Learning Sensory Correlations for 3D Egomotion Estimation

Cristian Axenie(✉) and Jörg Conradt(✉)

Neuroscientific System Theory Group, Department of Electric and Computer Engineering, Technische Universität München, Karlstrasse 45, 80333 Munich, Germany
{cristian.axenie,conradt}@tum.de

Abstract. Learning processes which take place during the development of a biological nervous system enable it to extract mappings between external stimuli and its internal state. Precise egomotion estimation is essential to keep these external and internal cues coherent given the rich multisensory environment. In this paper we present a learning model which, given various sensory inputs, converges to a state providing a coherent representation of the sensory space and the cross-sensory relations. The developed model, implemented for 3D egomotion estimation on a quadrotor, provides precise estimates for roll, pitch and yaw angles.

Keywords: Egomotion estimation · Cross-modal learning · Multisensory fusion · Mobile robots

1 Introduction

Human perception improves through exposure to the environment. A wealth of sensory streams which provide a rich experience continuously refine the internal representations of the environment and own state. Furthermore, these representations determine more precise motor planning [1]. An essential component in motor planning and navigation, in both real and artificial systems, is egomotion estimation. Given the multimodal nature of the sensory cues, learning cross-modal correlations improves the precision and flexibility of motion estimates.

Various methods, ranging from neural circuitry implementations to statistical correlation analysis, have been developed to extract correlational structure in sensory data. Related work [2] used a combination of simple biologically plausible mechanisms, like WTA circuitry, Hebbian learning, and homeostatic activity regulation, to extract relations in artificially generated sensory data. After learning, the model was able to infer missing quantities given the learned relations and available sensors. Moreover, due to recurrent connectivity, the sensory representations were continuously refined, de-noising the encoded real-world variable. Finally, due to the constraints imposed by the learned relations, the model was able to combine consistent and correlated data (i.e. cue-integration) and discriminate and penalize inconsistent data contributions (i.e. decision making).

S.P. Wilson et al. (Eds.): Living Machines 2015, LNAI 9222, pp. 329–338, 2015.
DOI: 10.1007/978-3-319-22979-9_32

Using a different neurally inspired substrate, [3] combined competition and cooperation in a self-organizing network of simple processing units to extract coordinate transformations in a robotic scenario. Inspired by sensorimotor transformations in the prefrontal cortex, the algorithm produced invariant representations and a topographic map representation of the scene guiding a robot's behaviour.

Going away from biological inspiration, [4] used nonlinear canonical correlation analysis to extract relations between sets of multi-dimensional random variables. The model implemented a change of representation from the variables input space to a new space of canonical variants. Subsequently, the model mapped the representations back to the initial space minimising the relative mismatch between the original data and the mapping. The extracted relation was encoded in the weights configuration maximising the correlation between the canonical variants.

Although these methods provide good results for dedicated scenarios, they lack the capability to handle non-uniform sensory data distributions and, at the design stage, need judicious parametrisation and prior information about sensory data. The proposed model tries to address these aspects. It uses relatively simple mechanisms to provide a flexible way to learn sensory correlations for precise egomotion estimation.

2 A Neurally Inspired Model for Learning Sensory Correlations

During development, the biological nervous system must constantly combine various sources of information and moreover track and anticipate changes in one or more of the cues. Furthermore, the adaptive development of the functional organisation of the cortical areas seems to depend strongly on the available sensory inputs, which gradually sharpen their response, given the constraints imposed by the cross-sensory relations [5].

Following this principle, we propose a model based on Self-Organizing Map (SOM) and Hebbian Learning (HL) as main ingredients for extracting underlying relations in sensory data. In order to introduce the proposed model, we provide a simple example in Figure 1. In the basic dual-modality scenario, the relation between sources of sensory data is extracted between the pair of sensory modalities, as shown in Figure 1b. In a multisensory scenario, all-to-all connections between modalities are considered, and similar dynamics applies to each possible modality pair.

The input SOMs are responsible for extracting the statistics of the incoming data, depicted in Figure 1a, and encoding sensory samples in a distributed activity pattern, as shown in Figure 1c. This activity pattern is generated such that the closest preferred value of a neuron to the input sample will be strongly activated and will decay, proportional with distance, for neighbouring units. Figure 2 provides a detailed depiction of processing stages which take place when sensory input samples are presented to the network. Using the SOM distributed

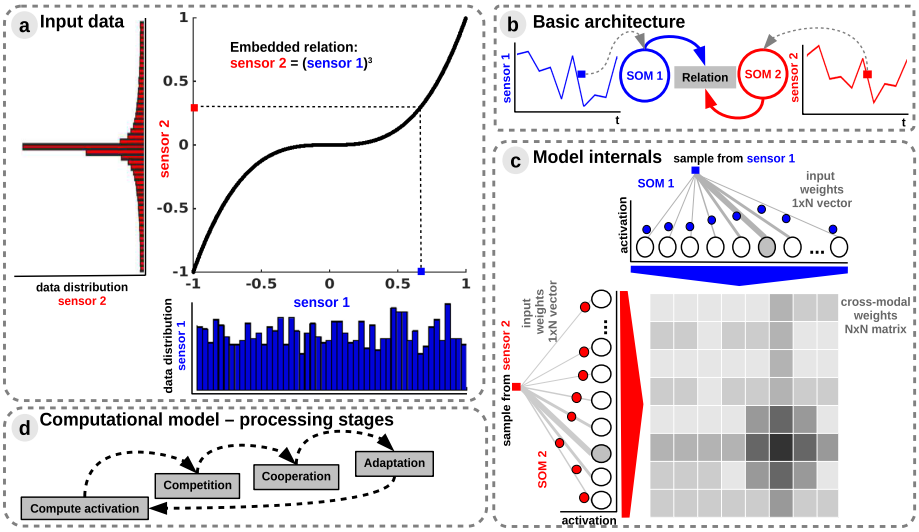

Fig. 1. Model architecture. a) Input data resembling a nonlinear relation and its distribution. b) Basic architecture; c) Model internal structure. d) Processing stages.

representation, the model learns the boundaries of the input data, such that, after relaxation, the SOMs provide a topological representation of the input space. We extend the basic SOM in such a way that each neuron not only specialises in representing a certain (preferred) value in the input space, but also learns its own sensitivity (i.e. tuning curve shape). Given an input sample, $s^p(k)$ at time step k, the network follows the processing stages depicted in Figure 1d and explicitly presented in Figure 2. For each $i-th$ neuron in the $p-th$ input SOM, with the preferred value $w^p_{in,i}$ and $\xi^p_i(k)$ tuning curve size, the sensory elicited activation is given by

$$a^p_i(k) = \frac{1}{\sqrt{2\pi}\xi^p_i(k)} e^{\frac{-(s^p(k)-w^p_{in,i}(k))^2}{2\xi^p_i(k)^2}}. \quad (1)$$

The winner neuron of the $p-th$ population, $b^p(k)$, is the one which elicits the highest activation given the sensory input at time k

$$b^p(k) = \underset{i}{argmax}\ a^p(k). \quad (2)$$

During self-organisation, at the input level, competition for highest activation is followed by cooperation in representing the input space (second and third step in Figure 1d). Given the winner neuron, $b^p(k)$, the interaction kernel,

$$h^p_{b,i}(k) = e^{\frac{-||r_i - r_b||^2}{2\sigma(k)^2}}. \quad (3)$$

allows neighbouring cells (found at position r_i in the network) to precisely represent the sensory input sample given their location in the neighbourhood $\sigma(k)$.

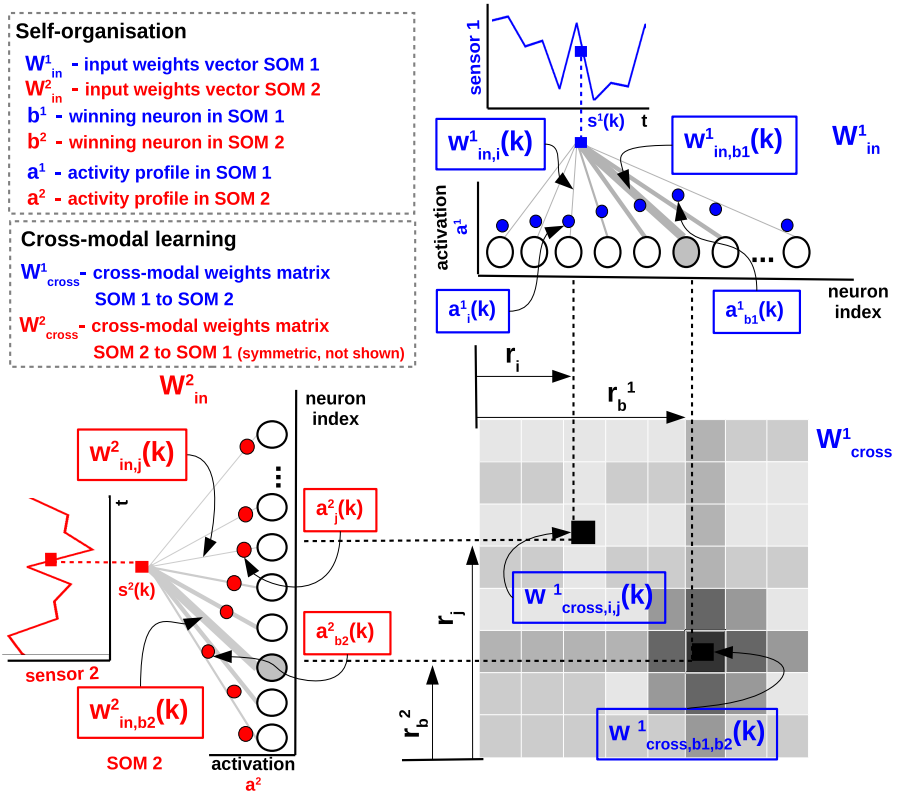

Fig. 2. Detailed architecture of the model and processing stages

The interaction kernel in Equation 3, ensures that specific neurons in the network specialise on different areas in the sensory space, such that the input weights (i.e. preferred values) of the neurons are pulled closer to the input sample,

$$\Delta w^p_{in,i}(k) = \alpha(k) h^p_{b,i}(k)(s^p(k) - w^p_{in,i}(k)). \tag{4}$$

This corresponds to the adaptation stage in Figure 1d and ends with updating the tuning curves. Each neuron's tuning curve is modulated by the spatial location of the neuron, the distance to the input sample, the interaction kernel size, and a decaying learning rate $\alpha(k)$,

$$\Delta \xi^p_i(k) = \alpha(k) h^p_{b,i}(k)((s^p(k) - w^p_{in,i}(k))^2 - \xi^p_i(k)^2). \tag{5}$$

If we consider learned tuning curves shapes for 5 neurons in the input SOMs (i.e. neurons 1, 6, 13, 40, 45), depicted in Figure 3, we notice that higher input probability distributions are represented by dense and sharp tuning curves. Whereas lower or uniform probability distributions are represented by more sparse and wide tuning curves. Using this mechanism, the network optimally allocates resources (i.e. neurons): a higher amount to areas in the input space,

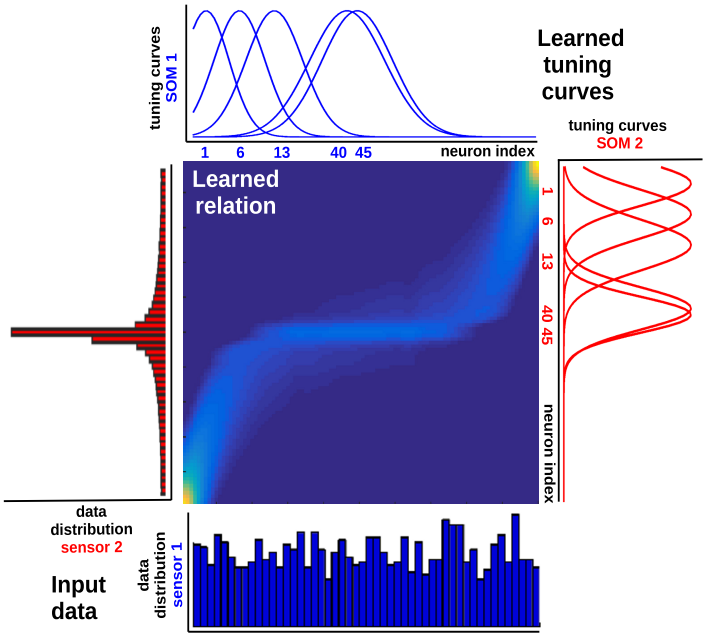

Fig. 3. Extracted sensory relation and data statistics using the proposed model

which need a finer representation; and a lower amount for more coarsely represented areas. This feature emerging from the model is consistent with recent work on optimal sensory encoding in neural populations [6]. This claims that, in order to maximise the information extracted from the sensory streams, the prior distribution of sensory data must be embedded in the neural representation.

The second component of the proposed model is the Hebbian linkage. This consists of a fully connected matrix of synaptic connections between neurons in each input SOM. Using an all-to-all connectivity pattern each SOM unit activation is projected to the Hebbian matrix. The Hebbian learning process is responsible for extracting the co-activation pattern between the input layers (i.e. SOMs), as shown in Figure 1c, and for eventually encoding the learned relation between the sensors, as shown in Figure 3. As one can see in Figure 3 - central panel, the connections between uncorrelated (or weakly correlated) neurons in each population (i.e. w_{cross}) is suppressed (i.e. darker color) while correlated neurons connections are enhanced (i.e. brighter color). The effective correlation pattern encoded in the w_{cross} matrix, imposes constraints upon possible sensory values. Moreover, after the network converges, the learned sensory dependency will make sure that values are "pulled" towards the correct (i.e. learned) corresponding values, will detect outliers, and will allow inferring missing sensory quantities. Formally, Hebbian connection weights, $w^p_{cross,i,j}$, between neurons i, j in each of the input SOM population are updated using

$$\Delta w^p_{cross,i,j}(k) = \eta(k)(a^p_i(k) - \overline{a}^p_i(k))(a^q_j(k) - \overline{a}^q_j(k)), \tag{6}$$

where

$$\overline{a}_i^p(k) = (1 - \beta(k))\overline{a}_i^p(k-1) + \beta(k)a_i^p(k), \tag{7}$$

and $\eta(k)$, $\beta(k)$ are monotonic decaying functions. The original Hebbian postulate only allows for an increase in synaptic weight between synchronously firing neurons. In order to prevent unlimited weight growth, we use a modified Hebbian learning rule (i.e. covariance rule, Equation 6) to allow for weight decreases when neurons fire asynchronously. The proposed mechanism uses a time average of pre- and postsynaptic activities, $\overline{a}_i^p(k)$, defined in Equation 7. When neurons fire synchronously in a correlated manner their connection strengths increase, whereas if their firing patterns are anticorrelated the weights decrease.

Self-organisation and correlation learning processes evolve simultaneously, such that both representation and correlation pattern are continuously refined. Moreover, the timescales of the two processes align, such that once the representations are learned in the SOMs the correlation pattern in the Hebbian connection matrix becomes sharper.

In the initial example we consider a set of values drawn from a uniform random distribution (i.e. sensor 1), Figure 1a, to which we apply a power-law, and we compute a second input (i.e. sensor 2) drawn from a Gaussian distribution. The network is fed with random pairs from the two datasets. After learning, the Hebbian connectivity matrix encodes the input data relation, as shown in Figure 3. Moreover, the tuning curves encode the input data distribution: narrower spaced for higher probability distributions and widely spaced for lower (or uniform) distributions of the input data.

In order to develop and test the specific hypotheses of the proposed model, we apply it for a quadrotor 3D egomotion estimation, briefly depicted in Figure 4.

3 Instantiating the Model

For the basic testing scenario, a quadrotor hovers (remote controlled) in an uncluttered environment, while an overhead camera system keeps track of its position and orientation.

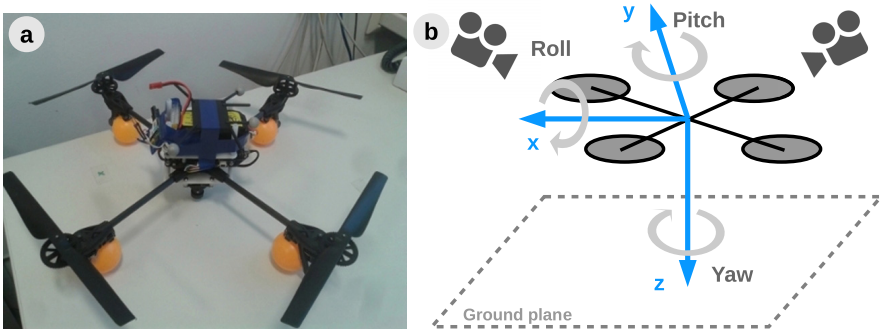

Fig. 4. Experimental setup: a) Quadrotor platform; b) Reference system alignment and ground truth camera tracking system

After the flight, preprocessed data from the available sensors (i.e. gyroscope, accelerometer and a magnetic sensor, Figure 5a) is fed to the model to extract the relations between the sensors for each of the three degrees of freedom (i.e. roll, pitch and yaw). The reference system setup is depicted in Figure 4b. Although initially the model considers all sensory contributions for the estimation of all motion components, as shown in Figure 5b, it will enforce only those connections providing a coherent relation for each degree of freedom, as shown in Figure 5c. As mentioned earlier, all-to-all connections between sensors are considered, but only those encoding coherent relations (i.e. contributions to same degree of freedom estimate) are enforced and considered for subsequent fusion. For roll and

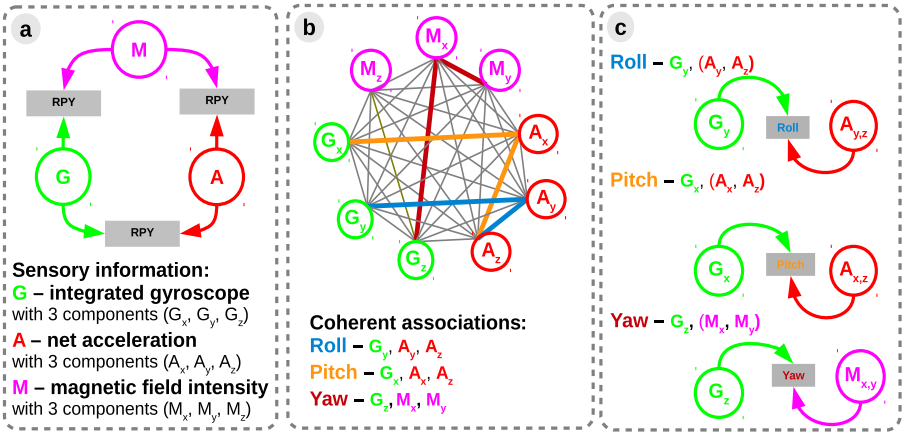

Fig. 5. Network instantiation for 3D egomotion estimation: network structure and sensory associations. a) Sensors configuration; b) Network connectivity; c) Sensory associations for learning.

pitch angles (i.e. rotation around the x and y reference frame axes), the network learns the relation between the roll and pitch angle estimates from integrated gyroscope data and rotational acceleration components (i.e. orthogonal x and y with respect to z reference frame axes). Similarly, the yaw angle is extracted by learning the relation between the yaw angle estimate from integrated gyroscope data (i.e. absolute angle) and aligned magnetic field components from the magnetic sensor (i.e. projected magnetic field vectors on orthogonal x and y reference frame axes). The sensory associations are not arbitrary, but rather represent the dynamics of the system and are consistent with recently developed modelling and control approaches for quadrotors [7], [8]. To make use of the learned relations we decode the Hebbian connectivity matrix using a relatively simple optimisation method [9]. After learning, we apply sensory data from one source and compute the sensory elicited activation in its corresponding (presynaptic) SOM neural population. Furthermore, using the learned cross-modal Hebbian weights and the presynaptic activation, we can compute the postsynaptic activation.

Given that the neural populations encoding the sensory data are topologically organised (i.e. adjacent values coding for similar places in the input space), we can precisely extract (through optimisation) the sensory value for the second sensor given the postsynaptic activation pattern. Without using an explicit function to optimise, but rather the correlation in activation patterns in the input SOMs, the network can extract the relation between the sensors.

4 Experimental Results

In order to validate the extracted relations, we use the aforementioned mechanism to extract the roll, pitch, and yaw estimates for the quadrotor scenario. Figure 6 presents a decoupled view for each degree of freedom, depicting the learned relations and estimation accuracy.

We observe in Figure 6a that the learned relations resemble the nonlinear functions (i.e. arctangent) used in typical modelling approaches, although preserving irregularities in the cross-sensory relations. The learned cross-sensory relations, encoded in the Hebbian matrix, provide the intrinsic constraints between the sensory cues contributing to the estimate of each degree of freedom. For roll estimation, Figure 6b - upper panel, the network learns the relation between net rotational acceleration provided by the accelerometer and the absolute roll angle estimate provided by the gyroscope. Given that accelerometer data is noisy and gyroscope data drifts, as a consequence of integration process, the network is able to "pull" the values of the two cues towards the "correct" value of the roll angle as given by ground truth (accelerometer RMSE: $< 2\%$, gyroscope RMSE: $< 3\%$).

For pitch estimation the network extracts the nonlinear dependency between the accelerometer data and the gyroscope data. Although both cues follow the trend of change in angle, as shown in Figure 6b - middle panel, the accelerometer is overestimating, due to the noisy signal and the overall limited motion of the drone on this axis. The gyroscope contribution was able to modulate the accelerometer contribution such that the overall estimates are acceptable (accelerometer RMSE: $< 7\%$, gyroscope RMSE: $< 3\%$).

Finally, for yaw estimation the network uses the gyroscope absolute angle and the magnetometer contribution, based on magnetic field readings on the other two axes. Interestingly, albeit the fact that the yaw estimate of the magnetometer follows the trend, Figure 6b - lower panel, there is an intrinsic offset (RMSE:$\sim$ 15%) visible from $t = 5s$. Investigating during many test flights, we noticed that the current change generated when arming the rotors introduced a significant modification in magnetic field distribution, subsequently reflected in the magnetometer readings. In the current setup, the network is not able to explicitly compensate for the offset, as one can see in Figure 6a - lower panel, where the co-activation pattern is not sharp like for roll and pitch.

As our results show, the model is able to extract the underlying data statistics without any prior information, as depicted in Figure 3, such that the sensory data distribution was learned directly from the input data. Moreover, following

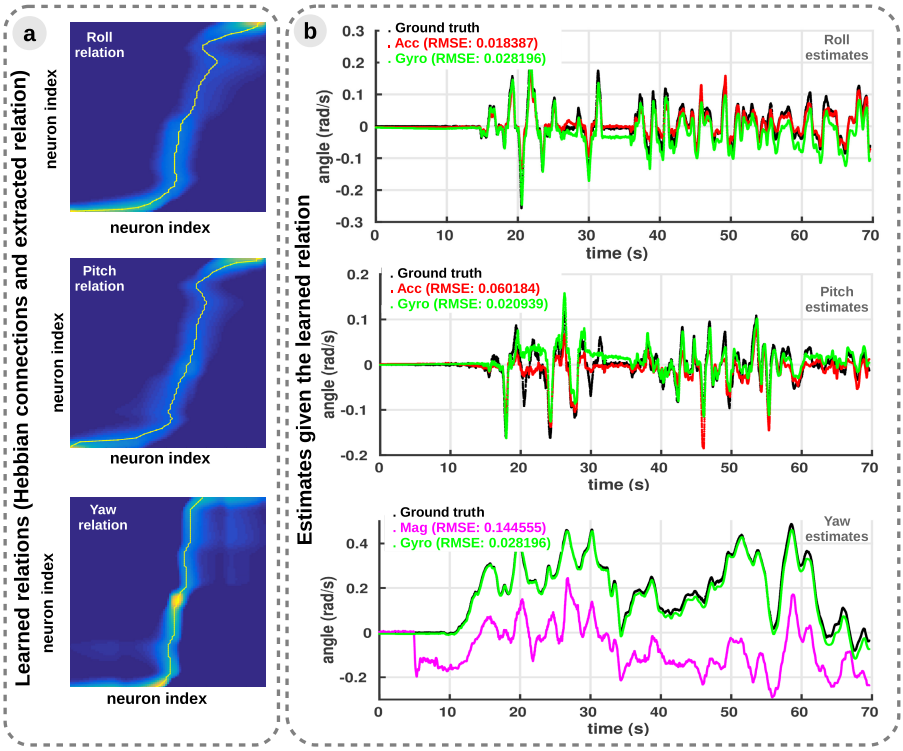

Fig. 6. Network instantiation for 3D egomotion estimation: a decoupled view analysis. a) Learned relations; b) Estimation quality using learned relations.

the statistics of the data, the network allocates more neurons to represent areas in the sensory space with a higher density such that the cross-sensory relations are sharpened, visible in Figure 6a.

As shown in Section 2, there is no specific parameter tuning routine (Figure 1d) to handle different kinds of input data for different scenarios. The generic processing elements (i.e. SOM, Hebbian learning) and their extensions (i.e. tuning curve adaptation, covariance update) ensure that the network first learns (in an unsupervised manner) the structure of the data, and then uses this representation to sharpen its correlational structure. Moreover, given the learned relations, the network is able to infer missing quantities in the case of sensor failures. As the relation is encoded as a synaptic weight, after learning, it is enough to provide samples from one sensor, encode them in the SOM, and project the activity pattern through the Hebbian matrix. The resulting activity pattern, subsequently decoded, will provide the missing real-world sensory value.

The proposed learning scheme extends [10], in which given various sensory inputs and simple relations defining inter-sensory dependencies, the model infers a precise estimate of the perceived motion. Now, by alleviating the need to explicitly encode sensory relations in the network dynamics, we propose a

model providing flexible and robust multisensory fusion, without prior modelling assumptions, and using only the intrinsic sensory correlation pattern.

5 Conclusions

Given relatively complex and multimodal scenarios in which robotic systems operate, with noisy and partially observable environment features, the capability to precisely and rapidly extract estimates of egomotion critically influences the set of possible actions. Utilising simple and computationally effective mechanisms, the proposed model is able to learn the intrinsic correlational structure of sensory data and provide more precise estimates of egomotion. Moreover, by learning the sensory data statistics and distribution, the model is able to judiciously allocate resources for efficient representation and computation without any prior assumptions and simplifications. Alleviating the need for tedious design and parametrisation, it provides a flexible and robust approach to multisensory fusion, making it a promising candidate for robotic applications.

References

1. Gibson, E.J.: Principles of Perceptual Learning and Development, pp. 369–394. ACC Press (1969)
2. Cook, M., Jug, F., Krautz, C., Steger, A.: Unsupervised learning of relations. In: Diamantaras, K., Duch, W., Iliadis, L.S. (eds.) ICANN 2010, Part I. LNCS, vol. 6352, pp. 164–173. Springer, Heidelberg (2010)
3. Weber, C., Wermter, S.: A self-organizing map of sigma-pi units. Neurocomputing **50**, 2552–2560 (2007)
4. Mandal, A., Cichoki, A.: Non-Linear Canonical Correlation Analysis Using Alpha-Beta Divergence. Entropy **15**, 2788–2804 (2013)
5. Westermann, G., Mareschal, D., Johnson, M.H., Sirois, S., Spratling, M.W., Thomas, M.S.: Neuroconstructivism. Dev. Sci. **10**, 75–83 (2007)
6. Ganguli, D., Simoncelli, E.P.: Efficient Sensory Encoding and Bayesian Inference with Heterogeneous Neural Populations. Neural Computation **26**, 2103–2134 (2014)
7. Hyon, L., Park, J., Lee, D., Kim, H.J.: Build your own quadrotor. IEEE Robotics and Automation Magazine, 33–45 (2012)
8. Lee, J.K., Park, E.J., Robinovich, S.N.: Estimation of attitude and external acceleration using inertial sensor measurement during various dynamic conditions. IEEE Transactions on Instrumentation and Measurement **61**, 2262–2273 (2012)
9. Brent, R.P.: An Algorithm with Guaranteed Convergence for Finding a Zero of a Function. Algorithms for Minimization without Derivatives. Dover Books on Mathematics, pp. 47–58 (2013)
10. Axenie, C., Conradt, J.: Cortically inspired sensor fusion network for mobile robot egomotion estimation. Robotics and Autonomous Systems (2014)