

Tratamiento de multitrazas GNSS 3D para la obtención de ejes medios

Francisco Javier Ariza-López^a, Antonio T. Mozas-Calvache^a, Paula Gil de la Vega^a

^a *Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría, Universidad de Jaén
E-mail: fjariza@ujaen.es.*

Resumen

El uso de los sistemas de navegación se ha difundido mucho y la explotación de sus datos es compleja debido tanto al volumen como a su variabilidad. Este trabajo de revisión presenta una definición de los elementos (trazas) y un análisis de las fuentes de variabilidad, así como los criterios usuales de filtrado de los datos originales. Otro aspecto presentado son los atípicos, éstos se definen y clasifican para el caso de trazas GNSS. La obtención del eje medio se aborda en dos pasos, primero se presentan algoritmos que, a partir de dos polilíneas 3D permiten obtener una solución media y, posteriormente, se presentan otras opciones que permiten obtener una solución de eje medio a partir de un conjunto de trazas.

Palabras clave: minería de datos; trazas GNSS; ejes medios; datos espaciales

1. Introducción

Dentro de los datos espaciales las vías de comunicación terrestre son de interés dado el conjunto de optimizaciones que permiten (tiempo, desplazamiento, energía, flotas, etc.). Por otra parte, esta tipología de elementos es muy abundante en los productos de carácter general (bases topográficas). Tradicionalmente su captura se ha realizado por métodos fotogramétricos, pero en la actualidad hay nuevas posibilidades como son las aportaciones de voluntarios (VGI, Goodchild, 2007), donde los ciudadanos actúan como sensores del mundo real y aportan datos objetivos (p.e. procedentes de sensores GNSS –*Global Navigation Satellite Systems*–) y subjetivos.

En la línea de la VGI es de interés la posibilidad de obtener resultados cartográficos a partir de las multitrazas GNSS que capturan los ciudadanos. Las aportaciones VGI "en bruto" se corresponden con trazas de navegadores GNSS de los vehículos particulares. Se pueden obtener cientos de trazas de calles y carreteras con la ventaja de ofrecer múltiples usos (Biagioni y Eriksson 2012, Quddus y col., 2007): inferencia de mapas de carreteras, procesos de actualización y/o levantamiento automatizado de información de vías de comunicación con rapidez y frecuencia, posibilidad de uso en control posicional, etc. Para aprovechar estas trazas hay que derivar un eje promedio preciso. En este trabajo se presentan retos y soluciones, entre ellos:

- a) las necesidades de filtrado previo de los datos GNSS (p.e. calidad, velocidad, tiempo, etc.),
- b) los problemas que se presentan en este tipo de datos (p.e. atípicos puntuales, atípicos lineales).
- c) las distintas soluciones posibles para la minería y obtención de un eje medio (p.e. rasterización 3D, clusterización de líneas, ajuste mínimo cuadrático, condensado, promediado, etc.),

2. Multitrazas GNSS de navegación

Definimos una traza GNSS de navegación T_i a una secuencia de posiciones 3D, ordenadas en el tiempo, que ocurren en el espacio constreñido de una vía de comunicación determinada (p.e. carretera, calle, etc.), sin auto-intersecciones, y obtenida por un equipo GNSS de bajo coste situado en un vehículo y que es tomada por cualquiera durante un proceso de conducción sin precaución especial a esta captura posicional. Las trazas pueden estar afectadas por múltiples circunstancias, entre otras:

- La geometría instantánea de la constelación de satélites. De ella depende la calidad geométrica de la captura.
- Equipo de captura. Aspectos importantes en este caso son el chip del sistema, la sensibilidad, la frecuencia de captura, el número de canales y las posibilidades de GPS asistido y DGPS (WAAS/EGNOS).
- Configuración del equipo. Los equipos pueden ser configurados en su software atendiendo diversos aspectos como: i) información a almacenar, ii) la frecuencia de captura, iii) el modo de trabajo, iv) el uso de sistemas para la mejora de la exactitud posicional, etc.
- Posición en el vehículo. La posición del sensor es otro elemento clave, pensando en turismo la posición de la antena puede ser exterior, interior (sobre el salpicadero o sobre la batea del maletero).
- Elementos externos. El efecto cañón en las calles estrechas o con edificios altos, la presencia de taludes altos, más o menos cercanos en las carreteras, la presencia de arboledas altas, etc., puede afectar la recepción de la señal.

Definimos un conjunto multitrazas GNSS $CMT_{VT-d(A-B)}=\{...T_i...\}$ a una agrupación de trazas T_i que se corresponden con una ventana temporal VT de interés (p.e. del día 01/01/2014 al 15/01/2014), y que representan el conjunto de desplazamientos $d(A-B)$ entre dos posiciones A y B de interés. Si pensamos en aportaciones de trazas GNSS por voluntarios (p.e. en OSM), este conjunto puede ser muy heterogéneo en cuanto a los elementos que lo componen.

3. Necesidades de filtrado

En todos los trabajos con sistemas de captura de datos por medio de GNSS en navegación y con alta frecuencia se plantea la necesidad del filtrado de los datos, los principales motivos son de tres tipos:

- Reducción del número de elementos. En estos casos es usual quedarse con 1 de cada tantos puntos (p.e. para pasar de 5 hz a 1 hz nos quedamos con 1 de cada 5 puntos de una traza). También se puede aplicar algoritmos de simplificación para eliminar detalles irrelevantes (p.e. Lima y Ferreira, 2009, aplican Douglas-Peucker con $tol = 1m$).

- Calidad de las observaciones individuales (puntos). Se pueden eliminar puntos individualizados a partir de la información relativa a su calidad particular y/o al contexto en el que ocurren.
- Calidad en la traza. Se pueden eliminar conjuntos de puntos que afecten a la traza como elemento con unas propiedades específicas (p.e. continuidad, verosimilitud, exactitud posicional, etc.).

Algunos procesos comunes de filtrado centrados en observaciones individuales (puntos y trazas GNSS) son:

- Eliminación de puntos con mala geometría de los satélites (HDOP, VDOP o GDOP). Por ejemplo, Lima y Ferreira (2009) eliminan puntos con $HDOP > 2$. Afecta a la exactitud posicional.
- Eliminación de puntos con pocos satélites. Lima y Ferreira (2009) eliminan puntos con menos de 5 satélites, en general el criterio asumido es eliminar con menos de 3 para 2D y con menos de 4 para 3D. Es un criterio complementario al anterior. Afecta a la exactitud posicional.
- Eliminación de X puntos de inicio (p.e. 10) tras un salto temporal de T segundos (p.e. 10 s, 20 s) que dependerá de los datos. El propósito es eliminar el ruido existente cuando se fija la posición tras una pérdida de señal.
- Eliminación de puntos que exceden una aceleración/velocidad máxima a la del vehículo de toma. Zhang y col. (2010) toman como criterios de eliminación por velocidad aquellos que exceden 250 km/h en carreteras, y 100 km/h en entornos urbanos. Liu y col. (2012) consideran como velocidad máxima los 120 km/h. Lima y Ferreira (2009) establecen un criterio de velocidad mínima (6 km/h) por debajo del cual eliminan las posiciones pues consideran que las bajas velocidades proceden de observaciones de mala calidad (p.e. inicio) o con ruido (p.e. maniobras lentas).
- Eliminación de puntos por saltos en distancia, tiempo o ángulo. Se consideran sospechosos los puntos de una secuencia en la que se da un salto excesivo. Así, Zhang y col. (2010) consideran 300 m pero en Agamennoni y col. (2010) es 100 m. También se pueden considerar redundantes aquellos puntos demasiado cercanos, p.e. en Cao y Krumm (2009) 30 m (10 m si hay cambio de dirección mayor de 10 grados). Como criterio angular de eliminación, Zhang y col. (2010) proponen un salto angular superior a 45° . El salto en distancia también puede analizarse desde la perspectiva del tiempo, Lima y Ferreira (2009) que eliminan puntos con separación temporal mayor a 7 s y Agamennoni y col. (2010) se considera más de 3 s.
- Rotura de la traza por salto espacial o temporal. Fathi y Krumm (2010) y Cao y Krumm (2009) rompen la traza si hay salto temporal mayor de 10 s. Los primeros también rompen la traza si la velocidad < 5 mph ó > 90 mph. En Cao y Krumm (2009) también es criterio de rotura un salto de más de 100 m entre puntos adyacentes.
- Eliminación de trazas respecto a una referencia. Zhang y col. (2010) aplican los siguientes criterios de eliminación de trazas: a) si hay dirección, se eliminan las de dirección contraria, b) si hay una diferencia angular mayor de 20 grados, c) si la distancia supera 30 m.

4. Los atípicos

Se ha indicado que cada T_i puede ser considerada como un suceso aleatorio y, así, como en cualquier conjunto de datos aleatorios, en un CMT se pueden encontrar observaciones que presentan comportamientos extraños o atípicos. A modo de ejemplo, la Figura 1 presenta un CMT con algunas trazas atípicas.

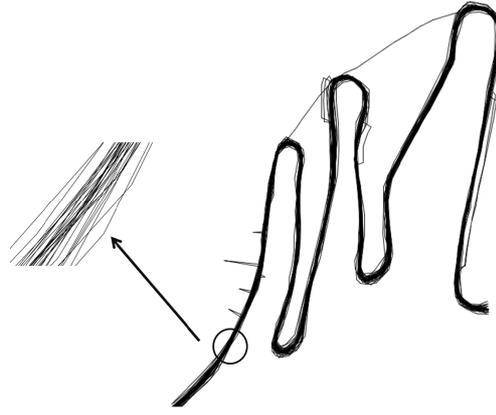


Fig. 1. Trazas GNSS atípicas

Una observación atípica es aquella que se distancia sensiblemente de otras observaciones y que genera sospechas de haber sido generada por un mecanismo distinto (Hawkins 1980). Según la Wikipedia (2014) la decisión sobre si una observación es atípica o no es, en última instancia, un ejercicio subjetivo. Consideramos que cualquier T_i sigue un proceso de generación "normal", y que las trazas atípicas T_i' son aquellas que proceden de un mecanismo extraño. Este mecanismo no tiene por qué ser único (p.e. afectaciones ionosféricas, situaciones de oclusión, multicamino, etc.). Al observar trazas GNSS los atípicos posicionales pueden ocurrir como (Figura 2):

- Puntos. Se trata de puntos aislados dentro de una traza GNSS (Figura 2.a).
- Secciones. Es una secuencia seguida de puntos de una traza que presentan un sesgo evidente respecto al conjunto mayoritario de trazas (Figura 2.b).
- Trazas. Es una traza que en toda su longitud, o en gran parte de ella, puede ser considerada como atípico al separarse del conjunto mayoritario de trazas (Figura 2.c).

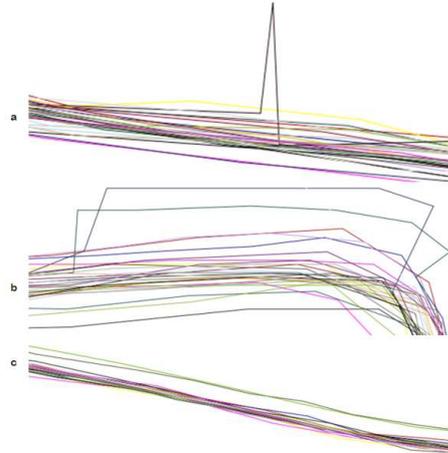


Fig. 2. Categorías de atípicos en trazas GNSS: a) puntos, b) secciones, c) trazas

Las tres categorías indicadas pueden darse tanto en 2D como en 3D. El proceso de detección y eliminación de atípicos más sencillo es equivalente a los que se aplican a datos unidimensionales (p.e. método de las k-sigmas).

5. Soluciones para la determinación del eje medio de un par de trazas

La determinación del eje medio de un CMT puede alcanzarse por un proceso iterativo a partir de parejas de trazas. Este proceso es interés en Producción Cartográfica y en Sistemas de Información Geográfica, donde se puede encontrar abundante bibliografía. Esta es la base de la fusión de geometrías y no siempre es directa pues se necesita previamente tener determinadas parejas de vértices homólogos o puntos medios. Por ello tiene interés conocer algunos algoritmos para conseguirlo:

- Segmento deslizante. Se basa en el concepto de homotopía. Dadas dos trazas se establece un sistema local tal que la longitud se considera en el intervalo $[0, 1]$. De esta forma el eje medio está en la mitad de un segmento deslizante de tamaño variable que se discurre uniendo posiciones homólogas.
- Distancia mínima. Dada una traza **A** y un conjunto de puntos pertenecientes a esa traza (p.e. vértices y puntos equiespaciados entre los vértices), se considera el conjunto el segmentos de distancia mínima hasta otra traza **B** y el punto medio. Rogers y col. (1999) proponen que la interpolación sea ponderada según criterios de calidad.
- Triangulación. Se triangula entre las dos trazas, en este caso el eje medio se interpola a partir de las aristas de los triángulos (Figura 3.a). Es una solución robusta si se filtra adecuadamente las trazas originales para evitar triángulos muy alargados y se construye adecuadamente.
- Transformación de eje medio (esqueletización topológica o *straight skeleton*). En la versión vectorial puede entenderse como la intersección de un conjunto de orlados crecientes realizado a ambas trazas.
- Distancia de Frèchet. Permite determinar puntos homólogos entre geometrías. El eje medio pasa por la mitad del segmento que une puntos homólogos (Figura 3.b).

Un problema que presentan estos métodos de cálculo del eje medio en el caso del trabajo con trazas GNSS es que, debido a su dispar procedencia, estas trazas no tienen siempre la misma calidad posicional, ni una traza frente a otra, ni todos los puntos que conforman una traza dada. La inclusión de este aspecto es necesaria, pero no es sencilla en todos los casos aunque, como se ha indicado anteriormente, ya existen algunas propuestas (p.e. Rogers y col., 1999).

Con independencia de la opción adoptada para la obtención del eje medio, la solución final podrá ser filtrada (p.e. Douglas Peucker) y suavizada, por ejemplo, con un ajuste por splines (Hasberg y Hensel, 2008).

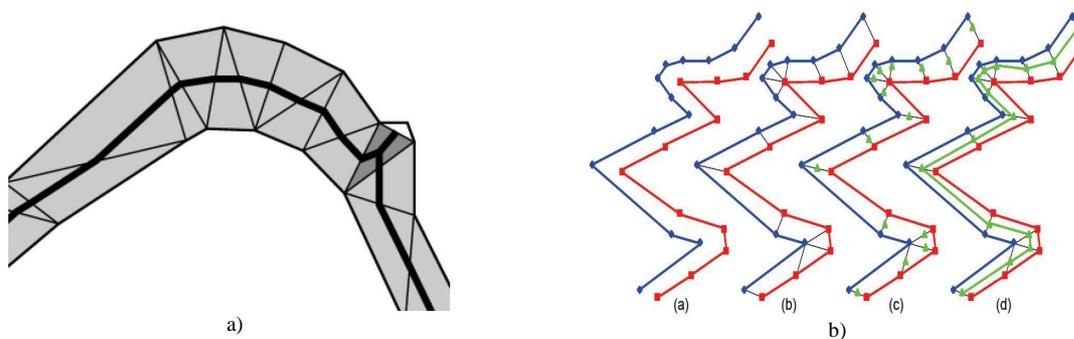


Figura 3. Algoritmos de determinación de eje medio: a) triangulación, b) mediatriz de segmentos entre vértices homólogos

6. Soluciones para la determinación del eje medio de un conjunto de trazas

La determinación del eje medio de un CMT de n trazas es un proceso de minería de datos. Este tema se ha tratado mayoritariamente en 2D pero todas las soluciones ya desarrolladas son extensibles a 3D. Algunas de las opciones se pueden basar en la iteración de las soluciones expuestas en el apartado 5. Otras opciones son:

- Rasterización (*kernel density estimation*). Se rasterizan las trazas en una matriz de celdas 3d (*voxels*) (Figura 4.a), posteriormente se aplica un filtro de paso bajo, luego se vectoriza y finalmente se filtra. Se ha aplicado con éxito en 2D por Davies y col. (2006), Chen y Cheng (2008). También en 2D Shi (2009) describe el uso de morfología matemática.
- Inferencia geométrica. Hay técnicas estadísticas para determinar geometrías, un caso son las curvas principales (Figura 4.b) para las que hay varias definiciones (p.e. Hastie y Stuetzle 1989; Kégl y col. 2000). Una definición es: curva de longitud dada L que minimiza las distancias a un conjunto X . Agamennoni y col. (2010) la aplican a trazas GNSS 2D.
- Agrupamiento (cluster). Se aplican técnicas estadísticas de agrupamiento. A partir de unas semillas, mediante un proceso iterativo convergente se determinan grupos y sus posiciones centrales:
 - K-medias (Edelkamp y Schrödl, 2003). Básicamente consiste en realizar un conjunto de agrupamientos a partir de unas semillas distribuidas en el espacio a lo largo de las trazas. Esta propuesta se plantea como una solución a un problema (detección de carriles o de calzadas) ligeramente distinto al que aquí se presenta (eje medio de un grupo de trazas).
 - Núcleos espaciados. Modificación del algoritmo de las k-medias que se centra en la obtención del eje medio, este algoritmo se describe en otra comunicación (ver Gil de la Vega y col., 2014).
 - Planos espaciados (Zhang y col. 2010). Sobre una traza "semilla", se van tomando cortes perpendiculares al eje longitudinal sobre los que se determinan las intersecciones del resto de las trazas.
- Condensación (Mozas-Calvache y Ariza-López, 2014). Aleatoriamente se van promediando dos a dos las trazas del CMT inicial, se genera un conjunto modificado (CMMT). El proceso se itera en los sucesivos CMMT un cierto número de veces o tal que se alcanza una tolerancia.
- Modelos físicos (Cao y Krumm, 2009). Hay un modelo de fuerzas contrapuestas sobre los vértices, una gravitatoria (ley gaussiana invertida) y otra lineal (ley de Hook). Wang y col. (2011) proponen ponderar para obtener una solución más eficaz y eficiente.
- Teoría de la información (Chazal y col., 2011). Se toma una traza semilla del CMT y se lleva a un espacio de fases de mayor dimensión (EFMD), en el EFMD la traza se mueve hacia los elementos del CMT y tras este ajuste se recupera la traza después de promediar sus coordenadas en el EFMD.

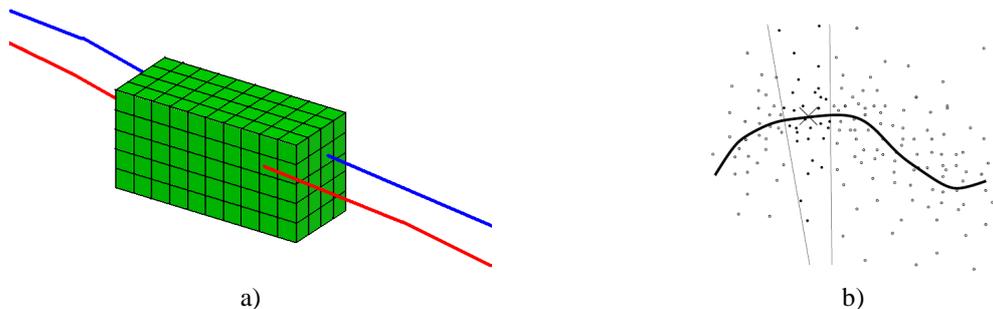


Fig. 4. Idea sobre los algoritmos de rasterización (a) y curvas principales (b)

7. Conclusiones

En este trabajo se han presentado diversos aspectos relacionados con la minería de trazas GNSS 3D: definición de términos, aproximación a la variabilidad de los datos, definición de tipos de atípicos y presentación de algoritmos para obtener ejes medios, aportando numerosas referencias actuales. Este trabajo facilita el camino a aquellos que quieran iniciarse en la minería de multitrizas GNSS 3D al dar una visión general del problema y de las soluciones aplicadas y aplicables.

Agradecimientos

Este trabajo se ha realizado dentro del proyecto de investigación “Evaluación 3D de elementos lineales de información geográfica” del Ministerio de Ciencia e Innovación (BIA2011-23271) cofinanciado por el Fondo Europeo de Desarrollo Regional.

Referencias

- Agamennoni, G., Nieto, J., Nebot, E. (2010). Technical Report: Inference of Principal Roads Paths Using GPS Data.
- Axhausen, K.W., Schönfelder, Wolf, J., Oliveira, M., Samaga, U. (2003). 80 Weeks of GPS traces: approaches to enriching the trip information. Institute for Transport Planning and System, Zurich
- Biagioni, J., Eriksson, J. (2012). Inferring road maps from GPS traces: survey and comparative evaluation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2291, pages 61-71, 2012.
- Cao, L., Krumm, J. (2009). From GPS Traces to a Routable Road Map. *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 3-12.
- Chazal, F. (2011). Data-Driven trajectory smoothing. *ACM SIGSPATIAL GIS' 11*, Chicago.
- Chen C., Cheng, Y. (2008). Roads Digital Map Generation with Multi-track GPS Data. *In Education Technology and Training, 2008*, pp: 508 - 511.
- Davies, JJ., Beresford, AR., Hopper, A. (2006). Scalable, Distributed, Real-Time Map Generation. *IEEE Pervasive computing*, Vol. 5, No. 4, pp:47-54.
- Edelkamp, S., Schrödl, S. (2003). Route planning and map inference with Global Position Traces. *Comp. Sci. In Perspective, LNCS 2598*, pp:128-151.
- Fathi, A., Krumm, J. (2010) Detecting Road Intersections from GPS Traces. *GIScience 2010, 6th International Conference on Geographic Information Science*, Zurich.
- Gil de la Vega, P., Ariza-López, F.J., Mozas-Calvache, A.T. (2014). Nucleos espaciados: algoritmo para la determinación de un eje medio a partir de multitrizas GNSS. *XVI Congreso Nacional de tecnologías de la Información Geográfica*, Alicante, España.
- Hasberg, C., Hensel, S. (2008). Online-estimation of road map elements using spline curves. *In Fusion, IEEE* p:821-827.
- Hastie, T., Stuetzle, W., (1989). Principal Curves. *Journal of the American Statistical Association*, vol 89(406):502-516.
- Hawkins, D. (1980). *Identification of Outliers*. Chapman and Hall.
- Kégl, B., Krzyzak, A., Linder, T., Zenger, K. (1999). A polygonal Line Algorithm for Constructing Principal Curves. *Advances in Neuronal Information Processing Systems*, vol 11, pp:501-507.
- Lima, F., Ferreira, M. (2009) Mining spatial data from GPS traces for automatic road network extraction. *En 6th International symposium on mobile mapping technology*, Presidente Prudente, Brazil.
- Liu, X., Biagioni, J., Eriksson, J., Wang, Y., Forman, G., Zhu., Y. (2012). Mining Large-Scale, Sparse GPS traces for Map Inference: Comparison of Approaches. *KDD'12*, China.
- Mozas-Calvache, AT, Ariza-López, FJ. (2014). Iterative method for obtaining a mean 3D axis from a set of GNSS traces for use in positional controls. Pendiente de aceptación.
- Quddus, M., Ochieng, W., Noland, R. (2007). Current map-matching algorithms for transport applications: state-of-the art and future research directions. *Transportation Research Part C* (15).312-328.
- Rogers, S., Langley, P., Wilson C., (1999). Mining GPS Data to Augmented Road Models. *En KDD-99 San Diego Ca, USA*
- Shi, W. (2009). Automatic generation of road network map from massive GPS vehicle trajectories.
- Wang, J., Rui, X., Song, X., Wang, C., Tang, L., Li, C., Raghvan, V. (2011). A Weighted Clustering Algorithm for Clarifying Vehicle GPS Traces. *In IGRSS*, pp: 2949-2952.
- Wikipedia (2014). Outlier. <http://en.wikipedia.org/wiki/Outlier>
- Zhang, L., Thiemann, F., Sester, M. (2010). Integration of GPS traces with road map. *En ACM SIGSPATIAL'10*, San José, USA.