
Shape Recognition Through Dynamic Motor Representations

Navendu Misra and Yoonsuck Choe

Department of Computer Science, Texas A&M University
College Station, TX 77843-3112, USA

Summary. How can agents, natural or artificial, learn about the external environment based only on its internal state (such as the activation patterns in the brain)? There are two problems involved here: first, forming the internal state based on sensory data to reflect reality, and second, forming thoughts and desires based on these internal states. (Aristotle termed these passive and active intellect, respectively [1].) How are these to be accomplished? Chapters in this book consider mechanisms of the instinct for learning (chapter PERLOVSKY) and reinforcement learning (chapter IFTEKHARUDDIN; chapter WERBOS), which modify the mind's representation for better fitting sensory data. Our approach (as those in chapters FREEMAN and KOZMA) emphasizes the importance of action in this process. Action plays a key role in recovering sensory stimulus properties that are represented by the internal state. Generating the right kind of action is essential to decoding the internal state. Action that maintains invariance in the internal state are important as it will have the same property as that of the represented sensory stimulus. However, such an approach alone does not address how it can be generalized to learn more complex object concepts. We emphasize that the limitation is due to the reactive nature of the sensorimotor interaction in the agent: lack of long-term memory prevents learning beyond the basic stimulus properties such as orientation of the input. Adding memory can help the learning of complex object concepts, but what kind of memory should be used and why? The main aim of this chapter is to assess the merit of memory of action sequence linked with a particular spatiotemporal pattern (skill memory), as compared to explicit memory of visual form (visual memory), all within an object recognition domain. Our results indicate that skill memory is (1) better than visual memory in terms of recognition performance, (2) robust to noise and variations, and (3) better suited as a flexible internal representation. These results suggest that the dynamic nature of skill memory, with its involvement in the closure of the agent-environment loop, provides a strong basis for robust and autonomous object concept learning.

1 Introduction

What does the pattern of activity in the brain mean? This is related to the problem of semantics [2] (also see chapters FREEMAN and KOZMA, this volume) or symbol grounding [3]. The question, as straight-forward as it seems, becomes quite complex as soon as we realize that there can be two different interpretations, as shown in Fig. 1. In (a), the task is to understand what is the meaning of the internal brain state of someone else's brain, while in (b), one wishes to understand, sitting within one's own brain, what the internal state means. The first task seems feasible, and it reflects how neuroscientists conduct their research. The second task seems impossible at first (it is reminiscent of Plato's allegory of the cave [4] or Searle's Chinese room [5]), but since this is how the brain operates, it should not be a problem at all. What is missing from this picture? In our previous work, we have argued that action plays an important role in recovering sensory stimulus properties conveyed only by the internal state (Fig. 2(a)) [6, 7]. Furthermore, we showed that action that maintains invariance in the internal state will have the same property as that of the represented sensory stimulus (Fig. 2(b)). Thus, generating the right kind of action (the kind that maintains internal state invariance) amounts to decoding the internal state.

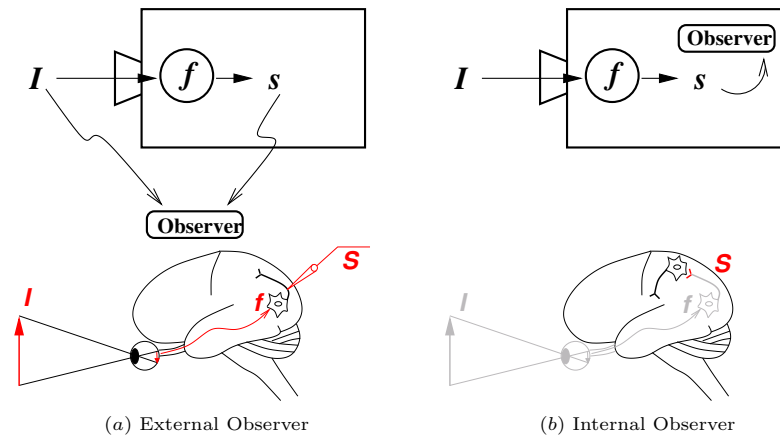


Fig. 1. External vs. Internal Perspective on Understanding Internal Brain State (a) The diagram for external observer, on the left, demonstrates how the input I creates a spike pattern s . Since the observer has access to both of these one can infer what stimulus property is conveyed by s . (b) However, in the internal observer model, on the right, there is no direct access to the outside environment and as such one can only monitor one's internal sensory state or spike pattern. (Adapted from [7].)

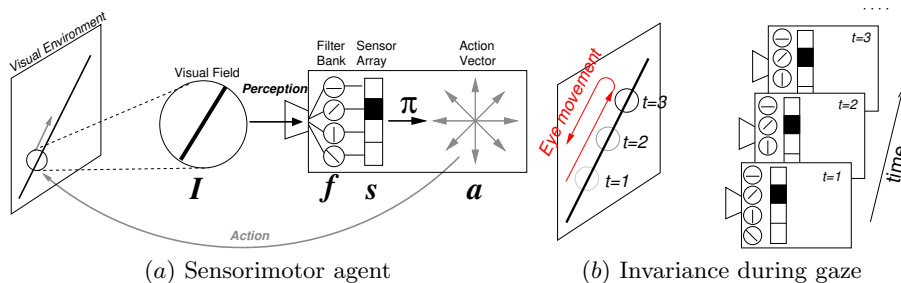


Fig. 2. Understanding Internal State in a Sensorimotor Agent (a) The visual field I receives stimulus from a small region of the visual environment. This is fed to the filter bank f that activates the sensor array (of orientation filters). The agent then performs certain actions a based on the sensory state s , which may in turn affect the sensory state in the next time step. (b) By moving in the diagonal direction, the internal state stays invariant over time. The property of the action (traversing diagonally) exactly matches the stimulus property (diagonal orientation) signaled by the invariant internal state. (Adapted from [7].)

One limitation of the sensorimotor agent is how the approach can generalize to learning more complex object concepts, since moment-to-moment invariance cannot be maintained while traversing the contour of a complex object. The limitation is due to the reactive nature of the sensorimotor interaction in the agent, i.e., the lack of long-term memory. Memory is needed for learning beyond the local orientation of the input. We will investigate how adding memory can help the learning of complex object concepts.

Let us consider what the internal state activation pattern will look like over time as the sensorimotor agent traverses the contour of a form, an octagon for example (Fig. 3). As we can see from the activation pattern of the internal state in (b) and (d), moment-to-moment invariance is maintained only when a straight stretch of the octagon is being traversed. One observation here is that if the changing spatiotemporal pattern of the internal state is treated as a representational unit, then a corresponding complex action sequence maintaining invariance in this repeating spatiotemporal pattern can be found that has a property congruent with that of the stimulus object (in this case, an octagon). Memory plays an important role here, since such an invariance in the spatiotemporal pattern has to be detected: You need to know your own action sequence over time, and the corresponding spatiotemporal pattern.

An important issue is to assess the merit of memory of such an action sequence linked with a particular spatiotemporal pattern (skill memory or motor representation; or “fixed action pattern” [8]), as compared to explicit memory of visual form (visual memory or sensory representation), all within an object recognition domain. (Note that these concepts are analogous to episodic vs. procedural memory in psychology research [9].) As we can see from Fig. 3(b) and (d), the spatiotemporal pattern and its associated action sequence, if

linearly scaled in time, will exactly match each other. Properties like these can be beneficial when serving as a representational basis for invariant object recognition.

In order to assess the relative merit of skill memory, we tested object recognition performance based on skill memory vs. visual memory, and tested how easy it is to map from an arbitrary representation to either skill memory or visual memory representation. Our results indicate that skill memory is (1) better than visual memory in terms of recognition performance, (2) robust to noise, and (3) better suited as a flexible internal representation. These results indicate that the dynamic nature of skill memory, with its involvement in the closure of the agent-environment loop, provides a strong basis for robust object concept learning and generalization.

The rest of this chapter is organized as follows: The following Sec. 2 briefly reviews related work, and Sec. 3 provides details about input preparation and training procedure. Sec. 4 presents the main computational experiments and results, and we finally conclude (Sec. 6) after a brief discussion (Sec. 5).

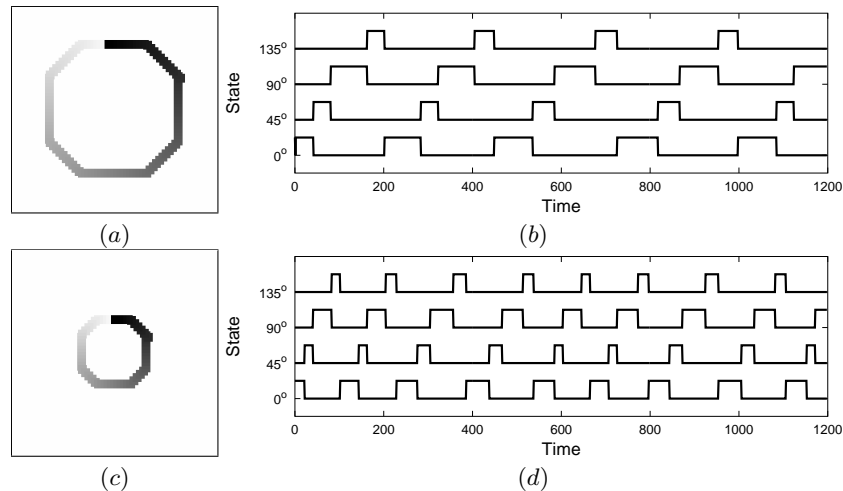


Fig. 3. Tracing a Complex Object. The sensorimotor agent’s gaze over time as it traces an octagon input and its corresponding internal state is shown. (a) The agent’s gaze is plotted over 250 time steps, where the grayscale indicates the time step (white is time 0, and black is time 249). (b) The activation state of the four neurons in the agent is shown over 1200 time steps. As the agent repeatedly directs its gaze around the octagon, the internal state also shows a repeating pattern, from 0° , 45° , 90° , 135° , back to 0° , for example (the pattern in the interval $[200, 449]$). Note that the trace in (a) corresponds to the interval $[0, 249]$ in (b). (c) and (d) show the same information as in (a) and (b), but for a smaller input. Note that the period of the activity is shorter in (d) than in (b), as the length to traverse is shorter in (c) compared to (a).

2 Background

The importance of action and interaction in perception has been identified early on, as documented in the works of Lashley on motor equivalence, the concept that different effectors can be used to generate the same kinematic pattern, [10] (also see [11, 12]) and those of Gibson on ecological perception [13]. The ensuing active vision movement [14, 15, 16] and embodied robotics [17] continued investigation in this area. In this section, we will review related works putting an emphasis on the role of action in sensory perception.

As discussed in the introduction, an action-oriented approach may be able to solve the problem of internal understanding. Even when the stimulus is not directly accessible, through action and the associated change in the internal state, key properties of the stimulus can be recovered [6, 7].

Experimental results suggest that action and the motor cortex actually play an important role in perception [18, 19, 20, 21, 22], providing support for the above idea. The implication is that sensorimotor coordination may be a necessity for autonomous learning of sensory properties conveyed through sensory signals.

There are ongoing research efforts in the theories of the sensorimotor loop [23, 24, 25, 26, 27, 28], developmental robotics [29, 30, 31, 32], bootstrap learning [33, 34], natural semantics [35], dynamical systems approach to cognition [36, 37, 38], embodied cognition [39], imitation in autonomous agents [40, 41, 42, 43, 44, 45, 46, 47], etc. that touch upon the issue of sensorimotor relationship. More recent works have specifically looked at the role of action in perception and perceptual organization [48, 49], information flow [50], and the role of internal models in action selection and behavior [51]. However, these approaches have not focused on the question of how the brain can understand itself. (The works by Freeman [52, 2] and Ziemke and Sharkey [53] provide some insights on this issue.)

3 Methods

3.1 Input preparation

The input to the neural network was prepared by randomly generating the action sequence and the 2D array representation for three different shapes (circles, triangles, and squares). Each of the representations was defined to have the same input dimension.

Shape generation

The generation of the three types of shape was done by following a simple algorithm. These algorithms were constructed using a LOGO-like language [54]. In this language instructions for navigating a space is provided by a

fixed range of actions, i.e., 'turn left', 'move forward', and so on that are then plotted. This language construct was adapted for the formation of the shape generation algorithms in the following manner. Initially a starting point was chosen for each of the figures. Then a series of steps were produced to allow for a full rendition of the entire shape. The algorithm was parameterized so as to produce images that were scaled and translated. As different coordinates were traversed the corresponding 2D array points were marked. The benefit of the LOGO-like algorithm was that it allowed for the easy capture of the action sequence for the particular shape. This was the case because the actions produced to traverse the object could be captured as the direct representative action sequence for a figure. Since each of the algorithms was parameterized, a sequence of random values for scaling and translating the images were provided.

Visual memory – 2D array

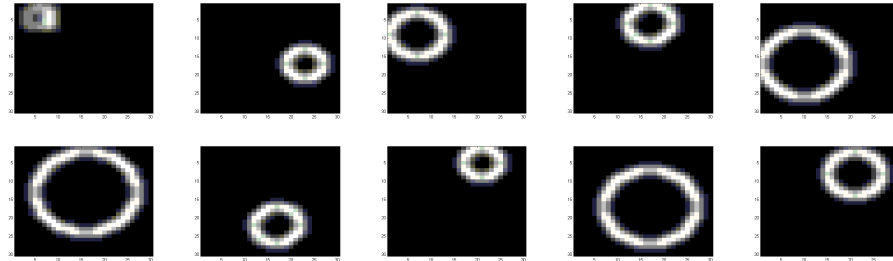


Fig. 4. The visual representation of circles. This sequence of figures illustrates the range of variations that are performed on the circle shape in the visual memory (2D array) representation.

Visual memory representation is a direct copy of the sensory data. As a result, when the 2D array representation for the figure is visualized it appears like the figure itself. All the shapes for the 2D arrays were generated using the algorithm described above. The output was a two dimensional array with the pixels constituting the contour of the figure set to one and the background to zero. On this a Gaussian filter was applied. This caused the values in the array to have more continuous values. The primary reason for this last step was to have a more fair comparison between 2D array and action sequence. The normalized range of values was between 0 and 1 and the resultant size of the array was 30×30 . Fig. 4, Fig. 5, and Fig. 6 provide examples of the different shapes in the 2D array representation.

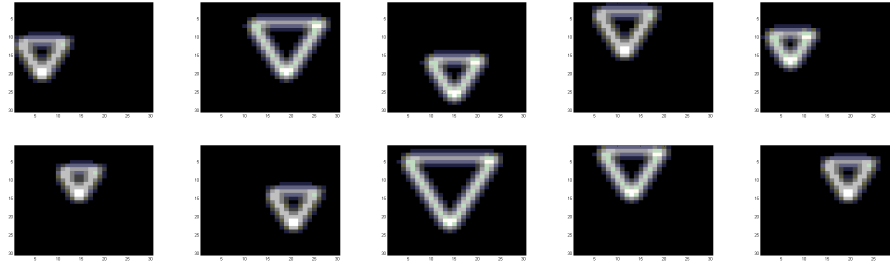


Fig. 5. The visual representation of triangles. This sequence of figures illustrates the range of variations that are performed on the triangle shape in the visual memory (2D array) representation.

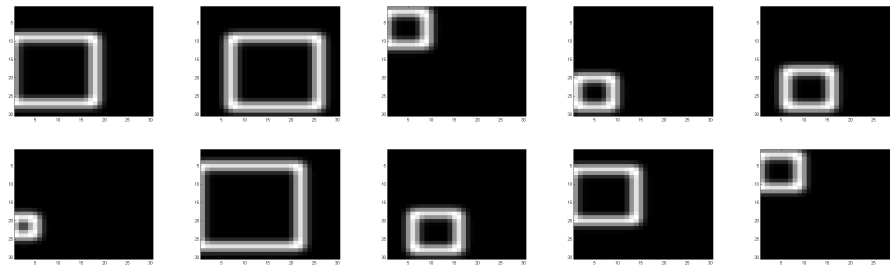


Fig. 6. The visual representation of squares. This sequence of figures illustrates the range of variations that are performed on the square shape in the visual memory (2D array) representation.

Skill memory – action sequence

Skill memory representation, action sequence, involves the retention of actions that an agent may have performed while navigating the environment. As explained earlier the action sequence was generated by utilizing an algorithm based on the LOGO language. The produced output action sequence had four actions; motion north, motion south, motion east, and motion west. This was represented by values 0, 90, 180, and 270 degrees. The values were subsequently normalized to lie between the range 0 to 1. Before normalization the action vectors were smoothed. Smoothing was accomplished by taking an average of values representing the neighboring action vectors, as specified by the size of the smoothing window. This resulted in a less discrete change of action vectors. This difference is illustrated in Fig. 7. In (a) the action sequence is not smoothed and the sequence of actions constructed has staircase-like variations. However, in (b) the actions get averaged to form a new action that is formed by averaging the neighboring action vectors. Fig. 7 shows the action vectors at the coordinates where these actions were performed. Figs. 8–9 show the same smoothing effect but for triangles and squares.

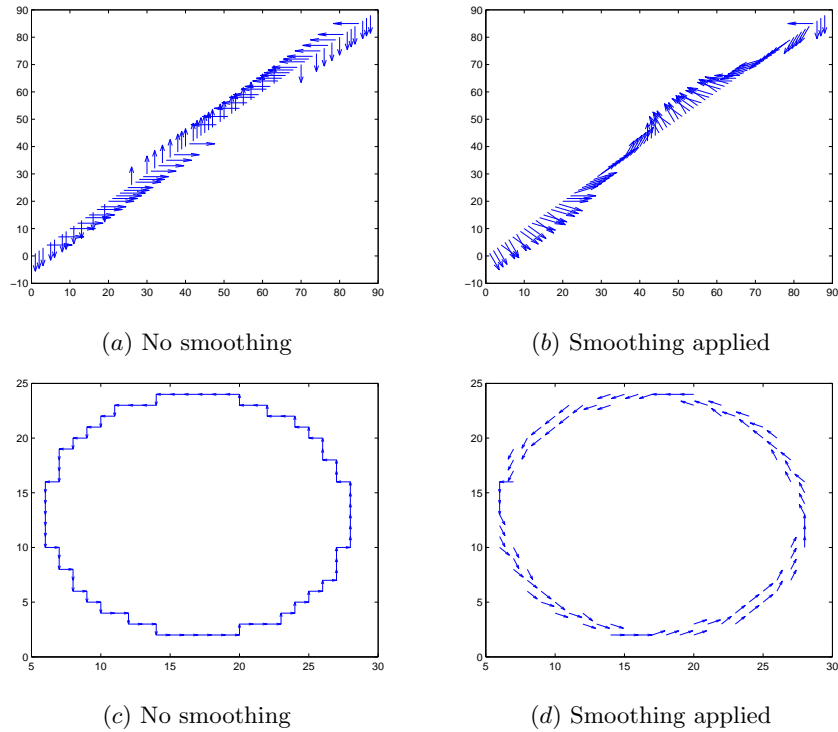


Fig. 7. The action sequence for a circle shape. (a) This plot demonstrates a linear ordering of the action sequence for a circle shape before any smoothing is applied. Note in this figure the discrete change in action vectors. (b) This plot demonstrates a linear ordering of the action sequence for a circle shape after smoothing is applied. Here the action vectors have been smoothed to display a continuous variation in actions. (c) This plot demonstrates a 2 dimensional view of the action sequence for a circle shape before any smoothing is applied. Note in this figure the discrete change in action vectors in the staircase-like formation. (d) This plot demonstrates a 2 dimensional view of the action sequence for a circle shape after smoothing is applied. The smoothed action vectors show a more intuitive sequence of actions.

Another issue with the action sequence was that its length (number of individual action steps in the sequence) was not fixed, unlike the visual representation where it was fixed to $30 \times 30 = 900$. This meant that resizing of the action sequence had to be done to match the input dimension ($30 \times 30 = 900$). A very simplistic algorithm was devised to resize or stretch the action sequence. The end result was that each action sequence size was of 900 dimensions.

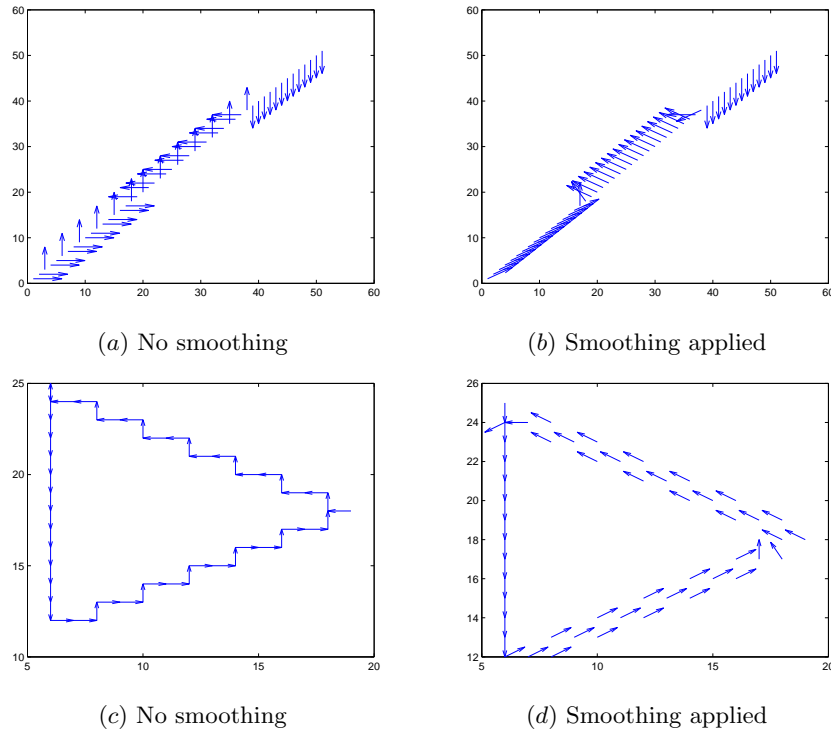


Fig. 8. The action sequence for a triangle shape. These plots demonstrate the same effect of smoothing as mentioned in Fig. 7, except that they show the case of a triangle.

3.2 Object Recognition and Representation Mapping

The experiments were formulated by creating one thousand randomly scaled and randomly translated figures for each shape category (triangle, circle, and square) with their corresponding action sequence and 2D array representations. The patterns generated for the action sequence and 2D array representations were stored in their respective data sets. A portion of this data set (75%) was used for training a neural network and the rest was used for testing (25%). Using the testing data set, the performance of each of the representations was recorded. A detailed explanation is given in the following sections.

There were a total of ten runs for each experiment. For each of these trials the training set and the test set were chosen at random from the data set for each class (circle, square, and triangle). This was composed of a thousand points for action sequence and 2D array for each of the three figures, resulting in action sequence and 2D array data set of three thousand each. For each run, at random, 75% of the data set was chosen as training data and the rest was used as the test data. The training set was provided to the neural

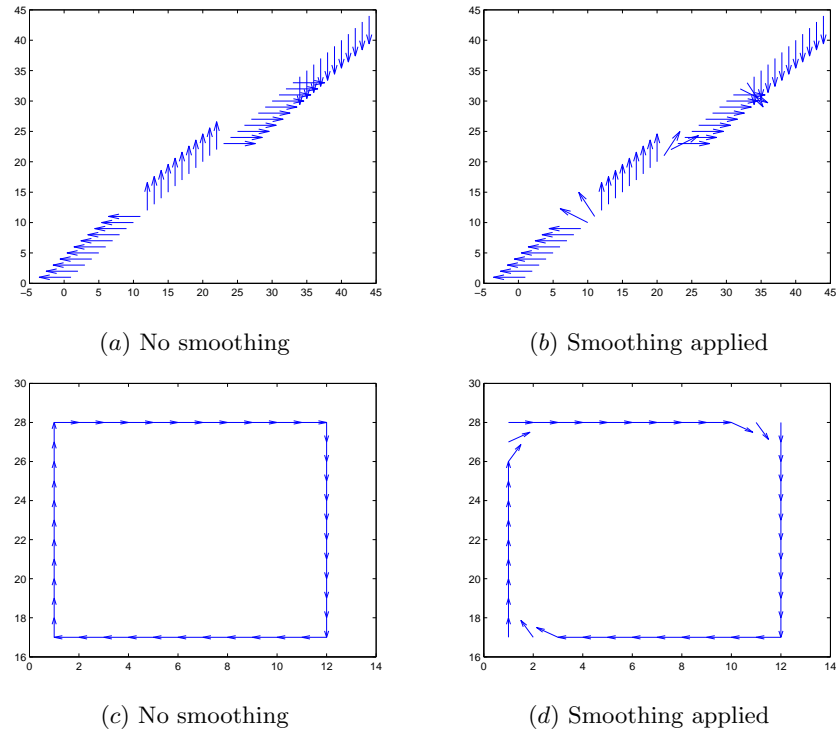


Fig. 9. The action sequence for a square shape. These plots demonstrate the same effect of smoothing as mentioned in Fig. 7, except that they show the case of a square.

network, and trained using backpropagation [55, 56, 57]. The neural network had 900 input neurons, 10 hidden neurons in the single hidden layer and 3 output neurons corresponding to the 3 shape classes. For the representation mapping experiment, the target vectors were modified to be the actual visual and skill memory representations. In this case the number of output neurons was increased to 900.

Average classification rate on the test set was a measure that was used to gauge the relative performance for both visual and skill memory. The classification rate for each trial recorded the average number of times the actual output deviated from the target vector of the three output neurons. A threshold of 0.5 was set so that if the deviation of the output neuron activation was within this value then the particular input was claimed to be properly classified. The average classification rate was then acquired by running the experiments ten times and taking the mean and standard deviation of the values. Student's t-test was used to measure the significance of the differences [58].

Another measure is the mean squared error (MSE). This value represents the average of all the squared deviations of the output values from the exact target values. MSE gives a general idea of how well the mapping was learned in case hard classification is not possible.

4 Computational Experiments and Results

In order to evaluate the effectiveness of each of the memory representations there needs to be a comprehensive evaluation of each of the memory systems with respect to the performance measures specified in the previous section. These performance measures were used to primarily demonstrate the relative difference between the two memory systems rather than provide a mechanism for absolute comparison with a general pattern recognition approach that may seek to maximize the performance.

4.1 Visual memory vs. skill memory in recognition tasks

The overall speed of learning, measured using MSE, is illustrated in Fig. 10. MSE values for each of the curves were calculated by taking an average of ten trials for each of the two memory representations. The neural network was allowed to train for one thousand epochs. As can be seen from Fig. 10, the error rate for skill memory is consistently lower than that of visual memory. Also after about 200 epochs the MSE comes close to zero for skill memory while visual memory can only reach an MSE value of about 0.1 after the full period of one thousand epochs. The results clearly demonstrate that the neural network can more easily learn the various action sequences in skill memory.

The differences between skill memory and visual memory are further emphasized in Fig. 11. Here the average classification rate on the test sets is shown using the bar chart with the error bars representing the 95 % confidence intervals. The average classification rate for visual memory was 0.28 while for skill memory it was almost four times higher, close to 0.97. The difference was significant under t-test ($p = 0, n = 10$). In sum, the action-based skill memory was significantly easier to learn than the bitmap-based visual memory, both in terms of speed and accuracy.

4.2 Skill memory with variations

The performance of skill memory was measured under variations in the formation of the action sequence. These variations included (1) changes in the smoothing window size, (2) variations to the number of starting points in object traversal for a particular action sequence, and (3) noise in action sequence.

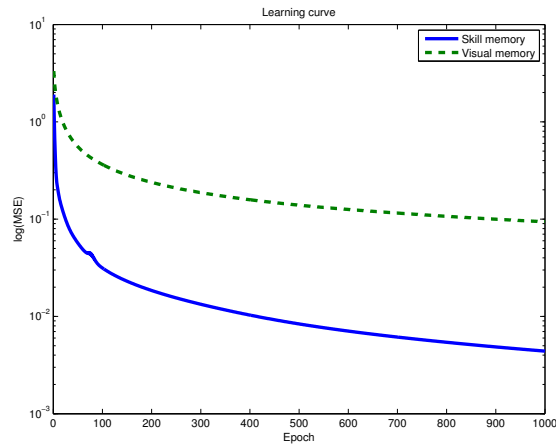


Fig. 10. Learning curve for visual and skill memory. This plot shows the average learning curve for both skill and visual memory (10 trials each). From this plot we can see that skill memory is learned faster and more accurately. On the other hand visual memory still has a higher MSE even after 1,000 epochs.

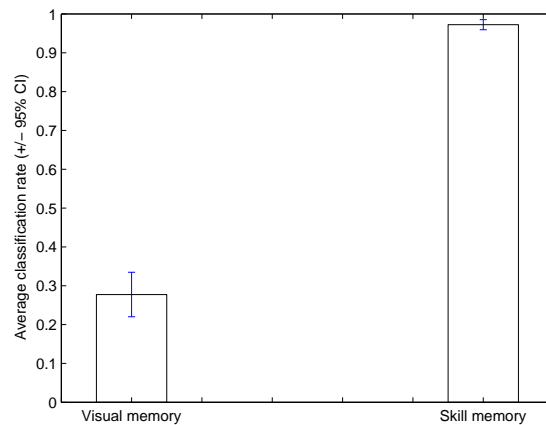


Fig. 11. The average classification rate of visual and skill memory. This bar chart shows the average classification rate of skill and visual memory on the test set (\pm 95 % confidence interval). Skill memory has a smaller variance and higher average classification rate representing a more consistently good performance as opposed to visual memory.

Smoothing window size

Smoothing was applied to all the action sequence that was generated. The effect of this has already been illustrated in the previous section. The average classification rate increases slightly with an increase in the window size. However, a further increase in window size causes the average classification rate to decrease slightly. As a result the default smoothing window size was chosen to be three. Fig. 12 shows how smoothing affects the average classification rate for skill memory. These values were averages taken by performing ten trials. Difference between window size 3 and size 0 was significant (t-test, $p < 0.0002, n = 10$), but the difference between window size 6 and 3 was not (t-test, $p > 0.35, n = 10$).

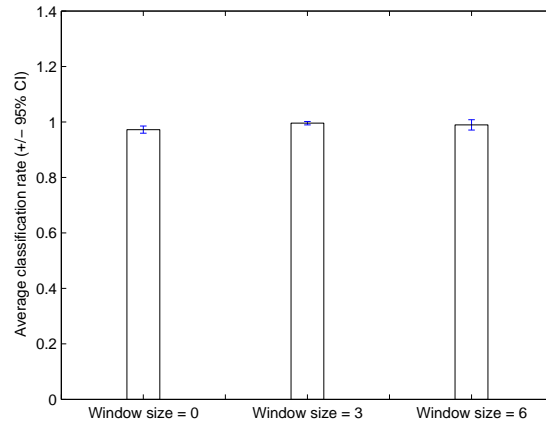


Fig. 12. The effect of smoothing on the classification rates for skill memory. This bar chart shows the effect of varying window size on the average classification rate ($\pm 95\%$ confidence interval). Window size 3 yielded the most optimal performance with the lowest variation.

Random starting points

Another variation was to test how the classification rate was impacted by different starting points on the input object chosen for the action sequence generation. This was implemented by varying the number of starting locations for each action sequence. Having different starting points is an important variation because the way the memory representation system was originally setup, all the action sequences were generated by having the trajectory start at the same relative location on the shape and as a consequence the action sequence generated would not have much variation. Fig. 13 shows a 2D view

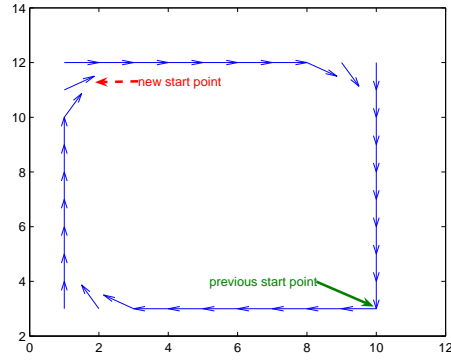


Fig. 13. The 2D view of the action sequence for a square with smoothing. The two arrows point to the two different locations where the action sequences may have started.

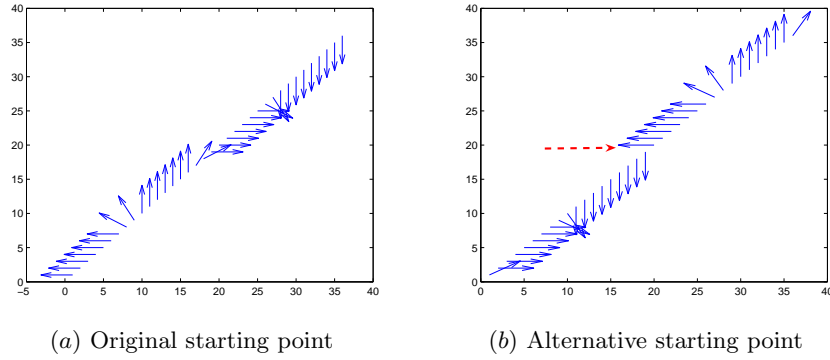


Fig. 14. The linear ordering of the action sequence for a square with differing starting points. The plot in (a) shows how the action sequence will appear if the action sequence generation started from the original starting point. The plot in (b) shows how the action sequence will appear if the action sequence generation started from the new starting point (shown in Fig. 13). The dashed arrow in (b) points to the original starting location. The only difference with the original version is that the action sequence is shifted, but that is enough to affect the classification accuracy.

of the action sequence for a square and possible positions where the action sequence may start. Fig. 14 shows the corresponding 1D view.

The overall effect of adding different starting points was that the average classification rate decreased with increasing number of alternative starting points (shown in Fig. 15). However, even with a high variation in the possible starting points the average classification rate for skill memory was still higher than visual memory. These values were averages taken by performing ten trials with varying training and test data and a constant smoothing window size of three. Differences between the visual against all the skill memory trials were significant under t-test ($p < 0.00015, n = 10$). In sum, skill memory was significantly easier to learn than visual memory, even when the task was made harder for skill memory. Note that the performance would suffer greatly if action sequence generation can start from an arbitrary point on the shape. However, humans typically start their trajectories from a stereotypical position, so the starting point may not be too arbitrary in practice. Here, the purpose was mainly to show how skill-based memory performs under typical conditions.

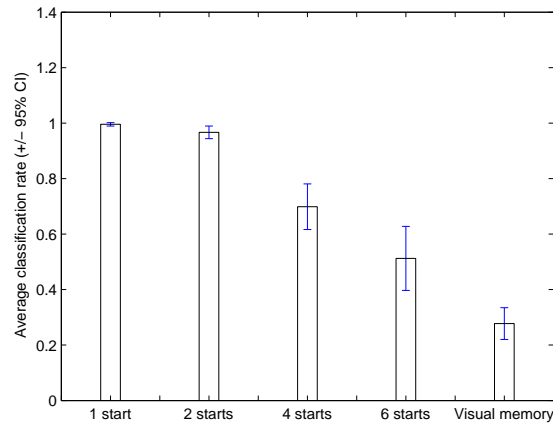


Fig. 15. Effect of increasing the number of alternative starting points in skill memory on the classification rate. This bar chart shows the change in the average classification rate as the number of trajectory starting points is increased ($\pm 95\%$ confidence interval). As the number of random starting points is increased, the average classification rate steadily drops, but in all cases skill memory performs better than visual memory.

Motor noise

In humans, tracing the contour of a form usually results in small variations in the traced trajectory. The last variation we tried was the introduction of noise in the action sequence. This means that a random error at some point in time during the creation of the action sequence occurred causing the trajectory to deviate from its normal course. The action sequences were generated as before, however, at random an angle between 0 and 360 was added based on the magnitude of the noise factor. An example of noise in the action sequence is shown in Fig. 16. In this figure a noise factor of 0.1 was used.

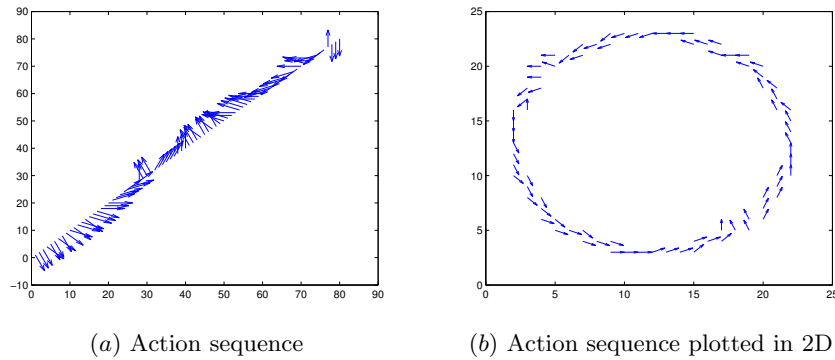


Fig. 16. The action sequence with noise for a circle shape. The plot shows a 1D and a 2D view of the action sequence for a circle after the application of random noise (noise factor 0.1) and after smoothing.

The noise factor is the probability that affects the magnitude by which an action vector's angle in space may be affected. Hence, a larger noise factor will mean a larger deformation of the shape that a particular action sequence may trace. Fig. 17 shows how the classification rate decreases with the increase in noise. However, even at higher noise levels, skill memory is still able to outperform visual memory. Differences between visual against all skill memory trials were significant under t-test ($p < 0.002, n = 10$). This demonstrates that skill-based memory is resilient to noise in motor sequence. The reason for the robustness is due to the fact that despite the noise, the components of skill memory (action vectors) do not change in number or position. Only their orientations change. In visual memory on the other hand, noise may increase or decrease the number and change the position of active pixels.

4.3 Representation mapping: Action as an intermediate representation

In order to test the hypothesis that action sequence (skill memory) may serve as a good intermediate representation of sensory information, the following

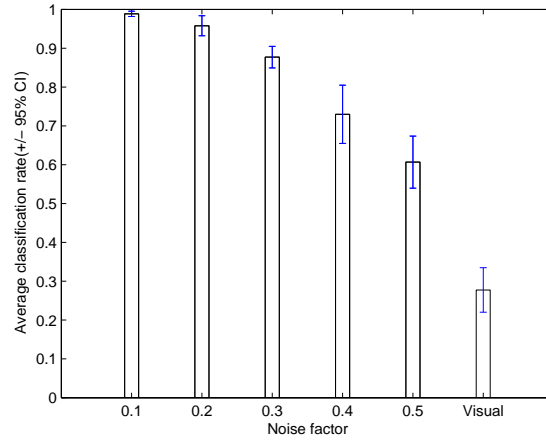


Fig. 17. The effect of motor noise on classification rate. This bar chart shows the effect of increasing noise on the average classification rate of skill memory ($\pm 95\%$ confidence interval). Visual memory is shown here as a baseline. The effect of noise can be observed clearly from this bar chart. However, notice that skill memory is quite resilient to noise and outperforms visual memory at noise factor of 0.5.

test was devised. The different mappings were; action to action, visual to action, action to visual, and visual to visual representation. If the learning for visual to action is easier with respect to sensory to sensory mapping (e.g. visual to visual), then that would indicate that in fact sensory information can be easily represented in terms of action. This idea coupled with the primary view that action-based memory may perform better at object recognition tasks further supports the idea that skill memory is a more ideal form of memory. The reason for this is that if the sensory to action mapping was very difficult then the performance advantage that skill memory holds may become less pronounced and there by limit the applicability of action-based memory.

To verify this tests involving the different mappings were conducted. This test was carried out on a neural network with 900 inputs and 900 outputs, corresponding to the 900 dimensional input for each representation. Fig. 18 shows the results of the experiment. The figure shows that the learning curve for visual to action mapping is as low as action to action mapping ($p = 0.37$, $n = 10$). The figure also shows that the action to visual memory is slightly easier to learn than visual to visual mapping (however, t-test showed that $p = 0.82$, $n = 10$, indicating that the difference was not significant). All other differences were significant under t-test ($p < 0.026$, $n = 10$). This bolsters our idea that action may be a good intermediate representation for sensory data. This support makes action-based memory more appealing. As we will discuss in detail in section 5.1, the demonstrated superiority of action-based memory is independent of the particular neural network learning algorithm

used. So, we expect the use of any standard supervised learning algorithm to give similar results as reported above. Furthermore, psychological experiments can be conducted to test our main claim, that action-based representation help improve object recognition performance. See section 5.2 for an expanded discussion on this matter.

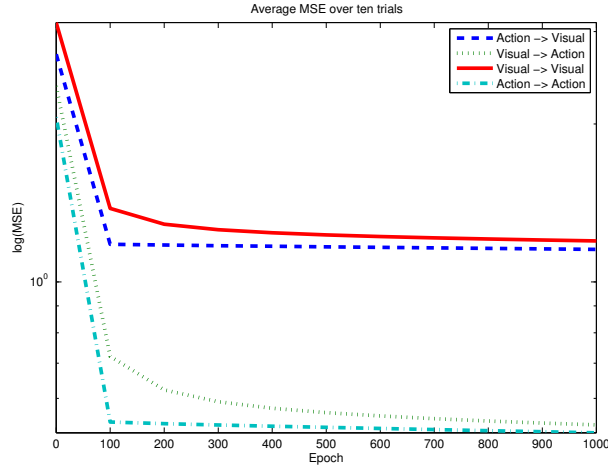


Fig. 18. The learning curve for the different mappings. This plot shows the learning curve for each representation mapping. From the learning curves we can infer how well each mapping performs with respect to other mappings. As expected mapping to action representation (skill memory) is easier to learn.

5 Discussion

Analysis of the results in the previous section indicates that skill-based memory representation performs better than visual memory in recognition tasks. It has been additionally demonstrated that even under quite severe variations skill-based memory is able to yield results which indicates its merits. It has been further demonstrated that action may serve as a good intermediate representation for sensory information. In the following, we will discuss why we believe these results may hold independent of the particular learning algorithm used (section 5.1) and how our general claims can be verified in psychological experiments (section 5.2). We will also discuss the relation of our work to human memory research, and provide some future directions.

5.1 Properties of action sequence

The primary reason why skill memory yields such impressive results is because of its ability to capture the core discriminating property of the respective shapes. This is the case because aspects such as size and location of the figure do not cause variations in the resized action sequence. Hence the action sequence for different-sized shapes was similar in the end. This makes it easy for the neural network to learn the skill-based representation. Variations introduced to the action sequence to compensate for the apparent advantage, i.e., using noise and random start methods did not affect the results. However, even large variations did not cause visual memory to outperform skill-based memory.

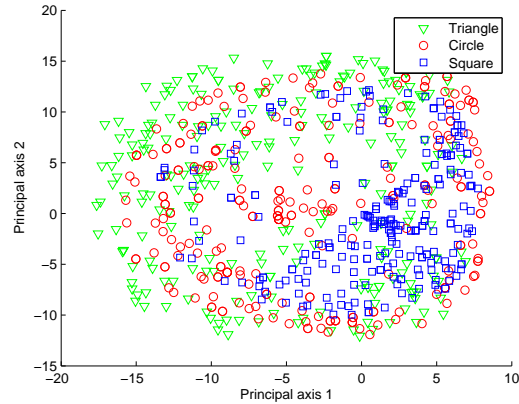
The properties of the action sequence and the 2D array representations can be further analyzed as in Fig. 19. In this figure, the Principal Components Analysis (PCA) plots along two principal component axes are shown for the data points in the action sequence and the visual representations. Fig. 19(b) shows that the PCA plot of skill-based representation has three distinct clusters for the three classes of input shapes. On the other hand, the PCA plot for visual memory (Fig. 19(a)) has all the data points almost uniformly scattered and overlapping, indicating that making proper class distinctions may be difficult. Such an analysis provides some insights on why skill memory performs better than visual memory in object recognition tasks.

Fig. 20 shows that with the introduction of noise the class boundaries become less pronounced for skill memory. With low noise the three distinct clusters for the corresponding classes are maintained. As the noise factor is increased the clusters become less compact and it becomes slightly harder to determine the class boundaries. However, even with high noise the class boundaries can still be determined more or less. These plots help us understand why the neural network was marginally less able to properly recognize skill-based representations when the noise was high.

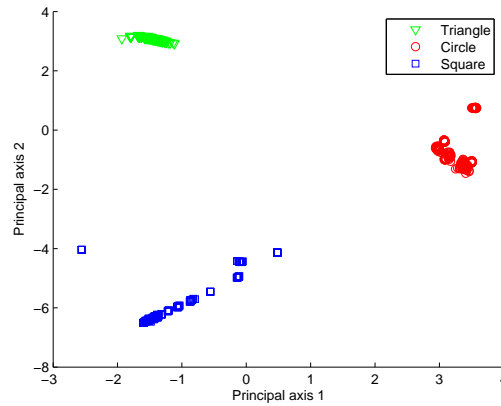
With the introduction of varying starting points many more clusters appear in the PCA plot, as illustrated in Fig. 21. However, it is interesting to note that the local clusters are more compact as opposed to the broader clusters that emerge with the addition of noise in Fig. 20. That is, data points from the same class are scattered around but they locally form tight non-overlapping clusters.

5.2 Related psychological experiments

One of the main assumptions of this research is that action can be represented along a time series that are scaled to be of the same length. However, one may question the validity of creating an action sequence and scaling such an action sequence. This also leads one to question, at what intervals are the actions stored and should this interval be long or short. All of these questions can be answered by the recent experiments performed by Conditt et al. [59]. The



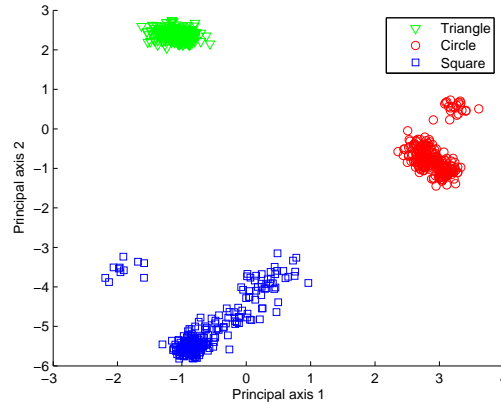
(a) PCA plot for visual memory



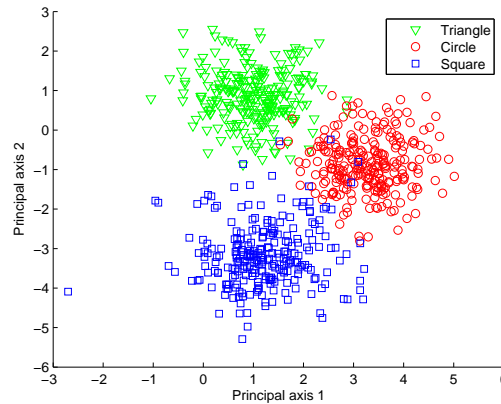
(b) PCA plot for skill memory

Fig. 19. The plot of PCA projection for visual and skill memory. (a) PCA projection for 2D array representation (visual memory) of the input data along the first two principal axes is shown. (b) PCA projection for action sequence representation (skill memory) of the input data along the first two principal axes is shown. Skill memory shows a much more separable clustering than visual memory.

result of their experiment suggests that when humans are asked to perform a series of actions, the actions tend to be represented as time invariant. This means that humans do not store the actions parameterized by absolute time. More precisely, humans do not have a timing mechanism that stores the exact duration between actions. This form of representation allows humans to counteract disturbances in the environment. A disturbance can result in the delay in the completion of an action for a given sequence. However, most humans are able to go along and complete the rest of action. Such experimental



(a) PCA plot for skill memory with low noise

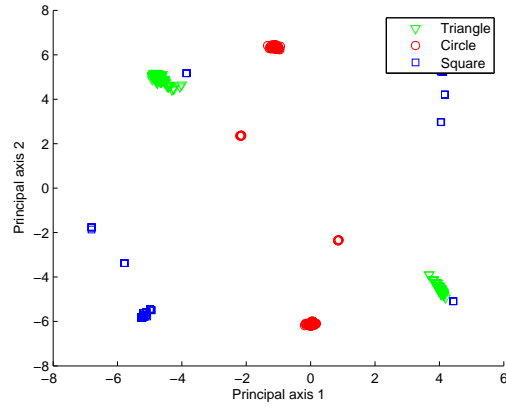


(b) PCA plot for skill memory with high noise

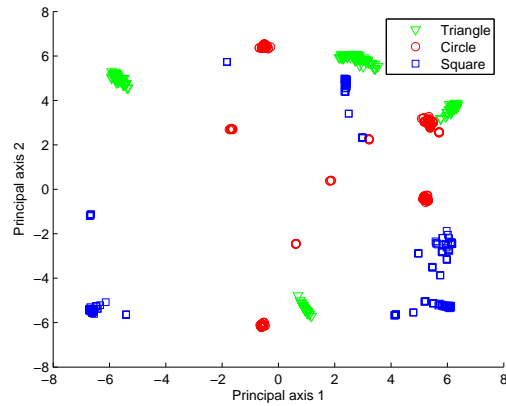
Fig. 20. The projection on the two principal axes for skill memory with varying noise factor. (a) PCA projection for skill memory with a lower noise factor of 0.1 along the first two principal axes is shown. (b) PCA projection for skill memory with a high noise factor of 0.5 along the first two principal axes is shown. Even at high noise, the clusters are still quite separable.

evidence allows us to be more confident about representing action sequence in the way we did in this chapter.

The action sequences produced in these experiments seem to be the same for each shape, thus making the job of learning trivial and therefore it appears that skill memory had an unfair advantage. Rather than this be a criticism against the validity of the research, it points out the fact that action sequence is not affected by size and translation of object. The similarity in action sequences further point out the core thesis of this chapter, that action sequence



(a) PCA plot for skill memory with few starting points



(b) PCA plot for skill memory with many starting points

Fig. 21. The projection on the two principal axes for skill memory with varying number of random starting points. (a) PCA projection for skill memory with two random starting points for action sequence generation along the first two principal axes is shown. (b) PCA projection for skill memory with four random starting points for action sequence generation along the first two principal axes is shown. With an increase in the number of starting points, the number of subclusters for each category increases, but there is minimal overlap across categories (at least locally).

as represented in skill memory may be an inherently superior representation scheme than the raw sensory information as in visual memory, because of its ability to capture properties of the object when time can be scaled with ease.

Finally, we can devise psychological experiments to test whether action-based skill memory can improve object recognition performance. One way to test the influence of the motor system in recognition is to immobilize the motor system during recognition tasks, or to dynamically eliminate motor-induced refreshing of sensory input (e.g., using eye-tracking devices). We predict that with the motor system temporarily turned off, object recognition performance will degrade significantly. There exists indirect evidence that the motor system plays an important role in perceptual tasks. Hecht et al. showed that pure motor learning enhanced perceptual performance, through an action-perception transfer [18]. Naito et al., on the other hand, showed that the primary motor cortex is involved in somatic perception of limb movement [19]. These results suggest an intimate role for the motor system in perceptual tasks.

5.3 Relation to memory in humans

The two memories, visual and skill, are analogous in many ways to the types of memory employed by natural agents. Natural agents have episodic and procedural memory [9]. Episodic memory is fact-based where certain information about events is stored. It is currently believed that these events are temporarily stored in the hippocampus [60] [61]. This may have similarities to visual memory described above. On the other hand, skill memory can be thought of as being similar to procedural memory. Procedural memory deals with the ability to recall sequential steps required for a particular task that an agent may perform [62]. It may be interesting to investigate if the main results derived in this chapter applies to the understanding of human memory and recognition.

Note, however, that the results presented here are insufficient to explain how the human memory actually works. Rather, what is presented here only suggests that skill-based memory may have theoretical virtue regarding perceptual understanding, as compared to visual memory.

5.4 Future work

The future expansion of this research topic involves the actual implementation of the skill-based memory system in an autonomous agent, as well as psychological experiments to systematically test the merits of skill-based memory. Other variations to the action sequence format can be studied such as methods that retain only the changes in the sequence of actions, i.e. only when there is a certain change in action, rather than retaining the total sequence. The resultant action sequence for shapes like square will have only four points, since in the traversal of a square the action vectors will need to be only changed four times. Another possibility is to use the relative difference between successive

action directions, which will give rotation invariance as well as the other two invariances (translation and scale) already achieved by our approach.

5.5 Contributions

The primary contribution of this research is the demonstration that skill-based memory has beneficial properties that can aid in perceptual understanding. These properties are in line with other research that suggested that action is a fundamental component for learning simple properties. However, in this chapter we were able to demonstrate that action plays an important role in learning complex objects when the system was allowed to have memory. This research clearly demonstrates how action can be incorporated into a powerful autonomous learning system. Another important observation is that when things are represented dynamically, then certain invariant properties can be naturally stored, i.e., simply changing the time scale of action generation is sufficient.

6 Conclusion

The study of memory systems, arose from the desire to develop a memory system that would allow autonomous agents to learn about complex object properties. The most basic memory system that an agent can have is the direct (raw) storage of the sensory data (such as visual memory). Another system is skill-based memory, which primarily involves the retaining of action sequences performed during a task. Skill memory was anticipated to be a better representation because of the crucial role action played in simple perceptual understanding [6, 7]. To test this hypothesis, we compared the two memory representations in object recognition tasks. The two primary performance measures, average classification rate and MSE, revealed the superior properties of skill memory in recognizing objects. Additionally, a related experiment demonstrated convincingly that action can serve as a good intermediate representation for sensory data. This result provides support for the idea that various sensory modalities may be represented in terms of action (cf. [63, 26]).

Based on the above results, we conclude that the importance of action in simple perceptual understanding of objects can successfully be extended to that of more complex objects when some form of memory capability is included. In the future, the understanding we gained here is expected to help us build memory systems that are based on the dynamics of action that enable intrinsic perceptual understanding.

Acknowledgments

The main results presented here is largely based on an unpublished thesis by NM [64]. The text was substantially revised for this chapter by the au-

thors. We would like to thank the editors of this volume for their constructive suggestions.

References

1. Barnes, J.: Aristotle. In Gregory, R.L., ed.: *The Oxford Companion to the Mind*. Oxford University Press, Oxford, UK (2004) 45–46
2. Freeman, W.J.: *How Brains Make Up Their Minds*. Wiedenfeld and Nicolson Ltd., London, UK (1999) Reprinted by Columbia University Press (2001).
3. Harnad, S.: The symbol grounding problem. *Physica D* **42** (1990) 335–346
4. Plato: *Plato's Republic*. Hackett Publishing Company, Indianapolis (1974) Translated by G. M. A. Grube.
5. Searle, J.: Is the brain a digital computer? In Grim, P., Mar, G., Williams, P., eds.: *The Philosopher's Annual*, Atascadero, CA, Ridgeview Publishing Company (1990) Presidential address (American Philosophical Association, 1990).
6. Choe, Y., Bhamidipati, S.K.: Autonomous acquisition of the meaning of sensory states through sensory-invariance driven action. In Ijspeert, A.J., Murata, M., Wakamiya, N., eds.: *Biologically Inspired Approaches to Advanced Information Technology*. Lecture Notes in Computer Science 3141, Berlin, Springer (2004) 176–188
7. Choe, Y., Smith, N.H.: Motion-based autonomous grounding: Inferring external world properties from internal sensory states alone. In Gil, Y., Mooney, R., eds.: *Proceedings of the 21st National Conference on Artificial Intelligence*. (2006) 936–941.
8. Llinás, R.R.: *I of the Vortex*. MIT Press, Cambridge, MA (2001)
9. Silberman, Y., Miikkulainen, R., Bentin, S.: Semantic effect on episodic associations. *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* **23** (1996) 934–939
10. Lashley, K.S.: The problem of serial order in behavior. In Jeffress, L.A., ed.: *Cerebral Mechanisms in Behavior*. Wiley, New York (1951) 112–146
11. Moore, F.C.T.: *Bergson: Thinking Backwards*. Cambridge University Press, Cambridge, UK (1996)
12. Bergson, H.: *Matter and Memory*. Zone Books, New York, NY (1988) Translated by Nancy Margaret Paul and W. Scott Palmer.
13. Gibson, J.J.: *The Perception of the Visual World*. Houghton Mifflin, Boston (1950)
14. Aloimonos, J.Y., Weiss, I., Bandopadhyay, A.: Active vision. *International Journal on Computer Vision* **1** (1988) 333–356
15. Bajcsy, R.: Active perception. *Proceedings of the IEEE* **76** (1988) 996–1006
16. Ballard, D.H.: Animate vision. *Artificial Intelligence* **48** (1991) 57–86
17. Brooks, R.A.: Intelligence without representation. *Artificial Intelligence* **47** (1991) 139–159
18. Hecht, H., Vogt, S., Prinz, W.: Motor learning enhances perceptual judgment: A case for action-perception transfer. *Psychological Research* **65** (2001) 3–14
19. Naito, E., Roland, P.E., Ehrsson, H.H.: I felt my hand moving: A new role of the primary motor cortex in somatic perception of limb movement. *Neuron* **36** (2002) 979–988

20. Held, R., Hein, A.: Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology* **56** (1963) 872–876
21. Bach y Rita, P.: *Brain Mechanisms in Sensory Substitution*. Academic Press, New York (1972)
22. Bach y Rita, P.: Tactile vision substitution: Past and future. *International Journal of Neuroscience* **19** (1983) 29–36
23. O’Regan, J.K., Noë, A.: A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* **24(5)** (2001) 883–917
24. Philipona, D., O’Regan, J.K., Nadal, J.P.: Is there something out there? Inferring space from sensorimotor dependencies. *Neural Computation* **15** (2003) 2029–2050
25. Philipona, D., O’Regan, J.K., Nadal, J.P., Coenen, O.J.M.D.: Perception of the structure of the physical world using unknown multimodal sensors and effectors. In Thrun, S., Saul, L., Schölkopf, B., eds.: *Advances in Neural Information Processing Systems 16*, Cambridge, MA, MIT Press (2004) 945–952
26. Humphrey, N.: *A History of the Mind*. HarperCollins, New York (1992)
27. Hurley, S.: Perception and action: Alternative views. *Synthese* **129** (2001) 3–40
28. Granlund, G.H.: Does vision inevitably have to be active? In: *Proceedings of the 11th Scandinavian Conference on Image Analysis*. (1999) 11–19
29. Weng, J., McClelland, J.L., Pentland, A., Sporns, O., Stockman, I., Sur, M., Thelen, E.: Autonomous mental development by robots and animals. *Science* **291** (2001) 599–600
30. Lugarella, M., Metta, G., Pfeifer, R., Sandini, G.: Developmental robotics: A survey. *Connection Science* **15** (2003) 151–190
31. Pfeifer, R., Scheier, C.: *Understanding Intelligence*. MIT Press, Cambridge, MA (1999)
32. Almásy, N., Sporns, O.: Perceptual invariance and categorization in an embodied model of the visual system. In Webb, B., Consi, T.R., eds.: *Biorobotics: Methods and Applications*. AAAI Press/MIT Press, Menlo Park, CA (2001) 123–143
33. Pierce, D.M., Kuipers, B.J.: Map learning with uninterpreted sensors and effectors. *Artificial Intelligence* **92** (1997) 162–227
34. Kuipers, B., Beeson, P., Modayil, J., Provost, J.: Bootstrap learning of foundational representations. *Connection Science* **18** (2006) 145–158
35. Cohen, P.R., Beal, C.R.: Natural semantics for a mobile robot. In: *Proceedings of the European Conference on Cognitive Science*. (1999)
36. Beer, R.D.: Dynamical approaches to cognitive science. *Trends in Cognitive Sciences* **4** (2000) 91–99
37. Cariani, P.: Symbols and dynamics in the brain. *Biosystems* **60** (2001) 59–83
38. Kozma, R., Freeman, W.J.: Basic principles of the KIV model and its application to the navigation problem. *Journal of Integrative Neuroscience* **2** (2003) 125–145
39. Varela, F.J., Thompson, E., Rosch, E.: *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge, MA (1993)
40. Schaal, S., Ijspeert, A.J., Billard, A.J.: Computational approaches to motor learning by imitation. *Philosophical Transactions of the Royal Society: Biological Sciences* **358** (2003) 537–547
41. Billard, A.: Imitation. In Arbib, M.A., ed.: *The Handbook of Brain Theory and Neural Networks*. 2nd edn. MIT Press, Cambridge, MA (2003) 566–569

42. Breazeal, C., Scassellati, B.: Robots that imitate humans. *Trends in Cognitive Sciences* **6** (2002) 481–487
43. Rao, R.P.N., Shon, A.P., Meltzoff, A.N.: A bayesian model of imitation in infants and robots. Cambridge University Press, Cambridge, UK (2004) In press.
44. Matarić, M.J.: Sensory-motor primitives as a basis for imitation: Linking perception to action and biology to robotics. In Dautenhahn, K., Nehaniv, C., eds.: *Imitation in Animals and Artifacts*. MIT Press, Cambridge, MA (2001) 391–422
45. Billard, A., Matarić, M.J.: A biologically inspired robotic model for learning by imitation. In: *International Conference on Autonomous Agents: Proceedings of the Fourth International Conference on Autonomous Agents*, New York, ACM Press (2000) 373–380
46. Ikegami, T., Taiji, M.: Imitation and cooperation in coupled dynamical recognizers. In: *Proceedings of the 5th European Conference on Advances in Artificial Life: Lecture Notes in Computer Science 1674*. Springer, London (1999) 545–554
47. Ito, M., Tani, J.: On-line imitative interaction with a humanoid robot using a dynamic neural network model of a mirror system. *Adaptive Behavior* **12** (2004) 93–115
48. Wyss, R., König, P., Verschure, P.F.M.J.: A model of the ventral visual system based on temporal stability and local memory. *PLoS Biology* **4** (2006) e120
49. Floreano, D., Suzuki, M., , Mattiussi, C.: Active vision and receptive field development in evolutionary robots. *Evolutionary Computation* **13** (2005) 527–544
50. Lungarella, M., Sporns, O.: Mapping information flow in sensorimotor networks. *PLoS Computational Biology* **2** (2006) 1301–1312
51. Bongard, J., Zykov, V., Lipson, H.: Resilient machines through continuous self-modeling. *Science* **314** (2006) 1118–1121
52. Freeman, W.J.: A neurobiological theory of meaning in perception. In: *Proceedings of the International Joint Conference on Neural Networks, IEEE* (2003) 1373–1378
53. Ziemke, T., Sharkey, N.E.: A stroll through the worlds of robots and animals: Applying Jakob von Uexküll’s theory of meaning to adaptive robots and artificial life. *Semiotica* **134** (2001) 701–746
54. Abelson, H.: *LOGO for the Apple II*. 1st edn. McGraw-Hill, Peterborough, N.H. (1982)
55. Werbos, P.J.: *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Department of Applied Mathematics, Harvard University, Cambridge, MA (1974)
56. Rumelhart, D.E., McClelland, J.L., eds.: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. MIT Press, Cambridge, MA (1986)
57. Werbos, P.J.: Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE* **78** (1990) 1550–1560
58. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: *Numerical Recipes in C*. Second edn. Cambridge University Press (1992)
59. Conditt, M.A., Mussa-Ivaldi, F.A.: Central representation of time during motor learning. *Journal of Neurobiology* **20** (1999) 11625–11630
60. Tulving, E., Markowitsch, H.J.: Episodic and declarative memory: role of the hippocampus. *Hippocampus* **8** (1996) 198–204
61. Buckner, R.L.: Neural origins of ‘i remember’. *Nature Neuroscience* **3** (2000) 1068–1069

62. Wise, S.P.: The role of the basal ganglia in procedural memory. *Seminars in Neuroscience* **8** (1996) 39–46
63. Humphrey, N.: *Seeing Red*. Harvard University Press, Cambridge, MA (2006)
64. Misra, N.: Comparison of motor-based versus visual representations in object recognition tasks. Master's thesis, Department of Computer Science, Texas A&M University, College Station, Texas (2005)