

# Learning Simultaneous Sensory and Motor Representations- (SESEMO)

Berkeley Kids

Redundancy reduction [1], edge detection [3] and hierarchical representations [5] have been the main stay for lot of work in computer vision and vision neuroscience to represent sensory data. But, these representations do not directly lend themselves to action. It has been hypothesized [6] that brain probably represents sensory information in a way that helps an organism to act in the world.

Philipona and O'Regan [7, 8] showed that the joint manifold of sensory and action space taken together has a lower dimensionality than that obtained by simply adding the dimensions of motor and sensory spaces. An observation which explains this is that motion of an object in the world and the equal but opposite motion of the organism with the object being stationary leads to the same sensory percept. It's the relative motion which matters. In other words, sensory and motor spaces are 'compensable'. This insight is probably employed by organisms to achieve stability of percept. For example, even when we humans are moving we perceive the world to be stationary.

In this work we consider the problem of joint sensory and motor representations with in the context of percept stabilization. This goal is interesting because it lets the agent explore building a sensory representation that is coupled with its actions. We assume that agent has no apriori knowledge of its kinematic model nor does it have any meaningful representations of its sensory stimulus, i.e. the agent is in the state of 'tabula rasa'.

In overactuated motor systems (high degrees of freedom/redundancy) training a control model becomes challenging. One way to approach this problem is through models proposed that explore the control space through inverse kinematics [9]. Other works have explored this idea using concepts of motor primitives [10], but they are not learnt from the statistics of the motion that an agent has to perform to achieve a goal. Here, we wish to explore the idea of learning motor primitives that can then be composed sequentially to perform actions.

An added advantage of such representations is that, once learnt, the problem is no longer solving an optimization problem but inferring coefficients on a learnt basis. That is to say, at each time step we no longer have to specify the control sequence for each controller but the problem is now of choosing a basis element that then provides us with a control sequence. Since this basis is in a far smaller space, we expect that this model will perform faster. This also mimics biological systems, where you have spinal reflexes and cortical motor systems that combine to give you complex motions[4].

Thus, we argue that for a robotic agent, a hierarchical control system with some low-level reflex circuits that are modulated by high level control systems that are task dependent can effect a more versatile system. This idea gains favor in the literature of optimal control theory as well [11]

We explore this problem by introducing the following ideas

- Learning a basis space that represent motor actions
- Joint estimation of both sensory and motor representations
- Representations that are flexible (e.g. faulty actuator, etc)

## I. MODEL

$\mathbf{I}(t)$  is the sequence of images (frames) that fall on the retina (say). We then learn a sparse generative sensory representation along the lines of Cadieu & Olshausen [2] that tries to account for both form and motion separately using complex basis elements. For the sake of brevity and simplicity, we begin with assumed a trained sensory representation. We refer the reader to [2] for more details for its training.

Thus, for a given sequence of images we infer our coefficients  $\alpha^t$  and  $\gamma^t$  that represent motion and form respectively. The matrix  $\mathcal{G}$  transforms the motion component of the sensory percept to the motor primitive space  $\beta^{t+1}$  thus, choosing the next action to be made for percept stabilization.

The action that is performed on the agent (self) is then given by product of the vector  $\beta^{t+1}$  and motor basis  $\mathcal{M}$ .

The last step that completes the loop is to get an estimate of how the action chosen effects the agent. The matrix  $\mathcal{F}$  transforms the action performed (actuator space) to the sensory (motion) percept space.

$$\min_{F, M, G, \tau} \|\alpha^{t+1} - \hat{\alpha}^{t+1}\|_2 \text{ s.t.} \quad (1)$$

$$\sum_i \lambda_2^i G^i \quad (2)$$

where  $\beta^{t+1}$  is computed as follows

$$\beta^{t+1} = \mathcal{G}\alpha^t \quad (3)$$

$$\hat{\alpha}^{t+1} = \sum_{i=0}^{t-1} e^{-\tau i} \alpha(t-i) + \mathcal{F}\mathcal{M}\beta^{t+1} \quad (4)$$

$$(5)$$

The objective that we wish to minimize is defined in Eq 1. Eq 2 defines the constraint that we require the columns of  $G$  (defined by  $G^i$  to be sparse so that our inferred  $\alpha$  results in a sparse  $\beta$ ). At this point, we explore no explicit constraints on  $F$  other than those specified above. We compute  $\beta^{t+1}$  as

a transformation of  $\alpha^t$  as described in Eq 3. We compute our expected value of  $\hat{\alpha}^{t+1}$  as a weighted sum of previous  $\alpha^{t-i}$  and a transformation of our current action into a sensory percept by the matrix  $\mathcal{F}$  as seen in Eq 4 where the exponent helps to impose a decaying weight for oldest time frame.

## II. DISCUSSION

## III. ACKNOWLEDGMENTS

Discussions with Jitendra Malik and Tony Bell motivated a lot of this work as well. Special thanks to Pavan Ramkumar, Northwestern University for helping clarify a lot of ideas through discussions. Special thanks to Bruno Olshausen for urging us to think about sensorimotor representations. We would also like to thank our respective funding agencies - Fulbright Scholar Program. NGA. NIH. UC Berkeley.

All the code for this project can be found on our github<sup>1</sup> link.

## REFERENCES

- [1] Horace B Barlow. Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234, 1961.
- [2] Charles F Cadieu and Bruno A Olshausen. Learning intermediate-level representations of form and motion from natural movies. *Neural computation*, 24(4):827–866, 2012.
- [3] David H Hubel and Torsten N Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968.
- [4] Eric R Kandel, James H Schwartz, Thomas M Jessell, et al. *Principles of neural science*, volume 4. McGraw-Hill New York, 2000.
- [5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, volume 1, page 4, 2012.
- [6] J Kevin O’Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(05):939–973, 2001.
- [7] David Philipona, J Kevin O’Regan, and J-P Nadal. Is there something out there? inferring space from sensorimotor dependencies. *Neural computation*, 15(9):2029–2049, 2003.
- [8] David Philipona, Jk O’regan, J-P Nadal, and Olivier Coenen. Perception of the structure of the physical world using unknown multimodal sensors and effectors. In *Advances in neural information processing systems*, page None, 2003.
- [9] Matthias Rolf and Jochen J Steil. Efficient exploratory learning of inverse kinematics on a bionic elephant trunk. 2012.
- [10] Stefan Schaal, Jan Peters, Jun Nakanishi, and Auke Ijspeert. Learning movement primitives. In *Robotics Research*, pages 561–572. Springer, 2005.
- [11] Emanuel Todorov. Optimality principles in sensorimotor control. *Nature neuroscience*, 7(9):907–915, 2004.

<sup>1</sup>[github.com/rctn/sesemo](https://github.com/rctn/sesemo)