Topological Consistency for Collapse Operation in Multi-scale Databases

Hae-Kyong Kang[◇], Tae-Wan Kim[†], and Ki-Joune Li[‡]

[◇]Department of GIS, [†] Research Institute of Computer Information and
Communication, [‡] Department of Computer Science and Engineering
Pusan National University, Pusan 609-735, South Korea
{hkkang,twkim,lik}@pnu.edu

Abstract. When we derive multi-scale databases from a source spatial
database, the geometries and topological relations, which are a kind of
constraints defined explicitly or implicitly in the source database, are
transformed. It means that the derived databases should be checked to
see if or not the constraints are respected during a derivation process. In
this paper, we focus on the topological consistency between the source
and derived databases, which is one of the important constraints to re-
spect. In particular, we deal with the method of assessment of topological
consistency, when 2-dimensional objects are collapsed to 1-dimensional
ones. We introduce 8 topological relations between 2-dimensional objects
and 19 topological relations between 1-dimensional and 2-dimensional
objects. Then we propose four different strategies to convert the 8 topo-
logical relations in the source database to the 19 topological relations in
the target database. A case study shows that these strategies guarantee
the topological consistency between multi-scale databases.

1   Introduction

Multi-scale databases are a set of spatial databases on the same area with dif-
ferent scales. In general, small-scale databases can be derived from a larger-
scale database by a generalization method[17]. During a generalization process,
both geometries and relations such as topology in the source database are to be
changed[11]. Although the relations should be respected, parts of them may not
be identical. Thus, it is needed to assess if the derived relations are consistent
or not.

The topological consistency can be checked in two ways. The first way is
that if the topologies in source and derived databases are identical[16], the two
databases are consistent as shown by figure 1-(1). In this figure, *Roads* in the
source database figure 1-(1)(A) does not intersect with *Buildings*, while *Roads*
in the derived database figure 1-(1)(B) intersects with *Buildings*. Consequently,
the two databases are inconsistent and the intersected *Buildings* with *Roads*
should be corrected. This approach can be applied when generalization oper-
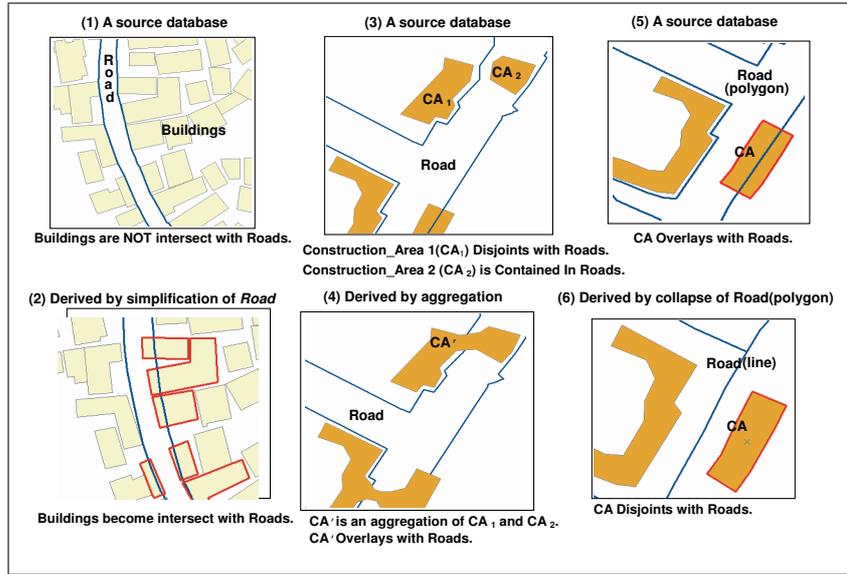ators change geometrical properties such as simplification. The second way is

Fig. 1. Conversion of topological relations in multi-scale databases

to check if or not topological relations in the derived database correctly correspond with the relations in the source database despite of their difference. In the figure 1-(3) and (4), for instance, *disjoint* in a source database can be converted into *overlay* in its derived database. Although the two relations are different, it may be a correct correspondence in a generalization process. While previous researches[4, 6, 8, 9, 13, 15] have focused on this field that the consistent conversion of topological relations between 2-dimensional objects, we consider the topological consistency in the case of the reduction of spatial dimension from 2-dimension to 1-dimension. Figure 1-(5) and (6) shows the motivation of our work.

The *overlap* relation between *Road* and CA in the source database becomes a *disjoint* relation in the derived database by a collapse operation. *Overlap* may be converted into another relations such as *meet* instead of *disjoint* depends on collapse algorithms. However, all of these conversions may not be acceptable. For instance, *disjoint* relations in the derived database is not acceptable since $Road(line)$ and CA are intersect whereas $Road(polygon)$ and CA in the source are not intersect. Thus, we must determine the possible conversions of relations to guarantee the consistency between the source and derived databases. The goal of our work is to discover the adequate topology conversions from the source database to the derived database by a *collapse* operator. In this paper, we propose four topology conversion strategies and give a comparison between these strategies.

The rest of this paper is organized as follows. In section 2, we will give a survey on the related work and explain the topological relations between two polygons and between a polygon and a line. Based on these topological relations, four topology conversion strategies will be proposed in section 3. A case study will be presented to compare these strategies in section 4. Finally, we will conclude this paper in section 5.

2 Related Work and Preliminaries

2.1 Related Work

We focus on the topological consistency on multi-scale databases. Thus our work is related to (i) topological relations, (ii) derivation or inference of new relations from existing relations in a source database, (iii) consistent conversion of topological relations on multi-scale databases. We will introduce a few novel researches on these themes briefly here and mention the limitation. Topological relations will be described in the section 2.2, separately.

• **Inference of spatial relations** : Relations in derived databases can be defined automatically by inference of relations in source databases. Thus we need to investigate the researches about inference of spatial relations. In [4], the inference of new spatial information has been formalized with properties of the binary topological relations such as symmetry, transitivity, converseness, and composition. For example of transitivity, *A contains B* and *B contains C* implies that *A meet C* is obviously inconsistent, therefore, {*contains, contains*} can not be transited to *meet*. Although this example shows part of the idea of [4], it proposed an assessment way for inferred spatial information. In [8], the transitivity of relations with set-inclusion was introduced to derive new topological relations. For instance, *A includes B* and *B intersect C* implies *A intersect C*. This approach can not be applied when relations are not represented by the set-inclusion. In [9], topology, cardinal directions like north, approximate distance like far and near, and temporal relations such as before and after were considered whereas previous researches were treated topological relations. The composition of different kinds of relations may have meaningful results since inferences can be made from not only isolated individual relations. For instance, *A is north from B* and *B contains C* implies that *A is north from C* and *A is disjoint with C*. During the derivation of a multi-scale database, spatial objects in a source database can be distorted by aggregation or collapse operators. However, these researches considered spatial objects which are not distorted unlike objects in multi-scale databases, and relations between regions(2-dimensional objects).

•**Consistent conversion of relations** : Relations in source database are, occasionally, converted to different but adequate ones on multi-scale databases. In this case, similarity or consistency between converted relations and its original ones needs to be evaluated. In [6], a boundary-boundary intersection was proposed to assess similarity of two relations on multi-scale representations. The boundary-boundary intersection is part of 9-intersection model[5], and if boundary-boundary intersections of two relations are same each other, the two

relations are same. The idea was developed based on the monotonicity assumption of a generalization, under any topological relations between objects must stay the same through consecutive representations or continuously decrease in complexity and detail. In [13], a systematic model was proposed to keep constraints that must hold with respect to other spatial objects when two objects are aggregated as shown in Figure 1-(3) and (4). This work can be a solution when a multi-scale database is derived by aggregation. However, we still need a solution for multi-scale databases derived by collapse from a source database.

2.2   Polygon-Polygon and Polygon-Line Topological Relations

The topological consistency is a core concept of our work. In order to explain it, we introduce the topological model $[1, 2, 5, 7, 10, 14]$ in this section. The most expressive way to describe topological relations is *9-intersection model*. It describes binary topological relations between two point sets, A and B, by the set intersections of interior($\circ$), boundary(∂) and exterior($-$) of A with the interior($\circ$), boundary(∂), and exterior($-$) of B.

$$R(A, B) = \begin{pmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{pmatrix}$$

With 9-intersections being empty or non-empty, 512 topological relations are possible and 8 of which in figure 2-(1) can be realized for two polygons with connected boundaries if the objects are embedded in R^2. For lines and polygons, 19 distinct topological relations in figure 2-(2) are realized if lines are not branching and self-intersections and polygons are simple, connected and no holes in R^2 [1]. These categories of topological relations are an important basis of our work. Based on the categories, we will compare topological relations in source and derived databases respectively if they are consistent or not.

3   Consistent Correspondences between Topological Relations

In this section, we introduce four approaches to correspond 8 topological relations in figure 2-(1) to 19 topological relations in figure 2-(2). We call the 8 polygon-polygon topological relations P-P relations and 19 polygon-line topological relations P-L relations for short. Four approaches are the matrix-comparison, the topology distance, the matrix-union, and hybrid approaches. The results of these correspondences are a partial set of 19 P-L relations, which is consistent with the set of P-P relations. In figure 3, we show the results of three approaches for comparisons. For example, *contain* in P-P relations corresponds to a R14 in P-L relations by a matrix-comparison and topology distance approach and R14, R15, R16, R17, R18, and R19 by the matrix-union approach. In the following, we describe how each of these approaches determines consistent relations.

h
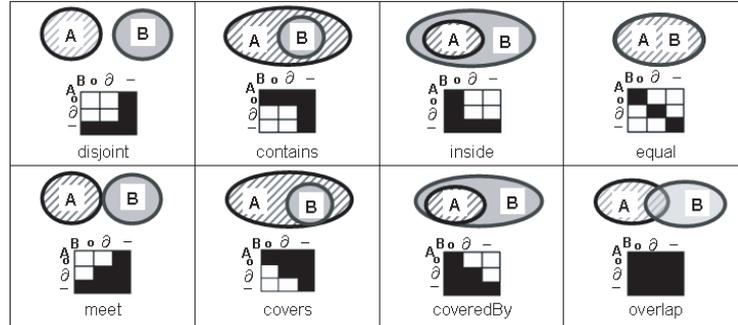
(1) 8 *P-P* Topological relations between polygons.
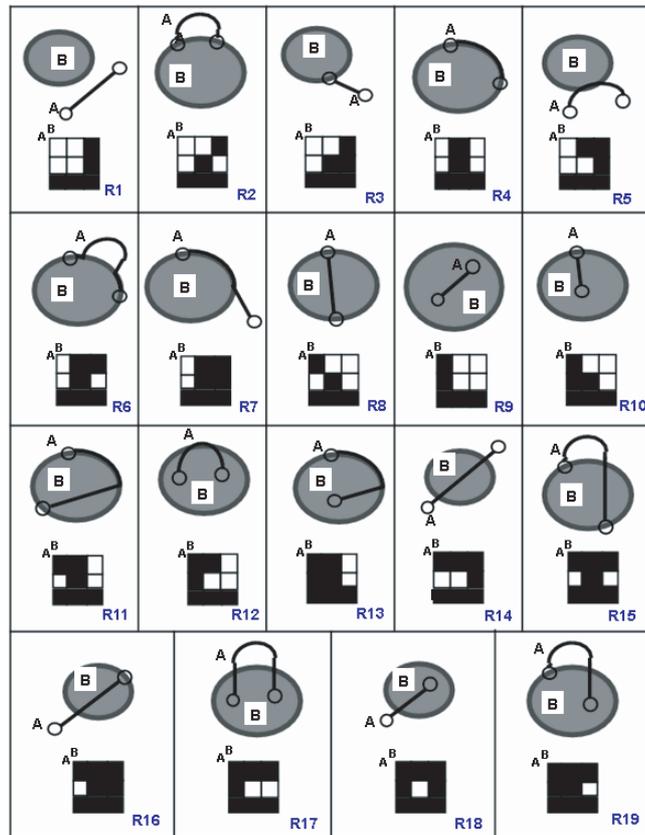
(2) 19 *P-L* Topological relations between polygons and lines.

Fig. 2. Topological Relations

8 Relations		19 Relations		
		Matrix-Comparison	Topological Distance	Matrix-Union
Disjoint		R1	R1	R1
Contains		R14	R14	R14, R15, R16, R17, R18, R19
Inside		R9	R9	R9
Equal		R8	R8	R8, R10, R11, R12, R13
Meet		R3	R3	R2, R3, R5, R6, R7
Covers		R16	R16	R14, R15, R16, R17, R18, R19
coveredBy		R10	R10	R8, R10, R11, R12, R13
Overlap		none	R16, R18, R19	R14, R15, R16, R17, R18, R19

Fig. 3. Corresponences between 8 *P-P relations for Polygons and 19 P-L relations for Polygon and Line*

3.1 Matrix-Comparison Approach

We represent a topological relation R as the 9-intersection matrix M_R. When M_Rs of two relations are equal, the two relations are the same. It is the same with the basic idea of the matrix-comparison approach. The difference is that a 6-intersection matrix is considered in the matrix-comparison approach since intersections of L^- and P^o, ∂P or P^- are always not-empty.

Definition 1. *: Line and Polygon Constraint*
A topological relation between a line L and a polygon P has the following constraint:

$$(L^- \cap P^\circ = \neg\emptyset) \wedge (L^- \cap \partial P = \neg\emptyset) \wedge (L^- \cap P^- = \neg\emptyset)$$

The 6-intersection matrix, $M_{R'_{1D}}$, for a polygon and a line L which is a simplification of a polygon P_A are constructed as follow:

$$M_{R'_{1D}(L,P_B)} = \begin{pmatrix} L^\circ \cap P_B^\circ & L^\circ \cap \partial P_B & L^\circ \cap P_B^- \\ \partial L \cap P_B^\circ & \partial L \cap \partial P_B & \partial L \cap P_B^- \end{pmatrix}$$

The 6-intersection matrix, $M_{R'_{2D}}$ for two polygons, P_A and P_B, are constructed as follow:

$$M_{R'_{2D}(P_A,P_B)} = \begin{pmatrix} P_A^\circ \cap P_B^\circ & P_A^\circ \cap \partial P_B & P_A^\circ \cap P_B^- \\ \partial P_A \cap P_B^\circ & \partial P_A \cap \partial P_B & \partial P_A \cap P_B^- \end{pmatrix}$$

Matrix-Comparison is that the two relations are consistent if and only if $M_{R'_{1D}}$ and $M_{R'_{2D}}$ are identical. For example, *equal* and R_8 are consistent since their 6-intersection matrix representations are the same.

$$M_{equal} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = M_{R_8} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

3.2 Topology Distance Approach

Similarity between two relations, R_1 and R_2, can be measured by a topology distance T_{R_1,R_2} as follows [3]:

$$T_{R_1,R_2} = \sum_{i=0}^{2} \sum_{j=0}^{2} | M_{R_1}[i,j] - M_{R_2}[i,j] |$$

The next example shows the calculation of a topology distance between *contains* and R_{16} in figure 2.

$$M_{contains} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \ M_{R_{16}} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \ M_{contains} - M_{R_{16}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ -1 & -1 & 0 \end{bmatrix}$$

$$\| M_{inside} - M_{contains} \| = 3$$

According to the topological distance approach, the correspondance of relations is shown figure 4. Smaller number of topology distances implies higher similarity of two relations. Therefore, the correspondent relations on the minimum topology distance should be selected as consistent relations.

Topology Distance / Relations	0	1	2	3	4	5	6	7	8
Disjoint	r1	r3, r5	r2, r7, r14	r6, r16, r18	r4, r8, r9, r15, r17	r10,r11, r12,r19	r13	X	x
CoveredBy	r10	r8, r9, r13	r11, r12, r19	r2, r4, r15, r17					
Cover	X	X	r16	r2,r4,r15,r17	r3, r6, r16, r18	r1, r7, r14, r15			
Overlap	x	r16, r18, r19	r7, r13, r14, r15, r17	r3, r5, r6, r10, r11, r12	r1, r2, r4, r8, r9				
Meet	r3	r1, r2, r7	r5,r6, r16	r4,r8,r14,r15	r10, r11, r18, r19	r9,r13,r17	r12		
Equal	X	X	r8	r10, r11	r2, r4, r9, r13, r15	r3, r6, r12, r16, r19	r1,r7, r14, r17	r5, r18	
Contain			r14	r5,r16, r18	r1, r7, r15, r17	r3,r6, r11,r12, r19	r2, r4, r8, r9, r13	r10	X
Inside	r9	r10, r12	r8, r13, r17	r11, r18, r19	r1, r2, r4, r14, r15	r3,r5,r6, r16,	R7		

Fig. 4. Topology Distance between 8 *P-P relations for polygons and 19 P-L relations for polygons and lines*

This approach has more expressive power than the matrix comparison approach since all of 8 P-P relations correspond to 19 P-L relations. But it has ambiguity in correspondences, that is has multiple corresponding relations in case of *overlap*. That is, *overlap* relations has three corresponding relations R_{16}, R_{18}, and R_{19}. Arbitrary selection of one of three possible relations does not still suggest satisfiable solution yet. That is, a relation whose distances are larger than another relations, can be a better relation in the context of generalization.

3.3 Matrix Union Approach

The main purpose of the matrix-union approach is to overcome the limitations of the topological distance approach, we described above, by proposing multiple consistent relations. Our purpose is to propose a method that finds consistent correspondences between P-P relations and P-L relations. A polygon could be collapsed to a line in generalization process. In other words, the interior and boundary of a polygon is collapsed to a line. Thus we can redefine a line as follows:

Definition 2. *: Line(collapsed polygon)*
A line L is a subset of the union of the interior and the boundary of a polygon P. That is,

$$L \subseteq P^{\circ} \cup \partial P$$

Based on the definition 1 and 2, we compare 3-intersection matrixes of two relations. We call the 3-intersection matrix for a line and a polygon as $M_{R_{1D}}$ and for two polygons as $M_{R_{2D}}$. $M_{R_{1D}}$ and $M_{R_{2D}}$ are constructed as follow:

$$M_{R_{1D}(L,P_B)} = \begin{pmatrix} L^{\circ \cup \partial} \cap B^{\circ} & L^{\circ \cup \partial} \cap \partial B & L^{\circ \cup \partial} \cap B^{-} \end{pmatrix}$$

$$M_{R_{2D}(P_A,P_B)} = \begin{pmatrix} (A^{\circ} \cup \partial A) \cap B^{\circ} & (A^{\circ} \cup \partial A) \cap \partial B & (A^{\circ} \cup \partial A) \cap B^{-} \end{pmatrix}$$

Matrix-Union is that the two relations are consistent if and only if $M_{R_{1D}}$ and $M_{R_{2D}}$ are identical. This approach allows multiple candidates when we generalize the P-P relation to the P-L relation. However, it may fail to suggest the most appropriate relation in some cases due to a information loss during a union process, unlike the matrix-comparison and the topological-distance approaches. In the next section, we suggest another approach, *the hybrid approach*, which resolve ambiguity in the matrix-union approach by proposing the way to select the best one.

3.4 Hybrid Approach : Matrix-Union and Topological Distance

For candidates relations found by the matrix-union approach, this approach applies topological distance to each of these relations and thus orders each of them. In generalization process, this approach suggests to select a relation among multiple candidate relations by this ordering. In figure 5, we show a set of consistent

Consistency	disjoint	contains	inside	equal	meet	covers	coveredBy	overlap	Topology Distance
Weaker									6
		R19		**R12**		R17			5
		R15,R17		**R13**		R18,R19			4
		R16		**R10,R11**		R14,R15			3
		R14			R5,R6	R16	R11,R12	R14,R15,R17	2
					R2,R7		R8,R13	R16,R18,R19	1
Stronger	R1		R9	**R8**	R3		R10		0

Fig. 5. A Hybrid Approach(Matrix-Union and Topology Distance)

relations which are ordered by topological distance. Axis y denote the degree of consistency and axis x shows each of P-P relation. For example, the topological distance between *equal* and R_8 is 0 and thus this pair has the highest consistency among $\{R8, R10, R11, R13, R12\}$. But this fact does not mean we always choose R_8 when we generalize *equal* in $P - P$ relation to one of P-L relation. Since we do not know exact generalization process, e.g. threshold value to collapse polygons, criteria to select objects to be generalized, algorithms used to generalize, etc. Main purpose of the hybrid approach is to suggest a set of consistent correspondent relations, and is to reject relations which are not included in this set.

4 A Case Study

In the precedent section, we have proposed four approaches to define the consistent correspondences between 8 P-P topological relations and 19 P-L topological relations. These approaches, first, aims to detect erroneous or inconsistent conversion of topological relations on multi-scale databases and to let users correct the inconsistent relation conversions. On the second, they are useful in choosing the best topological relations among several possible topology conversions.

In this section, we present an example of application of the topology distance and matrix-union approaches to detect inconsistent topology conversions during a derivation of multi-scale database. For this example, we prepared a large-scale source database containing roads and road facilities, as shown by figure 6-A. Roads, which are 2-dimensional objects in the source database are collapsed into lines in the derived database. During the collapse, the topologies between roads and road facilities are converted as depicted in figure 6-B.

We can detect inconsistent relations with the consistent correspondences proposed in the previous section. Figure 7 shows consistent and inconsistent topo-
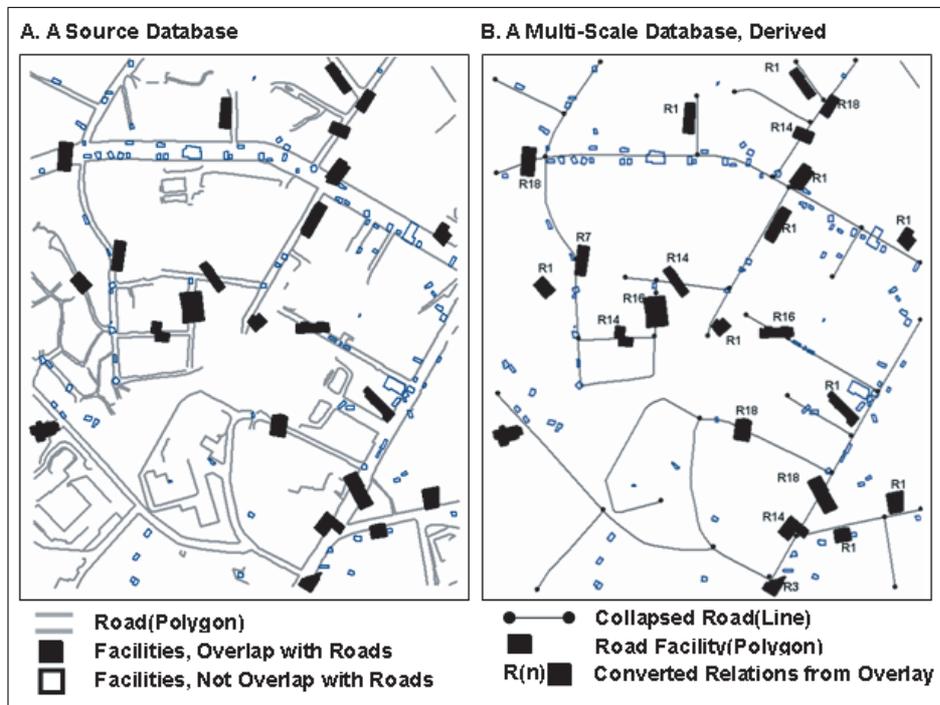
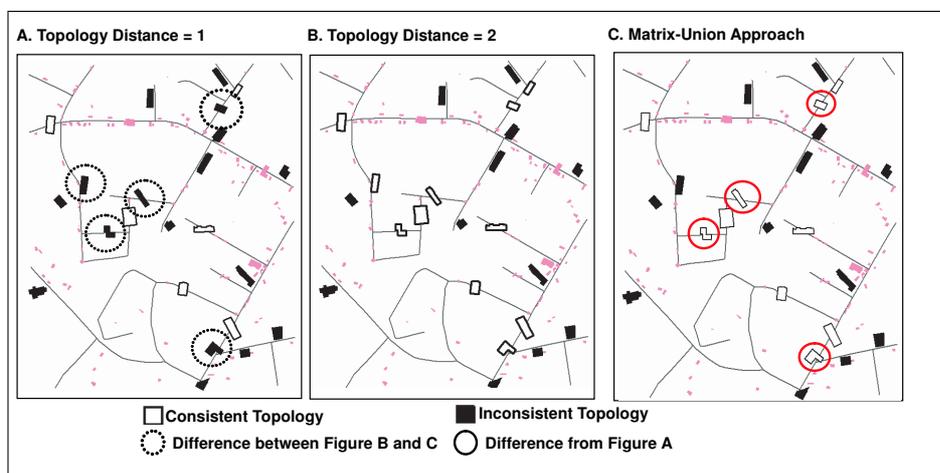**Fig. 6.** A Source and A Multi-scale Databases

Fig. 7. Topology Distance and Matrix-Union Approaches

logical relations detected by the topological distance method and matrix-union method, respectively. By topological distance method and matrix-union method, we obtain different detections of inconsistent topological conversions. With small values of topological distance, we can allow only a small set of topologies in the derived database. For example, we observe that only 6 topological relations are considered as consistent when topological distance is 1, while 11 topological relation are detected as inconsistent relation on topology distance 2. It means that we enforce the topological consistency by giving a small value of topological distance. Figure 7-C shows the result of the matrix-union approach. The road facilities with circle indicate the difference between the results of topological distance method, when the topological distance is 1. They are considered as inconsistent by topological distance method, while they are consistent by matrix-union method. But in reality, it is obvious that they are not inconsistent unlike figure 7-A, and the results of matrix-union are correct as shown by figure 7-C since *overlap* topology between polygons must be still *overlap* after the collapse operation as well. On the contrary, when topological distance is given as 2, the results of two methods are almost similar as shown by figure 7-B and 7-C, except the road facilities with dotted circle. But there is no reason to consider it as inconsistent relation, since *overlap* relation can be converted to *meet* relation like the object of dotted circle in figure 7-B. It means that topological distance method is better than matrix-union method in this case.

From this case study, we conclude that none of the proposed methods exactly discovers inconsistent and consistent relations in derived database. The reason is that we cannot handle perfectly all cases in the real world by any method without intervention of users. And an appropriate method can be chosen by trial and errors of users, which satisfies the most requirements of users. For example, it is preferable to select the method detecting as many inconsistent relations as possible, in cases where topological relations are important.

5   Conclusion

When we generalize a source spatial database of a large scale to another database of a small scale, not only topological relations but also geometries are changed from the source database. Some topological conversions from a source database may be incorrect and we must detect erroneous conversions.

In this paper we proposed several methods to assess the consistency of topological conversions when polygonal object is collapsed to a line. Our methods are based on 8-topological relation model between polygon and polygon, and 19-topological relation model between polygon and line [1]. In particular, we defined the consistent topological correspondences between the source and derived topologies by four different approaches. With these correspondences, inconsistent or erroneous topology conversions can be detected and we can maintain the consistency between source and derived databases during generalization or updates on the source database.

We have considered only the case where 2-dimensional objects are collapsed to 1-dimensional one, and the rest types of collapse operations have not been fully investigated in this paper. For example, we should define a similar type of correspondences between topologies in source and derived databases, where 2-dimensional objects are collapsed to 0-dimensional ones. Future work therefore includes the completion of the consistent correspondences considering all cases of collapse operations.

References

1. M. J. Egenhofer and H. Herring, *Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases*, Technical Report, Department of Surveying Engineering, University of Maine, 1990.
2. M. J. Egenhofer, *Point-Set Topological Spatial Relations*, International Journal of Geographical Information Systems 5(2):161-174, 1991.
3. M. J. Egenhofer and K. K. Al-Taha, *Reasoning about Gradual Changes of Topological Relationships*, Theory and Methods of Spatio-Temporal Reasoning in Geographic Space, LNCS, VOL. 639, Springer-Verlag, 196-219, 1992.
4. M. J. Egenhofer, and J. sharma, *Assessing the Consistency of Complete and Incomplete Topological Information*, Geographical Systems, 1(1):47-68, 1993.
5. E. Clementini, J. Sharma, and M. J. Egenhofer, *Modeling Topological Spatial Relations : Strategies for Query Processing*, Computer and Graphics 18(6):815-822, 1994.
6. M. Egenhofer, *Evaluating Inconsistencies Among multiple Representations*, 6th international Symposium on Spatial Data Handling, 902-920, 1994.
7. M. J. Egenhofer, E. Clementini and P. Felice, *Topological relations between regions with holes*, International Journal of Geographical Information Systems 8(2):129-144, 1994.
8. M. J. Egenhofer, *Deriving the Composition of Binary Topological Relations*, Journal of Visual Languages and Computing, 5(2):133-149, 1994.
9. J. Sharma, D. M. Flewelling, and M. J. Egenhofer, *A Qualitative Spatial Reasoner*, 6th International Symposium on Spatial Data Handling, 665-681, 1994.
10. D. M. Mark, and M. J. Egenhofer, *Modeling Spatial Relations Between Lines and Regions:Combining Formal Mathematical Models and Human Subjects Testing*, Cartography and Geographical Information System, 21(3):195-212, 1995.
11. Muller J. C., Lagrange J. P. and Weibel R., *Data and Knowledge Modelling for Generalization in GIS and Generalization*, Taylor & Francis Inc., 73-90, 1995.
12. M. J. Egenhofer, *Consistency Revisited*, GeoInformatica, 1(4):323-325, 1997.
13. N. Tryfona and M. J. Egenhofer, *Consistency among Parts and Aggregates: A Computational Model*. Transactions in GIS, 1(3):189-206, 1997.
14. A.Rashid, B. M. Shariff and M. J. Egenhofer, *Natural-Language Spatial Relations Between Linear and Areal Objects: The Topology and Metric of English Language Terms*, International Journal of Geographical Information Science, 12(3):215-246, 1998.
15. J. A. C. Paiva, *Topological Consistency in Geographic Databases With Multiple Representations*, Ph. D. Thesis, University of Maine, 1998, http://library.umaine.edu/theses/pdf/paiva.pdf.
16. A. Belussi, M. Negri and G. Pelagatti, "An Integrity Constraints Driven System for Updating Spatila Databases", Proc. of the 8th ACMGIS, 121-128, 2000.
17. H. Kang, J. Moon and K. Li, *Data Update Across Multi-Scale Databases*, Proc. of the 12th International Conference on Geoinformatics, 2004.