Towards an Interactive Personal Care System driven by Sensor Data

Stefano Bragaglia, Paola Mello, and Davide Sottara

DEIS, Facoltà di Ingegneria, Università di Bologna
Viale Risorgimento 2, 40136 Bologna (BO) Italy
stefano.bragaglia@unibo.it, paola.mello@unibo.it,
davide.sottara2@unibo.it

Abstract. This demo presents an integrated application, exploiting technologies derived from various branches of Artificial Intelligence. The prototype, when triggered by an appropriate event, interacts with a person and expects them to perform a simple acknowledgement gesture: the following actions, then, depend on the actual recognition of this gesture. In order to achieve this goal, a combination of data stream processing and rule based programming is used. While the more quantitative components exploit techniques such as Clustering or (Hidden) Markov Models, the higher levels, more qualitative and declarative, are based on well known frameworks such as Event Calculus and Fuzzy Logic.

Keywords: Complex Event Processing, Event Calculus, Sensor Data Fusion, Activity Recognition

1 Introduction

Intelligent beings need senses to interact with the environment they live in: the same principle holds for artificially intelligent entities when they are applied outside of purely virtual contexts, in the real world. Much effort is being spent in emulating smell (processing signals acquired through electronic noses), taste (using electronic tongues) and touch, although hearing and especially sight are arguably the subject of an even larger amount of research and application development. Focusing on computer vision, observing the position or the shape of people and objects is extremely relevant when dealing with problems such as pathfinding, tracking, planning or monitoring. Moreover, analysing the visual information is often the prelude to a decisional process, where actions are scheduled depending on the sensed inputs and the current goals. In our case study, part of the DEPICT project, we are addressing the monitoring of an elderly person during their daily activity, recognizing, if possible, the insurgence of undesired short and long term conditions, such as frailty, and ultimately trying to prevent severely detrimental events such as falls. Should a fall or similar event actually happen, however, it is essential to recognize it as soon as possible, in order to take the appropriate actions. To this end, we are planning to use a mobile platform equipped with optical and audio sensors and a decisional processing unit: the device would track or locate the person as needed, using the sensors

to acquire information on their current status and capabilities. In this paper, however, we will not discuss the mobile platform, but focus on the sensors it is equipped with, and the collected information. This information will be analyzed to identify and isolate, in ascending order of abstraction, *poses*, *gestures*, *actions* and *activities* [1]. A pose is a specific position assumed by one or more parts of the body, such as "sitting", "standing" or "arms folded"; a gesture is a simple, atomic movement of a specific body part (e.g. "waving hand", "turning head"); actions are more complex interactions such as "walking" or "standing up", while activities are composite actions, usually performed with a goal in mind. Recognizing patterns directly at each level of abstraction is a relevant research task of its own, but it may also be necessary to share and correlate information between the levels. From a monitoring perspective, the more general information may provide the necessary context for a proper analysis of the more detailed one: for example, one might be interested in a postural analysis of the spine when a person is walking or standing up from a chair, but not when a person is sleeping.

2 AI Techniques

In order to analyse a person's movements in a flexible and scalable way, we propose the adoption of a hybrid architecture, combining low level image and signal processing techniques with a more high level reasoning system. The core of the system is developed using the "Knowledge Integration Platform" Drools[1], an open source suite which is particulary suitable for this kind of applications. It is based on a production rule engine, allowing to encode knowledge in the form of `if-then` rules. A rule's premise is normally a logic, declarative construction stating the conditions for the application of some consequent action; the consequences, instead, are more operational and define which actions should be executed (including the generation of new data) when a premise is satisfied. Drools, then, builds its additional capabilities on top of its core engine: it supports, among other things, temporal reasoning, a limited form of functional programming and reasoning under uncertainty and/or vagueness [6]. Being object oriented and written in Java, it is platform independent and can be easily integrated with other existing components. Using this platform, we built our demo application implementing and integrating other well known techniques.

Vision sub-system. At the data input level, we used a Kinect hardware sensor due to its robustness, availability and low price. It combines a traditional camera with an infrared depth camera, allowing to reconstruct both 2D and 3D images. We used it in combination with the open source OpenNI[2] middleware, in particular exploiting its tracking component, which allows to identify and trace the position of humanoid figures in a scene. For each figure, the coordinates of their "joints" (neck, elbow, hip, etc...) are estimated and sampled with a frequency of 30Hz.

[1] http://www.jboss.org/drools
[2] http://www.OpenNI.org

Vocal sub-system. We rely on the open source projects FreeTTS[3] and Sphinx-4[4] for speech synthesis and recognition, respectively.

Semantic domain model. Ontologies are formal descriptions of a domain, defining the relevant concepts and the relationships between them. An ontology is readable by domain experts, but can also be processed by a (semantic) reasoner to infer and make additional knowledge explicit, check the logical consistency of a set of statements and recognize (classify) individual entities given their properties. For our specific case, we are developing a simple ontology of the body parts (joints) and a second ontology of poses and acts. The ontologies are then converted into an object model [5], composed of appropriate classes and interfaces, which can be used to model facts and write rules.

Event Calculus. The Event Calculus [2] is another well known formalism, used to represent the effects of actions and changes on a domain. The base formulation of the EC consists of a small set of simple logical axioms, which correlate the happening of *events* with *fluents*. An event is a relevant state change in a monitored domain, taking place at a specific point in time; fluents, instead, denote relevant domain properties and the time intervals during which they hold. From a more reactive perspective, the fluents define the overall state of a system through its relevant properties; the events, instead, mark the transitions between those states.

Complex Event Processing. In addition to flipping fluents, events may become relevant because of their relation to other events: tipical relations include causality, temporal sequencing or aggregation [4]. Causality indicates that an event is a direct consequence of another; sequencing imposes constraints on the order events may appear in; aggregation allows to define higher level, more abstract events from a set of simpler ones. Exploiting these relations is important as the number and frequency of events increases, in order to limit the amount of information, filtering and preserving only what is relevant.

Fuzzy Logic. When complex or context-specific conditions are involved, it may be difficult to define when a property is definitely true or false. Instead, it may be easier to provide a vague definition, allowing the property to hold up to some intermediate degree, defined on a scale of values. Fuzzy logic [3] deals with this kind of graduality, extending traditional logic to support formulas involving graded predicates. In this kind of logic, "linguistic" predicates such as `old` or `tall` can replace crisp, quantitative constraints with vague (and smoother) qualitative expressions. Likewise, logical formulas evaluate to a degree rather than being either valid or not.

3 Demo Outline.

The techniques listed in the previous section have been used as building blocks for a simple demo application, demonstrating a user/machine interaction, based on vision and supported by the exchange of some vocal messages. The abstract

[3] http://freetts.sourceforge.net/docs/index.php
[4] http://cmusphinx.sourceforge.net/sphinx4/

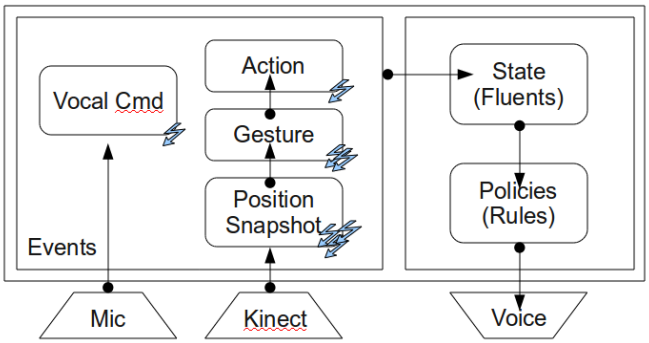

Fig. 1. Architectural outline

system architecture is depicted in Figure 1. The simple protocol we propose as a use case, instead, is a simplified interaction pattern where a monitoring system is checking that a person has at least some degree of control over their cognitive and physical capabilities. This is especially important when monitoring elderly patients, who might be suffering from (progressively) debilitating conditions such as Parkinson's disease and require constant checks. Similarly, in case of falls, estimating the person's level of consciousness may be extremely important to determine the best emergency rescue plan.

To this end, for each person tracked by the Kinect sensor, we generate an object model (defined in the ontology) describing its skeleton with its joints and their position in space. The actual coordinates are averaged over a window of 10 samples to reduce noise: every time a new average is computed, an event is generated to notify the updated coordinates. The coordinates are then matched to predetermined patterns in order to detect specific poses: each snapshot provides a vector of 63 features (3 coordinates and 1 certainty factor for each one of the 15 joints, plus the 3 coordinates of the center of mass) which can be used for the classification. The definition of a pose may involve one or more joints, up to the whole skeleton and can be expressed in several ways. In this work, we adopted a "semantic" [1] approach, providing definitions in the form of constraints (rules) on the joints' coordinates. To be more robust, we used fuzzy predicates (e.g. `leftHand.Y is high`) rather than numeric thresholds, but the infrastructure could easily support other types of classifiers (e.g. neural networks or clusterers). Regardless of the classification technique, we assume the state of being in a given pose can be modelled using a fuzzy fluent, i.e. a fluent with a gradual degree of truth in the range $[0, 1]$. A pose, in fact, is usually mantained for a limited amount of time and, during that period, the actual body position may not fit the definition perfectly. In particular, we consider the maximum degree of compatibility (similarity) between the sampled positions and the pose definition over the minimal interval when the compatibility is greater than 0 (i.e. it is not impossible that the body is assuming the considered pose). Technically, we keep

a fluent for each person and pose, declipping it when the compatibility between its joints and the pose is strictly greater than 0 and clipping it as soon as it becomes 0. Given the poses and their validity intervals and degrees, denoted by the fluents, it is possible to define gestures and actions as sequences of poses and/or gestures. While we are planning to consider techniques such as (semi-)Hidden Markov Models, currently we continue adopting a "semantic", CEP-like approach at all levels of abstraction. Gestures and actions, then, are defined using rules involving lower level fluents. Their recognition is enabled only if they are relevant given the current context: the event sequencing rules, in fact, are conditioned by other state fluents, whose state in turn depends by other events generated by the environment.

Usage. The user is expected to launch the application and wait until the tracking system has completed its calibration phase, fact which is notified by a vocal message. From this point on, until the user leaves the camera scope, the position of their joints in a 3D space will be continuously estimated. When the application enters a special recognition state, triggered pressing a button or uttering the word "*Help*", the user has 60 seconds to execute a specific sequence of minor actions, in particular raising their left and right hand in a sequence or stating "*Fine*". With this example, we simulate a possible alert condition, where the system believes that something bad might have happened to the person and wants some feedback on their health and their interaction capabilities.

If the user responds with gestures, the system will observe the vertical coordinates of the hands: a hand will be raised to a degree, which will be the higher the more the hand will be above the shoulder level. A hand partially lifted will still allow to reach the goal, but the final result will be less than optimal. At any given time, the recognition process considers the highest point a hand has reached so far, unless the hand is lowered below the shoulder (the score is reset to 0). If both hands have been completey lifted, the user has completed their task and the alert condition is cancelled; otherwise, the system will check the current score when the timeout expires. In the worst case scenario, at least one of the hands has not been raised at all, leading to a complete failure; finally, in intermediate situations, at least one hand will not have been lifted completely, leading to a partial score. The outcome will be indicated using a color code: green for complete success, red for complete failure and shades of orange for the intermediate cases, in addition to a vocal response message. If the user answers vocally, instead, the alarm will be cleared and no further action will be taken.

4 Conclusions

We have shown an example of a tightly integrated, leveraging both quantitative and qualitative AI-based tools. Despite its simplicity, it proves the feasibility of an interactive, intelligent care system with sensor fusion and decision support capabilities.

Acknowledgments

This work has been funded by the DEIS project DEPICT.

References

1. J.K. Aggarwal and M.S. Ryoo. Human activity analysis: A review. *ACM Comput. Surv.*, 43(3):16:1–16:43, April 2011.
2. F. Chesani, P. Mello, M. Montali, and P. Torroni. A logic-based, reactive calculus of events. *Fundamenta Informaticae*, 105(1):135–161, 2010.
3. Petr Hájek. *Metamathematics of Fuzzy Logic*, volume 4 of *Trends in Logic: Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 1998.
4. David C. Luckham. *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.
5. G. Meditskos and N. Bassiliades. A rule-based object-oriented OWL reasoner. *IEEE Transactions on Knowledge and Data Engineering*, 20(3):397–410, 2008.
6. D. Sottara, P. Mello, and M. Proctor. A configurable rete-oo engine for reasoning with different types of imperfect information. *IEEE Trans. Knowl. Data Eng.*, 22(11):1535–1548, 2010.