Segmentation quality evaluation for large scale mapping purposes in Flanders, Belgium

F.M.B. Van Coillie [a], N. .A. F. Van Camp [b], R.R. De Wulf[a], L. Bral [c] and S. Gautama [d]

[a] Laboratory of Forest Management and Spatial Information Techniques, Faculty of Bio-science Engineering, Ghent University, Belgium – (frieke.Vancoillie, robert.Dewulf)@ugent.be
[b] Flemish Geographical Information Agency, Gent, Belgium – nancy.vancamp@agiv.be
[c] Department of Geography, Faculty of Sciences, Ghent University – lander.bral@ugent.be
[d] Department of Telecommunications and Information Processing, Faculty of Engineering, Ghent University, Belgium – sidharta.gautama@telin.ugent.be

ABSTRACT:

In Flanders the large scale reference database called GRB, takes care of the layout, exchange and management of large scale geographic information with respect to, amongst others, roads, buildings and parcels. As Flanders is extremely urbanized (average population density of about 450 inhabitants per square kilometer), the large scale maps need to be highly accurate. Currently, accuracies at the centimeter level are guaranteed due to topographic field measurements aided by standard photogrammetry based on analogue aerial photographs. In order to speed up the GRB production and to ensure large scale map products at the long term, it is essential to automate this labour-intensive, but highly accurate production process. Segmentation of very high resolution digital images could be an alternative approach for maintaining and updating the Flemish GRB as long as high accuracy segmentation results are obtained. Based on DMC images (8 cm ground resolution) and several reference buildings, a comprehensive sensitivity analysis is performed testing different segmentation parameter settings in order to gain insight into their impact on segmentation accuracy. The segmentation quality is evaluated using similarity measures focusing on aspects of presence, shape and positional accuracy where emphasis is placed on interpretability of the measures with respect to operational conditions put on the reference data. The end user should be able to read the measures and link this to the return-on-investment he will gain by using a given segmentation process on his data.

1. INTRODUCTION

1.1 The Flemish GRB

Flanders is characterized by a rapid and continuously changing environment mainly due to bustling economic activities e.g. inducing extension of dock areas or the construction of HST-railway beddings. Many initiatives are undertaken renovating the global traffic infrastructure: construction of roundabouts, speed ramps and expansion of cycle track networks are some examples. On the other hand, Flanders is also characterized by a lot of building, rebuilding and pull down activities: recent allotments, new industrial zones and major city renovation projects are everyday fare. The large scale geographic reference database, also called GRB, exactly contains data about these ground objects that could possibly change due to the above stated activities: buildings, parcels, watercourses, roads and their layout. Only terrain elements relevant for use within the scale limits of 1:250 to 1:1250 are considered (AGIV, 2004). Because of continuous and numerous ground alterations, the GRB reference database dates fast and consequently threatens to antiquate rapidly. Next to a high accuracy, the operational value of the GRB is precisely determined by its degree of topicality. As the GRB must represent a sustained answer to the continuous call for large scale maps, attention should be paid to an enduring production process.

1.2 GRB production: current status

The GRB is initially produced by a combination of photogrammetric and terrestrial techniques, whereas the update is usually still tallied by terrestrial measurements. Both production and maintenance are pricey but highly accurate processes. Based on experiences from home and abroad, the cost of maintenance and upkeep of a database with geographic data is yearly estimated at 15% of the total initial production cost. With an estimated initial production cost of 114 million euro, the update cost soon amounts 17 million euro on a yearly basis. Next to the high production cost, both processes are labour-intensive and consequently very time-consuming. To this day, about 28% (April, 2010) of the Flemish territory is totally mapped (AGIV, 2010).

1.3 GRB automation

In order to speed up GRB production and to ensure large scale map products at the long term, it is essential to automate this expensive and time-consuming production process. As large format digital aerial cameras become readily available and are continuously improving with regard to their geometric and radiometric accuracy, segmentation of very high resolution (VHR) digital images could serve as an alternative approach for maintaining and updating the Flemish GRB. Naturally high accuracy segmentation results are mandatory in order to live up with the high accuracy standards of the Flemish GRB (15 cm

planimetric accuracy for house fronts to 30cm for roads and 50 cm for an outbuilding).

In this paper segmented VHR DMC images (8 cm ground resolution) are evaluated on their usefulness for large scale mapping purposes. Hereby we rely on similarity measures evaluating segmentation quality introduced in earlier research (Van Coillie et al., 2008). The applied similarity measures focus on aspects of presence, shape and positional accuracy where emphasis is placed on interpretability of the measures with respect to operational conditions put on the reference data. In the current paper a comprehensive sensitivity analysis is performed testing different segmentation parameter settings in order to gain insight into their impact on segmentation accuracy. Based on the results we like to formulate an answer to the question whether image segmentation can be incorporated into an operational workflow of the GRB update process in Flanders, Belgium.

2. DATA

2.1 *Study site and imagery*

Aerial DMC images flown over the city of Ghent by Hansa Luftbild (Muenster, Germany) in March 2006 are orthorectified and used in this study. They are recorded on a height level of 796m and feature an end lap of 60% and a side lap of 80%. Together with the average ground height of 12m, the photo scale amounts to 1:6440, corresponding to a ground sampling distance (GSD) of 77mm.

2.2 *Reference data set*

For now we concentrated on one specific GRB entity i.e. buildings. Logically, the highly accurate large scale GRB product should serve as reference database. However, there is an apparent difference between the GRB gold standard (Zhang et al. 1996, 2008) and the ground truth that semantically comes closest to the image content of the segmentation. The distinct base material that is used in the GRB production process and the presented automation method are at the basis for that ground truth disparity. As mentioned before the initial GRB is produced by topographic field measurements (facades of houses, roads) aided by standard photogrammetry (completion of house polygons) based on analogue aerial photographs (scale 1:4000), while the presented automation procedure is an elementary segmentation of aerial DMC digital images (8 cm ground resolution, scale 1:6440). In order for image segmentation to qualify for maintaining and updating the GRB, we first evaluate to what degree segmentation results are corresponding to the ground truth semantically matching the image content the closest. At a later stage we will aim to incorporate the GRB as absolute gold standard. So in order to obtain semantically matching ground truth for the studied orthophoto, a reference set of 37 manually digitized buildings was created (a subset is visualized in Figure 1). Using this set as reference image, we made abstraction of large relief displacements inducing buildings to lean away from the image center.
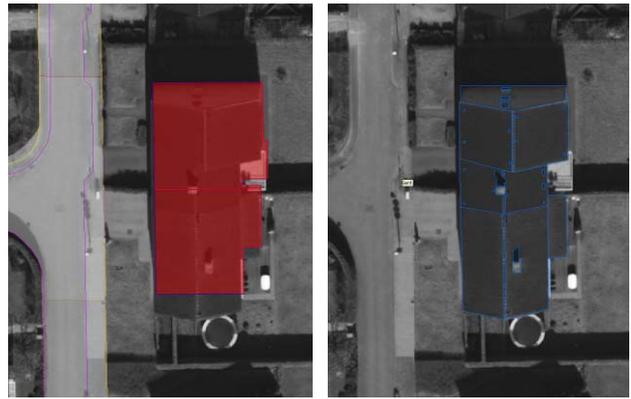
Figure 1: subscene of DMC orthophoto with left the GRB entities (buildings on the ground, outbuilding and façade) and right the manually digitized reference buildings.

3. METHOD

In order to investigate whether image segmentation could offer an alternative to the current GRB update process we should puzzle out if, and to what degree, segmentation parameter settings influence segmentation accuracy. Therefore a comprehensive sensitivity analysis is set up, measuring segmentation quality for a large number of parameter combinations (using the widely applied region-based multiresolution segmentation algorithm incorporated in the Definiens software (Developer 7.0)). To start with, only panchromatic images were processed. Scale was set to values of 10, 20, 80, 320 and 640. Shape varied stepwise (0.2) from 0.1 to 0.9 (with constant scale and compactness) and compactness also altered stepwise (0.2) from 0 to 0.9 (with constant scale and shape). Scales below 10 did not grow regions beyond pixel level; scales above 640 were not representative anymore for the GRB entity under study. The above parameter combinations were implemented in a process rule generating 155 segmentations with varying parameter combinations (scale, color-shape, compactness-smoothness).

The obtained segmentation results were exported to ArcGIS where a Visual Basic macro was written calculating seven different and straightforward quality measures assessing the discrepancies between the actually segmented and reference buildings while accounting for presence, shape and positional agreement with the reference data (Table 1, Figure 2) (based on Van Coillie et al., 2008)

Consecutively the seven quality measures are combined in a normalized weighted segmentation quality measure, DQM_{norm} in order to evaluate the impact of the parameter settings on the segmentation accuracy compared to the reference polygons (Table 2). Interpretation of the error rate per individual measure enables to pronounce upon the reliability of the segmentation results.

Table 1: Discrepancy quality measures (DQM)

DQM	Description
nr_centroid	Number of segments having their centroid in the reference polygon
diff_area	Difference in total area (%) (percentage of total reference area)
diff_per	Difference in total perimeter (%)
diff_ff	Difference in form factor (%) where form factor = perimeter²/4*π*area (average form factor of reference buildings = 1.77)
hist_1	Cumulative* at 1 pixels (%)
hist_2	Cumulative* at 2 pixels (%)
hist_5	Cumulative* at 5 pixels (%)

* For all segments having their centroid within a reference polygon, the number of pixels on the wrapping perimeter is determined. Next, the cumulative distribution per segmentation run (number of pixels in relation to the distance from the reference perimeter) is calculated.
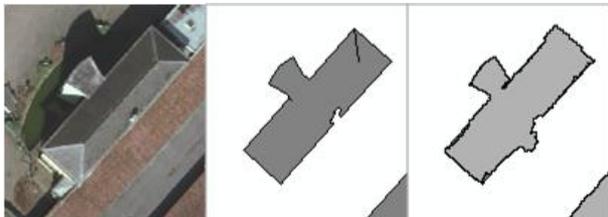
Figure 2: Left: the DMC image, middle: an exemplar reference building (1 out of 37), and right: an exemplar segmentation result.

Table 2: User defined weight optima and settings

	nr_centroid	diff_area (%)	
optimal	37	0	
weight	0.001	0.01	
	diff_per (%)	**diff_ff(%)**	
optimal	0	0	
weight	0.01	0.01	
	hist_1 (%)	**hist_2 (%)**	**hist_5 (%)**
optimal	100	100	100
weight	0.0005	0.0005	0.0005

4. RESULTS AND DISCUSSION

Over all 155 segmentation runs, on average the smallest DQM_{norm} (1.97) was reached with a scale setting of 20, a shape of 0.7 and a compactness value of 0.7. These parameters settings yield the segmentation result with on average minimal deviation from the 37 reference buildings. The total normalized quality measure thus provides a good indication of the preferred parameter settings that are usually selected by trial and error.

Next to an overall quality impression each individual component makes a valuable contribution in evaluating the segmentation quality. Table 3 shows the discrepancy measures for two exemplar segmentation runs: 1) de overall best run with scale, shape and compactness at values of resp. 20, 0.7 and 0.7, and 2) a sub-optimal segmentation with scale=80 and shape and compactness kept constant at 0.7.

Table 3: Discrepancy quality measures for segmentation runs with shape=0.7, compactness=0.7 and varying scale

scale	nr_centroid	diff_area (%)	DQM_{norm}
20	278	-2.84	**1.97**
80	115	-4.89	**2.31**
	diff_per (%)	**diff_ff(%)**	
20	-41.66	-48.63	
80	-63.94	-66.52	
	hist_1 (%)	**hist_2 (%)**	**hist_5 (%)**
20	29.85	45.70	65.50
80	25.20	42.10	57.90

With a scale of 20 there is on average 2.84% overestimation of the total reference area. Increasing the scale to 80 also induces an increase in area overestimation to on average -4.89%. Moreover using a larger scale is at the expense of average edge agreement: this is reflected by increased overestimations of perimeter (63.94% versus 41.66%) and form factor (66.52% versus 48.63%). Proportionally and on average, there are more wrapping perimeter pixels lying at a distance of 1, 2 or 5 pixels from the reference polygon perimeters (hist_1, hist_2, hist_5) using a scale factor of 20 compared to applying a larger scale of 80 (Table 3). This implies that, with a scale factor of 20, next to area, the edges/contours of the reference buildings are better respected (Figure 3). However, a scale to 20 yields considerably more but smaller segments (278 versus 111). This might be unwanted, because this could to a certain extent hamper the interpretability of the segmentation results. Therefore the number of segments has been taken up as a component of the weighted quality measure, DQM_{norm}. Consequently, although both area and edge deviation is smaller using a scale of 20 compared to 80, the respective DQM_{norm} values do not differ that much. It is thus the user who decides whether a big amount of resulting segments should be fined severely or not.
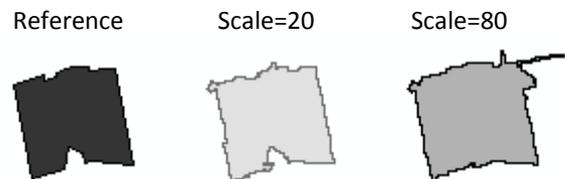
Figure 3: Segmentation results for 1 of the 37 reference building with shape=0.7, compactness=0.7 and varying scale

Although the entire result data set holds a lot of valuable information, only two segmentation results were highlighted and discussed (the best one and a sub-optimal result). It is up to the operator to interpret the individual components and weigh them according to his/her preferences. This methodology thus allows for incorporation of expert knowledge in the segmentation evaluation which is a great advantage in the context of the GRB maintenance and update. Although on average the errors are still too large to meet the high GRB accuracy standards, the results already point to an operational use of image segmentation when incorporated in the workflow of the GRB update process in Flanders.

5. CONCLUSION

The presented sensitivity analysis is based on seven simple measures of matching, each indicating the percentage of aberration from the ground truth (the reference). Overall quality estimation is performed using a normalized weighted combination of the individual quality measures.

Primarily, the sensitivity analysis provides insight on how parameter settings influence the number of resultant segments and how well these segments match the contour, area and shape of the reference. Consequently, based on a priori knowledge of the ground truth the best parameter set for segmentation can be selected. Furthermore, according to the user's preference (e.g. shape must be correct, or a minimum of segments is preferred rather than a fair shape) the weights of the individual discrepancy measures can be set.

Secondly, this method allows for defining to what extent the selected segmentation result is accurate compared to the reference. As such it seems even possible to apply this methodology in detecting changes in the GRB reference dataset, solely based on segmentation efforts. While in this paper manually digitized reference buildings were applied, the aim of a next step in this research is to apply the GRB as absolute gold standard.

As the trial and error stage in selecting suitable segmentation parameter settings can be bypassed, this method facilitates the incorporation of image segmentation into an operational workflow of the GRB update process in Flanders, Belgium.

ACKNOWLEDGEMENTS

REFERENCES

References from Journals:

Zhang, Y. J., 1996. A survey on evaluation methods for image segmentation. *Pattern Recognition,* 29(8), pp. 1335-1346.

Zhang, H., Fritts, J. E. and Goldman, S. A., 2008. Image segmentation evaluation: a survey of unsupervised methods. *Computer Vision and Image Understanding,* 110(2), pp. 260-280.

References from Other Literature:

Van Coillie, F. M. B., Pirez, R., Gautama, S., Van Camp, N. & De Wulf, R. R., 2008, Quantitative segmentation evaluation for large scale mapping purposes. Proceedings of GEOBIA2008, Pixels, Objects, Intelligence: Geo-Object Based Image Analysis for the 21st Century, 04-09/08/2008, Alberta, Canada (http://homepages.ucalgary.ca/~gjhay/geobia/Proceedings/)

References from websites:

AGIV, 2004. Nieuwsbrief GIS-Vlaanderen. Periodieke uitgave - maart 2004 - NR 17. http://www.agiv.be/gis/getDownload.ashx?id=85 (accessed 11 October 2009)

AGIV, 2010. Basisdocumenten GRB, http://www.agiv.be/gis/downloads/?SID=82 (accessed 02 February 2010).