Combining Multimodal Sensory Input for Spatial Learning

Thomas Strösslin[1], Christophe Krebser, Angelo Arleo[2], and Wulfram Gerstner[1]

[1] Laboratory of Computational Neuroscience, EPFL, Lausanne, Switzerland
[2] Laboratoire de Physiologie de la Perception et de l'Action, Collège de France-CNRS, Paris, France

Abstract. For robust self-localisation in real environments autonomous agents must rely upon multimodal sensory information. The relative importance of a sensory modality is not constant during the agent-environment interaction. We study the interrelation between visual and tactile information in a spatial learning task. We adopt a biologically inspired approach to detect multimodal correlations based on the properties of neurons in the superior colliculus. Reward-based Hebbian learning is applied to train an active gating network to weigh individual senses depending on the current environmental conditions. The model is implemented and tested on a mobile robot platform.

1 Introduction

Multimodal information is important for spatial localisation and navigation of both animals and robots. Combining multisensory information is a difficult task. The relative importance of multiple sensory modalities is not constant during the agent-environment interaction, which makes it hard to use predefined sensor models.

The hippocampal formation of rats seems to contain a spatial representation which is important for complex navigation tasks [1]. We propose a spatial learning system in which external (visual and tactile) and internal (proprioceptive) processed signals converge onto a spatial representation. Here we focus on the dynamics of the interrelation between visual and tactile sensors and we put forth a learning mechanism to weigh these two modalities according to environmental conditions. Our system is inspired by neural properties of the superior colliculus, a brain structure that seems to be involved in multimodal perception [2,3,4].

2 Related Work

The rat's hippocampal formation receives highly processed multimodal sensory information and is a likely neural basis for spatial coding [1,5,6,7]. Hippocampal place cells discharge selectively as a function of the position of the rat in the environment.

J.R. Dorronsoro (Ed.): ICANN 2002, LNCS 2415, pp. 87–92, 2002.

The superior colliculus (SC) is involved in oculomotor responses and in the processing of multimodal information (visual, tactile and auditory) [2]. There is also evidence that SC contains neurons with spatial firing properties [3,4].

Robotic models of multimodal integration [8,9,10] are mostly based on probabilistic sensor fusion techniques in the framework of occupancy grids which cannot easily be transposed into biological models. Most current biological models of localisation and navigation [11,12] focus on a single external modality and neglect the problem of combining multimodal information.

3 Proposed Model

We adopt a hippocampal place code similar to [12] as a spatial map. Hebbian learning is used to correlate idiothetic (path integration) and allothetic (visual and tactile) stimuli with place cell activity.

Here we model the integration of visual and tactile signals into a common allothetic representation. The weight of each sense is modulated by a gating network which learns to adapt the importance of each sense to the current environmental condition. Intermodal correlations are established using uni- and multimodal units inspired by neurons in the superior colliculus. The model is implemented on a Khepera mobile robot platform. Figure 1 shows the architecture of the system.

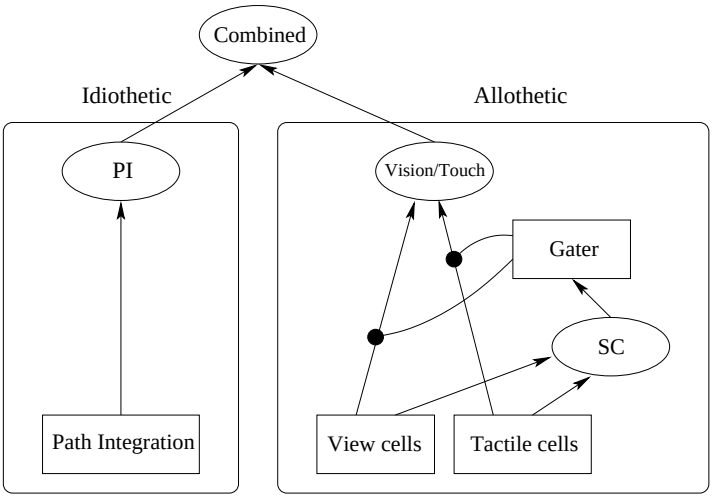

Fig. 1. Architecture of our spatial localisation system

3.1 Neural Coding of Sensory Input

During exploration, the hippocampal place code is established. At each new location, *view cells* (VCs) and *tactile cells* (TCs) convert the agent's sensory input to neural activity.

Sensory cell (VC and TC) activity r_i depends on the mean distance between the current sensor values x_i and the stored values w_{ij} at creation of cell i.

$$r_i = \exp(-\frac{(\frac{1}{N}\sum_{j=1}^{N}|w_{ij} - x_j|)^2}{2\sigma^2}) \tag{1}$$

Visual input from the greyscale camera is processed using a set of Gabor filters. The magnitudes of the complex filter responses are stored in the weights w_{ij} of the created VC (see [13] for more details).

Tactile input from each of the eight infrared proximity sensors is scaled to $[0, 1]$ and stored in the weights w_{ij} of the created TC.

3.2 Unimodal and Multimodal Cells in Superior Colliculus

Intermodal correlations between visual and tactile input are established in the exploration phase using uni- and multimodal neurons inspired by the superior colliculus (SC) [2,3]. Sensory cells project to the input layer of SC which consists of unimodal visual (UVs) and tactile (UTs) cells. Those unimodal cells project to multimodal cells (MMs) in the output layer of SC. The architecture of our SC model is shown in Figure 2 (a).

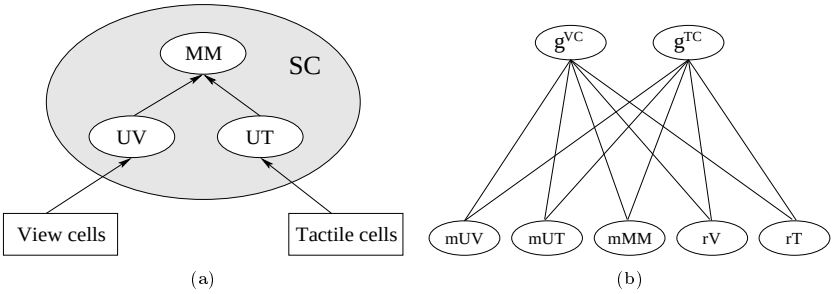

Fig. 2. Architecture of our superior colliculus model (a) and the gating network (b)

Whenever the agent receives strong visual and tactile input simultaneously, it creates a tactile and a visual unimodal cell. Synapses between sensory cells (TCs and UTs) are established and adapted using a Hebbian learning rule

$$\Delta w_{ij} = \eta \; r_i(r_j - w_{ij}) \tag{2}$$

where r_i is the postsynaptic unimodal cell, r_j is the sensory cell and η the learning rate. The same happens for VCs connecting to UVs. The firing rate r_i of an unimodal cell is given by the weighted mean activity of its presynaptic neurons j.

$$r_i = \frac{\sum_j w_{ij} r_j}{\sum_j w_{ij}} \tag{3}$$

The most active unimodal cells connect to a new multimodal output cell and synapses are learnt according to equation 2. The firing rate r_i of a multimodal cell i differs from equation 3 in that both UTs and UVs need to be active to trigger the firing of a multimodal cell

$$r_i = \tanh\left(k\,(\frac{\sum_{j\in \mathrm{UV}} w_{ij} r_j}{\sum_{j\in \mathrm{UV}} w_{ij}})(\frac{\sum_{j\in \mathrm{UT}} w_{ij} r_j}{\sum_{j\in \mathrm{UT}} w_{ij}})\right) \tag{4}$$

where k is a constant.

3.3 Learning the Gating Network

During exploration, the most active sensory cells establish connections to a newly created allothetic place cell. The synaptic strengths evolve according to equation 2.

The firing rate r_i of an allothetic place cell i is the weighted mean activity of its presynaptic neurons j where all inputs from the same modality are collectively modulated by the gating value g^{VC} or g^{TC} respectively.

$$r_i = g^{\mathrm{VC}}\left(\frac{\sum_{j\in \mathrm{VC}} w_{ij} r_j}{\sum_{j\in \mathrm{VC}} w_{ij}}\right) + g^{\mathrm{TC}}\left(\frac{\sum_{j\in \mathrm{TC}} w_{ij} r_j}{\sum_{j\in \mathrm{TC}} w_{ij}}\right) \tag{5}$$

The gating network consists of five input neurons which are fully connected to two output neurons as shown in figure 2 (b). mUV, mUT and mMM are the mean UV, UT and MM activity. rV is the mean pixel brightness of the unprocessed camera image. rT is the mean proximity sensor input. The output neurons g^{VC} and g^{TC} provide the gating values of equation 5

We train the gating network to adapt its output values to the current environmental condition. During learning, the agent moves randomly in the explored environment and tries to localise itself. At each timestep, illumination in the environment is turned off with probability P_L or left unchanged otherwise. The weights are updated according to equation 2, but Δw_{ij} is also modulated by a reward signal q.

The reward q depends on two properties of the allothetic place code: (a) variance around centre of mass σ_{pc} and (b) population activity act_{pc}. Positive reward is given for compact place cell activity (ie. small variance σ_{pc}) and reasonable mean population activity. Negative reward corresponds to very disperse place coding or low activity. The equation for the reward q is as follows

$$q = \Big[\exp(-\frac{(\sigma_{pc}-\sigma_{opt})^2}{2\sigma_{size}^2}) - d_{size}\Big]\Big[\exp(-\frac{(act_{pc}-act_{opt})^2}{2\sigma_{act}^2}) - d_{act}\Big] \tag{6}$$

where $\sigma_{opt}, \sigma_{size}, d_{size}, act_{opt}, \sigma_{act}, d_{act}$ are constants.

4 Results and Conclusions

Experiments are conducted on a Khepera mobile robot. An $80 \times 80cm$ boarded arena placed on a table in a normal office serves as environment. A rectangular-shaped object is placed in the arena to increase the amount of tactile input to the system.

Figure 3 (a) shows the gating values for the visual and tactile senses after learning the gating network. Most of the time, visual input is the only activated modality. Everytime the robot is near an obstacle however, the tactile sense is assigned a slightly higher importance than vision. The abrupt changes are due to the binary nature of the tactile sensors.

Figure 3 (b) shows the gate values when the illumination is reduced by 80%. Most of the time, vision and tactile senses receive equal importance. Whenever an obstacle is near, however, the agent relies mostly on its tactile input.

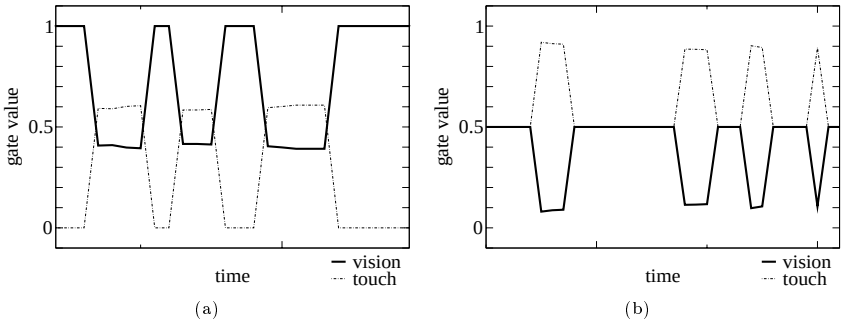

Fig. 3. Gate values in openfield and border positions. (a) good illumination. (b) almost no light

The main difficulty in learning the importance of sensory input lies in determining the reliability and uncertainty of a percept. We use the mean place cell activity and the activity variance around the centre of mass as a quality measure to change the weights of the gating network. Accessing the variance in spatial representations might be difficult to motivate biologically. Plausible neural mechanisms that measure place code accuracy should be found.

The brain certainly uses various methods to evaluate the relevance of sensorial input. We are working on more biologically plausible methods to assess place code quality.

References

1. O'Keefe, J., Nadel, L.: The Hippocampus as a Cognitive Map. Clarendon Press, Oxford (1978)

2. Stein, B.E., Meredith, M.A.: The Merging of the Senses. MIT Press, Cambridge, Massachusetts (1993)
3. Cooper, B.G., Miya, D.Y., Mizumori, S.J.Y.: Superior colliculus and active navigation: Role of visual and non-visual cues in controlling cellular representations of space. Hippocampus 8 (1998) 340–372
4. Natsume, K., Hallworth, N.E., Szgatti, T.L., Bland, B.H.: Hippocampal thetarelated cellular activity in the superior colliculus of the urethane-anesthetized rat. Hippocampus **9** (1999) 500–509
5. Quirk, G.J., Muller, R.U., Kubie, J.L.: The firing of hippocampal place cells in the dark depends on the rat's recent experience. Journal of Neuroscience **10** (1990) 2008–2017
6. Save, E., Cressant, A., Thinus-Blanc, C., Poucet, B.: Spatial firing of hippocampal place cells in blind rats. Journal of Neuroscience **18** (1998) 1818–1826
7. Save, E., Nerad, L., Poucet, B.: Contribution of multiple sensory information to place field stability in hippocampal place cells. Hippocampus **10** (2000) 64–76
8. Elfes, A.: Using occupancy grids for mobile robot perception and navigation. IEEE Computer **22** (1998) 46–57
9. Castellanos, J.A., Neira, J., Tardos, J.D.: Multisensor fusion for simultaneous localization and map building. IEEE Transattions on Robotics and Automation **17** (2001) 908–914
10. Martens, S., Carpenter, G.A., Gaudiano, P.: Neural Sensor susion for spatial visualization on a mobile robot. In Schenker, P.S., McKee, G.T., eds.: Proceedings of SPIE, Sensor Fusion and Decentralized Control in Robotic Systems, Proceedings of SPIE (1998)
11. Burgess, N., Recte, M., O'Keefe, J.: A model of hippocampal function. Neural Networks **7** (1994) 1065–1081
12. Arleo, A., Gerstner, W.: Spatial cognition and neuro-mimetic navigation: A model of hippocampal place cell activity. Biological Cybernetics, Special Issue on Navigation in Biological and Artificial Systems 83 (2000) 287–299
13. Arleo, A., Smeraldi, F., Hug, S., Gerstner, W.: Place cells and spatial navigation based on 2d visual feature extraction, path integration, and reinforcement learning. In Leen, T.K., Dietterich, T.G., Tresp, V., eds.: Advances in Neural Information Processing Systems 13, MIT Press (2001) 89–95