# Classification of Physical Interactions between Two Subjects

Ruzena Bajcsy, Alessandro Borri, Maria Domenica Di Benedetto, Annarita Giani, Claire Tomlin*

## Abstract

*This work addresses the problem of detecting and classifying the interaction between two subjects. While the use of Wireless Sensor Networks (WSN) for human action detection has been widely investigated, the use of this technology for detecting actions that involve multiple subjects has not been explored yet. We explain what we mean for interaction between subjects and we also give the motivation behind this research. Our approach is then presented wherein we perform classification in a particular case study.*

## 1. Introduction

The growth of WSNs and advances in camera technologies and image measurements have encouraged many researchers to focus on the study of human actions. Advanced methodologies for the classification of actions in various scenarios have been developed. Challenging situations such as moving backgrounds, non-stationary cameras, change in light and viewpoint, have been widely investigated [5] and applications have been developed involving video retrieval, surveillance, human-computer interaction and health-related situations.

The University of California at Berkeley and, in particular, the Department of Electrical Engineering and Computer Sciences have experience in distributed human action recognition via wearable motion sensor networks [6]. In fact they developed a novel platform called DexterNet [2] to support real-time human monitoring in both indoor and outdoor environments. In this system, five sensors (3 accelerometers and 2 gyroscopes) are embedded in a node. Users wear 5 nodes (on the waist, ankles and arms) that communicate wirelessly to a local device connected to a stationary wireless receiver. Also heart rate and heart rate variability are measured through four small electrodes attached to the side of the ribcage and are connected to a biosensor. The system combines a Body Sensor Layer (BSL), a Personal Network Layer (PNL) and a Global Network Layer (GNL). A mobile station that supports Linux OS and the IEEE 802.15.4 protocol receives and processes sensor data that are measured by the body sensors.

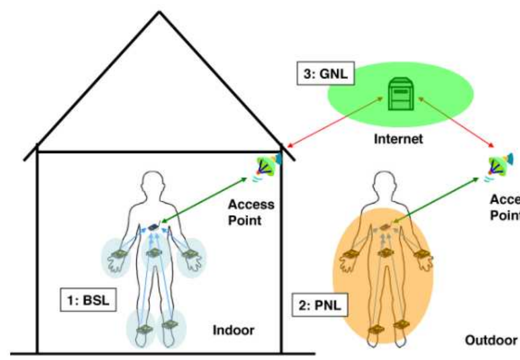


**Figure 1.** The Three-Layer DexterNet Architecture.

DexterNet is versatile and supports many applications. Among the applications supported are: human action/activity recognition, avatar visualization, integration with global positioning sensors and air pollution detection for asthma studies and monitoring of human motions in correlation with the energy expenditure. Within the action recognition application, a Wearable Action Recognition Database (WARD) was developed and made public [1]. Bending, laying down, standing up, walking, running and jumping are some of the actions that are stored in the database.

We have not found any literature addressing the problem of classifying the physical interaction between subjects. DexterNet provides all the data needed for studying the interaction between two subjects. Two people wear body sensors and perform actions and movements while being physically connected. Being physically connected means that the two subjects touch each other or are connected through a rigid object. Imagine the subjects walking and holding hands or holding the top and the bottom of an umbrella. In both cases there is a two way effect. In fact the two subjects influence each other's action, i.e. the action of one subject

*Ruzena Bajcsy, Annarita Giani and Claire Tomlin are from the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA, USA.

Alessandro Borri and Maria Domenica di Benedetto are from the Department of Electrical and Information Engineering, University of L'Aquila, Italy.

produces an effect on the way the other subject moves.

This project extends previous work on body sensor networks, but also presents new interesting challenges like the modeling of two nonlinear and dynamical very complex systems. We see great potential for theoretical research as well as interesting and useful applications.

### 1.1 Motivations

Our research focuses on measuring and interpreting data from two people that physically interact. The motivation behind this research is driven both by application and theoretical challenges.

There is a delicate balance between helpful and abusive interaction in both directions (care giver versus care receiver and vice versa). For example in health care it is very important to disambiguate between good practice and malpractice. Medical professional have liability insurance that financially covers them in the event they are sued for malpractice.

From a theoretical point of view, the problem of understanding the coupling of two highly nonlinear and dynamically changing systems poses scientific challenges that we start investigating.

Situations that will gain advantages from this research are the following.

1. *Health Care*: As mentioned above, an immediate application of this research consists in disambiguating between actions that help a subject and actions that cause discomfort. This is especially important to monitor the nursing care given to people that are not able to perform everyday actions completely independently. Examples are elderly or younger subjects with physical disabilities that need help to perform everyday actions.
2. *Training*: Training athletes implies to have information about the correct movement to optimize the performance in the game. The interaction with other subjects is essential for training in sports that require a contact between players. Sport teams and video games companies may be interested in this application.
3. *Personal Signature*: Each subject has a very peculiar way to perform a movement. It depends on the age, weight, agility and psychological state, among other factors. If we can model a movement as a function of all these parameters, such function becomes a signature for the subject. This is useful to find if an action has been performed by a specific person.
4. *Individual Movement vs Two People Movement*: Performing a movement independently or interacting with another subject is very different. In the second case the needed force can be lower and differently distributed, in case of help from another subject, or can be higher and more concentrated in a specific area, in case of discomfort caused by others. The disambiguation between these two cases is very important in determining if a movement was performed independently or in conjunction with other subjects.

Let us consider a situation where a subject (care giver) helps another (care receiver) to stand up from a sitting position. The helper stands in front of the care receiver and, holding his/her hands, pulls the person up. The person that gives help spends energy pulling the first person up. On the contrary, the care receiver performs the action of getting up with less force than he/she would use if he/she were completely independent. The study of such energy exchange and the consequences on each other's total body movement are not trivial problems. The weights and the heights of the two people are involved and that affects also the threshold at which the situation switches from giving help to giving disconcert. Imagine, for example, if a person is pulled too hard. A similar question can be posed in the situation in which a person helps another person to sit down holding his/her hands so that he/she does not fall too hard into the chair. In this case it is important to detect when a violent action like pushing into the chair is taking place.
The main contributions of this paper are as follows:

- we introduce the problem of classifying the way two subjects physically interact and give motivation for this research;
- we provide a description of the system that contains discrete as well as continuous states (hybrid model);
- we present a classification algorithm and some preliminary classification results;
- we give directions and ideas for further research and experiments in this area.

The paper is organized as follows. Section 2 explains the experimental setup and the description of the system. Section 3 specifies how sensor data were collected. Our classification algorithm and some simulation results are presented in Section 4. Section 5 concludes the paper with some open problems that still need to be investigated.

## 2 Experimental Setup and System Description

Each person involved in the experiment wore 5 sensor nodes, one at the waist, one at each ankle and one at the each wrist. The sensor nodes communicated with a base station connected to a computer server via a USB port. Tmote Sky boards running TinyOS on a 8MHz microcontroller with 10K RAM were used as nodes. Each sensor included a 3-axis accelerometer and a 2-axis gyroscope. For now we did not include biosensors. Each axis was reported as a 12bit value to the node, indicating values in the range of $\pm 2g$ and $\pm 500°/s$ for the accelerometer and gyroscope, respectively. The protocol used for wireless communication was 802.15.4. Data were transmitted at 40Hz with minimal packet loss, thanks to a time division multiple access (TDMA) protocol.

We will describe the system as a hybrid model [3]. This is a dynamical model that involves the interaction of continuous (real-valued) states and discrete (taking values in a finite set) states. The states are the position, the velocity and the orientation, 9 quantities in total for each node. This corresponds to 90 continuous states for two subjects with five sensors each. The inputs are the 3D linear acceleration (provided by the accelerometer) and the 2D angular velocity (provided by the gyroscope). So for 5 sensors and 2 subjects we have 50 continuous inputs. We consider each sensor completely decoupled from other sensor as the nodes are not placed on fundamental joints as shoulders, knees, elbows that would allow strict coupling.

The discrete states describe a particular form of interaction. For example "Subject A sits down with help", "Subject A sits down without help", "Subject A is pushed into a chair".

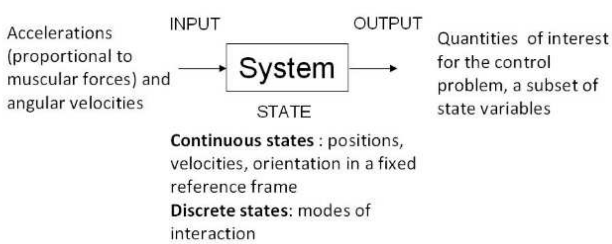


**Figure 2.** The model of the system.

In our initiative to classify the different activities we use the discrete part of the model described above. The continuous part will be crucial in performing a deeper mathematical analysis of each movement given the sensor data. This approach will be part of our next effort to detect the critical states of the system.

A graphical description of the system is given in figure 2. Notice that accelerations are inputs of the system and are proportional to the muscular forces that a subject applies through his nervous and muscular systems. Each movement has a peculiar input vector function (motion content) and includes variations dependent from the single person (motion style). At this stage we do not consider the body shape and strength of the subjects. Gravity is already included in the muscular reaction, so we did not consider it explicitly in our motion equations.

Table 1 shows six discrete states related to two actions, standing up and sitting down. We see how the same action can be performed with help, with discomfort caused by the helper, and without help.

| Correct Interaction | Wrong Interaction | No Interaction |
|---|---|---|
| B helps A to sit down | B pushes A into a chair | A sits down |
| B helps A to stand up | B pulls A forcefully | A stands up |

**Table 1.**

The set of critical states corresponds to the wrong interaction between the two subjects. For these first experiments we concentrated on the action of sitting down so that we did not try to identify the critical state corresponding to pulling a subject forcefully from a chair.

## 3 Data Collection

We collected data from two healthy subjects. The subjects first performed the actions independently. Then they repeated the same actions with a physical connection between them. In particular three trials for each of the following actions were repeated.

1. One subject.
   (a) The subject stands up
   (b) The subject carries an object while walking
   (c) The subject sits down
   (d) The subject walks one step forward
   (e) The subject throws and catches an object
2. Two subjects.
   (a) Person A helps Person B stand up
   (b) Person A and B carry an object together while walking
   (c) Person A helps Person B sit down
   (d) Person A pushes Person B from behind
   (e) Person A pushes Person B into a chair
   (f) Person A and B throw and catch an object

Of particular interest are the trials involving the same action repeated independently, with help and with discomfort.

Figure 4 shows the output of 5 sensors worn by a subject (subject A) while performing the action of sitting down from an initial standing position with the help of another person (subject B). The different axes of acceleration and rotation are color-coded accordingly to figure 3.

Figure 4 and 5 represent the cases in which subject A is helped to sit down, and is pushed down onto a chair. In the first situation (figure 4), subject B stands in front of subject

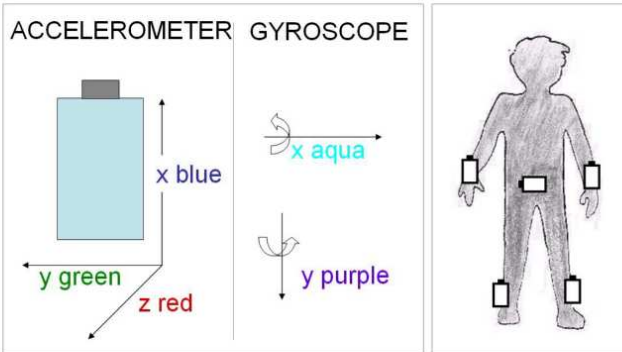


**Figure 3.** Meaning of the different colors that appear in the sensor readings and position of the body sensors.

A and, holding his/her hands, gives help in the movement of sitting down. In the second situation (figure 5) subject A is unexpectedly pushed onto a chair.

The sensor outputs that mostly distinguish the three cases are the ones related to the nodes attached to subject A ankles. It is graphically visible from figure 6 that shows the three cases next to each other. We see that for actions without help, the feet tend to move more, in order to stabilize the whole body. In the case of discomfort the feet move even more. We also see how being pushed results in shorter duration accelerations and rotations. Another difference is noticeable at the wrists that present different types of accelerations, which are indicative of the interface between the two subjects (assistance through holding hands, arms).

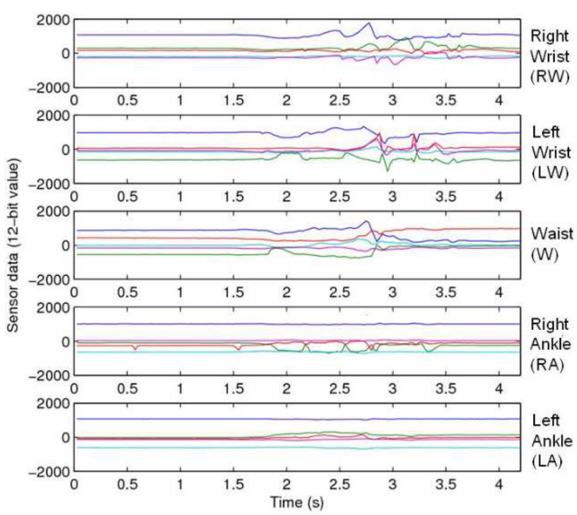


**Figure 4.** Sensor outputs of subject A sitting down with the help of subject B.

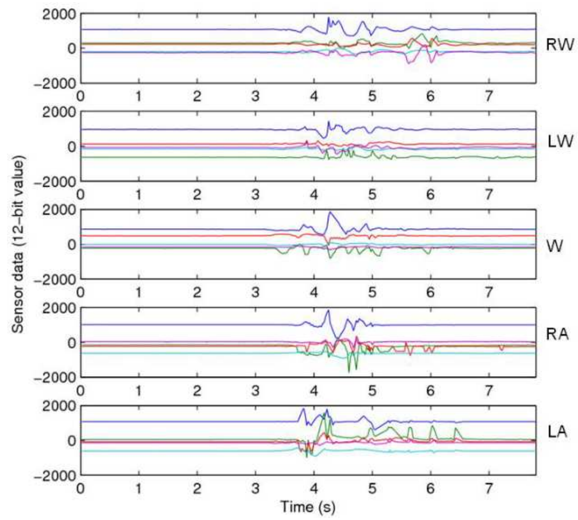


**Figure 5.** Sensor outputs of subject A pushed onto a chair by subject B.

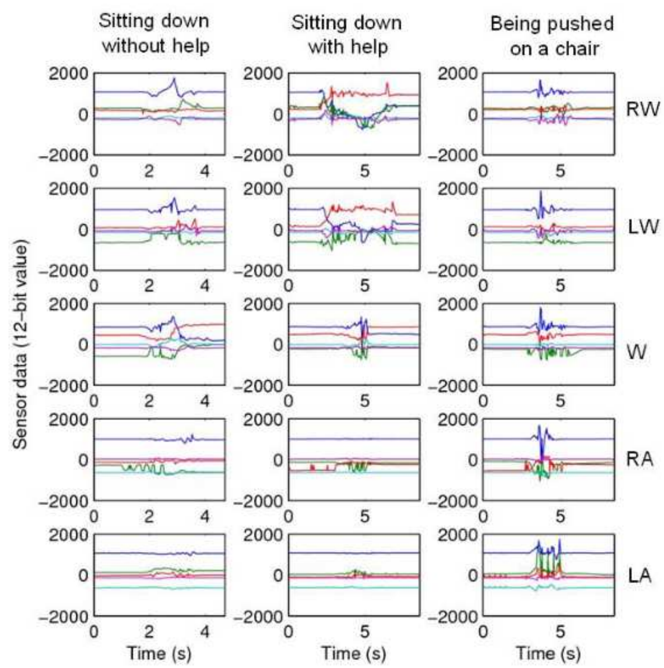


**Figure 6.** Sensor outputs for the action of sitting down in three situations.

For convenience, we decided to start our study with five specific behaviors as from Table 2.

In particular we want to distinguish between the correct interaction (2,4), the critical interaction (5) and the situations with no interaction at all (1,3). We can consider different levels of safety.

- A *safe state* corresponds to a correct interaction between the two subjects. The two subjects help each other or one subject helps the other. In this situation detection would not be necessary.
- A *warning state* corresponds to the absence of interaction, when it is needed (detection is recommended).
- A *critical state* corresponds to an incorrect interaction (detection is necessary). The critical state is the one in which a subject creates discomfort for the other subject.

## 4 Classification

Classification is the process of assigning sensor output to groups within groups of movements. An approach is based on processing the signals without taking into account physical or dynamic properties of the movement. A powerful way for clustering mixed data is to use *generalized principal component analysis* (GPCA). This technique allows to cluster data with structures that cannot be described using a single model [4]. GPCA is widely applied to computer vision, image processing, system identification.

Another approach is to use *motion primitives* that divide complex behaviors in local movements. This technique is promising when the degree of freedom is high, because the constraints are used to reduce complexity.

A more generative approach consists of building models of movements and check for the best model that justifies the sensor data. Here there is no need for training data, but building accurate models is not trivial.

Our approach consists of using some of the sensor outputs to learn the behaviors. In fact, the sample data are divided into 2 groups: the first one is used for the learning phase, the second one is used for recognition. We considered as *comparison functions*, characterizing each movement, either the accelerations in the relative reference frame of each sensor (accelerometers measurements), or the accelerations in an absolute (inertial) frame (computed by integrating accelerations with the data provided from the gyroscopes). Sensor output includes not only oscillations characterizing the motion but also some high frequency data that are due to unwanted muscle motion. For this reason in the learning phase, the functions are LP-filtered (the cut-off frequency of the filter is chosen at 20 Hz), normalized and sampled. Sample vectors are concatenated in order to have only one vector characterizing each learning experiment. Then the mean among such vectors gives us a vector (called *reference vector*) characterizing each movement. Similarly, the sample functions of each recognition experiment are normalized, sampled and concatenated in a *feature vector*. The recognition phase is based on the concept of ”minimum distance” among vectors. Each experiment is mapped into the discrete state whose reference vector has minimum distance from its feature vector.

| Discrete States | |
|---|---|
| 1 | Subject A stands up without help |
| 2 | Subject A stands up with help |
| 3 | Subject A sits down without help |
| 4 | Subject A sits down with help |
| 5 | Subject A is pushed into a chair |

**Table 2.**

First we performed analysis using sensor data from a subject for learning and data from the other subject for recognition. We noticed that this choice does not lead to exceptional results. We built a *recognition percentual rate vector* (with as many components as the number of movements), where the components show the recognition rate for each movement. The results are shown in table 3.

| Standing up without help | Standing up with help | Sitting down without help | Sitting down with help | Being pushed into a chair |
|---|---|---|---|---|
| | | | | Critical State |
| 100 | 83 | 100 | 46 | 50 |

**Table 3. Recognition percentual rate vector when the analysis was performed using sensor data from a subject for learning and data from the other subject for recognition**

The mean recognition rate is $75.8\%$. The results show that individual movements in the two subjects are very similar ($100\%$ recognition), while the interaction movements are not the same ($59.6\%$ recognition). This supports the idea that the physical connection between subject deeply changes the way actions are performed. Notice that the critical location state is not detected in the $50\%$ of cases ($50\%$ of false negatives).

Using $1/3$ samples (from both subjects) for training and $2/3$ for recognition, we were able to achieve better results. Table 4 shows the classification results in this case.

| Standing up without help | Standing up with help | Sitting down without help | Sitting down with help | Being pushed into a chair |
|---|---|---|---|---|
| | | | | Critical State |
| 100 | 71 | 100 | 96 | 88 |

**Table 4. Recognition percentual rate vector when $1/3$ samples (from both subjects) were used for training and $2/3$ for recognition**

The mean recognition rate of $90.8\%$. Notice that $12\%$ of critical situations are not detected. With more learning samples $(2/3)$, there is a further improvement (table 5) and the average recognition rate of $98.2\%$. The critical situations are always detected. Using as comparison functions the accelerations in the absolute frame instead of the ones in the relative frames (raw data), the average recognition rate increases from $92.4\%$ to $96.6\%$.

| Standing up without help | Standing up with help | Sitting down without help | Sitting down with help | Being pushed into a chair |
|---|---|---|---|---|
| | | | | Critical State |
| 100 | 100 | 100 | 91 | 100 |

**Table 5. Recognition percentual rate vector when $2/3$ samples (from both subjects) were used for training and $1/3$ for recognition**

In our experiments we considered the case of recognition with partial information to address the issue of lack of real data (i.e. packet loss) and in order to save power resources on nodes or servers. This reflects the fact that recognition is reasonable if it is performed in real time during the execution of the action.

We called *primary* the sensors on the person who needs to be helped and *secondary* the sensors on the helper. Our experiments show that considering only the primary sensors ($50\%$ save in computation and memory occupation) does not affect recognition results at all.

## 5 Conclusion and Open Problems

In this paper we addressed the problem of classifying the interaction between two subjects. At first, we described an experimental setup and provided a hybrid model of the interaction, in which the continuous states are kinematic quantities, while the discrete states are "modes" of behavior/interaction, which describe the type of interaction. We defined critical states as the discrete states in which a subject causes discomfort to the other subject. We addressed the problem of the classification of interactions and provided a simple algorithm for the discrete state recognition. Experimental results show a good behavior of the recognition algorithm, with particular regard to the detection of the critical state. Best results are achieved when raw acceleration data are integrated with rotation informations, provided by the gyroscopes.

Our approach has some limitations we will address in our future work. These include modeling and experimental issues. With respect to the model, the use of a complete hybrid modeling, integrating both discrete and continuous dynamics, will be crucial in performing a deeper mathematical analysis of each movement and an improvement in the detection of critical states. From the experimental point of view, a current limitation is the lack of data. To improve our results we plan to have a larger set of possible interactions. In fact, we will create a database of typical interactions between two subjects (collecting many classes of individual actions). It is important, in fact, to have subjects who are different in size and weight. Further issues can be: real-time reconstruction and visualization of motion based on state-space modeling, integration with Avatar [2], observation/detection of more complex situations relating health-care. We also plan to integrate data from the bio-sensors and detection of complex actions in 3 different steps, at the level of Body Sensor Layer, Personal Network Layer and Global Network Layer, with different levels of elaboration.

## 6 Acknowledgments

This work has been partially supported by the Centre of Excellence for Research DEWS (Design methodologies for Embedded controllers, Wireless interconnect and System-on-chip), L'Aquila (Italy), and by TRUST (Team for Research in Ubiquitous Secure Technology) which receives support from the National Science Foundation (NSF award number CCF-0424422). The authors would like to thank Edmund Seto, Posu Yan, Allen Yang and the other members of their group who helped in the data collection and provided precious advices.

## References

[1] http://www.eecs.berkeley.edu/ yang/software/war/. Last visited: 01/27/2009.

[2] P. Kuryloski, A. Giani, R. Giannantonio, K. Gilani, R. Gravina, V. Seppa, E. Seto, V. Shia, C. Wang, P. Yan, A. Yang, J. Hyttinen, S. Sastry, S. Wicker, and R. Bajcsy. Dexternet: An open platform for heterogeneous body sensor networks and its applications. *Body Sensor Network (submitted)*, 2009.

[3] J. Lygeros. Lecture notes on hybrid systems. *http://robotics.eecs.berkeley.edu/ sastry/ee291e/lygeros.pdf*, Last visited: 01/27/2009, 2004.

[4] Y. Ma, A. Yang, H. Derksen, and R. Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Review*, 2008.

[5] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: A local svm approach. *17th International Conference on Pattern Recognition (ICPR'04)*, 3, 2002.

[6] A. Yang, R. Jarafi, P. Kuryloski, S. Iyengar, S. Sastry, and R. Bajcsy. Distributed segmentation and classification of human actions using a wearable motion sensor network. *Workshop on Human Communicative Behavior Analysis, CVPR*, 2008.