

تناظریابی عوارض چندضلعی بر اساس معیارهای هندسی در مجموعه داده‌های چندمقیاسی

مژگان چمنی^۱، رحیم علی عباسپور^{۲*}، علیرضا چهرقان^۳

^۱ کارشناس ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - دانشگاه تهران
m.chamani@ut.ac.ir

^۲ استادیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - دانشگاه تهران
abaspour@ut.ac.ir

^۳ دانش آموخته دکتری سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - دانشگاه تهران
achehregan@ut.ac.ir

(تاریخ دریافت خرداد ۱۳۹۶، تاریخ تصویب مهر ۱۳۹۶)

چکیده

شناسایی عوارض با ماهیت یکسان در مجموعه داده‌های مختلف تحت عنوان تناظریابی شناخته می‌شود که می‌تواند کاربردهای متفاوتی نظیر تلفیق، به‌روز رسانی و ارزیابی کیفیت داشته باشد. هدف از این مطالعه ارائه راهکاری جهت شناسایی عوارض چندضلعی متناظر با استفاده از خصوصیات هندسی عوارض می‌باشد. بدین منظور راهکاری متشکل از چهار مرحله پیش‌پردازش، محاسبه شباهت مکانی، استخراج روابط متناظر و ارزیابی نتایج پیشنهاد می‌گردد. در این راهکار یک الگوریتم تناظریابی احتمالی مبنا ارائه شده که در آن از چهار معیار اندازه‌گیری شباهت فاصله، مساحت هم‌پوشانی، جهت و شکل عارضه استفاده شده است. پیاده‌سازی بر روی منطقه شش تهران به عنوان منطقه مورد مطالعه با دو مجموعه داده متفاوت با مقیاس‌های ۱:۲۵۰۰۰ و ۱:۲۰۰۰ انجام شده است. نتایج نشان می‌دهد که الگوریتم ارائه شده با ترکیب چهار معیار معرفی شده به دقت F1-score برابر با ۹۹ درصد بر روی منطقه مطالعاتی رسیده است. همچنین در این تحقیق میزان تأثیر هریک از معیارهای مورد بررسی هم به صورت مجزا و هم در حالت‌های مختلف ترکیب با یکدیگر مورد مطالعه قرار گرفت که نشان می‌دهد لزوماً با افزایش تعداد معیارهای شباهت دقت افزایش نمی‌یابد.

واژگان کلیدی: تناظریابی، عوارض چندضلعی، هندسه عوارض، مجموعه داده‌های چند مقیاسی

۱- مقدمه

با رشد سریع فناوری‌های جمع‌آوری داده‌های مکانی و کاربرد وسیع آن‌ها در حوزه‌های مختلف علوم مکانی، حجم عظیمی از داده‌های مکانی در مقیاس‌ها و دقت‌های مختلف در دسترس می‌باشد. این داده‌های مکانی با پوشش جغرافیایی یکسان به دلایل مختلفی همچون اختلاف مقیاس، تنوع مراکز تولید داده و یا اشتباهات ناگزیر نقشه‌برداری، دارای موقعیت و شکل متفاوتی هستند. به عبارت دیگر یک عارضه یکسان در دنیای واقعی ممکن است در مجموعه داده‌های مختلف با روابط مکانی، معنایی، مقیاس و هندسه متفاوتی نمایش داده شود [۱]. در نتیجه هنگام استفاده از مجموعه داده‌های چند مقیاسی در کاربردهایی همچون تلفیق، غنی‌سازی داده‌های مکانی معتبر، به‌روزرسانی داده‌ها و شناسایی تغییرات، نیاز به راهکارهایی جهت شناسایی عوارض مشابه تحت عنوان مسائل تناظریابی می‌باشد [۲-۵].

تناظریابی داده‌های مکانی به عنوان فرآیند شناسایی عوارض مشابه بین مجموعه داده‌های مختلف تعریف می‌شود که این عوارض مشابه نشان‌دهنده یک موجودیت یکسان در دنیای واقعی هستند. مسائل تناظریابی در مدل داده برداری بر اساس ماهیت داده تناظریابی شده، به سه دسته نقطه‌ای، خطی و چندضلعی تقسیم‌بندی می‌شوند [۱]. همچنین معیارهای اندازه‌گیری شباهت که جهت مقایسه میزان نزدیکی دو عارضه به یکدیگر در نظر گرفته می‌شوند، به سه دسته هندسی، معنایی و روابط مکانی تقسیم‌بندی می‌شوند [۱، ۶، ۷]. البته در برخی از مسائل امکان دسترسی به معیارهای معنایی مانند نام عارضه وجود ندارد و یا ناقص می‌باشد، در نتیجه استفاده از معیارهای توصیفی (معنایی) کاربرد کمتری دارد.

هدف از این تحقیق ارائه رویکردی جهت تناظریابی عوارض چندضلعی در مجموعه داده‌های چندمقیاسی می‌باشد که در آن تنها از خصوصیات هندسی عوارض استفاده شده است. دلیل استفاده از خصوصیات هندسی این است که امکان دسترسی به آن‌ها به راحتی فراهم می‌باشد و نیاز به اطلاعات اضافی که برخی از مجموعه داده‌ها فاقد آن هستند، نمی‌باشد. در این رویکرد با ارائه الگوریتم تناظریابی احتمالی مبنا از چهار معیار هندسی اندازه‌گیری شباهت شامل فاصله، مساحت هم‌پوشانی،

جهت و شکل عارضه استفاده شده است. وزن‌دهی به هر معیار که میزان اهمیت آن معیار را در شناسایی عوارض مشابه مشخص می‌کند، به کمک نظر کارشناس انجام شده است. همچنین در بخش ارزیابی نتایج، تاثیر و حساسیت هر معیار اندازه‌گیری شباهت در دقت الگوریتم به صورت مجزا و در ترکیب با سایر معیارها مطالعه شده است که نقطه تمایز این پژوهش را با مطالعات پیشین که در ادامه به شرح کامل آن‌ها پرداخته شده، نشان می‌دهد.

ادامه این تحقیق به شرح ذیل است: در بخش دوم پیشینه تحقیق ارائه شده است. در بخش سوم روش پیشنهادی به تشریح بیان می‌شود. بخش چهارم شامل پیاده‌سازی روش پیشنهادی بر روی منطقه مورد مطالعه می‌باشد و نتایج این پیاده‌سازی را نشان می‌دهد. در انتها نیز در بخش پنجم نتیجه‌گیری و پیشنهادهای ارائه می‌گردد.

۲- پیشینه تحقیق

تناظریابی عوارض برداری برای اولین بار در سال ۱۹۸۸ به هنگام تلفیق نقشه در اداره سرشماری آمریکا بررسی شده است تا یک نقشه یکپارچه با صحت بالا حاصل شود [۸]. پس از آن، مسائل تناظریابی با تمرکز بیشتر بر روی دو نوع عارضه نقطه‌ای و خطی گسترش یافت به طوریکه امروزه پیشرفت وسیعی در مطالعات تناظریابی عوارض خطی حاصل شده است که می‌توان در آخرین مطالعات به تحقیقات Kim و همکاران [۹]، Tong و همکاران [۱۰]، Chehrehghan و Abbaspour [۱۱]، Olteanu-Raimond و همکاران [۱۲] و Chehrehghan و Abbaspour [۱۳] اشاره نمود. تناظریابی عوارض نقطه‌ای بر خلاف دو نوع عارضه خطی و چندضلعی ساده‌تر و اغلب از فاصله موقعیت‌های عوارض نقطه‌ای به عنوان معیار شباهت آن‌ها استفاده شده است؛ اما در مطالعات کمتری به تناظریابی عوارض چندضلعی که دارای پیچیدگی‌های به مراتب بیشتری نسبت به دو نوع دیگر عوارض هستند، پرداخته شده است. در ادامه جمع‌بندی مطالعات تناظریابی عوارض چندضلعی بیان شده است.

مطالعات تناظریابی عوارض چندضلعی را می‌توان از سه دیدگاه نوع الگوریتم تناظریابی، معیارهای اندازه‌گیری شباهت مورد استفاده و یا نوع روابط متناظر استخراج شده بررسی نمود. Xavier و همکاران [۱۴] مرور جامعی بر

انواع این دسته‌بندی‌ها برای هر سه نوع عارضه نقطه‌ای، خطی و چندضلعی انجام داده‌اند. در این مقاله تقسیم‌بندی مطالعات تناظریابی عوارض چندضلعی از منظر نوع الگوریتم تناظریابی مطرح می‌شود. الگوریتم‌های تناظریابی عوارض چندضلعی به سه دسته بر مبنای کاهش بعد، ماتریس مبنا و احتمالی مبنا تقسیم‌بندی می‌شوند [۱، ۶]. در الگوریتم‌های تناظریابی بر مبنای کاهش بعد، عارضه چندضلعی تبدیل به عارضه خطی یا نقطه‌ای می‌شود و سپس از الگوریتم‌های عوارض خطی یا نقطه‌ای جهت حل مسئله استفاده می‌شود. در الگوریتم‌های ماتریس مبنا، ویژگی‌های هندسی و یا روابط مکانی بین عوارض چندضلعی در قالب یک ماتریس مدل می‌شود و سپس این ماتریس به صورت تکراری بروز می‌شود. در الگوریتم‌های احتمالی مبنا، احتمال انطباق دو عارضه چندضلعی بر اساس میزان شباهت یا فاصله بین معیارهای مختلف تناظریابی محاسبه می‌شود و در رابطه احتمال انطباق، برای هر معیار وزنی در نظر گرفته می‌شود.

تناظریابی بر مبنای کاهش بعد در مطالعات Kiehl و همکاران [۱۵] در سال ۲۰۰۹ جهت تناظریابی مجموعه داده‌های چندمقیاسی رودخانه‌ها، استفاده شده است. در این روش ابتدا خطوط مرکزی چندضلعی‌های رودخانه‌ها به کمک الگوریتم اسکلت‌بندی استخراج می‌شوند و سپس تناظریابی گره‌ها و یال‌های شبکه خطوط استخراج شده پیاده‌سازی می‌شود. همچنین Safra و همکاران [۱۶] در سال ۲۰۱۰ روشی جهت تناظریابی مکان‌های مختلف ارائه دادند که در این روش، مرکز جرم عوارض چندضلعی به جای کل عارضه در نظر گرفته شده است.

تناظریابی ماتریس مبنا برای عوارض چندضلعی با مطالعات Samal و همکاران [۴] در سال ۲۰۰۴ جهت تلفیق مجموعه داده‌های مکانی آغاز شد. در این روش ماتریس مشابهت اولیه با محاسبه شباهت‌های مستقل زمینه مقداردهی شده است. سپس این ماتریس در یک روند تکراری به کمک معیار شباهت زمینه محاسبه شده توسط نمودار نزدیکی^۱، به‌روز می‌شود. Kim و همکاران [۱۷] در سال ۲۰۱۰ به کمک نمودار ورونوی و مثلث‌بندی، محاسبات شباهت زمینه ارائه شده توسط Samal و همکاران را بهبود بخشیدند.

الگوریتم‌های تناظریابی احتمالی مبنا که اکثر مطالعات انجام شده در این گروه قرار می‌گیرند، بر اساس نوع وزن‌دهی به معیارهای شباهت به دو روش نظر کارشناس و آموزش هوشمند به کمک الگوریتم‌های هوش مصنوعی انجام می‌شوند. در استفاده از الگوریتم‌های احتمالی مبنا به کمک نظر کارشناس، می‌توان به مطالعات Wu و Fu [۱۸]، Tong و همکاران [۱۹]، Wang و همکاران [۵] و Fan و همکاران [۳] اشاره نمود. Wu و Fu [۱۸] در سال ۲۰۰۸ از معیارهای شباهت مساحت عارضه، مساحت هم‌پوشانی و فاصله Barycentre جهت تناظریابی عوارض خطی و چندضلعی استفاده کردند. Tong و همکاران [۱۹] در سال ۲۰۰۹ یک الگوریتم احتمالی مبنا جهت تناظریابی عوارض نقطه‌ای، خطی و چندضلعی ارائه دادند که در این روش، احتمال انطباق میانگین وزن‌داری از معیارهای شباهت موقعیت، شکل، جهت و روابط توپولوژی می‌باشد. Wang و همکاران [۵] در سال ۲۰۱۴ به منظور به‌روزرسانی انتشار یافته مجموعه داده‌های چندنمایشی از الگوریتم تناظریابی بر مبنای شباهت‌های هندسی و توصیفی دو عارضه چندضلعی استفاده کردند. Fan و همکاران [۳] در سال ۲۰۱۴ جهت ارزیابی کیفی داده‌های OSM از معیارهای مساحت هم‌پوشانی و تابع چرخش برای محاسبه مشابهت عوارض استفاده نمودند.

روش‌های تناظریابی احتمالی مبنا به کمک استفاده از الگوریتم‌های هوش مصنوعی در مطالعه Wang و همکاران [۱] در سال ۲۰۱۵ بررسی شده است. در این مطالعه، تناظریابی به کمک شبکه عصبی پس انتشار^۲ جهت انتشار به‌روزرسانی مجموعه داده‌های چندنمایشی مطالعه شده است. همچنین Wang و همکاران [۶] در سال ۲۰۱۵ الگوریتم بهینه‌سازی تناظریابی عوارض چندضلعی جهت ادغام داده‌ها ارائه دادند که در آن وزن‌دهی بر اساس شبکه عصبی بهینه‌سازی ازدحام ذرات^۳ پیاده‌سازی شده است.

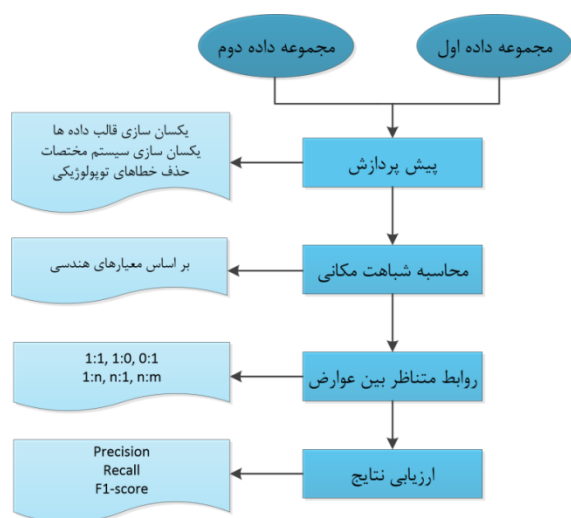
بررسی پیشینه تحقیق نشان می‌دهد که استفاده از الگوریتم‌های بر مبنای کاهش بعد کارایی بالایی ندارد چرا که از خصوصیات خاص عوارض چندضلعی همانند مساحت عارضه استفاده نمی‌شود. همچنین الگوریتم‌های ماتریس مبنا حساسیت بالایی به انتخاب وزن‌دهی به

^۲ Back Propagation Neural Network

^۳ Particle Swarm Optimization Neural Network

^۱ Proximity graph

کیفیت Precision، Recall و F1-score ارزیابی می‌گردد که شرح کامل پیاده‌سازی، استخراج روابط متناظر و کارایی الگوریتم در بخش ۴ ذکر شده است.



شکل ۱- مراحل پیاده سازی روش پیشنهادی

۳-۱- الگوریتم تناظریابی احتمالی مبنا

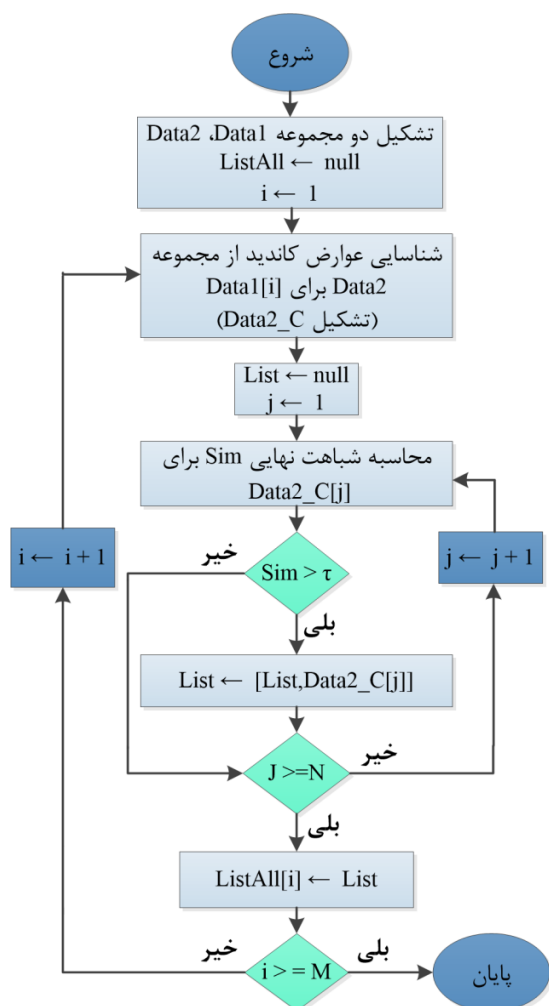
الگوریتم‌های تناظریابی عوارض چندضلعی به سه دسته بر مبنای کاهش بعد، ماتریس مبنا و احتمالی مبنا تقسیم‌بندی می‌شوند که اکثر مطالعات انجام شده در گروه الگوریتم‌های احتمالی مبنا می‌باشد [۶، ۱]. در الگوریتم‌های احتمالی مبنا، احتمال انطباق برای تمام عوارض کاندید به صورت ترکیب وزن‌داری از معیارهای شباهت مکانی محاسبه می‌گردد. سپس اگر این احتمال انطباق بیش از حدآستانه‌ای (τ) محاسبه شود، آن دو عارضه متناظر می‌باشند. فلوچارت الگوریتم ارائه شده در شکل ۲ نمایش داده شده است. همچنین شبه کد این الگوریتم با جزئیات بیشتری در شکل ۳ نشان داده شده که در ادامه توضیحات تکمیلی ارائه می‌گردد.

مطابق با شکل ۲، در ابتدای الگوریتم تناظریابی نیاز به شناسایی عوارض کاندید مجموعه داده دوم برای هر عارضه از مجموعه داده اول می‌باشد تا شباهت مکانی تنها برای عوارض کاندید محاسبه شود و زمان محاسباتی اجرای الگوریتم کاهش یابد. در اینجا فرض شده است که تعداد عوارض مجموعه اول شامل M عارضه و تعداد عوارض کاندید از مجموعه دوم برای هر عارضه از مجموعه اول، N عضو می‌باشد. جهت محاسبه عوارض کاندید از محدوده بافر استفاده می‌شود که فاصله آن از طریق دقت

معیارها و حدآستانه‌ها دارد و انتخاب نامناسب این ضرایب به راحتی منجر به واگرایی الگوریتم می‌شود. اکثر مطالعات انجام شده در ارتباط با الگوریتم‌های احتمالی مبنا می‌باشد که وزن‌دهی به کمک نظر کارشناس انجام شده است و رویکرد این مقاله نیز این نوع از الگوریتم‌های تناظریابی می‌باشد. به طور کلی در این مقاله سعی شده است هم‌زمان از چهار معیار مهم اندازه‌گیری شباهت هندسی که شرح کامل هر کدام در بخش ۳ ارائه شده، در بستر یک الگوریتم تناظریابی احتمالی مبنا استفاده شود. نقطه تمایز این پژوهش با مطالعات پیشین در ترکیب هم‌زمان این چهار معیار هندسی می‌باشد که در مطالعات پیشین به صورت مجزا استفاده شده‌اند. همچنین تأثیر هر معیار اندازه‌گیری شباهت هندسی ارائه شده به تنهایی و در ترکیب با سایر معیارها بررسی شده است تا میزان تأثیرگذاری هر گروه از معیار(ها) در دقت نهایی الگوریتم مطالعه گردد و میزان حساسیت آن (ها) بررسی شود؛ در حالی که تا کنون در مطالعات پیشین این موضوع بررسی نشده است.

۳-۲ روش پیشنهادی

همان‌طور که بیان شد، هدف از این تحقیق حل مسئله تناظریابی عوارض چندضلعی به کمک معیارهای هندسی می‌باشد. جهت رسیدن به این هدف، فرآیندی مطابق با شکل ۱ پیشنهاد می‌گردد که شامل ۴ مرحله اصلی می‌باشد. طبق این ساختار در ابتدا دو مجموعه داده اول و دوم با دو مقیاس متفاوت به عنوان ورودی وارد فرآیند تناظریابی می‌شوند. سپس به دلیل تنوع منابع تولید داده نیاز به مرحله پیش‌پردازش می‌باشد. این مرحله شامل یکسان‌سازی قالب داده‌ها، تبدیل به سیستم مختصات یکسان و حذف خطاهای توپولوژیکی می‌باشد. پس از مرحله پیش‌پردازش، یک الگوریتم تناظریابی احتمالی مبنا ارائه شده که در آن، درجه شباهت مکانی عوارض چندضلعی بر اساس معیارهای هندسی محاسبه می‌گردد (بخش ۳-۱). سپس در مرحله استخراج روابط متناظر، هر ۶ نوع رابطه متناظر یک به یک، یک به صفر، صفر به یک، یک به چند، چند به یک و چند به چند شناسایی می‌شود (بخش ۳-۲). در انتها با پیاده‌سازی این فرآیند بر روی منطقه مورد مطالعه، نتایج الگوریتم به کمک معیارهای



شکل ۲- فلوچارت الگوریتم تناظریابی احتمالی مبنای ارائه شده

مکانی مجموعه داده بدست می‌آید. بدین ترتیب برای هر عارضه چندضلعی از مجموعه داده اول، بافر عارضه با فاصله معین در نظر گرفته می‌شود و عوارض چندضلعی از مجموعه داده دوم که حداقل یکی از رئوس آن‌ها درون این محدوده بافر قرار گیرد، به عنوان عوارض کاندید انتخاب شده و شباهت مکانی تنها برای آن عوارض با عارضه مذکور از مجموعه داده اول محاسبه می‌شود.

پس از محاسبه عوارض کاندید، شباهت مکانی میان هر عارضه کاندید با عارضه مجموعه داده اول محاسبه می‌شود. معیارهای شباهت مکانی در الگوریتم‌های تناظریابی، میزان نزدیکی دو عارضه را از نظر مکانی بررسی می‌کنند که در حالت کلی شباهت دو عارضه چندضلعی به کمک معیارهای هندسی، توصیفی و روابط مکانی عارضه با عوارض مجاور قابل بررسی می‌باشد [۱، ۶، ۷]. در این مطالعه تنها از معیارهای هندسی استفاده می‌شود، زیرا در تمامی مجموعه داده‌ها امکان دسترسی به این معیارها وجود دارد و نیاز به اطلاعات اضافه به جز هندسه عوارض نیست. در این مقاله از معیارهای فاصله، مساحت هم‌پوشانی، جهت و شکل به عنوان معیارهای هندسی محاسبه شباهت عوارض چندضلعی استفاده شده است که ترکیب وزن دار آن‌ها در نهایت توسط متغیر Sim محاسبه می‌شود.

Algorithm 1: Proposed matching algorithm

```

01: 'Data1[]: polygons of dataset1 with (x,y)
02: 'Data2[]: polygons of dataset2 with (x,y)
03: 'listAll[]: list of all dataset2 matched polygons for each dataset1 polygon
04: function matching (Data1[],Data2[])
05:   for i ∈ Data1[] do
06:     list=null
07:     Data2_C[] = Candidate(Data2[])
08:     for j ∈ Data2_C[] do
09:        $\sigma_{Dis}$  = Distance (Data[i].Data2_C[j])
10:        $\sigma_{Area}$  = OverlappingArea (Data1[i].Data2_C[j])
11:        $\sigma_{Dir}$  = Direction (Data1[i].Data2_C[j])
12:        $\sigma_{Shape}$  = Shape (Data1[i].Data2_C[j])
13:        $Sim = \frac{W_{Dis} \times \sigma_{Dis} + W_{Area} \times \sigma_{Area} + W_{Dir} \times \sigma_{Dir} + W_{Shape} \times \sigma_{Shape}}{W_{Dis} + W_{Area} + W_{Dir} + W_{Shape}}$ 
14:       if Sim>τ then
15:         list[] = [list,j]
16:       end if
17:     end for
18:     listAll[i] = list[]
19:   end for
20:   return listAll
21: end function

```

شکل ۳- شبه کد الگوریتم تناظریابی احتمالی مبنای ارائه شده

همان‌طور که در رابطه (۱) ملاحظه می‌شود، شباهت نهایی دو عارضه A و B $(S(A, B))$ ، برابر با ترکیب وزن‌داری از معیارهای شباهت مکانی $(S_i(A, B))$ می‌باشد که در این‌جا ۴ معیار فاصله، مساحت هم‌پوشانی، جهت و شکل در نظر گرفته شده‌اند ($n=4$). همچنین وزن هر معیار (W_i) به کمک نظر کارشناس و بر اساس نوع منطقه مورد مطالعه انجام می‌پذیرد. پس از محاسبه شباهت مکانی برای تمام عوارض کاندید، جفت عارضه‌ای به عنوان دو عارضه متناظر معرفی می‌گردد که شباهت مکانی آن دو بزرگتر از حدآستانه‌ای (τ) باشد که انتخاب این حد آستانه نیز وابسته به منطقه مورد مطالعه است.

$$S(A, B) = \frac{\sum_{i=1}^n W_i \cdot S_i(A, B)}{\sum_{i=1}^n W_i} \quad (1)$$

در انتها نیز روند ذکر شده برای تمامی چندضلعی‌های مجموعه داده اول اجرا می‌گردد تا خروجی listAll که شامل عوارض متناظر از مجموعه داده دوم برای هر عارضه از مجموعه داده اول می‌باشد، حاصل شود. در ادامه هر یک از معیارهای اندازه‌گیری شباهت به تفصیل تشریح شده‌اند.

۳-۱-۱- فاصله

فاصله مکانی یکی از معیارهای اندازه‌گیری موقعیت نسبی بین اشیاء مکانی است که درجه‌ای از شباهت آن‌ها را نشان می‌دهد. معروف‌ترین فاصله مورد استفاده در GIS، فاصله اقلیدسی است که فاصله بین دو نقطه منفرد را اندازه‌گیری می‌کند. جهت محاسبه فاصله اقلیدسی بین دو عارضه چندضلعی، می‌توان فاصله بین مراکز ثقل دو شی را محاسبه نمود که در اکثر مطالعات تناظریابی عوارض چندضلعی از این روش محاسبه شده است [۴-۶، ۲۰]. در نتیجه با وجود مراکز ثقل دو چندضلعی A و B که با مختصات $p_1 = (x_1, y_1)$ و $p_2 = (x_2, y_2)$ نمایش داده می‌شوند، درجه شباهت بر اساس فاصله اقلیدسی (σ_{Dis}) برای این دو چندضلعی طبق رابطه (۲) تعریف می‌گردد.

$$\sigma_{Dis}(p_1, p_2) = 1 - \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{U} \quad (2)$$

که در آن از U به عنوان فاکتور نرمال‌سازی رابطه استفاده می‌شود تا شباهت بر اساس فاصله اقلیدسی در

بازه (0,1) قرار گیرد. به عبارتی دیگر، U برابر با حداکثر فاصله ممکن است که دو چندضلعی در یک مجموعه داده می‌توانند داشته باشند. در نتیجه این مقدار بر اساس مجموعه داده مورد مطالعه محاسبه می‌گردد [۴].

۳-۱-۲- مساحت هم‌پوشانی

مساحت هم‌پوشانی دو چندضلعی یکی از معیارهایی است که در اکثر مطالعات تناظریابی در مرحله شناسایی شباهت دو عارضه یا در مرحله انتخاب عوارض کاندید استفاده شده است [۱۸]. همچنین ممکن است به صورت ترکیبی نیز از این معیار استفاده شود. به عنوان نمونه از نسبت مساحت هم‌پوشانی بر روی کمترین مساحت دو عارضه و یا بر روی میانگین مساحت دو عارضه استفاده می‌شود [۳، ۲۱]. در این مقاله درجه شباهت بر اساس مساحت هم‌پوشانی طبق رابطه (۳) تعریف شده است.

$$\sigma_{Area}(A, B) = \frac{Area(A \cap B)}{\min(Area(A), Area(B))} \quad (3)$$

در این رابطه، $Area(A \cap B)$ مساحت ناحیه هم‌پوشانی دو چندضلعی A و B می‌باشد و $\min(Area(A), Area(B))$ برابر با حداقل مقدار بین مساحت دو چندضلعی A و B است.

۳-۱-۳- جهت

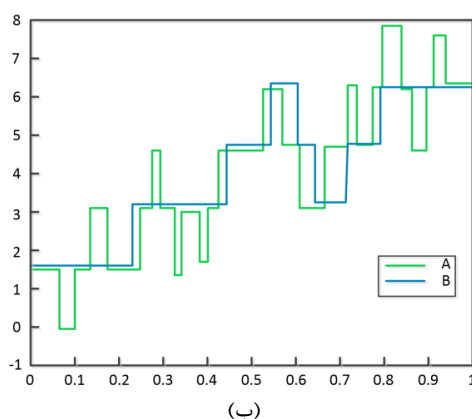
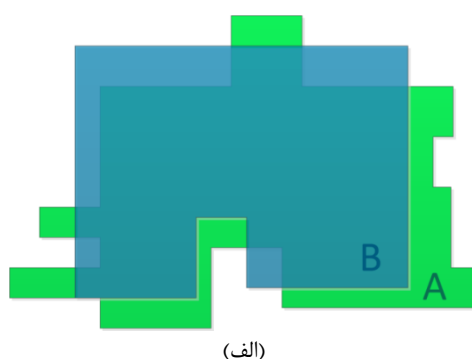
جهت یک چندضلعی جزء معیارهایی است که در رابطه با ویژگی‌های هندسی یک عارضه تعریف می‌شود. از دیگر ویژگی‌های هندسی یک عارضه می‌توان محیط، مساحت و یا طول قطر کوچکترین مستطیل محیطی^۱ (MBR) شکل را نام برد که همگی در کنار جهت شکل، مبنای مقایسه میزان شباهت دو عارضه چندضلعی می‌باشند [۱۴]. جهت یک عارضه خطی می‌تواند معادل زاویه‌ای باشد که نقطه شروع و انتهای خط با محور افقی مختصات می‌سازد [۱۹]. حال مشابه با عارضه خطی، جهت یک عارضه چندضلعی طبق شکل برابر با زاویه‌ای تعریف شده است که قطر MBR عارضه (با نقطه شروع $(\min(x), \min(y))$ و نقطه انتهایی $(\max(x), \max(y))$) با محور افقی می‌سازد [۱]. از آنجایی که این زاویه در

^۱ Minimum Bounding Rectangle

در این رابطه T_A و T_B به ترتیب برابر با تابع چرخش چندضلعی A و B می‌باشند. اپراتور $\|p\|$ نیز نشان‌گر نرم L_p می‌باشد که در این مقاله به ازای $p = 2$ محاسبه شده است. حال از این فاصله بین دو تابع چرخش می‌توان به عنوان معیار شباهت شکل استفاده نمود. از آنجایی که تمام معیارها به بازه $(0,1)$ نرمال شده‌اند، معیار شباهت شکل نیز طبق رابطه (۶) نرمال می‌شود که در آن، U برابر با حداکثر اختلاف در دو تابع چرخش است که با وجود آن، دو شکل از لحاظ شکل مشابه هستند و بر اساس منطقه محاسبه می‌شود.

$$\sigma_{shape}(A, B) = 1 - \frac{d_{TF}(A, B)}{U} \quad (۶)$$

به عنوان نمونه شکل تابع چرخش دو چندضلعی A و B را نشان می‌دهد. این تابع برای هر کدام از رئوس چندضلعی میزان چرخش یال متصل به آن را نسبت به محور افقی در نظر می‌گیرد. در انتها مساحت محصور بین دو نمودار تابع چرخش مطابق با شکل (ب) برابر با اختلاف دو شکل عارضه می‌باشد.

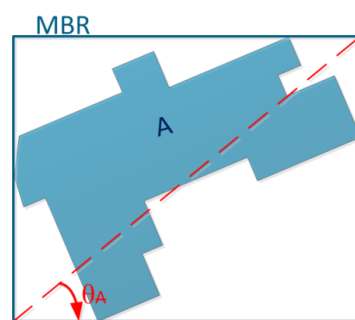


شکل ۵- تابع چرخش دو چندضلعی A و B با نمایش شکل (الف) و نمودار آن (ب)

محدوده $(0, \frac{\pi}{2})$ قرار دارد، درجه شباهت دو عارضه بر اساس رابطه (۴) تعریف می‌شود.

$$\sigma_{Dir}(A, B) = 1 - \frac{|\theta_B - \theta_A|}{\frac{\pi}{2}} \quad (۴)$$

در این رابطه، θ_A و θ_B به ترتیب برابر با جهت چندضلعی A و جهت چندضلعی B می‌باشد.



شکل ۴- نمایش MBR و جهت چندضلعی A

۳-۱-۴- شکل

توابع مختلفی جهت بررسی شکل عوارض با دقت و جزئیات بالا وجود دارد که در مسائل تناظریابی از آن‌ها استفاده شده است. از جمله معیارهای مورد استفاده جهت تشابه شکل، می‌توان به استفاده از تابع چرخش [۳]، میزان مستطیلی بودن^۱ [۳]، کشیدگی شکل [۲۲]، نسبت مساحت به محیط عارضه [۶] و تابع فوریه [۲۳] در مسائل تناظریابی اشاره نمود. در این مقاله از معیار تابع چرخش که یکی از پرکاربردترین توصیفگرهای مرتبط با شکل عوارض می‌باشد، استفاده شده است.

تابع چرخش توسط Arkin و همکاران [۲۴] ارائه شده است که در آن، زاویه پادساعتگرد خط مماس بر شکل به عنوان تابعی از طول تجمعی نرمال شده اندازه‌گیری می‌شود. به عبارتی دیگر، زاویه تجمعی خط مماس بر شکل با چرخش اضلاع شکل به سمت چپ افزایش می‌یابد و با چرخش به سمت راست کاهش می‌یابد. فاصله بین توابع چرخش دو چندضلعی A و B جهت اندازه‌گیری میزان شباهت آن‌ها طبق رابطه (۵) محاسبه می‌شود.

$$\begin{aligned} d_{TF}(A, B) &= \|T_A - T_B\|_p \\ &= \left(\int_0^1 |T_A(l) - T_B(l)|^p dl \right)^{\frac{1}{p}} \quad (۵) \end{aligned}$$

^۱ Rectangularity

۳-۲- روابط متناظر

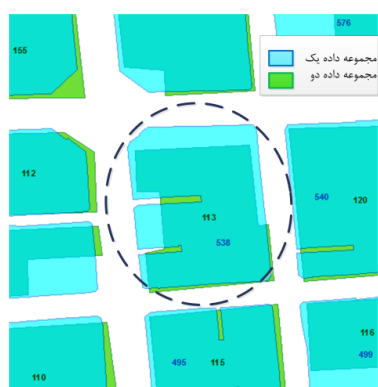
به طور کلی در مسائل تناظریاتی شش نوع رابطه یک به یک (1:1)، یک به صفر (1:0)، صفر به یک (0:1)، یک به چند (1:N)، چند به یک (N:1) و چند به چند (N:M) قابل استخراج می‌باشد. وابسته به نوع روابط متناظر قابل استخراج، تناظریاتی را می‌توان در سه حالت کلی یک به یک، یک به چند و چند به چند به شرح ذیل تقسیم‌بندی نمود [۱۴] که در این مقاله، هر سه دسته روابط کشف و شناسایی شده است.

۱- تناظریاتی یک به یک^۱: تناظریاتی یک به یک یکی از ساده‌ترین دسته‌ها می‌باشد که از اولین مطالعات وجود داشته است. در این گروه هر سه رابطه 1:1، 1:0 و 0:1 قابل شناسایی می‌باشد. همان‌طور که در شکل (الف) ملاحظه می‌شود، در حالت 1:1، یک عارضه از مجموعه داده اول تنها با یک عارضه از مجموعه داده دوم متناظر است. مطابق با شکل (ب)، در حالت 1:0 (و یا برعکس آن حالت 0:1)، برای یک عارضه از یک مجموعه داده هیچ عارضه متناظری از مجموعه داده دوم پیدا نشده است.

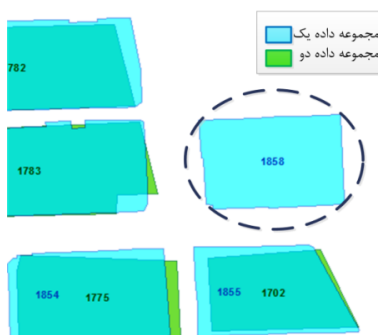
۲- تناظریاتی یک به چند^۲: این تناظریاتی منجر به شناسایی روابط 1:N و N:1 می‌شود که در مجموعه داده‌های چندمقیاسی با سطح جزئیات متفاوت بسیار کاربرد دارد. در این دو حالت، یک عارضه از یک مجموعه داده با چند عارضه از مجموعه داده دیگر متناظر است. شکل (ج) رابطه 1:N را برای دو مجموعه داده نشان می‌دهد که در آن یک عارضه از مجموعه داده یک با چهار عارضه از مجموعه داده دو متناظر است. رابطه N:1 دقیقاً برعکس رابطه 1:N می‌باشد و زمانی است که چند عارضه از مجموعه داده یک متناظر با یک عارضه از مجموعه داده دو هستند.

۳- تناظریاتی چند به چند^۳: در این تناظریاتی رابطه N:M قابل شناسایی می‌باشد که جزو پیچیده‌ترین روابط است و تمام شرایط ممکن را بررسی می‌کند. در این رابطه N عارضه از یک مجموعه داده با M عارضه از مجموعه داده دوم متناظر است. به عنوان

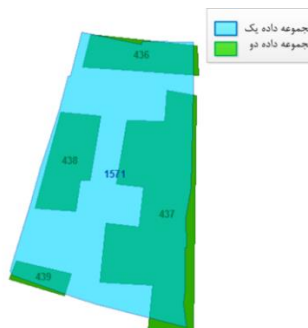
نمونه شکل (د) نوعی از رابطه N:M را نمایش می‌دهد که در آن، دو عارضه از مجموعه داده یک با سه عارضه از مجموعه داده دو متناظر هستند.



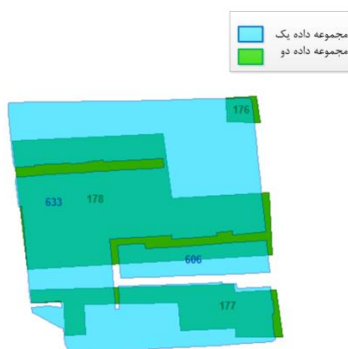
(الف)



(ب)



(ج)



(د)

شکل ۶- انواع روابط متناظر؛ (الف) رابطه 1:1؛ (ب) رابطه 1:0؛ (ج) رابطه 1:N؛ (د) رابطه N:M

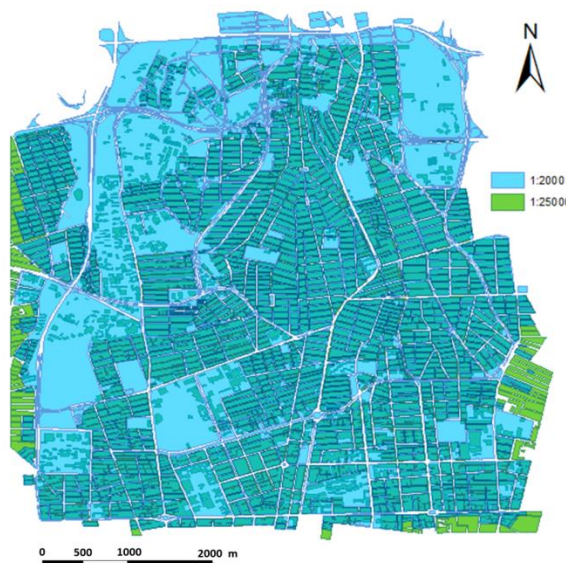
^۱ One-to-one matching

^۲ One-to-many matching

^۳ Many-to-many matching

۴- پیاده‌سازی

جهت ارزیابی دقت الگوریتم ارائه شده و همچنین بررسی ترکیب مختلفی از معیارهای هندسی در دقت نهایی الگوریتم، دو مجموعه داده با دو مقیاس متفاوت در نظر گرفته می‌شود که تناظریابی دستی آن‌ها انجام شده است. در این مقاله از دو مجموعه داده عوارض چندضلعی مربوط به منطقه شش تهران به عنوان منطقه مورد مطالعه استفاده شده است که مجموعه داده اول با مقیاس ۱:۲۰۰۰ در سال ۲۰۰۸ و مجموعه داده دوم با مقیاس ۱:۲۵۰۰۰ در سال ۲۰۰۱ می‌باشد. شکل عوارض چندضلعی برای این دو مجموعه داده را در منطقه شش تهران نشان می‌دهد. پس از انجام مرحله پیش‌پردازش، تعداد عوارض مجموعه داده اول و دوم به ترتیب برابر ۱۸۷۵ و ۱۹۵۸ گردید. علاوه بر آن، مجموع مساحت مجموعه داده اول و دوم به ترتیب برابر با ۱۲/۰۳۶ و ۱۸/۹۵۶ کیلومتر مربع می‌باشد.



شکل ۷- مجموعه داده‌های منطقه شش تهران با مقیاس و منابع متفاوت

پس از انجام تناظریابی دستی دو مجموعه داده منطقه شش تهران که در مرحله ارزیابی نتایج الگوریتم ارائه شده ضروری می‌باشد، تعداد صحیح هر شش نوع رابطه تناظر برای این دو مجموعه داده مطابق با جدول ۱ استخراج گردید. همان‌طور که ملاحظه می‌شود، بیشترین تعداد روابط تناظر از نوع 1:1 می‌باشد و پس از آن روابط N:1 بیشترین تعداد را شامل می‌شود. تعداد نسبتاً بالای روابط 0:1 در مقایسه با 1:0 نشان می‌دهد که برای بسیاری از

عوارض مجموعه داده دوم، عارضه معادلی از مجموعه داده اول وجود ندارد که خود بیان‌گر کامل‌تر بودن مجموعه داده دوم به لحاظ سطح پوششی منطقه می‌باشد. همچنین تعداد روابط نسبتاً بالای N:M نشان‌دهنده وجود سطحی از پیچیدگی در منطقه مورد مطالعه است.

جدول ۱- تعداد صحیح روابط تناظر برای دو مجموعه داده منطقه

روابط تناظر					
1:1	1:0	0:1	1:N	N:1	N:M
۸۳۱	۷۹	۴۸۰	۳۳۸	۶۹۸	۲۷۸

جهت انجام تناظریابی، شعاع بافر جستجو عوارض کاندید، وزن‌های نهایی هر یک از معیارهای شباهت و همچنین حدآستانه شناسایی جفت‌های متناظر (T) مجهول می‌باشند که به کمک نظر کارشناس و یا با توجه به نوع و دقت منطقه مورد مطالعه انتخاب می‌گردند. بر این اساس شعاع بافر جستجو به کمک دقت دو مجموعه داده و مطابق بانشریه مشخصات فنی نقشه‌برداری شماره ۹۵ محاسبه گردید. رابطه (۷) محاسبه شعاع بافر جستجو را نشان می‌دهد [۲۵].

$$W_{buffer} = \sqrt{\sigma_a^2 + \sigma_b^2} \quad (۷)$$

در این رابطه پارامترهای σ_a و σ_b به ترتیب برابر با دقت مجموعه داده اول و دوم می‌باشند. مطابق با بازه اطمینان ۹۰٪ نشریه مشخصات فنی نقشه‌برداری، این پارامترها برای دو مجموعه داده با مقیاس‌های ۱:۲۰۰۰ و ۱:۲۵۰۰۰ محاسبه گردید و در نهایت شعاع بافر جستجو برابر ۱۶/۳۳ متر انتخاب گردید.

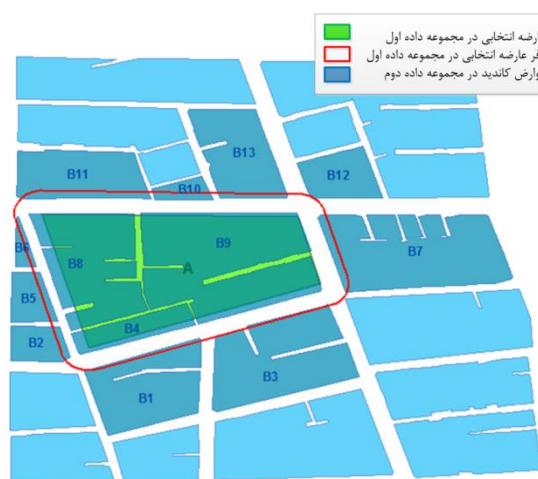
همچنین وزن نهایی هر یک از معیارهای هندسی از منظر ادراک مکانی انسان به کمک پرسشنامه قضاوت کارشناسان نقشه‌برداری در خصوص میزان اهمیت هر معیار حاصل شد. جدول ۲ وزن نهایی هر یک از چهار معیار شباهت را نشان می‌دهد که برابر میانگین وزن‌های اخذ شده از کارشناسان برای هر معیار می‌باشد. همچنین جهت تعیین حدآستانه شناسایی شباهت از مقدار تجربی ۰/۷ با توجه به منطقه مورد مطالعه استفاده شده است.

جدول ۲- وزن نهایی هر یک از معیارها

معیارهای هندسی			
W_{Dis}	W_{Area}	W_{Dir}	W_{Shape}
۰/۴۵	۰/۷۳	۰/۵۰	۰/۶۲

۴-۱- محاسبه شباهت مکانی

پس از مقداردهی وزن‌ها و حدآستانه‌ها به کمک نظر کارشناس و بر اساس منطقه مطالعاتی، شباهت مکانی عوارض کاندید برای هر عارضه از مجموعه داده اول مطابق با الگوریتم تناظریابی ارائه شده در بخش ۳-۱ محاسبه می‌شود تا جفت عوارض متناظر شناسایی شوند. در این زیربخش به عنوان نمونه، محاسبه شباهت مکانی عوارض کاندید در مجموعه داده دوم برای یکی از چندضلعی‌های انتخابی در مجموعه داده اول بیان شده است. همان‌طور که در شکل مشاهده می‌شود، با ترسیم بافر ۱۶/۳۳ متری برای عارضه انتخابی در مجموعه داده اول، ۱۳ عارضه در مجموعه داده دوم درون این محدوده بافر قرار می‌گیرند و به عنوان عوارض کاندید انتخاب می‌شوند. سپس شباهت مکانی بر اساس چهار معیار معرفی شده فاصله، مساحت هم‌پوشانی، جهت و شکل برای این عوارض کاندید در مقایسه با عارضه انتخابی محاسبه می‌شود.



شکل ۸- نمایش عوارض کاندید در مجموعه داده دوم برای عارضه منتخب در مجموعه داده اول

جدول ۳ نتایج محاسبه شباهت مکانی به تفکیک هر معیار را برای هر کدام از این عوارض کاندید در مقایسه با عارضه انتخابی مجموعه داده اول نشان می‌دهد. همان‌طور که بیان شد، حدآستانه درصد شباهت انتخاب جفت عارضه متناظر برابر با ۷۰ درصد در نظر گرفته شد. بر این اساس تنها سه عارضه B4، B8 و B9 درصد شباهتی بیش از ۷۰ درصد دارند و از میان عوارض کاندید انتخاب می‌شوند. در انتها سه جفت عوارض A-B8، A-B4 و A-B9 به عنوان سه جفت متناظر شناسایی شده‌اند.

۴-۲- ارزیابی دقت الگوریتم پیشنهادی

در بررسی کارایی روش‌های تناظریابی، هدف افزایش تعداد عوارضی است که صحیح منطبق شده‌اند و همچنین کاهش تعداد عوارضی که یا منطبق نشده‌اند و یا اشتباه منطبق شده‌اند. بدین منظور معیارهای مختلفی جهت ارزیابی این روش‌ها با در نظر گرفتن دو ملاک ذکر شده وجود دارد. دو معیار بسیار متداول جهت ارزیابی کارایی این مسائل، Precision و Recall می‌باشند که در اکثر مطالعات به آن‌ها اشاره شده است [۲، ۴، ۶، ۲۲]. معیار Precision نسبت تعداد عوارض صحیح منطبق شده (Nc) به کل عوارضی را که صحیح یا اشتباه منطبق شده‌اند (Na)، طبق رابطه (۸) بررسی می‌کند. همچنین معیار Recall نسبت عوارض صحیح منطبق شده (Nc) به کل عوارضی را که صحیح منطبق شده‌اند و یا منطبق نشده‌اند (Nt)، طبق رابطه (۹) ارزیابی می‌کند. از دیگر معیارهای مورد استفاده، معیار F1-score است. دو معیار Precision و Recall ممکن است در تضاد با یکدیگر باشند [۱]. به عبارت دیگر هر دو معیار به صورت هماهنگ بهبود نمی‌یابند و ممکن است به عنوان نمونه هنگامی که Recall کمتر شود، مقدار Precision بهبود یابد. در نتیجه نمی‌توان دید جامعی از ارزیابی روش مورد استفاده داشت. بدین منظور در برخی مطالعات از معیار F1-score طبق رابطه (۱۰) که ترکیبی از دو معیار Precision و Recall است، استفاده می‌شود [۱، ۲۱، ۲۶].

$$Precision = N_c / N_a \quad (8)$$

$$Recall = N_c / N_t \quad (9)$$

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (10)$$

بر اساس این سه معیار دقت الگوریتم تناظریابی ارائه شده در منطقه مورد مطالعه بررسی شده است. هنگامی که از چهار معیار هندسی فاصله، مساحت هم‌پوشانی، جهت و شکل به عنوان معیارهای اندازه‌گیری شباهت بر اساس وزن‌های نهایی ارائه شده استفاده می‌شود، معیار F1-score برابر با ۹۹/۰۰٪ حاصل می‌شود. همچنین دو معیار Precision و Recall به ترتیب برابر با ۹۸/۹۳ و ۹۹/۰۸ درصد می‌باشند که دقت بسیار مطلوبی در منطقه مورد مطالعه می‌باشد و نشان‌دهنده کارایی بالای الگوریتم ارائه شده بر اساس معیارهای هندسی در نظر گرفته شده است.

جدول ۳- مقادیر محاسبه شده هر یک از معیارها برای عوارض کاندید در مقایسه با عارضه انتخابی

جفت عوارض	مقادیر معیارهای هندسی				درصد شباهت
	اختلاف شکل (رادیان متر)	اختلاف جهت (رادیان)	مساحت هم پوشانی (متر مربع)	فاصله (متر)	
A-B1	۱/۱۷۴۶	۰/۲۶۱۷	۰	۱۴۰/۰۹۶۸	۵۶/۳۳
A-B2	۰/۲۶۹۸	۰/۰۹۵۲	۰	۱۵۳/۶۲۸۸	۵۷/۳۴
A-B3	۱/۰۱۶۳	۰/۲۰۱۳	۰	۱۴۳/۶۸۶۶	۵۶/۹۰
A-B4	۰/۶۰۴۲	۰/۰۶۱۷	۱۹۴۷/۸	۷۹/۶۵۲۴	۸۶/۹۹
A-B5	۰/۵۸۱۹	۰/۳۸۳۴	۰	۱۳۷/۸۵۵۴	۵۵/۳۴
A-B6	۰/۸۵۲۳	۰/۷۱۱۰	۰	۱۴۱/۶۷۷۸	۵۰/۲۳
A-B7	۱/۰۱۸۹	۰/۰۷۷۵	۰	۲۲۴/۱۹۶۱	۵۰/۰۴
A-B8	۱/۳۵۶۲	۰/۳۴۲۳	۸۲۷۵/۶	۷۶/۲۲۶۹	۸۷/۷۷
A-B9	۰/۲۵۸۴	۰/۱۱۲۷	۱۴۵۶۴	۴۷/۵۵۳۵	۹۵/۰۵
A-B10	۰/۲۹۱۷	۰/۰۹۱۶	۰	۸۷/۷۰۶۰	۶۴/۱۸
A-B11	۰/۰۲۹۳	۰/۱۱۶۰	۰	۱۳۶/۲۵۳۶	۵۹/۲۴
A-B12	۰/۷۵۱۷	۰/۱۱۳۲	۰	۱۸۶/۲۹۵۱	۵۳/۰۸
A-B13	۰/۴۱۹۹	۰/۲۹۲۸	۰	۱۴۲/۱۹۴۱	۵۶/۰۸

۳-۴- بررسی تاثیر معیارهای هندسی در دقت نهایی

در این بخش به بررسی تاثیر هر معیار هندسی ارائه شده به صورت مجزا و همچنین در حالت های مختلف ترکیب با سایر معیارها در دقت نهایی الگوریتم پرداخته می شود. هر کدام از معیارهای هندسی ارائه شده بخشی از ابهامات موجود در مسئله تناظریابی را حل می کنند، اما میزان تاثیرگذاری هر معیار وابسته به ماهیت معیار به لحاظ شناسایی شباهت و همچنین منطقه مورد مطالعه است. جدول ۵ دقت تناظریابی را برای همه حالت های ممکن ترکیب میان این چهار معیار نشان می دهد.

همان طور که جدول ۵ نشان می دهد، هنگامی که تنها از یک معیار شباهت در الگوریتم تناظریابی استفاده شود، معیار مساحت هم پوشانی بیشترین مقدار F1-score را در مقایسه با سه معیار دیگر (فاصله، جهت و یا شکل) فراهم می آورد. به عبارتی دیگر، هنگام استفاده از تک معیار، بهتر است از معیار مساحت هم پوشانی جهت شناسایی عوارض متشابه استفاده شود چرا که به عنوان نمونه در مقایسه با تک معیار فاصله که مقدار F1-score آن برابر با ۸۲/۹۱ درصد می باشد، حدود ۱۴ درصد دقت را بهبود می بخشد.

جدول ۴ تعداد روابط متناظر شناسایی شده اعم از روابط شناسایی شده صحیح، روابط شناسایی شده اشتباه، روابط شناسایی نشده و جمع کل روابط شناسایی شده را به تفکیک هر رابطه نشان می دهد. همان طور که ملاحظه می شود، بیشترین و کمترین تعداد روابط شناسایی شده به ترتیب از نوع 1:1 و 1:0 می باشد. همچنین مجموع کل روابط شناسایی شده ۲۷۰۶ رابطه می باشد که در مجموع ۲۷ رابطه آن به اشتباه شناسایی شده است و ۲۵ رابطه نیز شناسایی نشده است. این آمار و ارقام نیز نشان می دهد که الگوریتم ارائه شده با دقت بالایی هر شش نوع رابطه متناظر را در منطقه مطالعاتی شناسایی کرده است.

جدول ۴- تعداد روابط متناظر شناسایی شده در منطقه مورد مطالعه

کل روابط شناسایی شده	روابط شناسایی نشده	روابط شناسایی شده اشتباه	روابط شناسایی شده صحیح	روابط متناظر
۸۳۱	۱	۱	۸۳۰	1:1
۸۵	۱	۷	۷۸	1:0
۴۶۸	۱۳	۱	۴۶۷	0:1
۳۴۶	۰	۸	۳۳۸	1:N
۶۹۴	۴	۰	۶۹۴	N:1
۲۸۲	۶	۱۰	۲۷۲	N:M

هنگام استفاده از ترکیب دو معیار مطابق با جدول ۵، شش حالت ممکن برای ترکیب دو معیار از میان چهار معیار مورد بررسی وجود دارد. در میان این شش حالت، ترکیب دو معیار مساحت هم‌پوشانی و جهت با مقدار F1-score برابر با ۹۸/۰۷ درصد، بیشترین تأثیرگذاری در راستای رسیدن به بهترین دقت را فراهم می‌کند. همچنین هنگام ترکیب سه

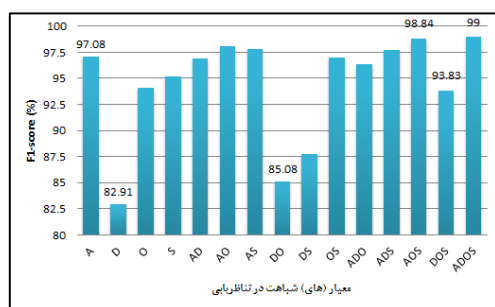
معیار، چهار حالت ممکن وجود دارد که بیشترین مقدار دقت تناظریابی با ترکیب سه معیار مساحت هم‌پوشانی، جهت و شکل حاصل می‌شود. در آخر نیز هنگام ترکیب هر چهار معیار مساحت هم‌پوشانی، فاصله، جهت و شکل، مقدار F1-score برابر با ۹۹ درصد می‌باشد که بیشترین مقدار در میان تمام حالات بررسی شده (۱۵ حالت ممکن) است.

جدول ۵- تأثیر ترکیب معیارهای اندازه‌گیری شباهت بر روی دقت تناظریابی

معیار(های) مورد بررسی	Precision	Recall	F1-score
مساحت هم‌پوشانی (A)	۹۷/۶۱	۹۶/۵۶	۹۷/۰۸
فاصله (D)	۸۲/۹۴	۸۲/۸۸	۸۲/۹۱
جهت (O)	۹۳/۶۶	۹۴/۴۹	۹۴/۰۷
شکل (S)	۹۴/۴۶	۹۵/۹۳	۹۵/۱۹
مساحت هم‌پوشانی و فاصله (AD)	۹۷/۴۲	۹۶/۴۵	۹۶/۹۳
مساحت هم‌پوشانی و جهت (AO)	۹۸/۵۱	۹۷/۶۳	۹۸/۰۷
مساحت هم‌پوشانی و شکل (AS)	۹۸/۰۳	۹۷/۵۶	۹۷/۷۹
فاصله و جهت (DO)	۸۴/۹۸	۸۵/۱۷	۸۵/۰۸
فاصله و شکل (DS)	۸۷/۵۵	۸۷/۸۷	۸۷/۷۱
جهت و شکل (OS)	۹۶/۵۶	۹۷/۴۵	۹۷/۰۰
مساحت هم‌پوشانی، فاصله و جهت (ADO)	۹۶/۵۸	۹۶/۰۸	۹۶/۳۳
مساحت هم‌پوشانی، فاصله و شکل (ADS)	۹۷/۸۹	۹۷/۶۳	۹۷/۷۶
مساحت هم‌پوشانی، جهت و شکل (AOS)	۹۸/۶۷	۹۹/..	۹۸/۸۴
فاصله، جهت و شکل (DOS)	۹۳/۴۷	۹۴/۱۹	۹۳/۸۳
مساحت هم‌پوشانی، فاصله، جهت و شکل (ADOS)	۹۸/۹۳	۹۹/۰۸	۹۹/۰۰

همچنین شکل مقدار F1-score را به ازای ۱۵ حالت بررسی شده که به اختصار بر اساس جدول ۵ تعریف شده‌اند، بر روی نمودار نشان می‌دهد. همان طور که ملاحظه می‌شود، حداقل دقت به هنگام استفاده از تک معیار فاصله حاصل می‌شود و بیشترین دقت مربوط به استفاده از چهار معیار معرفی شده می‌باشد که رویکرد اصلی این تحقیق نیز بر این اساس بوده است. به عبارتی دیگر، معیار دقت F1-score از ۸۲/۹۱ درصد تا ۹۹/۰۰ درصد بهبود یافته است. با وجود این، نکته قابل تامل در این بررسی عدم افزایش یکنواخت دقت با افزودن معیارهای هندسی است. برای مثال، طبق شکل، استفاده از تک معیار مساحت هم‌پوشانی (A) دقتی به مراتب بالاتر از استفاده ترکیبی از دو معیار فاصله و جهت (DO) و یا حتی استفاده از سه معیار فاصله، جهت و شکل (DOS)

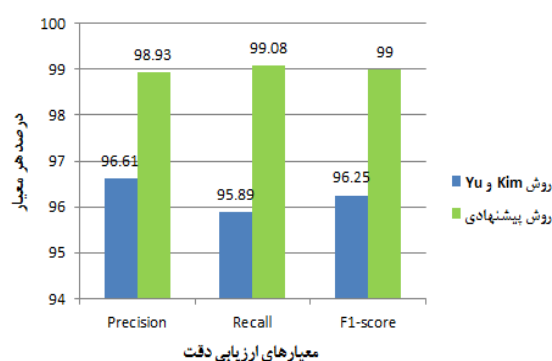
دارد. این امر نشان می‌دهد که لزوماً با افزایش تعداد معیارهای شباهت دقت افزایش نمی‌یابد. همچنین دقت استفاده از هر چهار معیار (ADOS) بسیار نزدیک به دقت استفاده از تنها سه معیار مساحت هم‌پوشانی، جهت و شکل (AOS) است. در نتیجه برای کاهش حجم محاسبات و زمان پردازش برنامه می‌توان با حفظ نسبی دقت مطلوب از معیارهای کمتری در مسائل تناظریابی استفاده نمود.



شکل ۹- دقت تناظریابی به ازای ترکیب مختلفی از معیارهای شباهت

۴-۴- ارزیابی نهایی روش پیشنهادی

جهت ارزیابی نهایی روش پیشنهادی، نتایج این روش با آخرین مقاله تناظریابی عوارض چندضلعی با استفاده از معیارهای هندسی ارائه شده توسط Kim و Yu در سال ۲۰۱۵ [۲۷] مقایسه شده است. معیارهای شباهت هندسی استفاده شده در روش Kim و Yu [۲۷] شامل سه معیار شباهت موقعیت، شکل و هم‌پوشانی می‌باشد که جزئیات روابط مربوط به هر معیار در متن مقاله مذکور تشریح شده است. در روش پیشنهادی این مطالعه نیز استفاده از تمام چهار معیار مساحت هم‌پوشانی، فاصله، جهت و شکل مبنای ارزیابی بوده است. لذا جهت مقایسه نتایج این دو پژوهش، روش تناظریابی Kim و Yu [۲۷] نیز بر روی منطقه مطالعاتی پیاده‌سازی شد. شکل مقادیر سه معیار دقت Precision، Recall و F1-score را به ازای پیاده‌سازی روش Kim و Yu [۲۷] در مقایسه با روش پیشنهادی این پژوهش نشان می‌دهد.



شکل ۱۰- ارزیابی روش پیشنهادی در مقایسه با روش Yu و Kim

همان‌طور که ملاحظه می‌شود، دقت روش پیشنهادی در هر سه معیار دقت بهبود یافته است. به عنوان نمونه مقدار F1-score در روش پیشنهادی ۲/۷۵ درصد در مقایسه با روش Kim و Yu [۲۷] بهبود یافته است. این نتایج نشان می‌دهد که نوع معیار اندازه‌گیری شباهت، رابطه استفاده شده جهت محاسبه معیار و ترکیب معیارهای انتخابی در دقت نهایی الگوریتم بسیار تاثیرگذار می‌باشد. در این تحقیق به کمک چهار معیار شباهت مذکور و روابط ریاضی تشریح شده جهت محاسبه هر معیار، دقت در مقایسه با آخرین روش ارائه شده بهبود یافته است.

۵- نتیجه‌گیری و پیشنهادها

هدف این پژوهش ارائه روشی جهت تناظریابی عوارض چندضلعی در مجموعه داده‌های چندمقیاسی بود که ضمن رسیدن به دقت مطلوب، تنها از معیارهای اندازه‌گیری شباهت هندسی جهت شناسایی عوارض متشابه استفاده شود. همچنین میزان تاثیرگذاری هر معیار هندسی به صورت مجزا و در ترکیب با سایر معیارها در دقت الگوریتم ارزیابی گردد و در انتها نیز این روش با آخرین روش ارائه شده در حوزه مورد مطالعه مقایسه شود. بدین منظور از منطقه شش تهران با دو مجموعه داده با مقیاس‌های منطقه مورد مطالعه استفاده شد تا دقت الگوریتم پیشنهادی در این منطقه ارزیابی شود.

روش پیشنهادی شامل چهار مرحله اصلی پیش‌پردازش، محاسبه شباهت مکانی، استخراج روابط متناظر و ارزیابی نتایج است. در مرحله پیش‌پردازش، عملیاتی هم‌چون یکسان‌سازی قالب داده‌ها و سیستم مختصات و حذف خطاهای توپولوژیکی صورت می‌پذیرد. سپس در مرحله محاسبه شباهت مکانی، الگوریتم احتمالی‌مبنا بر اساس چهار معیار اندازه‌گیری شباهت فاصله، مساحت هم‌پوشانی، جهت و شکل پیشنهاد گردیده که وزن‌دهی به هر معیار و انتخاب حد‌آستانه‌ها به کمک نظر کارشناس و بر اساس دقت منطقه مورد مطالعه انجام می‌شود. در انتها شش نوع رابطه 1:1، 1:0، 0:1، 1:N، N:1 و N:M به عنوان نتیجه نهایی الگوریتم تناظریابی استخراج می‌گردد. در مرحله ارزیابی نتایج نیز به کمک سه معیار Precision، Recall و F1-score دقت الگوریتم پیشنهادی در حالت‌های ترکیب مختلف چهار معیار هندسی ارزیابی شده است. نتایج نشان می‌دهد که حداکثر دقت الگوریتم زمانی است که از ترکیب هر چهار معیار اندازه‌گیری شباهت ارائه شده شامل فاصله، مساحت هم‌پوشانی، جهت و شکل استفاده می‌شود. به عبارتی دیگر، حداکثر دقت تناظریابی به هنگام در نظر گرفتن هم‌زمان چهار معیار شباهت هندسی حاصل شده و الگوریتم به دقت F1-score برابر با ۹۹/۰۰ درصد می‌رسد که دقت بسیار مطلوبی برای منطقه مورد مطالعه است. این مقدار F1-score نسبت به زمانی که تنها از یک معیار شباهت فاصله استفاده شود، حدود ۱۶٪ بهبود یافته است. اما بررسی حالات مختلف

ترکیب معیارهای شباهت نشان می‌دهد که لزوماً با افزایش تعداد معیارهای اندازه‌گیری شباهت، دقت تناظریابی افزایش نمی‌یابد و این مسئله به نوع معیار مورد استفاده، ترکیب آن با سایر معیارها و منطقه مورد مطالعه بستگی دارد. به عنوان نمونه نتایج نشان می‌دهد که استفاده از تک معیار مساحت هم‌پوشانی، دقتی به مراتب بالاتر از استفاده ترکیبی از دو معیار فاصله و جهت و یا حتی استفاده از سه معیار فاصله، جهت و شکل دارد. به بیان دیگر در این پژوهش اثبات شده است که همواره افزایش تعداد معیارهای اندازه‌گیری شباهت سبب بهبود دقت تناظریابی نمی‌گردد. در نتیجه جهت بهینه‌سازی زمان پردازش و کاهش حجم محاسبات پیشنهاد می‌گردد که تنها از معیار(هایی) استفاده شود که بیشترین تاثیرگذاری را در کشف عوارض متشابه در منطقه مورد مطالعه دارد. از طرفی دیگر، نتایج مقایسه این تحقیق با آخرین روش

پیشنهادی در حوزه مورد مطالعه نشان می‌دهد که دقت در روش پیشنهادی با چهار معیار شباهت مورد استفاده نسبت به تحقیقات پیشین بهبود یافته است. در این پژوهش با وجود رسیدن به دقت بسیار مطلوب در منطقه مورد مطالعه و همچنین بهبود دقت تحقیقات پیشین، پیشنهاداتی به تحقیقات آتی ارائه می‌گردد تا عملکرد و کارایی الگوریتم را بهبود بخشد و دقت بالای الگوریتم را در مناطق مطالعاتی با پیچیدگی بیشتر حفظ کند. لذا در این راستا توصیه می‌گردد که تاثیر سایر معیارهای اندازه‌گیری شباهت شامل معیارهای توصیفی و روابط مکانی هم به صورت مستقل و هم در کنار معیارهای هندسی در مسائل تناظریابی بررسی شود. همچنین راهکاری جهت شناسایی مهم‌ترین معیار(های) اندازه‌گیری شباهت تاثیرگذار در منطقه مورد مطالعه به صورت خودکار ارائه شود.

مراجع

- [1] 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.
- [2] H. Du, N. Alechina, M. Jackson, and G. Hart, "A Method for Matching Crowd-sourced and Authoritative Geospatial Data," Transactions in GIS, 2016.
- [3] 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.
- [4] 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.
- [5] 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, ed: Springer, 2014, pp. 139-153.
- [6] 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.
- [7] H. Du, "Matching disparate geospatial datasets and validating matches using spatial logic," PhD Thesis, University of Nottingham, 2015.
- [8] A. Saalfeld, "Conflation automated map compilation," International Journal of Geographical Information System, vol. 2, pp. 217-228, 1988.
- [9] I.-H. Kim, C.-C. Feng, and Y.-C. Wang, "A simplified linear feature matching method using decision tree analysis, weighted linear directional mean, and topological relationships," International Journal of Geographical Information Science, vol. 31, pp. 1042-1060, 2017.
- [10] 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.
- [11] A. Chehreghani and R. Ali Abbaspour, "A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm," Cartography and Geographic Information Science, pp. 1-15, 2017.
- [12] A.-M. Olteanu-Raimond, S. Mustiere, and A. Ruas, "Knowledge formalization for vector data matching using belief theory," Journal of Spatial Information Science, vol. 2015, pp. 21-46, 2015.
- [13] A. Chehreghani and R. Ali Abbaspour, "A new descriptor for improving geometric-based matching of linear objects on multi-scale datasets," GIScience & Remote Sensing, pp. 1-26, 2017.

- [14] E. Xavier, F. J. Ariza-López, and M. A. Ureña-Cámara, "A Survey of Measures and Methods for Matching Geospatial Vector Datasets," *ACM Computing Surveys (CSUR)*, vol. 49, p. 39, 2016.
- [15] B. Kieler, W. Huang, J.-H. Haunert, and J. Jiang, "Matching river datasets of different scales," in *Advances in GIScience*, ed: Springer, 2009, pp. 135-154.
- [16] E. Safra, Y. Kanza, Y. Sagiv, C. Beerli, and Y. Doytsheer, "Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets," *International journal of geographical information science*, vol. 24, pp. 69-106, 2010.
- [17] 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.
- [18] F. Zhonglianga and W. Jianhuaa, "Entity matching in vector spatial data," in *Proceedings of the XXI ISPRS Congress*, 2008, pp. 1467-1472.
- [19] 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.
- [20] L. Huang, S. Wang, Y. Ye, B. Wang, and L. Wu, "Feature matching in cadastral map integration with a case study of Beijing," in *2010 18th International Conference on Geoinformatics*, 2010, pp. 1-4.
- [21] 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.
- [22] X. Zhang, T. Ai, J. Stoter, and X. Zhao, "Data matching of building polygons at multiple map scales improved by contextual information and relaxation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 92, pp. 147-163, 2014.
- [23] T. Ai, X. Cheng, P. Liu, and M. Yang, "A shape analysis and template matching of building features by the Fourier transform method," *Computers, Environment and Urban Systems*, vol. 41, pp. 219-233, 2013.
- [24] E. M. Arkin, L. P. Chew, D. P. Huttenlocher, K. Kedem, and J. S. Mitchell, "An efficiently computable metric for comparing polygonal shapes," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 13, NO. 3, MARCH 1991.
- [25] A. Farahanipooya, R. Abbaspour, S. 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.
- [26] Y. Huh, K. Yu, and W. Park, "Detecting Corresponding Vertex Pairs between Planar Tessellation Datasets with Agglomerative Hierarchical Cell-Set Matching," *PloS one*, vol. 11, p. e0157913, 2016.
- [27] J. Kim and K. Yu, "AREAL FEATURE MATCHING BASED ON SIMILARITY USING CRITIC METHOD," *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1, pp. 75-78, 2015.