

Spatial content-based scene similarity assessment

Caixia Wang^{*}, Anthony Stefanidis, Peggy Agouris

Department of Geography and Geoinformation Science, Center for Geospatial Intelligence, George Mason University, Fairfax, VA 20109, USA

ARTICLE INFO

Article history:

Received 4 April 2011

Received in revised form 25 February 2012

Accepted 27 February 2012

Available online 7 April 2012

Keywords:

Scene matching

Similarity metrics

Image matching

ABSTRACT

Scene comparison and matching is a fundamental operation in geoinformatics. However, existing solutions are rather inadequate to support scene similarity assessment when comparing datasets collected from diverse sources especially ones that are available in diverse modalities (e.g. comparing image to vector datasets), or represent different time instances and thus differ partially in their content. In this paper we introduce a two-stage scene similarity assessment and matching framework that makes use of spatial scene content to compare and match two scenes as they may be captured in two different datasets (e.g. an aerial image and a map). At first stage our approach makes use of a matching algorithm based on the comparison of attributed graphs, where linear feature networks (e.g. road networks) are transformed into graphs and network properties are expressed through graph-embedded invariant attributes. By matching these graphs we can assess the similarity between two scenes. At the second stage, we proceed with an invariant scene comparison metric that incorporates additional scene content in the form of object configurations present within individual road network loops (e.g. building arrangements within city squares). By combining diverse but co-located pieces of information (e.g. roads and buildings) in an integrated process, our algorithm supports scene comparison and matching even when comparing heterogeneous datasets. In this paper we present key theoretical concepts and provide experimental results to demonstrate the performance of the proposed approach.

© 2012 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.

1. Introduction

Scene comparison and matching is a fundamental operation in geoinformatics, supporting for example the co-registration or georegistration of geospatial datasets, and the integration of multi-source and multitemporal datasets for site monitoring and analysis. This challenging process entails the identification in one or more datasets of scenes that correspond to (i.e. match) a reference scene. In this paper we are addressing the problem of comparing entities extracted from digital imagery (e.g. through photogrammetric or remote sensing techniques) to the corresponding records of these features in a geospatial database. In the context of this publication we are referring to this as an image-to-vector matching. While a wide variety of image-to-image matching algorithms exist that make use of image intensity values to compare and match two image windows (Babbar et al., 2010), vector datasets lack such intensity information, thus rendering the existing intensity-based algorithms not suitable for image-to-vector matching. The matching problem becomes even more complicated when the compared datasets (and thus the representations of entities extracted from them) vary in scale, orientation, or coverage, or when

they are affected by occlusions, shadows and other comparable artifacts. Thus, image-to-vector comparison often requires a human operator to manually identify corresponding features from the image and the GIS dataset and use them for matching. These features can be points, objects, or networks of linear features (e.g. road networks).

As point features are easily extracted and can be widely available, point-based solutions have long been popular approaches. For instance, Drewniok and Rohr (1997) proposed to automatically match the constellations of manhole covers extracted from large-scale urban imagery to those from a cadastral database for georegistration. In the work of Holm et al. (1995), natural objects such as islands and lakes have been used as point features for matching SPOT imagery. Brown and Lowe (2007) use scale invariant features derived through scale space analysis as input for a RANSAC solution to stitch panoramic images. More recent approaches deal with point cloud matching through the use of objective functions (e.g. Kaminsky et al., 2009), through a merge-and-mesh approach (Liu et al., 2010), or through a plane reconstruction approach (Pathak et al., 2009). Point-based approaches prove inadequate in scenes where the number and spatial distribution of extracted point features is low and irregular, and they also tend to be error-prone, as points themselves have minimal content information. Regarding object shape, position, size and orientation we can mention the work of Zhao et al. (2011) on spatial congruence metrics. By

^{*} Corresponding author.

E-mail addresses: cwanggg@gmu.edu (C. Wang), astefani@gmu.edu (A. Stefanidis), pagouris@gmu.edu (P. Agouris).

modeling geometric characteristics of elongated road segments, Wang et al. (2008) proposed a promising matching method in areas where features are not well-distributed. Efforts that employ more complex features for matching include the approach of Schickler (1992) to use 3D wireframe building models, the work of Hild et al. (2000) who used polygon features representing land-use classes to reconstruct the orientation of aerial imagery, and work on shape-based 3D object retrieval from large databases (Ferreira et al., 2010). Multiple features are proposed in recent literature to increase the number of tie points, e.g. edges and corners (Tipdecho, 2002). However, these approaches often require substantial prior or auxiliary information (e.g. relatively-accurate initial orientation information). More recently our community witnessed the emergence of scene matching techniques that make use of road networks as matching features. For instance, Chen et al. (2004) developed a specialized point pattern matching algorithm using road intersections. The proper transformation between the two datasets is determined from a fraction of the detected road intersections using a brute-force point pattern algorithm. The direction and relative distance from the two datasets are assumed available and used as prior knowledge to prune the search space of possible mapping. Wu et al. (2007) proposed an automatic matching by breaking the global alignment problem into a set of localized domains (tiles). Within each tile, a least square optimization is applied to estimate the best translation locally. This group of work on road network-based matching (Chen et al., 2004; Doytsher et al., 2001; Filin and Doytsher, 2000; Walter and Fritsch, 1999; Yu et al., 2004; Zhu et al., 2009) gives some promising results but they require that the datasets are georeferenced and based on approximate transformation. Their misalignments are assumed due to projection errors, inaccurate camera models, absence of precise terrain models, etc. These assumptions may not be satisfied in general scenarios. Furthermore, these techniques typically require an approximate transformation of the two compared scenes.

Relational graph matching has been widely adopted to solve correspondence problems in computer vision and image processing. Objects or images are represented as graphs, where vertices usually represent features of the object, and edges between them represent the relations between features. The classic early reference is the work of Barrow and Popplestone (1971), where a relational graph is designed to represent scene structure for matching. Two major categories of approaches can be identified. The first group involves the construction of structural graph model where geometric attributes of components are not taken into consider. These matching techniques are developed solely based on structure pattern to calculate graph and sub-graph isomorphism (Bunke, 1999; Jain and Wysozski, 2002; Zampelli et al., 2010). The major drawbacks in these graph-theoretical methods are their computational complexity and inability to handle any forms of inexactness within their exact framework (Wilson, 1996). Later works in Wilson and Hancock (1997) and Luo and Hancock (2001) exemplify some enhancements based on pure structural graph model. The second class of approaches compute graph similarity based on network component measurements and network relational structure. Examples include relaxation labeling (Gautama and Borghraef, 2003; Li, 1992; Rosenfeld et al., 1976; Zhao et al., 2009), the pioneering of (Vosselman, 1992) on relational matching for geoinformatics, and later extensions of this direction (Shi and Malik, 1998), Markov Random Field method (Li, 1994). In these approaches, invariant measurements of network components are essential for the matching as they can reduce ambiguities in local similarity and the corresponding searching space. But, due to the differences in scope between computer vision applications (e.g. face recognition, content-based image retrieval) and geoinformatics, one could argue that research in this direction has addressed geometric and topological attributes of the network in a

rather limited manner, focusing instead more on performance metrics (e.g. faster convergence). For example, topology is typically limited to the level of adjacency within a network. In the geoinformatics community we have witnessed that transition of some of these techniques (Mustière and Devoegele, 2008; Volz, 2006), but their impact is still considered rather limited.

Regarding scene comparison using more abstract relationships like topology, we have the work of Bruns and Egenhofer (1996) on an aggregation of gradual transitions of various geometric and topological properties to define similarity among two scenes assuming a priori pairwise correspondence of objects comprising a scene. The work of Stefanidis et al. (2002) presented a geometry-oriented approach to the problem, identifying similarities in direction and distance relations among objects comprising a scene. More recently Nedas and Egenhofer (2008) presented a cognitively motivated approach to scene comparisons. They make use of explicit domain knowledge (in the form of initial object class labels) about objects comprising a scene, and proceed with an association graph and a relaxation of spatial query constraints to select the maximal correspondence between query and database.

In this paper we propose an integrative, two-stage approach to the problem of scene similarity assessment which proceeds by sequentially filtering potential matches (Fig. 1). At first stage our approach matches road networks through a matching algorithm based on the comparison of attributed graphs: linear feature networks (e.g. road networks) are transformed into graphs, and network properties are expressed through graph-embedded invariant attributes. This generates a list of potential matching candidates who are then compared in more detail, using additional scene content in the form of object configurations present within individual road network loops (e.g. building arrangements within city squares). In this manner, we introduce a novel successively refined scene matching framework, combining the advantages of graph matching (for rapidly going over a larger area of interest to identify potential matches) with detailed scene similarity assessment (for more precisely comparing object configurations in specific scenes). The innovation of the approach lies on both the two-stage framework itself, as well as on the development of appropriate metrics to support its implementation. This innovative two-stage approach is equivalent to introducing additional geospatial layers to the matching process: while in the first stage we consider networks of linear features (e.g. road networks) to identify matching candidates, in the second stage we introduce addi-

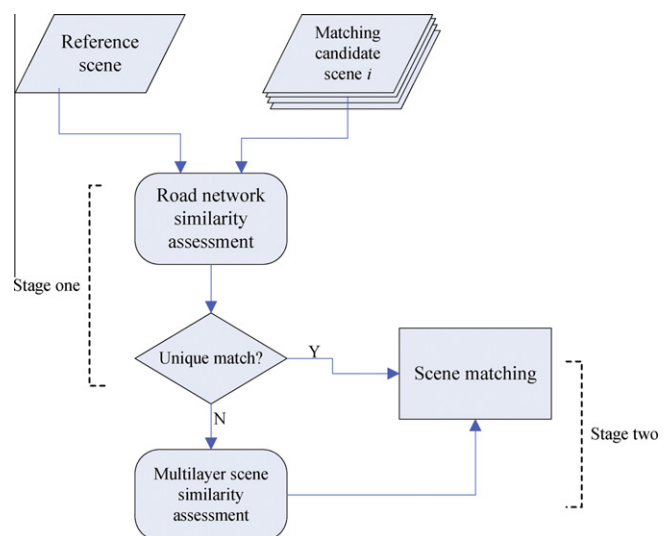


Fig. 1. Schematic flowchart of the proposed approach.

tional layers (e.g. buildings) to solve ambiguities and identify the correct match.

The ability to use spatial content for matching in a two-stage approach offers several distinct advantages. Firstly, a road network has very rich structure. It comprises line segments, point intersections and loops, and each one of these primitives has its own properties (e.g. length, width, material). By using such substantial contextual information, inherently present in our networks, we can minimize gross errors associated with erroneous point matches, thus aiming at reliable solutions that do not require extensive post-matching edits to eliminate mismatches. Furthermore, by making use of road networks (or networks of other comparable features, e.g. bodies of water) we take advantage of the natural tessellation of space offered by them, segmenting a very broad search space to small, manageable grid-like windows. Thus, our framework allows us to combine diverse but co-located pieces of information (e.g. roads and buildings) in an integrated similarity assessment process for scene similarity assessment and matching. This two-stage approach represents a hierarchical analysis of a scene, which is comparable to human visual perception, moving from edges and object outlines that form a primal sketch representation of the scene to increasingly added content that forms a 2.5D sketch and a 3D model (Henderson, 2003; Mannan et al., 1997; Marr, 1982). An argument can be made that the road network provides a natural first approximation to a primal sketch, with pronounced and elongated edges, while object configurations within the tessellation provided by this primal sketch add content. These unique advantages serve both as the motivation for our work and constitute the main contributions of this paper.

The remainder of the paper is structured as follows: Section 2 describes our approach based on attributed graph-matching scheme using road network. The next section advances the model of a road network to incorporate additional scene information in the form of object configurations present within individual road network loops (e.g. as they may become available from other GIS layers). Using this information in addition to the road network itself, an integrated matching technique for complex scene is developed. Experiments are presented in Section 4, which shows our approach is successful. In the last section we provided some conclusion remarks and discussed the need for future related research.

2. Matching using the road network

We start to present our detailed techniques of using road networks to automatically establish the matching, which perform robustly under the above-mentioned noises and are invariant to translation, rotation and scale differences between the two to-be-matched datasets. The extraction of road networks is not a topic addressed by this paper, as automatic road extraction from remotely sensed data is a well-researched topic in computer vision, photogrammetry and remote sensing. In this work, we assume that the data has been preprocessed using digital image processing and analysis techniques, e.g. Poullis and You (2010), and road networks have been detected in both datasets to be matched. A general overview of our approach framework is described as follows: The network components (junctions and their connectivity) are modeled as graphs with vertices representing junctions, edges their connectivity. Novel invariant metrics to describe networks are developed from the data. The approach proceeds by iteratively re-labeling one-graph vertices with vertices of the other graph by changing their corresponding weights. For each labeling, a local similarity is measured for the current pair of matching vertices using derived attributes. After each iteration, the global matching (i.e. global compatibility) is measured by optimizing these weights according to the local similarity between current corresponding vertices. The

process reaches an optimal matching when the global compatibility measurement becomes unchanged or varies to a limited threshold.

2.1. Network modeling and some definitions

A road network presents rather complex structure comprising segments, loops, and spatial relations. Intrinsically it contains geometric and topological information, which is of vital importance for matching. In view of the powerful representation of graphs in modeling objects and relations among them in various computer vision tasks, road networks acquired from both datasets are first transformed into *graphs* as input to our approach. In the graph, a *vertex* (also termed *nodes*) models each road intersection, *edges* of the graph represent the fact that there exist road segments joining two intersections, and an attributes set contains unary attribute attaching to each node, binary attribute attaching to each edge, ternary attribute attaching to every three related nodes, and if necessary, *n*-ary ($n \geq 3$) attribute associating to every related *n* nodes. Fig. 2 provides a graphic view of the graph over an image, where the road network is extracted. It is noted there is no edge joining the two solid nodes as the two road intersections represented by these vertices have no road segments joining them. The advantages of such modeling include that it models not only the topological structure of the road network (e.g. connectivity) but also its non-structural properties with the use of attributes. In addition, the important information for the edge set is whether or not there is a road joining two intersections. We do not need to know how many road segments connect them, which is significantly affected by the gap problems since one road segment between intersections may be extracted as several ones.

This type of graph is also termed *attributed relational graph* (ARG) extended from the ordinary graph by attaching discrete or real-valued attributes to its vertices, edges, and subset of vertices. By transforming road networks into graphs, we abstract the image-to-vector matching problem as graph matching. An important issue for graph matching is to develop attributes which are invariant from orientation difference. In this work, a set of invariant attributes associated with the graph is developed to model intrinsic properties among the graph vertices (e.g. distance between nodes), which will be elaborated in next subsection.

Here, we formalize some of the discussion and notations of ARGs from images (scenes) and vector data. Let us denote the ARG from the image as $G^{im} = (V^{im}, E^{im}, R^{im})$. In this notation, $V^{im} = \{v_1, v_2, \dots, v_s\}$ is the set of *s* vertices in the graph representing road intersections and $E^{im} = \{e_1, e_2, \dots, e_t\}$ is the set of *t* edges in the

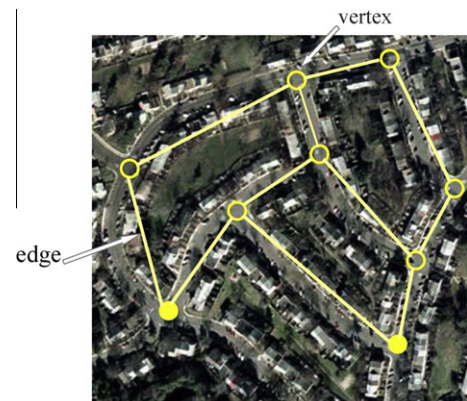


Fig. 2. A graph superimposed on the image (the graph models the extracted road networks from the image and the two solid nodes show an example of not-connected pair in the model).

graph representing relationships between road intersections. The degree $d(v_i)$ of the vertex v_i is defined as the number of edges with v_i as an endpoint. If there is an edge between vertex v_i and v_j , then v_i and v_j is said to be *adjacent* to each other and denoted by $v_i \sim v_j$. Vertices adjacent to v_i are termed *neighbors* of v_i . The *neighborhood* of v_i (denoted by $N(v_i)$) is the set of all vertices adjacent to v_i (Wallis, 2007). The *adjacency matrix* of the vertex set V^{im} is defined as follows:

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \sim v_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The third element R^{im} is a set of attributes include unary attributes r_1 defined over V^{im} , binary attributes r_2 defined over E^{im} , ternary attributes $r_3(i, j, k)$ defined over any vertex with its two neighbors and n -ary attributes $r_n(i, j, k, \dots, p)$ defined over any vertex with its n neighbors. The road network, hence, is defined in this manner through sets of vertices, edges and attributes among nodes. Similarly, the corresponding vector data can also be defined as $G^{\text{db}} = (V^{\text{db}}, E^{\text{db}}, R^{\text{db}})$ where members in each set are in uppercase notation. Using the above notations for these two networks, our aim in matching is to optimally correspond (label) vertices $V^{\text{im}} = \{v_1, v_2, \dots, v_s\}$ in graph G^{im} to those from the set $V^{\text{db}} = \{V_1, V_2, \dots, V_u\}$ in graph G^{db} satisfying certain matching criteria.

2.2. Invariant attributes for local similarity

With the general representation introduced in Section 2.1, road networks from image and vector data can be represented with graph structures associated with attributes. Invariant attributes are essential for matching as they can reduce ambiguities in local similarity and the corresponding search space. Developing invariant attributes, however, is a non-trivial issue. In one hand, as the involved imagery and GIS datasets may differ in terms of scale, coverage, and orientation in general, the conjugate features may also differ to a certain extent. On the other hand, as road networks usually involve high volume of matching components, it is important to develop attributes that require less computational efforts. In this section, we introduce attributes derived from the geometry and topology of road networks, which are invariant to translations, rotations and scale changes.

As the topological properties of a graph describe its structural characteristics and are not altered by 2D transformations (such as scaling or rotation), it is straightforward that the *degree*, a topological invariant of graphs, is an ideal unary attribute associated with each node. In our defined ARGs for the road networks, every node represents a road intersection where at least two roads join. The degree of any node must be equal to or greater than two.

Typically the mathematical Euclidean metric is an important measurement of the geometry. It is invariant to translations and rotations, but not to scale changes. The Euclidean distance between two nodes joined by an edge can not be used directly as a binary property. The *angle* formed by one node and its two neighbors, however, is invariant to 2D translations and appropriate to be applied as a ternary property. To overcome the variation of Euclidean distance to scales, a *relative distance* is proposed and can be used as one type of ternary property:

$$r_{ij,j \in N(i)} = \frac{D_{ij,j \in N(i)}}{\frac{1}{2}(D_{ij,j \in N(i)} + D_{it,t \in N(i) \cap t \neq j})} \quad (2)$$

where D_{ij} is the Euclidean distance between vertices i and j (j is a neighbor of i) and D_{it} is the Euclidean distance in i and t (t is a neighbor of vertex i and $t \neq j$). The denominator is the average of the Euclidean distances between vertexes i and its two neighbors. As vertices in the graphs denote road intersections, every vertex will have at least two adjacent vertices. In the case of more than two

neighbors to current vertex i , j and t in the relative distance are selected randomly from its neighborhood. Apparently, the property can be extended for higher-ordered property as:

$$r_{ij,j \in N(i)} = \frac{D_{ij,j \in N(i)}}{\frac{1}{m} \sum_{p_q \in N(i), q=1}^m D_{ip_q}} \quad (3)$$

Using adjacency information, we derive another attribute termed *basic loop* formed by some vertices and edges. Apart from above discussed attributes which quantitatively describe geometric properties, basic loop models higher topological structure of the network. It is invariant to translations, rotations, and scale changes. Basic loop attribute is very useful in, for instance, integrating multilayer information (e.g. buildings) bounded by a closed loop for local similarity. Let us go over some definitions (Wallis, 2007) in graph theory before we proceed to define the basic loop. A *walk* in a graph G is defined by a finite sequence:

$$x_0, a_1, x_1, a_2, \dots, a_n, x_n \quad (4)$$

where x_0, x_1, \dots, x_n are vertices and a_1, a_2, \dots, a_n are edges of G ; for each i , the endpoints of a_i are x_{i-1} and x_i . The number of edges in a walk is its *length*. If $x_0 = x_n$, we say the walk is *closed*. A *cycle* of length n is a closed walk of length n , $n \geq 3$, with no repeated vertices. Accordingly, a *basic loop* is defined as:

If a cycle in a graph contains no cycles other than itself, it is said to be a basic loop; otherwise, the cycle with minimum length is said to be a basic loop.

In the case of graphs modeling road networks, the basic loops are of triangle, quadrangular or more complex polygons. Consider the graph in Fig. 3. It includes three cycles ($abcde$), ($abce$), (dec), of lengths 5, 4, 3 respectively. Since the cycle ($abcde$) contains ($abce$) and has greater length, the cycle ($abce$) is a basic loop. Meanwhile, the cycle (dec) contains no other cycles, and thus is a basic loop of the graph.

These developed attributes, invariant to orientation variations, characterize the local patterns in terms of geometry and relationship structure. More or complex attributes (e.g. road width or average curvature), if desired, may be derived and used. Potentially additional attributes may contribute to improving performance of the next phase – matching extracted features. The derivation process and computing local matching, on the other hand, may be computationally expensive, particularly when the number and complexity of proposed attributes increase substantially. What's worst is that these additional attributes may be redundant to achieve a robust matching. At this point, proper attributes used for matching are critical and further analysis is necessary in terms of sufficiency. To conclude, proposed attributes are preferred to be capable of describing sufficient patterns of the road networks geometrically and/or topologically, invariant to any 2D transformation, required for a robust matching, yet without comprising much on the complexity in deriving and exploring them. Unfortun-

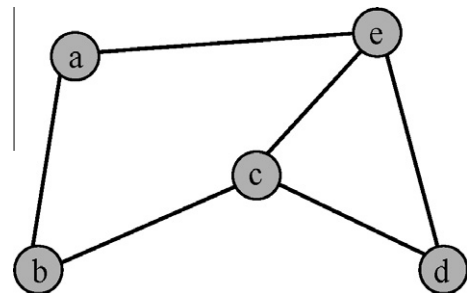


Fig. 3. A sample graph for cycles, and basic loops.

nately, this issue has not given rise to much attention in popular matching literature. Attributes are usually selected based on ad hoc decisions.

2.3. Attribute entropy analysis

In this section, we discuss the above-mentioned issue using entropy concept. Originated from classical thermodynamics (Clausius, 1879), Entropy is a quantitative entity defined fundamentally via an equation. It has been extended to various new domains ever since as a measurement. In Shannon information theory (Shannon, 1948), the entropy is used to measure the uncertainty over the true content of a message (a string of binary bits). Mathematician Alfréd Rényi constructed the proper entropy for fractal geometries (Jizba and Arimitsu, 2001). In statistical mechanics, the entropy is defined as a function of statistical probability to measure the probability for a given macrostate. Following the information-theoretic measure by Boyer and Kak (1988), Vosselman (1992) presented a new evaluation measure for comparing relational descriptions. In this work, we introduce an attribute entropy function as follows to measure how well a type of attribute describes a given data in terms of its pattern for matching:

$$Ep = \frac{1}{n-1} \sum_{i=1}^n |R_i - \bar{R}| \quad (5)$$

where R_i is each measured attribute value, \bar{R} is the arithmetic mean of the population, n is the number of the R_i . A high entropy indicates salient pattern represented by such attribute, and a low entropy indicates weak pattern represented by the attribute. The defined attribute entropy Ep is similar to unbiased estimator of sample variance σ^2 where the two parameters μ and σ^2 are estimated from the data itself. Attributes that has most distributed values is useful for salient patterns and should be selected for matching.

2.4. Matching of two networks via attributed graphs

Accordingly, the road networks from the image and vector data are respectively defined through a graph embedded topological (e.g. basic loop) and geometric attributes (e.g. relative distance). Using the above notations for these two networks from two datasets, our aim in matching is to optimally correspond (label) nodes v_i in graph G^{im} to those in graph G^{db} satisfying certain matching criteria. Now the problem of matching an image to a vector data becomes a matching of attributed graphs.

Establishing such matching is not a straightforward task. In general, the giving matching task requires an effective search mechanism in order to prevent a combinatorial explosion. The matching process can also be further complicated by errors in road extraction due to the presence of noise in scenes, like building-induced shadows and occlusions. Furthermore, land covers may vary in different time of data acquisition. Specifically, roads may change geometrically (e.g. in width, curvature), emerge or disappear in imagery compared to vector data. The matching, therefore, can be inexact correspondence: elements (vertices or edges) in one graph may not have their counterparts in the other.

In this work, we propose an optimal matching that maximizes similarities between two graphs both in topological relations among vertices and their geometric attributes to overcome the matching ambiguity and graph corruption. The matching process, based on continuous relaxation labeling, iteratively re-labels the data nodes with model nodes by changing their corresponding weights. The weights are optimized according to their local geometric and topological similarity. After each iteration, the global matching (i.e. global compatibility) is measured. The process reaches an optimal matching when the global compatibility mea-

surement becomes unchanged or varies to a limited threshold. We detail the matching process in the following subsections.

2.4.1. Local similarity

Once we have constructed the attributed graphs from two networks, we proceed with their local similarity. Our aim at this stage is to measure the similarity in structure and geometry associated with each mapping nodes. The local similarity is also termed the *goodness of the local fit*, *local compatibility* or the *goodness of local mapping*. They are used interchangeably in this paper. Given V_j from G^{db} as the current counterpart (label) of v_i in G^{im} , let $\{v_s, \dots, v_q\}$ be neighbor vertices of v_i and $\{V_p, \dots, V_t\}$ be any neighbor vertices of V_j . We introduce an exponential function, a modified version from the work of Li (1992), to measure the goodness of such mapping ($v_i \rightarrow V_j$). Using the relative distance attribute as an example, the goodness of the local fit can be measured with $H(v_i, V_j)$:

$$H(v_i, V_j) = \exp \left(- \sum \frac{\min |r_{i,\{s,\dots,q\}} - R_{j,\{p,\dots,t\}}|}{\sigma} \right) \quad (6)$$

where σ is some parameter, $r_{i,\{s,\dots,q\}}$ is the relative distance from v_i to its neighbors in G^{im} , $R_{j,\{p,\dots,t\}}$ is the relative distance from V_j to its neighbors in G^{db} .

Use Fig. 4 as an example. Let us assume we are considering labeling v_2 for V_1 in the process. According to Eq. (6), the summation component in $H(v_2, V_1)$ includes two ingredients since v_2 has two adjacent vertices v_1 and v_4 . Specifically, relative distance r_{21} is compared with all relative distance $R_{1,t,t \in \{2,3,4\}}$ and the most proximate one (say $t = 2$) is used to calculate the absolute difference as one ingredient. The relative distance r_{24} is then compared with the rest of relative distance $R_{1,t,t \in \{3,4\}}$ and the most proximate one (say $t = 4$) is used to calculate the absolute difference as the second ingredient. With relative distance attribute and adjacency constraints, the vertex ($t = 3$) in neighborhood of V_1 matches no neighbors of v_1 . We assign it to match null, a special label in V^{im} .

The novel feature of this local consistency measure H is its compound exponential structure, which distinguishes it from many alternatives in the literature. The underlying advantages are that the constructed H function will not be affected by the presence of noise (i.e. the additional link V_3 in Fig. 4) and the ambiguity will be reduced as low as possible. Similarly, the presence of noise (i.e. additional links) in V^{im} would not affect our measurement.

From entropy analysis, more than one type of attributes may be needed. For a general purpose, $H(v_i, V_j)$ can be defined as:

$$H(v_i, V_j) = \omega_1 \exp \left(- \sum \frac{\min |r^1 - R^1|}{\sigma^1} \right) + \omega_2 \exp \left(- \sum \frac{\min |r^2 - R^2|}{\sigma^2} \right) + \dots + \omega_n \exp \left(- \sum \frac{\min |r^n - R^n|}{\sigma^n} \right) \quad (7)$$

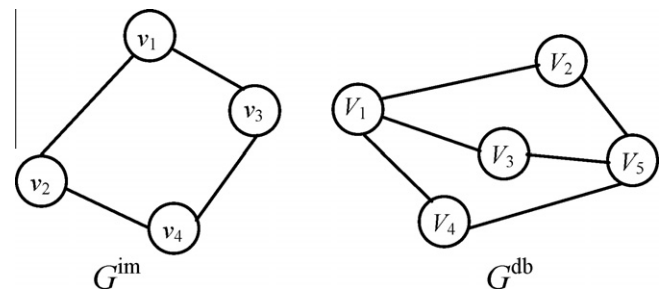


Fig. 4. Two graphs used for exemplifying local similarity measure.

where $\omega_1, \omega_2, \dots, \omega_n$ are weights that denote the importance of their associated attribute and the superscripts on r, R or σ represent the cardinal number of attributes. For instance, number 1 denotes unary attributes and 2 binary attribute, 3 ternary attributes, etc. The proposed function H corresponds to the compatibility function of Hummel and Zucker (1983). With function H , the local difference between V^{im} and V^{db} under their associated geometric and topological constraints is mapped into a similarity measure for assigning V^{im} to V^{db} and special node null.

2.4.2. Optimal mapping

As discussed before, matching two attributed graphs that model road networks requires an optimal solution, which maximizes a sort of global compatibility. With the constructed local similarity H , we use the continuous relaxation labeling method introduced in (Hummel and Zucker, 1983; Rosenfeld et al., 1976) which require no threshold to determine whether a mapping is acceptable. As the continuous relaxation-labeling framework, probability values other than logical assertions (1 or 0) are attached to all possible assignments for each vertex in G^{db} and null vertex. The probability with which label V^{im} is assigned to vertex V^{db} (including null) is denoted by $p_\mu(\lambda)$ and satisfies:

$$0 \leq p_\mu(\lambda) \leq 1, \quad \lambda \in V^{im}, \quad \mu \in \{V^{db}, \text{null}\} \quad (8)$$

and

$$\sum_{\mu(\lambda)} p_\mu(\lambda) = 1, \quad \lambda \in V^{im}, \quad \mu \in \{V^{db}, \text{null}\} \quad (9)$$

Let Θ be all available assignments with V^{im} to V^{db} and null. The global compatibility function (using relative distance attribute as an example) can be formed as:

$$A(\Theta|d, m) = \sum_{i,j,k,a,b,c} H(v_i \rightarrow V_a, v_j \rightarrow V_b, v_k \rightarrow V_c) p_a(i) p_b(j) p_c(k) \quad (10)$$

where $v_i, v_k \in N(v_j)$ and $V_b, V_c \in N(V_a)$. This function is close in nature to the global gain in (Li, 1992). Yet, with the data structure and attributes we proposed in this work, the neighborhood can easily be identified through edges of graphs. This is important to network matching as it allows our algorithm to determine the local similarity based on the topological structure intrinsic inside the data itself, not any ad hoc information. Thus, the optimal labeling of G^{im} with G^{db} will be the one that maximizes the above function:

$$A(\Theta^*) = \max(A) \quad (11)$$

The gradient projection algorithm by Hummel and Zucker (1983) is used. The advantage of the algorithm lies in its projection operator which is based on a theory of consistency and still widely used in solving constrained optimization problems. An exhaustive discussion and proof can be referred to (Mohammed et al., 1983). With the algorithm, we can iteratively compute the length and direction of the updating vector to update p such that the global compatibility function A will increase with each updating of p . The iteration terminates when the algorithm converges, generally producing an unambiguous labeling (or matching).

3. Matching using multilayer data

In this section we introduce the second stage of our approach to incorporate additional scene information in the form of object configurations present within individual road network loops (e.g. as they may become available from other GIS layers). By modeling this information in addition to the road network itself we developed an approach for scene matching based on its complex multilayer content. By *complex*, we mean that we make use of more than

one geospatial feature class or multiple layers of GIS data (e.g. roads and buildings). This allows us to incorporate additional scene information in our model, thus enhancing its descriptive power, and improving the potential accuracy of matching. In addition, we assume these objects on both data sets have been extracted, using standard digital image analysis techniques. Among various geospatial features, road networks and buildings are selected in our approach for matching due to their common availability. Furthermore, substantial research efforts have been dedicated and progress has been made for their automated identification and delineation. Recent examples include the ridge detection efforts of Gautama and Borghgraef (2003), or the template detection approach of Tipdecho (2002). It is easily understood though that we could have any other layers in addition to, or instead of these types, including for example vegetation outlines, or land lot outlines.

3.1. Content measure: spatial configuration

As the two compared representations of a scene may vary in scale, coverage, or orientation, it is important to use invariant metrics. Let us use the same notation as in Section 2: consider the graph of the image data to be represented as G^{im} and the one from database as G^{db} . Correspondingly, V^{im} is the vertex set of G^{im} and E^{im} the edge set of G^{im} , V^{db} the vertex set of G^{db} , E^{db} the edge set of G^{db} .

The comparison of two scenes based on their content requires the development of quantitative metrics to assess the similarity of the corresponding spatial configurations. Considering buildings in particular, their arrangement in a scene is considered equivalent to the configuration of their corresponding Minimum Boundary Rectangles (MBRs). In terms of topology, two major spatial relations between buildings may be identified: disjoint or meet. It is this type of information that we model in a topology similarity metric.

In Fig. 5 we show a scene comprising buildings A and B. The topology similarity metric F_{AB} expresses their relative position as a function of the external minimum distance (d_e^{AB}) between their MBR boundaries, the diagonal of the MBR of the union of building MBRs (diagonal($A \cup B$)) the common length (l_o^{AB}) of the building MBRs along their parallel direction and the average length of the two MBRs' sides along the parallel direction.

$$F_{AB} = \frac{d_e^{AB}}{\text{diagonal}(A \cup B)} + \frac{l_o^{AB}}{L_A + L_B} \quad (12)$$

If $d_e^{AB} = 0$, the expression of Eq. (12) is reduced to its *meet relation equivalent*. This topology similarity metric measure represents a modification of the measure introduced by (Godoy and Rodrigues, 2002). By accounting for the characteristics along the direction of building parallels, Eq. (12) embeds more expressive power in the topological measure, allowing it to differentiate extreme cases like those shown in Fig. 6:

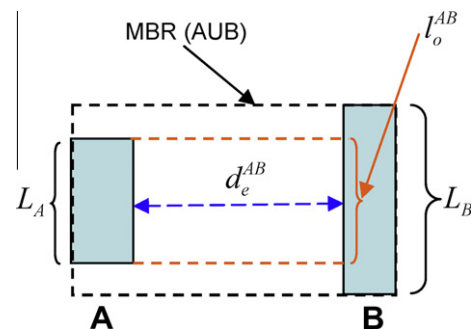


Fig. 5. Modeling the relative position of two objects.

- Comparable meet configurations composed of two similar buildings that differ only on the overlapping length of the meeting sides (Fig. 6 top).
- Comparable disjoint configurations composed of two similar buildings that are similar in their relative (horizontal) distance but differ in their vertical arrangement (Fig. 6 bottom).

3.2. Association graph for content similarity

Association graphs are constructed and used to assess content similarity between an image and candidate database scenes. Initial correspondences of objects are necessary to form the graph nodes. Accordingly, and without any initial knowledge, when matching a scene with a number of buildings to another candidate entry, one building can have any counterpart in the matching scene, leading to very large numbers of potential permutations, and imposing computational constraints. In order to overcome this problem we introduce in this section a *view-consistent* approach to select potential candidates.

As discussed before, our content matching proceeds iteratively with road network matching. Furthermore, during road network matching, each node from G^{im} is iteratively compared with every node from G^{db} . This node comparison assigns a local similarity score to this pairing, taking into account two components: (1) the similarity of local road structure around these nodes (see details in Section 2); and (2) the similarity of scene content in the road loops that correspond to the compared nodes.

Thus, each content similarity assessment is associated with a pair of nodes considered as potentially corresponding from the two analyzed data sets. The two compared nodes (e.g. v_1 from G^{im} and V_2 from G^{db}) provide a reference for each scene and are referred to as *reference points* (or *view points*). From the local similarity of road network structure, we can identify two other nodes (e.g. v_2 from G^{im} and V_1 from G^{db}) which are closest to v_1 and V_2 . With these four nodes, we can create reference axes for each scene: $v_1 \rightarrow v_2$ and $V_2 \rightarrow V_1$.

For each building in the image, a bounding view angle (refer to α in Fig. 16a as an example) for the building's MBR can be measured relative to the reference point v_1 (U_1 in Fig. 16a) and reference line $v_1 \rightarrow v_2$ ($U_1 \rightarrow U_2$ in Fig. 16a) in the image. The angle is positive when the view is clockwise from the reference axis, negative if counterclockwise. In the database scene, the same view angle is reconstructed, but using the reference point (V_2) and reference line ($V_2 \rightarrow V_1$) in this GIS scene. All buildings inside the database loop that overlap the bounding view angle by more than a preset threshold are considered to be potential matches for the corresponding building in the aerial image. This is equivalent to

having a viewer positioned in the ground at the road intersection and looking at the buildings in the corresponding city square. The viewing angles show the relative position of buildings within this square. If another scene is comparable to it, the corresponding building viewing angles are expected to match.

This bounding view angle analysis generates nodes of the association graph comparing the two scenes. We measure the similarity between corresponding buildings using area ratio, and assign the score to the association graph node:

$$S_{X:A} = \begin{cases} \frac{R_Area(A)}{R_Area(X)} & \text{if } R_Area(X) > R_Area(A) \\ \frac{R_Area(X)}{R_Area(A)} & \text{if } R_Area(X) \leq R_Area(A) \end{cases} \quad (13)$$

$S_{X:A}$ is the similarity score (between 0 and 1) for building X from the image, compared to building A from the GIS scene and $R_Area()$ is the footprint area ratio of the corresponding building. The larger the value of $S_{X:A}$ is, the more similar the two buildings.

Two nodes (e.g. node (X:A) and node (Y:B) in the association graph are linked by an edge if and only if the spatial relation of X and Y is qualitatively same (i.e. disjoint) as the spatial relation of A and B. Thus the value associated with the edge ((X:A)(Y:B)) can be calculated using Eq. (12) to represent the similarity of the spatial relation (XY) in the image, with respect to the relation (AB) in the GIS scene:

$$S_{X:A \sim Y:B} = 1 - |F_{XY} - F_{AB}| \quad (14)$$

F is the expression of content spatial configuration introduced in Eq. (12). The larger the value of $S_{X:A \sim Y:B}$ is, the more similar the relation is.

Buildings in the image may not have their counterpart in the GIS scene (or vice versa) due to physical changes of the buildings between two datasets (e.g. representing the same scene at different time instances) or object extraction errors (e.g. due to shadows or intrusions). We extract *maximal clique*, i.e. a maximal complete subgraph (Nedas and Egenhofer, 2008), as the solution for the constructed weighted association graph. The clique-enumerating algorithm (Tomita et al., 2006) can be used to extract the maximal clique.

3.3. Scene completeness

While successful matches provide input in the assessment of similarity, we want to ensure that unmatched buildings will also contribute to alter accordingly (negatively) the result. This is meant to reflect the manner in which humans assess similarity among scenes: the presence of unaccounted objects in two scenes reduces our certainty of their similarity. (Nedas and Egenhofer, 2008) have addressed this issue and we use the *completeness* metric they developed, while considering each object to be of equal importance as they are the same type of objects

$$S_{comp(im,db)} = \frac{M}{M + (n - M) + (N - M)} \quad (15)$$

where M is the number of matched objects, n is the number of all objects in the image and N is the number of all objects in the GIS.

This completeness measure penalizes scenes with unmatched objects and thus reduces the scene similarity. Given the extracted maximal clique and scene completeness, the final scene similarity measure is then defined as:

$$S_{scene} = \left(\sum S_{nodes} + \sum S_{edges} \right) * S_{comp} \quad (16)$$

It should be pointed out that, for the final scene similarity, a general version should consider assigning weights for object similarity (i.e. nodes), object relation similarity (i.e. edges) and the completeness measure. This has been discussed extensively in (Nedas and Egenhofer, 2008).

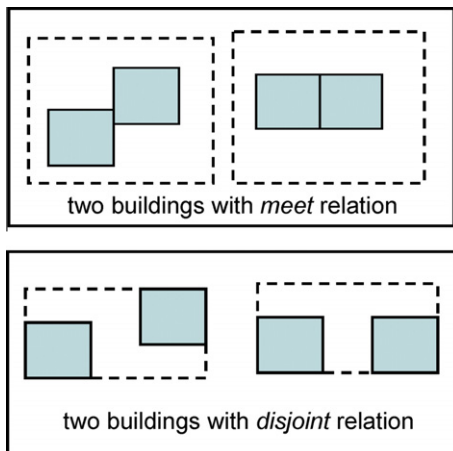


Fig. 6. Extreme cases affecting topological analysis.

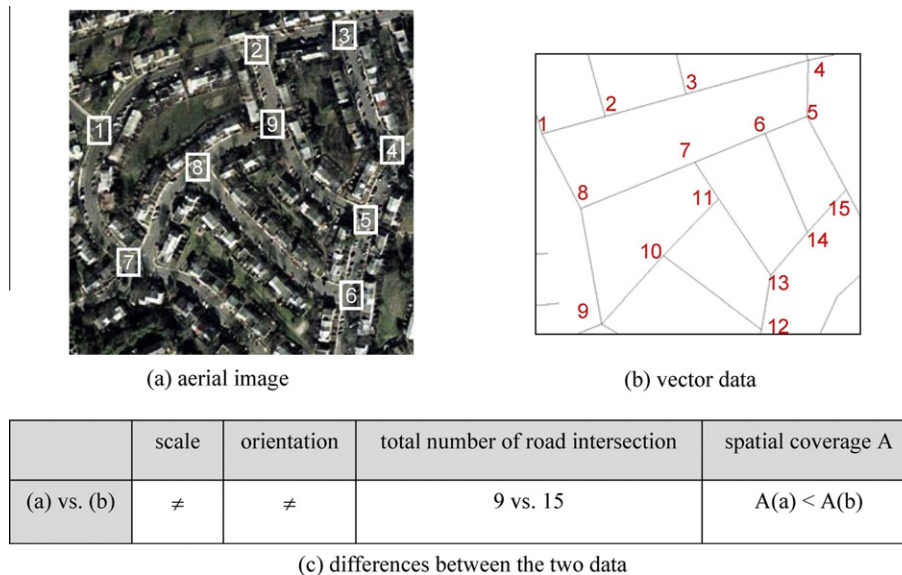


Fig. 7. The original imagery and vector data with differences in scale, orientation and coverage (numbers are corresponding to those in their graph models).

3.4. Optimal matching

This optimal matching process is the same as that using single layer, but with a new function H for the local similarity. Given V_k from G^{db} as the current label of v_i in G^{im} , the goodness of such mapping ($V_k \rightarrow v_i$) in Eq. (7) should be measured with $H(v_i, V_k)$ including both the geometric information and building information:

$$H(v_i, V_k) = \frac{\alpha S_{rd}^{v_i V_k} + \beta S_{scene}^{v_i V_k}}{\alpha + \beta} \quad (17)$$

where $S_{rd}^{v_i V_k}$ is the local similarity measure between v_i and V_k based on road network structure; $S_{scene}^{v_i V_k}$ is the local similarity measure between v_i and V_k based on partitioned scene content (i.e. building configuration); α and β are the weight given to the network structure measure and partitioned scene content measure respectively.

With function H , the local difference between V_k and v_i under the constraints in road structure and the content partitioned inside the closed loop is mapped into a similarity measure for assigning V_k to v_i . Based on relaxation labeling, the matching process iteratively re-labels the image nodes with database nodes by changing their corresponding weights. As the continuous relaxation labeling framework, weighted values other than logical assertions (1 or 0) are attached to all possible assignments for each vertex in G^{im} . The weight (denoted by $p(\lambda)$) with which label V_λ is assigned to vertex v_i belongs to $[0, 1]$. In addition, the sum of the weights for all possible assignments to any vertex should be equal to 1. Thus, optimal labeling will be the assignment that maximizes the global compatibility defined as in Section 2. The weights are updated using the gradient projection algorithm, which iteratively computes the length and direction of the update vector such that the global compatibility will increase with each updating of p . The iteration terminates when the algorithm converges, generally producing an optimal matching.

4. Experiments

The scene matching framework presented in this paper has been implemented in a MATLAB environment. In order to evaluate the performance of our approach and algorithms, we present three sets of experiments in this section. The *first set of experiments* offers an assessment of the performance of our single-layer network

comparison component introduced in Section 2. We consider data sets with representative variations encountered in registration tasks: vector data covering larger area than the image and differing in translation, rotation, and scale. Additionally, noise is introduced (e.g. missing links between road intersections, demolished/new buildings) to better approximate the effects of object extraction from aerial imagery. In the *second set of experiments*, we demonstrate the performance of our scene similarity metrics (introduced in Section 3), as a means to identify similar configurations of objects. There we consider the use of more than a single layer (road network and buildings) in this case, and demonstrate the matching mechanism. Finally, in the *third set of experiments* we demonstrate the overall performance of our approach in a larger area, to provide a better understanding of its overall functionality.

4.1. Matching using road networks

The used imagery and vector data are queried from the National Map Seamless Server.¹ The imagery has spatial resolution of 0.5 m that covers some area of the county of Prince William, VA and has been orthorectified. The vector data covers larger area than the imagery and presents unknown differences. The two datasets are shown in Fig. 7a and b respectively. Fig. 7c summarizes aforementioned differences.

From the datasets, two attributed graphs can be derived and they are marked as G and H below (Fig. 8), where G represents the attributed graph built from the road network from the satellite image; H represents the attributed graph built from the road network in the vector data.

For each graph, we derive two types of attribute based on the datasets: degree on each node and relative distance between nodes. Before proceeding to the matching process, we analyze these properties using entropy analysis. The analysis is significant as it allows us to quantitatively measure the distribution of attribute values, an important factor for reducing local ambiguity. Based on the analysis, it comes to a decision about whether or not both types of attribute should be used or additional attributes are needed. Fig. 9 shows their distribution graphically ordered by nodes and their corresponding entropy using Eq. (5). Node 7 in

¹ <http://seamless.usgs.gov/>

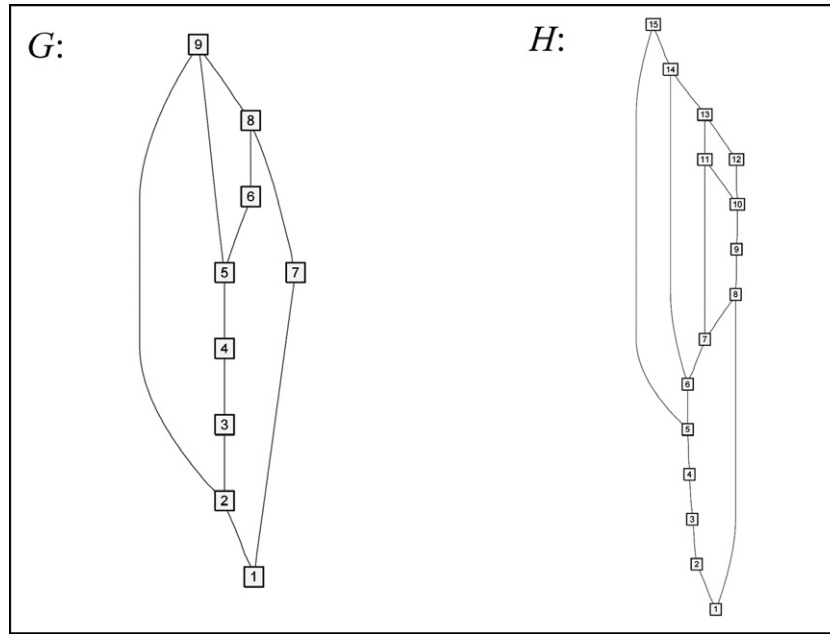


Fig. 8. Attributed graphs G from the imagery and H from the vector data.

Fig. 7a shows a salient degree-4 (there are four road segments in the image incident to the intersection represented by node 7) from others. Similarly, node 4 and 9 in Fig. 7b. Because the majority of nodes in (a) and (b) have close degrees (same degree in this case), accordingly their entropy used to evaluate the distribution is very small, close to zero. This means the node degree attribute is not appropriate to use for matching due to their proximate values. The relative distance attribute, however, has great entropy (96.92 and 67.76), representing distinguishable patterns locally and thus appropriate and sufficient for matching.

It should be noted that, although the degree attributes score low entropy, they still include useful information for initiating the assignments to start the matching process. The underlying rationale is that a matched pair mostly likely has the same number of degree. Our initial assignment is shown in Table 1. The numbers in the first row denotes nodes from graph G and the numbers in the first column are nodes from graph H . The special node in H is represented as null. Note that the sum of each column is equal to 1, satisfying the condition addressed by Eq. (9).

We demonstrate the novel feature of the local consistency measures in Fig. 10, where stars are global compatibility over iteration with relative distance attribute. One may note that the global compatibility has a steep increase with first iteration. This characterizes the local similarity measurement we proposed having the advantage to fast approach the optimal matching locally without much ‘confusion’.

To evaluate the performance of our approach when errors present in the data, we assume there is a new road appearing in the image between intersections denoted by node 1 and node 8 in graph G , but the vector data has not been updated yet. There is no edge connecting node 8 and 10 in graph H . Furthermore, there is a road in the vector data between intersections denoted by node 3 and 5 in graph H . This road on the image, however, is missed due to extraction errors. Fig. 11 shows that, under presented errors, the global compatibility present a similar characteristic (steep increase) as in Fig. 10, representing a distinct feature of our approach.

Both matching results are summarized in Table 2, where the correspondences between nodes as a result of the network match-

ing are shown column-wise. It can be easily seen that all nodes were matched correctly despite of coverage and orientation differences, as well as presented errors between these two networks.

4.2. Performance in complex scenes

In this subsection, we perform our algorithms for complex scene matching with representative examples to demonstrate its performance.

4.2.1. Assessing similar object configurations

The objective of the experiment presented in this section is to demonstrate the function of the scene similarity metrics introduced in Section 3, as a means to identify similar configurations of objects. Thus we consider as an example a scene comprising three buildings (X , Y , and Z) located inside a road loop (U_1, \dots, U_4) as they may be identified in a GIS (Fig. 12a). This is the reference scene that will be matched to the content of 4 aerial images (Fig. 12b–e) to identify the image that contains a combination of buildings that is most similar to the one in the reference GIS scene.

To better serve the purpose of this experiment the four aerial image datasets have been selected to have a road loop that is similar in structure to the road loop in the reference scene. However, the four aerial images contain comparable but different sets of buildings in them.

We assume all buildings have been detected (see e.g. the work of Tseng and Wang (2003) and Suveg and Vosselman (2004)) and are represented by their MBRs. We also assume that the corresponding road networks have been extracted, using standard digital image analysis techniques, e.g. Hinz and Baumgartner (2003). In order to minimize the potential effect of semantic noise in the scene (e.g. tree-induced shadows and occlusions) we consider buildings that have a spatial footprint that is larger than a certain threshold (e.g. 2% of the corresponding road loop). We also consider all weights in the similarity assessment equation (Eq. (17)) to be equal.

Let us consider a comparison of the reference scene (a) to candidate image (b). We construct the association graph by selecting

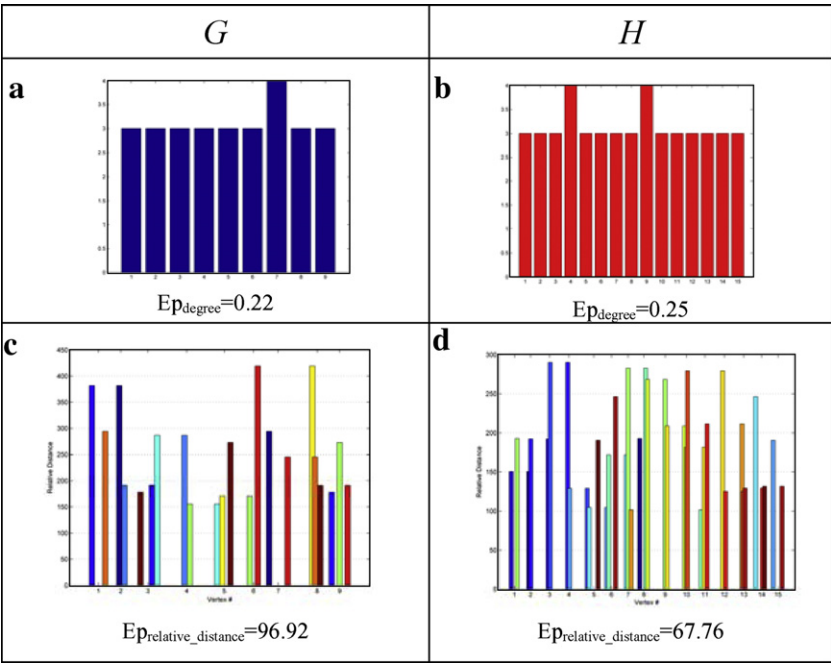


Fig. 9. Derived attribute populations (x – the node # as shown in Fig. 8; y – attribute values): (a) degree over each node in graph G ; (b) degree over each node in graph H ; (c) relative distances for each node to its connected neighbors in graph G ; (d) relative distances for each node to its connected neighbors in graph H .

Table 1
Initial matching assignments with degree attribute.

	1	2	3	4	5	6	7	8	9
1	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
2	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
3	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
4	0	0	0	0	0	0	0.5	0	0
5	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
6	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
7	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
8	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
9	0	0	0	0	0	0	0.5	0	0
10	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
11	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
12	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
13	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
14	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
15	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0	0.0769	0.0769
Null	0	0	0	0	0	0	0	0	0

the first object X in the GIS data and finding comparably located objects in the candidate image. In the reference scene we determine a *view angle* for object X relative to reference point U_1 and reference line $U_1 \rightarrow U_2$. We select as reference point the northwest (upper-left) node of the road loop that encloses the objects in the reference scene, and as reference line the north component of the enclosing road loop. Using this viewing angle we identify in image (b) two buildings that may be considered as conjugate images of building X : buildings b_1 and b_2 .

This step generates two nodes for our association graph: $(X:b_1)$ and $(X:b_2)$, indicating in essence that X may correspond to b_1 or b_2 . Similarly, the remaining components of the reference scene (buildings Y and Z) are analyzed and their potential counterparts are identified to construct the complete association graph as it is shown in Fig. 13. The association graph summarizes the potential matches between the reference scene and the scene contained in image (b). The linking lines in the association graph (e.g. $(X:b_1)$ and $(Y:b_4)$ are joined by a graph edge) indicate combinations of

associations that may co-occur. For example, $(X:b_1)$ and $(X:b_2)$ cannot be linked in the graph, because X can either match b_1 or b_2 , but not both of them at the same time.

The complete association graph contains 6 *maximal cliques*, or maximal complete subgraphs, which show potential scene configurations in the image that may match the reference scene. Fig. 14 summarizes these potential configurations for image b. We applied the metrics presented in Section 3 to assess the similarity between subgraph nodes (ratio of corresponding building footprints) and edges (topological relations between buildings) and thus assign a similarity score for each maximal clique. Clique 5 is the optimal match, with a similarity score of 0.343, whereby X is matched to b_2 , Y to b_5 , and Z to b_6 . It should be noted that clique 6 generates a score of 0.342, where X is matched to b_2 , while Y and Z are assigned to b_6 and b_5 .

The reason for having comparable results for these two configurations is as follows. The best selection among the ones available (clique 5) represents a topologically identical situation as the

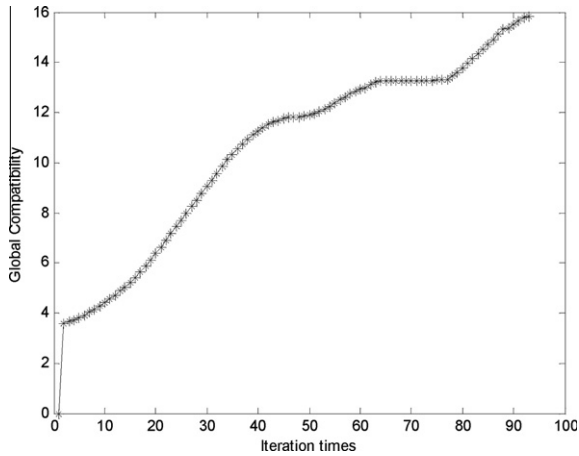


Fig. 10. Variation of the global compatibility over iterations.

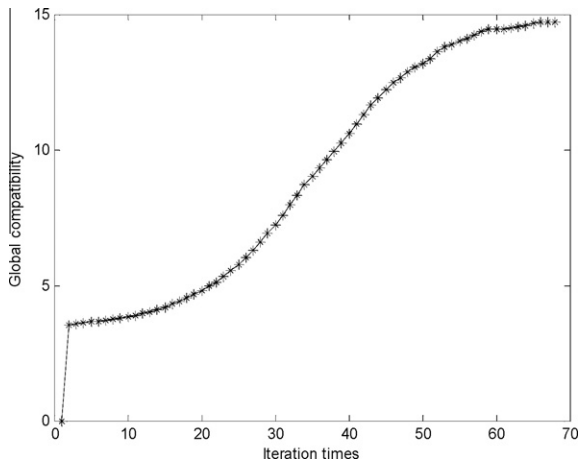


Fig. 11. Global compatibilities over iterations when errors present in data.

Table 2
Matching results from both experiments.

Nodes in G	1	2	3	4	5	6	7	8	9
Nodes in H	8	7	6	14	13	12	9	10	11

query. However, the two buildings on the right of the scene (Y and Z) display differences in their area to their potential counterparts (b_5 and b_6 , respectively). This causes the similarity metric among these two scenes to remain relatively low. In clique 6 we have the opposite situation, with an inverted topological relation but a better area match. It is easily understood that users may choose to alter the weights of these various metrics to emphasize more properties that are more important to the application they pursue.

By repeating this process using images (b) through (e) we find the optimal match for each and summarize the results in Table 3. We can see that scene (e) has the highest similarity score when compared to our reference GIS scene. Meanwhile, similarity among scenes may be impacted by the inaccuracy in feature extraction process and selected feature models. Accordingly, similarity measures could be slightly overestimated or underestimated. For instance, the occlusion introduced by trees near buildings may make two buildings, which are very differently shaped, look highly similarity with the minimum bounding rectangle model. In order to select a conclusive matching, we use standard deviation to

measure the confidence in our matching conclusion. These falling outside the range of chosen standard deviation of the population are considered statistically important and thus are considered to be the match. In this experiment, the similarity score of scene (e) is 0.6, which is the only one falling outside the range of one standard deviation (i.e. $\sigma = 0.1071$) of the score population ($X = \{0.343, 0.429, 0.473, 0.6\}$). It is therefore selected as the match to (a). A visual inspection confirms the correctness of this selection.

4.2.2. Assessing multilayer scene similarity

The objective of this experiment is to demonstrate the performance of our scene matching method using more than a single layer. More specifically, we consider the use of both a road network and buildings in assessing the similarity of two scenes. The datasets are shown in Fig. 15: (a) represents a high-resolution aerial image depicting a road loop and buildings within this loop; (b) is a broader scene from a database comprising roads and buildings. The scene in Fig. 15(b) contains three road loops that are potential matching candidates for the configuration of Fig. 15(a). The objective is to identify among these loops the best match for Fig. 15(a) in terms of both the structure of the road network that delineates it, and the buildings that exist within the loop.

The aerial image in Fig. 15(a) has been preprocessed, extracting roads and buildings (represented through their corresponding MBRs). The basic loop in this image (U_1, U_2, U_3, U_4, U_1) is referred to as W_{aerial} .

Comparing image Fig. 15(a) to the three basic loops of scene Fig. 15(b) we can make the following observations:

- The upper-left loop of (b), formed by the node sequence (V_1, V_2, V_5, V_6, V_1) is referred to as W_{db}^1 . It has a similar network structure in terms of connectivity and geometry to W_{aerial} . Furthermore, buildings within W_{db}^1 display high similarity to the certain building configurations within W_{aerial} .
- The upper-right loop of (b), formed by the node sequence (V_2, V_3, V_4, V_5, V_2) is referred to as W_{db}^2 . It has a similar network structure in terms of connectivity and geometry to W_{aerial} . However, the potential configurations of buildings within W_{aerial} display some differences from the buildings within W_{db}^2 .
- The lower loop of (b), formed by the node sequence ($V_4, V_5, V_6, V_7, V_8, V_4$) is referred to as W_{db}^3 . It displays dissimilarities when compared to W_{aerial} in terms of its network structure, namely different numbers of nodes and differences in the relative lengths of the corresponding edges. However, its building content is quite similar to building configurations within W_{aerial} .

Thus, the objective of this experiment is to demonstrate the combined effects of variations in road network structure or building configurations on the assessed accuracy. Similar to Section 4.2.1 we consider buildings that have a spatial footprint that is larger than a certain threshold (e.g. 2% of the corresponding road loop), and assign equal weights to the similarity assessment equation.

As the matching process is initialized we have each image node U_i assigned an equal probability to match any of the database nodes V_j . The iterative process presented in Section 2 allows us to update these probabilities until we reach a final solution with U_i nodes matched uniquely to V_j ones. Each updating loop assesses the similarity in road network structure and in the corresponding building configurations.

Let us consider for example that we assess an iteration where U_1 may match V_2 , in which case W_{aerial} may be compared to any of the two loops in which V_2 belongs, namely W_{db}^1 and W_{db}^2 (Fig. 16). In this setup, U_1 is the viewing position and $U_1 \rightarrow U_2$ is the reference line for the aerial image. We will identify potential building configurations within W_{db}^1 and W_{db}^2 that may match the one in W_{aerial} in

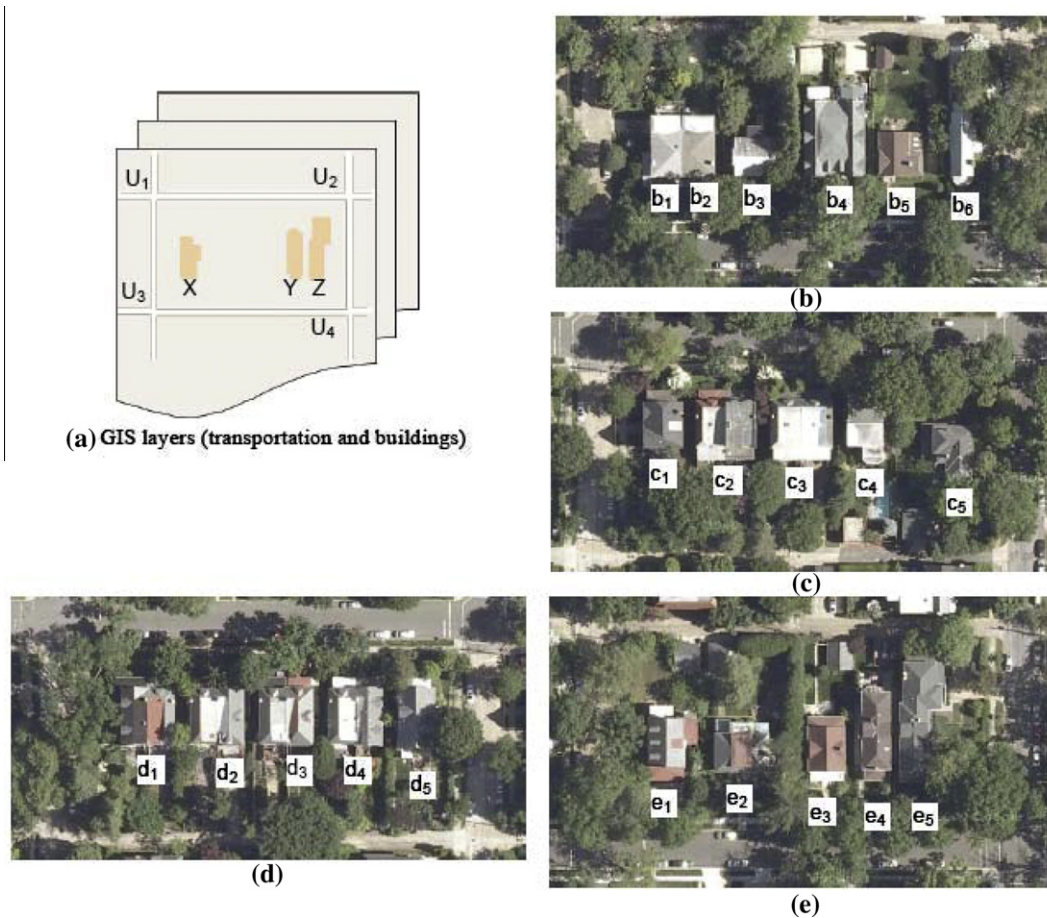


Fig. 12. A reference scene (a) and four aerial images (b–e) that are matched to it.

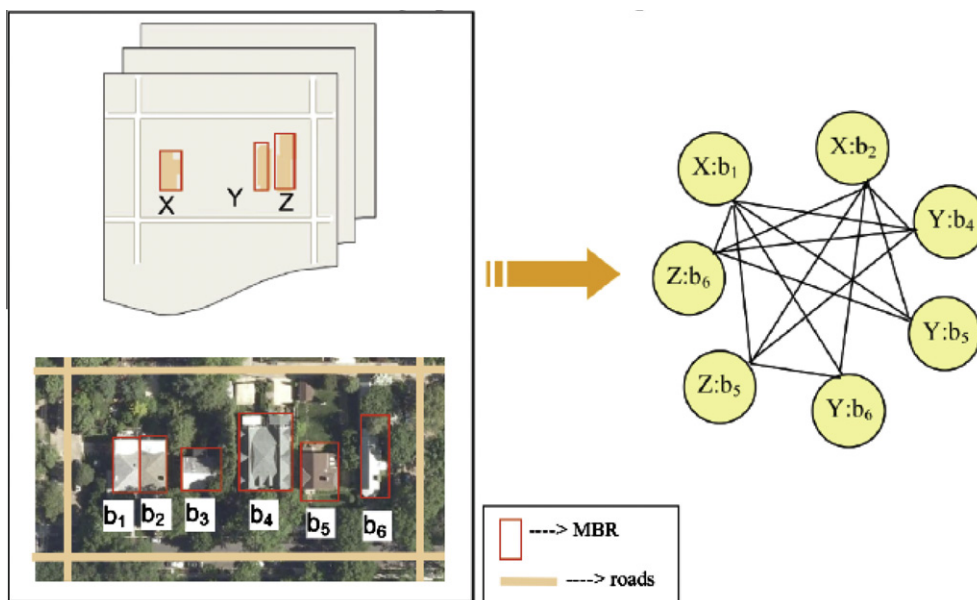


Fig. 13. Constructing the association graph for reference scene (a) and candidate image (b).

the following manner: for each building in the loop W_{aerial} , a *bounding view angle* can be measured from the reference axis, for example, $+\alpha$ for building 1 in Fig. 16(a). The angle is positive when the view is clockwise from the reference line. A similar bounding view angle is formed in W_{db}^1 using V_2 as viewing point and

$V_2 \rightarrow V_1$ or in W_{db}^2 using $V_2 \rightarrow V_3$ as reference line. In the GIS scene, all buildings that overlap the bounding view angle by more than a pre-set threshold (e.g. 75%) are considered to be potential matches for the corresponding aerial image building. This bounding view angle analysis generates nodes of the association graph comparing

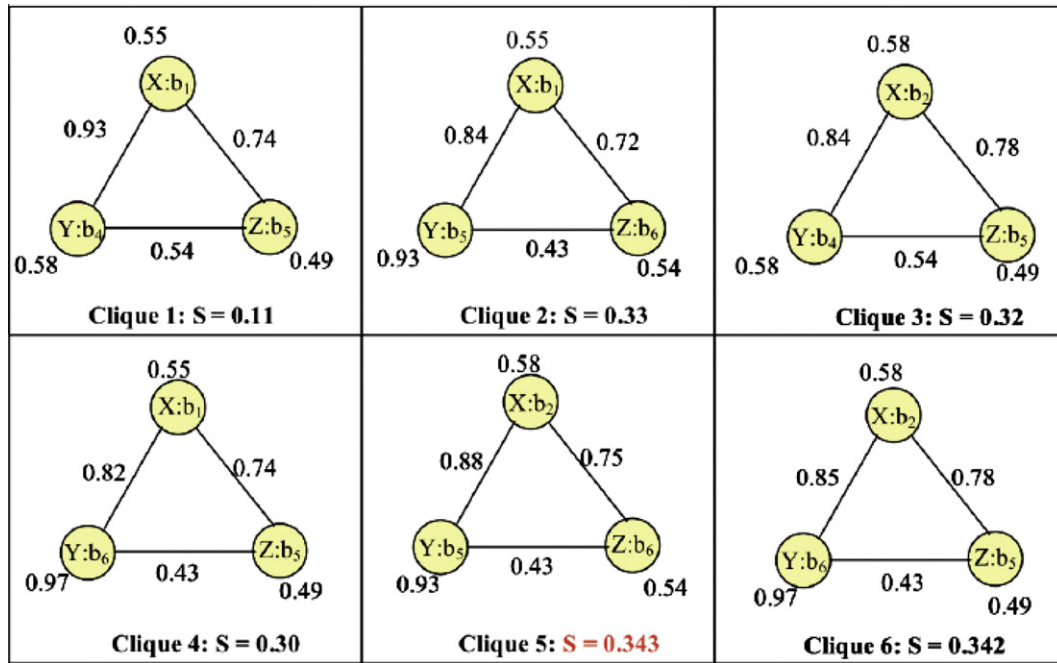


Fig. 14. The six similarity cliques and their corresponding similarity scores for each clique.

Table 3
Similarity scores between the reference scene and each of the (b)–(e) images.

Datasets	Association graph	Max similarity score	Optimal match
(a) vs. (c)		0.429	(X: c ₁) (Y: c ₄) (Z: c ₅)
(a) vs. (d)		0.473	(X: d ₁) (Y: d ₅) (Z: d ₄)
(a) vs. (e)		0.600	(X: e ₁) (Y: e ₄) (Z: e ₅)

the two scenes. In this example the new node would be (1:D), indicating that building 1 in W_{aerial} may match building D in W_{db}^2 . Similarly, a node (4:D) would be generated, and linked to (1:A) to express that these are two matches that could co-exist in the same loop.

Fig. 17 shows the constructed association graph for this example. After applying Eqs. (12)–(15) sequentially to the nodes and edges of the association graph, we measure the scene similarity using Eq. (16). The result with highest score represents the best match and is used for the local similarity in Eq. (17).

We perform the same procedure for the local similarity of each pair and optimize the matching probability iteratively. In this specific example, after the global compatibility function converges and gives the correct matching result shown in Table 4. The corresponding association graph and similarity scores are also shown in the table. This similarity score is the highest among all pairs of potential subscenes. We see that the process has successfully

matched the aerial image loop W_{aerial} and its building content to loop W_{db}^1 . Differentiating it successfully from the other potential matches, W_{db}^2 and W_{db}^3 which demonstrated deviations in terms of their building scene content and road network respectively.

4.3. Performance using larger datasets and operational considerations

This third set of experiments uses larger datasets, to offer a better understanding of the overall performance of our approach. In this study, a scene extracted from a query image (e.g. by an analyst delineating it on-screen) is compared against a search window selected in Washington DC in order to identify a match and geolocate the image.² Fig. 18 shows the problem at hand, with a scene comprising four city blocks delineated by solid lines in the query image. Within the broader DC metro area (an area of 177 km²) we have identified a search window of approximately 7 km² within which we will search for the query scene (identified by the dashed-line window in the DC map, and enlarged to show detail on the upper right of the figure).

We present in this section results from matching the query scene to data within this specific search window, but it is easily understood that this can be part of a parallelized process: when needed, a very large search area can be tessellated arbitrarily (or otherwise, if additional information is known a priori) into smaller search windows (25 in our case would cover the complete DC area) and each of them can be processed separately, even using different computers if time is of the essence. Needless to say, this tessellation can be made into overlapping windows to minimize (at a computational cost) the adverse effects of having the query scene falling between two search windows. While there may exist some scale variations between query and dataset we expect them to be global in scale, and not vary locally. Accordingly, the extent of the overlap area can be determined as a function of the size of the query scene itself (e.g. 1 × to 2 × the equivalent size of the query window in the search dataset). Also, window sizes could vary to adjust to the available computational resources. This ability

² Note that roundabouts in the vector data are reduced to a single point.

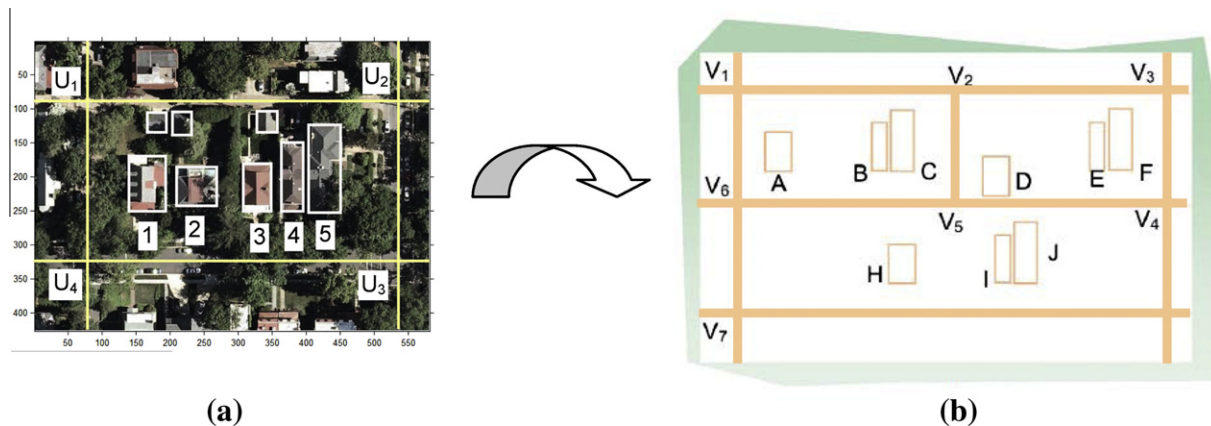


Fig. 15. (a) Image data with extracted road network and MBRs of buildings; (b) GIS data composed of a road layer and a building layer (buildings are shown in MBRs).

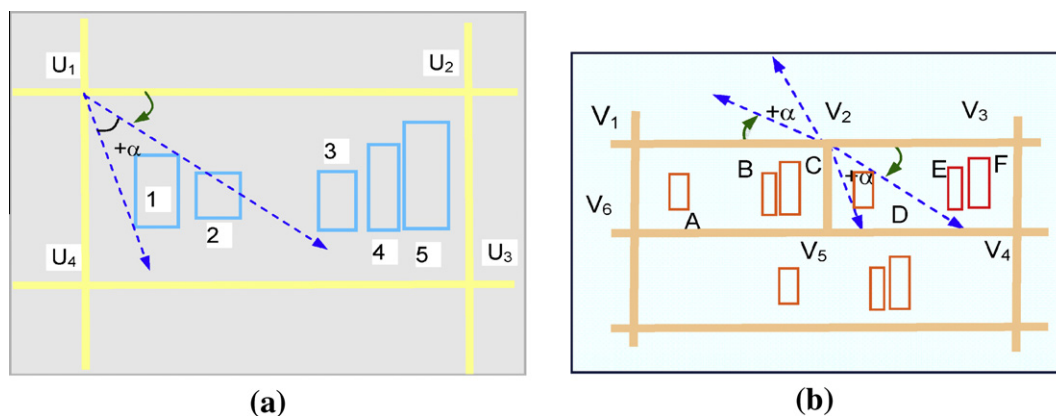


Fig. 16. An example for scene similarity where U_1 is assigned a label of V_2 : (a) U_1 as viewpoint and $U_1 \rightarrow U_2$ as the reference line; (b) V_2 as viewpoint and $V_2 \rightarrow V_1$ or $V_2 \rightarrow V_3$ as the reference line (in both figures, the dotted lines are view lines).

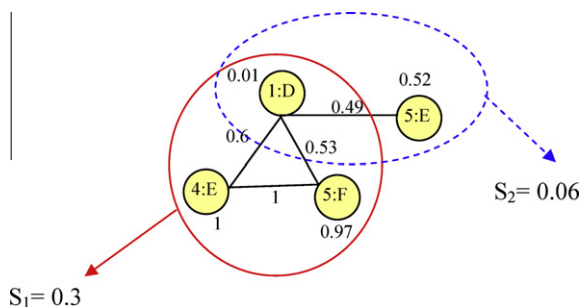


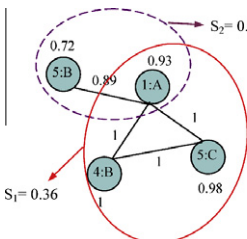
Fig. 17. The association graph for the example: the maximal clique S_1 (enclosed in solid circle) gives the highest scene similarity score 0.3.

to parallelize this operation is an important aspect when considering how the approach scales for large-scale applications. In the experiments presented in this paper we used what can today be considered as an average laptop configuration (HP Notebook with i7 CPU, Quad 720 @1.6 GHz).

As we see in Fig. 18, the query scene comprises various elements, including a triangular portion at its top. This is an example where an analysis of the content of the query can be used to speed up the search, as triangular road loops are much rarer than the traditional rectangular ones. Searching triangles in a graph (a case of 3-clique in graph theory) is a widely used tool in network analysis. Examples can be found in (Brandes and Erlebach, 2005;

Eubank et al., 2004; Watts and Strogatz, 1998). We proceed by applying a fast search algorithm proposed in (Schank and Wagner, 2005) to identify triangular road loops in the available road network for the area of interest. This can be augmented by adding angular or distance constraints in the loop parameters, depending on the accuracy with which this information is available. In our case the search for triangular segments leads to the selection of potential match candidates and corresponding search sub-regions containing this type of loop (Fig. 19). This is a search space refinement step, and thus we are not interested in delineating precisely these search windows, just have them large enough to eliminate the possibility that a portion of the query will be left outside them. As we see in Fig. 19, within the search area of Fig. 18 we have identified 11 sub-regions within which we will proceed with more precise matching. Within each one we have an element (triangular road loop) that serves as anchor for the subsequent search: we will attempt to expand around it in order to identify broader scenes containing it that match the particular configuration. As the corresponding elements may differ in scale, these sub-regions may also be of different size. This portion of the process took only 3 s in the above-specified laptop. For each sub-region we can identify a list of the included road intersection (corresponding to graph nodes), and for this specific experiment this is tabulated in Table 5. Thus the query search space has been progressively reduced from the 7 km² area with which we started in Fig. 18 to a set of 11 sub-regions with less than 50 road intersection nodes each. It should be noted here that while in our case

Table 4
Matching result and the association graph.

Matches to	U_1 V_1	U_2 V_2	U_3 V_5	U_4 V_6
The association graph with similarity measures				

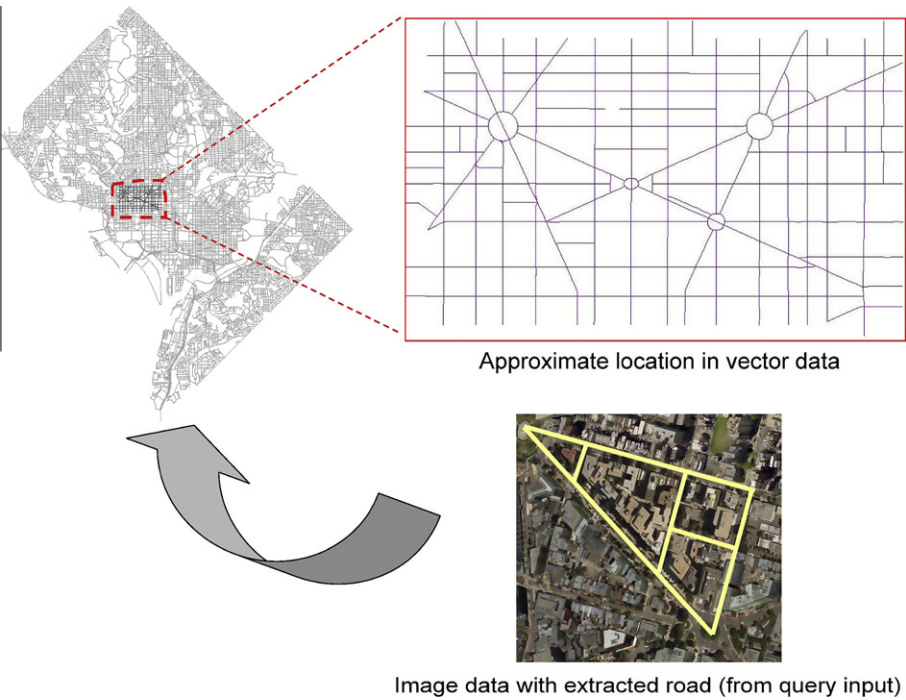


Fig. 18. A scene comprising four blocks (delineated in solid lines) selected in a query image (lower right); it is compared against a search window (top right) selected in Washington DC area (top left).

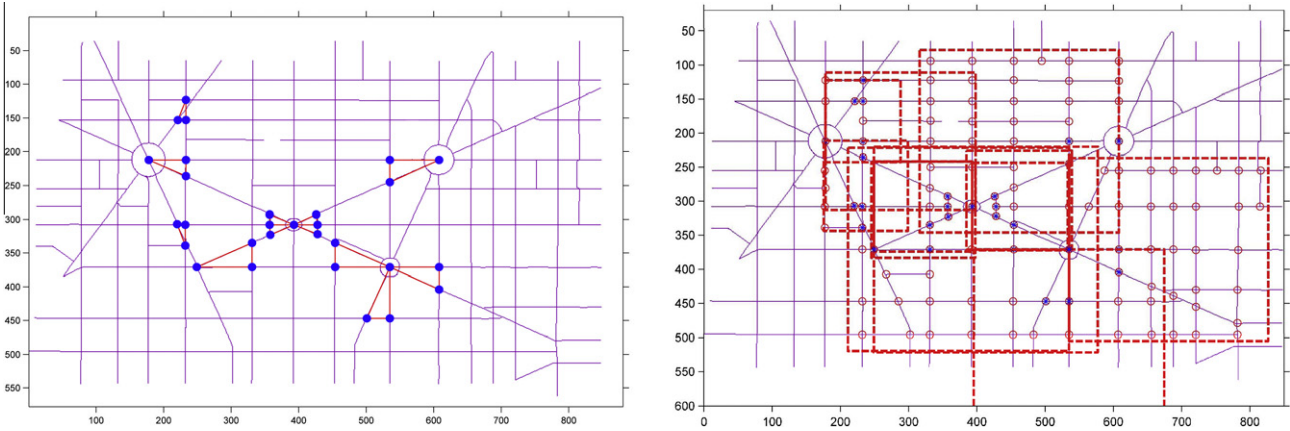


Fig. 19. Identified potential matching loops (left) and corresponding search sub-regions (right).

we used a triangular loop element to do so, any other configurations of properties can be used during this process (e.g. searching

for rectangular elements of particular proportions arranged in specific manners). Depending on how unique these properties

Table 5
Sub-regions and included road intersections.

Sub-regions	Bounded road intersection ID by each sub-region	Total #
1	19,20,32,33,34,49,57,58,75	9
2	19,20,21,22,32,33,34,35,36,49,50,51,57,58,59,60,75,76,77,90,91,97,98,99,100,101,103	27
3	6,7,8,9,10,11,21,22,23,24,25,35,36,37,38,39,50,51,52,53,59,60,61,62,63,76,77,78,79,80,81,91,92,99,100,101,102,103,104,105,106,107,108,109,110,170,171	47
4	57,58,74,75,90,96,97,98,116,117	10
5	76,77,91,99,100,101,102,103,122,123,124,170	12
6	77,78,79,92,103,104,105,106,107,108,124,125,126,171	14
7	77,78,92,103,104,105,106,107,125,171	10
8	76,77,78,79,91,92,99,100,101,102,103,104,105,106,107,108,109,122,123,124,125,126,132,133,140,141,142,143,144,145,157,158,159,160,161,162,170,171	38
9	75,76,77,78,79,91,92,97,98,99,100,101,102,103,104,105,106,107,108,117,121,122,123,124,125,126,132,133,139,140,141,142,143,144,145,156,157,158,159,160,161,162,170,171	44
10	79,80,81,82,83,84,85,86,87,108,109,110,111,112,113,114,115,126,127,128,129,130,131,134,135,145,146,147,148,149,150,151,152,162,163,164,165,166,167	39
11	125,126,127,128,134,135,143,144,145,146,147,160,161,162,163,164	16

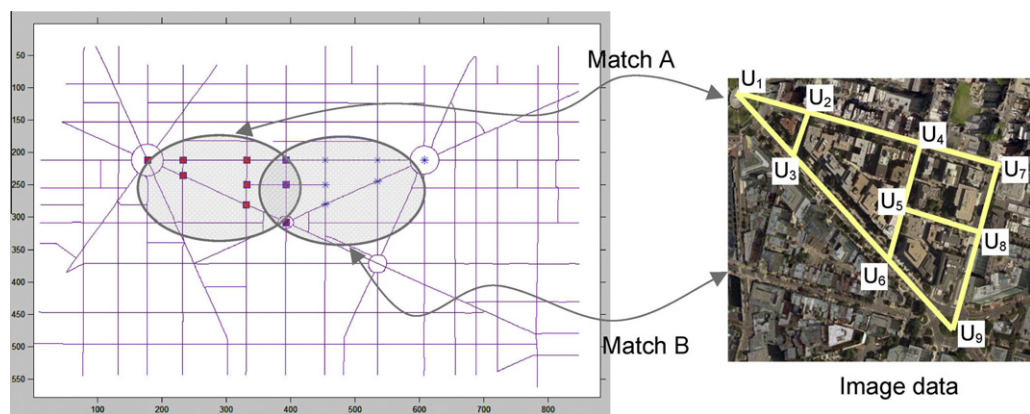


Fig. 20. The matching results from the first stage of our approach (building layouts in both potential matches are shown in gray).

are compared to the overall search space, the resulting gains in search space may be large or limited.

Once the initial anchor locations and search sub-spaces have been identified we can proceed with the broader road network matching using road network properties presented in Section 2.2. The attributes we use are invariant to scale, orientation, and rotation. No matter how the patterns are translated, rotated and scaled, these descriptions are required to hold in order for a correct match to be considered correct. Here, in view of the small entropy (<40) of relative distance of the data, we follow the invariant attribute for point patterns introduced in (Li, 1992) and use interior angles of the triangle formed by three points. Accordingly, Eq. (2) is revised as:

$$\begin{aligned} r_{ijk,j \in N(i),k \in N(i),j \neq k}^1 &= \text{angle}(i,j,k) \\ r_{ijk,j \in N(i),k \in N(i),j \neq k}^2 &= \text{angle}(i,k,j) \\ r_{ijk,j \in N(i),k \in N(i),j \neq k}^3 &= \text{angle}(k,i,j) \end{aligned} \quad (18)$$

Using the interior angle attributes to construct attributed graphs for the image and 11 sub-regions respectively, we carry out our proposed road network matching algorithm.

In this particular experiment we chose to highlight the performance of our approach, the road matching provides two perfect matches, as we see in Fig. 20, one being the mirror version of the other. This ideally illustrates a situation where existing network-based comparisons would be unable to differentiate between these two candidates. However, by proceeding with the comparison of building configurations (as presented in Section 3) we will be able to resolve this ambiguity. It should be pointed out that since we

Table 6
Similarity in spatial configurations of buildings.

Case A (selected block)				Case B (selected block)			
	U_4	U_5	U_7		U_4	U_5	U_7
Match to	V_{59}	V_{76}	V_{60}	Match to	V_{61}	V_{78}	V_{77}
$S = 0.91$				$S = 0.72$			

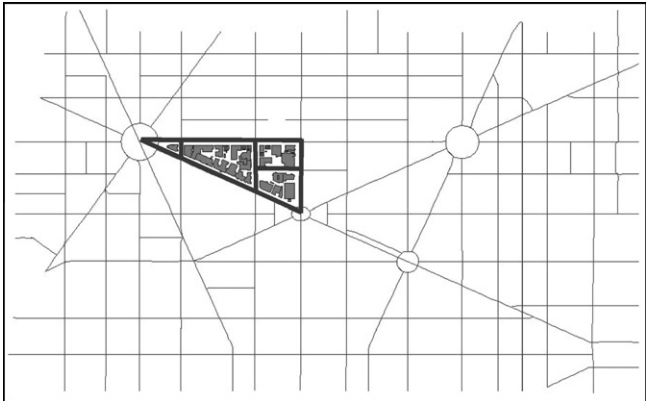


Fig. 21. Best match in the vector data to the queried image.

have already established correspondences between road intersections in the network matching step, we now have specific road loop correspondences, thus we can compare building configurations be-

tween specific loops, and not consider all potential configurations. Accordingly, by using the street block formed by U_4 , U_5 , U_7 and U_8 in the imagery and its matches in the vector data: V_{59} , V_{76} , V_{60} , V_{77} in A and V_{61} , V_{78} , V_{60} , V_{77} in B we can proceed to identify correctly that candidate A gives the best similarity configuration (a similarity score of 0.91 compared to 0.72 for B) and thus it is the best match to our queried imagery. Table 6 shows the similarity score for the selected block in the two cases. And Fig. 21 shows the best match in the search window to the image using both road network and building layers. With the MBR constructed for each building in selected blocks for case A and B , our approach took about 5 s in computing time to generate the similarity score for each case shown in Table 6.

5. Conclusions and future work

In this work, we presented a novel approach for the image-to-vector scene similarity assessment and matching based on the contextual information among salient features (road networks and buildings). Our approach proceeds by first comparing road networks in a scene in order to initially identify similarities at this coarse scene representation level, and then uses additional content in the form of object configurations within the network basic loops in order to solve ambiguities and find the best match. In this process we consider both local compatibility and global consistency metrics. We presented similarity assessment metrics for these comparisons that are invariant to scale and orientation variations, are able to accommodate small scene variations across time, and show robustness to inaccuracy commonly associated with feature extraction (especially from imagery). The road network comparison is based on a graph matching scheme that models networks' components (junctions and their connectivity) as graphs with embedded invariant attributes. It utilizes information about the topology and geometry of a network to assess similarity and establish the correspondences among its components (matching network junctions). The ability to utilize both topology and geometry allows us to reduce the ambiguity of local consistency, especially when inexact matching takes place as demonstrated in the first group of our experiments. Recognizing that road networks often look very similar we introduced in the second stage of our approach a view-consistent metric to compare object configurations within road loops. This introduction of additional geospatial information in our scene similarity assessment resembles the manner in which the human visual system moves from primal sketch to more complex visual perception. We used our view-consistent metrics in associated graphs and optimal similarity was selected as the maximal clique with the highest similarity score.

We considered a GIS environment for our work as presented in this paper, and information that may be available in various layers, namely road networks and buildings in our case. It is easily understood that our approach is also suitable for other comparable features (e.g. using water body networks, or networks of powerlines, and parking lots, or land use information). We are currently considering its extension of our approach to a multimodal environment, where some of the scene content may become available in the form of verbal descriptions of objects and their topological and geometric relations. Additional research into the properties of the association networks that could lead to more efficient searches may also improve the computational performance of our approach.

By using abstract representations of a scene, in the form of attributed relational graphs (ARGs), our approach performs reasonably well in situations where some information may be missing. We provided an example in Section 4.1, where two road networks were successfully matched even though one of them was missing

one or two segments. This may represent an example of comparing datasets captured at different time instances (i.e. before and after new road construction). It is easily understood though that if the magnitude of the missing information becomes overwhelming, the algorithm would not be able to overcome the challenge. We may also have other types of $n:m$ matching challenges, for example when the two datasets are represented at different levels of detail. For example, a roundabout with four road intersections may be represented as a loop in one dataset, whereas it may be reduced to a point in another. This $n:m$ matching challenge has been and remains an active research area for our community (Huang et al., 2010; Li and Goodchild, 2011; Sester et al., 1998). The use of ARGs offers some flexibility towards this challenge as well, as it avoids the limitations of using geometrically detailed descriptors of spatial objects such as street centerlines (see e.g. Chen et al., 2004; Li and Goodchild, 2010; Walter and Fritsch, 1999). Future work to derive hierarchical ARG scene descriptions could aid towards overcoming the substantial challenge of $n:m$ matching.

Regarding further future extensions of our work, we can identify some emerging opportunities as they relate to weight assignment in the overall scene similarity assessment. In the implementation we presented here, we assigned similar weights to the road network and building arrangement comparison metrics. While this is working well in the presented examples, eventually it would be interesting to consider a learning process which can be trained to automatically and robustly determine the weights based on feature properties and/or subject matter expertise. For example, certain properties may be considered semantically much more important than others, and therefore should be assigned higher weight when comparing two scenes that contain them. This would be a situation where analyst knowledge may be used when deciding on partial weight assignment in order to best address the challenges of the specific process that employs the scene similarity tool.

Acknowledgment

This work was supported by the National Geospatial-Intelligence Agency through a NURI grant.

References

- Babbar, G., Bajaj, P., Chawla, A., Gogna, M., 2010. Comparative study of image matching algorithms. *International Journal of Information Technology and Knowledge Management* 2 (2), 337–339.
- Barrow, H.G., Popplestone, R.J., 1971. Relational descriptions in picture process. In: Meltzer, B., Michie, D. (Eds.), *Machine Intelligence VI*. American Elsevier, New York, pp. 377–396.
- Boyer, K.L., Kak, A.C., 1988. Structural stereopsis for 3-D vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10 (2), 144–166.
- Brandes, U., Erlebach, T., 2005. *Network Analysis: Methodological Foundations*, first ed. Springer-Verlag, Berlin, Heidelberg.
- Brown, M., Lowe, D.G., 2007. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision* 74 (1), 59–73.
- Bruns, T., Egenhofer, M.J., 1996. Similarity of spatial scenes. In: Kraak, M.-J., Molenaar, M. (Eds.), *Proc. 7th International Symposium on Spatial Data Handling*, Delft, The Netherlands, pp. 4A.31–42.
- Bunke, H., 1999. Error correcting graph matching: on the influence of the underlying cost function. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (9), 917–922.
- Chen, C.-C., Knoblock, C.A., Shahabi, C., Chiang, Y.-Y., Thakkar, S., 2004. Automatically and accurately conflating ortho imagery and street maps. In: *Proc. 12th Annual ACM International Workshop on Geographic Information Systems*, Washington DC, USA, pp. 47–56.
- Clausius, R., 1879. *The Mechanical Theory of Heat*, second ed. Macmillan and Co., London.
- Doytsher, Y., Filin, S., Ezra, E., 2001. Transformation of datasets in a linear-based map conflation framework. *Surveying and Land Information Systems* 61 (3), 159–169.
- Drewniok, C., Rohr, K., 1997. Exterior orientation – an automatic approach based on fitting analytic landmark models. *ISPRS Journal of Photogrammetry and Remote Sensing* 52 (3), 132–145.

- Eubank, S., Kumar, V.S.A., Marathe, M.V., Srinivasan, A., Wang, N., 2004. Structural and algorithmic aspects of massive social networks. In: Proc. 15th annual ACM/ SIAM Symposium on Discrete Algorithms, pp. 718–727.
- Ferreira, A., Marini, S., Attene, M., Fonseca, M.J., Spagnuolo, M., Jorge, J.A., Falcidieno, B., 2010. Thesaurus-based 3D object retrieval with part-in-whole matching. *International Journal of Computer Vision* 89 (2–3), 327–347.
- Filin, S., Doytsher, Y., 2000. Detection of corresponding objects in linear-based map conflation. *Surveying and Land Information Systems* 60 (2), 117–128.
- Gautama, S., Borghgraef, A., 2003. Detecting change in road networks using continuous relaxation labeling. In: Proc. International Workshop on High Resolution Mapping from Space, Hannover, p. 6 (on CDROM).
- Godoy, F., Rodrigues, A., 2002. A quantitative description of spatial configurations. In: Richardson, D., Oosterom, P.V. (Eds.), Proc. 10th International on Spatial Data Handling, Ottawa, Canada, pp. 299–311.
- Henderson, J.M., 2003. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences* 7 (11), 498–504.
- Hild, H., Haala, N., Fritsch, D., 2000. A strategy for automatic image to map registration. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science* 33 (Part B2), 313–321.
- Hinz, S., Baumgartner, A., 2003. Algorithms and techniques for multi-source data fusion in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (1–2), 83–98.
- Holm, M., Eija, P., Kaj, A., Arto, V., 1995. Nationwide automatic satellite image registration system. In: McKeown, D.M., Dowman, I.J. (Eds.), Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II. SPIE, pp. 156–167.
- Huang, L., Wang, S., Ye, Y., Wang, B., Wu, L., 2010. Feature matching in cadastral map integration with a case study of Beijing. In: Proc. 18th International Conference on Geoinformatics, Beijing, China, pp. 1–4.
- Hummel, R.A., Zucker, S.W., 1983. On the foundations of relaxation labeling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5 (3), 267–287.
- Jain, B.J., Wysozki, F., 2002. Fast winner-takes-all networks for the maximum clique problem. In: Proc. 25th Annual German Conference on AI: Advances in Artificial Intelligence, Aachen, pp. 163–173.
- Jizba, P., Arimitsu, T., 2001. The world according to Renyi: thermodynamics of fractal systems. In: Proc. International Symposium on Non-equilibrium and Nonlinear Dynamics in Nuclear and Other Finite Systems, Beijing, pp. 341–348.
- Kaminsky, R., Snavey, N., Seitz, S.M., Szeliski, R., 2009. Alignment of 3D point clouds to overhead images. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Miami, pp. 63–70.
- Li, L., Goodchild, M.F., 2010. Automatically and accurately matching objects in geospatial datasets. In: Proc. Joint International Conference on Theory, Data Handling and Modelling in Geospatial Information Science, Hong Kong, pp. 26–28.
- Li, L., Goodchild, M.F., 2011. An optimization model for linear feature matching in geographical data conflation. *International Journal of Image and Data Fusion* 2 (4), 309–328.
- Li, S.Z., 1992. Matching: invariant to translations, rotations and scale changes. *Pattern Recognition* 25 (6), 583–594.
- Li, S.Z., 1994. A Markov random field model for object matching under contextual constraints. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, pp. 866–869.
- Liu, Y., Dai, Q., Xu, W., 2010. A point-cloud-based multiview stereo algorithm for free-viewpoint video. *IEEE Transactions on Visualization and Computer Graphics* 16 (3), 407–418.
- Luo, B., Hancock, E.R., 2001. Structural graph matching using the EM algorithm and singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (10), 1120–1136.
- Mannan, S.K., Ruddock, K.H., Wooding, D.S., 1997. Fixation patterns made during brief examination of two-dimensional images. *Perception* 26 (8), 1059–1072.
- Marr, D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, New York.
- Mohammed, J., Hummel, R., Zucker, S., 1983. A feasible direction operator for relaxation methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5 (3), 330–332.
- Mustière, S., Devogele, T., 2008. Matching networks with different levels of detail. *Geoinformatica* 12 (4), 1384–16175.
- Nedas, K., Egenhofer, M.J., 2008. Spatial-scene similarity queries. *Transactions in GIS* 12 (6), 661–681.
- Pathak, K., Vaskevicius, N., Poppinga, J., Pfingsthorn, M., Schwertfeger, S., Birk, A., 2009. Fast 3D mapping by matching planes extracted from range sensor point-clouds. In: Proc. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, pp. 1150–1155.
- Poullis, C., You, S., 2010. Delineation and geometric modeling of road networks. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (2), 165–181.
- Rosenfeld, A., Hummel, R., Zucker, S., 1976. Scene labelling by relaxation operations. *IEEE Transactions on Systems, Man and Cybernetics* 6 (6), 420–433.
- Schank, T., Wagner, D., 2005. Finding, counting and listing all triangles in large graphs, an experimental study. In: Nikolettseas, S.E. (Ed.), Proc. 4th International Workshop on Experimental and Efficient Algorithms, Santorini, Greece, pp. 606–609.
- Schickler, W., 1992. Feature matching for outer orientation of single images using 3D wireframe controlpoints. *International Archives of Photogrammetry and Remote Sensing* 30 (B3/III), 591–598.
- Sester, M., Anders, K.-H., Walter, V., 1998. Linking objects of different spatial data sets by integration and aggregation. *Geoinformatica* 2 (4), 335–358.
- Shannon, C., 1948. A mathematical theory of communication. *Bell System Technical Journal* 27 (7&10), 379–423, 623–656.
- Shi, J., Malik, J., 1998. Self inducing relational distance and its application to image segmentation. In: Burkhardt, H., Neumann, B. (Eds.), Proc. 5th European Conference on Computer Vision, Freiburg, Germany, pp. 528–543.
- Stefanidis, A., Agouris, P., Georgiadis, C., Bertolotto, M., Carswell, J.D., 2002. Scale- and orientation-invariant scene similarity metrics for image queries. *International Journal of Geographical Information Science* 16 (8), 749–772.
- Suveg, I., Vosselman, G., 2004. Reconstruction of 3D building models from aerial images and maps. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (1), 202–224.
- Tipdecho, T., 2002. Automatic image registration between image and object spaces. In: Proc. Open Source GIS – GRASS Users Conference, Trento, Italy, p. 10 (on CDROM).
- Tomita, E., Tanaka, A., Takahashi, H., 2006. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science* 363 (10), 28–42.
- Tseng, Y.-H., Wang, S., 2003. Semiautomated building extraction based on CSG model-image fitting. *Photogrammetric Engineering and Remote Sensing* 69 (2), 171–180.
- Volz, S., 2006. An iterative approach for matching multiple representations of street data. In: Hampe, M., Sester, M., Harrie, L. (Eds.), Proc. ISPRS Workshop on Multiple Representation and Interoperability of Spatial Data, Hannover, vol. XXXVI-2-W40, pp. 101–110.
- Vosselman, G., 1992. *Relational Matching*. Springer-Verlag, Berlin, Heidelberg.
- Wallis, W.D., 2007. *A Beginner's Guide to Graph Theory*, second ed. Birkhäuser, Boston.
- Walter, V., Fritsch, D., 1999. Matching spatial data sets: a statistical approach. *International Journal of Geographical Information Science* 13 (5), 445–473.
- Wang, C., Stefanidis, A., Croitoru, A., Agouris, P., 2008. Map registration of image sequences using linear features. *Photogrammetric Engineering & Remote Sensing* 74 (1), 25–38.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature* 393 (6), 440–442.
- Wilson, R.C., 1996. *Inexact graph matching using symbolic constraints*. Ph.D. Dissertation Thesis, The University of York, York, UK.
- Wilson, R.C., Hancock, E.R., 1997. Structural matching by discrete relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (6), 634–648.
- Wu, X., Carceroni, R., Fang, H., Zelinka, S., Kirmse, A., 2007. Automatic alignment of large-scale aerial rasters to road-maps. In: Proc. 15th International Symposium on Advances in Geographic Information Systems, November 7–9 Seattle, WA, p. 8 (on CDROM).
- Yu, Z., Prinnet, V., Pan, C., 2004. A novel two-steps strategy for automatic GIS-image registration. In: Proc. 11th International Conference of Image Processing, pp. 1711–1714.
- Zampelli, S., Deville, Y., Solnon, C., 2010. Solving subgraph isomorphism problems with constraint programming. *Constraints* 15 (3), 327–353.
- Zhao, H., Chan, K.L., Cheng, L.M., Yah, H., 2009. A probabilistic relaxation labeling framework for reducing the noise effect in geometric biclustering of gene expression data. *Pattern Recognition* 42 (11), 2578–2588.
- Zhao, Z., Stough, R.R., Song, D., 2011. Measuring congruence of spatial objects. *International Journal of Geographical Information Science* 25 (1), 113–130.
- Zhu, C.-F., Li, S.-X., Chang, H.-X., Zhang, J.-X., 2009. Matching road networks extracted from aerial images to GIS data. In: Proc. 2009 Asia-Pacific Conference on Information Processing, July 18–19 Shenzhen, China, pp. 63–66.