نشریه علمی- پژوهشی علوم و فنون نقشه برداری، دوره سوم، شماره ۱، مرداد ماه ۱۳۹۲

تناظریابی معابر برون‌شهری با استفاده از خط کمترین مربعات در پایگاه داده مکانی چندمقیاسی

علی فراهانی‌پویا[۱]، رحیم علی‌عباسپور*[۲]، سیدکاظم علوی‌پناه[۳]، منصور جعفربیگلو[۴]

[۱]دانشجوی کارشناسی ارشد سنجش از دور و سیستم‌های اطلاعات مکانی - گروه کارتوگرافی - دانشکده جغرافیا - دانشگاه تهران
farahanipooya@ut.ac.ir

[۲]استادیار گروه مهندسی نقشه‌برداری - پردیس دانشکده‌های فنی - دانشگاه تهران
abaspour@ut.ac.ir

[۳]استاد گروه کارتوگرافی - دانشکده جغرافیا - دانشگاه تهران
salavipa@ut.ac.ir

[۴]استادیار گروه کارتوگرافی - دانشکده جغرافیا - دانشگاه تهران
mgbeglou@ut.ac.ir

(تاریخ دریافت آذر ۱۳۹۱، تاریخ تصویب بهمن ۱۳۹۱)

چکیده

با پیشرفت سریع تولید و جمع‌آوری داده‌های مکانی و تکنیک‌های پردازش آن‌ها، همواره حجم عظیمی از داده‌ها که ناحیه جغرافیایی یکسانی را می‌پوشاند، در دسترس قرار می‌گیرد. این حجم داده، متخصصین حوزه علوم ژئوماتیک را به ایجاد پایگاه داده چندمقیاسی با اهدافی مانند به روز رسانی مجموعه داده‌ها و آنالیزهای چندمقیاسی ترغیب کرده است. هسته اصلی تولید پایگاه‌های چندمقیاسی، فرآیند تناظریابی داده‌ها می‌باشد. در این مقاله الگوریتم جدیدی برای تناظریابی داده با هدف تولید پایگاه‌های داده چندمقیاسی ارائه شده است. الگوریتم پیشنهادی در سه فاز اصلی پیاده‌سازی شده است. در فاز اول پیش‌پردازش انجام شده و در فاز دوم به جستجوی و شناسایی عوارض کاندیدا در بین دو مجموعه داده برمبنای ویژگی‌های توپولوژیک پرداخته می‌شود. در فاز سوم، با استفاده از روش کمترین مربعات، خطی به عنوان بهترین نماینده برای هر عارضه راه محاسبه شده و با مقایسه ویژگی‌های هندسی نماینده عارضه راه، بین عوارض کاندیدای مرحله قبل، بهترین جفت متناظر انتخاب می‌شود. این الگوریتم بر روی معابر برون‌شهری نقشه‌های در مقیاس ۱:۲۵۰۰۰ و۱:۵۰۰۰۰ تولیدی سازمان نقشه‌برداری کشور پیاده سازی شده است. نتایج ارزیابی موفق الگوریتم را نشان می‌دهد، بطوری که ۹۴/۲ درصد عوارض خطی راه به درستی به عنوان جفت عارضه متناظر شناخته شدند.

واژگان کلیدی: تناظریابی داده، عوارض متناظر، ویژگی‌های توپولوژیک، کمترین مربعات، پایگاه داده مکانی چندمقیاسی.

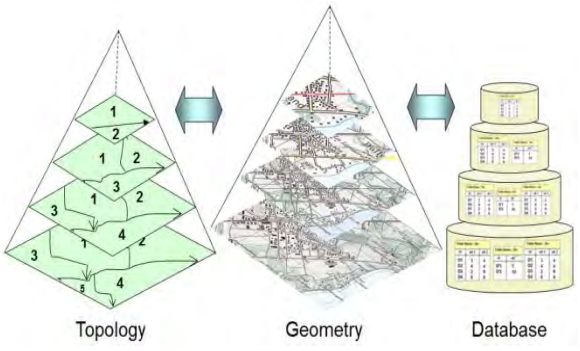
* نویسنده رابط

۱- مقدمه

با پیشرفت سریع روش‌های گردآوری داده‌های مکانی و تکنیک‌های پردازش آن‌ها، حجم عظیمی از داده‌ها از سازمان‌های دولتی و خصوصی به سهولت در دسترس کاربران قرار می‌گیرند. این داده‌ها ممکن است محدوده جغرافیایی یکسانی را پوشش دهند و در عین حال درمقیاس، موضوع و محتوا، دقت، صحت و مدل داده متفاوت باشند. یک مجموعه داده می‌تواند از برخی جنبه-ها نظیر سطح کیفی یا محتوای اطلاعات نسبت به دیگر مجموعه داده‌ها متمایز بوده و برتری داشته باشد؛ اماقطعاً این برتری در همه جنبه‌ها امکان‌پذیر نیست. بنابراین، به منظور بهره‌مندی از حداکثر مزایای ممکن، بهتر است مجموعه داده‌های متفاوت با هم مورد استفاده قرار گیرند تا کارآمدتر بوده و کاربردهای گوناگونی را در برگیرند [۱].

مقیاس پارامتر تعیین‌کننده‌ای در تهیه و تولید مجموعه داده‌ی جغرافیایی است [۵، ۶، ۷]. به تبع مقیاس، دقت و سطح جزئیات[۱] متفاوت است. یکپارچه سازی مجموعه داده‌ها در مقیاس‌های متفاوت منجر به تولید پایگاه داده-های چندمقیاسی[۲] می‌گردد (شکل۱). یک پایگاه داده چندمقیاسی به عنوان پایگاه داده‌ای مطرح می‌شود که عوارض یکسانی از جهان واقعی را در سطوح مختلفی از مقیاس و دقت و صحت نگهداری می‌کند [۲، ۳]. یک پایگاه داده چندمقیاسی اجازه آنالیزهای چند مقیاسی را به ما می‌دهد [۲، ۳، ۴، ۲۱]؛ بدین معنی که اطلاعات در یک مقیاس می‌توانند با کمک اطلاعات مقیاس(های) دیگر تجزیه و تحلیل شوند. همچنین، پایگاه داده چندمقیاسی می‌تواند فرآیندهای به‌روز رسانی اتوماتیک را پشتیبانی کند [۱، ۲، ۳، ۵]. در این حالت، عوارض در مجموعه داده‌های مختلف به یکدیگر متصل شده و زمانی که عوارض در یکی از مجموعه داده‌ها به روز رسانی می‌شود، این به روز رسانی در مجموعه داده دیگر اعمال می‌شود. از جمله مزایای دیگر پایگاه داده چندمقیاسی حل مسأله جنرالیزاسیون آنی برای سرویس‌های تحت وب است [۲، ۵].

[۱] Level of Details
[۲] Multi Resolution/Representation Databases(MRDB)



شکل ۱- پایگاه داده مکانی چندمقیاسی

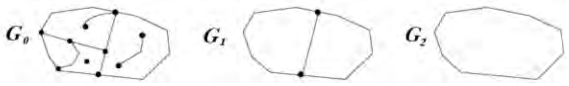
فرض کنیم اطلاعات مکانی با بالاترین سطح جزئیات (یا دارای بزرگترین مقیاس) باشد و مجموعه توابع تبدیل f_i ، توابعی باشند که اطلاعات مکانی را از سطح $i-1$ به سطح i منتقل نمایند. به عبارت دیگر در ساختار پایگاه داده مکانی چندمقیاسی با $k+1$ سطح داریم:

$$f_1 : G_0 \mapsto G_1 \quad ..... \quad f_k : G_{k-1} \mapsto G_k$$

بدین ترتیب می‌توان بطور مستقیم، با ترکیب توابع تبدیل، از G_0 به هر سطح دلخواه رسید:

$$F_i = f_i \circ f_{i-1} \circ ... \circ f_2 \circ f_1 \quad : G_0 \mapsto G_i$$

توالی اطلاعات مکانی $G_i \quad (i = 0,...,k)$ را توالی چندمقیاسی و تابع F_i را تابع تبدیل از ریشه می‌نامیم (شکل ۲).



شکل ۲- توالی لایه‌ها در ساختار چندمقیاسی

هر پایگاه داده مکانی چندمقیاسی دارای دو ویژگی اساسی می‌باشد:

- سطوح مختلف جزییات که در یک پایگاه داده ذخیره شده‌اند.
- عوارض که در سطوح مختلف به یکدیگر متصل شده‌اند.

اتصال عوارض در سطوح مختلف جزئیات یکی از مسائل اصلی در ایجاد پایگاه داده چندمقیاسی است که از آن به عنوان تناظریابی داده یاد می‌شود [۱، ۲، ۳، ۵، ۷]. در

واقع تناظریابی را می‌توان فرآیند جستجوی عوارض متناظر از میان مجموعه داده‌های مختلف که دارای ارتباط مکانی و مشابهت معنایی قوی هستند، دانست[۱]. به این ترتیب تناظریابی علاوه بر اینکه در ساختن پایگاه داده چندمقیاسی نقشی اساسی ایفا می‌کند، کاربردهای دیگری نیز دارد که شرح آن در ادامه می‌آید:

i. تناظریابی داده امکان انتقال اطلاعات غیرمکانی متصل به اطلاعات مکانی از یک مجموعه داده به مجموعه داده دیگر با سطح جزئیات مختلف را فراهم می‌سازد [۴].
ii. تناظریابی داده می‌تواند به ارزیابی و بهبود کیفیت داده‌ها کمک کند. به عبارت دیگر به روز بودن نقشه و صحت و دقت یک مجموعه داده در برابر مجموعه داده‌های دیگر کنترل می‌شود [۱۸].
iii. تناظریابی داده در فرآیند تثبیت تصویر و زمین‌مرجع نمودن آن نیز کاربرد دارد. تصویرهای خام بدست آمده از سنجنده‌های ماهواره‌ای معمولاً دارای خطاهای گوناگونی می‌باشند. یک تصویر واپیچیده می‌تواند با نقشه‌های موجود تثبیت شود، به عبارت دیگر به طور اتوماتیک تصحیح شود [۱۹].

در مدل داده میدان‌مبنا، داده‌ها به سه گروه اصلی نقطه‌ای و خطی و سطحی تقسیم‌بندی می‌شوند. در این میان، داده‌های خطی برای نمایش طیف گسترده‌ای از عوارض طبیعی همچون رودخانه و آبراهه و عوارض ساخت بشر مانند راه و خطوط انتقال نیرو بکار گرفته می‌شود. از بین عوارض مذکور، راه‌ها و معابر از اهمیت خاصی برخوردارند، به طوری که همواره سازمان‌های متولی تولید نقشه مبادرت به تولید انواع نقشه با موضوع راه کرده و تحقیقات متعددی نیز منحصراً بر روی مدلسازی و تحلیل شبکه گذرگاهی انجام می‌گیرد [۶، ۷، ۸، ۱۲، ۱۳].

در این مقاله سعی شده است روش جدیدی برای تناظریابی عارضه راه در پایگاه داده مکانی چندمقیاسی ارائه شود. برای این منظور بعد از مرور تحقیقات پیشین (بخش ۲)، به معرفی روش پیشنهادی جهت تناظریابی عوارض خطی راه در پایگاه داده چندمقیاسی پرداخته می‌شود (بخش۳). در ادامه، روند پیاده‌سازی تشریح شده (بخش ۴) و در پایان، نتایج حاصل از پژوهش بیان می‌گردد (بخش ۵).

۲- پیشینه تحقیق

تحقیقات فراوانی در حوزه تناظریابی داده‌های برداری انجام شده است. این تحقیقات را می‌توان در سه گروه اصلی تقسیم‌بندی کرد. موضوع مورد بحث در گروه اول، ساختن پایگاه داده چندمقیاسی است که تناظریابی داده از بخش‌های اصلی آن محسوب می‌گردد. در گروه دوم، تناظریابی داده در مجموعه داده‌های با مقیاس‌های مختلف مورد بررسی قرار گرفته است و گروه سوم شامل مقالاتی است که صرفاً به تناظریابی داده برای عوارض خطی راه پرداخته‌اند.

سکونی [۵] به ایجاد پایگاه داده چندمقیاسی تحت وب پرداخته است. وی بر روی دو مسأله بهبود کیفیت داده و تناظریابی داده تمرکز نموده است. در مرحله بهبود کیفیت داده از اطلاعات کمکی استفاده می‌کند تا فرآیند جنرالیزاسیون در مرحله بعد ساده‌تر شود. در مرحله تناظریابی، فرآیند جنرالیزاسیون برای چندین کلاس عارضه مطرح می‌شود. از دو تکنیک ساده‌سازی شبکه‌ای[۱] و تبدیل مورفینگ[۲] برای تولید نقشه جدید از سطوح مختلف جزئیات استفاده می‌کند. وی تبدیل مورفینگ را برای ارتباط بین پایگاه داده چندمقیاسی تحت وب استفاده کرده است.

همپ و همکاران [۲] پایگاه داده چندمقیاسی را بر مبنای سیستم جنرالیراسیون قاعده‌مبنا ایجاد کردند. در ایجاد مدل پایگاه چندمقیاسی از نرم‌افزار ArcGIS و پایگاه داده Oracle استفاده کرده‌اند. دستورات قاعده‌مبنا در تناظریابی برای عارضه‌های ناحیه‌ای و تبدیل عوارض ناحیه‌ای به عوارض خطی و نقطه‌ای بکار رفته است.

اندرس[۳] در کاری مشابه با کار همپ یک پایگاه داده چندمقیاسی برای داده‌های ATKIS ایجاد نموده است. هدف اصلی وی، فراهم کردن بستری برای فرآیند به روز رسانی اتوماتیک بوده است.

مقالات مذکور با وجود بیان و تاکید فراوانی که برمسأله تناظریابی داده در پایگاه داده چندمقیاسی داشته‌اند، مسأله تناظریابی را بطور کلان مطرح نموده‌اند.

---

[۱] Mesh Simplification
[۲] Morphing Transformation

دیواجل و مستییر [۶] بر روی یکپارچه‌سازی شبکه خطی کار کرده‌اند که دارای سطوح مختلفی از جزئیات بوده است. در این کار، فاصله هاسدروف بین چندخطی‌ها و ویژگی‌های توپولوژیک شبکه که کمتر متأثر از مقیاس است، بکار گرفته شده است. در این پژوهش، تنها از ویژگی هندسی فاصله استفاده شده و ویژگی‌های توپولوژیک نیز به طور کامل مستقل از مقیاس در نظر گرفته شده است، بطوری که این فرض همه جا پاسخگوی حل مسأله نیست.

ژانگ وهمکاران [۷] الگوریتمی برای تناظریابی شبکه خطی درمقیاس‌های مختلف ارائه داده‌اند. ایشان الگوریتم خود را بر مبنای فرآیندی تکراری و خودفراگیر با افزایش حریم به صورت انطباقی ارائه داده و از ویژگی‌های هندسی و توپولوژیک برای تناظریابی استفاده نموده‌اند. در این پژوهش، اگرچه برای انتخاب بهترین کاندیدا چند مرحله افزایش حریم اتفاق می‌افتد و از مزایای این پژوهش محسوب می‌شود، ولی هیچ‌گونه راهکاری برای تعیین حدود آستانه در استفاده از ویژگی‌های هندسی وتوپولویک ارائه نگردیده و فقط به بیان حدود آستانه اکتفا شده است.

لوشار و همکاران [۸] نیز الگوریتمی برای تناظریابی عوارض راه در مقیاس‌های مختلف و با اختلاف زیاد جهت تولید پایگاه داده چندمقیاسی ارائه داده‌اند. در این روش، ضمن تعیین حریم در اطراف گره‌ها، متناظر با آنها کاندیداهایی انتخاب و این کاندیداها با ویژگی‌های معنایی، هندسی و توپولوژیک فیلتر شده تا به بهترین جفت متناظر برسند. سپس یال‌های متناظر با استفاده از الگوریتم کوتاه‌ترین مسیر بین دو گره انتخاب شده‌اند. از ویژگی‌های این کار می‌توان به اختلاف زیاد بین مقیاس-های دو مجموعه داده اشاره کرد.

گودچایلد و لی [۱۰] تناظریابی شبکه خطی راه را در قالب مسأله بهینه‌سازی مطرح کردند. آنها تابع هدف را بر مبنای میزان شباهت بین عوارض دو مجموعه داده کمینه می‌کنند. طرح مسأله تناظریابی در قالب بهینه‌سازی گرچه روشی بدیع و نوآورانه است، ولی حجم محاسباتی بالا و سرعت پایین، کارایی روش را کاهش می‌دهد.

سافارا و همکاران [۱۱] روشی تک‌منظوره برای تناظریابی عوارض خطی راه بکار برده‌اند. در این روش، عارضه راه به صورت چندخطی تعریف می‌شود و موقعیت نقاط آغازی و پایانی برای تناظریابی عوارض بکار می‌رود. در این پژوهش به جای استفاده از شاخص‌های مکانی از شاخص شبکه استفاده می‌شود.

یانگ و همکاران [۱۲] با رویکرد احتمالاتی به حل مسأله تناظریابی شبکه راه پرداخته‌اند. این روش از سازگارسازی ماتریس احتمالی اولیه (که عناصر آن بیانگر میزان شباهت شکل عوارض است) در یک فرایند تکراری برای انتخاب مناسبترین جفت عارضه متناظر استفاده می-کند. به عبارت دیگر ایجاد ماتریس سازگار بیانگر یافتن جفت عارضه‌ی متناظر است.

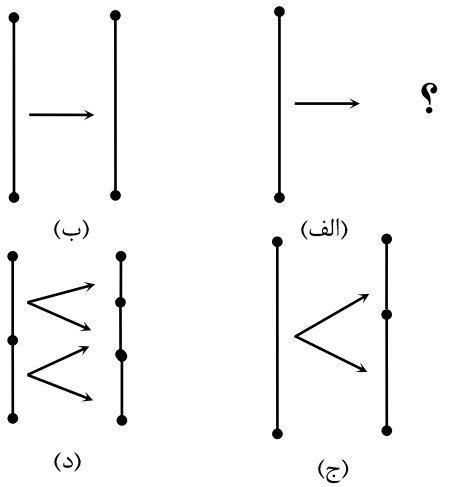
ژانگ [۱] برای تناظریابی شبکه راه، الگوریتمی به نام DSO را پیشنهاد داده است. وی الگوریتم خود را در پنج مرحله معرفی کرده است: ۱.پیش‌پردازش شامل کاهش نویزها و حذف ابهامات توپولوژیک، ۲. تعریف شبکه راه با گراف‌ها، ۳. اتصال DS، ۴. انطباق DS، ۵. انطباق عوارض در جاهایی که در چهار مرحله قبل انطباق نیافته‌اند. اگرچه این الگوریتم دارای نوآوری در خواص توپولوژیک و هندسی می‌باشد، اما پیچیده و زمانبر می‌باشد.

الگوریتم‌های ارائه شده در مقالات مذکور، هرچند راهکارها و راه‌حل‌های متنوع و گوناگونی را برای حل مسأله تناظریابی ارائه داده‌اند، اما هر یک دارای کاستی-هایی نظیر کلی‌نگری به مسأله، بخشی‌نگری به مسأله و عدم قابلیت تعمیم الگوریتم، عدم توجه به ظرفیت‌ها و قابلیت‌های شبکه (اعم از ویژگی‌های توپولوژیک و ویژگیهای هندسی عارضه راه)، عدم توجه کافی به تعیین حدود آستانه و تحلیل‌های مناسب آن برای حل مسأله و پیچیدگی الگوریتم‌ها و حجم بالای محاسباتی (و در نتیجه زمان‌بر بودن حل مسأله) می‌باشند. در این مقاله سعی شده است تا در عین ارائه راهکاری جدید، تا حد ممکن مشکلات الگوریتم‌های پیشین رفع و تکمیل گردد.

۳- روش پیشنهادی

در تناظریابی مجموعه داده‌هایی با مقیاس‌های مختلف، عوارض موجود در یک مجموعه داده ممکن است راهی را در جهان واقعی نمایش دهند که در مجموعه داده دیگر نمایش داده نشود. از طرف دیگر ممکن است یک عارضه در یک مجموعه داده با گروهی از عوارض در مجموعه داده دوم متناظر باشد. از این رو باید دسته‌بندی صحیحی از انواع ارتباطات بین عوارض متناظر وجود داشته باشد. در

پیشینه تناظریابی راه‌ها [۱۲، ۱۳، ۱۴] ارتباطات بین عوارض متناظر از مجموعه داده‌های مختلف در چهار گروه رابطه یک به هیچ، یک به یک، یک به چند و چند به چند دسته‌بندی می‌شوند (شکل ۳).



شکل ۳- انواع ارتباطات بین عوارض متناظر: (الف)یک به هیچ ،(ب) یک به یک، (ج)یک به چند، (د) چند به چند

به منظور تناظریابی شبکه راه‌ها در دو مقیاس مختلف در پایگاه داده مکانی چندمقیاسی، در این مقاله الگوریتمی در ۳ فاز به شرح ذیل طراحی و پیاده‌سازی گردید:

i. پیش پردازش: در این مرحله، مجموعه داده‌ها برای ورود به مرحله تناظریابی آماده‌سازی شده و شرایط مورد نیاز را تأمین می‌کنند.
ii. الگوریتم جستجو: در این مرحله، بر مبنای ویژگی‌های توپولوژیک شبکه، کاندیداهای متناظر انتخاب می‌شوند.
iii. تعیین درجه شباهت: خطی به کمک روش کمترین مربعات به کاندیداهای انتخاب‌شده در بخش قبل برازش داده می‌شود؛ بطوری‌که چندخطی عارضه راه را نمایندگی می‌کند و ویژگی‌های هندسی برای خط نماینده محاسبه می‌شود. با مقایسه این ویژگی‌ها، میزان شباهت بین دو عارضه تعیین شده و چنانچه از حد آستانه که برای درجه شباهت تعریف شده بیشتر باشد، دو عارضه به عنوان عوارض متناظر به یکدیگر متصل می‌شوند.

در ادامه، مجموعه فعالیت‌های انجام گرفته در هر یک از این سه فاز به اجمال توضیح داده می‌شود.

۳-۱- پیش‌پردازش

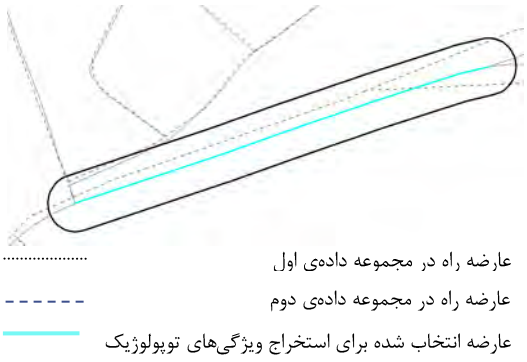
هدف اصلی در تناظریابی مجموعه داده‌های مکانی شناسایی زوج عوارض متناظر در دو لایه است. در این مقاله، شبکه راه‌ها به شکل داده برداری، بصورت گراف $G=(V,E)$ مدلسازی می‌شود که مجموعه V برای گره‌ها و مجموعه E، شامل زیرمجموعه‌های دو عضوی از مجموعه V، برای نمایش یال‌ها استفاده می‌شود. گره‌ها (که گره‌های توپولوژیک نیز نامیده می‌شوند) شامل تقاطع‌ها (مکان‌هایی که دو یا چند راه به یکدیگر می‌رسند) و نقاط ابتدا و انتهای راه‌ها می‌باشند و یال‌ها عارضه راه را نمایندگی می‌کنند. با این نحوه نمایش، یک راه می‌تواند از تقاطع شروع شده و یا در آن پایان یابد، اما هرگز شامل تقاطع نمی‌شود.

برای نمایش هندسی یال‌ها چند حالت مختلف وجود دارد که در حالت کلی یک عارضه راه به صورت یک چندخطی نمایش داده می‌شود. یک *چندخطی* در $\mathbb{R}^2$ به صورت مجموعه محدودی از پاره خطها به نحوی تعریف می‌شود که بجز احتمالاً برای دو نقطه که *نقاط انتهایی* چندخطی نامیده می‌شوند، نقطه انتهای هر یال دقیقاً بوسیله دو یال به اشتراک گذاشته شود. علاوه بر این، اگر دو یال در هیچ جایی غیر از احتمالاً نقاط انتهایی همدیگر را قطع نکنند، آن چندخطی را چندخطی *ساده* می‌نامیم [۲۰]. ظرفیت یک گره مانند N_i تعداد چندخطی‌هایی است که این گره، نقطه انتهایی آنها می‌باشد.

به منظور ایجاد شبکه راه‌ها بر اساس ساختار مورد اشاره، مجموعه داده‌های راه‌ها در نرم‌افزار ArcGIS ویرایش شده و پردازش‌های لازم روی آنها انجام گرفت. در طول این فاز تمام خطوطی که با یکدیگر تلاقی داشتند، در محل تقاطع به دو عارضه تبدیل شده و خطاهای ترسیمی همچون از روی هم رد شدن عوارض خطی، نرسیدن عارضه‌های خطی به یکدیگر و عدم همپوشانی لایه خطی رفع شدند. برای این منظور با ایجاد لایه توپولوژی برای هرکدام از مجموعه داده‌ها و تعریف قواعد توپولوژیک برای لایه‌ها، خطاها شناسایی شده و بر حسب مورد ویرایش شده‌اند.

۳-۲- الگوریتم جستجو با ویژگی‌های توپولوژیک

الگوریتم جستجو با انتخاب عارضه از مجموعه داده با مقیاس کوچکتر شروع شده و برای آن در مجموعه داده دوم عارضه‌ای متناظر یافته می‌شود. با انتخاب هر عارضه چندخطی از مجموعه داده اول نقاط آغازی و پایانی آن جدا شده و ویژگی‌های توپولوژیک این نقاط استخراج می‌گردد. برای یافتن عارضه متناظر در مجموعه داده دوم نیاز است تا حریمی پیرامون عارضه مجموعه داده اول تعریف شود. با ایجاد این حریم تمامی عوارضی که درون محدوده حریم قرار دارند، انتخاب می‌شوند. عوارض انتخابی درون مجموعه‌ای به نام مجموعه *عوارض کاندیدا* قرار می‌گیرند. پس از استخراج نقاط آغازی و پایانی تمام عوارض کاندیدا، ویژگی‌های توپولوژیک آنها استخراج می‌شود (شکل۴).



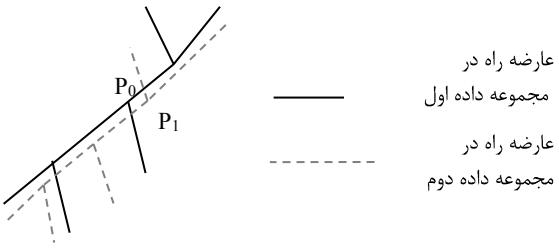
شکل ۴- فرآیند انتخاب عوارض کاندیدا

زمانی‌که ظرفیت گره برابر با یک است، آن گره قطعاً نقطه پایانی یا آغازی است؛ چرا که با دیگر قطعه‌ها ارتباط ندارد. گره‌ای با ظرفیت ۲ یک نقطه اتصال کاذب[۱] است و یک تقاطع واقعی نیست؛ به دلیل اینکه دو قسمت از یک راه را به یکدیگر متصل نموده است و نه دو راه مختلف را. این موضوع علل مختلفی دارد که یکی از آن‌ها خطای ترسیم عامل کارتوگرافی می‌باشد که این موضوع نیز در مرحله‌ی پیش پردازش به کمک ابزار Topology برطرف می‌شود. در گره‌هایی با ظرفیت بیش از دو، تعداد گره‌ها به عنوان خاصیت توپولوژیک در فرآیند جستجوی عوارض در دو مجموعه داده استفاده می‌شوند. در شکل ۵ انواع گره‌ها به همراه ظرفیت آنها نشان داده شده است.

ظرفیت	سه	چهار	پنج	شش	نوع گره

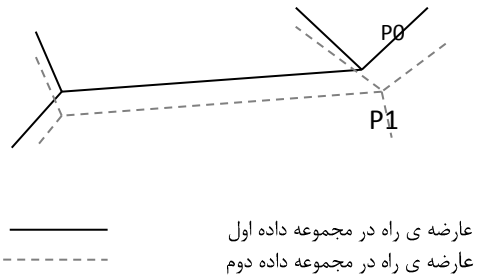
شکل۵- نمایش گره‌ها به همراه ظرفیت آنها

با این وجود، این پارامتر به تنهایی برای توصیف توپولوژیک تقاطع‌ها کافی نیست. مطابق با شکل ۶ گره‌های P_0 و P_1 ظرفیت‌های یکسانی دارند، اما در ظاهر با یکدیگر شباهتی ندارند.



شکل ۶-گره‌هایی با ظرفیت یکسان وغیرمتشابه[۱]

از این رو، زاویه بین یال‌های متصل به گره را برای تمایز بیشتر گره‌های متشابه در نظر می‌گیرند. اما در تناظریابی عوارض راه بین مجموعه داده‌هایی با مقیاس‌های متفاوت این پارامتر نیز کارآیی لازم را ندارد. شکل ۷ نمونه‌ای از این ناکارآمدی را نشان می‌دهد.



شکل ۷-گره‌هایی متشابه با ظرفیت‌های متفاوت

همانطور که در شکل۷ دیده می‌شود، گره‌های P_0 و P_1 ظرفیت یکسانی ندارند و همچنین زاویه بین خطوط آنها نیز متشابه نیست، اما در واقع یک تقاطع را در دنیای واقعی نشان می‌دهند. از این رو، به جای استفاده از زاویه

[۱] pseudo junction

بین آنها از آزیموت زاویه هر عارضه راه که در گره مورد نظر مشترک می‌باشد، استفاده می شود. این پارامتر از این جهت کمک کننده می‌باشد که به عنوان مثال در مجموعه داده با مقیاس کوچکتر، گرهی با ظرفیت ۳ و آزیموت‌های به ترتیب ۵۰ و ۱۰۰ و ۱۵۰ درجه وجود دارد و در مجموعه داده با مقیاس بزرگتر، گرهی با ظرفیت ۴ و آزیموت‌هایی به ترتیب برابر با ۱۰ و ۵۰ و ۱۰۰ و ۱۵۰ درجه موجود است. در این حالت، این دو گره متشابه در نظر گرفته شده و در واقع شرط لازم برای ورود به مرحله بعد حاصل می‌شود.

از سوی دیگر، یک عارضه راه که با چندخطی معرفی شده است، فقط شامل قطعاتی از خط نیست، بلکه همانطور که در تعریف عارضه راه نیز گفته شد، ممکن است چندخطی منحنی را در برگیرد و چنانچه این منحنی شامل نقطه‌ی پایانی و آغازی باشد، تعیین آزیموت آن به روش قبلی کمی مشکل است و لذا در این حالت از بردار یکه مماس بر منحنی استفاده می‌شود.

در جمع بندی تشریح خواص توپولوژیکی یک گره باید مراحل ذیل را در نظر گرفت:

i. تعیین ظرفیت گره
ii. تعیین آزیموت بردار یکه مماس بر هر یک از چندخطی‌های (شامل منحنی و قطعات پاره‌خط) متصل به گره

عوارض کاندیدی که به جهت توپولوژیک با عارضه اصلی مشابهت دارند، وارد مرحله تعیین درجه شباهت با کمک روش کمترین مربعات و ویژگی‌های هندسی می‌شوند.

۳-۳- تعیین درجه‌ی شباهت

ویژگی‌های هندسی کاربرد فراوانی در تناظریابی داده دارد و پایه آشکارسازی عوارض متناظر در مجموعه داده‌های متفاوت است. مهمترین ویژگی‌های هندسی توصیف کننده یک چندخطی در ادامه تشریح خواهند شد.

فاصله بین دو عارضه: یک عارضه راه در دنیای واقعی با نمایش‌های متفاوتی در دو مجموعه داده ظاهر می‌شود، هر چه فاصله بین دو عارضه راه در دو مجموعه داده کمتر باشد، نمایشی دقیق‌تری را ارائه می‌دهند. تعاریف مختلفی برای فاصله بین دو چندخطی بیان شده است که از میان آنها می‌توان به فاصله هاسدروف [۱، ۱۵] و فاصله فریشت [۶، ۱۶] اشاره کرد.

آزیموت عارضه: معمولاً در تناظریابی عوارض چندخطی، آزیموت پاره‌خط فرضی که نقاط آغاز و پایان یک چندخطی را به یکدیگر وصل می‌کنند، به عنوان آزیموت چندخطی در نظر گرفته می‌شود [۱، ۱۷].

طول عارضه: طول یک چندخطی از ویژگی‌های هندسی می‌باشد که بدون واسطه و به راحتی قابل محاسبه است. طول یک چندخطی عبارت است از مجموع فاصله‌ی اقلیدسی بین نقاط میانی یک چندخطی.

اندازه عارضه: اندازه از جمله پارامترهایی است که در تناظریابی عوارض به کار می‌رود. در برخی موارد، طول یک چندخطی را با اندازه آن یکسان در نظر می‌گیرند. اما به طور دقیق‌تر فاصله اقلیدسی بین دو نقطه آغاز و پایان چندخطی به عنوان اندازه چندخطی معرفی می‌شود [۱].

در مجموع، پارامترهای بیان شده به طور جدا از هم به بیان ویژگی‌های هندسی یک چندخطی اقدام می‌کنند. دو عارضه در مجموعه داده در واقع نماینده یک عارضه در جهان واقعی هستند و یکی از علت‌های اصلی که منجر به نمایش متفاوتی از دو عارضه می‌شود، بحث خطاها می‌باشد که در طی مراحل مختلف تهیه مجموعه داده‌ها مطرح می‌شود. از طرف دیگر، مجموعه داده‌ها با مقیاس‌های مختلف دارای سطوح مختلفی از جزئیات هستند. به طور مثال، طول یک عارضه‌ی راه در مقیاس بزرگتر بیشتر از طول همان عارضه در مقیاس کوچکتر است. دلیل این مسأله آن است که عارضه در مجموعه داده با مقیاس بزرگتر جزئیات بیشتری را اعم از شکستگی‌ها و پیچ و خم‌ها نشان می‌دهد که در مقیاس کوچکتر سطح این جزئیات کاهش می‌یابد. در نتیجه، عوامل بیان شده ما را به این سمت سوق داد تا ابتدا برای یک چندخطی، خط کمترین مربعاتی را تعریف کنیم تا چندخطی عارضه راه را نمایندگی کند.

۳-۲-۱- خط کمترین مربعات

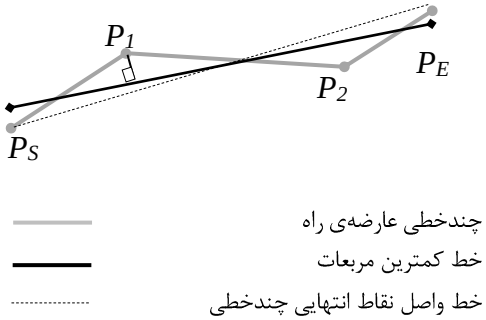
خط کمترین مربعات به عنوان نماینده یک چندخطی برای آن به صورت (۱) تعریف می‌شود:

$$Y = AX + B \quad Xs < X < Xe \qquad (۱)$$

که در این رابطه، Xs مولفه $\boldsymbol{X}$ نقطه آغاز و $\boldsymbol{Xe}$ مولفه $\boldsymbol{X}$ نقطه پایانی چندخطی است و پارامترهای A و B از روابط ۲ بدست می‌آیند.

$$A = \bar{Y} - B\bar{X} \quad و \quad B = \frac{\sum XY - N\bar{X}\bar{Y}}{\sum X^2 - N\bar{X}^2} \qquad (۲)$$

که در آن $\bar{\boldsymbol{X}}$ و$\bar{\boldsymbol{Y}}$ به ترتیب میانگین مولفه‌های $\boldsymbol{X}$ و$\boldsymbol{Y}$ نقاط میانی و نقاط آغازی وپایانی چندخطی است. شکل ۸ عارضه راه، خط نماینده کمترین مربعات آن عارضه و خط واصل نقاط انتهایی (که در سایر مقالات استفاده شده است) را نشان می‌دهد.



شکل ۸– عارضه ی راه و خط کمترین مربعات به نمایندگی از آن

برای خط تعریف شده چند ویژگی هندسی تعریف می‌شود.

اختلاف نسبی آزیموت (توجیه): زاویه بین آزیموت دوخط کمترین مربعات به عنوان پارامتر توجیه در نظر گرفته می‌شود.

$$d(\Delta Az) = \frac{(Az(pl_1) - Az(pl_2))}{\Delta Az_t} \qquad (۳)$$

که در این رابطه، ΔAz_t اختلاف آزیموت بین چندخطی‌های تناظریابی شده است که بطور تجربی تعیین می‌شود.

اختلاف نسبی طول: اختلاف طول دو خط کمترین مربعات که دو چندخطی کاندیدای متناظر نمایندگی می‌کند به صورت ۴ بیان می‌شود:

$$d(\Delta l) = \frac{l(pl_1) - l(pl_2)}{\Delta l_t} \qquad (۴)$$

که در این رابطه، $\boldsymbol{\Delta l_t}$ پارامتری است که به صورت تجربی با کمک چندخطی‌های تناظریابی شده بدست می‌آید.

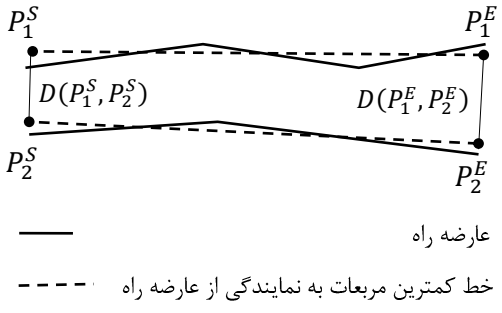
فاصله دو عارضه: میانگین فاصله اقلیدسی بین دو نقطه آغازین و پایانی خط کمترین مربعات به عنوان فاصله در نظر می‌گرفته می‌شود و از رابطه ۵ بدست می‌آید.

$$d(Dis) = Dis(pl_1, pl_2) / Dis_t \qquad (۵)$$

که در این رابطه،$\boldsymbol{Dis_t}$فاصله بین چندخطی‌های تناظریابی شده که به طور تجربی در نظر گرفته می‌شود و $Dis(.,.)$ از رابطه ۶ بدست می‌آید.

$$Dis(pl_1, pl_2) = \big(D(P_1^S, P_2^S) + D(P_1^E, P_2^E)\big)/2 \qquad (۶)$$

که در این رابطه، $D(P_1^E, P_2^E)$ فاصله اقلیدسی بین نقاط پایانی دو خط کمترین مربعات و $D(P_1^S, P_2^S)$ فاصله اقلیدسی بین نقاط آغازی دو خط کمترین مربعات است (شکل ۹).



شکل ۹- فاصله دو عارضه

پس از اینکه دو عارضه راه در مرحله جستجو در دو مجموعه داده به عنوان عوارض متناظر کاندید شدند، برای هر کدام از آنها خط کمترین مربعات به نمایندگی از آن عارضه راه (چندخطی راه) برازش داده می‌شود و ویژگی‌های هندسی بیان‌شده محاسبه می‌گردد. چنانچه هر سه مقدار بدست آمده شامل (اختلاف نسبی طول، اختلاف نسبی آزیموت، فاصله بین دو عارضه)کمتر از مقدار یک بود (چون نرمال شده‌اند)، دو عارضه به‌عنوان عوارض متناظر هم شناخته شده و به یکدیگر مرتبط می‌گردند.

برای بدست آوردن حدود آستانه به صورت دستی تناظریابی انجام می‌شود. بعد از تناظریابی مجموعه داده-

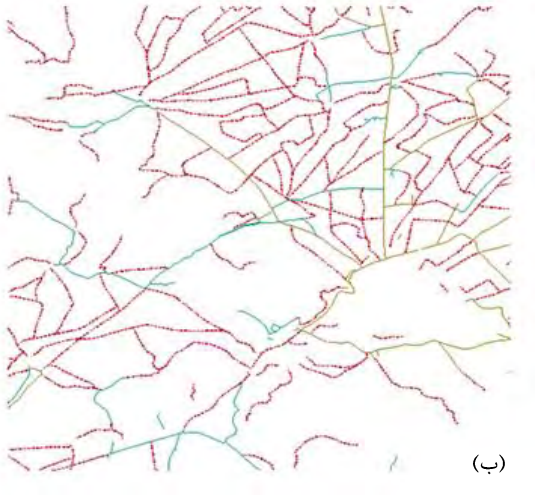
های برداری به صورت دستی پارامترهای هندسی برای داده‌ها محاسبه شده و با توجه به مقادیر بدست آمده حدود آستانه برای قبول یا رد شباهت دو عارضه‌ی تعیین می‌گردد.

۴- پیاده‌سازی

برای پیاده‌سازی الگوریتم از برنامه نویسی اشیای ArcObjects در فضای VBA نرم‌افزار ArcGIS 9.3 استفاده شده است. مرحله پیش پردازش با کمک ابزارهای Editor و Topology همین نرم‌افزار انجام شده و مراحل جستجوی توپولوژیک و تعیین درجه شباهت بطور کامل برنامه‌نویسی شده است.

مجموعه داده‌هایی که برای آزمون الگوریتم پیشنهادی مورد استفاده قرار گرفته‌اند، نقشه‌های پایه توپوگرافی تولید شده سازمان نقشه‌برداری کشور می‌باشند. این داده-ها شامل چهار شیت نقشه توپوگرافی با مقیاس ۱:۲۵۰۰۰ و یک شیت نقشه توپوگرافی با مقیاس ۱:۵۰۰۰۰ در ناحیه جغرافیایی یکسانی با طول‌های جغرافیایی ″۴۹°۴۵′۰۰ و ″۴۹°۳۰′۰۰ و عرض‌های جغرافیایی ″۳۴°۰۰′۰۰ و ″۳۴°۱۵′۰۰ درجه می‌باشند. نقشه‌های ۱:۲۵۰۰۰ در فرمت Shapefile و نقشه‌های ۱:۵۰۰۰۰ در فرمت DGN ارائه می‌شود (شکل ۱۰).

(الف)

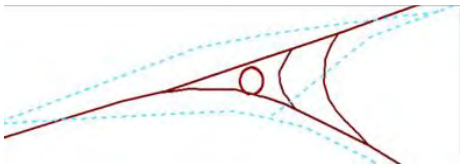


(ب)

شکل ۱۰ –(الف) نقشه در مقیاس ۱:۲۵۰۰۰ و (ب) نقشه در مقیاس ۱:۵۰۰۰۰

برای ارزیابی الگوریتم، ابتدا عوارض راهی که به دلیل اختلاف مقیاس در مجموعه داده‌های مقیاس کوچکتر نمایش داده نمی‌شدند، کنار گذاشته شدند. سپس نسبت عوارض درست تناظریابی شده به کل عوارضی که می-بایست در فرآیند اتوماتیک تناظریابی به عنوان عوارض متناظر شناسایی می‌شدند، محاسبه شد که این مقداربرابر با ۹۴/۲٪ بدست آمد. این درصد، نسبت به تحقیقات مشابه همچون پژوهش ژانگ [۷] که ۹۳/۷٪ بوده وکار کیلر [۹] که ۹۰/۵٪ و کار لوشار [۸] با ۸۹/۳٪ بوده است، مقایسه شده که برتری روش پیشنهادی را نشان می‌دهد.

با توجه به نرخ تناظریابی، برای حدود ۵٪ از عوارض، تناظریابی به درستی صورت نگرفته، که این مربوط به تقاطع‌های بزرگراهی متصل با ورودی و خروجی شهرها است که شرایط توپولوژیک خاصی دارند (شکل ۱۱).



عارضه‌ی راه در مجموعه داده اول

عارضه‌ی راه در مجموعه داده دوم

شکل ۱۱- بخشی از داده های ارزیابی که درفرآیند تناظریابی به عنوان عوارض متناظر شناسایی نشده اند.

در مرحله پیاده‌سازی، الگوریتم بدون بکارگیری خط کمترین مربعات نیز برنامه‌نویسی شد و پارامترهای آماری برای دو ویژگی طول عارضه و آزیموت عارضه در دو حالت محاسبه شد. نتایج نشان داد که میانگین اختلاف بین ویژگی‌های هندسی عوارض راه (طول و آزیموت) در زمان استفاده از خط کمترین مربعات کمتر بود (جدول ۱).

جدول ۱- اختلاف حدود آستانه در دو حالت استفاده از روش کمترین مربعات و بدون استفاده از روش کمترین مربعات

حدود آستانه / ویژگی‌های هندسی	با کمک روش کمترین مربعات	بدون استفاده از روش کمترین مربعات
میانگین اختلاف طول(متر)	۱۱/۳	۱۵/۳۴
میانگین اختلاف آزیموت (درجه)	۱/۲۵۲۲	۱/۷۲۹۵

این مسأله باعث شد در زمان تعیین حدود آستانه برای تعیین درجه شباهت بین ویژگی‌های هندسی حدود دقیق‌تری را در نظر بگیریم که در نتیجه با درجه اطمینان بیشتری می‌توانیم از عوارض متناظر دفاع کنیم. از طرفی با معرفی خط کمترین مربعات به عنوان نماینده‌ی یک چندخطی یک شاخص جامع و یکپارچه برای مقایسه مورد استفاده قرار می‌گیرد.

۵- نتیجه‌گیری

با پیشرفت سریع تولید و جمع‌آوری داده‌های مکانی و تکنیک‌های پردازش آن‌ها، همواره حجم عظیمی از داده‌ها که ناحیه جغرافیایی یکسانی را می‌پوشاند، در دسترس قرار می‌گیرند. این حجم داده، متخصصین حوزه علوم ژئوماتیک را به ایجاد پایگاه داده چندمقیاسی با هدف به روز رسانی مجموعه داده‌ها و جنرالیزاسیون تحت وب و آنالیزهای چندمقیاسی و... ترغیب می‌کند. هسته‌ی اصلی تولید پایگاه‌های چندمقیاسی تناظریابی داده است.

در این مقاله سعی شده است مسأله تناظریابی عوارض در یک پایگاه داده چندمقیاسی با ارائه روشی جدید مورد بررسی و پیاده‌سازی قرار گیرد. در الگوریتم بکار رفته در فاز جستجو رویکردی جدید برای بکارگیری ویژگی‌های توپولوژیک ارائه گردید تا فرآیند جستجو به سهولت و سریع صورت پذیرد.از دیگر نوآوری‌های این مقاله معرفی خط کمترین مربعات به عنوان نماینده‌ی عارضه‌ی خطی راه است که برای مقایسه عوارض راه کاندیدا به کار رفته است، به طوری که تناظریابی با کمک این نماینده دقیق‌تر صورت می‌پذیرد. ویژگی بارز این خط نماینده، حذف اثرات اختلاف هندسی عوارض راه که ناشی از اختلاف مقیاس بوده، می‌باشد. همچنین تحلیلی آماری در رابطه با بدست آوردن حدود آستانه برای تعیین درجه شباهت ویژگی‌های هندسی عارضه راه در دو حالت، یکی در زمان استفاده از خط نماینده (کمترین مربعات) و در حالت دیگر، بدون استفاده از خط نماینده بود، که معیاری را برای تعیین حدود آستانه بدست آوردیم و از طرفی نقش بکارگیری خط نماینده به خوبی مشخص شد. همچنین اضافه نمودن خط نماینده با توجه به نقش مهمی که در این الگوریتم ایفا نمود، کمترین تاثیر را بر روی سرعت اجرای الگوریتم داشته است که از جمله مزیت‌های روش پیشنهادی می‌باشد.

در پایان پیشنهادهایی برای کارهای آینده به نظر می‌رسد که شرح آن به صورت ذیل است:

- عوارض خطی به دو گروه عوارض خطی طبیعی مانند رودخانه‌ها و عوارض مصنوعی و دست‌ساز بشر مانند راه و تاسیسات انتقال نیرو تقسیم‌بندی می‌شوند. این دو گروه عارضه به جهت ماهیت و نحوه شکل‌گیری با یکدیگر تفاوت دارند. لذا تناظریابی عوارض خطی طبیعی مانند رود خانه‌ها نیز نیازمند تحقیق و بررسی می‌باشد.
- در پژوهش انجام گرفته تناظریابی بین مقیاس‌های ۱:۲۵۰۰۰ و ۱:۵۰۰۰۰ انجام گرفته است. تناظریابی بین مقیاس‌هایی با اختلاف بیشتر مانند مقیاس‌های ۱:۲۵۰۰۰ و۱:۲۰۰۰۰۰ چالش‌ها و مطالعات خاص خود را می‌طلبد.
- حل مسئله تناظریابی در قسمت تقاطع‌های بزرگراهی که الگوریتم پیشنهادی در این زمینه به خوبی عمل نمی‌کند.

مراجع

[1] Zhang, M. (2009). “Methods and implementations of road-network matching.”Doktorarbeit, Lehrstuhl für Kartographie, TU München.

[2] Hampe, M., Anders, K. H., & Sester, M. (2003, August).”MRDB applications for data revision and real-time generalisation.”In Proceedings of 21st International Cartographic Conference (Vol. 10, p. 16).

[3] Anders, K. H., & Bobrich, J. (2004, August). MRDB approach for automatic incremental update. In ICA workshop on generalisation and multiple representation (pp. 20-21).

[4] Song, W., Haithcoat, T. L., & Keller, J. M. (2006).”A snake-based approach for TIGER road data conflation.” Cartography and Geographic Information Science, 33(4), 287-298

[5] Cecconi, A. (2003).”Integration of cartographic generalization and multi-scale databases for enhanced web mapping” (Doctoral dissertation, Universität Zürich).

[6] Mustière, S., & Devogele, T. (2008).”Matching networks with different levels of detail”. GeoInformatica, 12(4), 435-453.

[7] Zhang, M., Shi, W., & Meng, L. (2005, July).” A generic matching algorithm for line networks of different resolutions.”In Workshop of ICA Commission on Generalization and Multiple Representation Computering Faculty of A Coruña University-Campus de Elviña, Spain.

[8] Lüscher, P., Burghardt, D., & Weibel, R. (2007, August).”Matching road data of scales with an order of magnitude difference.”In Proc. XXIII International Cartographic Conference, Moscow, Russia.

[9] Kieler, B., Huang, W., Haunert, J. H., & Jiang, J. (2009).”Matching River Datasets of Different Scales.” Advances in GIScience, 135-154.

[10] Li, L., & Goodchild, M. F. (2010).”Optimized feature matching in conflation.” In Geographic Information Science: 6th International Conference, GIScience (pp. 14-17).

[11] Safra, E., Kanza, Y., Sagiv, Y., & Doytsher, Y. (2012).”Ad hoc matching of vectorial road networks.” International Journal of Geographical Information Science,DOI:10.1080/13658816.2012.667104

[12] Yang, B., Zhang, Y., & Luan, X. (2012).”A probabilistic relaxation approach for matching road networks.” International Journal of Geographical Information Science, DOI:10.1080/13658816.2012.68348

[13] Hu, Y., Chen, J., Li, Z., Zhao R. (2008):” Road Data Updating Using Tools of Matching and Map Generalization.” ISPRS Vol. XXXVII, Part B4. Beijing, 2008

[14] Parent, C., & Spaccapietra, S. (2000).” Database integration: the key to data interoperability.” Advances in Object-Oriented Data Modeling, 221-253.

[15] Min, D., Zhilin, L., & Xiaoyong, C. (2007).”Extended Hausdorff distance for spatial objects in GIS.” International Journal of Geographical Information Science, 21(4), 459-475.

[16] Eiter, T. &Mannila H. (1994).” Computing discrete Fréchet distance,” Technical report of Christian DopplerLaborfürExpertensensyteme, Technical University of Vienna, Austria, http://www.kr.tuwien.ac.at/staff/eiter/et-archive/cdtr9464.ps.gz (accessed on 2012-10-27)

[17] Raimond, A.-M. O. &Mustière, S. (2008).”Data Matching - a Matter of Belief”, Headway in Spatial DataHandling, 13th International Symposium on Spatial Data Handling (SDH), pp. 501-519.

[18] Safra, E & Doytsher, Y. (2006).” Using matching algorithms for improving locations in cadastral maps”,XXIII FIG Congress, Munich, Germany, October 2006

[19] Butenuth, M., Gösseln, G.v., Tiedge, M., Heipke C., Lipeck, U. & Sester M. (2007).” Integration ofheterogeneous geospatial data in a federated database,” ISPRS Journal of Photogrammetry &Remote Sensing 62 (2007), 328-346

[20] Worboys M. and Duckham M. 2004, GIS: A Computing Perspective, Second Edition, CRC Press.

[21] Breunig, M., Thomsen, A., Broscheit, B., Butwilowski, E., & Sander, U. (2007).”Representation and Analysis of Topology in Multi-Representation Databases.”In Proceedings of PIA (Vol. 7, p. 6).