L' étude de la fréquence d' échantillonnage des mouvements des humains

Panagiota Katsikouli, Aline Carneiro Viana, Marco Fiore, Alberto Tarable

L'étude de la fréquence d'échantillonnage des mouvements des humains

Panagiota Katsikouli[1] et Aline C. Viana[2] et Marco Fiore[3] et Alberto Tarable[3]

[1]*Université de Lyon, Inria, INSA-Lyon, CITI,* [2]*Inria,* [3]*CRN-IEIIT*

Des études récentes ont profité des techniques de suivi basées sur les technologies de positionnement d'étudier afin la mobilité humaine. Ces recherches ont révélé, entre autres, une grande régularité spatio-temporelle des mouvements individuels. Sur la base de ces résultats, nous visons à répondre à la question "à quelle fréquence doit-on échantillonner les mouvements humains individuels afin qu'ils puissent être reconstruits à partir des échantillons recueillis avec un minimum de perte d'information ?". Notre quête d'une réponse a conduit à la découverte de (i) propriétés spectrales apparemment universelles de la mobilité humaine, et (ii) une loi de mise à l'échelle linéaire de l'erreur de localisation par rapport à l'intervalle d'échantillonnage. Nos résultats sont basés sur l'analyse des trajectoires GPS de 119 utilisateurs dans le monde entier. Les applications de nos résultats sont liées à un certain nombre de domaines pertinents pour l'informatique omniprésente, tels que l'informatique mobile économe en énergie, les opérations de service basées sur l'emplacement, le sondage actif des positions des abonnés dans les réseaux mobiles et la compression des données de trajectoire.

Mots-clefs : Mobilité humaine, trajectoires spatio-temporelles, fréquence d'échantillonnage

---

1 Introduction

Over the past few years, the pervasive usage of smart devices and location-tracking systems has made it possible to study and understand human mobility at unprecedented scales. Fig. 1, shows heatmaps and time-series of the locations visited by two random users in our employed dataset. Although these plots convey three weeks of data, a small set of frequently visited places emerges along with systematic paths connecting them. We investigate whether the regularity of human mobility entails the possibility of sampling individual movements at reduced constant frequencies, while allowing for the reconstruction of trajectories that retain a vast portion –if not all– of their original level of detail. Identifying suitable frequencies for human mobility sampling would have applications in a number of fields, including but not limited to *(i)* mobile computing, where overly frequent GPS localization unnecessarily reduces the battery life of mobile devices, *(ii)* location-based service design, where unwarranted users' position data collection raises significant privacy concerns, *(iii)* cellular networks, where active probing of subscribers' positions is a costly task whose rate must be duly optimized, *(iv)* trajectory data compression, where information loss must be minimized.

Our problem is equivalent to asking *"at what frequency should one periodically sample individual human movements so that they can be reconstructed from the collected samples with minimum loss of information ?"*. To respond, we consider mobility patterns as signals over time, and carry out a spectral analysis of human mobility. We find that the spectra of the movements of 119 individuals have very similar, flat shapes that suggest the absence of convenient sampling frequency thresholds. We then carry out a quantitative analysis of the user localization error in movements reconstructed from regular sampling at different periodicities. Our results unveil a linear scaling law of the error with respect to the span of the constant sampling interval. This law corroborates the outcome of the spectral analysis and has significant practical implications, as it controls the trade-off between accuracy and cost of measurements of human mobility.

Related Works. The objective in spatial data trajectory compression is to maintain the trajectory shape (see [KSG14, MHLR10] and references therein). We aim at also preserving the temporal dimension of movements. Our problem is also different from sampling to detect important locations [TCK13, CBT$^+$13].

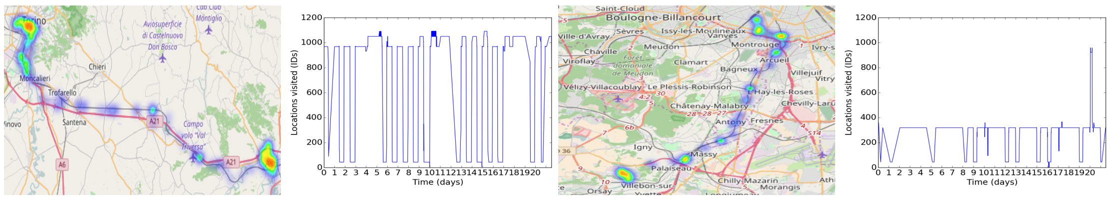


FIGURE 1: Heatmaps and time-series of locations visited by two distinct users during three weeks : humans tend to commute between a limited set of specific locations and revisit locations in a periodic fashion.

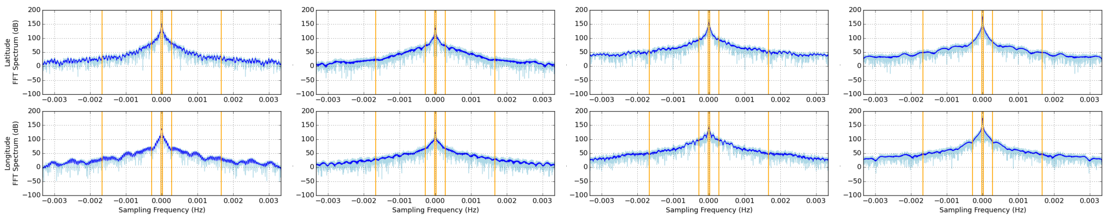


FIGURE 2: Frequency spectra of unidimensional movement signals, for a selection of one-week trajectories.

Indeed, we are not interested in simplifying a pre-recorded GPS trajectory but to find convenient sampling frequencies for human trajectory data. To the best of our knowledge, this is the first work to thoroughly study the problem of finding a good constant sampling frequency at which to sample human mobility so that users' complete movements can be accurately reconstructed.

2 Reference dataset

Our study employs a dataset of real-world individual mobility data extracted from three different sources.

- MACACO data : collected as part of the EU CHIST-ERA MACACO project †
- OpenStreetMap (OSM) data : collected by volunteers who recorded and uploaded their trajectories as a contribution to the OSM database ‡
- Geolife data : collected in Beijing by Microsoft Research Asia [ZLC$^+$08].

The GPS trajectories in our dataset represent human mobility with various means of transportation (e.g., walking, trams, airtravel, cycling etc.) and at various scales ; they cover sensibly different geographical and temporal spans with the quality of the data for a single user being typically very heterogeneous over time, with periods of days or weeks where GPS logs are completely absent. To build a consistent reference dataset, we segmented the mobility traces of all users into one-week trajectories, and analysed them separately. During each week, we bounded the mobility of each individual to the regions where the activity is concentrated. Then, we retained only the trajectories that contain complete GPS records in at least six out of seven distinct week days, resulting in 1,052 weeks of mobiliy. We finally argue that the sampling intervals in the one-week trajectories of our reference dataset, which are upto 15 minutes, are sufficient to capture human movements, and we consider them as our ground-truth in the remainder of the study.

3 Spectral analysis of human mobility

From a spectral analysis viewpoint, answering the question *"at what frequency should one periodically sample individual human movements so that they can be reconstructed from the collected samples with minimum loss of information ?"*, means to consider human movements as a signal in time, and study its spectrum in frequency by applying Fast Fourier Transform.

First, we need to transform individual GPS trajectories into unidimensional time series. We opted for a parallel study of the two dimensions of the geographical space, by considering them in isolation. Instead of

†. https ://macaco.inria.fr/macacoapp/
‡. https ://www.openstreetmap.org

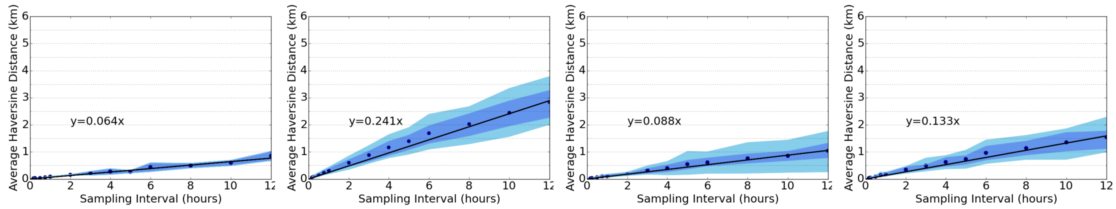


FIGURE 3: Average Haversine distance between the original and the reconstructed trajectories of four users, versus sampling intervals between 10 minutes and 12 hours.

using the absolute values of latitude and longitude, we replace them with the signed latitude and longitude displacements from the corresponding centre of mass of the one-week trajectory. Our transformations have the property of generating zero-mean signals whose frequency spectra have no DC components. We also analysed the correlation between the isolated latitude or longitude displacements and the actual travelled distance in the bidimensional space. The results we obtained let us conclude that both dimensions, when taken separately, still provide decent approximations of the overall mobility.

Fig. 2 shows the spectra of the latitude (top) and longitude (bottom) displacement signals of a representative selection of one-week trajectories. The original spectra are in light blue, while a moving-average that better displays the overall trends is in dark blue. Vertical orange lines outline the frequencies that correspond to sampling intervals of 10 minutes (farthest from the central frequency), 1 hour and 12 hours (closest to the central frequency). We make two important remarks : *(i)* despite the diversity across the different one-week trajectories, all spectra have very similar shapes ; *(ii)* the shapes do not show evidence of a bandwidth threshold beyond which the spectra become clearly negligible, making it impossible to identify an operational point for effective sampling. We found that the observations above hold in the overwhelming majority of our user-base. We can explain both facts by considering that human mobility is a sequence of long periods where individuals are almost static and fast transitions between such important locations. While positions during stationary time intervals contribute to low-frequency spectral components and are hence easily captured by a sparse sampling, travelling causes discontinuities in the mobility signal and is much harder to sample. Thus, the spectra do not reveal whether, e.g., collecting samples at every 10 minutes is obviously more efficient than sampling at every hour.

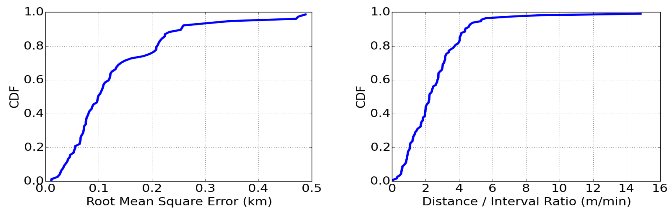
4 A quantitative analysis of mobility sampling

We experimentally investigate the exact trade-off between the quality and cost of sampling in the context of human mobility. We create downsampled versions of the one-week trajectories in our reference dataset, using a wide range of sampling intervals, from 10 minutes to 12 hours. We then linearly interpolated the samples, and assess how such reconstructed trajectories compare to the original ones. We measure the error in retrieving a complete individual trajectory from sampled data by computing the average Haversine distance, which is the mean of all Haversine distances between the points of the reconstructed and original mobility recorded at the same time instant.

Fig. 3 shows the evolution of the average Haversine error against the sampling interval, for a representative set of individuals in our reference dataset. Each plot presents results for all of the one-week trajectories of a specific user : as multiple one-week trajectories are aggregated in every plot, we outline the mean (dots), 25-75% quantiles (dark blue region) and 10-90% quantiles (light blue region) of the error measured over all trajectories of one user.

A surprisingly clear linear relationship characterizes all curves. Fittings on a simple linear model (solid lines in Fig. 3) show an excellent match for all users in our dataset. Fig. 4a portrays the CDF of the Root Mean Square Error (RMSE) between the linear fitting and the mean values (solid lines and dots respectively in Fig. 3) of the average Haversine distance, for all users, over sampling intervals that range from 10 minutes to 12 hours. The probability mass of the distribution is below 250 meters – a very reasonable RMSE for people travelling tens of km per day.

We also highlight that the slope α characterizes the ratio between the average Haversine distance and the



(a) Quality of fitting (b) Fitting slope across users

FIGURE 4: (a) Distribution of the RMSE due to the linear approximation of the relation between the average Haversine distance and the sampling interval. (b) Distribution of the ratio between the average Haversine distance and the sampling interval. Plots are for all users.

sampling interval, or, equivalently, it explains the mean additional error of the reconstructed trajectory when increasing the time that intercurs between samples. Hence, it can be measured in meters per minute (m/min). When looking at the value of α, we remark that it is not identical across users. We study the heterogeneity of α in Fig. 4b, which portrays the CDF of the distance/interval ratio associated to all individuals in our reference dataset. Over 90% of users have slopes that are uniformly distributed between 1 and 4 m/min. Hence, *for the vast majority of individuals, the inaccuracy of their recorded trajectory grows of 1 to 4 meters for each minute added to their movement sampling interval.*

5 Discussion and Conclusions

Summarizing our findings, we assert that *the average error incurred by trajectories reconstructed from periodic samples scales linearly with the constant sampling interval*. This result is well aligned with the outcome of our spectral analysis in Sec. 3 : the linearity of the relationship between error and sampling interval explains the absence of an operational point for the effective sampling of human movements. In addition, we find that the linear scaling law is characterized by a comparable parameter, i.e., the error-to-interval ratio, across all our user base depending on the individual, the error typically grows 1 to 4 meters when adding one minute to the inter-sample time. We are currently investigating the physical reasons behind the diversity among different individuals as well as working on an adaptive algorithm that will tailor the sampling frequency to the user mobility over time. A full version of this work is published in IEEE Globecom 2017.

Références

[CBT+13] Balázs Cs. Csáji, Arnaud Browet, V.A. Traag, Jean-Charles Delvenne, Etienne Huens, Paul Van Dooren, Zbigniew Smoreda, and Vincent D. Blondel. Exploring the mobility of mobile phone users. *Physica A : Statistical Mechanics and its Applications*, 392(6) :1459 – 1473, 2013.

[KSG14] Panagiota Katsikouli, Rik Sarkar, and Jie Gao. Persistence based online signal and trajectory simplification for mobile devices. In *Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '14, pages 371–380, New York, NY, USA, 2014. ACM.

[MHLR10] Jonathan Muckell, Jeong-Hyon Hwang, Catherine T. Lawson, and S. S. Ravi. Algorithms for compressing gps trajectory data : An empirical evaluation. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, pages 402–405, New York, NY, USA, 2010. ACM.

[TCK13] Benoit Thierry, Basile Chaix, and Yan Kestens. Detecting activity locations from raw gps data : a novel kernel-based algorithm. *International Journal of Health Geographics*, 12(1) :14, 2013.

[ZLC+08] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. Understanding mobility based on gps data. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 312–321, New York, NY, USA, 2008. ACM.