NEEDS AND APPLICATIONS FOR DATA MINING IN LARGE SERIES OF REMOTELY SENSED IMAGES

Wietske Bijker

International Institute for Geo-information and Earth Observation – ITC
P.O. Box 6; 7500AA Enschede; The Netherlands
Phone: +031-534874203; e-mail: bijker@itc.nl

ABSTRACT

We see that several large suppliers of remote sensing images have opted to make images available at very low cost or for free. A notable example was the switch from "maximum profit" to "maximum use" made in the data distribution policy of SPOT-Vegetation, which indeed led to a steep increase in its use as measured in the number of images downloaded per year. Changes in data policy can also be seen in a larger context of use of geo-information, where several researchers noted a switch from "market" to "polis" or from "maximum profit" to "maximum benefit for society". This kind of more open data policies fueled research on large image archives, including image mining techniques. This research was also supported by society, as increasing visibility of remote sensing images and derived products increased awareness of the potential of remote sensing to help solving questions about state and change of our planet.

In data mining of large series of images, there are basically four generic types of searching (mining) which are used alone or in combination for an application: a) Presence and location; b) Temporal patterns; c) Spatio-temporal patterns; d) Moving objects. In the search for presence and location of a certain feature, the feature is usually defined on a spectral basis, criteria on size and shape can be added. A typical example of this kind of mining is detection of forest- or wildfires from images with high temporal and low spatial resolution. These fires are usually detected by applying a spectral threshold on one or more thermal bands. Another example is the detection of oil spills at sea from radar images, based on backscatter and texture.

For the second type of search, detection of a temporal pattern of a known object, the focus is on the temporal variation of pixel values or derivatives (indices), either per pixel or for the group of pixels forming an object. A typical example is the study NDVI patterns over cropland or rangeland, to assess for each pixel or field whether at a certain moment the vegetation condition is above- or below average, or whether there are any trends in vegetation condition over the years (for degradation and desertification studies). The third type of mining considers spatio-temporal changes in a known object with known location. Monitoring changes in the extension of a lake or the extension of a glacier can be studied in this way. Where the two previous types were mainly pixel-oriented, the focus is here on the changing extent (size and shape) of an object. The fourth type deals with moving objects, like clouds, icebergs, or ships. Depending on the object of interest, the object may also change shape and/or size during movement. For these types of data mining, the algorithms exist and many are already used operationally and (partially) automated. Main issues for the users of the data are continuity of image supply, longer than the lifetime of a particular platform/sensor combination, geometric accuracy and radiometric calibration. These issues determine the time period over which image mining can be performed and the accuracy of the information.

While algorithms to retrieve information by for image mining are already well-developed, models that describe uncertainty in mining of large image archives are still largely under construction. Much work has been done on description of spatial distribution of uncertainty, and the first steps have been set to extend these models into the time domain, first for archives of images from the same sensor, recorded at regular intervals. Then, these models can be extended to archives with different types of images, recorded at irregular time intervals, which will lead to increased heterogeneity of uncertainty in space and time. For each part of the study area and each time span, uncertainty will depend on (varying) spatial and temporal resolution, detectability and speed of the process. When we are able to describe uncertainty of a certain image mining study in both space and time, the next step should be to find good and meaningful ways to communicate these uncertainties together with the results of the image mining to the general public.