Task Phase Recognition for Highly Mobile Workers in Large Building Complexes

Allan Stisen¹, Andreas Mathisen¹, Søren Krogh Sørensen¹, Henrik Blunck², Mikkel Baun Kjærgaard¹, Thor Siiger Prentow¹
¹Department of Computer Science, Aarhus University, Denmark
²Department of Computer Science, Münster University, Germany

e-mail: {allans,mathisen,krogh89,mikkelbk,prentow}@cs.au.dk, henrik.blunck@wwu.de

Abstract—Being aware of activities of co-workers is a basic and vital mechanism for efficient work in highly distributed work settings. Thus, automatic recognition of the task phases the mobile workers are currently (or have been) in many applications, e.g., efficient coordination of tasks by visualizing co-workers’ task progress, automatic notifications based on context awareness, and record filing of task statuses and completions.

This paper presents methods to sense and detect highly mobile workers’ tasks phases in large building complexes. Large building complexes restrict the technologies available for sensing and recognizing the activities and task phases the workers currently perform as such technologies have to be easily deployable and maintainable at a large scale. The methods presented in this paper consist of features that utilize data from sensing systems which are common in large-scale indoor work environments, namely from a WiFi infrastructure providing coarse grained indoor positioning, from inertial sensors in the workers’ mobile phones, and from a task management system yielding information about the scheduled tasks’ start and end locations. The methods presented have low requirements on the accuracy of the indoor positioning, and thus come with low deployment and maintenance effort in real-world settings.

We evaluated the proposed methods in a large hospital complex, where the highly mobile workers were recruited among the non-clinical workforce. The evaluation is based on manually labelled real-world data collected over 4 days of regular work life of the mobile workforce. The collected data yields 83 tasks in total involving 8 different orderlies from a major university hospital with a building area of 160,000 m². The results show that the proposed methods can distinguish accurately between the four most common task phases present in the orderlies’ work routines, achieving F1-Scores of 89.2%.

I. INTRODUCTION

The ability to recognize human activities is an important challenge within pervasive computing and has a multitude of applications, especially within work domains, e.g. car manufacturing [1], and assembly and maintenance tasks [2] have been in focus, extending the scope and potential applications of Human Activity Recognition (HAR) research.

In addition, highly collaborative work environments, such as hospitals, have shown to be of particular interest for applied activity recognition. In hospitals, being aware of co-workers’ activities is an important aspect of the patient care [3]. However, it is challenging for hospital workers to establish and maintain such an awareness when they are distributed across the hospital and are not co-located [3]. Within the hospital domain, the HAR research has had a particular focus on the clinical workers’ activities, including wards [4], emergency departments [5], and surgical operating rooms [6].

Previous work relies on prerequisites such as highly accurate location systems [6], [2], wearable sensors attached to equipment or workers [1], or video data [7]. To provide HAR for highly mobile workers in large building complexes, though, these prerequisites are hard to provide, maintain, calibrate, and deploy, which is recognized as an issue for real deployments [8], including hospitals [9]. In this paper we thus propose novel methods which allow for accurate task phase recognition of mobile workers in large building complexes and with comparatively low and domain-adequate demands on sensing and tracking technology. Concretely, these methods and their evaluation utilize the following: i) a nearly calibration-free indoor WiFi positioning system; ii) a task’s start and end points obtained from a task management system; iii) co-movement data for associated entities, such as the worker and equipment involved, estimated based on raw WiFi-signals; and iv) inertial sensors available in common smartphones. Thus, the methods rely on easily deployable and maintainable sensing systems requiring little or no calibration effort.

We evaluate the proposed methods on real-world sensor data, including raw WiFi signals and derived location estimates of the workers, the patients and the equipment, and accelerometer data from workers’ smartphones. The data was collected through 4 days (83 tasks) of observations of 8 hospital orderlies during their daily tasks, consisting mostly of service tasks and assistance to clinical workers, e.g., transporting patients and equipment across the hospital’s buildings. Our evaluation thus also provides one of the first reports on the applicability of HAR methods on a yet underrepresented work domain, namely, highly mobile non-clinical hospital work. Applying our task phase recognition methods for orderlies allows for a wide range of applications, e.g., to support the coordination of hospital orderlies, by enhancing the awareness of their colleagues, but also context aware notifications. The provided aid in coordinating tasks is crucial as any delays, errors or breakdowns of coordination in the non-clinical work affect the overall hospital efficiency, as they result in, e.g., delayed or canceled operations. Moreover, we deem the proposed methods applicable also to other types of highly mobile work in large building complexes, e.g., at airports, smart manufacturing plants [10], and large warehouses.
II. HIGHLY MOBILE AND DISTRIBUTED WORKERS

The task phases considered in this paper are general in nature to most tasks involving transporting resources, and little is assumed about the structure of the task’s phases. Throughout this paper we assume a task has a known start and end point, which also define a task’s 4 primary phases, i.e., (i) moving towards the task’s start location, (ii) loading the resource at the start point, (iii) transporting the resource to the end point, (iv) offloading (and potentially utilizing) the resource at the end point. We illustrate this particular view on task phases by presenting a concrete use case, hospital orderlies, based on a detailed understanding of the orderlies’ tasks and phases, which was acquired through empirical investigations. These involved two major university hospitals and included observations (130 hours), interviews with orderlies (n=72) and workshops conducted with orderlies. Furthermore, the orderlies’ managers (n=2) and clinical workers (n=5) have participated in interviews. For more details regarding the empirical investigations, see [11], [12]. Moreover, we have acquired logs of all tasks obtained from a deployed task management system, PLog, from two hospitals for a period of 248 and 300 days. Table I presents the distributions of task types and their total durations based on the logs from the latter hospital, revealing that patient transport tasks constitute 48.5% of the total work time and 46.2% of all tasks.

<table>
<thead>
<tr>
<th>Task type</th>
<th>% of all tasks</th>
<th>Average time</th>
<th>Total time (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient transports</td>
<td>46.2%</td>
<td>09:57 min.</td>
<td>48.5%</td>
</tr>
<tr>
<td>Patient mobilizations</td>
<td>13.7%</td>
<td>10:14 min.</td>
<td>13.6%</td>
</tr>
<tr>
<td>Other transports</td>
<td>21.1%</td>
<td>09:26 min.</td>
<td>19.4%</td>
</tr>
<tr>
<td>Other tasks</td>
<td>16.3%</td>
<td>11:02 min.</td>
<td>16.9%</td>
</tr>
<tr>
<td>Bed transport orders</td>
<td>2.7%</td>
<td>14:06 min.</td>
<td>1.5%</td>
</tr>
<tr>
<td>Overall</td>
<td>16595 tasks</td>
<td>10:36 min.</td>
<td>2%</td>
</tr>
</tbody>
</table>

TABLE I: Overall distribution of task types from a task management system’s log running for 300 days.

Every time a task is finished, the orderly will consult PLog’s mobile interface again, probably acquire a new task, and manually stop the current task (since the current work flow does not yet incorporate on automatic task phase recognition). If there are no available tasks for the orderly, he will likely return to the coffee room and wait for the next task, and PLog will consider the task to be *idle*. However, being idle is not as such a phase but rather a state of the orderly, and the orderlies have to anyway check for new tasks and manually select the next task. This means the idle state is defined and can be detected implicitly as starting when an orderly ends his task manually and ending when he selects a new task. Information about which of the four primary phases of a task an orderly is currently in is then used e.g. to inform colleagues in a visual manner in the (already deployed) colleague overview, which contains also the colleague’s location, his task deadline, his task start and end points. A mockup of such an overview interface is given in Figure 1b. While shown for a smartphone, the mock up could also be adapted for wearable devices, as investigated for the given domain [12].

III. OVERVIEW OF PROPOSED METHODS

In this section, first the sensing modalities which are considered both useful and applicable for task phase recognition in large building complexes such as hospitals are presented. Secondly, the features we extract from the respective sensor data are presented. Throughout, the described feature types rely on timestamped sensor measurements that are aggregated into time windows of specific durations, e.g., 10 seconds.

A. Sensor data and preprocessing

Building on investigations into supporting hospital service workers via ICT [11], [13], [14], we identified time-varying characteristics that are essential to track in order to recognize the task progress and phases in the chosen use-scenario.

- The location of the orderly and the patient
- The location of the equipment the orderly uses
- The above entities’ locations relative to the departments involved in the task
- The co-movement of the orderly, patient and equipment

Throughout this paper we assume that it is the mobile worker (here: the orderly), who is responsible for choosing (booking) a task, and afterwards for completing it. Our prior investigations into the work practices of orderlies suggest it favorable that these decisions are made by the orderly, in order to induce responsibility for which task to pick and commit to [11].
In order to track these characteristics, we made use of the existing WiFi infrastructure, as it is utilized by an already deployed indoor positioning system. Moreover, we exploit that the orderlies to be tracked are instructed to keep a work mobile phone in their jacket pocket. Additionally, the patient beds and the vehicles used by the orderlies were attached with WiFi-tags—as per the long term plan of the hospitals’ management—during the collection period. The mobile phone’s WiFi and the WiFi tags’ sampling frequency was set to 0.5 Hz and 0.1 Hz, respectively. The WiFi tags’ sampling frequency was chosen based on the WiFi-tag’s manufacturer’s recommendation on the life expectancy, estimating a year of battery life without changing at 0.1 Hz. For the smartphone the 0.5Hz sampling frequency is set to allow for a battery life of a work shift, i.e., 8 hours.

The algorithmic choices for deriving indoor positions are based on the work by Kjærgaard et al. [8], who also report that in large buildings, e.g., hospitals, it is vital to have a low level of maintenance and a high level of coverage. Therefore, the indoor positioning system is using a simple model-based algorithm using weighted centroid lateration of the locations of the access points in range. Hence, the only prerequisite of the positioning method is knowledge of the locations (as well as their emitting power) of the access points, and no calibration is needed—which is in contrast to, e.g., fingerprinting the whole building complex or to methods based on detailed building maps. The method has successfully been utilized in prior studies utilizing indoor positioning systems [13], [15], [14].

The following sections describe the proposed task phase recognition features, based on the respective utilized technologies and sensor modalities just presented.

### B. WiFi based spatial features

Central characteristics to a task itself is its start and end location, in our case the dispatching and receiving departments, respectively, since the evolving proximity or distance to the two departments is a good indicator for the current task status. In the rest of this paper we work with two granularities for start and end locations; department- and room-based granularity. The department-based granularity consists either of the bounding box of the dispatching or receiving department or (depending on the feature type) just the bounding box’s central location, whereas the room-based granularity associates a location instead with the central location of the start, resp. end, location room of the task. From the position estimates for orderlies, patients and equipment (of the two departments involved in a task), we derive spatial features describing their relative positions. Naturally, the usefulness of these features depends on the accuracy of the obtained position estimates. We specify two categories of spatial features: features based on Euclidean distance and boolean features capturing proximity to a department. To this end, the following is calculated for all position estimates in a give time window:

**Distances:** We consider two distances, the direct distance and the shortest path distance. The shortest path is computed using a simple building floor map. For a position \( p \) we compute both actual distances and normalized distances relative to the distance between the departments:

\[
\frac{\text{dist}(d_{\text{dept}}, p)}{\text{dist}(d_{\text{receiving}}, d_{\text{dispatching}})} \cdot 100.0
\]

where \( \text{dept} \) can be either the receiving or dispatching department. To illustrate the time-varying values of these features during transport tasks, Figure 2 illustrate the time-varying values of these features during transport tasks, here with respect to both the dispatching (blue) and receiving (green) department for for both the orderly (a) and the patient (b).

**Proximity:** We estimate the probability of an actual position being inside either the dispatching or the receiving department. The probability is computed based on the intersection area between the bounding box of a department \( b_{\text{dept}} \) and a circle coverage area \( c_p \), around the position estimate \( p \) with the expected error \( \epsilon \) as radius:

\[
\frac{\text{area}(c_p \cap b_{\text{dept}})}{\text{area}(c_p)} \cdot 100.0
\]

Based on our empirically gained knowledge about the deployed positioning system, an expected error of 15 m is used.

---

**Fig. 1:** Visualizing the four primary task phases (a) and a mockup of how to utilize the task phase recognition (b)
Besides the department proximity, we also calculate proximity to start and end destination, if room-based granularity of the latter is available. Here, instead of checking against a bounding box of the department, we define a circle around the destination $c_{destination}$ with radius $\alpha$ and compute the probability based on the intersecting area between the two circles $c_{destination}$ and $c_p$. We found that using $\alpha = 10 \text{ m}$ yielded the best results.

Since each window contains several position estimates, common statistical methods like average, min, max, median etc. are used to describe the computed measurements for the orderly, the patient and the equipment's distances, respectively.

C. WiFi Signal strength based features

The WiFi infrastructure can not only be used to estimate absolute positions, but also proximity of entities to each other as well as motion data, such as speed, by analyzing the received signal strength indicator (RSSI) values, as received for each access point in range, over time. For detecting motion of entities one can exploit, e.g., that the variability of incoming radio signal sources and their RSSI will be higher when the transmitting device is moving as opposed to when it is stationary. To obtain features exploiting such phenomena, the following data processing scheme is used. From the RSSI provided with each access points received by the respective target, i.e. either a WiFi-scanning smartphone of an orderly, or a WiFi-tag attached to a patient’s bed or to equipment, we compute the Euclidean distance in signal space between the samples, by treating each access point as a dimension in RSSI signal strength space and then apply Dynamic Time Warping (DTW) and Derivative DTW to quantify the differences between two entities within a window [17]. We employ these features to check for co-movement primarily of the following combinations of entities: orderly-patient, and orderly-equipment.

D. Inertial sensor features

The previous features which are all based on the WiFi infrastructure are prone to noise: e.g., when a user is standing still the received signal profiles may still vary substantially and thus the user behavior may be mistaken for erratic movement (even after temporal data smoothing). Thus, we complement the features based on noisy WiFi signals, with features using the accelerometers mostly commonly found in smartphones. Based on this sensor we use three common feature types, for the first two see Figo et. al [18], for details on the concrete data processing and feature computations used here see also [19]:

Time domain features which includes the maximum, minimum, standard deviation, average and other common statistical functions over a window.

Frequency domain features which based on Fourier Transformation report the histogram of the frequencies present in the accelerometer data. Lastly, we also consider a more novel feature type, or rather a feature learning scheme, namely:

Empirical Cumulative Distribution Function (ECDF) features [20]: these features correspond to the window’s distributions quantile function. The computation only requires a single parameter $d$ to adjust the granularity of the function’s shape, by adjusting the number of points of which the function is interpolated in the computation.

E. Historical features

Since the transport tasks often yield the same sequence of phases, we compute historical features, inspired by [6], to incorporate the sequential aspect in the classification. Some windows may not allow correct recognition of the task phase when only considering features describing the window itself,
hence knowledge about previous windows becomes essential, e.g., for distinguishing to task phases from transport phases.

The most basic historical feature is task duration, i.e., for how long a current task has already been started, i.e. accepted, which can be computed regardless of sensor availability. The second historical feature we consider is a summation of proximity probability. We compute the probability of being in the dispatching and receiving department historically across all time windows belonging to the current task. The intuition behind this feature is to describe how phases relate sequentially, e.g., during a transport phase the proximity sum for the dispatching department should be growing until leaving the department and afterwards not increase further. Figure 3 depicts an example of how an orderly’s historical proximity probability develops over time. Similar trends exist for the historical proximity of the patient and the equipment. In an ideal setting, the temporal evolution of the historical proximity probabilities should distinguish the phases; however, due to the sums being unbound and based on noisy data, phase changes may often be ambiguous, and thus a combination with other feature types is necessary.

IV. EVALUATION

We evaluated the proposed task phase recognition methods based on data collected from 4 work days, yielding in total 30 hours of sensor data, observing the normal work of in total 8 hospital orderlies. A focus in the analysis was to check how and to which extent combining feature data from different sources improves the recognition performance. The phase labelling was done manually by an observer in collaboration with the orderly as he was doing his tasks. The collected dataset’s distribution of task time (on average 10 minutes and 21 seconds including the idle mode and 9 minutes and 21 seconds when excluding it), see Table II, closely resembles the average time task logged by the task managements system (on average 10 minutes and 6 seconds), see Table I.

To systematically evaluate the performances of the features for recognizing the task phases, we consider four popular classifiers used in HAR research, namely nearest neighbor classifier, C4.5 decision tree, support vector machines (SVM), and random forest classifier. As primary evaluation metric the $F_1$-score is used:

$$ F_1\text{-score} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} $$

Moreover, we report the weighted average of the individual $F_1$-scores of all classes, in line with previous HAR research [21], which is more resilient to class imbalances in the dataset, i.e.:

$$ \text{Avg. } F_1\text{-score} = \frac{\sum_{i=1}^{c} w_i \cdot F_1^i\text{-score}}{\sum_{i=1}^{c} w_i} $$

where, $F_1^i\text{-score}$ is the $F_1$-score of the $i$th class and $w_i$ is the number of samples of class $i$ in the test dataset.

### TABLE II: Statistics collected on the phase durations from the dataset used for evaluating the phase recognition methods

<table>
<thead>
<tr>
<th>Overall task time</th>
<th>To task</th>
<th>Preparing patient</th>
<th>Transporting patient</th>
<th>Handing patient over</th>
<th>Idle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (mean)</td>
<td>10:21 min.</td>
<td>01:56 min.</td>
<td>02:31 min.</td>
<td>02:51 min.</td>
<td>02:03 min.</td>
</tr>
<tr>
<td>Percentage</td>
<td>100 %</td>
<td>18.67 %</td>
<td>24.31 %</td>
<td>27.53%</td>
<td>19.81 %</td>
</tr>
</tbody>
</table>

Evaluations have been conducted using 10-fold cross-validation. Note that we do not use random folds, as this approach often yields overconfident results when used with time-series data. When measurements are collected continuously and thereby close in time, the probability of neighboring measurements being similar is high, hence if neighboring measurements are placed in different folds, the probability of overfitting will increase. Furthermore we fold the data by placing complete tasks in each fold to achieve a higher robustness.

To achieve roughly even folds we always add a task to the current smallest fold until all tasks have been placed. Besides approximating even sized folds, this method also distributes tasks evenly with respect to task length. Using randomly assigned fold yields an $F_1$-score of 92.10% compared to 78.79% using task folding, when using random forests. Placing complete tasks in each fold and models realistically real world use cases, where training data will be of this form and where short and long tasks will be evenly distributed.

V. RESULTS

Given the logs of a task management system such as PLog, but no sensors or existing infrastructure, the only available data is the distance between the dispatching and receiving departments and a task’s time lapsed. Using the random forest classifier with only these features yields an $F_1$-score of 45.43% in recognizing a patient transport task’s four phases, hence motivating the need for additional features. In the following, we will first present which features and feature combinations prove superior for task phase recognition in the hospital use case. Secondly we will present how various factors, such as window size, task distance and destination granularity, influence the phase recognition performance.

A. Comparing classifiers

We compare the performance of the classifier types listed above for the use case at hand. For each classifier type, we varied both input features and parameters in order to justify the choices made in this project. Based on the five best performing feature combinations, random forest outperforms C4.5, SVM and K-NN by an average of 7.81, 3.95 and 7.98 percentage points (pp), respectively. These results are likely due to the low-precision positioning system and to random forests being more robust to noisy data (resulting from its common verdict approach using several separately trained classifiers). For the rest of this paper we focus on the results achieved by the random forest classifier.

B. Assessing and selecting feature types

Essential for designing a task phase recognition system is determining which sensors and features provide the most accurate classifications. First we evaluate the usefulness of each individual feature type separately, and subsequently we evaluate combinations of features.
Single feature types: Figure 4 presents the resulting confusion matrices of the four feature types providing the best results, respectively: Direct line distance features provide the best single feature result. However, classification of segments close to transition, e.g., in areas where two zones overlap, are difficult to classify based on distance alone, see Figure 4b; to task and transport segments will at some point have similar distances to the dispatching department. Similar problems exist with department proximity features which suffer from noise. Notably, the historical department features provide an increased certainty in separating to task from other phases, since the cumulative effect separates to task from transport more clearly, see Figure 4b.

Most feature types do not provide very accurate classifications when utilized alone: Time based features, signal strength features, co-movement features and inertial sensor features all yield an $F_1$-score below 41% when used separately. This is foremost an effect of the noise in the respective sensor data sets—which propagates to the calculated feature vectors. Secondly, the low $F_1$-scores are also due to the use scenario and environment: As often in indoor travel scenarios, there may not be much difference in e.g. accelerations or movement patterns between different transport activities or modes [13], [14] or phase types, thus resulting in similar feature values across phase types. Interestingly, distance features based on shortest path calculations also suffer from inaccurate position estimates: When attempting to replace the direct line distances with the hopefully more accurate shortest path distances, the $F_1$-score actually drops in all feature combinations. Unfortunately, outliers in the position estimates often enhance the errors in the shortest path distances as compared to direct line distances; this is mostly due to erroneous positioning into rooms adjacent to the true position, which then leads to significantly different shortest paths to a given destination. Smoothing of traces, e.g. with particle filtering, might increase the effectiveness of the shortest path distance feature type.

Combinations of feature types: Table III shows the resulting $F_1$-scores for selected feature type combinations. Notably, while many feature types yield low accuracies when used alone they still increase the accuracy when combined with distance features. Moreover, combining too many feature types can impair the recognition performances. Without historical features, combining direct line distance, time domain and co-movement features gives the highest $F_1$-score of 82.30%. Note, that WiFi signal strength features almost provide the same gain in $F_1$-score as co-movement features. When using historical features, combining direct line distance, historical department proximity and co-movement features yields the highest $F_1$-score of 84.43%, however even with the best performing combination significant uncertainties exist, especially among temporally adjacent phases, see Figure 4c. When utilizing historical features, all of time domain, co-movement and frequency domain feature types provide an increase in $F_1$-score, however combining several of these with historical proximity will then lower the $F_1$-score.

C. Impact of varying window sizes

Given the sampling frequency of 0.1 Hz used for the WiFi tags, we evaluated different window sizes before fixing a size of 10 seconds with a 50% overlap. No significant differences in performance result from increasing the window size, see Table IV: A gain of 2.62 pp for increasing the window size from 10 to 25 seconds may not outweigh the drawback of the resulting increased lag in recognition. The issue of a less responsive recognition can be addressed though by increasing the percentage of overlap between windows along with the window size. Furthermore, increasing the window size may result in not detecting short phases, or phases with little available data. This phenomenon is also likely a main explanation for the variances in the evaluation results.

<table>
<thead>
<tr>
<th>Features</th>
<th>Historical department proximity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct line distance</td>
<td>✓</td>
</tr>
<tr>
<td>Shortest path distance</td>
<td>✓</td>
</tr>
<tr>
<td>Co-movement</td>
<td>✓</td>
</tr>
<tr>
<td>Time domain</td>
<td>✓</td>
</tr>
<tr>
<td>Frequency domain</td>
<td>✓</td>
</tr>
<tr>
<td>ECDP</td>
<td>✓</td>
</tr>
<tr>
<td>WiFi signal strength</td>
<td>✓</td>
</tr>
<tr>
<td>$F_1$-score</td>
<td>73.52%</td>
</tr>
</tbody>
</table>

TABLE III: $F_1$-scores of selected feature type combinations with (right) and without (left) historical proximity features.
TABLE IV: $F_1$-scores of different window sizes using distance, historical proximity and co-movement features

D. Long vs short tasks

To investigate whether the spatial features are affected by the task’s length or duration, the tasks were clustered into two groups, i.e., long and short tasks. The clustering was done by providing the task duration and direct line distance between the task’s start and end points to the k-means clustering algorithm with $k = 2$. The clusters consisted of 75% short tasks and 25% long tasks. Moreover, as long tasks contain more windows compared to short tasks, the distribution of feature vectors is more even with 66.6% of feature vectors for short and 33.3% for long tasks, respectively. Table V shows the recognition performance for the two clusters. Comparing these, also with those shown for all tasks combined, in Table III, reveals only small performance differences (with the exception of long duration tasks, with a performance drop of 9.3 pp compared to the general results). For example clustering using the task time has a negative impact on the recognition performances with $F_1$-Scores of 75.67% and 69.49%, for short time and long task time, respectively using the direct line distances and department granularity only. Overall, the results indicate that the proposed methods can provide accurate classifications for both long and short distance tasks, and that clustering these does not provide further gains.

E. Destination granularity

We evaluated the impact of the granularity with which start and end locations are specified; concretely we compare department- vs. room-based granularity. Note that a finer-grained destination precision can be obtained manually or grown organically with active users [22]. When only the departments are known as destinations, the department centers are used as from or to destination, hence the precision of the distance features are biased by the size of the department which varies from around 600 $m^2$ to around 9840 $m^2$ in our data set, when estimated using a bounding rectangle. We approximate room accuracy by manual labelling the dispatching and receiving destinations based on an overview map of the hospital. Table VI depicts the $F_1$-scores of the best feature combination for each destination granularity. Our results show that more accurate knowledge of from and to destination yields accuracy gains despite position estimates being inaccurate; the

<table>
<thead>
<tr>
<th>Department-granularity</th>
<th>Room-granularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time $&lt; T_1$</td>
<td>Time $&lt; T_1$</td>
</tr>
<tr>
<td>75.67 %</td>
<td>81.47 %</td>
</tr>
<tr>
<td>82.66 %</td>
<td>86.05 %</td>
</tr>
<tr>
<td>Time $\geq T_1$</td>
<td>Time $\geq T_1$</td>
</tr>
<tr>
<td>69.49 %</td>
<td>84.11 %</td>
</tr>
<tr>
<td>83.75 %</td>
<td>90.85 %</td>
</tr>
<tr>
<td>Distance $&lt; T_d$</td>
<td>Distance $&lt; T_d$</td>
</tr>
<tr>
<td>75.60 %</td>
<td>83.20 %</td>
</tr>
<tr>
<td>82.26 %</td>
<td>89.11 %</td>
</tr>
<tr>
<td>Distance $\geq T_d$</td>
<td>Distance $\geq T_d$</td>
</tr>
<tr>
<td>75.25 %</td>
<td>84.95 %</td>
</tr>
<tr>
<td>78.90 %</td>
<td>84.46 %</td>
</tr>
</tbody>
</table>

TABLE V: $F_1$-scores for long and short tasks with thresholds: $T_1 = 628.08$ seconds and $T_d = 256.28$ meters using respectively direct line distance features and direct line distance features combined with historical proximity features.

TABLE VI: $F_1$-score of varying destination granularities

<table>
<thead>
<tr>
<th>Granularity</th>
<th>Dept granularity</th>
<th>Room granularity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>84.43 %</td>
<td>79.73 %</td>
</tr>
<tr>
<td></td>
<td>86.80 %</td>
<td>89.21 %</td>
</tr>
</tbody>
</table>

TABLE VII: $F_1$-score for orderly, equipment and patient data with the optimal feature type combinations for respectively department and room granularity.

F. Usefulness of patient and equipment positions

As mentioned, data was collected from WiFi tags placed on the equipment used and on the patient being transported. Table VII depicts the $F_1$-score for different combinations of data sources. Using data from only the orderly’s phone yields a relatively high $F_1$-score of 81.91 for department granularity. Combining all three data sources yields an performance increase for both department and room granularity, with a gain of up to 3.06 pp compared to using solely the order’s phone.

G. Idle state detection

For the sake of also detecting the event type task end, we modeled an additional fifth phase called idle. However, adding this phase impacts the prediction accuracy. The idle phases are often short since orderlies are able to pick a new task immediately after a patient has been handed over and the movement patterns in these phases are varying substantially. Figure 5 depicts both the recognized as well as the true phases for a set of consecutive tasks—without modelling the idle phases (Figure 5c) and with modelling them (Figure 5d). Most noticeable is the amount of noise the additional phase introduces. The uncertainty generally increases when introducing the phase idle, which is reflected by a decrease in $F_1$-score. For both department granularity and room granularity the best feature type combinations remain the same with $F_1$-scores of 77.32% and 83.90% respectively.

VI. RELATED WORK

The recognition of human work activities has been investigated using various sensor combinations and in various work domains ranging from industrial environments to hospital work settings: Activity recognition in industrial work settings has been investigated, e.g. by Stiefmeier et al. who used wearable sensors, e.g., RFID sensors, to support the training of assembly workers [1]. Koskimaki et al. presented the assembly line work domain a proactive instruction system focusing on simple
activities such as hammering and screwing. Recognition was based on a wrist-worn sensor which provided acceleration and angular speed to achieve accuracies at 90% [23].

Moreover, activity recognition research within work settings has also focused on the clinical work in hospitals. For example, Favela et al. detect the clinicians’ high-level activities in medical wards using manually labeled contextual information from a long observation study [24]. Neural network and a Hidden Markov Models (HMMs) can then be trained using contextual information including the clinicians’ locations and the artifacts, see also Sanchez et al. [4].

Parlak et al. focused on exploring activity recognition methods in the emergency department using RFID sensors and RFID tagged medical items that could detect interactions with certain objects. These interactions were used to recognize various tasks in trauma resuscitation [5]. Others have aimed at recognizing activities of nurses such as walking and carrying wheelchairs, based on a combination of an infrared location system, and body movements [2].

Furthermore, several studies focused on approaches for recognition of surgical steps within operation rooms, see, e.g., Padoy et al. [25], [7], who proposed two approaches to recognize surgical steps. The first approach is based on contextual data extracted from the laparoscopic camera’s images. Additionally, information about instruments being used was manually extracted from video recordings [25]. In the second approach, image analysis of 3D motion-flows is used for phase recognition. The activities inside a mock-up (simulated) OR are captured with a multiple-camera system and the activity inference is combined with workflow information [7].

Using a combination of different types of embedded and body-worn sensors, e.g., RFID readers and tags, locations of the staff involved in the operation, and the location of the instruments inside an operation room, Bardram et al. can predict the phases, e.g., sterilization, of an operation [6].

VII. DISCUSSION

In this section we will discuss the presented methods obtained in terms of feasibility and generalizability to other task types and work domains, and we discuss how the methods’ performance may be improved.

A. Task phases beyond hospital service work

In this paper we have focused on recognizing four phases of mobile service workers’ tasks. For recognition of these phases, we have exploited that the respective tasks have a specified start and end location which are known via a task management system, and that the phases relate directly to these locations (or to the transition to or between these). We expect the presented models and methods to be applicable for other domains where service workers are responsible for transporting resources across building complexes or where services are to be carried out at known locations. However, there are also task types that are not covered by this model such as various clinical functions at a hospital, e.g., by doctors and nurses, where the workers’ tasks consist of various stationary activities or activities that include several destinations which may be unknown a priori. To extend the work presented here to such activities one can borrow from existing HAR research, a large amount of which utilizes on-body motion sensors as given by the smartphones used in our study.

B. Improving the recognition performance

We have presented promising results on recognizing the task phases from mobile workers, which though leave room for improvement, especially in instances close to a phase transitions. To improve the results we plan to investigate learning schemes commonly used for sequence data, e.g., Hidden Markov Models (HMMs) and particle filtering. Such techniques allow for exploiting the temporal aspect of phases and the transitions between phases. However, the collected data used for training only contained tasks with phases that move forward in time. In this work, instead of applying a HMM, we added historical features to incorporate the temporal
aspects of phases, by summarizing the distances and proximity probabilities to both the start and end points of a task. These features are potentially limited by the same issues as the HMM or other sequential methods, but we show that complementing with other features achieves promising results. For future work we plan to focus on spatio-temporal analysis aspects and we plan to integrate these, e.g., by using the IntraTime system [13] which can accurately predict travel times by taking into account the building structures. Additionally, we would like to integrate transportation mode detection as a feature, since this has a high impact on the workers’ travel times, phase transitions and movements. Recent results have shown that it is possible to achieve highly accurate results in indoor environments [14].

VIII. CONCLUSION

In this paper we have presented methods for recognizing basic task phases in the work routines of mobile workers. The methods utilize sensing opportunities commonplace in many large indoor work environments: foremost a coarse grained WiFi indoor positioning system, which is easy to maintain and deploy, aided by inertial sensors common in smartphones, and by task meta-data as provided by a task scheduling system such as the task’s start and end points. As a result, in contrast to previous work, the methods proposed here have a sufficiently low deployment and maintenance effort, rendering them suitable for real-world application in large building complexes, such as hospitals.

The proposed methods were evaluated on real-world data from more than 80 tasks conducted by hospital orderlies. Using a standard classifier, random forest, we showed that by using coarse grained positioning in connection with task information and inertial sensors high accuracy of task phase recognition can be achieved.

Aiding future designs of task phase recognition systems, we have furthermore investigated several factors that can influence the accuracy of the task phase recognition based on our experiments and give indications on how to limit such factors and make the recognition more robust; our proposed methods are in particular robust to variations, e.g., in the distances between task start and end. The presented experiments investigate how to design the next generation of task phase recognition. Lastly, we are currently working on extending the results to accurately predict the remaining of a task and have already promising results. This would enable the mobile workers with a higher granularity information about the progress of their co-workers.

ACKNOWLEDGEMENTS

The authors acknowledge the support granted by the Danish Advanced Technology Foundation under J.nr. 076-2011-3.

REFERENCES


