Designing and sharing activity recognition systems across platforms: methods from wearable computing

Daniel Roggen, *Member, IEEE,* and Stéphane Magnenat, *Member, IEEE,* and Markus Waibel, *Member, IEEE,* and Gerhard Tröster, *Senior Member, IEEE*

***Abstract*—In robotics, activity recognition systems can be used to label large robot-generated activity datasets. It also enables activity-aware human-robot interactions, and opens ways to self-learning autonomous robots. The recognition of human activities from body-worn sensors is also a key paradigm in wearable computing. In that field, the variability in human activities, sensor deployment characteristics, and application domains, have led to the development of best practices and methods to enhance the robustness of activity recognition systems.**

We argue that these methods can benefit many robotics use cases. We review the activity-recognition principles followed in the wearable computing community and the methods recently proposed to improve their robustness. These approaches aim at the seamless sharing of activity recognition systems across platforms and application domains. Finally, we outline current challenges in wearable activity recognition.

***Index Terms*—Activity recognition, Wearable computing, Robotics, Machine learning, Domain transfer.**

I. Introduction

Recognizing, sharing and reusing robot behaviors across multiple robot platforms with varying similarity is challenging. While descriptions for objects (e.g., CAD models, recognition models) and environments (e.g., geo coordinates, local coordinates, feature maps) are largely interchangeable across different robot hardware, robot task descriptions are typically highly hardware-dependent. This has prevented the generation of generic datasets for robot behaviors. However, such datasets are important and underpin many of the algorithmic advances e.g. in object recognition [1], [2], [3], or in the creation of joint world-models [4], [5], [6], [7]. It has also hindered the progress in the field of robot cognition and robot learning, by preventing robots to understand and learn from each other's actions.

Driven by the rapid progress in mobile sensing and computing, wearable computing has developed powerful methods for the automatic recognition, categorization, and labeling of human actions and behaviors from sensor data. Due to the stringent requirements dictated by user acceptance, these methods are typically robust to human variability and to hardware-dependent factors including variability in sensor type and placement. This makes them a potentially useful tool for the automatic recognition and labeling of robot behaviors, and may lead to new opportunities for research in robot cognition and robot learning.

Daniel Roggen and Gerhard Tröster are with the Wearable Computing Laboratory, Swiss Federal Institute of Technology Zürich. Stéphane Magnenat is with the Autonomous Systems Lab, Swiss Federal Institute of Technology Zürich. Markus Waibel is with the Institute for Dynamic Systems and Control, Swiss Federal Institute of Technology Zürich. The contact author is D. Roggen, e-mail: droggen@gmail.com

Annotation of large-scale activity datasets

Due to the ease of systematic data collection from robots and their potential usefulness for data mining, a future WWW for robots is likely to include datasets for a large number of behavioral strategies for different robotic platforms in different situations. Such datasets may, for example, include sensor readings for the walking behavior of a humanoid robot on different terrains, or trajectory information for the grasp behavior of a pick-and-place robot for various target objects. While individual robots are typically aware of their current behavior and may hence partially label such data, difficulties in creating comprehensive naming conventions and precise definitions for behaviors make such labels too vague to support comparative performance evaluation. Current methods for human activity recognition may allow to automatically supplement such labels by providing systematic and hence comparable categories for behaviors. In addition, they may be used to automatically identify underlying motion primitives, further increasing their precision and potential for data mining.

Human-Robot Interaction (HRI)

Activity recognition supports human-robot interaction [8]. Here we take the viewpoint that human activities are infered from sensors worn by the user and broadcasted to surrounding robots. The typical applications are in the domain of assistive robotics. This form of activity recognition is the one typically researched in the wearable computing community.

Robot self-learning

Imagine that a human teaches a robot by demonstration [9] (see figure 1). Using activity recognition, the robot can supplement its known set of behaviors with an internal model to recognize different demonstrated activities [10], [11], and even build a repertoire of sub-goals and obtain a hierarchical decomposition of its actions [12]. The inferred model can provide feedback for a self-learning process [13]; for example using self-perception [14], reinforcement learning [15], or evolutionary techniques [16]. Such a self-learning process would allow the robot to develop its own realization of the motor-commands in light of the goal to reach. In addition, it may lead to increases in behavioral robustness. For instance, a damaged robot like NASA's famous "Spirit" rover may

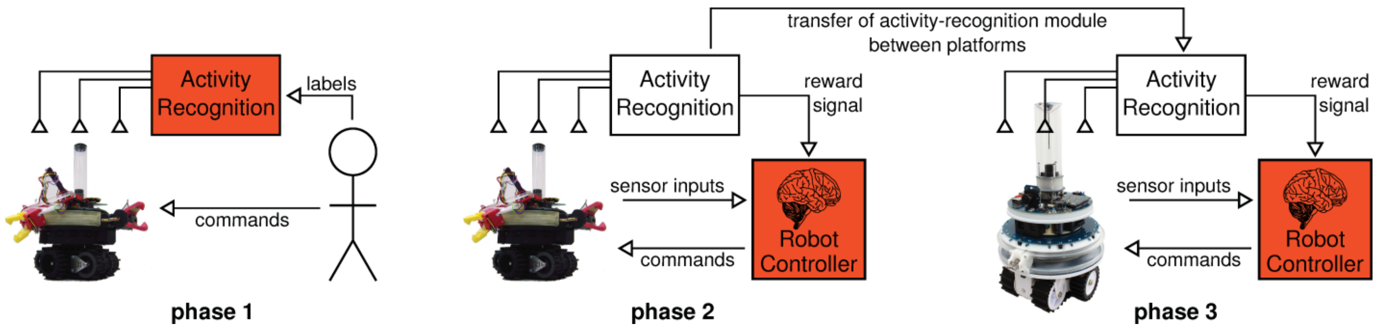

Fig. 1. Activity-recognition can be used to self-learn robot motor control. The colored elements are those that undergo learning/adaptation. In phase 1, an activity recognition system learns to recognize the activities demonstrated by the user using the robot's sensors. In phase 2, activity recognition guides the adaptation of the robot controller. This allows the robot to discover or improve its motor skills. In phase 3, the recognition system is transfered to a new robot that has a compatible perception subsystem but different actuators. The new robot self-learns the same motor skills as the previous robot.

not have been pre-programmed to cope with all modes of failures. However, should it be able to perceive its own actions, activity recognition could provide it with insight into its actual behavioral performance in spite of the damage. Using continuous self-learning, this may allow it to develop a novel, effective motor-control program, overcoming the problem.

Robot self-learning can also be applied across platforms. Rather than demonstrating all activities to all robots, activity recognition may allow robots to learn from each other. Given compatible perception systems, robots can directly share and reuse activity recognition systems across platforms. Differences such as parametrizations of motor control programs can then be self-learnt. While such direct sharing of activity recognition is only possible for identical perception systems, it is noteworthy that efforts in activity recognition which we review in this paper aim at developing methods that are robust across different perception systems [17]. In particular a reduced set of sensors may nevertheless allow to recognize a common subset of activities [18]. Since many of today's robots share common sensors like cameras or a laser scanners, this potentially allows reuse of activity recognition programs across platforms. Moreover, a robot capable of recognizing an activity using some of its sensors can learn the information content with respect to that activity in its other sensors [19]. This could be used to learn to use a new sensor. This could also lead to automatic calibration when replacing robot parts, and to transparent sensor substitution in case of failures. In the context of networked robots, activity recognition may help robots identify semantic relevance of their peer's activities, enhancing collaboration in heterogeneous robot teams.

Contribution

Activity recognition is a key principle underlying wearable computing. Body-worn sensor data is interpreted to infer the user's activities and realize activity-aware application [20], [21]. One realizes the direct parallel to robotics: the mechanical body is replaced by a human body, and wearable systems and autonomous mobile robots alike sense and interpret their environment from a first person perspective.

Activity recognition in wearable computing is challenging due to a high variability along multiple dimensions: human action-motor strategies are highly variable; the deployment of sensors at calibrated locations is challenging; the environments where systems are deployed are usually open-ended. This has led the wearable computing community to enhance existing, and investigate new, recognition methods that cope with such variability. This paper aims to present the principles developed for wearable activity recognition to the robotics community, emphasizing the issue of transfer and sharing of activity recognition systems between platforms.

We introduce wearable computing in section II. We give an overview of the approaches used for activity recognition in wearable computing along sensors and processing techniques in section III. In section IV we categorize and illustrate a few of the most relevant approaches recently proposed to share activity recognition systems across platforms and application domains. We conclude summarizing the key insights and ongoing research challenges, and indicating resources where further information about wearable activity recognition can be found in section V.

II. WHAT IS WEARABLE COMPUTING?

Wearable computing, as originally presented by Mann in 1996, emphasized a shift in computing paradigm [22]. Computers would no longer be machines separate from the persons using them. Instead, they would become an unobtrusive extension of our very bodies, providing us with additional ubiquitous sensing, feedback and computational capabilities. As implied by its name, wearable computing never considered implanting sensors or chips into the body. Rather, it emphasizes the view that clothing, which has becomes an extension of our natural skin, would be the substrate that technology could disappear into (figure 2). The prevalence of mobile phones now offers an additional vector for on-body sensing and computing [23].

Mann [24] and Starner [25] were among the first to show that complex contextual information can be obtained by interpreting on-body sensor data and that this would lead to novel adaptive applications. A wearable system can perceive activities, here defined to include both gestures and behaviors, from a "first person perspective". This leads to new forms of applications known as activity-based computing or interaction-based computing [20], [21]. Such applications can offer information or assistance proactively based on the user's situation,

Fig. 2. The vision of wearable computing in the mid 1990s (from [22]). The system comprises a head-worn camera, a see-through head-up display in the goggles, an Internet connection, and an on-body computer.

as well as support explicit interaction in unobtrusive ways through natural gestures or body movements.

A few application domains include e.g.: industrial assistance [26], gestural inputs for human-computer interaction [27], behavior monitoring for personalized healthcare [28], movement analysis for sports assistants [29].

The kind of activities or gestures that are recognized are wide ranging, but also depend on the available sensors. Examples include:

- the recognition of complex manipulative gestures performed by industrial workers on a car body to check its functioning [30], with gestures including checking the hood latch mechanism, checking the seat sliding mechanism, and checking the spacing between doors and car body by using sensors including inertial measurement units (see also figure 5);
- the recognition of seven modes of locomotion (sit, stand, walk, walk upstairs, walk downstairs, ride elevator up, ride elevator down) from a single acceleration sensor [31];
- the recognition of the assembly steps of a shelf or a mirror from acceleration sensors [32], and the recognition of nine wood-making activities (hammering, sawing, filing, drilling, sanding, grinding, screwing, using a vise, operating a drawer) [33], using accelerometers and microphones;
- the recognition of five hand gestures (square, cross, circle, fish, bend) for human-computer interaction (HCI) from a single accelerometer [27];
- the recognition of sports activities in a fitness room by intertial sensors [34].

Activity recognition in wearable computing shares a number of similarities to mobile robotics:

- sensing is performed on the human or robot body from a “first person” perspective;
- recognition of activities is directly relevant for the application at hand;
- continuous recognition of activities is essential to allow for adaptation of the system’s behavior;
- activities typically have a clear semantic description (e.g. reaching, grasping).

III. Wearable activity recognition

Activity and gesture recognition is generally tackled as a problem of *learning by demonstration* [35], [33]. The user is instrumented with the selected sensors and put into a situation where he performs the activities and gestures of interest. The sensor data are acquired with ground-truth annotations describing what the user performs or experiences. The resulting dataset is used to train the recognition system and test its performance. The training process consists of identifying the mapping between the user’s activities or gestures and the corresponding sensor signals.

Some terminology commonly used in wearable activity recognition differs from the one used in robotics:

- annotation or labelling: this is the process by which the experimenter manually provides ground-truth information about the activities of the subject, generally when collecting an activity dataset;
- recognition or spotting: this is the actual machine identification of an activitiy in the sensor stream. Activities are said to be “recognized” or “spotted”.

A. *Sensors for activity recognition*

Sensors are used to acquire signals related to the user’s activities or gestures. User comfort is paramount. Thus, sensors must be small, unobtrusive and ideally invisible to the outside. The sensors are selected according to a tradeoff between wearability, computational needs, power usage, communication requirements, and information content for the activities and contexts of interest. For instance cameras are currently seldomly used in wearable computing due to the computational requirements for video analysis. Instead sensor modalities that are computationally lighter are prefered.

Common sensor modalities are body-worn accelerometers and inertial measurement units (IMUs). Accelerometers are extremely small and low-power. IMUs contain accelerometers, magnetometers and gyroscopes and allow to sense the orientation of the device with respect to a reference. IMUs are typically placed on each body segment and allow to reconstruct a body model of the user. On-body microphones are also successfully used for activity recognition, as many human activities generate characteristic sounds (e.g. using a coffee machine, brushing teeth) [33]. Typical sensor modalities are listed in [36].

Clothing is a major platform to deploy sensors on-body unobtrusively. For instance, IMUs can be integrated in a worker’s jacket (see figure 3). There are also ongoing efforts to develop sensorized textile fibers, which allows for truly unobtrusive garment-integrated sensing [37].

Nowadays the trend goes towards an increased use of multiple multimodal sensors, as this tends to increase recognition performance (see e.g. [30]). Wearable systems are also complemented by object-integrated and ambient sensors. We recently coined the term *opportunistic activity recognition* in the EU FP7 project OPPORTUNITY [17]. It describes systems that make use of sensors that just happen to be available, rather than requiring specific sensor deployment. This will further address comfort issues and emphasizes the need for new machine learning techniques to share activity recognition systems across different sensor domains to reach this goal [17].

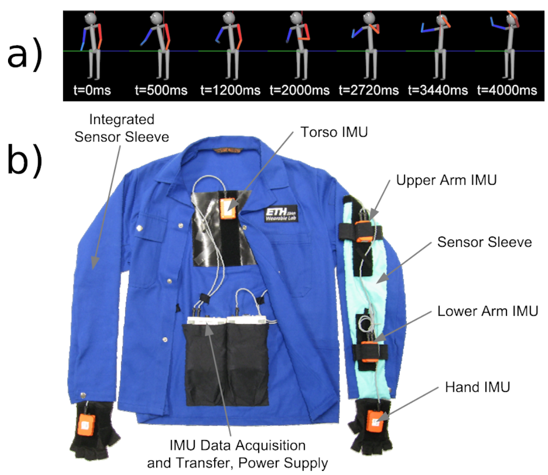

Fig. 3. This *MotionJacket* allows the unobtrusive capture of the upper-body movements using seven IMUs placed on each body segments. (b). The IMUs allow to reconstruct the user's instantaneous posture (a). From [30].

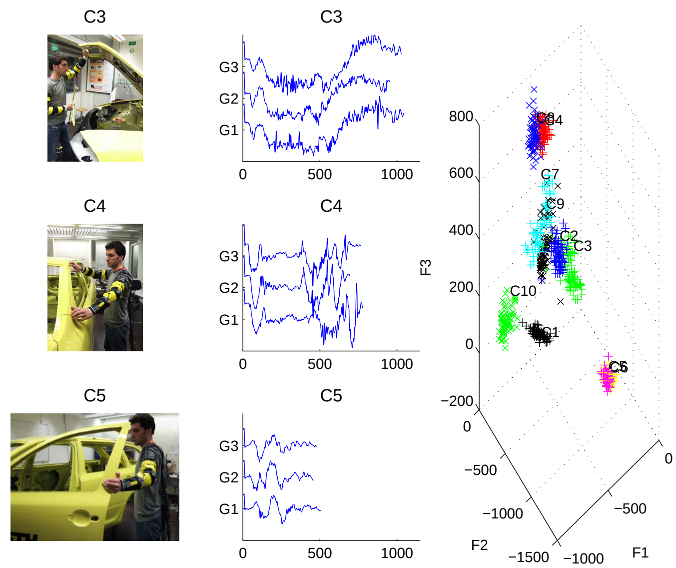

Fig. 5. Three activities of a car assembly scenario are shown: checking the engine hood (C3), checking the gap spacing between doors and car body (C4), and checking the opening of the front door (C5). The data of an acceleration sensor placed on the right wrist is shown for three repetitions (G1-G3) of the activity (center). Note the variability in the gesture execution length and signal shape. After feature extraction, the sensor signals are projected into a feature space (right). Note that some activities are well separated, lending them to robust classification (e.g. C10, C1) while others overlap as they are more similar (e.g. C5, C6, bottom right). During training of the recognition chain, the selection of pre-processing steps and features aims at increasing the separation between the activity classes. Sensor data from [18].

B. Activity recognition chain

We refer to the *activity recognition chain* (ARC) as a set of processing principles commonly followed by most researchers to infer human activities from the raw sensor data [33], [35], [38], [39] (see figure 4).

The *sub-symbolic* processing maps the low-level sensor data (e.g. body-limb acceleration) to semantically meaningful action primitives (e.g. grasp). Meaning is attributed to the sensor data streams by "comparing" them to known activity prototypes. This is realized by streaming signal processing and machine learning techniques. The outcome of the sub-symbolic processing are events indicating the occurrance of action primitives. The ARC terminates at this stage when the activities of interest consist of simple gestures, for instance used for gestural interfaces [27].

The *symbolic* processing maps sequences of action primitives (e.g. grasping, cutting) to higher-level activities (e.g. cooking). This may be realized by reasoning, expert knowledge, or by statistical approaches applied to the occurrences of action primitives.

Sub-symbolic processing ought to be robust to the large observed variability in sensor-signal to activity-class mapping due to human behaviors or sensor deployments. In wearable computing, sub-symbolic processing is usually co-optimized with sensor selection to maximize comfort and recognition performance. The sub-symbolic processing stages are usually [33], [35], [38] (see fig. 4):

- *sensor-data acquisition:* A stream of sensor samples $\boldsymbol{S}$ is obtained;
- *signal pre-processing:* The sensor data stream is pre-processed. Typical transformations are calibration, de-noising, or sensor level data fusion;
- *segmentation of the data stream.* The data stream is segmented into sections $\boldsymbol{W}$ that are likely to contain a gesture; Segments are identified by their start and end time in the data stream. A common type of segmentation technique is the sliding window, usually for periodic movements, or energy-based or rest-position based segmentation, when the user performs isolated gestures or returns to a rest position between gestures;
- *feature extraction:* Features are computed on the identified segments to reduce their dimensionality, yielding a feature vector $\boldsymbol{X}$;
- *classification:* A classifier, trained at design-time, maps the feature vector into a pre-defined set of output classes (activities, gestures): $\boldsymbol{X} \rightarrow c, p$. Usually a ranked likelihood of the output classes is obtained which can be used for decision fusion;
- *decision fusion:* Combines multiple information sources (multiple sensors, or multiple classifiers operating on one sensor) into a decision about the activity that occurred;
- *"null-class" rejection:* In cases where the confidence in the classification result is too low, the system may discard

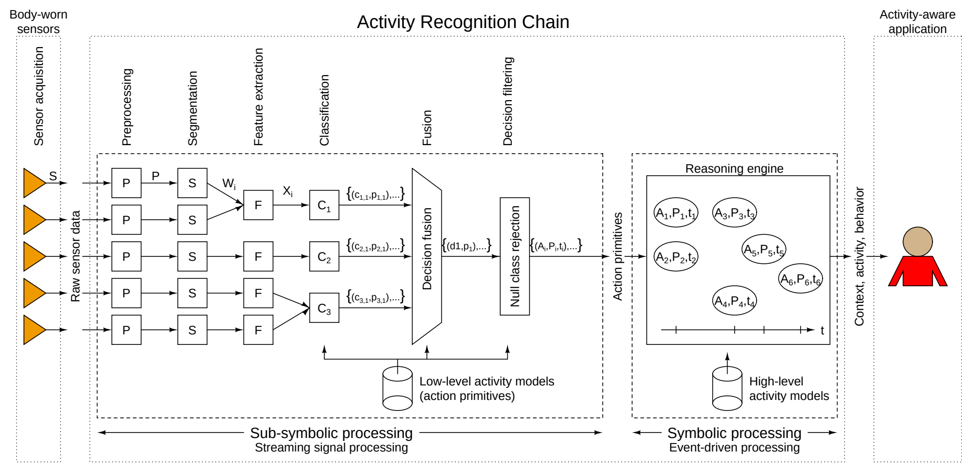

Fig. 4. Processing steps used to infer activities from on-body sensors. The raw sensor data is mapped to the occurrence of action primitive (events) with signal processing and machine learning techniques. Here 5 sensors deliver data. Data fusion is illustrated at the feature, classifier and decision level. Symbolic processing infers higher-level activities from the occurrence of action primitives usually with reasoning or statistical approaches.

the classified activity based on its likelihood. At this stage, the outcome is the detection of an action primitive A_i with likelihood p_i at time t_i.

Before operation, the classifiers used in the ARC are trained using a training set containing data instances (feature vectors) $\boldsymbol{X}$ and the corresponding activity label γ. Other parameters, such as the thresholds to segment activities or reject the null class, or the set of features, are also optimized prior to operation.

Classifiers commonly used for activity recognition have been reviewed in [38] together with typical features derived from acceleration signals. If the features corresponding to activities form clusters in the feature space (see figure 5), then the classifiers that are typically used include Support Vector Machines [40], decision trees, k-Nearest Neighbor or Naive Bayes classifiers [41]. This is usually the case with isolated gestures and when static postures are recognized with features such as limb angles. It is also the case with periodic activities when frequency domain features are used (e.g. walking leads to energy in specific frequency bands). When the temporal unfolding of the gesture must be analyzed, such as with sporadic gestures, approaches such as Dynamic Time Warping [42] or hidden Markov models (HMMs) [25] are used. Other methods include neural networks [43] or fuzzy systems [44].

In figure 5 we illustrate a set of activities and the corresponding sensor signals. With simple statistical features, the sensor signals can be projected in a feature space where the activities form clusters suitable for classification.

Symbolic level processing is usually event-driven, with events corresponding to activity occurrences. Higher-level activity models are thus built on event occurrences instead of raw sensor data. Approaches typically used for symbolic processing include ontological and statistical reasoning, probabilistic and temporal logic, Bayesian networks, fuzzy logic, Dempster-Shafer and hybrid approaches [45], [46], [47]. Modeling and reasoning methods used for human context inference are further reviewed in [39].

High-level models are usually also derived from data recordings. Alternative approaches include the use of expert knowledge. For instance in [30] we relied on a documented step-by-step guide for industry workers to detect a high-level task - the assembly of a car lamp - from a sequence of action primitives.

Few work has attempted to use expert knowledge to detect complex gestures from raw sensor data, such as accelerometer readings [48]. The main challenge faced is the large inter- and intra-user variability which is better captured by learning by demonstration approaches.

IV. Sharing activity-recognition chains

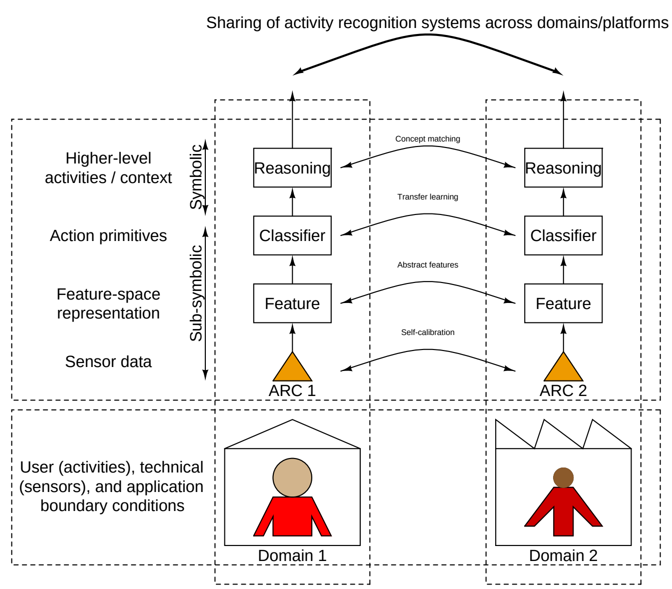

Fig. 6. Representation of the level at which a common representation is assumed to share a recognition system between users (platforms) or domains.

Human activity recognition in wearable computing is challenging due to a large variability in the mapping of sensor signals to activity classes. This variability has multiple origins:

- semantically identical action primitives (e.g. drinking from a glass) can be executed in a large number of ways (e.g. grasp with the left or right hand, while seated, standing or walking, at various speeds). This is referred to as *intra-user variability*. These variations come from personal preferences. Moreover, aging, injuries, or increased proficiency at a task also lead to variability. Figure 5 illustrates intra-user variability;

- although different persons may be considered as "robots of identical make" in practice there is an even higher variability in action-motor strategies between users (*inter-user variability*) than for a single user. Personal preferences, differences in expertise, body proportions or fitness level explain this variability;
- the placement of the sensors on body cannot be done with a high precision, especially when the users deploy the sensors themselves. For comfort reasons the user must be able to detach sensors when not needed (e.g. during sleep) and reattach them when needed, or to displace them when incomfortable. The placement of sensors in loose-fitting clothing is affected by the deformation of the garment depending on the user's activities and posture [49]. Figure 7 illustrates the effect of sensor placement on the projection of sensor data into the feature space;
- abstracting the specific environment in which the system can recognize activities is important to ensure cost-effective deployment on a large scale. Thus, activity recognition methods should work for a generic class of problems (e.g. in "any smart home") rather than a specific instance of the problem class (e.g. a specific smart home);
- To further increase unobtrusiveness, we argue in the EU FP7 FET-Open project OPPORTUNITY[1] to use *opportunistically discovered sensors* for activity recognition [17]. Thus, the available sensor configuration depends on the sensorized objects users take with themselves, on the smart-clothing they wear, and on the environment in which they are located. For each sensor kind and placement there is a different sensor-signal to activity-class mapping that an opportunistic activity recognition system should be able to abstract.

The wearable computing community has developed best practices and novel methods to deal with some forms of variability. In the following subsections we present a selection of methods developed by various groups and ours. In order to share an ARC, there must be a common representation at some stage in the recognition chain. We organize the methods along the level at which methods assume the common representation. We describe methods operating at the sensor-level, at the feature level, at the classifier-level, and at the reasoning level (see figure 6).

A. Sensor-level sharing

This level focses on training an ARC on the first platform and re-using it on the second platform. This assumes that the sensor signal to activity class mappings are statistically identical on the two platforms. This is usually not the case in practice due to slight variations in sensor placement and in human action-motor strategies. Training an ARC on one system is referred to as a *user-specific* system, and it is known to show degraded performance when deployed to another user [33]. Training user-specific ARCs is costly and thus not adequate for the deployment of wearable system on a large scale.

The best practice to realize an ARC that generalizes to new situations consists in training it on a dataset containing the variability likely to be seen when the system is deployed. By collecting a dataset from multiple user the ARC can be trained to be *user-independent* [33]. By collecting a dataset comprising multiple on-body sensor positions the ARC can be trained to be *sensor-placement independent* [31].

A similar approach in learning by demonstration in robotics could lead to platform independent activity recognition models by demonstrating a task to multiple platforms.

The previous approach requires to foresee all the variations likely to be encountered at run-time. Thus, we proposed an unsupervised self-calibration approach that removes this requirement [50]. The self-calibration approach operates as follows:

- the ARC continuously operates and recognizes the occurrence of activities/gestures;
- upon detection of an activity/gesture, the corresponding sensor data is stored as training data;
- the classifiers are re-trained including this new training data, using an incremental learning algorithm.

Thus, the activity models are optimized upon each activity instance to better model that activity. We demonstrated the benefits of this approach on the recognition of 6 fitness activities (see figure 7) when the position of sensors on the body are displaced between the training and testing phases. The figure illustrates the sensor placement, and the mapping of the activiy classes in the feature space. During adaptation, the method tracks the displacement of the activity clusters in the feature space. The assumptions underlying the approach are that activities form distinct clusters in the feature space, and that the speed of adaptation is matched to the speed at which clusters shift. In [50] we argue that this approach may also be applied to cope with slight changes in action-motor strategies, due e.g. to ageing or change of user.

A translation to robotics of these principles may allow activity models to adapt when sensors or actuators deteriorate.

B. Feature-level sharing

At this level, the ARC devised for the first platform is translated to the second platform from the feature level onwards. Thus, the ARC must abstract from the specific sensors. The use case for sharing ARCs at this level include systems where the sensor modalities on the two platforms do not coincide, or show large on-body displacement, for which a placement-independent ARC cannot be envisonned.

Kunze et al. have explored in [34], [51] approaches to elevate the processing of the activity recognition chain to "abstract" features. They show that features that are robust to on-body displacement can be designed using body models and fusing multiple modalities, such as accelerometer and

[1] http://www.opportunity-project.org

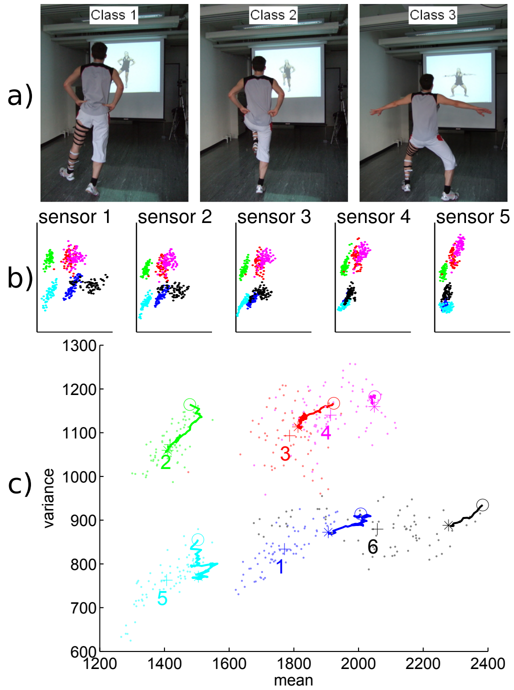

Fig. 7. **a)** Three of the 6 fitness activities performed by the subject to assess unsupervised classifier self calibration (flick kicks, knee lifts, and jumping jacks are depicted). Ten on-body sensors at regular interval are visible on the subject's left leg. **b)** Distribution of the 6 fitness activity classes in the feature space. **c)** Adaptation dynamics of the nearest class center classifier trained on one sensor position and deployed on another position.

gyroscope modalities [34]. They also show that a specific sensor modality (magnetic field sensor) can be replaced by another specific modality (gyroscope) [51].

In [17] we argue that other such transformation may be feasible, such as performing activity recognition on a 3D body model (see figure 3 (a)), which can be obtained from modalities such as IMUs, fitting a body model in video sequences, or inferring limb angles using clothing integrated elongation sensors [52].

A hybrid approach between sensor-level and feature level sharing was further proposed by Kunze et al. who demonstrated that sensors can autonomously self-characterize their on-body placement [53] and orientation [54] using machine learning techniques. They propose to use on-body sensor placement self-characterization as a way to select, among a number of pre-programmed ARCs, the one most suited to the detected sensor placement.

Similarly in robotics data from different sensors can be converted into identical abstract representations. For instance 3D point clouds can be measured by stereovision or laser range finder.

C. Classifier-level sharing

Transfer learning allows to translate a classification problem from one feature space to another [55] and was used to transfer perceptual categories across modalities in biological and artificial systems [56]. Conceptually, transfer learning may thus be used to translate the capability to recognize activities from one platform to another without enforcing a similar input space (i.e. sensors, features). Thus, the transfer does not affect higher-level reasoning.

Practical principles allowing a system A to confer activity recognition capabilities to another system B are outlined in [19]. Each system A and B is composed of a set of sensors S_A, S_B, ARCs ARC_A, ARC_B, and a unified communication protocol. The process of transfer learning works as follows (see figure 8):

- the user employs an activity-aware system A with ARC_A and sensor set S_A. For instance: a set of instrumented drawers capable of reporting which one is being opened or closed, in a storage management scenario;
- a new system is deployed in the user's personal area network comprising a set of unknown new sensors S_B (on body and/or in the user's surroundings) and a yet untrained ARC_B. For instance, the user wears a new sensorized wristband with integrated acceleration sensor;
- as the user performs activities, the ARC_A recognizes them and broadcasts this information;
- the new system B receives the class labels of the recognized activities. The ARC_B incrementally learns the mapping between the signals of the sensor set S_B and the activity classes;
- eventually, the system A can be removed. The activity recognition capability is now entirely provided by the system B.

The underlying assumptions are that the two systems coexist for long enough to operate the transfer learning. In figure 8 we show that, as the user interacts with the set of drawers, the body-worn system incrementally learns to recognize drawer activities.

In robotics, this sharing approach may be used to allow robots with different sensory inputs to learn to recognize semantically identical activities, or to learn how to use a new sensor when robot parts are upgraded, thus easing programming.

D. Symbolic-level sharing

The reasoning program to infer higher-level activities from spotted action primitives is shared between platforms. As the environment in which the two platforms operate may lead to the detection of semantically different action primitives, a direct transfer of the reasoning is not always possible. Carrying out a prior concept matching can address this.

For instance, to reason about the activity of a user, one needs first to know in which room he is located. One environment may have a sensor allowing to detect the action primitive "room door activated". Another environment may have a proximity infrared sensor allowing to detect "movement in the room". The interpretation of the sensor data requires different features and classifiers in each case. However, although the classifiers deliver semantically different action primitives, they may both be found to indicate the presence of a user in a room. Thus, higher-level reasoning may remain identical if these two different concepts are first matched.

Van Kasteren et al. extended transfer learning methods to operate on time series resulting from the activation of

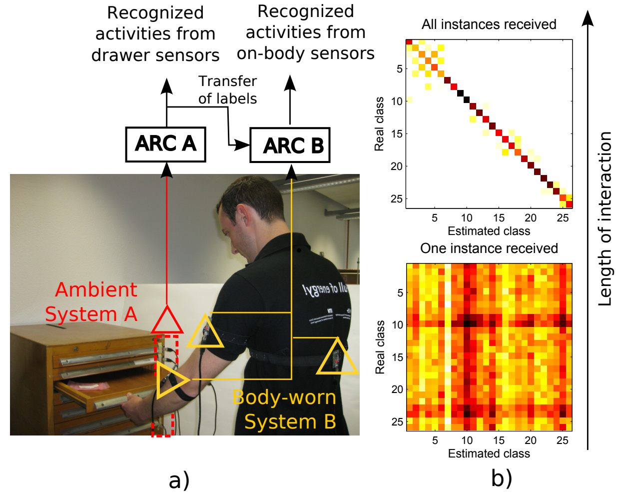

Fig. 8. **a)** An ambient system A consists of 13 drawers equipped with acceleration sensors and an ARC capable of recognizing which drawer is being opened or closed. A wearable system B consists of three on-body acceleration sensors. Its ARC is initially untrained. As the user interacts with the drawers, the system A provides activity labels to the system B, that incrementally learns to recognize opening and closing gestures. **b)** The activity recognition confusion matrix of system B before and after transfer learning. It indicates for each user activity (rows, here opening or closing a drawer) how the ARC classifies the activity. The closer to a diagonal distribution, the higher the recognition accuracy.

simple binary sensors [57]. They applied this method to transfer behavior recognition capabilities from one smart home kind to another smart home kind with different and a-priori unknown number and placement of sensors. The system first automatically finds how sensor activations in different environments relate to identical higher-level concepts using statistical approaches. A recognition system can also learn internal hierarchical representation of activities or "concepts" [58], upon which reasoning is performed. Advances in merging concepts in ontologies [59], [60] supports the transfer of activity recognition reasoning across different conceptual spaces.

In robotics these principles may allow robots to exchange the knowledge they have individually gained about the world. This may be especially relevant when principles of autonomous mental development are used, as robots can develop distinct world representations according to their capabilities.

E. Other approaches

Some approaches do not fit in the taxonomy above. Schiele et al. proposed a form of transfer learning to reuse action primitives across different but related application domains. Action primitive spotting (hammering, screwing, cutting) was trained on the dataset of a shelf-assembly task. These primitives were re-used as-is to detect higher-level steps of a mirror-assembly task, thus reducing considerably the amount of training data needed for the new task [32]. Further result support this approach [61].

Most of the approaches described previously attempt to reduce or eliminate the need for training data for activity recognition on the new platform. Beigl et al. proposed to "crowd-source" the acquisition of training data. They addressed the issues related to shared data labeling by developing a framework suitable for end-users operating on a mobile phone [62]. Semi-supervised learning allows to combine a limited number of labeled data with a large amount of unlabeled data to train classifiers. It was successfully used to train activity recognition systems using only sparse activity labels [63]. Recent trends seek further reduction in the data collection efforts by automatically generating activity recognition models from on-line sources by data mining [64].

Calatroni et al. argue that many existing sensors can be repurposed for activity recognition, even though they were initially deployed for other uses [65]. They show, for instance, how reed switches placed in windows for security purposes can be used to infer standing or walking, by means of assumptions about human behavior when interacting with the instrumented object. They indicate several other sensors and behavioral assumptions that allow to obtain sporadic information or "labels" about the modes of locomotion of the user or his gestures. They suggest to incrementally train the body-worn recognition system, whenever such labels are obtained, with the transfer learning method described above. Eventually the wearable system becomes capable of activity recognition even when the user does not interact with an instrumented object. Since this process can be continuous, the system can perform activity recognition with many different and unforeseen combinations of on-body sensors, as long as they provide discriminative signals.

V. Conclusion and Outlook

Activity recognition enables a WWW of robots by providing a tool to label large robot-generated activity datasets, by enabling activity-aware human-robot interaction (HRI) in hybrid teams, and by opening the way to self-learning autonomous robots capable of monitoring their own proficiency at a task.

Human activity recognition has been a major object of research in the wearable computing community since the mid-nineties. We summarized the methods developed in the community along the *activity recognition chain* - a set of processing principles followed in most wearable activity recognition research. Since human activities are highly variable we reported some of the recent advances to enhance the robustness of activity recognition systems when they are shared among different users or deployed in different application domains. Many methods can potentially be beneficial to robotic use cases.

Human activity recognition from on-body sensors is far from a solved problem. Some of the continuing challenges include:

- finding more efficient sensor modalities for activity recognition. They should satisfy multiple requirements: minimize obtrusiveness, be highly discriminative of the activities of interest, and minimize subsequent computational complexity;
- spotting rare events and short activities in large stream of data. This is still a challenging segmentation and null-class rejection problem;
- despite recent advances surveyed in this paper, coping with human motion variability remains an open area of research;

- deploying activity recognition to new problem domains without an expensive training phase is still elusive;
- shared reference activity recognition dataset are important for benchmarking purposes. We reviewed a few activity recognition dataset and proposed a new benchmark dataset in [36];
- building and updating the state of a world-model according to the user's actions. For instance when a user displaces a cup, this changes the meaning of a "grasp" gesture performed at the prior location of the cup. Most current approaches assume stateless world models;

Other challenges relate to the use of activity recognition systems in robotics. The annotation of large-scale datasets or the recognition of human activities for HRI must take into account that machine recognition of activities is not perfectly accurate. Thus, probability distributions on the recognized activity classes needs to be taken into account for further processing, for instance in a Bayesian framework.

Using activity recognition in a robotic self-learning paradigm builds on the assumption that it is preferable to translate an activity recognition system between robots, rather than a motor program. Translation between robots of identical make is relatively straightforward and may allow robots to learn new motor strategies when actuators are damaged. Translation across heterogeneous platforms assumes a greater invariance in the activity recognition system than in the motor program. The methods presented in this paper show how to approach this issue. The coming years will see whether self-learning in heterogeneous platforms driven by a common activity recognition system can be reliably achieved.

We invite the interested readers to look for further information on activity recognition in wearable and pervasive computing in the following conference proceedings: Int. Symp. on Wearable Computers (ISWC), Int. Conf. on Pervasive Computing (Pervasive), Int. Conf. on Ubiquitous Computing. The following journals also cover the topic: IEEE Pervasive Computing Magazine, Personal and Ubiquitous Computing (Springer), and Pervasive and Mobile Computing (Elsevier).

ACKNOWLEDGMENT

We acknowledge the financial support of the Seventh Framework Programme for Research of the European Commission, under the project OPPORTUNITY with grant number 225938, and RoboEarth with grant number 248942.

REFERENCES

[1] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[2] J. Geusebroek, G. Burghouts, and A. Smeulders, "The Amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005.

[3] A. Torralba, R. Fergus, and W. Freeman, "80 million tiny images: a large data set for nonparametric object and scene recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1958–70, 2008.

[4] J. Fenwick, P. Newman, and J. Leonard, "Cooperative concurrent mapping and localization," in *Robotics and Automation, 2002. Proceedings. ICRA 2002. IEEE International Conference on*, vol. 2, 2002, pp. 1810 – 1817.

[5] R. D'Andrea and P. Wurman, "Future challenges of coordinating hundreds of autonomous vehicles in distribution facilities," in *International Conference on Technologies for Practical Robot Applications*. IEEE, 2008, pp. 80–83.

[6] M. Roth, D. Vail, and M. Veloso, "A real-time world model for multi-robot teams with high-latency communication," in *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 3. IEEE, 2003, pp. 2494–2499.

[7] L. Parker, K. Fregene, Y. Guo, and R. Madhavan, "Distributed heterogeneous sensing for outdoor multi-robot localization, mapping, and path planning," in *Multi-robot systems: From swarms to intelligent automata: Proceedings from the 2002 NRL workshop on multi-robot systems*. Kluwer Academic Pub, 2002, p. 21.

[8] C. Martin, F.-F. Steege, and H.-M. Gross, "Estimation of pointing poses for visually instructing mobile robots under real world conditions," *Robot. Auton. Syst.*, vol. 58, pp. 174–185, February 2010. [Online]. Available: http://dx.doi.org/10.1016/j.robot.2009.09.013

[9] S. Calinon, F. D'halluin, E. L. Sauser, D. G. Caldwell, and A. G. Billard, "Learning and reproduction of gestures by imitation: An approach based on hidden Markov model and Gaussian mixture regression," *IEEE Robotics and Automation Magazine*, vol. 17, no. 2, pp. 44–54, 2010.

[10] F. Bertsch and V. Hafner, "Gesture recognition as a prerequisite of imitation learning in human-humanoid experiments," in *Proc. of 9th Int. Conf. on Epigenetic Robotics*, 2009, pp. 197–198.

[11] V. Hafner and F. Bachmann, "Human-humanoid walking gait recognition," in *Proc. 8th IEEE-RAS International Conference on Humanoid Robots*, 2008, pp. 598–602.

[12] T. Duong, D. Phung, H. Bui, and S. Venkatesh, "Efficient duration and hierarchical modeling for human activity recognition," *Artificial Intelligence*, vol. 173, no. 7–8, pp. 830–856, 2009.

[13] J. Tani, R. Nishimoto, J. Namikawa, and M. Ito, "Codevelopmental learning between human and humanoid robot using a dynamic neural-network model," *IEEE Trans. on Systems, Man, and Cybernetics Part B: Cybernetics*, vol. 38, no. 1, pp. 43–59, 2008.

[14] J. Sturm, C. Plagemann, and W. Burgard, "Unsupervised body scheme learning through self-perception," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 3328–3333.

[15] P. Varshavskaya, L. P. Kaelbling, and D. Rus, "Automated design of adaptive controllers for modular robots using reinforcement learning," *The International Journal of Robotics Research*, vol. 27, no. 3–4, pp. 505–526, 2008.

[16] I. Harvey, E. Di Paolo, R. Wood, M. Quinn, and E. Tuci, "Evolutionary robotics: A new scientific tool for studying cognition," *Artificial Life*, vol. 11, no. 1–2, pp. 79–98, 2005.

[17] D. Roggen, K. Förster, A. Calatroni, T. Holleczek, Y. Fang, G. Tröster, P. Lukowicz, G. Pirkl, D. Bannach, K. Kunze, A. Ferscha, C. Holzmann, A. Riener, R. Chavarriaga, and J. del R. Millán, "Opportunity: Towards opportunistic activity and context recognition systems," in *Proc. 3rd IEEE WoWMoM Workshop on Autononomic and Opportunistic Communications*, 2009.

[18] P. Zappi, C. Lombriser, E. Farella, D. Roggen, L. Benini, and G. Tröster, "Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection," in *5th European Conf. on Wireless Sensor Networks (EWSN 2008)*, R. Verdone, Ed. Springer, 2008, pp. 17–33.

[19] A. Calatroni, C. Villalonga, D. Roggen, and G. Tröster, "Context cells: Towards lifelong learning in activity recognition systems," in *Proceedings of the 4th European Conference on Smart Sensing and Context (EuroSSC)*. Springer, 2009.

[20] N. Davies, D. P. Siewiorek, and R. Sukthankar, "Special issue: Activity-based computing," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 20–21, 2008.

[21] P. Lukowicz, O. Amft, D. Roggen, and J. Cheng, "On-body sensing: From gesture-based input to activity-driven interaction," *IEEE Computer*, vol. 43, no. 10, pp. 92–96, 2010.

[22] S. Mann, "Smart clothing: The shift to wearable computing," *Communications of the ACM*, no. 8, pp. 23–34, 1996.

[23] R. Want, "When cell phones become computers," *IEEE Pervasive Computing*, vol. 8, no. 2, pp. 2–5, 2009.

[24] S. Mann, "Humanistic computing: "wearcom" as a new framework and application for intelligent signal processing," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2123–2151, 1998.

[25] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371–1375, 1998.

[26] V. Stanford, "Wearable computing goes live in industry," *IEEE Pervasive Computing Magazine*, vol. 1, no. 4, pp. 14–19, 2002.

[27] S. Kallio, J. Kela, P. Korpipää, and J. Mäntyjärvi, "User independent gesture interaction for small handheld devices," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 20, no. 4, pp. 505–524, 2006.

[28] A. Pentland, "Healthwear: Medical technology becomes wearable," *Computer*, vol. 37, no. 5, pp. 42–49, 2004.

[29] E. H. Chi, J. Song, and G. Corbin, ""killer app" of wearable computing: wireless force sensing body protectors for martial arts," in *Proc. of the 17th Annual ACM Conference on User Interface Software and Technology*. New-York: ACM Press, 2004, pp. 277–285.

[30] 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.

[31] J. Lester, T. Choudhury, and G. Borriello, "A practical approach to recognizing physical activities," in *Lecture Notes in Computer Science : Pervasive Computing*, 2006, pp. 1–16.

[32] U. Blanke and B. Schiele, "Remember and transfer what you have learned - recognizing composite activities based on activity spotting," in *Proc 14th IEEE Int Symp on Wearable Computers (ISWC'10)*, Seoul, Korea, 10/2010 2010.

[33] J. Ward, P. Lukowicz, G. Tröster, and T. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1553–1567, 2006.

[34] K. Kunze and P. Lukowicz, "Dealing with sensor displacement in motion-based onbody activity recognition systems," *Proc. 10th Int. Conf. on Ubiquitous computing*, Sep 2008.

[35] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing: Proc. of the 2nd Int'l Conference*, Apr. 2004, pp. 1–17.

[36] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, M. Creatura, and J. del R. Millán, "Collecting complex activity data sets in highly rich networked sensor environments," in *7th Int. Conf. on Networked Sensing Systems*. IEEE Press, 2010, pp. 233–240.

[37] A. Tognetti, N. Carbonaro, G. Zupone, and D. De Rossi, "Characterization of a novel data glove based on textile integrated sensors," *28th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, 2006 (EMBS '06).*, pp. 2510–2513, Aug. 30-Sept. 3 2006.

[38] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso, "Preprocessing techniques for context recognition from accelerometer data," *Pervasive and Mobile Computing*, vol. 14, no. 7, pp. 645–662, 2010.

[39] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, and D. Riboni, "A survey of context modelling and reasoning techniques," *Pervasive and Mobile Computing*, vol. 6, no. 2, pp. 161–180, 2010.

[40] H. Qian, Y. Mao, W. Xiang, and Z. Wang, "Recognition of human activities using svm multi-class classifier," *Pattern Recognition Letters*, vol. 31, no. 2, pp. 100–111, 2010.

[41] C. Randell and H. Muller, "Context awareness by analysing accelerometer data," in *ISWC 2000: Proc. of the 4th Int'l Symposium on Wearable Computers*, Oct. 2000, pp. 175–176.

[42] M. Ko, G. West, S. Venkatesh, and M. Kumar, "Online context recognition in multisensor systems using dynamic time warping," in *Proc. Conf. Intelligent Sensors, Sensor Networks and Information Processing Conference*, 2005, pp. 283–288.

[43] J.-Y. Yang, J.-S. Wang, and C. Y.-P., "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers," *Pattern Recognition Letters*, vol. 29, no. 16, pp. 2213–2220, 2008.

[44] G. Bailador and G. Trivinño, "Pattern recognition using temporal fuzzy automata," *Fuzzy Sets and Systems*, vol. 161, pp. 37–55, 2010.

[45] D. Riboni and C. Bettini, "Context-aware activity recognition through a combination of ontological and statistical reasoning," in *Ubiquitous Intelligence and Computing*, ser. Lecture Notes in Computer Science, D. Zhang, M. Portmann, A.-H. Tan, and J. Indulska, Eds. Springer, 2009, vol. 5585, pp. 39–53.

[46] A. Ranganathan, J. Al-Muhtadi, and R. H. Campbell, "Reasoning about uncertain contexts in pervasive computing environments," *Pervasive Computing, IEEE*, vol. 3, no. 2, pp. 62–70, 2004.

[47] S. McKeever, J. Ye, L. Coyle, and S. Dobson, "Using dempster-shafer theory of evidence for situation inference," in *Proceedings of the 4th European conference on Smart sensing and context*, ser. EuroSSC'09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 149–162.

[48] A. Benbasat and J. Paradiso, "An inertial measurement framework for gesture recognition and applications," in *Gesture and Sign Language in Human-Computer Interaction*, ser. Lecture Notes in Computer Science, I. Wachsmuth and T. Sowa, Eds. Springer, 2002, vol. 2298, pp. 77–90.

[49] H. Harms, O. Amft, and G. Tröster, "Modeling and simulation of sensor orientation errors in garments," in *Bodynets 2009: Proceedings of the 4th International Conference on Body Area Networks*. ACM, 2009, pp. 1–8.

[50] K. Förster, D. Roggen, and G. Tröster, "Unsupervised classifier self-calibration through repeated context occurences: Is there robustness against sensor displacement to gain?" in *Proc. 13th IEEE Int. Symposium on Wearable Computers (ISWC)*, 2009, pp. 77–84.

[51] K. Kunze, G. Bahle, P. Lukowicz, and P. K., "Can magnetic field sensors replace gyroscopes in wearable sensing applications?" in *Proc. Int. Symp. on Wearable Computers*. IEEE Press, 2010, pp. 1–4.

[52] C. Mattmann, F. Clemens, and G. Tröster, "Sensor for measuring strain in textile," *Sensors (MDPI)*, vol. 8, no. 6, pp. 3719–3732, 0 2008.

[53] K. Kunze, P. Lukowicz, H. Junker, and G. Troester, "Where am i: Recognizing on-body positions of wearable sensors," *LOCA'04: International Workshop on Location and Context-Awareness*, Jan 2005.

[54] K. Kunze, P. Lukowicz, K. Partridge, and B. Begole, "Which way am i facing: Inferring horizontal device orientation from an accelerometer signal," in *Proc. of Int. Symp. on Wearable Computers (ISWC)*. IEEE Press, 2009, pp. 149–150.

[55] W. Dai, Y. Chen, G. Xue, Q. Yang, and Y. Yu, "Translated learning: Transfer learning across different feature spaces," in *Proc. NIPS*, 2008, pp. 353–360.

[56] F. Orabona, B. Caputo, A. Fillbrandt, and O. F.W., "A theoretical framework for transfer of knowledge across modalities in artificial and biological systems," in *Proc. 8th Int. Conf. on Development and Learning (ICDL)*, 2009, pp. 1–7.

[57] T. van Kasteren, E. G., and B. Kröse, "Transferring knowledge of activity recognition across sensor networks," in *Proc. 8th Int. Conf on Pervasive Computing*, 2010, pp. 283–300.

[58] T. Huynh, M. Fritz, and B. Schiele, "Discovery of activity patterns using topic models," in *Proceedings of the 10th international conference on Ubiquitous computing*. ACM New York, NY, USA, 2008, pp. 10–19.

[59] D. Corbett, "Comparing and merging ontologies: A concept type hierarchy approach," in *14th Int. Symp. on Methodologies for Intelligent Systems*, 2003, pp. 75–82.

[60] N. F. Noy, "Semantic integration: a survey of ontology-based approaches," *SIGMOD Rec.*, vol. 33, pp. 65–70, December 2004.

[61] V. W. Zheng, D. H. Hu, , and Q. Yang, "Cross-domain activity recognition," in *Proc. 11th Int. Conf. on Ubiquitous Computing (Ubicomp)*, 2009, pp. 61–70.

[62] M. Berchtold, M. Budde, D. Gordon, H. Schmidtke, and M. Beigl, "Actiserv: Activity recognition service for mobile phones," in *Proc. 14th Int. Symp. on Wearable Computers (ISWC)*, 2010.

[63] M. Stikic, D. Larlus, and B. Schiele, "Multi-graph based semi-supervised learning for activity recognition," in *Proceedings of the 13th IEEE International Symposium on Wearable Computers (ISWC'09)*, Linz, Austria, 09/2009 2009, best paper award.

[64] M. Perkowitz, M. Philipose, K. Fishkin, and D. J. Patterson, "Mining models of human activities from the web," in *Proceedings of the 13th international conference on World Wide Web*, ser. WWW '04. New York, NY, USA: ACM, 2004, pp. 573–582. [Online]. Available: http://doi.acm.org/10.1145/988672.988750

[65] A. Calatroni, D. Roggen, and G. Tröster, "A methodology to use unknown new sensors for activity recognition by leveraging sporadic interactions with primitive sensors and behavioral assumptions," in *Proc. of the Opportunistic Ubiquitous Systems Workshop, part of 12th ACM Int. Conf. on Ubiquitous Computing*, 2010, doi:10.3929/ethz-a-006286309.