

Cortically Inspired Sensor Fusion Network for Mobile Robot Heading Estimation

Cristian Axenie and Jörg Conradt

Fachgebiet Neurowissenschaftliche Systemtheorie, Fakultät für Elektro- und
Informationstechnik, Technische Universität München, 80290 München, Germany
{cristian.axenie, conradt}@tum.de

Abstract. All physical systems must reliably extract information from their noisily and partially observable environment, such as distances to objects. Biology has developed reliable mechanisms to combine multi-modal sensory information into a coherent belief about the underlying environment that caused the percept; a process called sensor fusion. Autonomous technical systems (such as mobile robots) employ compute-intense algorithms for sensor fusion, which hardly work in real-time; yet their results in complex unprepared environments are typically inferior to human performance. Despite the little we know about cortical computing principles for sensor fusion, an obvious difference between biological and technical information processing lies in the way information flows: computer algorithms are typically designed as feed-forward filter-banks, whereas in Cortex we see vastly recurrent connected networks with intertwined information processing, storage, and exchange. In this paper we model such information processing as distributed graphical network, in which independent neural computing nodes obtain and represent sensory information, while processing and exchanging exclusively local data. Given various external sensory stimuli, the network relaxes into the best possible explanation of the underlying cause, subject to the inferred reliability of sensor signals. We implement a simple test-case scenario with a 4 dimensional sensor fusion task on an autonomous mobile robot and demonstrate its performance. We expect to be able to expand this sensor fusion principle to vastly more complex tasks.

Keywords: Cortical inspired sensor fusion, graphical network, local processing, mobile robotics.

1 Introduction

Environmental perception enables a physical system to acquire and build an internal representation of significant information within its environment. As an example of such an internal state, accurate self-motion perception is an essential component for spatial orientation, navigation and motor planning for both real and artificial systems. A system can build its spatial knowledge using a combination of multiple sources of information, conveyed from self-motion related signals (e.g. odometry or vestibular signals), but also from static external environmental cues (e.g. visual or auditory) [1].

In this paper we focus on one component of self-motion: heading estimation. The key aspect in this problem is to minimize interference or conflicts between multiple dynamic and static sources of perceived spatial information.

1.1 State-of-the-Art in Sensor Fusion Algorithms

Sensor fusion finds wide application in many areas of robotics, starting from object recognition to environmental mapping and localization. Many of the state-of-the-art sensor fusion methods are based on a probabilistic framework, typically using Bayes' rule to optimally combine information [2-4]. This rule defines the way to compute the posterior probability distribution of a state/quantity of interest from prior knowledge about the state and a likelihood function (e.g. knowledge about how likely different values of the state will give rise to observed sensory data). Incoming sensory data will progressively push the belief "away" from the initial prior towards beliefs (posteriors) that better reflect the data. Different methods translate the posterior into a quantity [4-5], but need to balance the trade-off between algorithmic complexity and available computing resources.

1.2 Brains Aim at Flexibility and Robustness versus Optimality

Engineering approaches for sensor fusion typically aim for optimal solutions, as demonstrated in various amazing robotic demonstrators [6]. Handling dynamic and complex features in real-world scenarios, however, engineered approaches typically require intense modifications (in the form of hand tuning), and often lack the ability to easily translate to new settings. With respect to flexibility and robustness neurobiology demonstrates vastly superior performance over today's engineered systems.

Building on principles known from information processing in the brain our proposed model for sensor fusion migrates away from classical computing paradigms: tightly coupled processing, storage and representation is replaced with a loosely interconnected network of distributed computing units, each with local memory, processing and interpretation abilities. Mutual influence among the different units that represent sensor data implicitly performs the integration of multiple sensory information.

In this paper Section 2 provides a broad description of the proposed neural network model. Following, Section 3 presents results from a simple real-world mobile robotic scenario (heading estimation), and section 4 highlights the main features of the developed model and presents ideas to extend the developed architecture.

2 Model Description

Large-scale cortical networks provide a framework to integrate evidence from neuro-anatomical and neuro-physiological studies on distributed information processing in the cerebral cortex. Elementary sensory processing functions are localized in discrete recurrently interacting cortical areas, [7-8], whereas complex functions (e.g. cue integration) are processed in parallel involving large parts of the brain [9].

Building up on such cortical architectural details we design a network model capable of finding a consistent representation of self-motion using a distributed processing scheme. Each unit in the network encodes a certain feature of self-motion, while the connectivity determines mutual influence between the units, so that information in each unit is consistent with each other. For the given task of heading estimation different sensors encode the movement in their own coordinate system and units. To extract a global motion estimate the network needs to learn correlations to maintain all information in agreement.

2.1 Generic Network Architecture

The proposed model is based on a recurrent network with dynamics given by information exchange between its nodes. Each node in the network internally stores a representation of some perceived aspect of self-motion, which is a simplified form of a more generic multi-dimensional map based representation and processing paradigm successfully applied to fast visual scene interpretation in [10]. In this work a network with multiple maps encoding various aspects of visual interpretations (e.g. light intensity, optic flow, and camera rotation) found global mutual consistency, by each map independently trying to be consistent with information in neighboring maps. Such map based representation and processing paradigms are supported by results from neuro-physiology and computational neuroscience, in terms of spatio-temporal correlations used to learn topographic maps for invariant object recognition [8], [11].

Following these principles our proposed 4D sensor fusion network is composed of four fully connected nodes, which mutually influence each other. Information stored in each node is represented by a neural population code [13], in which each neuron represents a preferred angular value. Nodes' mutual influence is characterized by the physical / mathematical relationship between the four sensor representations. Relationships between nodes can be interpreted as modulating neuronal connectivity between the network's different populations [13]. To minimize mismatch between values encoded in the nodes, the network will push all representations towards an equilibrium state in which simultaneously all relationships are - approximately - satisfied. This resulting state shows the inferred quantities in each of the nodes' representation with respect to both sensor connection and internal network belief.

2.2 Dynamics

Network dynamics is described by a random gradient descent update process at every convergence step, so that the value in a selected node will take a small step towards minimizing the mismatch with a relationship in which the node is involved. Each node represents a single value in form of a neuronal population code. Relations between the nodes in the network are arbitrary functions of v variables, $f_{m_i, m_j, \dots, m_v}(t) = 0$, where $v < N$ (the number of nodes). The generic update rule for a node m_i with respect to the relationship with node m_j in a network with N nodes, given E (the mismatch between node m_i and its relationship with m_j) and the update rate $\eta(t)$ is:

$$m_i(t+1) = m_i(t) - \eta(t) \cdot E_{m_i, m_j}(t), \quad E_{m_i, m_j}(t) = m_i(t) - f_{m_i, m_j}(t) \quad (1)$$

The convergence speed is determined by the update rate $\eta(t)$. In every iteration the update rate takes into account the external information from the sensor or the other node respectively, and the network's internal belief, to modulate the influence of that sensor or node:

$$\eta_{i,j}(t+1) = \eta_{i,j}(0) \cdot \frac{\sum_{k=1}^{N-\{i\}} E_{m_i, m_k}(t)}{N \cdot E_{m_i, m_j}(t)} \quad (2)$$

Eq. 2 shows that the update rate $\eta(t)$ adjusts itself proportionally to the mismatch from the network's belief (numerator) and penalizes the influence from a sensor or node if the values it provides conflicts with the network's belief.

2.3 Sensor Fusion Network

For our heading estimation scenario the four nodes in the network represent four sources of information providing heading measurements. We consider inertial heading (from gyroscope yaw axis, "SG"), magnetic heading (from compass, "SC"), odometry heading (computed from wheel encoder information, "SW") and vision heading (from an on-board camera tracking a distal marker on the ceiling, "ST"); see Fig. 1, gray outside boxes and Fig. 2 robot setup. Each of the sensors provides raw data, which is preprocessed to align to a common unit (e.g. by integration, shift/offset). After the preprocessing stage all sensors are connected to their respective network node and data flows into the network. Each node in the network (G, C, T, W) maintains an independent heading estimate. The network computes a global heading estimate by balancing its internal belief and new sensor contributions. Every convergence step each node's information is updated as given in eq. 1. For example, the gyroscope representation update rules for the given network topology are:

$$G(t+1) = (1 - \eta_{G,C}(t)) \cdot G(t) + \eta_{G,C}(t) \cdot C(t), \quad (3)$$

$$G(t+1) = (1 - \eta_{G,W}(t)) \cdot G(t) + \eta_{G,W}(t) \cdot W(t), \quad (4)$$

$$G(t+1) = (1 - \eta_{G,T}(t)) \cdot G(t) + \eta_{G,T}(t) \cdot T(t), \quad (5)$$

$$G(t+1) = (1 - \eta_{G,SG}(t)) \cdot G(t) + \eta_{G,SG}(t) \cdot SG(t), \quad (6)$$

The network architecture used in our scenario is depicted in Fig. 1.

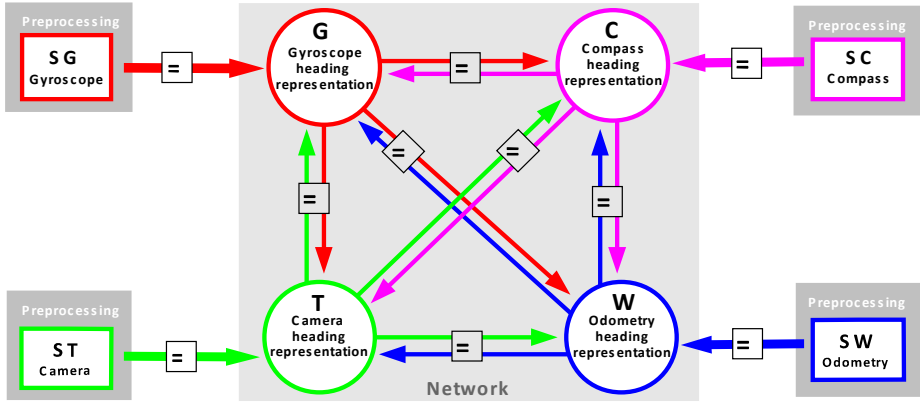


Fig. 1. The network architecture used in the 4 dimensional sensor fusion scenario

One can easily observe the recurrent flow of information within the network and from sensors into the network. The chosen structure allows balancing contributions from each sensor within the network, resulting in a global consistent representation.

2.4 Comparison with State-of-the-Art

Relating our method with state-of-the-art probabilistic sensory fusion mechanisms we address three key aspects: complexity, flexibility and robustness (Table 1).

Table 1. Comparison between state-of-the-art and our proposed model for sensor fusion

Criteria	State-of-the-art approaches (Bayesian)	Proposed model
Complexity computational costs	large number of probabilities to apply probabilistic inference [2].	compute multiple simple update rules (similar to eq. 1).
Flexibility possibility to add further sensory modalities	requires parameters adjustments for additional sensory modalities; adding sensors improves performance but increases complexity [3].	sensor addition (adding more update rules/constraints) is straightforward and without complexity increase.
Robustness handling sensor failures, conflicts, and uncertainty	dedicated means to detect failures, not generally applicable; challenges in assigning probabilities in an uncertain context [2].	abnormal sensor activity can be detected and penalized by adapting η (e.g. the influence of that sensor in the global estimate).

3 Model Evaluation

We designed a test case scenario to evaluate the proposed model. This simple scenario serves as proof-of-concept; we expect to extend the system within the same framework to other (more complex!) scenarios. A customized omnidirectional mobile robot (Fig. 2) equipped with inertial measurement unit (IMU, gyroscope and compass), wheel encoders, and a vision sensor provides a minimalistic system for heading estimation.

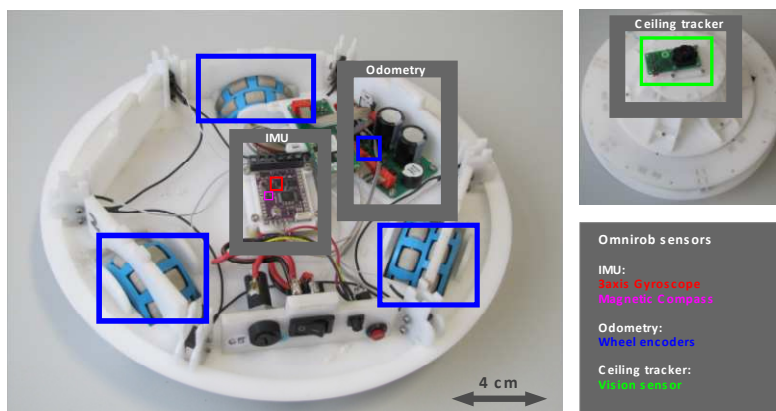


Fig. 2. Test setup for the 4 dimensional sensor fusion network: omnidirectional mobile robot

Each sensor alone is unable to provide precise heading angle measurements, as they all are affected by noise and systematic errors. The main challenge is computing a global estimate of heading direction given all unreliable sources of information. The test scenario is depicted in Fig. 3, left: the robot follows a simple predefined rectangular trajectory from start to target (top down view). Both, raw sensor data and inferred network data, are shown in Fig. 3, right.

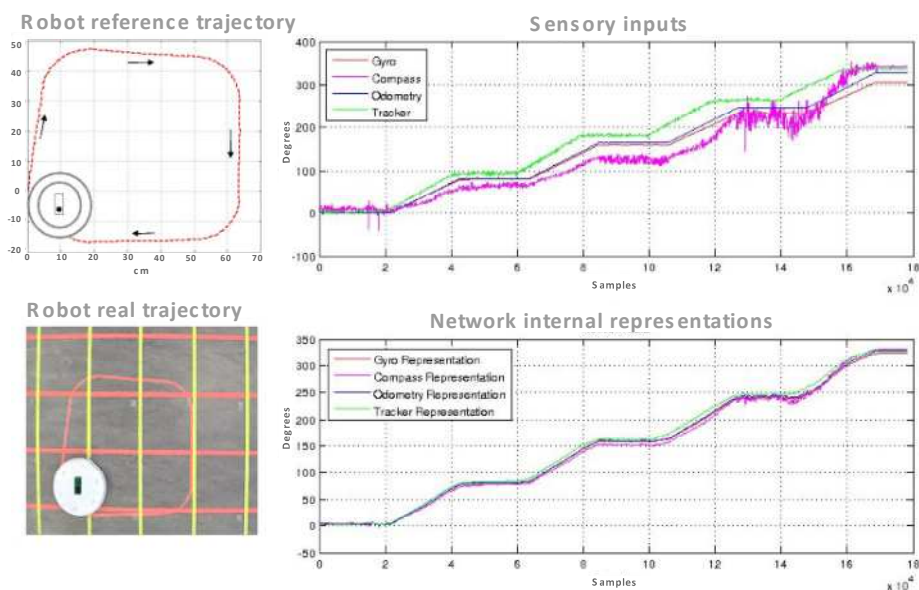


Fig. 3. Robot trajectory, raw sensors data and inferred network representations

During operation the sensors' raw data was preprocessed to align the sensors' coordinate systems (square boxes in Fig. 1). As shown in the upper right diagram of

Fig. 3 data from different sensors describe the robot’s heading, but each sensor data shows a different error. The preprocessed sensor data is fed into the network, which tries to balance each sensor’s contribution, as we see in the lower right diagram in Fig. 3. The network “pulls” the represented values of the nodes towards a common heading, thereby compensating for drifts and inaccuracies in individual sensor readings. The update rate is adapted (refer to eq. 2), such that “plausible” values are contributing stronger to the global estimate, while “implausible” values are penalized. Fig. 4 displays the adaptation process of the update rate in detail: to allow relaxation of the network, each sensor input data was presented for 100 network iterations. Each node in the network is fully connected to all other nodes and to an individual sensor.

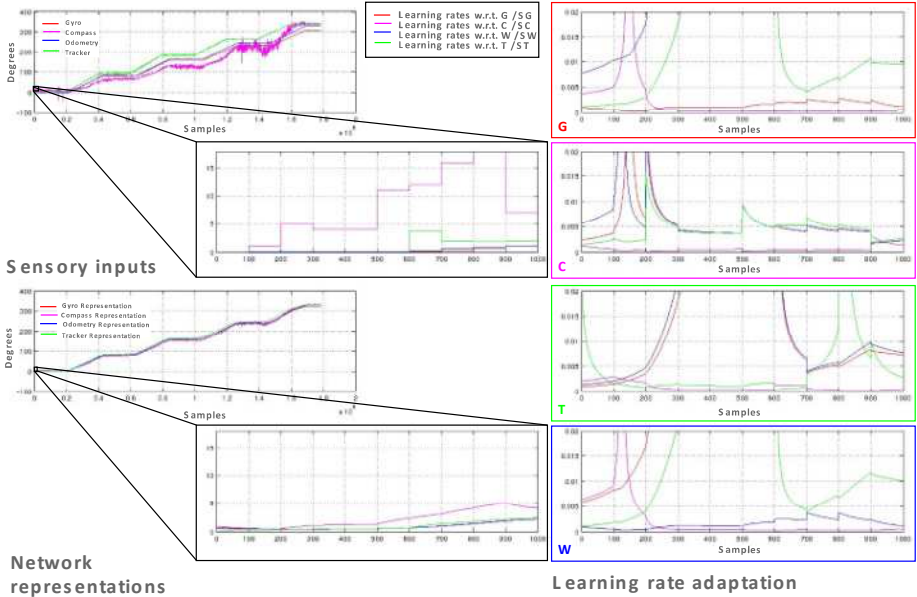


Fig. 4. Network analysis: inputs, inferred representations and learning rate adaptation

The network is able to infer which sources of information to trust by considering the mismatch (Fig. 4, right side, panels G, C, T, and W) between each local representation in a node and current representation in other nodes or its sensor. The network continuously re-evaluates and balances contributions from each sensor, so that the different representations in the network stay consistent with each other. As there is no global ground truth data source, and each of the sensors might be affected by noise and systematic errors, the network attempts to settle in a solution in which each relationship in the network is satisfied as good as possible.

4 Conclusions and Future Work

By taking inspiration from cortical distributed processing principles the presented implementation for real-world sensor fusion shows to be an alternative to state-of-the-art

methods for sensor fusion, with advantages in terms of complexity, flexibility and robustness. To improve the network architecture we are currently investigating temporal relationships between different nodes in the network, allowing to remove a preprocessing step and instead feeding raw sensor data directly into the network. Such preprocessing (e.g. integration, differentiation of sensor signals) will then happen inside the network as dual relations between two representations: one node representing raw data and a further node representing the derived quantity. We will extend the network to support sensors of different types (such as pure visual input in relation to ego motion as shown here), which requires more complex relations and the representation of multidimensional data. Finally, we envision learning the network topology and thereby the underlying relations between represented quantities based on consistent real-world sensory input obtained from mobile robot exploration tasks.

Acknowledgments. The authors would like to thank Matthew Cook of INI, ETH/University Zürich for intense and fruitful discussions about map based processing and networks of relations.

References

1. Arleo, A., Rondi-Reig, L.: [Multimodal sensory integration and concurrent navigation strategies for spatial cognition in real and artificial organisms. J. Integrative Neuroscience 6\(3\), 327–366 \(2007\)](#)
2. Siciliano, B., Khatib, O. (eds.): [Springer Handbook of Robotics. Springer, Berlin \(2008\)](#)
3. Hall, D.L., Llinas, J.: [An Introduction to Multisensor Data Fusion. Proc. of the IEEE 85\(1\), 6–23 \(1997\)](#)
4. Griffiths, T.L., Yuille, A.L.: [A primer on probabilistic inference. Trends in Cognitive Sciences 10\(7\) \(2006\)](#)
5. Körding, K.P., et al.: [Causal Inference in Multisensory Perception. PLoS ONE \(2007\)](#)
6. Thrun, S., Burgard, W., Fox, D.: [Probabilistic Robotics. MIT Press \(2005\)](#)
7. Bressler, S.L.: [Understanding Cognition Through Large-Scale Cortical Networks. Current Directions in Psychological Science 11\(2\), 58 \(2002\)](#)
8. Swindale, N.V.: [How different Feature Spaces may be Represented in Cortical Maps. Network: Computation in Neural Systems 15 \(2005\)](#)
9. Sporns, O.: [Networks of the Brain. MIT Press \(2011\)](#)
10. Cook, M., Gugelmann, L., Jug, F., Krautz, C., Steger, A.: [Interacting maps for fast visual interpretation. In: Proc. of International Joint Conference on Neural Networks, pp. 770–776 \(2011\)](#)
11. Michler, F., Eckhorn, R., et al.: [Using Spatiotemporal Correlations to Learn Topographic Maps for Invariant Object Recognition. J. of Neurophysiology 102, 955–964 \(2009\)](#)
12. Carreira-Perpinan, M.A., Lister, R.J., et al.: [A Computational Model for the Development of Multiple Maps in Primary Visual Cortex. J. Cereb. Cortex 15, 1222–1233 \(2005\)](#)
13. Averbeck, B.B., Latham, P.E., Pouget, A.: [Neural correlations, population coding and computation. Nature Review Neuroscience \(7\), 358–366 \(2006\)](#)