

Programming CHIP for the IJCAI-95 Robot Competition

R. James Firby, Peter N. Prokopowicz, Michael J. Swain,
Roger E. Kahn, and David Franklin

December 8, 1995

The University of Chicago's robot, CHIP, is part of the Animate Agent Project, aimed at understanding the software architecture and knowledge representations needed to build a general-purpose robotic assistant. CHIP's strategy for the Office Cleanup event of the 1995 Robot Competition and Exhibition was to scan an entire area systematically and, as collectible objects were identified, pick them up and deposit them in the nearest appropriate receptacle. This article describes CHIP and its various systems and the ways in which these elements combined to produce an effective entry to the robot competition.

The University of Chicago's robot, CHIP, is part of the Animate Agent Project aimed at understanding the software architecture and knowledge representations needed to build a general purpose robotic assistant [7]. The Animate Agent Project defines two primary levels of knowledge representation: the skill level consisting of modular processes that can be configured in a variety of ways, and the reactive task execution level consisting of a goal interpreter and a library of Reactive Action Packages, or reactive plans, that describe ways of achieving system goals.

One of the primary goals of the Animate Agent Project is to develop a vocabulary of skills and plans that can be used to achieve a wide variety of everyday tasks. Thus, while addressing the office cleanup task, we are attempting to build up "generic" skills and RAPs that can be reused as building blocks in RAPs for future tasks.

CHIP went into the clean-up competition with a general-purpose strategy for cleaning up a room. The strategy, broadly, was to scan an entire area systematically, and as collectible objects were identified, pick them up and deposit them in the nearest appropriate receptacle. This plan for cleaning up a room is built up from a collection of more general purpose operations, such as 'find-object-type', 'move-to-target', and 'pick-up-target'. In this sense, only CHIP's highest level goals were programmed specifically for the IJCAI event; the rest are generic and are the building blocks for CHIP's future behavior.

1. CHIP the Robot

CHIP is the working robot at the Animate Agent Laboratory at the University of Chicago. As illustrated in Figure 1, CHIP is built on a Real World Interface three-wheeled synchro-drive mobile base. This base is surrounded by an octagonal bumper and supports a body roughly three feet high. Around the middle of the body is a ring of eight sonar sensors and on top is a pair of color cameras mounted on computer-controlled pan-tilt platform. Mounted in front of CHIP's body where it can reach the floor, is a Heathkit Hero robot arm that has been augmented with force and contact sensors to get tactile feedback from the gripper. Onboard, CHIP carries a 68000 computer to manage the sensors and control the arm, and a 68030 computer to run the action skills discussed below. Video from CHIP's cameras is broadcast offboard to a DataCube image processing system attached to a Sparc-20 workstation, where all vision processing is done.

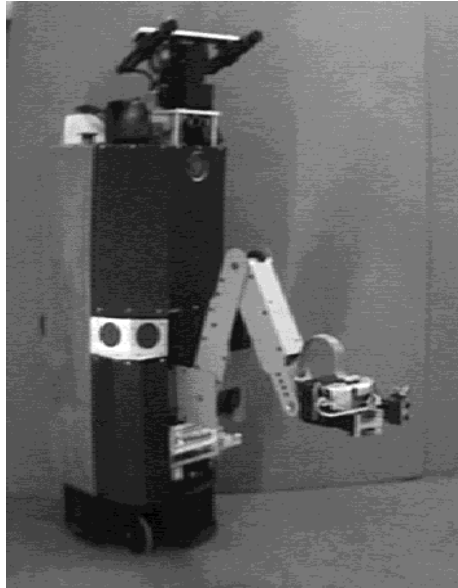


Figure 1: CHIP the robot

The software used for CHIP consists of a variety of connected systems. The low-level motor and sensor processing software is written in 'C' and runs onboard. The concurrent, perceptual-motor skills that make up CHIP's modular control system are written in CRL [4] with some running onboard the robot and others running offboard on a Macintosh computer. The onboard and offboard components are linked via a radio-ethernet. The RAP system [6, 5, 3] for task-level sequencing of the modular skills is written in LISP and runs offboard on the Macintosh where it interfaces to the CRL skill system. Vision processing [9] is done offboard on the Sparc-20 and communicates with other skills and the RAP system via ethernet. The vision system, the perceptual-motor skills, and the RAP task execution system make up the Animate Agent Architecture for intelligent, real-time robot control shown in Figure 2.

2. The Animate Agent Architecture

The RAP system is designed to deal with achieving goals in a dynamic environment. Each RAP task description encodes a set of methods for carrying out the task in different situations, a success check to tell when the task has accomplished its purpose, and notations that describe when things are not going as expected. At run time, a RAP task examines its methods and selects one that is appropriate in the current situation. By doing method selection at run time, RAPs are more likely to select the best method even if the world is changing or contains details that cannot be predicted in advance. RAP methods also can include other RAP tasks as subgoals, and the resulting hierarchy helps the system to cope with complexity by specializing actions through many different levels of detail. RAP methods also include annotations that specify assumed constraints and expected subtask results. These annotations are monitored during the execution of the method and cause the method to fail when they are not met. Monitoring expectations helps the system recognize early that a selected method has become invalid due to changing circumstances or unexpected events. To further increase robustness, a RAP task includes a success test to confirm that a method has succeeded in achieving its goal. If the success test is false after a method has completed, the RAP task continues to run, selecting alternative methods until the success test is satisfied or there are no methods to try. This tenacious execution allows a task to recover when the information it uses to select a method is incorrect, when actions don't have their intended effects, or when the situation changes while a method is executing.

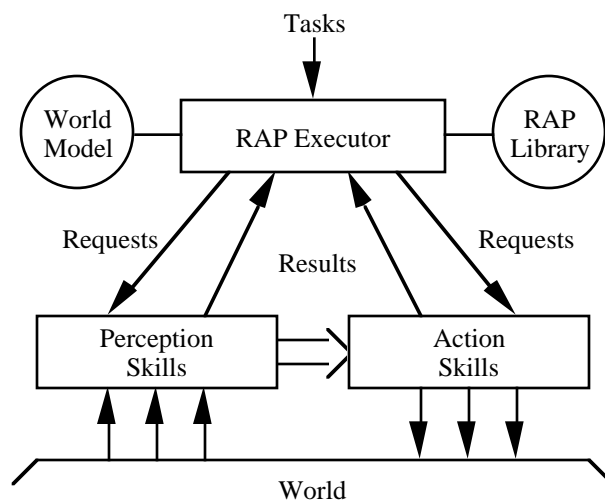


Figure 2: The Animate Agent Architecture

The skill system is designed to deal with soft real-time control of the robot's actuators. The skill system is a network of sensing and action processes, defined by separate programs, that can be enabled or disabled at will. When a skill is enabled it receives a set of parameters and begins to run. Running skills can access current sensor values, control actuator set-points, and communicate with one another through global channels. The actual be-

havior of the robot at any given moment is the result of the enabled skills acting on input from the environment. Thus, the behavior of the robot (i.e., its apparent immediate goal) can be changed by changing the set of active processes.

The RAP system refines tasks into commands that “program” the skill system. These commands enable skills and connect them into control networks to carry out activities in the situations encountered at run time. Goal-directed behavior results from task refinements that generate a sequence of different skill system configurations.

3. Basic Skills

The basic skills used by CHIP during the competition can be broken down into perception skills for extracting visual information from the world, and action skills for effecting changes in the world. While CHIP is carrying out a given activity, a combination of these skills will typically be running concurrently and working together to control the robot.

3.1 Perception

One of the key features of the architecture that makes the trash cleanup task possible is the ability to easily use different sensing strategies for different tasks. This is particularly important when building a system that uses vision for a wide variety of tasks because different tasks will typically require visual data to be processed in different ways. For the office cleanup task CHIP used three different vision processing algorithms: an algorithm for finding and classifying small objects on the floor, an algorithm for tracking a selected small object, and an algorithm for finding objects against more complex backgrounds.

3.1.1 Finding and Tracking Trash

CHIP can recognize a broad class of small objects when seen against a relatively clean background. The algorithm for this process has four steps. First, an edge operator is run over the visual scene to create an edge image and that image is segmented into regions that may be possible objects (see Figure 3). Second, the size, aspect ratio, edge density, average color, fraction of white, and contour regularity is computed for each possible object. Third, the resulting feature vector for each object is classified against a set of fuzzy exemplars by choosing the nearest neighbor within a maximal distance. We entered the competition with exemplars for generic cans that could handle essentially any drink can, cups (either white or colored) and balls of paper - a vestige of the 1994 competition. Finally, the relative location of each object with respect to the robot is calculated using the camera height, orientation with respect to CHIP’s body, and the location of the object in the image. The end result is that CHIP knows the type and location of all of the pieces of trash in the current scene.



Figure 3: Trash: Raw Image, Edge Image, Filtered Image

However, the location of each piece of trash found above is only approximate. Calibration errors and low visual resolution combine to make a trash location more uncertain the farther away it is. Furthermore, as CHIP moves toward a piece of trash, actuator error has the effect of increasing the uncertainty of that piece's location. To compensate for these errors, CHIP must track the piece of trash it is approaching while it is moving.

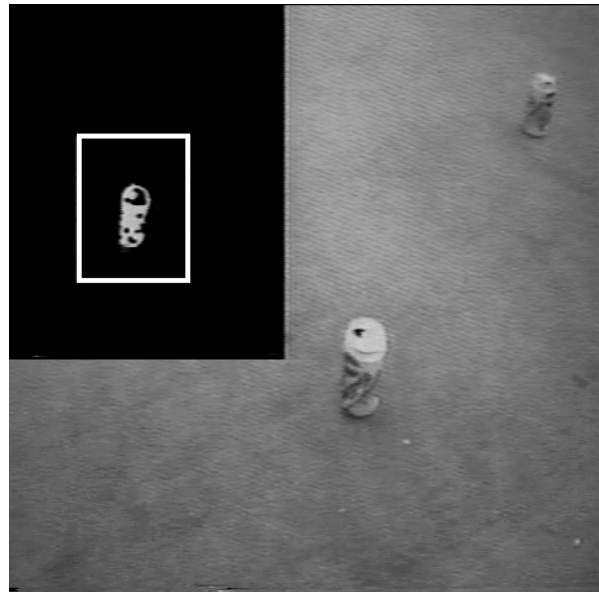


Figure 4: Tracking Trash: Target and Next Search Area

The visual process for tracking a piece of trash is simpler than for identifying trash because we want the process to run as fast as possible. The trash tracker uses two heuristics to enable it to do less processing and run faster. First, CHIP knows where the trash is to start with so the tracking algorithm only looks at a small portion of the image around that area (see Figure 4). Second, the trash tracker only performs the first step of the trash finding algorithm; computing the edge image over the small region and segmenting out a single object region. The region is assumed to correspond to the piece of trash being tracked and the tracker returns its location. This process is repeated as fast as possible so that as CHIP moves, the piece of trash stays within the small search region of the tracker. If the tracker fails to find a possible segmented object in the small search region, it doubles the size of the search region and

tries again, repeating the process until it finds an object or it has searched the whole scene.

3.1.2 Finding Trash Cans

During the competition, trash cans could not easily be segmented from the background using edge data alone because they were along the walls in the office near corners, furniture, and other objects. To find and recognize objects that are not easily separated from the background, CHIP used a two step algorithm illustrated in Figure 5. First, the scene was examined using color histogram back-projection [10] to find areas containing the same colors as the object. An edge operator was then applied to these areas followed by a soft edge-template matching algorithm that includes adjustments for viewpoint transform. This algorithm uses the Hausdorff distance [8] to compare different possible template matches, and the best match gives the exact location, size, and pose of the object. For the competition, we used the distinctive colors of the trash can and recycling bin labels to locate possible cans, and a set of six known views of the object to match against (see Figure 6). Unfortunately, even using fast color-backprojection to reduce the area searched by the Hausdorff template matcher, this algorithm took several tens of seconds.

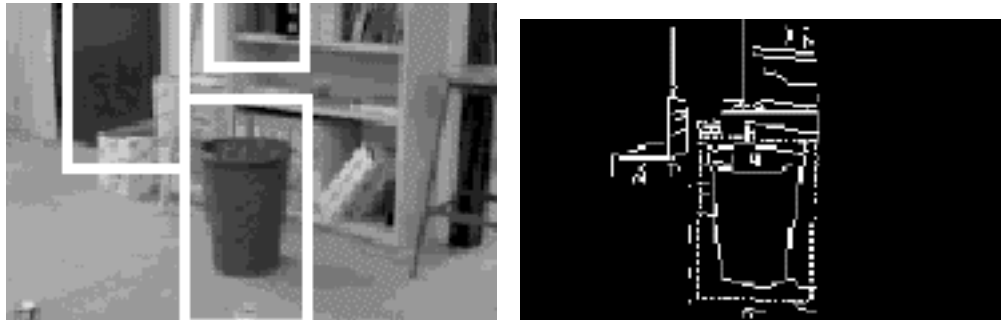


Figure 5: Trash Cans: Original scene with color areas boxed, Edges for areas of interest and Hausdorff match.

The same problems of low resolution and calibration error arise when finding trash cans as when finding trash. Trash can locations from the template matching routine are only approximate and must be refined as CHIP gets closer and lines up to drop in a piece of trash. For this process, the same template matching skill is used but at a much lower resolution. Starting with a lower resolution image speeds up the matching process significantly but prevents CHIP from seeing the trash can unless it is very close at hand.



Figure 6: Trash Can Hausdorff Models

3.2 Action

The basic skills CHIP uses for taking action fall into three categories: skills for moving CHIP's arm, skills for moving CHIP from one place to another, and skills for moving the pan/tilt head that carries CHIP's cameras.

3.2.1 Manipulating Objects

The skills used to control CHIP's arm include a skill to move the wrist to a given point in front of CHIP, a skill to tilt and twist the wrist, and several skills to open and close the gripper. Multiple gripper skills are used because of the variety of different stopping conditions required in different situations. For example, a generic skill for setting the gripper width to a specific value is used to open the gripper in preparation for picking up a piece of trash. This skill will open or close the gripper to get to the desired width and will signal when that goal is reached or when the gripper gets stuck and can't move. Another skill is used to grasp objects by closing the gripper until a specific pressure is generated. A different skill is used to let go of an object by opening the gripper until either its finger sensors signal the gripper is empty, the gripper is open as far as it will go, or the gripper gets stuck and can't move. A fourth skill is used to close the gripper until the fingers touch so that the arm can be tucked away while CHIP is moving with its hand empty. The skills associated with CHIP's arm do not currently use visual feedback or account for obstacles in the arm's workspace.

CHIP's arm can only move in a vertical plane directed forward, so CHIP must be lined up very precisely with objects on the floor to be able to reach them. This requirement is met by a special skill for aligning the center of CHIP's body with a target location at a precise stand-off distance. The skill signals when the alignment is good, or when the robot cannot move without hitting something. This skill relies on other skills to generate the target location and it can be used to align with a variety of objects or even a dead reckoned location. We use the trash tracking skill to align with a piece of trash and the low resolution Hausdorff template matching skill to align with a trash can.

Once CHIP is aligned with a piece of trash, the arm is down near the ground, and the gripper is open, another skill is used to move CHIP straight forward until the gripper detects an object between its fingers. This skill signals completion when an object is detected in the gripper, after a certain distance has been traversed without detecting an object, or when the fingers bump into something.

3.2.2 Moving to Locations

CHIP uses two basic skills for moving from place to place: one to turn to a particular heading and one to move to a particular location. The turn skill turns CHIP in the appropriate direction and signals completion when CHIP is facing the right way. It also monitors the sonar sensors and will move CHIP backward, if possible, to avoid bumping the arm while turning. If CHIP cannot move without bumping the arm or backing into something, the skill signals that CHIP is stuck. The move skill drives CHIP towards the goal location, monitors sonar values, and uses a simple potential field obstacle avoidance strategy as it goes. This skill signals when CHIP reaches the goal or when the obstacle avoidance strategy gets stuck. Like the aligning skill, both turning and moving rely on other skills to supply the target for their actions. That way, different target generating skills can be used to get CHIP to perform different actions at different times. For example, a target generated via dead reckoning can be used to move CHIP to a particular location in space, or a visual tracking skill can be used to get CHIP to approach a specific object and follow it if it moves.

3.2.3 Moving the Head

Skills for moving the pan/tilt head come in both feedback and non-feedback driven varieties. There is a skill for moving the head to a specified pan, which signals when the movement is complete, and a skill for panning the head to a goal direction generated by another skill. Combining this latter skill with a visual tracking skill will keep the head pointed toward a moving target. There are similar skills for controlling tilt.

4. Map Building and Object Memory

The basic algorithm used by CHIP for picking up trash is to look for a piece of trash, pick it up, look for a trash can, drop off the trash, and repeat. However, we also want CHIP to remember the locations of trash and trash cans, and to be able to tell when all of the trash is gone. In a small office like the one used in the competition, people don't need to remember where trash is because they can quickly scan the whole floor to see if there is any around. The same scan will also tell them when the task is done because the whole floor will be clean. However, CHIP is very near-sighted and cannot accurately identify a piece of trash more than 8 feet away. Therefore, CHIP must rely on memory to know where trash has already been seen and which areas of the office have been cleaned up.

4.1 Mapping Areas Searched

To remember where CHIP has looked for trash and trash cans, a binary grid in global coordinates is maintained for both trash and cans. Each time CHIP uses the visual skill for finding trash, the wedge of floor space imaged by the skill is recorded in the grid by setting the points covered by the image. The grid is then used by CHIP to search the office *systematically* for trash. When CHIP knows of no piece of trash on the floor, it consults the grid to find the nearest area that has not been examined yet. CHIP then moves to bring this area into view and execute the trash finding visual skill. The visible area is updated in the grid and the next time CHIP is searching for trash a different, unexplored area will be selected to look at. Simple heuristics are used to prefer to explore areas that can be seen by turning the pan/tilt head, then those that can be explored by turning the robot, and finally those that require moving to a new vantage point. A similar algorithm is used for keeping track of where trash cans have been sought. The grid showing where CHIP looked for trash during the final event of the competition is shown in Figure 7.

4.2 Mapping Accessible Areas

In addition to tracking where visual skills have looked, CHIP uses a simple occupancy grid technique [1, 2] to map office obstacles using sonar data. This mapping happens passively, in parallel with the basic task of cleaning the office and results in the construction of a map of reachable areas. This map is used to ensure that all vantage points chosen by the algorithm for looking for trash and trash cans are actually accessible. Furthermore, because the office environment was sealed by walls and closed doorways, once CHIP had checked every accessible area for trash, it knew that no more trash existed in the office and the task was complete. During the competition CHIP further constrained where it needed to look for trash using the knowledge that the office held four trash cans in the four corners. Thus, once three trash cans were found, a rectangle enclosing those cans demarked the area enclosed by the office. This heuristic also helped to tell CHIP when to stop cleaning up even if a door was open and there was accessible space outside of the office area. The occupancy grid generated during the final event of the competition is shown in Figure 8.

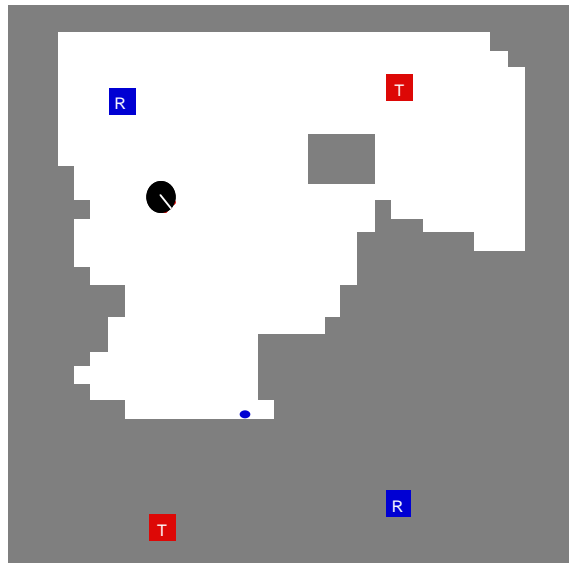


Figure 7: Areas Searched for Trash During Competition. The trash cans and recycling bins are marked as squares. A known can is also shown as a dot.

4.3 Remembering Item Locations

Of course, along with remembering those areas that have been explored, CHIP also remembers where it has seen trash and trash cans. The locations of these items are recorded in a simple database that supports looking up the closest item of a particular type. CHIP then uses the heuristic of picking up trash it knows about before exploring new areas.

Unlike pieces of trash, CHIP can see trash cans all the way across the office and knew that there were exactly two trash cans and two recycling bins. This information led to the heuristic that CHIP use a known trash can if it was facing towards one after picking up a piece of trash, or look around first if it was facing away from a known can and some trash cans had not yet been found. The result of this heuristic was for CHIP to find all four trash cans early and then always move to the closest can of the appropriate type.

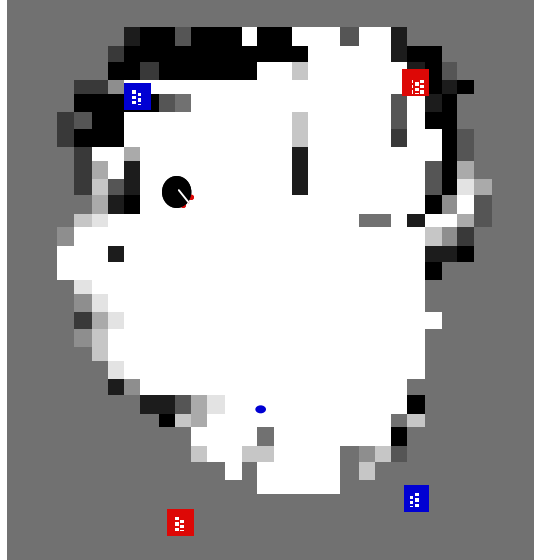


Figure 8: Occupancy Grid Generated During Competition

The locations of trash pieces and trash cans were recorded in the database as a direct result of using the corresponding visual routines.

5. Putting it All Together

The skills and mapping processes described above are not enough to do the trash cleanup task by themselves. The RAP task execution system is used to decide what action CHIP should perform next and hence chooses the sequence of active skills that actual control CHIP's activities.

The RAP library used for the cleanup task is organized as much as possible around reusable subtasks that can be used in future tasks as well. These subtasks include moving the robot from place to place, searching for items in the world, manipulating objects, and doing the trash cleanup task itself.

5.1 Moving to a New Location

The basic RAP plans for moving CHIP to a given goal location consist of sequences of the turning, moving, and tracking skills described above. These plans are variations on the sequence:

1. Enable a skill to "track" the goal location using odometry and dead-reckoning. Use this skill as the target generator for the skills to follow.
2. If CHIP is not facing within 30° of the goal location, enable the turn skill to turn CHIP in roughly the right direction. Signal completion when facing the right way.
3. Enable the skill to move CHIP toward the goal location. Signal completion when CHIP is close to the goal.
4. If CHIP is not facing toward the goal, enable the turn skill again to turn CHIP in the right direction. Signal completion when facing the right way.

5. Disable the dead-reckoning skill generating the target location for the motion skills.

For the office cleanup task, all goal locations are either object locations from the item database, or vantage points for areas yet to explore. However, in other situations, different goal tracking skills can be used in step one to get CHIP to follow a visual target or to approach a moving object.

These RAP methods can fail if the goal location cannot be reached. In particular, when moving to an unexplored area, CHIP might encounter new obstacles that prevent achieving the goal. The movement skill recognizes this condition by monitoring progress toward to goal and signaling when progress is not being made. The RAPs used in the competition simply fail in this case.

5.2 Searching for Trash and Trash Cans

One of the key aspects of the office cleanup task is the need to search the world for trash and trash cans. This behavior is encoded on CHIP as a collection of RAPs that use the spatial memory described above to execute the basic plan:

1. Select an unexplored area nearby using the memory of areas already explored looking for this object type
2. Use the movement subtasks described above to move to a good vantage point for viewing the area.
3. Execute the appropriate visual skill for finding the desired item and continue when the skill completes.
4. If an item of the right type was added to memory, succeed. Else repeat.

Steps one and four in this plan are actions on memory, step two is an instantiation of a RAP subtask, and step three enables a visual skill and waits for its completion.

This basic method can fail if there are no unexplored, accessible areas for the object, or if the selected area cannot be reached. In the first case, the searching RAP simply reports failure because there is nowhere left to look. In the second case, the searching RAP catches the movement failure and simply repeats. The mapping processes will have detected the obstacles blocking progress and the next iteration of the RAP will select a new area that is accessible based on this new information as well.

5.3 Picking Up and Dropping Off Trash

Once a piece of trash, or a trash can, has been found and approached, the next step is to pick up the trash or drop it off. The sequence of activities required to pick up a nearby piece of trash are:

1. Begin tracking the piece of trash with the visual trash tracker, and use output of that skill as the goal for aligning the robot near enough for a good grasp.

2. Lower the arm near the ground keeping the wrist horizontal and open the gripper as wide as possible.
3. Move forward until the object comes between the fingers of the gripper.
4. Close the gripper until an appropriate pressure is reached.
5. Move the arm up into the position for holding the trash while moving.

The first step is accomplished with a collection of concurrent skills for tracking the object and orienting the robot. The next step is really three steps done in parallel by activating concurrent skills for moving the arm, the wrist, and the gripper. The third, fourth, and fifth steps each correspond to activating individual skills and waiting for them to signal completion or some problem. As with the movement RAPs, this plan is parameterized by object type and step one can select different tracking skills for different types of objects. Therefore, the plan can be used for a variety of objects besides trash.

Possible problems with this basic plan are that the trash tracker might lose the target object, moving forward might miss placing the object in the gripper, or closing the gripper might squeeze the object out from between the fingers. In the first case, losing track of the object probably means that someone snatched it away, so the object is removed from the memory database and the RAP fails. The second case may also be the result of someone taking the object, or it may be an alignment error. If someone took the object it should be removed from memory, but if its an alignment error, CHIP should try to pick it up again. At the competition, we treated both of these cases as if the object had vanished and removed it from the database. We also did the same thing in the third case when the gripper closes without actually getting the object. By treating all of these failures as the same, CHIP will “forget” about problem pieces of trash and never try to pick them up again. What should happen is that the area of the world in which a grasp failure occurs should be reexamined after the failure to see if the piece of trash in question is still there.

A similar sequence of RAP steps is used to drop the trash into the trash can once the can is nearby:

1. Begin tracking the trash can with the low resolution template tracker, and use output of that skill as the goal for aligning the robot.
2. Raise the arm high enough to clear the trash can keeping the wrist horizontal.
3. Move forward enough to remove the offset between the hand and the trash can put there by the aligning skill.
4. Open the gripper until nothing is between the fingers.
5. Back up the same distance moved forward.
6. Move the arm and wrist down and close the gripper to get things ready for looking around again.

Again, each of these steps corresponds to a set of one or more skills being enabled to actually control the activity.

This plan is subject to the same potential tracking failure as the pickup plan while aligning to the trash can. If that happens, the trash can in question is assumed to have moved, or more likely, to have been something incorrectly recognized as a trash can. It is removed from the memory database and the method fails. Another possible error is that the piece of trash being held does not drop out of the gripper when it is opened. When that happens, CHIP is really in trouble and simply asks for help.

5.4 The Complete Cleanup Task

The top level RAP methods for the cleanup task define the plan:

1. Select a piece of trash. If one is known in the database use it, else search.
2. Go to the trash.
3. Pick it up.
4. Select an appropriate trash can. If a good one is known, use it, else search.
5. Go to the trash can.
6. Drop off the trash.
7. Repeat.

All of these steps are parameterized by object type and correspond to the instantiation of RAP tasks already described. In particular, going to a piece of trash, a trash can, or a search vantage point, all instantiate the same RAP methods. Selecting and searching for trash and trash cans also use the same underlying methods.

This basic plan can have a number of problems: there may be no pieces of trash to be found, a piece of trash may be unreachable, pickup might fail, the trash might be dropped while moving, or there may be no accessible trash can. The first problem is caught by this plan as a failure of the first step, and it means the cleanup task is complete. The second problem is dealt with by removing the piece of trash from the memory database and going back to step one. If pickup fails, or the piece of trash is dropped, the plan simply restarts at step one. If no trash cans can be found, then the plan fails and so does the cleanup task itself.

6. Summary

This task required implementing algorithms for robot navigation, simple object manipulation, and visual sensing for finding, identifying, and tracking objects on the floor. It also required integrating these pieces into coherent plans to systematically pickup all the trash in the office and to deal with contingencies like missing a grasp, dropping something being carried, or seeing ghost trash or trash cans because of sensing errors. The resulting implementation used 34 skills and 80 RAPs. Of these, only three top-level RAPs were specific to the cleanup task — we believe most of the RAPs and skills are reusable in other tasks because they encode relatively generic actions like finding an

object, picking an object up, and moving to a new location, while coping with changes and failures in a real-world environment.

References

- [1] Johann Borenstein and Yoram Koren. The vector field histogram — fast obstacle avoidance for mobile robots. *IEEE Transactions on Robotics and Automation*, 7(3):278–288, June 1991.
- [2] Alberto Elfes. Using occupancy grids for mobile robot perception and navigation. *IEEE Computer*, pages 46–57, June 1989.
- [3] R. James Firby. Adaptive execution in complex dynamic worlds. Technical Report YALEU/CSD/RR #672, Computer Science Department, Yale University, January 1989.
- [4] R. James Firby. The crl manual. Animate Agent Project Working Note AAP-3 Version 1, University of Chicago, April 1994.
- [5] R. James Firby. Task networks for controlling continuous processes. In *Second International Conference on AI Planning Systems*, pages 49–54, Chicago, IL, June 1994.
- [6] R. James Firby. The rap language manual. Animate Agent Project Working Note AAP-6 Version 1, University of Chicago, March 1995.
- [7] R. James Firby, Roger E. Kahn, Peter N. Prokopowicz, and Michael J. Swain. An architecture for vision and action. In *Fourteenth International Joint Conference on Artificial Intelligence*, pages 72–79, Montreal Canada, August 1995.
- [8] D. P. Huttenlocher and W. J. Rucklidge. A multi-resolution technique for comparing images using the hausdorff distance. Technical Report CUCS TR #92-1321, Cornell University Department of Computer Science, 1992.
- [9] Roger E. Kahn and Michael J. Swain. Understanding people pointing: The perseus system. In *International Symposium on Computer Vision*, pages 569–574, November 1995.
- [10] Michael J. Swain and Dana H. Ballard. Color indexing. *International Journal of Computer Vision*, 7:11–32, 1991.