MERGING OF HETEROGENEOUS DATA FOR EMERGENCY MAPPING: DATA INTEGRATION OR DATA FUSION?

Florin Savopol and Costas Armenakis

Center for Topographic Information
Geomatics Canada, Natural Resources Canada
615 Booth Str., Ottawa, Ontario, Canada K1A 0E9
fsavopol@NRCan.gc.ca ; armenaki@NRCan.gc.ca

Commission IV, WG IV/7

KEY WORDS: integration, fusion, heterogeneous data, emergency mapping

ABSTRACT:

Many terms are used to name and define these data operations: “fusion” and “integration” of geospatial data or “integration (or fusion) of digital images and geospatial information”, as well as “revision (or updating) of geospatial (or topographic) information (or data bases). The present paper will try first to delimitate the use of these terms in the context of the research work done for the CIT-O (Centre for Topographic information – Ottawa, Natural Resources Canada).
In an emergency situation the authorities in charge of mapping support will face two major challenges:
1) to deliver ‘immediately’ up-to-date existing topographical information showing the situation before the emergency occurs (position of existing roads, bridges, community facilities, strategic buildings, etc.);
2) to get as quick as possible digital images from the disaster area in order to understand and monitor the situation, to evaluate the damages and the risk for injuries or more damages and to support the rescue operations.
To meet these challenges there is a need to deal with a range of heterogeneous geodata consisting for example of various sources, geometries, scales, resolutions, types, accuracies and dates. In an emergency mapping situation, the choice of data sources to be integrated / fused could be limited and the user can be forced to use data and images with a resolution outside the normal limits. The present work evaluates the fusion of images with a significant difference in spatial resolution in the typical framework of an emergency mapping project. It also investigates the fusion possibilities of the various data with respect to their enhancement of feature interpretation and extraction as well as the integration of imagery with existing topographic data. Relations and criteria are established for the evaluation of the fusion processes, while certain relations can be established between the resolution of the imagery and the scale of the original cartographic product.

1. INTRODUCTION

In an emergency situation resulting from a major natural disaster or from another event (a major ecological or industrial accident, an act of war, etc) the authorities in charge of mapping support will face two major challenges:

1) to deliver ‘immediately’ up-to-date existing topographical information on both digital and classical (paper) support showing the situation before the emergency occurs (position of existing roads, bridges, community facilities, strategic buildings, etc.);

2) to get as quick as possible digital (eventually metric) images from the disaster area in order to understand and monitor the situation, to evaluate the damages and the risk for injuries or more damages and to support the rescue operations.

To meet these challenges there is a need to deal with a range of heterogeneous geodata consisting for example of various sources, geometries, scales, resolutions, types, accuracies and dates.

Starting from the experience accumulated during the effort for updating the National Topographic Database of Canada (NTDB) using satellite imagery, the present research investigates the possibilities integration and fusion of different kind of digital imagery with exiting topographic data in the context of emergency mapping.

The availability of cartographic data in digital form had encouraged users to manipulate, to merge and combine geospatial data from different origins, scales and content in order to solve theirs needs.

2. DEFINITIONS

Many terms are used to name and define these data operations: “fusion” and “integration” of geospatial data or “integration (or fusion) of digital images and geospatial information”, as well as “revision (or updating) of geospatial (or topographic) information (or data bases).
Some definitions are:

On images fusion:
..”image fusion is the combination of two or more different images to form a new image by using a certain algorithm” Van Genderen, Pohl (1994).

On fusion of information in general:
The data fusion is ...”set of methods, tools, and means using data coming from various sources of different nature, in order to increase the quality (in a broad sense) of the requested information” Mangolini (1994).

US DoD (Wald, 1998) uses more general definition without reference to the quality:

“data fusion is a multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation and combination of data and information from multiple sources”

Wald uses also the improvement of quality al primary scope of the fusion operation. He has introduced the “conceptual framework” in the definition of the fusion:
“data fusion is a formal framework in which are expressed means and tools for the alliance of data originating from different sources, in order to obtain information of greater quality”, Wald (1998).

Also, regarding the quality he states:
“Here the quality has not a specific meaning. It is a generic word denoting that the resulting information is more satisfactory for the <customers> than before performing the fusion process. For example, a better quality may be an increase in accuracy, or in the production of more relevant data.” Wald (1998).

Some authors state that data fusion is the result of input from at least two different sensors. Some accept also the result of two different data sets, event if these are generated from the same sensor.

In general, integration / fusion aim to achieve the following:

1.- To enhance and improve the overall quality and reliability of the final output to support better geospatial operations.

2.- To enhance the information content of the final output to support better decision-making operation.

3.- To augment insufficient data and information to support better geospatial representation and to achieve better solutions.

To satisfy the requirements of the above objectives we see that the operations to bring diverse, distinct or separate elements into a unified whole can be applied to either methods or to data or to tools or to any combination of these. The blending mode of operations uses terms such as integration, merging, combination and fusion, which often are interchangeable in practice. We will try to make an effort to clarify these terms in the realm of geomatics and put them into some perspective.

Starting from the input space with at least two different inputs we mathematically map them to the output space. Depending on the mapping operation and mapping output we propose the following use of terms:

If the mapping operations or the output of them result in a separable type of output, where the individual characteristics of each of the input elements are preserved then the terms “combination” or “integration” may be the more appropriate ones (Fig. 1).

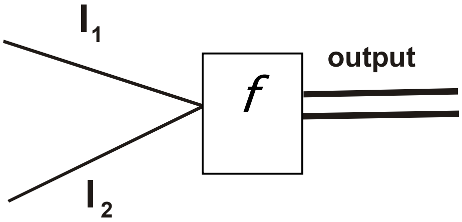

Figure 1. Combination / integration

If the mapping operations or the output of them result in a unified blended type of output, where the individual characteristics of each of the input elements are not preserved, then the terms “merging” and “fusion ”may be the more appropriate ones (Fig. 2).

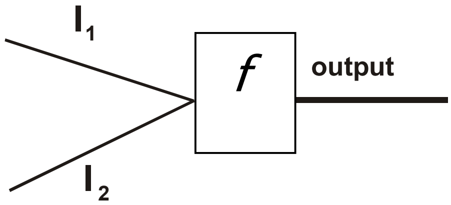

Figure 2. Merging / fusion

3. INTEGRATION OR FUSION?

The increased use of Geospatial Information Systems (GIS) for new thematic mapping application is pushing the need for quality, up-to-date data ready for end-user consumption. Historically, in mapping applications we have been working with heterogeneous data. Map compilation from aerial photographs is a combination of using control points determined by geodetic means and the analogue recorded images on film. In this case we see that the image data is augmented by the introduction of the control points to provide the spatial reference. Today we see the use of GPS and INS measurements to provide the information needed for the exterior orientation. This is an example of data integration, which provides the positional and orientation elements for each image. For cartographic purposes we see examples of using one set of updated features as a reference to correct the geometry of the rest of the features.

If we look now at the operation of image classification, the input sources are the various multispectral channels and the output is a layer of thematic classes. This merging of input sources to an entire new entity is considered as data fusion.

Another example of data fusion is the pan sharpening operation, which takes advantage of both the high resolution panchromatic image and lower resolution of the multispectral image. The result of this operation is an entire new enhanced multispectral or synthetic image and is considered as a typical data fusion process.

The National Topographic Data Base (NTDB) was created by Natural Resources Canada from the National Topographic Series (NTS) maps at 1:50 000 and 1:250 000 scales. The 1:50000 can be considered as the general, base scale in Canada. The original reprographic material was scanned, vectorized and the resulting data was structured in order to produce the NTDB. An important number of these NTS maps were already aged at the moment of conversion and the need of updating the NTDB became soon evident. Different methods were used in the past decade for the updating of NTS maps and NTDB files in the frame of research projects, experimental production and partially in normal cartographic production. Many data source were used as aerial photography, orthophotos and ortho-images from SPOT, IRS and Landsat imagery. Let us examine the case of one of the currents method in use: Landsat 7 images are converted into ortho-images using DEM derived from existing NTDB files and control points extracted from the original aerotriangulation. A fusion operation is performed using the panchromatic Landsat7 image (15 m pixels resolution) and the multispectral channels of Landsat 7 (30 m pixels resolution). The resulting multispectral (synthetic) image is used partially for classifications (water, forest) and partially for interactive change detection and updating process perform by superimposing the old (NTDB) vector information over the new, enhanced “synthetic” multispectral ortho-image.

Because of the important difference between the resolution and precision of the original images used to compile the map and the resolution and precision of the new imagery, some categories of objects (themes) represented on the old map and in NTDB file are difficult or impossible to detect /identify on the Landsat 7 imagery. Consequently, they are not subject of the updating process.

The resulting, updated NTDB file and the derived printed map have an enhanced quality as a result of the updating process. The updated NTDB file and the derived printed map offer a better representation of the terrain reality than the old one. Some of the represented objects have been extracted from the new Landsat 7 imagery and others from the original aerial photography. Is the resulting (updated) NTDB file the result of a fusion process between the old digital topographic data (or data base) and the new Landsat7 imagery?

As a new improved cartographic product this can been seen as the result of fusion operation. On the other hand, from the database point of view the characteristics of the elements are preserved, therefore it can be seen as an integration process.

In other cases only updated layers are replaced in a database, leaving the rest of the information as it was. This is the case when only the roads from a NTDB file are replaced (“updated”) by the data obtained from the Road Network data (RN). This DB is obtained by direct measurements using GPS methods. This is clearly a case of data integration.

One of the most popular approaches or products is the superimposition of vector data over image data (Figure 3).

As both inputs are put together by the process of “addition” without any change in their properties, this is consider the result of data combination or integration and not as data fusion.

Figure 3. Example of vector / image data integration by superimposition

4. CONSIDERATIONS

When integration or fusion occurs at the data level a number of factors need to be considered. These aspects include the type of datasets (vector-vector, vector-raster, raster-raster), scale and resolution of datasets, co-registration and the temporal decorrelation.

Merging vector data of different scales both the database (e.g., content, attributes) and the cartographic generalization aspects must be considered. If the scale of the output is the same with the scale of one of the input datasets, then we may have a case of data assimilation. If not, then the output is the result of data merging. In both cases the result of this data conflation is a new fused product.
In vector-raster case we usually deal with the superimposition of old vector datasets over newly acquired ortho-imagery. The relationship (ratio) between the scale of the vector data versus the image resolution needs to be considered. For example, what is the impact in spatial sense of superimposing small scale vector data over high resolution image data.

In the raster-raster cases the ratio of pixel size between the high resolution panchromatic image and the low resolution multispectral image is important. It may result in having erroneous spectral data in the resulting high resolution synthetic image, leading to classification errors.

In every of the above case, proper co-registration of the two data sets is a prerequisite. Examples of fusion between a high resolution panchromatic image and a low resolution multispectral image are shown in figure 4 and figure 5.

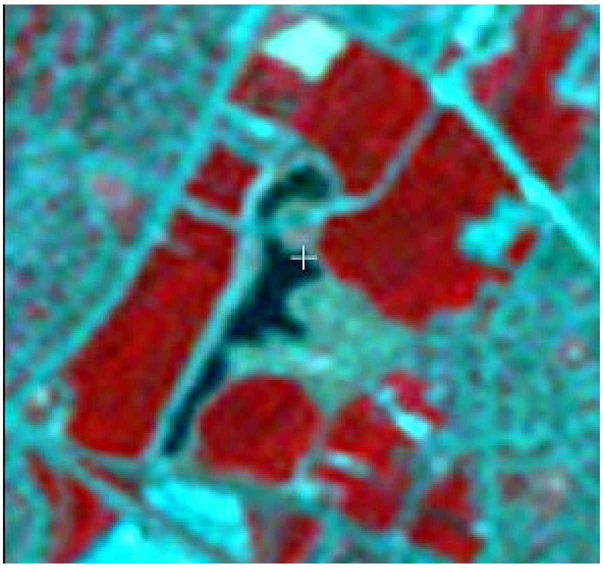

Figure 4. Original IRS multispectral image

Figure 5: Fused image

5. EMERGENCY MAPPING

In an emergency-mapping situation, the choice of data sources to be integrated (or fused) could be limited and the user may be forced to use cartographic data and imagery with a resolution outside the normal accepted limits. But how these 'normal limits' are defined?

In this case, it is useful to distinguish between cartographic data as compiled, edited and symbolize data and imagery.
In the first case, the 'normal' limit in resolution (or scale) difference could be considered the moment when the magnitude changes produced by the cartographic editing process and the symbolization of one of the data sets reaches a critical level at the scale of the data set with the higher resolution. If the mismatching is significant, this integration operation is to heavy, impossible ore will produce distorted, unacceptable results. Figure 6 shows the case of the integration of a more recent 1:50 000 scale NTDB file with an old 1:250 000 scale NTDB file. The discrepancies produced by the cartographic processes are evident.

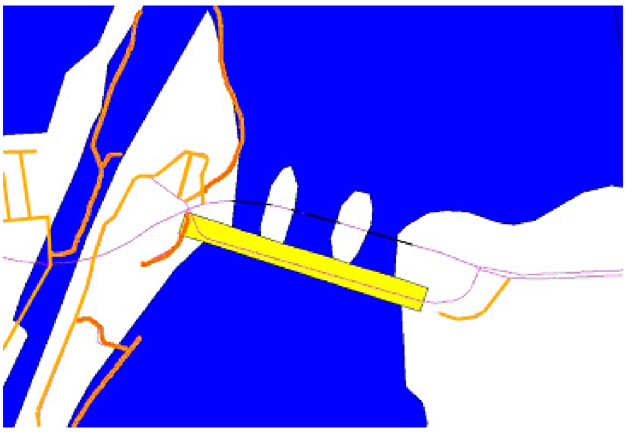

Figure 6: Example of problems produced by scale differences between two vectors maps (courtesy Jean Brodeur, NRCan).

This type of problems could be present in a situation of emergency mapping when, i.e., there is a need to integrate topographic data from the federal government with local / municipal large-scale vector data in order to offer recent, detailed and more complete information to the user. In both situations, it seems that we cannot talk about a data fusion situation.

In other situations, it is required to integrate a high-resolution digital image (raster image data) with an existing, eventually aged map (vector data). Typically, the high-resolution image will offer recent, up-to-date general topographic information or, in the case of an emergency situation, it will show the results of an accident, a natural disaster or an act of war. This will help to understand the situation, to evaluate the damages, to detect the potential hazards and to plan rescue and protection operations.

We can distinguish between several situations:

I- Similar scales. In an ideal situation, the resolution of the digital image should be in the same order of magnitude with the accuracy and resolution of the vector map (data consistency). These can be considered as a rule for the limit in scale difference. At the first stage. the image can be oriented using control points selected using a correlation between details points easy to identify on the vector map as road intersections and the corresponding point on the image. By superimposing the vectors over the image, the vectors representing each object should coincide with the image of that corresponding points in the image. More images could be oriented and assembled together as a mosaic. Is this 'mosaic' the result of a fusion or a merging process?

A more elaborated photogrammetric product, as a rigorous ortho-photo could be produced under normal circumstances. This high resolution, new ortho-photo could be augmented be superimposing selected features from one or from several existing (older) vector maps. We have in this case an <image-map> type product. In some situations, it could be important to have more information about a extended surrounding area. If the existing vector maps are not updated, it should be necessary to update these maps (what is described in the first section of this paper as a 'type 1 tasks' for emergency mappers). The specific emergency situation and the critical time limitations will not permit to use a normal cartographic updating process. A solution is to use any available recent imagery, as medium or high resolution satellite images. The conditions are that these images are significant more recent than the existing vector maps and their resolution meets the specific requirements. Finally, we my have a complex cartographic product that includes: 3-D data visualisation, a multi-source image-map with images from two or more sources with significant differences in spatial and spectral resolution, and one or more selected vector maps. From the point of view of the user, it seems to be a fused product. Figure 7 shows an example of fusion for 3-D visualisation using Landsat 7 and IRS (panchromatic) imagery, CanMatrix (scanned NTS maps) 250K CDED data (DEM derived from NTS 1:250 000 scale maps).

II- Large scale difference. It is possible that the new digital image has a much higher resolution than the old vector map. This situation is favourable for the point of view of the team in charge with the emergency situation, such images offering an increased potential for objects identification and helping for other tasks (situation understanding, response planning, etc.). From the point of view of the cartographic team, some additional problems will emerge. The volume of image data will increase. The spatial representation of features will be different.

It will be more difficult to correlate control points between the vector maps and the images. If the higher resolution images are controlled using the low resolution map, there will be probably some misalignments between the digital images. This situation is expected and we can accept that these misalignments are inside the error "budget" of the vectors map. Assuming that there are no problems at the cartographic information level, this is more a question of "visual comfort" and aesthetic aspect.

We can further distinguish between 2 situations regarding the image sensor and the working environment:

II-a : If there is not enough information about the geometry of the sensor or no appropriate photogrammetric software is available, the result of mixing and geo-referencing the image date will have the precision of the cartographic product used for the control (as geo reference) or less due to relief displacements. That is much lower as one can be expect from images of this type. This is a case of data integration.

II-b : If there is enough information about the geometry of the sensor and modern photogrammetric tools are available, the initial precision / accuracy of the original vector map(or topographic DB) could be improved significantly. The control obtained from the vectors map (or topographic DB) will be used only for the absolute orientation of the new images but the internal, relative precision of the element visible on the images will have a higher precision, in the same order of magnitude than the resolution of the digital imagery.
From the users point of view in an emergency mapping environment, this is considered as a data fusion situation.

III - Image-to-image fusion. In cases where we deal with images of various resolutions and wavelength bands, the objective is to end up with output pixels consisting of the highest possible resolution and the richest multispectral information. Considering that usually emergency situations have a local coverage we may face the case to blend images with significant differences in resolution. For example, IKONOS with 1m resolution with multispectral IRS LISS 23m resolution (1:23 ratio factor) (Figure 8).

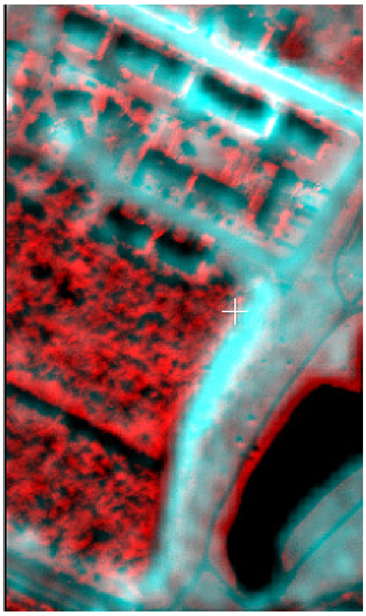

Figure 8: IKONOS (1m) with IRS LISS (23m) fusion

We can find examples, reporting ratio factor of up to 1:32 (Ersboll et al., 1998).

The large difference in resolution between the fused images creates a fuzzy effect. However, some additional thematic information can be extracted. For example the presence of vegetation (grass) can be seen on the right part of the image near the water body.

6. CONCLUDING REMARKS

In a complex, multi-source cartographic product it is difficult to establish very precise, absolute limits between fusion and integration type processes. If we consider the most accepted definitions (as Wald, 1998) and the improvement of quality from the aspect of a more relevant, more 'satisfactory' data for the user, most of the described products are considered as fusion. But on the other hand, from the database point of view, if the characteristics of the original elements are preserved, most of them can be seen as the result of an integration process.

Considering the differences in scale and resolution of the original geospatial data no precise limit can be imposed. It depends very much on the properties and type of the final product. It is seen as a general use mapping product as a printed map, more restrictions should be imposed in order to preserve the homogeneity of the product.

In an emergency mapping environment, the limits should be more flexible in order to meet critical time requirements. Further theoretical studies are required to understand and validate empirical results.

7. REFERENCES

Ersboll. B.K. and all, 1998: Fusion of SPOT HRV XS and Orthophoto Data Using a Markov Random Field Model - Proceedings of the Second International Conference Fusion of Earth Data, Sophia Antipolis, 28-30 January 1998.

Mangolini, M. 1994: Apport de la fusion d'images satellitaires multicapteurs au niveau pixel en télédétection et photo-interprétation . – Thèse de Doctorat, Université de Nice – Sophia Antipolis.

Ranchin, T., 2001: Data Fusion in Remote Sensing: Examples. – Proceedings of 4th Annual Conference on Information Fusion, Montréal, 7 – 10 August 2001.

Van Genderen, Pohl, 1994: Image Fusion, issues, techniques, and applications. Proceedings of the EARSeL workshop on intelligent image fusion, Genderen J. van and Capppellini Eds.

Wald, L., 1998: Data Fusion: A Conceptual Approach for an efficient Exploitation of Remote Sensing Images - Proceedings of the Second International Conference Fusion of Earth Data, Sophia Antipolis, 28-30 January 1998.

Figure 7: Example of fusion for 3-D visualisation (courtesy John A. Ells, NRCan).