THE EVALUATION OF SPATIAL DISTRIBUTION DENSITY IN MAP GENERALIZATION

Zhang Xiang[1,2,*], Ai Tinghua[1], Jantien Stoter[2]

[1]School of Resource and Environment Sciences, Wuhan University, 430072, China - xiangzhangchina@gmail.com, - tinghua_ai@tom.com
[2]International Institute for Geo-Information Science and Earth Observation P.O. Box 6, 7500 AA Enschede, the Netherlands – stoter@itc.nl

Commission II, Working Group II/2

KEY WORDS: Cartography, Generalization, Computer Vision, Spatial modeling, Quality control

ABSTRACT:

Spatial distribution density is an important constraint in map generalization, which acts on a group of objects at meso level or a whole thematic class at macro level. However, the spatial distribution density is difficult to be formalized and evaluated, due to the lack of both common understanding and appropriate measurements. The paper proposes an object-oriented density measurement based on skeletonization of gap space so that density can be calculated with each object and its spatial territory, in contrast to other approaches. Using adaptive space partitioning technique, the measurement can be applied on network feature (e.g. road network, drainage network) or a group of map objects (e.g. buildings) to give insight into their spatial distribution density. The measure is then integrated into an evaluation process for the assessment of the preservation of the constraint. In the process, spatial distribution density is measured both on generalized objects and their corresponding non-generalized objects. Statistic analysis on the distance between target value and measured value is adopted for assessing whether the balance of object densities are preserved or not. Different types of evaluation of density are also discussed. At last, the paper evaluates the constraint on two test cases (building generalization) as a proof-of-concept.

1. INTRODUCTION

In GIS and Cartography domain, quality evaluation has always been an important aspect in map generalization, through which the satisfactory of generalization with respect to various map requirements can be validated. Traditionally, the quality of map generalization is assessed manually by cartographers. Nevertheless, it is a time-consuming and subjective process (Brazile, 2000), since the evaluation varies strongly from person to person according to their own knowledge. On the other hand, generalized databases are not free from errors (João, 1998), since the generalization operators have various effects on the quality of map. Further more, some errors are hard to be identified visually but are crucial to the consistency of geo-databases (e.g. topological relation between building and road). It is therefore worthwhile to develop automated methodologies for an objective and holistic assessment of generalization quality.

To date, automated evaluation of map generalization is still an open issue and comes to draw more nowadays attentions. A few studies have been undertaken on different aspects of evaluation (Jansen and Kreveld, 1998; João, 1998; Brazile, 2000; Cheung and Shi, 2004). Bard (2004) was the first one to address the whole process of the automated evaluation. In some of the studies, cartographic constraints were identified as an important step towards the formalization of the requirements of intended maps as in case of automated generalization. Evaluation of properties of map objects can be made at three different levels, namely micro, meso and macro (Ruas, 2000, Bard, 2004). The micro level describes properties of individual objects independent of other objects (e.g. position, orientation, size, shape). The meso level concerns the contextual information among grouping objects (e.g. topological relation, density). The macro level deals with information concerning the population and distribution of characteristics across a whole thematic class. The fact is that, evaluation problems at three levels were not equally addressed by previous research. For instance the legibility constraints on minimum size for one object and distance between two objects can be easily formalized and evaluated automatically, as well as some preservation constraints on topological relation, whereas the distribution and density constraints for example are neither formalized nor measured straightforwardly (Burghardt et al., 2007).

Spatial distribution and density (black/white ratio) constraints at meso or macro level are key characteristics for the preservation of relative order between micro- or meso-objects (Mackaness and Ruas, 2007). The evaluation of such constraints needs to be further addressed. Since firstly, the constraint has a higher priority than legibility constraints in some situations. For instance, imperceptibility of buildings is acceptable in case of high building density (AGENT, 1998). Here, spatial distribution density plays a role of identifying specific situation (context), which is helpful to the selection of operators and parameters. Besides, map requirements from textbooks address the point that the differences in density on scattered objects should be retained or stressed according to different thematic classes, in order to reflect the characteristics of the reality (SSC, 2005). For example, density differences in building clusters reflect more or less the distinction between rural and urban settlements. As the spatial distribution and density also belong to structural constraint based on the classification proposed by Weibel (1997), they can be used to facilitate spatial related applications other than map generalization, by providing enriched structural and semantic information to the comprehension of specific geographical phenomena. Borruso (2003) explored the capability of using density of road network to enable the definition of city center by highlighting the peaks of distribution.

In order to address the evaluation of spatial distribution density constraint, three major questions are open:

- How can we define the concept and implement the measurement of the constraint?
- How can the constraint be integrated into the evaluation process and be evaluated?
- How can we formalize the target value of the constraint?

In the paper, a novel definition of spatial distribution density is presented, after a review of related work. The corresponding object-oriented measurement is introduced which is based on skeletonization of gap space among objects. The measurement aims to be generic to all geometric types in order to measure their distribution density in a uniformed manner. The detailed information about the measurement is discussed in section 3. Section 4 mainly focuses on automated evaluation framework aiming at addressing the latter two questions mentioned above. After a discussion of the experiment and result, the paper ends up with a conclusion.

2. RELATED WORK

Several work has be carried out related to distribution and density measurements. Measurements concerning the characteristics distribution (i.e. the distribution of properties like semantics, size, and orientation within certain spatial scope) have been used for evaluating the preservation of characteristics at macro level, for example the size distribution, as used in Bard (2004) and semantic-distribution in Ruas (1999). The measurement that only deals with the distribution of characteristics is not capable of describing the spatial aspect. In case of density measurement, different results can be obtained at different spatial levels. Since the spatial partitions used for the measurement are very different in size. In the literature, 4 major spatial partitions at different levels can be identified for density calculation:

- Map extent or convex hull of all the objects
- Feature-based partitioning
- Grid-based partitioning
- Object-oriented partitioning

Density calculation based on map extent or convex hull is the roughest one and only makes sense at global level, which is not of much help to describe the spatial distribution across the map. Feature-based partitioning is used to compute differences in density between different partitions, and it gains deeper insights into spatial distribution density. For example, Ruas (1998; 1999) discussed the evaluation of block density change before and after generalization. Brazile (2000) also proposed an evaluation model based on the partition formed by road and street to reduce the computation effort. However, the approach faces two major problems: one is that there is no partition always available in every situation (e.g. roads and streets in some rural areas are not enclosed), and the other is that if the distribution of objects within a partition is important and easy to be observed, the feature-based partitioning and density calculation will fail to capture it. Grid-based partitioning for density computing was studied by several authors. The main idea of the method is to transform the density estimate into a continuous surface by dividing the space into raster cells. Jansen and Kreveld (1998) used arbitrary grid to evaluate the consistency of map generalization, based on a clutter function which shares some ideas of density calculation. As stated in Borruso (2003), one has to choose the size of grid upon which the density estimation will be influenced greatly. In order to avoid this problem, other strategies have to be looked for. There is also an object-oriented density estimate using buffer-based measures (Steiniger et al., 2007). The calculation of density takes into account the context of each object by buffering operation. Likewise, the approach also encounters the same problem as in grid-based approach - the choice of buffer size.

Up to date, few studies exist yet as to the evaluation of density constraint at meso or macro level, except for a framework proposed by Bard (2004). The author pointed out that in order to evaluate constraints at group level, n-1 and n-m (n > m) relations have to be dealt with by the evaluation methodologies if the data is generalized by aggregation or typification. This point will be addressed in the evaluation section.

3. OBJECT-ORIENTED DENSITY MEASUREMENT BASED ON SKELETONIZATION OF GAP SPACE

3.1 Object-Oriented Density Measurement

Traditionally, spatial distribution density is measured as the number of objects per unit, or cartographically as ink-to-paper area ratios in local neighborhoods (AGNET, 1998). But it is not easy to apply this definition into practice as the reasons discussed in section 2. If we have look at a map of rural buildings with heterogeneous distribution of density, it is easy to find that the higher density of buildings the smaller territory (space surround each building) they have, due to the competition of space. This observation reveals an intrinsic relation between density per object and its territory.

Based on the analysis, we propose to define the object-oriented density measure in such a way as the space involved in the territory per object. The object-oriented density takes local neighborhoods into account, so that the territory of each object can be regarded as an indicator of spatial distribution density in its context, i.e. a smaller territory implies a higher density in local neighborhood, and vise verse. The proposed definition on different geometric types will be slightly different. The calculation of density on different geo-types is formalized as:

- *Density(point) = 1/TerrArea(point)*
- *Density(line) = Length(line)/TerrArea(line)*
- *Density(poly) = Area(poly)/TerrArea(poly)*

In the above measurement functions, the measuring of linear and polygonal features takes into account the length or the area of objects, to provide higher precision to the measurement. Section 3.2 will address the technical issues on territory computation.

3.2 Territory Computation based on Skeletonization of Gap Space

Territory referred in the paper is formed by partitioning the gap space between neighbour objects. Among all partitioning methods, partitioning based on skeleton is the most frequently used one. This idea has been applied in many applications:

commercial sites' serving areas grow outward and their boundaries ultimately coincide (approximately) with Voronoi Diagram generated from those commercial sites (Ai and Liu 2001); For generalizing river system, watershed area is determined by the spatial competition model applying the Voronoi-like Diagram partitioning, and basin polygon of each river channel is thus obtained (Ai et al 2006); Skeleton is firstly generated and then used to determine where minor ridges and valleys exist (Gold et al 1999, Ai 2007); In aggregation of urban building clusters, each building is surrounded by a partitioning polygon based on skeleton which can be deemed as the territory of the building (Ai and Zhang 2007).

Due to the many situations coming along as computing territory on network feature, we decide to use road network as an example to illustrate how the Delaunay based skeletonization of gap space work. The generation of territory on disjoint points, lines and polygons can be seen as degenerated cases of network feature.

3.2.1. Constructing Delaunay Triangulation

Constrained or conformed triangulation is first applied on the line segments in road network. The basic idea is that no triangles are allowed to cross any segments. The triangles play an important role of modeling the proximity relation between objects in their local neighborhoods, by which the shared space between objects can be divided equally and territory of each object is then formed.

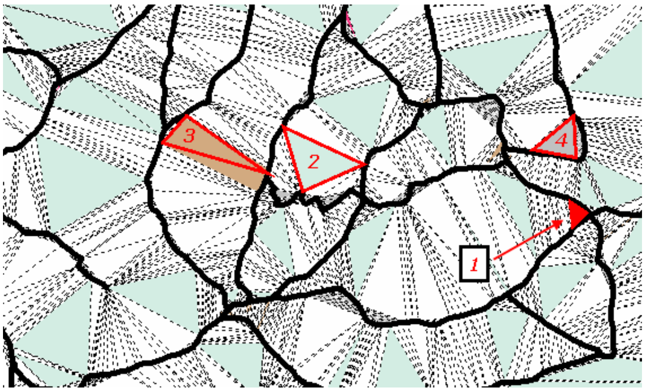

Figure 1: Delaunay Triangulation on road network feature and the classification of different types of triangles

Before carrying out skeletonization, classification of triangle is a necessary step to perform. 4 types of triangles are identified to be critical to the skeletonization process (also see Figure 1):

- *Type 1*: triangles that have at least one vertex on a junction point of segments, and the two triangle edges related to the vertex are rested on the edges of two segments respectively
- Type 2: triangles that have 3 vertices associated with at least 3 different segments
- Type 3: triangles that have at least one vertex on a junction point of segments, and at least one related edge is not on the edge of a segment
- Type 4: triangles that have 3 vertices associated with the same segment

The gap space among objects can be understood better with the help of the triangulation. According to the classification of triangle, two kinds of space can be recognized. One is the private space occupied by individual segments exclusively, which is marked with Type 4 triangles; the other one is the shared space by proximate objects which is formed by other types of triangles. Specifically, the triangles adjacent to each other enclosed by connected segments constitute a frame of gap space.

3.2.2. Skeleton Generation and Territory Computation

Every skeleton can be traced by walking through the triangle pass and portioning the space equally into two parts. The tracing process is briefly described as:

1> Start tracing from Type 1 triangle and end up at another Type 1 or Type 2 triangle;
2> Start tracing from Type 2 triangle and end up at another Type 2 or Type 1 triangle (tracing from Type 2 has three potential directions);
3> Split up the skeleton into 2 separate parts in case of passing through any Type 3 triangle (means it enters into the new territory of another segment).

By following the tracing process, a complete set of skeletons can be generated.

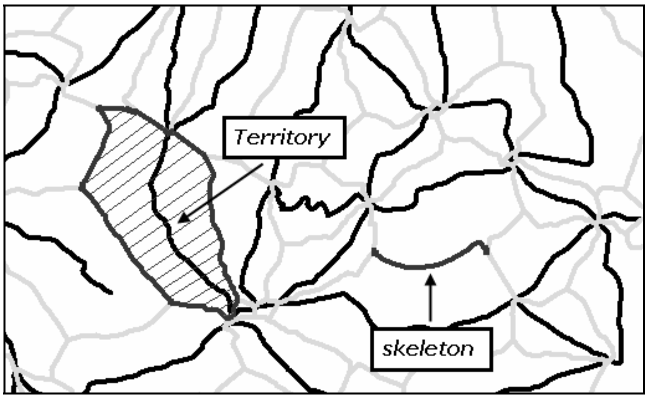

Figure 2: territory computation based on skeleton

Based on the generated skeleton, territory of each segment can be derived simply by connecting all skeletons surrounding the segment. To facilitate the constructing of territories, relations like topology are maintained between skeletons and segments during the tracing process. The derived territories and skeletons are illustrated in Figure 2. The territories calculated by our approach exhaust the region of interest, and no overlap between territories is observed. The approach is used for generating the territories of groups of points, lines and polygons as well.

3.3 Results of Object-Oriented Density Measurement

Figure 3: visualization of object-oriented density measurement

The object-oriented density measurement of points, lines and polygons can be calculated by the functions formalized in section 3.1 based on the generated territories of corresponding objects. The outputs of the measurement include the area of territory and density per object, as well as the visualization of the spatial distribution density (Figure 3 displays the density distribution of road network and buildings, the darker the higher).

4. EVALUATION PROCESS MODELING

4.1 Formalization of Evaluation Process

As described in Mackaness and Ruas (2007), a representation (map) is a reflection of the reality and it seems that the only reference of the evaluation is the non-generalized data, in order to check if the generalization retains specific properties of that reality. An assumption is made that the non-generalized data have an acceptable quality and that it reflects the reality at certain scale. In case of spatial distribution density constraint, the aim is to retain the balance of the density distribution. Thus the density distribution of initial dataset should also be analyzed.

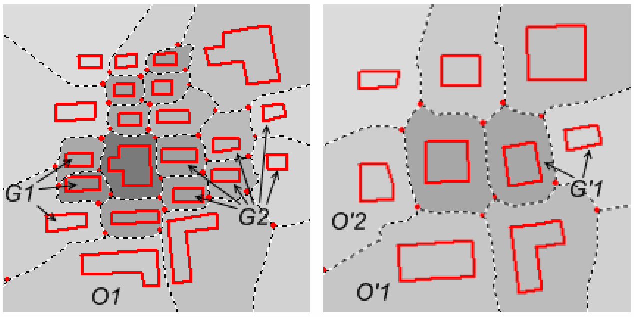

Figure 4: n-1 and n-m relations before and after generalization and the change of territories

As mentioned at the end of section 2, the object-oriented density evaluation has to deal with vertical relations between generalized and non-generalized objects. In Figure 4, non-generalized objects are on the left; the right picture represents the objects after generalization. Three kinds of vertical relations can be observed: through selection and simplification *O'1* is generalized from *O1*, which represents the 1-1 relation; *O'2* is aggregated from a group of objects *G1* representing the n-1 relation; by typification operation *G2* (6 objects) becomes *G'1* (2 objects), in which the n-m relation is maintained. The change of territories is obvious.

The proposed object-oriented density measurement means that the evaluation can be performed per object. In this sense, the density of each object becomes a character of the object. Hence, the evaluation can be treated as at micro level. Let us denote a non-generalized object as O_i. O_{gi} is an object generalized from O_i. $\{O_i\}$ and $\{O_{gi}\}$ are referred to as groups of objects and $\{O_{gi}\}$ are generalized from $\{O_i\}$. Measure-d() is the object-oriented density measurement function. As such, the density evaluation can be formalized as:

Evaluation = f (Measure-d(O_{gi}), Measure-d(O_i), *threshold*) (1)

Evaluation = f (Measure-d(O_{gi}), Measure-d($\{O_i\}$), *threshold*) (2)

Evaluation = f (Measure-d($\{O_{gi}\}$), Measure-d($\{O_i\}$), *threshold*) (3)

The evaluation functions are capable of assessing object(s) on 3 types of vertical relations. Function (1) evaluates the density changes on 1-1 relation; function (2) and (3) on n-1 and n-m relations respectively. Another problem is that, how can we get the density of a group of objects? In our research, the density is computed through aggregation of the density of every object in that group. The aggregation is formulated as:

$$meso\_density = \sum\left(d_i \times terrarea_i\right) / \sum terrarea_i \qquad (4)$$

where meso_density = density of $\{O_i\}$ or $\{O_{gi}\}$
d_i = object-oriented density of each object in the group
terrarea = area of the territory of each object

4.2 Formalization of Target Value of Density Constraint

The aim of the evaluation of density constraint is to preserve the balance or distribution of object densities across a map, which means that the order of object densities at target scale should be the same as at initial scale. The goal can be archived by specifying a target ratio of density change (*TRD*) before and after generalization process which should be followed by every object in a dataset. With this *TRD* the evaluation methodology can tell user which objects violate the density constraint, or to which extent it is violated.

However, it is difficult to formalize the *TRD*. Since user has no explicit knowledge of the threshold that can be specified directly for automated evaluation, neither do the experts. More over, the *TRD* cannot easily be deduced from some general principles. Töpfer and Pillewizer's Radical Law, for example, is able to give indicates of how many features to be selected during scale transition. It fails to answer how the spatial distribution can be maintained during the selection. Also because a map is not homogeneous in levels of detail, coefficients for the Radical Law that are appropriate for one part of the map are not likely to be suited for every part of the map (Buttenfield, 1991).

Our solution is that, we first compute a ratio of density change (*RD*) by dividing the density of target object by the density of its corresponding non-generalized one(s), then apply the calculation to every pair of measurements (on target and corresponding initial objects). At last we use the mean of all the *RDs* as a *TRD*, instead of struggling to figure out the *TRD* that are ground true. Without the absolute true value, the preservation of the relative order of densities across a map is also feasible.

The evaluation is performed by statistical analysis across all measured *RDs* in a dataset and their distance to *TRD*:

1> Deviation (distance) between each *RD* and *TRD*, denoted as $Dev(i)=|RD_i-TRD|$

2> Statistical deviation of all the *RDs* away from *TRD*, denoted as $StatDev = \sqrt{\sum (Dev(i))^2}$

Note that *StatDev* is different from the standard deviation in that the latter measure the distance of all variables to their mean, while in *StatDev* the *TRD* can also be other value than the mean value of all the *RDs*. Using *StatDev* analysis, the degree of the satisfactory of retaining the relative balance of densities distribution can be obtained. The *Dev(i)* is used to detect the exceptional violated objects by maximum operation or specifying a tolerance.

4.3 Types of Evaluation of Spatial Distribution Density

The proposed evaluation of spatial distribution density can be used for various purposes and scopes.

Purposes (as in Mackaness and Ruas, 2007):

- *Evaluation for editing*: detect high density area (objects) and inappropriate change in density distribution
- *Evaluation for grading*: compare different solutions by computing *StatDev* of them respectively, to grade the preservation of density balance of each solution

Scopes:

- Density change at global level (one value of all objects)
- Density differences of different spatial partitions by incorporating the partition features
- Density differences between different thematic classes via aggregating the object-oriented density of all the objects in each theme based on equation (4) (e.g. density differences of major, secondary, minor road and street classes)
- Density distribution across a map per object

5. EXPERIMENTS AND RESULTS

5.1 Test Cases

(a) (b) (c)

Figure 5: (a) initial data, (b) solution A, (c) solution B

We set up the test cases to proof the concept of the evaluation of spatial distribution density. In figure 5, the test cases consist of one initial data (a) and two solutions (b) and (c). Both solutions are generalized through selection, simplification, aggregation, displacement and typification. We do not apply any enlargement or enhancement operations. Clearly, the general target density in the cases is bound to be reduced meaning that *TRD* < 1. Generally speaking, generalized dataset A is better than B in case of preserving the balance of densities.

5.2 Evaluation Process

The object-oriented density measurement is carried out on objects at two datasets before and after generalization, using density = *Area(poly)/TerrArea(poly)* function. The density per object is obtained, and the measure unit is terrain unit (mm^2/mm^2). The measured density between linked objects (vertical relation) is joined together and the n:1 and n:m issue is done by *meso_density* calculation.

5.3 Results Analysis

In both target dataset, the results are presented as density pairs <INITDENSITY, TARGET DENSITY> and are stored in a relational table. And then, all the results are sorted ascendant according to INITDENSITY value.

Figure 6 shows the preservation of density distribution in solution A is better than in solution B. In the upper chart, the generalization has a trend that the density change on all objects is more or less the same. The lower chart shows that many objects have the same density before and after generalization. This point can be observed visually in Figure 5 (c).

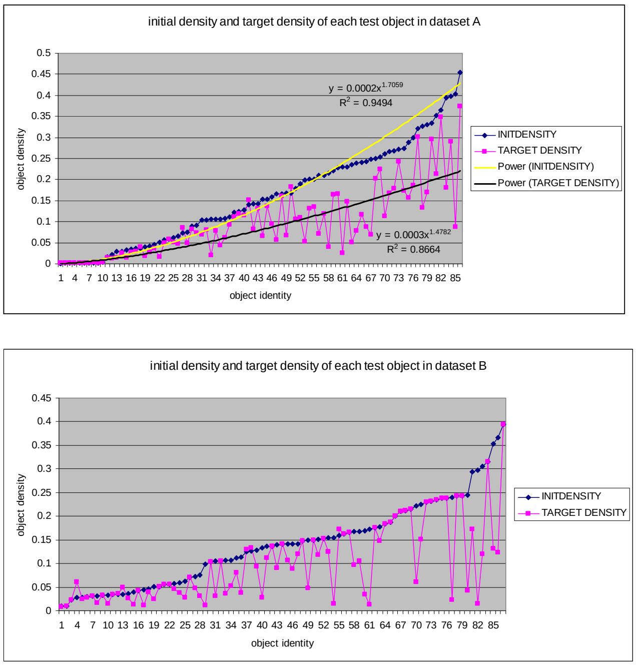

Figure 6: initial and target density comparison in solution A and solution B

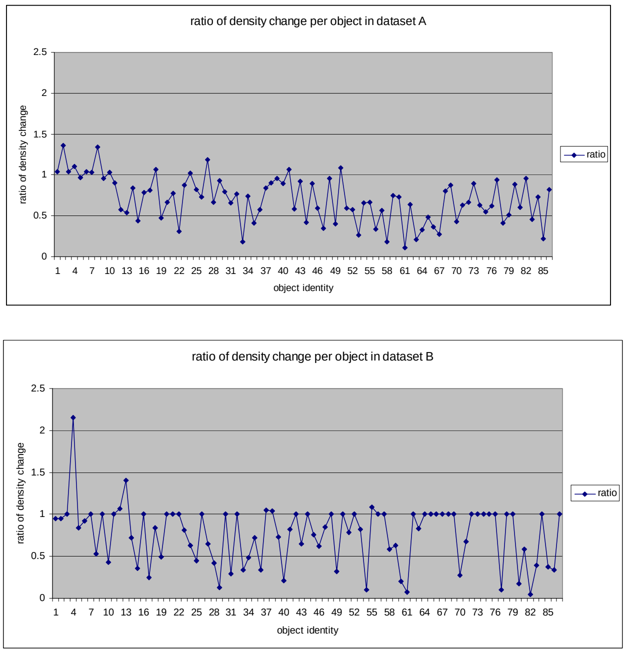

Figure 7: density change ratio in solution A and solution B

Figure 7 demonstrates from another perspective that solution A it better than solution B, since more *RDs* in solution A is lower than 1. In the lower chart however, many *RDs* equal to 1 or even higher which means that the density of these objects remains unchanged after generalization. Further more, *StatDev(A)*= 0.2759; *StatDev(B)* = 0.3510 reveals that all the RDs in solution A concentrate on their *TRD* more than in solution B. Because every variable in the above charts is related to an object in a dataset, it is easy to identify which objects in which dataset is inappropriate in sense of spatial density distribution. And this obviously facilitates the *evaluation for editing* discussed in section 4.3.

6. CONCLUSIONS

The paper developed an object-oriented density measurement to measure the distribution of density across a map. The result of the measurement is satisfactory as it is in line with the density feeling perceived by human. It is capable of measuring the densities distribution on network and groups of objects (points, lines, and polygons). But a limitation of the method shows that the mix up of different geo-types and density measure on them is probably problematic. It is because that the measurement functions used to describe density of 3 types of geometry do not incorporate with each other, although the territory extraction method supports this kind of application.

We carried out evaluation of spatial distribution density based on the proposed measurement. It proves that the measurement can be integrated into evaluation process and elaborates the question of how. The experiment results show the capabilities of this approach. But the formalization of absolute target value of the density constraint remains an open issue, which is hopefully addressed through training sample companied with reverse engineering techniques.

The further research and validation of the methodologies is of much need. Transforming the object-based representation of density into continuous surface is interesting and could be promising in providing more insights into the generalization quality. The reduction of computation consuming is also one of the greatest problems in practice. Moreover, an evaluation prototype is necessary to be developed in order to carry out holistic evaluation of various constraints.

REFERENCES

AGENT, 1998. Constraint Analysis. Deliverable A2.
http://agent.ign.fr/deliverable/DA2.html

Ai, T., Liu Y., 2001. Constraction of Street Center-line Network in Urban GIS – A Method Based on Delaunay Triangulation. *Proceedings of International conference of Urban Geo-informatics*, Surveying and Mapping Press House.

Ai, T., Liu, Y. and Chen, J., 2006. The Hierarchical Watershed Partitioning and Data Simplification of River Network. *12th International Symposium on Spatial Data Handling*, pp. 617-632.

Ai, T., 2007. The drainage network extraction from contour lines for contour line generalization. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62 (2007), pp. 93-103.

Ai, T., and Zhang, X., 2007. The Aggregation of Urban Building ClustersBased on the Skeleton Partitioning of Gap Space. *The European Information Society: Leading the way with Geo-Information.* Lecture Notes in Geoinformation and Cartography, Springer, Berlin, Germany.

Bard, S., 2004, Quality Assessment of Cartographic Generalisation. *Transactions in GIS*, 8(1): pp. 63–81.

Burghardt, D., Schmidt, S. and Stoter, J.E., 2007. Investigations on cartographic constraint formalization. ICA Workshop on Generalisation and Multiple Representation, Moscow.

Buttenfield, B., 1991. A rule for describing line feature geometry. In: Buttenfield, B. and McMaster, R. (eds): *Map generalization:*
Making rules for knowledge representation. Longrnan.

Borruso, G., 2003. Network Density and the Delimitation of Urban Areas. *Transactions in GIS*, 7(2), pp. 177-191.

Brazile, F., 2000. Semantic Infrastructure and Methods to Support Quality Evaluation in Cartographic Generalization. PhD Thesis, Department of Geography.

Cheung, C. K. and Shi, W., 2004. Estimation of the Positional Uncertainty in Line Simplification in GIS. *The Cartographic Journal*, 41(1), 37-45.

Gold, C.M., D.Thibault and Z.Liu., 1999. Map Generalization by Skeleton Retraction. ICA Workshop on Map Generalization, Ottawa.

Jansen, M. and van Kreveld, M., 1998. Evaluating the consistency of cartographic generalization. *8th International Symposium on Spatial Data Handling.*

João, E. M., 1998. Causes and Consequences of Map Generalization. Taylor & Francis.

Mackaness, W.A. and Ruas, A., 2007. Evaluation in the map generalisation process. In: Ruas, A., Mackaness, W.A. and

Sarjakoski, L.T. (eds.). *Challenges in the Portrayal of Geographic Information: Issues of Generalisation and Multi Scale Representation*. Elsevier Science, pp. 89–111.

Ruas A., 1998. OO-constraint modelling to automate urban generalisation process. *8th International Symposium on Spatial Data Handling* (SDH'98, Canada / Vancouver), pp. 225-235.

Ruas, A. 1999. Modèle de généralisation de données géographiques à base de contraintes et d’autonomie. PhD Thesis, Univ. De Marne La Vallée / IGN.

Ruas, A. (2000), The Roles Of Meso Objects for Generalisation. *9th International Symposium on Spatial Data Handling*, Beijing, 2000, pp.3b50-3b63.

Steiniger, S., Lange, T., Burghardt, D., Weibel, R., 2008. An Approach for the Classification of Urban Building Structures Based on Discriminant Analysis Techniques. Transactions in GIS 2008, 12(1), pp. 31-59.

SSC - Swiss Society of Cartography, 2005. Topographic Maps - Map Graphic and Generalization. Swiss Society of Cartography, Publikations Series, Nr. 17. Wabern.

Weibel, R., 1997. A Typology of Constraints to Line Simplification. In: Kraak, M.J. and Molenaar, M. (eds.): *Advances in GIS Research II (7th International Symposium on Spatial Data Handling)*. London: Taylor & Francis, pp. 533-546.