



HAL
open science

Trajectoires médianes

Laurent Etienne, Thomas Devogele

► **To cite this version:**

Laurent Etienne, Thomas Devogele. Trajectoires médianes. 14 ème conférence Extraction et Gestion des Connaissances, Ateliers fouille de données spatiales et temporelles & construction, enrichissement et exploitation de ressources géographiques pour l'analyse de données, Jan 2014, Rennes, France. hal-02110086

HAL Id: hal-02110086

<https://hal.science/hal-02110086>

Submitted on 25 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Trajectoires médianes

Laurent Etienne*, Thomas Devogele**

*Institut de Recherche de l'Ecole navale, Brest, France

laurent.etienne@ecole-navale.fr

<http://www.letienne.net>

**Université de Tours, Laboratoire d'informatique, Blois, France

thomas.devogele@univ-tours.fr

<http://thomas.devogele.free.fr>

Résumé. Des masses de données représentant des trajectoires sont aujourd'hui disponibles. Ces trajectoires sont nombreuses et très diverses (piétons, véhicules, navires, avions, animaux...). Cet article présente un algorithme générique et robuste pour obtenir une trajectoire médiane. Ce motif permet de synthétiser un grand nombre de trajectoires partant d'une zone d'intérêt et se rendant dans une autre zone. Il permet de déduire des connaissances sur les comportements d'objets mobiles de même type. Il est basé sur l'appariement des positions des trajectoires à l'aide de la distance de Fréchet discrète. Deux jeux de trajectoires volumineux permettent d'illustrer cet algorithme.

1 Introduction

Des masses de données de localisation sont aujourd'hui collectées. En particulier, la technologie GPS permet de disposer de volumes importants de trajectoires d'objets mobiles. Pour comprendre le mouvement de ces objets, il est nécessaire de définir des motifs synthétisant ces masses de trajectoires et d'en extraire des connaissances. Dans cet article nous nous focalisons sur des groupes de trajectoires partant d'une même zone initiale et arrivant à une même destination. L'objectif est de synthétiser un groupe de trajectoires d'objets de même type suivant un même chemin à l'aide d'une trajectoire centrale. Plusieurs solutions ont été proposées (Buchin et al. (2010); Freeman et al. (2011); Vermeulen (2013)). Hélas, ces solutions ne sont pas robustes. Elles sont sensibles à la présence de trajectoires aberrantes (outliers) et traitent difficilement les trajectoires réalisant des boucles.

L'utilisation d'algorithmes basés sur la technique du Dynamic Time Warping (DTW) permet de s'affranchir de ces problématiques car il respecte l'ordonnancement temporel des positions (Petitjean et al. (2011)). Différentes techniques d'alignement de séries temporelles appliquées aux trajectoires produisent des suites de couples de positions pouvant être utilisées pour définir le concept de trajectoire médiane proposé dans cet article.

Ces travaux, issus de la thèse de Etienne (2011), étendent ceux de Devogele (2002) qui définissent des lignes centrales par appariement de points homologues. L'approche proposée est similaire à celle de Petitjean et al. (2011) qui propose une méthode pour définir une séquence moyenne à partir d'un ensemble de séquences temporelles.

La structure de cet article est la suivante. Dans le chapitre suivant, l'algorithme permettant de définir la trajectoire médiane est détaillé. Dans le chapitre 3, deux exemples de motifs s'appuyant sur 500 trajectoires de navires et 800 trajectoires de pigeons permettent d'illustrer cet algorithme. Le chapitre 4 conclut en rappelant les apports de cette technique et les principales perspectives.

2 Trajectoires médianes

2.1 Approche générale

L'étude d'un grand nombre de jeu de trajectoires montre que les clusters de trajectoires ne sont pas aisément modélisables par une loi statistique de distribution connue. Par exemple, une loi symétrique n'est pas adaptée pour modéliser des motifs de trajectoires (Etienne (2011)). En effet, un objet mobile se déplaçant entre deux points a tendance à optimiser son déplacement en termes de temps, de sécurité et de coûts. Il est par ailleurs contraint dans son déplacement par certains obstacles (fixes ou mobiles) ainsi que par son environnement (pluie, vent, courant, ...). L'ensemble de ces critères impactent son déplacement et sont difficilement modélisables par une loi statistique.

D'un point de vue statistique descriptive, il est donc nécessaire de se tourner vers des approches de type médianes. De plus, la médiane est plus robuste à la présence de données aberrantes. Dans cet objectif, pour un ensemble de points, il est possible de définir un point médian. Il existe un grand nombre d'extensions de la médiane aux données en deux dimensions (Small (1990)). Le choix dépend essentiellement des propriétés recherchées pour ce point et de la complexité algorithmique pour l'obtenir. De même, pour une position (point ayant une estampille temporelle), une position médiane peut être choisie. Une autre contrainte forte à prendre en compte est qu'il ne faut pas que cette valeur centrale soit positionnée à l'intérieur d'un obstacle. Il faut donc mieux s'appuyer sur des points existants dans le jeu de données.

Pour définir une trajectoire centrale, il a donc été décidé que le processus doit :

- étendre une approche médiane,
- utiliser un processus d'appariement robuste entre les points de lignes,
- définir une trajectoire centrale (la trajectoire médiane) comme une suite ordonnée de positions où chaque point associé est un point réel et où l'estampille temporelle est le temps médian pour arriver à ce point depuis la zone de départ.

Dans cet article, l'approche consiste à apparier les positions des trajectoires afin de définir des clusters de positions. Pour chaque cluster, une position médiane peut alors être définie.

Pour apparier les données, une mesure de similarité basée sur une distance spatiale est utilisée. Pour les lignes ou trajectoires, différentes mesures sont envisageables :

- Dynamic Time Warping (DTW) (Berndt et Clifford (1994)) qui est un algorithme de mesure de similarité entre deux séquences temporelles et peut être étendu à des trajectoires. DTW met l'accent sur les différences locales entre les points correspondants de lignes.
- La distance de Fréchet (Alt et Godau (1995)) entre deux courbes qui tient compte de l'ordre des points le long des courbes. Cette mesure s'appuie sur la distance maximale entre les points appariés. Une programmation dynamique discrète a été proposée dans (Devogele (2002)) pour deux polygones. Cette version discrète calcule une approxima-

tion de la distance de Fréchet maximale, une distance moyenne proche de celle renvoyée par DTW et des appariements entre les points.

- La technique du Time Warp Edit Distance (Marteau (2009)) étend le concept de la distance d'édition (Chen et Ng (2004)) afin de définir la similarité entre deux séries de données temporelles (Marteau (2009)).

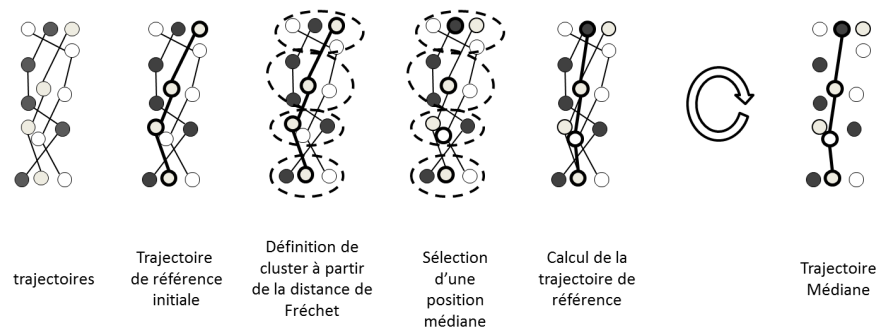


FIG. 1 – Les différentes étapes de l'algorithme de définition de la trajectoire médiane.

Ces méthodes sont robustes. Elles tiennent compte de l'ordre des points des lignes. De cette manière, les lignes avec des boucles ou des lignes sinueuses avec des décalages peuvent être appariées. Les étapes principales de l'algorithme de définition de la trajectoire centrale sont les suivantes :

1. Une trajectoire de référence est choisie ;
2. toutes les positions de chaque trajectoire sont appariées avec la position de la trajectoire de référence, ce processus génère une suite ordonnée de clusters de positions ;
3. pour chaque cluster, une position médiane est définie (cette position doit être une position réelle d'une trajectoire) ;
4. une nouvelle trajectoire de référence est construite à partir des positions médianes ;
5. les étapes 2 à 4 sont répétées jusqu'à ce qu'il n'y ait plus de changement dans les positions médianes ;
6. pour chaque point de la trajectoire de référence, une estampille temporelle médiane est calculée à partir du cluster de positions appariées à ce point.

La figure 1 illustre ces principales étapes avec un exemple très simple. Une trajectoire centrale est calculée pour un groupe de trois trajectoires différentes. Les positions de référence de la trajectoire de référence sont cerclées de gras. Les clusters de position sont délimités par des ellipses en pointillés. Plusieurs positions d'une même trajectoire peuvent appartenir à un même cluster. Le déroulement de cet algorithme est très proche de l'algorithme de clustering k-medoids (Kaufman et Rousseeuw (1987)).

Néanmoins, pour traiter les trajectoires de nouvelles contraintes doivent être prises en compte.

Trajectoires médianes

- Pour chaque cluster, au moins une position de chaque trajectoire doit être associée ;
- Toutes les positions p_{i_j} (position j de trajectoire i) sont dans au moins un cluster C_k ;
- l'ordre des positions doit être pris en compte. Si une position p_{i_j} est dans le cluster C_k , la position $p_{i_{j+1}}$ doit être dans le cluster C_k ou dans C_{k+1} .

Le processus d'appariement choisi doit répondre à toutes ces contraintes. Ainsi, la trajectoire médiane obtenue est composée d'une suite ordonnée temporellement de positions médianes minimisant la somme des écarts (Fréchet, DTW ou TWED) avec toutes les autres trajectoires du groupe. Il est important de noter que les positions qui composent cette trajectoire sont des positions réelles appartenant à différentes trajectoires du groupe.

2.2 Phase de prétraitement

La trajectoire médiane est calculée pour définir la trajectoire typique entre deux zones d'intérêt pour un même type d'objet mobile. Les trajectoires suivant ce chemin entre ces deux zones doivent donc au préalable être extraites.

L'algorithme de calcul de la trajectoire médiane proposé est basé sur des techniques d'alignement de série temporelles (DTW, Fréchet, TWED) dont la complexité dépend du nombre T de positions d'une trajectoire. L'ordre de complexité de ces algorithmes d'alignement est $\Theta(T^2)$. L'algorithme de calcul de la trajectoire médiane est itératif et réalise à chaque itération un nombre N d'appariements entre la trajectoire de référence et les N trajectoires du groupe. Ainsi pour un nombre I d'itérations réalisées, la complexité de l'algorithme de calcul de la trajectoire médiane est de l'ordre de $\Theta(I.N.T^2)$ (Petitjean et al. (2011)).

La complexité de l'algorithme de calcul de la trajectoire médiane étant directement liées au nombre T de positions des trajectoires du groupe, un filtrage et un ré-échantillonnage permet de réduire le temps de calcul. Un filtre spatio-temporel de Douglas et Peucker (Meratnia et de By (2004)) est utilisé. Le but de ce filtre est de ne retenir que les positions significatives de chaque trajectoire. Il faut donc conserver les positions qui permettent de conserver les informations sur les changements de directions ou de vitesse. Dans un deuxième temps, afin d'améliorer le processus d'appariement, si certains segments des trajectoires sont trop longs, elles sont ré-échantillonnées. Ces deux pré-traitements permettent de diminuer le temps CPU nécessaire au calcul de la trajectoire médiane.

2.3 Sélection de la trajectoire de référence

Le choix de la trajectoire initiale de référence a également un impact sur le nombre d'itérations nécessaires à l'algorithme avant de se stabiliser. Un choix judicieux de la trajectoire initiale de référence permet d'optimiser le processus.

En effet, en cas de sélection aléatoire, la trajectoire de référence peut potentiellement être une trajectoire aberrante du groupe. Ceci engendrera un nombre important d'itérations de l'algorithme et de répétitions des étapes 2 à 4.

Afin de sélectionner une trajectoire de référence proche de la trajectoire médiane, une heuristique est proposée. Cette heuristique consiste à sélectionner comme trajectoire initiale de référence celle dont la longueur est la plus proche de la valeur médiane des longueurs de toutes les trajectoires du groupe.

La complexité du calcul de la longueur d'une trajectoire est proportionnelle au nombre de segments composant cette trajectoire ($\Theta(T)$). Ainsi la complexité algorithmique de cette étape

d'optimisation est de $(\Theta(N.T))$ et n'impacte donc pas la complexité globale de l'algorithme $\Theta(I.N.T^2)$.

2.4 Processus de clustering des positions

Une fois la trajectoire de référence sélectionnée, les positions de chaque trajectoire peuvent être appariées aux positions de référence. Les différentes méthodes (DTW, distance de Fréchet et TWED) peuvent être employées. Les résultats de ces processus d'appariement sont proches et la complexité reste la même. Dans cet article, la distance de Fréchet discrète a été retenue et va être décrite. L'illustration classique de la distance de Fréchet est la suivante : un homme marche avec un chien en laisse. Il marche sur une courbe, le chien sur l'autre. Les deux peuvent faire varier leur vitesse, mais le retour en arrière n'est pas permis.

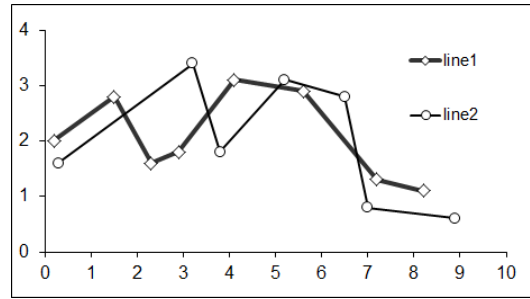


FIG. 2 – Exemple de couple de lignes.

La distance Fréchet entre ces deux courbes est la longueur minimale de la laisse qui est nécessaire pour assurer ce cheminement. La distance de Fréchet discrète entre L_a et L_b (d_{Fd}) est calculée de manière récursive :

$$d_{Fd}(L_a, L_b) = \max \left(\begin{array}{l} d_E(L_{a_n}, L_{b_m}) \\ \min \left(\begin{array}{l} d_{Fd}(\{L_{a_1} \dots L_{a_{n-1}}\}, \{L_{b_1} \dots L_{b_m}\}) \quad \forall n > 1 \\ d_{Fd}(\{L_{a_1} \dots L_{a_n}\}, \{L_{b_1} \dots L_{b_{m-1}}\}) \quad \forall m > 1 \\ d_{Fd}(\{L_{a_1} \dots L_{a_{n-1}}\}, \{L_{b_1} \dots L_{b_{m-1}}\}) \quad \forall n > 1 \\ d_{Fd}(\{L_{a_1} \dots L_{a_{n-1}}\}, \{L_{b_1} \dots L_{b_{m-1}}\}) \quad \forall m > 1 \end{array} \right) \end{array} \right) \quad (1)$$

Où L_{i_j} est le point j de la ligne i et d_E est la distance euclidienne. A l'aide d'une méthode de programmation dynamique, ce problème récursif complexe peut être résolu en calculant une matrice intermédiaire. Ainsi, deux matrices sont employées :

- La matrice de distance euclidienne (MD). La valeur de la cellule $MD_{i,j}$ correspond à la distance euclidienne entre L_{a_i} et L_{b_j} .
- La matrice Fréchet (MF) qui permet de calculer de manière itérative la distance de Fréchet (Eiter et Mannila (1994)).

La formule utilisée pour calculer $MF_{i,j}$ est :

$$MF_{i,j} = \max(d_E(L_{a_i}, L_{b_j}), \min(MF_{i-1,j}, MF_{i,j-1}, MF_{i-1,j-1})) \quad (2)$$

Trajectoires médianes

L'exemple présenté dans la figure 2, illustre le calcul de la distance de Fréchet discrète. La ligne 1 (L_a) se compose de 8 points (L_{a_1} à L_{a_8}) et la ligne 2 (L_b) est composée de 7 points (L_{b_1} à L_{b_7}). La distance de Fréchet discrète est contenue dans la dernière case de la matrice de Fréchet discrète ($MF_{n,m}$). Le tableau 1 donne la distance de Fréchet discrète entre les deux lignes de la figure 2 (1,80).

		i	1	2	3	4	5	6	7	8	
La		x	0,20	1,50	2,30	2,90	4,10	5,60	7,20	8,20	
		y	2,00	2,80	1,60	1,80	3,10	2,90	1,30	1,10	
Lb		Matrice de distances									
j	x	y	MD	1	2	3	4	5	6	7	8
1	0,30	1,60	1	0,41	1,70	2,00	2,61	4,09	5,46	6,91	7,92
2	3,20	3,40	2	3,31	1,80	2,01	1,63	0,95	2,45	4,52	5,50
3	3,80	1,80	3	3,61	2,51	1,51	0,90	1,33	2,11	3,44	4,46
4	5,20	3,10	4	5,12	3,71	3,26	2,64	1,10	0,45	2,69	3,61
5	6,50	2,80	5	6,35	5,00	4,37	3,74	2,42	0,91	1,66	2,40
6	7,00	0,80	6	6,91	5,85	4,77	4,22	3,70	2,52	0,54	1,24
7	8,90	0,60	7	8,81	7,72	6,68	6,12	5,41	4,02	1,84	0,86
		Matrice de Fréchet									
		MF	1	2	3	4	5	6	7	8	
		1	0,41	1,70	2,00	2,61	4,09	5,46	6,91	7,92	
		2	3,31	1,80	2,00	2,00	2,00	2,45	4,52	5,50	
		3	3,61	2,51	1,80	1,80	1,80	2,11	3,44	4,46	
		4	5,12	3,71	3,26	2,64	1,80	1,80	2,69	3,61	
		5	6,35	5,00	4,37	3,74	2,42	1,80	1,80	2,40	
		6	6,91	5,85	4,77	4,22	3,70	2,52	1,80	1,80	
		7	8,81	7,72	6,68	6,12	5,41	4,02	1,84	1,80	

TAB. 1 – Matrice euclidienne et Matrice de Fréchet.

L'ensemble des couples de points est défini par back-tracking à l'aide des deux matrices. Premièrement, le couple L_{a_n} et L_{b_m} est sélectionné. Un nouveau couple est choisi parmi les trois suivants : $(L_{a_{n-1}}, L_{b_m})$, $(L_{a_n}, L_{b_{m-1}})$ et $(L_{a_{n-1}}, L_{b_{m-1}})$. Les valeurs $MF_{n-1,m-1}$, $MF_{n-1,m}$ et $MF_{n,m-1}$ sont utilisées en premier. Ces trois valeurs permettent de sélectionner le couple de positions à appairer. Le couple sélectionné est celui ayant la plus petite valeur dans la matrice de Fréchet. Si plusieurs cellules ont des valeurs minimales identiques, alors la distance euclidienne est utilisée. Enfin, si les distances euclidiennes sont aussi identiques, le parcours en diagonale est privilégié (couple $n - 1, m - 1$).

Cette étape est répétée jusqu'au couple de positions de départ des deux lignes L_{a_1}, L_{b_1} . Pour cet exemple, les valeurs $MF_{6,7}$ et $MF_{6,8}$ sont identiques (1,80), nous avons donc besoin de comparer $ME_{7,6}$ et $ME_{8,6}$. La distance $ME_{7,6}$ est inférieure à $ME_{8,6}$, le couple (L_{a_7}, L_{b_6}) est donc ajouté au début de la liste des couples de points appariés. L'étape suivante consiste à sélectionner le prochain couple de points appariés entre (L_{a_6}, L_{b_6}) , (L_{a_6}, L_{b_5}) et (L_{a_7}, L_{b_5}) . Le couple de points (L_{a_6}, L_{b_5}) est sélectionné en appliquant le même algorithme de sélection.

La figure 3 montre les 9 couples de points appariés correspondant aux cellules surlignées dans la matrice de la table 1. Nous pouvons remarquer que dans certains cas, plusieurs points de la ligne L_a sont appariés à un même point de la ligne L_b et vice versa. Par exemple le point L_{a_6} est assorti avec L_{b_4} et L_{b_5} .

La figure 4, illustre la robustesse de l'algorithme aux lignes ayant des boucles. Dans les figures 3 et 4, nous pouvons également remarquer que les points ne sont pas toujours appariés

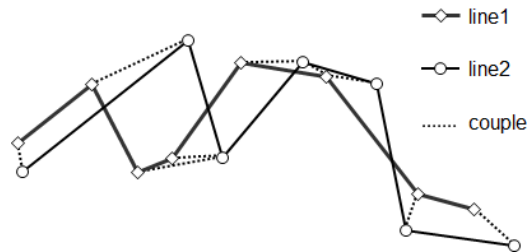


FIG. 3 – Couples de points appariés à l'aide de la distance de Fréchet discrète.

avec le point le plus proche. Les clusters de positions appariés sont définis à partir de ces couples de positions appariées entre la ligne de la trajectoire de référence (L_r) et les autres trajectoires. Pour chaque L_{r_i} de la trajectoire de référence, un cluster de points appariés est construit. Une position peut être dans plusieurs clusters. Par exemple, si les couples sont les suivants : (L_{r_1}, L_{1_1}) , (L_{r_1}, L_{1_2}) , (L_{r_2}, L_{1_2}) , (L_{r_2}, L_{1_3}) , (L_{r_3}, L_{1_3}) et (L_{r_1}, L_{2_1}) , (L_{r_2}, L_{2_2}) , (L_{r_3}, L_{2_3}) , trois clusters sont définis :

- Cl_1 : Positions appariées à L_{r_1} : L_{1_1} L_{1_2} L_{2_1}
- Cl_2 : Positions appariées à L_{r_2} : L_{1_2} L_{1_3} L_{2_2}
- Cl_3 : Positions appariées à L_{r_3} : L_{1_3} L_{2_3}

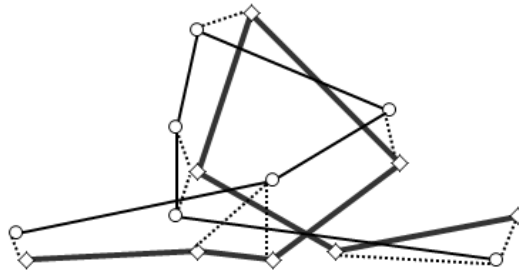


FIG. 4 – Couples de points appariés de lignes ayant une boucle.

2.5 Définition de la nouvelle trajectoire de référence

Pour chaque cluster une position médiane peut alors être définie. Cette position doit être une position réelle d'une trajectoire. Plusieurs médianes multidimensionnelles (Small (1990)) existent. La médiane marginale définie par (Puri et Sen (1971)) est utilisée. Cette position centrale est simple à calculer : X et Y médians sont calculés indépendamment. Pour éviter une position irréaliste, une position réelle proche de la médiane marginale doit être choisie. En effet, une position centrale synthétique pourrait être localisée dans une zone impossible (par

Trajectoires médianes

exemple l'emprise d'une rivière pour une trajectoire de piétons). La position la plus proche de la médiane marginale selon la distance euclidienne est choisie comme position centrale.

Des médianes multidimensionnelles ayant des meilleures caractéristiques auraient pu être employées, mais le temps CPU aurait été plus important et les résultats sont très similaires. De même, le médoïde (Struyf et al. (1997)) n'est pas employé. Ce point est souvent utilisé comme position centrale. Malheureusement, il n'est pas robuste dès que le nombre de points aberrants est important. Une fois la position médiane de chaque cluster défini, la trajectoire de référence peut être construite en réutilisant l'ordre des clusters pour ordonner les positions médianes.

2.6 Réitération des étapes

Les étapes de constructions des clusters, de définition des positions médianes et de la trajectoire de référence sont réitérées jusqu'à ce que la trajectoire de référence converge. Cette dernière est alors considérée comme la trajectoire médiane.

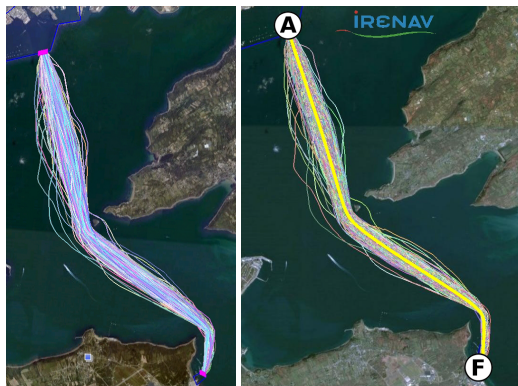


FIG. 5 – 554 trajectoires entre Brest et l'Ecole navale et la trajectoire médiane associée.

2.7 Estampille temporelle

À l'issu de cette phase, nous obtenons une médiane qui prend en compte uniquement la dimension spatiale de la trajectoire. Pour chaque position médiane, le temps médian nécessaire pour aller de la zone de départ à cette position est calculé.

Par construction cette médiane est robuste. Les outils employés ne sont pas sensibles aux données aberrantes et tiennent compte de l'ordre des points. De même, les trajectoires peuvent avoir des boucles. En revanche, la complexité de cet algorithme est importante et demande une phase d'optimisation préalable.

3 Applications à des trajectoires réelles

3.1 Trajectoire médiane de navires

Le premier jeu de données décrit 554 trajectoires de navires à passagers entre le port de Brest et l'Ecole navale (Etienne (2011)). La durée entre deux positions varie entre 1 à 6 secondes. La figure 5 (partie gauche) illustre ce jeu de trajectoires. Ce jeu de données peut être téléchargé librement sur internet (fichier AIS Brest, www.chorochronos.org).

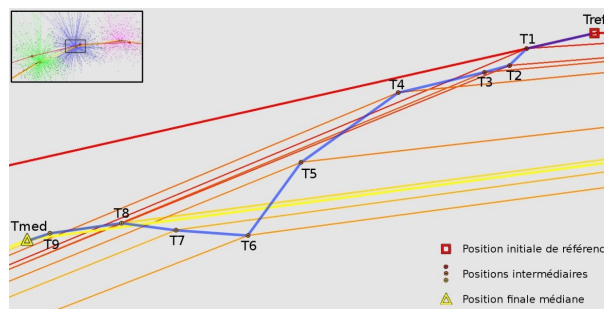


FIG. 6 – Evolution d'une position de la trajectoire de référence.

La figure 6 montre l'évolution du calcul d'une position de référence. 10 itérations ont été nécessaires pour définir cette trajectoire médiane. Sur la superposition du jeu de trajectoire et de la trajectoire médiane (figure 5 partie droite), nous pouvons constater que la trajectoire médiane n'est pas centrée sur le virage au milieu de la figure. En effet, la trajectoire passe plus près de la petite île. De même au niveau du virage du bas, la trajectoire médiane est proche de la rive.

3.2 Trajectoire médiane de pigeons

Le deuxième jeu de données concerne les déplacements de pigeons voyageurs. Ce jeu de données est présenté en détail dans Pettit et al. (2013). Il est composé de trajectoires de 23 pigeons différents équipés de balises GPS enregistrant leurs positions toutes les 200 ms (UTM Zone 30U). Les pigeons étudiés ont tous été entraînés de la même façon préalablement à l'expérience et connaissent déjà l'environnement aux alentours du pigeonnier. Lors de l'expérience, ces pigeons sont relâchés, un par un ou en couple, à partir d'un même endroit puis rentrent ensuite à leur pigeonnier. Ce trajet est réalisé 21 fois par pigeon afin de lui permettre de reconnaître l'environnement. Seul les 5 derniers trajets d'un pigeon sont enregistrés à l'aide du GPS.

L'étude de Pettit et al. (2013) montre que la distance moyenne entre les points appariés les plus proches des trajectoires consécutives diminue lorsque le pigeon a réalisé un nombre important de trajet identiques. Cela signifie que plus le pigeon réalise un trajet, plus il a tendance à suivre une route proche de son précédent trajet. L'article de Pettit et al. (2013) s'intéresse à la comparaison des déplacements des pigeons lorsqu'ils sont relâchés seuls ou en couple.

Trajectoires médianes

L'algorithme de Freeman et al. (2011) est utilisé pour synthétiser la trajectoire moyenne des 5 trajectoires d'un même pigeon avec un échantillonnage de 1000 points.

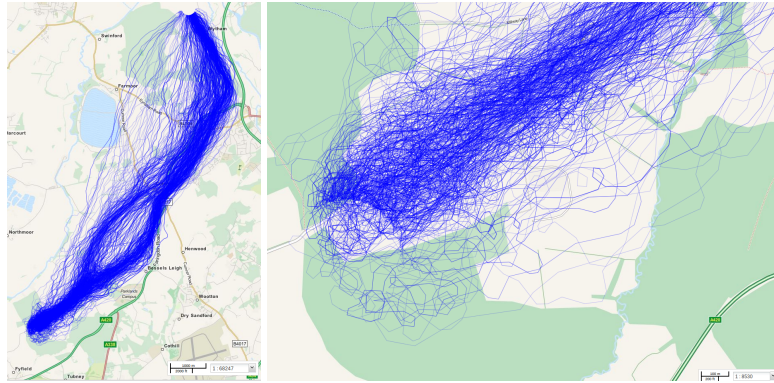


FIG. 7 – Cluster de trajectoires de pigeons voyageurs.

Comme indiqué précédemment, l'inconvénient de cette technique réside dans le fait qu'elle se base principalement sur une analyse purement spatiale en minimisant la distance aux plus proches voisins sans tenir compte de l'ordre de relation temporel entre positions consécutives. Ceci pose problème en cas de boucles dans les trajectoires. Dans ce jeu de données, nous avons pu observer de nombreuses petites boucles locales réalisées le long du parcours par certains pigeons (figure 7). Ces boucles sont problématiques pour l'analyse et la génération d'une trajectoire médiane car elles introduisent de nombreuses positions appariées à une même position de référence.

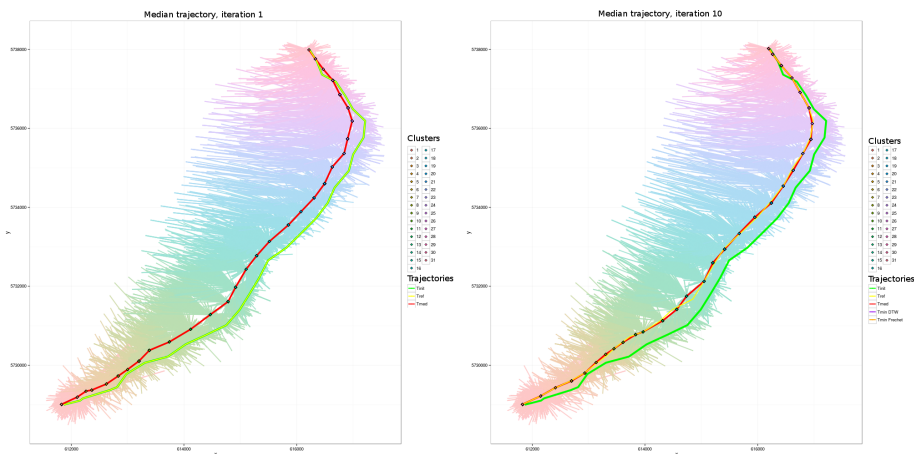


FIG. 8 – Trajectoires médianes obtenues après 1 et 10 itérations.

L'algorithme de calcul de la trajectoire médiane a été réalisé sur une version sous échantillonnée (une position toute les 10 s) du cluster de trajectoires de pigeons pour des raisons de temps de calculs. La trajectoire réelle initiale de référence sélectionnée dans le cluster est présentée en vert sur la figure 8. Les positions centrales des clusters de positions appariées sont représentées par des losanges colorés. Les résultats des itérations 1 et 10 sont présentés sur la figure 8. La trajectoire rouge correspond à la trajectoire médiane calculée à partir de la précédente trajectoire de référence représentée en jaune.

L'inconvénient du sous-échantillonnage est qu'il a tendance à supprimer les petites boucles locales dans les trajectoires. L'application d'un filtre spatio-temporel de Douglas et Peucker (Meratnia et de By (2004)) pourrait également minimiser les temps de calculs. En revanche, le nombre de positions multiples d'une même trajectoire affectées à un cluster serait plus important.

4 Conclusion

Cet article a proposé un algorithme robuste pour définir une trajectoire médiane. Ce motif permet d'obtenir une trajectoire centrale pour un groupe de trajectoires entre deux zones d'intérêt. Les deux exemples illustrent le potentiel et les avantages de cet algorithme sur des larges jeux de données. Cet algorithme est polynomial du fait qu'il tient compte de l'ordre des positions. Néanmoins, il peut être optimisé grâce à l'utilisation d'algorithmes de filtrage et il est également parallélisable. Les temps de calcul pour des jeux de données réels très conséquents se mesurent alors en heures. Ce motif ne devant pas être défini fréquemment, cela reste acceptable. La trajectoire médiane est un bon estimateur d'une trajectoire centrale. Afin de pouvoir distinguer les trajectoires inhabituelles, la trajectoire médiane doit être étendue à des représentations statistiques intégrant les moments d'ordre supérieurs. Des réflexions portant sur l'extension spatio-temporelle des boîtes à moustache ont été avancées dans (Etienne (2011)). Elles doivent encore être affinées. De même, pour des clusters de trajectoires qui se séparent en deux groupes, tel l'exemple des trajectoires des pigeons, des outils automatiques de détection d'une distribution bimodale des groupes de trajectoires doivent être mis en place. Finalement, une étude comparative approfondie de l'impact des différentes techniques d'alignement de trajectoires sur l'algorithme de génération de la trajectoire médiane reste à réaliser.

Références

- Alt, H. et M. Godau (1995). Computing the fréchet distance between two polygonal curves. *International Journal of Computational Geometry and Applications* 5(1), 75–91.
- Berndt, D. et J. Clifford (1994). Using dynamic time warping to find patterns in time series. In *AAAI94 workshop on knowledge discovery in databases*, pp. 359–370.
- Buchin, K., M. Buchin, M. van Kreveld, M. Löffler, R. Silveira, C. Wenk, et L. Wiratma (2010). Median trajectories. *Algorithms–ESA 2010*, 463–474.
- Chen, L. et R. Ng (2004). On the marriage of lp-norms and edit distance. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pp. 792–803. VLDB Endowment.

Trajectoires médianes

- Devogele, T. (2002). A new merging process for data integration based on the discrete fréchet distance. In *Advances in Spatial Data Handling : 10th International Symposium on Spatial Data Handling*, pp. 167–181. Springer Verlag.
- Eiter, T. et H. Mannila (1994). Computing discrete fréchet distance. Technical report, Technische Universität Wien.
- Etienne, L. (2011). *Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime*. Ph. D. thesis, Université de Bretagne Occidentale.
- Freeman, R., R. Mann, T. Guilford, et D. Biro (2011). Group decisions and individual differences : route fidelity predicts flight leadership in homing pigeons. *Biology letters* 7.1, 63–66.
- Kaufman, L. et P. Rousseeuw (1987). *Clustering by means of medoids*, pp. 405–416.
- Marteau, P.-F. (2009). Time warp edit distance with stiffness adjustment for time series matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(2), 306–318.
- Meratnia, N. et R. A. de By (2004). Spatiotemporal compression techniques for moving point objects. In *Advances in Database Technology - EDBT 2004*, Volume 2992 of *Lecture Notes in Computer Science*, pp. 561–562. Springer Berlin / Heidelberg.
- Petitjean, F., A. Ketterlin, et P. Gançarski (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition* 44(3), 678–693.
- Pettit, B., A. Perna, D. Biro, et D. J. Sumpter (2013). Interaction rules underlying group decisions in homing pigeons. *Journal of The Royal Society Interface* 10(89).
- Puri, M. L. et P. K. Sen (1971). *Nonparametric methods in multivariate analysis*. John Wiley & Sons.
- Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Review/Revue Internationale de Statistique* 58, 263–277.
- Struyf, A., M. Hubert, et P. Rousseeuw (1997). Clustering in an object-oriented environment. *Journal of Statistical Software* 1(4), 1–30.
- Vermeulen, T. (2013). Algorithms for finding a middle trajectory. Master's thesis, Eindhoven University of Technology.

Summary

Many different huge trajectories database are nowadays available. These trajectories are collected in different contexts (pedestrians, vehicles, ships, planes, animals...). This article introduce a robust algorithm to compute median trajectory that can be applied to these different contexts. This median trajectory depicts the spatio-temporal behaviour of a cluster of trajectory moving between two zones of interest. The algorithm rely on trajectories positions matching using discrete Fréchet distance. Two important trajectories datasets illustrates how this algorithm perform on real world examples.