Distributed Consensus for Interaction between Humans and Mobile Robot Swarms (Demonstration)

Alessandro Giusti, Jawad Nagi, Luca M. Gambardella, Gianni A. Di Caro
Dalle Molle Institute for Artificial Intelligence (IDSIA)
Lugano, Switzerland
{alessandrog,jawad,luca,gianni}@idsia.ch

ABSTRACT

The purpose of the demonstrator is to present a novel system for gesture-based interaction between humans and a swarm of mobile robots. The human interacts with the swarm by showing hand gestures using an orange glove. Following initial hand glove detection, the robots move to adapt their positions and viewpoints. The purpose is to improve individual sensing performance and maximize the gesture information mutually gathered by the swarm as a whole. Using multi-hop message relaying, robots spread their opinions and the associated confidence about the issued hand gesture throughout the swarm. To let the robots in the swarm integrate and weight the different opinions, we developed a distributed consensus protocol. When a robot has gathered enough evidence, it takes a decision for the hand gesture, and sends it into the swarm. Different decisions compete with each other. The one assessed with the highest confidence eventually wins. When consensus is reached about the hand gesture, the swarm acts accordingly, for example by moving to a location, or splitting into groups.

The working of the system is shown and explained in the video accessible at the following address: `http://www.idsia.ch/~gianni/SwarmRobotics/aamasdemo.zip`.

Categories and Subject Descriptors

I.2.9 [**Robotics**]; I.2.11 [**Distributed Artificial Intelligence**]: Coherence and coordination; C.2.4 [**Computer Communication Networks**]: Distributed applications

General Terms

Algorithms

Keywords

Gesture recognition, Distributed consensus, Swarm robotics

1. INTRODUCTION

We consider the problem of the *interaction between humans and robotic swarms*. The purpose is to let a human communicating commands to be executed by the swarm (e.g., split in two groups). The problem is particularly challenging since the robots in the swarm can be spread in different positions in the environment and be engaged in tasks of their own when the command is issued. Moreover, the robots typically used in swarm robotics are relatively simple and have limited processing capabilities. The task of the robots is to *detect* and *understand* the command, and collectively reach a *distributed consensus* about it in order to actuate its execution.

We use *hand gestures* as mean for human-swarm communication. In our scenario, a hand gesture encodes a command, that the swarm will execute. Hand gestures are a powerful and intuitive way to communicate, and do not require the use of additional devices. However, real-time *vision-based* recognition of hand gestures is a challenging task for the single robot, due to the limited processing power and field of view of robots that we use, the *foot-bots* (see next section).

We investigated how to exploit *robot mobility*, *swarm spatial distribution*, and *multi-hop wireless communications*, to let the robots in the swarm: (i) implement a *distributed and cooperative sensing* of hand gestures, and (ii) robustly reach a *consensus* about a gesture.

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), June, 4–8, 2012, Valencia, Spain.

2. THE ROBOTS IN THE SWARM

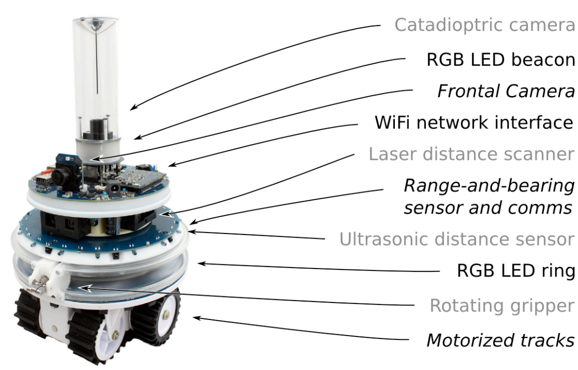

Figure 1: The foot-bot mobile platform. Italicized text indicates features we use; remaining features are either for monitoring (black), or are not used (gray).

We use *foot-bot* robots (Figure 1), developed in the *Swarmanoid* project [1] specifically for swarm robotics applications. The foot-bot is based on an on-board ARM 11 533MHz with a Linux-based operating environment.

We use a subset of the sensors and actuators available on such platform. In particular, the *frontal camera* is used for recognizing gestures, and acquires 512×384 RGB images. *Motorized track-based wheels* allow robots to move at speeds up to 5cm per second. The infrared-based *range-and-bearing* sensor and actuator allows a robot to detect its line-of-sight neighbors up to a range of few meters, and to recover their distance and bearing; messages can be broadcast to neighbors through a low-bandwidth (100 bytes/sec), low-reliability communication channel; in our implementation, all messages propagate to the swarm using multi-hop communication on this system. *RGB LEDs* are used to display the state of the system and for notifying the user about the decision the swarm has taken.

3. GESTURE RECOGNITION AND DISTRIBUTED CONSENSUS

We consider the two sets of gestures represented in Figure 2, namely finger counts (from 0 to 5) and four ad-hoc gestures representing furniture-like shapes, designed for interacting with *roombot* robots [2].

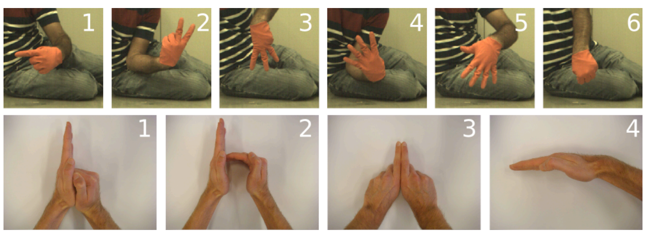

Figure 2: The six finger-count gestures (first row), and the four furniture-like gestures (second row).

3.1 Training of the Gesture Classifier

The first step was to use 13 foot-bots to collect a total of 70,000 hand gesture images from 65 different points of view. Figure 3 shows the acquisition setup.

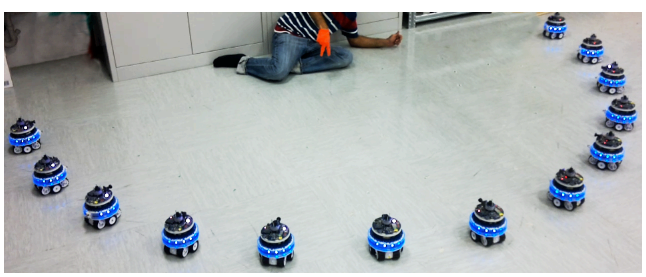

Figure 3: Setup for the training dataset acquisition.

For each acquired image, robots use color-based segmentation to detect the glove and obtain a binary mask, from which 20 shape features are computed.

With this data set we trained a *Support Vector Machine*, which is used by the robots for *individual gesture classification* and generation of an *opinion vector*, assigning a probability to each known gesture [3]. The resulting classifier performs correctly independently on the orientation of the hand since it was trained from images obtained from different points of view.

3.2 Distributed Gesture Recognition

In our scenario, robots search for the glove in an environment. When a robot detects the glove, the rest of the robots *moves* to adapt their viewpoint for better sensing.

Robots then start acquiring hand images at a rate of roughly one per second. Immediately after each acquisition, the image is processed as described above. The resulting opinions (i.e., classification vectors representing the probability of each gesture) are spread throughout the swarm through *multi-hop message relaying*. Each robot records its own opinions (deriving from successive acquisitions) as well as opinions received from the rest of the robots in the swarm.

The full set of available opinions (which may be conflicting) are additively taken into account by each robot, that incrementally builds a decision vector $\mathbf{D}$ as the component-wise *sum* of all the classification vectors (opinions) it has locally generated and/or received from other robots. $\mathbf{D}$'s component with the highest value, i', indicates the gesture class in favor of which most evidence is available at the moment to the robot. The robot also calculates a measure of its *confidence* about the true class being i' as $\lambda = \mathbf{D}_{i'} - \mathbf{D}_{i''}$, where i'' is the index of the second highest component of $\mathbf{D}$.

When a robot has gathered enough evidence, i.e., when λ exceeds a predefined threshold, it takes a *decision* for the hand gesture, and sends it into the swarm, where it is propagated through wireless line-of-sight multi-hop communication. Robots receiving a decision immediately adopt it. If different decisions are generated in a swarm, the one assessed with the highest confidence overrides and shutdowns the propagation of the others.

The linked video shows two examples of command *execution*. For the furniture-like shapes, robots send their decisions to the simulated *Roombot* system, where modular robots 'build' the furniture [2]. For the finger count, after receiving a 'two', the swarm splits in two groups moving in opposite directions.

4. DEMONSTRATION

The system will be demonstrated with a swarm composed by at least 5 foot-bots, as shown in the video. The full scenario described above will be operational and the participants will be able to interact with the swarm (one person at a time) using an orange glove.

5. ACKNOWLEDGMENTS

This research was supported by the Swiss National Science Foundation (SNSF) through the National Centre of Competence in Research (NCCR) Robotics (`www.nccr-robotics.ch`). The Authors participate to Project 4: *Distributed Robotics*, sub-project 4.5, *Symbiotic human-swarm cooperation*.

6. REFERENCES

[1] *Swarmanoid: Towards humanoid robotic swarms.* FET-OPEN project funded by the European Commission, `http://www.swarmanoid.org`.

[2] *Roombots: Modular robotics for adaptive and self-organizing furniture.* `http://biorob.epfl.ch/roombots`.

[3] J. Nagi et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *Proc. of IEEE ICSIPA*, Kuala Lumpur, Malaysia, 2011.