Mapping on Demand - A Dream

Wolfgang Förstner

Autumn 2002

1 Goal

1.1 Motivation

Classical maps are used for decision making. They have been graphical representations of the earths topography or of some specific theme mainly using symbols. Their production took several decades, updating was done locally. Their production was motivated by military purposes. Users were trained to read the maps.

Yesterdays maps are graphical representations of the content of a Geoinformation System (GIS) either on paper or digital displays. Photogrammetry enabled to manually produce and update maps within a few years. Maps became multi-purpose maps, serving a variety of users within society.

Todays maps are visual representations of the content of GIS integrating annotated images and videos. Remote sensing and semiautomated photogrammetric techniques allow to produce and update maps within a few months. Visualization of processes can be based on powerful simulation and computer graphics tools and be produced in *hours or days* and still require skilled users.

As many planning processes took weeks, months or even years, real time requirements could be fulfilled with current technologies, containing only little automation.

1.2 The Dream

Tomorrows maps are audio-visual representations integrating *annotated* images, video, speech, sound. They are used for disaster management, for vehicle control, for guidance of tourist with an intelligent personal digital assistant, for geo-scientists and others.

They need to be *generated on the fly* depending on the *specifications of an unexperienced user* and possibly need to transparently transfer the desired information by audio and video to the user.

They have access to various *modes* of available data: Existing maps, especially all pre-interpreted information already contained in Geoinformation Systems, but also intensity-, color-, range-, -images, image sequences, possibly in corresponding data bases or distributed in the Internet. In all cases it will be useful to go beyond pre-interpreted data and make all hidden information avaliable.

1.3 Embodiment

The generation and use of maps always were embodied into a decision making process referring to some spatial region within a certain time frame. The control loop always contained humans, who specified the purpose of the maps, who specified the legend the map, who generated the map and used the map, often as an interactive planning medium.

In contrast to seeing machines, where the control loop between the visual system and the actors might we very short, the *embodiment of the specification, the generation and the use of maps* needs to model the human as actor to such an extent, that the automatic generation and presentation of information is *adapted to the human cognitive abilities*, obviously depending on the training of the user.

Though this might reduce the speed requirements in terms of reaction times, the problems to be solved are similar to visually guided machines. On the contrary, due to the possibly longer response times fusion of data of various modes becomes necessary, increasing the complexity of the models for problem solving.

1.4 Technological and Scientific Situation

The foci of future maps are, current car navigation systems already containing some of these features:

1. limited transmission rate

   This will only gradually change in future. It may overcome in case of *high-level coding of the semantics*, requiring a cognitive approach

2. limited display resolution

   This will certainly be overcome in the next decade

3. multi-medial output (images, video, sound, audio)

   Technology will allow any type of integrated presentation in the future. The main problem is the transfer from information to audio-visual output such that the human perception deviates minimally from the desired intention.

4. untrained users for specification and reading of maps.

   There are two steps where *cognitive abilities of the user* need to be modeled:

   (a) the *specification* of the desired output
   A natural language interface would be desirable. In the most simple case the user could select from a set of pre-specified options. In a far more advance case the user would specify his intentions, which then would be transferred into requirements for the desired information and mode of presentation needed.

   (b) the *generation* of the audio-visual output
   In contrast to the cognition of natural scenes, the human visual system is optimized for, we have to *model the ability of the human visual system with respect to artificial visual stimuli.* Especially the visualization of non-visual information, usually *realized by symbolization*, needs to be modeled.

5. short response times, down to a few seconds.

   In case of pre-interpreted information this is state of the art.

6. *integration of un-interpreted image data*

   This is probably the most challenging problem and central for the cognitive vision part of the dream.

2 Scientific Issues

2.1 Context modeling

There are several places where modeling context is a central issue:

- the context in *existing maps* (GIS or other spatial data bases) is both:
  - related to the legend, i. e. to the data model (in data base terms)
  - related to the spatial neighborhood, e. g. when identifying a road crossing in a large scale map, where road crossings are no basic entity
- the context of *image related partial interpretations* similarly requires the world and image model for interpretation together with the spatial neighborhood relations.
- the specific task defines the context for the information fusion process.
- the *type of viewer* influences the way the various contexts, i. e. neighborhood relations between different spatially related facts is visualized.

Main problems appear to be:

- the definition or the learning or the contexts
- the adequate representation of the contexts
- the efficient use of the context within the real-time task in concern
- finding useful observables triggering control based on context

2.2 Spatio-temporal Multi-scale Representation

One can assume that tera-byte of images are available, both terrestrial and aerial, both intensity and range. Spatio-temporal multi-scale representation is a central issue. It appears at several places:

- Terrestrial images show an extreme range of scales already within one image
- Aerial images occur in scales between 1:1000 to 1:10000, satellite imagery shows pixel sizes between 1 m and several hundreds, the scale however does not change too much within one image.
- image sequences contain events which are measured in seconds (crash), hours (rush hour) or even weeks (weather).

Scale can refer to many modes:

- numerical scales such as map or image scale, resolution or smoothing width referring to the geometrical or physical extension of objects.
- ontological scales such as derivable from containment or abstraction hierarchies

Main problems appear to be:

- Finding adequate multi scale representations for spatial/temporal objects
- understanding the relation between numerical and ontological scales
- Finding useful observables in data related to one scale for objects having a different scale

2.3 Adaptive Control

Real time requests need efficient control adapted to the available sources:

- data (not all types of data will be available for every request)
- models and algorithms (they evolve over time)
- hardware (server and client; computer, screen, audio)
- time (depending on internal speed of process to be served, navigation of walking persons vs. driving persons vs. driving cars)

Main problems of the request for this type of scalability appear to be:

- An explicit model for meta data of available resources
- transparent, theoretically founded performance characteristics of all algorithms for task planning

2.4 Interactive Learning

Models certainly will have to be learned. This may be done off-line training the system. This is a problem in its own. But it *should use the knowledge of the users of the system*, e. g. in case a specialist on modern architecture is using the system as tourist guide. It could be seen as a type of life-long learning. *Mapping on demand gives an ideal situation of mutual learning of humans and machines.*

Main problems appear to be:

- The language for representing things to learn common to machine and human. This requires detailed *modeling of human related specification languages.*
- A seamless communication between machine and non-specialist. This requires detailed modeling of the human perception system w. r. t. seeing audio/visual annotated images or image sequences.

2.5 Cue Identification and Integration

The multi-modality of the available data requires an adequate identification of cues for the various categories to be identified. This can be seen as *finding observables.* These are realized as image analysis algorithms (taking image as a category here) which either detect certain objects or estimate attributes of certain objects.

Main problems appear to be:

- Find cues as relations between observables, realized as algorithms, and partial interpretations.
- Formally describe the value of the cues for initiating hypotheses. As the context of each of the different modes of data sources is different, homogenization of the representation of the various cues appears to be crucial.

2.6 Uncertainty Management

All data and informations sources uncertain. This refers to:

- The original data (statistical variations)
- the models for evaluating the data (simplifications, bias, phenomenological models in place of physically based models, assumed relations, ...)
- the specification (ambiguity of natural language, limitations of man-machine interface used for specification)
- ambiguities in the perception of maps as annotated images or image sequences.
- uncertainty about the performance of algorithms.

Uncertainty management is a prerequisite for reliable predictions, for self-awareness of algorithms and for trustful visualization.

Main problems appear to be:

- Representation of the various modes of uncertainty
- Modeling the uncertainty of user defined specifications
- Modeling the uncertainty propagation through reasoning chains with algorithms whose theory is understood only limited.
- Visualization of automatically derived uncertain categories, together with their uncertain attributes such that the *perception of an unexperienced user* is correct.