difference lying in the fact that the training set was now labeled with the *complex activity label*.

2. *Naive-SAMMPLE (NS)* In this two-tier approach, we essentially replicate the approach in Yan et al. (2012). In this approach, instead of the raw accelerometer data, we use the stream of inferred micro-activity labels as the input to the classifier. More specifically, each instance of a complex activity label is associated with a six-dimensional feature-vector consisting of the number of frames (effectively the total duration) of each of the six micro-activities considered in our study. For example, if an instance of 'cooking' consisted of three frames of 'sitting', four frames of 'standing' and seven frames of 'walking', the corresponding feature vector would be [3 4 7 0 0 0], as the last three micro-activities do not have any occurrences in this instance of 'cooking'.

3. *Infra-Mobile (IM)* This is the first infrastructure-augmented approach. Here, we associate with each *frame* of complex activity instance, a feature vector corresponding to the accelerometer data, plus the location estimated by our Viterbi algorithm. This is effectively a one-tier approach, as we try to classify the complex activity directly based on accelerometer features.

4. *Infra-Mobile-SAMMPLE (IMS)* This combines both the two-tier classification strategy and the additional 'location' context inferred by our Viterbi algorithm. This is effectively an extension of the Naive-SAMMPLE approach, in that we now have a seven-dimensional feature vector, with the first six elements corresponding to the frequency of the underlying micro-activities and the 7th element corresponding to the indoor location inferred by our Viterbi algorithm.

Figure 12 plots the accuracy of the different approaches (using tenfold cross validation) for a randomly selected set of five subjects. (The other subjects have similar results and are omitted for space reasons.) We see, as reported in prior literature, that classifying complex activities (which can vary significantly in duration and in the precise low-level activities undertaken) is very difficult using purely phone-based features: both Naive-Mobile and Naive-SAMMPLE report very poor classification accuracy—an average of 45 and 61 %, with values as low as 35 and 50 % respectively. In contrast, our ability to infer and provide the room-level location in the smart home setting leads to an increase (over 30 %) in the classification accuracy using the one-tier Infra-Mobile approach, as high as 79 %. Finally, the Infra-Mobile-SAMMPLE approach performs even better by using micro-activity features for classification, attaining classification accuracy as high as 85 %. The results indicate both the importance of location as a feature for complex ADL discrimination in smart homes (not an unexpected finding) and the ability of our approach to correctly infer this location in the presence of multiple inhabitants (a major improvement). Figure 13 represents the standard errors depicting a measure of variability of the sampling distributions of Naive-Mobile, Naive-SAMMPLE, Infra-Mobile and Infra-Mobile-SAMMPLE approach across the five users. The confidence intervals for Naive-Mobile, Naive-SAMMPLE, Infra-Mobile, and Infra-Mobile-SAMMPLE has been varied in between 97.24 and 96:78 %.

Table 6 provides the best-first tree confusion matrix for the six pre-defined complex activities, for both the Naive-Mobile approach and our suggested Infra-Mobile-SAMMPLE

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

approach. We can see that pure locomotion/postural features perform very poorly in classifying complex activities (such as medication, washing hands or watering plants) in the absence of location estimates; when augmented with such location estimates, the ability to classify such non-obvious activities improves.

5.7 Multiple users complex activity recognition

We also report the complex activity recognition accuracy in presence of multiple users. The observation sequences obtained from each user based on the ambient and smartphone sensors are used for training and testing purpose. The feature vectors from these observation sequences are derived based on our one tier and two-tier classification respectively for Infra-Mobile and Infra-Mobile-SAMMPLE approach along with the location information as obtained from our constraint reasoning. The activity models for each user are trained and built separately using the corresponding feature vectors. In case of multiple users the features vectors corresponding to all the users are fed into the CHMM model. The model is first trained from multiple training sequences associated with multiple users using the forward–backward algorithm and then tested in presence of combined testing sequences to infer the most likely activities of individual users based on the Viterbi algorithm. Figure 14 plots the complex activity recognition accuracy in case of Infra-Mobile and Infra-Mobile-SAMMPLE based methods considering a group of two users jointly. Figures 15 and 16 also plot the complex activity recognition accuracy respectively for a group of three and four users respectively. We do observe that as the number of users being considered in a group has been increased the complex activity recognition accuracy for the Infra-Mobile and Infra-Mobile-SAMMPLE based approaches has been decreased. Nevertheless, the Infra-Mobile-SAMMPLE approach performs better by using the micro-activity features and location metric obtained from the Viterbi algorithm, attaining a classification accuracy as high as 90 % in presence of a group of two, three and four users in the smart home setting. Figures 17, 18 and 19 plot the confidence interval for complex activity recognition accuracy in case of Infra-Mobile and Infra-Mobile-SAMMPLE based methods in presence of a group of two, there and four users respectively. It is noted that the confidence interval has been deteriorated for both the ambient-augmented mobile sensing based approaches, namely Infra-Mobile and Infra-Mobile-SAMMPLE, as the number of users inhabiting in the smart home environment has been increased.

In our future work we have been planning to exploit the correlations between multiple users to prune the state space model. We plan to combine the spatiotemporal correlations and constraints across multiple users to build an effective state space model for our activity model building, training and testing phase. While a rule based mining approach helps build the correlations in presence of multiple users and their contexts, a probabilistic graphical model will be augmented with the prior to jointly harness the state space reduction as the number of users increase in a smart home environment.

5.8 Computation complexity of Viterbi algorithm

We now report some micro-benchmark results on the performance of the Viterbi algorithm. In particular, we show the performance of our constrained pruned-HMM approach and evaluate it using two metrics: (a) estimation accuracy, measured as the log likelihood of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

resulting model predictions (effectively indicating how much improvement in accuracy the constraint-based pruning provides). (b) Execution speed (effectively indicating how much computational overhead may be saved by our pruning approach).

Figure 20 depicts the training and testing log-likelihoods of our coupled model which establishes that train-test divergence is very minimal. Figure 21 shows the computation time of our algorithms with a fixed number of states and increasing number of data sequences, whereas Fig. 22 plots the computation time with a fixed number of data sequences and increasing number of states. Clearly, pruning the state space can reduce the computational overhead. For example, if the joint number of states is reduced from $10 \times 10 = 100$ to $7 \times 7 = 49$, we would obtain a fivefold savings in computation time ($2500 \rightarrow 500$ ms).

6 Conclusions

In this work, we have outlined our belief in the practicality of a hybrid mobile-cum-infrastructure sensing for *multi-inhabitant* smart environments. This combination of smart-phone-provided personal micro-activity context and infrastructure-supplied ambient context allows us to express several unique constraints, and show how to use these constraints to simplify a coupled HMM framework for the evolution of individual context states. Results obtained using real traces from a smart home show that our approach can lead to ~ 70 % accuracy in our ability to reconstruct individual-level *hidden* micro-context ('room-level location'). This additional context leads to significant improvements in the accuracy of complex ADL classification.

These initial results are promising. However, we believe that the additional sensors on smartphones can provide significantly richer observational data (for individual and ambient context). We plan to explore the use of the smartphone audio sensor to enable capture of different 'noise signatures' (e.g., television, vacuum cleaner, human chat); such additional micro-context should help to further improve the accuracy and robustness of complex ADL recognition.

Acknowledgments

The work of Nirmalya Roy is partially supported by the National Science Foundation Award #1344990 and Constellation E^2: Energy to Educate Grant. The work of Archan Misra is partially supported by the Singapore Ministry of Education Academic Research Fund Tier 2 under research Grant MOE2011-T2-1-001. The work of Diane Cook is partially supported by NSF Grants 1064628, 0852172, CNS-1255965, and NIH Grant R01EB009675.

References

Acampora G, Cook D, Rashidi P, Vasilakos A. A survey on ambient intelligence in healthcare. Proc IEEE. 2013; 101(12):2470–2494.

[Accessed June 2013] Activity Recognition Challenge. 2013. http://www.opportunityproject.eu/challengeDataset

[Accessed Jan 2014] Activity Recognition Code. 2014. http://ailab.wsu.edu/casas/ar/

Alam, M.; Pathak, N.; Roy, N. Mobeacon: an iBeacon-assisted smartphone-based real time activity recognition framework; Proceedings of the 12th international conference on mobile and ubiquitous systems: computing, networking and services; 2015. in press

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Alam M, Roy N. Gesmart: a gestural activity recognition model for predicting behavioral health. Proceeding of the IEEE international conference on smart computing. 2014

Almashaqbeh G, Hayajneh T, Vasilakos A, Mohd B. QoS-aware health monitoring system using cloud-based WBANs. J Med Syst. 2014; 38(10):121. [PubMed: 25123456]

[Accessed Jan 2015] Android Wear: Information that Moves with You. 2015. http://googleblog.blogspot.co.uk/2014/03/sharing-whats-up-our-sleeve-android.html

Bergmann J, McGregor A. Body-worn sensor design: what do patients and clinicians want? Ann Biomed Eng. 2011; 39(9):2299–2312. [PubMed: 21674260]

Brand, M. Technical report 405. MIT Lab for Perceptual Computing; 1996. Coupled hidden Markov models for modeling interacting processes.

Chavarriaga R, Sagha H, Calatroni A, Digumarti S, Trster G, Milln J, Roggen D. The opportunity challenge: a benchmark database for on-body sensor-based activity recognition. Pattern Recognit Lett. 2013; 34(15):2033–2042.

Chen L, Hoey J, Nugent C, Cook D, Yu Z. Sensor-based activity recognition. IEEE Trans Syst Man Cybern-Part C. 2012; 42(6):790–808.

Chen M, Gonzalez S, Vasilakos A, Cao H, Leung V. Body area networks: a survey. MONET. 2011; 16(2):171–193.

Clarkson B, Mase K, Pentland A. Recognizing user context via wearable sensors. Proceedings of the 4th international symposium on wearable computers. 2000

Dernbach S, Das B, Krishnan N, Thomas B, Cook D. Simple and complex acitivity recognition through smart phones. Proceedings of the international conference on intelligent environments. 2012

Feng Z, Zhu Y, Zhang Q, Ni L, Vasilakos A. Trac: truthful auction for location-aware collaborative sensing in mobile crowdsourcing. INFOCOM. 2014:1231–1239.

Fortino G, Fatta G, Pathan M, Vasilakos A. Cloud-assisted body area networks: state-of-the-art and future challenges. Wirel Netw. 2014; 20(7):1925–1938.

Gong S, Xiang T. Recognition of group activities using dynamic probabilistic networks. Proceedings of international conference on computer vision. 2003

Gyorbiro N, Fabian A, Homanyi G. An activity recognition system for mobile phones. Mob Netw Appl. 2008; 14(1):82–91.

Hayajneh T, Almashaqbeh G, Ullah S, Vasilakos A. A survey of wireless technologies coexistence in wban: analysis and open research issues. Wirel Netw. 2014; 20(8):2165–2199.

Hossain H, Roy N, Khan M. Sleep well: a sound sleep monitoring framework for community scaling. Proceeding of the IEEE international conference on mobile data management. 2015

[Accessed Feb 2015] Huawei Smart Bracelet. 2015. http://www.huawei.com/us/index.htm

Huynh T, Blanke U, Schiele B. Scalable recognition of daily activities from wearable sensors. LNCS LoCA. 2007; 4718

[Accessed Apr 2014] Intel Make it Wearable. 2014. https://makeit.intel.com/

Intille S, Larson K, Tapia E, Beaudin J, Kaushik P, Nawyn J, Rockinson R. Using a live-in laboratory for ubiquitous computing research. Proceedings of 4th international conference on pervasive computing. 2006; 3968

Kasteren T, Noulas A, Englebienne G, Krose B. Accurate activity recognition in a home setting. Proceedings of the 10th international conference on ubiquitous computing. 2008; 3968

Khan, M.; Hossain, H.; Roy, N. Infrastructure-less occupancy detection and semantic localization in smart environments; Proceedings of the 12th international conference on mobile and ubiquitous systems: computing, networking and services; 2015a. in press

Khan, M.; Hossain, H.; Roy, N. Sensepresence: infrastructure-less occupancy detection for opportunistic sensing applications; IEEE international conference on mobile data management; 2015b. in press

Khan M, Lu S, Roy N, Pathak N. Demo abstract: a microphone sensor based system for green building applications. IEEE international conference on pervasive computing and communications (PerCom). 2015c

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Kwapisz, J.; Weiss, G.; Moore, S. Activity recognition using cell phone accelerometers; International workshop on knowledge discovery from sensor data; 2010.

Lee C, Hsu C, Lai Y, Vasilakos A. An enhanced mobile-healthcare emergency system based on extended chaotic maps. J Med Syst. 2013; 37(5):9973. [PubMed: 24018958]

Lester J, Choudhury T, Borriello G. A practical approach to recognizing physical activities. PERVASIVE LNCS. 2006; 3968

Lin D, Labeau F, Vasilakos A. QoE-based optimal resource allocation in wireless healthcare networks: opportunities and challenges. Wirel Netw. 2015a

Lin D, Wu X, Labeau F, Vasilakos A. Internet of vehicles for e-health applications in view of EMI on medical sensors. J Sens. 2015b

Logan B, Healey J, Philipose M, Tapia E, Intille S. A long-term evaluation of sensing modalities for activity recognition. UbiComp LNCS. 2007; 4717

Oliver N, Rosario B, Pentland A. A bayesian computer vision system for modeling human interactions. IEEE Trans Pattern Anal Mach Intell. 2000; 22(8):831–843.

Pathak N, Khan M, Roy N. Acoustic based appliance state identifications for fine grained energy analytics. IEEE international conference on pervasive computing and communications (PerCom). 2015

Philipose M, Fishkin K, Perkowitz M, Patterson D, Hahnel D, Fox D, Kautz H. Inferring activities from interactions with objects. IEEE Pervasive Comput. 2004; 3(4):50–57.

Plotz, T.; Flink, G. Tech rep. Faculty of Technology, University of Bielefeld; 2004. Accelerating the evaluation of profile hmms by pruning techniques. Report 2004-03.

Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989; 77(2):257–285.

Rahimi MR, Ren J, Liu C, Vasilakos A, Venkatasubramanian N. Mobile cloud computing: a survey, state of art and future directions. MONET. 2014; 19(2):133–143.

Rahimi MR, Venkatasubramanian N, Mehrotra S, Vasilakos A. Mapcloud: mobile applications on an elastic and scalable 2-tier cloud architecture. IEEE/ACM UCC. 2012

Roy N, Das SK, Julien C. Resource-optimized quality-assured ambiguous context mediation in pervasive environments. IEEE Trans Mob Comput. 2012; 11(2):218–229.

Roy N, Julien C. Immersive physiotherapy: challenges for smart living environments and inclusive communities. Proceeding of the 12th international conference on smart homes and health telematics. 2014

Roy N, Kindle B. Monitoring patient recovery using wireless physiotherapy devices. Proceeding of the 12th international conference on smart homes and health telematics. 2014

Roy N, Misra A, Cook D. Infrastructure-assisted smartphone-based adl recognition in multi-inhabitant smart environments. Percom. 2013:38–46.

Roy N, Misra A, Das SK, Julien C. Quality-of-inference (qoinf)-aware context determination in assisted living environments. ACM SIGMOBILE workshop on medical-grade wireless networks. 2009

Roy N, Misra A, Julien C, Das SK, Biswas J. An energy efficient quality adaptive multi-modal sensor framework for context recognition. Percom. 2011:63–73.

Roy, N.; Pathak, N.; Misra, A. Aarpa: combining pervasive and power-line sensing for fine-grained appliance usage and energy monitoring; IEEE international conference on mobile data management; 2015. in press

Roy N, Roy A, Das S. Context-aware resource management in multi-inhabitant smart homes: a nash h-learning based approach. Proceedings of IEEE international conference on pervasive computing and communications (PerCom). 2006:372–404.

Sheng Z, Yang S, Yu Y, Vasilakos A, McCann J, Leung K. A survey on the ietf protocol suite for the internet of things: standards, challenges, and opportunities. IEEE Wirel Commun. 2014; 20(6):91–98.

Wang L, Gu T, Tao X, Chen H, Lu J. Recognizing multi-user activities using wearable sensors in a smart home. Pervasive Mob Comput. 2011; 7(3):287–298.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Wilson D, Atkeson C. Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors. Pervasive Comput. 2005; 3468:62–79.

Witten, L.; Frank, E. Data mining: practicial machine learning tools and techniques with Java implementations. Morgan Kaufmann; San Francisco: 1999.

Yan Z, Chakraborty D, Misra A, Jeung H, Aberer K. Sammple: detecting semantic indoor activities in practical settings using locomotive signatures. International symposium on wearable computers. 2012

Yan Z, Zhang P, Vasilakos A. A survey on trust management for internet of things. J Netw Comput Appl. 2014; 42:37–40.

Yi-Ting C, Kuo-Chung H, Ching-Hu L, Li-Chen F, John H. Interaction models for multiple-resident activity recognition in a smart home. IROS. 2010:3753–3758.

Zhang Z, Wang H, Vasilakos A, Fang H. ECG-cryptography and authentication in body area networks. IEEE Trans Inf Technol Biomed. 2012; 16(6):1070–1078. [PubMed: 22752143]

Zheng Y, Li D, Vasilakos A. Real-time data report and task execution in wireless sensor and actuator networks using self-aware mobile actuators. Comput Commun. 2013; 36(9):988–997.

Zhou L, Xiong N, Shu L, Vasilakos A, Yeo S. Context-aware middleware for multimedia services in heterogeneous networks. IEEE Intell Syst. 2010; 25(2):40–47.

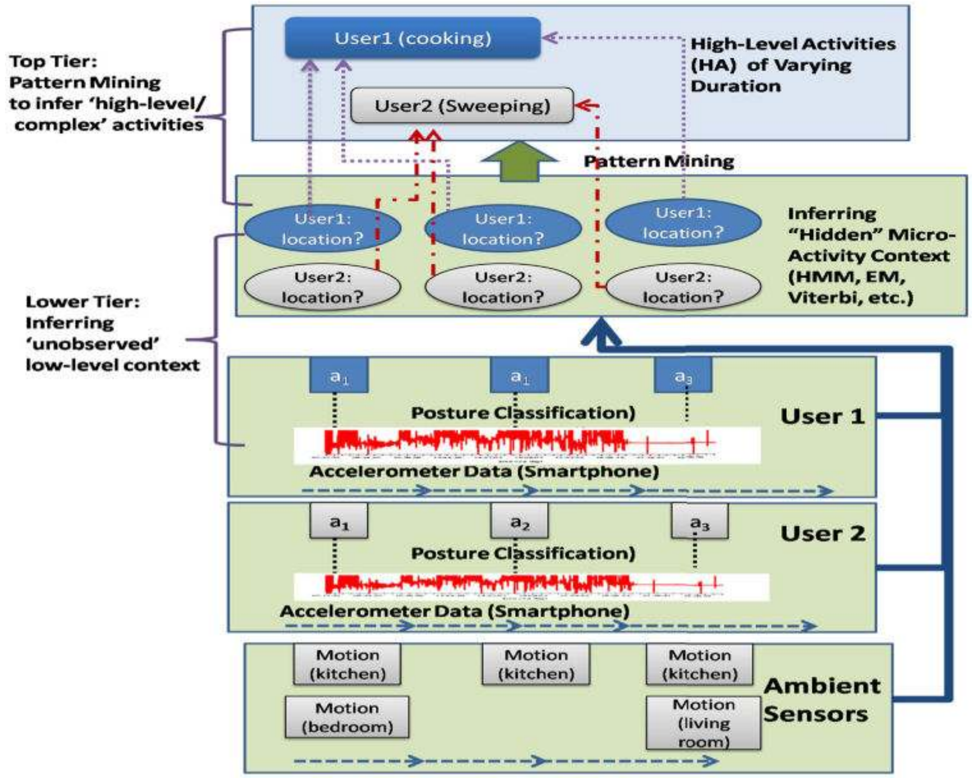

Fig. 1. Illustration of our two-tier approach to combining smartphone and ambient sensor data

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

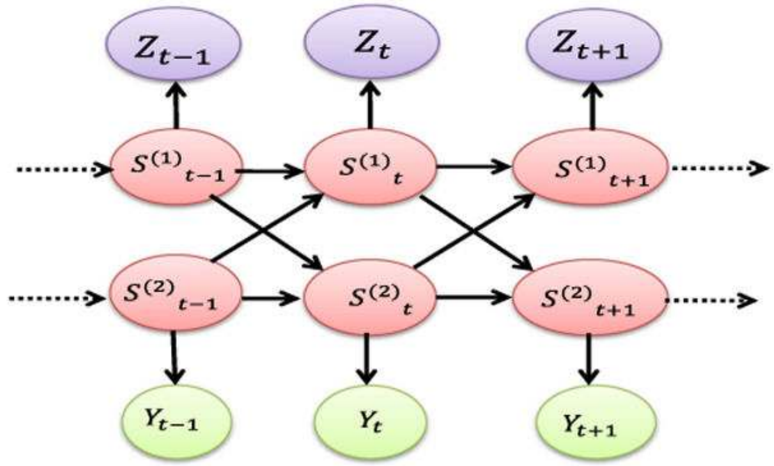

Fig. 2. CHMM structure

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

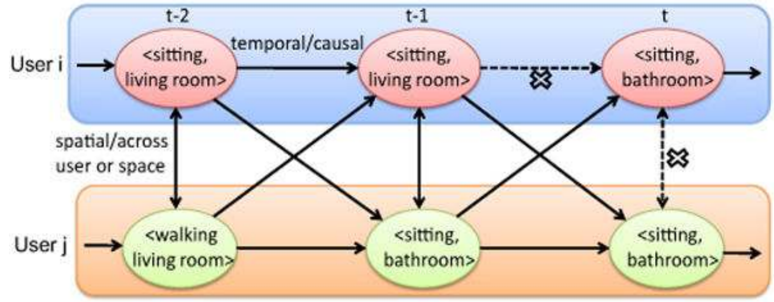

Fig. 3. CHMM with inter-user and intra-user constraints

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

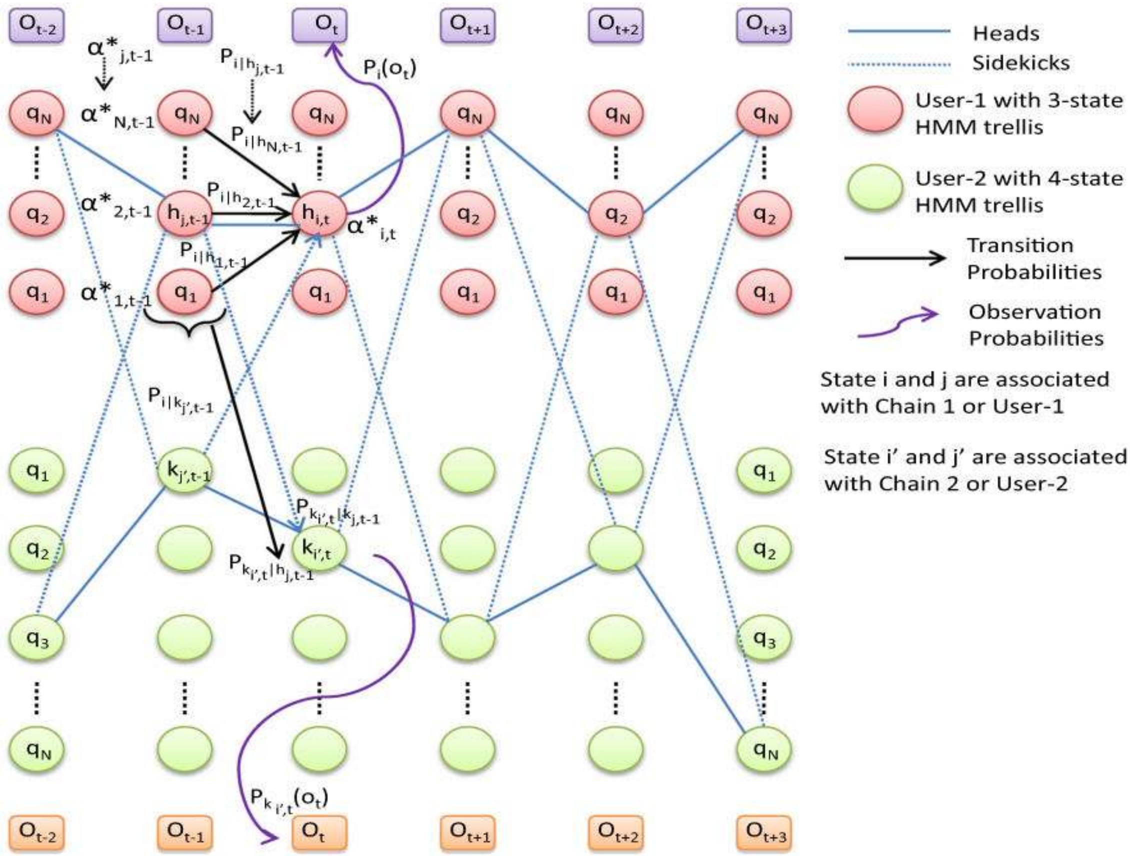

Fig. 4. Search through the state trellis of a 3-state HMM for User-1 and 4-state HMM for User-2 for state probabilities, transition, coupling and spatial probabilities and most likely path

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Procedure Forward (input: observation of length T, state-graph of length N; output: forward probability $\alpha_{i,t}$)
1. Procedure **State_Filter()**; // not shown due to lack of space
//prune state-space based on constraints
2. Initialize partial posterior probability matrix $pp[N,T]$; full posterior probability matrix $\alpha^*[N,T]$ and a forward probability matrix $\alpha[N+2,T]$
//Consider a dummy start (q_0) and final state (q_f)
3. **For** state i = 1 to N
4. $\alpha^*[i,1] \leftarrow P_{q_0|i} \times P_i(o_1)$; //start state is q_0
//Calculate all partial posteriors (pp) for selecting best sidekick (k) in each chain
5. **For** time step t = 2 to T //for T observations in chain 1
6. **For** state i = 1 to N //for $h_{i,t}$ in chain 1
7. **For** state j = 1 to N // for $h_{j,t-1}$ in chain 1
8. $pp[i,t] \leftarrow \sum_{j=1}^{N} \alpha^*[j,t-1] \times P_{i|h_j} \times P_{i|k_{j'}} \times P_i(o_t)$;
9. **For** state j'=1 to N//for sidekicks in chain 2
10. $k_{j',t-1} = arg\max_j pp[i,t]$
//best sidekick from chain 2 for a head in chain 1
//Calculate full posteriors for each path considering head $h_{i,t} = i$ and $h_{j,t-1}$ as a sidekick $k_{j,t-1}$ in chain 1 and sidekick $k_{j'|t-1}$ and $k_{i'|t}$ in chain 2
11. **For** time step t = 2 to T //for T observations in chain 1
12. **For** state i = 1 to N // chain 1
13. **For** state j = 1 to N // chain 1
14. $\alpha^*[i,t] \leftarrow \sum_{j=1}^{N} \alpha^*[j,t-1] \times P_{i|h_j} \times P_{i|k_{j'}} \times P_{k_{i'}|h_j} \times P_{k_{i'}|k_j} \times P_i(o_t) \times P_{k_{i'}}(o_t)$;
//Calculate marginalized α variables
15. **For** time step t = 2 to T //for T observations in chain 1
16. **For** state i = 1 to T //for T heads in chain 1
17. **For** state g' = 1 to N//for sidekicks in chain 2
18. $\alpha[i,t] \leftarrow \sum_{j=1}^{N} \alpha^*[j,t-1] \times P_{i|h_j} \times P_{i|k_{j'}} \times P_{k_{g'}|h_j} \times P_{k_{g'}|k_j} \times P_i(o_t) \times P_{k_{g'}}(o_t)$;
19. $\alpha[q_f,T] \leftarrow \sum_{i=1}^{N} \alpha[i,T] \times P_{i|q_f}$; //final state is q_f
20. **return** $\alpha[q_f,T]$. //likelihood

Fig. 5. Forward algorithm pseudocode for coupled activity model

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Procedure Viterbi (input: observation of length T, state-graph of length N; output: best-path)

1. Initialize a path probability matrix $viterbi[N+2, T]$ and a path backpointer matrix $backpointer[N+2, T]$
2. **For** state i = 1 to N
3. $\alpha[i, 1] \leftarrow P_{q_0|i} \times P_i(o_1)$; //forward variable
4. $backpointer[i, 1] \leftarrow 0$;

//for each antecedent path in $t-1$ select MAP sidekicks

5. **For** state j = 1 to N // for path $h_{j,t-1}$ in chain 1
6. **For** state i' = 1 to N // for sidekick in chain 2
7. $k_{i',t} = arg\max_i \alpha[i, t]$

//best sidekick from chain 2 for a head in chain 1

8. **For** time step t = 2 to T
9. **For** state i = 1 to N
10. **For** state j = 1 to N

//for each head in t, select antecedent path and sidekick that maximizes the new head's posterior

11. $viterbi[i, t] \leftarrow \max_{j=1}^{N} viterbi[j, t-1] \times P_{i|h_j} \times P_{k_{i'}|h_j} \times P_i(o_t)$;

//backpointer keeps track of whichever state was the most probable path to the current state

12. $backpointer[i, t] \leftarrow arg\max_{j=1}^{N} viterbi[j, t-1] \times P_{i|h_j} \times P_{k_{i'}|h_j}$;
13. $viterbi[q_f, T] \leftarrow \max_{i=1}^{N} viterbi[i, T] \times P_{i|q_f}$;

//final state is q_f

14. $backpointer[q_f, T] \leftarrow arg\max_{i=1}^{N} viterbi[i, T] \times P_{i|q_f}$;
15. **return** the path by following backpointers from $backpointer[q_f, T]$.

Fig. 6. Viterbi algorithm psuedocode for multiple users

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

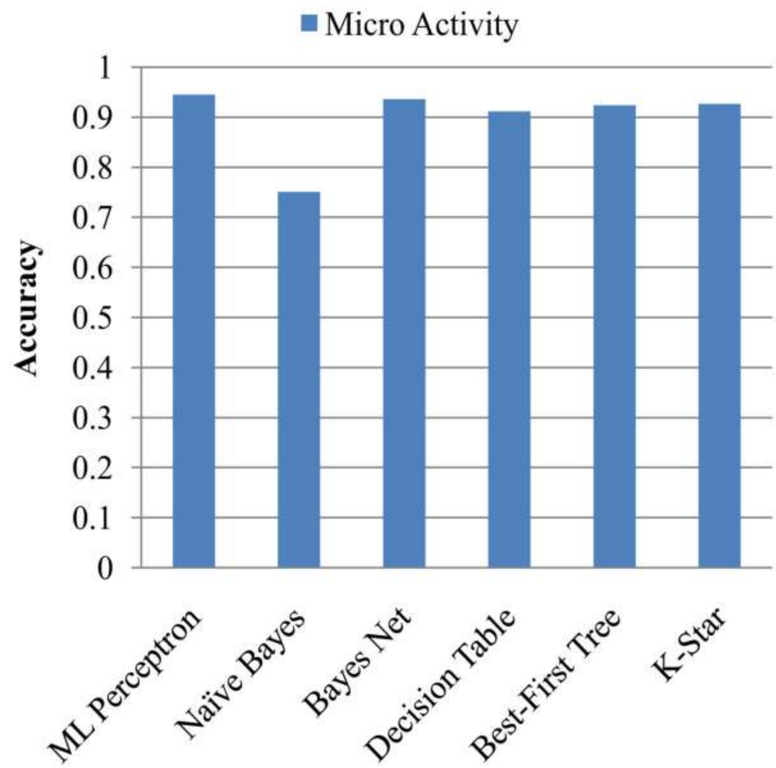

Fig. 7.
Micro-activity classification accuracy based on mobile sensing

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

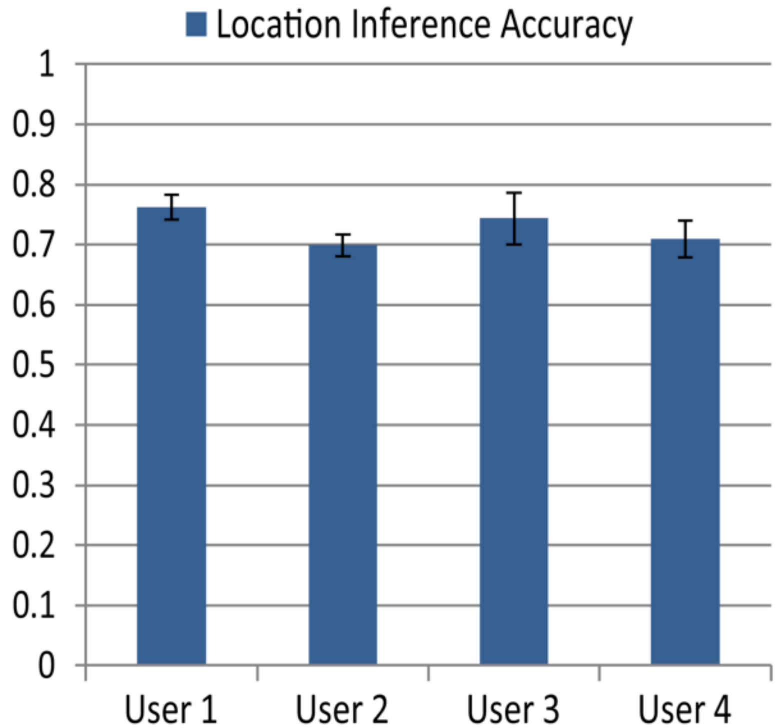

Fig. 8. Location inferencing accuracy using ambient sensor data

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

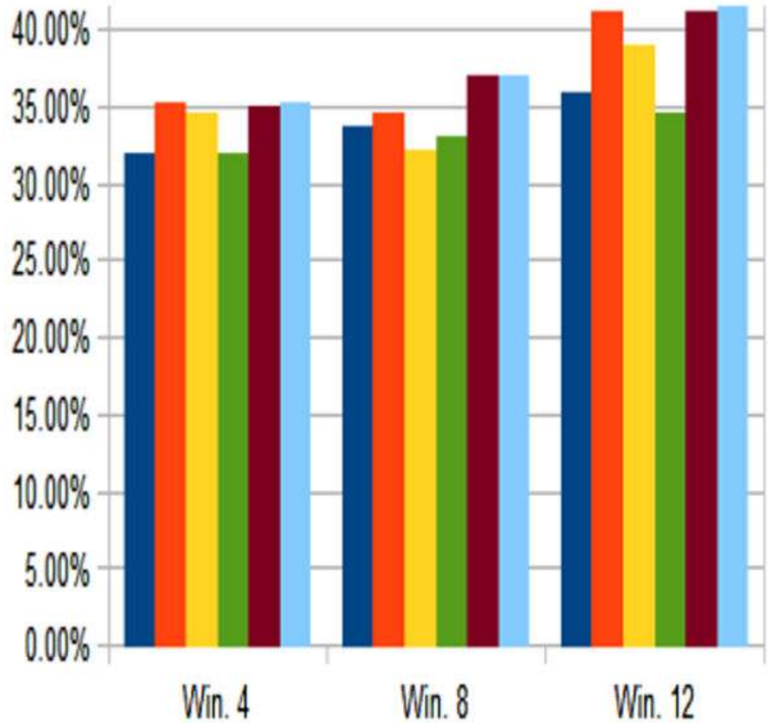

Fig. 9. Micro assisted macro activity recognition accuracy (PUCK dataset)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

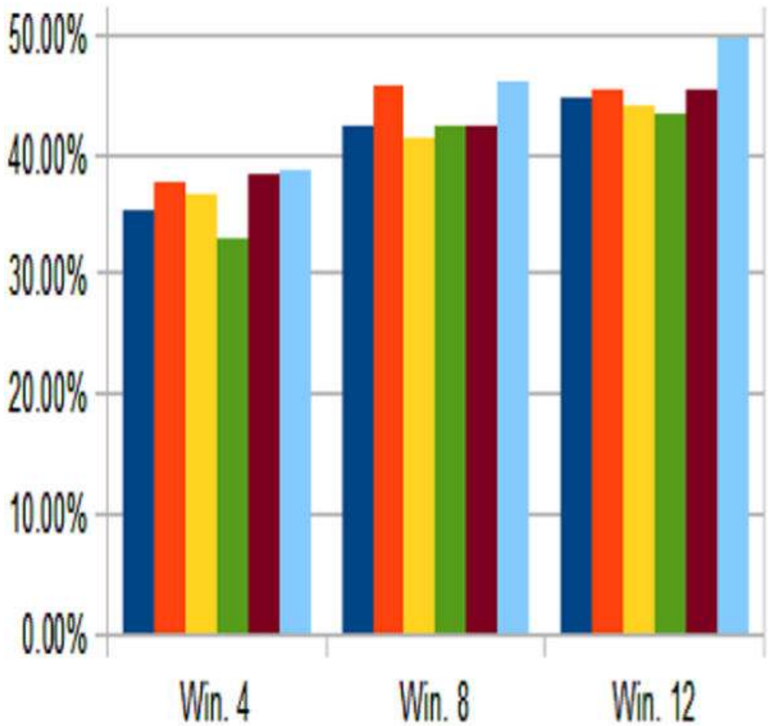

Fig. 10. Pruned micro assisted macro activity recognition accuracy (PUCK dataset)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

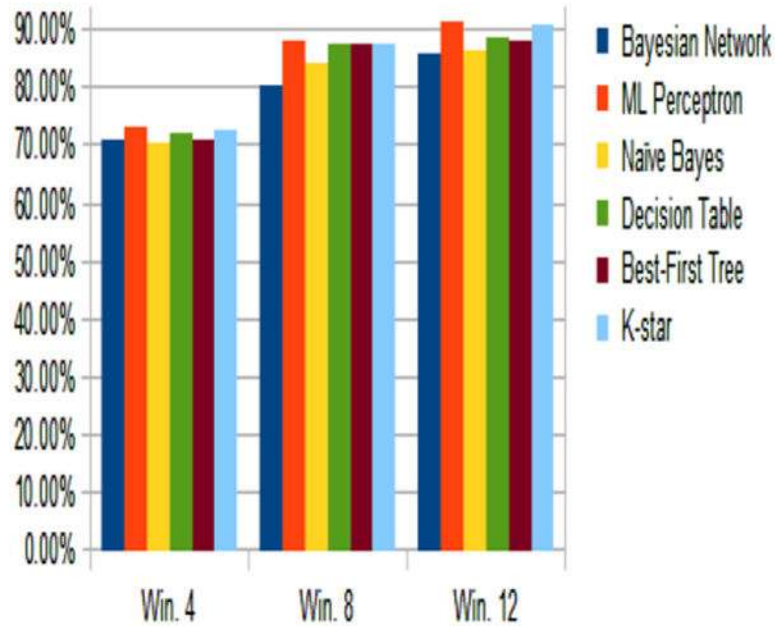

Fig. 11. Micro assisted macro activity recognition accuracy (opportunity dataset)

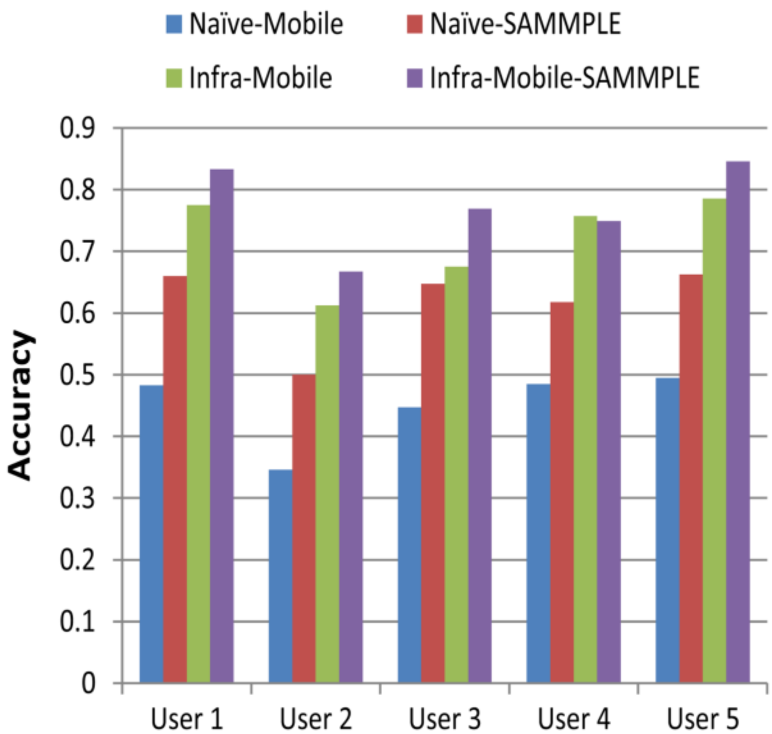

Fig. 12. Complex activity classification: mobile vs. ambient-augmented mobile sensing for multiple users

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

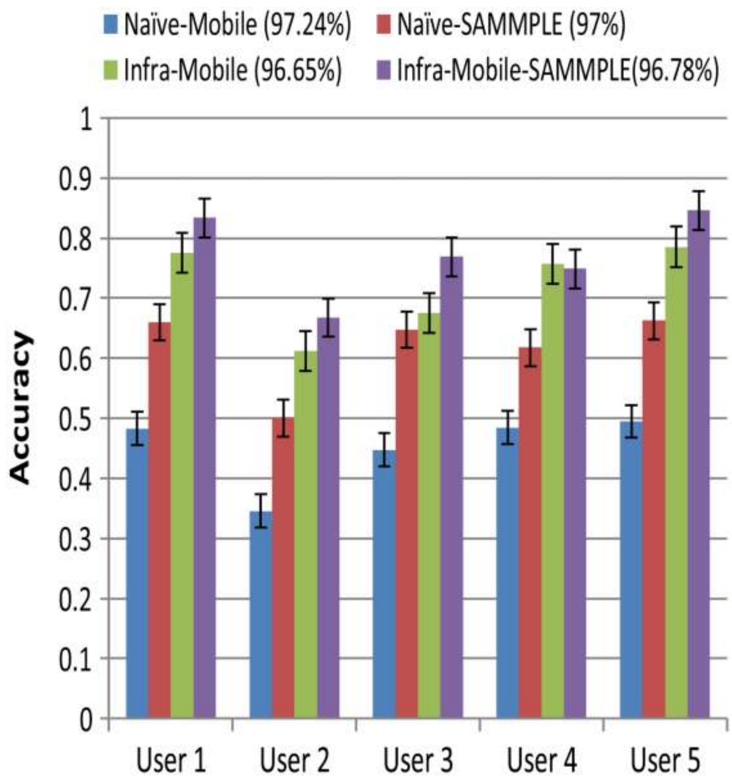

Fig. 13. Standard errors of mobile and ambient-augmented sensing for multiple users

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

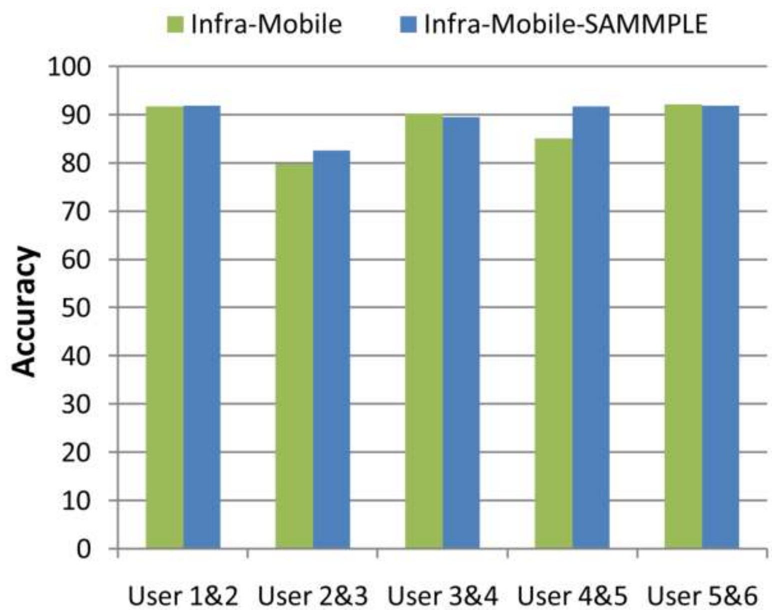

Fig. 14. Complex activity classification accuracy: ambient-augmented mobile sensing for a group of two users

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

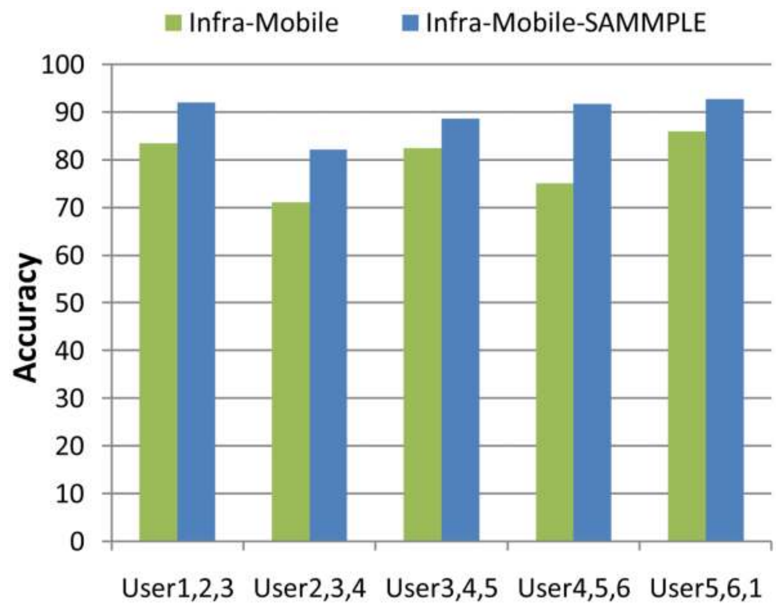

Fig. 15. Complex activity classification accuracy: ambient-augmented mobile sensing for a group of three users

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

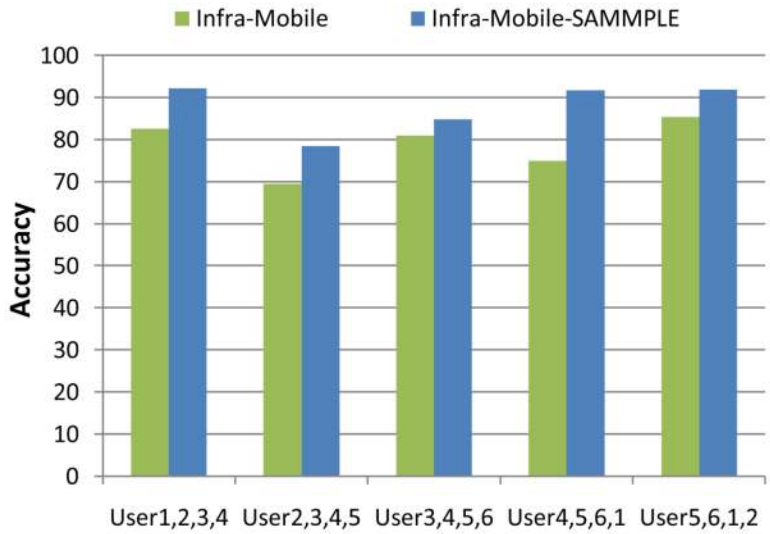

Fig. 16. Complex activity classification accuracy: ambient-augmented mobile sensing for a group of four users

Author Manuscript

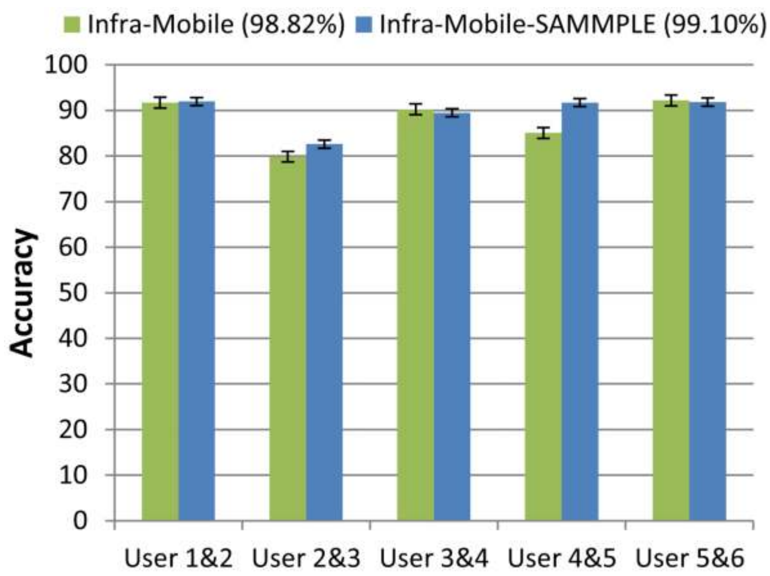

Fig. 17. Confidence interval of ambient-augmented mobile sensing for a group of two users

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

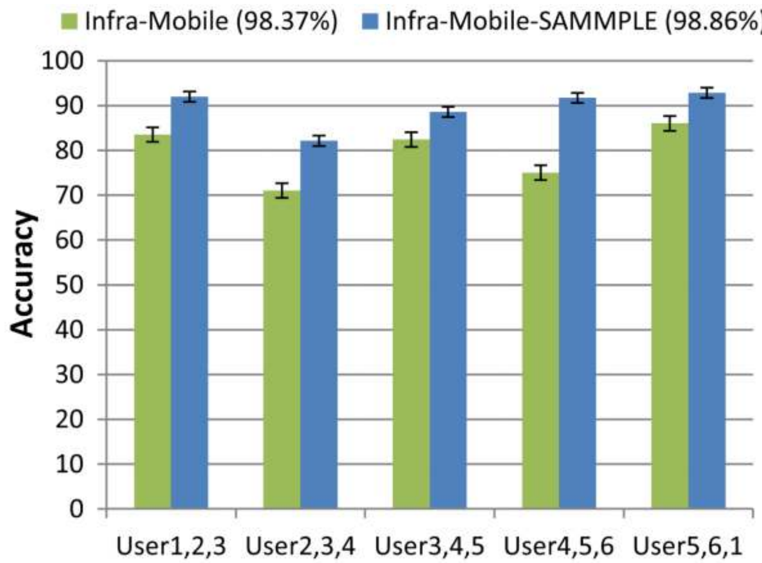

Fig. 18. Confidence interval of ambient-augmented mobile sensing for a group of three users

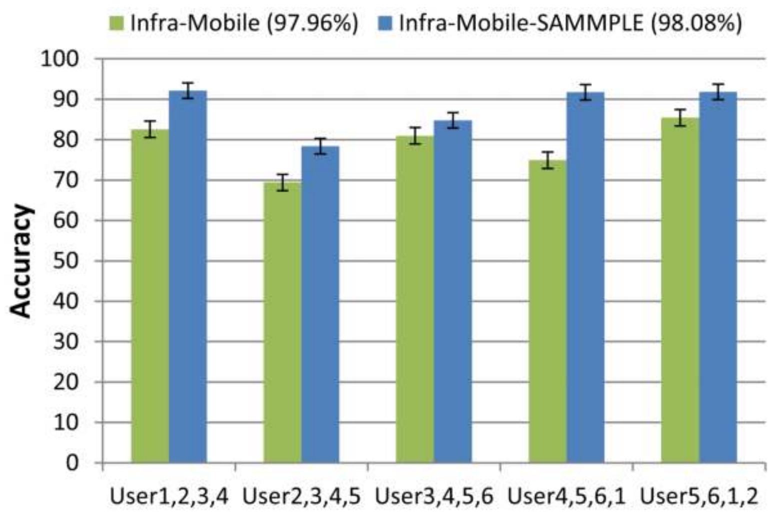

Fig. 19. Confidence interval of ambient-augmented mobile sensing for a group of four users

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

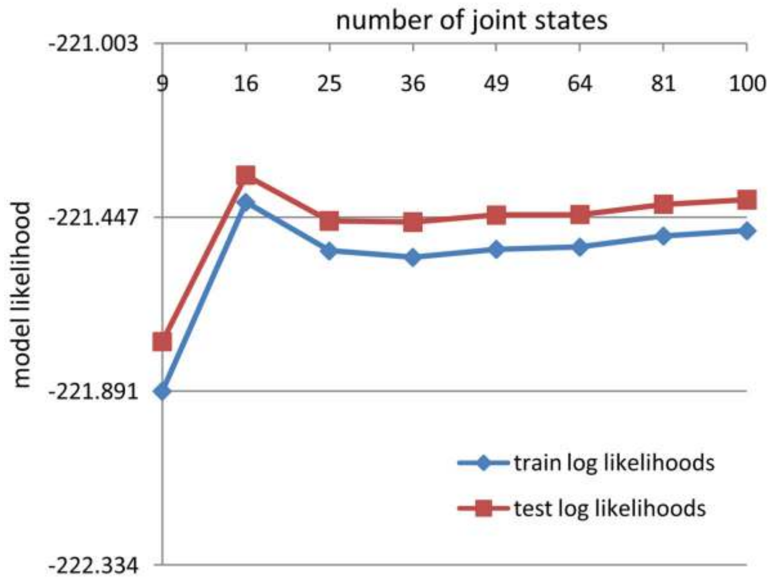

Fig. 20. Coupled activity model: training and testing log-likelihoods with # of joint states

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

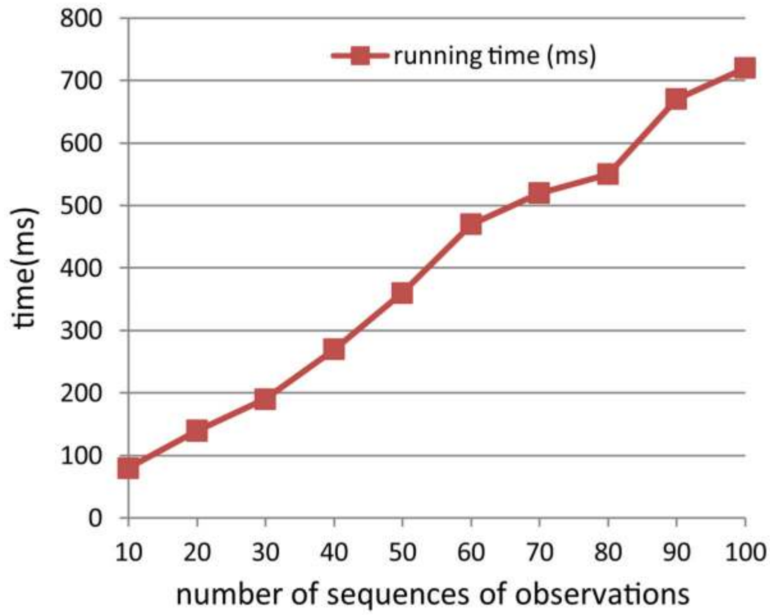

Fig. 21. Time of forward algorithm/Viterbi analysis with increasing # observation sequences

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

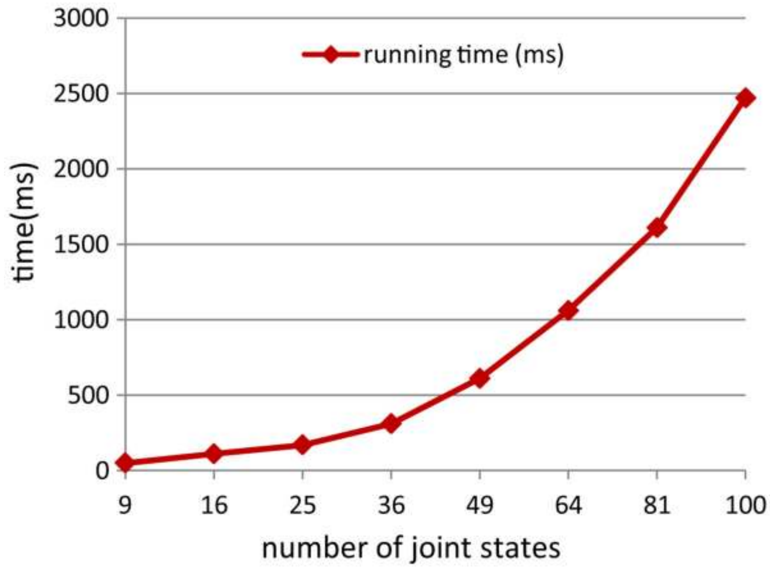

Fig. 22. Time of forward algorithm/Viterbi analysis with increasing # states

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 1

Feature extracted from the RAW data

Feature name	Definition
Mean	AVG(Σx_i); AVG(Σy_i); AVG(Σz_i)
Mean-magnitude	$\mathrm{AVG}(\sqrt{x_i^2+y_i^2+z_i^2})$
Magnitude-mean	$\sqrt{\overline{x}^2+\overline{y}^2+\overline{z}^2}$
Max, min, zero-cross	*max*, *min*, *zero-cross*
Variance	VAR(Σx_i); VAR(Σy_i); VAR(Σz_i)
Correlation	$\mathrm{corr}(x, y)=\frac{\mathrm{cov}(x, y)}{\sigma_x, \sigma_y}$

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 2

Micro activity frames

Micro activities	Number of frames
Sitting	54
Standing	109
Walking	324
Running	118
Lying	289
Climbing	165

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 3

Macro activity frames

Macro activities	Number of frames
Cleaning kitchen	904
Cooking	1485
Medication	1658
Sweeping	1366
Washing hands	758
Watering plants	843

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 4

Opportunity micro activity frames

Micro activities	Number of frames
Door	183
Fridge	57
Dishwasher	14
Drawer	60
Clean table	53
Drink	267
Toggle switch	1

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 5

Opportunity macro activity frames

Macro activities	Number of frames
Relaxing	326
Coffee time	701
Early morning	1084
Cleanup	650
Sandwich time	1239

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 6

Confusion matrix for complex activity set for both Naive-Mobile (NM) and Infra-Mobile-SAMMPLE (IMS)

Macro-activity (NM/IMS)	**a**	**b**	**c**	**d**	**e**	**f**
Cleaning kitchen = a	90/101	63/62	27/0	39/76	11/0	14/0
Cooking = b	53/61	251/315	59/0	111/151	22/0	39/0
Medication = c	26/0	65/0	383/580	60/0	24/0	30/0
Sweeping = d	27/45	114/106	69/0	359/476	35/0	31/0
Washing hands = e	29/0	31/0	37/0	48/0	49/207	14/0
Watering plants = f	11/0	56/0	34/0	54/0	10/0	85/248