

# Vers le milliard d'éléments et au-delà

## Billion and Beyond Visualization

**Abstract**—Recent advances in information visualization have shown that building proper structures to allow efficient lookup in the data can reduce significantly the time to build graphical representation of very large data sets, when compared to the linear scanning of the data. We present BiVis, a visualization technique that shows how such techniques can be further improved to reach a rendering time compatible with a continuous interaction. To do so, we turn the lookup into an *anytime* algorithm compatible with a progressive visualization: a visualization presenting an approximation of the data and an estimation of the error can be displayed almost instantaneously and refined in successive frames until the error converges to zero. We also leverages the *spatial coherency* of the navigation: during the interaction, the state of the (possibly partial) lookup for the previous frames is reused to bootstrap the lookup for the next frame despite the view change. We show that those techniques allow the interactive exploration of *out of core* data sets consisting of billions of items.

---

### 1 RÉSUMÉ

Ces dernières années ont vu l'émergence de techniques qui ont renversé le processus traditionnel de la visualisation — faire passer les éléments tour à tour au travers du *pipeline* de la visualisation d'information— pour adopter un modèle pour lequel on va chercher dans des structures d'indexation efficaces les données nécessaires à la construction de chaque pixel de la visualisation. Ces approches permettent d'envisager un passage à l'échelle sur des très gros jeux de données en produisant des techniques dont la complexité croît avec la taille de l'écran et le *log* de la taille des données, plutôt qu'avec la taille des données, ouvrant une porte vers la visualisation interactive de données massives. Dans cette lignée, nous proposons d'avancer encore vers l'idéal d'un rendu en temps interactif pour des jeux de données très volumineux, en vue de permettre leur exploration par une interaction fluide.

Pour cela, nous montrons que ce type de visualisation peut être rendue progressive, c'est-à-dire qu'elle peut être produite incrémentalement, d'abord grossièrement, puis raffinée au cours du temps, et cela au cours de l'exploration interactive de l'utilisateur. Notre contribution consiste donc à rendre un algorithme de visualisation interruptible, fournissant le meilleur résultat possible, ainsi qu'une estimation de l'erreur commise, dans le temps qui lui est imparti. Ce résultat est utilisé pour produire une image encodant ce résultat partiel et son imprécision. Il est aussi réutilisé comme point de départ à l'itération suivante pour affiner poursuivre l'algorithme jusqu'à ce que l'erreur converge vers 0. Cette réutilisation est également exploitée si la vue est modifiée par l'interaction en exploitant la cohérence spatiale de cette dernière.

Concrètement, nous montrons qu'il est possible de visualiser et d'explorer des jeux de données contenant des milliards d'éléments, i.e. ne tenant pas en mémoire. Cette

technique est implémentée dans deux prototypes, l'un montrant des séries temporelles, et l'autre des données géographiques.

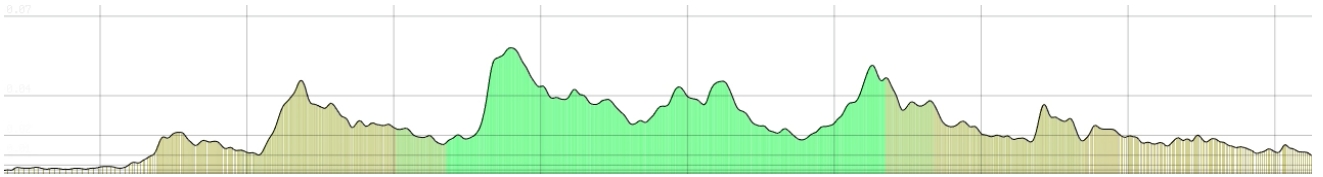


FIGURE 1. Série temporelle comportant des milliards d'évènements visualisée incrémentalement (l'abscisse encode le temps, l'ordonnée le débit de messages, la couleur l'erreur commise par l'approximation).

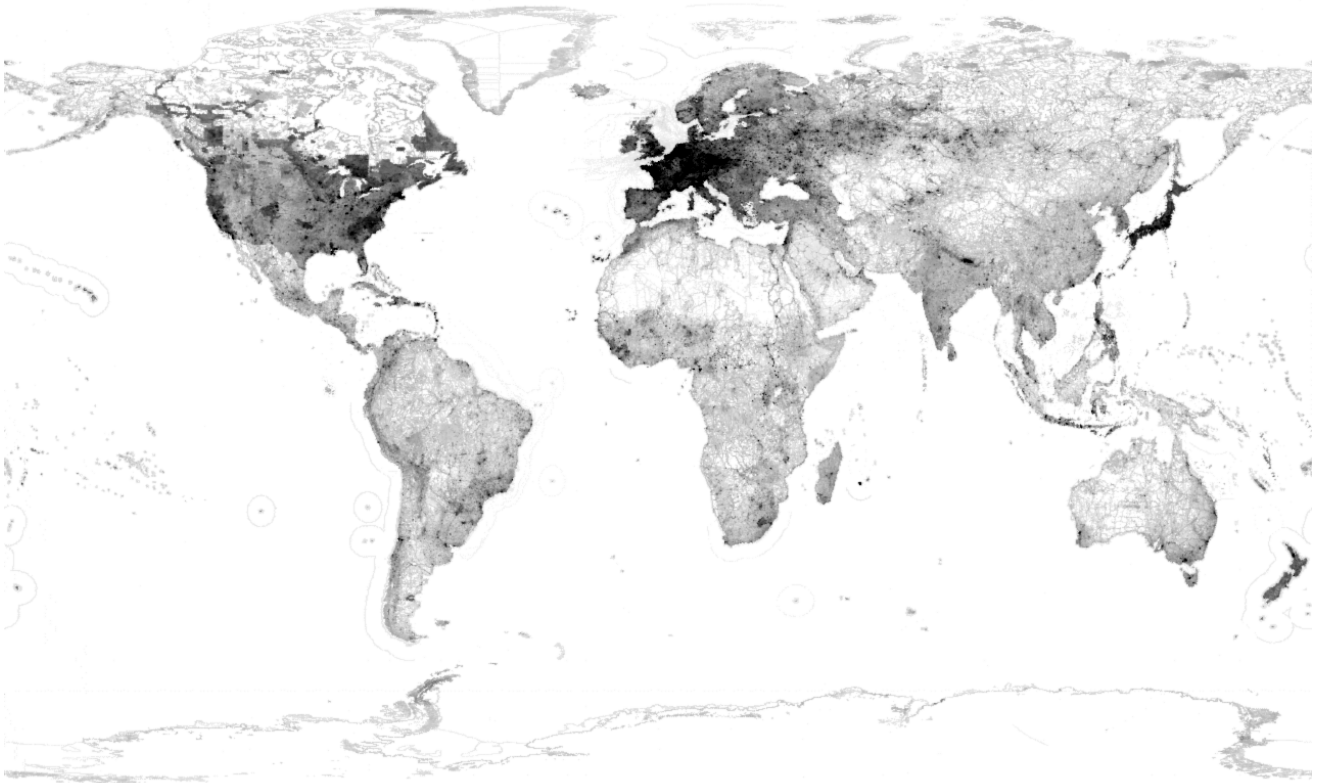


FIGURE 2. Le jeu de données *planet* d'OpenStreetMap (2863 152 191 points) discrétisé dans une image  $1280 \times 800$ ,  $1 \text{ px} \approx 32 \text{ km}$ .