Human and Machine Perception

Information Fusion

Human and Machine Perception

Information Fusion

Edited by

Virginio Cantoni
University of Pavia
Pavia, Italy

Vito Di Gesù
University of Palermo
Palermo, Italy

Alessandra Setti
University of Pavia
Pavia, Italy

and

Domenico Tegolo
University of Palermo
Palermo, Italy

SPRINGER SCIENCE+BUSINESS MEDIA, LLC

Library of Congress Cataloging-in-Publication Data

Human and machine perception : information fusion / edited by Virginio Cantoni ... [et al.].
p. cm.
"Proceedings of the Second International Workshop on Human and Machine Perception: Information Fusion, held July 21-25, 1996, in Trabia, Italy"--T.p. verso.
Includes bibliographical references and index.
ISBN 978-1-4613-7734-4 ISBN 978-1-4615-5965-8 (eBook)
DOI 10.1007/978-1-4615-5965-8
1. Human-computer interaction--Congresses. 2. Human-machine systems--Congresses. 3. Multisensor data fusion--Congresses.
I. Cantoni, V. II. International Workshop on Human and Machine Perception: Information Fusion (2nd : 1996 : Trabia, Italy)
QA76.9.H85H84 1997
004'.01'9--dc21 97-33281
CIP

Proceedings of the Second International Workshop on Human and Machine Perception: Information Fusion, held July 21 – 25, 1996, in Trabia, Italy

ISBN 978-1-4613-7734-4

Originally published by Plenum Press, New York in 1997

10 9 8 7 6 5 4 3 2 1

To Cristina Omet Levialdi
for her presence amongst us

PREFACE

The following are the proceedings of the Second International Workshop on Human and Machine Perception held in Trabia, Italy, on July 21-25, 1996, under the auspices of two Institutions: the Cybernetic and Biophysics Group (GNCB) of the Italian National Research Council (CNR) and the 'Centro Interdipartimentale di Tecnologie della Conoscenza' of Palermo University. A broad spectrum of topics are covered in this series, ranging from computer perception to psychology and physiology of perception (visual, auditory, tactile, etc.).

The theme of this workshop was: "Human and Machine Perception: Information Fusion". The goal of information and sensory data fusion is to integrate internal knowledge with complementary and/or redundant information from many sensors to achieve (and maintain) a better knowledge of the environment. The mechanism behind the integration of information is one of the most difficult challenges in understanding human and robot perception.

The workshop consisted of a pilot phase of eight lectures introducing perception sensorialities in nature and artificial systems, and of five subsequent modules each consisting of two lectures (dealing with solutions in nature and machines respectively) and a panel discussion.

The lectures focused on presenting the state-of-the-art and outlining open questions. In particular, they sought to stress links, suggesting possible synergies between different cultural areas. The panel discussion was conceived as a forum for an open debate, briefly introduced by each panellist, and mainly aimed at deeper investigation of the different approaches to perception and strictly related topics. The panellists were asked to prepare a few statements on hot-points as a guide for discussion. These statements were delivered to the participants together with the final program, for more qualified discussion.

The number of participants to the workshop was limited to 50. Besides the 30 invited lecturers and panellists, 20 more participants were admitted. Priority for these positions was given to young researchers who made significant contributions to the open discussions. Both the lectures and the contents of the panel discussions are included in the present proceedings.

Table 1 shows the scheduled program for the pilot phase including the four most important senses applied in artificial systems. The following modules, listed in Table 2, consider differences between turnouts from the perceived data.

The chapters of the book vary somewhat from the scheduled workshop program, to accommodate the positions that emerged from the debates. In what follows each subject is briefly introduced and commented.

Table 1. Sensors for perception.

Sensorialities	Nature	Machine
sound	auditory system	sonar techniques
image	visual pathways	visual sensors
posture	vestibular system	trim control
touch	somatosensory system	tactile sensors

Table 2. Perception versus X subjects.

Subject	Nature	Machine	Panel
Perception and Integration	Sensory motor computation in the cerebellum	Sensory fusion	Integrability and adaptability in sensory systems
Perception and Decision	Redundancy and deficiency of information	Perception and decision making	Pragmatic observation and evaluation
Perception and Action	Reflexes and voluntary behaviour	Predictive and reactive behaviour	Comportmental models
Perception and Representation	Multiple knowledge sources integration	Symbolic and conceptual representation	Knowledge model representations
Perception and Communication	Communication by words and pictures	Multimedia interfaces	Characters, pixels and phonemes

Sensorialities

The description of both natural and artificial senses represents a natural starting point for a workshop which intends to analyse the process of information fusion from different viewpoints. The lectures were organised by alternating the functional descriptions of natural and artificial senses related to sounds, vision, touch-proximity and space attitude.

From the debate emerged three main features common to natural sensors: redundancy, the ability to perform early computation, and plasticity. The first two features are strictly related; indeed, redundancy is necessary to realise fault tolerance mechanisms and to perform early computations at the acquisition level. This leads to a new representation of compressed information. For instance, perceptual grouping could also be partially exploited at an early stage, and used to interpolate data to fill-in missed information (in Moore's chapter this property is referred to as perceptual restoration). The third feature characterises the capacity for functional replacement among sensor elements, and it can be explained only if a sort of intra-element interaction holds. On the other hand, natural systems suffer from some limitations such as narrow frequency range, fixed topology, etc..

Artificial systems exhibit complementary properties, indeed they can be designed to detect a wider frequency range. A single artificial system could be equipped with as many sensors as it needs. Accuracy and resolution could be tuned, in principle, at our convenience. Today, however, artificial sensors mainly implement data acquisition, only a few of them have been designed to perform pre-analysis (active sensors), and integration in the early stages when signals are collected by different sensors is quite rare.

The vestibular system and the trim control theory complete the description of sensorialities in natural and artificial systems, as both are characterized by a particular implementation of the feedback principle.

Summarising, the first fact that emerges from this sensorialities overview, is the ability of natural sensory systems to perform intensive and distributed computation. Moreover, it is worthwhile to highlight the self-tuning capacity of natural systems, and the functional graduality that exists between acquisition and computational levels. These features should guide the design of next generation artificial acquisition systems.

Perception and Integration

The second session of the workshop investigated how information from various sensory systems is integrated. This is one of the most difficult challenges in understanding natural and artificial perception and cognition (see D'Angelo's and Morasso's contributions). Auditory-visual integration is treated as a problem of correspondence between scenes perceived by auditory and visual modalities. Auditory and visual information is integrated to generate an exploitable perception of the environment, which can usefully be applied to human and robot navigation.

Moreover, the excellent vision of humans when compared to artificial systems cannot be explained only in terms of the individual feature extracting function, but also by an ability to adequately integrate sensorial information. Taking an example from biology, when responding to visual stimulation, the cells of the cat visual cortex fire rhythmically at frequencies between 30 and 60 Hertz. This rhythmic firing can be synchronised among the cells in widespread areas of the visual cortex. The visual stimulus conditions causing this process to occur, suggest that such synchronisation contributes to the integration of information across broadly displaced parts of the visual field (see B. Jagadees, Visually evoked oscillations of membrane potential in cells of cat visual cortex, Science, 1992). This experiment suggested artificial visual systems and machine-vision architectures could be designed based on massively interacting processing elements (see also Morasso's presentation regarding robot navigation).

Two fundamental questions arise when designing artificial systems: what is the useful information that must be integrated? What are the rules of integration? The first question is also known as the data selection problem; the answer is not simple and requires in-depth investigation of the concept of conditional information. The second problem involves the chemistry of the integration process, where different sources of information are combined as in a chemical reaction to generate new information, whose nature can differ from the single compounds. A straightforward method of combining information from separated sources considers conditional class probabilities for each channel independently. These probabilities are then combined according to Bayes' rule under the assumption of conditional independence. Other approaches create sensory fusion by using Dempster-Shaffer's rule of combination on belief functions generated from input sensory data.

Multilevel mechanisms can be usefully employed to model the integration of information in both natural and artificial systems, including providing for sensory pre-processing, sensory fusion and high-level decision making. The comparison between natural and artificial systems again shows that flexibility and reconfigurability characterise natural systems, which could explain their high performance. Moreover, the biological "hardware" and its related genetic programs are fully matched, allowing sensorial data to be combined with fusion strategies and system actions.

Perception and Decision

It is well known that the quality of our decision-making is not always improved by increasing the data acquired. The reasons are manifold: data may carry contradictory information; some data is not suitable requiring extra-work to process it.

Moreover, even when all collected data brings useful information, the computational power of natural and artificial systems is physically limited. At present, all theories of perception accept two data reduction mechanisms: attention and fusion. Careful analysis of the meaning of information is required to fully understand them. Data are containers of hidden information that can be extracted only after performing a decision process. In other words, the quality of the decision and its closeness to the goal are plausible measurement criteria for the information contained in the data.

Attention was first studied by psychologists (see for example L. Broadbent, Selective and control processes, Cognition, 1981). Attention is responsible both for data selection and filtering. These however are not stable operations, as Gerbino highlights in his contribution. Indeed, using examples borrowed from vision psychology, he shows how data redundancy is often useful in confirming internal models. At the same time strict selection and reduction of data can destroy contextual information leading to misunderstanding.

The fusion process combines data to extract useful features. The combination rules are suggested by objective and subjective evaluation, and are affected by the goal. Integrated decision algorithms are usually based on inference-rules, that can be combined in graphs and semantic networks. For this purpose several approaches have been proposed, having in common the search for extreme solutions.

Human experience may play a fundamental role whenever data is vague. Moreover, cloudy quantities may exist which are not always describable in terms of probability distributions: examples are evaluating the beauty of something or the tallness of a man. In all these cases, education, fashion and global knowledge play a crucial role in making decisions. Fuzzy logic, which allows possibility values to range between 0 (false) and 1 (true), has been introduced to handle such types of vague concepts. The notion of possibility and probability distribution is central to the development of integrated decision systems. Note that, each possibility value also ranges in the interval [0,1], but their sum could be greater or less than 1. The meaning of possibility depends on the experimental situation. For example it could be related to the accuracy of each sensor or it could be determined by the user on the basis of his experience. Some of these aspects are discussed in Zavidovique's chapter.

It is hard to compare natural and artificial decision processes, because of the gap still present between the two physical systems. Natural systems have a distributed decision apparatus that maps in a many to many relation, namely perceived data set into goal space. Artificial systems, in their most common implementation, are mapping much data to a single goal (i.e. many to one relation). We emphasise once more that this limit is due to the present hardware limitation that does not allow fully distributed architectures equipped with active sensors.

Perception and Action

According to some experts a robot can be called intelligent when it is able to react to external stimuli in real time. The ability of industrial robot systems to analyse visual scenes allows action to be planned starting from different operative environments, as well as in unpredictable situations: e.g. the case where a grasped component is accidentally lost, sensorial information on the location of part can guide the recovery of the component and complete the goal. The interaction between perception and planning allows auto-

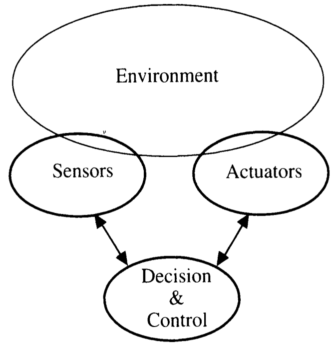

Figure 1. The perception-planning-action functional loop.

adaptation, even with limited resources. However artificial systems are still far from the high level performance of human beings.

In the current technology, three main functional aspects are outlined (see Figure 1): i) interaction with the external world, that could even include human beings; ii) carrying out decisional functions and control; iii) acting on the external world.

The perception of external surroundings is supported by sensorial channels. The information gathered, combined with other information possibly supplied by human interaction, is made available to the decisional block. During the accomplishment of a complex task very often unexpected situations may occur that must be resolved. In such cases the predictive models that provide long range planning, and operate with representations at the higher level of abstraction are not sufficient. They must be conjugated with the reactive models to guarantee the survival of the system by operating directly and immediately on the environment. These models operate with representations close to the raw data, with extreme timeliness.

The comparative analysis of the final objective and the external conditions, allows the intelligent kernel to update a temporal plan to complete the assignment. Each unforeseeable event may prevent the activity from being properly carried out: the sensoriality must detect the new situation, and must consent new planning, thus allowing the objective to be achieved in the altered external conditions. The perception-planning-action cycle which practices an adaptive and intelligent control on the autonomous system, is closed through the external world.

In planning the activities of the decisional block, it is important to point out that besides knowledge of the environment received through the sensors, meta-knowledge must also be made available, that is the ability of the system to know its own functionality. Moreover, the auto-adaptive behaviour of artificial systems cannot only be based on the knowledge provided to the system at the design stage. Remarkable technological attention must also be given to learning strategies (see the Becker panel and Becker-Zavidovique's dialogue) aiming at automatic run-time acquisition of the knowledge necessary for the current application. There are important examples of artificial systems (in nature this is guaranteed by plasticity as discussed in the Vallerga panel) that automatically improve their throughputs by a tuning process on the workload, or even systems able to generate autonomous operational strategies on the way.

An important approach in building mobile systems is based on the use of "autonomous agents" or modules, each carrying a sub-goal as a part of the common goal that the robot has to achieve. An intuitive example, is the approach to data fusion which employs a set of independent sensor neural nets, one for each sensor, coupled to a fusion net.

It is interesting that many of the tasks that are difficult for humans can be tackled easily by computers and vice-versa. Perception and motion-control, that we perform without any effort (see Gauthier's chapter), are in fact more complex and difficult to embed into a computer than many other more cognitive tasks. Indeed, evolution took several millions of years to perfect sensory perception mechanisms, while only some dozens of thousands of years were needed to improve conscious thought and language.

Perception and Representation

Both natural and artificial perception systems maintain various levels of internal representation of sensorial data. In both cases three different stages are envisaged (see Andreani-Dentici, Gärdenfors and Gaglio). The corresponding levels are characterised not only by the modality of the internal abstraction of external sensorial events, but also on the basis of functional characteristics, as well as on the temporal persistency of the representation itself. Nevertheless, decisional processes are applied to the three levels. These are usually pre-attentive in the lowest level and attentive in the highest.

The first level is called physical code or retinotopic representation in psychology and in the computer vision respectively (there are equivalent descriptions in the other sensorial modalities: echoic, tactile, somatosensory, etc. memories). At this level, direct reference to sensor neurobiology (Andreani-Dentici) is maintained; the level of abstraction is limited and the memory involved is consistent. This is the reason why these sensorial memories are the most rapidly decaying. From the functional point of view, in nature, this is a "pre-attentive" level. In artificial implementation it is also characterised by automatic control. As in all pre-attentive stages, a high degree of parallelism is exploited. Reactive controls (reflexes in biology) are applied, but usually without interfering with the other sensorial modalities and without having the ability to interrupt or avoid the execution of these early automatic stages of analysis. Due to the large amount of data, the complete information gathered is maintained only for these basic computations, soon passing to higher levels of abstraction which consent strong data compression. In biology too, the physical code is preserved only for a few seconds (1-2 s).

The second representation level corresponds to the naming code in psychology and the intermediate, feature representation, sub-symbolic level in sensorial artificial processing. This corresponds to an "attentive" stage, with a higher processing time due to the serial nature of this type of information manipulation. The control is no longer automatic, but extended flexibility allows the reaction to be adapted to the content of the data and to the events which involve the processes, and then to drive behaviour acording to the acquired attentive observations. At this stage much interference exists between the different sensorial modalities with which external information is collected; these cooperative interferences are essential in operational artificial systems. The serial process modality, typical of the attentive stage, requires extended timing which in biology takes up to 30 s before decaying and losing the collected information. This activity in biology is located usually in the so-called Short Term Memory (STM).

The third stage corresponds to the semantic code or to symbolic and functional representations, respectively in natural and artificial systems. This is a conceptual, linguistic level, in which the knowledge acquired through the sensors is associated with internal knowledge acquired by skill and experience. The high level of abstraction suggests possible solutions for knowledge representation, adapted to the current goals (the purposive representational approach in computer vision). The abstracted representation, following judgement and consciousness, can be added to the Long Term Memory (LTM). This complicated process which involves beliefs, credibility, importance, plausibility, reliability, supports, etc. is discussed in the contribution by Castelfranchi.

It is worthwhile noting that representation stages are directly integrated with the functional modalities of the behaviour of the system even in the early cases (reactive-reflex comportment) and obviously in the highest stages with predictive and purposive acts.

Perception and Communication

To communicate information to a user in a more effective and expressive way, advanced computer interfaces combine the characteristics of different media and modalities such as graphics, natural language, animation, sound, etc.. The considerations presented here, about the usage of multiple media and their organisation, do not pretend to be exhaustive: communication often requires many senses (the same medium may be perceived by more than one sense which differs from the others in lexicon, syntax and pragmatic) (see Levialdi and Mussio in particular). However, here and in Rossi's chapter and in the Levialdi panel, we will refer mainly to the workstation modalities such as text-graphics, pictures, and only partially sounds (see Mastronardi's contribution). Other topics discussed were the following:

- *inclusion of a new medium.* The inclusion of a new medium within a communication can be exploited to achieve different ends (see P. Duchastel, Illustrating instructional texts, Educational Technology, 1978): to capture attention (attentive role); to support information understanding (explicative role); to improve information retention (mnemonic role). In the active management of the interface, the inclusion of a new medium plays an important role, helping to drive the user's attention towards the currently relevant information. When something changes on the screen, the attention is focused on the changing 'loci'.

- *inclusion of icons.* According to Levin (see J.R. Levin, On functions of pictures in prose, Neurophychological and Cognitive Processes in Reading, F.J. Pirozzolo and M.C. Witrock eds., Academic Press, 1981), the insertion of an icon can assume different roles, specifically: to insert a redundancy in the information transmitted by recalling the same ideas in a different form (representation role); to illustrate the relationship between ideas and to show their structures (organisation role); to supply analogies and visual metaphors to increase understanding (interpretation role); to adjust the information to make it easier to be memorised (transformation role), in reality "Mnemonic" transformation has been adopted since the Egyptian era to elicit the perception of movement; to add aesthetic elements for the pleasure of the user (decoration role).

- *inclusion of text.* Images are generally considered to be self-explicative: information conveyed through images is generally more compact and directly accessible, and requires less effort to be understood. Unfortunately, this may lead to a superficial interpretation: text may then be important in elaborating the matter more deeply, so helping to retain a clearer memory of it.

- *media complementarity.* Complementarity is not only a matter of content: the same contents may produce different effects when transmitted through different media. Nevertheless, if the designer of a multimedia presentation selects two or more media using the same type of expression, the probability that what is said in the different media represents mere duplicates of each other increases greatly. On the other hand, by creating a seamless integration of multiple media using a different type of expression for each, the impact of the presentation will be strengthened.

- *media organisation.* The way in which different media are combined together may itself transmit specific information (see M. Gray, J.D. Foley and K. Mullet, Grouping and ordering User Interface Components, Technical Report Series, GVU Center, Georgia Tech., 1994). This is well known by graphic designers, who can benefit from significant literature about the visual composition of text and graphics (see Rossi's chapter and H. Lieberman, The Visual Language of Experts in Graphic Design, Proc. 11th IEES

Symposium on Visual Languages, 1995). Another design issue which is worthy of further investigation is the order of use of different media.

- *sequence effects.* An image which precedes a difficult text will generally make reading easier by providing a useful "platform" for the interpretation of that text where most of the ambiguities have been removed. Another important role for a figure which precedes a text is to provide a "criterion of analysis" for the text itself. Sometimes, instead, it is quite important to conclude a description with a picture which summarises and stores the concept previously introduced in the memory. This redundancy may be very effective if properly organised in connection with past and subsequent section contents.

- *representation grammars.* The considerations expressed so far refer to the organisation of different media within the same process as a way to transmit specific information. Even the organisation of a mono-media representation is very meaningful, as shown in verbal communication and music. A "representation grammar" explicits the set of rules and techniques to be used to properly accomplish knowledge transfer. Studies on iconic representations have been published recently (see W. Winn, The design and use of instructional graphics, Knowledge Acquisition from Text and Pictures, H. Mandl and J.R. Levin eds., Elsevier Science Publ., 1989).

Summarising, multimedia/multimodal interfaces may improve communication through more effectiveness, expressiveness and accessibility, provided that the different media are combined synergetically. Nevertheless, how to integrate the different media so that they play their intended role in exchanging information via computer is still a subject for research. We expect that research regarding these topics will increase rapidly, especially concerning the structure and the semantics of non-textual media used in hypermedia applications.

Acknowledgements

The workshop, and thus indirectly this book, was made possible through the generous financial support of the universities involved, and the research and industrial organisations that are listed separately. Their support is gratefully acknowledged.

The editors would also like to express their appreciation to the Program Committee of the workshop for their advice: Ornella Andreani-Dentici, Jörg D. Becker, Ivo De Lotto, Salvatore Gaglio, Peter Gärdenfors, Giorgio Grisanti, Walter Kropatsch, Stefano Levialdi, Brian Moore, Vito Roberto, Jean Pierre Rossi, Silvana Vallerga, Yehezkel Yeshurun, Bertrand Zavidovique.

Special thanks should go to Mrs. Clara Randazzo, for her precious help and patience in organising the workshop.

Virginio Cantoni, Vito Di Gesù, Alessandra Setti, and Domenico Tegolo

SPONSORING INSTITUTIONS

The following institutions are gratefully acknowledged for their contributions and support to the Workshop:

- Gruppo Nazionale di Cibernetica e Biofisica del C.N.R.
- Elsag Bailey
- Regione Siciliana
- Palermo University - Centro Interdipartimentale di Tecnologia della Conoscenza
- Pavia University - Centro Interdipartimentale di Scienze Cognitive

The Workshop has been held in co-operation with:

- Italian Presidency of the European Union - 1996
- ACM/SIGs: SIGARCH, SIGART, SIGCHI

CONTENTS

SENSORIALITIES

PERCEPTION AND INTEGRATION

PERCEPTION AND DECISION

PERCEPTION AND ACTION

PERCEPTION AND REPRESENTATION

PERCEPTION AND COMMUNICATION