Chapter IV
Different Roles and Definitions of Spatial Data Fusion

Patrik Skogster
Rouaniemi University of Applied Sciences, Finland

ABSTRACT

Geographic information is created by manipulating geographic (or spatial) data (generally known by the abbreviation geodata) in a computerized system. Geo-spatial information and geomatics are issues of modern business and research. It is essential to provide their different definitions and roles in order to get an overall picture of the issue. This article discusses about the problematic of definitions, but also the technologies and challenges within spatial data fusion.

INTRODUCTION

Due to the rapid advances in database systems and information technology over the last decade, researchers in information systems, decision science, artificial intelligence (AI), machine learning, and data mining communities are facing a new challenge: discovering and driving useful and actionable knowledge from massive data sets. During the last decade, many researchers have also studied how to exploit the synergy in information from multiple sources. This phenomenon

includes terminology such as spatial data fusion, information fusion, knowledge (and/or belief) fusion, and many more.

TERMINOLOGY

Geospatial data has many definitions, but one point of view is that it is data consisting of geographical information, geostatistics and geotextual information. This theme was handled already in the mid 80's by Crist & Cicone (1984a). According to Crist and Cicone (1984a), geostatistics are data that is related to a national or subnational unit and can be georeferenced. Geotextual data are defined as text databases (like treaty databases) that are linked to some geographic entity. Crist and Cicone also (1984b) argue that data fusion is not just overlaying maps.

Information fusion is a term defined as "a formal framework in which are expressed the means and tools for the alliance of data originating from different sources" (Wald 1999). Wald continues (2000) that spatial data fusion is therefore "the formal framework that expresses the means and tools for the alliance of data originating from different sources". It must be remembered, though, that every definition always reflects the current subject. Wald´s (2000) focus is mainly on the prominent vision of remote sensing data, where discussion is about pixel fusion, image fusion, sensor fusion and measurement fusion.

The term "information fusion" can, in other words, be used when different information and data is used to solve problems. Locational data can be added to this information fusion context and the result is spatial data fusion. As Kim (2005) describes, "information fusion can be implemented at two different levels: raw data and intermediate data". Information fusion at raw data basically means taking advantage of the synergy from considering multiple of the same pattern (i.e., considering two temporal series based on different measurement systems), while information fusion at immediate data is to take the synergy from utilizing multiple patterns (i.e., utilizing both temporal and spatial patterns)(Hall & Llinas 2001). The information fusion at the raw data level becomes important for example when two different measurements have recorded the same activities or events (e.g., flood level or market share) at the same location on a regular basis (Vanderhaegen & Muro 2005; Pereira 2002).

Knowledge fusion is the process by which heterogeneous information from multiple sources is merged to create knowledge that is more complete, less uncertain, and less conflicting than the input. Knowledge fusion can be seen as a process that creates knowledge. Knowledge fusion can also involve annotating the output information with meta-level information about the provenance of the information used and the mode of aggregation (Hunter and Liu 2005; Hunter and Summerton 2004).

Spatial data fusion is a combination of the above mentioned with the dimension of spatiality. It is by definition an enormous and complex field, comprising issues ranging from registration and pixel-level fusion of data for improving the spatial resolution of managerial decision level fusion by using previously computed information stored in geographic information systems (Malhotra 1998).

TECHNOLOGIES WITHIN SPATIAL DATA FUSION

It has been estimated that up to 80% of all data stored in corporate databases may have a spatial component (Franklin 1992). To support analytical processes, today's organizations deploy data warehouses and client tools such as OLAP (On-Line Analytical Processing) to access, visualize, and analyze integrated, aggregated and summarized data. The term "multidimensional" was established in the mid-1980s by computer scientists who were involved in the extraction of meaningful