Grounding imitative behavior

Paul Fitzpatrick[⋆]
[⋆]MIT AI Lab
Cambridge, Massachusetts, USA
paulfitz@ai.mit.edu

Giorgio Metta[⋆†]
[†]Lira Lab, DIST, Uiversity of Genova
Genova, Italy
pasa@dist.unige.it

Abstract

Foo.

1 Introduction

We begin like this...

And continue by breaking up our sections under sub-headings. We can refer to papers parenthetically (**?**), or by direct reference to the work, as in the next sentence.

? present some ideas[1].

1.1 A Subsection heading

The sub-headings themselves can be distributed under sub-sub-headings.

1.1.1 And one subsubsection heading

This would be
a
figure if I had
anything interesting to
put here.

Figure 1: An example figure

1.1.2 Followed by another sub-headings

Table 1: This is an example table

Again, I don't really
have anything to put in here.

An interesting question then is whether the system could extract useful information from seeing an object manipulated by someone else. In the case of poking, the robot needs to be able to estimate the moment of contact and to track the arm sufficiently well to distinguish it from the object being poked. We are interested in how the robot might learn to do this. One approach is to chain outwards from an object the robot has poked. If someone else moves the object, we can reverse the logic used in poking – where the motion of the manipulator identified the object – and identify a foreign manipulator through its effect on the object. The next experiment was designed to explore this aspect.

The first obvious thing the robot can do is to identify the action just observed with respect to its motor vocabulary. It is easily done, in this case, by comparing the displacement of the object with the four possible actions and by choosing the action whose effects are closer to the observed displacement. Indeed it allows – even if in this limited setting – recognizing a complex action by interpreting its consequences on the environment. This is orders of magnitude simpler than trying to completely characterize the action in terms of the observed kinematics of the movement. Here, the complexity of the data we need to obtain from the observations is somehow proportional to the complexity of the goal rather than that of the structure/skills of the foreign manipulator. In our case, because the action, the goal, and the object are relatively simple, the only information required is about the displacement of the object.

Therefore, the next question is whether we can use this "understanding" of observed actions to implement mimicry behavior. It would be easy now to try to replicate the action just observed if the same object were presented again. However, there is still a bit of ambiguity in that we can choose to mimic either the observed displacement of the object or the way the object was poked with respect to its rolling affordance.

We chose to implement the latter. It is clear that poking along a particular observed direction requires trivial modifications. In practice, after an action is observed the angle between the affordance (see table **??**) and the actual displacement is measured and stored. If it happens to see the same object again, the robot chooses the action that has the greatest probability of poking the object along the previously stored angle. Figures 2 and 3 show examples of such mimicry.

This response is exactly what we would expect from a

[1]And this is what a footnote looks like

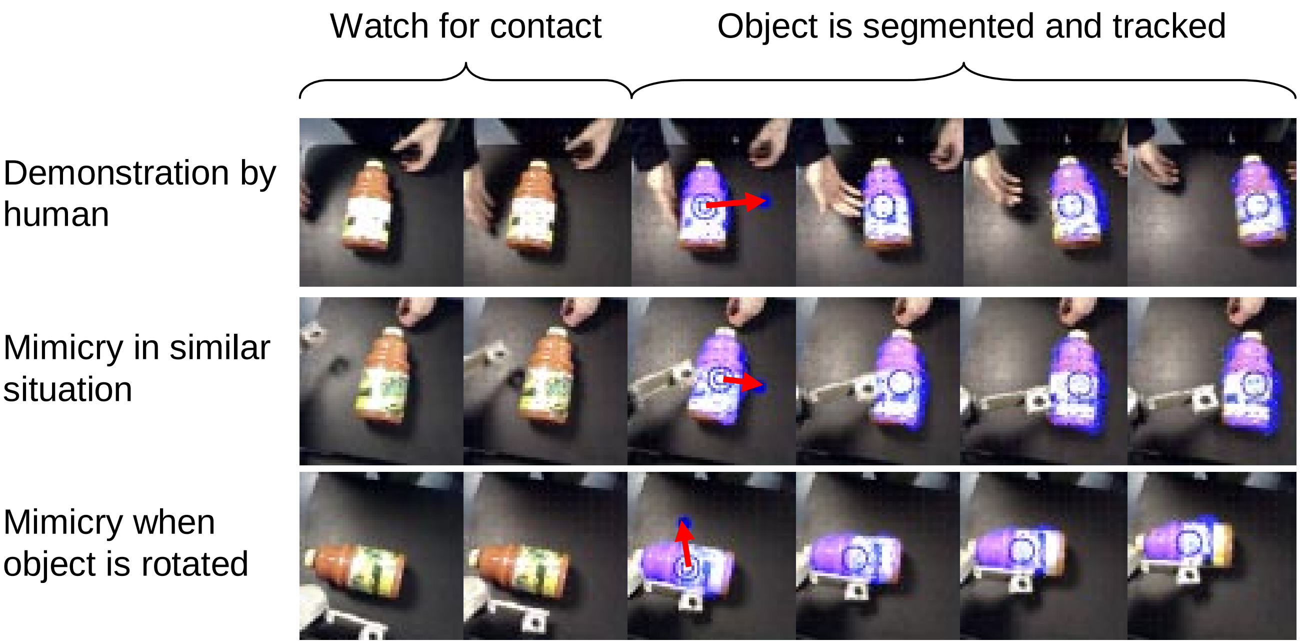

Figure 2: Basic mimicry. The first step in mimicking an action is to actually be able to observe it. The first sequence shows a human demonstration of a poking operation. Frames around the moment of contact are shown. The object, after segmentation, is tracked for 12 frames using a combination of template matching and optic flow. The big circles represent the tracked position of the bottle in successive frames. The arrow displayed on the frame of contact (3^{rd} from the left) projects from the position at the time of contact and at the 12^{th} frame respectively. In the second sequence, the bottle is presented to the robot in the same orientation it had in the demonstrated action and the robot repeats the observed action, a "side tap". In the third sequence, the car is presented at a different angle. The appropriate action to exploit the affordance and make the bottle roll is now a "back slap".

"mirror-type" representation. The observed action is interpreted on the basis of the robot own motor code. The same data structure is also used/activated when performing an action in response to the sight of a known object. The causal link between the two events that could be separated by several seconds is the object, the goal, and the object's affordances. There is considerable precedent in the literature for a strong connection between viewing object manipulation performed by either oneself or another **?**. There is also a growing evidence that imitation is goal-directed **?** and that the object of the action is explicitly coded (e.g. during reaching) **?**.

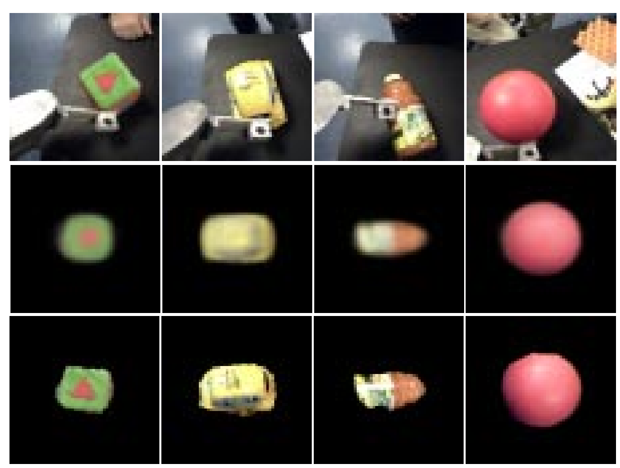

Figure 4: The top row shows the four objects used in this experiment, seen from the robot's perspective. The middle row shows prototypes derived for those objects using a naïve alignment procedure. None of the prototypes contain any part of the robot's manipulator, or the environment. These prototypes are used to find the best available segmentations of the objects (bottom row).

Acknowledgements

Some people are just so wonderful.

References

A. Wohlschläger and H. Bekkering. Is human imitation based on a mirror-neurone system? Some behavioural evidence. *Experimental Brain Research*, 143:335–341, 2002.

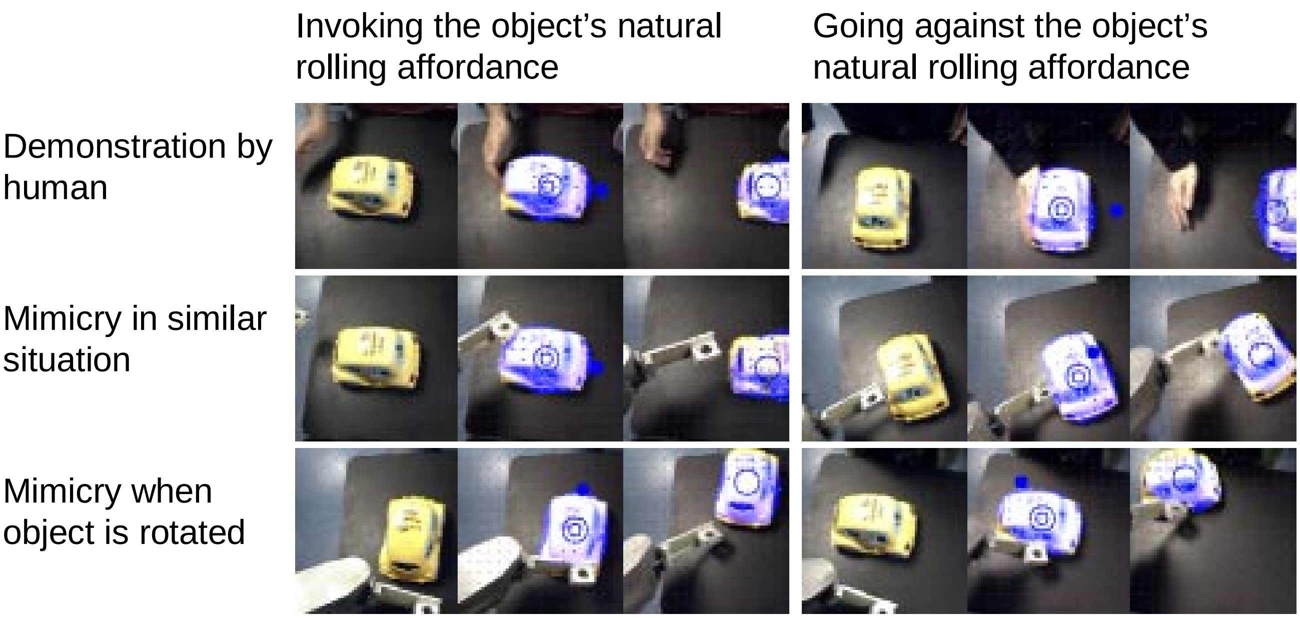

Figure 3: An extended mimicry example using the toy car. The sequences on the left show the robot mimicking a human exploiting the car's rolling affordance. The sequences on the right show what happens when the human hits the car in a contrary fashion, going against its preferred direction of motion. The robot mimics this "unnatural" action, suppressing its usual behavior of trying to evoke rolling.

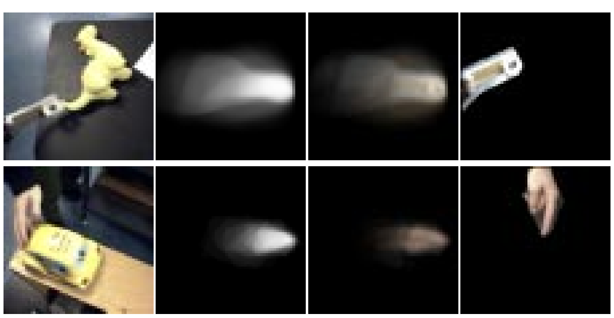

Figure 5: The robot manipulator (top left) was automatically segmented during 20 poking sequences. The segmentations were aligned and averaged, giving the mask and appearance shown in the adjacent images. The best matching view is shown on the top right. A similar result for the human hand is shown on the bottom, based on much less data (5 poking sequences, hands of two individuals).

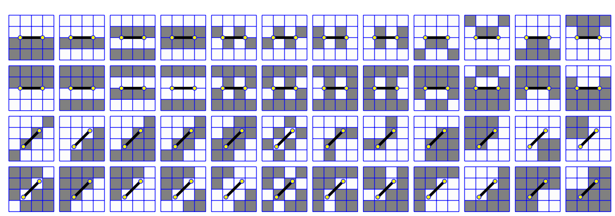

Figure 6: Edges have diverse appearances. This figure shows the orientations assigned to a test suite prepared by hand. Each 4×4 grid is a single test edge patch, and the dark line centered in the grid is the orientation that patch was observed to have in the training data. The oriented features represented include edges, thin lines, thick lines, zig-zags, corners etc. It is difficult to imagine a set of conventional filters that could respond correctly to the full range of features seen here – all of which appeared multiple times in object boundaries in real images.

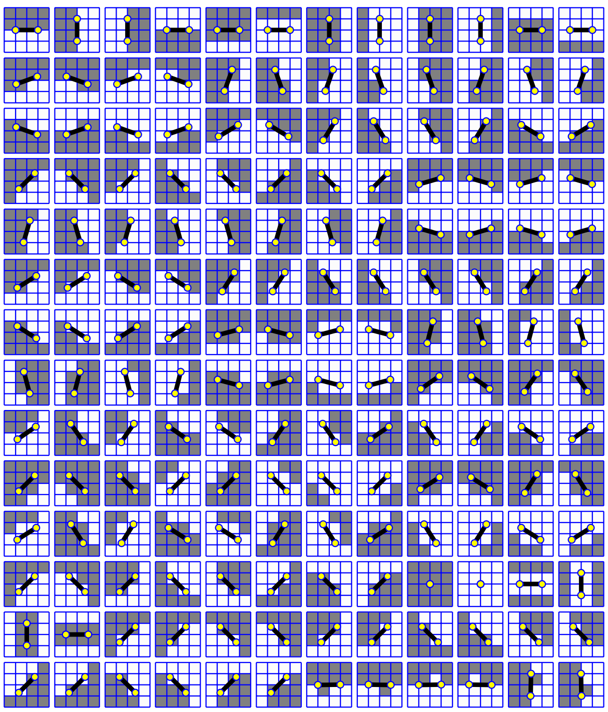

Figure 7: The most frequently observed edge appearances. All patches observed are replicated for all 90° rotations, mirror flips, and inversion of foreground/background. The most frequent (top) are simple straight edges. The line in the center of each patch shows the orientation associated with that patch. After the straight edges, the completely empty patch is common (produced in saturated regions), followed by a tube-like feature (third-last row) where the boundary is visually distinct to either side of the edge. This is followed corner-like features and many thousands of variations on the themes already seen.