

# Infrastructure-Assisted Smartphone-based ADL Recognition in Multi-Inhabitant Smart Environments

Nirmalya Roy\*, Archan Misra†, Diane Cook\*

\*School of Electrical Engineering and Computer Science, Washington State University

† School of Information Systems, Singapore Management University

nroy@eecs.wsu.edu, archanm@smu.edu.sg, cook@eecs.wsu.edu

**Abstract**—We propose a hybrid approach for recognizing complex Activities of Daily Living that lie between the two extremes of intensive use of body-worn sensors and the use of infrastructural sensors. Our approach harnesses the power of infrastructural sensors (e.g., motion sensors) to provide additional ‘hidden’ context (e.g., room-level location) of an individual and 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 spatiotemporal constraints can be used to significantly improve the accuracy and computational overhead of traditional coupled-HMM based approaches. Experimental results on a smart home dataset demonstrate that this approach improves the accuracy of complex ADL classification by over 30% compared to pure smartphone-based solutions.

**Keywords:** multi-modal sensing, context recognition

## I. INTRODUCTION

Research on activity recognition 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., [18]).
- *Infrastructure-based:* 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 [2]) are reluctant to the idea of continually wearing 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 [9] or mounting them on the ceiling has challenging operational costs and battery-life issues. Video sensors are also simply 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, coarse-grained infrastructural 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-based classification of smartphone-embedded accelerometers [7], [11]. 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 Activities of Daily Living (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 accurate 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 [2].

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?) as inferred from infrastructural sensors. 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 **inter-personal** (different individuals are more or less likely to have certain ‘hidden context’ concurrently). We then build such a CHMM through appropriate modifications to the standard Expectation Maximization (EM) algorithm, and use a modified Viterbi algorithm during the testing phase to determine the most likely *temporal evolution* of each person’s activity. We present performance results, on real-life activity traces collected from a smarthome facility, that suggest that such a hybrid infrastructure-augmented mobile sensing model can significantly improve the accuracy of both low-level and high-level context recognition.

Our investigations in this paper address several key research questions. First, given the reality of an indoor multi-inhabitant environment with cheap infrastructural 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?

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, 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.,  $\approx 5$ ). Finally, we use *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  $\approx 30\%$ .

## II. RELATED WORK

Much existing work on multi-user activity recognition relies primarily on video data. HMMs and CHMMs for modeling and classifying interactions between multiple users have been addressed [14], [18], while others [6] have developed a dynamically multi-linked HMM model to interpret group activities. Activity recognition in smart environments using unsupervised clustering of data collected by a rich set of wearable sensors has been explored [4]. 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 [7], [11].

Sensor-based activity recognition strategies can be categorized into static and temporal categories [22]. Naive Bayes [13], Decision Trees [13], K-Nearest Neighbors [8] and SVM [8] have been used extensively as static classifiers; temporal classification approaches infer the values of hidden context states using approaches such as HMMs [12], Dynamic Bayesian Networks [15], Conditional Random Fields [10] and CHMMs [18]. SAMMPLE [21] utilizes a layered approach, where the lower layer classifies low-level micro-activities and 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 infrastructural sensors.

The notion of using infrastructural sensors to infer individualized context in a multi-inhabitant smart environment was first studied by Wilson [19], which uses a particle filtering approach to infer the evolution of coupled HMMs based on events generated by multiple infrastructure-embedded sensors. Unlike earlier work [19], we exploit the pervasiveness of body-worn smartphone sensors to infer person-specific context; additionally, previous approaches focus only on inferring whether an individual is moving or stationary, our focus is on inferring complex ADLs.

## III. 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 [21].

Consider a smart environment (such as an assisted living facility) with  $N$  distinct individuals. The  $i^{th}$  individual’s **micro-context**, at a given time instant  $t$ , is captured by a  $M$ -dimensional tuple  $Context^i(t) = \langle c_1^i(t), c_2^i(t), \dots, c_M^i(t) \rangle$ , 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  $\{walking, sitting, standing, \dots\}$ ) and location can assume values such as  $\{bedroom, bathroom, kitchen, \dots\}$ . In general, assuming time to be discretely slotted, an individual  $i$ ’s activity pattern may be represented by a micro-context stream, i.e.,  $Context^i(t), Context^i(t+1), \dots$ . 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 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 [7], [11] 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 infrastructural sensors. Assume that the environment has a total of  $K$  such sensors, each of which is deployed at a well-known location. The  $k^{th}$  :  $k = 1, \dots, K$  sensor, 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 sensor ( $J = 1$ ), each of which is placed in a location such as  $\{bedroom, bathroom, kitchen, \dots\}$ .

#### A. Two-Tier Inferencing for Individual/Multiple Inhabitants

Given our formulation above, the evolution of the micro-activities of the  $i^{th}$  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 [17] 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 [21]). 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 inferable) 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.

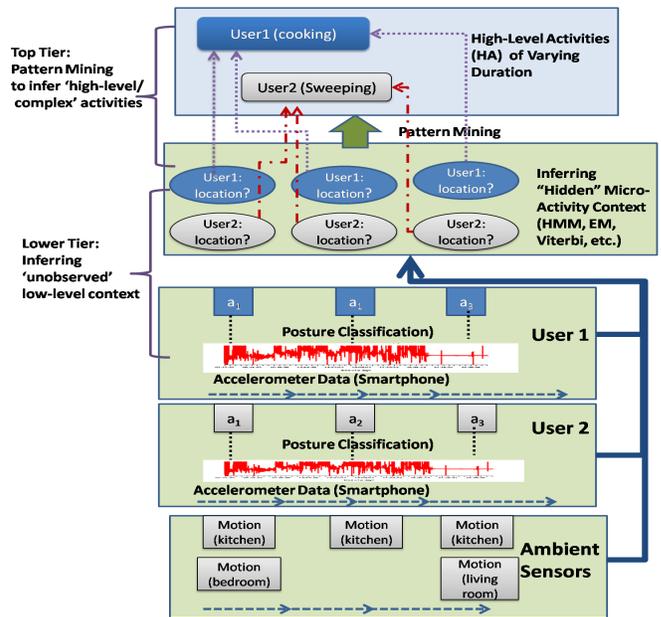


Figure 1. Illustration of our two-tier approach to combining smartphone and infrastructural sensor data.

After inferring these hidden states, we now have a complete set of  $Context^i(t)$  observations for each individual. As 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, the previous results in SAMMPLE [21] have used only the low-level ‘postural activity’ stream to identify an individual’s complex activity. Based on the inferencing performed in the lower-tier, we now have the joint (postural activity, location) stream. 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?*

## B. 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 ambient context sensed by the infrastructural 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. 2):

- intra-user temporal constraints*: For a specific user  $i$ , if  $Context^i(t-1) = (sitting, livingroom)$ ,  $Context^i(t)$  cannot equal  $(sitting, bathroom)$  (the user cannot simply change rooms while remaining in a ‘sitting’ state!)
- 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!

## C. Coupled HMM for Multiple Inhabitants

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 coupled Hidden Markov Model [3]. To define this HMM, let  $O(t)$  denote the “observable stream” (in practice, the accelerometer readings on the different 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:

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

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:

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

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.

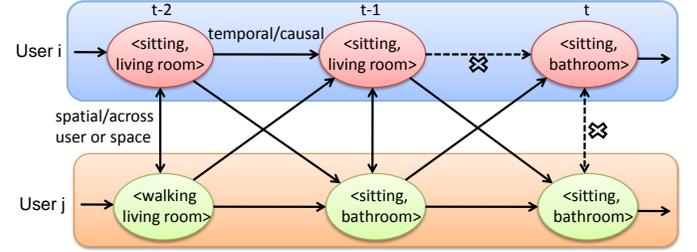


Figure 2. CHMM with inter-user and intra-user constraints.

Simplifying the  $N$  chain couplings as shown in Eqn. 2 by considering two users, the posterior of the CHMM for any user can be represented as follows.<sup>1</sup>

$$P(Context|O) = \frac{\pi_{s_1} P_{s_1}(o_1) \pi_{s'_1} P_{s'_1}(o'_1)}{P(O)} \prod_{t=2}^T P_{s_t|s_{t-1}} P_{s'_t|s'_{t-1}} P_{s'_t|s_{t-1}} P_{s_t|s'_{t-1}} P_{s_t}(o_t) P_{s'_t}(o'_t) \quad (3)$$

where  $\pi_{s_1}$  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 2, we modify the posterior of the state sequence for two users by:

$$P(Context|O) = \frac{\pi_{s_1} P_{s_1}(o_1) \pi_{s'_1} P_{s'_1}(o'_1)}{P(O)} \prod_{t=2}^T P_{s_t|s_{t-1}} P_{s'_t|s'_{t-1}} P_{s'_t|s_{t-1}} P_{s_t|s'_{t-1}} P_{s_t|s_t} P_{s'_t|s_t} P_{s_t}(o_t) P_{s'_t}(o'_t) \quad (4)$$

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.

## IV. SOLVING THE COUPLED ACTIVITY MODEL

Having defined the coupled Hidden Markov model (CHMM), we now discuss how we can solve this model to infer the ‘hidden’ context variables for multiple occupants simultaneously. Unlike prior work [3] 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

<sup>1</sup>We interchangeably use  $Context$  as a state  $s$  in our HMM model. For brevity we denote  $Context^i(t) = s_t$  and  $Context^j(t) = s'_t$  in equations.

joint observation of smartphone and infrastructural 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 [17] to determine the best parameters for the constrained CHMM, and subsequently describe a modified Viterbi algorithm to infer the probability of the temporal context of each user.

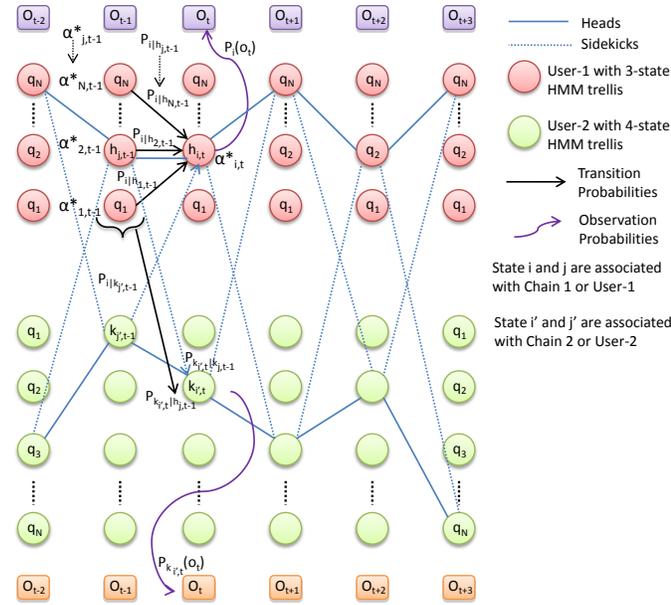


Figure 3. Search through the state trellis of a 3-state HMM for User-1 and 4-state HMM for User-2 for state probabilities, transition, coupling and spatial probabilities and most likely path

### A. 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., [16]) where such pruning is performed before the Viterbi inferencing process, we perform our pruning during the process of learning the model as well.

To illustrate our approach, consider the state-trellis for two users, User-1 and User-2, illustrated in Fig. 3. 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 4 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 3 possible values:  $\{(sitting, livingroom), (sitting, bathroom), (sitting, kitchen)\}$  in the trellis for User-1, only the  $(sitting, livingroom)$  state is possible at time  $t - 2$ . Likewise, of the 4 possible values:  $\{(standing, livingroom), (standing, bathroom), (standing, kitchen), (walking, corridor)\}$  for User-2, only the  $(standing, livingroom)$  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  $(2*2 =) 4$  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  $(sitting, livingroom)$  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  $M * N$  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  $(M * N)^K$  choices can be significantly curtailed by both (a) starting with those time instants where the context is deterministic (in our example, the  $t - 2$  choices are unambiguous as shown in Fig. 3) and (b) keeping the dimension  $T$  small (for our experimental studies,  $T = 2$  provided good-enough results).

### B. Model Likelihood Estimation

To intuitively understand the algorithm, consider the case where we have a sequence of  $T$  observations ( $T$  consecutive time instants), with  $\overline{M}$  underlying states (reduced from the  $M * N$  original states by the pruning process) at each step.

As shown in Fig. 3, 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  $\lambda = (\text{transition, emission probabilities})$ , 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. 3. 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 19 in Fig 4).

In each step of the forward analysis we calculate the *maximum a posterior* (MAP) for  $\{Context^i(t), Context^{j'}(t-1) = \text{head, sidekick}\}$  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 11 in Fig 4). 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 26 in Fig 4). This forward analysis algorithm pseudocode is articulated in Fig. 4 and explained with a pictorial diagram in Fig. 3 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  $\alpha^*[j, t-1]$ .

### C. 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: i) Select MAP sidekicks in  $t$  for each antecedent path in  $t-1$  and

```

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  $\alpha$  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

```

Figure 4. Forward Algorithm pseudocode for Coupled Activity Model

ii) Select the antecedent path and associated sidekick that maximizes the new head's posterior for each head in  $t$ . Fig. 5 presents the pseudocode for our modified Viterbi algorithm, developed for multi-inhabitant environments.

## V. 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 infrastructural testing. Our experiments are conducted using 10 participants at the WSU CASAS smart home.

### A. Data Collection

To validate our approach, we collected data from 10 subjects, each of whom carried an Android 2.1 OS based Samsung Captivate smart phone (containing a tri-axial accelerometer and a gyroscope) [5]. Each subject carried the phone while performing different activities of daily living. 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

```

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.  $viterbi[i, t] \leftarrow \max_{j=1}^N viterbi[j, t-1] \times P_{i|h_j} \times P_{k_{i',t}|h_j} \times P_i(o_t)$ ;
//backpointer keeps track of whichever state was the most
probable path to the current state
12.  $backpointer[i, t] \leftarrow \arg \max_{j=1}^N viterbi[j, t-1] \times P_{i|h_j} \times P_{k_{i',t}|h_j}$ ;
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
 $backpointer[q_f, T]$ .

```

Figure 5. Viterbi Algorithm psuedocode for Multiple Users

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.

### B. 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 6 labels:  $\{sitting, standing, walking, running, lying, climbing stairs\}$ .
- 2) **High-level** (or *complex*): These consisted of semantically meaningful ADLs, and included 6 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.
  - *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.

### C. 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  $\{1,2,4,8,12\}$  secs and report results here for the representative case of 4 seconds), and a 30-dimensional feature vector (see Table I) was computed over each frame. The ground-truth annotated training set (aggregated across all 10 users) was then fed into the Weka toolkit [20] and used to train 6 classifiers: Multi-layer Perceptron, Naive Bayes, Bayesian network, Decision Table, Best-First Tree, and K-star. The accuracy of the classifiers was tested using 10-fold cross-validation.

Table I  
FEATURE EXTRACTED FROM THE RAW DATA

Feature Name	Definition
Mean	$AVG(\sum x_i); AVG(\sum y_i); AVG(\sum z_i)$
Mean-Magnitude	$AVG(\sqrt{x_i^2 + y_i^2 + z_i^2})$
Magnitude-Mean	$\sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2}$
Max, Min, Zero-Cross	$max, min, zero-cross$
Variance	$VAR(\sum x_i); VAR(\sum y_i); VAR(\sum z_i)$
Correlation	$corr(x, y) = \frac{cov(x, y)}{\sigma_x \cdot \sigma_y}$

Fig. 6 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 Section V-E).

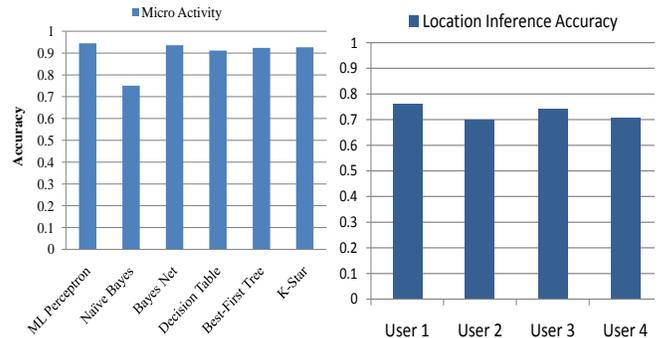


Figure 6. Micro-Activity Classification Accuracy based on Mobile Sensing

Figure 7. Location Inferencing Accuracy using Ambient Sensor Data Sensing

### D. 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 I but micro activity features as explained as Naive-SAMMPLE in subsection V-E) and the infrastructure (motion sensor) specific location feature into our activity recognition (ar version 1.2) code [1] 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. Fig. 7 reports on the accuracy of the location estimate of 4 individuals randomly chosen. 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  $\approx 5 - 6\%$  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.

### E. 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 4 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 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 [21]. 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 6-dimensional feature-vector consisting of the number of frames (effectively the total duration) of each of the 6 micro-activities considered in our study. For example, if an instance of ‘cooking’ consisted of 3 frames of ‘sitting’, 4 frames of ‘standing’ and 7 frames of ‘walking’, the corresponding feature vector would be [3 4 7 0 0 0], as the last 3 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 and gyroscope 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 7-dimensional feature vector, with the first 6 elements corresponding to the frequency of the underlying micro-activities and the 7<sup>th</sup> element corresponding to the indoor location inferred by our Viterbi algorithm.

Fig. 8 plots the accuracy of the different approaches (using 10-fold cross validation) for a randomly selected set of 5 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).

Table II provides the Best-First Tree confusion matrix for the 6 pre-defined complex activities, for both the Naive-Mobile approach and our suggested Infra-Mobile-SAMMPLE 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.

### F. Computation Complexity of the 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 resulting model predictions

Table II  
CONFUSION MATRIX FOR COMPLEX ACTIVITY SET FOR BOTH NM & 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

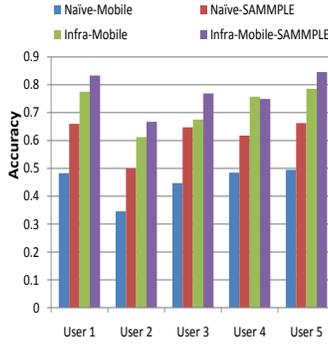


Figure 8. Complex Activity Classification: Mobile vs. Ambient-Augmented Mobile Sensing for Multiple Users

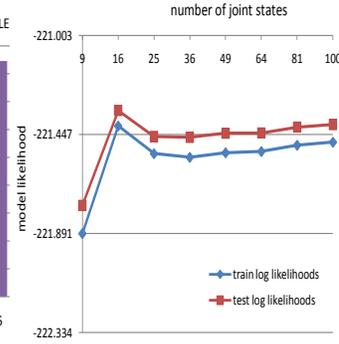


Figure 9. Coupled Activity Model: Training and Testing log-likelihoods with # of joint states

(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).

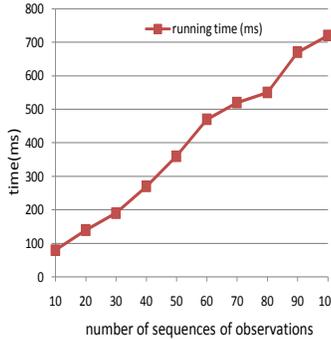


Figure 10. Time of forward algorithm/viterbi analysis with increasing # observation sequences

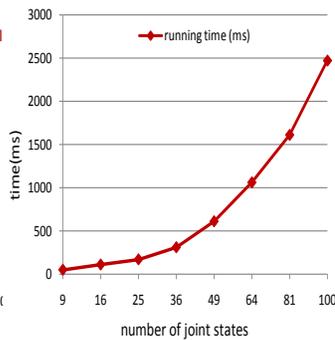


Figure 11. Time of forward algorithm/viterbi analysis with increasing # states

Fig. 9 depicts the training and testing log-likelihoods of our coupled model which establishes that train-test divergence is very minimal. Fig. 10 shows the computation time of our algorithms with a fixed number of states and increasing number of data sequences, whereas Fig. 11 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 5-fold savings in computation

time ( $2500ms \rightarrow 500ms$ ).

## VI. 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 smartphone-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  $\sim 70\%$  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.

## ACKNOWLEDGEMENT

This work is partially supported by NSF grants 1064628, 0852172, CNS-1255965, and NIH grant R01EB009675 and by the Singapore Ministry of Education Academic Research Fund Tier 2 under research grant MOE2011-T2-1-001.

## REFERENCES

- [1] AR Activity Recognition Code <http://ailab.wsu.edu/casas/ar/>
- [2] JH Bergmann and AH McGregor, “Body-worn sensor design: What do patients and clinicians want?”, *Annals of Biomedical Engineering*, 39(9):2299-2312, 2011.
- [3] M. Brand., “Coupled hidden markov models for modeling interacting processes”, *Technical Report 405, MIT Lab for Perceptual Computing, 1996*
- [4] B. Clarkson, K. Mase, and A. Pentland, “Recognizing user context via wearable sensors”, *Proceedings of the 4th International Symposium on Wearable Computers (ISWC 2000)*, 2000, pp. 69-76.
- [5] S. Dembach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, “Simple and Complex Activity Recognition Through Smart Phones”, *Proceedings of the International Conference on Intelligent Environments, 2012*.
- [6] Gong, S., Xiang, T., “Recognition of group activities using dynamic probabilistic networks”, *In Proceedings of International Conference on Computer Vision (ICCV 2003)*
- [7] N. Gyobiro, A. Fabian, and G. Homanyi, “An activity recognition system for mobile phones”, *Mobile Networks and Applications*, vol. 14, pp. 82-91, 2008.
- [8] Huynh, T., Blanke, U., Schiele, B., “Scalable recognition of daily activities from wearable sensors”, *LoCA 2007. LNCS*, vol. 4718, pp. 50-67
- [9] S. S. Intille, et al., “Using a live-in laboratory for ubiquitous computing research”, *PERVASIVE 2006, LNCS*

- [10] van Kasteren, T., Noulas, A., Englebienne, G., Krose, B., "Accurate activity recognition in a home setting", *In Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp 2008*, pp.1-9
- [11] J. Kwapisz, G. Weiss, and S. Moore, "Activity recognition using cell phone accelerometers", *International Workshop on Knowledge Discovery from Sensor Data, 2010*, pp. 10-18.
- [12] Lester, J., Choudhury, T., Borriello, G., "A practical approach to recognizing physical activities", *PERVASIVE 2006. LNCS, vol. 3968*, pp. 1-16.
- [13] Logan, B., Healey, J., Philipose, M., Munguia-Tapia, E., Intille, S., "A long-term evaluation of sensing modalities for activity recognition" *UbiComp 2007. LNCS, vol. 4717*, pp. 483-500.
- [14] Oliver, N., Rosario, B., Pentland, A., "A Bayesian computer vision system for modeling human interactions", *IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8)*, pp. 831-843 (2000)
- [15] M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Hahnel, D. Fox, and H. Kautz, "Inferring activities from interactions with objects," *IEEE Pervasive Computing, vol. 3, no. 4*, pp. 50-57, 2004.
- [16] T. Plotz and G. A. Flink, "Accelerating the evaluation of profile HMMs by pruning techniques", *In Tech rep. University of Bielefeld, Faculty of Technology; 2004. [Report 2004-03]*
- [17] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE, vol. 77, no. 2*, pp. 257-285. 1989.
- [18] L. Wang, T. Gu, X. Tao, H. Chen, Jian Lu, "Recognizing multi-user activities using wearable sensors in a smart home", *Pervasive and Mobile Computing, 7(3): pp. 287-298 (2011)*
- [19] Wilson, D., Atkeson, C., "Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors", *Pervasive Computing, pp. 62-79 (2005)*
- [20] I. Witten, and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations," *Morgan Kaufmann, 1999*.
- [21] Z. Yan, D. Chakraborty, A. Misra, H. Jeung, K. Aberer, "SAMMPLE: Detecting Semantic Indoor Activities in Practical Settings using Locomotive Signatures", *International Symposium on Wearable Computers (ISWC), 2012*
- [22] L. Chen, J. Hoey, C.D. Nugent, D.J. Cook and Z. Yu, "Sensor-Based Activity Recognition", *IEEE Transactions on Systems, Man, and Cybernetics-Part C*, vol.42, no.6, pp.790-808, 2012.