



HAL
open science

Visualisation de données spatiotemporelles imprécises : application en archéologie

Cyril de Runz, Frédéric Blanchard, Philippe Vautrot, Éric Desjardin

► To cite this version:

Cyril de Runz, Frédéric Blanchard, Philippe Vautrot, Éric Desjardin. Visualisation de données spatiotemporelles imprécises : application en archéologie. Atelier Fouille de Données Complexes, conférence Extraction et Gestion des Connaissances 2010, 2010, Hammamet, Tunisie. pp.69–80. hal-00584809

HAL Id: hal-00584809

<https://hal.science/hal-00584809v1>

Submitted on 10 Apr 2011

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.

Visualisation de données spatiotemporelles imprécises : application en archéologie

Cyril de Runz*, Frédéric Blanchard*
Philippe Vautrot*, Eric Desjardin*

* CReSTIC
IUT de Reims Châlons Charleville
Rue des Crayères, BP 1035, 51687 Reims Cedex 2
cyril.de-runz@univ-reims.fr

Résumé. Dans cet article, nous proposons d'exploiter une technique spécifique d'exploration visuelle d'un ensemble d'objets archéologiques dont les composantes spatiales et temporelles sont représentées par des ensembles flous convexes et normalisés. Pour cela, en nous basant sur la définition de vecteurs multidimensionnels issus de défuzzifications ou de comparaison entre deux nombres flous, nous construisons une image couleur dans laquelle chaque pixel représente un objet. L'image couleur donne un rendu synthétique de l'information permettant à l'utilisateur de l'observer et de l'analyser.

1 Introduction

L'étude intuitive et visuelle de l'ensemble des données associées aux objets d'une base de données archéologiques est complexe dans les systèmes d'information géographique (SIG). En effet, bien que l'on puisse actuellement avoir une légende combinant un certain nombre d'attributs, ce nombre est limité. L'exploration visuelle nécessite alors d'utiliser une technique spécifique de visualisation. Pour cela, il faut considérer ces composantes de l'information archéologique comme une collection de données multidimensionnelles (Guptill, 2005).

L'approche générale de l'exploration visuelle de grands volumes de données multicomposantes consiste à présenter un résumé en image de ces informations à l'instar de la démarche proposée par Keim (2000). Afin de visualiser la plus grande quantité d'information possible, nous utilisons une technique de visualisation qui construit une image couleur à partir de ces informations et qui fut introduite dans Blanchard et al. (2005).

Cette technique orientée-pixel consiste à représenter une collection par une image où chaque pixel correspond à une et une seule donnée. Les couleurs des pixels sont déterminées « objectivement ¹ ». La couleur et la spatialisation fournissent alors une image qui constitue un résumé des données en permettant de voir de manière intuitive les principales structures. Ce travail a montré son efficacité sur des bases de données classiques (Blanchard et al., 2005).

Pour cela, les données sont préalablement réduites par une Analyse en Composantes Principales à des données tridimensionnelles regroupant les trois composantes principales. En uti-

1. La couleur est calculée à partir des données sans utilisation d'une échelle de couleurs.

lisant ces données réduites, on associe un pixel de l'image couleur à chaque donnée ; la couleur est affectée objectivement et calculée par la transformée inverse de celle proposée dans (Ohta et al., 1980). Afin de regrouper au maximum, dans l'image de visualisation, les données proches selon leurs trois premières composantes principales, les pixels représentant les données sont organisés spatialement à l'aide d'une courbe de remplissage dite de Peano-Hilbert (Moon et al., 2001).

Dans cet article, deux processus d'exploration visuelle sont étudiés. Le premier a pour but de visualiser les composantes temporelles de l'information archéologique. Ces informations sont difficiles à visualiser dans le cas de grands volumes de données et rendent presque impossible l'exploration intuitive et directe de ces composantes. Dans le second processus, l'objectif est de visualiser les dissimilarités à un objet sélectionné dans la base de données.

Dans le cas du projet SIGRem (Desjardin et de Runz, 2009), les objets de *BDFRues* représentent des tronçons de rues romaines trouvées à Reims. Les composantes temporelles, spatiales et orientationnelles des données sont modélisées en tenant compte de leurs imprécisions par des ensembles flous (de Runz et al., 2008). Il faut donc pré-traiter l'information afin d'en dégager des évaluations quantitatives qui seront dès lors considérées comme des vecteurs multidimensionnels. Ainsi, la technique proposée est appliquée aux vecteurs multidimensionnels pour visualiser les objets archéologiques.

Nous proposons, dans cet article, de visualiser les composantes temporelles des objets de *BDFRues*. Pour cela, comme énoncé précédemment, il est nécessaire de quantifier les données avant même de lancer le processus de visualisation. Dans ce but, nous construisons un vecteur d'évaluation pour chaque nombre flou représentant la période d'activité d'un objet archéologique. Les différentes valeurs de ces vecteurs seront déterminées par différents estimateurs de quantités floues.

Pour visualiser les dissimilarités entre objets archéologiques, les vecteurs multidimensionnels nécessaires sont issus de trois indices de dissimilarité entre objets archéologiques présentés dans de Runz et al. (2008). Ces indices de dissimilarités permettent d'évaluer les dissimilarités temporelles, d'orientation et de localisation entre objets archéologiques. Pour un objet sélectionné, l'image couleur de visualisation regroupe alors spatialement, autour de sa représentation pixel, les pixels associés aux objets qui lui sont le moins dissimilaires d'un point de vue spatial, directionnel et temporel.

Nous présenterons dans la section 2 le contexte applicatif. La section 3 sera dédiée à la description de la technique de visualisation choisie. La section 4 exposera le processus de visualisation des composantes temporelles associées aux tronçons de rues trouvés à Reims datant de l'époque romaine. Dans la section 5, nous proposerons de visualiser la dissimilarité des objets de la base vis-à-vis d'un objet présélectionné.

2 Projet SIGRem

Dans la problématique de la valorisation et de la gestion du patrimoine archéologique, la démarche développée par l'Université de Reims Champagne Ardenne, l'Institut National de Recherches Archéologiques Préventives et Ministère de la Culture et de la Communication dans le Centre Interinstitutionnel de Recherches Archéologiques de Reims peut être considérée comme novatrice par l'intégration de la géomatique au cœur de l'analyse urbaine et régionale.

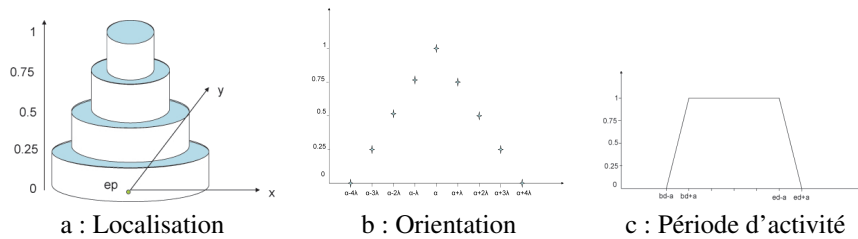


FIG. 1 – Modèles flous pour la localisation, l'orientation et les périodes d'activité des tronçons de rues romaines

Au-delà de l'élaboration de la cartographie archéologique de la cité des Rèmes², le projet SIGRem, soutenu par la région Champagne Ardenne, l'État et la ville de Reims, et cadre applicatif de ce travail, porte sur la mise en place d'un SIG pluridisciplinaire intégrant les données archéologiques recueillies depuis les 30 dernières années. Dans cet article, nous proposons d'appliquer le processus exploratoire proposé sur la base de données *BDFRues*, partie intégrante du projet SIGRem. Cette base est dédiée aux éléments de rues romaines à Reims. Elle est constituée de 33 objets à l'heure actuelle. Son enrichissement est en cours. Les tronçons de rues sont caractérisés par des points ayant une orientation et une période d'activité.

La datation de la période d'activité des objets est généralement issue d'interprétations ou d'estimations dépendantes de l'environnement de la découverte (lieux de fouilles, stratigraphie, comparaison aux objets se situant dans la même pièce. . .). Elle est donc largement imprécise. Le géoréférencement est lui aussi sujet à de l'imprécision liée à différents facteurs : positionnement du point de fouilles, position par rapport à la route, référentiel utilisé, mouvement de terrain. L'orientation de la route est aussi à redéfinir dans ce cadre. En effet, l'orientation est notamment dépendante de la technique d'estimation utilisée à l'époque de la fouille.

Nous représentons les orientations, les périodes d'activité, et les localisations par des ensembles flous convexes et normalisés à savoir respectivement par des nombres flous, des intervalles flous et des ensembles flous spatiaux (2D). On peut ainsi prendre en compte cette incertitude (voir Figure 1).

Afin d'obtenir un rendu synthétique visuel et pertinent de ces données, nous proposons d'exploiter une technique de visualisation orientée-pixel présentée dans la section suivante.

3 Visualisation de données par une image couleur

Afin de fournir une image couleur des objets, nous nous intéressons plus particulièrement à la visualisation statique et plane de données multidimensionnelles quantitatives. L'utilisation des outils de visualisation se heurte alors à deux difficultés principales : la dimension et l'effectif des échantillons de données.

La dimension de l'espace dans lequel se situent les données peut être importante (la dimensionnalité peut être supérieure à 100 dans certains cas). Ceci conduit à un ensemble de phénomènes dissimulant l'information pertinente que l'on recherche. Ces phénomènes sont

2. Cité des Rèmes : Reims et ses environs à l'époque romaine

connus sous le nom de « malédiction de la dimensionnalité » (Donoho, 2000). Par ailleurs, l'effectif de l'échantillon peut être considérable : il peut dépasser le million d'individus. Les techniques de visualisation ont alors tendance à masquer l'information pertinente du fait de cet effectif.

Dans ce cadre, la méthode de Blanchard et al. (2005) est une approche orientée-pixel qui résume les données, et en fournit un résumé sous forme d'une image couleur. Cette approche permet de s'affranchir de la première difficulté par la réduction de la dimensionnalité et de la seconde par l'association d'un pixel à chaque donnée permettant ainsi de visualiser autant de données qu'il y a de pixels affichables.

3.1 Réduction de la dimensionnalité

L'analyse de données multidimensionnelles nécessite une réduction de dimensionnalité pour des raisons pratiques et théoriques (représentations des données, la malédiction de la dimensionnalité). Dans l'approche de visualisation présentée ici, les données sont dans un espace initial de dimension supérieure à trois.

Une approche classique, simple et généralement efficace de la réduction de dimensionnalité est utilisée : on conserve les trois premières composantes générées par une Analyse en Composantes Principales. Une revue des techniques d'ACP est proposée dans Hyvärinen (1999).

Le principe est de projeter les données dans un sous-espace de dimension trois, les axes de projection étant orthogonaux et décorrélés. L'avantage de l'ACP est de déterminer les composantes, itérativement, par ordre décroissant de l'information portée. Ainsi, la première composante contient plus d'information que la seconde, qui en contient plus que la troisième, et ainsi de suite. Ainsi, en réduisant les données de dimension $n > 3$ à des données de dimension 3 par la sélection des trois premières composantes (C_1, C_2, C_3) de l'ACP, on maximise l'information contenue dans ces trois composantes.

Les données réduites guident ensuite le processus de visualisation. À chaque donnée de dimension trois est affecté un pixel que l'on place spatialement dans l'image à l'aide d'une courbe de Peano-Hilbert.

3.2 Remplissage de l'image de visualisation

Pour construire une image d'un échantillon de données, chaque donnée est associée à un pixel de l'image. Cette approche de la visualisation orientée-pixel permet de représenter des échantillons de grande taille (Keim, 2000). La construction de l'image consiste à déterminer les coordonnées des pixels (i.e. des représentations des données) dans l'espace image.

Si les pixels sont placés arbitrairement ou dispersés dans l'image, il devient difficile d'effectuer des rapprochements entre les données. Pour que l'image soit un outil de visualisation efficace, lisible au premier coup d'oeil de manière très intuitive, il faut que les proximités entre données soient faciles à déterminer. Pour cela, il faut que, dans l'image résultat, des données similaires soient spatialement très proches. La construction de l'image s'effectue en deux étapes : les pixels (i.e. les représentations des données) seront d'abord triés de manière à former une suite de pixels successifs ; ensuite cette « ligne » sera utilisée pour remplir l'image.

Ainsi la première étape du remplissage de l'image de visualisation consiste à trier les pixels associés aux données afin de produire une liste de pixels. Le tri des pixels est effectué en utilisant les résultats de la réduction de dimension des données. Les trois composantes obtenues

(C_1, C_2, C_3) donnent trois clefs pour effectuer un tri sur l'ensemble des données de l'échantillon. En effet, les composantes issues de l'ACP sont classées selon la quantité d'information qu'elles fournissent. Ainsi, le tri se fera en majeur sur la première composante car elle contient le plus d'information, puis sur la seconde composante et enfin en mineur sur la troisième composante. La récupération des trois premières composantes de l'ACP et leur tri sont d'ordre polynomial.

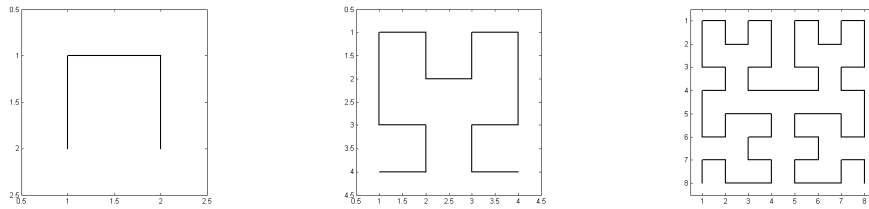


FIG. 2 – Étapes de la construction d'une courbe de Peano-Hilbert

L'étape suivante consiste à remplir l'image avec cette liste de pixels successifs. La courbe de Peano-Hilbert constitue le moyen le plus classique pour effectuer cette construction (Moon et al., 2001) (voir sur la Figure 2 la description de la procédure récursive de construction d'une telle courbe). Le principal avantage de cette courbe est de préserver au mieux les regroupements des classes de données.

Avec ces deux étapes de tri des données puis le remplissage de l'image par une courbe de Peano-Hilbert, on évite de disperser les pixels dans l'image construite. Cette approche tend à préserver la cohérence spatiale des données permettant ainsi une visualisation très intuitive des échantillons de données. Il faut maintenant déterminer la couleur de chaque pixel en fonction des valeurs contenues dans la donnée réduite associée au pixel.

3.3 À propos de la couleur

En imagerie, la couleur est souvent définie par un triplet (R, V, B) de trois valeurs Rouge, Vert et Bleu, codées sur 8 bits (entre 0 et 255). Après avoir réduit la dimension de l'échantillon, chaque donnée est représentée par un triplet (C_1, C_2, C_3) . Cependant, les trois composantes principales ne peuvent former directement le triplet RVB.

La technique de visualisation proposée se base pour l'affectation de la couleur sur l'étude statistique de la couleur de Ohta et al. (1980). Ces derniers proposent d'approximer l'ACP d'une image couleur par une transformation linéaire. A partir des données (R, V, B) des pixels couleur, ils calculent les triplets (C_1, C_2, C_3) qui approximent les trois composantes de l'ACP (Ohta et al., 1980). La technique de visualisation exposée ici cherche à associer une couleur à chaque triplet (C_1, C_2, C_3) . Par la transformation inverse de celle de Ohta *et al.*, elle associe à chaque donnée une couleur (R, V, B) . Ainsi les composantes couleurs R, V et B sont liées aux composantes C_1, C_2 et C_3 par la relation suivante :

$$\begin{cases} R &= (6 \times C_1 + 3 \times C_2 - 2 \times C_3)/6 \\ V &= (3 \times C_1 + 2 \times C_3)/3 \\ B &= (6 \times C_1 - 3 \times C_2 - 2 \times C_3)/6 \end{cases}$$

Visualisation de données spatiotemporelles imprécises

Ce type d'approche présente l'avantage d'être objectif et non supervisé contrairement aux méthodes traditionnelles de détermination de palettes ou d'échelles de couleurs. Cette approche de la couleur dépend de l'échantillon de données. Si l'échantillon change, les couleurs changent. Elle propose un résumé coloré associé à un échantillon. Cette technique de visualisation a été appliquée avec succès à des bases de données classiques, des données simulées et à des images de fluorescence X.

Notre idée est d'utiliser cette technique pour visualiser les objets selon l'information des composantes temporelles, et les dissimilarités des objets archéologiques en tenant compte de leurs imperfections.

4 Visualisation des objets selon leurs périodes d'activité par une image couleur

Les données archéologiques sont spatiotemporelles et imparfaites. Afin de donner une lecture intuitive des données, il est nécessaire de les visualiser. Lorsqu'elles sont stockées dans une base de données associée à un SIG, une visualisation classique consiste à la production d'une ou plusieurs cartes thématiques. Cependant ces cartes ne permettent pas de rapprocher spatialement les objets aux localisations éloignées, et les couleurs dépendent d'une échelle fixée (d'une légende). Afin de rapprocher les données archéologiques selon le temps en prenant en considération l'imperfection, il est nécessaire d'utiliser une autre approche pour la visualisation.

La visualisation d'un grand nombre de données spatiotemporelles imprécises est difficile. En effet, dans le cadre de données représentées par des quantités floues, représenter l'ensemble des fonctions d'appartenance sur un même repère complique la lecture des quantités floues à considérer. Cependant, de nombreuses techniques de visualisation de données multi-composantes, à l'instar de celle présentée précédemment, utilisent les informations quantitatives contenues dans ces données afin de les visualiser. Une solution à la visualisation de quantités floues consiste à visualiser les données par l'intermédiaire d'évaluations des différentes quantités floues. Ces évaluations forment des données quantitatives multi-composantes décrivant l'information à visualiser.

L'analyse de données floues nécessite généralement une défuzzification des données. La défuzzification est le processus qui amène à produire un résultat quantifiable à partir de données floues. Ainsi, par exemple, les méthodes de comparaison de quantités floues rangent le plus souvent celles-ci par le biais d'évaluations (Wang et Kerre, 2001).

Le principe de la visualisation consiste en premier lieu à décrire une quantité floue par plusieurs évaluations quantitatives obtenues avec différentes méthodes de défuzzification. Une quantité floue est alors représentée par un vecteur d'évaluations. Les données sont ensuite visualisées en utilisant les vecteurs d'évaluations.

Dans *BDFRues*, les objets archéologiques sont temporellement modélisés par des nombres flous. Le but de la visualisation est alors d'associer à chaque objet archéologique un pixel couleur de l'image résultat en fonction du nombre flou représentant sa période d'activité.

L'objectif ici est d'abord d'évaluer chaque nombre flou séparément puis de le positionner par rapport aux autres par la technique de visualisation précédente via ses évaluations. Les

évaluations des nombres flous doivent donc ne prendre en compte que le nombre flou devant être visualisé.

4.1 Méthodes de défuzzification des nombres flous

Les méthodes de défuzzification présentées ici ne considèrent que le nombre flou à évaluer. VanLeekwijck et Kerre (1999) les séparent en trois classes. Bien que chaque méthode ait ses particularités, les classes proposées suggèrent à des utilisations différentes. Le choix de l'utilisation de l'une de ces méthodes dépend donc fortement de l'analyse voulue.

Les méthodes de type maxima et les méthodes dérivées forment la première classe. Elles sélectionnent un élément du cœur de la quantité à évaluer comme valeur de défuzzification. Selon Van Leekwijck et Kerre, l'utilisation première de ces méthodes se situe dans le cadre des systèmes de connaissances floues. De plus, ces méthodes sont efficaces d'un point de vue calculatoire.

Dans la seconde classe, les opérateurs de défuzzification convertissent d'abord les fonctions d'appartenance en distribution de probabilités afin de calculer la valeur espérée. Au regard du manque de fondement théorique de ces conversions, la principale raison de leur utilisation est que ces méthodes vérifient l'hypothèse de continuité, essentielle pour les contrôleurs flous.

Dans la troisième classe, les méthodes utilisent les aires sous les fonctions d'appartenance afin d'évaluer les quantités floues. Comme pour les méthodes de la seconde classe, elles sont principalement dédiées au contrôle flou.

Afin d'explorer visuellement les objets archéologiques selon les représentations de leurs périodes d'activité, nous souhaitons définir pour chacun des nombres flous, modélisant la période d'activité de ces objets, un vecteur d'évaluation le représentant dans le processus de visualisation.

4.2 Construction du vecteur multidimensionnel d'évaluation de la représentation d'une période d'activité

Afin de construire simplement un vecteur avec des méthodes de chaque classe, nous proposons de n'utiliser que des méthodes de défuzzification sélectionnées parmi celles présentées dans VanLeekwijck et Kerre (1999) afin qu'elles ne prennent pas en considération de paramètre autre que l'ensemble à considérer, l'objectif final étant de proposer une visualisation temporelle non supervisée.

Pour la première classe, nous choisissons les méthodes suivantes : le "first of maximum" (FOM) qui retourne le plus petit élément du cœur d'un nombre flou ; le "last of maximum" (LOM) qui renvoie le plus grand élément du cœur d'un nombre flou ; le "middle of maximum" (MOM) qui permet de récupérer l'élément médian du cœur d'un nombre flou.

En ce qui concerne la seconde classe, nous sélectionnons les méthodes suivantes : le "center of gravity" (COG) qui donne en sortie le centre de gravité de la fonction d'appartenance d'un nombre flou ; le "mean of maxima" (MeOM) qui calcule la moyenne du cœur d'un nombre flou ; le "mean of support" (MeOS) par lequel on obtient la moyenne du support d'un nombre flou.

Enfin, pour la dernière classe, nous prenons le "center of area" (COA) car celui-ci permet d'obtenir l'élément du support minimisant la différence des aires de la fonction d'appartenance avant et après ce dernier.

Visualisation de données spatiotemporelles imprécises

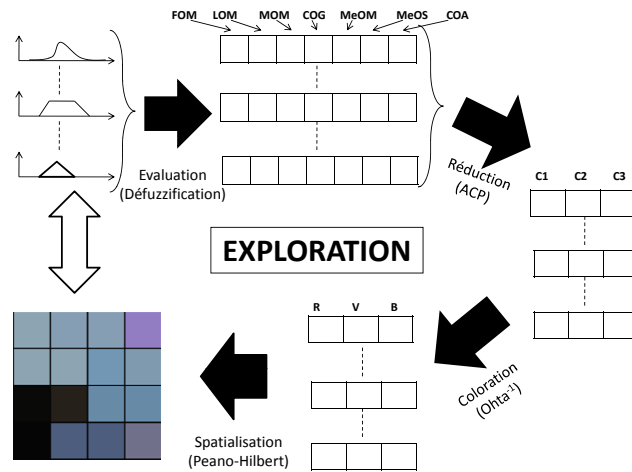


FIG. 3 – Visualisation des objets selon les représentations floues des périodes d'activité — schéma récapitulatif

Nous associons donc à chaque période d'activité un vecteur d'évaluation de dimension 7. C'est par le prisme de ce vecteur que nous explorons les données. Pour cela, nous utilisons l'ensemble des vecteurs en entrée du processus de visualisation de Blanchard et Herbin. Le processus général de l'exploration est présenté dans la figure 3.

4.3 Visualisation des objets selon les représentations de leurs périodes d'activité

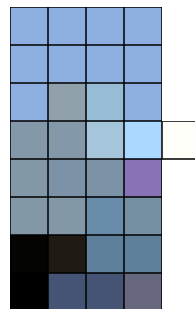


FIG. 4 – Visualisation des objets de BDFRues par une image couleur selon les représentations de leurs périodes d'activité

L'image résultat est présentée sur la figure 4. Elle contient 33 pixels en couleurs. Chaque pixel représente un objet selon l'ensemble flou associé à la période d'activité de l'objet. L'organisation spatiale et l'information couleur des pixels permettent d'observer de façon immédiate des informations de structuration de cet ensemble de périodes. Cette image suggère des regroupements des données par couleurs semblables.

Les regroupements observés correspondent à des objets ayant des représentations de leurs périodes d'activité de profils proches. En effet, les objets, dont les représentations des périodes d'activité ont les plus larges supports, sont visualisés par des pixels de couleur bleue claire (haut de l'image de visualisation), tandis que ceux dont les cœurs des représentations des périodes d'activité sont de cardinalité moyenne, sont coloriés dans les gris (milieu de l'image).

Par ailleurs, les composantes principales calculées sur l'ensemble des vecteurs décrivant les représentations des périodes d'activité des objets issus de *BDFRues* permettent d'expliquer plus de 99% de la variance totale. Cette opération de projection conserve donc la quasi totalité de l'information apportée par les différentes évaluations. La visualisation porte donc sur l'essentiel de l'information temporelle contenue dans *BDFRues*.

Ainsi, l'image couleur résultant de la visualisation permet une lecture intuitive — par proximité spatiale et de couleur — d'un grand nombre d'objets archéologiques selon la proximité entre les représentations de leurs périodes d'activité.

La section suivante porte sur une visualisation analogue (par le biais d'un vecteur d'évaluation) des dissimilarités entre les objets archéologiques de la base et un objet sélectionné.

5 Visualisation des dissimilarités à un objet sélectionné

L'objectif de ce processus exploratoire est de visualiser par une image couleur les objets selon leur dissimilarité à un objet sélectionné. On utilise pour cela les mêmes indices de dissimilarité que dans de Runz et al. (2008) en termes d'orientation, de localisation, de période d'activité, c'est à dire D_{date} , D_{orien} et D_{loc} .

5.1 Construction du vecteur multidimensionnel d'évaluation des dissimilarités à un objet sélectionné

Nous proposons d'utiliser une distance classique (Grzegorzewski, 1998) entre nombres et/ou intervalles flous comme mesure de dissimilarité. Soit F et G deux nombres et/ou intervalles flous, soit $F_{\alpha-}$ (resp. $G_{\alpha-}$) et $F_{\alpha+}$ (resp. $G_{\alpha+}$) les bornes inférieure et supérieure de l' α -coupe F_{α} de F (resp G_{α} de G), alors la distance entre F et G est obtenue par :

$$D(F, G) = \int_0^1 |F_{\alpha-} - G_{\alpha-}| + |F_{\alpha+} - G_{\alpha+}| d\alpha.$$

Nous utiliserons cette mesure pour le calcul de la dissimilarité d'orientations (D_{orien}) et de périodes d'activité entre éléments (D_{date}).

Pour le calcul de la dissimilarité de localisation, en raison du caractère cylindrique de la fonction d'appartenance des ensembles flous spatiaux associés aux données, nous calculons la mesure de dissimilarité D_{loc} à partir de leurs projections floues sur le plan passant par les centres des localisations (voir Figure 5).

Visualisation de données spatiotemporelles imprécises

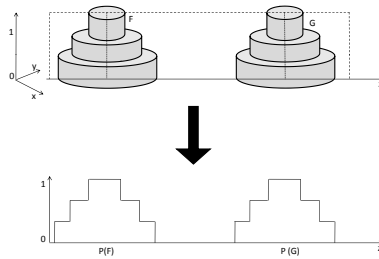


FIG. 5 – Projections pour le calcul de dissimilarité des localisations

Dans *BDFRues*, une fois l'objet A_j sélectionné, le vecteur d'évaluation $v_{A_j}(A_i)$ de chaque objet A_i à évaluer est déterminé par les dissimilarités de cet objet avec A_i . Ainsi, $v_{A_j}(A_i)$ est défini de la manière suivante :

$$v_{A_j}(A_i) = (D_{date}(A_j, A_i), D_{orien}(A_j, A_i), D_{loc}(A_j, A_i)).$$

Ce vecteur contient trois informations *a priori* décorrélées — D_{date} , D_{orien} et D_{loc} .

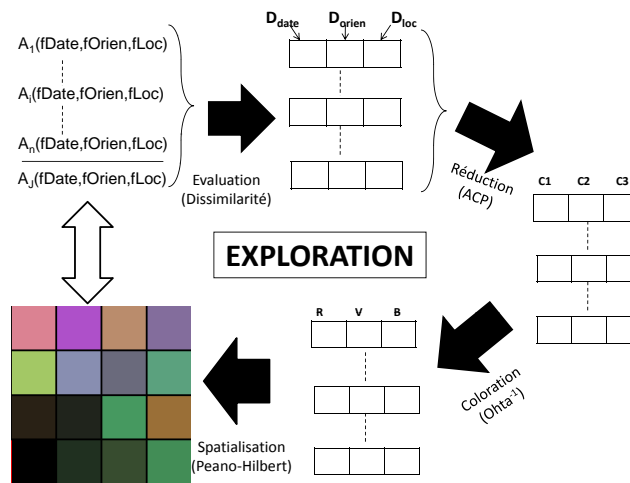


FIG. 6 – Visualisation des objets selon leurs dissimilarités à un objet sélectionné — schéma récapitulatif

Nous proposons de visualiser la dissimilarité des objets à un objet sélectionné dans la base, en donnant en entrée du processus de visualisation ces vecteurs de dimension 3. Le processus exploratoire est présenté dans la figure 6.

5.2 Visualisation des dissimilarités

Le processus de visualisation des dissimilarités des objets à un objet sélectionné donne la Figure 7.

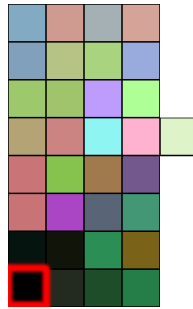


FIG. 7 – Visualisation de la dissimilarité des objets à un objet sélectionné (contour rouge) par une image couleur

Comme cette application prend à la fois en compte les dissimilarités d'orientation, de localisation et de datation, celle-ci ne fait plus apparaître les classes observées dans la visualisation strictement temporelle.

Les objets les plus similaires à l'objet sélectionné sont dans l'image les plus proches spatialement de celui sélectionné (de contour rouge dans la figure). Il y a donc *a priori* trois objets très similaires, en couleur sombre, à celui sélectionné. En effet, de par le fait que leurs dissimilarités vis-à-vis de l'objet sélectionné est faible, cela se traduit par une faible disparité et donc une faible variance. Les composantes couleurs des pixels associés aux dits objets auront donc des valeurs faibles ce qui donne une teinte proche du noir.

6 Conclusion

Nous avons présenté dans cet article une méthode originale d'exploration visuelle intuitive d'un ensemble d'objets archéologiques dont les composantes spatiales et temporelles sont représentées par des ensembles flous convexes et normalisés. Cette méthode s'est basée sur la construction de vecteurs dont les valeurs furent obtenues soit par plusieurs défuzzifications des représentations de la composante observée, soit par calcul des indices de dissimilarité des objets à un objet en entrée. L'étape de visualisation a consisté à affecter à chaque objet archéologique une couleur pour obtenir des pixels que l'on organise spatialement dans une image. Dans ce but, nous avons réduit les vecteurs d'évaluations par une ACP à des vecteurs de dimension 3. Par la transformée inverse de celle d'Ohta et al. (1980), nous avons déterminé les couleurs des pixels représentant les objets. L'image fut alors construite en utilisant une courbe de Peano-Hilbert.

Cette visualisation est strictement exploratoire. Elle permet de faire des rapprochements entre données et de les regrouper pour aider à les interpréter. C'est un outil qui présente d'autant plus d'intérêt que le nombre de données augmente (il offre la possibilité de visualiser plusieurs

millions de données). L'image résultante fournit une carte synthétique de la base archéologique étudiée en fonction de l'objectif du processus exploratoire. Cette image peut être considérée comme une légende organisée de l'information multidimensionnelle visualisée qui associe à chaque objet une couleur de manière objective.

Références

- Blanchard, F., M. Herbin, et L. Lucas (2005). A New Pixel-Oriented Visualization Technique Through Color Image. *Information Visualization* 4(4), 257–265.
- de Runz, C., F. Blanchard, E. Desjardin, et M. Herbin (2008). Fouilles archéologiques : à la recherche d'éléments représentatifs. In *Atelier Fouilles de Données Complexes - Conférence Extraction et Gestion des Connaissances - EGC'08*, Sophia Antipolis, France, pp. 95–103.
- Desjardin, E. et C. de Runz (2009). Gissar : de la saisie de fouilles à l'analyse spatiotemporelle en archéologie. In *Spatial Analysis and GEomatics*, Paris, France.
- Donoho, D. L. (2000). High-Dimensional Data Analysis : The Curses and Blessings of Dimensionality. In *AMS Conference Mathematical Challenges of the 21st Century*.
- Grzegorzewski, P. (1998). Metrics and orders in space of fuzzy numbers. *Fuzzy Sets and Systems* 97, 83–94.
- Guptill, S. C. (2005). Metadata and data catalogues. In P. A. Longley, M. F. Goodchild, D. J. Maguire, et D. W. Rhind (Eds.), *Geographical Information Systems. Principles, Techniques, Management and Applications*, Volume 2, Chapter 49, pp. 677–692. Wiley. Seconde Edition.
- Hyvärinen, A. (1999). Survey on independent component analysis. *Neural Computing Surveys* 2, 94–128.
- Keim, D. A. (2000). Designing Pixel-oriented Visualization Techniques : Theory and Applications. *IEEE Transaction on Visualization and Computer Graphics (TVCG)* 6(1), 59–78.
- Moon, B., H. V. Jagadish, C. Faloutsos, et J. H. Saltz (2001). Analysis of the Clustering Properties of the Hilbert Space-Filling Curve. *IEEE Transactions on Knowledge and Data Engineering* 13(1), 124–141.
- Ohta, Y., T. Kanade, et T. Sakai (1980). Color Information for Region Segmentation. *Computer Graphics and Image Processing* 13, 222–241.
- VanLeekwijck, W. et E. E. Kerre (1999). Defuzzification : criteria and classification. *Fuzzy Sets and Systems* 108, 159–178.
- Wang, X. et E. E. Kerre (2001). Reasonable properties for the ordering of fuzzy quantities (I). *Fuzzy Sets and Systems* 118, 375–385.

Summary

In this article, we use a specific technique for the visualization of an archaeological dataset in which object components are modeled using normalized convex fuzzy sets. In order to build a color image of data, we use definition of multidimensional vector for each object. The color image gives us a resume of information allowing users to observe and analyze it graphically.