An Integration Principle for Multimodal Sensor Data Based on Temporal Coherence of Self-Organized Patterns.

Emilia I. Barakova

GMD - Japan Research Laboratory, AIM Building 8F, 3-8-1,
Asano, Kokurakita-ku, Kitakyushu-city, 802-0001, Japan,
Work: +81 93 512 1566, Fax: +81 93 512 1588,
emilia.barakova@gmd.gr.jp, http://www.gmd.gr.jp.

Abstract. The world around us offers continuously huge amounts of information, from which living organisms can elicit the knowledge and understanding they need for survival or well-being. A fundamental cognitive feature, that makes this possible is the ability of a brain to integrate the inputs it receives from different sensory modalities into a coherent description of its surrounding environment. By analogy, artificial autonomous systems are designed to record continuously large amounts of data with various sensors. A major design problem by the last is the lack of reference of how the information from the different sensor streams can be integrated into a consistent description. This paper focuses on the development of a sinergistic integration principle, supported by the synchronization of the multimodal information streams on temporal coherence principle. The processing of the individual information streams is done by a self organizing neural algorithm, known as Neural gas algorithm. The integration itself uses a supervised learning method to allow the various information streams to interchange their knowledge as emerged experts. Two complementary data streams, recorded by exploration of autonomous robot of unprepared environments are used to simultaneously illustrate and motivate in a concrete sense the developed integration approach.

1 Motivation.

The ability of a brain to integrate the inputs it receives from different sensory modalities into a consistent description of its surrounding world is its basic feature, that helps us orient in tasks with different complexity. It has been widely argued how and whether at all the integration takes place [5][6][9][10][14], and many models has been suggested therefore[3][11][12].

The integration principle, that is featured in this paper is based on the understanding, that there are two aspects of the integration process: (1) achieving a sinergistic integration of two or more sensor modalities and (2) actual combination (fusion) of the various information streams at particular moments of their processing.

The sinergistic integration relies on a hypothesis of how different percepts unify in the brain. It is based on some evidences from temporal registration and binding experiments [14]. For the actual combination the hypothesis is concreticised so that the different sources of sensory information are brought to one coherent representation. For this purpose a synchronisation on a temporal principle is proposed.

This paper focuses on an information combination method on a temporal coherence principle. The combination is made within the framework of an integration strategy proposed, and is widely intertwining with the application domain of concurrent mapping and navigation[1][5].

Two complementary data streams, recorded during the exploration of unprepared environments by an autonomous robot are used to simultaneously illustrate and motivate in a concrete sense the developed integration approach. They provide information about the movement of an autonomous robot from two perspectives: absolute - the robot movement with respect to the surrounding objects (recorded by laser range finders) and relative (recorded by the build-in gyroscope).

The neurobiological experiments have shown, that information from one type of sensors is processed separately on a certain time interval[6][14]. Accordingly, the processing of the individual data streams is done separately, by a self-organizing neural structures (neural gas algorithm in particular [8]) each. The integration of the different information streams ensues the hypothesis made, as well as the outcome of the experiments, of Triesh at al. [13] and uses a backpropagation algorithm for ensuing the different processing streams learn from each other.

This paper is organized as follows: First, an integration hypothesis chooses the scope, that the integration principle will follow. Further on the integration principle as determined by the hypothesis is narrowed down to an practically implementable approach in section 3. Simultaneously, the applications domain is briefly introduced. The flow-chart, shown in the next section follows the information transformations, which bring the information from two orthogonal data streams into a coherent description. Some results illustrate the plausibility of temporal integration principle.

2 Integration hypotheses.

It has been widely argued how the results of different processing systems come together in the brain, to give an unitary perception of the surrounding world.

Chronologically first comes the hypothesis that there are one or more areas in the brain, where integration of different processing streams physically takes place .

Neurophysiological experiments have revealed that there is not a single area in the brain to which different specialized areas uniquely connect. Instead, the brain activities, caused by perception, as well as those, related to memory experiences are simultaneously active in different, highly interconnected functionally specialized areas.

The other group of attempts to reveal the mechanisms that relate various activations is based on the hypothesis, that there is a temporal relation of operations, performed in different processing streams. A precise temporal registration of the results of this operations is possible for intervals of time bigger than one second. The brain is therefore not capable of binding together information entities from different modalities in real-time; instead, it binds the results of its own processing systems.

By far, there is not a single theory, that explains exactly how integrating takes place in the brain. Instead of trying to answer to the question *how* the integration takes place, the approach, suggested in this paper will be build on the hypothesis *why* the integration takes place.

There are variety of answers to this question. The following reasoning will suggest one, that gives a constructive basis for an integration strategy. On a level of a separate sensor modality channel, the brain operates as a self-organizing information system. It obtains inputs from various sensors and in any separate sensor modality stream it clusters the information from its inputs in a self-organizing manner into asymmetric patterns. Since every separate modality brings a different level of generality and scope of information about the external world, the information from one modality can furthermore serve as a "teacher" for the other modality.

In the static world we could use the answers we know as a teacher or expert knowledge. Instead, in a changing world routines and category judgments from the past may be inadequate or misleading. Integrating the on-line, up-date information which brings different level of generality and is sensed from different scope, can give us a key of how to adapt to the new situation and deal with it, and not to solve problems from the past in the new reality. Therefore, the information integration is the mechanism, which allows us to learn in a changing world.

This hypothesis and its preliminaries suggest that we can process separately the information from one type of sensors on a certain time interval in a self-organizing manner. Evaluating the "superiority" of a certain sensor channel to judge more generally about a specific aspect or feature of the reality, we can make it instead give the major notion about the new encountered event. The other sensor stream can tune the certainty of the information from the first stream and to enrich it with the nuances of the novelty.

3 The integration approach.

In the previous sections of this paper an integration concept has been suggested. Here, the conceptual considerations will be brought to a concrete, technically plausible approach.

In the world of the artificial autonomous systems various sensors asynchronously provide information that has different meaning and sampling characteristics. In addition, there are not established ways of combining the information from different information sources. To achieve an actual combination of the multimodal information sources, the following arguments will be used as a starting point:

- Data, that are perceived (recorded) at the same time relate to the same situation (event).
- Processing of different data streams is done in separate modalities, followed by synchronization on a temporal principle.
- The temporal synchronization is event-based, (in contrast to fixed time interval based).

In addition, according to the conceptual considerations outlined so far, first, an event-based time intervals have to be defined. Second, the information, recorded within this intervals has to be brought to entities, that can be combined technically. And third, the actual combination has to take place.

To get a better intuition about the multimodal sensor integration approach, suggested in this paper, the application task of mapping of unknown environments for the purpose of navigation will be used.

The mapping task is to be solved by using the data, that a mobile robot records during its exploration of an unprepared environment. Figure 1 shows the experimental environment. With black points on the floor are shown some places, which are encountered by the robot as novel, and are clustered in different classes (situations), on the basis of the sensor information, as it will be outlined further on.

Figure 1 Experimental environment.

One can hardly think of a group of sensors, which can imitate the consummate description of the environment, that biological systems can create. As a plausible alternative, a set of orthogonal sensors that can complement the perception of each other views on the surrounding world, can be found.

In [2] is elaborated on the relevance of the egocentric perspective of an autonomous robot in spatial modelling of previously unknown environments. The egocentric model in [2] combines two types of information: absolute and relative with respect to the robot motion.

As an absolute source of information are used the “views” that the robot perceives with a laser range finder. The individual ‘view’ is formed by the record of 720 samples per 360 degrees. A snapshot of a polar representation of such a record is shown at figure 2a). Snapshots are recorded at frequency of 4.7 Hz. The distances are presented in milimeters.

Sequence of such snapshots, recorded during the robot exploration and stored in a short-term memory (STM) - like manner, form a dynamical trajectory, which represents the first information stream, used for the integration. It represents the absolute perspective of the robot about its own motion (i.e. robot motion with respect to the surrounding objects). More details on dynamical trajectory formation can be found in [1]. Here only the final description of the dynamical trajectory will be given:

$$f_k(t) = (t/k)^{\alpha}(e^{\alpha(1-\frac{t}{k})}) \qquad k = 1, \ldots, K \tag{1}$$

As a relative perspective of the robot is taken the information from the build-in gyroscope. It reflects the way in which the robot perceives its own motion. As a most informative is decided to be the curve, describing the angular velocity of the robot, since it reflects the changes in the direction of the trajectory of the robot and is usually associated with qualitatively novel situations in the surrounding environment, which have caused this changes.

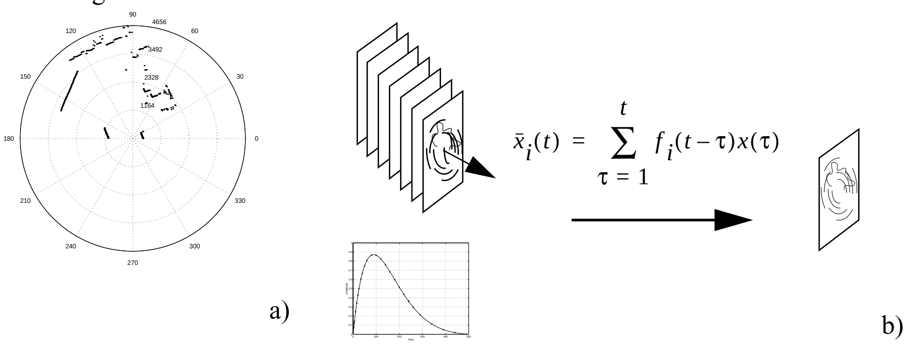

Figure 2 a)A sensor sample; b)Dynamic trajectory formation.

The temporal synchronisation of the two information streams is performed in the following way. After exploiting the information, that the egocentric perspective that a robot can provide about its movement (based on path integration information - angular velocity data) and on landmark-type information, dynamic events have been created. This dynamic events are developed with minimal processing or interpretation of the recorded data. They contain information from two highly orthogonal sensor sources (relative and absolute).

In addition, information about the time cooccurence of the two sequences of dynamic events (i.e. two time-dependant segments of measurements or representations that happened simultaneously) is used in order to make more complete final representation.

Figure 3 illustrates the implementation of the principle of temporal synchronization over the two information streams. The first one represents the dynamic trajectory, that a robot takes during its exploration of an environment. The qualitatively different “view”, that the robot observes define every new segment by this exploration (figure 3b). The duration of this segments determine the division of the other sensor data stream, that reflects the changes in the angular velocity as recorded by a gyroscope, as follows:

$$T_{ir} = \frac{f_{ia}}{T_{ir}}(T_{ia}) \tag{2}$$

Where T_{ia}, T_{ir} are correspondingly the lengths of the i-th segment of the absolute and the relative streams of data, and f_{ia}, f_{ir} are the corresponding sampling frequencies.

a)

b)

c)

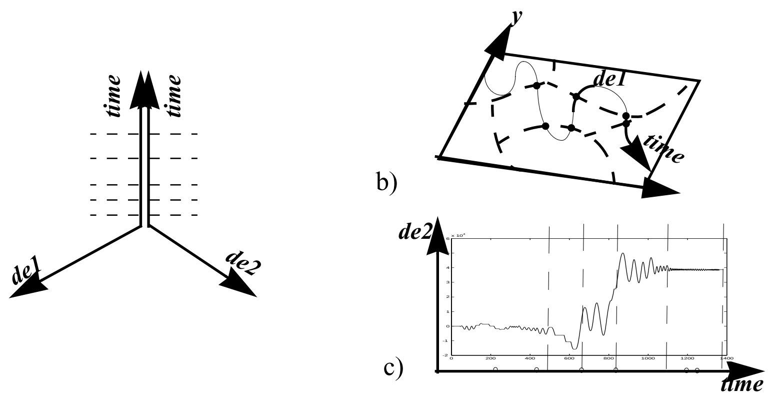

Figure 3 The temporal synchronization principle.

This way, the synchronization of the two sensor streams is completed. Further on the data streams have to be brought to the same representational format and integrated.

4 Information flow chart.

In this section in short will be explained the actual steps that developed approach takes to accomplish integration of multimodal information streams.

After representing the information from the laser range finders in a dynamical way, as discussed in the previous section, the first dynamical event is formed. It is prefered the term ‘event‘ to ‘feature‘, because feature is usually associated with some kind of processing of the underlying information, so that some of its essential properties are extracted. In this work neither processing of the information that presumes any sort of its interpretation has taken place, nor extraction of some essentials is made. Instead, the percieved information is coded as compact as possible, by using ideas from biological systems.

As mentioned in the previous section, we distinguish exploration (learning) and testing phase in our experiments. During the learning phase, a Neural gas (NG) algorithm clusters the sequences of views, recorded during the exploration of the mobile robot of previously unknown environment. The moments, when a qualitively new view is encountered, are used for a temporal division of the second dynamical stream of information, recorded by the build-in gyroscope. The distinctively new view is defined by a distance measure, which by now is empirically defined.

The velocity trajectories, recorded by the gyroscope are divided into segments on an event basis. The appearance of a new event is determined (as explained) by the absolute information stream. The so defined segments are clustered in different classes with a help of another NG network.

The testing phase takes similar operations, performed on the testing data sets. Instead of clustering, here is performed classification, according to the clusters, defined over the exploration data.

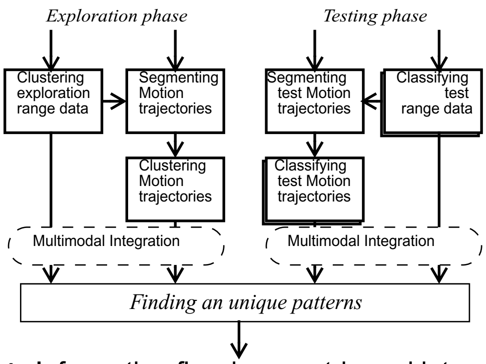

Figure 4 Information flow by event-based integration.

For a better observability, the so described processing steps will be shown in the following information flow chart (figure 4).

5 Integration results.

To show the plausibility if the suggested integration principle, two processing strategies are compared. The first one simply combines the clustering results from the both processing streams, while the second uses the integration principle for the combination. Both strategies are tested on recognizing passed (during the exploration phase) itineraries, while operating on the test data sets.

By the processing strategy that involves integration on the basis of the developed principle, the clustering/classification results from the relative processing stream are used as a teacher for distinguishing similarly looking dynamic trajectories. After the exploration data stream is clustered and the dynamic trajectories are defined, groups with similarly looking final trajectories were distinguished. The similarities could be caused either because the robot practically never passes the same itinerary by free exploration, or because there are similarly looking places in the environment. Therefore, the velocity information, represented as a sequence of classes was used as a teacher information, so the similarly looking places are discriminated as different scenarios, while the differ-

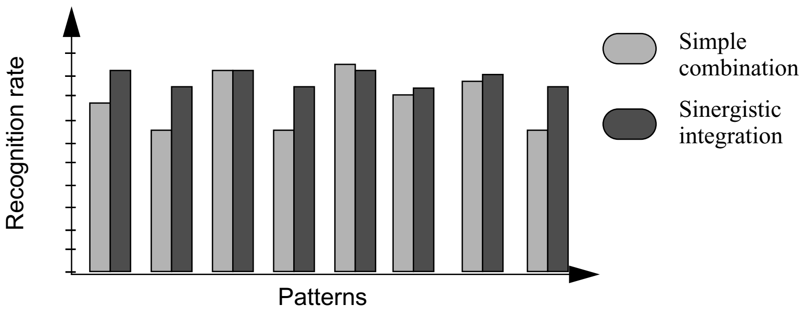

Figure 5 Recognition rate by simple combination and by integration.

ences in the dynamical trajectory representation, caused by slight variations of the underlying itineraries are considered as the same dynamical scenario.

6 Discussion.

The elaborated integration for concurrent mapping and navigation explains the perceptual hierarchy in the following way: the knowledge about the instant movement by humans and many animals contributes a lot to the short-time navigation, while in longer time span they use their perceptions of the surrounding world. In brief analogy, the information that the robot perceives with his range sensors resembles the orientation according to the remembered views, while the velocity information has an influential similarities with the instant movement information by humans.

The important outcomes from the developed hypothesis and its implementation to the considered integration case are as follows:

- The developed method takes into account measurements from a separate sensor source on intervals that does not allow accumulation of errors which can affect the modelling process.
- There is a dynamical way of coding both: the consequent perceptions as well as the transitions between them. This is made possible also by the partially independent ways of processing of the separate information streams.
- The only interpretation of the information, contained in the data streams is made only with respect to defining the priority of the sensor judgements.

The experiments, analysed in [13] suggest, that the ways in which different modalities are integrated depend on the information cues involved, and the nature of the task. The reason, according to him that there is not an unique integration strategy developed yet is that the biological systems do not use a single immutable strategy to combine different percepts.

The hypothesis, made in this paper presumes, that the different percepts about an observed event are processed separately in different sensor modalities. Therefore they acquire different scope and generality of the information about the event they describe. As a result, the outcome of one processing stream can be used as a teacher to the other processing stream's outcome. Considering the argumentation of Triesh [13], we do not tend to show, that this hypothesis is valid for any combination of percepts. Anyway, a large range of validity of this hypothesis is expected.

The shown results reflect actual experiments that aim concurrent mapping and navigation of unknown environment. In a way the shown results are preliminary: the integration principle argues, that the both information streams can be used as teacher of each other, while the made experiments show how the second information stream, used as a teacher of the first one improves the recognition rate with respect to a simple combination of the two information streams after being processed. Our current work concentrates on the mutual learning of the two streams of each other and may need a development of a novel supervised algorithm.

7 References.

[1] Barakova, E. I. & Zimmer, U. R.Dynamical situation and trajectory discrimination by means of clustering and accumulation of raw range measurements, Proc. of the Intl. Conf. on Advances in Intelligent Systems, Canberra, Australia, February 2000.

[2] Barakova, E.I. & Zimmer, U. R. "Global spatial modelling based on dynamics identification according to discriminated static sensations" Underwater Technology 2000, Tokyo, Japan, May 23-26 2000.

[3] Becker, S. and Hinton, G. E. (1992). A self-organizing neural network that discovers surfaces in random-dot stereograms. Nature, 355:161-163.

[4] Grossberg, S. The complementary brain :A Unifying View of Brain Specialization and Modularity.

[5] Elfes, A. “Sonar-based real-world mapping and navigation”, IEEE Journal of Robotics and Automation, RA-3(3):249-265, June 1987.

[6] Eagleman, D.M. & Sejnowski, T.J. (2000) Motion integration and postdiction in visual awareness. Science. 287(5460): 2036-8.

[7] Cramer, K.C. and Mriganka Sur. Activity-dependent remodeling of connections in the mammalian visual system. Current Opinion in Neurobiology, 5:106,111, 1995.

[8] Martinetz,T.M., S.G. Berkovich, K.J. Schulten, ‘Neural-Gas’ Network for Vector Quantization and its Application to Time-Series Prediction, IEEE Transactions on Neural Networks, Vol. 4, No. 4, July 1993, pp. 558-569.

[9] von der Malsburg, C. The correlation theory of brain function., in Models of Neural networks II, edited by E. Domany, J.L. van Hemmen, and K. Schulten (Springer, Berlin, 1994) Chapter 2, pp. 95, 119.

[10] von der Malsburg, C. Binding in Models of Perception and Brain Function. Current Opinion in Neurobiology, 5:520, 526, 1995.

[11] McGurk, H. and MacDonald, J. (1976)., Hearing lips and seeing voices., Nature, 264:746-748.

[12] de Sa, V.R. and Dana H. Ballard, Category Learning through Multi-Modality Sensing, In Neural Computation10(5) 1998.

[13] Triesch, J. (2000). Democratic Integration: A theory of adaptive sensory integration. Technical Report NRL TR 00.1, Nat'l. Resource Lab'y. for the Study of Brain and Behavior,U. Rochester.

[14] Zeki, S. Functional specialization in the visual cortex of the rhesus monkey, Nature (Lond.)274,423-8.