Phase Recognition during Surgical Procedures using Embedded and Body-worn Sensors

Jakob E. Bardram, Afsaneh Doryab, Rune M. Jensen, Poul M. Lange, Kristian L. G. Nielsen, and Søren T. Petersen
IT University of Copenhagen
Rued Langgaards Vej 7, DK-2300 Copenhagen, Denmark
Email: {bardram,adoyrab,rmj,pmla,klgn,stpe}@itu.dk

***Abstract*—In Ubiquitous Computing (Ubicomp) research, substantial research has been directed towards sensor-based detection and recognition of human activity. This research has, however, mainly been directed towards activities of daily living of a single person. This paper presents a sensor platform and a machine learning approach which are able to sense and detect phases of a surgical operation. Automatic detection of the progress of work inside an operating room has several important applications, including coordination, patient safety, and context-aware information retrieval. We verify the platform during a surgical simulation. Recognition of the main phases of an operation was done with a high degree of accuracy. Through further analysis, we were able to reveal which sensors provide the most significant input. This can be used in subsequent design of systems for use during real surgeries.**

***Keywords*-Phase Recognition, Activity Recognition, Pervasive Healthcare, Operating Room, Machine Learning, Sensors**

I. INTRODUCTION

Activity detection is core to Ubiquitous Computing (Ubicomp) research. Even though context-aware systems may be implemented based on simple context information like location, it is widely recognized that knowing what a user is doing (i.e., the activity of the user) is important to move context-aware computing further. So far, much research has been directed at recognizing daily activities done by a single person while at home [1], [2], [3], or outdoor activities [4], [5], [6]. Even though activity detection in car manufacturing has been researched [7], activity detection in collaborative workplaces is very rare.

In this paper, we present an approach to activity detection inside a surgical operating room (OR). Compared to existing work on activity detection, activity detection inside ORs has two distinct characteristics. First, work and hence activities inside an OR is executed by several clinicians working together at the same time in the same location. This means that an activity can be performed by one person; it can be done by two or more persons in collaboration; and several activities can take place simultaneously. Any sensor platform and activity detection systems should be able to handle this nature of co-located, concurrent, and collaborating activities inside the OR.

Second, in a surgical situation, the core activities to detect are related to the phases of an operation. Hence, in most cases what is important to recognize is the progression of an operation. Recognition of the progression and phases of an operation is important to a wide range of applications [8], [9]. For example, in peri-operative coordination and communication it is essential for clinical staff outside the OR to know the phase of the operation [10]. But specifying the phase has currently to be done manually by the circulating nurse inside the OR, and is hence prone to errors, delays, and negligence. Another example is an Electronic Medical Encounter Record (EMR), which automatically detects and records important medical events during surgery [11].

The system for phase recognition during surgical procedures presented in this paper was designed based on field studies of operations and interviews with surgical staff. The system consists of a range of sensors that are both embedded into the operating room and the instruments used during surgery, as well as a body-worn sensors that senses what surgical instrument a clinician is holding. Moreover, fine-grained location tracking is used to track the clinicians while inside the OR.

This sensor platform was designed and implemented together with clinical staff, and then verify its feasibility during a series of simulated operations. The results show that it is possible to achieve high classification accuracy of OR activity phases when using standard feature vector classifiers. The experiment also shows that accumulating sensor activity in historical features is important, since accuracy falls dramatically without historical information.

Besides establishing evidence for a high recognition accuracy, the experiment also helps to analyze the weight and hence of the different sensors. This analysis is essential since it would be challenging and costly to deploy the current sensor platform in an OR and use it during real surgeries. Hence, the paper also provides important input for the optimal design of a sensor platform for activity recognition inside an OR.

II. RELATED WORK

The general approach in the field of activity recognition is to use sensor technologies and machine learning algorithms to infer human activities. Most existing work has addressed the recognition of daily activities of a single person in a home or personal setting using different combinations of sensing technologies, such as embedded sensors [1], location tracking [3], wearable sensors [2], [12], and video analysis [13]. Some approaches have focused on activity recognition based on detecting and analyzing the sequence of objects that are being used by the user [14], [13]. Others have tried to combine the sensor readings with commonsense information extracted, e.g., from the web [15], [16] to minimize labeling overhead. In another type of work, the information from accelerometers and body-worn sensors has been used to identify physical activity, such as walking and sitting [17], [18], [19]. Location-based techniques for activitiy recognition have mainly explored the information about the location of people to infer the activity. Signals from GPS is usually used for estimation of outdoor activities [4], [5], and mobile phone [6] is used to work both indoor and outdoor.

In a hospital domain, Favela et al. [20], [21] estimate high level activities of clinicinas in a medical ward using neural network and Hidden Markov Models (HMM). In [22], logic programming is used to detect healthcare activities in a pervasive hospital environment where positions of people and things are tracked. Inside operating rooms, Agarwal et al. [11] have proposed a rule-based system for detecting significant medical events during surgery. Based on sensor input from RFID tags on patients and real-time data from the patient monitoring system and anesthesia machine, they are able to infer higher level events, like the onset of anesthesia. A so-called Electronic Medical Encounter Record (EMR) is automatically constructed. This EMR records and correlates medical events and video streams with the inferred higher level event model of the surgery.

Studies by Padoy et al. [9], [8] propose HMM-based approaches for online recognition of surgical steps. In the first approach, the contextual data was extracted by processing images from the laparoscopic camera and manually extracting information about instruments being used from video recordings [9]. In the second approach, image analysis of 3D motion-flows are 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 [8].

Our work seeks to take activity recognition as done in domestic or traveling settings and apply this for activity recognition in an OR. Due to the nature of surgeries, our activity recognition focuses on activity detection inside a room with multiple actors working both collaboratively and concurrent. This is in contrast to most work in the domestic domain, which only looks at activity detection of a single person. In relation to the research on activity detection in an OR setting, the most close and relevant is the work by Padoy et al. Our work differs from their work in the sense that we use a combination of different types of sensors including RFID, indoor location tracking, and wearable sensors. Our sensor setup is based on an analysis of the activities inside the OR which points out that sensing the usage of objects together with the location of people and objects are powerful indicators for recognition of activities inside the OR.

Figure 1. A typical setup inside the Operating Room (OR) during surgery, involving two surgeons, a scrub nurse, and a circulating nurse (using the PC in the back). The anesthesia nurse is behind the curtain.

Although an extensive number of different machine learning techniques for activity recognition has been proposed, the dominant approach has been to use HHMs. The main focus of our work is, however, to investigate the feasibility of the sensor platform. Therefore we have chosen to use Decision trees as a technique that can show the impact of each feature attribute in the learning process. This allows us to choose a suitable set of sensing features, which would lead to the most optimal sensor setup. For this purpose we have used decision trees as the analysis technique while adding historical features to address the temporal aspect of a surgery.

III. Activities in Surgery

In order to get a detailed understanding of the activities and phases involved in surgery, we conducted detailed observations of 7 operations at a gastric-surgical department in a university hospital. Since laparoscopic procedures are becoming more and more prevalent, we selected laparoscopic appendectomy – i.e. removal of the appendix using non-invasive surgery – as the focus of our research. The operations were done by different surgical teams, in different operating rooms, performing the same operation on different patients. As such, we argue that the dataset is generic enough for this specific operation. All operations were video-recorded and transcribed for further analysis.

The flow of a typical laparoscopic appendectomy is illustrated in Figure 2. The operation follows the standardized way all operations are performed. Before the operation is scheduled to start, the nurse anesthetist prepare for surgery by checking the anesthesia devices, and arrange medicine and the anesthesia instruments. Meanwhile, the scrub nurse and the circulating nurse prepares the surgical instruments

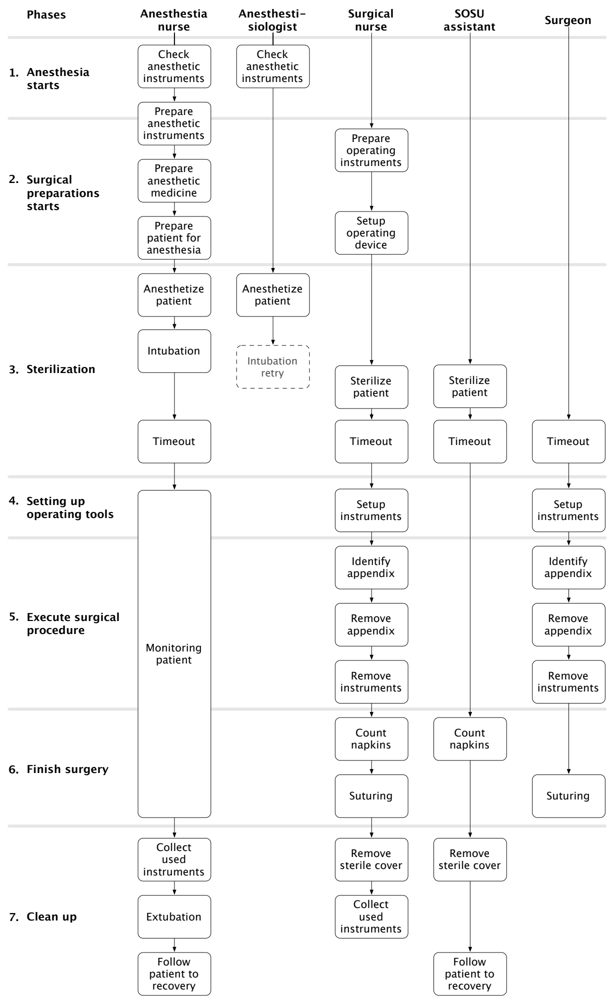

Figure 2. The activities involved in a laparoscopic appendectomy. The leftmost column lists the seven phases of the operation. The other columns show the flow of activities for each participant grouped into the phases.

and devices by placing them on the operating trolley next to the operating table. When the patient arrives, s/he is being anesthetized. When the surgeon(s) enters the OR, the operation starts. During the operation, the nurse anesthetist constantly monitors the patient's condition, and transfuses blood and medicine as needed. The scrub nurse assists the surgeon and hands him the instruments and material. When approaching the end of the operation, the nurse anesthetist starts waking the patient, and the scrub nurse starts collecting all surgical instruments. Finally, the patient is moved to the recovery department.

In order to draw patterns of the surgery activity, we did a detailed temporal analysis of the surgeries. We transcribed our video recordings using a coding schema that helped identified the specific actions performed in the OR and their detailed manual tasks, as well as the involved actors, physical instruments, and locations inside the OR. Figure 3

Figure 3. A snapshot of the intubation action

shows the intubation action, which consists of three tasks:

- The anesthesiologist holds the patient's head and opens the patient's mouth using a laryngoscope.
- The nurse anesthetist gives him the ventilation tube.
- The anesthesiologist puts the ventilation tube into the patient's trachea.

The instruments and objects involved in this action include a Laryngoscope, a ventilation tube, and the patient. A nurse anesthetist and an anesthesiologist participate in this action, and all operations take place at the operating table.

We identified 36 actions and 137 tasks in the laparoscopic surgery. The number of (human) tasks in each action varied from 1 to 13. Some actions had several actors (33%), and others were done individually. Almost all actions (97%) in the laparoscopic surgery involved using at least one physical instrument, and 78% involved several instruments. Only few coordination and communication actions did not involve visible or physical tools, e.g., talking about the remaining time of the surgery. We also noticed that there was a direct relation between the physical instruments and the actions, e.g., a laryngoscope is only used in intubation action. We identified the use of 39 different types of anesthesia instruments and medicine, and 37 different types of operation instruments.

Despite being co-located inside the OR, the actions are performed in certain areas in the room. We identified 4 important zones, i.e., specific areas where collections of actions were carried out. These zones were the anesthesia machine zone ($l1$), the anesthesia cabinet zone ($l2$), the operating table zone ($l3$), and the operating trolley zone ($l4$). Figure 4 shows a schematic view of these areas in the OR.

Anesthesia related actions are done in $l1$, $l2$, and $l3$; and operation related actions are performed in $l3$ and $l4$. For example, prepartion of anesthetic and anesthesia instruments is done near the anesthesia cabinet ($l2$) whereas preparation of operation instruments is carried out in $l4$. The team members move between these zones. For example, depending on the action, the anesthesia nurse switches between $l1$, $l2$, and $l3$. The frequency of movements between different zones depends on the operation phase. During the preparation and ending phases the clinicians move between zones more frequently than during the surgery.

Our observations of different types of surgeries together with interviews with clinicians indicates the fact that despite the type, all surgeries follow the same standard procedure, i.e., they always start with preparation followed by the operation and end with cleaning up. For example, the operation can only be executed if the patient is prepared and anesthetized, or the clean up process can only start if the operation is finishing. We identified three types of actions in all different types of surgeries we observed:

- Actions that are common in all types of surgeries and are done in certain phases of the surgery. These actions are prerequisite for initiating other actions. For instance, anesthetization is always done in the preparation phase and before the surgery starts.
- Actions that happen in some surgeries and in particular phases of the surgery. For example, the intubation is not necessary for all types of surgeries, but if the patient should be intubed, it will be done before starting the surgery.
- Actions that are carried out in some surgeries and can be done in all phases. For example, the nurse anesthetist can document the ordering of the blood either during the surgery or after the surgery is finished.

The duration, the type and number of instruments, the number of tasks, and the actors involved in an action vary in different surgeries. It depends on many factors especially the patient's condition and the type of surgery. For example, if a patient has a neck problem, the ventilation should be done by putting a mask on the patient's mouth instead of sending a tube into the patient's lung.

Some actions do not follow specific orders. For instance, it does not matter if the anesthetist checks the devices before preparing the medicine. The only thing that matters is, that both medicine and devices should be ready before the patient enters the OR. The order of doing these actions depends on individuals and their routines. The actor of an action or an operation can change. For example, if during the intubation the operation of putting the ventilation tube in the patient's lung does not succeed by a nurse anesthetist, it will be taken over by another nurse or the anesthesiologist.

Sensing the use of an instrument by a team member can be an indicator for other members' actions. For example, the anesthetist always needs to know when the surgery is finished, so she or he can start preparing for the ending phase. If the surgeon starts to use the suture needle and thread, that usually means the surgery is ending, so the anesthetist can start the process of waking the patient.

Getting involved in an action usually depends on the participants' roles and specializations. An anesthetist is concerned about the anesthesia and the patient's general condition during the surgery, whereas a surgeon concentrates on the surgery. This division of labour causes division of tools and instruments used by the members. An anesthetist would rarely touch the surgery instruments, partly because she or he is not scrub, and partly because there is no direct link between her/his tasks and the surgery instruments.

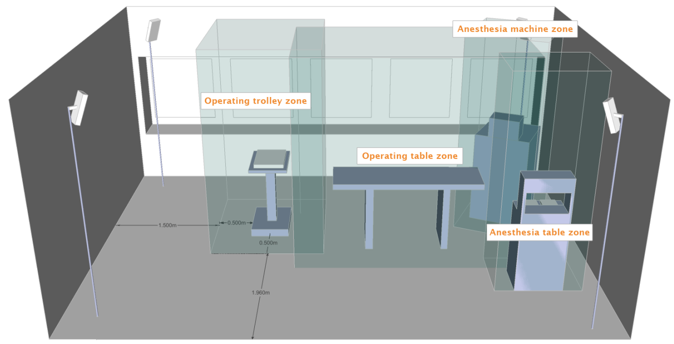

Figure 4. The main zones inside the OR: (i) Anesthesia table, (ii) Anesthesia machine; (iii) Operating table; and (iv) Operating trolley.

IV. Sensor Platform

Based on our detailed study, the following parameters seemed important to track:

- The location of clinicians and the patient
- The location of objects on different tables
- The use of objects and instruments by the clinicians.

Based on this insight we created a sensor platform with three sub-sensor systems sensing each of the items listed above, and a central server for collecting, filtering, time stamping, and storing sensor readings.

To track the location of people inside the OR, we used the Ubisense realtime location tracking system (RTLS). When a person is wearing a tag, this RTLS is able to track him with a theoretical accuracy of 10 cm. in space (i.e., x, y, z coordinates). Moreover, the tracking system allow us to divide the OR into different zones, reflecting the four main areas identified above. This is illustrated in Figure 4. Because the Ubisense system is not completely accurate, we introduced a buffer in the order of 50 cm. on each side of the zones. Furthermore, since the staff member will be positioned for instance in front of a table, the zone is extended in the direction where the person is most likely to be positioned. This makes the zones overlap and it is therefore possible for a staff member to be in more than one zone at the same time.

To track the instruments and objects on the tables, all instruments were tagged with passive RFID tags, and tables had built-in RFID readers. Instruments are only placed on the anesthesia table and the operating trolley. Since the OR contains many metallic instruments and tables, ultra-high frequency (UHF) RFID technology was used. UHF is more robust in such an environment, but issues with reflection and shielding still exist. Figure 5 shows the anesthesia nurse

Figure 5. The anesthesia nurse preparing instruments at the anesthesia table. All instruments are tagged with RFID tags and the table has a build-in RFID reader. Moreover, the nurse carries a palm-based RFID reader that reads which instruments she is holding.

preparing instruments at the anesthesia table, which has a build-in RFID reader.

The third sensor is a wireless, palm-based RFID sensor that is able to detect which instruments and objects a clinician is holding in his or her hand. This sensor is composed of a micro controller board (Arduino Duemilanove[1]), an RFID reader module (ID-12 Innovations[2]), and a wireless unit composed of a shield (Arduino XBee Shield Empty) and a wireless module (XBee 1mW Chip Antenna[3]), that allows the micro controller board to communicate wirelessly over a modified ZigBee protocol to the server. Figure 5 shows how this palm-based sensor is attached to the upper arm of the anesthesia nurse while the RFID reader is attached to her palm. Compared to the UHF readers, this smaller ID-12 reader can only detect one tag at a time and it operates in the low frequency (LF) and hence requires a different tag than the UHF tags. Therefore, in our experiments, all instruments and objects needed both LF and UHF tags. In total, four palm-based sensors were used by the surgeon, anesthesiologist, anesthesia nurse, and the scrub nurse respectively.

V. Experimentation

In order to verify the feasibility of the sensor platform for phase recognition in an OR, we conducted an experiment using a simulated setup. At the current stage of our research, the sensor platform is not suitable for deployment during real surgeries. There are some ergonomic and hygiene related issues that needs to be addressed first. Moreover, due to safety regulation, even experimental surgical instruments needs to be approved before use in surgery on real patients. Our approach was therefore to test the sensor platform and activity recognition system during a surgical simulation. Surgical simulation is a common method in medical practice and is used to educate clinical staff, to test new procedures, and to evaluate new clinical equipment.

The purpose of the experiment was to verify that the sensor design was sufficient; how accurate phase recognition could be done based on the sensed data, whether standard machine learning classifiers could be used for phase recognition, and lastly to identify which features coming from the sensor system are most important for achieving high accuracy. The latter is important in the further design and development of such sensing technologies for ORs.

The simulation took place in a laboratory which was rigged to resemble an OR. The setup included four palm-based sensors, two table-based sensors, and the Ubisense location tracking system setup to recognize the four main zones. The setup is illustrated in Figure 4. We tagged real surgical instruments and performed the operations on a fictive patient.

We executed four experiments that simulated a laparoscopic appendectomy operation. Each simulation scenario was designed as a set of steps as outlined in Figure 2. The scenarios were based on the video recorded operations and designed in close collaboration with domain experts, i.e. surgeons, anesthesiologists, and nurses. Exact timing of the different step was extracted from the original video recordings. The scenarios varied in terms of the exact timing of the steps, e.g., when the surgical nurse starts the preparations. Moreover, some of the activities were performed with slight variations. For example, simulating that intubation fails and is completed with a laryngeal mask instead. The simulations were performed by the researchers.

A. Feature Processing

The raw sensor readings were sampled, synchronized and transformed into feature instances by the sensor platform on the fly. A representative subset of the logged features is shown in Table I. Since a person can be in two zones at the same time and an RFID tag can be read by two sensors at the same time, a boolean value is used for each feature.

Among many machine learning techniques that were tested including Bayesian Networks, Logistic regression, and Neural Network, we found the results of Decision Trees most useful for two reasons. First, because our objective was to understand the impact of the sensors on the classification, which can be established by looking at the generated decision trees. Second, because the system should work in real

[1] http://www.arduino.cc/
[2] http://www.id-innovations.com/
[3] http://www.digi.com/

Name	Description
LRSH	Laparoscopic retractor in surgeon's hand
SSNH	Suture in surgical nurse's hand
NHSH	Needle holder in surgeon's hand
VCOT	Verres cannula on operating trolley
NHSNH	Needle holder in surgical nurse's hand
TOT	Trocar on operating trolley
NHOT	Needle holder on operating trolley
SAT	Syringe on anesthesia table
AAMZ	Anesthesiologist in anesthesia machine zone
SSH	Scalpel in surgeon's hand
ANAMZ	Anesthesia nurse in anesthesia machine zone
SNOTZ	Surgical nurse in operating table zone
PMFANH	Pre-medication form on in anesthesia nurse's hand
SOTZ	Surgeon in operating table zone
...	...

Table I
A REPRESENTATIVE SUBSET OF THE LOGGED SENSOR FEATURES.

Activity	1	2	3	4	5	6	7
1	**1**	0	0	0	0	0	0
2	0.04	**0.82**	0.14	0	0	0	0
3	0	0.26	**0.57**	0.18	0	0	0
4	0	0	0.07	**0.93**	0	0	0
5	0	0	0	0	**1**	0	0
6	0	0	0	0	0	**0.5**	0.5
7	0	0	0	0	0	0.41	**0.59**
1	**0.79**	0	0	0	0	0.02	0.19
2	0.03	**0.33**	0.21	0	0.17	0.05	0.22
3	0	0.48	**0.28**	0.02	0.11	0	0.11
4	0	0.41	0.04	**0.25**	0.28	0	0.02
5	0	0.41	0.01	0.11	**0.43**	0.01	0.02
6	0.05	0.18	0.01	0	0.01	**0.5**	0.25
7	0.29	0.09	0.04	0	0	0.16	**0.42**

Table II
CONFUSION MATRIX OF A DECISION TREE (J48) USING ALL FEATURES (TOP) AND USING NO HISTORICAL FEATURES (BOTTOM).

time during operations, and having a short inference time is thus critical. Using decision trees have a constant execution time. One problem with decision trees is, however, that they assume independence between different labels and as a result the temporal dependency among the procedure steps is uncovered. While this temporal aspect of surgical procedure could be addressed in more complex models such as Hidden Markov Models, these techniques are often expensive to train and require a large dataset and structure learning.

As the sensor readings only provide information about the current state of the operation within a given second it is difficult to distinguish two identical states. For instance, it is not really possible to know whether a surgeon is picking up the scalpel in the start or at the end of a surgery. One way to address this problem is to add wall clock time to the feature set. However, the wall clock time interval of a phase can vary a lot between surgeries. It depends on how fast the staff is, as well as how difficult the patient is to operate.

Instead, we added a so-called historical feature for each sensor feature that is equal to the number of times $\{0, 1, 2, \ldots\}$ that the sensor feature has been true. The classifiers do not distinguish between sensor features and historical features. An example of a historical feature is the total number of seconds that the anesthesiologist has been in anesthesia machine zone, at the point in time the feature instance is logged. The sensor platform logs historical features simultaniously with ordinary sensor features.

In order to train and evaluate the sensor platform, we labeled the correct phases for each collected feature instance. For this purpose, we used an application that is able to display the collected data as well as show the video recordings.

VI. PHASE RECOGNITION RESULTS AND ANALYSIS

We used a leave-one-out cross validation on our four data sets $D_1, \ldots, D_4$. The instances of each data set were the logged feature instances of a unique surgery simulation. The size of each data set was 1051, 990, 1116, and 1172, respectively.

Formally, for each validation experiment, we train a classifier four times $C_1, \ldots, C_4$, where C_i is the classifier trained on $D_1 \cup \ldots \cup D_4 \setminus D_i$ and validated on D_i. Thus, each classifier is validated on instances from an unknown surgery as would happen in a real application. Experiments vary in the type of classifier and subset of features considered. The result of an experiment is a 7×7 confusion matrix $M = [m_{ij}]$, where m_{ij} is the fraction of phase i instances classified as phase j when classifying the validation sets $D_1, \ldots, D_4$ using their associated classifiers $C_1, \ldots, C_4$. Notice that the statistical significance of each experiment is quite high due to the large number of instances classified in the validation sets (4329 in total).

A. Phase Recognition

The primary objective of the experiment was to examine the performance of our approach in phase recognition using a decision tree. As mentioned above, the issue of using standard classification methods in labeling time series data was addressed by adding historical feature attributes to the feature vector. Recall that a historical feature accumulates the time that an associated sensor feature is active and thus summarizes the past states of the OR. The bottom part of Table II shows the results achieved by the decision tree classifier when removing historical features from the data sets. As can be seen, using historical features is essential to achieve a high classification accuracy. The fraction of instances classified correctly has dropped significantly and the deviation of the classification has increased a lot.

B. Sensor Significance

The secondary objective of the experiment was to evaluate the effect of the different sensors in achieving accurate phase recognition. An advantage of the DT classifier is that it produces a decision tree that can be inspected manually. Essentially, DT algorithms choose features greedily according to how well they individually classify the training data. If a feature is absent from the produced decision tree, it is therefore reasonable to conclude that it is a poor predictor of the target feature.

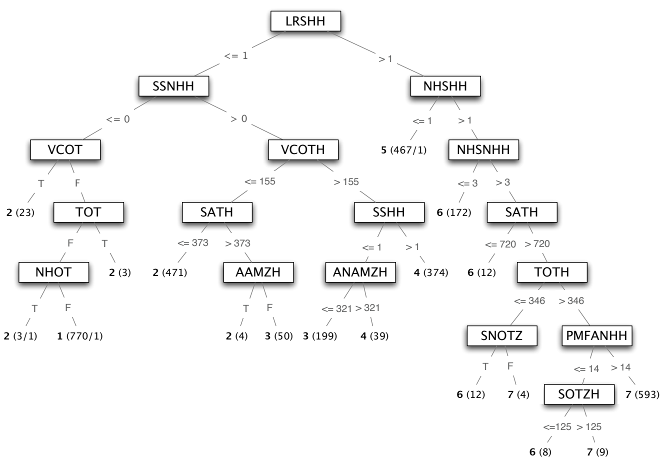

Figure 6. The pruned decision tree constructed by the DT classifier using data set 1–3 for training and data set 4 for test.

Figure 6 shows one of the four decision trees produced by the DT classifier in the experiment shown in Table II. The remaining three decision trees were similar in structure. The tree shows that features from all sensor systems are present in the tree: wristband sensors (e.g., LRSHH, SSNHH, and NHSHH), table sensors (e.g., VCOT and SATH), and Ubisense sensor (e.g., SOTZH and AAMZ). The wristband sensor features are higher in the tree than the table sensor features, which in turn are higher in the tree than the Ubisense features. This indicates that the wristband sensor features are the strongest phase predictors while the Ubisense features are the weakest.

However, the absolute difference in prediction strength between the sensor systems cannot be determined from the decision tree. To investigate this further, we conducted three experiments with: 1) only features from the Ubisense sensor, 2) features from the wrist band and table sensors, and 3) only features from wristband sensors. The results are shown in Table III.

We first note that since we have reduced the number of available features for classification, none of the results achieved in this experiment could have higher classification accuracy than the DT result shown in Table II . We attribute the fact that higher classification accuracy actually was

Activity	1	2	3	4	5	6	7
1	**0.92**	0.08	0	0	0	0	0
2	0.42	**0.53**	0.05	0	0	0.01	0
3	0.01	0.19	**0.5**	0.21	0.08	0	0
4	0	0	0.15	**0.47**	0.34	0.03	0
5	0	0	0.05	0.25	**0.36**	0.13	0.21
6	0	0	0	0.04	0.54	**0.25**	0.18
7	0	0	0	0	0.42	0.17	**0.41**
1	**1**	0	0	0	0	0	0
2	0.04	**0.83**	0.13	0	0	0	0
3	0	0.24	**0.59**	0.18	0	0	0
4	0	0	0.07	**0.93**	0	0	0
5	0	0	0	0	**1**	0	0
6	0	0	0	0	0	**0.82**	0.18
7	0	0	0	0	0	0.37	**0.63**
1	**1**	0	0	0	0	0	0
2	0.04	**0.86**	0.09	0	0	0	0
3	0	0.03	**0.68**	0.29	0	0	0
4	0	0	0.07	**0.93**	0	0	0
5	0	0	0	0	**1**	0	0
6	0	0	0	0	0	**0.59**	0.41
7	0	0	0	0	0	0.26	**0.74**

Table III
CONFUSION MATRIX FOR UBISENSE ONLY (TOP), WRISTBAND AND TABLE SENSORS (MIDDLE), AND WRISTBAND ONLY (BOTTOM).

achieved when using only wristband and table sensors to the sub-optimality of the DT algorithm due to its greedy selection of nodes in the decision tree.

VII. DISCUSSION

This section discusses the results of our research in terms of whether phase recognition is feasible to use during surgeries; how our sensor platform can be improved; and how the classification accuracy can be improved.

A. Phase Recognition in ORs

The results show that the proposed sensor platform can successfully recognize the seven phases of a laparoscopic appendectomy in a simulated setup. Using a decision tree classifier, we were able to achieve a high recognition accuracy, and the labeled phase was never more than one phase wrong. The results also show that creating a real-time activity detection system in the OR is feasible. One should note that the recognition rates in, e.g., Table II are results for each sampling, i.e., for every second. Hence, if we assume a uniform distribution, it should be easy to make a sliding average over, e.g. 10-20 seconds intervals, which would result in very high recognition rates. Moreover, these good results are obtained using a decision tree approach. This means that once the decision tree is built from training data, the classification can happen really fast.

These results show that it is feasible to build applications that take into account automatically sensed phases during an operation. This is important because, studies have shown that accurate knowledge about the progress of an operation is essential for coordination in an OR suite on a large hospital, which again has significant impact on efficiency [23].

B. Improving the Sensor Platform

Table III shows that the highest degree of phase recognition is obtained using the RFID-based sensors in the tables and worn on the wrist by the clinicians. This result is challenging since this part of the sensor setup is the most demanding sensors to develop into a version acceptable for real use in an OR. It has two main challenges. First, the sensor reader needs to be deployed on the surgeons and nurses, which inevitably will be intrusive to them. Moreover, the sensors will pose a hygiene risk and need to be embedded in the gloves used by clinicians. Hence, the ergonomic of this sensor is a challenge to design in a satisfying way. Second, all instruments, tools, and utensils used in the OR need to be tagged with RFID tags. This is challenging due to several factors; many surgical instruments are metallic which does not work well with RFID technology, and some surgical instruments like a needle and a thread are very small and hence difficult to tag.

In order to investigate why location data did not play any role in the phase recognition, we compared the manually labelled position data obtained from the video recordings with the Ubisense features. This shows that even when averaging the sensor output in each second, the position logged by the Ubisense sensor are unstable and inaccurate with a high degree of noise. We believed this to be the main reason for the low prediction value of this sensor. Therefore, we performed a small experiment, where we ran the phase recognition using only manually labeled position data, thereby obtaining a high location accuracy. This experiment showed that phase recognition was very high with an average of 92%. Hence, we should not disqualify location data as such to be irrelevant, but location data should be both very precise and accurate in order to work. And since location tracking sensors have the advantage of not being touched by people, thereby reducing hygiene-related issues, it seems like improving the indoor location tracking using other technologies might be useful to investigate further.

C. Improving the Classification

Even though the classification results from the simulated operations are promising, real operations contain much more complexity and variation in data. The characteristics of a machine learning method to be used for OR phase recognition include a short inference time despite large amount of time series data. As mentioned in previuos sections, generative models such as HMMs are the most common approach for sequential learning. However, despite their broad use, these models suffer from complexity in structure and expensive training. Our approach in adding historical features to the feature vector and using DT is a simple but promising solution. DT is among the fastest candidates when it comes to learning and inference. We also demonstrated the impact of the accumulated features on accuracy in Table II. However, we expect that in order to have more accurate results using data from real surgeries, we might need to improve the classification. We can do this in three ways. First, by adding other useful and significant features into the feature vector we can reach a higher accuracy rate. Second, because the data in real-world scenarios will have much more variations than simulated data and in order to avoid the risk of over-fitting in small datasets, we will need to collect more data for training to improve the accuracy and achieve more stable results. In this paper, we used four data sets, but we intend to experiment with at least 10 more data sets and compare the classification results. Finally, as the real-world data evolves over time, using the same classifier might not be appropriate. Instead, we consider to examine a weighted ensemble of different classification methods such as naive Bayes and Decision trees. The weight of each classifier can then be calculated and used to vote for the final output. This will increase the accuracy of the classification results and address the issue of concept drift in the evolving data.

VIII. Conclusion

This paper presented a sensor platform and classification system for phase detection during laparoscopic appendectomy procedures in an operating room (OR). The system consists of a fine-grained real-time location tracking system inside the OR, a range of sensors embedded into the surgical tables, and a body-worn sensor that is able to sense what surgical instruments a clinician is using.

This sensor platform was designed and implemented together with clinical staff, and then verified during a series of simulated operations. Using a standard classifier, the results showed that it is possible to achieve a high classification accuracy of OR activity phases when accumulating sensor activity in historical features.

Moreover, the experiment showed that ignoring data from the indoor location tracking system didn't affect the classification accuracy. Equally high accuracy was achieved when solely using the body-worn wristband sensor. Thus, besides providing evidence for a high recognition accuracy, the experiment also helps to analyze the weight and hence importance of the different sensors. This is essential since it would be challenging and expensive to deploy the current sensor platform in an OR and use it during real surgeries.

Recognizing phases during surgery is important for a range of applications in an OR. Such applications include systems for peri-operative coordination and communication, context-aware medical information management, and for general logging and safety systems. Based on our experiment, we have valuable information of how to design the next generation of the sensor platform and have hence come an important step closer to be able to deploy activity detection technology during real surgical operations.

References

[1] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," *Pervasive Computing*, pp. 158–175, 2004.

[2] T. Gu, Z. Wu, X. Tao, H. K. Pung, and J. Lu, "epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition," in *PerCom*, vol. 0. Los Alamitos, CA, USA: IEEE Computer Society, 2009, pp. 1–9.

[3] S. Das and N. Roy, "Learning, prediction and mediation of context uncertainty in smart pervasive environments," in *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer-Verlag, 2009, ch. 107, pp. 820–829.

[4] B. D. Ziebart, A. L. Maas, A. K. Dey, and J. A. Bagnell, "Navigate like a cabbie: probabilistic reasoning from observed context-aware behavior," in *UbiComp '08: Proceedings of the 10th international conference on Ubiquitous computing*. New York, NY, USA: ACM, 2008, pp. 322–331.

[5] D. Patterson, L. Liao, D. Fox, and H. Kautz, "Inferring high-level behavior from low-level sensors," in *UbiComp 2003: Ubiquitous Computing*, 2003, pp. 73–89.

[6] D. Choujaa and N. Dulay, "Tracme: Temporal activity recognition using mobile phone data," in *2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*. IEEE, December 2008, pp. 119–126.

[7] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster, "Wearable activity tracking in car manufacturing," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 42–50, 2008.

[8] N. Padoy, D. Mateus, D. Weinland, M.-O. Berger, and N. Navab, "Workflow monitoring based on 3d motion features," in *Proceedings of the International Conference on Computer Vision Workshops, IEEE Workshop on Video-oriented Object and Event Classification*, 2009.

[9] N. Padoy, T. Blum, H. Feussner, M.-O. Berger, and N. Navab, "On-line recognition of surgical activity for monitoring in the operating room," in *IAAI'08: Proceedings of the 20th national conference on Innovative applications of artificial intelligence*. AAAI Press, 2008, pp. 1718–1724.

[10] J. E. Bardram, T. R. Hansen, and M. Soegaard, "Awaremedia: a shared interactive display supporting social, temporal, and spatial awareness in surgery," in *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*. New York, NY, USA: ACM, 2006, pp. 109–118.

[11] S. Agarwal, A. Joshi, T. Finin, Y. Yesha, and T. Ganous, "A pervasive computing system for the operating room of the future," *Mob. Netw. Appl.*, vol. 12, no. 2-3, pp. 215–228, 2007.

[12] Y.-J. Hong, I.-J. Kim, S. C. Ahn, and H.-G. Kim, "Activity recognition using wearable sensors for elder care," *Future Generation Communication and Networking*, vol. 2, pp. 302–305, 2008.

[13] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. M. Rehg, "A scalable approach to activity recognition based on object use," in *In Proceedings of the International Conference on Computer Vision (ICCV), Rio de*, 2007.

[14] M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, and D. Hahnel, "Inferring activities from interactions with objects," *IEEE Pervasive Computing*, vol. 3, no. 4, pp. 50–57, January 2004.

[15] M. Perkowitz, M. Philipose, K. Fishkin, and D. J. Patterson, "Mining models of human activities from the web," in *WWW '04: Proceedings of the 13th international conference on World Wide Web*. New York, NY, USA: ACM, 2004, pp. 573–582.

[16] S. Wang, W. Pentney, A. M. Popescu, T. Choudhury, and M. Philipose, "Common sense based joint training of human activity recognizers," in *IJCAI'07: Proceedings of the 20th international joint conference on Artifical intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007, pp. 2237–2242.

[17] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1553–1567, 2006.

[18] J. He, H. Li, and J. Tan, "Real-time daily activity classification with wireless sensor networks using hidden markov model," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007.

[19] T. Choudhury, G. Borriello, S. Consolvo, D. Haehnel, B. Harrison, B. Hemingway, J. Hightower, P. P. Klasnja, K. Koscher, A. LaMarca, J. A. Landay, L. LeGrand, J. Lester, A. Rahimi, A. Rea, and D. Wyatt, "The mobile sensing platform: An embedded activity recognition system," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 32–41, 2008.

[20] J. Favela, M. Tentori, L. A. Castro, V. M. Gonzalez, E. B. Moran, and Ana, "Activity recognition for context-aware hospital applications: issues and opportunities for the deployment of pervasive networks," *Mob. Netw. Appl.*, vol. 12, no. 2-3, pp. 155–171, 2007.

[21] D. Sanchez, M. Tentori, and J. Favela, "Activity recognition for the smart hospital," *Intelligent Systems, IEEE*, vol. 23, no. 2, pp. 50–57, 2008.

[22] H. B. Christensen, "Using Logic Programming to Detect Activities in Pervasive Healthcare," in *Proceedings of International Conference on Logic Programming ICLP 2002*, ser. Lecture Notes in Computer Science 2401, P. Stuckey, Ed. Copenhagen, Denmark: Springer Verlag, July 29–August 1 2002.

[23] J. E. Bardram and T. R. Hansen, "Why the plan doesn't hold - a study of situated planning, articulation and coordination work in a surgical ward," in *CSCW '10: Proceedings of the 2010 conference on Computer supported cooperative work*. New York, NY, USA: ACM Press, 2010, pp. 331–340.