Where do I move my sensors?
Emergence of an internal representation from the sensorimotor flow

Valentin Marcel, Sylvain Argentieri and Bruno Gas

Abstract—In autonomous robotics the question of the representation of the world is of crucial importance for the realization of complex tasks. However, building such a representation is often rooted on human-crafted a priori about the world. But as complexity increases this idea is not adapted anymore: fully autonomous agents in the real world require generalized representations. These must be built from experience and possibly with minimal external assumptions.

This context is perfectly suited to the approach of sensorimotor perception, where the agent has to interpret the effects of naive actions on its inputs that come from exteroceptive and proprioceptive sensors. By exploiting basic sensory invariants, we show that it is possible to project the highly dimensional motor configurations into an internal representation of the sensors' configuration space initially unknown to the agent. This allows the agent to build an internal model of the sensitive configurations.

I. Introduction

Space perception is a central issue in mobile robotics. Indeed, many abilities depend on it, as moving, trajectory planning or obstacle avoidance. Traditional approaches consider that space is something that exists objectively. But the sensorimotor contingencies theory (SMC) [1], [2] claims that it has not to be the case. Space has not to be an established substrate *per se*, but something that an agent may experience via the determination of sensorimotor invariants called *contingencies*. The underlying idea is based on the notion of *compensable sensory changes* proposed initially by Poincaré [3], [4]. Since then, substantial works have been published about considering action in the structuring of perception [5]–[9], some of them aiming to verify Poincaré's idea, but more recently with a growing interest for robotics applications. For instance, [10] introduces "interactive perception" very recently as a set of approaches in robotics concerned with the implication of action in perception.

In the early 2000s, Philipona [11] proposed a first mathematical formalism by defining the sensorimotor law as $s = \Psi(e, m)$ where s denotes the sensation vector, e the vector representing the state of the environment and m the motor state of the agent. Laflaquière et al. [12] have also shown that, beyond the dimension of space, it is also possible to obtain an external space representation by using appropriate partitions of the motor space, resulting in a much more motor oriented than sensor oriented framework. In [13] we have further shown

V. Marcel, S. Argentieri and B. Gas are with Sorbonne Universités, UPMC Univ. Paris 06, UMR 7222, ISIR, F-75005 Paris, France and CNRS, UMR 7222, ISIR, F-75005 Paris, France

that the topological properties of the space of sensory invariant on a sensitive body can be well captured. In this article, we will extend these previous works and provide a framework to describe the process that goes from a quasi-uninterpreted sensorimotor flow to a structured internal representation of the external physical space in a changing environment.

The first section is devoted to the presentation of the set-up priors and notations in the formalization of the sensorimotor refinement. The second section presents an experimental description of the building of the internal representation. Finally, we propose a set of simulations as proofs of concepts illustrating the reconstruction of sensors states in the physical space.

II. Formalization of sensorimotor refinement

A. Priors and notations

Let's consider a naive agent which can interact with its environment by generating commands in the space $\mathcal{M}$ called the motor configuration space. This space is described by the latent variables $m \in \mathcal{M}$ that parameterize the actuators sates: joint angles, relative positions, etc. This agent is also composed of sensors that are rigidly placed on parts of its body. These sensors send a signal $s \in \mathcal{S}$, where $\mathcal{S}$ is the sensory space. The sensory input $s \in \mathcal{S}$ depends both on the current motor state $m \in \mathcal{M}$ and the current physical state of the environment denoted by $\epsilon \in \mathcal{E}$ where $\mathcal{E}$ is the set of all environment states. The function $\Psi_\epsilon(m) = s$ depicts the sensorimotor law. It is assumed here that the absolute position of the agent is fixed in the physical space and can not be displaced by any commands in $\mathcal{M}$ so that the environment physical state ϵ can not depend on m. Of course all considered variables are also function of time t such that one can write $m(t)$, $\epsilon(t)$ and the current sensation is given by $s(t) = \Psi_{\epsilon(t)}(m(t))$.

To go further, the agent requires a level of interpretation of the sensory input; this can be given through multiple structure type: the sensory input is a binary vector or it is composed of real values or it is an image, more generally a complex tensor, etc. These semantics usually enable the agent to perform complex computations that directly provide high level information. However, in this article we will just require the agent to be able to perceptually assess if two sensory inputs are equal or not. This is the only a priori information needed here to build the sensorimotor structure. So let's call

$\delta : \mathcal{S} \times \mathcal{S} \to \{0, 1\}$ the comparison map for sensory inputs which gives the boolean 1 if they are different and 0 if equal. For the sake of simplicity we will generally denote by $s = s'$ the statement $\delta(s, s') = 0$. The sensory comparison already gives a structure to the sensory space, the so-called "sensory invariants". Thanks to the sensorimotor law Ψ we can already state that there are two kinds of sensory invariants.

B. Configuration invariance and sensorimotor points

Two motor configurations are "configuration invariant" when they generate an equal sensory input for all environment states explored by the agent. More formally: take a motor configuration $\mathbf{m} \in \mathcal{M}$ and let $\mathscr{E}$ be the set of environments explored by the agent, then the set $[\mathbf{m}]_{\mathscr{E}} = \{m' \in \mathcal{M}; \Psi_{\epsilon}(\mathbf{m}') = \Psi_{\epsilon}(\mathbf{m}), \forall \epsilon \in \mathscr{E}\}$ is called the *configuration invariant set* of $\mathbf{m}$. These sets are sensory redundant to the agent, without loss of information they can be reduced to "sensorimotor points". This gives the *refined motor set*: $\mathcal{M}/_{\mathscr{E}} = \{[\mathbf{m}]_{\mathscr{E}}; \mathbf{m} \in \mathcal{M}\}$ as the set of points that cannot be discriminated from the obtained sensorimotor flow. There are three possibilities for two motor configurations to be configuration invariant. Firstly, they can correspond to redundant configurations of the kinematic structure of the agent. Secondly, they can correspond to different sensors' poses but identical receptive fields (temperature sensors are generally not sensitive to pure rotation). Finally, it is possible that the environment states have not been rich enough: a portion of the physical space may have not changed its physical properties. As an example, suppose the environment is composed of a cloudless sky which is always blue, any color sensor pointing at any part of this sky will always return the same sensory input.

C. Environment invariance and similarity

Two motor configurations, even if not in the same configuration invariant set, may generate equal sensory inputs for a set of environment states. More formally, for a pair $\mathbf{m}$, $\mathbf{m}' \in \mathcal{M}$, let's call the set: $E(\mathbf{m}, \mathbf{m}') = \{\epsilon \in \mathscr{E}, \Psi_{\epsilon}(\mathbf{m}) = \Psi_{\epsilon}(\mathbf{m}')\}$ the *environment invariant set* of $(\mathbf{m}, \mathbf{m}')$. It is assumed here that the physical properties in the physical space are structured in space so that close points in space share similar physical properties. Therefore a pair of motor configurations, that gives two sensors spatial configurations with close receptive fields, may have a high probability to generate equal sensory inputs. This property is captured in the size of their environment invariant sets, which can be measured using as an example an *a priori probability measure*. The closer they are, the bigger are their environment invariant set. This can be used by the agent as a similarity measure between motor configurations.

To subsume the agent can, from the "configuration invariants" build an internal representation closely related to points in space and from the "environment invariants" add a measure of similarity between these points which are likely to represent continuity in space. The next section is devoted to show how the agent can experimentally build this internal representation.

III. EXPERIMENTAL CONSIDERATIONS AN INTERNAL REPRESENTATION

A. Notations

Let's consider now that the agent has only access to a discrete time sensory input, so that the sensory input is sent every Δt. Let's write $\mathbf{m}[k] = \mathbf{m}(t_0 + k\Delta t)$, $\epsilon[k] = \epsilon(t_0 + k\Delta t)$ and $\mathbf{s}[k] = \mathbf{s}(t_0 + k\Delta t)$ the sampled variables with $k \in \mathbb{N}$ the current sample and t_0 an arbitrary starting time. If the agent sends a new command it will obtain sensorimotor input $(\mathbf{m}[k + 1], \mathbf{s}[k + 1])$ which naturally correspond to $\epsilon[k + 1]$. Thanks to the comparison map δ the agent can compute $\delta(s[k], s[k+1])$. However in order to state that motor configurations $\mathbf{m}[k]$ and $\mathbf{m}[k + 1]$ are in a "configuration invariant" situation for the current environment state, one needs to postulate that the environment has not changed, or very little, during Δt: $\epsilon[k] \sim \epsilon[k + 1]$, otherwise sensory equality can be provoked by a change in physical state and not because of spatial closeness. More generally, this leads to an assumption about environment states dynamics. Moreover, in order to build an internal representation, the agent needs to perform the following *sensorimotor refinement process*.

B. Sensorimotor refinement process

This process consists in repeating again and again the exploration of a specific set $\mathcal{M}_N \subset \mathcal{M}$ of N motor configurations. So, assuming that the sample increments after each movement we can write $\mathcal{M}_N = \{\mathbf{m}[k + 1], \mathbf{m}[k + 2], \ldots, \mathbf{m}[k + N]\}$ and the repetition yields $\mathbf{m}[k + N + 1] = \mathbf{m}[k + 1]$. This exploration should be the most rapid as possible so that the assumption about environment states dynamics is verified: $\epsilon[k + 1] \sim \epsilon[k + 2] \sim \ldots \epsilon[k + N]$. After the exploration, it is possible for the agent to wait for changes in the environment states such that $\epsilon[k + N] \nsim \epsilon[k + N + 1]$ which allows the exploration to be repeated for a different physical state and enable refinement. From L repetitions of the exploration of the motor exploration set $\mathcal{M}_N$ the agent obtains the matrix $S[L]$ of sensations of size $L \times N$ composed of:

$$S[L] = \begin{bmatrix} \mathbf{s}[1] & \ldots & \mathbf{s}[N] \\ \mathbf{s}[N + 1] & \ldots & \mathbf{s}[2N] \\ \vdots & \vdots & \vdots \\ \mathbf{s}[(L-1)N + 1] & \ldots & \mathbf{s}[LN] \end{bmatrix}. \quad (1)$$

The first dimension of $S[L]$ corresponds to the repetitions of the exploration which ideally corresponds to different environment states, it is the index of refinement for $\mathcal{M}_N$. The second dimension corresponds to the number of motor configurations in the explored set.

C. Internal representation structure

Then, the next step consists in the computation of statistics on sensory invariance which allows to build a structured internal representation. Let's apply the comparison map δ to comparable sensations, the pairs of sensations that are on the same row because the environment state is assumed stationary. We obtain pairwise comparisons for all refinement indexes $l \in [1, L]$ in the matrix $\Delta[l]$ of size $N \times N$ such that:
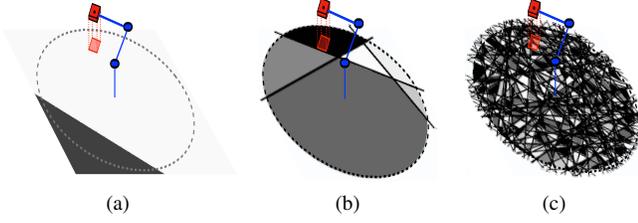
Fig. 1. Simulated set-up and projection of sensory invariants in the working space. (a) corresponds to the sensory invariants after 1 random environment state. (b) corresponds to the sensory invariants after having seen 3 random environment states, the gray shades correspond to the sequences of the 3 possible successive sensory inputs at each point in the working space. (c) corresponds to the sensory invariants after 100 environments.

$$\Delta[l] = \begin{bmatrix} 0 & \delta_{12}[l] & \cdots & \delta_{1N}[l] \\ \delta_{21}[l] & 0 & \ldots & \delta_{2N}[l] \\ \vdots & \vdots & \ldots & \vdots \\ \delta_{N1}[l] & \delta_{N2}[l] & \ldots & 0 \end{bmatrix}. \quad (2)$$

where $\delta_{ij}[l] = \delta(\mathbf{s}[(l-1)N+i], \mathbf{s}[(l-1)N+j])$ is the comparison of the sensations obtained from motor configuration i and j of the exploration set $\mathcal{M}_N$ at repetition l. One can obtain the current similarity between all motor configurations by computing the size of the observed *invariant environment sets* for all pairs. However, we are more interested in the dissimilarity, which gives a notion of distance between points. Let's call $D[L]$ the dissimilarity matrix. It is composed of elements d_{ij} corresponding to the mean number of repeated explorations for which the motor configurations i and j have generated different sensations. Simply put, $\forall i, j \in [1, N]$

$$d_{ij}[L] = \frac{1}{L} \sum_{l=1:L} \delta_{ij}[l]. \quad (3)$$

Or in matrix form:

$$D[L] = \frac{1}{L} \sum_{l=1:L} \Delta[l]. \quad (4)$$

It can be shown that the dissimilarity is in fact a metric on the set of current discrete equivalence classes $[i]_L = \{j \in [1, N]; \delta_{ij}[l] = 0, \forall l \le L\}$.

At the beginning all points are not sensory distinguished. As long as the refinement process goes, the number of distinguished points increases until the finest refinement is obtained for a sufficiently large number of repetitions L. Then the dissimilarity $D[L]$ can be used as a structure on the set of finest points corresponding to the "configurations invariants" of the current sampled motor exploration set. If the number of sampled motor configuration is dense enough, the obtained structure is likely to be representative of the topology of the configuration invariants in the physical space.

IV. SIMULATIONS

We can illustrate the approach with a simple system. The presented agent is the serial arm in figure 1 with two degrees of freedom controlled by two joint actuators through motor configurations $\mathbf{m} = [m_1, m_2] \in [0, 2\pi]^2$. The generated
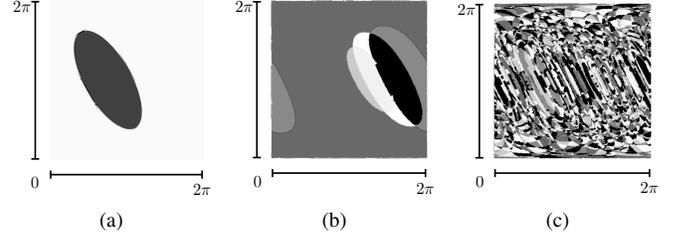

Fig. 2. Configuration invariants in the motor space. In (a), (b) and (c) colors respectively corresponds with colors in (1a), (1b) and (1c)

sensations are given by a single pixel black and white camera $\mathbf{s} \in \{0, 1\}$ placed at the end-effector of the serial arm so that it is displaced inside a disk forming the working space. The environment is a black and white background image that is randomly changed.

A. Simulated refinement

To show the evolution of the refinement, consider the most simple environment set-up for the background: a black and a white half planes separated by a random straight line which intersects with the working space, one possible state is shown in subfigure (1a). Keeping track of the successive sensory invariant during the repeated explorations of the motor space with different environment states, the agent should be able to refine it as shown in subfigures (2a), (2b) and (2c) for respectively 1, 3 and 100 different environment states. The configuration invariant areas depicted in shades of gray in the motor space correspond to the portion in the working space with the same shade in subfigures (1a), (1b) and (1c). The more repetitions there is, the more there will be "sensorimotor points" in the refined motor set.

Let's now show the structure obtained from the statistics on sensory invariants.

B. Structure from statistics on sensory invariants

Let's first consider the same agent with the same environment. As an exploration set $\mathcal{M}_N$, the agent sends $N = 400$ naive random commands in $[0, 2\pi]^2$ and gets the corresponding motor configurations into the set $\mathcal{M}_N$. The N motor states are each associated to their corresponding sensor poses (i.e. points in the working space), which are shown in Subfigure (3a) with a color depending on their distance to the center. After running the refinement process with $L = 1000$ repetitions of the $N = 400$ generated motor configurations, the agent obtains a dissimilarity between the sensorimotor points which can be compared to the Euclidean distance matrix between the $2D-$poses of the sensor. The comparison is shown in a variogram in Subfigure (3b) as a scatter plot. In this figure, one can directly verify if small distances correspond to small dissimilarities and if so the representative space has the same topology than the $2D-$euclidean pose space. Here we see a linear correlation between dissimilarity and distance, therefore topological properties are indeed conserved by the refinement process. Moreover, one can use Classical Multidimensional Scaling [14] on the dissimilarity matrix in order to visualize

(a) Poses ground truth (b) Variogram dissimilar-ity/Euclidean distance. (c) Projected representa-tive space with cMDS.
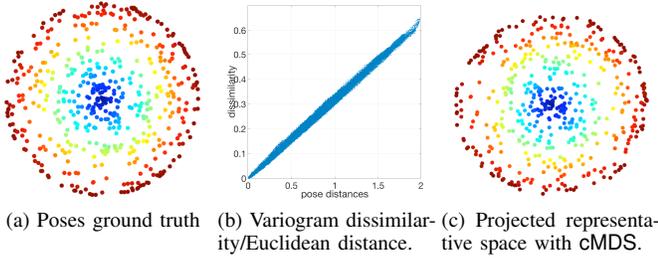
Fig. 3. Results for the illustrative example. LEFT: 400 target poses in the plane. Colors correspond to the radius of poses. RIGHT: Internal representation with preserved topological structure. Colors correspond to the radius of poses in (a).

(a) Poses ground truth. (b) Variogram dissimilar-ity/Euclidean distance. (c) Internal representa-tion with Isomap.

Fig. 5. Results for the more challenging example. LEFT: 400 target poses, the color are arbitrarily drawn to show the topological structure. RIGHT: Internal representation with preserved topological structure.
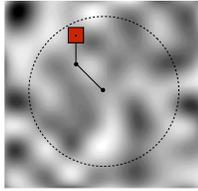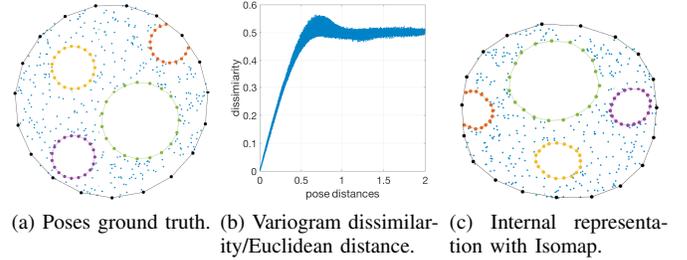

Fig. 4. A more challenging environment: random black and white noise.

the internal representation. The projection in 2 dimensions is given in the subfigure (3c) where the obtained points are colored with the corresponding Euclidean distance to the center. This qualitatively confirms that the topological structure of poses in space are preserved in the internal representation.

Let's now take a more challenging environment for which sensory invariant statistics are not linear in spatial distance and test if the internal representation still captures topological properties. In the following simulations, the environment is now made of background images of normalized spatially coherent noise whose statistics are invariant through translation or rotation. The implemented noise function is very similar to Perlin noise [15] and is shown in figure 4.

The exploration is repeated for $L = 1000$ different environments on $N = 400$ motor configurations. In order to show the conservation of the topology, the motor exploration set has been modified to add some holes and contours which are represented in subfigure (5a). After the repetitions, one can compare the obtained dissimilarity to the pairwise Euclidean distances in the working space in subfigure (5b). From the subfigure, one can see that the small distances are preserved; therefore the topology is also preserved. However the dissimilarity saturates for high Euclidean distances due to low sensory correlation between far poses. Instead of MDS, one need to use an algorithm that preserves small distances such as $\epsilon-$Isomap [16] which gives the satisfying result of subfigure (5c).

V. CONCLUSION

In this article, the authors have a proposed an approach for the emergence of a spatially-related structure from the sensorimotor flow with very weak assumptions on the sensory structure or a priori knowledge about the physical world. It has been shown that from a simple comparison rule between sensations, one can derive an internal representation which captures continuity in the physical space. A description of the process has been provided as well as some simple results to illustrate the interest of the approach. The internal representation can be improved using information-based principles as in [17]. Extended work could focus on the generalization to more complex set-up and the use of the internal representation to guide intrinsically motivated sensorimotor behaviors.

REFERENCES

[1] J. K. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness." *The Behavioral and brain sciences*, vol. 24, no. 5, pp. 939–973; discussion 973–1031, Oct. 2001.

[2] J. K. O'Regan, "Solving the "real" mysteries of visual perception: The world as an outside memory," *Canadian Journal of Psychology*, vol. 46, no. 3, pp. 461–488, 1992.

[3] H. Poincaré, "L'espace Et la Géométrie," *Revue de Métaphysique Et de Morale*, vol. 3, no. 6, pp. 631–646, 1895.

[4] ——, "Science and hypothesis," *London: Walter Scott Publishing*, 1905.

[5] J. J. Gibson, *The senses considered as perceptual systems*. Boston: Houghton, 1966.

[6] P. Bach-y Rita, C. Collins, F. Saunders, B. White, and L. Scasdden, "Vision substitution by tactile image projection," *Nature*, vol. 221, no. 5184, pp. 963–964, mar 1969.

[7] J. J. Gibson, *The ecological approach to visual perception*. Boston: Houghton, 1979.

[8] A. Berthoz, *Le sens du mouvement*. Odile Jacob, 1997.

[9] A. Noë, *Action in perception*. The MIT Press, 2004.

[10] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme, "Interactive perception: Leveraging action in perception and perception in action," *IEEE Transactions on Robotics*, vol. PP, no. 99, pp. 1–19, 2017.

[11] D. Philipona, J. K. O'Regan, and J.-P. Nadal, "Is there something out there?: Inferring space from sensorimotor dependencies," *Neural Comput.*, vol. 15, no. 9, pp. 2029–2049, 2003.

[12] A. Laflaquière, J. K. O'Regan, S. Argentieri, B. Gas, and A. Terekhov, "Learning agents spatial configuration from sensorimotor invariants," *Robotics and Autonomous Systems*, vol. 71, pp. 49–59, September 2015.

[13] V. Marcel, S. Argentieri, and B. Gas, "Building a sensorimotor representation of a naive agent's tactile space," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 2, pp. 141–152, June 2017.

[14] W. S. Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

[15] K. Perlin, "An image synthesizer," *ACM Siggraph Computer Graphics*, vol. 19, no. 3, pp. 287–296, 1985.

[16] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[17] A. S. Klybin, D. Polani, and C. L. Nehaniv, "Keep your options open: An information-based driving principle for sensorimotor systems," *PLOS ONE*, vol. 3, no. 12, 2008.