

## Thesis's subject

### **Mining and Analysis of enriched trajectories – Application to the trajectories of the visitors of a museum**

### **Fouille et Analyse de trajectoires enrichies – application aux trajectoires des visiteurs d'un musée**

#### Summary

Movement of users in indoor settings introduces new challenges, which are not addressed by the existing research in the area. The challenges stem from the fact that we have the ability to capture and analyze massive users' trajectories that take place indoors, but also to provide and link them with enriched information due to the advanced technological devices that can provide additional data on user's activities. We have also the ability to process now massive datasets of user paths (trajectories) of their visits. One application domain is museums, as visitors are offered aid through devices (e.g. audio guides, mobile applications) and the data collected can be used to allow museums to learn more about their visitors and their visiting trajectories and behavior patterns.

So the main challenges of the thesis are related to the:

- modeling of users' trajectories in indoor environments (geometric level), enriched with information about the visitors and their activities (e.g. use of related multimedia, consultation of texts or other descriptive works, etc.) (symbolic and semantic level),
- development of analytical methods to analyze these enriched trajectories that would also scale to handle the increased size of the data. These methods might extend the search techniques of spatiotemporal data for the detection of recurrent movement patterns to indoor environments,
- interpretation of test results in a museum setting, in order to better understand visitor's behavior,
- suggest in real time paths or points to visit based on the analysis of user's behavior.

The work builds on a collaboration with the Louvre, which will provide the data, support and validation of the use cases' analysis and interpretation.

#### Context

The thesis takes place within the "Trajectories" project, funded by the Heritage Science Foundation (LabEx Patrima) carried by the ETIS laboratory at the University of Cergy-Pontoise (UCP), in collaboration with the DAVID laboratory at the University of the Versailles Saint-Quentin-en-Yvelines (UVSQ), the AGORA laboratory (UCP) and the Louvre Museum. The thesis will start in October 2016 will be co-directed by Dimitris Kotzinos (ETIS) and Karine Zeitouni (DAVID), in collaboration with all project partners.

#### Application

Applicants should send a CV, a cover letter and transcripts of their academic qualifications (including a master thesis and/or any published scientific papers, if available) in PDF format to [Dimitrios.Kotzinos@u-cergy.fr](mailto:Dimitrios.Kotzinos@u-cergy.fr) in an e-mail titled "PhD Museum Trajectories". They can provide, if they wish, letters of recommendation. Further enquiries are also welcome. Knowledge of French is not a requirement.

A first selection will be made for applications arrived before 20 July 2016 if necessary a second selection will take place in September.

## Detailed Description

### Context

Museums need to better know their audience, to check the relevance of the system in place and improve access to information for visitors during the visit. New electronic guides (usually in the form of mobile phones apps) tend to go beyond the limited scope of conventional audio guides and offer advanced capabilities based on the real-time location of the visitor, allowing him/her to obtain contextual information, customize the visiting path or schedule, etc. The Louvre Museum is already part of such activities through the guide using Nintendo 3DS and the Pyramid smartphone application will be added in the summer of 2016. In both cases, apart from information on visitor profiles and content information related to the visit, there is the capability for automatic collection of location data, as positions in the museum space at specific times, forming a trajectory of the visit. The basic idea of this thesis is to use these trajectories as a central element in the analysis of visitor behavior.

### Scientific challenges

We distinguish in the literature three types of moving objects trajectory models: *geometric*, formed by positions specified as coordinates; *symbolic*, formed by positions specified as reference to an object in the environment known in advance, and *semantic*, enriched with information describing their context / domain [1]. We are interested in modeling users' trajectories combining symbolic and semantic approaches in the museum context. A particular challenge is the design of several models with semantic aspects (location, type of work) and several levels of granularity in both space (work, room, wing) or meaning (painting, Italian painting, Italian painting XIII fifteenth century), allowing analysis at several levels.

Behavioral analysis methods based on trajectories [2] have been proposed for geometric or symbolic trajectories [3], but the integration of semantic dimensions related to the visitor, to his actions, the environment remains a challenge. To this is added the aspect of processing massive data sets, which means that we need to develop trajectories' mining algorithms that scale well with the number of trajectories and the size and type of desired patterns.

In a closed environment (indoor), modeling and analysis are made more difficult by the absence of reference maps/points and the unavailability of space models. In the state of the art, only some mining approaches of frequent patterns [4] have been proposed. Moreover, the difficulty of having a precise location relative to points of interest and navigation constraints make the development of models in these environments more complex [5]. The developed model should be able to represent the uncertainty of the indoor location and should be generalizable to different contexts related to the accuracy of the tracking device and the semantic model.

### Objectives

- Develop a model of users' trajectories in the museum, at the symbolic and semantic level. Consider several semantic facets and multiple levels of granularity in this model. Consider the uncertainty of indoor location while modeling the geometries of the trajectories.
- Develop methods of analysis for this type of trajectories, focusing on the search for recurring patterns. Develop methods to enrich the position with several semantic dimensions, e.g. the characteristics of places and visitors, visitor actions, etc. Develop methods of mass analysis that scale well and their efficiency can be validated experimentally. Model the different types of the visitors' profiles, define measures and develop interpretations for deviations for these types of profiles.

- Explore the extension of these models and methods to the analysis of the trajectories in real time, to dynamically suggest a suite of courses, business objectives, content consulting, etc., depending on the visitors' behavior(s).

#### Références

- [1] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot. 2008. A conceptual view on trajectories. *Data Knowl. Eng.* 65, 1 (April 2008), 126-146.
- [2] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, Y. Theodoridis, and Z. Yan. 2013. Semantic trajectories modeling and analysis. *ACM Comput. Surv.* 45, 4, Article 42 (August 2013),
- [3] Kharrat A., Sandu Popa I., Zeitouni K., Faiz S., Clustering Algorithm for Network Constraint Trajectories, in 13th International Symposium on Spatial Data Handling, SDH 2008, pp.631-647.
- [4] Li, K. J., Yoo, S. J., Han, Y. S., Geocoding Scheme for Multimedia in Indoor Space, In Proc. ACM SIGSPATIAL Conference, pp. 434--437, ACM, 2013
- [5] Lu, H., Guo, C., Yang, B., & Jensen, C. S. Finding Frequently Visited Indoor POIs Using Symbolic Indoor Tracking Data. In Proc. of 19th International Conference on Extending Database Technology (EDBT), 2016, pp. 449-460.

## Sujet de la thèse

### **Fouille et Analyse de trajectoires enrichies – application aux trajectoires des visiteurs d’un musée**

#### Résumé

L’analyse du déplacement des utilisateurs en milieu fermé (indoor) introduit de nouveaux défis par rapport à la problématique de recherche dans ce domaine. Ces défis viennent en partie des modèles spatiaux contraints spécifiques à l’indoor, mais surtout de la richesse d’informations fournies par les dispositifs associés à ces déplacements. Par exemple, les musées offrent au public des équipements d’aide à la visite (audioguides, applications mobiles) et sont intéressés à exploiter les données d’utilisation recueillies pour mieux connaître leurs visiteurs et comprendre leur comportement. Ces dispositifs captent non seulement la localisation en temps réel des visiteurs, ce qui permet de reconstituer les trajectoires de visite, mais aussi des informations sur les activités pendant le déplacement, par exemple la consultation des contenus multimédia décrivant les œuvres.

Les principaux défis de la thèse concernent :

- La modélisation des trajectoires de visite en milieu fermé (au niveau géométrique), enrichies avec des informations sur les visiteurs et leurs activités (au niveau symbolique et sémantique),
- L’élaboration de méthodes d’analyse et fouille de ces trajectoires enrichies qui passent à l’échelle lorsqu’on analyse un très grand nombre de trajectoires. Ces méthodes doivent étendre les techniques de fouille de données spatio-temporelles pour la détection de motifs de déplacement récurrents en milieu fermé.
- L’interprétation des résultats d’analyse dans un cadre muséal afin de mieux comprendre le comportement des visiteurs.
- La suggestion en temps réel de parcours ou points de visite, basée sur l’analyse du comportement de l’utilisateur.

Le travail sera réalisé en collaboration avec le musée du Louvre, qui fournit les données de visite, la problématique d’analyse et l’expertise d’interprétation.

#### Cadre

La thèse se déroule dans le cadre du projet Trajectoires, financé par la Fondation des Sciences du Patrimoine (LabEx Patrima), porté par le laboratoire ETIS de l’Université de Cergy-Pontoise (UCP), en collaboration avec le laboratoire DAVID (UVSQ), le laboratoire AGORA (UCP) et le Musée du Louvre. La thèse débutera en octobre 2016 et sera codirigée par Dimitris Kotzinos (ETIS) et Karine Zeitouni (DAVID), en collaboration avec tous les partenaires du projet.

#### Candidature

Les candidats doivent envoyer un CV et une lettre de motivation à [Dimitrios.Kotzinos@u-cergy.fr](mailto:Dimitrios.Kotzinos@u-cergy.fr) un dans un email intitulé “[PhD Museum Trajectories]”. Ils peuvent fournir, s’ils le souhaitent, des lettres de recommandation et tout élément lié à leurs travaux de recherche. Une première sélection sera faite pour les candidatures arrivées avant le 20 juillet 2016, si nécessaire une seconde sélection aura lieu en septembre.

## Descriptif détaillé

### Contexte

Les musées ont besoin de mieux connaître leur public, afin de vérifier la pertinence du dispositif de visite mis en place et d'améliorer l'accès à l'information durant la visite. Les nouveaux guides tendent à dépasser le cadre limité des audioguides classiques et à offrir des applications avancées, s'appuyant sur une localisation en temps réel du visiteur, lui permettant de s'orienter, d'obtenir une information contextuelle, de personnaliser les parcours, etc. Le musée du Louvre s'inscrit déjà dans cette tendance à travers le guide sur console Nintendo 3DS et l'application Pyramide sur smartphone (été 2016). Dans les deux cas, en dehors des informations sur le profil du visiteur et sur les contenus consultés, on dispose de données de localisation automatique, sous forme de positions dans l'espace muséal à des moments précis, formant une trajectoire de visite. L'idée de base de cette thèse est d'utiliser les trajectoires comme élément central dans l'analyse du comportement des visiteurs.

### Problématique

On distingue dans la littérature trois types de modèles de trajectoires d'objets mobiles : *géométrique*, formé des positions horodatées, *symbolique*, par référence à un environnement géographique connu d'avance, et *sémantique*, enrichi avec des informations décrivant leur contexte/domaine [1]. Nous nous intéressons ici à la modélisation des trajectoires de visite en combinant les approches symbolique et sémantique dans le contexte muséal. Un défi particulier est la conception de modèles à plusieurs facettes sémantiques (localisation, type d'œuvres) et plusieurs niveaux de granularité à la fois dans l'espace (œuvre, salle, aile) ou sémantique (peinture, peinture italienne, peinture italienne XIII-XV siècle), permettant des analyses à plusieurs niveaux.

Des méthodes d'analyse du comportement basées sur des trajectoires [2] ont été proposées pour des trajectoires géométriques ou symboliques [3], mais l'intégration des dimensions sémantiques liées au visiteur, à ses actions, à l'environnement, reste un défi. A cela se rajoute l'aspect masse de données, avec la nécessité de développer des algorithmes de fouille de trajectoires qui passent à l'échelle avec le nombre de trajectoires et la taille des motifs recherchés.

En milieu fermé (indoor), la modélisation et l'analyse sont rendues plus difficiles par l'indisponibilité de cartes de référence et de modèles de l'espace. Dans l'état de l'art, seules quelques approches de fouille de points d'intérêt fréquents [4] ont été proposées. De plus, la difficulté d'avoir une localisation précise par rapport aux points d'intérêt, ainsi que les contraintes de navigation rendent l'élaboration de modèles dans ces environnements plus complexe [5]. Le modèle élaboré doit permettre de représenter l'incertitude de localisation indoor et être généralisable à différents contextes liés à la précision du dispositif de localisation et au modèle sémantique.

### Objectifs

- Concevoir un modèle de trajectoires de visite dans le musée, au niveau symbolique et sémantique. Considérer plusieurs facettes sémantiques et plusieurs niveaux de granularité dans ce modèle. Tenir compte de l'incertitude de localisation indoor dans le modèle de trajectoires.
- Développer des méthodes d'analyse pour ce type de trajectoires, centrées sur la fouille de motifs récurrents. Concevoir des méthodes qui permettent de croiser la position avec plusieurs dimensions sémantiques, par exemple les caractéristiques des endroits et des visiteurs, les actions des visiteurs, etc. Elaborer des méthodes d'analyse massive qui

passent à l'échelle et valider de façon expérimentale leur efficacité. Modéliser des profils types de visiteurs, définir des mesures et élaborer des interprétations pour les déviations relatives à ces profils types.

- Explorer l'extension de ces modèles et méthodes à l'analyse de la trajectoire en temps réel, afin de pouvoir suggérer dynamiquement une suite du parcours, des objectifs de visite, des contenus à consulter, etc., en fonction du comportement du visiteur.

#### Références

- [1] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot. 2008. A conceptual view on trajectories. *Data Knowl. Eng.* 65, 1 (April 2008), 126-146.
- [2] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, Y. Theodoridis, and Z. Yan. 2013. Semantic trajectories modeling and analysis. *ACM Comput. Surv.* 45, 4, Article 42 (August 2013),
- [3] Kharrat A., Sandu Popa I., Zeitouni K., Faiz S., Clustering Algorithm for Network Constraint Trajectories, in 13th International Symposium on Spatial Data Handling, SDH 2008, pp.631-647.
- [4] Li, K. J., Yoo, S. J., Han, Y. S., Geocoding Scheme for Multimedia in Indoor Space, In Proc. ACM SIGSPATIAL Conference, pp. 434--437, ACM, 2013
- [5] Lu, H., Guo, C., Yang, B., & Jensen, C. S. Finding Frequently Visited Indoor POIs Using Symbolic Indoor Tracking Data. In Proc. of 19th International Conference on Extending Database Technology (EDBT), 2016, pp. 449-460.