Updating authoritative spatial data from timely sources a multiple representation approach

Zhang, Xiang; Yin, Weijun; Yang, Min; Ai, Tinghua; Stoter, Jantien

DOI
10.1016/j.jag.2018.05.022

Publication date
2018

Document Version
Final published version

Published in
International Journal of Applied Earth Observation and Geoinformation

Int J Appl Earth Obs Geoinformation 72 (2018) 42–56

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Updating authoritative spatial data from timely sources: A multiple representation approach

Xiang Zhang[a,b,c,*], Weijun Yin[a], Min Yang[a], Tinghua Ai[a], Jantien Stoter[d]

[a] School of Resource and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
[b] Key Laboratory of Geographic Information System, Ministry of Education, Wuhan, China
[c] Key Laboratory of Digital Mapping and Land Information Application Engineering, National Administration of Surveying, Mapping and Geo-information, Wuhan, China
[d] Faculty of the Built Environment and Architecture, Delft University of Technology, The Netherlands

ARTICLE INFO

Keywords:
Volunteered geographic information
OpenStreetMap
Incremental update
Multiple representation databases
Data matching
Cartographic generalization

ABSTRACT

Integrating updates from timely sources such as volunteered geographic information (VGI) into the spatial data maintained at official agencies is becoming a more demanding requirement but presents many challenges. This paper proposes an approach to addressing the technical challenge of propagating updates from timely sources (e.g. OpenStreetMap) to spatial data maintained at separate map scales. The main idea is to establish a multiple representation database (MRDB) for datasets at different scales and time to facilitate incremental update, where linkages between corresponding objects at different datasets are made explicit. First, two ways in which the timely sources can be integrated into official data for incremental update are discussed. To derive the linkages between different datasets, a data matching procedure based on computer vision is presented and fine-tuned to match data in different scale ranges. Furthermore, the generalization history used to produce smaller scale data from the larger ones in official data is inferred based on the linkages, and is then used to guide the update propagation. Finally, a framework for incremental generalization in MRDBs is proposed, where crucial issues like strategies for update propagation, cartographic generalization, and the so-called 'chain reaction' are addressed. The framework is implemented as a fully automated process where operators like simplification, enlargement, compression, displacement and typification are incorporated into the incremental update process. By testing the framework against real world data sets (i.e. OpenStreetMap and official data at 1:10k, 1:50k and 1:100k), we show that the updates are integrated consistently into existing data in terms of spatial relations and cartographic quality. Our work suggests that making use of timely sources by official mapping agencies and companies in a continuous or event-driven data update is technically feasible, with further improvement and extensions discussed.

1. Introduction

Over decades, the acquisition, maintenance and update of geospatial framework data are in the domain of professionals such as national mapping agencies (NMAs) and companies. Although the update cycle of framework data in NMAs has reduced significantly, the update still relies largely on field survey, aerial photo interpretation and interactive generalization (Stoter, 2005; Stoter et al., 2009b) and therefore becomes the bottleneck for data/map production and services.

Recent years have witnessed the proliferation of volunteered geographic information (VGI) (Goodchild, 2007). OpenStreetMap (OSM) is one of the most prominent VGI projects to date and provides access to the open, free and up-to-date digital map data covering the world. In some countries/regions, OSM has received large imports[1] from professional data providers in its early days so that a basic level of data coverage and quality can be guaranteed. Moreover, OSM is constantly updated by large amount of volunteers worldwide and has grown to be a timely data source that covers a rich set of features and semantics (Heipke, 2010; Dorn et al., 2015). However, given its successful applications in the public and scientific domains (Goetz, 2012; Hagenauer and Helbich, 2012), the role of VGI in governmental sectors has not been properly identified until more recently (Haklay et al., 2014).

The major concern lies in data quality. Although comparing VGI with authoritative datasets for quality assessment is informative (Haklay, 2010; Girres and Touya, 2010), some researchers argued for the Linus's law principle from the perspective of big data (Haklay et al.,

* Corresponding author at: School of Resource and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China.
E-mail address: xiang.zhang@whu.edu.cn (X. Zhang).

[1] https://wiki.openstreetmap.org/wiki/Import

https://doi.org/10.1016/j.jag.2018.05.022
Received 2 March 2018; Received in revised form 25 May 2018; Accepted 29 May 2018
Available online 05 June 2018
0303-2434/

2010; Foody et al., 2015), suggesting a different quality assurance approach in VGI communities as compared with NMAs (Elwood et al., 2012). This might explain to some extent NMAs' concerns with VGI. On the other hand, some studies revealed that VGI is accurate enough for use by official agencies in map production (Parker et al., 2012; Olteanu-Raimond et al., 2016). For example, OSM shows at least comparable quality to authoritative datasets in terms of spatial accuracy and coverage in densely populated areas (Haklay, 2010). Besides, since most crowdsourced geographic data (e.g. OSM) have the Open Database license (ODbL), there are many restrictions in using such data.

Given the above issues, there is a recent trend for professional communities to consider the use of crowdsourced geographic information for production (Mooney and Morley, 2014). In a recent survey, Olteanu-Raimond et al. (2016) identified that VGI have been engaged in various degrees in European NMAs. Their survey shows further that NMAs such as Kadaster, the Netherlands and IGN, France have used OSM for change detection, which reduces work load and improves efficiency in traditional methods. This is perhaps in line with the adaptation of the Open Data Policy in many European NMAs (Olteanu-Raimond et al., 2016). Similar progress is reported in North America, where NMAs and companies have used, or plan to use, updates from VGI sources such as OSM instead of surveyors to speed up their production lines (Elwood et al., 2012). Begin (2014) for instance shows that, by providing data to OSM, Canadian NMA could therefore receive updates from OSM.

In this paper, we will discuss the use of timely sources in updating data sets maintained by official agencies from a technical point of view. As discussed previously, framework data are maintained at multiple scales and updated by interactive (manual) generalization. Our general question is therefore: with the timely sources available (e.g. OSM), how can we incorporate the changes and propagate them to data at different scales more consistently and efficiently? This is especially an issue when cartographic generalization has to be used and graphic conflicts need to be handled (Stoter et al., 2009b). Here we propose a multiple representation approach for incremental update of separately maintained datasets. Some concepts and related work is reviewed in the next section.

1.1. Concepts and related work

A multiple representation database (MRDB) is a spatial database that consists of datasets with different levels of abstraction (i.e. map scales), where multiple representations of the same real-world objects are linked by inter-scale connections (Kilpeläinen, 2000). *Incremental update* is a major application of MRDBs.

While incremental update is a more general term from software and database field, *incremental generalization* is used specifically for spatial databases, meaning the propagation of updates across different scales in an MRDB (Kilpeläinen and Sarjakoski, 1995). MRDBs and Incremental generalization are very attractive ideas for both researchers and practitioners (Devogele et al., 1996; Harrie and Hellström, 1999; Hampe et al., 2003; Haunert and Sester, 2005; Müller et al., 2012). However, MRDBs have seldom been implemented for production, especially when the linkage between representations is concerned (Burghardt et al., 2010).

As a key element in MRDBs, the linkage (hereafter referred to as *vertical relation* as in Bobzien et al., 2008) can be used to assess the quality, maintain the consistency between representations, facilitate incremental update, and is useful for multi-scale analysis and visualization (Stoter et al., 2009a; Burghardt et al., 2010). These relations can be complicated, since many-to-many correspondences are possible due to the use of cartographic generalization. For example, objects can be aggregated (resulting in a relation cardinality of *n*-to-1), deleted (1-to-0), or typified (*n*-to-*m*), where *n* is the number of larger scale objects and *m* the smaller scale ones.

The vertical relations can be explicitly modeled and recorded in the generalization process, with which the generalization history can also be stored (Burghardt et al., 2010). For instance, Zhou et al. (2009) proposed a model and prototype to log such metadata during the generalization to support incremental update. However, in the absence of such linkages (as in most software systems and NMA datasets), data matching can be used. As for update propagation in MRDBs, Haunert and Sester (2005) identified the issue of 'chain reaction' where an update may reform the linkage structure and hence more objects in the vicinity of the update are influenced. More recently, Müller et al. (2012) proposed to build an MRDB for OSM for potential applications in different domains, but this requires considerable efforts to generalize representations from a single database.

Our work differs from previous research and hence contributes in several aspects. First, we establish the MRDB and hence the vertical relations between data at different scales and time (i.e., OSM and official data at 1:10k, 1:50k, 1:100k) by data matching (Section 2.2). This fits better into current practices in NMAs. The matching is able to handle ambiguous situations caused by cartographic generalization (e.g. displacement and typification). Second, the generalization history of official data is extracted based on the vertical relations and is used to guide the update propagation at individual and group levels (Section 2.3). More importantly, we proposed a framework for propagating updates incrementally to datasets at multiple scales, where cartographic generalization and the so-called 'chain reaction' can be properly handled (Section 2.4).

As a proof-of-concept, we focus on building features in this paper and using OSM as a timely data source. Building footprints in OSM are shown to exhibit higher levels-of-detail than official datasets (Touya and Brando, 2013; Touya and Reimer, 2014; Fan et al., 2014). Also, when compared with our Top10nl (1:10k) we found that OSM buildings have greater details and is more current. Besides, building features require more efforts of generalization and hence pose a more challenging case for incremental generalization. After describing our approach in Section 2, the implementation and results are presented in Section 3 and discussed in Section 4. We draw our conclusions in Section 5.

2. Methodology

2.1. MRDB approach to incremental update of spatial data

2.1.1. An assumption of consistency for multiple representation databases

As a premise, we assume that all representations in an MRDB are logically consistent and synchronous in time. That is, data objects at different map scales represent the identical physical entities of the same time; no change exists between them. This is reasonable because smaller scale data were usually generalized from larger scale data. As a result, differences between data sets in an MRDB are due to map generalization no matter how significant the differences can be (e.g. Fig. 1). This is also a condition to guarantee a logical propagation of updates from larger to smaller scales. If the datasets in an MRDB are not synchronous, it is hard to know whether the differences were caused by the generalization or physical changes, and there is no way in which reliable vertical relations can be established.

2.1.2. Choice of scale in an MRDB for change detection

In an MRDB, changes occur to the data are only reliably detected by comparing the timely data source (e.g. OSM) with data of identical, or similar, map scales. If the scale or LoD of the data deviate considerably from one another, it can be hard (if not impossible) to distinguish physical changes from the discrepancies caused by map generalization.

This can be better illustrated with Fig. 1, where changes can be more reliably identified when the two data (e.g. OSM and Top10nl) are of similar scales (Fig. 1a; areas indicated by arrows are considered physical changes). When comparing OSM and top50nl (Fig. 1b), it becomes obscured if the observed differences qualify as changes or not, as the difference can be a result of map generalization or physical changes,

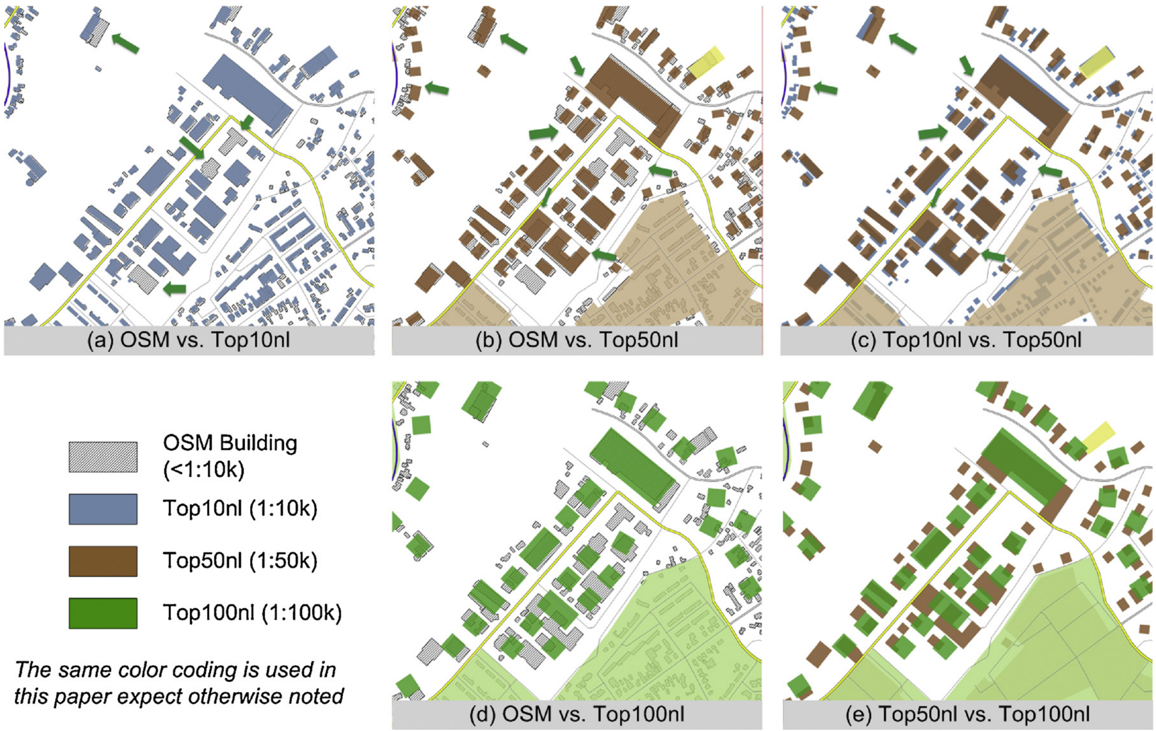

Fig. 1. Comparison between datasets at different scales for change detection: (a) comparison between data sets of similar scales reveals reliable changes; (c), (e) differences between datasets of adjacent scales in an MRDB are due to generalization (e.g. displacement, enlargement, aggregation and typification), and hence no physical change presents; change comparison is not reliable if the source and target data differ too much in scales (b), (d).

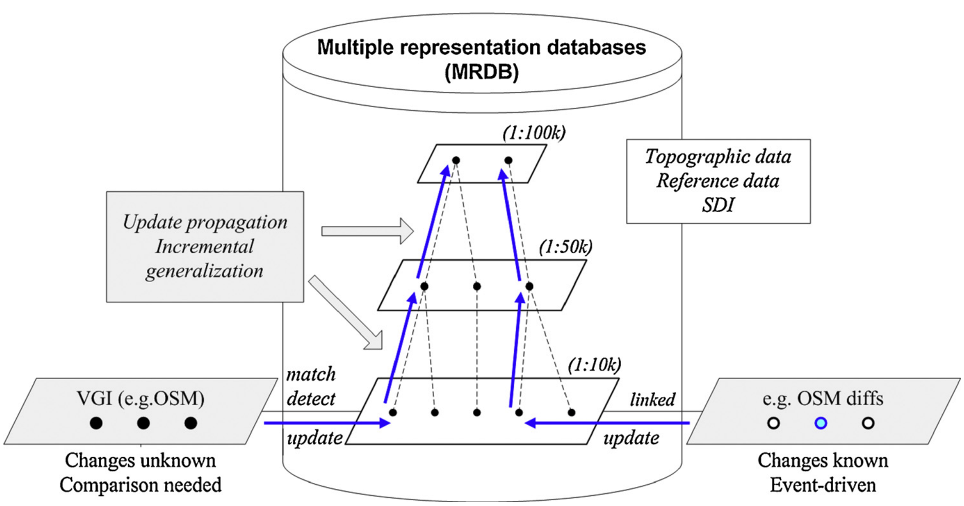

Fig. 2. Two modes of change detection in a schematic view: changes are normally detected by comparing the source (VGI) and the target (authoritative) data sets (left to the MRDB); if the VGI source is already linked/synchronized with the target data, changes are known every time user edits occur (right to the MRDB); in an MRDB approach, the detected changes are firstly updated to larger scale target data sets and then propagated via the vertical relations (dashed) to smaller scale ones (update propagation); scales shown in the brackets are for illustration only.

or both. However, since the datasets in an MRDB are synchronous in time, the differences between the two (Fig. 1c) are caused solely by generalization. For instance, buildings indicated by arrows were enlarged to snap the nearby roads, displaced or even shrunk to maintain a legible map at 1:50k. The problem becomes more intractable when comparing OSM with data of even smaller scale (e.g. Top100nl in Fig. 1d). Again we know that Top100nl is a generalized version of Top50nl and no change presents between the two (Fig. 1e). As a result, changes are better identified by comparing OSM with the MRDB data that has an identical or similar scale (e.g. the largest map scale). Then, by establishing the vertical relations between datasets in the MRDB (Section 2.2), initial updates could be propagated from larger scales to smaller scales (Fig. 2).

2.1.3. Two modes of change detection (snapshot-based vs. event-driven)

OSM has its own internal mechanism to maintain the history of user edits and to keep the data up-to-date. It hence provides a potential source for updating framework data. Specifically, the OSM project has a running database that accumulates user edits (additions, deletions, modifications to the geometry and attributes) in a continuous manner, so that users always get access to the up-to-date map. The latest data are made accessible to users every week. On the other hand, user edits are made available as small compressed XML files (known as Diffs) on a minutely, hourly, or daily basis[2], which are used as packets to incrementally update the OSM central database.

Hence, we can conceive two basic ways in which changes are identified and updated to the data sets in an MRDB. The first is what we

[2] https://wiki.openstreetmap.org/wiki/Planet.osm/diffs

call a snapshot-based approach, where the source data and data to be updated are two snapshots in time of the same geographic area. In this mode, we do not know which OSM objects have changed, and therefore comparison between the two snapshots has to be carried out to detect dissimilarities that qualify as changes.

On the other hand, if the MRDB database has already been synchronized with OSM and vertical relations established (e.g., after a cycle of the above-mentioned update), the subsequent update cycles can be carried out regularly, relying solely on the change sets delivered by OSM in the form of Diffs files (an event-driven approach).

As the use of OSM is still in its early stage in official agencies (Olteanu-Raimond et al., 2016), the event-driven approach where authoritative data need to be tightly coupled with OSM does not seem to be practical at the moment. The snapshot-based change detection is therefore a more natural choice.

2.2. A multi-scale database structure via feature matching

Reconstructing the vertical relations between objects of different scales is not straightforward. Due to the use of cartographic generalization, map objects can be eliminated, simplified, aggregated, typified and displaced. We use a hybrid strategy which is detailed as follows.

2.2.1. A hybrid feature matching strategy

From a practical point of view, there seems to be no single technique that is adequate for matching map objects at multiple scales. In this paper we will use two main procedures in different situations to establish the linkage structure for multiple representations: a relaxation labeling based approach (Zhang et al., 2014) and an overlap-based approach. A common dichotomy between individual buildings and built-up areas is used here as a guideline.

2.2.1.1. Relaxation labeling-based matching in ambiguous situations. The relaxation labeling-based matching is a computer vision technique and can be viewed as relational matching. That is, it finds the best possible correspondence pairs relying not only on the similarity between individual objects but also on the similarity between spatial relations, which are available in the spatial context of the object pairs to be compared. It usually gives a best match between two groups of objects.

Formally, we denote any smaller scale object as $i \in I$, larger scale object as $j \in J$, and some neighbors of i and j as $h \in N_i$ and $k \in N_j$ respectively. Any pair of objects $(i, j) \in I \times J$ can be classified as 'matched' or 'not matched'. The use of binary relations is in line with Bobzien et al. (2008). In our approach, the result is described by a matching probability $p_{ij} \in [0,1]$. The basic principle of the process is briefly described as follows (see also Fig. 3).

To determine if the pair (i, j) is a good match, the process quantifies to what extent the spatial relations between (i, h) are compatible/similar to those between (j, k), with h's and k's in the neighborhood of i and j respectively. Let us consider the relations (i, h_1) and (j, k_1) in Fig. 3a: their relative positions are similar (h_1 and k_1 lie relatively close to i and j; both are northeast to i and j); the orientation of h_1 in relation to i is compatible to that of k_1 in relation to j; i is a bit bigger than h_1, which is similar for (j, k_1). These relational similarities are quantified in the form of a so-called compatibility coefficient, $r_{ij}(h, k)$. This coefficient, when weighted by the matching probability p_{hk}, indicates the support that (i, j) get from (h, k). So, given the compatible relations, the matching of (i, j) is strongly supported by (h_1, k_1) if the latter is also a highly probable match (i.e. high p_{hk}). The support (positive or negative) is gathered from all the neighbors (e.g. h_1, h_2, h_3, h_4) of i to summarize a total support; the stronger the total support the more probably (i, j) can be matched. Eq. (1) gives a useful way to calculate the total support q_{ij}.

$$q_{ij}^{(t)} = \sum_{h \in N_i} \max_{k \in N_j} \{ r_{ij}(h, k) \bullet p_{hk}^{(t)} \} \tag{1}$$

The relaxation labeling is an iterative process. It starts with an initial matching matrix P, which can be populated randomly or by input from a simple matching method; in each iteration p_{ij} is updated using the following formula:

$$p_{ij}^{(t+1)} = \frac{p_{ij}^{(t)} + q_{ij}^{(t)}}{1 + \sqrt{\sum_{j=1}^{J} (q_{ij}^{(t)})^2}} \tag{2}$$

The denominator in Eq. (2) serves as a normalizing factor, so that the constraint $\sum_{j=1}^{J} (p_{ij})^2 = 1$ can be satisfied. The iterative updating of p_{ij} is dynamic since p_{hk} is updated as well and hence the support q_{ij} received by (i, j) varies each iteration. The process will converge to a consistent solution (i.e. all p_{ij} no longer change). That is, ambiguity in the matching is largely reduced with the help of contextual information. In our implementation, we used relative position, relative orientation, relative size, and relative shape (elongation in particular) between an object and its neighbors which are formulated as compatibility coefficient $r_{ij}(h, k)$. Final links are selected based on their matching probabilities (e.g. $p_{ij} \geq 0.5$). Note that the similarity between i and j has also a role to play in calculating the support, and Delaunay triangulation is used to define the neighbors of i (N_i). For a thorough discussion of this approach, its algorithm, quantification of the compatibility coefficient $r_{ij}(h, k)$, and many other options, one is referred to Zhang et al. (2014).

When reconstructing the MRDB structure, we found that the relaxation labeling-based approach suits better for the matching of individual buildings no matter if they are displaced, simplified, aggregated or typified during the generalization, i.e., in situations where matching ambiguity is high and many-to-many correspondences are present. Nevertheless, this technique is currently set to yield a maximum of 4 links for a target object[3]. Although it is possible to modify the normalization factor in Eq. (2) to accommodate more possible links, modeling arbitrary number of vertical relations as in the cases of built-up areas is not well understood. It may lead to undesirable side-effect. Besides, spatial relations that are essential to relational matching become obsolete in this situation. Hence, for large aggregated objects (e.g. built-up areas) which are normally aggregated from many larger-scale objects (giving n-to-1 relations, n > > 4), relaxation labeling failed to give satisfactory results. Nevertheless, an overlap-based approach to matching built-up areas with its larger-scale footprints will suffice due to the low ambiguity involved. Although it does not mean that the relaxation labeling method restricts to the matching of the same conceptual type but in practice, separating conceptual types would give better results in general.

2.2.1.2. Overlap-based matching in simple situations. In two simple cases we used overlap-based matching. One is the matching between data of similar map scale such as OSM and Top10nl (Fig. 9a); the other is the matching between a built-up area and its larger scale building footprints in the MRDB. The matching is based simply on the ratio between the overlapped area and the original areas, and if at least one of the two ratios is larger than a threshold (e.g. 20%), we identify (i, j) as a match pair. We find that in simple situations this strategy is sufficient to produce satisfactory results (e.g. Fig. 9).

2.2.2. Fine-tuning of the matching parameters for different scale ranges

The relaxation labeling-based matching should be parameterized differently for each scale range because relative relations are preserved in varying degrees for different scales. For instance, as most individual buildings are represented by squares at 1:100k, elongation of buildings becomes largely useless when matching between 1:50k and 1:100k data. The fact that which relative relations are still kept for which scales are not known a priori, so in our work the contribution of each relative

[3] According to the constraint $\sum_{j=1}^{J} (p_{ij})^2 = 1$ imposed in Eq. (2), and if we are to pick up links with $p_{ij} \geq 0.5$ as accepted matches, the maximum number of matches to object i is 4 (with each p_{ij} equals to 0.5)

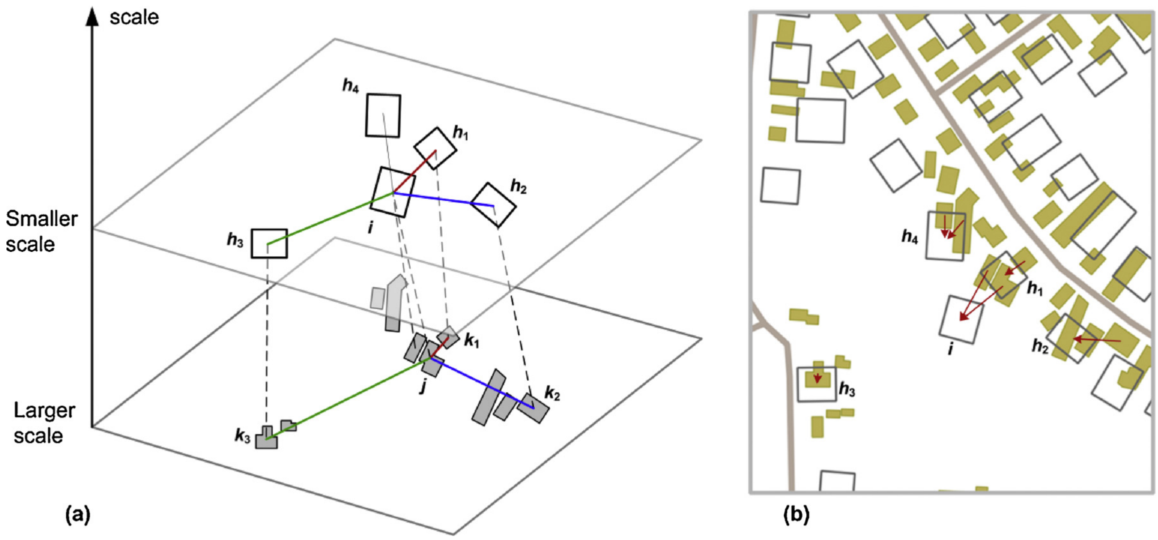

Fig. 3. Relaxation labeling-based matching: (a) the rationale illustrated (dashed lines are tentative corresponding relations to be verified and dynamically adjusted during the iterative ambiguity reduction process); (b) larger and smaller scale data being superimposed on top of each other and the final links obtained for the objects highlighted in (a) making use of relative relations available in the spatial context.

Table 1
Tuning of the relaxation labeling-based matching algorithm and its parameters for different scale ranges (data: TopNL data from 1:10k to 1:100k).

	Relative relations used								Parameter values used	
	Position		Orientation		Size		Elong.		Searching distance (d_r)	Min area ($Area_{min}$)
Scale range	rel	abs	rel	abs	rel	abs	rel	abs		
1:10k ~ 1:50k	×	×	×	×	×	×	×		60m	400m^2
1:50k ~ 1:100k	×	×	×	×	×	×			120 m	1600m^2

relations in the matching were determined empirically. The choices of the relations are shown in Table 1, which gave satisfactory overall matching results for TopNL data of the Netherlands.

Besides, the minimum area of buildings ($Area_{min}$) imposed in the generalization are also incorporated into this matching algorithm and is adapted here for different scale ranges.

2.2.3. The partition scheme in the matching

Another issue is that the partition scheme used in the relaxation labeling-based matching can affect its performance in several ways. First, a well-formed neighboring structure may obtain a better result (higher accuracy) as candidates to be matched are thereby surrounded by enough neighbors, providing rich contextual information to fertilize the matching approach. If there was no neighboring object presented, the matching probability would not change during the iterative matching process (Zhang et al., 2014). Second, objects can fall into different partition units, yielding more false positive matching pairs (Fig. 4). However, since the iterative matching process converges more slowly as the partition unit grows, one has to balance between the two factors.

As a rule-of-thumb, there should be at least three smaller scale objects (i.e. $i \in I$ in the above notation) in a partition so that each object have more than one neighbor. Here, we set this minimum number to six in order to increase the average number of neighbors per object. On the other hand, including too many objects will not constantly increase the performance of the relaxation labeling technique. This is because the use of 2D Delaunay triangulation to model an object's proximity graph restricts the average number of its immediate neighbors (≤ 6). Meanwhile, this also introduces extra computational overheads, both in the proximity graph computation and in the relaxation labeling process (Zhang et al., 2014).

We use the following procedure to merge neighboring partitions until certain criteria are met.

1 Mark the common boundary to be merged: if there is a building across the boundary between two partitions, mark this boundary so that it can to be removed in a later stage.
2 Progressive merging of small partitions: (a) after the above step, rank all the partitions in an increasing order of $Num_p(i)$ (i.e. the number of buildings i contained in p); (b) in each loop, select the smallest partition ($Num_p(i) < 6$) and merge it with the second smallest partition in its immediate neighborhood, then update the rank-order list; (c) proceed with sub-step (b) until there is no partition with $Num_p(i) < 6$.

2.3. Reverse engineering of generalization history

An advantage of updating spatial data in an MRDB structure is that, knowledge used to generate smaller scale data can be inferred from the vertical relations. The knowledge already contains the complex judgement and decision made during previous generalization. For instance, operators used to generalize the objects in a specific area reveal the spatial and semantic contexts under which the generalization decision was made. This can be used to guide the incremental generalization of updates, and to reduce its complexity. By analyzing corresponding objects and vertical relations in the MRDB, it is possible to make (part of) that knowledge explicitly available.

2.3.1. Inference of individual-level generalization operators

For this work, we only outline the inference for a limited set of operators (typification, aggregation, displacement, exaggeration and compression) (Roth et al., 2011). Since some buildings are compressed

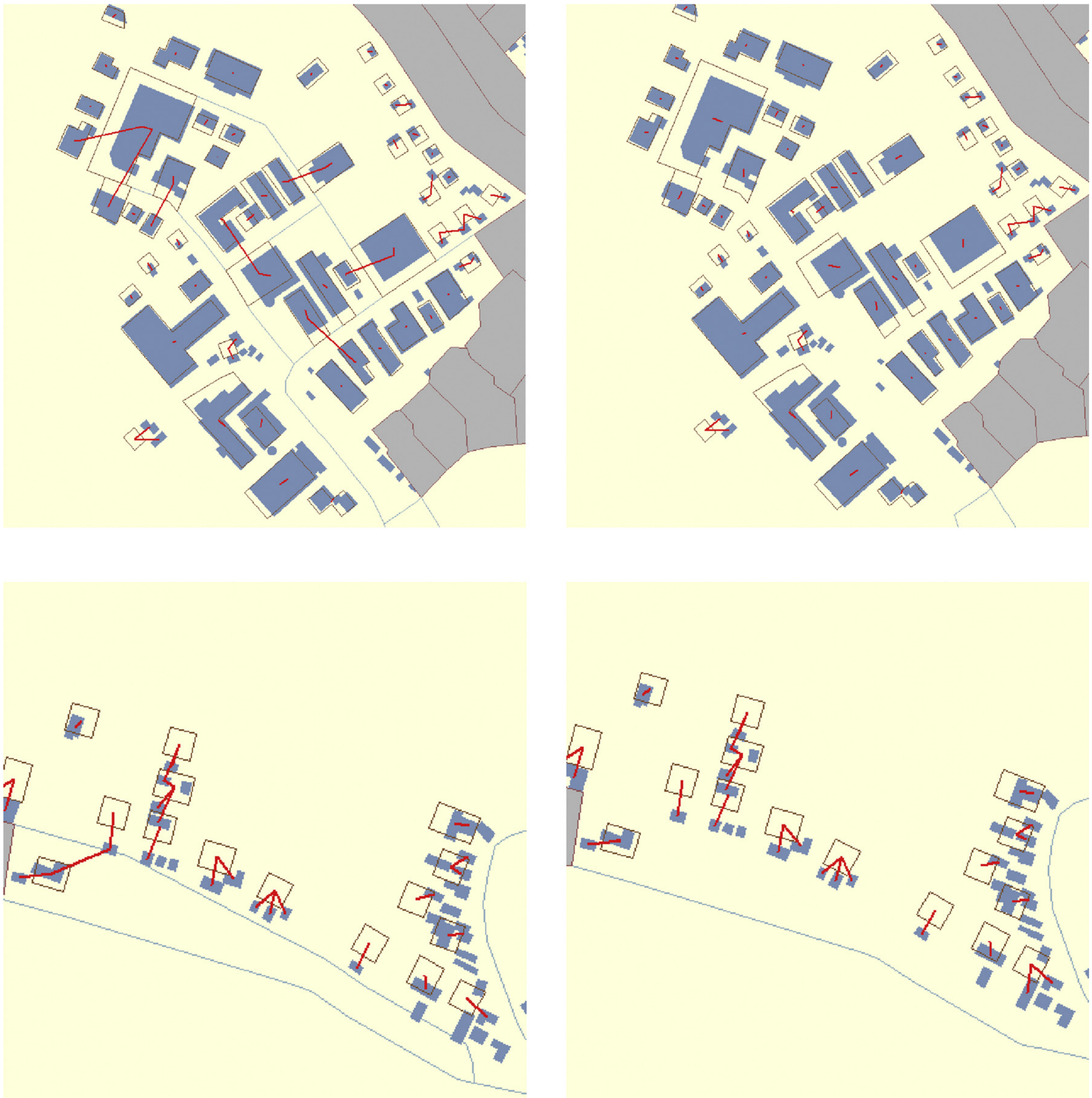

Fig. 4. Vertical relations (red links) resulted from the matching can be affected by the partitions used: finer units partitioned by local street network lead to some false positives (left); after the partition aggregation (right) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

by the widening of nearby symbols, compression is also considered here. Selection and elimination are implied in the vertical relations, and simplification is always applied during the update propagation, so there is no need for explicitly modeling. We use a set of rules to infer these operators and their parameter values using the vertical relations and related objects.

Basically, a building can be characterized by its size, length, width and orientation (Zhang et al., 2014). The length and width are obtained based on its rotated minimum bounding rectangle (RMBR), and are measured on the longer and shorter sides respectively. Building orientation can be further utilized to determine the use of rotation during map generalization.

2.3.1.1. Displacement. Displacement is extracted for all types of relations, which is the vector (magnitude and direction) between centroids of the two sides of a vertical relation (Fig. 5a). For 1-to-*m*, *n*-to-1, and *n*-to-*m* relations, centroid of the group of buildings in the *n*- or *m*-part is used to calculate the displacement vector (Fig. 5b). If the distance is greater than 0.05 mm at the target scale, we consider it a valid displacement.

2.3.1.2. Typification. Typification naturally results in *n*-to-*m* relations (Fig. 5e). For 1-to-1 and n-to-1 relations, typification usually replaces the initial group of buildings with a rectangle or square (Fig. 5c–d). The rectangles or squares are templates in that their width and/or the length equals to the minimum dimension ($Width_{min}$) at the target scale. In these cases, we examine if the generalized buildings are template geometries by comparing their width and length to $Width_{min}$ and their number of vertices. If a small building is replaced as a rectangle/square, it can also be regarded as enlargement. But here we consider this as a special case of typification for 1-to-1 relations.

2.3.1.3. Aggregation. Aggregation mostly applies for *n*-to-1 relations. It can be hard to distinguish aggregation from typification solely relying on the cardinality of the vertical relation. For aggregation, the generalized building should be at least larger than the generalized version (i.e. to meet the min dimension of $Width_{min}$) of the biggest building in the *n* initial buildings (Fig. 5f–h). In situations of small

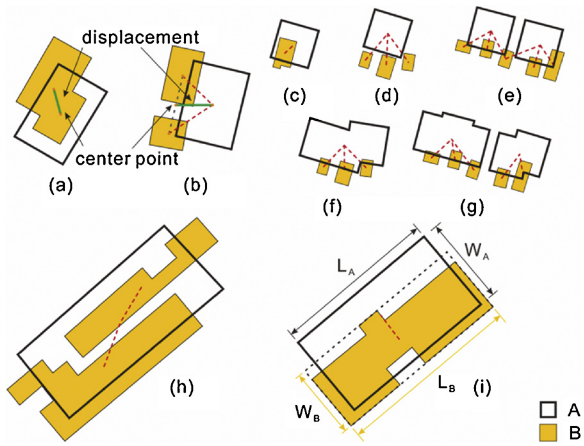

Fig. 5. Illustration of the inference of generalization operators and parameter values (dashed red links are vertical relations; A: small-scale buildings; B: large-scale buildings) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

buildings, the generalized footprint should be significantly larger than the area of a template rectangle or square.'

2.3.1.4. Enlargement and compression. Enlargement and compression are extracted based on the change of width and length of buildings before and after generalization. The change in width and length is measured independently, so that we are able to describe an object being enlarged or shrunk in either direction. The enlargement or compression of building width is determined by $W_A/W_B > \lambda$ and $W_B/W_A > \lambda$ (we used $\lambda = 1.1$), respectively; the change of length can be determined similarly (Fig. 5i).

Usually, small buildings are represented as point symbols of the same size at smaller scale (e.g. 1:50k and 1:100k). The point symbols, however, are encoded by polygons (squares in particular) in the TopNL data sets of the Netherlands. We note here that replacing a small building with a square symbol is a special case of enlargement or typification. Besides, displacement is an important compliment for handling enlargement or compression properly during update propagation. Take Fig. 5c for example, knowing that the initial object is enlarged/typified means only that you can enlarge it to a standard square with respect to its centroid; while the square can be moved to the right location with proper displacement.

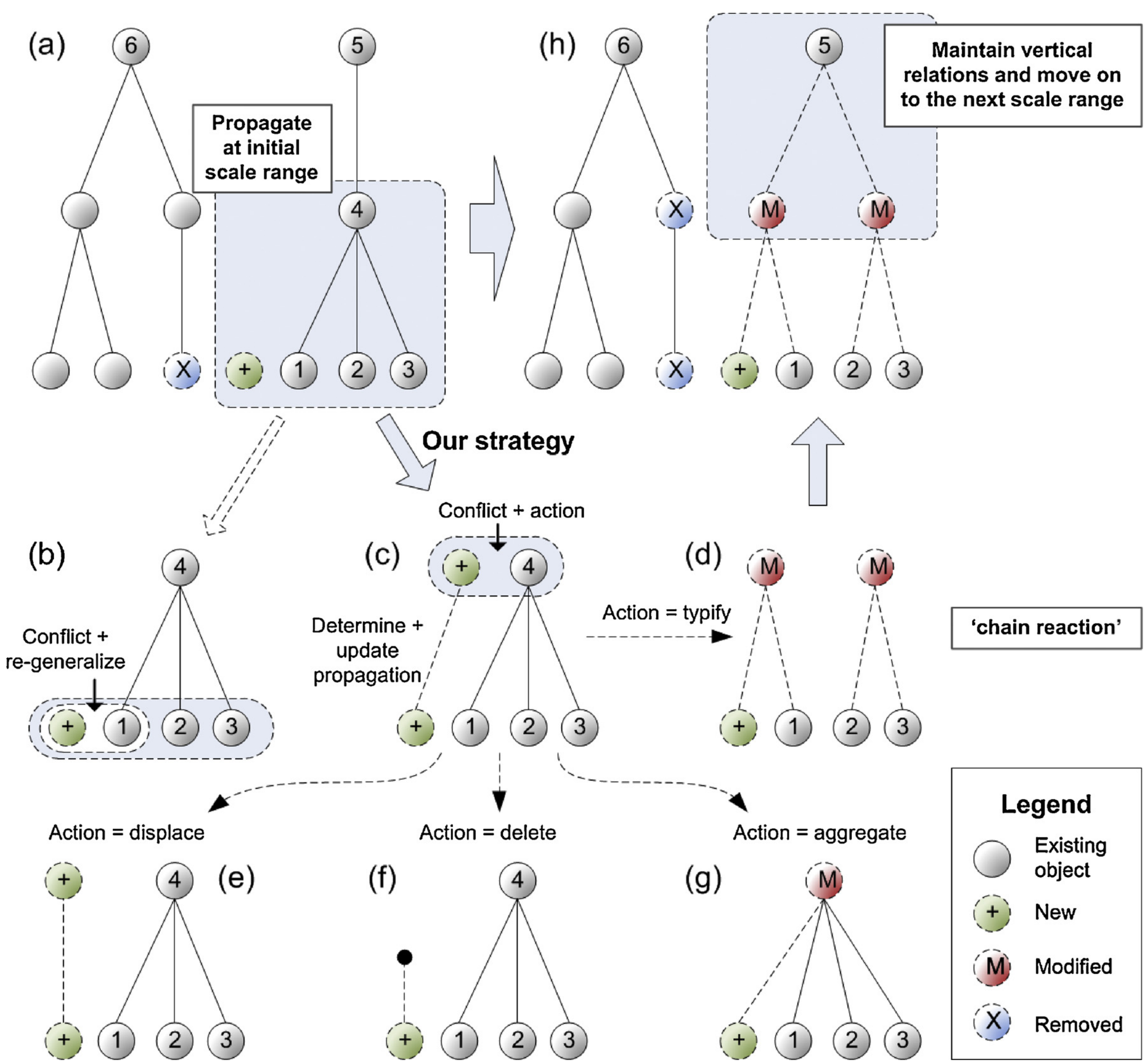

Fig. 6. Our framework for update propagation and incremental generalization in MRDBs: (a) the initial structure before the propagation; updates are presented at the larger scale (map scales are ordered in the hierarchy with the largest scale at the bottom); (b) (c) two possible strategies for conflict handling in the propagation; in our strategy (c) different actions can be taken depending on the conflicting situations and generalization history; examples are shown in (d)-(g); the pathway of the propagation examplified is: a- > c- > d- > h, and the reformed MRDB structure in (h).

2.3.2. Group-level generalization statistics for high-level decisions

We assume that in building generalization, blocks are elementary spatial units where the generalization operators used are more or less homogeneous. That is, there may be a dominant set of operators (e.g. aggregation or typification) used for some local groups or blocks of buildings. This can be helpful if individual-level guidance for the update is less relevant, e.g., when new objects are added to the database.

2.4. Update propagation and incremental generalization

In this work, changes were identified by firstly searching for candidate buildings in top10nl that differ significantly from OSM buildings using the overlap-based analysis (see e.g. Fig. 1a). Then visual inspection was used to refine the change set from the candidates for our test area. The changes identified here are by no means complete and are only used to show the potential of the update propagation process. A fully automated method is still under development and needs to be tested in the future. The identified changes in OSM are transferred to data at 1:10k and simplified (with a small simplification threshold in ArcGIS) as initial updates. The propagation of initial updates is detailed in this section.

2.4.1. The general framework

To design our update propagation framework, we have to choose from among alternative options. The first concern is: do we use a ladder or star approach (Stoter, 2005) for update propagation? In the former approach, the process is separated into different scale ranges in which updates are propagated from the next larger scale. In the star approach, every time an object (e.g. object 5 in Fig. 6a) needs to be updated, it looks for changes at the largest scale through the chaining of vertical relations (e.g. {1, 2, 3} via object 4). We prefer the ladder approach since otherwise the computational complexity increases as hierarchy levels increases, and consistency is hard to maintain between adjacent scales.

The second consideration is: when do we consider spatial conflicts caused by the updates, during the propagation or after it? If we handle conflicts during the propagation, horizontal relations should be considered at the source scale (Fig. 6b). We choose to handle conflicts after the propagation (Fig. 6c) for the following reasons. First, new or modified updates do not usually lead to conflicts at the source scale (e.g. 1:10k); conflicts only become apparent at the target scale (e.g. 1:50k) due to use of displacement and enlargement. Second, when an update is propagated to the smaller scale, it may not cause conflicts even when it might seem to at the larger scale, possibly because the elimination of its neighbors at the smaller scale gives more space. However, if we choose to handle conflicts at the larger scale, re-generalizing the conflicting objects (e.g. {+, 1} in Fig. 6b) becomes the only choice, which expands the influence of the update and reforms the tree structure. We term this a 'chain reaction' (Haunert and Sester, 2005) because by handling e.g. {+, 1}, objects {2, 3} needs also to be re-generalized and therefore object 4 modified (Fig. 6b). When conflicts are handled after the propagation at the smaller scale, we are able to provide more flexible solutions (e.g. Fig. 6d–g) with knowledge available at that scale (e.g. detected conflicts and generalization history). This does not necessarily cause 'chain reactions'.

Based on the design choice we derive the following main steps in update propagation:

1. Determine if the update can be propagated to the smaller scale; if the update is not significant with respect to the requirements at the target scale, the update is suppressed and not propagated to the target scale;
2. Propagate and generalize the updates (initial propagation) guided by the generalization history;
3. Handle spatial conflicts (horizontal influence) of the propagated updates and maintain vertical relations in the MRDB.

When we move the update to the next scale range (e.g. 1:50k ~ 1:100k), the above reasoning applies too. The larger scale in this range become the new source where data are updated and conflicts resolved in the previous update (e.g. 1:10k ~ 1:50k). So the above three steps can be repeated in subsequent update cycles (scale ranges). We elaborate on the second and third steps in the following sections.

2.4.2. Initial propagation guided by the generalization history

The initial propagation consists of the following steps: (1) updates are transferred (copied) to the target (smaller) scale as is; (2) the updates transferred to the target scale are displaced, and then (3) generalized accordingly. The latter two steps can be guided the generalization history previously extracted. In our initial propagation, the displacement is performed in the first place to avoid potential conflicts with existing objects, and to facilitate the subsequent generalization. We now discuss details for three types of updates.

2.4.2.1. Modified updates. Modified objects are re-generalized according to the generalization history. The major concern is the scope in which the re-generalization is performed. For the 1-to-1 relation, modified object is copied to the target (smaller) scale, displaced according to the offset and direction extracted (Section 2.3), and generalized according to its previous operators; if there is

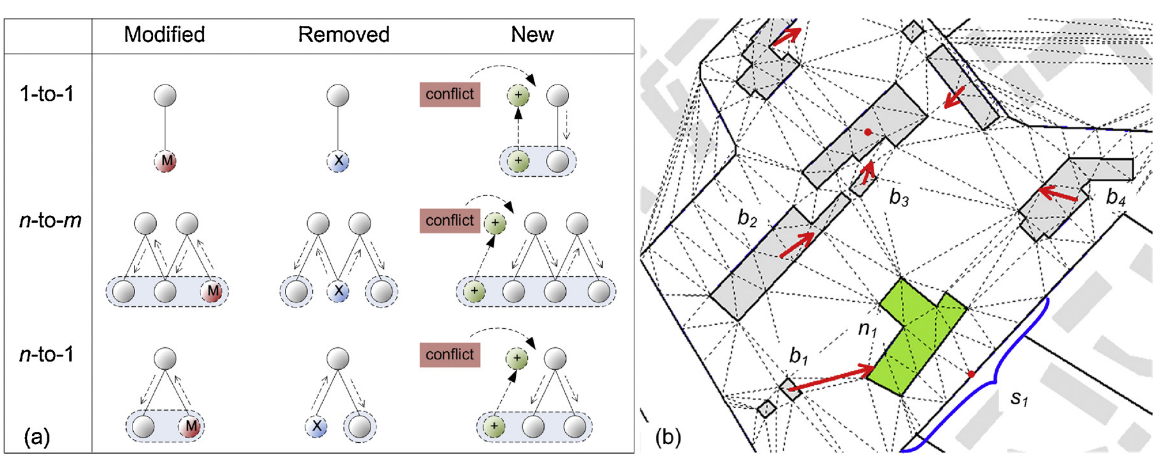

Fig. 7. Horizontal spread of an update for different update types and vertical relations during initial propagation and conflict handling; objects involved in the context of a 'chain reaction' is enclosed by dashed areas (a); displacement of a new update n_1 is interpolated from surrounding objects {b_1, b_2, b_3, b_4, s_1} with known displacement parameters (red arrows) (b) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

no specific operator extracted for the 1-to-1 relation, the updated object at the target scale is simplified and enlarged. For the *n*-to-*m* or *n*-to-1 relation, objects linked to the modified object by the vertical relation as enclosed by dashed circles, Fig. 7a, are transferred to the target scale altogether ('chain reaction'), displaced and generalized according to the operators inferred. The updated objects are then marked as 'modified' and their vertical relations to objects at even smaller scale maintained (Fig. 6h). For the 1-to-0 relation, if the modified object is significant enough to be represented at the target scale, it is treated as the propagation of a new object.

2.4.2.2. Removed updates. For the 1-to-1 relation, the object to be updated is marked as 'removed' rather than being deleted straight away. This is useful for keeping the integrity of the MRDB so that the information can be used for propagation in the next scale range. For the object removed from the *n*-to-*m* or *n*-to-1 relation, objects linked by the vertical relation (enclosed by dashed circles, Fig. 7a) are transferred, displaced and re-generalized at the target scale using the generalization history (similar to the above process). The updated objects are marked as 'modified' updates.

2.4.2.3. New updates (additions). For new objects added to the database, there is no direct reference for its generalization. We use the dominant operators inferred from their spatial context (e.g. the street block). Specifically, the displacement can be interpolated using the idea of vector fields (Haunert, 2005; Ai et al., 2015). This is done by searching for immediate neighbors within a space constrained by streets, and then by interpolating among the neighboring displacements:

$$\overline{disp_i} = \sum_{h \in N_i} \left(\frac{1}{d^2(i,\,h)} \bullet \overline{disp_h} \right) / \sum_{h \in N_i} \frac{1}{d^2(i,\,h)} \tag{3}$$

where $disp_i$ is the displacement vector of the new object *i*, $disp_h$ is the displacement vector of any object *h* (a building or a part of street) in the neighborhood of *i* ($h \in N_i$); $d(i,h)$ is the distance between outlines of *i* and *h*.

The neighborhood can be defined in a number of ways. We used Delaunay triangulation in our implementation (Fig. 7b) because it is able to identify parts of a street to which a building is in proximity and to calculate $d(i,h)$ based on outlines instead of centroids. Here we set the displacement of streets to zero to restrict the interpolated displacement for new objects. If there is more than one new update in the block, displacement is interpolated separately for each new object. After the displacement, the new objects are simplified and enlarged to meet the requirements at the target scale and marked as 'new'.

2.4.3. Dealing with the 'chain reaction' under spatial conflicts

Although spatial conflicts can be reduced by the displacement in the initial propagation, new updates (additions) may lead to new conflicts with surrounding objects at the target scale. To resolve the conflicts, generalization is needed and sometimes source scale data influenced have to be re-generalized. Which operators will be used to re-generalize the conflicts is guided by the main operators used previously for the conflicting situation or in the block (Section 2.3.2). Note that re-generalization can be optimized if displacement, aggregation, deletion is used to resolve the conflicts, where only objects at the target scale are processed.

Fig. 8 shows an example: (a) before update, two initial objects were aggregated into a bigger object; (b) when a new object (update) is added near the two initial objects, the update is transferred to the target scale and then (c) displaced using the above-mentioned method; (d) if there is no conflict, the updated situation is finished as in (c); otherwise the two objects at the target scale is re-generalized. Since the existing object (the white building in Fig. 8c) was previously generalized by aggregation, the new update is aggregated with this existing object at the target scale. Due to the displacement interpolated in the initial propagation, we do not have to re-generalize the situation from the objects at the source scale here.

However, when the conflicting situation is to be re-generalized by typification, we should look back for objects at the source scale. Source objects that are necessary for the re-generalization cannot always be found by the vertical relations, especially when multiple groups of objects are involved. This is because some objects may have been removed by typification and hence cannot be linked by the vertical relations (Fig. 13d). Without these objects, re-generalizing the influenced objects via typification would be problematic. We use the following procedure to find the source objects that are potentially influenced:

1. Let $C = \{c_i\}$ be the conflicting objects at the target scale that can be reached by the new updates within a buffer radius of minimum spacing ($Spac_{min}$), and that can be linked by vertical relations (Fig. 7a), $L = \{l_j\}$ be their corresponding objects at the source (larger) scale;
2. Calculating another buffer b_c around *C* with a radius of *R*; *R* is set to the largest displacement between *C* and *L*;
3. The set of objects at the source scale covered by b_c and still within the same block is denoted as $F = \{f_k\}$; by removing f_k that is not related to *C* by the vertical relations, we obtain the influenced set of objects *F'* at the source scale ($L \subseteq F'$).

The objects in *F'* are transferred (copied) to the target scale, and displaced using the information extracted. With the above information we are able to derive e.g. the ratio of selection ($|C|/|F'|$), which can be used to guide the typification of *F'*. The results are marked as 'modified'. An example is shown in Fig. 13.

3. Implementation and results

To update authoritative data at different scales incrementally, we populated our MRDB with OSM, Top10nl, Top50nl and Top100nl data. Since Top10nl and the more up-to-date OSM are of similar scales, our MRDB maintains multiple representations at different spatial scales and time. The data matching and update propagation were implemented as an extension to ArcGIS 9.2 and the update was carried out firstly from 1:10k to 1:50k and then from 1:50k to 1:100k. Within the same scale range, the update was performed independently for street blocks. Therefore every time an object is updated, its context such as street block and surrounding buildings is available. The generalization and supporting functionalities used in the update propagation were partially available as ArcGIS tools such as simplification and typification (available as 'resolve building conflicts') and partially self-developed (triangulation for proximity graph and neighborhood analysis, geometric measures, aggregation, enlargement, compression, and replacement of smaller buildings by template rectangles).

3.1. Vertical relations reconstructed

Fig. 9 exemplifies the corresponding relations computed between different datasets. It reveals that the matching between OSM and Top10nl is easier, and therefore changes are easier to detect between them. Ambiguities arise when matching data of different scales (e.g. 1:10k ~ 1:50k and 1:50k ~ 1:100k), due to the use of cartographic generalization such as displacement and typification. Our matching procedure yields satisfactory results for the uneven displacement and many-to-many relations.

3.2. Generalization operators inferred

Our result suggests that at the street block level, operators used are homogeneous (Fig. 10) and can be used to guide the incremental generalization if the individual-level guidance is not available. Table 2

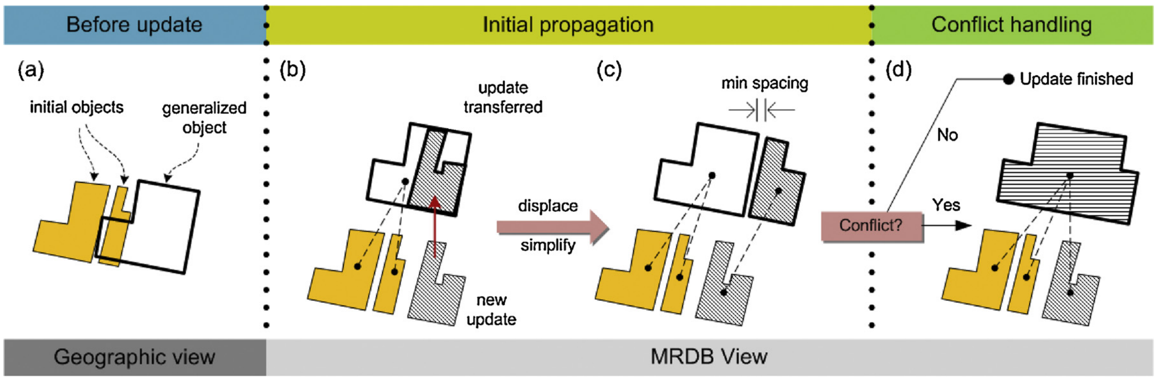

Fig. 8. The initial propagation and conflict handling illustrated for the case of aggregation.

confirms that typification is mostly used for suburban and rural areas while enlargement and compression are more frequently used in industrial areas. At the data set level, displacement is a universal operator and typification is second to it (Table 3). As map scale decreases, these two operators become more frequently used. Aggregation is rarely used in our datasets.

3.3. Typical situations of update propagation

The first situation shows the propagation of three types of updates: new, modified, and removed updates (Fig. 11). Propagation of the modified update (Fig. 11b) made use of its own generalization history (i.e. displacement); In our implementation, buildings that are reconstructed and hence change their identities, but still occupy the same location, are also considered modified updates. For the new updates in Top10nl (Fig. 11b), they were transferred to Top50nl, displaced by interpolating among surrounding objects in the block following the method in Section 2.4.2. The displaced buildings were finally simplified with the threshold $ShortSide_\tau = 20$ m for 1:50k. The interpolated displacement prevents the updates from being in conflict with existing Top50nl objects in Fig. 11c.

For the block where objects are removed from Top10nl (yellow buildings in Fig. 11b), the Top10nl buildings and the built-up area in Top50nl form an *n*-to-1 relation. According to our chaining approach in Fig. 7 for removed updates, the remaining larger scale objects in the n-to-1 relation were re-generalized. Since the building density in that block become low, they should not be converted to a built-up area. So there is no reference for how to generalize this block. We used typification (i.e. 'resolve building conflict' in ArcGIS, with $ShortSide_\tau = 20$ m, $Spac_{min} = 10$ m) to generalize the remaining buildings in the block to 1:50k (Fig. 11c). This is because globally typification is dominantly used our data besides displacement (Table 3), and the 'resolve building conflict' algorithm naturally involves displacement.

In the second situation (Fig. 12a), a Top10nl object is compressed in its right direction (with a ratio of $\lambda = 1.16$) to avoid the widening of the green space to the right. This building is later modified (extended) in Fig. 12b. To propagate this modified update, it was transferred to Top50nl, displaced with its previous history, and compressed by the same amount in one direction. The result is shown in Fig. 12c. The new update in Fig. 12b was handled properly by interpolating the displacement, which resulted in consistent update in Top50nl (bottom-right in Fig. 12c).

In the third example, the new update in Top10nl (Fig. 13 a) was propagated to Top50nl following the steps in Section 2.4.2, where it was displaced by interpolating among the neighboring displacements and was then typified (replaced by a square) because typification is dominant in this block. This did not cause conflict with existing Top50nl buildings (Fig. 13b).

Then in the next scale range, the update in Top50nl (Fig. 13c) was propagated to Top100nl following the same procedure. Although the update was displaced, conflict with existing Top100nl objects still resulted (Fig. 13d), due to the even larger template rectangles and the over-crowded space. Hence the conflicting objects in Top100nl, and thereby their larger scale correspondences in Top50nl, needs to be re-generalized ('chain reaction'). The influenced objects in Top50nl were identified using the steps described in Section 2.4.3 (a buffer with R = 28.3 m was derived from the vertical relations, Fig. 13e), and then re-generalized by typification with $ShortSide_\tau = 40$ m, $Spac_{min} = 15$ m (Fig. 13f). Again typification has been the main operator for previously generalizing the street block from 1:50k to 1:100k.

Note that the updated situation in 1:50k is a bit denser than its 1:10k counterpart. The small building in 1:10k was deliberately propagated to 1:50k here to demonstrate the above-mentioned 'chain reaction' in the propagation over two scale ranges and how it is handled (Fig. 13(d)–(f)). We can avoid creating such an unnatural dense area by filtering out this initial update in 1:10k in the first step of the proposed framework (Section 2.4.1), and as a result this update would not be propagated to 1:50k and 1:100k.

The last example shows further what happens if more objects are added to the database. There are two situations. In the lower block in Fig. 14a, the propagation of the many new updates was guided by previous generalization history of the existing objects in the block. Hence, these new objects were firstly transferred, displaced, and enlarged to template rectangles. However, the propagation created a conflict (highlighted in Fig. 14b). The influenced objects were identified with the same way and, because the block has previously been typified, they were re-generalized by typification with $ShortSide_\tau = 20$ m, $Spac_{min} = 10$ m (Fig. 14c). Note that after the initial propagation, the bottom-right building was not in conflict with other buildings, so it remains in the final result.

In the upper blocks in Fig. 14a, there are no reference for the incremental generalization of the new updates. We use guidelines specific to TopNL data (Stoter et al., 2009b) to generalize the situation. That is, densely populated urban buildings, except for detached houses, were changed to built-up areas (Fig. 14b). The detached houses are determined if they are individual polygons in OSM and tagged as 'apartment', 'commercial', or 'industrial'. Those changed to built-up areas are small buildings adjacent (sharing boundary) to each other, and are merged into bigger footprints in Top10nl as initial updates.

4. Discussion

The results for update propagation suggest that, with the help of the generalization history inferred and the way in which we deal with 'chain reactions', the updated data exhibit high cartographic quality, and the updated objects can be integrated into existing data in a more consistent way. Furthermore, the proposed strategy of initial propagation followed by conflict handling seems to be effective for incremental

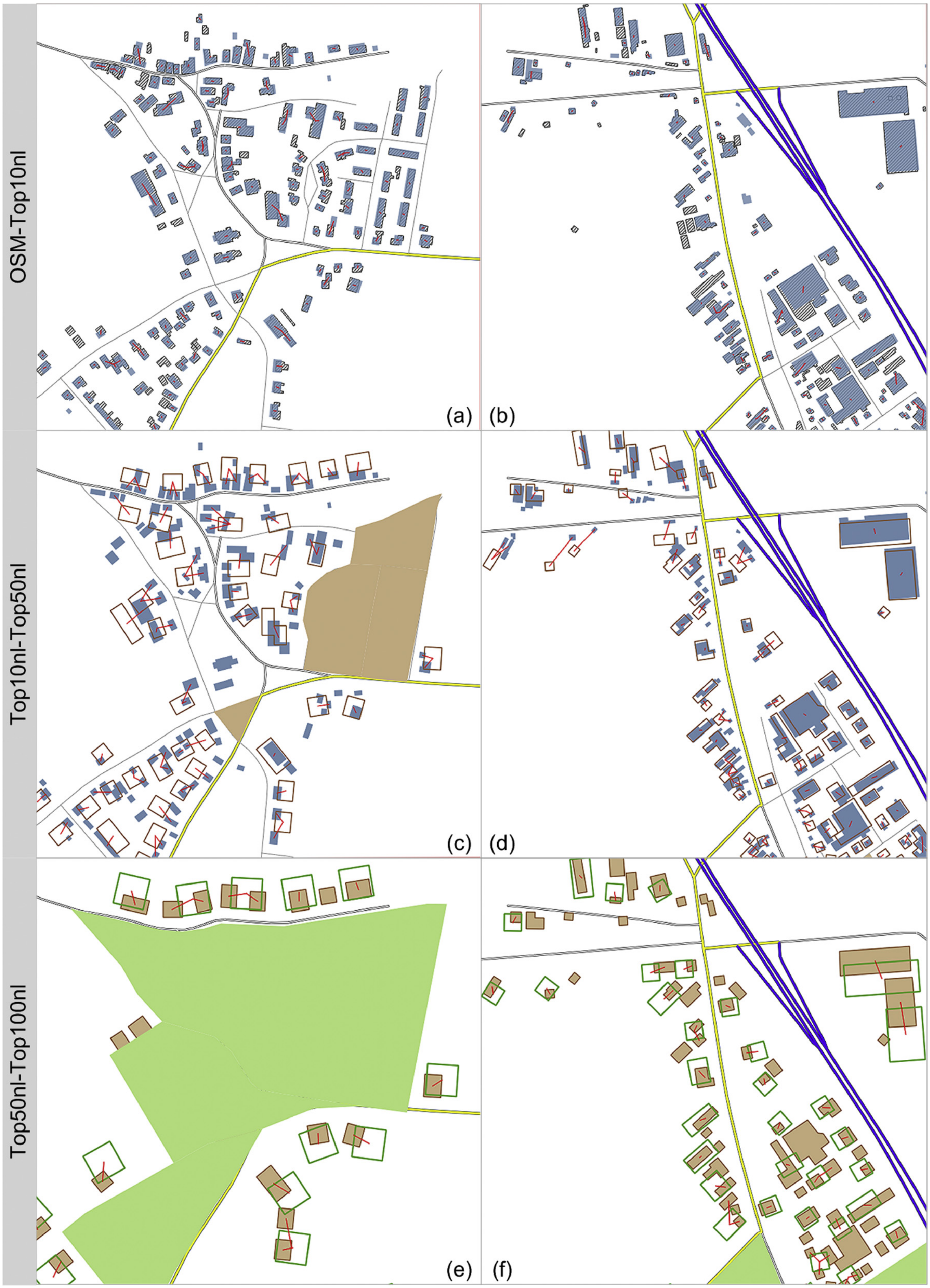

Fig. 9. Vertical relations (red links) between datasets of the same area but at different scales (links between built-up areas and their corresponding objects are hidden for clarity) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

generalization in different scale ranges.

4.1. Incremental vs. batch generalization in update propagation

To update topographic datasets at multiple scales, however, there is a debate over the choice between incremental generalization and re-generalization of the entire dataset in both production and research fields (Stoter, 2005; Stoter et al., 2014). But we note that there is no sharp distinction between the two. Indeed, incremental generalization is performed by re-generalization occurring at different granularities:

Fig. 10. Generalization operators inferred: suburban-rural (a) and industrial areas (b); note that a relation can be associated with more than one operator, which explains the color variation.

Table 2
Operator usage statistics for the two areas (1:10k ~ 1:50k) in Fig. 10 (when a relation is classified as 'typification' it is removed from 'enlarge & comp'.).

	displacement	typification	enlarge & comp.	aggregation
Area (1)	92.2%	89.2%	8.8%	2.0%
Area (2)	84.9%	30.2%	47.9%	0.0%

Table 3
Operator usage statistics for the entire datasets at different scales (buildings converted into built-up areas are excluded from our statistics).

	displacement	typification	enlarge & comp.	aggregation
1:10k ~ 1:50k	89.5%	84.8%	14.5%	< 1%
1:50k ~ 1:100k	97.6%	89.2%	8.2%	< 1%

an object, a group of objects, or a certain region; it can be a waste of time to re-generalize the entire data every time changes occur. At its extreme, incremental update can be performed in an event-driven manner, e.g., when official data is closely coupled with OSM and gets notified by user edits (as discussed in Section 2.1.3). In any cases, the proposed MRDB approach to incremental update should be appliable.

On the other hand, the efficiency gain in the fine-grained incremental generalization also depends on the rate of change and can be counteracted by the horizontal influence and the degree of 'chain reactions'. At some point, batch generalization may prove to be more suitable, though it seems to be rarely the case (when the entire dataset is drastically changed).

4.2. 'Chain reaction' in the propagation

Different from the modularization idea in Kilpeläinen and Sarjakoski (1995), our results indicate that incremental generalization can be complicated by 'chain reactions' which blur the boundary of a modular process. 'Chain reactions' occurred mainly in two situations: (1) within the scope of a many-to-many relation during the initial propagation (e.g. Figs. 7 and 11), and (2) under spatial conflicts caused by the initial propagation (e.g. Figs. 13 and 14). A 'chain reaction' usually enlarges the scope of a re-generalization (Haunert and Sester, 2005) and hence reduces the performance. This is especially the case if new buildings are added to a densely populated area (e.g. Fig. 13), though the possibility of constructing new buildings in such areas is low. Further analysis indicates that during a real update (e.g. from 1:10k to 1:50k), the process did not invoke a high level of 'chain reaction', and more objects were influenced at the source (larger) scale than at the target (smaller) scale (Table 4).

4.3. Effectiveness of the generalization history

The knowledge used to generalize the MRDB before the update is extracted and useful in several ways. Operators like enlargement and

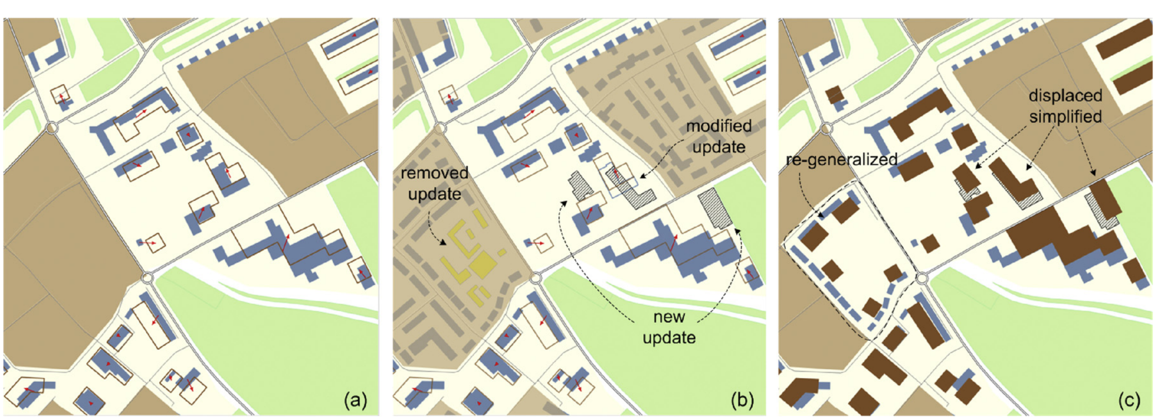

Fig. 11. Initial Top50nl and Top10nl with vertical relations (a); initial updates in Top10nl are gray and yellow polygons (b); the updates are either propagated using generalization history or re-generalized to Top50nl in case of a 'chain reaction' (c) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

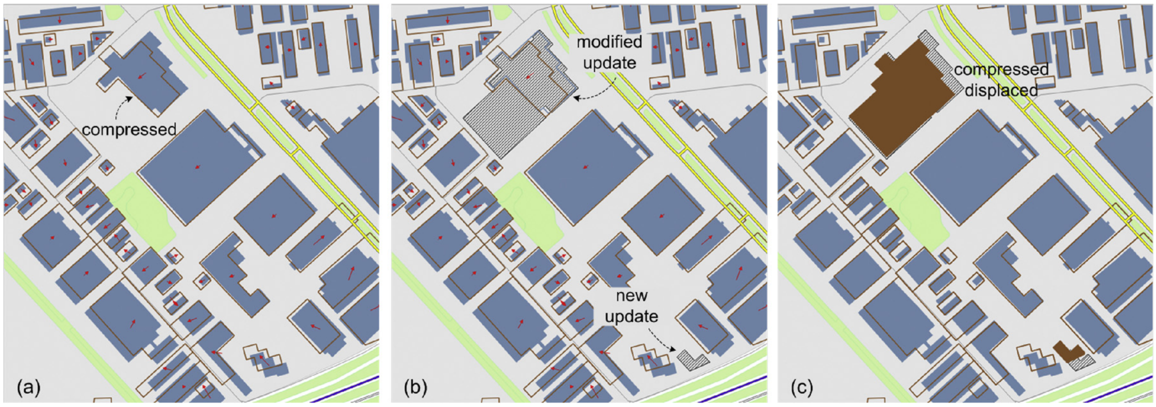

Fig. 12. The same as in Fig. 11 (a) (b); the updates are propagated to Top50nl with compression and displacement in generalization history; the untouched buildings are not shaded in dark brown for clarity (c).

compression (Fig. 12), and typification (Figs. 13 and 14) extracted in the history were effective to maintain the legibility constraints and to suppress the graphical conflicts. Since our MRDB datasets have been generalized interactively by cartographers, the knowledge extracted has the flexibility not to follow the rules strictly in the specifications. This would produce updated representations of higher cartographic quality than those generalized by following strictly the rules (Stoter et al., 2009b).

Specifically, interpolated displacement is essential to resolve (part of) the conflicts and to integrate more consistently the updates with existing data. For instance, n_1 in Top10nl is transferred to Top50nl with displacement (dashed building in Fig. 15a), which was in conflict with the existing object and caused a 'chain reaction'. This means that $\{b_1, n_1, b_2, b_3\}$ are influenced (Section 2.4.3) and need to be re-generalized. Object n_2 is transferred to the correct location and kept as is. In Fig. 15b, the two new updates are transferred to Top50nl without displacement. Here n_2 causes conflict whereas n_1 does not, which is not correct and would yield inconsistent result. In our experiments, interpolated displacement resolved a large part of the conflicts that may have been caused by new updates transferred to 1:50k, and helped to keep the spatial relations after the propagation.

In principle, the generalization history is more valuable in regions consisting mainly of continuous, incremental changes. For abrupt changes (e.g. a land-use change from residential to commercial and hence drastic reconstruction of the area), previous knowledge of generalizing the area becomes less useful. However, we demonstrate similar cases in Fig. 11 and Fig. 14 where previous generalization knowledge is not available. For such cases we used either statistics for the entire data or rules specific to our test data (Kadaster, the Netherlands) to generalize the updates. The processing seems to give promising results.

Stoter et al. (2014) suggest that predefined pipelines and rules

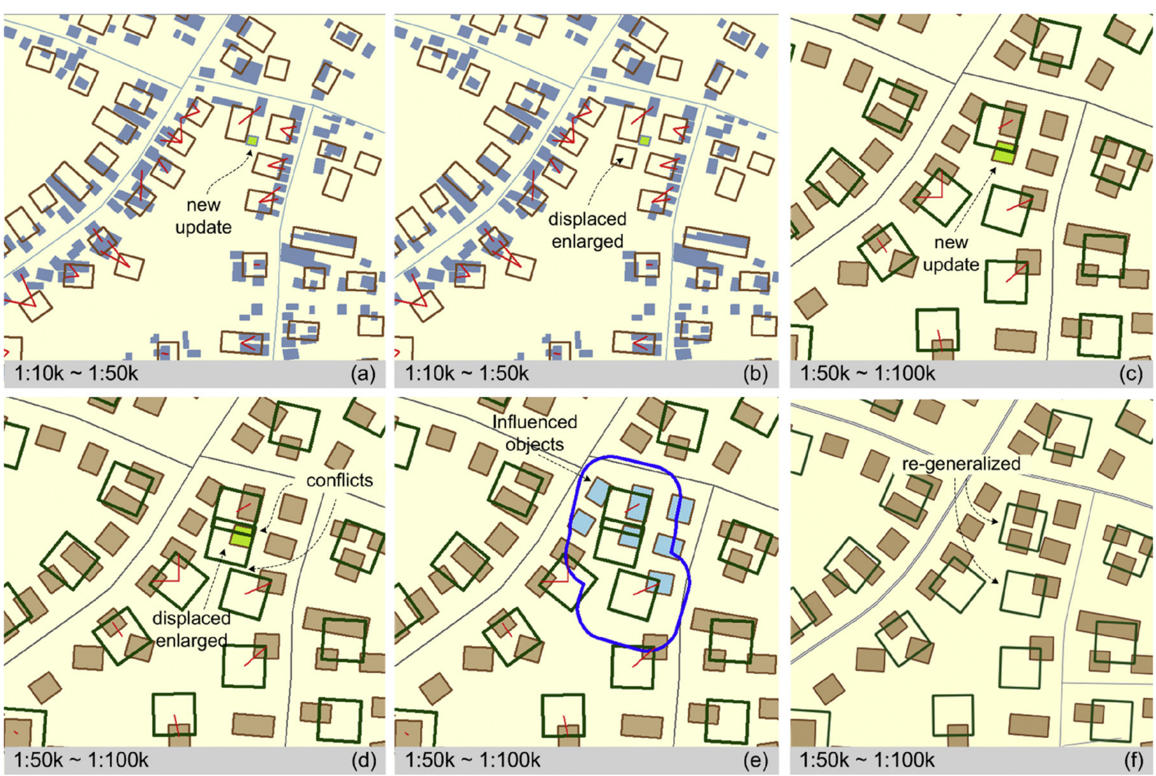

Fig. 13. Update propagation through two scale ranges and the 'chain reaction': initial update in Top10nl (a) is propagated to Top50nl (b); the update in Top50nl (c) is then propagated to Top100nl leading to graphical conflicts (d); Top50nl objects influenced by the conflicts are identified (e) and re-generalized by typification (f).

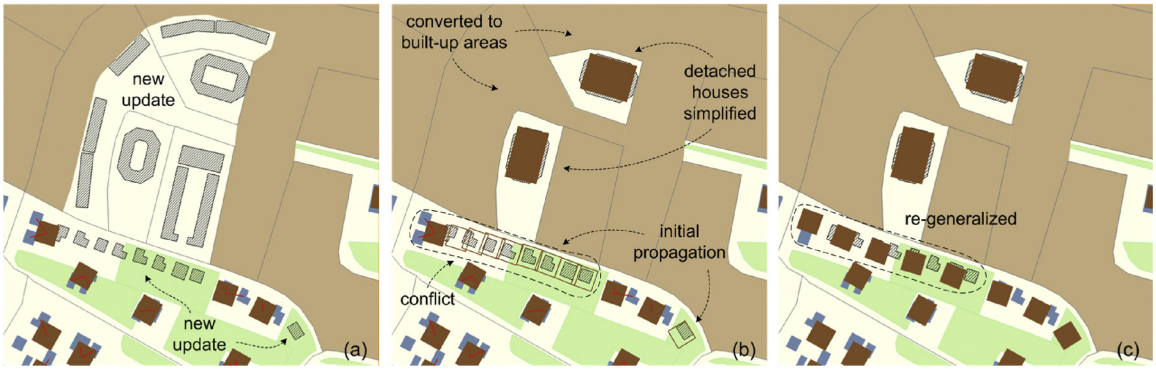

Fig. 14. Initial Top50nl and updated Top10nl, where the updates are shaded in gray (a); after initial propagation (b); after re-generalizing the influenced objects by typification.

Table 4
Statistics of the number of objects that are physically changed and those influenced by the 'chain reaction' for our test data when updating 1:50k from 1:10k.

Top10nl objects			Top50nl objects		
existing	changed	influenced	existing	changed	influenced
5784	325	113	2250	229	28

should suffice for automated generalization in relaxed requirements for, e.g., timely provision. Therefore, a question remains if there is still a need for a more holistic process (e.g. including the analysis of the data characteristics and the decision of appropriate strategies for generalization) for incremental generalization in MRDBs.

4.4. Further improvement

The focus of this paper is on the update framework that tries to make best use of the MRDB structure and the generalization history thereof, and to streamline the handling of 'chain reaction' in the propagation. There are many ways in which the framework can be refined. Currently we implemented part of the operators (e.g. typification) in ArcGIS, but this framework is generic enough to be integrated with existing generalization systems and services (Burghardt et al., 2005). For example, we tested the typification in WebGen (Neun et al., 2008) which needs the output object number as an input parameter (Burghardt and Cecconi, 2007). It can be obtained when calculating the influenced objects under conflict (Section 2.4.3). In other cases, deletion should also be an option for conflict resolution (Fig. 6f), and should be further studied as to when and how to use it.

Secondly, the degree of change should be analyzed at the change detection phase. This should be implemented as an integrated part of a functional updating system in the future to qualify if an area can be classified as continuous or abrupt change, if it is a completely new area, and how much it has changed.

In current implementation, 'chain reaction' is avoided in certain cases (e.g. the aggregation in Fig. 8). The propagation can be further optimized in the future. For instance, we notice that the effect of the re-generalization in Fig. 13f is hardly discernable (cf. Fig. 13c); the propagated update at 1:100k (Fig. 13d) can simply be deleted. Such a situation was typical when updating 1:100k from 1:50k in our datasets. The key problem is that if we can distinguish the situation in Fig. 13c from Fig. 14a by an analysis of change. Besides, the update of many-to-many relations produced by typification may be further optimized. For example, is it possible to only remove the relations linked to a removed object in the relation and complete the update? Or when only one of the objects is modified, can we minimize the scope of the re-generalization (as discussed in Bobzien et al., 2008)? There are many situations that need to be carefully examined.

Additionally, the corresponding relations between objects at multiple representations are not without uncertainty. Since the relations are calculated with probabilities in our algorithm, the reliability of an update can be notified to human under certain conditions, e.g., when the vertical relation is of a lower probability. Admittedly, as the updating proceeds, more and more vertical relations are replaced by those

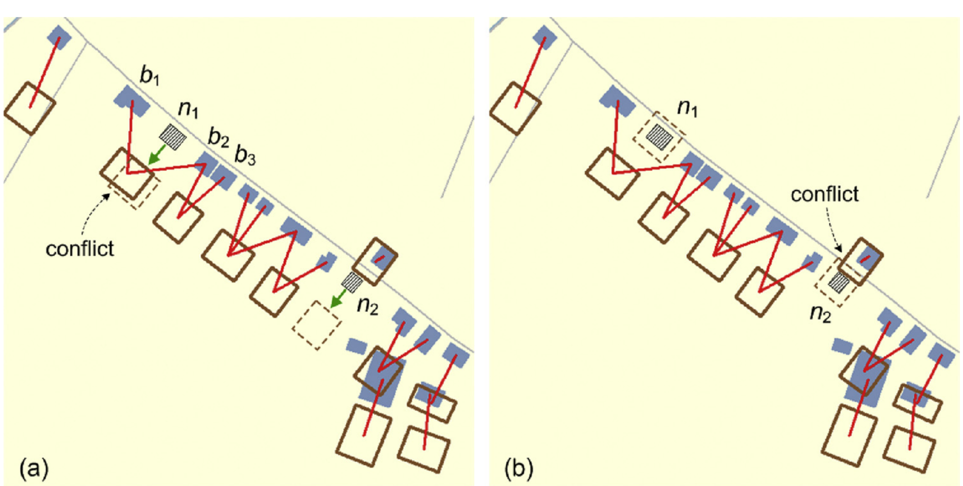

Fig. 15. Importance of displacement in the initial propagation: with (a) and without (b) displacement (red links are vertical relations) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

generated by the incremental generalization with no uncertainty.

In the future, the proposed methodology (i.e. matching features, extracting generalization history, and update propagation) should be adapted to other feature types such as road networks. On the surface, the generalization of road feature in our dataset seems to be easier than that of buildings, and hence it may be conceivable that some of the processes can be simplified. Displacement history can be useful to guide the resolution of coalescence. Because topology of road network is an important character, so another emphasize should be on maintaining the topological consistency during the process.

5. Conclusions

This paper presents a fully automated process for the establishment of MRDBs and incremental update. With updates from timely sources, spatial data at multiple scales can be revised at an unprecedented frequency and low cost. This now becomes important in the context where more and more NMAs plan to make better use of timely sources such as VGI and to adopt a continuous revision policy.

Specifically, a hybrid data matching is presented to calculate the vertical relation between the timely source and official data at several scales based purely on geometrical and contextual information (i.e. spatial relations). The matching results were quite satisfactory for different ranges of scale, based on which reasonable generalization history was inferred from the vertical relation. Then, generalization such as simplification, displacement, enlargement, compression and typification are incorporated into the incremental update process. The update propagation is guided by the generalization history extracted. Together with the suggested strategy of initial propagation followed by conflict handling, consistent data of high cartographic quality resulted.

In the future, more effort should be dedicated to a local generalization mechanism that can make use of the generalization history, minimize side-effects, and integrate updates seamlessly with existing data in an MRDB.

It is worth noting that we do not have to rely on OSM as demonstrated in this paper, in order to apply the proposed methodology. For example, in many countries municipalities have more accurate and recent data collected from the building permit process. Such data could be used as sources of updates in our approach directly. On the other hand, it is possible to approach an event-driven update of authoritative data in the future with the proposed incremental update. However, whereas our results show its technical possibilities, the practical implications of integrating the event-driven updates from timely sources (e.g. VGI) into the production lines of official agencies need to be further considered.

Acknowledgements

We thank for the reviewers for their constructive suggestions which substantially improved this paper. This work was supported by the National Natural Science Foundation of China (grant 41671384 and 41301410) and the National Key Research and Development Program of China (grant 2017YFB0503500).

References

Ai, T., Zhang, X., Zhou, Q., Yang, M., 2015. A vector field model to handle the displacement of multiple conflicts in building generalization. Int. J. Geogr. Inf. Sci. 29 (8), 1310–1331.

Begin, D., 2014. Towards integrating VGI and national mapping agency operations: a Canadian case study. In: Proceedings of the Workshop on the Role of Volunteered Geographic Information in Advancing Science: Quality and Credibility. Columbus, Ohio.

Bobzien, M., Burghardt, D., Petzold, I., Neun, M., Weibel, R., 2008. Multi-representation databases with explicitly modeled horizontal, vertical and update relations. Cartography Geogr. Inf. Sci. 35, 3–16.

Burghardt, D., Cecconi, A., 2007. Mesh simplification for building typification. Int. J. Geogr. Inf. Sci. 21, 283–298.

Burghardt, D., Neun, M., Weibel, R., 2005. Generalization services on the web – a classification and an initial prototype implementation. Cartography Geogr. Inf. Sci. 32 (4), 257–268.

Burghardt, D., Petzold, I., Bobzien, M., 2010. Relation modelling within multiple representation databases and generalisation services. Cartographic J. 47 (3), 238–249.

Devogele, T., Trevisan, J., Raynal, L., 1996. Building a multi-scale database with scale-transition relationships. In: Kraak, M., Molenaaar, M. (Eds.), Advances in GIS Research II. Taylor & Francis, London, pp. 559–570.

Dorn, H., Törnros, T., Zipf, A., 2015. Quality evaluation of VGI using authoritative data—A comparison with land use data in Southern Germany. ISPRS Int. J. Geo-Inf. 4 (3), 1657–1671.

Elwood, S., Goodchild, M.F., Sui, D.Z., 2012. Researching volunteered geographic information: spatial data, geographic research, and new social practice. Ann. Assoc. Am. Geogr. 102 (3), 571–590.

Fan, H., Zipf, A., Fu, Q., Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. Int. J. Geogr. Inf. Sci. 28 (4), 700–719.

Foody, G.M., See, L., Fritz, S., van der Velde, M., Perger, C., Schill, C., Boyd, D.S., Comber, A., 2015. Accurate attribute mapping from volunteered geographic information: issues of volunteer quantity and quality. Cartographic J. 52 (4), 336–344.

Girres, J.F., Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. Trans. GIS 14 (4), 435–459.

Goetz, M., 2012. Using crowdsourced indoor geodata for the creation of a three-dimensional indoor routing web application. Future Internet 4, 575–591.

Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69, 211–221.

Hagenauer, J., Helbich, M., 2012. Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. Int. J. Geogr. Inf. Sci. 26 (6), 963–982.

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. Environ. Plan. B 37, 682–703.

Haklay, M., Basiouka, S., Antoniou, V., Ather, A., 2010. How many volunteers does it take to map an area well? The validity of linus' law to volunteered geographic information. Cartographic J. 47, 315–322.

Haklay, M., Antoniou, V., Basiouka, S., Soden, R., Mooney, P., 2014. Crowdsourced Geographic Information Use in Government. World Bank Global Facility for Disaster Reduction and Recovery, London.

Hampe, M., Anders, K.-H., Sester, M., 2003. MRDB applications for data revision and real-time generalization. In: In 21st International Cartographic Conference. Durban, Aug 10-16. pp. 192–202.

Harrie, L., Hellström, A.-K., 1999. A prototype system for propagating updates between cartographic data sets. Cartographic J. 36 (2), 133–140.

Haunert, J.-H., 2005. Link based conflation of geographic datasets. In: Proceedings of the 8th ICA Workshop on Generalisation and Multiple Representation. A Coruña, Spain, 7–8 July. p. 7.

Haunert, J.-H., Sester, M., 2005. Propagating updates between linked datasets of different scales. In: In the Proceedings of the 22nd International Cartographic Conference (ICC2005). A Coruña, Spain, 11–16 July. p. 8.

Heipke, C., 2010. Crowdsourcing geospatial data. ISPRS J. Photogramm. Remote Sens. 65, 550–557.

Kilpeläinen, T., 2000. Maintenance of multiple representation databases for topographic data. Cartographic J. 37 (2), 101–107.

Kilpeläinen, T., Sarjakoski, T., 1995. Incremental generalization for multiple representations of geographical objects. In: Müller, J.C., Lagrange, J.P., Weibel, R. (Eds.), GIS and Generalization, GISDATA 1. Taylor & Francis, London, pp. 209–218.

Mooney, P., Morley, J., 2014. Crowdsourcing in national mapping. Eur. Spatial Data Res. 64, 147–166.

Müller, M., Wiemann, S., Grafe, B., 2012. A framework for building multi-representation layers from OpenStreetMap data. In: Proceedings of 15th ICA Workshop on Generalisation and Multiple Representation. Istanbul, Turkey, 13–14 Sept.. p. 10.

Neun, M., Burghardt, D., Weibel, R., 2008. Web service approaches for providing enriched data structures to generalisation operators. Int. J. Geogr. Inf. Sci. 22 (1–2), 133–165.

Olteanu-Raimond, A.-M., Hart, G., Foody, G.M., Touya, G., Kellenberger, T., Demetriou, D., 2016. The scale of VGI in map production: a perspective on European national mapping agencies. Trans. GIS 21 (1), 74–90.

Parker, C., May, A., Mitchell, V., 2012. Understanding design with VGI using an information relevance framework. Trans. GIS 16, 545–560.

Roth, R.E., Brewer, C.A., Stryker, M.S., 2011. A typology of operators for maintaining legible map designs at multiple scales. Cartographic Perspect. 68, 29–64.

Stoter, J.E., 2005. Generalization within NMA's in the 21st century. In: International Cartographic Conference (ICC2005). A Coruña, Spain, 11–16 July. p. 11.

Stoter, J.E., Burghardt, D., Duchêne, C., Baella, B., Bakker, N., Blok, C., Pla, M., Regnauld, N., Touya, G., Schmid, S., 2009a. Methodology for evaluating automated map generalization in commercial software. Comput. Environ. Urban Syst. 33 (5), 311–324.

Stoter, J.E., van Smaalen, J., Bakker, N., Hardy, P., 2009b. Specifying map requirements for automated generalisation of topographic data. Cartographic J. 46 (3), 214–227.

Stoter, J., Post, M., van Altena, V., Nijbuis, R., Bruns, B., 2014. Fully automated generalization of a 1:50k map from 1:10k data. Cartography Geogr. Inf. Sci. 41 (1), 1–13.

Touya, G., Brando, C., 2013. Detecting level-of-detail inconsistencies in volunteered geographic information data sets. Cartographica 48, 134–143.

Touya, G., Reimer, A., 2014. Inferring the scale of OpenStreetMap features. In: Jokar Arsanjani, J., Zipf, A., Mooney, P., Helbich, M. (Eds.), OpenStreetMap in GIScience: Experiences, Research, Applications. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, pp. 81–99.

Zhang, X., Ai, T., Stoter, J., Zhao, X., 2014. Data matching of building polygons at multiple map scales improved by contextual information and relaxation. ISPRS J. Photogramm. Remote Sens. 92, 147–163.

Zhou, S., Regnauld, N., Roensdorf, C., 2009. Generalisation log for managing and utilising a multi-representation spatial database in map production. Comput. Environ. Urban Syst. 33, 334–348.