

Medidas de similitud entre caminos: Una propuesta

J. A. Álvarez-García^a, L. González-Abril^b, J. A. Ortega^a y F. Velasco^b

^a Dpto. Lenguajes y Sistemas Informáticos. Universidad de Sevilla. Avda Reina Mercedes s/n.

^b Dpto. de Economía Aplicada I. Universidad de Sevilla. Avda. Ramón y Cajal, s/n.

E-mail: {jaalvarez,luisgon}@us.es, ortega@lsi.us.es,velasco@us.es

Abstract

En este trabajo se presentan y describen diferentes medidas de similitud específicamente diseñadas para caminos realizados en exteriores y registrados a través de dispositivos receptores de GPS. Dado que la precisión del "Sistema de Posicionamiento Global" oscila entre 5 y 10 metros en exteriores y a veces se produce el efecto cañón con la siguiente pérdida de datos validados, nuestras medidas se adaptan a este tipo de problemas y permite comparaciones efectivas. Además también se proponen técnicas de pre-procesado para reducir la cantidad de puntos almacenados de manera que no se pierda la información fundamental del recorrido.

1. Introducción

Los avances en comunicaciones inalámbricas y tecnologías de computación ubicua están permitiendo desarrollar aplicaciones con conocimiento del contexto y localización cada vez más complejas. Además, tras la orden de la Comisión de Comunicaciones Federal de Estados Unidos (FCC) y la recomendación europea E112 que permite a los proveedores inalámbricos localizar dentro de decenas de metros a usuarios que envían llamadas broadcast de emergencia E911/E112, la investigación sobre localización y los servicios que se pueden ofrecer gracias a ésta está en auge.

La facilidad para recuperar y tratar datos de localización, por ejemplo con receptores GPS, permite trabajar y gestionar en tiempo real objetos en movimiento. Enmarcado en este ámbito, nuestra investigación se centra fundamentalmente en la comparación de caminos finalizados con aquéllos parcialmente recorridos por parte de los usuarios. De este modo podremos predecir los lugares a los que se dirige un determinado usuario y tomar algunas decisiones al respecto. Por ejemplo, si el ponente de una determinada sesión de un congreso se dirige a la sala donde presentará sus trabajos, y ese recorrido ha sido generado anteriormente, el proyector de la sala se encenderá, las persianas se bajarán y las luces reducirán su intensidad antes de que llegue. Para este fin debemos cuantificar el parecido entre diferentes caminos y de aquí dar una medida de similitud y/o de distancia.

La búsqueda de similitudes ha sido estudiada en diferentes contextos como seguimiento de movimiento en videos [Keogh *et al.*, 2004; Little and Gu, 2001], análisis de series temporales [Cuberos, 2005; Faloutsos *et al.*, 1994] y trayectorias [Lin, 2006; Lin and Su, 2005; Chen *et al.*, 2005; Lee *et al.*, 2000; Vlachos *et al.*, 2004]. Debido a la imposibilidad de tener una medida de "similitud" universal y a la

diversidad de aplicaciones existentes, los resultados obtenidos a partir de una medida de similitud previamente dada para un problema concreto no son satisfactorios en la mayoría de contextos y por ello, pensamos que cada problema debe ser investigado individualmente y de las características de éste, obtener una adecuada medida de similitud. Por ello, en el presente trabajo presentamos dos propuestas de medir parecidos entre caminos obtenidos a partir de dispositivos GPS.

El presente documento se estructura de la siguiente forma: en la Sección 2 se presentan los conceptos básicos, se proponen dos formas de filtrar caminos y se define una relación de inclusión entre caminos. A partir de esta inclusión se definen dos medidas de similitud en las Secciones 3 y 4. En la Sección 5 se muestra una medida de similitud clásica y se comenta sus características. Se finaliza con unos comentarios sobre lo desarrollado y las referencias.

2. Caminos entre dos puntos. Similitud

Sea $R^\delta(AB)$ un conjunto de caminos entre dos puntos A y B de radio δ , i.e., $R^\delta(AB) = \{R_1^\delta, R_2^\delta, \dots\}$ donde

$$R_i^\delta = \{B^\delta(x_0), B^\delta(x_1), \dots, B^\delta(x_{n_i})\}$$

siendo $B^\delta(x_i) = \{x \in \mathbb{R}^2 : d(x, x_i) < \delta\}$, $d(\cdot, \cdot)$ una distancia, $A = x_0$, $B = x_{n_i}$, y $\delta > 0$.

Respecto al conjunto de caminos es necesario hacer algunos comentarios:

- El conjunto $R^\delta(AB)$ puede ser infinito no numerable, sin embargo para estudios prácticos e incluso teóricos podemos suponer que es a lo sumo infinito numerable.
- No es estrictamente necesario suponer que $x_0 = A$ y $x_{n_i} = B$, sería suficiente que $A \in B^\delta(x_0)$ y $B \in B^\delta(x_{n_i})$. De hecho los datos obtenidos de la constelación de satélites una vez activado el dispositivo receptor GPS tarda un tiempo en recibir la información validada de un número suficiente de satélites y generalmente si se enciende el dispositivo estando en movimiento el primer punto válido diferirá del inicial. Precisamente esta es la una de las ideas que subyace en la definición de bolas en lugar de simples puntos. Otra de las características que hacen más eficiente el uso de bolas en vez de puntos es la precisión. Aunque en el año 2000 se suprimió la "disponibilidad selectiva" que introducía un error aleatorio, la precisión de los receptores GPS actuales oscilan entre los 5 y los 10 metros.
- Los caminos R_i^δ se suponen que están formados por un número finito de bolas (después del proceso de filtrado) ya que de otra forma se estaría indicando que nunca se llega al destino, lo cual no resulta lógico ni operativo.

- Realmente, el conjunto de caminos se identifica mediante puntos ya que el δ se puede elegir a posteriori (siempre en función de la precisión del medidor de posición). Incluso δ puede ser variable si se considera un recorrido mixto en el sentido de utilizar diferentes medios de locomoción (andar, bicicletas, coche,...). Como tema a tratar, podríamos considerar δ como una función de la velocidad de desplazamiento puntual, lo cual pasaría a engrosar el número de datos por punto obtenido (actualmente únicamente tratamos la longitud, latitud e instante de tiempo ¹ en el que se recupera el dato).

Respecto a los caminos $R_i^\delta \in R^\delta(AB)$, puede ocurrir que tengan una gran cantidad de bolas, algunas de las cuales pueden ser eliminadas sin disminuir la precisión en el posterior tratamiento. Así pues, el primer problema que debemos solucionar es cómo reducir el número de bolas en cada camino. Para este fin, nosotros proponemos dos tipos de filtrados:

Filtrado 1: Dado un valor ϵ definimos el ϵ -filtrado de un camino como sigue: Sea $R \in R^\delta(AB)$, si existe un índice i tal que $d(x_i, x_{i+1}) \leq \epsilon$ entonces $B^\delta(x_{i+1})$ se elimina ² de R (ver Figura 1). Como consecuencia, una vez llevado a

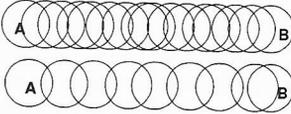


Figura 1: Ejemplo de un camino ϵ -filtrado.

cabo el ϵ -filtrado, $d(x_i, x_{i+1}) > \epsilon$ para todo camino definido en $R^\delta(AB)$. Respecto al valor de ϵ , éste debe ser menor que el valor de δ con objeto de que las bolas no sean disjuntas (aunque en la figura anterior vemos representadas las bolas como circunferencias, también podrían representarse como rectángulos si en vez de distancia euclídea usamos la distancia del máximo). Evidentemente de esta forma se consigue no tener señales excesivamente cercanas unas de otras.

Por otro lado, para evitar bolas disjuntas dentro de cada recorrido, este filtro no debe ser usado cuando $d(x_i, x_{i+1}) \leq \epsilon$ y $d(x_i, x_{i+2}) > \delta$. Aún así pueden existir regiones de bolas disjuntas si los puntos muestreados distan entre sí una distancia mayor a δ , un caso particular es cuando se pierde la conexión de manera momentánea con la constelación de satélites (un túnel, efecto cañón, etc.). La solución en ese caso es elegir un valor mayor de δ , eliminar este camino del conjunto de trabajo, o en casos como el túnel (que se sabe la entrada y salida) elegir puntos intermedios u otra solución.

Respecto a tener algún tipo de argumento para la elección de los valores de ϵ y δ , indicar que si las mediciones se hacen en trayectos de coche, los valores de ϵ y δ deben ser mayores que si se hacen andando o en bicicleta (de ahí, el elegir δ en función de la velocidad de desplazamiento).

Con este filtrado se evitan los problemas de atascos, la espera ante semáforos y otros situaciones en las que se realizan mediciones muy cercanas unas a otras. Sin embargo, el filtrado no evita los bucles (situación común en la

búsqueda de aparcamiento cuando se conduce un vehículo). Una solución para este problema es:

Filtrado 1a: Sea $R \in R^\delta(AB)$, si dado un índice i existe un índice j con $j > i$ tal que $d(x_i, x_j) \leq \epsilon$ entonces las bolas $B^\delta(x_{i+1}), \dots, B^\delta(x_j)$ son eliminadas de R (ver Figura 2).

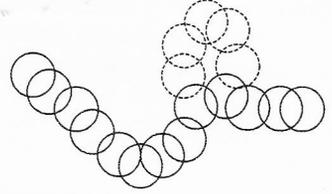


Figura 2: Ejemplo de recorrido con bucle filtrado. El camino inicial está dibujado con trazos discontinuos y la filtrada aparece ya con trazos continuos.

Nota 2.1 Evidentemente, este filtrado elimina los bucles pero claramente pueden considerarse situaciones en las que se desee recoger este hecho para lo cual el primer filtrado sería el adecuado.

Filtrado 2: En los anteriores filtrados se debe elegir a priori un valor para ϵ . Una forma de evitar tener que realizar este supuesto queda recogida en el siguiente pseudocódigo

Dado el conjunto de puntos pertenecientes a un recorrido $\{x_1, \dots, x_n\}$ y el radio de las bolas δ
 $A = \{x_i\}; i = 1;$

Mientras $i \neq n$

$$A_i = \{x_j : d(x_i, x_j) < \delta, j = i + 1, \dots, n\}$$

$$\text{si } A_i = \emptyset \text{ entonces } A = A \cup \{x_{i+1}\} \text{ y } i = i + 1$$

(*1)

$$\text{en otro caso, } A = A \cup \max A_i \text{ y } i = i + 1 \quad (*2)$$

(*1) Sería el caso de una pérdida de conexión

(*2) Se selecciona el índice del punto más alejado del x_i pero a una distancia inferior a δ

Sea desde ahora, $R_1 = \{B_{1j}\}_{j=1}^n$ con $B_{1j} = B^\delta(x_{1j})$ y $R_2 = \{B_{2j}\}_{j=1}^m$ con $B_{2j} = B^\delta(x_{2j})$, donde previamente en cada camino se ha llevado a cabo un filtrado. Para la siguiente definición es posible que los valores de δ sean diferentes en R_1 y R_2 pero es necesario que dicha diferencia sea muy pequeña (para obtener bolas homogéneas). Sin embargo, como este valor es determinado a priori, con objeto de simplificar, puede suponerse que el radio de las regiones de bolas es idéntico para cada recorrido.

A continuación buscamos establecer algún tipo de relación de inclusión entre los elementos de $R^\delta(AB)$. De esta forma damos la siguiente definición:

Definición 1 ((ϵ, α) -inclusion) Se dice que R_1 es (ϵ, α) -incluido en R_2 y se denota por $R_1 \subset_\alpha^\epsilon R_2$ si al menos el 100 α % de B_{1i} satisface que existe un B_{2j} tal que $d(x_{1i}, x_{2j}) \leq \epsilon$ donde $\epsilon > 0$ y $0 \leq \alpha \leq 1$.

Se dice que R_1 es ϵ -incluido en R_2 y se denota por $R_1 \subset_\epsilon R_2$ si $R_1 \subset_1^\epsilon R_2$.

Esta definición es fácil de interpretar ya que si se considera ϵ como el límite del "campo visual de un punto" desde

¹Inicialmente este instante se utiliza para etiquetar la serie

²Si existen varios i , se elimina el primero y se vuelve a reevaluar.

un camino, el valor de $\alpha 100$ nos indica el porcentaje de puntos del otro camino que se encuentran en el campo visual del primer camino.

Comentemos el ejemplo que se puede ver en la Figura 3. Se tiene dos caminos, R_1 (10 bolas) en trazos continuos y R_2 (15 bolas) en trazos discontinuos. Se observa que si se toma $\epsilon = 1$ y una medida de distancia adecuada, 9 de las 15 bolas de R_2 están a menos de una separación de 1 de las bolas de R_1 , luego $R_2 \subset_{\alpha}^1 R_1$ con $\alpha \leq 0,60 = \frac{9}{15}$ (si hay un 60% hay más de un 40%, es la propiedad 1 siguiente). Por otro lado, 8 de las 10 bolas de R_1 están a menos de una separación de 1 de las bolas de R_2 (en orden las bolas 4 y 5 son las que se escapan), luego $\alpha = 0,8$, es decir $R_1 \subset_{\alpha}^1 R_2$ con $\alpha \leq 0,80 = \frac{8}{10}$.

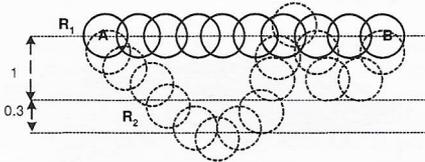


Figura 3: Ejemplo de inclusión.

Nota 2.2 Podría pensarse, desde una perspectiva frecuentista de la probabilidad, en la (ϵ, α) -inclusión como una probabilidad de una intersección, sin embargo no es así ya que ésta no es simétrica. Pero sí puede interpretarse como una probabilidad condicionada.

A partir de esta inclusión hemos definido dos medias de similitud entre caminos.

3. α -Similitud

Por simplicidad, obviaremos en esta sección el superíndice ϵ en la inclusión. Las propiedades de esta inclusión son:

Propiedad 1: Si $R_1 \subset_{\alpha} R_2$ entonces $R_1 \subset_{\alpha'} R_2$ para todo $\alpha' \leq \alpha$.

Propiedad 2: $\max_{\{R_1 \subset_{\alpha} R_2\}} \alpha$ existe dado que $0 \leq \alpha \leq 1$.

Propiedad 3: Es posible considerar la doble inclusión entre caminos i.e. $R_1 \subset_{\alpha_1} R_2$ y $R_2 \subset_{\alpha_2} R_1$. Sin embargo, no existe una relación general entre α_1 y α_2 .

Definición 2 (α -cuasiequivalencia e inclusión máxima)
 Dos caminos son α -cuasiequivalentes y se denota por $R_1 \overset{\alpha}{\equiv} R_2$ si $R_1 \subset_{\alpha} R_2$ y $R_2 \subset_{\alpha} R_1$.

Se define la máxima inclusión de R_i en R_j , y se denota por α_{ij} , como $\alpha_{ij} = \max_{\{R_i \subset_{\alpha} R_j\}} \alpha$.

De la definición anterior se sigue que si $\alpha = \min \{\alpha_{12}, \alpha_{21}\}$ entonces $R_1 \overset{\alpha}{\equiv} R_2$.

La relación $\overset{\alpha}{\equiv}$ no es una relación de equivalencia dado que $R_1 \overset{\alpha}{\equiv} R_2$ y $R_2 \overset{\alpha}{\equiv} R_3$ no implican que $R_1 \overset{\alpha}{\equiv} R_3$ como puede verse en la Figura 4. Pero se cumple algo similar ya que dado ϵ si $R_1 \overset{\alpha}{\equiv} R_2$ y $R_2 \overset{\alpha}{\equiv} R_3$ entonces $R_1 \overset{\alpha}{\equiv} R_3$ para 2ϵ .

Definición 3 Dados dos caminos R_1 y R_2 , se define la α -similitud entre ellos, y se denota por $Sim_{\epsilon}(R_1, R_2)$, como sigue:

$$Sim_{\epsilon}(R_1, R_2) = \frac{\alpha_{12} + \alpha_{21}}{2} \quad (1)$$

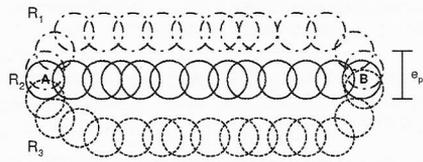


Figura 4: Ejemplo de tres caminos tales que $R_1 \overset{\alpha}{\equiv} R_2$ y $R_2 \overset{\alpha}{\equiv} R_3$ pero donde no es cierto que $R_1 \overset{\alpha}{\equiv} R_3$ para $\epsilon = 2\epsilon_p$.

Ejemplo 3.1 Los caminos dibujados en la Figura 3 tienen una similitud igual a 0,7 dado que $\alpha_{12} = 0,6$ y $\alpha_{21} = 0,8$, para $\epsilon = 1$.

Las propiedades fundamentales de esta similitud son:

- $0 < Sim_{\epsilon}(R_1, R_2) \leq 1$
- $Sim_{\epsilon}(R_1, R_2) = Sim_{\epsilon}(R_2, R_1)$
- $R_1 \overset{\alpha}{\equiv} R_2$ entonces $Sim_{\epsilon}(R_1, R_2) \geq \alpha$.
- $Sim_{\epsilon}(R_1, R_2) = 1 \Leftrightarrow R_1 \overset{\alpha}{\equiv} R_2$.
- $Sim_{\epsilon}(R_1, R_2) = 0 \Leftrightarrow \alpha_{12} = \alpha_{21} = 0$ pero esta situación no es posible debido a que ambos caminos comienzan en A y terminan en B. Por tanto, si $N = \min \{n, m\}$ y $M = \max \{n, m\}$ entonces

$$\min_{R_1, R_2 \in \mathcal{R}(AB)} Sim_{\epsilon}(R_1, R_2) = \frac{2}{M}$$

dado que $\min \alpha_{12} = \frac{2}{n}$ y $\min \alpha_{21} = \frac{2}{m}$.

4. ϵ -Similitud

En esta sección se considera la $(\epsilon, 1)$ -inclusión i.e. $R_1 \subset_{\epsilon}^1 R_2$ si para alguna B_{1i} existe una B_{2j} tal que $d(x_{1i}, x_{2j}) \leq \epsilon$ donde $\epsilon > 0$. Por simplicidad, se denota \subset_{ϵ}^1 como \subset^{ϵ} . Esta inclusión verifica las siguientes propiedades:

Propiedad 1: Si $R_1 \subset^{\epsilon} R_2$ entonces $R_1 \subset^{\epsilon'} R_2$ para todo $\epsilon' \geq \epsilon$.

Propiedad 2: $\min_{\{R_1 \subset^{\epsilon} R_2\}} \epsilon$ existe dado que $0 \leq \epsilon$. (siempre es posible elegir un ϵ tan grande como se quiera para contener al otro camino).

Propiedad 3: Es posible considerar la doble inclusión entre caminos i.e. $R_1 \subset^{\epsilon_1} R_2$ y $R_2 \subset^{\epsilon_2} R_1$. Sin embargo, no existe una relación general entre ϵ_1 y ϵ_2 .

Propiedad 4: Sea $R_1 \subset^{\epsilon_1} R_2$ y $R_2 \subset^{\epsilon_2} R_1$ y $\epsilon = \max \{\epsilon_1, \epsilon_2\}$, entonces $R_1 \subset^{\epsilon} R_2$ y $R_2 \subset^{\epsilon} R_1$ y se podría decir que R_1 y R_2 son caminos iguales de ancho ϵ y denotar $R_1 \overset{\epsilon}{=} R_2$.

Definición 4 (ϵ -cuasiequivalencia) Dos caminos son ϵ -cuasiequivalentes y se denota por $R_1 \overset{\epsilon}{=} R_2$ si $R_1 \subset^{\epsilon} R_2$ y $R_2 \subset^{\epsilon} R_1$.

Nota 4.1 La ϵ -cuasiequivalencia se interpreta fácilmente sin más que suponer que fijado un campo visual ϵ , un individuo que sigue un determinado camino, ve las trazas de todos los caminos equivalentes al suyo. Así por ejemplo se recoge la situación que se produciría si se va de un punto a otro pero por aceras o calles paralelas.

Esta idea está estrechamente relacionada con la envolvente de una trayectoria.

Definición 5 (ϵ -separación) Dados dos caminos R_1 y R_2 , se define la ϵ_{12} -separación como:

$$\epsilon_{12} = \min_{R_1 \subset^c R_2} \{ \epsilon : \epsilon \geq 0 \}.$$

Se prueba directamente que:

$$\epsilon_{12} = \max_{i=1, \dots, n} \min_{j=1, \dots, m} d(x_{1i}, x_{2j}).$$

Por simetría se sigue que:

$$\epsilon_{21} = \max_{j=1, \dots, m} \min_{i=1, \dots, n} d(x_{1i}, x_{2j}).$$

Definición 6 Dados dos caminos R_1 y R_2 , se define la ϵ -separación entre ellos, y se denota por $\epsilon(R_1, R_2)$, como sigue:

$$\epsilon(R_1, R_2) = \frac{\epsilon_{12} + \epsilon_{21}}{2}. \tag{2}$$

Ejemplo 4.1 Los caminos dibujados en la Figura 3 tienen una ϵ -separación igual a 1,15 dado que $\epsilon_{12} = 1$ y $\epsilon_{21} = 1,3$.

Las propiedades fundamentales de esta separación son:

1. $\epsilon(R_1, R_2) \geq 0$
2. $\epsilon(R_1, R_2) = \epsilon(R_2, R_1)$
3. $\epsilon(R_1, R_2) = 0 \Leftrightarrow R_1 = R_2$ dado que $\epsilon_{12} = \epsilon_{21} = 0$ implican que para cualquier x_{1i} existe x_{2j} tal que $x_{1i} = x_{2j}$ y para cualquier x_{2j} existe x_{1i} tal que $x_{2j} = x_{1i}$.
4. $\epsilon(R_1, R_2)$ verifica la desigualdad triangular.
5. Si $R_1 \overset{c}{\subseteq} R_2$ entonces $\epsilon(R_1, R_2) \leq \epsilon$ dado que $R_1 \subset^c R_2$ entonces $\epsilon_{12} \leq \epsilon$.

De las cuatro primeras propiedades se sigue que $\epsilon(R_1, R_2)$ es una medida de distancia.

Definición 7 Dados dos caminos R_1 y R_2 , se define la ϵ -similitud entre ellos, y se denota por $Sim(R_1, R_2)$, como sigue:

$$Sim(R_1, R_2) = \frac{1}{1 + \epsilon(R_1, R_2)} \tag{3}$$

5. Δ -Similitud

En esta sección se considera una medida clásica de distancia/similitud entre trayectorias (área definida entre dos funciones). Así, se supone que un camino está definido por puntos en vez de bolas, i.e.

$$R_i^\delta = \{A, x_{i1}, x_{i2}, \dots, x_{in_i} = B\}.$$

Dados dos caminos R_1 y R_2 , como ambos salen de A y llegan a B, definen un polígono cerrado a partir de las dos poligonales obtenidas uniendo los puntos de cada camino (ver Figura 5).

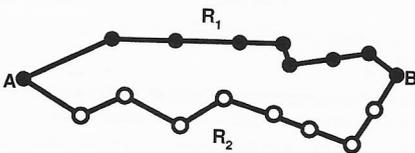


Figura 5: Ejemplo de dos caminos como un polígono.

De esta forma se da la siguiente definición:

Definición 8 (Equivalencia) Dos caminos son equivalentes y se denota por $R_1 \equiv R_2$ si definen el mismo polígono i.e. ambos caminos comparten los mismos puntos (igualdad entre conjuntos).

Claramente esta relación es una relación de equivalencia.

Definición 9 (Distancia y Similitud) Dados dos caminos R_1 y R_2 , se define una distancia entre ellos, y se denota por $darea(R_1, R_2)$, como el área cerrada entre R_1 y R_2 (ver Figura 6).

Se define también la similitud entre R_1 y R_2 , y se denota por $\Delta(R_1, R_2)$, como sigue:

$$\Delta(R_1, R_2) = \frac{1}{1 + darea(R_1, R_2)} \tag{4}$$

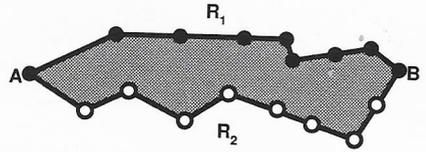


Figura 6: Distancia entre dos caminos como un área.

Ciertamente, $darea$ es una distancia dado que para cualquier $R_1, R_2 \in R(AB)$ se tiene que:

1. $darea(R_1, R_2) = 0$ ó $(\Delta(R_1, R_2) = 1) \Leftrightarrow R_1 \equiv R_2$.
2. $darea(R_1, R_2) = darea(R_2, R_1)$ ($\Delta(R_1, R_2) = \Delta(R_2, R_1)$).
3. $darea(R_1, R_2) \geq 0$ ($0 \leq \Delta(R_1, R_2) \leq 1$).
4. $darea(R_1, R_2) \leq darea(R_1, R_3) + darea(R_3, R_2)$ para cualquier $R_3 \in R(AB)$.

Esto es claro sin más que pensar que se ha identificado cada camino con una trayectoria (se puede pensar en el plano o en el espacio) o una función. Por tanto $R_1 = f_1(x)$ y $R_2 = f_2(x)$ y se ha definido $darea(R_2, R_1) = \int_A^B \|f_1(x) - f_2(x)\| dx$.

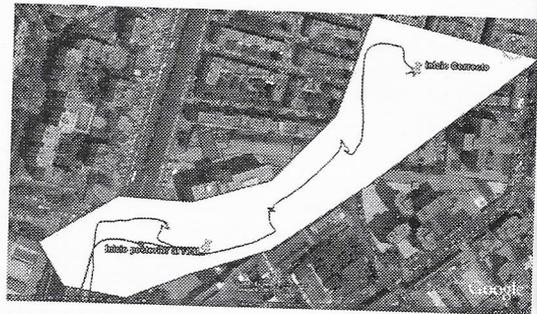


Figura 7: Recorrido con inicio posterior al real.

El principal inconveniente que presenta en la práctica esta aproximación se encuentra cuando se pretende aplicar este tipo de distancia a dos caminos que aún empezando y finalizando en los mismos lugares, los puntos iniciales y finales validados no coinciden, como podemos ver en la Figura 7, con lo cual no se obtendría un polígono (las poligonales no se cierran). Sin embargo, esta situación puede ser salvada si previamente se decide, por convenio, tomar

los lugares más frecuentados en los caminos como marcas de posiciones y de esta forma se cierra el polígono con la línea que une el anterior punto con el primer punto validado del camino. Por ello, en las futuras implementaciones se considerará también esta aproximación.

6. Comentarios

En este artículo se describen los estudios preliminares realizados sobre comparación de caminos realizados por usuarios mediante cualquier medio. Dado que la captura de datos del receptor de GPS no es exacta como hemos señalado, es necesario definir medidas de similitud válidas que permitan comparar de manera eficiente caminos, considerando la idea de que dos caminos se parecen si sus orígenes y destinos son los mismos y las trayectorias intermedias están cercanas espacialmente.

En futuros trabajos se aplicarán las diferentes medidas de similitud para poder comparar en tiempo real caminos inacabados con aquellos finalizados para poder predecir el lugar objetivo de la persona que se desplaza. Como posibles escenarios prácticos destacamos la detección de pérdidas de orientación en exteriores de personas con enfermedades psíquicas, la predicción de encuentros entre diferentes usuarios de los que se predice sus destinos y su momento de llegada a los mismos o la anticipación de acciones automáticas como la descrita anteriormente del orador que llega a la sala de presentaciones.

7. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por los proyectos FAMENET-InCare (TSI2006-13390-C02-02) del Ministerio de Educación y Cultura Español y CUBICO (P06-TIC-02141) de la Junta de Andalucía.

Referencias

- [Chen *et al.*, 2005] Lei Chen, M. Tamer Ozsu, and Vincent Oria. Robust and fast similarity search for moving object trajectories. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 491–502, New York, NY, USA, 2005. ACM Press.
- [Cuberos, 2005] Francisco Javier Cuberos. *Índices para la Comparación Cualitativa de Series Temporales*. PhD thesis, Universidad de Sevilla, 2005.
- [Faloutsos *et al.*, 1994] Christos Faloutsos, M. Ranganathan, and Yannis Manolopoulos. Fast subsequence matching in time-series databases. In *SIGMOD '94: Proceedings of the 1994 ACM SIGMOD international conference on Management of data*, pages 419–429, New York, NY, USA, 1994. ACM Press.
- [Keogh *et al.*, 2004] E. Keogh, Palpanas T., Zordan V., Gunopulos D., and Cardle M. Indexing large human-motion databases. In *Proceedings of the 30th International Conference on Very Large Data Bases*, 2004.
- [Lee *et al.*, 2000] Seok-Lyong Lee, Seok-Ju Chun, Deok-Hwan Kim, Ju-Hong Lee, and Chin-Wan Chung. Similarity search for multidimensional data sequences. In *ICDE*, pages 599–608, 2000.
- [Lin and Su, 2005] Bin Lin and Jianwen Su. Shapes based trajectory queries for moving objects. In *GIS '05: Proceedings of the 13th annual ACM international workshop on Geographic information systems*, pages 21–30, New York, NY, USA, 2005. ACM Press.
- [Lin, 2006] Yi Lin. Efficient human motion retrieval in large databases. In *GRAPHITE '06: Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, pages 31–37, New York, NY, USA, 2006. ACM Press.
- [Little and Gu, 2001] James Little and Zhe Gu. Video retrieval by spatial and temporal structure of trajectories. In *SIPE, Storage and Retrieval for Media Databases*, 2001.
- [Vlachos *et al.*, 2004] Michail Vlachos, D. Gunopulos, and Gautam Das. Rotation invariant distance measures for trajectories. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 707–712, New York, NY, USA, 2004. ACM Press.