World Embedded Interfaces for Human-Robot Interaction*

Mike Daily, Youngkwan Cho, Kevin Martin, Dave Payton
HRL Laboratories, LLC.
3011 Malibu Canyon Road, Malibu CA 90265
{mjdaily, ykcho, martin, payton}@hrl.com

ABSTRACT

Human interaction with large numbers of robots or distributed sensors presents a number of difficult challenges including supervisory management, monitoring of individual and collective state, and apprehending situation awareness. A rich source of information about the environment can be provided even with robots that have no explicit representations or maps of their locale. To do this, we transform a robot swarm into a distributed interface embedded within the environment. Visually, each robot acts like a pixel within a much larger visual display space so that any robot need only communicate a small amount of information from its current location. Our approach uses Augmented Reality techniques for communicating information to humans from large numbers of small-scale robots to enable situation awareness, monitoring, and control for surveillance, reconnaissance, hazard detection, and path finding.

Keywords: Augmented reality, robot swarm, human-robot interface

1. INTRODUCTION

Emerging miniaturization technologies (e.g. micro machining and MEMS) will someday enable the creation of large numbers of extremely small robots, with fully self-contained sensors, actuators, computation, and power. While such robots individually are of limited use, thousands of them, operating as a coordinated swarm, could conceivably accomplish a wide range of significant tasks [5,6,9]. Ultimately, swarms of small scale robots should be able to achieve large-scale results in tasks such as surveillance, reconnaissance, hazard detection, path finding, payload conveyance, and small-scale actuation. However, to fully exploit the prospects of miniaturization, we must first address the challenges posed by the need for humans to interact with, communicate with, and coordinate the activities of thousands of tiny cooperating entities.

Coordinating and interacting with a large collective of tiny robots involves many issues that are not encountered when dealing with one or a few robots [6,7,8]. Even something as trivial as turning them all on at the same time requires new interface approaches when dealing with many thousands of robots. Interaction schemes that require unique identities for each robot, direct control and communication between human operators and robots, monitoring specific robots, and using data in centralized representations for human consumption will not be feasible when dealing with extremely large numbers of robots.

Our focus in this paper is to address the extraction of useful information from a robot swarm with minimal requirements for communications bandwidth or accurate positional information.

Consider a search and rescue scenario where a search team enters an unfamiliar building after a disaster and needs to quickly locate survivors We envision the team opening a jar and emptying thousands of tiny robots into the building. This robot swarm quickly disperses throughout the building, with each individual robot maintaining communications contact only with nearby neighbors. A robot, upon detection of a survivor, emits a message signaling the discovery. This message is diffused throughout the distributed mesh of robots, propagating only along unobstructed paths, and ultimately, making its way back to the rescue team. As a result, each robot in the swarm has information about the best direction to head from its own location in order to reach the survivor. This, in effect, provides a gradient encoding of all possible paths to the survivor [13]. The human operator views this path information as a series of arrows superimposed on the position of each visible robot (see Figure 1). The path information holds the local gradient in the direction of the

* This work is supported by the Defense Advanced Research Projects Agency under contract N66001-99-C-8514. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Defense Advanced Research Projects Agency.

0-7695-1874-5/03 $17.00 © 2002 IEEE

survivor, so following this gradient will provide the team with the shortest unobstructed path to the survivor.

This approach has numerous advantages. From the standpoint of the user, information from the swarm is presented within the context where it is needed, enhancing situation awareness. Robot swarm operators do not need to turn attention from their local environment to understand information from the swarm. World embedded interfaces also take advantage of the large number of robots in a swarm in a completely distributed manner. This eliminates the need for positional information, robot identifiers, data collection, and aggregation into a single representation, which is not feasible with large numbers of small robots.

Previous publications describe in detail the methods, based on the concept of virtual pheromones, for performing distributed computations across the robot swarm, including behaviors such as disperse, go-hide, wake-up, follow-gradient, follow-wall, and others [14,15]. In this paper, we address the interaction and interface issues of a large robot/sensor swarm by constructing interfaces that take advantage of the location, context, large numbers, and limited processing and communications of robot swarms. Each robot acts as a "pixel" in the construction of a larger visual display. Robots within a user's view transmit short messages that are decoded and presented using an optical see-thru head-worn display such that the information appears superimposed over the corresponding robots. Hence, we consider the interface to be "world-embedded."

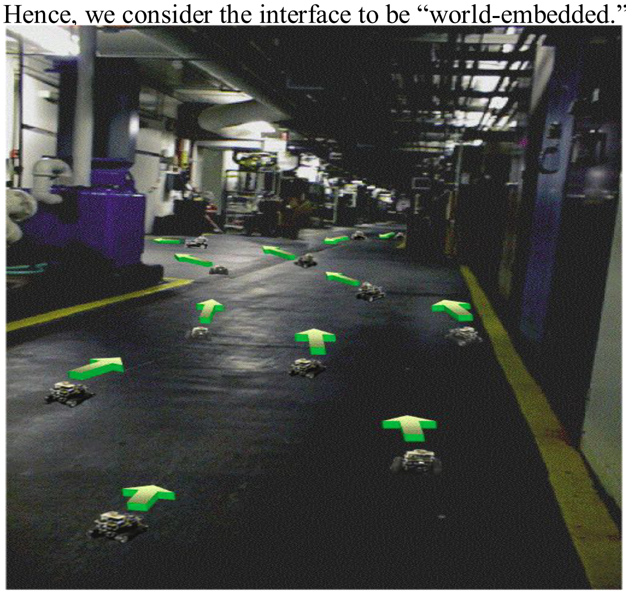

Figure 1. Shortest path to an intruder shown as a world embedded display.

Other research has addressed many issues in communication between individual robots and humans, however there is little work on methods of communicating between large numbers of robots and people. Brescia University Advanced Robotics discusses the use of robot swarms for mine detection including the use of odor sensors [3]. The University of Toronto proposes the use of a system called ARGOS (Augmented Reality through Graphic Overlays on Stereovideo) for communication with and control of telerobotic manipulators [10,11,12]. ARGOS provides virtual pointers for enhancing a user's depth judgement tasks, virtual tape measures for real-world measurements, virtual tethers for perceptual teleoperation enhancements, virtual landmarks for enhancing depth scaling, and virtual object overlays for on-object display superposition. Several other publications describe augmented reality systems for a variety of applications [1,2,4].

2. WORLD EMBEDDED INTERFACE

The world embedded interface provides a coherent information display from a collection of loosely coupled distributed display elements, called active fiducials, each typically mounted on a robot or sensor platform. Each fiducial both transmits information and provides a reference location where information should be presented to a user, augmenting the user's view of the world with computer generated graphics. A message is transmitted from each active fiducial and is received by a head-mounted camera worn by the user. The received message is decoded and then converted into a form that can be displayed as a graphical overlay within the user's field of view, positioned to appear coincident with the physical location of the active fiducial source. Below, we describe how several active fiducials can be made to work in unison such that each active fiducial acts as a single picture element of a much larger overall information display.

This approach makes it possible to display location-specific information from a collection of distributed sensors or mobile robots. Most conventional approaches to displaying information from multiple sensors require a map of the environment combined with known coordinates of each sensor. The data from each sensor can then be placed on a map display with corresponding information overlaid at appropriate map coordinates. In contrast, our approach works without the need for maps or sensor coordinates. By having each sensor or robot transmit its own local piece of information, and by displaying this information as a graphical overlay, suitably aligned on the physical world, it is possible to convey much of the same information that would otherwise require use of a map. Thus, the world-embedded display is well-suited to situations where no map is available, where it is not possible to maintain accurate position information for each robot/sensor, or where human situation awareness requires direct display of information in registration with the world.

Figure 2 illustrates how information can be received from robots and presented to a user. Robots equipped with special beacons signal with these beacons, encoding the local gradient vector and other information as spatial, temporal, or spectral patterns. The camera collects this

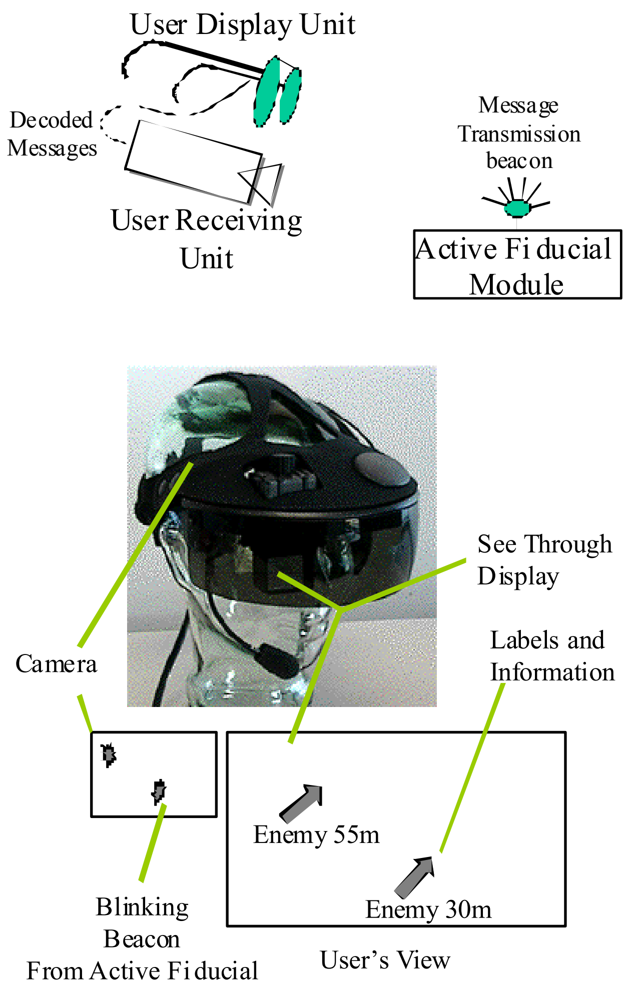

Figure 2. Active fiducials transmit messages that are decoded at the user.

information along with the 2-D location of each signal in the image plane. With proper alignment of the camera with the user's augmented reality display, positions in the camera's image plane map directly onto positions in the user's view plane.

3. IMPLEMENTATION

Active Fiducials

An important requirement for the world-embedded interface is to be able to depict gradient vectors that remain fixed with respect to the environment regardless of the user's viewpoint. For example, consider two users looking at a robot. If one user, standing in front of the robot, sees an arrow pointing to the right, then the other user, standing behind the robot, should see an arrow pointing to the left. For both users, the arrow will point in the same direction relative to their surroundings.

In order to transmit directional information of this type, the robot and user need to share a common reference frame. One way to do this is to use a compass. However, a compass is often ineffective in indoor environments, so we must have some other way of establishing a common reference frame. Another way is to design the beacons on each robot to transmit different messages in different directions. We do this by using a directional beacon system on each active fiducial, (see Figure 3). The directional beacon system consists of two or more directional beacons separated by baffles. A different message is transmitted from each directional beacon such that the message received by a user looking at the beacon from that direction is appropriate to his orientation. These directional messages are encoded such that the vector direction transmitted from the front of the robot is 180 degrees from the vector transmitted out the rear. Likewise, the vector direction transmitted out the side of the robot is 90 degrees from the direction transmitted out the front. This way, from whatever angle a user views a robot, the decoded gradient vector will always appear to be pointing the same way relative to the physical world. This approach has the advantage that it does not require user head position to be registered with some pre-determined reference frame. We are not seeking to retrieve data from a position-indexed database, instead we display the information transmitted by the active fiducial itself along the same line of sight that the fiducial beacon would ordinarily be seen. Consequently, multiple users can view a set of active fiducials from different directions, and each user will see a display that is appropriate for his own viewing angle.

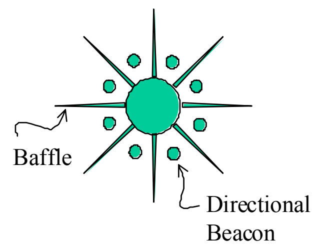

Figure 3. A directional beacon system

We implement the directional beacon system on our robots using an Augmented Reality Mast (ARM). The ARM provides directional information to the user's video camera via a ring of infrared LEDs that emit a 30 degree cone at an

Figure 4. The Augmented Reality Mast (ARM) projects information from the robots to a user's head-mounted display.

880nm wavelength. The LEDs blink coded signals that can be detected by the video camera. The ARM mounts on top of the robot (Figure 4a) and provides eight sectors of LEDs (Figure 4b). The ARM uses two LEDs per sector, which are stacked on top of each other. This is in order to increase brightness and enhance detection at different distances from the robots. In order to minimize potential interference from adjacent LEDs, we activate two alternating sets of LEDs in the ring. While one set of LEDs (eg. 0, 2, 4, 6) is sending messages, the other set (eg. 1, 3, 5, 7) remains off. We also cover each individual LED with a small cylindrical baffle to prevent unwanted illumination leakage.

The user wears an optical see-through AR Head-Mounted Display (HMD) with an AR Camera (ARC) mounted above the display (Figure 5). The ARC is a monochrome NTSC "lipstick" sized camera without the IR cutoff filter. This camera provides better detection of the IR LEDs than other camera units with the IR cutoff filter and extends the range of detection.

Figure 5: Augmented Reality head-mounted display.

IR Tracking/Decoding Algorithm

The ARC captures IR filtered images at the rate of 30 frames per second. The detection and tracking software detects bright spots from active fiducials in each image (see Figure 6). If the bright spot is new, then it is added to a message pool. Correspondence between subsequent bright

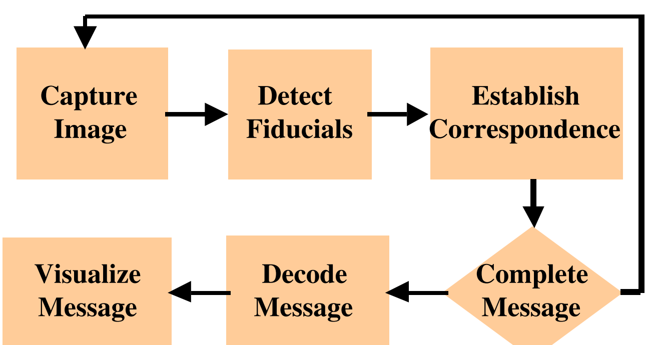

Figure 6. The structure of the AR tracking/decoding subsystem.

spots and previously detected messages in the pool is maintained regardless of user or robot motion. Each message is kept in an active state until the entire message has been received, at which time the message is decoded and used to index into a set of visual icons (e.g. gradient vectors) that will be placed in registration with the most recent location of the fiducial. Partial messages are discarded.

Since the user and robots are moving while the messages from the ARM are being sent, it is difficult to establish correspondence between the detected fiducials and fiducials in the message pool. Message *off* periods that correspond with image capture also increase the difficulty of tracking. To improve tracking performance, we linearly predict the potential locations of the active fiducials based on previous locations and determine correspondence based on the shortest distance in the image within a threshold.

Message Format

For this specific example, there are eight local gradient vectors, requiring a message format that can encode at least eight symbols in a short duration transmission. We are using code-39 barcode that has 9 pulses, 3 of which are wide pulses and 6 of which are narrow pulses, resulting in 44 symbols. Currently, to recover the original signal, the wide pulse uses 4 frame times (~132 ms) and the narrow pulse uses 2 frame times according to Nyquist's sampling theory. Since the ARC captures the video at 30 fps, it takes 24/30 seconds to transmit a complete barcode message. To separate messages, we use a six frame time pause between message transmissions.

We use only eight symbols out of the 44 possible symbols to approximate gradient vectors at 45 degree increments. We can use the rest of the symbols for different types of sensors, distance information, or other information. Since a pulse in Code-39 may either be bright or dark (beacon ON or OFF), wide pulses with the beacon OFF pose the greatest challenge for tracking. To simplify the tracking, we added a separate tracking beacon on the top of the ARM (Figure 7). With the tracking beacon constantly emitting a signal, software determines the robot location within every video frame. By adjusting the height of the tracking beacon with respect to the directional beacons, we are able to vary tracking performance. A three inch tall tracking beacon provides acceptable differentiation with an 8mm lens approximately two feet away. Figure 8 shows the effect of distance and mast height on differentiability between the tracking and blinking beacons. Mast heights of 3 inches are more reliably differentiated to a distance of over 20 feet.

Figure 7. Tracking Beacon

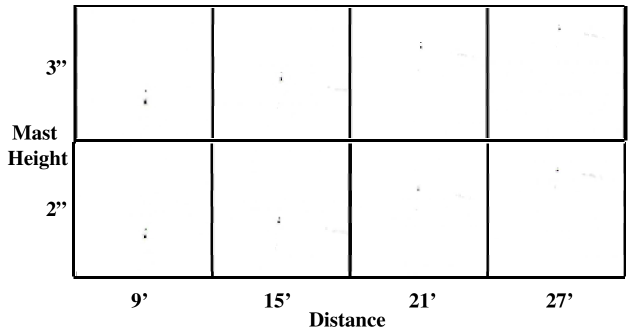

Figure 8: Separation of tracking beacon from blinking beacon with different mast heights for different camera distances from the beacon.

4. ROBOT SWARM RESULTS

We have demonstrated the use of the World Embedded Interface with a swarm of 20 Pheromone Robots (Pherobots). The gradient in the direction of a target computed by the swarm is superimposed on any of the robots whose tracking and message beacons are visible to the user’s head mounted camera. Figure 9 depicts select frames from a sequence recorded through the HMD. In this case, the user tasks the robot swarm by introducing a pheromone message into the swarm via a handheld PDA. The swarm disperses into the space, maintaining contact across the swarm. Once the target is detected, a virtual pheromone propagates through the swarm indicating the shortest path to the target. All of the beacons send coded, directionally specific gradient information that is decoded and displayed for those robots that are visible to the user.

5. FUTURE CONCEPTS AND RESEARCH

The capability described so far may be thought of as a *Local User Mode* (LUM) of the interface, which enables an operator to gain situation awareness by looking directly at the swarm. The LUM accurately superimposes computer-generated information on those robots that are within direct line of sight using the optical see-thru system described. We envision a wide range of new modes of interaction and visualization that will enable swarms of robots to provide qualitatively accurate information to human operators without the need for accurate position and maps, and for regions of space that are remote from the user. For example, the *Remote User Mode* (RUM) focuses on methods that depict robot state, sensor data, and gradients for occluded robots. The RUM might include topological features from the robot swarm for a space such as a building, including corners, t-junctions, corridors, and walls. Figure 10 shows a concept for visualizing topological information accurately registered with visible members of the swarm and qualitatively drawn for occluded robots.

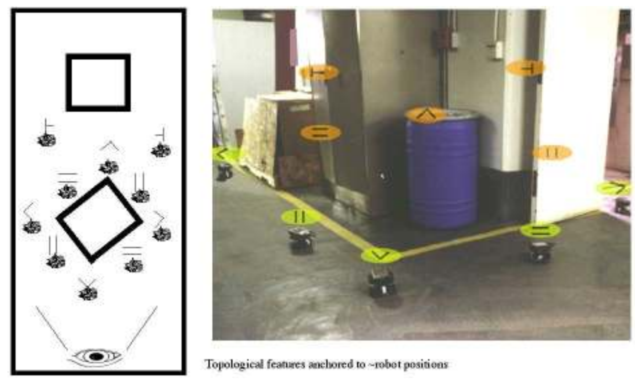

Figure 10. Topological features anchored to visible robots and qualitatively depicted for occluded robots.

By using gradients calculated by the swarm, we can linearly index into the swarm for the purpose of conveying situation information at specific locations as well as tasking portions of the swarm. We envision an animated *fly-through* along the gradient, providing users with a qualitative understanding of the path using very limited information about robot location and without robot IDs. Figure 11 depicts a sequence in an animation traveling along the gradient to a target where we construct a qualitative 3D representation and then fly along the gradient. Animated qualitative 2D representations are also possible, as are fly-throughs that implicitly reconstruct the space from robot sensor data such as imagery, audio, temperature, and movement. As the number and density of sensor equipped robots increases, this *telepresence* fly-through approaches a continuous animation along the gradient.

6. CONCLUSION

The user interface to our distributed robot swarm is itself distributed. Instead of communicating with each robot individually, the entire swarm works cooperatively to provide a unified display embedded in the environment. For example, robots that have dispersed throughout a

Figure 9. This sequence of frames, recorded through the optical see-thru display with a camera at the eyepoint, shows two new robots entering the field of view and moving in the direction of the gradient to an intruder (through the door top left).

building are able to guide a user toward an intruder by synchronizing to collectively blink in a marquee-style pattern to highlight the shortest path to the intruder. Using augmented reality, the robots can present information that is more complex. Users wearing a see-through head-mounted display and a head-mounted camera that detects and tracks encoded messages emanating from the robot beacons see a small amount of information superimposed over each robot. Each robot, is in effect, a pixel that paints information upon its local environment. The combination of our world-embedded interface with world-embedded, distributed computation directly maps information onto the world with no intermediate representations required.

REFERENCES

[1] Azuma, R., B. Hoff, H. Neely III, R. Sarfaty. “A Motion-Stabilized Outdoor Augmented Reality System,” Proceedings of IEEE VR '99, Houston, TX, pp. 252-259, March 1999.

[2] Azuma, R., Y. Baillot, R. Behringer, S. Feiner, S. Julier, B. MacIntyre. "Recent Advances in Augmented Reality," IEEE Computer Graphics and Applications vol. 21, #6, pp. 34-47, Nov/Dec 2001.

[3] Cassinis, R. “Landmines Detection Methods Using Swarms of Simple Robots”, Proceedings of The 6th International Conference on Intelligent Autonomous Systems (IAS2000), Venice, Italy, July 25-27, 2000.

[4] Cho, Y., "Multi-ring Fiducial Systems for Scalable Fiducial-Tracking Augmented Reality," Presence: TeleOperators and Virtual Environments, December, 2001.

[5] Flynn, A.M. "Gnat Robots (And How They Will Change Robotics)," In Proceedings of the IEEE Microrobots and Teleoperators Workshop, Hyannis, MA, November 1987. Also appeared in AI Expert, December 1987, pp. 34 et seq.

[6] Gage, D.W. "Command Control for Many-Robot Systems," In The Nineteenth Annual AUVS Technical Symposium (AUVS-91), Huntsville AL, 22-24, June 1992. Reprinted in Unmanned Systems Magazine, 10(4):28-34.

[7] Gage, D.W. "Sensor Abstractions to Support Many-Robot Systems," In Proceedings of SPIE Mobile Robots VII, Boston, MA, Vol. 1831, pp. 235-246, 18-20, 1992.

[8] Gage, D.W. "How to Communicate with Zillions of Robots," In Proceedings of SPIE Mobile Robots VIII, Boston, MA, Vol. 2058, pp. 250-257, 9-10, 1993.

[9] Lewis, M.A., and Bekey, G.A. “The Behavioral Self-Organization of Nanorobots Using Local Rules,” In Proceedings of the 1992 IEEE/RSJ International Conference on Intelligent Robots and Systems, Raleigh, NC, July 7-10, 1992

[10] Milgram, P., S. Zhai, D. Drascic and J.J. Grodski, "Applications of augmented reality for humanrobot communication", Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots & Systems (IROS), Yokohama, 1993.

[11] Milgram, P., Z. Shumin, D. Drascic “Applications of Augmented Reality for Human-Robot Communication”, In Proceedings of the 1993 International Conference on Intelligent Robots and Systems, Yokohama, Japan, July 26-30, 1993.

[12] Milgram P, Rastogi A, Grodski JJ: "Telerobotic control using augmented reality". Proc. 4th IEEE Intn'l Workshop Robot & Human Communication (Ro-Man'95), Tokyo, pp. 21-29, July 1995.

[13] Payton, D.W. “Internalized Plans: A Representation for Action Resources, in Designing Autonomous Agents,” ed. Pattie Maes, MIT Press, Cambridge, Mass, pp. 89-103, 1990.

[14] Payton, D., M. Daily, R. Estowski, M. Howard, C. Lee, "Pheromone Robots", in Autonomous Robots, Kluwer Academic Publishers, Boston, MA, 9(8), 2001.

[15] Payton, D., M. Daily, B. Hoff, M. Howard, C. Lee, "Autonomy-Oriented Computation In Pheromone Robots", in Workshop on Autonomy Oriented Computing, Autonomous Agents 2001 Conference, Montreal, Canada, May, 2001.

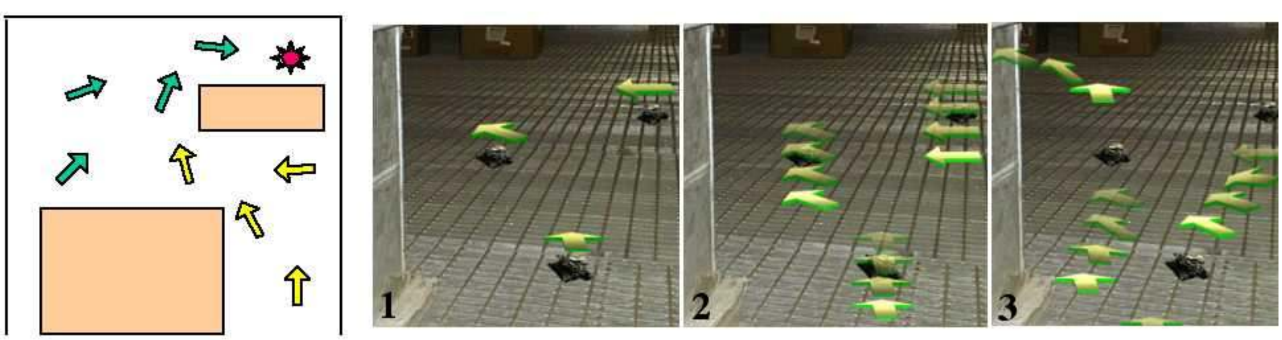

Figure 11. A conceptual animation of a sequence of frames in a 3D fly through along the gradient.