Gestural Activity Recognition for Canine-Human Communication

Giancarlo Valentin
Animal Interaction Lab
Georgia Institute of Technology
Atlanta, GA 30332
giancarlo@gatech.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

ISWC'14 Adjunct, September 13 - 17, 2014, Seattle, WA, USA
ACM 978-1-4503-3048-0/14/09.
http://dx.doi.org/10.1145/2641248.2642733

Abstract
Despite close collaboration existing between humans and working dogs, there are few options for reliable two-way communication between them. The main goal of the FIDO project is to explore and develop wearable technologies to support this communication [5]. In this manuscript, we describe work in progress regarding the use of intentional, motion-based dog gestures as a mechanism of communication. In particular, we are interested in gestures that can be identified with the use of inertial measurement sensors such as accelerometers and gyroscopes.

Author Keywords
Wearable technology; Animal-Computer Interaction

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction and Overview
Recognizing the need for working dogs to clearly communicate to their owners (or other humans), the FIDO project researches practical applications of wearable technologies to better facilitate this communication. We began exploring wearable interfaces for dogs with occupations ranging from assistance scenarios (e.g., guide
and hearing dogs) to field work (e.g. search & rescue or law enforcement) [5]. The field of dog training has also benefited from advances in wearable computing since passive sensors were used to estimate canine posture [4] during training. We propose a combination of these approaches for enhancing intentional communication between working dogs and their human companions.

Most canine communication (including with humans) seems to be based on unique body postures [1]. These gestures are used by experienced handlers to achieve an intricate form of communication with their service dogs. For example, many police dogs raise their nose in the air when they first perceive the scent of interest. Inexperienced handlers, on the other hand, may not perceive this gesture, and (unknowingly) request that the dog ignore this air current and continue searching the current location.

In addition to detecting existing gestures, new ones (performed on an environmental cue) can also be detected. For example, a guide dog can be trained to perform a certain head, nose, or paw movement to wirelessly alert a handler of an object to avoid. We propose translating these actions into an audio or vibration output that any human, regardless of limited proximity, hearing or sight, can perceive.

**Previous Work**

There have been two previous efforts to recognize dog activities from inertial measurement data. The first, alluded to earlier, attempted to do this as part of an automated training system [4]. By recognizing the dog’s posture, it would be possible for the system to determine if the correct action was performed and whether a reward should be dispensed. Although originally intended for posture estimation, this work has expanded to include non-static activities.

The second effort in this field was undertaken by researchers at the Culture Lab in Newcastle University [6]. Their experiments were focused on monitoring activities that correspond to healthy behavior traits of pet dogs. These previous efforts provide a solid foundation for our study of the use of gestures in communication. Although the ultimate goal is inherently different, we began our work by expanding upon their results before proceeding to gesture-specific activities which are not part of daily behavior. Additionally, we draw upon the vast literature on human activity recognition using inertial measurements.

**Problem Statement**

To address the informational asymmetry between the large amount of stimuli perceived by working dogs and the small amount of options to communicate them, we propose using wearable inertial measurement sensors to detect and mediate intentional communication between working dogs and the humans around them. This problem must be addressed on all three levels of communication described by Shannon’s information theoretic model [7]. These are: the accuracy of transmissions, the semantic interpretation of the messages, and their effectiveness in changing behavior. Some of the peculiarities that differentiate this problem from traditional work in activity recognition are the following:

1. Users are unable to annotate their own data.
2. Users are unable to reposition sensor if dislodged.
3. Activities are non-periodic and short in duration.
4. High reliability is required (no false-positives).
Current Work
The first method we employed was to gather data from dogs using the on-body Axivity Sensor Platform developed at Newcastle. This device includes a three-axis accelerometer (but not a gyroscope at this point). It was attached to the front of a service dog harness (Figure 1). The placement of the sensor on the neck (as a collar) was initially explored, but postponed until a mechanism to maintain the position (or compensate for its change) is determined.

Figure 1: The Axivity accelerometer (circled) was used in the first experiment. This will be replaced by a unit with six inertial measurements.

With the accelerometer on the harness, each dog was instructed to perform certain activities using behavior-reward scenarios. Multiple repetitions of these activities were video-recorded, synchronized, and hand annotated using the ELAN annotation toolkit. These labels were then associated with the corresponding raw data. The activities included: spin (clockwise), twirl (counterclockwise), jump, sit from stand, stand from downstay, sit from downstay, downstay from stand, downstay from sit, roll over (to the right), roll over (to the left), roll-return (to the right) and roll-return (to the left).

Originally, only a subset of these activities were intended to be detected (spin, twirl, jump and roll). As we labeled the data in the video recording, we noticed other activities that occurred frequently and decided to label them as well to explore whether enough discriminatory power existed to separate them (e.g. sit-from-stand, sit-from-downstay).

In other cases, we noticed different versions of an activity that we previously thought as the same, and decided to label them separately. The most notable example of this was roll-over to the left, and its half-roll counterpart, "troll-return". We decided to differentiate these, both in terms of direction and motion. The net result yielded four activities from what was previously considered a single one.

We began data collection with participants of different breeds (all male, age between 2-7). The activities were sampled at a rate of 100 Hz with a range of +/-8g. The data was segmented using a window size corresponding to 100 samples and 50% overlap. The vectors for each dimension were concatenated along a single dimension (1x300). No feature selection techniques were performed at this stage.

Performance and Evaluation
We currently use the WEKA toolkit to perform a preliminary analysis of different classification methods. WEKA is a freely available Java-based machine learning tool. For evaluation of classification methods, we first used a ten-fold cross validation method in continuous
streams of data without the null class. Classification by random forests yielded the highest accuracy across all techniques for within-subject training and testing (92% to 98% frame-level accuracy for ten activities). Nevertheless, our current use of explicit signal matching is not satisfactory for subject-independent classification. Considering that "of all the species on the planet, dogs have the largest variation in size" [3], improving on subject-independent classification is a critical next step.

Planned Work

Data Analysis

Unlike evaluating classification algorithms, the heuristic nature of the feature selection process makes the search for appropriate techniques less amenable to systematic analysis. This issue has been described at length by researchers in this domain[6]. Up to this point, we have (temporarily) avoided these considerations by examining the bounds of performance with the raw features only. As we move forward, achieving generalizable models will require appropriate methods for:

- Data pre-processing and normalization.
- Correcting for changes in orientation.
- Accounting for instances of the null class.
- Feature selection.

Experiments

We are planning to add less commonly performed activities to our data-collection experiments. The number and type of participants is to be increased as well to represent a broader population. For the purposes of finding appropriate gestures, 'new' gestures will be explored against a database of everyday activities performed by working dogs based on similar work performed with humans [2]. Nonetheless, we note that in current practice, the context of the gesture is critical. For example, canines in bomb sniffing squads perform alerts with behaviors as common as lying down. This behavior is acceptable because the context contains enough information to provide a clear indication of the alert. Such an alert, however, might not be acceptable for dogs in other lines of work. The extent to which each type of working dog will require specific gestures will be explored.

New data will be collected using a different platform containing both accelerometer and gyroscopic measurements. We expect the additional degrees of freedom will help compensate for the rotation of the sensor along the collar. The new placement for the sensor (Figure 2) will be along the collar since head movements are present in most gestures in canine communication. These improvements must be in place before we can begin porting the system into an embedded environment for real time implementation and deployment.

Figure 2: Although any direction could be accounted for, the sensor is front-facing for ease of illustration.
Acknowledgements
This work is funded by the National Science Foundation under grant No. IIS-1320690 and the GVU Center research grants program. I would additionally like to thank Professor Thad Starner, Daniel Kohlsdorf and Ryan Kerwin for their assistance at different stages of this work.

Biographical Sketch
I am currently a Ph.D. student at the School of Electrical and Computer Engineering at Georgia Tech. My interests are at the intersection of animal-computer interaction, wearable computing and pattern recognition. I joined the Ph.D. program in the fall of 2012, and tentatively plan to graduate in the spring of 2016. My advisors are Dr. Melody M. Jackson and Dr. Ayanna M. Howard.

References