Recurrent Slow Feature Analysis for Developing Object Permanence in Robots

Hande Çelikkanat
KOVAN Research Lab.
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
Email: hande@ceng.metu.edu.tr

Erol Şahin
KOVAN Research Lab.
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
Email: erol@ceng.metu.edu.tr

Sinan Kalkan
KOVAN Research Lab.
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
Email: skalkan@ceng.metu.edu.tr

***Abstract*—In this work, we propose a biologically inspired framework for developing object permanence in robots. In particular, we build upon a previous work on a slowness principle-based visual model (Wiskott and Sejnowski, 2002), which was shown to be adept at tracking salient changes in the environment, while seamlessly "understanding" external causes, and self-emerging structures that resemble the human visual system. We propose an extension to this architecture with a prefrontal cortex-inspired recurrent loop that enables a simple short term memory, allowing the previously reactive system to retain information through time. We argue that object permanence in humans develop in a similar manner, that is, on top a previously matured object concept. Furthermore, we show that the resulting system displays the very behaviors which are thought to be cornerstones of object permanence understanding in humans. Specifically, the system is able to retain knowledge of a hidden object's velocity, as well as identity, through (finite) occluded periods.**

I. Introduction

Humans are born into persistent worlds. Through years and countless interactions, we come to understand the world as a place that makes temporal and spatial sense. Objects do not appear out of nowhere, nor vanish into thin air, and as they move from point A to point B, they indeed have to exist for some time at every point in between. However, it is difficult to claim that we have so far built robots that truly make use of these basic axioms. Given that this understanding is a basis for us humans to act effectively in our persistent world, in this study we propose a model for building an understanding of object permanence in terms of a higher-order internal representation of the environment. Our ultimate goal is to build effective environment manipulation capabilities on top of this basis later on. But first, the robot needs to "understand" what it is to exist in a persistent world.

Against the complexity of a world abundant with continuously changing sensory signals, we take refuge in the "slowness principle" [1]: While the sensory signals are noisy and erratic, their underlying physical causes are relatively persistent in time. For instance, retinal signals can vary greatly from one moment to another due to lightning conditions, as well as saccadic movements of the eye, however the object which the eye sees is constant. Therefore we must be able to process these erratic sensory signals to extract meaningful high-level representations, which are characteristic of varying more "slowly", thereby containing more valuable "information", than the readily-available sensory signals.

In [1], Wiskott and Sejnowski propose Slow Feature Analysis (SFA). They show that sensory signals can be processed through successive steps of principal component analysis to extract optimally slow signals, which summarize the meaningful event in the scene. The solutions are guaranteed to be optimally slow within a predefined family of functions, while still conveying meaningful information. In forthcoming work, Wiskott et al. design a hierarchical visual architecture which can recognize objects through translational, orientational, and scaling transformations [2], distinguish known and novel objects, predict the type of the solutions if the transformation is known a priori [3], survive multiple co-occuring transformations, and even adapt themselves to behave like simple and complex visual neurons when trained with natural-life scenes [4]. As is, this architecture develops the object concept very plausibly. However, it is reactive in time, responding momentarily to inputs; and not being able to retain information through time, it cannot survive the object permanence problem. We propose a prefrontal cortex inspired extension to serve as a working memory.

The contributions of this paper are threefold: First, we apply SFA to real world images to demonstrate that the invariant object recognition capabilities can indeed survive real world data. (Note that with the exception of Zhang and Tao [5] and Berkes and Wiskott [4], SFA has not been used for real world images before. Furthermore, in these two studies, it has not been utilized for object recognition.) Second, we propose a quantitative method to estimate the sufficient number of slowly varying signals to represent a certain event. Finally, and most significantly, we propose an extension to develop an understanding of object permanence.

Our fundamental claim in building our extension on top of the SFA framework is that the object permanence can be regarded as a stage which develops on top of an already developed reactive object concept. In this sense, we claim that the SFA architecture fulfills the initially maturing object concept understanding, on top of which the object permanence understanding develops later in time on a par with the maturation of the prefrontal cortex. Last but not least, we show that the proposed framework demonstrates similar characteristics (and pitfalls) like an infant learning permanence of objects.

H. Celikkanat gratefully acknowledges support of TUBITAK 2211 program.

II. Related Work

A. Object Permanence

Piaget famously proposed that the cognitive functions of an infant progresses in developmental "stages" [6]. Within the first stage, he singles out the *object permanence* understanding as one of the cornerstones, at the end of which objects come to be identified as independent entities. Furthermore, he also claimed that object permanence similarly develops in substages. The object concept forms in the second substage (1–4 months), indicated by the infant starting to follow their movements. She reaches for partially hidden objects by 4–8 months, and for fully hidden objects 8–12 months. By this time, she makes the A-not-B error [1], which disappears by 12–18 months.

Once fully developed, Michotte identifies two indicators of the object permanence understanding [7]:

The Tunnel Effect is the infant's capability of judging when an object, having previously disappeared behind a screen, will reappear again. This indicates the ability to track the object's position even while it's not directly observable. It also depends on the length of the occluded period: Young infants (of 4 months) can track the objects behind sufficiently short screens (< 14.8 cm), but their performance degrades to chance level as the occluded time gets longer. Older infants (of 6 months) can handle longer periods.

The Screen Effect is the surprise of the infant when object A enters behind a screen, and reappears not as itself, but having transformed into object B, indicating she understands the integrity of the object's identity.

B. Neurological Bases of Object Permanence

The close relationship between the maturation of the prefrontal cortex, and the emergence of object permanence understanding, has attracted much attention from neuroscientists [8]–[12]. Diamond and Goldman-Rakic [8] are one of the first to the demonstrate the link between the maturation (or integrity) of the dorsolateral prefrontal cortex and successful performance at the A-not-B task. They conduct a longitudinal study of infants performing the A-not-B task, as well as of adult rhesus monkeys with bilateral prefrontal and parietal ablations. They note a significant performance increase between 7.5-9 and 12 months, since the delay necessary to elicit the A-not-B error increases from 2-5s to 10s. In addition, monkeys with bilateral ablations of DL-PFC perform at the level of 7.5-9-month-olds, while unoperated and parietally operated monkeys are as successful as 12-month-olds; showing the direct dependence of A-not-B task on DL-PFC maturation.

Imaruoka et al. [11] and Saiki [12] introduce a novel paradigm, called the multiple object permanence tracking task, in which objects are moving in a display. They are distinguishable by their features, such as color and shape. The participants are required to track the objects, while also maintaining their features mentally. In this dynamic environment, Saiki [12] shows that our ability to keep bindings of objects color, shape and spatiotemporal locations was significantly impaired when objects move. Even though the visual short-term memory is generally assumed to be capable of maintaining 3-5 feature bound object representations, when the objects are on the move, this ability regresses down to 1 or 2 objects. Employing the same paradigm in an fMRI experiment, Imaruoka et al. [11] demonstrates the activation of anterior prefrontal cortex.

One thing significant about the prefrontal cortex is that it is abundant with recurrent loops, both intrinsic [13], and through other brain areas [14]. The generally accepted hypothesis is that these recurrent loops are the key structure to keep track of time concept in sequential events [15].

C. Robotics Studies

Chen and Weng [16] propose a value-based behavior to develop a rudimentary object permanence. The system is hard-coded to (1) be "surprised" when events are incongruent with its predictions, and (2) gaze longer upon surprising events. After habituation, it gazes longer at events which violate object permanence principles. Roy et al. [17] propose a mental imagery system for the robot, with a global physical model of itself, the objects, and the human partner. The system has an object tracking module, which maintains invisible objects for some time, and dropping ones that are hidden for too long.

A highly relevant work is the MTRNN model by Yamashita and Tani [18]. MTRNN is composed of two groups of contextual neurons, one group with a slow learning timescale, and one with a fast learning timescale. The fast neurons adapt themselves to rapid changes in the environment, thus discovering motion primitives, while the slow neurons learn to discern the context, thereby learning the sequence of necessary primitives to perform a certain behaviour. The major downside is that the slow neurons must be set to a certain discriminative initial state, both to learn, and to reproduce a certain sequence. Therefore, even though it uses the same idea of separating fast and slow signals, MTRNN needs a level of supervision that is not available in the classic object permanence scenario.

D. Slow Feature Analysis

Wiskott and Sejnowski [1] take a novel approach to visual perception. Through a rigorous mathematical procedure called Slow Feature Analysis, they extract the slowest signals carrying most information about the scene. These signals have a total ordering, allowing the selection of the slowest and most informative ones. The resulting system turns out to be highly robust, with some extra features emerging as well. Many identical SFA modules can be stacked together hierarchically, enabling feasible processing and parallelization of high-dimensional images. The system develops invariant object recognition [2]. It can withstand (possibly multiple) transformations such as translation, rotation, and scaling, distinguishing known and novel objects, while also providing insight about the transformation. For instance, in case of multiple moving objects, the system not only distinguishes different objects, but it can also identify the position of any of them. It is not negatively affected by multiple co-occuring transformations (translation, rotation, and/or scaling), rendering it suitable for real-life scenarios, where transformations do not generally occur in isolation. It is also mathematically treatable [3], and it is possible to predict the exact shape of outputs that will result from each of these transformations.

An interesting feature is the biologically plausible prop-

[1] In the A-not-B task, an object is first hidden at a location A several times, until the infant learns to retrieve it successfully. Afterwards, it is visibly taken away from A and moved to a second hidden location B, however infants at this stage still try to retrieve it from the previously learnt location A.

erties that emerge. For instance, the nodes self-organize to behave like simple and complex cells of the visual cortex [4]. When trained with natural life-like scenes, they adapt to prefer Gabor-filter-like inputs, responding maximally to certain directions, and minimally to others. In addition, certain nodes self-specialize to display end-inhibition or side inhibition, again similarly to specialized V1 complex cells. These adaptations are purely due to the input characteristics: Since the visual sequences are natural, they bear spatial and temporal continuity, resulting in these preferences. In yet another study, Franzius et al. [19] show the emergence of hippocampal place cell, head direction cell, and spatial view cell-like formations, which have specialized in rodents to represent its spatial state. Again, they emerge completely due to the nature of the input.

SFA has also been successfully employed for practical purposes. Zhang and Tao [5] propose an SFA-based system to recognize human actions. They also introduce three variants: (1) supervised, (2) discriminative, and (3) spatial discriminative. Kompella et al. [20] devise incremental and online deduction of slow features, while retaining computational feasibility in the face of high-dimensional input. The original SFA approach requires a (costly) offline learning phase, therefore this is an important step for real-time applications.

To the best of our knowledge, there has been no a priori studies to enhance this system with recurrence, nor any attempts to carry a trace of activation through time. The previous studies of slow feature analysis are purely reactive in time.

III. Methods

A. Slow Feature Analysis

Wiskott and Sejnowski [1] formalize the following optimization problem: Given an I-dimensional input signal, $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_I(t)]^T$, the objective is to find a set of input-output functions, $\mathbf{g}(x)$, which will produce a J-dimensional output signal $\mathbf{y}(t) := \mathbf{g}(\mathbf{x}(t))$, whose components vary as slowly as possible, while still containing information. The objective is to minimize $\langle(\dot{y})_j^2\rangle$, $\forall j \in 1, ..., J$, with:

$$\begin{aligned} \langle y_j \rangle &= 0 \quad \text{(zero mean)}, & (1)\\ \langle y_j^2 \rangle &= 1 \quad \text{(unit variance)}, & (2)\\ \forall j' < j : \langle y_{j'} y_j \rangle &= 0 \quad \text{(decorrelation)}. & (3) \end{aligned}$$

The angular brackets indicate averaging over time.

The unit variance constraint avoids the trivial solution with zero information content. The decorrelation constraint ensures non-redundant signals. It also enforces a total order: The smaller the index j is, the more optimal is the solution y_j.

This optimization problem is difficult to solve, but it can be simplified by constraining the output functions to be linear combinations of a finite set of nonlinear functions, that is, $y_j(t) = g_j(\mathbf{x}(t)) := \mathbf{w}_j^T \mathbf{z}(t)$. The nonlinear functions $\mathbf{z}(t)$ can be obtained via applying a set of functions $\mathbf{h} = [h_1, ..., h_K]$ on the input signals, thus expanding them nonlinearly: $\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t))$. After this nonlinear expansion, the problem can be treated as linear in the expanded signal components $z_k(t)$, similar to using a kernel to linearize the classification problem.

Then the problem reduces to finding the weight vectors $\mathbf{w}_j = [w_{j1}, ..., w_{jK}]^T$ to minimize $\langle \dot{y}_j^2 \rangle = \mathbf{w}_j^T \langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \rangle \mathbf{w}_j$.

Assuming that the functions h_k are chosen such that the expanded signal $\mathbf{z}(t)$ has zero mean and unit covariance matrix ($\langle \mathbf{z} = 0 \rangle$ and $\langle \mathbf{z}\mathbf{z}^T = I \rangle$), the constraints:

$$\begin{aligned} \langle y_j \rangle &= \mathbf{w}_j^T \langle \mathbf{z} \rangle = 0, \\ \langle y_j^2 \rangle &= \mathbf{w}_j^T \langle \mathbf{z}\mathbf{z}^T \rangle \mathbf{w}_j = \mathbf{w}_j^T \mathbf{w}_j = 1, \\ \forall j' < j : \langle y_{j'} y_j \rangle &= \mathbf{w}_{j'}^T \langle \mathbf{z}\mathbf{z}^T \rangle \mathbf{w}_j = \mathbf{w}_{j'}^T \mathbf{w}_j = 0, \end{aligned}$$

are fulfilled if and only if the weight vectors form an orthonormal set. Therefore the set of eigenvectors of $\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \rangle$ gives us the weight vectors that satisfy the constraints. From these eigenvectors, we choose the ones with the smallest eigenvalues as the weight vectors, $\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \rangle \mathbf{w}_j = \lambda_j \mathbf{w}_j$,with $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_J$, resulting in the input-output functions: $g_j(x) = \mathbf{w}_j^T \mathbf{h}(x)$.

In other words, to find the slowest signal, we use the eigenvector of the smallest eigenvalue, corresponding to the direction of the least variance in the time derivative of the input. For other signals, orthogonal directions can be used, given by eigenvectors of increasing eigenvalues. They are found by a principle component analysis on the matrix $\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \rangle$.

For nonlinear expansion, Wiskott et al. use the first and second-degree monomials of the input: $\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t)) = [x_1(t), \cdots, x_I(t), x_1(t)x_1(t), x_1(t)x_2(t), \cdots, x_I(t)x_I(t)]^T$. Higher order expansions are possible, but not necessary, since a hierarchical architecture results in increasing complexity in higher-levels, performing this expansion in every layer.

Notice that the outputs signal are computed instantaneously, i.e., they are not a result of simple temporal low-pass filtering. Hence, the optimization problem is being solved by instantaneously calculating a higher level representation.

B. Recurrent SFA

We use a hierarchical architecture composed of SFA nodes (Figure 1) [2]. The input images have a resolution of 65x65. The bottom layer reads from the input, and is formed of 15x15 SFA nodes, each with a 9x9 receptive field, among which 5 pixels overlap. The higher 3 levels have 7x7, 3x3, 1x1 nodes respectively, all but the last one with 3x3 receptive fields. This part of the architecture is proposed in previous studies [2], and called thereupon **Feed-forward SFA** for clarity.

We extend Feed-forward SFA with an extra 1x1 layer on top, which feeds its output at time t is back to itself at $t + \Delta t$. The new architecture is called **Recurrent SFA**. The input to the top (n^{th}) layer at time t, $\mathbf{x}^n(t)$, becomes:

$$\begin{aligned} \mathbf{x}^n(t) = [&y_1^{n-1}(t), y_2^{n-1}(t), ..., y_J^{n-1}(t), \\ &y_1^n(t - \Delta t), y_2^n(t - \Delta t), ..., y_J^n(t - \Delta t)]^T. \end{aligned}$$

where $\mathbf{y}^n$ is the output of the n^{th} layer.

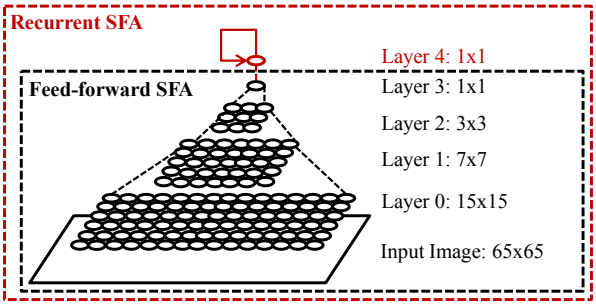


Fig. 1: The architecture of Feed-forward and Recurrent SFAs.

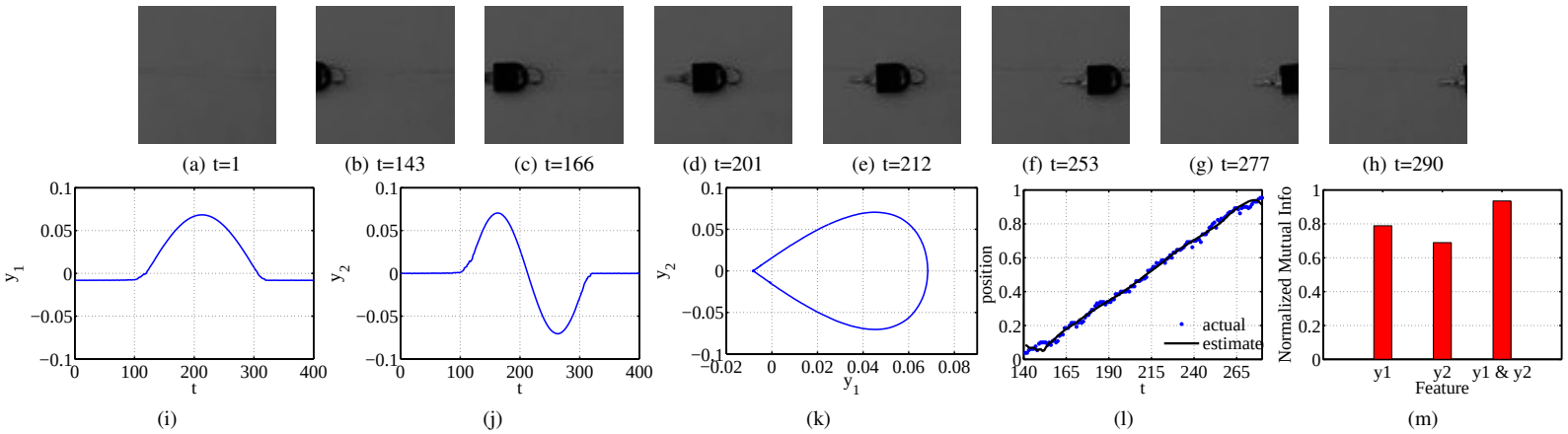


Fig. 2: The feed-forward SFA response to a single moving object. (a-h) Object enters into view from left at t=110, leaves from right at t=320. (i-j) The slowest two responses of the system, y_1 and y_2. (k) Phase diagram y_1 vs. y_2. Every point on the phase diagram corresponds to a unique location of the object on the retinal plane. (l) Actual versus estimated position values of the object. (n) Mutual information between the (actual) position values of the object, and y_1, y_2, and $y_1\&y_2$.

The time difference Δt, as the single parameter of the system, determines the maturity of the emulated prefrontal cortex, and hence called the "maturity" parameter. The smaller it is, the "younger" the system will be, and recall only the near past. As the system gets more mature, it can be increased to allow a longer window.

Note that a hierarchical structure: (1) maintains feasibility by restricting the input matrices of each node to a constant size, (2) enables parallel processing, (3) forms a biologically accurate model of the visual cortex, with strongly position-dependent lower-level cells, and position-independent higher-level cells. Furthermore, higher-level cells can represent increasingly more complicated input-output functions (starting with degree of 2 at the lowest-layer, and increasing as 4, 8, 16, and so on.)

IV. Experimental Results

The experiments are divided into two sets to distinguish capabilities that are already offered by feed-forward SFA, versus the newly introduced ones. In the first set, we demonstrate feed-forward SFA in various cases, such as a single moving object, a single object that disappears and reappears again, and multiple objects moving around. These are also interesting as a proof-of-concept that the original SFA approach is feasible for object recognition in real-world images. The second set demonstrates recurrent SFA in an object permanence scenario. Specifically, we show that, when recurrence is introduced, the tunnel and screen effects emerge. We further demonstrate how it is possible to model an increasingly mature prefrontal cortex, by manipulating the single parameter. For each set, same object and behavior was used for both training and testing.

A. Feed-forward SFA

The first experiment shows the response of the feed-forward SFA to an object traversing the x-axis from left to right (Figure 2a-h, data was grayscaled to remove the color cue, which makes classification too easy for different objects.) This set is important for establishing a basis of the output shapes. Figures 2i and 2j show the slowest two signals, whose shapes are exactly as predicted by the theoretical analysis [3]. Let $[t_A, t_B]$ denote the whole experiment duration, and $[t_a, t_b] \in [t_A, t_B]$ a time interval in the experiment during which the object is visible. A single pattern is visible during $[t_a, t_b]$, and is out of the view during $[t_A, t_B]\backslash[t_a, t_b]$ ($\backslash$ indicating set difference). The case with $t_a \neq t_A$ and $t_b \neq t_B$, is called a *bounded* case, since the output must equal to a constant c_1 all during the interval $[t_A, t_B]\backslash[t_a, t_b]$, given that the system sees the (approximately) same background all the while. Due to the zero mean constraint (Equation 1), c_1 tends to 0 in the limit $(t_B - t_A) \to \infty$. The analysis predicts that the slowest signal (y_1) should be a half cosinus, with the second slowest signal (y_2) being a sinus of a single oscillation. (The other signals which are not shown here are cosinuses and sinuses of increasing oscillations.)

The slowest two signals have a significance: They predict the object's 1D position uniquely. On the phase diagram of y_1 vs. y_2 (Figure 2k), every point corresponds to a single position on the x axis. This is because, as shown previously, the SFA outputs reflect the main underlying free variable causing the change in the system, called the **configuration variable**, which in this case is the position. Exact position values can be estimated via a simple regression [2]: Figure 2l depicts the actual and regressed position values.

Ideally, one would like to predict the states of the configuration variables based on the outputs. However which output combination would be necessary or sufficient is not automatically given by the network. For instance, in this case, notice that y_1 on its own is not enough to retrieve the position values, and neither is y_2, due to nonlinearities of both signals. Here, a combination of the two is sufficient. However different transformations need different outputs to be combined. When there is more than one configuration variable, this can be even more complicated: In one case in [1], where both position and identity are changing, a combination of y_1 and y_3 estimate the position, while y_2 and y_4 estimate the identity. So far, a qualitative (human-supervised) assessment have been used to decide. We propose using mutual information for a quantitative assessment, without supervision. Specifically, we calculate the mutual information between all the output combinations, and

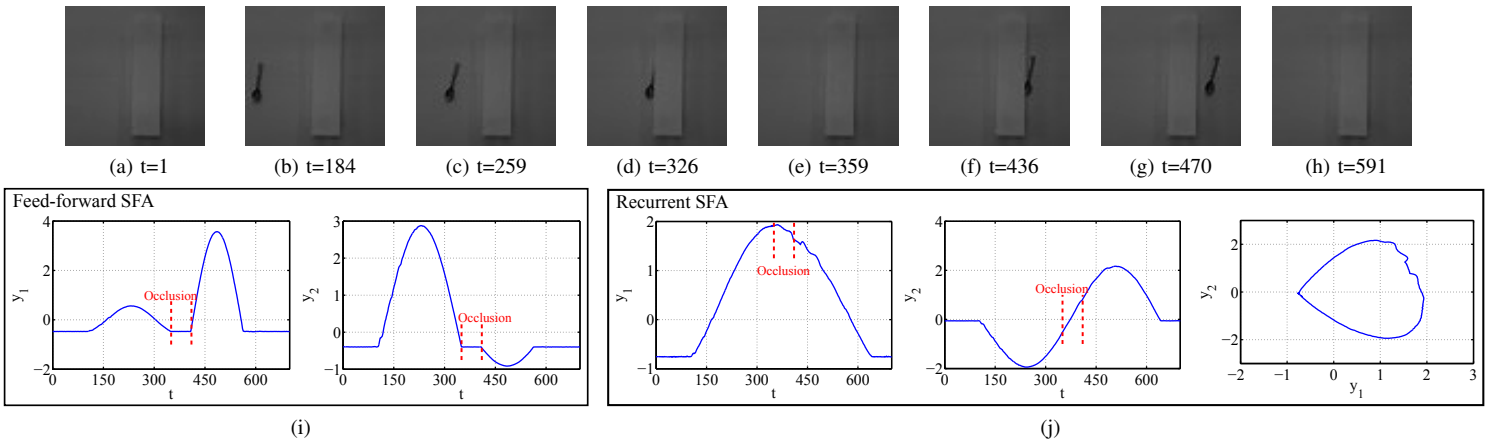


Fig. 3: A single moving object which is occluded for some time (between t=350 and 410) during the trial. (a-h) Snapshots from the input images. (i) Feed-forward SFA. (j) Recurrent SFA and the "Tunnel Effect".

the position values, then perform a thresholding to select the minimum sufficient number of outputs. Figure 2m displays the mutual information provided by y_1, y_2 and $y_1\&y_2$. As expected, $y_1\&y_2$ is sufficient for this case.

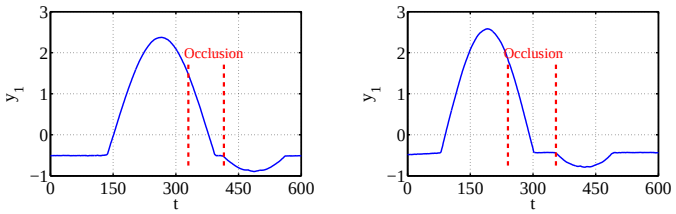
The second experiment stands as a proof-of-concept: As the object is traversing the retinal plane, it disappears behind a screen. It continues to move behind the screen with a constant velocity, and reappears in due time (Figure 3a-h). As expected, due to the reactive nature SFA, as soon as it disappears, the SFA outputs diminish to 0, and on its reappearance they increase again (Figure 3i).

A final issue is the response of the system to more than one object. In this case, there are two objects, the first one in view at t=100-360; the second one at t=710-880. Both are occluded shortly, the first between t=230-250, and the second between t=820-830 (Figure 4a-k). As predicted, the system develops highly object-dependent outputs (4l-n). It is still possible to estimate the position of the objects, but in addition, the outputs also code the identity of the object at any time. For instance, a positive y_1 response during the first visible interval distinguishes the first object from the second one, which has a negative y_1 response for that interval. As shown in [2], a kNN classifier with $\approx 95\%$ success rate can be trained to estimate the identity (Figure 4o).

B. Recurrent SFA

When a recurrent input is added, the system begins to behave similarly with infants with maturating prefrontal cortex. The first indicator is an ability of tracking the position of an object behind a screen. This is demonstrated by the child's ability to guess when it will become visible again (the **tunnel effect**). Figure 3j demonstrate the occluded object case with recurrent input. Recurrent SFA is able to retain its activation throughout occlusion, giving a comparable phase diagram y_1 vs. y_2 with the visible case. This means we can "track" the position of the object uniquely, even through occlusion.

Psychological studies indicate that the tunnel effect depends on the length of the "tunnel". Younger infants are successful for short tunnels only, while older infants can manage increasingly longer ones. A similar effect is observed in Figure 5 with two longer tunnels. Keeping the maturity parameter constant, there is a limit to the occluded period which can be compensated, similar to infants.



(a) Tunnel with size 10 cms. (b) Tunnel with size 14 cms.

Fig. 5: As the tunnel gets longer, Recurrent SFA is no more able to sustain the signals from vanishing.

The final indicator of a mature understanding of object permanence is the **screen effect**, in which the infant maintains the object's identity. She is surprised if a different object reappears from behind the screen. To demonstrate the effect, we show that the architecture has difficulty adjusting when a different object reappears, in which case its predictions collide with the apparent stimuli, resulting in a "surprise". Figure 6a demonstrates the feed-forward case: When Object A disappears behind the screen, and reappears having changed into Object B at time 150, the system responds immediately. Figure 6b demonstrates the recurrent case, with maturity parameters of $\Delta t = 20$ and $\Delta t = 40$, where the system needs time to adjust itself to the changed object. The delays, in which the system insists on seeing Object A, indicate an expectation that the object's identity should have been preserved.

V. Conclusion

We have shown how slow feature analysis, previously shown to develop the object concept, can be extended with a recurrent loop to retain information through time. The proposed extension mimics an important developmental stage, the object permanence understanding. We argue that the building of one ability on top of another is reminiscent of the way humans maturate. The resulting system can predict an occluded object's movements, as well as keeping in mind its identity. These abilities are not infinitely powerful: After a long enough occlusion, they give in, just as in infants. Our study also serves as a minor contribution to the SFA framework: We demonstrate

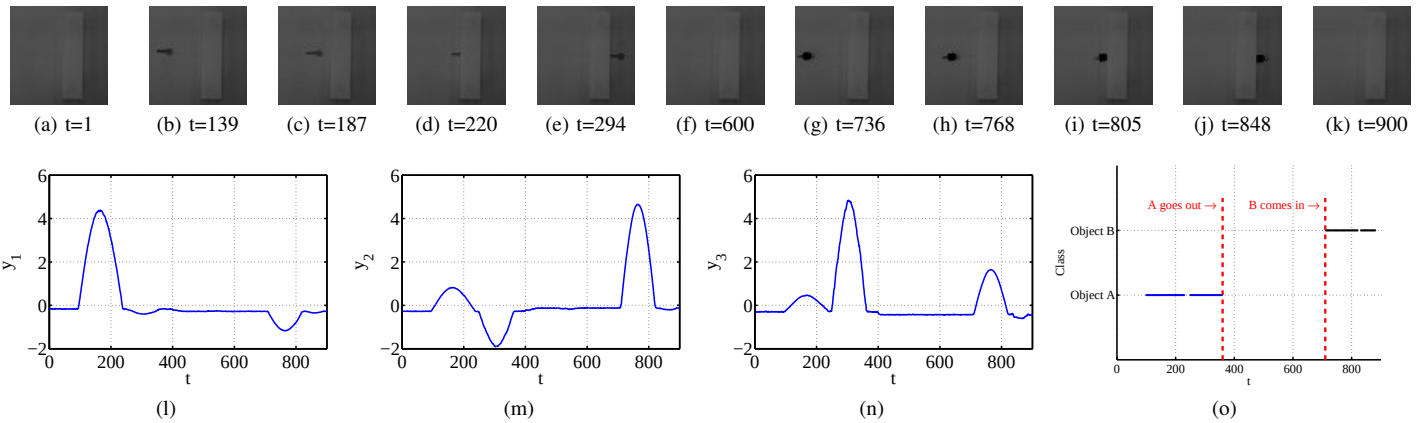
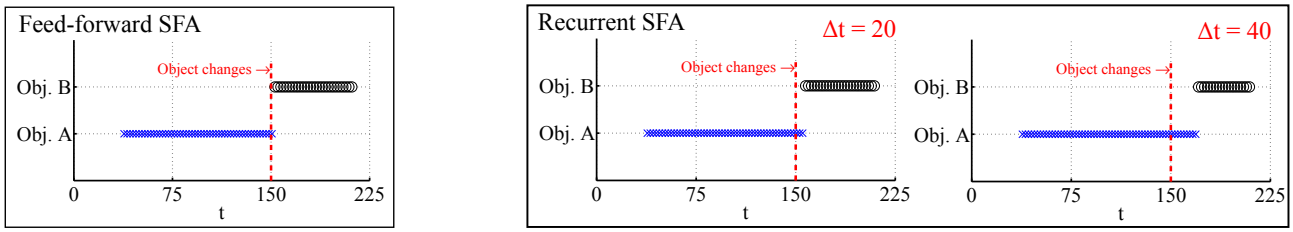


Fig. 4: The feed-forward SFA response to two objects presented sequentially, both of which are occluded for some time. (a-k) The first object is in view from t=100 to 360, occluded between t=230 and 250; the second object is in view between t=710 and 880, occluded between t=820 and 830. (l-n) The slowest three responses. (m) kNN classification of object identity.



(a) Feed-forward case

(b) Recurrent case

Fig. 6: The "Screen Effect". The identity classification made by the recurrent SFA. Object A is changed with object B at t=150. (a) Feed-forward case. (b) Recurrent case with maturity parameters $\Delta t = 20$ and $\Delta t = 40$. Notice that as the maturity parameter increases, the system takes more time getting over its reluctance to accept the chance.

how mutual information can estimate the sufficient outputs, as well as validating SFA for recognizing real-world objects.

An interesting question is whether a gradually increasing maturation parameter will boost cognitive development, since as shown repeatedly, initial limitations of our body promote development by restricting the complexity. The effect of working memory restrictions, other levels of recurrence, and real robotic applications are all promising future directions.

References

[1] L. Wiskott and T. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Comp.*, vol. 14, pp. 715–770, 2002.

[2] M. Franzius, N. Wilbert, and L. Wiskott, "Invariant object recognition and pose estimation with slow feature analysis," *Neural Comp.*, vol. 23, pp. 2289–2323, 2011.

[3] L. Wiskott, "Slow feature analysis: A theoretical analysis of optimal free responses," *Neural Comp.*, vol. 15, pp. 2147–2177, 2003.

[4] P. Berkes and L. Wiskott, "Slow feature analysis yields a rich repertoire of complex cell properties," *Journal of Vision*, vol. 5, pp. 579–602, 2005.

[5] Z. Zhang and D. Tao, "Slow feature analysis for human action recognition," *PAMI*, vol. 34, no. 3, pp. 436–450, 2012.

[6] J. Piaget, *The origins of intelligence in children*, 1952.

[7] A. Michotte, "Perception and cognition," *Acta Psychologica*, vol. 11, pp. 69–91, 1955.

[8] A. Diamond and P. Goldman-Rakic, "Comparison of human infants and rhesus monkeys on piaget's ab task: evidence for dependence on dorsolateral prefrontal cortex," *Exp. Brain Res.*, vol. 74, 1989.

[9] M. A. Bell, "Brain electrical activity associated with cognitive processing during a looking version of the a-not-b task," *Infancy*, vol. 2, 2001.

[10] A. Baird, J. Kagan, T. Gaudette, K. Walz, N. Hershlag, and D. Boas, "Frontal lobe activation during object permanence: Data from near-infrared spectroscopy," *NeuroImage*, vol. 16, pp. 1120–1126, 2002.

[11] T. Imaruoka, J. Saiki, and S. Miyauchi, "Maintaining coherence of dynamic objects requires coordination of neural systems extended from anterior frontal to posterior parietal brain cortices," *NeuroImage*, vol. 26, no. 1, pp. 277–284, 2005.

[12] J. Saiki, *Multiple Object Permanence Tracking: Maintenance, Retrieval and Transformation of Dynamic Object Representations*, August 2008, pp. 277–284.

[13] M. Pucak, J. Levitt, J. Lund, and D. Lewis, "Patterns of intrinsic and associational circuitry in monkey prefrontal cortex," *Journal of Comparative Neurology*, vol. 376, pp. 614–630, 1996.

[14] G. Alexander, M. Delong, and P. Strick, "Parallel organization of functionally segregated circuits linking basal ganglia and cortex," *Annual Review of Neuroscience*, vol. 9, pp. 357–381, 1986.

[15] P. F. Dominey, M. Hoen, J.-M. Blanc, and T. Lelekov-Boissard, "Neurological basis of language and sequential cognition: Evidence from simulation, aphasia, and erp studies," *Brain and Lang.*, vol. 86, 2003.

[16] Y. Chen and J. Weng, "Developmental learning: A case study in understanding object permanence," in *Fourth Int. Workshop on Epigenetic Robotics*, 2004, pp. 35–42.

[17] D. Roy, K.-Y. Hsiao, and N. Mavridis, "Mental imagery for a conversational robot," *IEEE Transactions on System, Man and Cybernetics, Part B: Cybernetics*, vol. 34, pp. 1374–1383, 2004.

[18] Y. Yamashita and J. Tani, "Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment," *PLoS computational biology*, vol. 4, 2008.

[19] M. Franzius, H. Sprekeler, and L. Wiskott, "Slowness and sparseness lead to place, head-direction, and spatial-view cells," *PLoS Computational Biology*, vol. 3, no. 8, pp. 1605–1622, 2007.

[20] V. R. Kompella, M. Luciw, and J. Schmidhuber, "Incremental slow feature analysis," in *Proc. of* 22^{nd} *Int. Jnt. Conf. on Artificial Intelligence*, vol. 2, 2011, pp. 1354–1359.