Learning motor dependent Crutchfield's information distance to anticipate changes in the topology of sensory body maps

Thomas Schatz, Pierre-Yves Oudeyer

▶ **To cite this version:**

Thomas Schatz, Pierre-Yves Oudeyer. Learning motor dependent Crutchfield's information distance to anticipate changes in the topology of sensory body maps. IEEE International Conference on Learning and Development, 2009, Shangaï, China. pp.n/a. inria-00420186

HAL Id: inria-00420186
https://inria.hal.science/inria-00420186

Submitted on 28 Sep 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning motor dependent Crutchfield's information distance to anticipate changes in the topology of sensory body maps

Thomas Schatz and Pierre-Yves Oudeyer

***Abstract*—What can a robot learn about the structure of its own body when he does not already know the semantics, the type and the position of its sensors and motors? Previous work has shown that an information theoretic approach, based on pairwise Crutchfield's information distance on sensorimotor channels, could allow to measure the informational topology of the set of sensors, i.e. reconstruct approximately the topology of the sensory body map. In this paper, we argue that the informational sensors topology changes with motor configurations in many robotic bodies, but yet, because measuring Crutchfield's distance is very time consuming, it is impossible to remeasure the body's topology for each novel motor configuration. Rather, a model should be learnt that allows the robot to predict Crutchfield's informational distances, and thus anticipate informational body maps, for novel motor configurations. We present experiments showing that learning motor dependent Crutchfield distances can indeed be achieved.**

***Index Terms*—Crutchfield distance, information theory, sensorimotor learning, body map**

THE DISCOVERY OF INFORMATIONAL TOPOLOGICAL BODY MAPS

In developmental robotics [13], one aims at building robot capable of learning progressively and continuously new skills in unknown changing bodies and environments. In particular, this involves mechanisms for discovering its own body and its relationships with the environment. In this context, learning body maps is a crucial challenge. Body maps are topological models of the relationships among body sensors and effectors, which human children learn progressively, abstract and build upon to learn higher-level skills involving the relationships between the shape of the body and the physical environment [3]. Accordingly, inferring and re-using body maps from initially uninterpreted sensors and effectors has been identified as an important objective in developmental robotics [10].

Studies in developmental psychology [11] showed that body babbling in infants, that is to say the systematic exploration of the consequences of their actions on their sensations, was an essential part in the building of a primary sense of bodily self. That sense being necessary to the ulterior development of complex sensorimotor abilities. Other studies in neuroscience and medical imaging unveiled the presence of sensoritopic and motor projections in the brain, i.e. neural circuits whose geometry reflects the topography of the different sensory and motor organs [5]. [10] proposed an heuristics to allow an artificial system to build sensoritopic maps of its sensory organs by exploring randomly its environment, i.e. by practicing a form of body babbling. This heuristics doesn't suppose any prior knowledge of the surrounding environment nor of the morphology of the system and its sensory and motor parts. It consists mainly in computing a metrics between the different sensory receptors from the series of successive values they take. [8] [9] then improved this heuristics, notably by using concepts from information theory in order to compute the Crutchfield information metrics between the different sensors. [6] computed the metrics while the system was performing a programmed activity (e.g. walking or dancing...) instead of randomly moving. They made measures for different activities and then they used a clustering algorithm to associate automatically the measures of the metrics between each pair of sensors to the corresponding activity. The idea being to characterize an activity, robust to variations in the particular external conditions of the performance, by a set of values of the metrics recorded while the system performs the activity.

The work described in [10], [8], [9] and, to a lesser extent, [6], is restricted by the fact that mean values of the metrics are computed without taking into account information characterizing motor commands and motor configurations. This boils down to assuming that the set of sensors of the body has a single fixed informational topology (or several discrete fixed topologies in the case of [6]), which is indeed the case in for the distance sensors of the Khepera robot or the pixel sensor matrices used as testbeds in these works. This is problematic since in many robot bodies, the informational topology of sensor will be dependant upon motor configurations: for example, the touch sensors in the fingers of two hands are informationally very close if the two hands are touching each other (double touch), but are informationally farther away if one hand is scratching the head (in which case the sensors in this hand will be informationally topologically close to the touch sensors of the head) and the other hand is freely hanging. Moreever, the number of possible motor configurations is generally very high (infinite with continuous actuators), and it is unpractical to measure Crutchfield metrics for each pair of sensors for each novel motor configuration. So it appears necessary to realize only a limited number of measures of Crutchfield metrics for certain motor configurations and to estimate from those the values of the metrics for other motor configurations. In what follows, we will first present and discuss in more details this previous work ([10] [8] [9]), and then we will show how a learning algorithm can be used to anticipate the Crutchfield distance between pairs of sensors

Thomas Schatz is with ENS de Cachan - antenne de Bretagne
Pierre-Yves Oudeyer is with INRIA Bordeaux - Sud-Ouest

for novel motor configurations that modify the informational topology of these sensors.

I. Inferring body maps from Crutchfield informational metrics

A. Introduction

We consider afterwards any system (e.g. a robot) possessing some sensors, which allow it to collect information about its state and its environment, some actuators, which allow it to change its state and to interact with its environment and a control system. The control system receives information from the sensors, then compute them to produce a command addressed to the actuators. Those information pass through real-valued communication channels. The control system can write on motor channels and read sensory channels. From the point of view of the different algorithms proposed in the studied works and of those proposed in this article, the system is a tabula rasa: the control system can only access the list of sensory channels and the list of motor channels, without any additional knowledge, such as the type, semantics or positions of the different sensory and motor parts.

Formally, the problem tackled in the presented works is to infer, from the sole knowledge of the values passing on the channels as time flows, a mathematical structure on the channel's set. The goal is to allow the system to structure automatically its perceptions and its actions. The inferred structure is a metrics on the set of the sensory channels for [10] and [8], a metrics on a set of such metrics for [6], and a mapping from a value of the motor channels to a metrics on the set of the sensory channels in this article. Other structures may be possible. For instance, a number associated to each sensor/actuator pair, which would account for the correlation between the use of the actuator and the variation of the sensor's value. Those numbers could then be used to map the actuators on the sensoritopic maps. Then an oriented graph on the set of the motor channels, corresponding intuitively to the skeletton of the system, may be derived. These inferred structured may then be used to improve the computation of sensory information and the derivation of actuators' commands. For instance, [9] use the obtained sensoritopic maps to compute the optical flow through visual sensors with visual field of complex shapes.

B. Uniform binning and adaptive binning

Data received on a sensory channel are often more precise than desired, that is to say the information on the channel can take too many different values. It is then necessary to group certain classes of values. *Pierce and Kuipers*, for whom the sensory channels took their values in (a finite subset of) the interval $[0;1]$, used an uniform partition of this interval in order to be able to arbitrarily set the number of values a channel can take (uniform binning). For instance, if we set this number to 3: let $I_1 = [0;\frac{1}{3}[$, $I_2 = [\frac{1}{3};\frac{2}{3}[$ and $I_3 = [\frac{2}{3};1]$, and let $x_1 \in I_1$, $x_2 \in I_2$ et $x_3 \in I_3$, then the value of the channel after binning for an initial value in I_1 would be x_1, x_2 for an initial value in I_2, etc. Notice that x_i can be chosen arbitrarily in I_i, as a matter of fact any symbol s_i such as for all i, j we have $i \neq j \Rightarrow s_i \neq s_j$ would fit, provided that the numerical properties of the channel's values are not used (as it is the case in [8] et [9]).

A drawback of this approach is that if, for instance, only ten values are distinguished for a given canal whose values are in $[0;1]$ and if it always take its values between 0.2 and 0.3, the observed value for the system will always be the same and the sensor will be useless (it won't bring any information in the sense of information theory). [8] proposed the use of adaptive binning rather than uniform binning. Adaptive binning allow an optimized use of the information provided by the sensors. The $[0;1]$ interval is then divided in such a way that the observed values on the sensory channels be dispatched as uniformly as possible in the different parts of the partition. This method maximizes the mean collected information for a partition of given size (in the sense of information theory). This is particularly useful for sensors whose characteristic isn't linear.

C. Creation of sensoritopic maps

[10] proposed an heuristics to allow an artificial system to develop an approximative representation of the topography of its sensors relatively to a certain metrics, i.e. a sensoritopic map. This metrics is computed from the sole knowledge of the sequences of values taken by the sensor channels while the system moves randomly. The proposed algorithm was tested with a simulation of different robots in a simple environment. [8] then introduced the concept of informational metrics (Crutchfield's metrics) and of adaptive binning in order to improve the algorithm. They tested the algorithm in simulation and on an AIBO robot.

There is three main steps in the heuristics. In a first step, a distance between each pair of sensors is measured, i.e. a positive real number, which grows bigger as the sensors are further apart. It doesn't have to be a metrics in the mathematical sense. In a second step, a well-fitted dimension for the sensoritopic map is computed. Eventually, the sensors are depicted by points in an euclidean space of that dimension, in such a way that the euclidean distance between two points be as close as possible to the distance between the two corresponding sensors.

Pierce and Kuipers consider sensory and motor channels with value in a uniform partition of $[0;1]$ of given size N. They propose two metrics on the set of sensory channels. For the first, two sensors are close if their values are close at any given time. Let $s1$ and $s2$ be two sensors, and $(s_1(1), s_1(2)..., s_1(T))$, $(s_2(1), s_2(2)..., s_2(T))$ the values taken by these sensors at T consecutive times, then:

$$d_T(s_1, s_2) = \frac{1}{T}\sum_{t=1}^{T} |s_1(t) - s_2(t)|.$$

For the second, two sensors are close if they take each of their value with close frequency. Let $s1$ and $s2$ be two sensors and $(s_1(1), s_1(2)..., s_1(T))$, $(s_2(1), s_2(2)..., s_2(T))$ the values taken by these sensors at T consecutive times. Let $(s_1^{freq}(1), s_1^{freq}(2)..., s_1^{freq}(N))$ and

$(s_2^{freq}(1), s_2^{freq}(2)..., s_2^{freq}(N))$ be, where $s_i^{freq}(f)$ stands for the ratio of values taken by s_i in the f^{th} interval. Then:

$$d_{N,T}(s_1, s_2) = \frac{1}{2}\sum_{f=1}^{N} |s_1^{freq}(f) - s_2^{freq}(f)|.$$

However, these metrics suffers several problems, notably they can't account for non-linear correlations between sensors, which are nonetheless very usual in the case of two sensors physically close, but of different modalities (e.g. cones and rod cells in the retina). Even an affine correlation won't be detected.

Olsson and al. introduced another metrics, originating from information theory: Crutchfield's metrics [2]. As a reminder, being given a random variable X of probability law p with values in E, X's Shannon entropy or mean information is:

$$\mathcal{H}(X) = \sum_{e\in E} -p(\{X = e\})log_2(p(\{X = e\})).$$

Being given a random vector (X, Y) of probability law p with values in $E \times F$, the conditional entropy of X knowing Y is:

$$\mathcal{H}(X|Y) =$$

$$\sum_{(e,f)\in E\times F} -p(\{(X, Y) = (e, f)\})log_2(\frac{p(\{(X, Y) = (e, f)\})}{p(\{Y = f\})})$$

(See [1] for an in-depth presentation of information theory). Let $s1$ and $s2$ be two sensors, and $(s_1(1), s_1(2)..., s_1(T))$, $(s_2(1), s_2(2)..., s_2(T))$ the values taken by these sensors at T consecutive times. We consider this couple of time series as a T-sample of a random vector of dimension 2 (S_1, S_2), S_1 and S_2 taking their values in finite sets. From the empiric probability law associated to the couple of samples, we then compute:

$$d_T(s1, s2) = \frac{\mathcal{H}_T(S_1|S_2) + \mathcal{H}_T(S_2|S_1)}{\mathcal{H}_T(S_1) + \mathcal{H}_T(S_2)}.$$

The only interest of the fraction's denominator is to normalize the metrics in $[0; 1]$. $\mathcal{H}_T(S_i|S_j)$ can be interpreted as the information brought in proper by S_i, i.e. the information contained by S_i, which isn't already in S_j. So Crutchfield's metrics gives a measure of the quantity of information which isn't common to S_i and S_j, which is indeed a form of distance. Besides it's a pseudo-metrics in the mathematical sense, i.e:

- it's symetric: $d_T(S_1, S_2) = d_T(S_2, S_1)$,
- it's positive: $d_T(S_1, S_2) \in \mathbb{R}^+$,
- it complies to triangular inequality:

$$d_T(S_1, S_3) \leq d_T(S_1, S_2) + d_T(S_2, S_3).$$

It lacks the property of being defined to be a real metrics. Indeed it is possible to find S_1, S_2 such as $d_T(S_1, S_2) = 0$ and $S_1 \neq S_2$. However, it can be shown that Crutchfield's metrics between two random variables is zero if, and only if, they are two equivalent encoding of a same source. See [2] for proofs.

Once the metrics over the set of sensors is measured, [10] propose to determine a well-fitted dimension for the sensoritopic map with the metric scaling method: the distance between the pairs of sensors are represented by a square matrix $(m_{i,j})$ such as $m_{i,j} = d_T(s_i, s_j)$. This matrix is symetric real, so it's diagonalizable and it's eigenvalues are reals, let's call them $\lambda_1, \lambda_2, ..., \lambda_n$, with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$. We shall then be able to select information in a number of dimension more restricted than n while losing a minimum of information by retaining only the main eigenvalues. The selected dimension m is such that $\lambda_{m+1}^2 - \lambda_m^2$ be maximal. Intuitively the selected dimension is such that there be the biggest possible difference between the quantity of information brought in by this dimension and the one brought in by a supplementary dimension. For a detailed presentation of principal component analysis, see [14].

Eventually, each sensor is depicted by a point in an euclidean space of dimension m, in such a way that the euclidean distance between two points coincide at best with the measured distance between the two corresponding sensors. There are numerous methods for this classic optimization problem, see [4]. One example of sensoritopic maps obtained by [8] is given on figure 1.

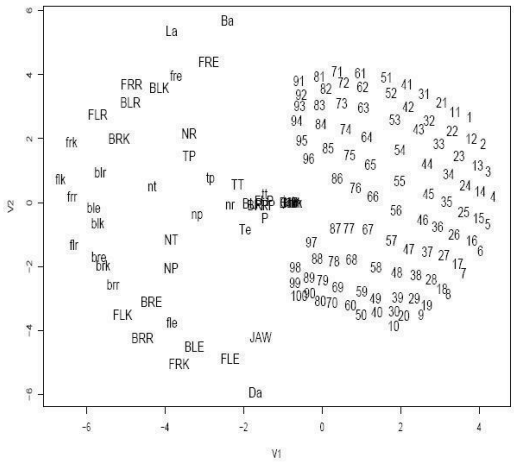

Fig. 1. **Sensoritopic map of the AIBO robot with Crutchfield's metrics.** Each pixel of the visual sensor is depicted by a number between 1 and 100. The infrared distance sensor is depicted by Ir, the accelerometers by Ba, La and Da, the temperature sensor by Te, the remaining power in the battery by P, the binary contact sensors of the paws by BLP, BRP, FLP and FRP, where FL means front right, BR back right, etc. Each paw has three degrees of freedom: E (elevation), R (rotation) et K (knee), for each degree of each paw there is a position sensor (for instance FLK for Front Left Knee) and a torque sensor, depicted in lower-case letters (for instance flk). The neck has three degrees of freedom too named tilt, pan and roll and so three position sensors NT, NP and NR and three torque sensors nt, np, nr. It's the same for the tail: TT, TP, TR, tt, tp, tr. At last the jaw has only one degree of freedom, and so only two sensors JAW and jaw. **[After [8]]**

II. Automated learning of Crutchfield's metrics

A. Introduction

In the work of [10] and [8] the informational Crutchfield distance over all pairs of sensors was computed based on a series of sensor values measurements corresponding to randomly changing motor commands: this implicitly assumed that there was only one possible distance per pair of sensor, i.e. it assumed that the informational structure of the set of

sensors was fixed. This is indeed the case in examples such as the distance sensors of Khepera robots or the pixel sensors of a camera. In [6]'s work, the motor commands changed according to one of several pre-defined activity: they identified the fact that a given motor activity or configuration, as well as a given dynamics of the environment itself, could modify the informational structure of a set of sensors. Yet, in their experiment, only a finite set of motor configurations was used and the pair wise Crutchfield metrics was computed for each configuration after many measurements of the sensor values. In most real robot bodies, a change in motor configuration will provoke a change in the informational structure of the set of sensors, and because with continuous motors there is an infinite (or very large) number of motor configurations, it is impossible to compute the Crutchfield metric for each of them. Rather, we argue that it is necessary that a model of Crutchfield distances, corresponding to all particular motor configurations, be learnt from a small finite set of really measured metrics, such that metrics for novel motor configurations can be accurately predicted. In the following, we will show that it is indeed possible to learn Crutchfield distances in motor dependant dynamically reconfigurable informational body maps.

B. Learning algorithm

A memory-based learning algorithm is chosen in this article, based on the use of a gaussian kernel on the k-nearest neighbours of a query point. Let a system with m motor channels be. From a database containing the distance obtained with Crutchfield's metrics between two sensors s_1 and s_2 of the system for n given motor command, a prediction of the distance between them for any command x is given by:

$$d_{s1,s2}(x) = \frac{\sum_{i=1}^{k} d_{s1,s2}(x_i) e^{-\frac{||x-x_i||_2^2}{\sigma^2}}}{\sum_{i=1}^{k} e^{-\frac{||x-x_i||_2^2}{\sigma^2}}},$$

where $(x_1, x_2, .., x_k)$ are the k closest command to x in the database in the sense of the euclidean metrics on $\mathbb{R}^m$ (with $k \ll n$). The algorithm is parameterized by σ and k. A decrease in σ accelerate the decreasing of the influence of the k nearest neighbours on the prediction with the distance to the prediction point. See [7] for more details on this algorithm. In the following experiments, optimal values of k and σ were chosen to maximize performances in generalizations (depending on the experiment, k is between 1 and 10).

C. Presentation of the simulation

To evaluate the performance of Crutchfield's metrics' automated learning, measures were made on different simulated systems. The simulated space is a plan. The bodies are made of rigid segments assembled together by pivot joints with a motor in each joint (cf. figure 2). Several types of sensors (distance, color, apparent angle and apparent diameter) can be placed on the segments. Figures 3, 4 and 5 show the principles of distance, apparent angle and apparent diameter sensors respectively. The color sensor takes the value of the RGB encoding on 24 bits of the color of the nearest object present in its perceptive field. The environment is made of colored disks of various radius moving in the plan. The movement of a disk is determined by two parameters, its speed and its instantaneous probability of changing its direction, the path of the disk being consequently a broken line (cf. figure 10). To preserve the simplicity of the model, collisions between the different elements in presence aren't taken into account. At any time it is possible to access the values of the sensors and to fixate the position of the joints.

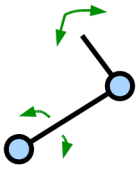

Fig. 2. **Example of a body made of two segments.** The segment in the bottom is linked to the ground by a pivot joint and the upper segment is linked to the segment in the bottom by another pivot joint. On this body, sensors can be attached at any place along the segments. There are four types of sensors, three of which are described in the following figures (the fourth one is a color sensor). These body are placed in an environment where colored disks move on a random path (see figure 6).

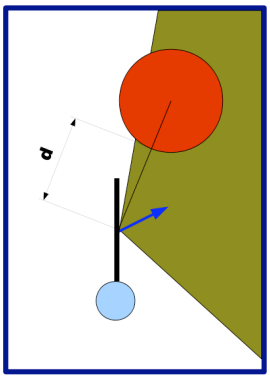

Fig. 3. **Distance sensor.** The location and orientation of the sensor are represented by the blue arrow. The perceived distance **d** correspond to the distance between the origin of the sensor and the nearest object of the environment within the perceptive field of the sensor (yellow-green area).

D. Experiments and results

We ran experiments on four different simulated bodies illustrated on figure 7. The types of sensors on these bodies are randomly chosen (but fixed once chosen). Figure 8 summarizes the evolution of performances in predicting Crutchfield's information distance between the two sensors of each of the four bodies for various numbers of learning examplars. A learning examplar consists in an association between a random motor configuration and the associated Crutchfield metrics between the two sensors of the body computed directly with 100000 samples. For testing the learnt model of Crutchfield distances, we use a test database composed of 100 uniformly distributed motor configurations associated to the corresponding recomputed Crutchfield distances. Of course, the training and test

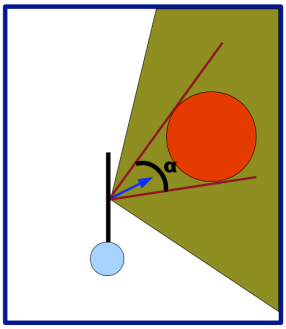

Fig. 4. **Apparent angle sensor.** The location and orientation of the sensor are represented by the blue arrow. Perceive the angle α under which is seen from the sensor the closest object of the environment within its perceptive field.

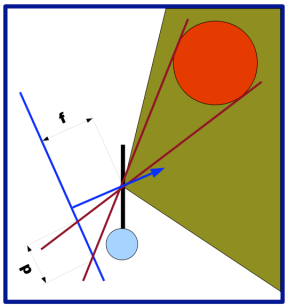

Fig. 5. **Apparent diameter sensor.** The location and orientation of the sensor are represented by the blue arrow. The perceived length **d** correspond to the size of the projection of the sensed object on a plan orthogonal to the sensor direction and situated at a focal distance **f** behind the sensor.

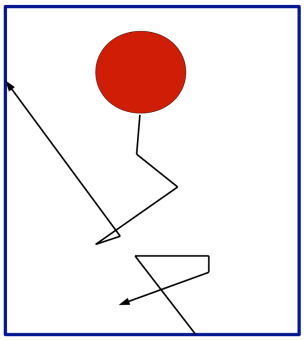

Fig. 6. **Path of a colored disk.** At every moment, it has a certain probability of moving toward another uniformly randomly chosen direction. Its speed remains constant.

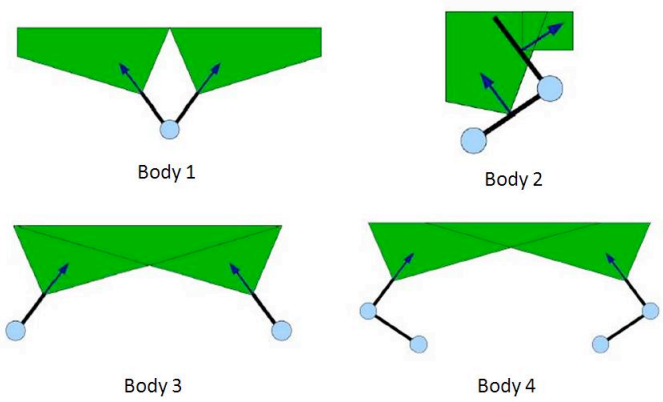

Fig. 7. Body 1 (with two rigid segments, two superposed joint and two sensors), Body 2 (with two chained rigid segments, two joints and two sensors), Body 3 (with two independant rigid segments, two joints and two sensors), Body 4 (with two independant chains of segments, four joints and two sensors)

databases are independant. As we can observe on figure 8, we see that the system is able to learn very efficiently to predict the Crutchfield distance between two sensors in all bodies and for motor configurations that were not encountered in the training database. Figures 9 and 10 compares all Crutchfield distances in the test database with the corresponding distances predicted by the learnt model after 1024 learning examplars have been provided. This allows us to see that relatively to the global amplitude of the Crutchfield distance variations when motor configurations are varied, the prediction errors are low. This shows that a model of Crutchfield distances over the whole space of motor configuration has been efficiently learnt, and thus, reusing the body map reconstruction techniques presented in [6], [8], could be used to dynamically anticipate informational sensory body maps when motor configurations are changing.

E. Discussion: Computational complexity and scalability

In the experiments presented above, 100000 measures were used for estimating the informational topology of a given motor configuration. This is a lot and may prevent the system to be applicable in realtime robotic experiments. Yet, on the one hand these measure can be done at a high-frequency without affecting the results, e.g. 100hz, if the sensori apparatus permits it, and we believe that accurate measures of informational distance can be realized with much less samples. This will be the topic of future experiments. Furthermore, it can be noted that increasing the number of sensors in the system would increase the complexity quadratically, since Crutchfield distances need only be measured, learnt and predicted pairwise to reconstruct the whole informational topology. As far as the motor space is concerned, increasing the number of degrees-of-freedom to several dozens should not be problematic since mathematically, the regression performed here is quite similar to high-dimensional regression problems of robot motor learning adressed by techniques recently developped, such as LWPR [12].

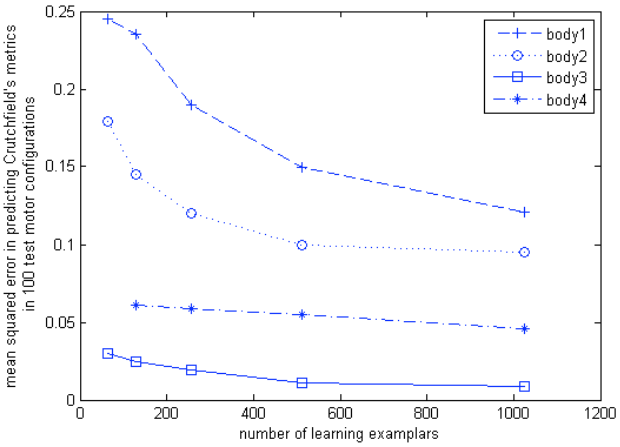

Fig. 8. Evolution of mean squared errors in predicting Crutchfield's metrics between the two sensors of bodies 1, 2, 3 and 4 when various number of learning examplars are available (each learning examplar is an association between a random motor configuration and the associated measured Crutchfield informational distance) and with a test database composed of 100 motor configurations uniformly spread in the space of configurations.

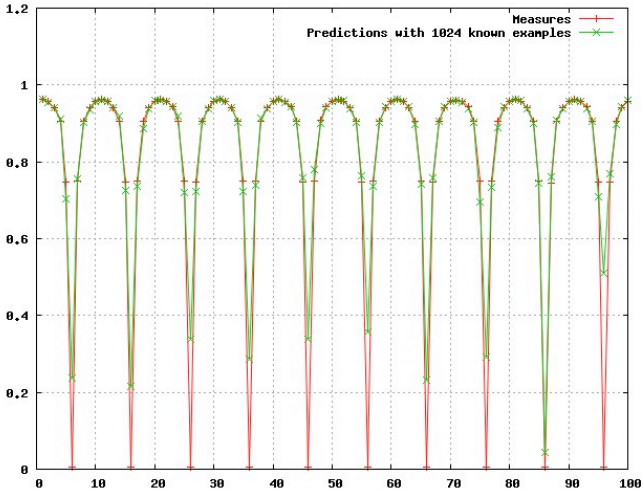

Fig. 9. Comparison between the Crutchfield distances in the test database of body 2 with the corresponding predicted informational distances after 1024 learning examplars have been provided: one observes that the system predicts accurately the Crutchfield information distance among the two sensors for motor configurations that were not experimented beforehand. The sequence of examplars is generated by fixing the position of the first effector, then rotating progressively the second effector around the clock, before rotating the first effector by adding it $\frac{2\pi}{10}$ and rotating again the second effector around the clock, which is repeated until the first effector has itself rotated by 2π. The peaks where the Crutchfield distance is zero corresponds to motor configurations in which sensors superpose and perceive exactly the same thing.

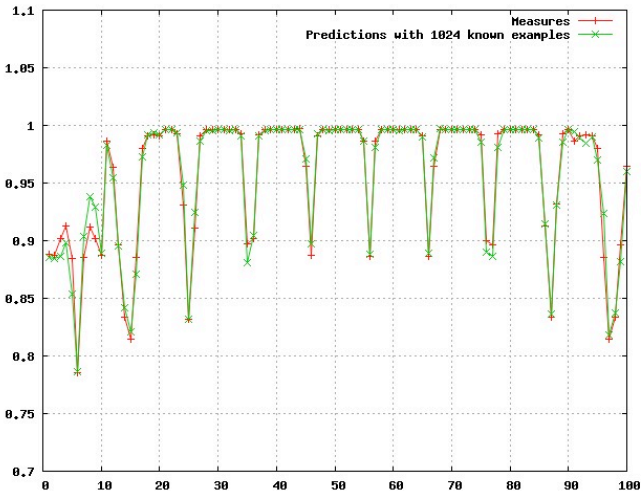

Fig. 10. Comparison between the Crutchfield distances in the test database of body 3 with the corresponding predicted informational distances after 1024 learning examplars have been provided: one observes that the system predicts accurately the Crutchfield information distance among the two sensors for motor configurations that were not experimented beforehand. The sequence of examplars is generated in the same way than in the previous figure. The shape of the curve is not equivalent since rotations of the second effectors are not symmetric in relationship to rotations of the first effector.The peaks with low values correspond to motor configuration for which sensors perceive overlapping regions, and the peaks with values close to 1 correspond to motor configurations for which sensors perceive non-overlapping regions.

REFERENCES

[1] T.M. Cover and Thomas J.A. *Elements of Information Theory*. John Wiley and Sons, Inc., N.Y., 1991.

[2] J.P. Crutchfield. Information and its metric. *Nonlinear Structures in Physical Systems - Pattern formation, Chaos and Waves*, pages 119–130, 1990.

[3] J.J. Gibson. *The theory of affordances*. NJ: Laurence Erlbaum Associates.

[4] G. Goodhill, S. Finch, and T. Sejnowski. Quantifying neighbourhood preservation in topographic mappings. *Institute for Neural Computation Technical Report Series, No. INC-9505*, 1995.

[5] E.R. Kandel, J.H. Schwartz, and T.M. Jessel. *Principles of Neural Science, 4th edition*. McGraw-Hill, N.Y., 2000.

[6] F. Kaplan and V.V. Hafner. Information-theoretic framework for unsupervised activity classification. *Advanced Robotics*, 20:1087–1103, 2006.

[7] T.M. Mitchell. *Machine Learning*. McGraw-Hill, N.Y., 1997.

[8] L. Olsson, C.L. Nehaniv, and D. Polani. Sensory channel grouping and structure from uninterpreted sensor data. In *2004 NASA/DoD Conference on Evolvable Hardware June 24-26, 2004 Seattle, Washington, USA*, 2004.

[9] L. Olsson, C.L. Nehaniv, and D. Polani. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. *Connection Science*, 18:121–144, 2006.

[10] D. Pierce and B. Kuipers. Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92:169–229, 1997.

[11] P. Rochat. Self-perception and action in infancy. *Experimental Brain Research*, 123:102–109, 1998.

[12] S. Vijayakumar, A. DSouza, and S. Schaal. Incremental online learning in high dimensions. *Neural Computation*, 17(12):119–130.

[13] J. Weng, J. McClelland, A. Pentland, and O. Sporns. Autonomous mental development by robots and animals. *Science*, 291:599–600, 2001.

[14] I.H. Witten and F. Eibe. *Data Mining*. 2000.