

5-1-2012

## Modeling spatial uncertainties in geospatial data fusion and mining

Boris Kovalerchuk  
*Central Washington University*

Leonid Perlovsky  
*Harvard University*

Michael Kovalerchuk  
*BKF Systems, Seattle*

Follow this and additional works at: <https://digitalcommons.cwu.edu/cotsfac>



Part of the [Computer Sciences Commons](#), and the [Geometry and Topology Commons](#)

---

### Recommended Citation

Boris Kovalerchuk, Leonid Perlovsky, Michael Kovalerchuk, "Modeling spatial uncertainties in geospatial data fusion and mining," Proc. SPIE 8396, Geospatial InfoFusion II, 83960N (May 1, 2012).  
<http://dx.doi.org/10.1117/12.920878>

This Article is brought to you for free and open access by the College of the Sciences at ScholarWorks@CWU. It has been accepted for inclusion in All Faculty Scholarship for the College of the Sciences by an authorized administrator of ScholarWorks@CWU. For more information, please contact [scholarworks@cwu.edu](mailto:scholarworks@cwu.edu).

## Modeling spatial uncertainties in geospatial data fusion and mining

Boris Kovalerchuk<sup>a\*</sup>, Leonid Perlovsky<sup>b</sup>, Michael Kovalerchuk<sup>c</sup>

<sup>a</sup> Department of Computer Science, Central Washington University, Ellensburg, WA 98926-7520,

<sup>b</sup> Harvard University, Charlestown, MA 02129 & AFRL, WPAFB, OH 45433

<sup>c</sup> BKF Systems, 8241 S. 123 St., Seattle, WA, 98178

### ABSTRACT

Geospatial data analysis relies on Spatial Data Fusion and Mining (SDFM), which heavily depend on topology and geometry of spatial objects. Capturing and representing geometric characteristics such as orientation, shape, proximity, similarity, and their measurement are of the highest interest in SDFM. Representation of uncertain and dynamically changing topological structure of spatial objects including social and communication networks, roads and waterways under the influence of noise, obstacles, temporary loss of communication, and other factors, is another challenge. Spatial distribution of the dynamic network is a complex and dynamic mixture of its topology and geometry. Historically, separation of topology and geometry in mathematics was motivated by the need to separate the invariant part of the spatial distribution (topology) from the less invariant part (geometry). The geometric characteristics such as orientation, shape, and proximity are not invariant. This separation between geometry and topology was done under the *assumption* that the topological structure is *certain* and does not change over time. New challenges to deal with the dynamic and uncertain topological structure require a reexamination of this fundamental assumption. In the previous work we proposed a dynamic logic methodology for capturing, representing, and recording uncertain and dynamic topology and geometry jointly for spatial data fusion and mining. This work presents a further elaboration and formalization of this methodology as well as its application for modeling vector-to-vector and raster-to-vector conflation/registration problems and automated feature extraction from the imagery.

**Keywords:** Uncertainty modeling, geospatial data, data fusion, geometry, topology, feature extraction, conflation, dynamic logic.

### 1. INTRODUCTION

Spatial distribution of the dynamic network is a complex and dynamic mixture of its topology and geometry [21]. The change of the geometric and topological structure of the network over time is challenge in many tasks including tracking objects that requires fusing and mining multi-source information. In cyber-physical space, Data Fusion and Mining (DFM) heavily depend on topological and geometrical representation and uncertainty [5, 8,9] and appropriate similarity measures [20] of spatial objects such as communication networks, roads, waterways, social networks.

In the map matching task [7] where trajectories are matched a road network, the uncertainty is twofold: (1) uncertainty of the trajectory and (2) uncertainty of the road network. The last one is caused by poor geo-registration, outdated road layers, inaccurate vectorization, feature extraction, generalization processes that produce roads from imagery and other factors [5]. Modeling such uncertainties requires techniques that go beyond cylinders and beads [17] that generalize uncertainty cylinders around the trajectory.

Figure 1 illustrated the issue of dynamically changing topology and geometry of famous in the graph theory Konigsberg's bridges. During World War II two of the seven original bridges were destroyed. Later two other bridges were replaced by a modern highway. The three bridges remain, one was rebuilt in 1935, and thus only two bridges are really original. Also Figure 1 shows on the left the green oval with three modern bridges with only one original bridge at the same location on the left, and red ovals show other changes in bridges.

---

\* [borisk@cwu.edu](mailto:borisk@cwu.edu); phone/fax 509 963-1438

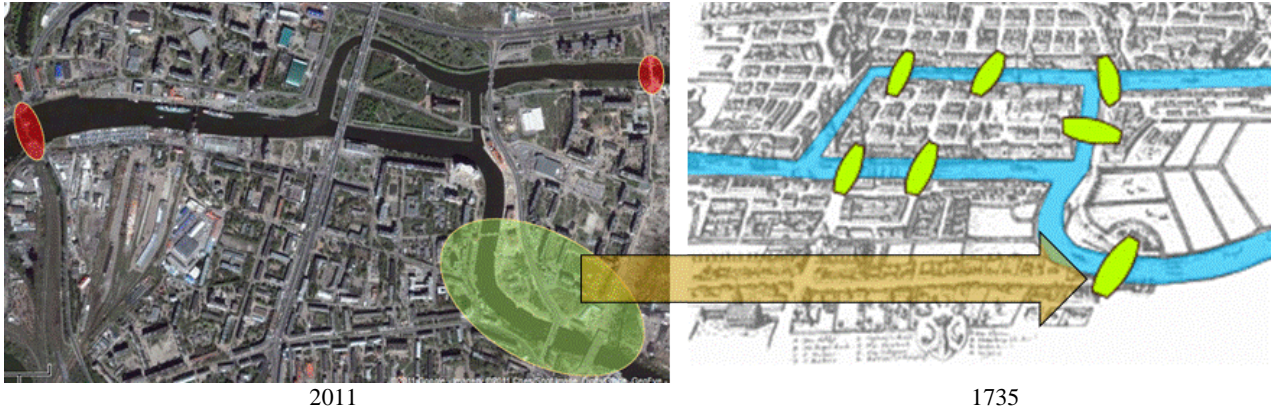


Figure 1, Königsberg's bridges: Which of three bridges matches the historic one?

This paper is organized as follows. Section 2 summarizes the main concepts of Dynamic Logic of Phenomena, where the concepts of uncertainty, generality, and simplicity for models, evaluation criteria are presented. Section 3 defines generalization of the dynamic logic for DMF of heterogeneous spatial data. Correlation with Intermediate Model Objects (CIMO) is presented in this section and is applied to the spatial data fusion problem of vector-to-raster conflation. It provides models to deal *jointly* with uncertain topology and geometry of spatial objects. Section 4 is devoted to the guided extraction of features from the imagery where GMTI serves as guidance. Section 5 summarizes the paper and discusses future research.

## 2. DYNAMIC LOGIC OF PHENOMENA (DLP)

The first concept of Dynamic Logic of Phenomena (DLP) is the concept of model (*model of phenomenon* or **P-model**). It is used in this paper as it is common in physics in contrast with how the term model is used in logic. "To model a phenomenon is to construct a formal theory that describes and explains it" [6]. Definitions that are more formal are given in [15, 2011, 2012]. The next DLP concept is a *similarity* (or *correspondence*) *measure*  $L(M,E)$  between the empirical data  $E$  and an a-priory model  $M$  that is assigned individually to each specific problem and data  $E$ :  $L: \{(M,E)\} \rightarrow R$ , where  $R$  is a set of real numbers. DLP used the concepts of uncertainty, generality, and simplicity as *partial order relations*.

An *uncertainty relation between P-models* is denoted as  $\geq_{Mu}$ , relation  $M_i \geq_{Mu} M_j$  is read: "Model  $M_i$  is equal in uncertainty or *more uncertain* than model  $M_j$ " or "Model  $M_i$  is no more certain than model  $M_j$ ". This relation is a partial order. If  $M_i >_{Mu} M_j$  then we simply say that  $M_j$  is more certain than  $M_i$ .

A *generality relation between P-models* is denoted as  $\geq_{Mg}$  and relation  $M_i \geq_{Mg} M_j$  is read: "Model  $M_j$  is a *specialization* of the model  $M_i$ " or "Model  $M_i$  is a generalization of the model  $M_j$ ".

A *simplicity relation between P-model* is denoted as  $\geq_{Ms}$  and relation  $M_i \geq_{Ms} M_j$  is read: "Model  $M_i$  is equal in simplicity of simpler than Model  $M_j$ ".

Similarly, uncertainty, generality and simplicity relations are defined for similarity measures that are matched with P-models in DLP. All these relations are partial orders. For instance, an uncertainty relation between similarity measures  $L_i$  and  $L_j$  is denoted as  $L_i \geq_{Lu} L_j$  and is read: "Measure  $L_i$  is equal to in uncertainty or more uncertain than measure  $L_j$ ".

**Definition.** Mapping  $F$  between a set of P-models  $\{M\}$  and a set of similarity measures  $\{L\}$ ,  $F: \{M\} \rightarrow \{L\}$ .

is called a *match mapping* if  $F$  preserves uncertainty, generality and simplicity relations between models and measures in the form of *homomorphism* from a relational system  $\langle \{M\}, \geq_{Mg}, \geq_{Mu} \rangle$  to a relational system  $\langle \{L\}, \geq_{Lg}, \geq_{Lu} \rangle$ , i.e.,

$$\forall M_a, M_b (M_a \geq_{Mg} M_b \Rightarrow F(M_a) \geq_{Lg} F(M_b)),$$

$$\forall M_a, M_b (M_a \geq_{Mu} M_b \Rightarrow F(M_a) \geq_{Lu} F(M_b)).$$

Two different models can be at the same level of uncertainty ( $M_1 =_u M_2$ ), one model can be more uncertain than another one ( $M_1 >_u M_2$ ), or these models can be incomparable for uncertainty. Thus, we have a *partial order* of models. Figure 2 illustrates the case when the chain of uncertainty of the models differs from the chain of generality of these models. In Figure 2 each model consists of 5 nodes with different level of uncertainties. Some links are not known. For instance, link  $c_1$  is uncertain in both models in the second row, but it is uncertain only on the left model in the third row. Each column forms a chain of models. The chain in Figure 2(b) contains three models at three levels:

Level 0: models  $M_0$  with all network nodes and connections are known.

Level 1: models  $M_1$  with one unknown connection,  $c_i$ .

Level 2: models  $M_2$  with two unknown connections,  $c_j, c_k$ .

If  $c_j = c_i$  in  $M_2$  then the models  $M_0, M_1$ , and  $M_2$  form a chain from a more specific and certain model  $M_0$  at level 0 to a less specific and certain model  $M_2$  (level 2). This is both a *certainty and generality chain*. Here  $M_2$  is more general than  $M_1$  and  $M_0$ . Note that if both  $c_j$ , and  $c_k$  are not equal to  $c_i$  then  $M_0, M_1$ , and  $M_2$  form a *certainty chain only*, but do not form a generality chain. In this case,  $M_2$  is not a generalization of  $M_1$  and  $M_2$  anymore.

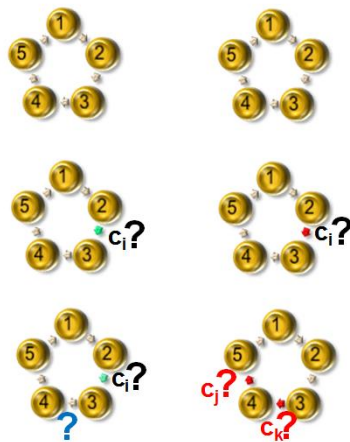
**Definition.** Mapping  $C_M: \{M\} \rightarrow \{M\}$  is called a *model learning (adaptation) operator*,  $C_M(M_i, E) = M_{i+1}$ , where E are data and  $M_i \geq_{Mu} M_{i+1}$ ,  $M_i \geq_{Mg} M_{i+1}$ .

**Definition.** Mapping  $C_L: \{L\} \rightarrow \{L\}$  is called a *similarity measure learning (adaptation) operator*,  $C_L(L_i, E) = L_{i+1}$ , where E are data and  $L_i \geq_{Lu} L_{i+1}$ ,  $L_i \geq_{Lg} L_{i+1}$ .

DLP process is illustrated in Figures 3 and 4. Figure 3 shows the mappings of models M and similarity measures L by F at different levels of models and similarity measures. It also shows transition of models by using  $C_M$  and transition of similarity measures by using  $C_L$  to more specific models and similarity measures. At each level  $i$  DLP process searches for models  $M_{imin}$ ,

$$M_{imin} = \arg \min_{\{M_i\}} L_i(M_i, E)$$

and test if  $L_i(M_{imin}, E) > T$ , i.e., is above the needed correspondence threshold to stop the process of generating more specific levels of DLP. In some cases a single initial model  $M_0$  can be given and used to generate the set of models  $\{M_i\}$  to start DLP process.



(a) Uncertainty and generality increase (b) Uncertainty increase only ( $c_i$  differs from  $c_j$  and  $c_k$ )

Figure 2. Comparison of uncertainty and generality orders.

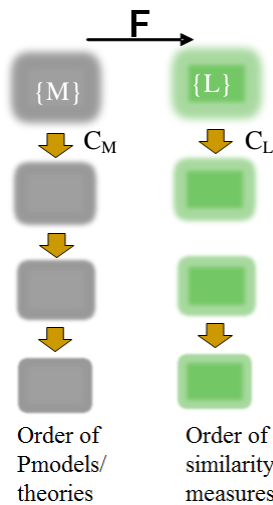


Figure 3. DLP process

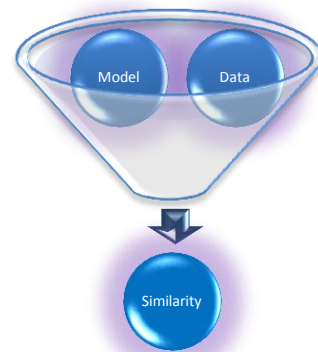


Figure 4. Main DLP step at the fixed level,  $L: \{(M,E)\} \rightarrow R$

Above we have considered separately the uncertainty of the models when the nodes are uncertain and when connections (links) are uncertain. Now we consider them together. In Figure 5, the upper row contains a fully certain 4-crossing

model, where all four links take place. The row directly below it shows all four less certain models where 3 out of 4 possible links to the crossing point are certain. The last row shows the most uncertain crossing model, where all links are uncertain and even the existence of the crossing point is uncertain. The complete 16-node lattice is parameterized as a 4-dimensional binary cube,  $E^4$ , as shown in Figure 5. Thus, we have the uncertainty sequence of models here:

$$M_3 \geq_{Mu} M_2 \geq_{Mu} M_1 \geq_{Mu} M_0,$$

where  $M_0$  is a top row model and  $M_3$  is a bottom row model.

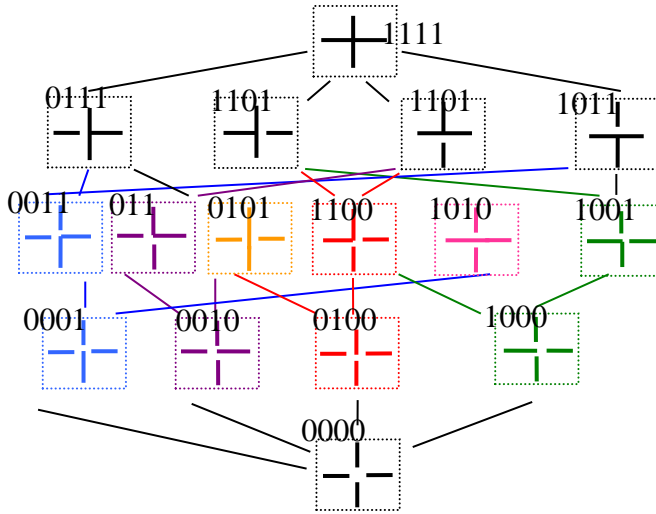


Figure 5. Uncertainty lattice of crossing models that use both geometry and topology

### 3. FUSION AND MINING OF HETEROGENEOUS DATA

Consider a road conflation task [1-4, 10, 12-14,19,22]. We need to match and conflate roads (marked A) and roads extracted from the image (marked B). Roads A should be transformed and aligned with the image using the extracted roads B. The standard topological match of the road crossings does not work here due to node mismatch, gaps in connections and differences in road coverage, which is quite typical in geospatial data matching. The matching is important because it enriches the raster image with geospatial database information from the vector data. Figure 6 shows a Tiger road network, A and a road network extracted from the overhead image B (Figure 7).

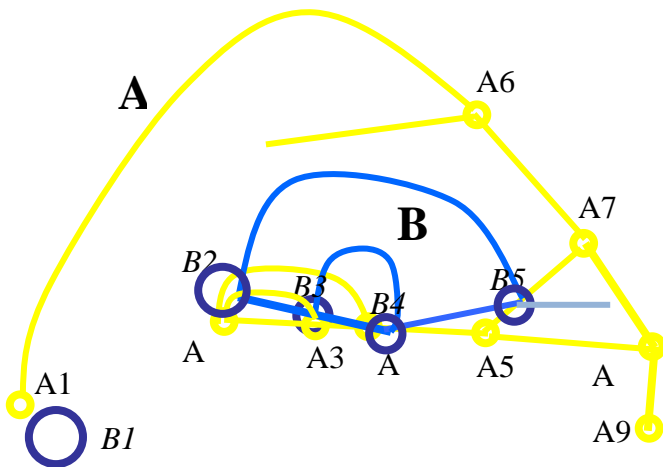


Figure 6. Matched uncertain road networks



Figure 7. Raster image

Below we use the following notation: X-nodes are connections of 4 lines, T-nodes are connections of 3 lines, L-nodes are connections of 2 lines, I-models are “connections” of a single line (I-model represents a gap or a node on single line), O-nodes are terminal nodes of the line that are not connected to other lines, and D-nodes are disconnected nodes. The network A contains two O-nodes (A1,A9), and seven T-nodes (A2- A8) and Roads B contain one D-node (B1), two L-nodes (B2,B5), and two T-nodes (B3,B4).

Below the steps of the SIMO algorithm are shown for the conflation task.

1. Identify all connections in roads A and B in the image. These are nodes A1-A9 in data A and B1-B5 in data B. In DLP terms we form a class of all nodes from A,  $\{M_{0A}\}$  and a class of all nodes from B,  $\{M_{0B}\}$ . It constitutes a highest level of uncertainty of the matching model. In formal terms we have a set of triples  $\{\langle\{N_{0A}\}, \{N_{0B}\}, f_k\rangle\}$ , where  $f_k$  is a mapping that matches nodes between A to B. At this *highest level of uncertainty* any  $f_k$  is considered to be possible. We denote this set of triples as models  $\{M_0\}$ . The similarity measure L of these models with data  $D=(A,B)$  is trivial, L values are equal for all models  $M_{0i}, M_{0j}$  with any  $f_k$ ,  $L_0(M_{0i},D)=L(M_{0i},D)$ .
2. Identify node types and label nodes with its type. Table 1 shows types of the nodes. I-nodes, X-nodes, are absent in the example. In the DLP terms this means that we construct a more specific set of models  $\{M_1\}$  where each node is labeled by its type.
3. Identify nodes as match candidates using node types. Only T-nodes A2-A8 and B2, B4 are direct candidates. In DL terms this means that a similarity measure  $L_1(M_{1i},D)=1$  if nodes in A are matched with nodes in B of the same type. Thus, we narrowed a set of matching models while still keeping a significant level of uncertainty, but lower than with  $\{M_0\}$ .
4. Measure similarity of orientation of nodes and distances between nodes of the same types. In DL terms this means that we start to generate the next more specific set of models  $\{M_2\}$  and respective similarity measure  $L_3$  based on orientation and distance as described below. The distance between T-nodes is computed as a distance between their centers (actual intersection points). All T-nodes have the same topology, thus their topological similarity is the same. Nodes A3 and A4 are similar to each other and more similar in orientation to B3 than A5 to B3. The orientation similarity measure  $L_{or}$  used in  $L_3$  is defined as a function of angles between edges,

$$L_{or}(A3,B3) = (\cos \alpha + \cos \beta + \cos \gamma)/3.$$

5. Match nodes using information from 1-4. The correct matches found are:  $A2 \rightarrow B2$ ;  $A2 \rightarrow B3$ ;  $A3 \rightarrow B4$  and  $A4 \rightarrow B5$  that is A2 is matched to two B nodes. Such complex match requires first to assume a rich model that allows one-to-many match (node 2 is linked with two nodes).
6. Change and enhance “rough” match criteria  $L_{R0}$  and  $L_{RD}$  to more certain criteria. Informally these criteria should maximize the number of nodes matched consistently with topology,  $L_{CT}$ . In the rough match each node has multiple unresolved matches. The algorithm starts with matching nearest nodes and evaluates the number of nodes matched consistently  $N_C$  relative to the total number  $N_T$  of nodes,  $N_C/N_T$ . The last process is forming a more specific set of models  $\{M_4\}$  and a more specific similarity measure  $L_4$ .

This produces matches  $A2 \rightarrow B2$ ;  $A3 \rightarrow B3$ ;  $A4 \rightarrow B4$  and  $A5 \rightarrow B5$ . The match of nodes A5 and B5 is not consistent with topology (A5 is not a part of the closed contour in contrast with B5). The match of nodes A2 and B5 is not consistent with node-types (A2 and B5 are of different types, but both belong to close contours). Thus,  $N_C/N_T=2/4$ .

Next match  $A2 \rightarrow B2$  and  $A3 \rightarrow B3$  is not consistent with topology. There are two edges between A2 and A3 that form a close a contour, but only one edge between B2 and B3. The opposite inconsistency takes place for matches  $A3 \rightarrow B3$  and  $A4 \rightarrow B4$  with two edges between B3 and B4, but only one edge between A3 and A4. This inconsistency degrades  $N_C/N_T$  to 0. An alternative match  $A2 \rightarrow B3$ ;  $A3 \rightarrow B4$ ,  $A4 \rightarrow B5$  is topologically more consistent,  $N_C/N_T=3/4$  where only B2 has no match.

The closest node to match B2 is A2 that is already matched to B3. This match is topologically consistent with two closed contours (B2, B5, B4, B3, B2) and (A2, A4, A3, A2). Having A2 node matched with two B nodes we can estimate the match  $A2 \rightarrow B2$ ;  $A2 \rightarrow B3$ ;  $A3 \rightarrow B4$  and  $A4 \rightarrow B5$  with  $N_C/N_T=3.5/4$ . This is the best SIMO match, which is consistent with the best manual solution. See also [8] on this task.

#### 4. GUIDED FEATURE EXTRACTION and GMTI

Advanced GMTI algorithms use road network constraint information to correct the target tracking [18]. However road network can be incorrectly extracted by an automatic feature extraction (AFE) algorithm and/or geo-referenced providing multiple uncertainties for moving object representations [17]. Traditional methods of road extraction *without guidance* often find multiple erroneous roads [15] illustrated in Figure 8 [10]. Such “correction” can exaggerate errors of the GMTI sensor. For instance, the overlay of a “corrected” GMTI track with aerial imagery can show a car “moving” on the roofs of the buildings. Figure 9 shows the case where a trail is not extracted from the imagery at all and cannot be used to correct GMTI data. In such cases we first need to extract a trail.

This paper *reverses* the GMTI approach and attempts to improve a trail extraction using GMTI as guidance. In the traditional GMTI approach the road network guides the GMTI track correction. The overall goal of this task is extracting the trail from imagery by using a guidance line obtained from the ground moving target identifier (GMTI) sensor. This situation is illustrated in Figure 10 with a red guidance line.



Figure 8. Scratch AFE output (blue) performed on test scene [10]. Figure 9. GMTI Guidance line without extracted trail

The *reversed approach* proposed in this paper is based on the assumptions that the actual trail is in the *vicinity* of the GMTI track and the trail has a *structure* similar to the structure of the track. Both these assumptions are uncertain. How large is the vicinity and what is the structure of the track? The next uncertainty is about *measuring* the structural and geometric similarity of the extracted trail and the GMTI track. The first one represents the topological uncertainty and the second one - the geometric uncertainty. Figure 9 shows an additional challenge to using GMTI as guidance. A red triangular tracklet in Figure 9 can be off the trail. It can show movement of a group of people in a rest area off the trail. Such tracklets do not represent the trail structure while they still can indicate that the trail is in the vicinity, but further than other tracklets from the trail. The alternative methods that optimize a weighted sum of different similarity measures have fundamental difficulties. Such methods do not control what is actually achieved in the result of optimization beyond optimizing a value of the weighted sum. Such “black box” methods including a popular snake algorithm have difficulties filtering out “rest areas” from the trail. In large part, this is a result of incremental small changes in the guidance line accomplished by the snake algorithm without using a global structure of the guidance. In other words, this is a fundamental difference between local (differential) and global (integral) approaches. This difference is also similar to the differences between top-down and bottom-up approaches. Moreover the abilities to discover a difference in the structure of the GMTI and the extracted road or trail can be used as a tool for identifying suspicious activities, such as planting an IED.

To deal with both topological and geometric uncertainties we use the Algebraic Structural Algorithm (ASA) approach implemented in MapSnap [10, 12, 14] and the DLP approach [7, 11, 16] that was outlined above. The proposed *ASA-DLP approach* is also applicable to traditional tasks of correcting the GMTI by using road or trail as a guidance. ASA-DLP approach has a mechanism to deal with the gradual change in similarity not only binary change (preserve, not preserve topological and geometric structures). In particular, we can gradually preserve such components of multilevel geometric structures as angles and relations between these angles. The data in Figure 9 is courtesy of Mark Pritt (Lockheed Martin Co). Below we illustrate the ASA-DLP algorithm on these data.



Figure 10. Road extraction using ASA-DLP approach with GMTI guidance

The first algorithm goal is *extracting the trail with preserving the global topological and geometrical structures of the guidance line* such as angles, orientation and connectivity in the search process of the trail location in the vicinity of the guidance. Consider a wide green line in Figure 10. This green line does not represent GMTI data, but it represents the *upper level structure* of it with the same general orientation, angles and connectivity. It consists of 3 linear parts that the algorithm is supposed to preserve in its first phase, while it may not preserve smaller structural elements of the red GMTI line. The *hypothesis* is that the global structure of three lines and angles between them is more likely to be present on the trail than the small structural details of the guidance line. The attempt to preserve the smaller structural elements of the guidance is made in the consequent phases of the ASA-DLP algorithm.

The second algorithm goal is to refine the extracted trail using substructures of the guidance line with more detailed information contained in the guidance. This dynamic refinement process continues several times. Our experiments had shown that no more than 8 iterations are needed to get an acceptable result in conflating vector roads and extracts from the imagery [12, 13]. It allows representing a trail sequentially with 2, 4, 8, 16, 32, 64, 128, and 258 linear segments. The process of producing the next set of segments is Binary Sequential Division (BSD) of previous segments [13].

In the first phase the ASA-DLP algorithm identifies the *global structure* of the guidance line (red line in Figure 10) by producing the *structure line of the upper level* (green line in Figure 10). Next this algorithm produces *extracted shapes* (blue lines in Figure 10) from imagery that are a set of rectangular shapes. The width of the rectangle should capture the width of the trail. These shapes have the same orientation and angles as structure lines (green lines in Figure 10), but without requiring to preserve connectivity (topological structure). As a result, the produced blue lines can be



unconnected as Figure 10 shows. At the next step of this dynamic process (building more certain models) the connectivity requirement is added to preserve the topology of the green line and to avoid “orphan” lines.

The steps of ASA-DLP are shown in Figure 11. In terms of DLP steps (1)-(5) constitute operations with a sets of models  $\{M_0\}$  and similarity measures  $\{L_0\}$ . Step (7) operates with more specific models  $\{M_1\}$  and associated similarity measures  $\{L_1\}$  and step (8) operates with further more specific models  $\{M_i\}$  and associated similarity measures  $\{L_i\}$  for  $i=2,3\dots n$ .

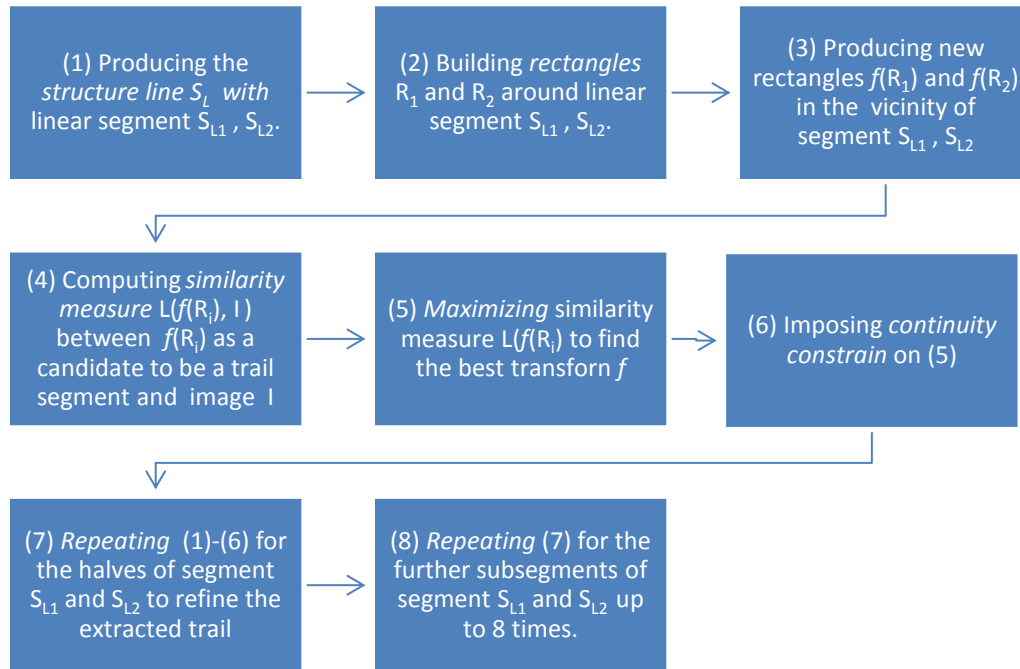


Figure 11. Steps of ASA-DLP algorithm

The ASA-DLP algorithm in more detail is as follows:

- (1) Producing the *structure line*  $S_L$  of the upper level with guidance line  $G$  interpolated by two linear segments  $S_{L1}$  and  $S_{L2}$  (green line in Figure 10). For more detail see [13].
- (2) Building *rectangles*  $R_1$  and  $R_2$  around each found linear segment  $S_{L1}$  and  $S_{L2}$ .
- (3) Mowing *rectangles*  $R_1$  and  $R_2$  in the vicinity of linear segment  $S_{L1}$  and  $S_{L2}$  and producing new rectangles  $f(R_1)$  and  $f(R_2)$ , where  $f$  is shift and rotation transform of the rectangles  $R_1$  More generally  $f$  can any affine transform. The time limitation for execution the algorithm is the major factor that limits the number of functions to be used.
- (4) Computing *similarity measure*  $L(f(R_i), I)$  between  $f(R_i)$  as a candidate to be a trail segment and overhead image  $I$ . The measure  $L$  takes into account the distribution of the pixel intensities within the rectangle  $f(R_i)$  and the area that surround this rectangle as well as the differences between these distributions. Simple versions of  $L$  can be generated by computing:
  - (i) the fraction of pixels in  $f(R_i)$  that have intensity above threshold  $T$ ,  $N_i(T)$ ,
  - (ii) the fraction of pixels in the area that surround  $f(R_i)$  that have intensity above threshold  $T$ ,  $N_{is}(T)$ ,
  - (iii) computing ratios of  $N_i(T)$  and  $N_{is}(T)$  that are similar conceptually to computing likelihood,

$$L_1(f(R_i)) = N_i(T) / N_{is}(T), \quad L_2(f(R_i)) = N_i(T) / (N_i(T) + N_{is}(T)).$$

$L_1(f(R_i))$  will be relatively large when the area of the rectangle  $f(R_i)$  is very distinct from the surrounding area, that is the surrounding area contains much smaller number of pixels with intensities above threshold  $T$ .

$L_2(f(R_i))$  will be close to 1 when the area of the rectangle  $f(R_i)$  is very distinct from the surrounding area, that is the surrounding area almost has not pixels with intensities above threshold T.

- (5) *Maximizing* a selected similarity measure  $L(f(R_i))$  with finding transform  $f$  that maximizes  $L$ ,

$$\arg \max_f L(f(R_i))$$

This is a maximization without a continuity constrain that can lead to the effect shown in Figure 10 where found  $f(R_i)$  can be disjoint.

- (6) *Imposing continuity constrain* on (5)  
(7) *Repeating* steps (1)-(6) for the halves of linear segment  $S_{L1}$  and  $S_{L2}$  to refine the extracted trail segments.  
(8) *Repeating* step (7) for the further sub-segments of segment  $S_{L1}$  and  $S_{L2}$  up to 8 times.

## 5. CONCLUSION AND FUTURE WORK

This paper continued a series of works aimed at dealing with the topological and geometrical uncertainty challenges in Spatial Data Fusion in a unified way by generalizing DLP for heterogeneous spatial data in cyber-physical space. This generalization includes: (1) “*coordinated in uncertainty*” concepts of data, models, and similarity measures, (2) simplicity, generality, and uncertainty order *relations* on data models and similarity measures, and (3) the integrated uncertainty levels for *topology and geometry* of spatial objects. The proposed concepts have been developed for solving guided automated feature extraction from imagery and the vector-to-raster conflation problem with topologically and geometrically uncertain geo-spatial data. Currently they are partially implemented in the experimental version of MapSnap The next version of MapSnap conflation software [14] will include this functionality. The alternative methods that optimize a weighted sum of different similarity measures have fundamental difficulties. Such methods have little control over what is actually achieved in the result of optimization, beyond optimizing a value of the weighted sum. The DLP generalization intends to mimic the humans’ capabilities of switching evaluation criteria and similarity measures instantaneously in a dynamic environment. This is a new frontier for future research in this area. Multiple fields of GIS, computer science, logic, mathematical modeling, and cognitive science can benefit each other in discovering the adaptive learning mechanisms of changing the spatial data models, and the similarity measures.

## REFERENCES

- [1] Chen, C., Knoblock, C., Shahabi, C., Thakkar, C., and Chiang Y., “Automatically and Accurately Conflating Orthoimagery and Street Maps”, Proc. the 12th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS'04), 47-56 (2004).
- [2] Doucette P., Kovalerchuk, B., Kovalerchuk, M., and Brigantic R., “An evaluation methodology for vector data updating”, Proc. SPIE 7334, 73341F (2009).
- [3] Doucette, P., Kovalerchuk, B., Brigantic, R., Seedahmed, G., and Graff, B., “A Method for Vector-to-Image Registration”, Applied Imagery Pattern Recognition Workshop (AIPR), IEEE, doi 10.1109/AIPR.2009.5466322 (2007).
- [4] Ghys, K., Kuijpers B., and Vaisman A., “Map matching and uncertainty: an algorithm and real-world experiments”, Proc. 17th ACM SIGSPATIAL Intern. Conf. on Advances in GIS, 468-471 (2009).
- [5] Giannotti, F. and Pedreschi, D., (eds) [Mobility, Data Mining, and Privacy: Geographic Knowledge Discovery], Springer, Berlin, (2008)
- [6] Hodes, W. "First-order model theory", Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/modeltheory-fo/> (2005)
- [7] Kovalerchuk B., and Perlovsky L., Wheeler G., "Modeling of Phenomena and Dynamic Logic of Phenomena," Journal of Applied Non-classical Logics, 22(1): 51-82 (2012).
- [8] Kovalerchuk, B. and Perlovsky, L., “Integration of geometric and topological uncertainties for geospatial Data Fusion and Mining”, Applied Imagery Pattern Recognition (AIPR) Workshop, doi 10.1109/AIPR.2011.6176346 (2011).
- [9] Kovalerchuk, B. and Perlovsky, L., “Uncertainty Modeling for Spatial Data Fusion and Mining”, Proc. 2011 IEEE Symposium Series on Computational Intelligence, Paris, doi 10.1109/CCMB.2011.5952126 (2011).

- [10] Kovalerchuk, B., Doucette P., Seedahmed, G., Tagestad, J. Kovalerchuk, S., Graff, B., “MapSnap System to Perform Vector-to-Raster Fusion”, *Proc. SPIE 8053, 805306*, doi 10.1117/12.8865 (2011)
- [11] Kovalerchuk, B., and Perlovsky L., “Dynamic Logic of Phenomena and Cognition”, in *IEEE World Congress on Computational Intelligence, Hong Kong, IEEE, 3529-3536* (2008).
- [12] Kovalerchuk, B., Doucette, P., Brigantic, R., Seedahmed, G., Kovalerchuk, M., and Graff, B., “Automated Vector-to-Raster Image Registration”, *Proc. SPIE 6966, 69660W* (2008).
- [13] Kovalerchuk, B. and Schwing, J., (Eds) [*Visual and Spatial Analysis: Advances in Data Mining, Reasoning, and Problem Solving*], Springer (2005).
- [14] MapSnap vector to raster conflation software (2011) <http://www.bkfsystems.com/mapsnap>
- [15] Doucette, P., Grodecki, J., Clelland, R., Hsu, A., Nolting, J., Malitz, S., Kavanagh, C., Barton, S., and Tang, M., “Evaluating Automated Road Extraction in Different Operational Modes”. *Proc. SPIE 7334, 73341A* (2009).
- [16] Perlovsky L., [*Neural Networks and Intellect: Using Model-Based Concepts*], Oxford University Press (2000).
- [17] Pfoser, D. and Jensen, C., “Capturing the uncertainty of moving-object representations”, in *Advances in Spatial Databases*, vol. 1651 LNCS, 111–132 (1999).
- [18] Pannetier, B., Dezert, J., Pollard, E., “Improvement of Multiple Ground Targets Tracking with GMTI Sensor and Fusion of Identification Attributes”, *IEEE Aerospace Conference*, doi 10.1109/AERO.2008.4526437 (2008)
- [19] Ruiz, J., Ariza F., Ureña, M., and Blázquez E., “Digital Map Conflation: A Review of the Process and a Proposal for Classification”, *Intern. J. of Geographical Information Science*, vol. 25, Issue 9, 1439-1466 (2011).
- [20] Santini S., and Jain R., “Similarity Measures”, *IEEE Trans. Pattern Analysis and Machine (Intelligence)*, vol. 21, 871-883 (1999).
- [21] Tøssebro, E. and Nygård, M., “A Discrete Model for Topological Relationships between Uncertain Spatial Objects”, in *Developments in Spatial Data Handling, 11th International Symposium on Spatial Data Handling*, Ed. P. F. Fisher, Springer, 395-406 (2006).
- [22] Wu, X., Carceroni, R., Fang, H., Zelinka, S., and Kirmse, A., “Automatic Alignment of Large-scale Aerial Rasters to Road-maps”, *ACMGIS'07*, doi 10.1145/1341012.1341035 (2007).