

استفاده از الگوریتم تطبیق نقشه مکانی-زمانی به منظور افزایش دقت اطلاعات ترافیکی حاصل از ردیابی GPS با نرخ پایین

وحید شکری^۱، رحیم علی عباسپور^۲

۱- دانشجوی کارشناسی ارشد مهندسی سیستم‌های اطلاعات مکانی GIS - دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس دانشکده‌های فنی - دانشگاه تهران

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

چکیده

افزایش محبوبیت دستگاه‌های مجهز به سیستم تعیین موقعیت جهانی امکان ردیابی اشیای متحرک مانند خودرو، حیوانات حیات وحش و افراد را برای کاربران فراهم ساخته است. اما اطلاعات به دست آمده از حسگر GPS همواره همراه با خطای تعیین موقعیت و نمونه برداری می‌باشد. از طرف دیگر، کاربردهای ترافیکی از جمله جمع‌آوری داده‌های ترافیکی، نیازمند مسیرهای ردیابی دقیق و منطبق بر شبکه جاده‌ای هستند. از آنجایی که درصد بالایی از داده‌های ردیابی GPS، به خصوص داده‌های تولید شده توسط GPS تلفن همراه و خودروها، با نرخ پایین هستند، به کارگیری یک الگوریتم پیشرفته برای خدمات مکان‌مبنا ضروری بنظر می‌رسد. در این تحقیق هدف به کارگیری الگوریتم تطبیق نقشه مکانی-زمانی به منظور تحلیل مسیرهای ردیابی شده با نرخ پایین می‌باشد. بدین منظور در این مقاله داده‌های ردیابی GPS مربوط به هشت خط اتوبوس در محدوده مناطق ۱ تا ۴ شهرداری با نرخ نمونه‌برداری دو دقیقه مورد استفاده قرار گرفت. این الگوریتم از تحلیل‌های مکانی و زمانی به منظور ساخت گراف کاندید که در آن مجموعه‌ای از نتایج منطبق شده با بالاترین مجموع رتبه به عنوان نتیجه تطبیق نقشه تعیین می‌شود، استفاده می‌کند. نتایج حاصل از این الگوریتم نشان می‌دهد الگوریتم موردنظر در تطبیق نقشه با ردیابی‌های GPS با نرخ پایین با دقت نسبتاً بالایی در حدود هفتاد درصد صحیح عمل می‌کند.

واژگان کلیدی: جمع‌آوری داده‌های ترافیکی، مسیر ردیابی GPS، تطبیق نقشه، شبکه جاده، نمونه برداری با نرخ پایین.

۱- مقدمه

و تحلیل جریان ترافیک نیازمند مجموعه داده‌های دقیق و کارای ردیابی مسیر خودرو هستند [۲] و [۳]. با توجه به حریم خصوصی و مشکلات هزینه، به دست آوردن مجموعه داده‌های ردیابی مسیر با مشکلاتی روبرو است [۴] و [۵]. تاکنون دستگاه سیستم موقعیت جهانی، منبع اصلی برای مجموعه داده‌های مسیر بوده است. به طور معمول یک مسیر GPS خودرو شامل دنباله‌ای از نقاط نمونه برداری با موقعیت مکانی و برچسب اطلاعات زمانی است [۶]. اطلاعات حسگر GPS همراه با خطای تعیین موقعیت و نمونه برداری است و خط سیر ردیابی شده توسط آن به ندرت عاری از خطا می‌باشد [۷].

از آنجا که اکثر دستگاه‌های GPS مورد استفاده توسط عموم مردم در تلفن‌های هوشمند دقت پایینی داشته و با توجه به فاصله نمونه‌برداری کم و همچنین DOP پائین (به خصوص

افزایش محبوبیت دستگاه‌های مجهز به سیستم تعیین موقعیت جهانی امکان ردیابی اشیای متحرک مانند خودرو، حیوانات حیات وحش و مردم را برای کاربران فراهم ساخته است. این امر موجب بوجود آمدن جنبه‌های جدید در حوزه‌های تحقیقاتی حمل و نقل شده است. تجزیه و تحلیل اطلاعات ردیابی سیستم تعیین موقعیت جهانی^۱ اطلاعات ترافیکی مفیدی را در اختیار قرار می‌دهد و امکان استفاده از خدمات مکان مبنای سامانه حمل و نقل هوشمند را فراهم می‌آورد [۱].

بسیاری از برنامه‌های مبتنی بر مکان، مانند ناوبری خودرو، مدیریت ناوگان، سیستم حمل و نقل هوشمند، کاربردهای مدیریت ترافیک مانند پیشنهاد مسیر، نظارت بر ترافیک

1 GPS trajectory

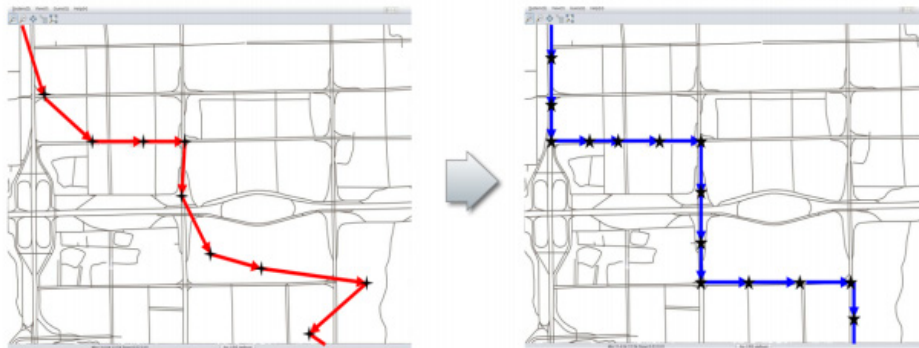
نرخ بالا سروکار داشته و بکارگیری آنها برای نمونه‌های داده با نرخ پایین موجب افزایش عدم قطعیت در داده و کاهش کارایی می‌گردد [۹_ENREF_۶]. لذا هدف این مقاله، معرفی و به‌کارگیری یک الگوریتم تطبیق نقشه مکانی-زمانی به منظور افزایش دقت اطلاعات ردیابی GPS با نرخ پایین می‌باشد. برای این منظور از داده‌های ردیابی GPS با نرخ نمونه‌برداری دو دقیقه که مربوط به هشت خط اتوبوس در محدوده مناطق یک تا چهار شهرداری تهران استفاده شده‌است. در ادامه، ابتدا مفهوم تطبیق نقشه و انواع روشهای آن تشریح می‌گردد. سپس الگوریتم مورد نظر و روابط آن بیان می‌گردد و در انتها این الگوریتم برای داده‌های مورد نظر به‌کارگیری می‌شود.

۲- تطبیق نقشه

مکان نقاط گزارش شده GPS، اغلب با خطایی حتی در مواردی تا محدوده صدها متر از مکان واقعی هستند [۴]. برای درک بهتر شکل ۱ الف ۹ نقطه GPS در محدوده چهار کیلومتر مربع را نشان می‌دهد در اکثر برنامه‌های کاربردی مسیر دقیق و کارآمدی مطابق شکل ۱ ب، مورد نیاز است مسیری که منطبق بر شبکه جاده‌ای باشد [۷].

در مناطق شهری، بلوک ساختمانهای بلند به طور جدی سیگنال GPS و در نتیجه دقت تعیین موقعیت را تحت‌الشعاع قرار می‌دهد (داده‌های خام به دست آمده، معمولاً شامل نقاط نمونه ناکافی خطادار است [۸]). تطبیق نقشه روند هماهنگی و تطبیق مجموعه‌ای از موقعیتهای مشاهده شده توسط کاربر با شبکه جاده بر روی یک نقشه دیجیتال است. اکثر الگوریتم‌های تطبیق نقشه بر مبنای مجموعه داده‌های ردیابی با نرخ بالا هستند. از آنجا که اخذ و نگهداری داده‌های با نرخ بالا نیازمند فضای ذخیره‌سازی و انرژی مصرفی بیشتر است، کاربران معمولاً در برداشت داده‌های GPS، فرکانس نمونه‌برداری را کاهش می‌دهند و در نتیجه مجموعه داده‌های بسیاری با نرخ پایین تولید می‌شود [۹].

با توجه به گسترش کاربردها و قابلیت‌های تلفنهای هوشمند مجهز به GPS و تجهیز ناوگان حمل‌ونقل عمومی از جمله اتوبوسها به GPS، در عمل داده‌های ردیابی GPS با نرخ پایین (تقریباً یک نقطه در هر ۲ تا ۵ دقیقه) بسیاری وجود دارد که می‌تواند به عنوان منابع داده‌های اطلاعات ترافیکی مناسب مورد استفاده قرار گیرد. متأسفانه، اکثر پژوهشهای تطبیق نقشه در حال حاضر تنها با داده‌های GPS با نمونه برداری



(الف)

(ب)

شکل ۱: نمونه ردیابی GPS [۷]

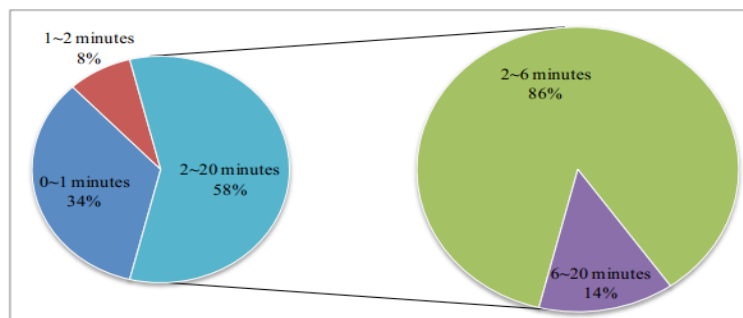


Figure 1. Distribution of the sampling interval

شکل ۲: توزیع فاصله زمانی نمونه برداری [۸]

می‌کنند [۱۱] و [۳]. گروه سوم از روشهایی دیگری و ترکیب روش‌ها و اطلاعات جانبی GPS مانند سرعت، جهت حرکت استفاده می‌کنند و معمولاً نیازمند زمان بیشتری هستند. برای نمونه از الگوریتمهای مورد استفاده در این روش می‌توان به الگوریتمهای مبتنی بر روشهای یادگیری ماشین مبتنی بر منطق فازی و کالمن فیلتر اشاره نمود [۱۳] و [۱۴].

۲-۲- مشکلات الگوریتمهای تطبیق نقشه

افزایش فاصله بین دو نقطه مجاور یک چالش بزرگ برای تطبیق نقشه به‌شمار می‌رود؛ چرا که این امر باعث کاهش دقت تعیین موقعیت اشیا متحرک می‌شود. این مسئله زمانی که شی در حال حرکت با سرعت بالا باشد و یا تقاطعهای زیادی بین دو نقطه مجاور مشاهده شده باشد، تشدید می‌گردد. به منظور رویارویی با چالشهای مذکور در الگوریتم ارائه شده از دو مشاهده دیگر استفاده می‌شود. نقاط Pa, Pb, Pc مطابق شکل ۳ الف سه نقطه متوالی از مسیر ردیابی GPS می‌باشند. بیشتر الگوریتمهای تطبیق نقشه، نقطه Pb را به نزدیکترین قطعه جاده عمودی در شکل اختصاص می‌دهند. اما با توجه به موقعیت مکانی نقطه قبل و بعد به نظر می‌رسد مسیر صحیح مسیر افقی مشاهده شده در شکل ۳ الف می‌باشد. بنابراین اطلاعات توپولوژیکی شبکه جاده به همراه اطلاعات هندسی موجب تطبیق بهتر نقشه می‌شود. یک مسیر دیگر ردیابی GPS

یک پیش پردازش اساسی برای بسیاری از کاربردهای ذکر شده، نگاشت اطلاعات به دست آمده از گیرنده GPS بر یک نقشه رقومی است و در نتیجه این عمل می‌توان به تحلیل رفتار تحرکی پرداخت و هم چنین به طور خاص کاربردهایی مانند محاسبه عوارض به صورت پویا و یا اطلاعات رسیدن اتوبوس‌ها به ایستگاه‌ها دست یافت. بنابراین یک الگوریتم مناسب و قابل اعتماد برای کاهش این خطا در خدمات مبتنی بر مکان بسیار ضروری است [۸].

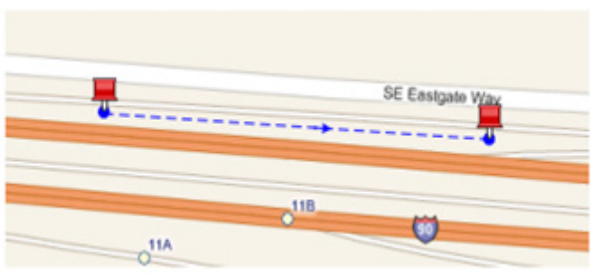
مطابقت دادن داده‌های ردیابی GPS با نقشه اصلی دیجیتال یا یک شبکه جاده‌ای اغلب به عنوان تطبیق نقشه یاد می‌شود. [۶]. هدف اصلی تطبیق نقشه به‌کارگیری اطلاعات شبکه جاده راه‌ها و اطلاعات مکانی-زمانی GPS به منظور تصحیح خطاها می‌باشد. شکل ۲ توزیع آماری فواصل نمونه برداری ردیابی GPS تولید شده توسط بیش از ده هزار تاکسی در پکن در یک هفته را ارائه می‌دهد. متوسط فاصله زمانی این مجموعه داده‌ها دقیقه ۳/۲۷ است. با توجه به نمودار، تنها ۳۴ درصد از داده‌های نمونه‌گیری با نرخ بالا (نرخ نمونه برداری به کمتر از ۱ دقیقه) است. بیش از ۶۰ درصد از داده‌ها داده‌های با نرخ پایین است [۸]. در ایران نیز اکثر داده‌های ردیابی موجود از جمله داده‌های ردیابی GPS موجود شرکت کنترل ترافیک که در این مقاله مورد استفاده قرار گرفته است، با نرخ نمونه‌برداری پایین هستند.

۲-۱- انواع الگوریتمهای تطبیق نقشه

الگوریتمهای متعددی برای تطبیق نقشه ارائه شده است که در سه گروه قرار می‌گیرند: الگوریتمهایی که از اطلاعات هندسی برای تعیین موقعیت استفاده می‌کنند، الگوریتمهایی که از روشهای توپولوژیکی کمک می‌گیرند و الگوریتمهایی که از روشهای یادگیری ماشین استفاده می‌کنند [۱۰]. الگوریتمهای هندسی از جمله الگوریتم تطبیق نقطه با نقطه، تطبیق نقطه با منحنی و تطبیق منحنی با منحنی تنها از اطلاعات هندسی برای تطبیق نقشه استفاده می‌کنند. به طور کلی الگوریتمهای هندسی تنها با توجه به شکل لینکهای جاده عمل می‌کنند و اتصال بین شبکه را در نظر نمی‌گیرند و به دلیل حساسیت بالا به داده‌های پرت و خطاهای بارز، چندان مطلوب نیستند [۱۱] و [۱۲]؛ و نیاز به روشهایی است که عمل تطابق را با توجه به نقطه قبلی و داده‌های ثبت شده GPS (مانند زمان) و هم‌چنین توجه بیشتر به داده‌های نقشه در دسترس انجام دهند [۱]. الگوریتمهای مبتنی بر روشهای توپولوژیکی با استفاده از هندسه، اتصال و مجاورت لینکها، مبادرت به تطبیق نقشه



(الف)



(ب)

شکل ۳: توصیف مشاهده اول و دوم [۶]

می تواند شامل چندین قطعه باشد.

- مسیر: از دو رأس V_j, V_j در شبکه جاده G ، مسیر P مطابق رابطه ۲ شامل مجموعه پیوسته‌ای از قطعات جاده که با V_i شروع و به V_j ختم می‌شوند.

$$P: e_1 \rightarrow e_2 \rightarrow \dots \rightarrow e_n \quad (2)$$

$$e_1.start = V_i, e_n.end = V_j, e_k.end = e_{k+1}.start, 1 \leq k < n$$

بنابر تعاریف فوق مسئله تطبیق نقشه به صورت زیر است:

- تطبیق نقشه ردیابی GPS خام T و شبکه جاده $G(V, E)$ موجود است و هدف یافتن مسیر P از شبکه G به نحوی که T با مسیر واقعی مطابق باشد.

۴- الگوریتم تطبیق نقشه مکانی زمانی ST-matching

با توجه به توضیحات بخش مبانی نظری، در این مقاله از الگوریتم ST-matching که برای ردیابی با نرخ پایین توسعه داده شده است، استفاده می‌شود. این الگوریتم از هر دو نوع اطلاعات هندسی و ساختار توپولوژیکی شبکه و محدودیتهای سرعت و زمان ردیابی GPS بهره می‌گیرد. با ترکیب تحلیل‌های مکانی و زمانی و انتخاب یک گراف کاندید الگوریتم ST-matching سعی بر آن دارد تا یک مسیر کلی با بالاترین رتبه را تشخیص دهد.

در مرحله اول اطلاعات GPS و شبکه معابر وارد پایگاه داده گردیده و نقاط کاندید تعیین می‌گردند. در مرحله دوم با بهره‌گیری از تحلیل‌های مکانی شامل کوتاهترین مسیر و

را مطابق شکل ۳ ب در نظر بگیرید، تشخیص مسیر صحیح از بین دو مسیر موجود در شکل ۳، تقریباً غیر ممکن است. اما با در نظر گرفتن سرعت بر اساس فاصله بین دو نقطه و برچسب زمانی می‌توان دریافت که سرعت مسیر ردیابی نزدیک به سرعت بزرگراه است. لذا اطلاعات زمان و سرعت نیز در فرآیند تطبیق نقشه بسیار مفید است [۶].

۳- مبانی نظری

قبل از ورود به معرفی الگوریتم پیشنهادی، لازم است تا برخی از پارامترهای مورد نیاز برای الگوریتم تشریح گردد که در این بخش به این امر مبادرت خواهد شد.

- GPS log: شامل مجموعه ای از نقاط GPS $L = \{p_1, p_2, \dots, p_n\}$ که هر نقطه آن دارای طول جغرافیایی، عرض جغرافیایی و برچسب زمانی است.

- ردیابی GPS: T به صورت رابطه ۱ شامل مجموعه‌ای از نقاط متوالی GPS که بازه زمانی آنها از یک آستانه مشخص ΔT تجاوز ننماید.

$$T = p_1, p_2, \dots, p_n, p_i \in L \quad (1)$$

$$0 < p_{i+1}.t - p_i.t < \Delta T, (1 \leq i < n)$$

شکل ۴ نمونه‌ای از ردیابی GPS را نشان می‌دهد که در آن می‌باشد.

- قطعه جاده: e مطابق شکل ۵ یک قسمت مستقیم از جاده که با شناسه مشخص $e.id$ ، سرعت متوسط $e.v$ با طول $e.l$ و نقطه شروع $e.start$ و نقطه پایان $e.end$ و لیست نقاطی که یک بخش از جاده را توصیف می‌کنند. یک جاده

Latitude,	Longitude,	Time
p_1 : 39.9848,	116.3319,	13:11
p_2 : 39.9841,	116.3472,	13:13,
.....
p_n : 39.9858,	116.3558,	13:25

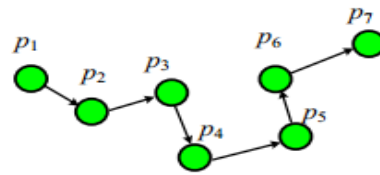


Figure 3. Illustration of GPS log and GPS trajectory

شکل ۴: نمونه GPS log و ردیابی GPS [۶]



Figure 4. Illustration of road segment

شکل ۵: نمونه قطعه جاده [۶]

نقطه همسایه نقطه نمونه برداری p_{i-1} و p_i احتمال انتقال از نقطه c_{i-1}^t به c_i^s به صورت شباهت مسیر واقعی از نقطه p_{i-1} به p_i با کوتاهترین مسیر c_{i-1}^t به c_i^s به صورت رابطه ۴ تعریف می‌شود:

$$V(c_{i-1}^t \rightarrow c_i^s) = \frac{d_{i-1} \rightarrow i}{w(i-1, t) \rightarrow (i, s)} \quad (4)$$

که در آن $d_{i-1} \rightarrow i$ فاصله اقلیدسی بین دو نقطه p_{i-1} و p_i و $w(i-1, t) \rightarrow (i, s)$ طول کوتاهترین مسیر از c_{i-1}^t به c_i^s است. با ترکیب رابطه ۳ و ۴ تابع تحلیل مکانی به صورت رابطه ۵ به دست می‌آید.

$$F_s(c_{i-1}^t \rightarrow c_i^s) = N(c_i^s) * V(c_{i-1}^t \rightarrow c_i^s), 2 \leq i \leq n \quad (5)$$

کوتاهترین مسیرها از c_{i-1}^t به c_i^s به صورت لیستی از قطعات جاده $[e_1, e_2, \dots, e_k]$ هستند در تحلیل زمانی سرعت متوسط کوتاهترین مسیر از رابطه ۶ محاسبه می‌گردد:

$$\bar{v}(i-1, t) \rightarrow (i, s) = \frac{\sum_{u=1}^k l_u}{\Delta t_{i-1} \rightarrow i} \quad (6)$$

که در آن $l_u = e'_u \cdot \Delta t_{i-1} \rightarrow i$ طول e'_u و $\Delta t_{i-1} \rightarrow i = p_i \cdot t - p_{i-1} \cdot t$ بازه زمانی بین دو نقطه نمونه برداری p_{i-1} و p_i است. بنابراین هر قطعه e'_u متناسب با یک مقدار سرعت $e'_u \cdot v$ است. با به کارگیری فاصله کسینوسی میزان شباهت سرعت متوسط واقعی و محدودیت سرعت مسیر از c_{i-1}^t به c_i^s به دست می‌آید. تابع تحلیل زمانی از ضرب دو بردار به شکل رابطه ۷ محاسبه می‌شود:

$$F_t(c_{i-1}^t \rightarrow c_i^s) = \frac{\sum_{u=1}^k (e'_u \cdot v \times \bar{v}(i-1, t) \rightarrow (i, s))}{\sqrt{\sum_{u=1}^k (e'_u \cdot v)^2} \times \sqrt{\sum_{u=1}^k \bar{v}(i-1, t) \rightarrow (i, s)^2}} \quad (7)$$

۲-۴-۲- تطبیق نتایج

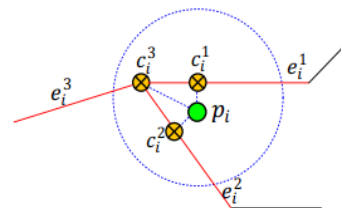
پس از تحلیل مکانی و زمانی اکنون قادر به ایجاد گراف کاندید شامل مجموعه‌ای از نقاط کاندید برای هر نقطه نمونه برداری $T: p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n, V_T$ برای ردیابی $G_T(V_T, E_T)$ شامل مجموعه‌ای از نقاط کاندید برای هر نقطه نمونه برداری E_T و GPS هر دو نقطه همسایه کاندید است. همانطور که در شکل ۷ مشاهده می‌شود هر گره در G' با $N(c_i^s)$ و هر یال آن با $V(c_{i-1}^t \rightarrow c_i^s)$ و $F_t(c_{i-1}^t \rightarrow c_i^s)$ در ارتباط است. با ترکیب رابطه ۵ و ۷ تابع ST c_{i-1}^t به c_i^s به صورت رابطه ۸ است

$$F(c_{i-1}^t \rightarrow c_i^s) = F_s(c_{i-1}^t \rightarrow c_i^s) * F_t(c_{i-1}^t \rightarrow c_i^s), 2 \leq i \leq n \quad (8)$$

اطلاعات توپولوژیکی شبکه معابر و تحلیل‌های زمانی شامل محاسبه سرعت متوسط، گراف کاندید انتخاب می‌گردد. نقاط این گراف شامل مجموعه نقاط کاندید برای هر مشاهده GPS و یالهای آن شامل مجموعه کوتاهترین مسیرها بین هر دو نقطه کاندید همسایه می‌باشد. بر اساس نتایج تحلیل‌های مکانی و زمانی به نقاط و یالهای گراف وزنی اختصاص می‌یابد. در مرحله سوم گراف کاندید با وزنه‌های اختصاص یافته ارزیابی می‌گردد. سپس ردیابی GPS با مسیری که دارای بالاترین وزن در گراف کاندید است، تطبیق می‌یابد [۶]. در نهایت اطلاعات خروجی می‌تواند در پایگاه داده حمل‌ونقل و ترافیک به منظور بهره‌گیری در کاربردهای مختلف ترافیکی از جمله استخراج پارامترهای ترافیک مانند زمان سفر، ذخیره گردد.

۱-۴-۱- تحلیل مکانی و زمانی

در مرحله تحلیل مکانی در ابتدا باید مجموعه‌ای از قطعات جاده در محدوده یک دایره مانند شکل ۶ کاندید شوند. و سپس بر اساس آن، مجموعه نقاط کاندید شوند. حال باید نقاط کاندید که بیشترین تطابق و نزدیکی با واقعیت را دارند، انتخاب شوند.



شکل ۶: نقاط کاندید نقطه نمونه برداری p_i [۶]

اطلاعات هندسی از طریق پارامتر احتمال مشاهده و اطلاعات توپولوژیکی از طریق پارامتر احتمال انتقال بیان می‌شوند که در ادامه به تعریف هر کدام از آنها پرداخته می‌شود.

احتمال مشاهده میزان شباهت نقطه نمونه برداری GPS p_i با نقطه کاندید c_i^j می‌باشد که بر اساس فاصله بین این دو نقطه مطابق رابطه ۳ محاسبه می‌گردد. در واقع میزان خطای اندازه‌گیری GPS از طریق توزیع نرمال فاصله بین این دو نقطه محاسبه گردد.

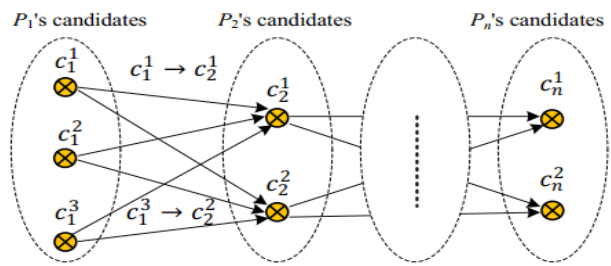
$$N(c_i^j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_j^i - \mu)^2}{2\sigma^2}\right) \quad (3)$$

که در آن x_j^i فاصله بین دو نقطه p_i و c_i^j است و توزیع نرمال با میانگین صفر و انحراف معیار ۲۰ متر بر اساس ارزیابی‌های تجربی در نظر گرفته شده است.

پارامتر احتمال انتقال: دو نقطه کاندید c_i^s به ترتیب برای دو

۵- منطقه مورد مطالعه و داده‌های مورد استفاده

در این مقاله منطقه مورد مطالعه شامل مناطق یک تا چهار شهرداری شهر تهران می‌باشد. این مناطق دارای خصوصیتی همچون وسعت بالا، شبکه معابر غنی و متنوع می‌باشد. به منظور تست و ارزیابی الگوریتم ارائه شده در فوق از داده‌های ردیابی GPS اخذ شده توسط اتوبوسهای هشت خط از این مناطق بهره‌گرفته شده است. که این هشت خط به صورت چهار خط رفت و برگشت می‌باشد داده‌های GPS اخذ شده از شرکت کنترل ترافیک شهر تهران شامل محدوده مناطق ۱ تا ۴ شهرداری، شبکه معابر، خطوط اتوبوسرانی و داده‌های مربوط به ردیابی GPS هشت خط اتوبوس با نرخ نمونه برداری پایین (دو دقیقه) و جداول پایگاه داده می‌باشد. مراحل آماده‌سازی داده، تشکیل شبکه راه (بالها و گره‌ها)، تشکیل جداول توپولوژی و پیاده سازی الگوریتم در پایگاه داده postgresql مجهز به افزونه PostGIS و به زبان plpgsql



شکل ۷: گراف کاندید $G'_T(V'_T, E'_T)$ [۶]

مسیر کاندید P_C برای کل ردیابی T شامل مسیری در گراف

کاندید به صورت $P_C: c_1^{s_1} \rightarrow c_2^{s_2} \rightarrow \dots \rightarrow c_n^{s_n}$ است.

هدف تعیین مسیر با بالاترین رتبه کلی است که مطابق ترین مسیر با ردیابی GPS است و از رابطه ۹ به دست می‌آید [۶]:

$$P = \arg \max_{P_C} F(P_C), \forall P_C \in G'_T(V'_T, E'_T) \quad (9)$$

$$F(P_C) = \sum_{i=1}^n F(c_{i-1}^{s_{i-1}} \rightarrow c_i^{s_i})$$



شکل ۸: نقاط نمونه ردیابی GPS خط پایانه لاله-پایانه افشار

جدول ۱: مقادیر N_j^i برای سه نقطه نمونه برداری متوالی ۱ و ۲ و ۳

$i = 5$	$i = 4$	$i = 3$	$i = 2$	$i = 1$	i, j
۰/۰۰۰۵۱۱	۰/۰۰۰۵۲۴	۰/۰۰۱۳۷۰	۰/۰۰۶۶۶۷	۰/۰۱۵۱۱۶	$j = 1$
۰/۰۰۰۱۱۸	۰/۰۰۰۲۲۸	۰/۰۰۱۸۱۸	۰/۰۱۰۱۵۷	۰/۰۱۹۳۲۷	$j = 2$
۰/۰۰۲۴۸۶	۰/۰۰۲۴۸۶	۰/۰۰۴۳۰۲	۰/۰۰۹۷۰۳	۰/۰۱۷۹۵۴	$j = 3$

در مرحله بعد با استفاده از رابطه ۷ و مقادیر شاخص مکانی F_s و شاخص زمانی F_t برای نقاط کاندید از c_{i-1}^t به c_i^s محاسبه گردید.

جدول ۲: مقادیر $F(c_{i-1}^t \rightarrow c_i^s)$ از نقاط کاندید نقطه نمونه برداری ۱ به ۲

c_4^t	c_4^s	c_3^t	c_3^s	c_2^t	c_2^s	$F = F_s * F_t$
۰/۰۰۸۳۵۱	۰/۰۱۵۰۷۵	۰/۰۱۲۸۱۸	۰/۰۰۷۹۶۶	۰/۰۰۷۸۹۹	۰/۰۰۷۸۹۹	c_1^1
۰/۰۰۲۲۲۰	۰/۰۰۵۵۸۹	۰/۰۰۶۶۵۰	۰/۰۰۲۹۴۱	۰/۰۰۲۰۴۳	۰/۰۰۲۰۴۳	c_1^2
۰/۰۰۰۷۵۴	۰/۰۰۰۷۰۹	۰/۰۰۰۷۱۲	۰/۰۰۱۳۶۸	۰/۰۰۷۱۲	۰/۰۰۷۱۲	c_1^3
۰/۰۰۰۳۵۲	۰/۰۰۰۲۵۷	۰/۰۰۰۲۲۷	۰/۰۰۰۵۲۱	۰/۰۰۳۳۷	۰/۰۰۳۳۷	c_1^4
۰/۰۰۰۳۱۶	۰/۰۰۰۵۰۴	۰/۰۰۰۵۰۹	۰/۰۰۳۰۵	۰/۰۰۳۰۲	۰/۰۰۳۰۲	c_1^5

جدول ۳: مقادیر $F(c_{i-1}^t \rightarrow c_i^s)$ از نقاط کاندید نقطه نمونه برداری ۲ به ۳

c_3^s	c_3^t	c_2^t	c_2^s	c_1^s	$F = F_s * F_t$
۰/۰۲۳۶۰۶	۰/۰۱۹۳۱۸	۰/۰۱۸۷۱۸	۰/۰۱۵۱۸۹	۰/۰۱۹۳۱۸	c_2^t
۰/۰۱۲۳۸۴	۰/۰۱۰۱۳۷	۰/۰۰۹۷۹۹	۰/۰۱۰۱۵۱	۰/۰۱۰۱۳۷	c_2^s
۰/۰۰۱۸۵۳	۰/۰۰۱۵۶۳	۰/۰۰۱۵۱۶	۰/۰۰۱۵۷۴	۰/۰۰۱۵۷۳	c_2^t
۰/۰۰۰۰۹۶	۰/۰۰۰۰۸۹	۰/۰۰۰۰۸۸	۰/۰۰۰۰۹۲	۰/۰۰۰۰۹۲	c_2^s
۰/۰۰۰۱۴۷	۰/۰۰۰۱۱۷	۰/۰۰۰۱۱۳	۰/۰۰۰۰۹۱	۰/۰۰۰۱۱۷	c_2^t

 جدول ۴: مقادیر $f[c_i^j]$ برای سه نقطه نمونه برداری متوالی ۱ و ۲ و ۳

$i=5$	$i=4$	$i=3$	$i=2$	$i=1$	i, j
۰/۰۰۰۵۱۱	۰/۰۰۰۵۲۴	۰/۰۰۱۳۷۰	۰/۰۰۶۶۶۷	۰/۰۱۵۱۱۶	$j=1$
					$j=2$
					$j=3$

 جدول ۵: مقادیر $f[c_i^j]N_{ij}^z = \frac{N_C}{N_A} A_L = \frac{L_C}{L_A}$ برای سه نقطه نمونه برداری متوالی ۱ و ۲ و ۳

$i=5$	$i=4$	$i=3$	$i=2$	$i=1$	i, j
۰/۰۰۰۵۱۱	۰/۰۰۰۵۲۴	۰/۰۰۱۳۷۰	۰/۰۰۶۶۶۷	۰/۰۱۵۱۱۶	$j=1$
۰/۰۲۳۴۶۸	۰/۰۳۰۱۹۲	۰/۰۲۷۹۳۵	۰/۰۲۳۰۸۳	۰/۰۲۳۰۱۵	$j=2$
۰/۰۴۶۶۲۲	۰/۰۴۲۳۳۴	۰/۰۴۱۷۳۴	۰/۰۳۸۲۰۵	۰/۰۴۲۳۳۴	

برنامه‌نویسی شده است. استراتژی گفته شده در بالا و با انتخاب دایره به شعاع ۱۰۰ متر، برای هر یک از نقاط نمونه‌برداری GPS با شماره‌های ۱ و ۲ و ۳ نقاطی منتخب شد ($k=5$) و بر مبنای رابطه ۳ برای هر یک از نقاط مقادیر به صورت جدول ۱ محاسبه گردید.

مقادیر $F(c_{i-1}^t \rightarrow c_i^s)$ از نقاط کاندید نقطه نمونه برداری ۱ به ۲ و ۳ به ۲ از طریق رابطه ۸ به صورت جداول ۲ و ۳ به دست آمد. برای یافتن دنباله تطبیق شده نهایی جدول ۴ نیمه تمام را در نظر بگیرید؛ از آنجا که نقطه‌ای قبل از نقطه اول وجود ندارد، برای نقاط منتخب نقطه GPS شماره ۱ مقادیر $f[c_i^j]$ برابر با مقدار N هریک از آنها در نظر گرفته می‌شود.

در ادامه برای محاسبه هر یک از $f[c_i^j]$ مقادیر $f[c_{i-1}^j]$ مبنای محاسبات قرار می‌گیرد. به عنوان مثال برای محاسبه مقدار $f[c_1^j]$ به صورت زیر عمل می‌شود:

۶- نتایج پیاده سازی الگوریتم

در این مقاله با توجه به نرخ نمونه برداری پایین، هر ۲ دقیقه یک نقطه برداشت می‌شود. با این نرخ نمونه برداری و با توجه به سرعت متوسط اتوبوس ۳۰ کیلومتر بر ساعت متوسط فاصله بین دو نقطه در طول حدود یک کیلومتر خواهد بود. در ادامه نتایج به کارگیری الگوریتم را برای یک نمونه ردیابی شامل پانزده نقطه مربوط به خط پایانه لاله-پایانه افشار بیان می‌گردد.

همانطور که در شکل ۸ مشخص است نقاط نمونه برداری با رنگ قرمز بر روی مسیر واقعی شبکه (مسیر سبز رنگ) منطبق نیست. در ادامه نحوه محاسبه بیشترین رتبه $f[c_i^j]$ و تعیین دنباله تطبیق شده نهایی برای سه نقطه متوالی از ردیابی GPS به عنوان نمونه تشریح می‌گردد. در ابتدا بر اساس

بر اساس طول مطابق با رابطه ۱۰ تعیین می‌گردد که در آن از تقسیم تعداد قطعات جاده درست تطبیق شده بر تعداد کل قطعات ردیابی و از تقسیم مجموع طول قطعات جاده درست تطبیق شده بر کل طول مسیر ردیابی محاسبه می‌شود.

$$A_N = \frac{N_C}{N_A}, A_L = \frac{L_C}{L_A} \quad (10)$$

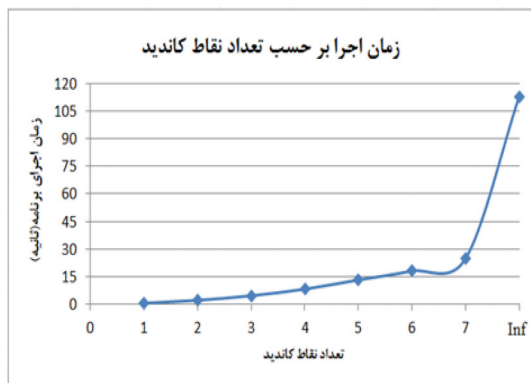
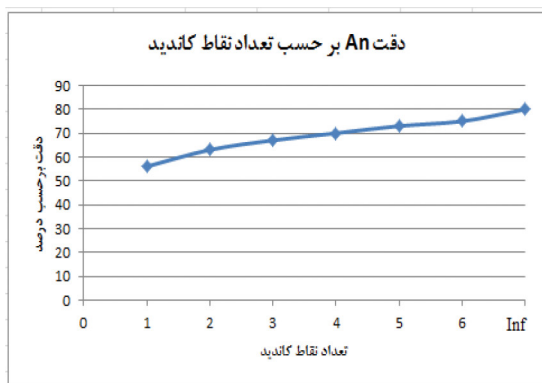
شکل ۹ تاثیر ماکزیمم تعداد نقاط کاندید را بر زمان اجرای برنامه نشان می‌دهد. در اینجا منظور از Inf این است که در تعداد نقاط کاندید محدودیتی وجود نداشته باشد. همانطور که در شکلها مشاهده می‌شود با افزایش تعداد نقاط کاندید دقت الگوریتم بالا می‌رود اما افزایش تعداد این نقاط باعث پیچیده شدن محاسبات مربوط به کوتاهترین مسیر و بالا رفتن زمان اجرای برنامه می‌گردد. در حالتی که محدودیتی برای تعداد نقاط کاندید وجود نداشت گاهی تعداد نقاط تا ۲۵ عدد نیز می‌رسید و این امر حجم محاسباتی بالایی را تحمیل می‌کند. با توجه

$$\begin{aligned} f[c_1^1] &= \max \{f[c_1^1] + F(c_1^1 \rightarrow c_1^2), f[c_1^2] \\ &+ F(c_1^2 \rightarrow c_1^3), f[c_1^3] + F(c_1^3 \rightarrow c_1^4), \\ &f[c_1^4] + F(c_1^4 \rightarrow c_1^5), f[c_1^5] + F(c_1^5 \rightarrow c_1^6)\} \\ &= \max \{0.023015, 0.008710, \\ &0.0002082, 0.000861, 0.000813\} \\ &= 0.023015 \end{aligned}$$

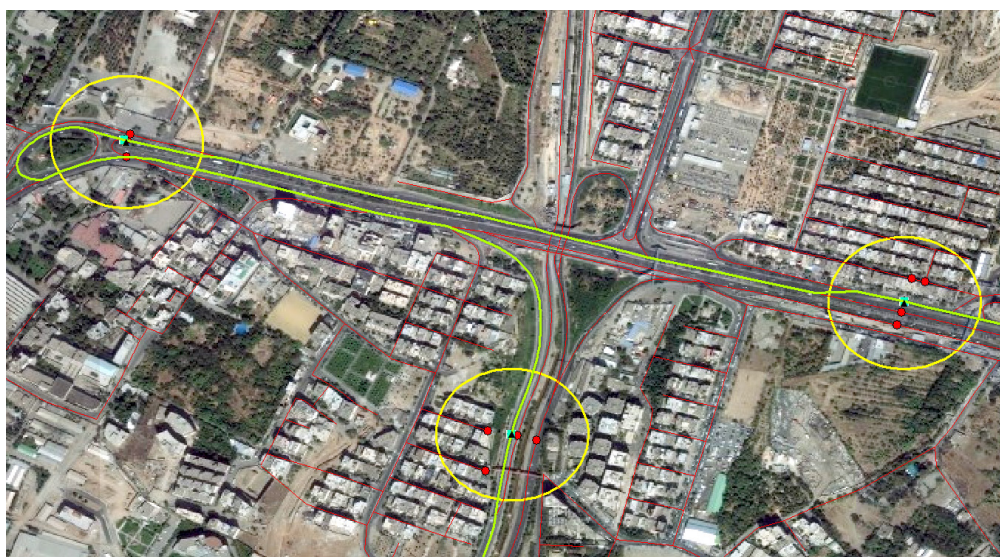
پس از تکمیل مقادیر $f[c_1^1]$ به صورت جدول ۵ دنباله ای از نقاط منتخب با بیشترین مقدار $f[]$ به عنوان دنباله تطبیق شده نهایی مشخص می‌گردد یعنی $c_1^1, c_1^2, c_1^3, c_1^4, c_1^5$.

۷- ارزیابی نتایج

ارزیابی نتایج بر حسب زمان اجرای برنامه و کیفیت تطبیق صورت می‌گیرد. زمان اجرا بر مبنای زمان به دست آمده از اجرای برنامه در محیط پایگاه داده PostGIS محاسبه می‌گردد. کیفیت تطبیق با دو پارامتر دقت براساس تعداد A_N و دقت



شکل ۹: زمان اجرا و دقت بر حسب تعداد نقاط کاندید



شکل ۱۰: نمونه نقاط خروجی تطبیق یافته

به نمودارهای شکل ۱۳ مشخص است که نظر گرفتن ماکزیمم تعداد نقاط کاندید ۵ مناسب است. و با این فرض مقادیر دقت به صورت زیر به دست آمد: $A_N = 73.3\%$, $A_L = 67.12\%$

شکل ۱۰ نتایج به کارگیری این الگوریتم را نشان می‌دهد که در آن مسیر اصلی شبکه جاده با رنگ سبز، نقطه نمونه برداری GPS با علامت مثلث، نقاط قرمز رنگی که با علامت دایره نشان داده شده‌اند نقاط کاندید، و نقاط مربع شکل نیز نقطه کاندید با بیشترین وزن اختصاص داده شده است و در نهایت به عنوان بهترین نقطه انتخاب شده است. همانطور که در شکل ۱۴ مشاهده می‌شود هر سه نقطه مربع شکل کاملاً بر مسیر شبکه جاده قرار دارند و این امر حاکی از آن است که عملیات تطبیق نقشه به درستی صورت پذیرفته است.

۸- نتیجه‌گیری

درسالهای اخیر استفاده از تلفن‌های هوشمند و PDAهای مجهز به GPS افزایش فراوانی یافته‌اند. گسترش این دستگاه‌ها افزایش مجموعه عظیمی از ردیابی‌های GPS که اکثراً با نرخ نمونه برداری پایین در حد دوتا پنج دقیقه و همراه با خطا هستند، را به دنبال داشته‌است. برنامه‌های کاربردی، مانند برنامه‌ریز مسیر، مسیر یاب، تجزیه و تحلیل جریان ترافیک، شبکه‌های اجتماعی مکان‌مبنا، برای دستیابی بهتر به کیفیت بالاتر نیازمند داده‌های ردیابی GPS دقیق به نحوی که بر شبکه جاده منطبق باشد، نیاز دارند.

از اینرو در این مقاله یک الگوریتم تطبیق نقشه مکانی-زمانی به منظور تطبیق داده‌های ردیابی GPS اتوبوس‌های هشت خط در مناطق یک تا چهار شهرداری تهران با نرخ نمونه‌برداری پایین (دو دقیقه) با شبکه معابر واقعی به کار گرفته شد. در این الگوریتم ابتدا برای هر نقطه نمونه‌برداری نقاطی به عنوان نقاط کاندید در نظر گرفته می‌شود و از میان آنها نقاطی که بیشترین تطابق و نزدیکی با واقعیت را دارند انتخاب می‌شوند. رؤس گراف کاندید شامل مجموعه نقاط کاندید برای هر مشاهده GPS و یالهای آن شامل مجموعه کوتاهترین مسیرها بین هر دو نقطه کاندید همسایه می‌باشد. بر اساس نتایج تحلیلهای مکانی و زمانی به نقاط و یالهای گراف وزنی اختصاص می‌یابد. سپس گراف کاندید با وزنهای اختصاص یافته ارزیابی می‌گردد و ردیابی GPS با مسیری که دارای بالاترین وزن در گراف کاندید است، تطبیق می‌یابد.

این الگوریتم از تحلیل‌های مکانی و زمانی به منظور ساخت گراف کاندید که در آن مجموعه‌ای از نتایج منطبق

شده با بالاترین مجموع رتبه به عنوان نتیجه تطبیق نقشه تعیین می‌شود، استفاده نموده است. نتایج حاصل از این الگوریتم نشان می‌دهد این الگوریتم در تطبیق نقشه برای ردیابی‌های GPS با نرخ پایین با دقت بالایی در حدود هفتاد درصد صحیح عمل می‌کند. مقادیر پارامترهای دقت $A_N = 73.3\%$, $A_L = 67.12\%$ می‌باشد. از ویژگی‌های بارز این الگوریتم در نظر گرفتن همزمان توپولوژی و ویژگی‌های مکانی ویژگی زمانی می‌باشد. اما به دلیل آنکه برای تطبیق هر نقطه نمونه برداری فقط یک نقطه قبل از آن را در نظر می‌گیرد و برای نقطه اول نقطه ماقبل وجود ندارد، این الگوریتم به نقطه شروع وابستگی زیادی دارد. با توجه به اینکه در داده‌های با نرخ نمونه برداری پایین (دو دقیقه) مسافت زیادی توسط ماشین پیموده می‌شود و همچنین به منظور یافتن بخش‌هایی از راه بین دو نقطه متوالی GPS از روشهای کوتاهترین مسیر استفاده می‌شود، الگوریتم در گردش‌های U شکل ممکن است دچار مشکل شود و باعث کاهش دقت الگوریتم گردد که در کارهای آتی به بهبود این امر پرداخته خواهد شد.

۹- مراجع

- ۱- ا. دوباش، «تطبیق مشاهدات جی‌پی‌اس بر نقشه به صورت برون خط»، کارشناسی ارشد، مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ۱۳۹۳.
- 2- I. Rauschert, P. Agrawal, R. Sharma, S. Fuhrmann, I. Brewer, and A. MacEachren, "Designing a human-centered, multimodal GIS interface to support emergency management," in Proceedings of the 10th ACM international symposium on Advances in geographic information systems, McLean, Virginia, USA, 2002, pp. 119-124.
- 3- M. A. Quddus, W. Y. Ochieng, and R. B. Noland, "Current map-matching algorithms for transport applications: State-of-the art and future research directions," Transportation Research Part C: Emerging Technologies, vol. 15, pp. 312-328, 2007.
- 4- S. Liu, Y. Liu, L. M. Ni, J. Fan, and M. Li, "Towards mobility-based clustering," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 2010, pp. 919-928.
- 5- H. Zhu, Y. Zhu, M. Li, and L. M. Ni, "Seer: Metropolitan-scale traffic perception based on lossy

- 10- S. Syed, "Development of Map-Aided GPS Algorithms for Vehicle Navigation in Urban Canyons," MSc, Department of Geomatics Engineering, University of Calgary, Department Of Geometrics Engineering, Calgary, Alberta, 2005.
- 11- J. S. Greenfeld, "Matching GPS observations to locations on a digital map," in Transportation Research Board 81st Annual Meeting, 2002.
- 12- S. Brakatsoulas, D. Pfoser, R. Salas, and C. Wenk, "On map-matching vehicle tracking data," in Proceedings of the 31st international conference on Very large data bases, 2005, pp. 853-864.
- 13- S. Kim and J.-H. Kim, "Adaptive fuzzy-network-based C-measure map-matching algorithm for car navigation system," *Industrial Electronics, IEEE Transactions on*, vol. 48, pp. 432-441, 2001.
- 14- S. Syed and M. Cannon, "Fuzzy logic-based map matching algorithm for vehicle navigation system in urban canyons," in ION National Technical Meeting, San Diego, CA, Alberta, Canada, , 2004, pp. 26-28.
- sensory data," in INFOCOM 2009, IEEE, 2009, pp. 217-225.
- 6- Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang, "Map-matching for low-sampling-rate GPS trajectories," in Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2009, pp. 352-361.
- 7- S. Liu, C. Liu, Q. Luo, L. Ni, and R. Krishnan, "Calibrating large scale vehicle trajectory data," in Mobile Data Management (MDM), 2012 IEEE 13th International Conference on, 2012, pp. 222-231.
- 8- J. Yuan, Y. Zheng, C. Zhang, X. Xie, and G.-Z. Sun, "An interactive-voting based map matching algorithm," in Proceedings of the 2010 Eleventh International Conference on Mobile Data Management, 2010, pp. 43-52.
- 9- D. Wang, Z. Wang, X. Li, and Z. Xiao, "A Map Matching Algorithm to Eliminate Miscalculation Based on Low-Sample-Rate Data," in 3rd International Conference on Computer Science and Service System, 2014.

Implement Spatial-Temporal map-matching algorithm to enhance accuracy of the traffic data extracted low sampling rate GPS trajectories

Vahid Shokri¹, Rahim Ali Abbaspour²

1- Master student, Dept. of Surveying and Geomatics Eng., Engineering Faculty, University of Tehran, 16th Azar

2- Associate Prof of Dept. of Surveying and Geomatics Eng., Engineering Faculty, University of Tehran, 16th

Abstract

With the increasing popularity of devices equipped with a Global Positioning System, users will be able to track moving objects such as cars, animals, and people. GPS sensors data has positioning and sampling error. On the other hand, traffic applications, such as traffic data collection requires accurate tracking GPS based on road network. Since most of the percentage of tracking data GPS, particularly GPS modules in smart phones are low-end and low accuracy, a sophisticated and reliable map matching algorithm is crucial for these location-based services. This study aimed to analyze and implement an spatial-temporal map matching algorithm for low sampling rates GPS trajectories. Therefore, in this article GPS tracking data related to eight bus lines in the range of 1 to 4 municipalities with a sampling rate of two minutes will be used. The algorithm employs spatial and temporal analysis to generate a candidate graph, from which a sequence of matched results with highest sum of score is identified as the matching result. The experiment results demonstrate that ST-matching algorithm for low-sampling trajectories significantly performs good and the accuracy is about seventy percent.

Key words: Traffic data collection, GPS trajectory, map matching, road network, low-sampling rate.