Robotics and Autonomous Systems 71 (2015) 69–82

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Cortically inspired sensor fusion network for mobile robot egomotion estimation

Cristian Axenie *, Jörg Conradt

Neuroscientific System Theory Group, Department of Electrical Engineering and Information Technology, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany

HIGHLIGHTS

- A cortically inspired sensor fusion network for robotic applications is introduced.
- Distributed graphical network, with combined feed-forward and recurrent connectivity.
- No global supervisor, only local processing, storage and exchange of information.
- Given sensory data, network relaxes into the best explanation of an underlying cause.
- Extensible to learn sensor correlations and adapt connectivity from incoming data.

ARTICLE INFO

Article history:
Available online 31 December 2014

Keywords:
Sensor fusion
Cortically inspired network
Egomotion estimation
Mobile robots

ABSTRACT

All physical systems must reliably extract information from their noisy and partially observable environment and build an internal representation of space to orient their behaviour. Precise egomotion estimation is important to keep external (i.e. environmental information) and internal (i.e. proprioception) cues coherent. The constructed representation subsequently defines the space of possible actions. Due to the multimodal nature of incoming streams of sensory information, egomotion estimation is a challenging sensor fusion problem. In this paper we present a distributed cortically inspired processing scheme for sensor fusion, which given various sensory inputs, and simple relations defining inter-sensory dependencies, relaxes into a solution which provides a plausible interpretation of the perceived environment. The proposed model has been implemented for egomotion estimation on an autonomous mobile robot. We demonstrate that the model provides a precise estimate of both robot position and orientation.

1. Introduction

An essential component in motor planning and navigation, for both real and artificial organisms, is egomotion estimation. Egomotion or self-motion refers to the combined rotational and translational displacement of a perceiver with respect to the environment. During motion organisms build their spatial knowledge and behaviours by continuously refining their internal belief about the environment and own state [1–3]. Our approach is motivated by three main aspects consistent with recent results in spatial processing for navigation and perception [4].

The first aspect addresses the importance of maintaining a precise position of the self. Building an internal representation of the environment and own state implies the coherent alignment of the acquired sensory cues. As sensory cues are conveyed from both dynamic egomotion related signals such as odometry and inertial signals, and static external environmental signals, such as visual or auditory, the precise position of the self links and keeps the representation coherent. In this context a coherent representation provides the ability to recognise and define "action possibilities" from all available sensory cues (e.g. distance to objects). Subsequently, egomotion defines the space of possible actions and impacts behaviour [2,3,5].

A second aspect refers to the capability of a real or artificial organism to understand space itself from its own state (in space). Egomotion estimation contributes to the understanding of high-level features of the environment, like structure and layout, such that the organism can direct actions and control its movement. Typically, with respect to position, the primary question is related to distances to key objects in the environment. In order to infer correct distances, the organism must traverse the environment and

* Corresponding author. Tel.: +49 89 289 26908.
E-mail addresses: cristian.axenie@tum.com (C. Axenie), conradt@tum.de (J. Conradt).
URL: http://www.nst.ei.tum.de (C. Axenie).

http://dx.doi.org/10.1016/j.robot.2014.11.019
0921-8890/

distinguish between its dynamic and static features as they lead to different consequences [6].

The third aspect points directly to the solution offered in this paper, namely how can precise egomotion perception be obtained given the complex multisensory environment. In order to handle environmental variability and complexity, continuous and simultaneous incoming sensory data streams from different sensors must be combined into a robust representation. But sensory cues are usually complementary and redundant and it is not clear how they describe the spatio-temporal properties of the environment. To disambiguate the complex scenario the global representation should combine all cues in an informative and plausible way. This combination process, termed sensor fusion, is not trivial, as current implementations [7–9] show. The primary objective is aligning reference systems of the different congruent and redundant sensory cues. After alignment, depending on inferred spatio-temporal correlations, interference and conflicts between the cues need to be minimised [2,3,10]. Finally, sensor fusion should not propagate biases or errors in the final (fused) estimate but compensate for them.

The paper is organised as follows. After a review of existing sensor fusion mechanisms and the motivation for the bioinspired paradigm shift in Section 2, we introduce our model in Section 3. Starting from the general neurally inspired processing model, we present the architecture and the specific instantiation for the mobile robot egomotion estimation. Section 4 provides the analysis and evaluation of our model and a comparison with state-of-the-art methods, whereas Section 5 provides a thorough discussion of the obtained experimental results. Finally, Section 6 concludes the work, by summarising the main features and advantages of the model and introducing future extensions.

2. Review of sensor fusion algorithms for egomotion estimation

Currently developed engineered approaches for sensor fusion typically aim at optimal solutions, and many results in robotic applications demonstrate this [7–9]. Real-world scenarios, however, are typically dynamic, prone to parameters changes and characterised by complex features. To cope with such aspects, typical approaches often need parameter tuning or model refinements. Their dedicated structure cannot handle the variability of the percepts or accommodate different scenarios. These algorithms cannot easily handle different contexts from those considered in the design. When considering adaptivity and robustness, neurobiology offers vastly superior performance over today's engineered systems. With different processing paradigms and distributed representations our brain solves the task of combining sensory information not only more effectively, but seemingly without much effort. Extremely flexible, the brain can easily accommodate and handle, by learning, new and different tasks and situations.

Having set up the framework, in the upcoming sections we first provide an overview of state-of-the art sensor fusion approaches and their instantiations using specific computational methods. Second, we mark the advantages of the paradigm shift towards bioinspired computational approaches. Supported by examples in robotic cognitive systems we introduce the main features of neural systems which transferred to technical systems will provide a higher degree of flexibility and robustness.

2.1. Standard computational methods for sensor fusion

Most state-of-the-art sensor fusion algorithms are based on probabilistic models of observations and processes. This framework has become attractive for both engineering and computational neuroscience as a powerful tool to describe sensory models and dynamics (in engineering) [7–9] and also inference and sensory integration in populations of neurons (in computational neuroscience) [11–15].

These algorithms use Bayes' rule to integrate observations and system's model into a unified estimate of the system's state. In addition, these methods replace point representations of perceptual estimates with probability distributions such that the statistics of the sensory estimates can be quantified and used for inference. Bayesian methods provide an optimal estimation scheme, in the sense that their estimate of the given state is unbiased (i.e. difference between this estimator's expected value and the true value of the state being estimated is zero/estimation is true on average) and has minimum variance [7–9]. In Bayesian inference belief is progressively updated as new data from the sensors is presented such that the initial belief evolves towards an informed posterior distribution. In many simple cases it is possible to analytically describe the posterior distribution. This is the case where the prior and likelihood function are given by Gaussian normal distributions [7–9,16]. However, in many real-world scenarios this is not the case [16,15].

As an alternative approach to probabilistic approaches other non-Bayesian approaches have been developed. Providing a strong framework to describe uncertainty by using the notion of partial membership, fuzzy logic, is a powerful tool for imprecise reasoning [8]. Although fuzzy theory is particularly useful to represent and fuse information provided by human experts, it is limited merely to fusion of vague data [17]. More often this method is integrated with probabilistic fusion algorithms [18]. Using a qualitatively different reasoning technique,the Dempster–Shafer theory of evidence, fuses information relying on probability mass to characterise belief and plausibilities in the data. Despite the ability to fuse uncertain and ambiguous data, evidential reasoning is inefficient in fusing highly conflicting data and has scaling problems in the case of high-dimensional state spaces [8,19]. Even though a large variety of methods to fuse sensory data were developed, they are usually combined to improve performance in more general scenarios.

Independent of the underlying computational method in use, current sensor fusion algorithms' implementations are described by a global and sequential processing scheme. Dictated by current computing architectures this paradigm constrains algorithms to obey to a pipelined sequence of filters and other feed-forward processing stages. Neuroscience studies postulate that neural processing is described by distributed and unsupervised computation mechanisms, with mixed feed-forward and recurrent flow of information and local storage and processing capabilities [11–14]. These mechanisms can adapt to novelty, by learning, and exhibit robustness in the face of uncertainty. Transferring these principles to technical systems will support this paradigm shift towards more flexible and adaptive systems.

2.2. Bioinspired approach to sensor fusion: changing paradigm

Sensor fusion is a process that influences major aspects of perception, cognition and behaviour in both physical and artificial systems [1,7,8,20]. Traditionally sensor fusion describes a mechanism to combine cues from the same or different modalities, converted to a common, internal representation which is subsequently used in the actual fusion process [21,22]. It is commonly agreed that the brain contains areas specialised for processing different types of information incoming from sensors [23–25]. A major determinant for a brain area's ability to process a certain type of information is the input it receives [23,26]. It is considered that the unique processing characteristic of each cortical area is defined in terms of the area's interactions with the other areas [27,23,26,12–14]. Hypotheses from cortical interareal coordination studies support evidence that these areas aim to reach a consensus and maintain mutually consistent information with the others resolving coherence or incoherence relations (i.e. constraints) [26,28,12]. Furthermore, neurobiological evidence supports the view that elementary sensory representation and functions are localised in discrete areas [11–14], whereas complex functions are processed in parallel in widespread networks [27,23,29,25,26].

Although the neural correlates for sensor fusion are complex and diverse there are some global principles that govern their information processing mechanisms [30,31,29]. Using both feed-forward and recurrent connections sensor data streams are processed and information is exchanged between discrete areas which are specialised to handle one sensory modality. As there is no supervisor to mediate the processing flow, coordination between these interacting sensory areas only depends on the dynamics of interareal coordination defined by the anatomical pathways linking the areas [23,29,32,33].

As real-world scenarios are typically dynamic, this processing framework is able to flexibly accommodate, the variability and uncertainty in the scene. Behavioural and psychophysical studies in human crossmodal phenomena provide a quantification of superior human performance in tasks where various neural substrates are involved to solve perceptual tasks in real time [34].

Robustness, is another important feature of the neural mechanisms for sensor fusion. It was generally believed that the brain builds up passively an internal representation while noisy sensory data is available and "idles" when there is no incoming stream of sensory data [35]. Evidence supports the current view that the brain is permanently active and generates predictions about incoming sensory inputs using previously learned spatio-temporal structure of the noisy stimuli [36,37,14].

Using a distributed processing paradigm, capable of learning and robustly adapting to under-constrained or conflicting sensory inputs, the brain can maintain and continuously refine its internal belief about the environment.

The underlying neural substrate for sensor fusion provided already inspiration for successful robotics implementations. Combining optic flow and inertial data for motion estimation in UAVs using a parallel Reichardt motion detection algorithm on FPGAs [38]; combining visual and inertial information in a model of the superior colliculus modulated by active attention and behavioural exploration for motion estimation [39,40]; and even a vestibulo-ocular reflex fusion model for camera stabilisation in driver assistance systems [41], are only some examples which support the usage of the aforementioned principles in real-world sensor fusion problems.

3. A cortically inspired network for egomotion estimation

The architecture and main computational features of our model are based on work in [27]. The model in [27] uses a distributed network in which independent neural computing nodes obtain and represent sensory information, while processing and exchanging exclusively local data, to infer an estimate of robot orientation. The scenario in [27] is now extended to full egomotion estimation. The generic processing framework in this paper is inspired by the neural processing paradigm introduced in Section 2.2, where cortical areas involved in sensory processing assume rapid resolution of a large number of mutually imposed constraints (i.e. coherence/incoherence relations), leading to a globally coherent estimate of the percept. Following similar high level cortical organisation and interaction principles with our model the work in [39,40] introduce and evaluate the capabilities of a neuromimetic Bayesian framework for multimodal sensor fusion for motion estimation and 3D structure extraction, motivating once again the advantages of the paradigm shift towards biological inspiration.

3.1. General processing model

The proposed model is a network of processing units whose connectivity is provided by relations defined between the units

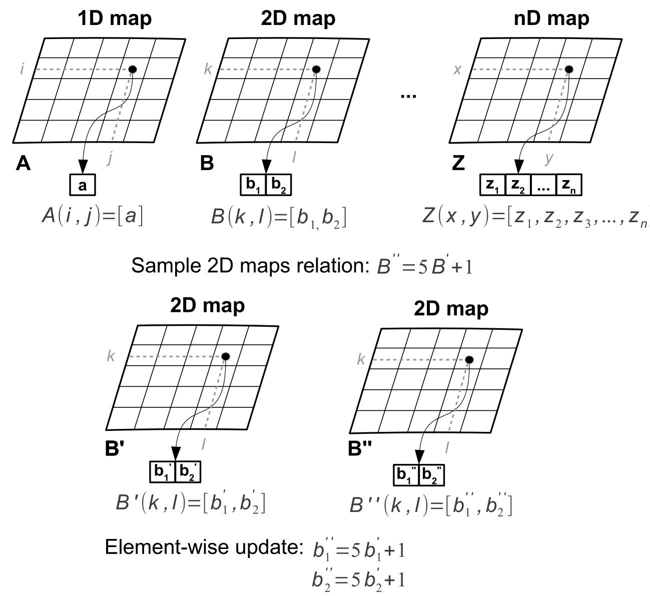

Fig. 1. Generic map based representation used in our model. Dimension is determined by the size of feature space a map represents. Sample implementation of an algebraic relation between two 2D maps: $B'' = 5B' + 1$.

(e.g. given by physics of the sensors or the inter-sensor interactions). The relations between the units can be explicitly encoded in the network or obtained as a result of a learning process. Even though in the current stage relations are embedded in the network at design time, the dynamics and integration capabilities of the network are the same for both hand-crafted and learned relations. Hence, our model separates relaxation dynamics (convergence towards a solution) from learning (connectivity setup).

Obeying constraints imposed by relations, each unit processes, stores and exchanges only local information, to the extent that each unit builds and refines its local belief about the represented sensory modality. Furthermore, both feed-forward and feedback processing pathways connect the units such that the mutual exchange of information is kept to a consistent state (i.e. fulfilled relations). A similar principle was successfully used for fast visual interpretation in [42]. The proposed network in [42] interprets input from a neuromorphic vision sensor by means of recurrently interconnected areas, each of which encodes a different aspect of the visual interpretation, such as light intensity, optic flow and camera calibration. This network of interacting maps is able to maintain its interpretation of the visual scene in real time. Using a similar representation, each unit in our network contains a map based representation. The maps are inspired by the topographic organisation in cortical and midbrain structures for sensor fusion [31,29,32,24,25], and share the same functional role of mapping the stimuli distribution to an internal representation. A similar approach has been also used in [39,40] modelling the superior colliculus and dorsal pathway and employing iterative Bayesian programming to refine spatial maps (i.e. probability distributions).

In our approach the dimension of a map is determined by the feature space it represents (e.g. 1D angular velocity, 2D optic flow, 3D rotation vector). To get a feeling about the map based representation and the way relationships are linking maps, we provide a toy example in Fig. 1.

The encoded quantity in each cell of a map can be represented by point estimates or using a sparse representation, encoded in neural population activity [42,33]. In a more general view, a map encodes the representation of a continuous stimulus parameter by a place-coded population response, whose peak reflects the mapped parameter [16]. Independent of representation (point estimate/population code) the core of the model is implementing

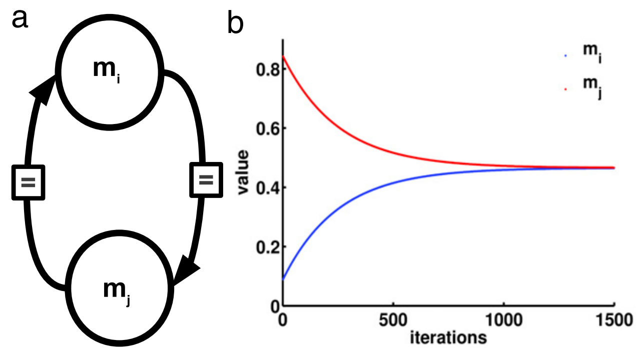

Fig. 2. Canonical network for identity. (a) Mutual influence between two units, m_i and m_j, obeying update rules to minimise the mismatch between the values stored in each unit's map; (b) Starting from initial random values the maps converge towards fulfilling the identity relation.

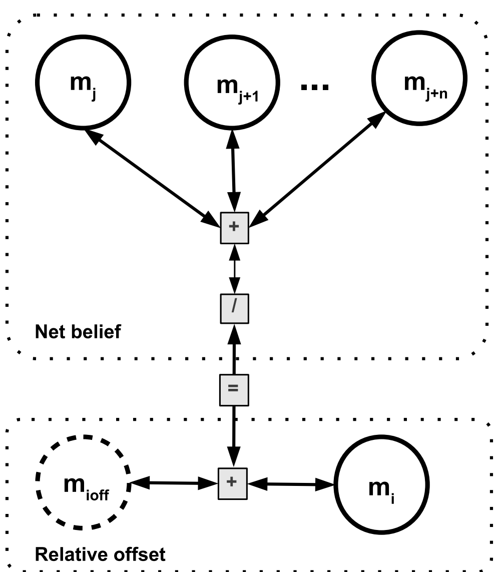

Fig. 3. Network for generic algebraic relations. Implementing summation, division, or difference using the proposed model is straightforward. Particular implementation of offset computation network.

relations that describe the connectivity of the network. Fig. 2 introduces a canonical network that implements the identity relation between two units.

Network dynamics is based on a random update process, using gradient descent. Values in each unit's map take small steps towards minimising the mismatch with the relations in which the unit is involved. In the case of absence of external sensory input (as shown in Fig. 2), given the random initialisation of each unit's map, the network will converge to a solution in which the relations are fulfilled. Each unit follows update rules given by

$$\Delta m_i(t) = -\eta_{i,j}(t)\frac{\partial E_{m_i,m_j}(t)}{\partial m_i(t)} \tag{1}$$

$$E_{m_i,m_j}(t) = (m_i(t) - m_j(t))^2. \tag{2}$$

Eq. (1) provides the update rule for map m_i. To minimise the mismatch with respect to m_j, given by E_{mi}, $m_j(t)$, the map takes a step proportional to the mismatch, modulated by a factor $\eta_{i,j}(t)$. The mismatch (error signal) computation in Eq. (2), is based on the squared error between the two units.

Applying relatively simple operations upon the locally stored estimate, each unit balances the influence from all the other units. Units can be linked using generic algebraic relations (e.g. summation, division, difference, or product), which can be used to implement diverse and more complex relations. The advantage is that these "atomic" operations are simple, keeping local processing fast enough to support fast network dynamics. A sample network to implement averaging is introduced in Fig. 3. This network is used extensively in our model as a core mechanism to quantify the relative mismatch (offset) between the quantities inferred in the network. The network computes the average activity of all other connected units, m_j, in the network, stores it in *net*, and isolates the contribution of a certain unit, m_i, to compute its relative offset, m_{ioff}. The offset is inferred in the network, in a separate unit, and each main unit has an associated offset node following the next generic update rules

$$net(t) = \frac{\sum_{k=0}^{n} m_{j+k}(t)}{n+1} \tag{3}$$

$$\Delta m_{ioff}(t) = -\eta_{m_i,net}(t)\frac{\partial E_{m_{ioff},net}(t)}{\partial m_{ioff}(t)} \tag{4}$$

$$E_{m_{ioff},net}(t) = (m_{ioff}(t) - (net(t) - m_i(t)))^2. \tag{5}$$

During operation, the network brings all quantities in agreement by satisfying all relations. The amplitude of the update step that each unit takes towards the correct value is computed on-line, on a per map basis, modulated by a confidence factor, η. This factor accounts as a measure of reliability of a certain source of incoming information into a unit. Using this mechanism the network is able to penalise strongly conflicting sources of information (a mechanism that improves fault tolerance) and enhance the contribution of consistent sources. Each contribution to a unit, m_i from another unit, for example m_j, in a network with n units, is modulated by the confidence factor, adapted using

$$\Delta\eta_{m_i,m_j}(t) = \eta_{m_i,m_j}(0)\frac{\sum_{k=1,k\neq j}^{n} E_{m_i,m_k}(t)}{(n-1)E_{m_i,m_j}(t)}. \tag{6}$$

Assuming that all n units in a network should contain the same value, Eq. (6), computes the confidence factor, η, for map m_i when receiving influence from map m_j, by comparing the expected mismatch (i.e. average error of m_i with respect to the network), $\frac{\sum_{k=1,k\neq j}^{n} E_{m_i,m_k}(t)}{(n-1)}$, with the error between m_i and m_j, $E_{m_i,m_j}(t)$.

Although many possible relations can be implemented using only the basic algebraic operations, to handle sensory data we also need relations which encode temporal dependencies. One example is temporal integration and temporal differentiation. As the network should combine contributions from different sources, they should be aligned and represent the same type of values (i.e. rate of change or absolute values). Our model implements temporal integration, locally, using two units, one which maintains the (persistent) absolute value, and one which provides the rate update. This persistence mechanism is necessary to keep a coherent absolute value in the presence of inputs from other units. A sample integration network used in our model is depicted in Fig. 4.

The preprocessing unit, m^{pp}, which maintains a persistent value in unit m_i, obeys the same update rules with other units, the only difference is just its limited connectivity. The update rules for m_i in the canonical integration network depicted in Fig. 4 are given by

$$\Delta m_i(t) = -\eta_{m_i,m^{pp}}(t)\frac{\partial E_{m_i,m^{pp}}(t)}{\partial m_i(t)} \tag{7}$$

$$E_{m_i,m^{pp}}(t) = (m_i(t) - \int_0^t m^{pp}(t)\mathrm{d}t)^2. \tag{8}$$

Following the generic update rule for units in the network, Eq. (7) computes the new value of m_i by using the mismatch from the relation with m^{pp}, depicted in Eq. (8). One can also use an error signal given by $E_{m_i,m^{pp}}(t) = m_i(t) - \int_0^t m^{pp}(t)\mathrm{d}t$ with differences

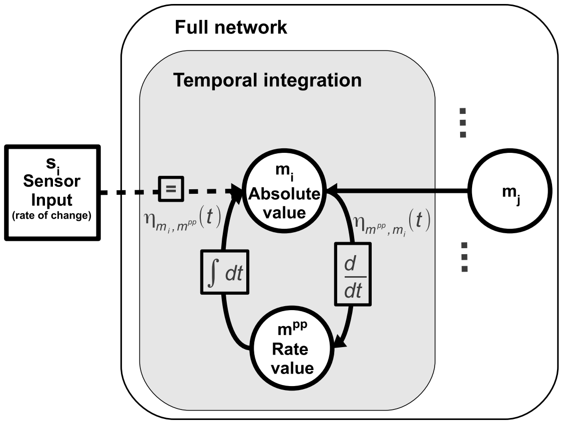

Fig. 4. Canonical network for temporal integration. The local integration network maintains a persistent accumulated value of sensor data in m_i. In the absence of sensory data, the preprocessing unit, m^{pp} ensures the proper rate update, avoiding large updates induced by other connected units, m_j.

in the magnitude of the update. The value in m_i converges to the accumulated absolute value of the sensory input s_i. As m_i is connected to m_j, there is an influence towards m_i from m_j unit, and the amplitude of the update, given in Eq. (9) is equal to the mismatch between the value in m_i and the value in m_j, as shown in Eq. (10).

$$\Delta m_i(t) = -\eta_{m_i,m_j}(t)\frac{\partial E_{m_i,m_j}(t)}{\partial m_i(t)} \tag{9}$$

$$E_{m_i,m_j}(t) = (m_i(t) - m_j(t))^2. \tag{10}$$

The differentiation map, m^{pp}, is updated using a single set of update rules (shown in Eqs. (11) and (12)) due to the unique connection to m_i. The m^{pp} map converges to the rate of change of s_i, given by $\frac{d}{dt}m_i(t)$, and provides a persistent contribution to m_i.

$$\Delta m^{pp}(t) = -\eta_{m^{pp},m_i}(t)\frac{\partial E_{m^{pp},m_i}(t)}{\partial m^{pp}(t)} \tag{11}$$

$$E_{m^{pp},m_i}(t) = (m^{pp}(t) - \frac{d}{dt}m_i(t))^2. \tag{12}$$

As mentioned in Section 2.2 one main feature of the neural substrate for sensor fusion is robustness. Our model exhibits robustness at the unit level through the confidence factor. The confidence factor provides a fault tolerance mechanism which detects errors in incoming noisy raw sensory streams. By adapting the confidence factor according to the measured mismatch (from the expected value and the current value) the network penalises conflicting sources and enhances congruent sources.

Using these canonical circuits as building blocks we instantiate our model for egomotion estimation. Taking advantage of relatively simple processing stages at unit level the network can rapidly relax in a solution, fulfilling all relations (constraints) between the fused quantities.

3.2. Sensor fusion network for mobile robot egomotion estimation

We tested our model in a real environment using an omnidirectional mobile robot depicted in Fig. 5. In the basic scenario the robot moves in an uncluttered environment while an overhead camera tracking system keeps track of its position and orientation. The robot is equipped with an inertial measurement unit, consisting of 3-axis gyroscope, 3-axis magnetometer which acts as vestibular input; wheel encoders acting as proprioceptive input; motor driver providing an efferent copy of the PWM signal; and a camera for visual input. Raw sensor data is fed to the network which updates its internal belief and infers an estimate of robot's position and orientation.

The main architecture of our network for egomotion estimation is depicted in Fig. 6. There is no explicit input or output in/from the network and sensor data just mildly influences the activity in the network. Based on the embedded relations the network is able, in the absence of one or more sensors, to infer the missing quantities.

In order to properly visualise the network connectivity we split the embedded functionality with respect to the task, namely position and orientation estimation. Most connections within the network are bidirectional and elicit mutual influence between the units linked by relations. The only unidirectional connections are the ones coming from the sensors, as the network cannot influence sensory readings. In order to quantify the performance of our model, we also added two readout units which provide an average of the estimated quantities. These units obey the same update rules and dynamics like all other inferred quantities in the network.

3.3. Experimental setup

The egomotion network consists of two main interacting components, one for orientation estimation, and the other for position estimation. We first introduce the heading estimation network, depicted in Fig. 7. This network is comprised of:

- 4 main units (G, C, W_h, V_h) encoding representations of modalities' heading estimates (gyroscope, compass (magnetometer), wheels encoders and camera) functionally similar to neural visual-somatosensory-vestibular integration models in [43–45];
- 4 preprocessing (pp) units (G^{pp}, C^{pp}, W_h^{pp}, V_h^{pp}) which transform raw sensor data performing offset compensation (for vision and magnetometer) or temporal integration (for gyroscope and encoders odometry);
- 4 heading angle offset (ho) units (G_o, C_o, W_{ho}, V_{ho}) which quantify sensors bias or drift;
- 1 global readout unit, H, which provides an average of the inferred quantities.

After internally preprocessing raw sensory data each main unit stores an estimate of absolute heading angle and tries to be consistent with the values in the other units. As the network infers multiple estimates of heading angle from different sources, it ensures that all yield the same value. In fact the fusion process assumes that all complementary modalities are combined yielding a more precise estimate. As each modality provides its own estimate of absolute angle, the network combines all contributions judiciously. Mutual influence between units is modulated by the confidence factor. Each interaction pathway of a unit has an associated confidence factor which adapts according to the level of trustworthiness of a source of information to which the unit is connected. Hence, the network benefits of a mechanism to detect and compensate for faults and abrupt changes in sensor data.

The distributed local representations inferred in network's nodes are integrated in the readout node, which provides an average of network inferred quantities, and can be used to quantify performance. The readout node contribution to each of the main units is proportional to the contribution of that specific unit in the global estimate, modulated by the confidence factor. The processing steps behind the global readout node are depicted in the lower right panel of Fig. 7.

The other component of the egomotion network, dedicated to position estimation, computes a global estimate of robot's position in the 2D plane as well as the travelled path. Using an integration scheme functionally similar to mechanisms presented in [5,46], the network, shown in Fig. 8, is composed of:

- 3 main units encoding representations of different modalities' 2D position (p) estimates (V_p — camera estimate, W_p — encoders estimate, M — position from efferent copy of motor command estimate);
- 3 position offset (po) units (M_o, W_{po}, V_{po}) encoding the relative mismatch between the inferred quantities;

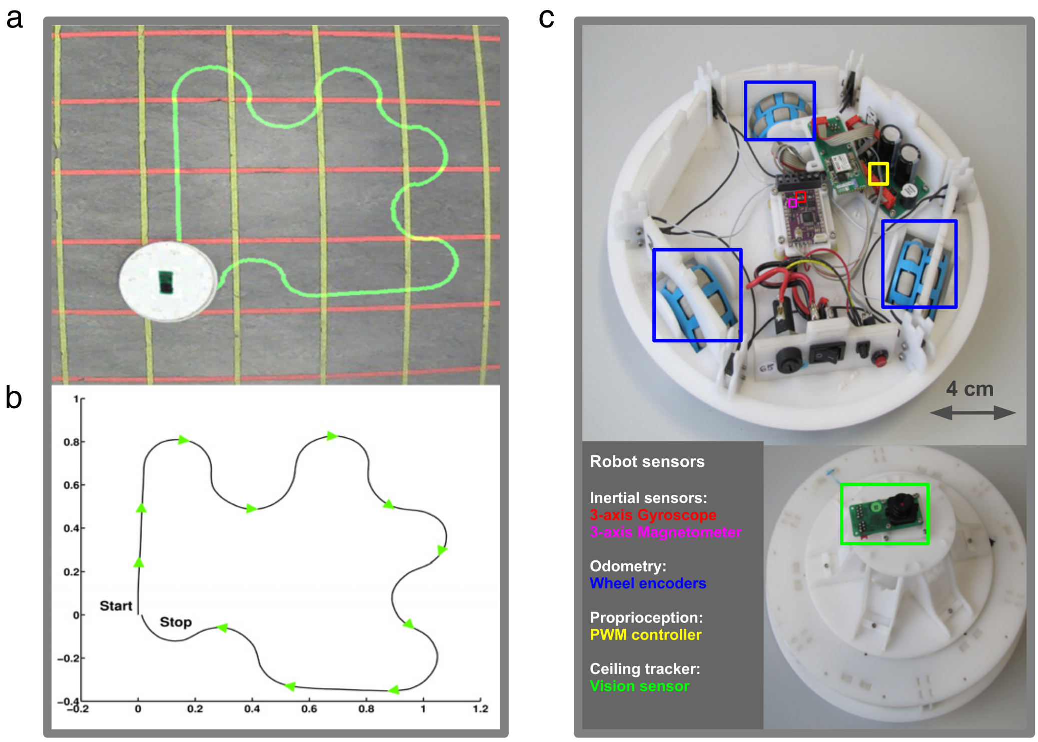

Fig. 5. Robot architecture and experimental setup: mobile robot and test trajectory. (a) Overhead tracker trajectory; (b) Robot reference trajectory; (c) Mobile robot sensors.

- 3 sensory preprocessing (pp) units (M^{pp}, W_p^{pp}, V_p^{pp}) which are tightly coupled with the main units and perform temporal integration or simple transformations from robot reference frame to world reference frame;
- 1 global readout node, P, providing an average of the network belief about robot's position;
- 1 global readout node, P_i, providing the travelled path, by accumulating changes in P unit.

The position estimation network in Fig. 8 infers a 2-dimensional estimate of position, given as (x, y) coordinates, and a 1-dimensional estimate of travelled path, using sensor data, the heading network estimate and inter-sensory relations.

The internal link between the two sub-networks is based on the transformation between robot reference system and world reference system. Computing the rotation matrix necessary to perform the coordinate transformation assumes that the absolute heading angle is known. Hence, the heading angle estimate from the heading network is fed into each of the main units of the position estimation network such that the quantities are re-encoded in the world reference frame. This coordinate transformation is necessary to evaluate the performance of the proposed model in a (understandable) world centred reference system representation.

At the current stage we performed our experiments in a relatively simple, uncluttered environment. Yet imagining a more complex scenario is straightforward as long as there are additional cues to measure and (if needed) internally build a map of occupancy. As a concrete idea, depth or ultrasonic sensors can be added to provide distance to objects. In this scenario the distance to objects will be computed in the egocentric reference frame of the robot, and will provide another cue which the network will fuse with the computed travelled distance (i.e. from the odometry, vision and efferent motor copy). This way the network can infer a more precise travelled distance given also the occupancy information of the environment.

4. Experimental results

We analysed the behaviour of the heading estimation network for the complex trajectory depicted in Fig. 9(a). After being randomly initialised in a stable state in which all relations are fulfilled, the heading estimation network receives raw sensory data samples from gyroscope, compass, wheel encoders and camera. In order to align the sensory data, the network preprocesses the raw samples to obtain an absolute heading angle, depicted in Fig. 9(d). We observe that the inferred absolute angle values are not perfectly matching. This is due to the fact that real sensors are unable to provide precise angle measurements due to noise and errors. The gyroscope is affected by drift which accumulates over time, the magnetometer is affected by disturbing magnetic fields in the environment or the robot's motors, odometry is affected by non-systematic errors (e.g. wheel slippage, imprecise sizes of the wheels), and the camera tracking is affected by low accuracy, changing illumination and strongly depends on the robot's velocity. Fig. 9(b) highlights the mismatch between the computed heading angle from sensory data.

When sensory data is continuously fed in, the network accommodates new observations by updating its own belief about the current state. At the low level, each unit modulates the influence from the units/sensors connected with, such that only consistent data is used for updating its state. The network combines all contributions and enforces that all the quantities are close to the same value, Fig. 9(c, f).

The confidence factor adaptation is presented in Fig. 9(g–j), on a per map basis. For each inferred main map (G, C, W_h, V_h) the confidence factor is computed as previously shown in Eq. (6), such that each source of information contributing to a unit's update is compared against the other sources. Depending on the mismatch, the confidence factor is adjusted proportionally. Fig. 9(c) depicts the inferred values for the G, C, W_h, V_h, and one can notice that the network brings all the quantities in agreement (fulfil the relations) and changes in confidence factor support that. For example, if we consider unit G between $t_1 = 60$ s and $t_2 = 64$ s, one can see that the confidence factor with respect to V_h is high (saturated to a maximum imposed value) and all the others are low. This behaviour is supported by the graph in Fig. 9(c), where we see that between t_1 and t_2, G and V_h values are overlapping while after t_1 maps store

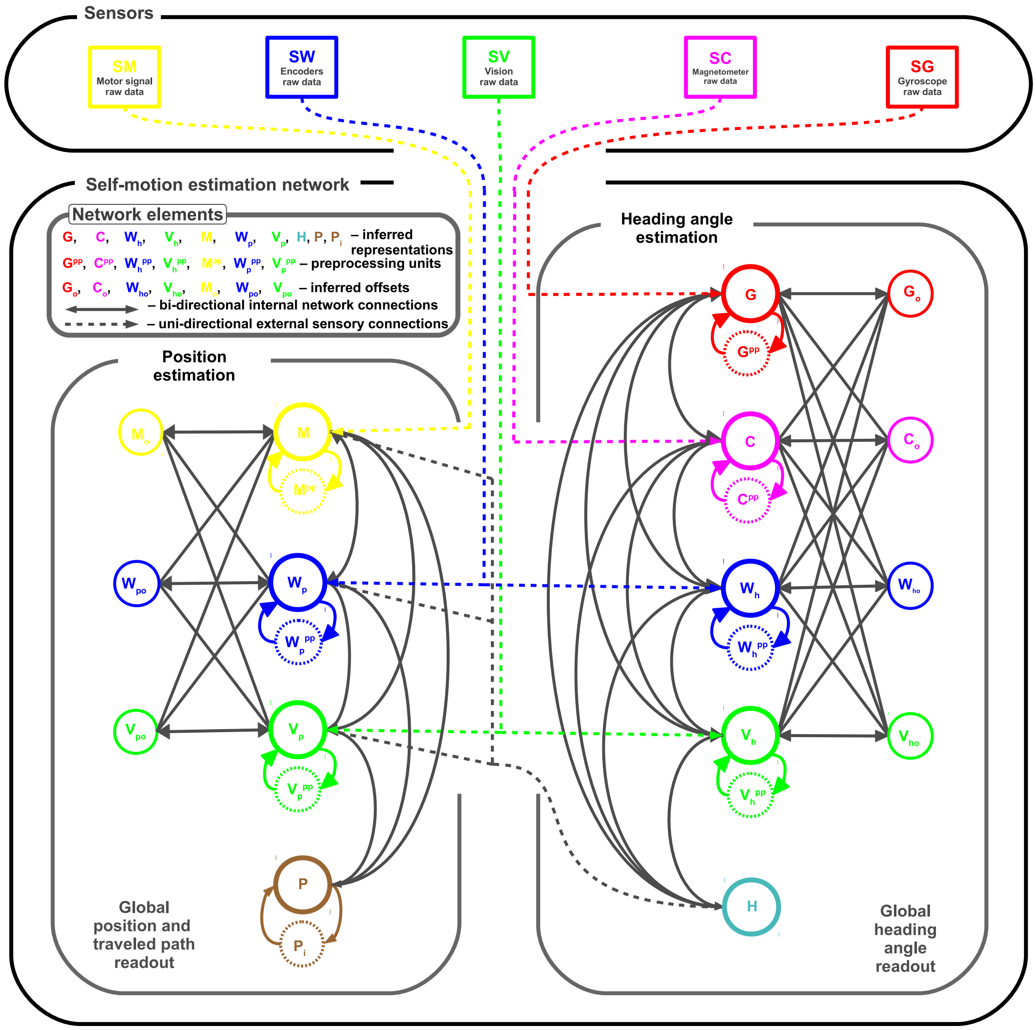

Fig. 6. Network architecture for egomotion estimation. Distributed network composed of interconnected sub-networks for heading and position estimation. All connections are bidirectional except those coming from the sensors. All-to-all connections within each network obeying same update dynamics.

slightly different values. An additional quantity inferred by the network, useful in detecting and compensating sensor errors or biases, is the offset, depicted in Fig. 9(e). Each main map in the network has an associated offset node, which will store a relative mismatch with respect to the other units. This quantity is used to provide an additional source of consistent data, that will use only knowledge inferred internally in the network and contribute to a unit's update. Given the sensory data the network finds a solution which fulfils all the embedded relations, and keeps all inferred quantities in agreement, as shown in Fig. 9(f).

The second component of our model is the position estimation network. Similar to the heading estimation network, raw data from wheel encoders, a copy of the PWM signal and vision data, are presented to the network. Subsequently, the network preprocesses raw data to obtain an estimate of 2D position. The position network operation is depicted in Fig. 10. As one can observe in Fig. 10(g), no modality estimate is able to provide a precise position estimate. Despite strongly conflicting estimates, the network brings all quantities in agreement, such that a more precise global estimate is inferred. Fig. 10(h) depicts the network data. Following the same principle with the heading angle estimation network, each map penalises contradictory sources of information (providing a low confidence factor) and enhance contribution from consistent sources (high value of confidence factor). The confidence factor adaptation is depicted Fig. 10(a–f), for both X and Y position estimation. In order to get a feeling on how each modality main map (V_{p}, W_{p}, M) is updated under the influence of the sensory data and local network belief, we can analyse the inferred offsets in Fig. 10(i–k) as they quantify a relative mismatch between the units. One can observe a mismatch in the inferred quantities, despite the judicious network processing, which emerges due to the fact that each sensor has a different response time and the network cannot influence (by prediction and correction) the sensory readings explicitly.

In order to measure the performance of our model we compare it with two state-of-the-art methods. A Kalman filter and a Maximum Likelihood Estimator (MLE) were implemented. The data fed into the two models is equivalent to the output of the preprocessing stage of the network (i.e. absolute angle measurements). Hence, for heading estimation both Kalman filter and MLE receive four sources of absolute angle. Using sensory observations, each model updates the modelled system state representation such that we can directly read out an estimate of heading angle, similar to typical implementations [8,9].

As one can see in Fig. 11(a,b), the network response is slower than both Kalman filter and MLE approaches. This is due to the preprocessing stage (temporal integration and offset compensation) which is performed in the network while new sensory observations are received. This complete approach is different from standard methods which need already preprocessed data for the estimation process. In Fig. 11(c) it can be noticed that the network overestimates or underestimates the values from the other estimators when significant changes in orientation take place. In the

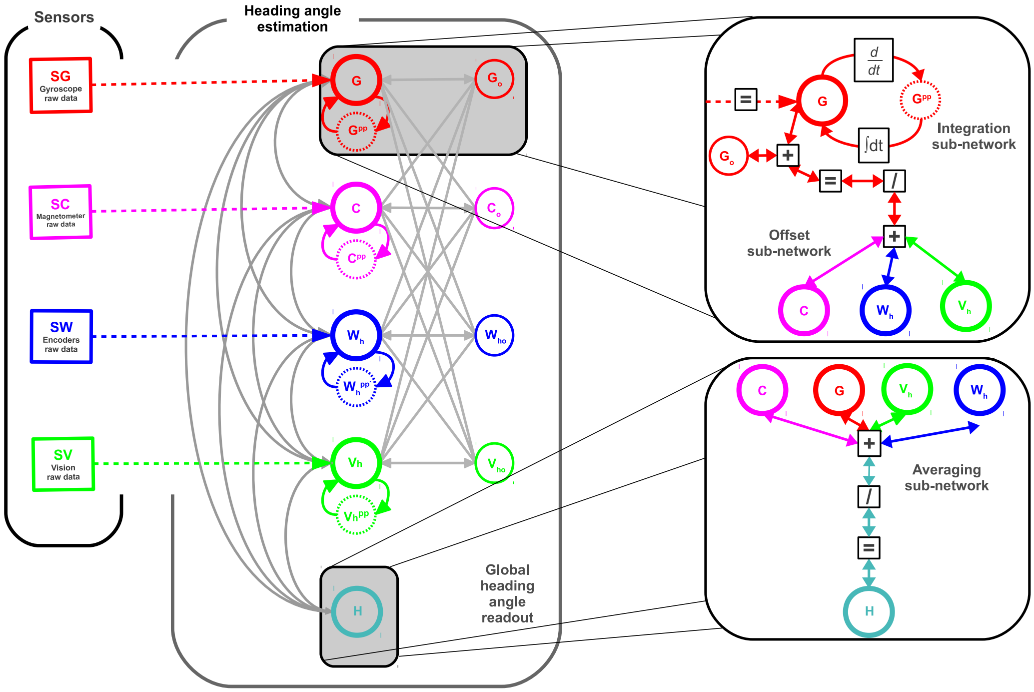

Fig. 7. Heading estimation network. Sensory data from gyroscope (SG), compass (SC), wheels encoders (SW) and camera (SV), flows in the network mildly influencing its activity. Local networks implement preprocessing, using G^{pp}, C^{pp}, W_h^{pp}, V_h^{pp} units, while all-to-all connections between main units determine interactions yielding a more precise fused estimate.

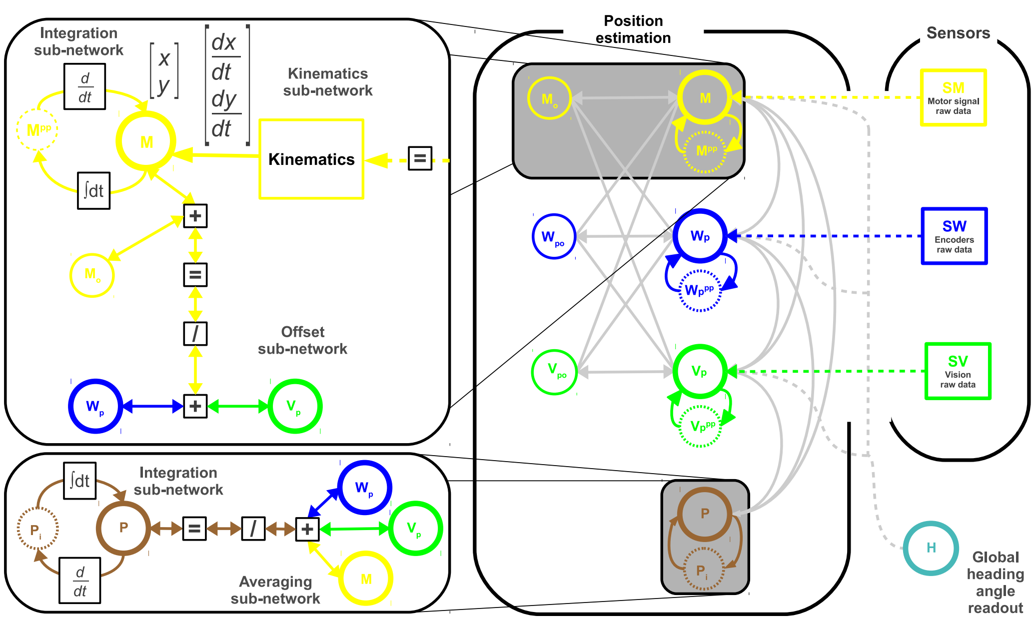

Fig. 8. Position and travelled distance estimation network. Sensors provide raw data to the network. Local circuits preprocess raw data by using algebraic or temporal (differentiation) relations to maintain a position estimate in each main unit.

considered scenario the robot often changed direction with different angles and the network needed time to accommodate the changes and balance different sensors' contributions. As sensors react differently to changes, the network identifies which contributions are consistent and accommodates them as new observations are available. Albeit the network does its best in balancing the represented quantities autonomously, it can also accommodate externally imposed constraints (i.e. user can set "preferred value") to exhibit slower responses with higher accuracy or faster responses with lower accuracy. In order to quantify the network performance, the Root Mean Squared Error (RMSE) was calculated against Kalman filter and MLE estimates with respect to ground truth data. In our scenario a smaller RMSE value describes a better estimation. Given available sensory data, our network estimate is comparable with state-of-the-art estimators with $<10\%$ RMSE, as shown in Fig. 11(c, d).

For position estimation, both Kalman filter and MLE receive three sources of (x, y) position, inferred from a copy of PWM motor command, wheel encoders data and vision data. The network infers a global estimate comparable with state-of-the-art

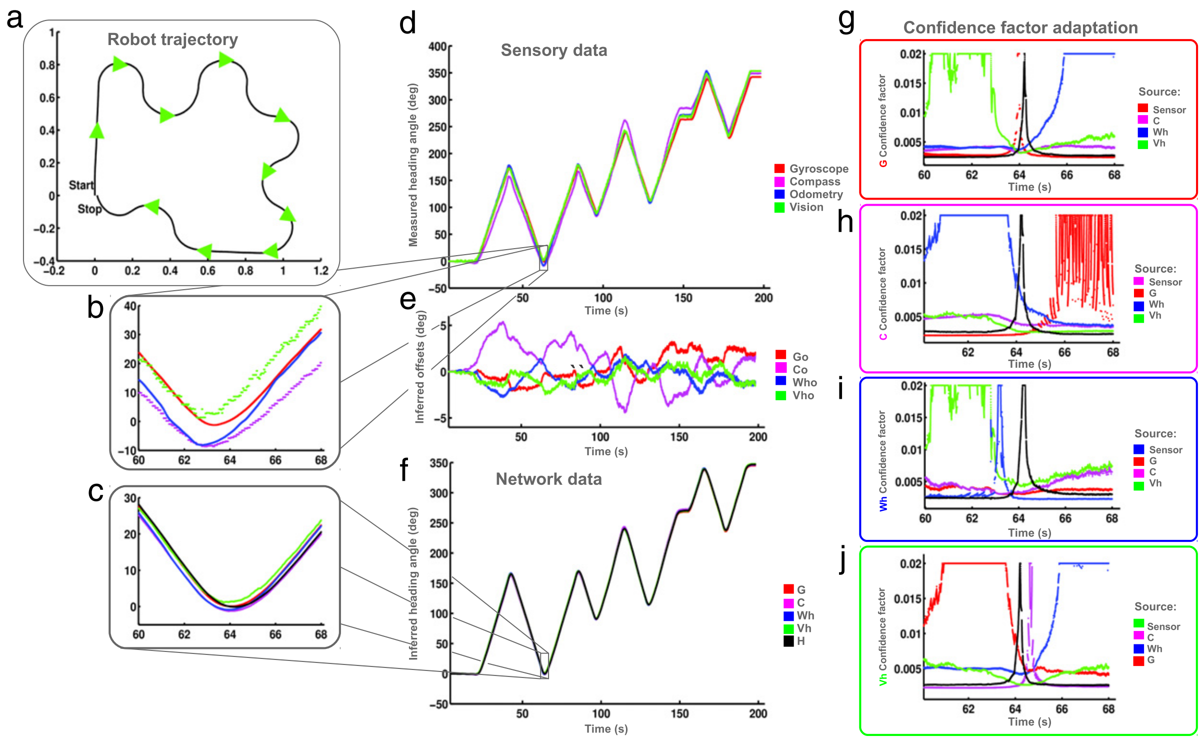

Fig. 9. Heading estimation network dynamics. (a) Robot trajectory; (b), (d) Measured heading angle from sensors; (c), (f) Inferred network quantities; (e) Inferred offset values; (g), (h), (i), (j) Confidence factor adapts according to the mismatch between local and incoming information in a unit. When the network relaxes, inferred quantities fulfil the relations given external sensory data.

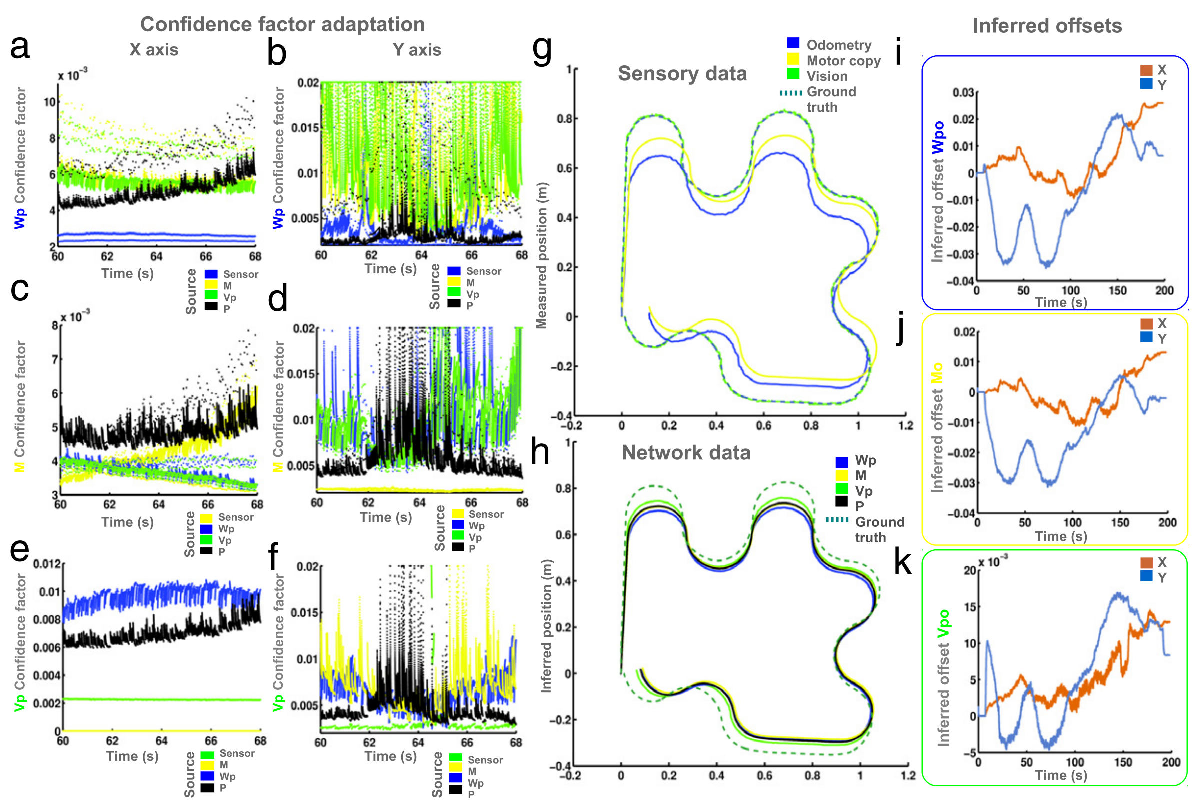

Fig. 10. Position estimation network dynamics. (a), (c), (e) Confidence factor adapts to minimise errors between the inferred X position estimates in the network; (b), (d), (f) Confidence factor adapts to minimise errors between the inferred Y position estimates in the network; (g) Comparison between measured position from sensors and ground truth data; (h) Network inferred quantities and ground truth data; (i), (j), (k) Inferred offsets for each map.

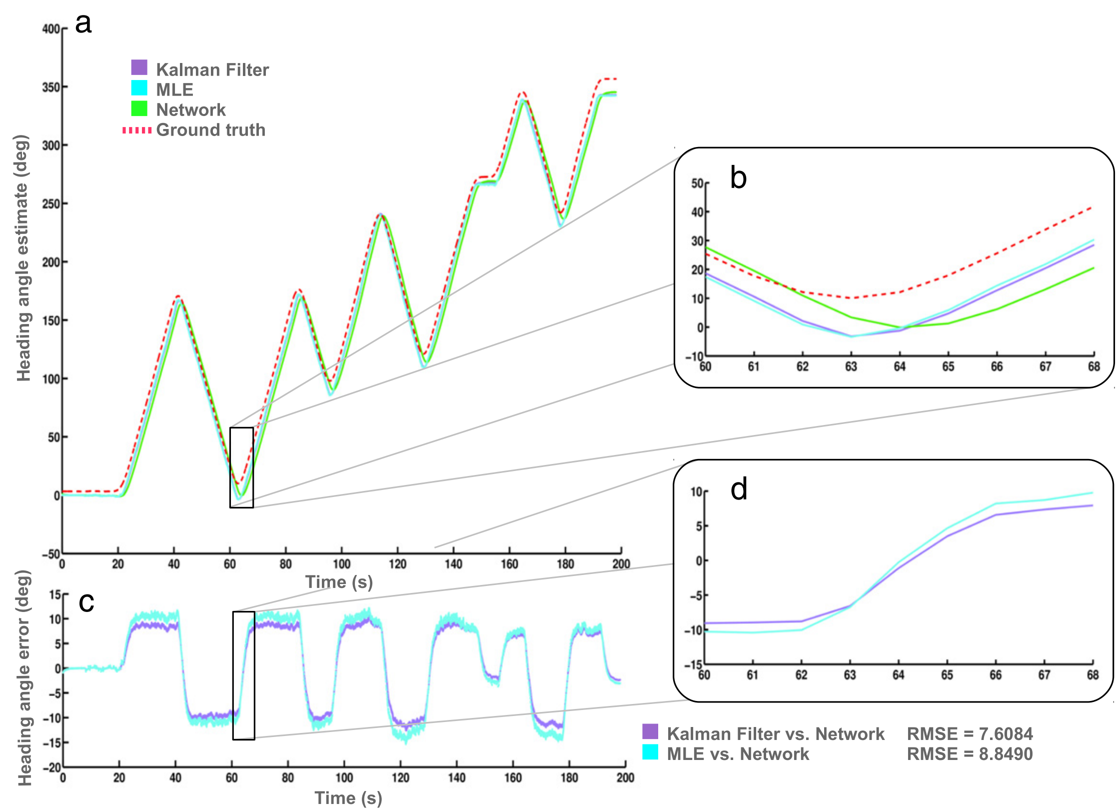

Fig. 11. Heading estimation evaluation. (a), (b) Comparison between network estimate, Kalman filter estimate, MLE estimate, and ground truth data; (c), (d) Heading angle estimation error and RMSE values.

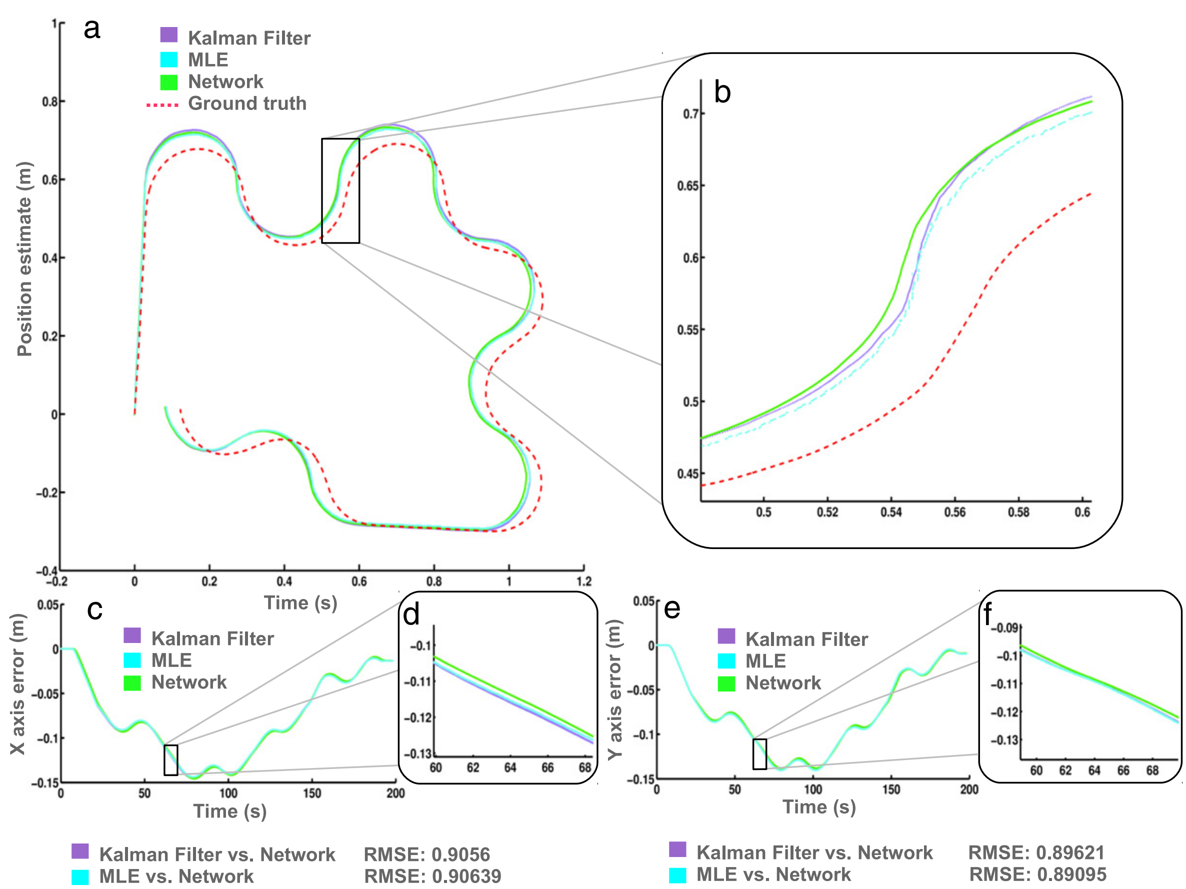

Fig. 12. Position estimation evaluation. (a) Comparison between network estimate, Kalman filter estimate, MLE estimate, and ground truth data. (c), (d) X axis position estimation errors for each estimator and relative RMSE values; (e), (f) Y axis position estimation errors for each estimator and relative RMSE values.

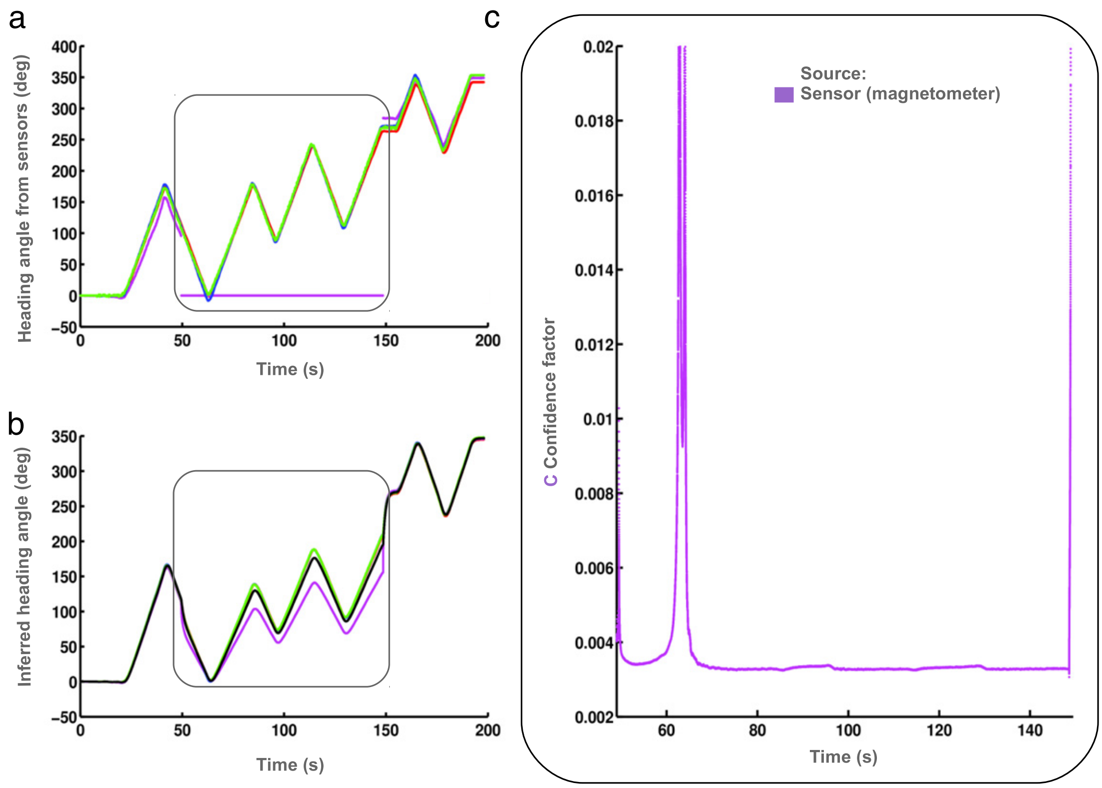

Fig. 13. Fault tolerance analysis. (a) Heading angle values from sensors. The magnetometer is faulty between $t1 = 50$ s and $t2 = 150$ s. (b) Inferred heading angle values in the network. Even if the sensor (i.e. magnetometer) is faulty the network compensates for that, and infers its representation from the other modalities. (c) Confidence factor adaptation w.r.t magnetometer.

Table 1
Inferred travelled distances from individual modalities.

Network unit	Inferred travelled path (m)
Vision, V_p	4.212
Odometry, W_p	4.160
Motor efference, M	4.177
Global average, P_i	4.183
Ground truth	4.190

estimators given the available sensory data, as shown in Fig. 12(a, b). Fig. 12(c, e) show that our network is comparable to both Kalman filter and MLE as the measured RMSE values for position estimation are smaller than 1%. Furthermore, one can see in the decoupled analysis on each axis shown in Fig. 12(d, f) that the network is close to estimates of the two state-of-the-art methods.

Another quantity inferred by the network is the travelled path, P_i, computed as an accumulation of all intermediate position estimates, P. The P_i unit integrates the average position provided by the global position estimate, P. Furthermore, to quantify the precision of the main maps in the network (V_p, W_p, M) we computed the corresponding travelled path values as given by the integration of the successive position estimates in each of the maps. The computed values are given in Table 1 and one can see that the values are consistent (with an error smaller than 3 cm).

In order to assess the fault tolerance capabilities of our network we performed an additional set of experiments in which we tested the network in the presence of faulty sensors. In order to do that, we clamped the sensor (i.e. magnetometer) readings to 0 for a certain amount of time during operation. After some time we re-activated the readings simulating just a temporary failure. Fig. 13(a) illustrates the sensory data and the faulty transient for the magnetometer while in Fig. 13(b) we can see that even if the sensor is not useable anymore the network does its best to infer its representation from the available sensors. This is fully supported by the dynamics of the confidence factor of map C (encoding the magnetometer representation for heading angle) with respect to the incoming sensory data. The confidence factor decreases immediately as the sensor is faulty at $t_1 = 50$ s and is restored to a high value once the readings are consistent with the network belief at $t_2 = 150$ s. One interesting aspect to mention is that during the faulty transient the confidence factor changes dramatically (around $t_3 = 65$ s) when as one can see in Fig. 13(b) the sensor is consistent with the other sensors and network belief. This experiment was meant to quantify the robustness against faulty sensors of our network and provide an insight on the simple yet efficient mechanism behind it.

The current section presented results supporting that our network is able to infer a position estimate from each modality, and combine them into a global estimate more precise than individual modalities.

5. Discussion

In the present paper we introduce a cortically inspired model for sensor fusion which uses principles of neural processing to combine contributions from different sensors and infer an estimate of both position and orientation of a mobile robot. It is unanimously accepted that an organism's possible actions and movements are conditioned by the environment. Egomotion estimation contributes actively in shaping this space of possibilities.

While moving, both real and artificial organisms receive a constant flow of information from parallel sensory channels, bind and compare the stimuli with previous experience and goals, and produce motor outputs to match the current circumstances. Yet combining sensory contributions is not a trivial task, sensory contributions must be aligned in a common reference frame depending on their spatio-temporal properties, and the resulting representation should be plausible and informative to disambiguate the scenario.

State-of-the-art methods for sensor fusion provide good results for dedicated scenarios but lack the generality and fail to accommodate different contexts from the ones considered at design time. The complexity of real-world scenarios goes over the prepared environment of the lab, and the impact of complex environments on sensor fusion is likely to become a major issue, as models become more and more sophisticated [47].

Alternatively, there is evidence that our brain is able to combine different information streams from available senses and use the combined representation in a flexible manner to robustly orient behaviour. Albeit a large number of putative models which were developed it seems that biological systems tend to combine not only exteroceptive, and interoceptive cues, but also psychological and cognitive cues when integrating senses for orienting behaviour [20].

In order to build and maintain a precise representation of the self in space sensory cues conveyed from egomotion must be combined to precisely orient behaviour. Psychophysical studies in human spatial cognition hypothesise that behaviour can arise from perception. A unifying theory introduced in [3] provides a possible mechanism that recognises and/or creates spatial identity or similarity between various sensory experiences (e.g. kinaesthetic, visual) to enhance cognition about the environment. Following this hypothesis, our model combines contributions from different sensors for estimating robot's position and orientation. For heading angle estimation the model enforces identity between the individual absolute angle estimates computed from the raw gyroscope, magnetic compass, wheel encoders and vision sensor data depicted in Fig. 9(d). Fusing the different sensors' contributions yields a more precise estimate than individual estimates, because the integration process compensates for sensor errors and noise in individual measurements, as shown in Fig. 9(c, f). In order to estimate position, the model receives wheel velocities from the wheel encoders, 2D position from vision and a PWM motor command copy. The raw data is preprocessed by the network which infers three different sources of robot absolute position as shown in Fig. 10(g). These individual estimates are kept in agreement by the network which computes also an average estimate from all contributions as shown in Fig. 10(h). The global average is a quantification of network's belief and can be used in planning more precise motor commands.

Egomotion estimation provides the organism the capability to understand the environment from its own state. Evidence in motion psychophysics and kinaesthesis [5] enforce the hypothesis that humans build their perceptions and conceptions of space when they learn their bodies and move based on some innate kinetic dynamics. Furthermore, they develop more complex notions of space (e.g. connectivity, distances to objects, occlusions, objectification, [5]) useful in conceiving themselves to spatial bounds and layout. The current instantiation of our model is able to estimate both egomotion components, position and orientation, from available sensory data. Derived quantities (i.e. do not have an associated sensor) can be also inferred by combining other quantities in the network. For example, each position estimate provided by the network is accumulated in a global travelled distance unit (i.e. P_i) and can be used to compute distances to objects in the environment. In a sample use case, the first step is to combine global (P) and camera (V_p) positions estimates, such that the network determines which areas of the environment were already traversed by simply matching the positions from the two sources. In the second step, the travelled distance (P_i) provides the absolute distance to occupied areas of the environment detected in the first step. To support higher level representations or derived quantities the network accommodates new sensors by simply defining additional connections. This is beneficial as the network can be easily extended such that complementary sensors can be fused to yield an occupancy map which along with the self-motion cues can provide a complete description of the environment (e.g. SLAM — Simultaneous Localisation and Mapping).

The perception of the environment and body orientation are influenced by multiple sensory and motor systems [10]. In order to handle the variability and complexity of the environment, available sensory are combined such that interference and conflicts between the individual measurements are minimised and a more precise estimate is obtained. For technical systems, in order to build such a representation the system designer must precisely describe (a) the system model, (b) the prior information about the sensory observations and the system, and (c) the preprocessing steps as shown in [8,9]. In order to relieve the system designer from the difficult task of describing the aforementioned aspects, our model simplifies the representation and fusion mechanism by using a different processing paradigm inspired by cortical computation principles. Basic mathematical relations link different processing units which use feed-forward and feedback connections to exchange information as shown in Fig. 6. The network tries to keep all quantities stored in the units in agreement given noisy sensory data, Fig. 9(b, d). Despite the fact that each source of information is affected by noise or systematic errors, the network is able to detect abnormal changes in sensory data, such that there is a small impact over its internal belief. Preprocessing is performed inside the network, such that sensory contributions are aligned to a common representation, without increasing the network complexity. The preprocessed data flows into the network and each unit balances contributions from all its connections. An adaptive mechanism (i.e. confidence factor) modulates the influence of external information sources, to penalise strongly conflicting estimates and enhance consistent values, as shown in Fig. 9(g–j) and Fig. 10(a–f). Moreover, the network uses relatively simple dynamics for unit update such that it converges rapidly to a solution, shown in Fig. 9(c,f) and Fig. 10(h), given the constraints imposed by the embedded relations and sensory data. As sensory data mildly influences the network activity, in the absence of one sensory modality the network can recover the missing quantity based on the other modalities and the connectivity. Relevant experimental results are illustrated in Fig. 13(b). This inference capability accounts for a fault tolerance mechanism. Assuming that temporarily a sensor does not provide any measurements, its value will be continuously inferred by the network such that when it will become online its impact will be modulated by the network belief and progressively accommodated in the network, as shown in Fig. 13(b, c). Continuously refining its own belief given the available sensory data, the network provides an estimate which is comparable with state-of-the-art methods, as shown in Fig. 11(c,d) and Fig. 12(c–f).

Although for the current scenario the network relations were hard-coded by the designer, we expect to extend this model such that relations emerge from learned correlations in sensory data. The network will no longer need a predefined structure as the incoming stream of sensory data will shape its connectivity given the cross-modal interaction patterns. These extensions are inspired by the underlying neural circuits for sensor fusion in Cortex and their experience based development and plasticity. Another intuitive extension focuses on self-organising-maps learning and specialisation principles, through competition and cooperation. This extension can be accommodated in the existing structure as representation and processing capabilities of the units can be modified without altering the network level dynamics.

The proposed processing scheme provides many advantages, in terms of implementation and complexity, being able to distribute computation, evaluate and balance contributions of the fused sensory data, while using only relatively simple operations representing cross-sensory relations.

6. Conclusions

Given noisy and sometimes conflicting sensory data, sensor fusion, is crucial for precise egomotion estimation. Our model introduces a new approach for sensor fusion. Using a cortically inspired processing paradigm our model provides results comparable with optimal state-of-the-art methods. Without precise modelling and parametrisation of the system model, our network is able to combine information from multiple sensors into a global estimate, more precise than individual estimates. Balancing external sensory contributions with its internal belief, the network is able to detect and compensate for sensor inconsistencies. By distributing computation such that each unit processes and stores only local information, using only basic mathematical relations, complexity is reduced. The current instantiation of the model for egomotion estimation provides comparable results with state-of-the-art methods in terms of estimate accuracy, but with less design challenges. Finally, our network is highly parallelizable, making it suitable for implementations on massively parallel hardware architectures for real-time robotics applications. Given its generality, computational efficacy, and ease of implementation our model is a promising candidate for sensor fusion in robotic applications.

Acknowledgements

The authors would like to thank Matthew Cook of INI, ETH/University Zurich for intense and fruitful discussions about map based processing and networks of relations, Marcello Mullas of NST, TU Munich for discussions and comments on writing the manuscript, and the Elite Network of Bavaria for supporting this research.

References

[1] A. Arleo, L. Rondi-Reig, Multimodal sensory integration and concurrent navigation strategies for spatial cognition in real and artificial organisms, J. Integr. Neurosci. 3 (2007) 327–366.
[2] T. Heed, B. Roeder, The body in a multisensory world, in: M.M. Murray, M.T. Wallace (Eds.), The Neural Bases of Multisensory Processes, CRC Press, Boca Raton, 2012, pp. 557–582.
[3] R.W. Mitchel, Understanding the body: spatial perception and spatial cognition, in: F.L. Dolins, R.W. Mitchell (Eds.), Spatial Cognition, Spatial Perception: Mapping the Self and Space, Cambridge University Press, Cambridge, 2010, pp. 341–364.
[4] F.W. Mast, L. Jancke (Eds.), Spatial Processing in Navigation, Imagery and Perception, Springer, Berlin, 2007.
[5] M. Sheets-Johnstone, Movement: the generative source of spatial perception and cognition, in: F.L. Dolins, R.W. Mitchell (Eds.), Spatial Cognition, Spatial Perception: Mapping the Self and Space, Cambridge University Press, Cambridge, 2010, pp. 323–340.
[6] T. Warren, Preliminary questions for the study of egomotion, in: R. Warren, A.H. Wertheim (Eds.), Perception & Control of Self-motion, Hillsdale, 1990, pp. 3–33.
[7] J.F. Ferreira, J. Dias, Probabilistic Approaches for Robotic Perception, in: Springer Tracts in Advanced Robotics (STAR), vol. 91, 2014.
[8] H. Durrant-Whyte, T.C. Henderson, Multisensor data fusion, in: B. Siciliano, O. Khatib (Eds.), in: Springer Handbook of Robotics, Springer, Berlin, 2008, pp. 585–608.
[9] S. Thrun, W. Burgard, D. Fox, Probabilistic Robotics, MIT Press, Cambridge, MA, 2005.
[10] J.R. Lackner, P. Dizio, Multisensory influences on orientation and movement control, in: G. Calvert, C. Spence, B.E. Stein (Eds.), The Handbook of Multisensory Processes, MIT Press, Cambridge, Massachusetts, 2004, pp. 409–424.
[11] K. Doya, et al., Bayesian Brain: Probabilistic Approaches to Neural Coding, MIT Press, 2007.
[12] A. Pouget, J.M. Beck, W.J. Ma, P.E. Latham, Probabilistic brains: knowns and unknowns, Nature Neurosci. 9 (2013) 1170–1178.
[13] P.E. Latham, S. Deneve, A. Pouget, Optimal computation with attractor networks, J. Physiol. (2003) Elsevier.
[14] S. Deneve, J. Duhamel, A. Pouget, Optimal sensorimotor integration in recurrent cortical networks: a neural implementation of Kalman filters, J. Neurosci. (2007).
[15] M. Ernst, Optimal multisensory integration: assumptions and limits from B.E. Stein (Ed.), in: The New Handbook of Multisensory Processing, MIT Press, 2012, pp. 527–543.
[16] J. Feldman, Bayesian models of perceptual organization, in: J. Wagemans (Ed.), Handbook of Perceptual Organization, Oxford University Press, Oxford, 2015.
[17] H. Zhu, O. Basir, A novel fuzzy evidential reasoning paradigm for data fusion with applications in image processing, J. Soft Comput. 10 (2006) 1169–1180.
[18] P.J. Escamilla-Ambrosio, N. Mort, Hybrid Kalman filter-fuzzy logic adaptive multisensor data fusion architectures, in: Proc. of the IEEE Conf. on Decision and Control, 2003, pp. 5215–5220.
[19] M. Oxenham, The effect of finite set representations on the evaluation of Dempster's rule of combination, in: Proc. of the Intl. Conf. on Information Fusion, 2008, pp. 1–8.
[20] B. Rowland, Computational models of multisensory integration, in: B. Stein (Ed.), The New Handbook of Multisensory Processing, MIT Press, Cambridge, 2012, pp. 511–514.
[21] M.S. Landy, M.S. Banks, D.C. Knill, Ideal-observer models of cue integration, in: J. Trommershuser, K. Krding, M.S. Landy (Eds.), Sensory Cue Integration, Oxford University Press, 2001, pp. 5–29.
[22] M.O. Ernst, H.H. Blthoff, Merging the senses into a robust percept, J. Trends Cogn. Sci. 8 (2004) 162–169.
[23] S.L. Bressler, Understanding cognition through large-scale cortical networks, J. Curr. Dir. Psychol. Sci. 2 (2002) 58–61.
[24] M.S.A. Graziano, C.G. Gross, C.S.R. Taylor, T. Moore, A system of multimodal areas in the primate brain, in: C. Spence, J. Driver (Eds.), Crossmodal Space and Crossmodal Attention, Oxford University Press, New York, 2004, pp. 51–68.
[25] N.V. Swindale, How different feature spaces may be represented in cortical maps, Network 5 (2005) 217–242.
[26] S.L. Bressler, E. Tognoli, Operational principles of neurocognitive networks, J. Psychophysiol. 60 (2006) 139–148.
[27] C. Axenie, J. Conradt, Cortically inspired sensor fusion network for mobile robot heading estimation, in: Proc. of Intl. Conf. on Artificial Neural Networks, 2013, pp. 240–247.
[28] S.L. Bressler, J.A.S. Kelso, Cortical coordination dynamics and cognition, Trends Cogn. Sci. 5 (2001) 26–36.
[29] F. Michler, R. Eckhorn, T. Wachtler, Using spatiotemporal correlations to learn topographic maps for invariant object recognition, J. Neurophysiol. 102 (2009) 955–964.
[30] O. Sporns, Networks of the Brain, MIT Press, Cambridge, MA, 2011.
[31] M.A. Carreira-Perpinan, R.J. Lister, G.J. Goodhill, A computational model for the development of multiple maps in primary visual cortex, J. Cerebral Cortex 15 (2005) 1222–1233.
[32] B.E. Stein, T.R. Stanford, M.T. Wallace, J.W. Vaughan, W. Jiang, Crossmodal spatial interactions in subcortical and cortical circuits, in: C. Spence, J. Driver (Eds.), Crossmodal Space and Crossmodal Attention, Oxford University Press, New York, 2004, pp. 25–50.
[33] A. Pouget, S. Deneve, J.-R. Duhamel, A computational neural theory of multisensory spatial representations, in: C. Spence, J. Driver (Eds.), Crossmodal Space and Crossmodal Attention, Oxford University Press, New York, 2004, pp. 123–140.
[34] G.A. Calvert, T. Thesen, Multisensory integration: methodological approaches and emerging principles in the human brain, J. Physiol. 98 (2004) 191–205.
[35] A.K. Engel, C. Gerloff, C.C. Hilgetag, G. Nolte, Intrinsic coupling modes: multiscale interactions in ongoing brain activity, Neuron 80 (2013) 867–886.
[36] L.H. Arnal, A.L. Giraud, Cortical oscillations and sensory predictions, Trends Cogn. Sci. 16 (2012) 390–398.
[37] R.P.N. Rao, D.H. Ballard, Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, Nature Neurosci. 2 (1999) 79–87.
[38] T. Zhang, et al. An FPGA implementation of insect-inspired motion detector for high-speed vision systems, in: Proc. of Intl. Conf. on Robotics and Automation, 2008, pp. 335-340.
[39] J.F. Ferreira, J. Lobo, P. Bessire, M. Castelo-Branco, J. Dias, A Bayesian framework for active artificial perception, IEEE Trans. Cybern. 2 (2013) 699–711.
[40] J.F. Ferreira, J. Lobo, J. Dias, Bayesian real-time perception algorithms on GPU — real-time implementation of bayesian models for multimodal perception using CUDA, J. Real-Time Image Process. 6 (2011) 171–186.
[41] W. Guenthner, et al., Biologically inspired multi-sensor fusion for adaptive camera stabilization in driver-assistance systems, in: Advanced Microsystems for Automotive Applications, Springer, Berlin, 2006.
[42] M. Cook, L. Gugelmann, F. Jug, C. Krautz, A. Steger, Interacting maps for fast visual interpretation, in: Proc. of Intl. Joint Conf. on Neural Networks, 2011, pp. 770–776.
[43] T. Mergner, W. Becker, Perception of horizontal self-rotation: Multisensory and cognitive aspects, in: R. Warren, A.H. Wertheim (Eds.), Perception & Control of Self-motion, Hillsdale, 1990, pp. 219–260.
[44] A.W. Wertheim, Visual, vestibular, and oculomotor interactions in the perception of object motion during egomotion, in: R. Warren, A.H. Wertheim (Eds.), Perception & Control of Self-motion, Hillsdale, 1990, pp. 171–210.
[45] G.C. DeAngelis, D.E. Angelaki, Visual-vestibular integration for self-motion perception, in: M.M. Murray, M.T. Wallace (Eds.), The Neural Bases of Multisensory Processes, CRC Press, Boca Raton, 2012, pp. 629–652.
[46] J.M. Wiener, A. Berthoz, T. Wolbers, Dissociable cognitive mechanisms underlying human path integration, J. Exp. Brain Res. 208 (2011) 61–71.
[47] B. Khaleghi, A. Khamis, F.O. Karray, Multisensor data fusion: a review of the state-of-the-art, J. Inf. Fusion 14 (2013) 562–578.

Cristian Axenie is a Ph.D. student and research assistant in the Neuroscientific System Theory Group at the Technische Universität München in the Faculty of Electrical Engineering and Information Technology, Institute of Automation and Control Engineering. He holds B.Sc. and M.Sc. degrees in Automation and Robotics from University of Galati, Romania. His main research interests are distributed neural computational architectures for multisensory processing instantiated both in software and hardware.

Jörg Conradt is a Junior Professor at the Technische Universität München in the Faculty of Electrical Engineering and Information Technology, Institute of Automation and Control Engineering. The laboratory is affiliated with TUM's Competence Center on NeuroEngineering and the Munich Bernstein Center for Computational Neuroscience. He holds an M.S. degree in Computer Science/Robotics from the University of Southern California, a Diploma in Computer Engineering from TU Berlin and a Ph.D. in Physics/Neuroscience from ETH Zurich. His research group on Neuroscientific System Theory (http://www.nst.ei.tum.de) investigates key principles by which information processing in brains works, and applies those to real-world interacting technical systems.