Multisensory Memory Representation of Robot Actions

Martin Weser
E-mail: weser@informatik.uni-hamburg.de

Motivation

An extensive grounded robot memory is one of the most interesting, important and challanging research topics in recent years. Apart from many applications there are several different learning approaches discussed in the scientific community. The common point of all differing learning methods is their ability to modify or adjust data within certain memory structures. Many learning methods adjust parameters which directly influence the behaviour of the executing algorithm but nevertheless these are at least task-specific types of memory. Apart from this learning methods an extensive memory system could be considered the most central theme within the modern AI.

History of Intelligent Systems

In the history of AI scientists have taken one of the three following approaches to build intelligent machines [1]:

1. The oldest approach is **knowledge based**. That means that a program is directly written to solve a given problem or a certain class of problems. The knowledge how to solve this problems is induced directly by the programmer within the source code.
2. The second approach is **learning based**: A task speciffic learning method is applied to a set of training data. Parameters used by an algorithm are adapted in order to achieve better results while processing test data.
3. The **genetic search** is the most recent approach. Here one tries to find the best method to solve a problem using a biologically inspired *mutation and survival of the fittest* method. In contrast to the privious learning-based approach the algorithm itself could be modified.

It is noticeable that the developement of structures (no matter whether they are of algorithmic or parametric nature) is shifted from the programmer to the artificial system. However, none of these approaches are capable of leading to diverse cognitive capabilities like the brain. Furthermore, it is noticeable that the first approach is still a promising one. To give just one prominent example the Loebner price for solving the Turing test was given to Jason Hutchins in 1996 who did not use any AI techniques at all.
The fact that all these approaches are only suitable to solve speciffic problems points to the need for a more general approach which can be used for more unspecified tasks. If we assume that this approach is learning based it leads to **autonomous mental developement** (see [1] for a definition).

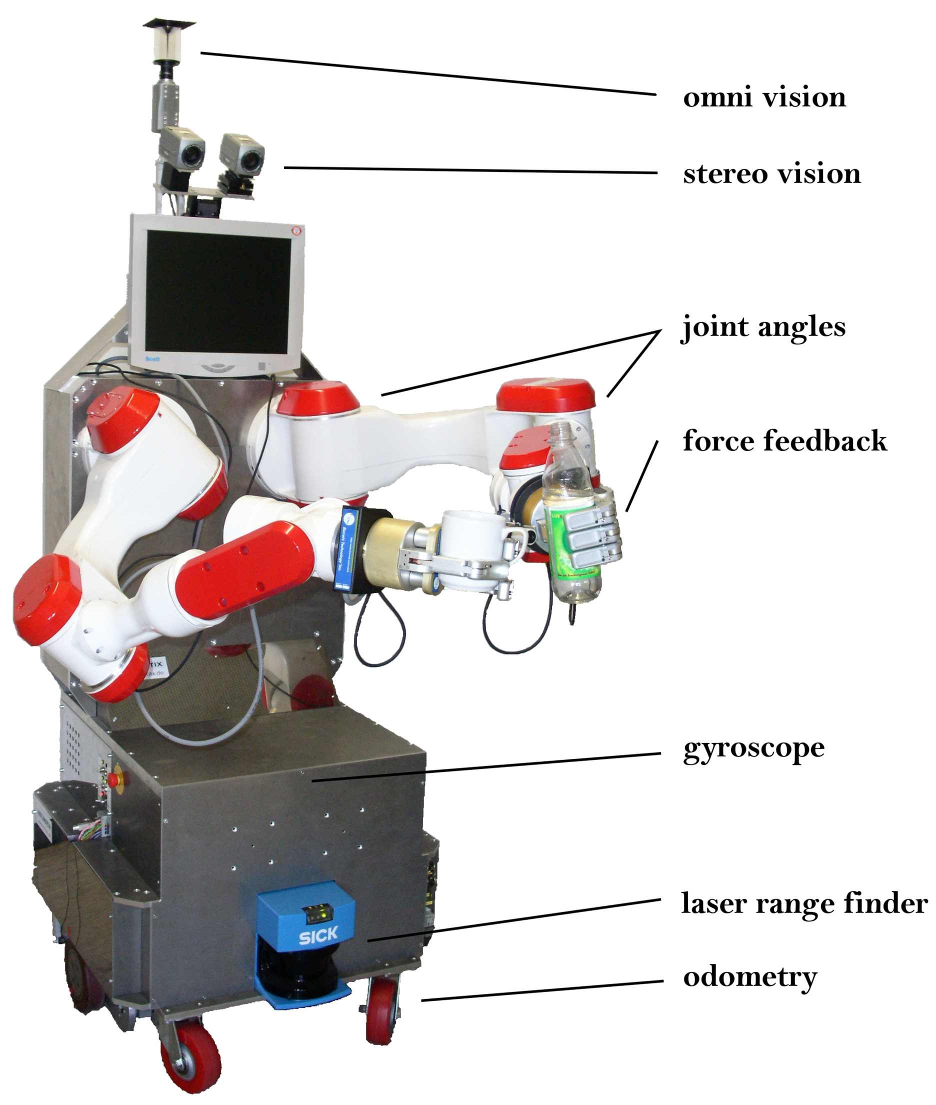

Figure 1: Our mobile service robot TASER.

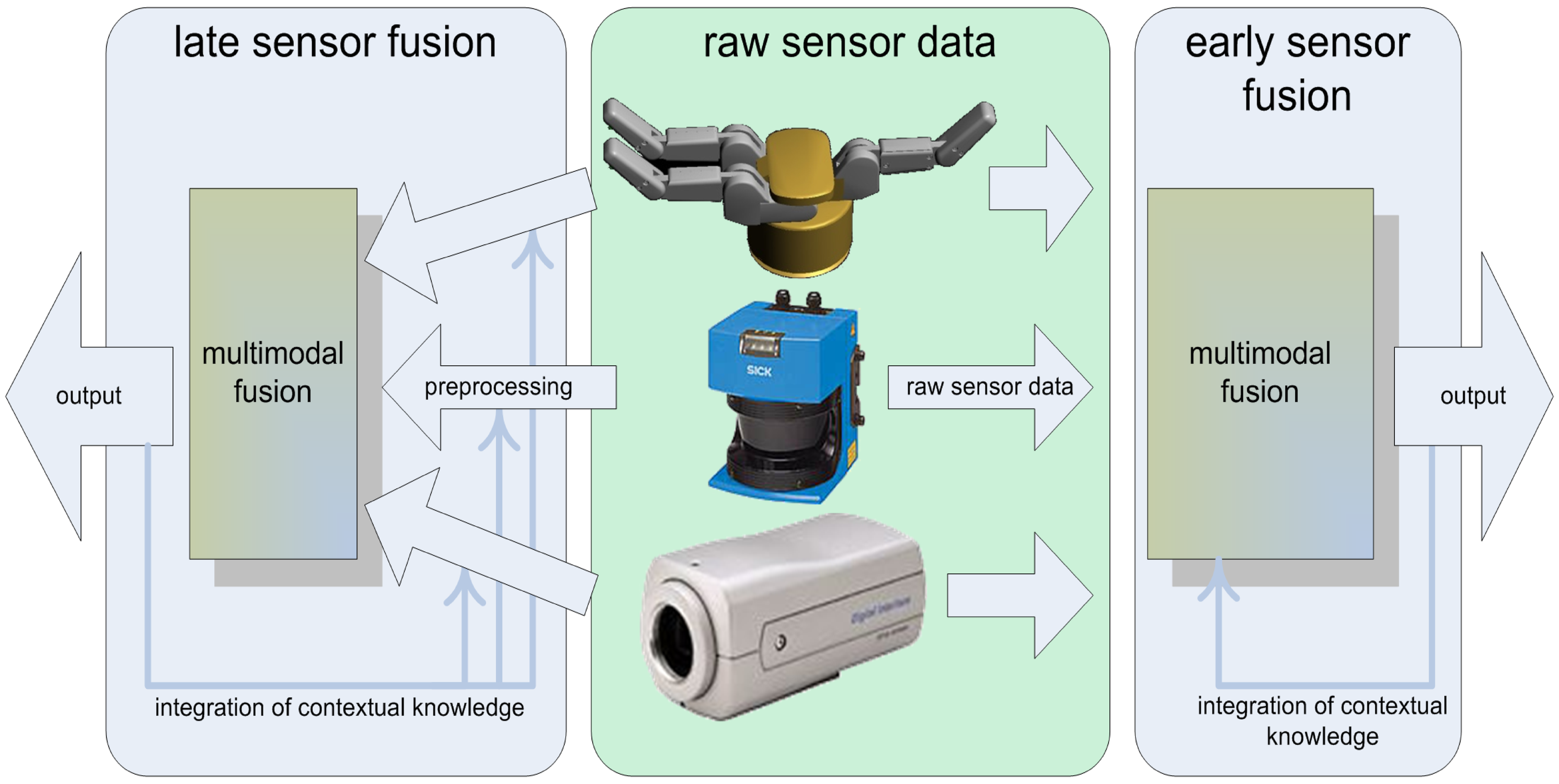

Figure 2: Two different methods for fusing multimodal sensor data: On the left the fusion is done after preprocessing each sensory data stream seperately; on the right the raw sensory data are fused.

Modeling requires Multisensor Fusion

The generation of an environment model out of sensor data is a nontrivial task since sensor calibration and measurements are susceptible to errors. In addition, the observed objects may be occluded and therefore not detectable by all sensors. We argue that this uncertainty about the environment decreases when different sensor-modalities with diverse qualities are combined. The advantages of the different sensors can complement one another using appropriate fusing methods.
The profit from combining the input of the different sensory systems is not only that each modality provides information about different aspects of the world but also because the different senses can jointly encode particular aspects of events, e.g. the location or meaning of an event. Our mobile service robot TASER offers a wide variety of sensor modalities. They are listed in figure 1.
Sensor fusion could be achieved at different levels of processing. The lowest level is considered to be at the stage of raw sensory data. The fusion of abstract object properities (or environmental situations) can be considered as high level integration of sensor data. This two variants are illustrated in figure 2.

Advagtages of Late Multimodal Integration

An advantage of the second approach is that several preprocessing methods for single sensor data streams are available or already build in hardware. Furthermore, computational costs can be ballanced in order to focus resources on modalities which seem to be most important for the currently processed application. The complexity of preprocessing sensor data could be reduced by deviding the combined multimodal data stream into distinct ones.

Disadvagtages of Late Multimodal Integration

One and the same event in the real world could cause almost unlimited sensory patterns. Thus the raw sensory data are fraught with ambiguities which are hard to be solved without contextual information. Furthermore the accuired data could contain noise and distorted information. To filter out this clutter and to solve the ambiguities one could use information coded within other sensor-modalities at a very early stage of computation. In turn this could save computational and memory resources.

The Biological Archetype

The human brain has the most complex, diverse and highly integrated capabilities in building and retrieving memory representations. Recent results of neuro-psychological investigations show that the integration of different modalities is done at a very early stage e.g. within the first 150ms. Due to the fact that evolution has evolved the most powerful algorithm over millions of years we try to understand and adapt its underlying mechanisms.

Hierarchical Framework of Functional Realistic Neurons

In [2] it was shown that a multiple-stage framework of neurons with unsupervised local learning rule is capable of learning high-level structures e.g. the position and location of an autonomous robot. They used a single front view camera as sensor modality.
An outline of our intended work is given by the following points:

- The same mechanisms will be applied to different types of sensors (e.g. laser range scanners) to show that simple neural mechanisms are capable of learning structures in arbitrary real world sensor data.
- The next step is to apply this network to different sensor modalities simultaneously. The different modalities will be fed into the network at different hierarchical stages.
- We want to investigate whether there is a benefit of integrated processing compared to independent processing and subsequent computational fusion methods (e.g. Kalmann Filter, Hidden Markov Model).
- Finally, we will compare results of the integration of multimodal data at different hierarchical stages within the network.

References

[1] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. In *Science, 291*, 2001.

[2] Wyss R., Koenig P., and Verschure P. F. M. J. A model of the ventral visual system based on temporal stability and local memory. *PLoS Biology*, 4(5), 2006.

CINACS – Cross-modal Interactions in Natural and Artificial Cognitive Systems
Dept. Informatics, Faculty of Mathematics, Informatics and Natural Sciences
http://www.cinacs.org