



## تنظیریابی هندسی رودخانه‌ها در مجموعه داده‌های چندمقیاسی برداری

علیرضا چهرقان<sup>۱\*</sup>، رحیم علی‌عباسپور<sup>۲</sup>

۱- دانشجوی دکتری GIS، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران

۲- استادیار گروه GIS، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران

### چکیده:

تنظیریابی به شناسایی عوارض با ماهیت یکسان در مجموعه داده‌های مختلف گفته می‌شود. در تحقیق حاضر با تمرکز بر عوارض خطی و انتخاب رودخانه‌ها به عنوان مطالعه موردی، شناسایی عوارض متناظر در دو مجموعه داده با مقیاس‌های ۱:۲۵۰۰۰ و ۱:۵۰۰۰ تولید شده از منابع تهیه متفاوت صورت گرفته است. در تحقیق حاضر با در نظر گرفتن معیارهای هندسی نظری فاصله، جهت، مساحت، شکل و ناحیه مشترک بین حریم و همچنین در نظر گرفتن ادراک مکانی کارشناسان در تعیین درجه اهمیت هریک از معیارها، درجه شباهت عوارض محاسبه می‌گردد. علاوه بر موارد فوق، در راهکار ارائه شده تمام روابط بین عوارض شامل یک به هیچ، هیچ به یک، یک به یک، یک به چند، چند به یک و چند به چند در نظر گرفته می‌شود. نتایج نشان می‌دهد که روش پیشنهادی با محاسبه مقدار F – Score ۶۱/۸۴ درصد به خوبی عوارض متناظر را شناسایی می‌کند.

واژه‌های کلیدی: تنظیریابی، معیارهای هندسی، مجموعه داده‌های چندمقیاسی، رودخانه‌ها.

## ۱- مقدمه

تئاظریابی به شناسایی عوارض با ماهیت یکسان در مجموعه داده‌های مختلف گفته می‌شود [۱]. با توجه به انواع مدل داده‌های برداری و شبکه‌ای، تئاظریابی نیز به انواع مختلفی نظیر تئاظریابی بردار-شبکه، بردار-بردار و شبکه-شبکه تقسیم می‌گردد. منتهی تحقیق حاضر با تمرکز بر تئاظریابی بردار-بردار و با تأکید بر داده‌های رودخانه ارائه شده است. شناسایی عوارض متراکم در مجموعه داده‌های مختلف می‌تواند کاربردهای مختلفی داشته باشد که ضرورت تئاظریابی را نشان می‌دهد. همچنین تئاظریابی کاربردهای مستقیم و غیر مستقیم در تلفیق، ارزیابی کیفیت، یکپارچه‌سازی داده‌ها و مدیریت داده‌ها دارد، منتهی در بسیاری از تحقیقات انجام شده تا به امروز [۲-۱۷]، هدف اصلی تئاظریابی یکپارچه‌سازی مجموعه داده‌های مختلف ذکر شده است. در علوم مکانی تمایل به یکپارچه‌سازی مجموعه داده‌های مکانی مختلف در حال افزایش می‌باشد. در بین مجموعه داده‌های مکانی مختلف می‌توان به محبوبیت داده‌های مکانی داوطلبانه (VGI<sup>۱</sup>) در کنار مجموعه داده‌های رسمی و همچنین گسترش پیاده‌سازی زیر ساخت اطلاعات مکانی اشاره نمود. در تحقیقات مختلف تئاظریابی عوارض را به عنوان هسته مرکزی یکپارچه‌سازی داده‌ها در نظر گرفته‌اند [۱۸]. مخصوصاً با توجه به افزونگی داده در وضعیت امروزی، برخی از محققین این فرایند را مرتبط با "کلان داده"<sup>۲</sup> دانسته‌اند که تئاظریابی دستی بین پایگاه‌های داده مکانی را غیر عملی توصیف کرده‌اند [۱۹، ۱۸]. در تحقیقی که توسط Kang و همکاران (۲۰۱۵)، بیان شده است که برای تئاظریابی دستی ۴۰۰۰ عارضه خطی به میزان ۹۰ ساعت زمان صرف شده است [۲۰]. از این رو اهمیت هر چه بیشتر ارائه روش‌های تئاظریابی نمایان است.

علاوه بر یکپارچه‌سازی مجموعه داده‌ها، در تحقیقات دیگری نظیر Saalfeld در سال ۱۹۹۳، Hastings در سال ۲۰۰۸، Li و Goodchild در سال ۲۰۱۱، Song و همکاران (۲۰۱۱)، Pourabdollah و همکاران (۲۰۱۳)، Touya و همکاران (۲۰۱۳)، Tong و همکاران (۲۰۱۴)، Yang و همکاران (۲۰۱۴) و Zhang و همکاران (۲۰۱۶) از تئاظریابی با هدف غنی‌سازی مجموعه داده‌های مختلف استفاده شده است که منجر به تلفیق مجموعه داده‌ها گردیده است [۳۰-۲۱]. همچنین تعدادی از تحقیقات از تئاظریابی به عنوان ابزاری برای ارزیابی کیفیت داده‌ها استفاده کرده‌اند. Brunz و Egenhofer در سال ۱۹۹۶ و Sheeren و همکاران (۲۰۰۹) برای ارزیابی ناسازگاری [۳۲، ۳۱]، Seo و Ohara در سال ۲۰۰۹ و Abdolmajidi و همکاران (۲۰۱۲)، Lendinez و همکاران (۲۰۱۲) برای ارزیابی کامل بودن [۳۶-۳۳، ۱۲]، Abdolmajidi و همکاران (۲۰۱۶) برای بررسی صحّت مکانی [۳۶-۳۳، ۱۲]، Ludwig و همکاران (۲۰۱۱)، Koukoltsos و همکاران (۲۰۱۱) و Cao و Fan و همکاران (۲۰۱۴) برای ارزیابی کاملاً صحّت اطلاعات توصیفی استفاده کردند [۴۰، ۳۹]. تحقیقاتی نظیر Wingaarden و همکاران (۲۰۱۶) برای ارزیابی صحّت اطلاعات توصیفی استفاده کردند [۴۰، ۳۹]. تحقیقاتی نظیر Wang و Hmكاران (۱۹۹۷)، Haunert و Sester در سال ۲۰۰۵ و همکاران (۲۰۰۸)، Mustiere در سال ۲۰۰۵ و همکاران (۲۰۰۹) و Wei و Wang در سال ۲۰۰۹ و Yuan در سال ۲۰۰۵ برای شناسایی تغییرات استفاده نمودند [۸، ۴۱، ۴۲، ۴۳، ۴۴، ۴۵ و ۴۶].

در حالت کلی جهت تئاظریابی عوارض سه خصوصیت هندسی، تپولوژیکی و معنایی در نظر گرفته می‌شود که بر اساس آنها میزان مشابهت عوارض به یکدیگر جهت شناسایی عوارض متراکم سنجیده می‌شود [۴۱]، اما در برخی مسائل امکان دسترسی به معیارهای معنایی وجود ندارد و یا اینکه در مجموعه داده‌های با تعداد عوارض بالا به دلیل زمان بالای محاسبات، نیاز است تا حداقل معیارها در نظر گرفته شود. از این رو تحقیق حاضر با در نظر گرفتن معیارهای مرتبط با هندسه عوارض به شناسایی عوارض متراکم در مجموعه داده‌های خطی با تمرکز بر رودخانه‌ها می‌پردازد. در راهکار ارائه شده سعی شده است با در نظر گرفتن معیارهایی هندسی نظیر فاصله، جهت،

<sup>1</sup> Volunteered Geospatial Information

<sup>2</sup> Big Data

مساحت، شکل و ناحیه مشترک بین حریم درجه شباهت عوارض محاسبه گردد. دلیل استفاده از معیارهای هندسی این است که در تمامی مجموعه داده‌ها امکان دسترسی به این معیارها وجود دارد و نیاز به داده اضافه‌ای به جز هندسه عارضه مورد نظر نمی‌باشد.

ادامه این تحقیق به شرح ذیل است: پس از مقدمه در بخش ۲ چارچوب پیشنهادی ارائه و جزئیات مورد استفاده تشریح می‌گردد. در بخش ۳ ضمن تشریح منطقه مورد مطالعه، چارچوب پیشنهادی بر روی این منطقه اعمال و نتایج مورد بررسی قرار می‌گیرد. در انتها نیز در بخش ۴ نتیجه گیری و پیشنهادات ارائه می‌گردد.

## ۲- راهکار پیشنهادی

جهت تئوریابی رودخانه‌ها در مجموعه داده‌های با مقیاس و منبع تولید متفاوت راهکاری مطابق با شکل (۱) ارائه شده است. برای تفسیر هریک از مؤلفه‌های شکل (۱) نیاز است تا دو مفهوم "عارضه مرجع" و "عارض کاندید" تشریح گردد. عارضه مرجع به هر یک از عوارض در مجموعه داده اول گفته می‌شود که هدف شناسایی عوارض متناظر برای آنها در مجموعه داده دوم می‌باشد. عارض کاندید نیز به تعدادی از عوارض در مجموعه داده دوم گفته می‌شود که هدف استخراج عوارض متناظر با عارضه مرجع از بین آنهاست. در ادامه هریک از مراحل راهکار پیشنهادی تشریح می‌گردد.



شکل ۱: راهکار پیشنهادی جهت شناسایی عوارض متناظر در مجموعه داده‌های مختلف

## ۲-۱- پیش پردازش

برای تئوریابی مجموعه داده‌های با مقیاس و منابع متفاوت، در ابتدا نیاز است که یک مرحله پیش پردازش بر روی مجموعه داده‌ها صورت گیرد. از آنجاییکه ممکن است دو مجموعه داده در دو قالب متفاوت باشد، در ابتدا هر دو مجموعه داده تبدیل به قالب یکسان می‌شوند. همچنین ممکن است دو مجموعه داده دارای سیستم مختصات‌های متفاوتی از یکدیگر باشند، از این رو نیاز است تا هر دو مجموعه به یک سیستم مختصات تبدیل گردد. در انتها نیز برای کاهش خطاهای سیستماتیک در مجموعه داده‌ها، خطاهای توپولوژیک حذف می‌گردد.

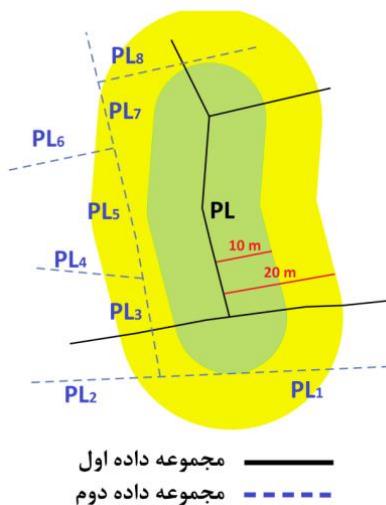
## ۲-۲- شناسایی عوارض کاندید برای عارضه مرجع

در این مرحله برای عارضه‌ای نظری PL حریمی به فاصله  $\beta$  در نظر گرفته می‌شود و در صورتیکه عوارض کاندید در این محدوده شناسایی نگردند فاصله  $\beta$  افزایش می‌باید. این الگوریتمی مقادیری را به عنوان بیشترین فاصله رشد و گام‌های رشد در نظر می‌گیرد. برای مثال در شکل (۵-۲) برای عارضه PL در مجموعه داده اول، در ابتدا حریمی به فاصله ۱۰ متر زده می‌شود. در این فاصله عوارض کاندید در مجموعه داده دوم شناسایی نمی‌گردد، از این رو فاصله

حریم‌گیری به ۲۰ متر افزایش می‌یابد. در فاصله حریم ۲۰ متر عوارضی نظیر PL<sub>1</sub>, PL<sub>2</sub>, PL<sub>3</sub>, PL<sub>4</sub>, PL<sub>5</sub>, PL<sub>6</sub>, PL<sub>7</sub> و PL<sub>8</sub> به عنوان عوارض کاندید برای PL انتخاب می‌گردد. چالش اصلی در این الگوریتم تعیین مقدار  $\beta$  می‌باشد که تأثیر مستقیمی بر زمان محاسبات و دقت تناظریابی دارد. این مقدار به دو صورت می‌توان در نظر گرفت: از طریق تجربه و از طریق دقت مکانی دو مجموعه داده. در حالت اول بسیاری از تحقیقات [۵, ۲۱, ۴۷] با توجه به مقیاس‌های ورودی و بر اساس تجربه مقداری را به  $\beta$  اختصاص می‌دهند. در حالت دوم تعدادی از تحقیقات [۱۱, ۵۱, ۵۲ و ۵۳] رابطه‌ای را با توجه به دقت مکانی دو مجموعه در نظر گرفته و از این طریق  $\beta$  محاسبه می‌گردند. در این تحقیق با استفاده از دقت مکانی دو مجموعه مقدار فاصله حریم در نظر گرفته می‌شود. رابطه (۱) نحوه محاسبه  $\beta$  را نشان می‌دهد.

$$\beta = \sqrt{\sigma_1^2 + \sigma_2^2} \quad (1)$$

در این رابطه  $\sigma_1$  و  $\sigma_2$  دقت مکانی دو مجموعه داده اول و دوم می‌باشد.



شکل ۲: الگوریتم BG برای عارضه PL در مجموعه داده اول

### ۳-۲- محاسبه درجه شباهت مکانی

پس از اینکه عوارض کاندید در مجموعه داده دوم برای عارضه مرجع در مجموعه داده اول شناسایی گردید، نیاز است تا میزان شباهت مکانی هریک از عوارض کاندید با عارضه مرجع محاسبه گردد. شباهت عوارض کاندید با عارضه مرجع از طریق محاسبه درجه شباهت مکانی و با در نظر گرفتن ویژگی‌های هندسی صورت می‌پذیرد. رابطه (۲) نحوه محاسبه درجه شباهت مکانی (به درصد) برای دو عارضه خطی PL<sub>1</sub> و PL<sub>2</sub> را نشان می‌دهد.

$$\text{Sim}(PL_1, PL_2) = \frac{\sum_{i=1}^n W_i \text{Sim}_{PL_1, PL_2}^{C_i}}{\sum_{i=1}^n W_i} \times 100 \quad (2)$$

در این رابطه  $\text{Sim}_{PL_1, PL_2}^{C_i}$  مقدار نرمال شده و  $W_i$  وزن اختصاص داده شده به معیار  $C_i$  می‌باشد. همچنین  $n$  تعداد معیارهای مورد استفاده و  $\text{Sim}(PL_1, PL_2)$  درجه شباهت مکانی بین عارضه PL<sub>1</sub> و PL<sub>2</sub> می‌باشد که دارای مقادیری بین صفر تا ۱۰۰ است. معیارهای مورد استفاده در این تحقیق شامل فاصله، جهت، مساحت، شکل و ناحیه مشترک حریم می‌باشد.

## - فاصله -

هاسدورف یکی از پرکاربردترین روش‌های محاسبه فاصله در علوم مکانی می‌باشد که مناسب بودن آن در محاسبه نزدیکی بین عوارض خطی به اثبات رسیده است [۵۴]. این فاصله به عنوان بیشترین فاصله بین کوتاهترین فاصله موجود بین هر نقطه از عارضه اول و مجموعه نقاط عارضه دوم تعریف شده است [۲۶]. با این حال فاصله هاسدورف نسبت به شکل دو عارضه، به خصوص نسبت به قسمت‌های دورافتاده از مرکز حساس می‌باشد. تحقیق [۲۵] با معرفی هاسدورف میانه بر مبنای طول نشان داد که این روش در مقایسه با روش‌های هاسدورف و هاسدورف میانه دارای واریانس کمتر و عملکرد کارامدتری در مواجه با داده‌ای پیچیده جهت اندازه‌گیری فاصله بین چندخطی‌ها می‌باشد.

رابطه (۳) فاصله هاسدورف میانه بر مبنای طول بین دو چندخطی مفروض را نشان می‌دهد [۲۵].

$$\overline{Sim}_{PL_1, PL_2}^{C_1} = \begin{cases} m(PL_1, PL_2), & \text{if } L(PL_1) < L(PL_2) \\ m(PL_2, PL_1), & \text{if } L(PL_1) \geq L(PL_2) \end{cases} \quad \text{رابطه (۳)}$$

که در این رابطه  $L(PL_1)$  و  $L(PL_2)$  طول دو عارضه خطی  $PL_1$  و  $PL_2$ ،  $m(PL_1, PL_2)$  و  $m(PL_2, PL_1)$  نیز از طریق روابط (۴) و (۵) بدست می‌آیند.

$$m(PL_1, PL_2) = \text{median}_{P_a \in PL_1} \left\{ \min_{P_b \in PL_2} \|P_a - L_b\| \right\} \quad \text{رابطه (۴)}$$

$$m(PL_2, PL_1) = \text{median}_{P_b \in PL_2} \left\{ \min_{P_a \in PL_1} \|P_b - L_a\| \right\} \quad \text{رابطه (۵)}$$

در این روابط  $L_a$  و  $L_b$  دو یال اختیاری از عوارض خطی  $PL_1$  و  $PL_2$ ،  $\|P_a - L_b\|$  فاصله عمودی یکی از نقاط عارضه  $PL_1$  از یکی از یال‌های عارضه  $PL_2$  ( $L_b$ ) و  $\|P_b - L_a\|$  فاصله عمودی یکی از نقاط عارضه  $PL_2$  ( $P_b$ ) از یکی از یال‌های عارضه  $PL_1$  ( $L_a$ ) می‌باشد.

## - جهت -

برای یک عارضه خطی، جهت به زاویه بین خط تشکیل شده از گره‌های ابتدایی و انتهایی عارضه و محور افقی گفته می‌شود. در این صورت برای دو عارضه خطی با جهت‌های  $\alpha$  و  $\beta$  اختلاف جهت برابر  $|\alpha - \beta|$  خواهد بود که در رابطه (۶) نمایان است.

$$\overline{Sim}_{PL_1, PL_2}^{C_2} = |\alpha - \beta| = \cos^{-1} \left( \frac{\vec{V}_{PL_1} \cdot \vec{V}_{PL_2}}{\|\vec{V}_{PL_1}\| \cdot \|\vec{V}_{PL_2}\|} \right) \quad \text{رابطه (۶)}$$

در این رابطه  $\vec{V}_{PL_1}$  بردار تشکیل شده از گره‌های ابتدایی و انتهایی عارضه اول،  $\vec{V}_{PL_2}$  بردار تشکیل شده از گره‌های ابتدایی و انتهایی عارضه دوم و عملگر  $\|\cdot\|$  فاصله اقلیدسی بین گره‌های ابتدایی و انتهایی عارضه مورد نظر می‌باشد.

## - مساحت -

از طریق اتصال گره ابتدایی به انتهایی در عارضه خطی یک چند ضلعی ایجاد می‌شود که می‌توان اختلاف بین مساحت‌های آن را به عنوان مبنای جهت مقایسه عوارض مختلف با یکدیگر مورد استفاده قرار داد.

## - شکل -

یکی از شناخته شده ترین و پرکاربردترین توابع مرتبط با شکل عارض، توصیفگر تابع تجمعی زاویه<sup>۱</sup> می‌باشد [۵۵]. رابطه (۷) نحوه محاسبه اختلاف شکل بین دو عارضه خطی  $PL_1$  و  $PL_2$  را نشان می‌دهد [۵۶]. در این رابطه  $(s)\theta$  تابع تجمعی زاویه عارضه مورد نظر می‌باشد.

<sup>1</sup> Cumulative Angle Function

(رابطه (۷)

$$\overline{\text{Sim}}_{\text{PL}_1, \text{PL}_2}^{\text{C}_4} = \int_0^1 f(\theta_{\text{PL}_1}, \theta_{\text{PL}_2}) ds, \quad f(\theta_{\text{PL}_1}, \theta_{\text{PL}_2}) = \begin{cases} |\theta_{\text{PL}_1} - \theta_{\text{PL}_2}|, & \text{if } |\theta_{\text{PL}_1} - \theta_{\text{PL}_2}| \leq \pi \\ 2\pi - |\theta_{\text{PL}_1} - \theta_{\text{PL}_2}|, & \text{if } |\theta_{\text{PL}_1} - \theta_{\text{PL}_2}| > \pi \end{cases}$$

### - مساحت مشترک حریم

از جمله معیارهای دیگر مورد استفاده در تحقیقات مختلف، در نظر گرفتن مساحت موجود بین منطقه مشترک بوجود آمده از حریم ایجاد شده بین عوارض می‌باشد [۳۶]. رابطه (۸) نحوه محاسبه اختلاف مساحت بین حریم عوارض را نشان می‌دهد که هرچه قدر مقدار این مساحت به یک نزدیکتر باشد دو عرضه از نظر هندسی به یکدیگر شبیه‌تر هستند.

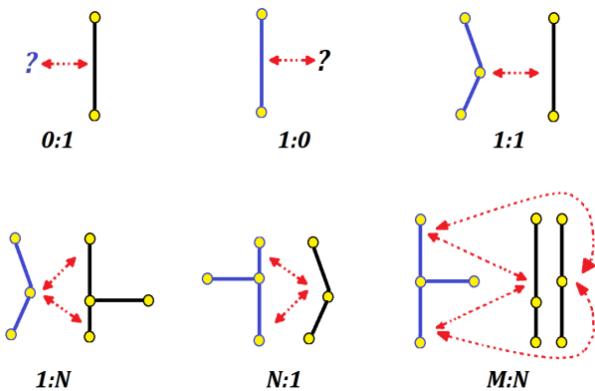
$$\overline{\text{Sim}}_{\text{PL}_1, \text{PL}_2}^{\text{C}_5} = \frac{2A_i}{A_{\text{PL}_1} + A_{\text{PL}_2}} \quad \text{رابطه (۸)}$$

در این رابطه  $A_{\text{PL}_1}$  مساحت حریم ایجاد شده برای عرضه اول،  $A_{\text{PL}_2}$  مساحت حریم ایجاد شده برای عرضه دوم و  $A_i$  مساحت منطقه مشترک بین دو حریم ایجاد شده می‌باشد.

### ۲-۲- تعیین عوارض متناظر

پس از اینکه برای جفت عوارض مرجع و کاندید درجه شbahت مکانی محاسبه گردید، از طریق مقدار بهینه حد آستانه شbahت ( $\tau = 80\%$ )، عوارض متناظر شناسایی می‌گردند. اگر مقدار درجه شbahت مکانی محاسبه شده برای دو عرضه بزرگتر از مقدار  $\tau$  باشد آن دو عرضه با یکدیگر متناظر هستند. در فرایند شناسایی عوارض متناظر نیاز است تا روابط موجود بین عوارض مرجع و کاندید مورد بررسی قرار گیرد. از این رو با توجه به شکل (۳) روابط بین عوارض در مجموع شامل شش حالت می‌شود که بصورت ذیل قابل شناسایی می‌باشد.

- در صورتیکه هیچ یک از مقادیر محاسبه شده برای درجه شbahت مکانی بزرگتر از مقدار  $\tau$  نباشد، حالت یک به هیچ رخ داده است.
- در صورتیکه تنها یکی از مقادیر محاسبه شده برای درجه شbahت مکانی بزرگتر از مقدار  $\tau$  باشد و همچنین این عرضه کاندید عارضه متناظر هیچ عارضه مرجع دیگری نباشد، حالت یک به یک رخ داده است.
- در صورتیکه بیش از یک مقدار بزرگتر از  $\tau$  باشد و همچنین این عوارض کاندید عارضه متناظر هیچ عارضه مرجع دیگری نباشد، حالت یک به چند رخ داده است.
- در صورتیکه تنها یکی از مقادیر محاسبه شده برای درجه شbahت مکانی بزرگتر از مقدار  $\tau$  باشد و همچنین این عرضه کاندید عارضه متناظر عارضه مرجع دیگری باشد، حالت چند به یک رخ داده است.
- در صورتیکه بیش از یک مقدار بزرگتر از  $\tau$  باشد و همچنین این عوارض کاندید عارضه متناظر عارضه مرجع دیگری باشند، حالت چند به چند رخ داده است.
- پس از بررسی تمامی عوارض موجود در مجموعه داده اول، در صورتیکه در مجموعه داده دوم عوارضی بدون عرضه متناظر در مجموعه داده اول باشند، این عوارض دارای رابطه هیچ به یک خواهد بود.



شکل ۳: روابط موجود بین عوارض مرجع و کاندید

### ۳-۲- ارزیابی

Precision و Recall دو پارامتری هستند که در تحقیقات بسیاری جهت ارزیابی نتایج حاصل از تئوریابی مورد استفاده قرار گرفته‌اند [۵۷, ۵۹, ۶۱, ۲۴, ۲۶] رابطه (۹) محاسبه Precision و Recall را نشان می‌دهد.

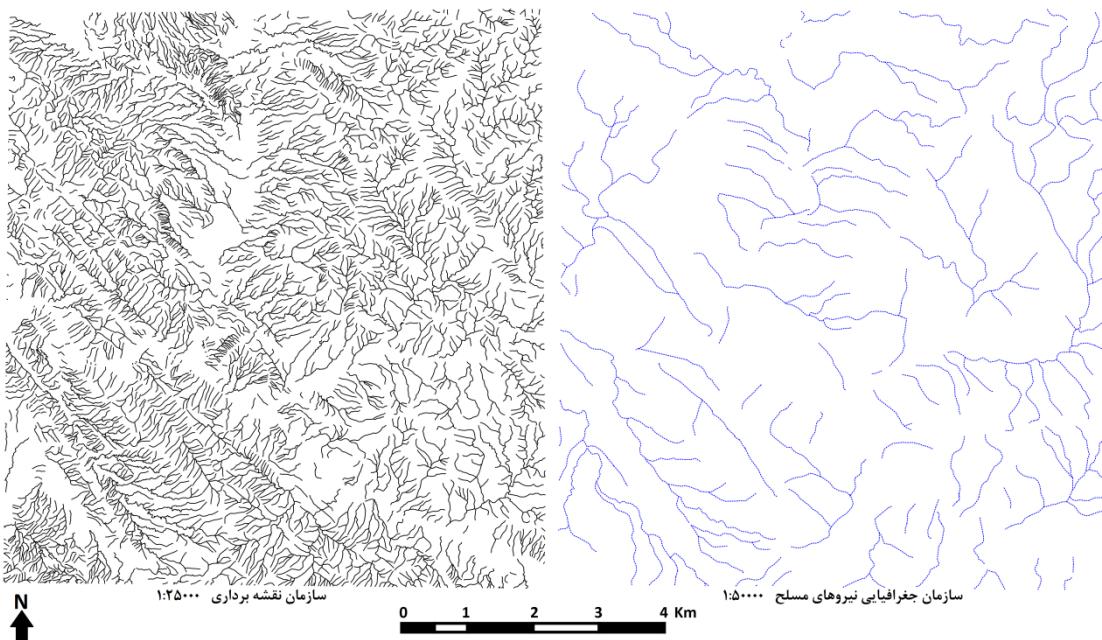
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad \text{رابطه (۹)}$$

در این روابط TP تعداد روابطی هستند که به درستی کشف شده‌اند، FP تعداد روابطی هستند که به اشتباه شناسایی شده‌اند و FN تعداد روابطی هستند که کشف نشده‌اند. اما ممکن است مقادیر Precision و Recall مخالف یکدیگر باشند، ممکن است Precision مقدار بالا و Recall مقدار پائین و یا Precision مقدار پائین و Recall مقدار بالا به دست آید، از این رو تعدادی از تحقیقات [۱۵, ۳۶, ۶۰, ۳۹] برای ارزیابی نهایی از مقدار F-Score بهره می‌برند. رابطه (۱۰) نحوه محاسبه مقدار F-score برای ارزیابی نهایی را نشان می‌دهد.

$$F - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{رابطه (۱۰)}$$

### ۳- پیاده‌سازی

برای تعیین عوارض متناظر از طریق رویکرد پیشنهادی از دو مجموعه داده با مقیاس‌ها و منابع تهیه متفاوت استفاده شده است. منطقه مورد مطالعه شامل قسمتی از منطقه ایلام می‌باشد که داده‌های مرتبط با رودخانه در دو مجموعه داده با مقیاس‌های ۱:۲۵۰۰۰ و ۱:۵۰۰۰۰ از این منطقه مورد استفاده قرار گرفته است. با توجه به شکل (۴) مجموعه داده اول در سازمان نقشه‌برداری کشور و از طریق عکس‌های هوایی اخذ شده در سال‌های ۱۳۸۱ و ۱۳۸۲ تهیه شده است. همچنین مجموعه داده دوم در سازمان جغرافیایی نیروهای مسلح و از طریق تصاویر ماهواره‌ای در سال ۱۳۸۵ تهیه شده است.



شکل ۴: مجموعه داده‌های مورد استفاده

همانطور که در راهکار پیشنهادی بیان گردید، در ابتدا برای اینکه بتوان داده‌های متضاظر را در هر دو گروه تعیین نمود، نیاز است بر روی داده‌ها عملیات پیش پردازش صورت گیرد. عملیات پیش پردازش شامل هم قالب نمودن داده‌ها، بردن به سیستم مختصات یکسان و حذف خطاهای توپولوژیکی در مجموعه داده‌ها می‌باشد. با توجه به جدول (۱) پس از انجام عملیات پیش پردازش بر روی داده‌ها، تعداد عوارض موجود در مجموعه داده اول شامل ۴۹۸۶ عارضه و در مجموعه داده دوم شامل ۳۶۰ عارضه می‌باشد.

جدول ۱: تعداد عوارض و مجموع طول عوارض در مجموعه داده‌های مورد استفاده

مجموعه داده‌ها	مقیاس	تعداد عوارض	طول عوارض
مجموعه داده اول	۱:۲۵۰۰۰	۴۹۸۲	۷۵۹۶۴۳/۶
مجموعه داده دوم	۱:۵۰۰۰۰	۳۶۰	۱۵۶۸۳۰

برای اینکه بتوان در انتها به ارزیابی نتایج پرداخت، نیاز است تا در ابتدای پیاده‌سازی روابط بین عوارض در بین مجموعه داده‌های اول و دوم شناسایی گردد. جدول (۲) روابط شش گانه شناسایی شده بین عوارض در دو مجموعه داده را نشان می‌دهد.

جدول ۲: روابط شش گانه شناسایی شده بین عوارض در دو مجموعه داده

مجموعه داده‌ها	هیچ به یک	یک به یک	یک به چند	چند به یک	چند به چند	چند به چند
۱:۲۵۰۰۰-۱:۵۰۰۰۰	۴۶۲۲	۴۰	۲۶	۲۸۵	۴	۵

در ادامه برای تعیین وزن هریک از معیارهای هندسی از منظر ادراک مکانی انسان از طریق پرسشنامه قضایت کارشناسان مرتبط (دانشجویان و فارغ التحصیلان مقطع کارشناسی (نقشه‌برداری)، کارشناسی ارشد و دکتری (GIS)) در مورد اهمیت هریک از معیارهای هندسی در سنجش میزان شباهت دو عارضه در نظر گرفته شد، میانگین وزنهای اخذ شده به عنوان وزن نهایی هریک از معیارها محاسبه گردید. جدول (۳) وزن نهایی استخراج شده را نشان می‌دهد.

جدول ۳: وزن نهایی بدست آمده برای هر یک از معیارها

معیارهای هندسی				
ناحیه مشترک حریم	شکل	مساحت	جهت	فاصله
۰/۶۵	۰/۷۰	۰/۴۵	۰/۳۱	۰/۶۲

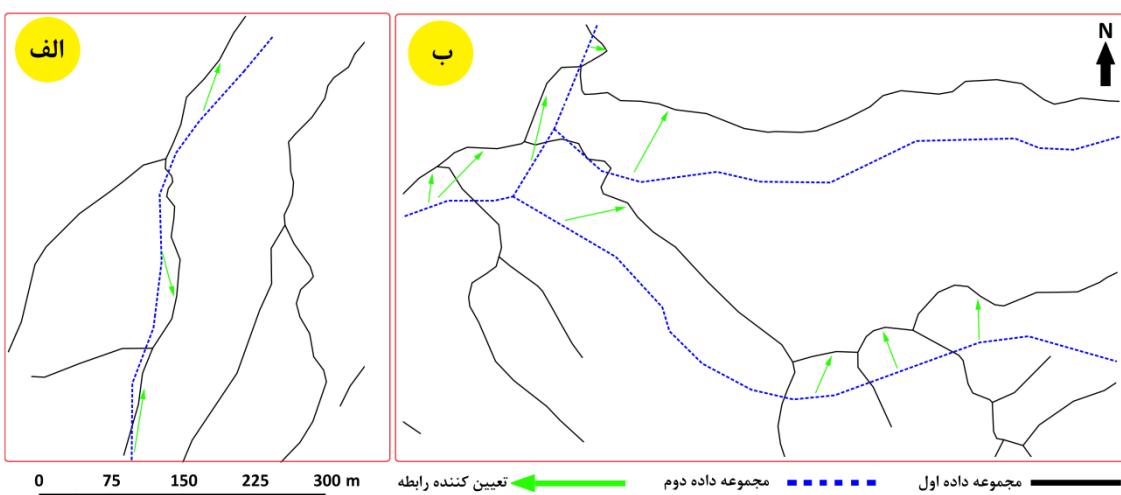
با توجه به رویکرد پیشنهادی در این مرحله نیاز است که برای عارضه در مجموعه دوم، عوارض کاندید در مجموعه اول انتخاب گردد. برای این نیاز است تا فاصله عرض حریم تعیین شود. برای تعیین عرض حریم از دقت دو مجموعه داده و نشریه مشخصات فنی نقشه‌برداری شماره ۹۵ استفاده می‌شود. طبق این نشریه ۹۰٪ نقاط مشخص شده روی نقشه و قابل انطباق با زمین نباید خطای بیش از ۰/۶۵ میلیمتر را داشته باشند و خطای ۰/۱۰ بقیه نیز نباید بیش از ۱/۳ میلیمتر گردد. از این رو طبق رابطه (۱) عرض حریم‌گیری با ۳۳/۱۴ متر شروع و با ۷۵ متر خاتمه می‌یابد.

پس از اینکه فاصله عرض حریم‌گیری تعیین شد، برای هر کدام از عوارض موجود در مجموعه داده دوم عوارض کاندید در مجموعه اول تعیین می‌شود. پس تعیین عوارض متناظر برای تمامی عوارض، ارزیابی صورت می‌گیرد. جدول (۴) نتیجه نهایی بدست آمده از اعمال راهکار پیشنهادی بر روی منطقه مورد مطالعه را نشان می‌دهد. مقدار F-score بدست آمده نشان دهنده این است که راهکار پیشنهادی بخوبی عوارض متناظر را در مجموعه داده‌های با مقیاس و منبع تهییه مختلف شناسایی می‌کند. برای مثال شکل (۵) عوارض متناظر شناسایی شده در دو منطقه از مجموعه داده‌های مورد مطالعه نشان می‌دهد. در شکل (۵-الف) راهکار پیشنهادی رابطه ۱:۳ را بخوبی شناسایی نموده است، همچنین در شکل (۵-ب) نیز روابط ۱:۱، ۱:۲، ۱:۴ و ۱:۱ بخوبی شناسایی شده‌اند.

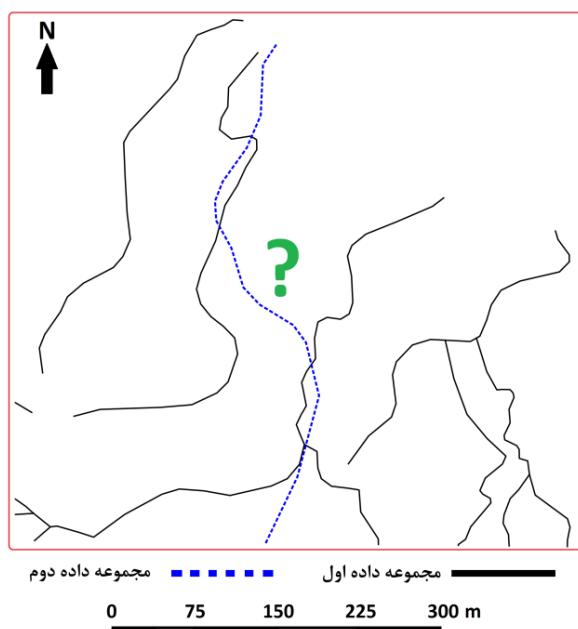
یکی از دلایل عدم شناسایی عوارض متناظر برای تعدادی از عوارض مرجع به ابهام در تخصیص عوارض بر می‌گردد. برای مثال در شکل (۶) الگوریتم پیشنهادی توانایی شناسایی عوارض متناظر برای عارضه مرجع در مجموعه داده دوم را ندارد. زیرا در مجموعه عوارض کاندید، عارضه‌ای با هندسه متناسب با عارضه مرجع وجود ندارد.

جدول ۴: تعداد روابط شناسایی شده

F-score (%)	Recall (%)	Precision (%)	FN	FP	TP
۹۸/۶۱	۹۸/۰۵	۹۹/۱۷	۹۷	۴۱	۴۸۸۵



شکل ۵: نمونه‌ای از شناسایی عوارض متناظر در منطقه مورد مطالعه



شکل ۶: نمونه‌ای از عدم شناسایی عارضه متناظر برای عارضه مرجع

#### ۴- نتیجه‌گیری و پیشنهادات

هدف از این تحقیق تناظریابی رودخانه‌ها در مجموعه داده‌های چند مقیاسی بر اساس معیارهای استخراج شده از هندسه عارض برای شناسایی عارض متناظر در مجموعه داده‌های چند مقیاسی می‌باشد. برای این امر معیارهایی نظیر طول، جهت، مساحت، شکل و ناحیه مشترک حاصل از در نظر گرفتن حریم عارض جهت محاسبه مقدار درجه شباهت در نظر گرفته شد. همچنین در تحقیق حاضر تمامی روابط موجود بین عارض شامل یک به هیچ، هیچ به یک، یک به یک، یک به چند، چند به یک و چند به چند منظور شد. در بخش پایانی نیز از طریق مقادیر Precision و Recall و F – Score تناظریابی مورد ارزیابی قرار گرفت.

نتایج این تحقیق نشان داد که در نظر گرفتن معیارهای مختلف هندسی منجر به شناسایی تعداد قابل توجهی از عارض متناظر در مجموعه داده‌های مختلف می‌گردد. از این رو برای دو مجموعه داده رودخانه مورد استفاده با مقیاس‌های ۱:۲۵۰۰۰ و ۱:۵۰۰۰۰ مقدار ۹۸/۶۱ F – Score درصد محاسبه گردید. نکته قابل توجه در این است که این مقدار بدون در نظر گرفتن سایر معیارها نظیر معیارهای معنایی و توبولوژیک و همچنین در مجموعه داده‌های دارای اختلاف مقیاس بالا و منبع تهیه متفاوت بدست آمده است.

در تحقیقات آتی توصیه می‌شود که با استفاده از راهکار ارائه شده در این تحقیق، کاربردهای نظیر به روز رسانی خودکار داده‌ها و تلفیق داده‌ها انجام گردیده و نتایج حاصل مورد ارزیابی قرار گیرد.

#### مراجع

- [1] A. Samal, S. Seth, and K. Cueto 1, "A feature-based approach to conflation of geospatial sources," International Journal of Geographical Information Science, vol. 18, pp. 459-489, 2004.
- [2] M. Al-Bakri and D. Fairbairn, "Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources," International Journal of Geographical Information Science, vol. 26, pp. 1437-1456, 2012.
- [3] C. Parent and S. Spaccapietra, "Database integration: the key to data interoperability," ed, 2000.
- [4] R. M. Pendyala, Development of GIS-based conflation tools for data integration and matching: Florida Department of Transportation, 2002.

- [5] M. Zhang and L. Meng, "An iterative road-matching approach for the integration of postal data," *Computers, Environment and Urban Systems*, vol. 31, pp. 597-615, 2007.
- [6] G. Gösseln and M. Sester, "Integration of geoscientific data sets and the german digital map using a matching approach," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 35, 2004.
- [7] A. Cecconi, "Integration of cartographic generalization and multi-scale databases for enhanced web mapping," *Universität Zürich*, 2003.
- [8] F. van Wijngaarden, J. van Putten, P. van Oosterom, and H. Uitermark, "Map integration—update propagation in a multi-source environment," in *Proceedings of the 5th ACM international workshop on Advances in geographic information systems*, 1997, pp. 71-76.
- [9] T. Devogele, "A new Merging process for data integration based on the discrete Fréchet distance," in *Advances in spatial data handling*, ed: Springer, 2002, pp. 167-181.
- [10] D. Xiong and J. Sperling, "Semiautomated matching for network database integration," *ISPRS Journal of photogrammetry and Remote sensing*, vol. 59, pp. 35-46, 2004.
- [11] E. Safran, Y. Kanza, Y. Sagiv, and Y. Doytsher, "Efficient integration of road maps," in *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, 2006, pp. 59-66.
- [12] E. Abdolmajidi, J. Will, L. Harrie, and A. Mansourian, "Comparison of Matching Methods of User Generated and Authoritative Geographic Data," in *ICA Workshop on Integration, Filtering and Abstraction of User Generated or Multi-Source Spatial Content*, 2014.
- [13] M. Butenuth, G. v. Gösseln, M. Tiedge, C. Heipke, U. Lipeck, and M. Sester, "Integration of heterogeneous geospatial data in a federated database," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, pp. 328-346, 2007.
- [14] H. Du, S. Anand, N. Alechina, J. Morley, G. Hart, D. Leibovici, et al., "Geospatial information integration for authoritative and crowd sourced road vector data," *Transactions in GIS*, vol. 16, pp. 455-476, 2012.
- [15] Y. Wang, H. Lv, X. Chen, and Q. Du, "A PSO-Neural Network-Based Feature Matching Approach in Data Integration," in *Cartography-Maps Connecting the World*, ed: Springer, 2015, pp. 189-219.
- [16] J. O. Kim, K. Yu, J. Heo, and W. H. Lee, "A new method for matching objects in two different geospatial datasets based on the geographic context," *Computers & Geosciences*, vol. 36, pp. 1115-1122, 2010.
- [17] M. Sester, J. J. Arsanjani, R. Klammer, D. Burghardt, and J.-H. Haunert, "Integrating and generalising volunteered geographic information," in *Abstracting Geographic Information in a Data Rich World*, ed: Springer, 2014, pp. 119-155.
- [18] E. Xavier, F. J. Ariza-López, and M. A. Ureña-Cámarra, "A Survey of Measures and Methods for Matching Geospatial Vector Datasets," *ACM Computing Surveys (CSUR)*, vol. 49, pp. 1-34, 2016.
- [19] J. W. Crampton, M. Graham, A. Poorthuis, T. Shelton, M. Stephens, M. W. Wilson, et al., "Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb," *Cartography and geographic information science*, vol. 40, pp. 130-139, 2013.
- [20] B. Kang, J. Y. Scully, O. Stewart, P. M. Hurvitz, and A. V. Moudon, "Split-Match-Aggregate (SMA) algorithm: integrating sidewalk data with transportation network data in GIS," *International Journal of Geographical Information Science*, vol. 29, pp. 440-453, 2015.
- [21] W. Song, J. M. Keller, T. L. Haithcoat, and C. H. Davis, "Relaxation-Based Point Feature Matching for Vector Map Conflation," *Transactions in GIS*, vol. 15, pp. 43-60, 2011.
- [22] B. Yang, Y. Zhang, and F. Lu, "Geometric-based approach for integrating VGI POIs and road networks," *International Journal of Geographical Information Science*, vol. 28, pp. 126-147, 2014.
- [23] A. J. Saalfeld, "Automated map conflation. Washington, DC: University of Maryland.,," 1993.
- [24] G. Touya, A. Coupé, J. L. Jollec, O. Dorie, and F. Fuchs, "Conflation optimized by least squares to maintain geographic shapes," *ISPRS International Journal of Geo-Information*, vol. 2, pp. 621-644, 2013.

- [25] X. Tong, D. Liang, and Y. Jin, "A linear road object matching method for conflation based on optimization and logistic regression," *International Journal of Geographical Information Science*, vol. 28, pp. 824-846, 2014.
- [26] L. Li and M. F. Goodchild, "An optimisation model for linear feature matching in geographical data conflation," *International Journal of Image and Data Fusion*, vol. 2, pp. 309-328, 2011/12/01 2011.
- [27] C.-C. Chen and C. A. Knoblock, "Conflation of geospatial data," in *Encyclopedia of GIS*, ed: Springer, 2008, pp. 133-140.
- [28] J. Hastings, "Automated conflation of digital gazetteer data," *International Journal of Geographical Information Science*, vol. 22, pp. 1109-1127, 2008.
- [29] M. Zhang, W. Yao, and L. Meng, "Automatic and Accurate Conflation of Different Road-Network Vector Data towards Multi-Modal Navigation," *ISPRS International Journal of Geo-Information*, vol. 5, p. 68, 2016.
- [30] A. Pourabdollah, J. Morley, S. Feldman, and M. Jackson, "Towards an Authoritative OpenStreetMap: Conflating OSM and OS OpenData National Maps' Road Network," *ISPRS International Journal of Geo-Information*, vol. 2, p. 704, 2013.
- [31] T. Bruns and M. Egenhofer, "Similarity of spatial scenes," in *Seventh international symposium on spatial data handling*, 1996, pp. 31-42.
- [32] D. Sheeren, S. Mustière, and J. D. Zucker, "A data-mining approach for assessing consistency between multiple representations in spatial databases," *International Journal of Geographical Information Science*, vol. 23, pp. 961-992, 2009.
- [33] S. Seo and C. G. O'Hara, "Quality assessment of linear data," *International Journal of Geographical Information Science*, vol. 23, pp. 1503-1525, 2009.
- [34] J. Ruiz-Lendínez, F. Ariza-López, and M. Ureña-Cámarra, "Automatic positional accuracy assessment of geospatial databases using line-based methods," *Survey Review*, vol. 45, pp. 332-342, 2013.
- [35] E. Abdolmajidi, A. Mansourian, J. Will, and L. Harrie, "Matching authority and VGI road networks using an extended node-based matching algorithm," *Geo-spatial Information Science*, vol. 18, pp. 65-80, 2015.
- [36] H. Fan, B. Yang, A. Zipf, and A. Rousell, "A polygon-based approach for matching OpenStreetMap road networks with regional transit authority data," *International Journal of Geographical Information Science*, vol. 30, pp. 748-764, 2016/04/02 2016.
- [37] T. Koukoletsos, M. Haklay, and C. Ellul, "Assessing data completeness of VGI through an automated matching procedure for linear data," *Transactions in GIS*, vol. 16, pp. 477-498, 2012.
- [38] I. Ludwig, A. Voss, and M. Krause-Traudes, "A Comparison of the Street Networks of Navteq and OSM in Germany," in *Advancing Geoinformation Science for a Changing World*, ed: Springer, 2011, pp. 65-84.
- [39] H. Fan, A. Zipf, Q. Fu, and P. Neis, "Quality assessment for building footprints data on OpenStreetMap," *International Journal of Geographical Information Science*, vol. 28, pp. 700-719, 2014/04/03 2014.
- [40] L. Cao, N. C. Codella, G. Hua, G. Leiguang, A. I. Natsev, and J. R. Smith, "Techniques for spatial semantic attribute matching for location identification," ed: Google Patents, 2016.
- [41] S. Mustière and T. Devogele, "Matching networks with different levels of detail," *GeoInformatica*, vol. 12, pp. 435-453, 2008.
- [42] Y. Wang, D. Chen, Z. Zhao, F. Ren, and Q. Du, "A Back-Propagation Neural Network-Based Approach for Multi-Represented Feature Matching in Update Propagation," *Transactions in GIS*, vol. 19, pp. 964-993, 2015.
- [43] Y.-h. Wang and F.-y. Wei, "A schema-matching-based approach to propagating updates between heterogeneous spatial databases," in *Geoinformatics 2008 and Joint Conference on GIS and Built environment: Advanced Spatial Data Models and Analyses*, 2009, pp. 714605-714605-10.
- [44] J.-H. Haunert and M. Sester, "Propagating updates between linked datasets of different scales," in *Proceedings of XXII International Cartographic Conference*, 2005, pp. 11-16.

- [45] W. A. Mackaness and G. A. Mackechnie, "Automating the detection and simplification of junctions in road networks," *GeoInformatica*, vol. 3, pp. 185-200, 1999.
- [46] J. McIntosh and M. Yuan, "Assessing similarity of geographic processes and events," *Transactions in GIS*, vol. 9, pp. 223-245, 2005.
- [47] V. Walter and D. Fritsch, "Matching spatial data sets: a statistical approach," *International Journal of Geographical Information Science*, vol. 13, pp. 445-473, 1999.
- [48] M. Zhang and L. Meng, "Delimited stroke oriented algorithm-working principle and implementation for the matching of road networks," *Geographic Information Sciences*, vol. 14, pp. 44-53, 2008.
- [49] A. Moosavi and A. A. Alesheikh, "Developing of Vector Matching Algorithm Considering Topologic Relations," *Proceedings of Map Middle East*, 2008.
- [50] X. Tong, W. Shi, and S. Deng, "A probability-based multi-measure feature matching method in map conflation," *International Journal of Remote Sensing*, vol. 30, pp. 5453-5472, 2009.
- [51] E. Safra, Y. Kanza, Y. Sagiv, and Y. Doytsher, "Ad hoc matching of vectorial road networks," *International Journal of Geographical Information Science*, vol. 27, pp. 114-153, 2013.
- [52] C. Beeri, Y. Doytsher, Y. Kanza, E. Safra, and Y. Sagiv, "Finding corresponding objects when integrating several geo-spatial datasets," in *Proceedings of the 13th annual ACM international workshop on Geographic information systems*, 2005, pp. 87-96.
- [53] A. A. Farahanipooya, R. A. Abbaspour, S. K. Alavipanah, and M. Jafarbeiglu, "Roads matching in a multi-scale spatial database using a least square line," *Journal of Geomatics Science and Technology*, vol. 3, pp. 87-104, 2013.
- [54] I. Abbas, "Base de donne'es vectorielles et erreur cartographique: proble'mes pose's par le contro'le ponctuel; une me'thode alternative fonde'e sur la distance de Hausdorff. Computer Science. Paris: Universite' de Paris VII.," 1994.
- [55] R. C. Veltkamp, "Shape matching: similarity measures and algorithms," in *Shape Modeling and Applications, SMI 2001 International Conference on*, 2001, pp. 188-197.
- [56] M. Zhang, "Methods and implementations of road-network matching," Unpublished PhD Dissertation, Technical University of Munich, 2009.
- [57] L. Li and M. Goodchild, "Automatically and accurately matching objects in geospatial datasets," in *Proceedings of joint international conference on theory, data handling and modelling in geospatial information science*, Hong Kong, 2010, pp. 98-103.
- [58] A.-M. Olteanu Raimond and S. Mustière, "Data Matching – a Matter of Belief," in *Headway in Spatial Data Handling*, A. Ruas and C. Gold, Eds., ed: Springer Berlin Heidelberg, 2008, pp. 501-519.
- [59] X. Tong, D. Liang, G. Xu, and S. Zhang, "Positional accuracy improvement: a comparative study in Shanghai, China," *International Journal of Geographical Information Science*, vol. 25, pp. 1147-1171, 2011.
- [60] Y. Wang, D. Chen, Z. Zhao, F. Ren, and Q. Du, "A Back-Propagation Neural Network-Based Approach for Multi-Represented Feature Matching in Update Propagation," *Transactions in GIS*, vol. 19, pp. 964-993, 2015.
- [61] Y. Wang, Q. Du, F. Ren, and Z. Zhao, "A Propagating Update Method of Multi-Represented Vector Map Data Based on Spatial Objective Similarity and Unified Geographic Entity Code," in *Cartography from Pole to Pole: Selected Contributions to the XXVIth International Conference of the ICA, Dresden 2013*, M. Buchroithner, N. Prechtel, and D. Burghardt, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 139-153.