On the influence of sensor morphology on sensory-motor coordination: an information theoretic view

Technical Report

H. Martinez, M. Lungarella, and R. Pfeifer
Artificial Intelligence Laboratory
Department of Informatics
University of Zurich, Switzerland

Abstract

In developmental robotics, there is a lot of work related to learning, which focuses on how to enable robots with adaptive capabilities. However, there is not much understanding of how the morphology of sensors and actuators affects the learning process. By taking inspiration from the human vision system, we explore how the sensor integration, its morphology together with the body of the robot, are the conditions that enable infants, in the first months of life, to achieve behaviors such as attention, vergence, reaching and grasping. Hence the sensor morphology is an important aspect that can be exploited, first to reduce the amount of inputs without disrupting perception, and second it limits the possible actions and behaviors that enable the agent to learn whether an environment model or a specific task. In this report, we present an information theoretic analysis of three experiments showing how foveation and the creation of a single image from two images demand the vergence behavior as a mechanism to increase the information structure not just among the pixels in the image but also among the motor actions and the pixels, establishing the prerequisite for learning. We also speculate on the implications of our results for theories of human development.

1. Introduction

In the field of developmental robotics, there has been increasing interest concerning the creation of algorithms that enable robots or artificial agents not just to cope with changes in the environment but also to acquire new task-independent skills as a living being (Schmidhuber 2009, Oudeyer et. al., 2007, Barto et. al., 2004). However, in all these works, the sensory input is not rich in states, on the contrary they are generally binary, far from what we can observe in nature, and there is little work related to how a rich sensory system should be implemented in order to increase the learning performance.

Living beings are embodied and embedded in their ecological niches. It follows presence of information structure and directed information flow induced by dynamically coupled sensory-motor activity, including effects of motor outputs on sensory inputs. The selection of an appropriate sensory morphology is going to provide an initial quality that restricts the possible actions and behaviors, because not all the actions executed by the agent can provide a complex structured information full of relations and dependencies among the sensory system and the actions, which are critical for learning, action selection, adaptability and developmental process (Körding et al., 2006, Thelen et al., 1994). This relation of the sensors, and the different kind of interactions that the agent can engage in, are the source of general principles to design robots able to extend their sensory-motor competences during their life.

To study what features are necessary to bootstrap adaptive behaviors, we used the active vision head of the robot platform iCub (Beira et al., 2006). In the first months of life, a child is able to develop sensory-motor competences almost from scratch (Smith et al., 1998). Behaviors such as tracking, saccadic movements and fixation start to develop at the beginning of a child's life and are mature after about three months (Tondel et al., 2007; Aslin, 1977). This report presents through different experiments why behavior such as vergence increases the information structure of the robot, which means that the statistically dependencies as well as causality (Pearl, 2000; Pearl, 2009) relations among the image pixels and the actions are strengthened, and therefore establishes all the preliminary relevant aspects for any kind of learning. However, this is just a feature that could be exploited thanks to the morphology of the sensor and the coupling between different sensory systems.

The learning allows the infant to predict because through its sensory stimulus and its life history intrinsically there is knowledge concerning the world, and about what could happen, this information is used to improve its actions and reactions. This drives us to think that at the end, our mind is just an approximator that exploits a relation between the actions and the sensory input. However, what is more interesting, is that not all the possible actions are going to give us the possibility to learn, and our mind has to select the appropriate set of actions that increase the relation between them. In addition these actions are going to depend in the characteristics and morphology of the sensors. Following this idea the relations among actuators and sensors and going to be constrained by theses aspects. In order to measure how much the agent can learn given a specific sensory morphology we use the information measurements based on Shannon entropy (Shannon, 1948, Cover and Thomas, 1991).

In the case of vergence the coupling exploited in our experiments is between the visual system and the propioceptive system. The results presented here enlighten the possible extension of this approach, in term of the development of the attention systems based not just in the visual data but in the relations among different sensory systems. This allows the robot to develop its own sensory-motor coordination to be able to learn, to predict and exploit the sensory system. The development of the attention system then enables the agent to extract the information relevant for its own tasks providing the substrate for the emerging of behaviors like eye hand coordination.

Even though the complexity of the robot head, or the data acquired with the cameras do not have all the variables of the human vision system, the principal features are taking into account in the development of the experiments. The results of the experiments present how the behavior is affected thanks to sensor morphology. This helps understanding why the features presented in our natural sensory system, our body, and the motor system are the key property that defines our behaviors such as vergence.

This report is organized as follows. First, we describe the robot head platform used for our experiments, the sensory morphology, and each informational measure employed to quantify the results in the experiments. Then, we present the experiments and the results of each one. Before concluding the report, we discuss our results and some of their implications for theories of infant development.

2. Materials and Methods

Robot. The RobotCub project developed an open humanoid platform, iCub. This platform is the tool to develop studies in cognitive systems and embodied cognition. Our experimental test bed was the 6 degrees of freedom (DOF) iCub robot head described in (Beira *et al.,* 2006). A difference among this and robots such as QRIO, ASIMO, HOAP-2, are the 3 DOF to pan and tilt the two DragonFly cameras. The image delivered by each camera has a resolution of 640x480 at 30 fps. Moreover, both eyes can pan independently thanks to a belt system with a motor behind the camera. The common tilt movement is actuated for another belt system placed between the cameras; all the belt systems have a tension adjustment mechanism. The motors used are Faulhaber DC micromotors, equipped with optical encoders and planetary gearheads. The other 3 DOF are used to control the neck of the head in a configuration that best represents the human neck movements, the motors used are Faulhaber DC micromotors with planetary gearheads and magnetic encoders. The motor control boards are integrated in the head and connected to the computer through a CANbus.

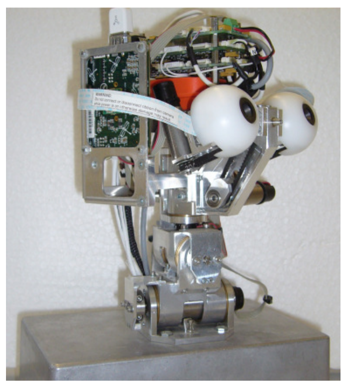

Figure 1: iCub robot head.

Sensor morphology and control. The human vision system has to interpret a 3D world from 2D projections, and in this process the ocular motions play an important role, these motions are not defined as an intrinsic feature, but are developed through the interaction with the environment, moreover abilities such as stereopsis (depth perception from binocular vision that exploits parallax disparities) that allow the depth perception are a result of this development in the first months of life (Birch et al. 1996 and Birch et all. 2005). The question is what mechanism drives this process, and what could be the contribution of eyes morphology and muscle to this.

The iCub robot head was selected as a test bed for the experiments. It has an appropriate number of DOF in order to emulate behaviors like vergence, smooth pursuit, and saccades, typical of the vision system. Taking advantage of the results that show that neurons respond to simple features such as intensity contrast, color, orientation, and motion (Nothdurf 1990), which defining the pre-attentive visual cues (Itti & Koch 2001), computed in parallel and in separated cortical streams (Dacey 1996) in the brain. Color was the main feature selected as a measure in all the experiments. In addition, there is also a well known visual illusion produce by the binocular single image phenomena (Wheatstone, C 1838), where if you focus your attention in an object in front of your nose and then you put your thump between your nose and the object, keeping your focus in the previews object then you should see two thumbs and if you now focus the thumb, then the previews object is going to appear twice. Because the image in our mind is a combination of the images from both eyes. In our implementation we applied the average of both cameras as the image function that happens in our mind. Another important characteristic is the foveation. Our eye has in the center a greater amount of receptors and these decreases with the radius. This was modeled with the log-polar transform [1], the transformation changes the coordinate system from Cartesian (x,y) to the logarithm of the magnitude and the angle:

$$\rho(x, y) = M \cdot \log(\sqrt{x^2 + y^2})$$
$$\phi = \arctan(y / x)$$

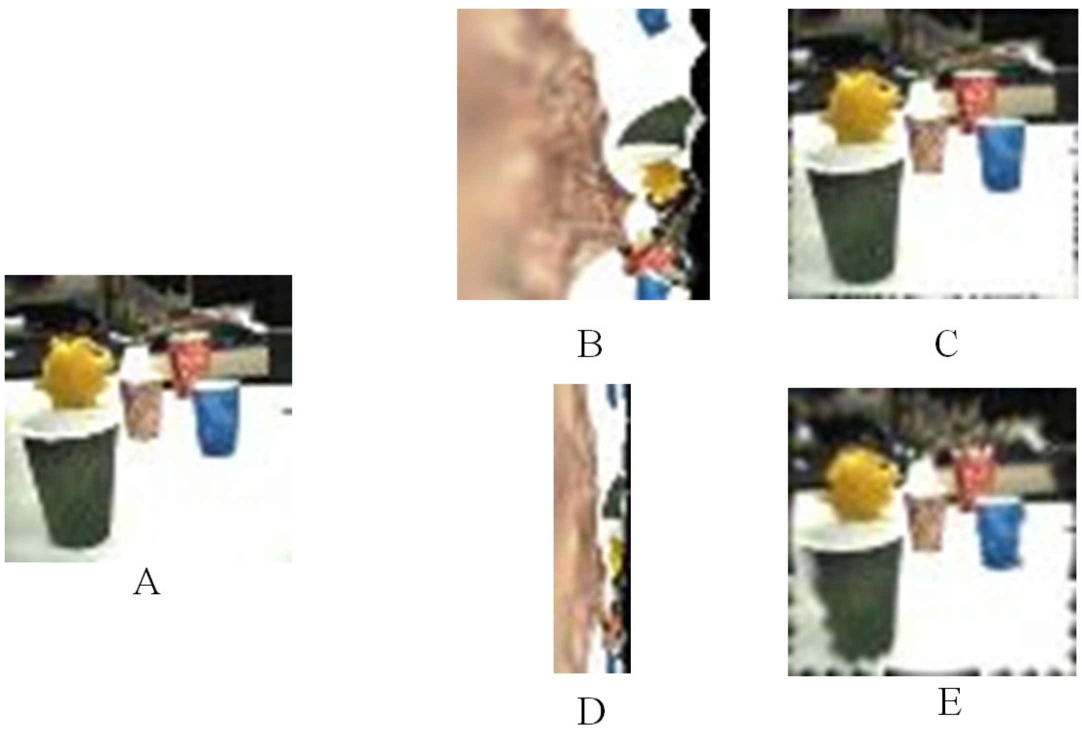

Figure 2: Log-Polar Transform of 60X60 Image. (A) Raw image. (B) Log-polar transform with M = 20. (C) Inverse log-polar transform. (D) Log-polar transform with M=12. (E) Inverse log-polar transform.

The parameter M was used to increase or decrease the amount of pixels used in the log-polar space. In our experiments, we used color, foveation and image composition from the two cameras to find out whether the vergence behavior increases the information structured and how the agent can learn to exploit it.

Information metrics. In order to present how the association and causal relations among the variables (actuators and sensors) are going to depend on the morphology and specific behaviors we have adopted five measurements, all of them fundamentally based on Shannon entropy (Shannon 1948, Cover and Thomas1991). Entropy, mutual information, integration, and complexity (Lungarella *et al.,* 2005), which measure statistical regularities among random variables without taking into account temporal precedence. Transfer Entropy is the fifth measurement used to quantify causal relations (Schreiber 2000). These measurements were

[1] Implemented in OpenCv

selected to compare the results in the experiments, because with them it is possible to find all the nonlinear statistical patterns and understand why a specific behavior could give better relations among the data.

Shannon entropy: measures the average uncertainty, or information. Given a discrete time series *x(t)* that can have N different states, it can be calculated using the state probability distribution according to:

$$H(x) = -\sum_{i=1}^{N} P_x(i) \log P_x(i)$$

Where $P_x(i)$ is the probability of *x(t)* being in the ith state. When the uncertainty is maximal the entropy is maximal (uniform distribution), while deviations from equiprobability states result in lowered entropy (increased order and decreased uncertainty).

Mutual information: measures the deviation from statistical dependence between two or more random variables, quantifying the error we make in assuming X and Y as independent variables. The formal definition of mutual information in terms of single and joint state probability distributions is

$$MI(x, y) = -\sum_i \sum_j P_{xy}(i, j) \log \frac{P_x(i) P_y(j)}{P_{xy}(i, j)}$$

If X and Y are two statistically independent random variables, $P_{xy}(i, j) = P_x(i)P_y(j)$ and MI(X,Y) =0. For this reason any statistical dependence between X and Y yields MI(X,Y) > 0. However in general, the mutual information is insufficient to disclose directed interactions (e.g., causal relationships) between X and Y, or between Y and X.

Integration: is the multivariate generalization of mutual information (McGill 1954) and captures the total amount of statistical dependency among a set of random variables X_i forming elements of a system **X**=$\{X_i\}$. Integration (Tononi et. all 1994) is defined as the difference between the individual entropies of the elements and their joint entropy:

$$I(X) = \sum_i H(X_i) - H(X)$$

As for mutual information, if all elements Xi are statistically independent, I(X) = 0. Any amount of statistical dependence leads to I(X) > 0.

Complexity: If a system X has positive integration, and also it has locally segregated dependencies we would expect to find statistical dependence among units at specific spatial scales. A system combining local and global structure has high complexity:

$$C(X) = H(X) - \sum_i H(X_i \mid X - X_i)$$

Where $H(X_i \mid X - X_i)$ is the conditional entropy of one element X_i given the complement $X - X_i$ composing the rest of the system.

Transfer entropy: is the measure used to disclose the directed flow or transfer of information (also referred to as "causal dependency") between time series (Schreiber 2000). Given two time series X_t and Y_t, transfer entropy essentially quantifies the deviation from the generalized Markov property: $p(x_{t+1} \mid x_t) = p(x_{t+1} \mid x_t, y_t)$, where p denotes the transition probability. If this deviation is small, then Y does not have relevance on the transition probabilities of system X. Otherwise, if the deviation is large, then the assumption of a Markov process is not valid, because Y influence the transition of system X.

$$T(Y \to X) = \sum_{X_{t+1}} \sum_{X_t} \sum_{Y_t} p(x_{t+1}, x_t, y_t) \log \frac{p(x_{t+1} \mid x_t, y_t)}{p(x_{t+1} \mid x_t)}$$

Where the sums are over all amplitude states, and the index T(Y → X) indicates the influence of Y on X. The transfer entropy is explicitly nonsymmetrical under the exchange of X and Y—a similar expression exists for T(X → Y)—and can thus be used to detect the directed exchange of information (e.g., information flow, or causal influence) between two systems. As a special case of the conditional Kullback-Leibler entropy, transfer entropy is non-negative, any information flow between the two systems resulting in T > 0. In the absence of information flow, i.e., if the state of system Y has no influence on the transition probabilities of system X, or if X and Y are completely synchronized, T(Y → X) = 0 bit.

3. Experiments

Experiment 1. In the first experiment we tested how the morphology of the sensor can help increase the information structure in one region of the image and in the region in the space defined by vergence. The robot looked from the side of a rotating table with a cup on it, where the angle of vergence α was fixed (Fig. 3). We tested four different sensor configurations: (1) the average of the left and right image. (2) The inverse log polar of the average of the left and right image. (3) The log polar of the average of the left and right image, and (4) a single image, the left camera.The results show that the center had more structure than the rest of the image due to the fact that the cup in the center is not changing its size with the rotational movement of the table. In Figure 4a we can appreciate how pixels outside the center have the cup small behind and big in the front.

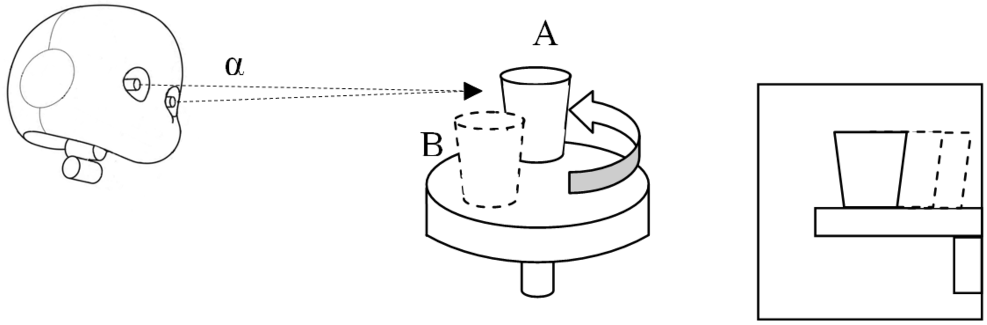

Figure 3: Experimental Setup. The robot is looking at a fixed area over the rotating table, when the cup is not in this area it is blurred.

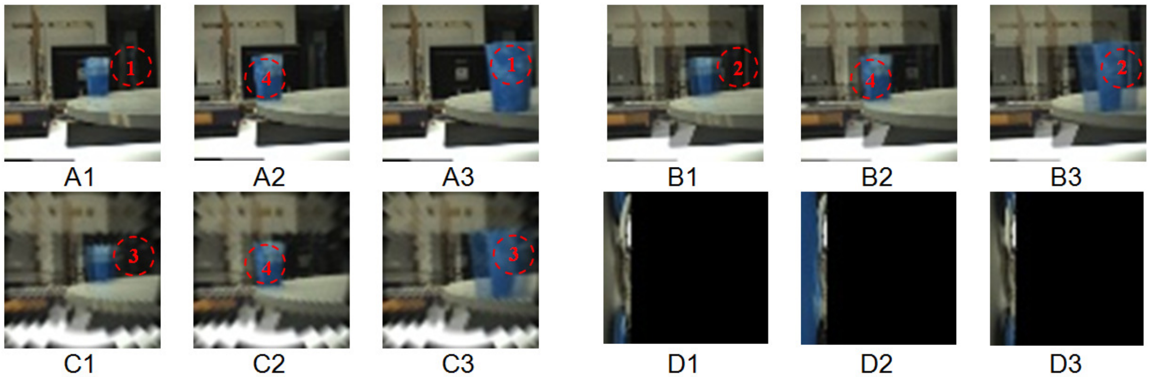

Figure 4: Data from the Experiment. These are the frames saved in the experiment, they show a cup on a rotating table that is coming from behind and going out to the front. (A) The frames from the left camera. (B) The average images from the left and right. (C) Average inverse log polar transform. (D) Average log polar transform. (1) The pixels far from the center (e.g. dotted circle) have less information structure than the pixels in the center because the object is changing the size in the image for the rotation. (2) The average between left and right images, introduce a distortion in the pixels that are not in the vergence area (e.g. dotted circle). (3) The log polar transform has less receptors far from the center increasing the distortion outside the center (e.g. dotted circle). (4) The center is not affected by the log polar transform neither for the average because the object is in the vergence region (e.g. dotted circle).

The second configuration was the average of both images left and right (Fig. 5), with this operation the pixels far from the vergence point are blurred (Fig. 4b). In our experimental setup we fixed the center of the image also in the vergence point. For this reason the pixels in the center are not distorted. In the Fig. 4c we can see inverse log polar transform of the average image, here the center has no distortion but the pixels far from the center are blurred, thanks to the log polar transformation, which takes a sub group of pixels (Fig. 4d), this transformation decrease the number of pixels far from the center decreasing the information of the image, but keeping the center not distorted.

We run 8 different trials for 4 different M parameters of the log polar transform in order to evaluate the impact of this transform on the information structure. In each trial we saved 12300 frames (60X60 pixels). The calculation of the entropy was done based on the probability density function (PDF) of the normalize colors (green, red, blue and yellow), which define the dimensions of the PDF. We used 8 bins for each dimension. To calculate mutual information, integration and complexity, we used statistical formulae (Cover et al. 1991) that allow the calculation of entropies from the covariance matrix, under the assumption that these covariances were generated by a stationary Gaussian random process. All samples were examined for Gaussian state distributions (by fitting state histograms) as well as stationarity (by ensuring stable means and standard deviations across time).

In this way using an object in the 3D space in front of the robot we were able to measure how the pixels' information structure changes, given different sensory morphologies that take in account the distribution of the receptors in the camera (log polar transform) and vergence (induced by the average of both images).

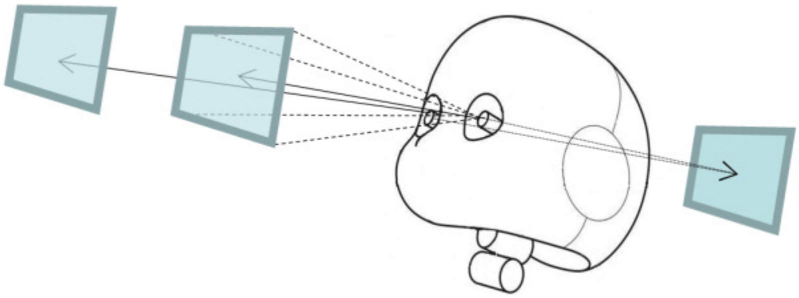

Figure 5: Image Combination. A new image is created by averaging the left and the right image.

The results of this experiment can be understood by taking into account the "sensor morphology transformations" (that is, the average of left and right image, the average of the inverse log-polar transform, and the average of the log-polar transform). On the one hand, the average is introducing distortions in the image proportional to how far the object is from the vergence point. As can be seen in the Figures 6 to 8, the pixels in the center keep its structure, but the pixels outside decrease its structure, and this is useful for the robot because reduce the variability in the structure given that what really matters is what it is happening in the vergence point. On the other hand the log-polar transform is helping reduce the amount of pixels that represent the image, and given that the bigger amount of receptors is also in the vergence point then these pixels are exactly selecting the pixels with more structure among them.

The variance in the left and right image have a bigger dispersion compare with the average, and the inverse log polar, given that those images are not restricted with the morphological transformations, as it is present in the Fig 9.

The log polar sampling in average has the biggest information structure compare with the rest of morphological transformations, and as well provides an important reduction in the number of pixels and therefore the resources needed for any possible learning

All this results show how when the sensory-morphology is defined by the average of the two cameras, the vergence point is going to be the place with the biggest information structure, and given that this area is going to be small in the image, there is no need to sample with the same distribution all over the image. Hence it makes more sense to increase the number of samples in the vergence area than the rest of the image. Now, given that the log polar sampling is fixed (the center is always sampled more) then the agent should be able to develop the skill to actuate its motors to move the cameras in an appropriate way to generate more information structure.

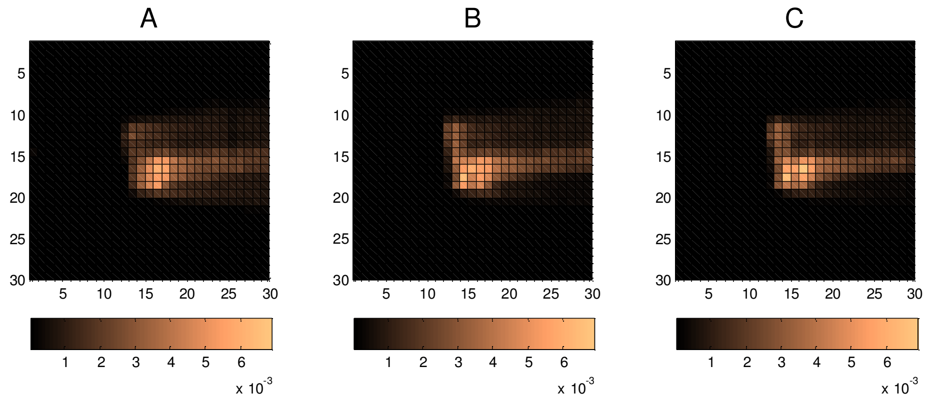

Figure 6: Mutual Information. The measure is over patches of 2X2 pixels, and the copper scale is in bits. (A) Left image (B) Average left and right image (C) Average inverse log polar image. M parameter 8

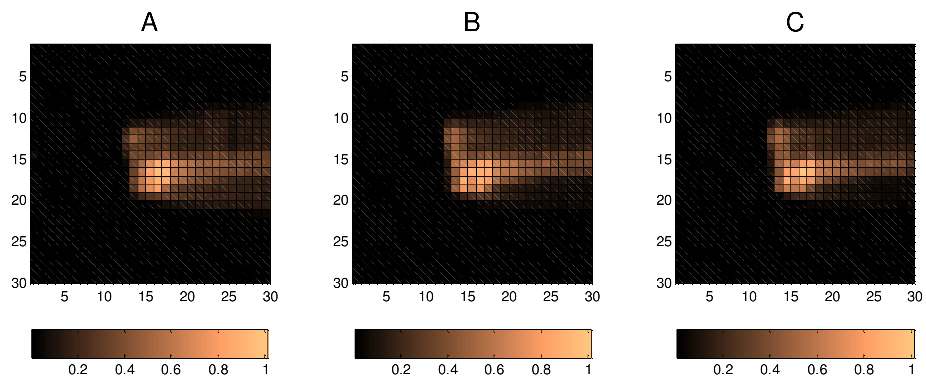

Figure 7: Integration. The measure is over patches of 2X2 pixels, and the copper scale is in bits. (A) Left image (B) Average left and right image (C) Average inverse log polar image. M parameter 8

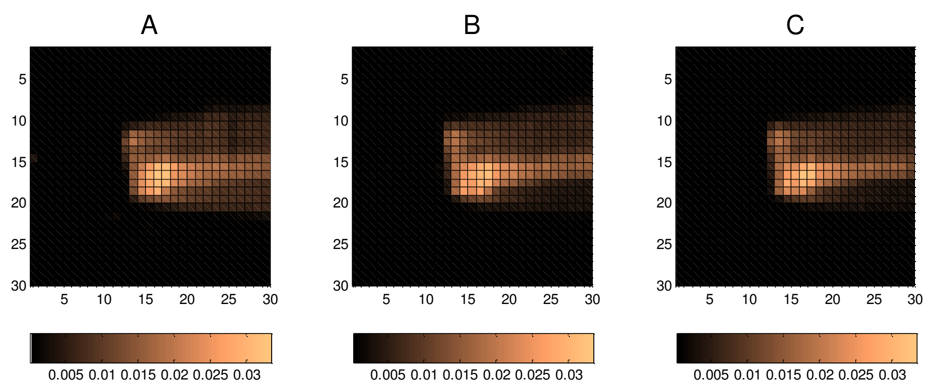

Figure 8: Complexity. The measure is over patches of 2X2 pixels, and the copper scale is in bits. (A) Left image (B) Average left and right image (C) Average inverse log polar image. M parameter 8

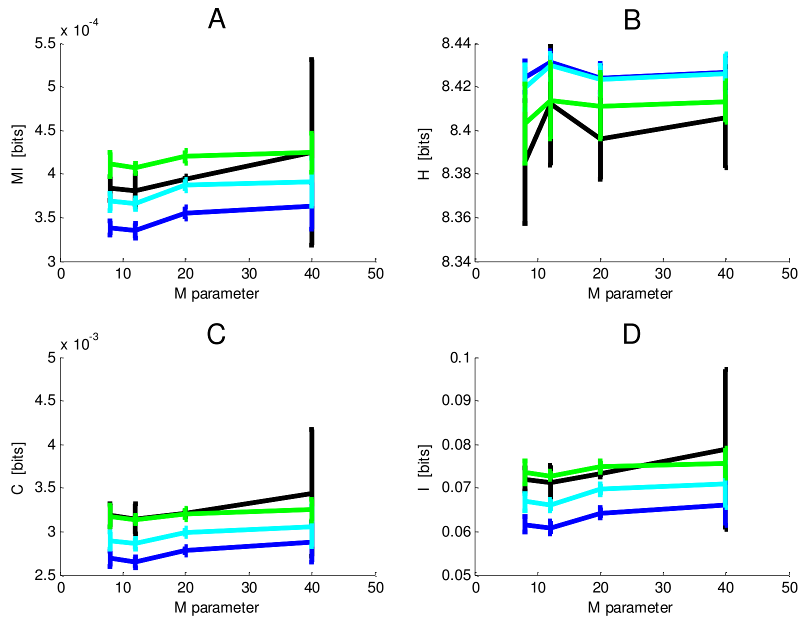

Figure 9: Experimental Results. In the figures A to D green is the right camera, black is the left camera, cyan is the average of both images and blue is the Average inverse log polar transform (A) Mutual information (B) Entropy (C) Complexity (D) integration.

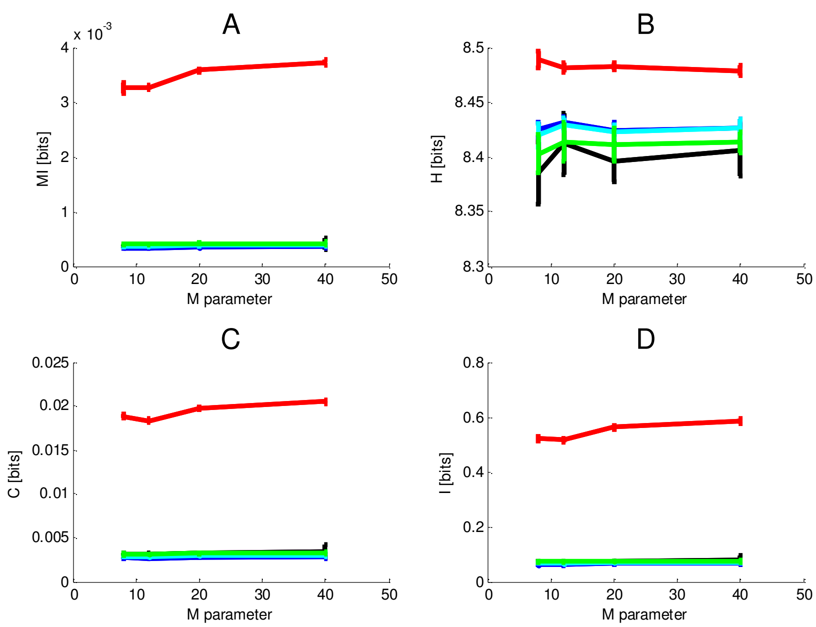

Figure 10: Log Polar Pixels Information Structure. In the figures A to D green is the right camera, black is the left camera, cyan is the average of both images, blue is the Average inverse log polar transform and red is the average log polar transform (A) Mutual information (B) Entropy (C) Complexity (D) integration.

Experiment 2. Our first experiment shows that the proposed sensory morphology structures the information in the visual stream. However, given that the log-polar transform is a fixed transformation, the robot has to move the cameras in order to keep the vergence point always in the center of the image. We expect that this coordinated behavior should create information structure among pixels and actions. This is important in the case where the robot has to learn this behavior, because it establishes a relation among sensor inputs and motor actions through the interaction that the agent carries out with the environment. The purpose of this second experiment is to test how these pixels are also strongly related to the actions. Here we used a color based tracker, where the object to track was modeled with a normalized color model (Breazeal et al., 1999). To follow the cup on the rotating table with both cameras and force the vergence behavior on the cup, the tracker controls the 3 DOF of the eyes of the robot. We recorded 12300 frames for each trial. Each frame had 60X60 pixels and for each morphological condition (the average of the left and right image, the inverse log polar of the average of the left and right image, the log polar of the average of the left and right image, and the left camera.) we run 8 trials for 4 different M parameters of the log polar transform.

In this experiment we analyzed the blue color component from the images (given that this component was the most relevant feature for the tracker) and the three values of the encoders. For all the calculations we took all the 12300 frames. To estimate the probability density function, we used 8 bins per dimension. As in the first experiment, to calculate mutual information, integration and complexity, we used statistical formulae that allow the calculation of entropies from the covariance matrix, under the assumption that these covariances were generated by a stationary Gaussian random process (Cover et al., 1991).

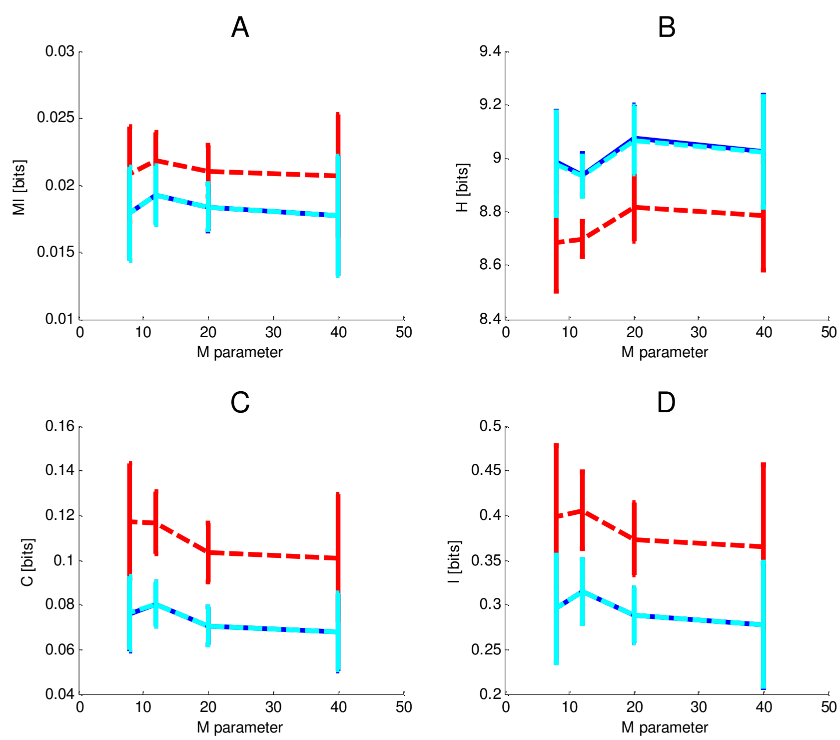

Figure 11: Results Information Structure Pixels and Encoders. In the figures A to D cyan is the average of both images, blue is the Average inverse log polar transform and red is the average log polar transform (A) Mutual information (B) Entropy (C) Complexity (D) integration.

The most important result here is that the actions and the pixels also present the same information structure distribution. The pixels in the center are more related to the encoders, and they are also the pixels with the data from the vergence region and are sampled at a bigger rate than the rest thanks to the log-polar image (Figs. 11 and 12).

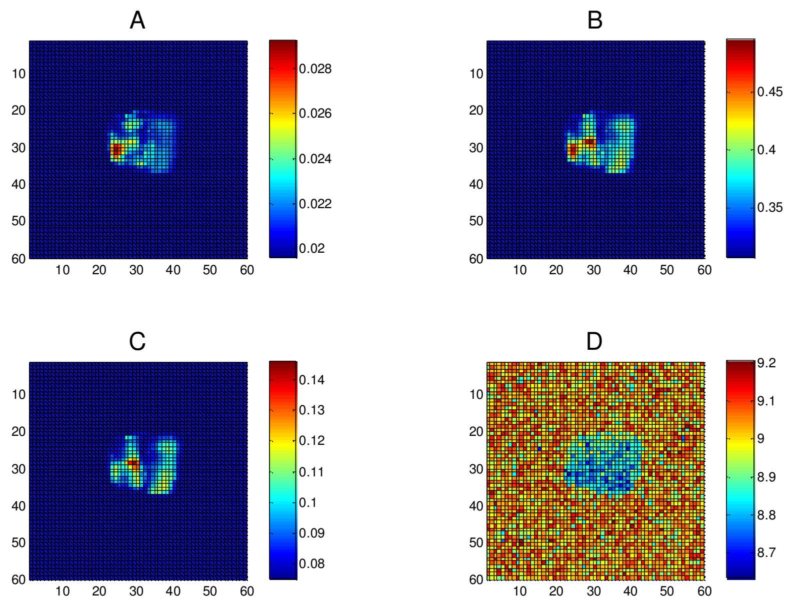

Figure 12: Information Structure in the Image. The measure is over the one pixel and the set of encoder values, color scale is in bits (A) Mutual information (B) Integration (C) Complexity (D) Entropy. This result is with M parameter equal to 8

Experiment 3. Even if there is information structure among actions and pixels and among the pixels themselves, it does not mean that the actions can explain how the pixels are going to behave, what enables the robot to learn a skill, like vergence, could be the necessity for better prediction. In this regard, the proposed sensory morphology is given a correct set of signals that are related but it is not clear if the future sensory input could be expressed in terms of the actual sensory input and actions. From the previous experiments we infer that the selection of the actions is important because the vergence point has to be synchronized with the log-polar transform, therefore the fixation point has to be always in the center of both cameras and all the 3DOF that we can manipulate with the control are crucial for this task.

In this experiment we measured the causal relation among sensor and motor actions in order to see if the control quality (behavior) can affect the causality. In this sense, the possibility to explain the future based in the actual data and actions, hence validating that the vergence procedure is an action capable of increasing the causal relations among the pixels and the actions. We used the color based tracker to change the attention of the robot to 4 different objects. The tracker changed the color model to enable the robot to verge on the different objects in a predefined sequence. The objects where distributed in the field of view to force the robot to change the value in the 3 DOF of the cameras. In order to be able to measure the influence of vergence, we developed three different controllers: (1) the left camera performed random movements while the right one followed the sequence; (2) a controller that allowed parallel motions of the left and right camera; and (3) a controller that forced the vergence with both cameras to focus the object.

For the three controllers we tested four different log-polar transformations (M parameter equal to 8,12,20 and 40); for each transformation we ran 8 different trials; and we recorded 12300 frames (60X60 pixels each) for all different kind of images (left, right, average, log polar transform and inverse log-polar transform). The causality was calculated using transfer entropy. We took the grayscale image and calculated the casualty against each DOF with lags in [-25, 25]. In the Figures 14 to 17, we present the average causality per pixel of the sum of all causalities per DOF.

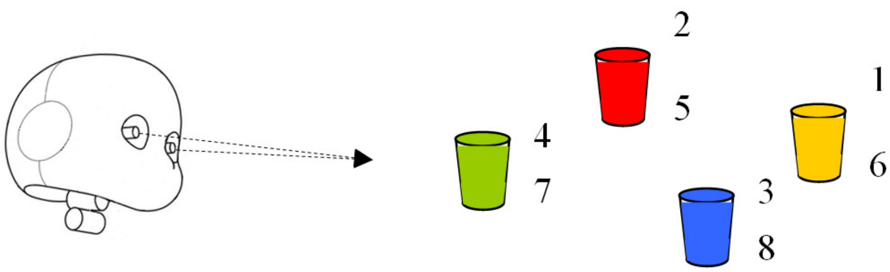

Figure 13: Experimental Setup. The robot is looking the deferent cups in the sequence represented with the numbers, after 7 the robots starts again with 1.

In this experiment happens something similar than in the previous ones, when the controller is doing an appropriate vergence the pixels with the highest causal relation stay also in the center of the image and in the vergence region, for this reason the inverse log polar transform and average got a similar result as it appears in Fig 14.

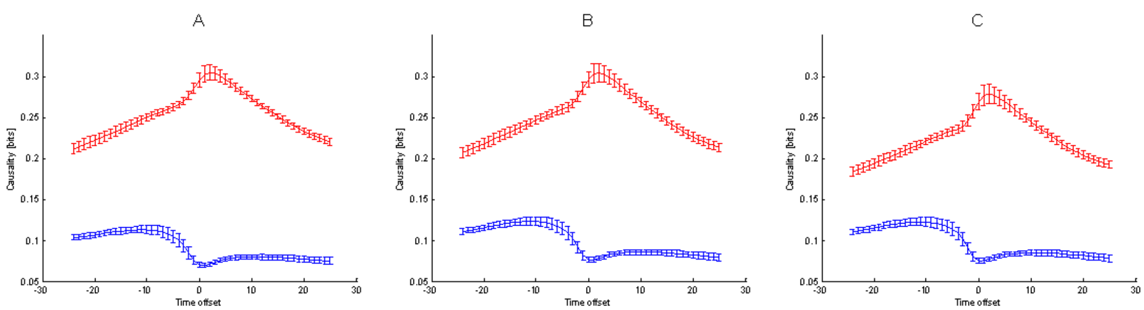

Figure 14: Causality in the Vergence Experiment. In the figures A to C appears the average per pixel of the 3DOF causality summation, blue is the sensor to motor and red is motor to sensor (A) Left Image (B) Average Image (C) Average Inverse Log Polar Image with M parameter equal to 8

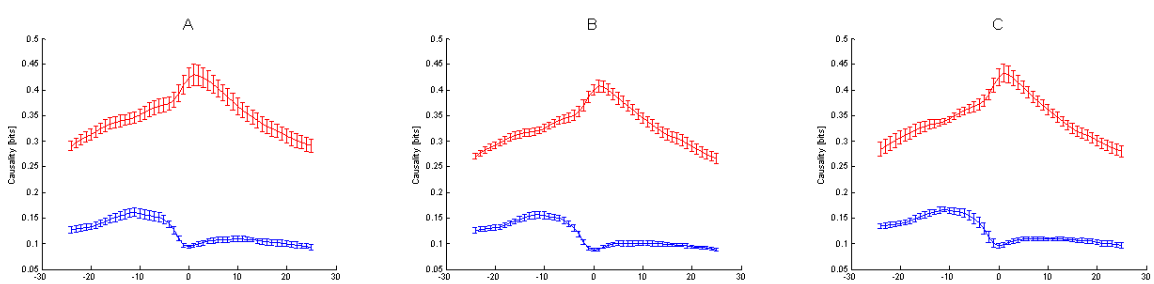

Figure 15: Causality Results with Vergence in the Log Polar Transform. In the figures A to C appears the average per pixel of the 3DOF causality summation, blue is the sensor to motor and red is motor to sensor (A) M parameter equal to 40 (B) M parameter equal to 12 (C) M parameter equal to 8

With the log polar transform we are sampling exactly those pixels for that reason in average we achieve more causality because even though when we reduce the amount of data with the sampling, this data has more causal relation like it is presented in the Fig 15.

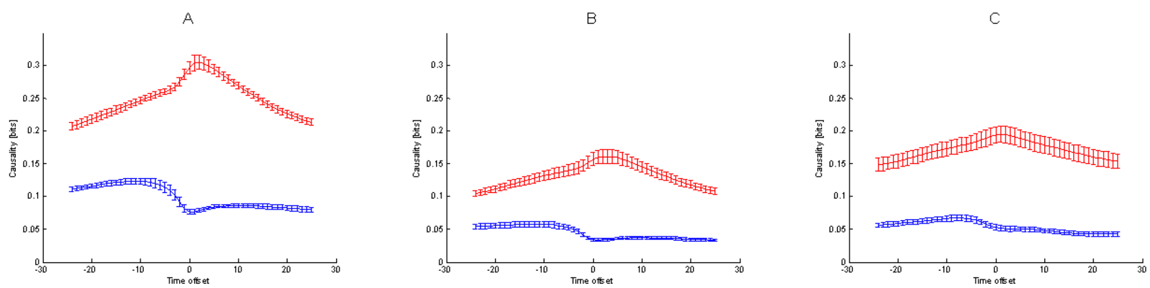

Figure 16: Causality Results with Different Controllers in Average Image. In the figures A to C appears the average per pixel of the 3DOF causality summation, blue is the sensor to motor and red is motor to sensor (A) Controlling all the 3 DOF (B) One camera is tracking while the second one is moving in the same way (C) adding noise to the motion of the camera

The different controllers allow us to see how different “qualities” in the behavior could help the robot to model the environment. Given first the structure among pixels themselves, second the relation among pixels and actions and third the causal relation among pixels and actions, an appropriate behavior that exploits the sensory morphology is going to produce the biggest structure and causal relation enabling the robot model the environment, and what is more important the robot is able to build up this model through the development of a coordinated behavior in this case keep the vergence and the center of the image in the object of interest.

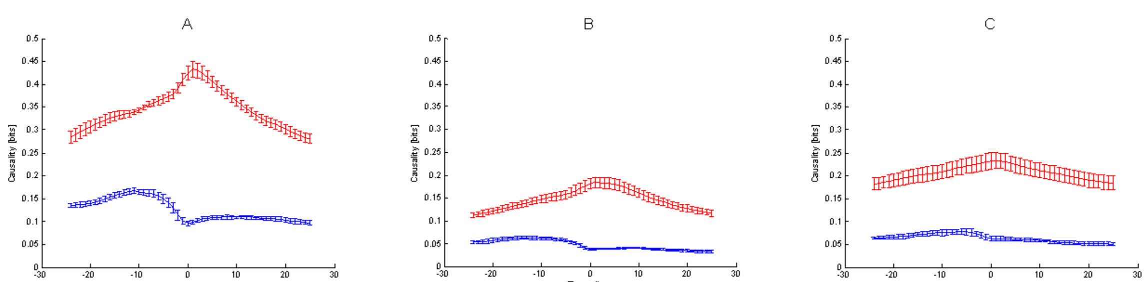

Figure 17: Causality Results with Different Controllers in Log Polar Image M parameter equal to 8. In the figures A to C appears the average per pixel of the 3DOF causality summation, blue is the sensor to motor and red is motor to sensor (A) Controlling all the 3DOF (B) One camera is tracking while the second one is moving in the same way (C) adding noise to the motion of the camera

4. Discussion

With the previews experiments we can see how the average and log polar transform have to be coordinated in order to get a good structure in the sensory data. This coordination is developed with the controller, which not just develop structure among pixels and actions but in addition the more precise the vergence and the center of the image (log polar origin) are kept, the bigger the causal relation among pixels and actions. These relations can be exploited by the robot to develop a model of the environment, and given that the learning capability is limited by the predictive capacity of the sensor and action, hence the robot is limited by the “quality” of its control. In this sense the sensor morphology and the combination of different sensory modalities restrict the possible behavior in order to be able to increase the predictability of the actions of the robot.

The vision system allows us to generate a belief of the environment beyond the simple 3D perception or spatial distribution. Thanks to the interaction with the world and the coupling with other sensory inputs, visual information permits the prediction of taste, weight, heat transfer or even how soft an object is. Our capacity to use our attention towards what it is needed, like a reflex, and the capacity of prediction of our visual system, are two features that makes our vision system a fascinating tool to handle the world, and it is an incredibly complex system that is not easy to isolate or emulate in an artificial platform.

In the recent years there is a lot of work in the development and implementation of attention systems based in the different features extracted from our brain (Itti & Koch 2001, Dankers et al., 2007, Ruesch et al., 2008). This bottom up approach produces a weighted average among all the features, called saliency map, which is used to select the next point of attention. In general this approach is decoupled from any possible task, and the final behavior with this kind of mechanism could be explained as just explorative. However, changing the gains of the weighted average using a top down modulation to drive the attention coupled to a given task. With this approach all the tasks have to be defined previously and they cannot be a result from the development of the artificial creature.

In comparison Bruce et al. (2009) presented a new approach defining the new focus of attention using a measurement of information gain. Here the attention is not the result of a linear combination of features, but a specific region with higher information compare with its own neighborhood. This approach allow us to think that the attention system could be a tool that maximizes the information needed in a particular task, and given that tasks like grasping or reaching involve a great variety of different sensory systems such as vision, propioception, haptics, and a motor action system for such tasks, we have to take in account the information gain, not just in the visual field, but also in the different modalities.

Early behaviors developed by human infants like vergence, therefore, could be explained as the result of such mechanism that is combining the propioceptive system and the vision in order to provide better information about the world, helping to develop superior models of the world.

Storck et. al, 1995 has shown how the information gain could be used as a measure of curiosity, and how this helps an artificial agent to develop a better model of the world. The disadvantage of this approach is the huge amount of data required to build the probability density functions (PDF) needed to calculate the information gain in this specific problem, given that the pixels are the variables used to calculate the PDF. Nonetheless, this is not the unique way to determine the artificial curiosity, and if the attention is driven for this mechanism, then in order to be able to use the attention system to perform a task such as grasping or reaching, it would be natural to think that the agent has to incorporate not just the pixels but also the other sensors involved in the task (haptic system, propioceptive system).

Curiosity in this approach is a tool that helps the agent to build a model of the world. The agent gains experience from the interaction, with the experience it delimits the possible actions or series of actions that compounds a skill. The attention system is a skill that the agent is able to develop and it is in charge of providing better information of the surroundings. In these terms this system looks for increasing the prediction level of the agent in terms of the relations with other sensors and its own actions. Assuming this, the control is defined then in terms of this goal of understanding better the world, and since all the actions of the control are defined in this framework, then a simple action as vergence could be analyzed in this way. Having this in mind we develop these experiments to show how the morphology of the sensory input push the agent to execute an specific behavior in order to achieve an information structure that enable the robot its own development.

In our first experiment we show how the log-polar transform and the average of the two images are beneficiated when the robot focuses to a specific object, and the vergence is in the center of the image, where the log polar transform has the biggest amount of receptors. These characteristics increase the information structured among the pixels. This means that when the agent with such sensors is using vergence is able to find relations that help to understand the environment. For this reason the agent has to develop this skill in order to have better information structure, which also enable the agent to model the sensory input better and therefore allows it to identify predicted errors.

In the second experiment, we present the relation between the sensory input and the actions, which are highly coupled as is also showed in Lungarella et al. (2006). This is important because the log polar transformation is fixed, all the time the center of the image has a bigger number of receptors, which should also be the vergence point, and in order to perceive better the surrounding world the robot does not have another choice than move its cameras. The information measures show that the actions are strongly coupled to the visual information.

However the agent should be able to explore the environment and through this exploration it has to find the appropriate actions to be able to model the world, reducing the possible space of actions towards those actions that give it a better explanation of the input. The viability of this task lies in the necessity of causal relations among the actions and sensors. We followed this thought in the third experiment. Here we show how different kind of movements can increase or decrease the causal relation among the set of variables. An important result here is that the vergence plays a fundamental role because, as appears in Fig 15 and 16. The causal relation is increased meaning that the future sensory input is better explained, enabling the agent to model better the world.

5. Future Work

In all the experiments we can see that the log polar transform samples the right pixels, those with higher information structured and causality. These pixel are fewer than the pixels in the raw image therefore the log polar sensor reduces the computational load, additionally improves the learning because those are the pixels with the higher structure even when the amount of inputs are reduced to 17%. With a normal Cartesian pixel array the rest of the pixels in the learning process are just noise, due to the lack of structure, and in this sense the perception of the agent is decreased. This bring the opportunity to apply directly all the possible artificial curiosity algorithms to handle directly the data.

This approach allows us to think in new models of attention that takes into account different sensory modalities in order to find behaviors that allow the agent develop models of the environment and with them beliefs built from the experience because the nature of the sensory input determines the possible set of actions and behaviors for the agent.

In the perspective of human infants our results show that the behavior is a result of better information structure, actions like vergence allow us to predict better to understand better the environment, and the integration of several sensory modalities can generate therefore more complex final behaviors in order to achieve structure in several sensory systems, this should be the approach to analyze the constitutive principles of the relation of the vision and hand in a task such as grasping.

References

Aslin, R. N. (1977) Development of binocular fixation in human infants. *Journal of Experimental Child Psychology*, 23(1),133–150

Barto, A., Singh, S., Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. *Proc. 3rd Int. Conf. Development Learn., San Diego, CA*, 112–119

Beira, Ricardo., Lopes, Manuel., Praça, Miguel., Santos-Victor, José., Bernardino, Alexandre., Metta, Giorgio., Becchi, Francesco., Saltarén., Roque. (2006). Design of the Robot-Cub (iCub) Head. *Proc. IEEE Conf. on* Robotics *and. Automation,* 94–100

Birch, E., Morale, S., Jeffrey, B., Oconnor, A., Fawcett, S.(2005) Measurement of stereoacuity outcomes at ages 1 to 24 months: Randot stereocards. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 9(1), 31-36

Birch, E., Petrig, B. (1996) FPL and VEP Measures of Fusion, Stereopsis and Stereoacuity in Normal Infants. *Vision Res.,*36(9), 1321-1327

Breazeal, C., Scassellati, B. (1999). A context-dependent attention system for a social robot. *Proc. of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI99*, 1146-1151

Bruce, N. D. B., Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3):5, 1-24

Cover, TM., Thomas, JA., (1991) Elements of information theory. New York: Wiley. 542

Dacey, M. (1996). Circuitry for color coding in the primate retina. *Proc. Nat. Acad. Sci.*, 93:582–588. 48, 217

Dankers, A., Barnes, N., Zelinksy, A. (2007). A Reactive Vision System: Active-Dynamic Saliency. *Proc. of the 5th International Conference on Computer Vision Systems, ICVS*

Itti, L., Koch, C. (2001). Computational Modeling of Visual Attention. *Nature Reviews Neuroscience* 2(3):194-203

Körding, KP., Wolpert, DM. (2006) Bayesian decision theory in sensorimotor control. *Trends Cogn Sci* 10: 319–326.

Lungarella, M. (2007). Developmental robotics. *Scholarpedia,* 2(8):3104

Lungarella, M., Pegors, T., Bulwinkle, D., Sporns, O. (2005) Methods for quantifying the information structure of sensory and motor data. *Neuroinformatics* 3: 243–262

Lungarella, M., Sporns,O. (2006). Mapping information flow in sensorimotor networks. *PLoS Computational Biology*, 2(10):e14

McGill WJ (1954) Multivariate information transmission. *IEEE Trans Inform Theory* 4: 93–111

Oudeyer, P-Y., Kaplan , F., Hafner, V. (2007) Intrinsic Motivation Systems for Autonomous Mental Development, *IEEE Transactions on Evolutionary Computation,* 11(2), 265—286

Pearl, J. (2000). Causality: Models, reasoning, and inference. *Cambridge (United Kingdom): Cambridge University Press.*

Pearl, J. (2009). The structural theory of causation. To appear in *P. McKay Illari, F. Russo, and J. Williamson (Eds.), Causality in the Sciences, Clarendon Press, Oxford, 2009.Cambridge University Press.*

Ruesch, J., Lopes, M., Bernardino, A., Hoernstein, J., Santos-Victor, J. and Pfeifer, R. (2008). Multimodal saliency-based bottom-up attention: a framework for the humanoid robot iCub. *Proc. of IEEE Int. Conf. on Robotics and Automation*, 962-967

Schmidhuber, J. (2009) Driven by Compression Progress: A Simple Principle Explains Essential Aspects of Subjective Beauty, Novelty, Surprise, Interestingness, Attention, Curiosity, Creativity, Art, Science, Music, Jokes. *Anticipatory Behavior in Adaptive Learning Systems, from Sensorimotor to Higher-level Cognitive Capabilities, Springer, LNAI*, 48-76

Schreiber, T. (2000) Measuring information transfer. *Phys Rev Lett* 85: 461–464

Shannon, C. (1948) A mathematical theory of communication. Bell Sys Tech J 27: 379–423

Smith, P., Cowie, H., and Blades, M. (1998). Understanding children's development. Blackwell.

Storck, J., Hochreiter, S., Schmidhuber, J. (1995) Reinforcement-driven information acquisition in non-deterministic environments. *Proc. ICANN'95, vol. 2,* 159-164

Thelen, E., Smith, L. (1994) A dynamic systems approach to the development of cognition and action. *Cambridge, MA: MIT Press/Bradford*.

Tondel, G., Candy, T. (2007) Human infants' accommodation responses to dynamic stimuli. *Investigative Ophthalmology & Visual Science*, 48(2), 949–956

Tononi, G., Sporns, O., Edelman, GM. (1994) A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proc Natl Acad Sci U S A* 91: 5033–5037

Wheatstone, C. (1838). Contributions to the physiology of vision.—Part the First. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London* 128: 371–394

Zothdurf, H. (1990). Texture discrimination by cells in the cat lateral geniculate nucleus. *Exp. Brain Res*, 82:48–66. 48, 197