Curiosity driven exploration of sensory-motor mappings

Karaoguz, Cem; Potapova, Ekaterina; Drix, Damien; Hülse, Martin

Publication date:
2011

Citation for published version (APA):
Karaoguz, C., Potapova, E., Drix, D., & Hülse, M. (2011). *Curiosity driven exploration of sensory-motor mappings*. Paper presented at Capo Caccia Cognitive Neuromorphic Engineering Workshop, Aberystwyth, United Kingdom of Great Britain and Northern Ireland. http://hdl.handle.net/2160/7197

Deliverable for the IM-CLeVeR Spring School at the Capo Caccia Cognitive Neuromorphic Engineering Workshop, 1-7 May 2011

Curiosity Driven Exploration of Sensory-motor Mappings

Cem Karaoguz
Research Institute for Cognition and Robotics
University of Bielefeld
33615 Bielefeld, Germany
ckaraogu@cor-lab.uni-bielefeld.de

Damien Drix
School of Informatics
University of Sussex
Falmer, Brighton
BN1 9QJ, UK
damien.drix@gmail.com

Ekaterina Potapova
Vienna University of Technology
1040 Vienna, Austria
potapova@acin.tuwien.ac.at

Martin Huelse
Department of Computer Science
Aberystwyth University,
SY23 3DB, UK
msh@aber.ac.uk

***Abstract*—Adaptive mechanisms present numerous benefits to artificial systems interacting with the environment. For instance learning sensory-motor mappings eliminates the necessity of (re)calibration process for active vision systems when the aspects of the system or the environment is changed. The amount of time spent on the adaptation process may be reduced if an efficient strategy is followed. We propose such a mechanism which uses intrinsic motivation for sensory-motor learning. We tested our framework in a series of experiments where the environment the system learned a mapping for was systematically changed. Our curiosity-driven framework yielded distinct exploration patterns where distorted areas were concentrated immediately after a change was applied.**

I. INTRODUCTION

The human eye has its highest resolution in the center of the retina, the fovea. In order to recognize objects the eye moves rapidly fixating different points of interest. This movement space is the gaze space of the eye. However, when an object has been located (at some location in the gaze space) a robot will want to move its hand to the same location in order to grasp the object. The movement space of the arm is called the reach space and a key question is how the reach space can be correlated (both in human brains and in robots) with the gaze space so that any point in space has the same correspondence in the two very different gaze and reach spaces. Several frameworks for learning such sensory-motor mappings were introduced [1, 2, 3, 4, 5]. Such systems can adapt learned mappings in response to changes in the perceived environment such as caused by camera replacement or changes in the physical environment. This adaptation of learned mappings eliminates the necessity of (re)calibration.

An interesting issue with such adaptive systems is how to guide the system to gather information from regions that have not yet been learned in the learning space and how to avoid getting stuck on areas that present samples with high signal to noise ratio. Intrinsic motivation can lead artificial learning systems towards situations in which the system maximises its learning progress [6]. Such a motivational drive makes the system focus on situations which are neither too predictable nor too unpredictable, thus permitting autonomous mental development. Such a framework may also improve the learning process significantly. We present a method that uses intrinsic motivation to drive the learning process and make the system explore interesting areas in the environment.

A variant of sensory-motor mapping learning scheme explained in [1] was used as the application platform. The system consists of a static camera and a laser device that can rotate around pan/tilt axes. The system learns a mapping between laser pan/tilt commands and the retinal position of the laser pointer in the right camera image (Fig. 1). We tested our method against two other methods: random sampling and gap method. In experiments systematic changes were applied to the scene that required the learning system to adapt the learned mapping. Our method provided system with an exploration strategy where the altered regions were focused immediately after changes. This may result in a faster adaptation process.

II. METHODS

A. System Overview

Out robotic scenario includes a laser system and an active vision system, their spatial organization is shown in Fig. 3. The active vision system consists of a single camera (which provides RGB 750x576 image data). Both camera and laser were mounted on a motorized pan-tilt-verge unit. The position of the motors is controlled by indicating the values of their absolute target position, or the change from the current position. The laser and active vision system operate independently. Relation between two independent physical systems - laser position on the screen and tilt-verge camera coordinates can be represented by sensory-motor mapping and can be learned for a given scenario. The final mapping should enable the overall system to look where it reaches to (i.e. enable transformation from laser coordinate space to camera coordinate space), based on the already established reach- and gaze control.

B. Sensory-motor Mapping Learning

The domains of the laser and camera movements are both represented as a two-dimensional coordinate system. The purpose of the gaze control is to move the camera in such a way that the visual stimuli, the projection of the laser pointer on the camera image, is situated in the centre of the image. The input RGB images are filtered with a colour filter, in our

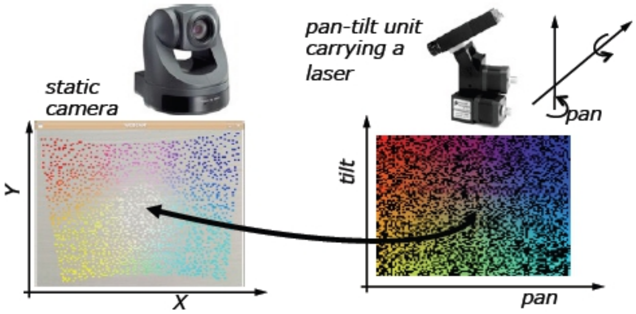


Fig. 1. Mapping scheme.

case red. Output images are gray-scale and non-zero values indicate the appearance of the filtered colour in the original image.

Assuming two spaces $\Theta \in R^n$ (in our case laser space and $n = 2$) and $I \in R^m$ (in our case camera space and $m = 2$) a mapping M stores the pairs of vectors $([\Theta_p, \Theta_t], [x, y])$, which represent real examples how one point in space Θ is related to one point in space I (Fig. 1). This is called a link. Because there is a modal relation between links, there is also an additional property of bi-directionality (i.e. Θ refers to I and vice versa).

Whenever a new training sample is encountered, a distance metric (here the euclidean distance) can be used to find the link stored in the mappings that is closest to that new sample. The existing link and the distance metric can be used a a prediction and prediction error, respectively. When the prediction error is above a certain threshold $T = 300$, a new link corresponding to the new sample is added to the mappings. Thus, the maximum prediction error also defines the minimum distance between two distinct links in the mappings. Additionally, links in the mapping have a maximum age $Q = 80$. The age of a link is set to 0 when it is created and when it results in a successful prediction, and increased by 1 at every timestep otherwise. The oldest link with an age above Q is removed from the mapping at every timestep.

The main difficulty with the described mapping method arises when the explored environment changes: a robust algorithm dealing with the appearance of new relations between the two spaces is required. It should update links in the mapping, that is delete links corresponding to the former relation and add links reflecting the new relation. It should also do so efficiently, spending more time in regions that have changed than in regions that have not.

C. Curiosity-driven Exploration Framework

The proposed curiosity-driven exploration framework for the sensory-motor mapping learning system defines a method for selecting learning examples from the sampling space (i.e. laser motor space) so as to concentrate on *interesting* regions of space. In the context of this experiment, interesting regions are those where the prediction accuracy can be improved. This includes areas where no mappings have yet been learned and areas where the previous mappings have been invalidated by change, but excludes regions where no accurate predictions can ever be made.

For this purpose, [6] defined a single metric reflecting the learning progress in a particular region of the sampling space as the difference between the past mean error in a $[t_{past} - n, t_{past}]$ time window, and the current mean error for the $[t - n, t]$ window. Higher values of this metric were associated with a higher probability of choosing the next learning example from that region of space, thus driving the system to explore preferentially the regions where the highest improvement in prediction accuracy is being achieved at any particular time. In their experiment, an adaptive partitioning of the sensori-motor space into discrete regions was used as well, with the aim of letting the curiosity-driven process partition the space according to its underlying structure. That structure, however, did not change over time.

Here, we chose to defined a fixed set of regions in laser motor space instead, and let the sensori-motor relations in each of these regions change over time. This prevented us from using the same metric as [6] to drive the exploration process: with the nearest-neighbour prediction scheme used, a change in the mapping to be approximated results in a steep increase of the online prediction error until the affected region has been substantially explored and remapped. It is therefore not possible to rely on an online estimation of the instantaneous learning progress; an offline estimation would be possible but impractical in our experimental setup as it would require performing a systematic exploration of the entire laser motor space, thus negating any benefits of the curiosity-driven method in terms of learning speed.

Instead, we chose to let the curiosity-driven exploratory behaviour emerge from the interaction of two distinct factors: a drive towards regions where the existing mappings are not yielding accurate predictions (reacting to change), and a drive away from regions where no long-term improvements can be obtained (analogous to boredom).

Fig. 2 illustrates the basic principle of this method. According to this training examples are sampled from a Gaussian distribution over the laser motor space:

$$G(\mu, \sigma) = \exp\left(\frac{-(\Theta_p - \mu)^2}{2\sigma^2}\right), \tag{1}$$

where μ is the mean of the Gaussian, defining the position of the Gaussian peak in the sample space and σ is the spread of the distribution. In this work the Gaussian distribution is defined over the pan dimension Θ_p of the sample space and tilt commands were sampled from a uniform distribution. This simplification was possible since distortions applied in the experiments were only in the pan dimension. Extending the framework to a multi-dimensional space is trivial. The Θ_p values were normalised to a $[0, 1]$ range.

The mean of the Gaussian (i.e. position of its peak in the sample space) is computed using the centroid of the last N links added to the mapping:

$$\mu = \frac{1}{N}\sum_{i}^{N}\Theta_p^i, \quad (2)$$

where $N = 3$ is the size of the history, Θ_p^i position of ith link from the last added to the mapping. This drives the exploration to focus on areas where new links have been recently added.

The spread of the Gaussian is coupled with the performance of the mapping in the region where the sample was taken. For this purpose, the pan dimension of the mapping was divided into $R = 5$ equal regions and the performance of the mapping in these regions is constantly monitored with a metric called success rate. Success rate s_i for region $i \in R$ is computed as:

$$s_r = s_r + t_s(L - s_r), \quad (3)$$

where $t_s = 0.2$ is a time constant for smoothing over time and $L = 0$ if the training sample at this timestep resulted in a new link being added (eg. the prediction error was above the threshold), $L = 1$ otherwise. After μ is computed the region where μ belongs to is derived and defined as r. Consequently, spread factor σ is computed as:

$$\sigma(t, s_r) = \sigma_0 + k_b t_b + k_f(t_f - s_r)^2, \quad (4)$$

where t_b is the number of timesteps since the last new link was added to the mappings and $k_b = 0.05$ is the boredom coefficient, s_r is the success rate of region r, $\sigma_0 = 0.03$ is the base sigma, $k_f = 8.0$ is the failure coefficient, $t_f = 0.4$ is the failure threshold. The term $(t_f - s_r)$ is called the failure metric and kept between 0 and 1. Also, the spread factor has a higher bound of $\sigma_{MAX} = 1.0$. These values were experimentally found to be satisfactory in a 1-dimensional faster-than-realtime simulation of the system and later transposed to the much slower hardware experiments.

Starting with blank or random mappings, the success rate s_r will initially be low in all regions and the $k_f(t_f - s_r)^2$ term will be high, resulting in a large σ value and a wide sampling distribution. Thus, a mostly uniform exploration of space will be performed.

As the success rate increases towards and above t_f, the σ value becomes smaller and the system concentrates on that particular region. When the maximum link density has been reached (eg. none of the prediction are above the error threshold), t_b will start to increase. This absence of improvement causes the spread σ to get wider and eventually the focus μ is shifted to another area.

If change occurs in a previously learned region, t_b will be reset to 0 and the high past success rate of that region will result in a narrow gaussian distribution of samples (determined by σ_0). As the links corresponding to the new mapping are being added, that narrow distribution will enhance the re-test of newly-added links and prevent the success rate in that region from being driven below the t_f threshold.

Conversely, newly-added links in unlearnable areas (high signal to noise ratio) will typically fail on re-test, which will drive the success rate for that region below the threshold and increase the width of the sampling distribution, driving the system away from that region.

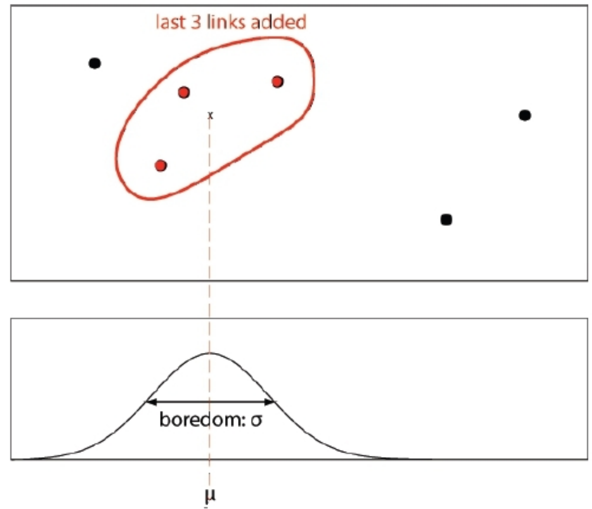


Fig. 2. Curiosity-driven exploration defines a Gaussian distribution on the sample space to select learning examples. The mean of the Gaussian depends on the last links created and the spread of the Gaussian is coupled with the performance of the mapping in a defined region.

III. EXPERIMENTS

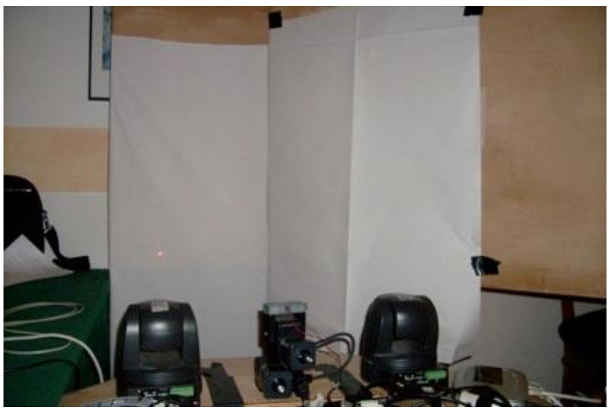
A. Experimental Setup

The experimental setup consists of a laser device mounted on a motorised unit with two degrees of freedom and an RGB camera (Fig. 3(a)). The motorised unit allow the laser to move around pan and tilt axes in the space. The pointer of the laser is reflected on a white board situated in front of the laser and camera. The plane of the white board is positioned orthogonal to the direction of the camera (i.e. parallel to the image plane) and laser in zero position. The task was to learn a mapping between the pan-tilt movements of the laser and the projection of the laser pointer on images from the camera. After 200 iterations of learning, right side of the workspace is distorted by placing a second white board in front of the laser/camera system in a slanted position (Fig. 3(b)). The slanted surface was positioned in the right side of the laser/camera system, left side was left unchanged. Such an alteration in the environment required system to relearn the areas in the mapping which were affected by the change. After iteration 600 the scene was returned to its original state. This introduced a change once more in the right side of the workspace, which has to be adapted again.

B. Results

We observed behaviour of the system using three different methods: random sampling, gap method and the proposed method. Random sampling chose random points in the laser pan/tilt space for learning. The gap method samples from regions where the number of links is less densely populated [1]. We compared which regions of the workspace was used

(a)

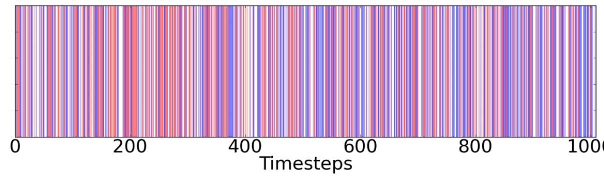


(b)

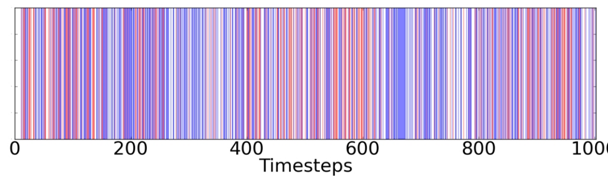
Fig. 3. The hardware setup used in the experiments.

to acquire learning samples throughout the experiments done with these three methods. For this purpose the laser pan/tilt space was divided into three equally sized clusters in pan direction. Since the distortion affected the mapping mostly in pan direction, no additional clustering was done in tilt direction. The clusters were defined as left (from -2300 to -1117 units in pan space, the whole tilt space), centre (from -1117 to 66 units in pan space, the whole tilt space) and right (from 66 to 1250 units in pan space, the whole tilt space).

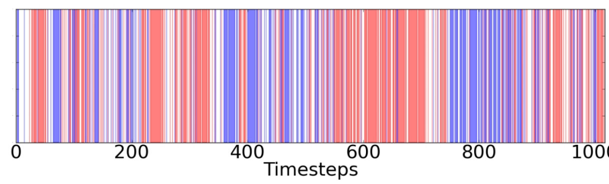
Fig. 4 shows which cluster chosen pan/tilt samples belong to at every timestep in the experiments using random sampling, gap method and the proposed method. The distribution of samples in the experiment using random sampling does not show any significant preference on the left, centre or right. Samples in the experiments using the gap method were mostly taken from the left side of the workspace after the changes although they were applied to the right side of the scene (iterations 200 and 600). This means that the system had tendency to sample away from the altered regions in the mapping. This can be explained with the fact that the gap method prefers regions with low number of links for sampling. As soon as the right side of the scene was changed, the learning mechanism started adding new links to the corresponding area in the mapping. This could result in a relatively low number of links in the opposite region from where the gap method preferred sampling. For the experiments with the proposed method the distribution of learning samples are more dense on the right side after distortions were applied. The system using the proposed method showed a significant preference on altered regions to select the learning samples from.



(a)



(b)



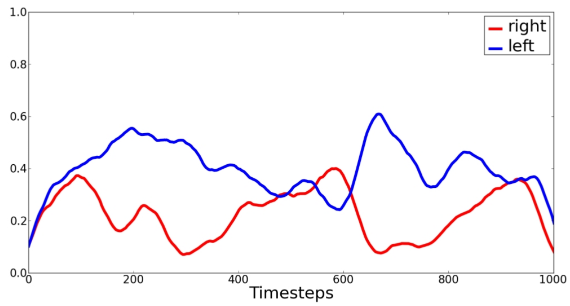
(c)

Fig. 4. Behaviour of the system in three different experiments. Plots show the distribution of selected learning samples in three clusters (left, right and centre clusters are shown in blue, red and white colours respectively) in experiments using (a) random sampling (b) the gap method and (c) the curiosity driven exploration method. Changes in the right side of the environment were applied after 200 and 600 timesteps.

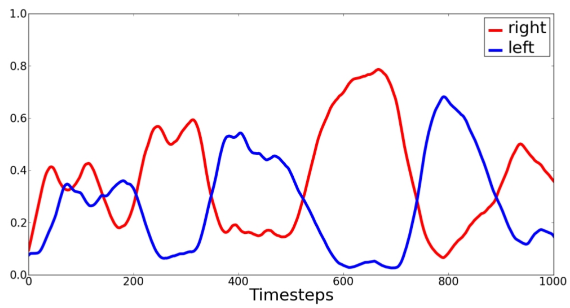
For a quantitative analysis normalised mean of number of samples in the clusters were computed using a sliding window (window size was 100 timesteps) was computed for experiments using the gap and proposed methods. The results for left and right clusters are shown in Fig. 5. Using the proposed method the rate of samples from the right side reaches up to 60% after the first distortion was introduced and 80% after the second. The tendency of the gap method was the opposite: the rate of samples from the right side dropped below 20% after the first modification in the scene and to 10% after the second.

IV. CONCLUSIONS AND FUTURE WORK

We presented a curiosity driven exploration method for learning systems and tested it against two different approaches in a sensory-motor mapping learning paradigm. The experiments with other methods showed either no preference (random sampling) or preference on the converse regions (gap method) for exploration when a specific region in the mapping was changed and had to be relearned. The proposed method results in significantly different exploration patterns. Our method showed tendency to explore the areas in the mapping, which



(a)



(b)

Fig. 5. Mean values of selected learning sample proportions in two clusters (computed with three clusters, left and right are plotted, centre is not shown) over the last 100 timesteps from the experiments with (a) the gap method and (b) the curiosity driven exploration method in experiments using (a) random sampling (b) gap method.

correspond to the altered part of the scene. The exploration behaviour observed in the gap method experiments maybe due to the fact that the gap method samples from regions with low number of links in order to improve parts that underfit in the mapping. While this may be a good strategy for learning a balanced mapping, adaptation periods may take longer using this approach. We showed that the curiosity driven exploration framework explained in this work can keep the spotlight of exploration on altered areas in case of changes in the environment. Moreover, it does not get stuck on regions,which are practically impossible to learn. Therefore, it may fasten the adaptation process significantly. More quantitative analysis is required to validate this hypothesis.

Acknowledgements

This work was supported by the PC-FP7 projects IM-CLeVeR and ROSSI, and through UK EPSRC grant EP/C516303/1. Author C.K. is supported by Honda Research Institute Europe GmbH.

References

[1] M. Hülse and M. Lee, "Adaptation of coupled sensorimotor mappings: An investigation towards developmental learning of humanoids," in *11th Int. Conf. on Simulation of Adaptive Behavior (SAB), LNAI 6226, Springer*, 2010, pp. 468–477.

[2] C. Karaoguz, M. Dunn, T. Rodemann, and C. Goerick, "Online adaptation of gaze fixation for a stereo-vergence system with foveated vision," in *Proceedings of Int. Conf. on Advanced Robotics 2009*, Jul. 2009, pp. 1–6.

[3] M. Pagel, E. Mael, and C. Von Der Malsburg, "Self calibration of the fixation movement of a stereo camera head," *Autonomous Robots*, vol. 5, no. 3-4, pp. 355–367, 1998.

[4] F. Chao, M. H. Lee, and J. J. Lee, "A developmental algorithm for ocular-motor coordination," *Robotics and Autonomous Systems*, vol. 58, pp. 239–248, 2010.

[5] M. Hülse, S. McBride, and M. Lee, "Fast learning mapping schemes for robotic hand-eye coordination," *Cognitive Computation*, vol. 2, no. 1, pp. 1–16, 2010.

[6] P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 265–286, Apr. 2007.