

Techniques de bundling : un cas d'étude pour l'exploration de grandes quantités d'informations

Bundling techniques applied to data exploration use cases

Christophe Hurter

Christophe.hurter@enac.fr

Ecole Nationale de l'Aviation Civile

English Abstract – We present a fast and simple method to compute bundled layouts of general graphs, dynamic graphs and temporal paths. For this, we first transform a given graph drawing into a density map using kernel density estimation. Next, we apply an image sharpening technique which progressively merges local height maxima by moving the convolved graph edges into the height gradient flow. We show how these techniques can produce simplified visualizations of static, streaming and sequence graphs. We illustrate our techniques with datasets from aircraft monitoring, software engineering, and eye-tracking of static and dynamic scenes.

1 INTRODUCTION

L'accès à l'information étant simplifié, de grandes quantités de données sont désormais disponibles dans tous les domaines d'applications. Ces données peuvent être de type personnelles, industrielles, scientifiques. L'exploitation de telles données peut poser des problèmes de par leur taille, et leur temps nécessaire d'analyse. Les outils existants peuvent manquer de réactivité ; il peut même arriver que l'analyse de ces données soit plus longue que leur temps de validité. Cette simple vérité montre à quel point il est important d'utiliser des techniques d'analyse de données pour ainsi extraire rapidement de l'information pertinente.

Cette problématique est d'autant plus vraie dans le domaine aéronautique où les informations à traiter ont un caractère de criticité. Dans le domaine du contrôle aérien, les actions doivent être prises rapidement pour assurer la sécurité des personnes et des biens transportés mais aussi pour fluidifier le trafic aérien (optimisation des coûts d'exploitation).

Dans ce document, nous nous intéressons à un sous ensemble de la visualisation d'information avec la visualisation de trajectoires et de leur exploitation. Ce sous

ensemble propose des challenges complexes comme la densité des informations qui empêche leur affichage direct. Un mois de trafic aérien directement affiché sur un écran ne produit qu'un ensemble compact de pixels uniformément allumés. Des techniques existent pour maximiser l'utilisation de l'espace d'affichage informatique [Keim 2000], [Fekete, Plaisant 2002]. Des nouvelles techniques ont émergés avec le bundling qui permet de simplifier la visualisation en affichant des distorsions dans les données au profit d'une meilleure lisibilité [Holten 2006]. Aujourd'hui de nouveaux algorithmes de bundling ont été développés qui permettent de traiter de grandes quantités de données en temps interactif [Ersoy et al. 2011], [Hurter et al. 2012]. L'utilisation de ce cet algorithm KDEEB (Kernel Density Edge Bundling) s'appuie Mena-shift [Comaniciu and Meer 2002] et ainsi permet le bundling de graphs dynamiques et séquentiels [Hurter et al. 2013].

2 DISCUSSION

Une des définitions possible du bundling est la production d'une visualisation abstraite qui propose un compromis entre des zones avec des fortes densités de données et des zones

avec peu de données. KDEEB utilise ce principe en créant des zones qui attirent les données pour créer des zones avec du vide et d'autre avec un fort taux de recouvrement.

Ce principe permet de clarifier la visualisation en augmentant le ration informations utiles par rapport à l'information inutile (ink ration [Tuft 83]), mais ceci au prix d'une forte déformation des données affichées.

Le bundling n'est pas forcément l'outil optimal pour l'exploration de grandes quantités d'informations mais il produit une vue simplifiée et abstraite et ainsi fait partie des outils prometteurs pour l'exploitation de données.

Dans cette présentation, les techniques de bundling vont être utilisées pour extraire de l'information (figure 1). Un jeu de données de trafic aérien va être exploré pour simplifier sa représentation avec une lentille sémantique interactive [Hurter et al. 2011]. Des données de routes aériennes, graphs temporels, vont être explorées avec une technique de bundling continue [Hurter et al. 2013]. Une illustration de cette même technique va être utilisée pour explorer les évolutions des révisions logiciels dans le domaine de la visualisation de code (révision de Firefox Mozilla) [Hurter et al. 2013]. Enfin, un jeu de données issue d'un eye tracker et d'un pilote en phase d'atterrissage dans un simulateur va être analysé pour extraire le comportement oculaire et valider la conformité du cheminement visuel dans une phase critique de vol (figure 2).

L'utilisation du bundling dans les domaines des trajectoires, de l'eye tracker, de la visualisation logicielle est avérée mais n'est pas limitée à ces seuls domaines. Beaucoup reste à faire pour découvrir d'autres algorithmes de simplification, de comprendre leurs utilisations potentielles. La fin de cette présentation sera dédiée aux enjeux et aux questions de recherche dans le domaine de la simplification de données.

3 REFERENCES

Holten, D. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE TVCG*, 12(5):741–748, 2006.

Hurter, C., Tissoires, B., Conversy, S., 2009. FromDaDy: Spreading Aircraft Trajectories Across Views to Support Iterative Queries. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (November 2009), 1017-1024.

Hurter, C., Ersoy, O., Telea, A., Smooth Bundling of Large Streaming and Sequence Graphs. *Proceedings of IEEE Pacific Visualization 2013 (PacificVis 2013)*.

Hurter, C., Telea, A., Ersoy, O., 2011. MoleView: An Attribute and Structure-Based Semantic Lens for Large Element-Based Plots. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (December 2011), 2600-2609.

Ersoy, O., Hurter, C., Paulovich, F., Cantareiro, G., Telea, A., 2011. Skeleton-Based Edge Bundling for Graph Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (December 2011), 2364-2373.

Keim, D.A. "Designing Pixel-Oriented Visualization Techniques: Theory and Applications", *IEEE Trans. Visualization and Computer Graphics*, 6(1), pp. 1-20, Mar 2000

Fekete, J-D. and Plaisant, C. "Interactive Information Visualization of a Million Items", *Proc. InfoVis'02*, pp. 117-124, 2002, IEEE

Comaniciu, D. and Meer, P.. Mean shift: A robust approach toward feature space analysis. *IEEE TPAMI*, 24(5):603–619, 2002.

Tuft, E.R., *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut, 1983.

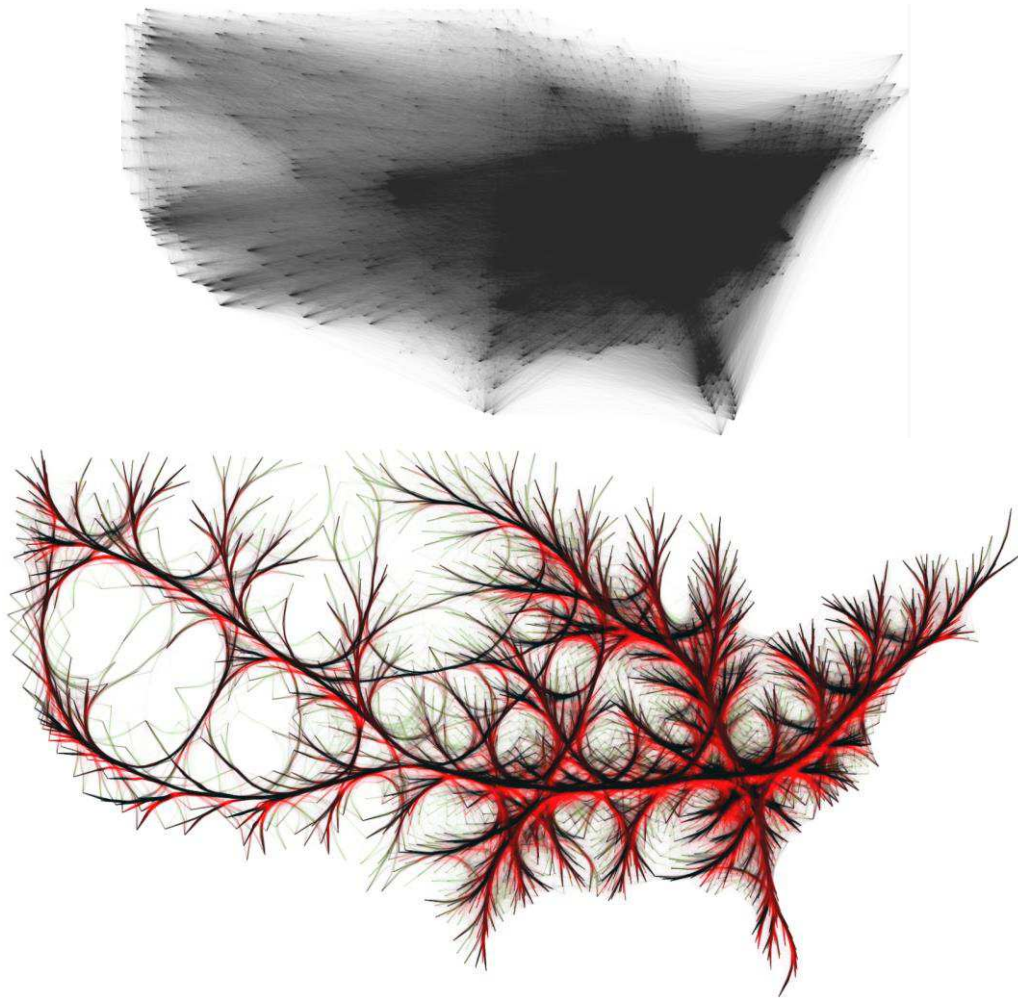


Figure 1 : Données de mouvement de population aux Etat Unis (<http://www.census.gov/population/www/cen2000/ctytoctyflow/>). Ces données proviennent du Census 2000: 5 ans de questionnaires sur les déménagements de personnes entre villes. Version originale en haut, version bundlée en bas.

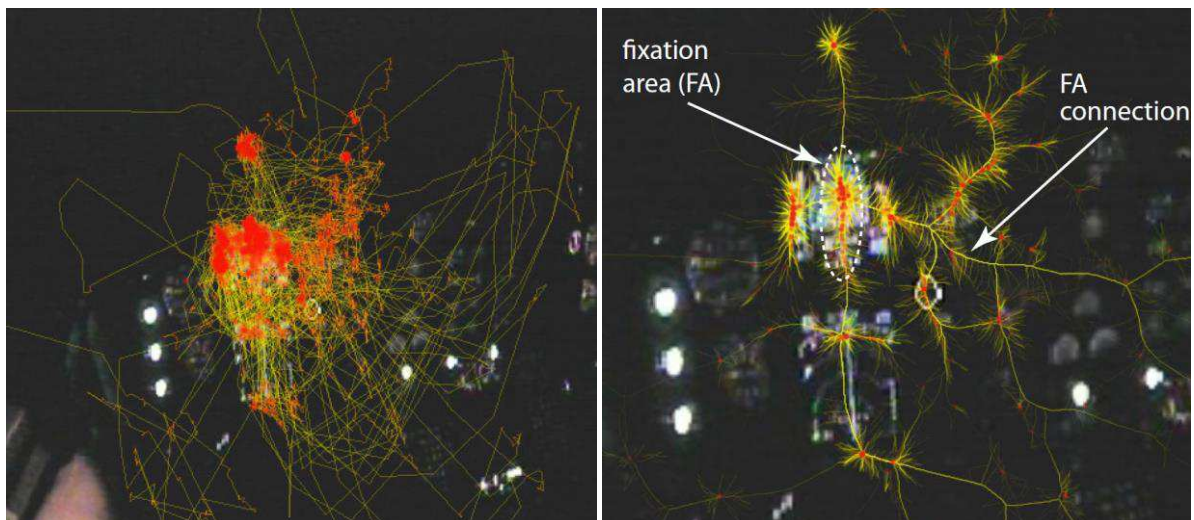


Figure 2 : Analyse de traces oculaires issues d'un pilote, à l'atterrissage en simulation, à l'aide d'outils d'agrégation de trajectoire (bundling, KDEEB [Hurter et al. 2012]).