Rémy RAMADOUR

Automatisation de procédures de Géoréférencement,

recalage d'images aériennes haute résolution sur BD Ortho de l'IGN

21 juin 2010

sous la direction de

Dorothée James,
Luc Vacher
Cellule géomatique de
l'UMR 6250 LIttoral, ENvironnement
et Sociétés CNRS-Université de La Rochelle

Christophe Saint-Jean,
Laurent Mascarilla
Laboratoire Mathématiques, Image et
Applications (EA 3165 MIA) de
l'Université de La Rochelle

Rapport de Stage de première année
Master Ingénierie Mathématique et Image
Université de La Rochelle

Année Universitaire 2010-2011

Abstract

En partenariat avec le LIENSs, nous avons tenté d'automatiser les procédures de géo-référencement d'images. Le problème se décompose en deux parties : trouver des points d'intérêts dans chaque image, générer des descripteurs locaux pour ensuite mettre en correspondance ces points et calculer les coefficients de la matrice de transformation qui permet de passer de l'image source à l'image référentielle.

Contents

1 Situation **3**

2 Modélisation **6**
2.1 Les matrices de transformation 6
2.2 Les points d'intérêts et les descripteurs locaux 9

3 Résolution **13**
3.1 Contexte 13
3.2 Les impasses de l'approche par image 16
3.3 Les avantages d'une approche par groupe d'images 18

4 Perspectives **25**
4.1 Le développement logiciel 25
4.2 Le signal monogène 28
4.3 Les espaces galactiques 31

1 Situation

L'Institut du Littoral et de l'Environnement héberge le laboratoire du LIENSs (Littoral Environnement et Sociétés). Parmi l'ensemble de ses missions, le LIENSs propose un ensemble de statistiques et d'informations géographiques générées à partir du traitement d'images du littoral de Charente-Maritime.

Ce travail - qui appartient au domaine de la géomatique - comporte trois phases distinctes :

- collecte de données
- extraction de l'information
- diffusion des résultats

Le contexte particulier au sein duquel a été élaborée la coopération entre le LIENSs et le MIA (Mathématiques, Images et Applications) :

- un photographe professionnel fournit des prises de vue aériennes de l'ensemble du littoral du département. Les différents clichés comportent des zones de superposition.
- l'équipe du LIENSs calcule par exemple la surface de plage occupable des plages du département, élabore des statistiques sur l'occupation de celles-ci.
- les informations traitées sont géoréférencées afin de pouvoir être mises en correspondance avec l'ensemble des informations géographiques disponibles.

Le MIA s'est engagé auprès du LIENSs a explorer des voies d'automatisation de la seconde et de la troisième partie. En ce qui concerne la phase d'extraction et de traitement de l'information, il s'agira entre autres de segmenter des régions de plage à partir par exemple de l'information couleur, voire couleur/chaleur, ou encore de mettre en place des outils de RdF afin d'automatiser le comptage des touristes sur les plages. Ces informations permettront par exemple aux communes d'ajuster leur politique de touristique et de surveillance de plages ; elles sont donc de première importance et leur automatisation permettra d'affiner la qualité du travail généré.

La troisième partie - le géoréférencement - est celle au sein de laquelle s'est inscrit mon stage. L'enjeu consiste à mettre en correspondance deux groupes d'images : les premières sont des images déjà référencées fournies par les institutions reconnues, les secondes sont les images fournies par le photographe auxquelles sont éventuellement ajoutées les informations générées par l'équipe du LIENSs.

Le géoréférencement consiste à donner une existence spatiale - géographique - à un ensemble d'informations. Il est dépendant d'un système de coordonnées et du système de projection utilisé. Par exemple, on peut utiliser la projection conique conforme de Lambert ou sa déclinaison utilisée pour les cartes de la métropole française, le Lambert93. Les valeurs calculées pour un type de projection sont ensuite exportables vers un autre système. Il est donc important de connaître le système de projection dans lequel on travaille, son point d'origine, l'échelle de mesure entre deux points, ...

Afin de géoréférencer une image, on doit fournir un minimum de trois points [1] que l'on estime être les mêmes dans chaque image - l'image de référence et l'image à référencer. Une fois ces points indiqués, on peut en déduire leurs coordonnées géographiques, ainsi qu'une transformation de l'image de premier ordre (rotation et mise à l'échelle) qui permettra d'avoir une approximation des coordonnées des autres points de l'image à référencer.

En pratique, le géoréférencement d'une image consiste en la détermination de six valeurs : les quatre premières nous renseignent sur l'échelle et l'orientation de l'image, les deux dernières sont les coordonnées d'un point particulier de l'image (par convention, il s'agit du point haut-gauche de l'image). En partant du point haut-gauche, et par approximation linéaire fournie par les quatre première valeurs, on peut avoir une approximation pertinente des coordonnées de tout point de l'image. On suppose cependant que l'image a correctement été projetée selon les normes du système choisie, que l'on possède donc une image rectifiée, et que le géoréférencement ne concerne que cette image projetée.

Afin d'automatiser le géoréférencement, il faut donc être capable de comparer deux images de natures (taille, colorimétrie, définition, ...) différentes et d'en extraire des points d'intérêts corrélés qui serviront de support à l'interpolation (de premier, second ou troisième ordre) des coordonnées de l'image dans son ensemble. Les informations produites par le LIENSs sont

[1] en pratique, on se fixe un minimum de 4 points par souci de robustesse

alors indexées à la base partagée par l'ensemble de la communauté géographe, et l'étape de diffusion est validée.

Mon stage s'est déroulé au sein des deux équipes, sous la double direction de Dorothée James (membre du LIENSs) et de Christophe Saint-Jean (membre du MIA), en collaboration avec Oumar Simagan (encadré par Dorothée James également, et par Laurent Mascarilla, membre du MIA). La dernière semaine, l'essentiel de mon travail s'est effectué sous la tutelle de Bernard Besserer. Tout au long du stage, j'ai été amené à solliciter l'aide et les conseils de Michel Berthier, Carl Frélicot, Renaud Peteri, José Mennesson et Guillaume Demarcq, ainsi que de Anne-Laure Bethe et Ulrich André, auxquels j'adresse tous mes remerciements, tant pour leur disponibilité que pour le plaisir de discuter avec eux.

2 Modélisation

Soient $\{R_i\}_{i\in I}$ une famille d'images dont le référentiel est connu - et dont l'indexation à ce référentiel est parfaitement connue également - et $\{S_j\}_{j\in J}$ une famille d'images sources à référencer. On suppose qu'il existe une relation T entre ces deux ensembles que l'on peut noter ainsi :

$$\forall j \in J, \quad \exists \tilde{I} \subset I, \quad T(S_j) \subset \underset{i\in\tilde{I}}{\cup} R_i$$

Cette transformation T est celle que nous essaierons de déterminer à partir des points d'intérêts corrélés que nous pourrons exhiber. Plus précisément, l'objectif est d'arriver à obtenir une approximation de son déve-loppement au premier, second, voire troisième ordre.

Dans le cadre de ce compte-rendu (et dans la limite de mes connaissances), je ne parlerai ici que de ce qui concerne le premier ordre qui - à l'essai - fournit une approximation pertinente, à condition que les points corrélés soient suffisamment bien répartis dans l'image S_i.

2.1 Les matrices de transformation

La suite de cette section s'inspire largement de *Multiple View Geometry*, Hartley and Zisserman, 2000.

Soit $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ un vecteur de coordonnées d'une surface plane. L'essentiel des transformations qui nous intéressent dans le cadre des applications du stage peut être codé dans une matrice 3×3 dont les facteurs dépendront de la transformation à appliquer.

La règle de transformation sera : $\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = T \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$, où $\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}$ est un vecteur à projeter ensuite sur le plan $Z = 1$.

Rotations :

$$T_\theta = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Dilatations et contractions isotropes :

$$T_\varepsilon = \begin{bmatrix} \varepsilon & \varepsilon & 0 \\ \varepsilon & \varepsilon & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Isométries :

$$T_{\theta,\varepsilon} = \begin{bmatrix} \varepsilon\cos\theta & -\varepsilon\sin\theta & 0 \\ \varepsilon\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

où $\varepsilon = \pm 1$

Translations :

$$T_{t_x,t_y} = \begin{bmatrix} 1 & 1 & t_x \\ 1 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

Les transformations du premier ordre d'Arcgis :

ArcGis se limitant à des opérations de contraction, rotation et translation pour le premier ordre, la matrice totale sera donnée par :

$$T = T_{t_x,t_y} \cdot T_\theta \cdot T_\varepsilon = \begin{bmatrix} \varepsilon\cos\theta & -\varepsilon\sin\theta & t_x \\ \varepsilon\sin\theta & \varepsilon\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

On remarque que la matrice est en fait décomposée en plusieurs blocs. Elle peut être représentée ainsi :

$$\begin{bmatrix} A & T \\ V^T & 1 \end{bmatrix}$$

où $A \in \mathcal{M}_2(\mathbb{R})$ représente les rotations et les contractions/dilatations, T est un vecteur 2×1 de translation, et V est un vecteur 2×1 pour l'instant nul.

Il est bien-sûr possible de considérer comme indépendants chaque facteur de la sous-matrice A, mais on rentre alors dans le cadre de transformations anistropes (transformation d'un carré en losange par exemple), et beaucoup plus de points sont nécessaires pour évaluer les coefficients.

La dernière transformation qui nous intéresse - non fournie par Arcgis pour le premier ordre, mais qu'AutopanoGiga propose par défaut - est codée

dans le vecteur V est représente une projection affine. Elle permet d'obtenir une coordonnée en Z nécessaire pour projeter ensuite le vecteur obtenu sur le plan $Z = 1$. En posant $V = \begin{bmatrix} v_x \\ v_y \end{bmatrix}$, on obtient $Z = 1 + v_x x + v_y y$, soit $Z = 1 + < (x, y) | (v_x, v_y) >$.

On définit alors une matrice de transformation affine de la manière suivante :

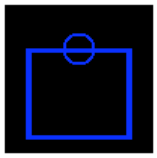
$$T_V = \begin{bmatrix} Id_2 & 0 \\ V^T & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ v_x & v_y & 1 \end{bmatrix}$$

On pourra enfin projeter les coordonnées obtenues en posant par exemple :

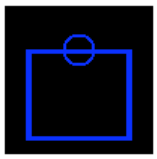
$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \frac{1}{1 + v_x x + v_y y} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Exemples :

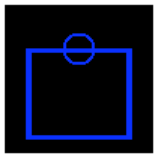
$T = \begin{bmatrix} .7\cos\theta & -.7\sin\theta & 0 \\ .7\sin\theta & .7\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$, avec $\theta = \frac{\pi}{6}$:



$T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ .009 & .003 & 1 \end{bmatrix}$:



$$T = \begin{bmatrix} .7\cos\theta & -.7\sin\theta & 0 \\ .7\sin\theta & .7\cos\theta & 0 \\ .009 & .003 & 1 \end{bmatrix} :$$



2.2 Les points d'intérêts et les descripteurs locaux

Afin d'étudier une fonction simple, continue, différentiable, on nous apprend assez rapidement à en dresser un tableau de variations sur lequel sont représentés les points pour lesquels elle s'annule, ainsi que les points qui annulent sa dérivée première, voire ceux qui annulent sa dérivée seconde. L'information que l'on peut retirer de ce tableau est suffisante pour en esquisser la forme. Ces points sur lesquels s'annule la fonction, son gradient et eventuellement sa Jacobienne, peuvent être considérés comme des points d'intérêt de la fonction. Si deux fonctions possèdent les mêmes points d'intérêts, alors elles ne devraient pas avoir - sur un compact donné - des comportements radicalement différents. En comparant les valeurs prises par les fonctions en ces points, éventuellement en ajoutant les valeurs de leurs dérivées, on obtient un vecteur descripteur pour chaque fonction qui pourra nous permettre de les comparer entre elles, à partir d'une mesure de similarité reposant sur le choix d'une distance entre les vecteurs associés.

De la même manière, on cherchera pour comparer deux images un échantillon de points auxquels seront associés leurs vecteurs descripteurs respectifs. Soit A un point de $R\,(\subset \bigcup R_i)$ et B un point de S_j pour un j fixé. Soit maintenant V_A un vecteur descripteur de A et V_B le vecteur descripteur associé

à B. Soit enfin δ une distance entre descripteurs. On supposera qu'il existe une valeur ϵ telle que :

$$\delta\left(V_A, V_B\right) < \epsilon \Longrightarrow A \in S_j \text{ et } B \in R$$

Sous cette hypothèse, on va sélectionner un échantillon de points satisfaisant à cette relation et les utiliser pour calculer les coefficients de la transformation qui nous permettra de projeter chaque S_j sur $R = \bigcup_{i \in I} R_i$.

Une méthode bien connue de sélection de poins d'intérêts est la méthode de Harris. Elle consiste à prendre en compte la géométrie de l'image et extraire les points qui forment des coins ou des points-cols (les deux étant regroupés sous le terme anglo-saxon de *cornerness*. Suivant l'espace colorimétrique dans lequel on se situe, la notion de coin prendra une signification différente, mais elle reste abstraitement la même et garde une pertinence quel que soit l'espace considéré.

Dans sa forme la plus répandue, un détecteur de points d'intérêts de Harris utilisera la première forme fondamentale de la géométrie différentielle. On commencera par fabriquer un tenseur métrique dont la valeur en chaque point sera donnée. Pour une image située dans un espace N-dimensionnel, on utilisera par exemple le tenseur suivant :

$$\begin{bmatrix} \sum_{i=1}^{N} I_x^2 \cdot e_i & \sum_{i=1}^{N} I_x I_y \cdot e_i \\ \sum_{i=1}^{N} I_x I_y \cdot e_i & \sum_{i=1}^{N} I_y^2 \cdot e_i \end{bmatrix}$$

où les e_i représentent la base de l'espace colorimétrique, et I_x & I_y représentent les dérivées spatiales de l'image.

Une fois le tenseur construit, sa valeur en chaque point de l'image calculée, on peut utiliser les propriétés d'invariance des matrices de dimension 2×2 pour calculer la somme des valeurs propres et leur produit, valeurs dont on sait qu'elles correspondent respectivement à la trace et au déterminant de la matrice, et qu'elles s'interprètent en terme de courbure intrinsèque et extrinsèque de la surface au point considéré.

Un ratio intelligent de ces valeurs nous permettra alors d'avoir un critère de forte orthogonalité de la surface en ces points, et les points fortement orthogonaux seront considérés comme des points d'intérêt de l'image.

La méthode de détection de points d'intérêt utilisée par l'algorithme SIFT reprend l'idée de sélection des points selon qu'ils présentent ou non l'analogue d'un coin, mais ajoute un paramètre d'échelle. Un point sera alors inclus dans l'échantillon final si sa caractérisation en tant que coin reste valide pour plusieurs échelles. Les images ne sont plus alors considérées comme prenant leurs valeurs dans $\Omega \in \mathbb{R}^2$, mais dans $\Omega \times \Theta \in \mathbb{R}^3$, où Ω reste le champ de coordonnées spatiales, et Θ représente les coordonnées d'échelle. Un point d'intérêt persistant en échelle sera alors aisément détectable et référençable.

Se donner un espace d'échelle consiste dans le choix d'une fonction qui - par convolution - lissera la surface, modélisant ainsi la disparition de détails et propriétés corrélativement à la distance simulée. Soit Ψ une telle fonction. En associant le paramètre d'échelle θ à la matrice de dispersion de Ψ (ou plus simplement à son écart-type), on construira la projection de l'image dans ce nouvel espace de la manière suivante :

$$I(x, y; \theta) = I(x, y) * \Psi_\theta$$

la fonction Ψ étant le plus souvent une gaussienne - c'est d'ailleurs le cas pour SIFT.

Les descripteurs de l'algorithme SIFT reposent sur la transformée de Hough qui est un outil mathématique utilisé pour repérer des formes géométriques particulières dans une surface. La distance utilisée sur les descripteurs sera généralement la distance euclidienne.

Finalement, l'utilisation d'un algorithme compétitif permettra à moindre coût de sélectionner les points ayant une probabilité forte d'être les plus pertinents, en fonction du modèle souhaité.

Le modèle sera généralement caractérisé par les propriétés géométriques au regard desquelles on le souhaite invariant. Les plus répandues - et celles qui par ailleurs nous intéressent ici, tant que l'on ne considère qu'une projection de premier ordre - sont l'invariance par translation, l'invariance par rotation et l'invariance d'échelle.

Cependant, si la méthode de détection SIFT - et plus récemment ASIFT - paraît être la méthode actuellement privilégiée par la communauté scientifique, nous avons pu durant le stage en observer les limites. En séparant la question de la détection des points de celle de la création de descrip-

teurs locaux, il devrait être possible de construire des méthodes adaptées à la problématique particulière qui est la nôtre. Il en sera question dans la section suivante, il est parfois nécessaire de forcer la détection de points d'intérêts en jouant sur les paramètres, ce qui demande une supervision de laquelle on souhaite justement s'affranchir - ou au moins que l'on cherche à diminuer radicalement.

3 Résolution

La durée du stage étant de sept semaines, il a été décidé que les deux premières semaines seraient consacrées exclusivement à la littérature, tandis que les cinq suivantes seraient utilisées pour les différents tests et l'écriture d'un programme utilisable en l'état par le LIENSs.

3.1 Contexte

Précisons tout d'abord le type d'images avec laquelle nous avons travaillé, ainsi que la manière dont se déroulait jusqu'ici le géoréférencement.

Les images sources - les images S_j à géoréférencer - sont issues d'une campagne aérienne effectuée en 2009 par un photographe professionnel, Michel Bernard, de ECAV-Aviation. L'ensemble des prises de vue couvre les plages de l'île de Ré, les plages de l'île d'Oléron, ainsi que les plage urbaines de la CDA rochelaise. Entre deux images successives, une zone de recouvrement minimale a été respectée permettant l'assemblage panoramique de ces images entre-elles. Elles ont été fournies dans un format compressé JPEG et ont une résolution d'environ 1 pixel pour 7 cm réels.

Les images de références - la famille R_i préalablement géoréférencée - sont issues de l'IGN (Institut Géographique National). Elles possèdent une résolution d'un pixel pour 50 cm. Elles couvrent l'ensemble du territoire départemental. Usuellement, on possède une image pour tout le département, à partir de laquelle on extrait via un logiciel des sous-images. Elles sont issues d'une campagne datant de 2006.

Actuellement, les chercheurs et employés du LIENSs sélectionnent à la main 4 points (ou plus) dans chaque image, en utilisant les outils de la suite logicielle ArcGis. Les coordonnées de ces points s'affichent dans un éditeur qui permet éventuellement de les corriger, ou de les supprimer.

Mais il est possible de charger un simple fichier texte contenant les coordonnées de points déjà calculées et de géoréférencer à partir de ceux-ci. C'est ici que réside l'objectif pratique du stage : fournir le fichier texte contenant les coordonnées des points mis en corrrespondance.

Dans un premier temps, nous avons travaillé sur des images sources isolées, considérées une à une.

Exemples d'images sources :

Aytre_2009_014.png

Boucholeurs_2009_001.png

Re_2009_0585.png, Re_2009_0586.png, Re_2009_0587.png

Exemple d'image de référence :

Zoom sur la zone sud-est de l'Île de Ré

Zoom sur des zones géographiques semblables :

Image de référence, 196 × 220 px

Image source, 1616 × 1136 px

3.2 Les impasses de l'approche par image

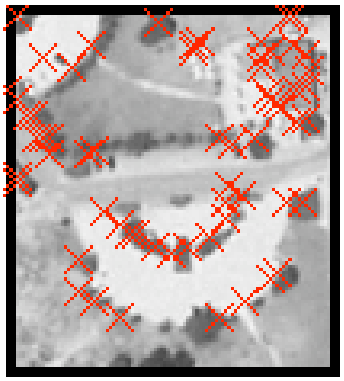
On remarque d'ores et déjà deux problèmes majeurs à l'approche *en l'état* par détection de points d'intérêts et descripteurs locaux :

- d'une part, la différence de résolution, de colorimétrie, ... pourrait fausser la mise en correspondance des valeurs de descripteurs locaux. On observera dans les deux images ci-dessus la différence de détails au niveau du toit de maison.

- d'autre part, les prises de vues sont faites à des époques différentes, des saisons différentes, des heures différentes, ... Dans les deux images présentes ci-dessus, la présence de voitures dans l'une et pas dans l'autre perturbe énormément la corrélation des vecteurs de descripteurs. Dans le cas de l'image de référence, les points d'intérêts sont majoritairement dans la végétation, tandis qu'ils se concentrent pour l'image source sur les voitures. De manière générale, tout artefact artificiel (voiture, route, maison, ...) aura tendance à prendre l'ascendant sur un environnement naturel et prendra à coup sûr les valeurs les plus importantes.

Différentes méthodes ont été testées, de SIFT à ASIFT, de Harris à Harris-Laplace, en passant par les différences de gaussiennes, des méthodes plus ou moins multi-échelles (au sens de l'espace d'échelle défini plus haut), jusqu'à l'implémentation d'une déclinaison de Harris faisant correspondre à un point de l'image source l'ensemble de son 7×7 voisinage (correspondant à la différence de résolution). Sans succès véritable.

détection par l'algorithme de Harris sur l'image source haute résolution



détection par l'algorithme de Harris sur l'image de référence basse résolution

Différentes pistes ont été envisagées alors : incursion dans la théorie du signal monogène couleur, utilisation de la logique floue pour une détection plus robuste des coins, ...

3.3 Les avantages d'une approche par groupe d'images

La solution du problème a émergée d'une considération plus simple : statistiquement parlant, il est plus probable de trouver des points d'intérêts comparables sur des zones plus larges. A l'essai, c'est effectivement ce qui s'est produit. En assemblant six images issues d'une même série de prises de vue, le matching des points d'intérêts s'est effectué sur des régions moins sensibles aux modifications (essentiellement des zones habitées)

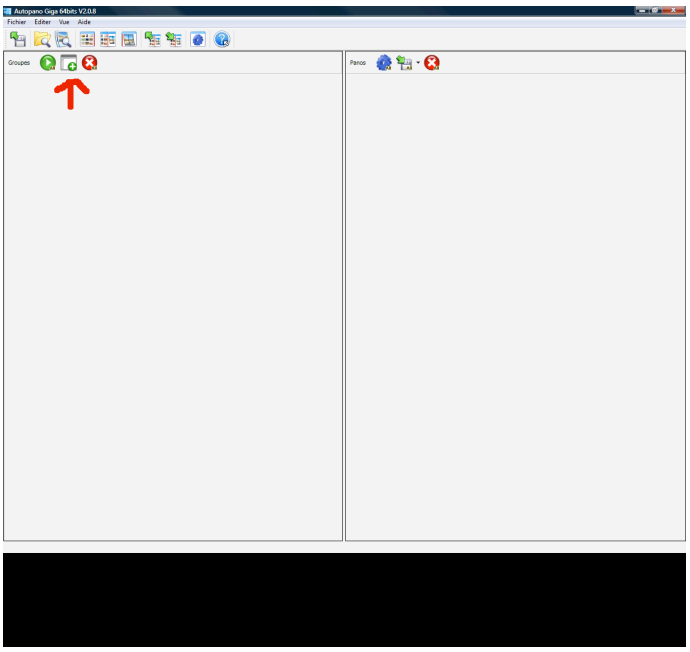
Points d'intérêts SIFT (amélioré) générés par le logiciel AutoPanoGiga :

Il y a un triple avantage à utiliser cette méthode :

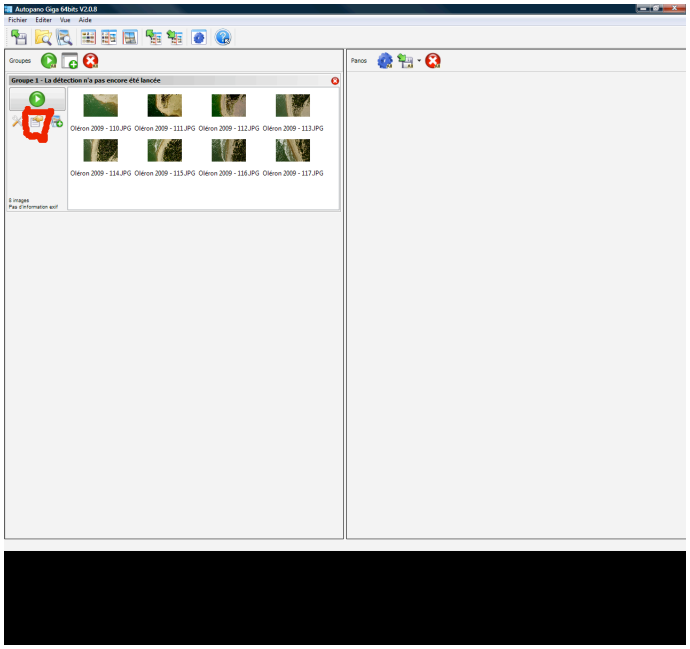
- la mise à l'échelle de l'image source nécessaire à la comparaison à l'image de référence - de moindre résolution - implique une détérioration de l'image qui atténuera éventuellement les objets artificiels contingents (voitures, bâteaux, ...).
- les images sont référencées par paquets ; gain de temps sans perte de qualité (au vu des résultats obtenus à ce jour).
- en couvrant une zone plus large, on maximise le potentiel d'avoir des points d'intérêts fortement corrélés. Il est parfois impossible d'en trouver plus de un ou deux sur une image isolée. En prenant des groupes de six images, je n'ai pas eu une seule fois moins de 30 points distribués sur l'image composite.

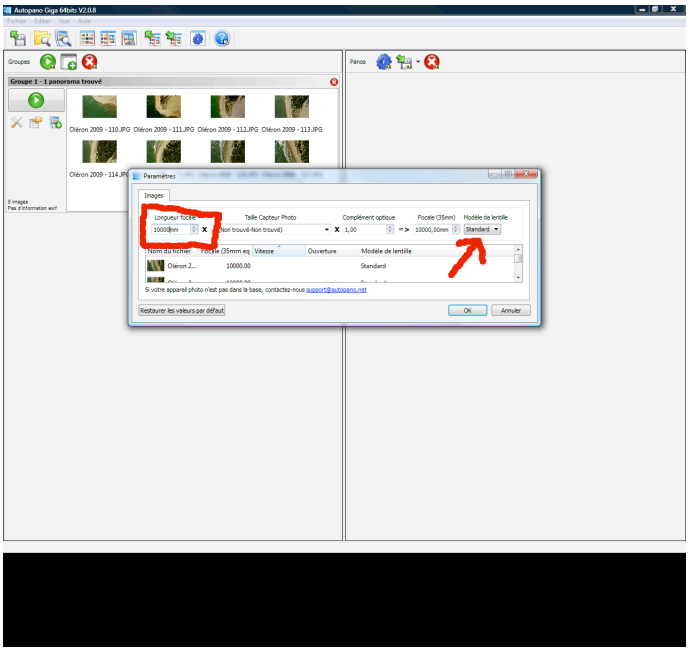
Dans l'état actuel, les étapes sont les suivantes :

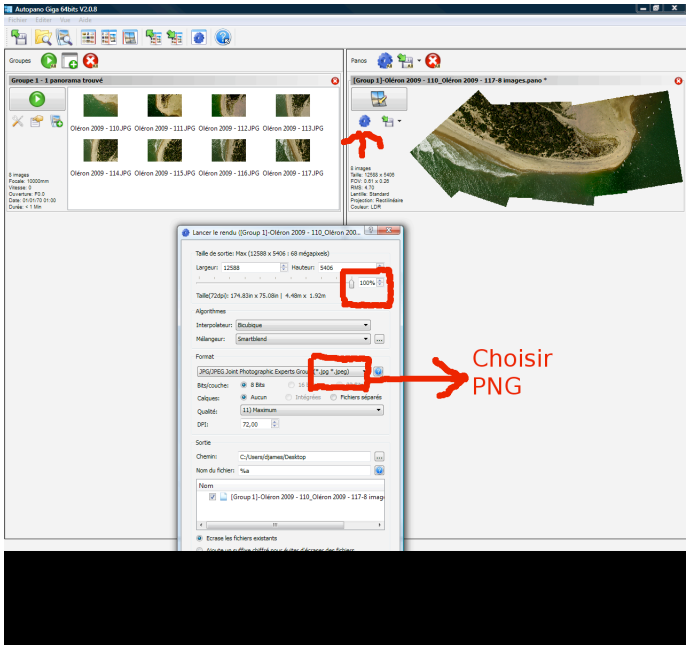
- on utilise le logiciel AutopanoGiga pour générer une image composite. Celle-ci est exportée au format .png [2] en deux versions : une image pleine résolution, une image à la résolution comparable à l'image de référence. Afin de déformer le moins possible les images sources, on fixera arbitrairement une focale importante (> 1000).

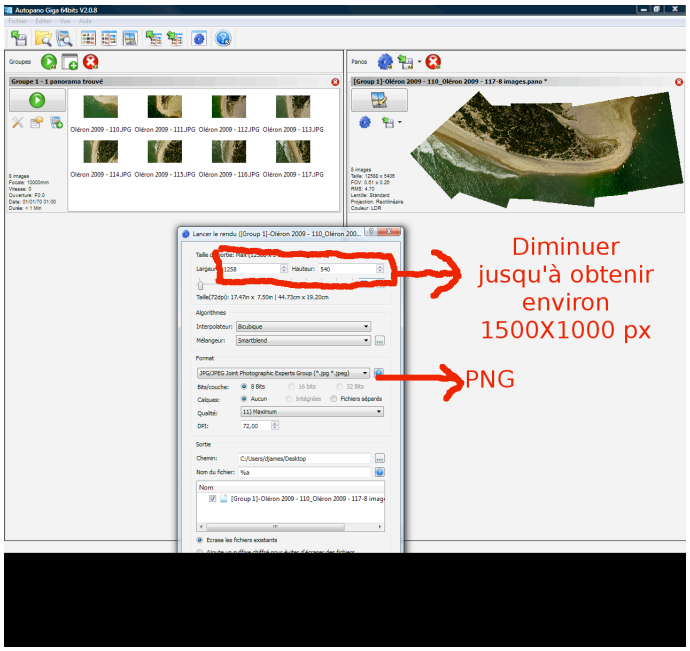


[2]prise en charge plus pratique sous ArcGis : bonne gestion de la transparence, un pixel représente dans un premier temps une unité de mesure, ce qui facilite les conversions à venir, coin haut gauche fixé à $(0, 0)$

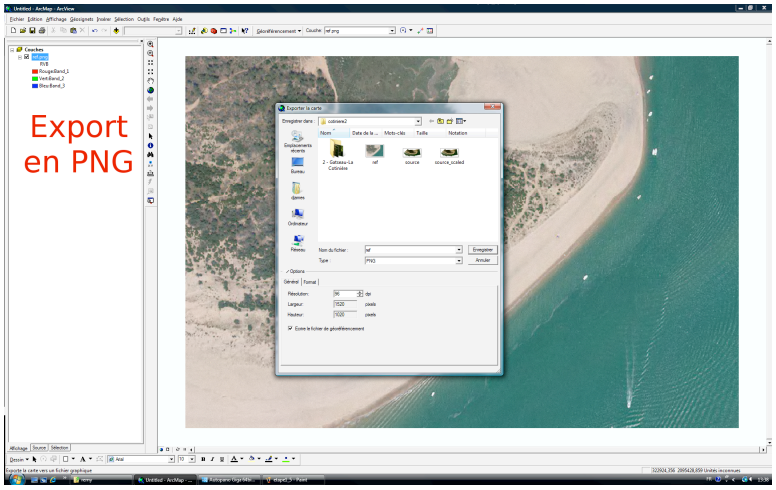

Autopano Giga 64bits V2.0.8
Fichier Editer Vue Aide
Groupes
Groupe 1 - La détection n'a pas encore été lancée
Oléron 2009 - 110.JPG Oléron 2009 - 111.JPG Oléron 2009 - 112.JPG Oléron 2009 - 113.JPG
Oléron 2009 - 114.JPG Oléron 2009 - 115.JPG Oléron 2009 - 116.JPG Oléron 2009 - 117.JPG
6 images
Pas d'information exif
Panos


Autopano Giga 64bits V2.0.8
Groupes
Groupe 1 - 1 panorama trouvé
Oléron 2009 - 110.JPG Oléron 2009 - 111.JPG Oléron 2009 - 112.JPG Oléron 2009 - 113.JPG
Oléron 2009 - 114.JPG
6 images
Pas d'information exif
Panos
Paramètres
Images
Longueur focale
Taille Capteur Photo
Complément optique
Focale (35mm)
Modèle de lentille
10000mm
(Non trouvé-Non trouvé)
1,00
=> 10000,00mm
Standard
Nom du fichier
Focale (35mm eq
Vitesse
Ouverture
Modèle de lentille
Oléron 2...
10000.00
Standard
Si votre appareil photo n'est pas dans la base, contactez-nous support@autopano.net
Restaurer les valeurs par défaut
OK
Annuler

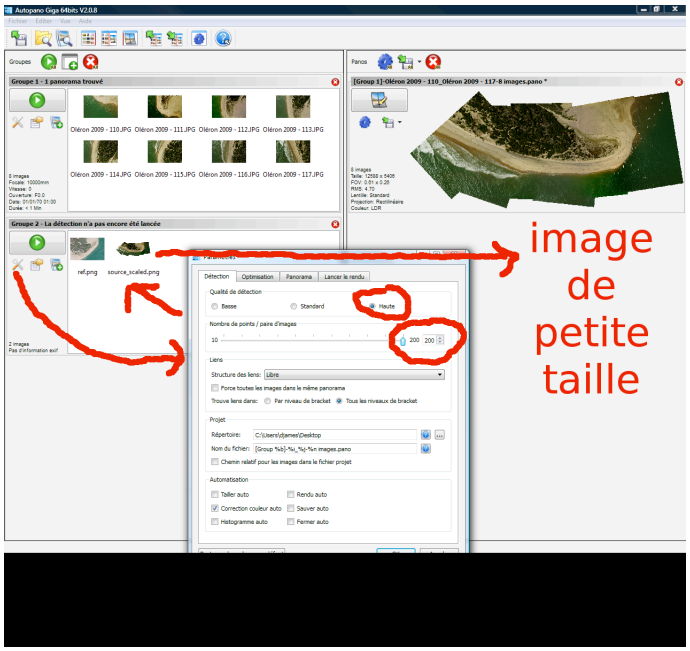

Choisir
PNG


Diminuer
jusqu'à obtenir
environ
1500X1000 px
PNG

- par le biais d'ArcGis, on exporte une carte englobant la région que l'on souhaite référencer. La taille de l'image produite est la même par défaut (1600×1050), ce qui permet de fixer l'échelle de la seconde version de l'image source.

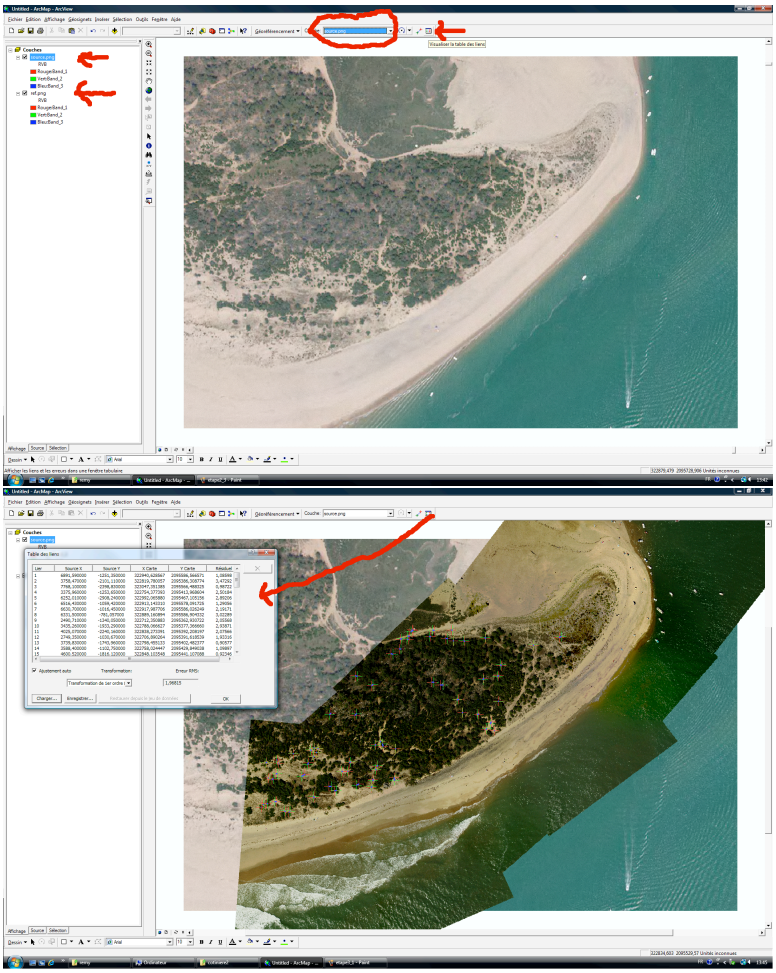


- on réinjecte les images dans AutopanoGiga ; le logiciel calcule les points d'intérêts des deux images et les corrèle. Un fichier .pano (format AutopanoGiga) ou .pto (format Hugin) est produit.



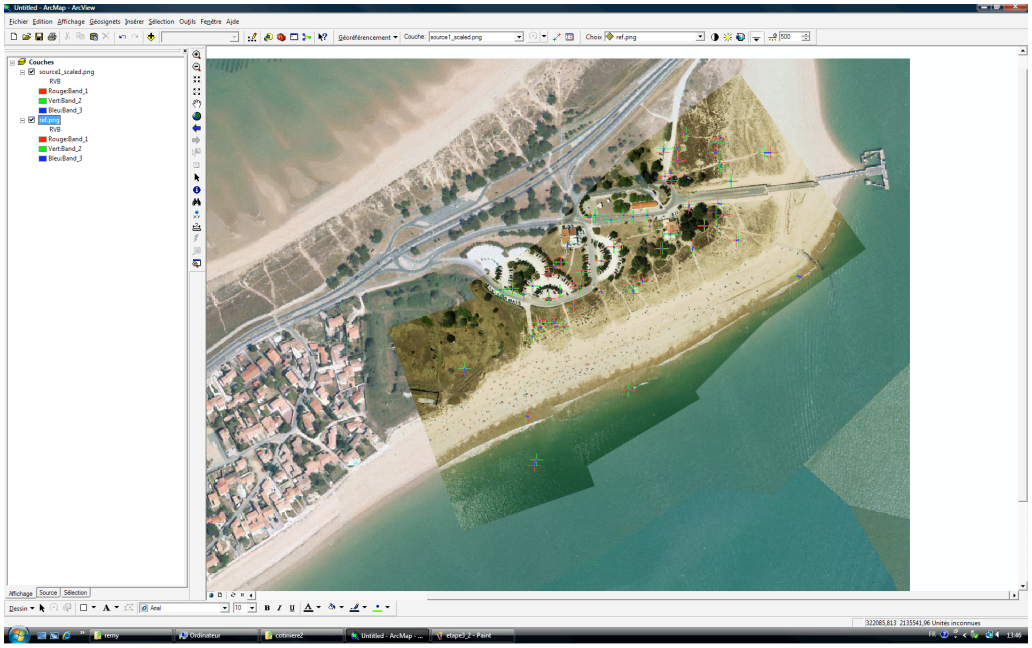
- trois fonctions écrites durant le stage prennent en paramètre le fichier .pano (ou .pto) ainsi que le fichier contenant les paramètres de géoréférencement de l'image IGN et génère un fichier texte comprenant les coordonnées actualisées pour chaque image de ses points d'intérêts corrélés.

- ce fichier est chargé dans ArcGis et le géoréférencement s'effectue.



La dernière image montre les résultats sur une zone sensible (marais, végétation mouvante, glissement des zones de sable, marées, ...).

L'image suivante montre le résultat sur une zone qui a posé problème lors de l'approche image par image :



On peut voir un décalage sur la droite au niveau de la construction portuaire, ce qui est dû à une mauvaise répartition des points d'intérêt sur l'image. Ce décalage sera corrigé en sélectionnant manuellement un point d'intérêt dans la zone concernée. Certes, cela va à l'encontre de l'automatisation de la procédure, mais il suffit désormais de sélectionner de temps en temps un point pour 6 images au lieu de 4 points par image. De plus, plusieurs pistes sont explorées pour évacuer ce désagrément.

4 Perspectives

4.1 Le développement logiciel

Maintenant que l'ordre des étapes et les différents paramètres sont connus ou aisément accessibles, il est possible de proposer une interface qui automatise la procédure. Plusieurs voies sont actuellement envisagées.

La première reprend les étapes précédentes. Elle est conditionnée à la possibilité d'appeler AutopanoGiga dans un script, à pouvoir lui passer au préalable un certain nombre de paramètres.

L'utilisateur indique les groupes d'images qu'il souhaite traiter. Le script les charge dans AutopanoGiga, et créé l'image composite et son clone de résolution inférieure. Les dimensions de ce dernier sont connues : elles doivent correspondrent au format par défaut des images exportées par Arcgis, soit environ $[1300, 1600] \times [900, 1100]$.

L'utilisateur exporte ensuite via ArcGis la zone correspondante à celle que l'on souhaite projeter et en fournit le chemin.

Le programme effectue une deuxième incursion dans AutopanoGiga afin de calculer les points de corrélations entre les deux images.

A l'aide du fichier de géoréférencement de l'image fournie par ArcGis, et sachant que le point haut-gauche d'une image PNG sera par défaut envoyé sur le point $(0, 0)$ du référentiel géographique, on applique l'algorithme suivant :

load fichier_georeferencement.pgw ;
%% récupération des valeurs de rotation et d'échelle
A = readline ;
B = readline ;
C = readline ;
D = readline ;
%% récupération des coordonnées du point $(0, 0)$
X = readline ;
Y = readline ;

load fichier.pano
create fichier.txt

WHILE points
$[x1, y1, x2, y2]$ = lire_ligne (fichier.pano, $\%.6f, \%.6f, \%.6f, \%.6f$) ;

$x1 = x1$;
$y1 = -y1$;
$X = A * x2 + B * y2 + X$;
$Y = -(C * x2 + D * y2) + Y$;

ecrire_ligne(fichier.txt, x1, y1, X, Y) ;

END WHILE

Le fichier texte généré peut alors être appelé dans ArcGis et le géoréférencement s'effectuer.

Toutefois, il est à noter que ArcGis n'offre que des transformations de rotation et d'échelle si on utilise le premier ordre, et les résultats avec le second et le troisième ordre n'ont pas paru plus pertinents.

Sachant qu'AutopanoGiga propose un plus grand panel de transformations (trapèze, coussinet et barillet, ...), il est à mon sens envisageable de l'utiliser afin d'automatiser le géoréférencement même, sans passer pour cela par ArcGis. Une interface de contrôle est à prévoir, afin que l'utilisateur puisse valider tant que l'on n'est pas certain de l'exactitude de nos méthodes, mais il paraît tout-à-fait envisageable d'effectuer le géoréférencement en dehors d'ArcGis, avec des images projetées avec plus d'exactitude. On éliminerait ainsi certainement des problèmes comme celui du second résultat proposé (décalage sur les régions en périphérie de l'image composite).

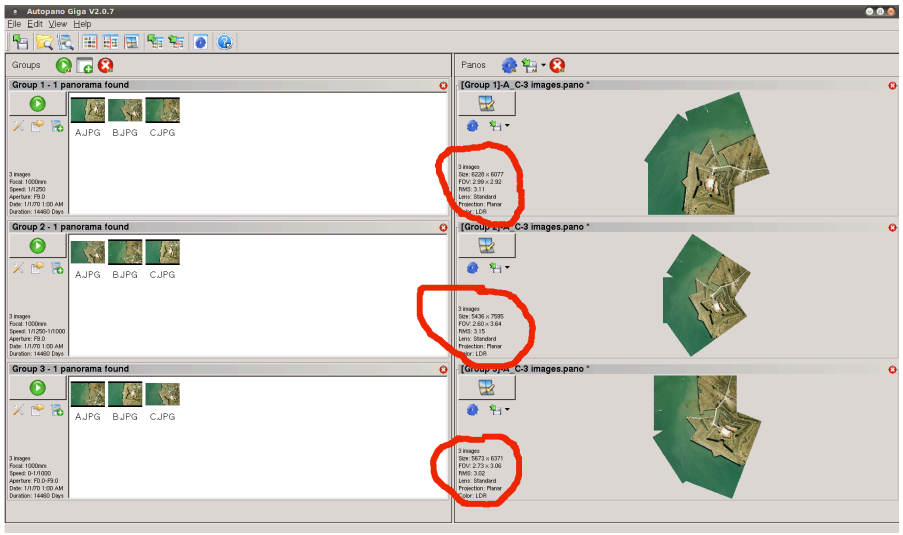
Afin de rendre l'opération plus performante, il est également souhaitable - pour les prochaines campagnes de prise de vue - d'équiper l'appareil photo d'un GPS intégré, permettant de connaître au préalable la latitude, la longitude et l'altitude de la prise de vue. En récupérant par ailleurs les images au format brut (RAW, sans pré-traitement ni compression), on pourra notablement améliorer la résolution de l'image, ce qui pourra se révéler utile pour l'analyse géomatique de celle-ci (segmentation des zones de plages utiles, comptage de population, ...).

Des tests enfin sont encore à envisager conernant l'assemblement des images composites. On peut dans un premier temps affiner les paramètres que l'on utilise (meilleure focale, meilleure réduction d'échelle), mais aussi

générer autant de mosaïques qu'il existe de combinaisons possibles d'images, et tester laquelle est la plus semblable à l'image référentielle.

Supposons par exemple que l'on souhaite traiter trois images en même temps. Notons les A, B et C. Les images composites ABC, BCA et CAB ne sont pas exactement les mêmes. C'est en effet la première qui influence le plus la géométrie globale de l'image.

On peut voir sur les images suivantes que les différences - bien qu'en apparence minimes - ne sont pas à négliger, et qu'il doit être envisagé de tester chacune des compositions possibles afin de sélectionner la plus performante.



La condition est bien-sûr de pouvoir lancer AutopanoGiga en ligne de commande avec des paramètres pré-chargés, ce sur quoi nous devrions être bientôt fixé.

En cas d'impossibilité, d'autres solutions sont à explorer.

4.2 Le signal monogène

(merci Guillaume !)

Le coeur du problème est la détection de points d'intérêts et le calcul de descripteurs pertinents pour chaque image. Nous savons que l'approche par les méthodes usuelles est inefficace dans notre situation avec des images isolées. On peut s'en sortir parce les images sont composables et qu'un petit nombre d'entre-elles suffit pour avoir l'information nécessaire à la bonne marche des algorithmes usuels. Mais, afin de rendre la méthode plus robuste, afin d'anticiper des situations plus problématiques, afin surtout de proposer une alternative à l'utilisation d'AutopanoGiga si celui-ci n'est pas intégrable à une interface externe, il est important d'examiner de potentielles pistes pour la détection et la description.

Concernant la première, il est possible par exemple de faire spécifier à l'utilisateur une teinte particulière qui contraindrait les régions d'exploration pour la génération de points d'intérêts. On pourrait ainsi forcer l'algorithme à ignorer des zones que l'on sait être trop inconstantes. On utilisera dans ce but filtre fréquentiel couleur, implémenté à partir de la transformée Fourier-Clifford quaternionique.

C'est un premier pré-traitement qui permettra de ne pas avoir à dégrader trop l'image pour la rendre comparable.

Ensuite, il sera non seulement nécessaire d'enrichir l'information que l'on possède, mais il faudra surtout rendre invariant les descripteurs par changement d'échelle. La méthode SIFT et ses dérivées clament cette invariance, mais à l'essai, nous avons vu qu'elles étaient bien insuffisante par rapport à nos besoins ici.

Le signal monogène peut répondre à cette double exigence. Il va effectivement plonger le signal dans un espace d'échelle sur lequel il sera possible de faire agir divers descripteurs, et associer à chaque pixel ou région des informations supplémentaires que l'on peut paramétrer.

Soit $S(x)$ un signal mono-dimensionnel. A chaque point du signal, nous allons associer les valeurs de sa transformée de Hilbert afin d'y ajouter des informations d'amplitude et de phase. On construit donc la fonction :

$$f(x) = S(x) + ih_H * S(x)$$

où $h_H(x) = \frac{1}{\pi x}$.

La fonction peut être réécrite de la manière suivante :

$$f(x) = A(x)e^{i\varphi(x)}$$

où $A(x) = \sqrt{S(x)^2 + (h_H * s(x))^2}$ définit une amplitude et $\varphi(x) = \text{atan}\left(\frac{h_H * S(x)}{S(x)}\right)$ une phase.

La fonction $S(x)$ est analytique, la fonction f est holomorphe, et tout ceci peut se réécrire dans un formalisme cliffordien de la manière suivante :

- $\mathbb{C} \simeq \mathbb{R}_{2,0}$, de base $\{1, e_1, e_2, e_1e_2\}$.
- $f(x) = S(x)e_2 + h_H * S(x)e_1 = A(x)e^{\varphi(x)e_{12}}$

Posons maintenant le système différentiel suivant :

$$(E) := \begin{cases} D^2u = \frac{\delta^2}{\delta x^2}u + \frac{\delta^2}{\delta y^2}u = 0, & y > 0 \\ \frac{\delta}{\delta y}u \cdot e_2 = S(x) \cdot e_2 \end{cases}$$

Ce système est lié à la résolution de l'équation de Dirac dans $\mathbb{R}_{n,0}$: $D = \sum_{i=1}^{n} \frac{\delta}{\delta x_i} \cdot e_i$, une fois convenu que le Dirac est quelque part une racine du Laplacien - pour le produit géométrique - donné par $D^2u = \sum_{i=1}^{n} \frac{\delta^2}{\delta x_i^2}u \cdot e_i$.

La solution de (E) s'écrit sous la forme :

$$u(x, y) = h_p * S(x)e_2 + h_p * h_H * S(x)e_1$$

où $h_p = \frac{y}{\pi(x^2+y^2)}$ est le noyau de Poisson est où y peut s'interpréter comme un paramètre d'échelle tel que :

$$\lim_{y \to 0} u(x, y) = u(x)$$

Cette nouvelle fonction $u(x, y)$ pourra alors s'interpréter comme un plongement du signal initial augmenté dans un espace d'échelle de Poisson.

Passons maintenant à un signal bi-dimensionnel, une image. On se place alors dans $R_{3,0}$ et le système devient :

$$(E) := \begin{cases} D^2u = \frac{\delta^2}{\delta x^2}u + \frac{\delta^2}{\delta y^2}u + \frac{\delta^2}{\delta z^2}u = 0, & z > 0 \\ \frac{\delta}{\delta z}u \cdot e_3 = S(x) \cdot e_3 \end{cases}$$

La solution de ce système s'écrit :

$$f(x,y;z) = h_p * S(x,y)e_3 + h_p * h_{R_x} * S(x,y)e_1 + h_p * h_{R_y} * S(x,y)e_2$$

où h_R est la transformée de Riesz : $h_R = \frac{xe_1+ye_2}{2\pi(x^2+y^2)^{\frac{3}{2}}}$, et où $h_p = \frac{z}{2\pi(x^2+y^2+z^2)^{\frac{3}{2}}}$ est le noyau de Poisson.

Alors f est une fonction monogène telle que f puisse se réécrire $f(x,y;z) = A(x,y)e^{\beta\varphi(x,y)}$ où β est un bivecteur.

En particulier :

$$f \cdot e_3 = \underbrace{h_p * S(x,y)}_{Sc} + \underbrace{h_p * h_{R_x} * S(x,y)e_{13} + h_p * h_{R_y} * S(x,y)e_{23}}_{B}$$

soit :

$$f \cdot e_3 = Sc + \frac{B}{|B|}|B| = A \exp \frac{B}{|B|}\varphi$$

où $\frac{B}{|B|}$ est isormorphe à I_3.

A chaque coordonnée (x,y) de l'image I, on peut désormais faire correspondre le vecteur suivant :

$$f(i,j;s) = \begin{pmatrix} h_p * I \\ h_p * h_{R_x} * I \\ h_p * h_{R_y} * I \end{pmatrix}$$

On enrichit ainsi la variété des descripteurs utilisables : on pourra par exemple utiliser des corrélations de blocs pour des bivecteurs donnés. On a surtout accès à un paramètre d'échelle qui nous permettra de rendre le matching d'images plus robuste que ce qu'il n'est pour l'instant.

Toutefois, la nature des images que nous avons eu à traiter est telle qu'il ne s'agit plus simplement d'une variation d'échelle, mais d'une différence de résolution telle que les objets peuvent être amenés à changer radicalement de nature topologique. La dernière perspective que je veux proposer est pour l'instant de nature purement théorique, mais je pense qu'en se penchant sur ses possibilités d'application et d'implémentation, on pourrait avancer de manière certaine sur notre problématique et de nombreuses autres.

4.3 Les espaces galactiques

Cette dernière section s'inspire d'un article de Guy Wallet sur les espaces métriques généralisés : *Contraction of a Generalized Metric Structure*, 2008.

L'article commence par relever que toutes les contractions ne sont pas nécessairement que des homothéties. Dans certains espaces, une contraction ne laissera pas invariante la topologie : des objets à distance non-nulle pour une échelle de perspective donnée fusionnent pour certains rapprochements un peu soudains.

Un changement d'échelle empirique, est-il écrit, se décompose en deux processus dépendants :

- une homothétie modifiant la taille des objets
- une simplification qui permet de négliger de petits détails

soit typiquement la situation dans laquelle nous nous trouvons dans ce travail.

La première étape consiste à se donner une extension $\mathbb{K}$ de $\mathbb{R}$ qui soit un corps, propre et ordonnée.

Ajoutant par exemple au corps $\mathbb{R}$ des nombres infiniment petits - tels qu'ils sont inférieurs à $\frac{1}{n}$ pour tout $n \in \mathbb{N}$ - et des nombres infiniments grands - tels qu'ils sont supérieurs à tout $n \in \mathbb{N}$.

Posons maintenant $Hal(0)$ l'ensemble des nombres infiniments petits. On a alors la relation de proximité suivante :

$$\forall (s,t) \in \mathbb{K}^2, (s \simeq t \Longleftrightarrow s - t \in Hal(0))$$

On peut alors définir pour tout $x \in \mathbb{K}$ son halo de la manière suivante : $Hal(x) = x + Hal(0)$.

Posons maintenant l'ensemble des éléments finis de $\mathbb{K}$: $Gal(0)$. Alors l'ensemble quotient $\frac{Gal(0)}{Hal(0)}$ est isomorphique à $\mathbb{R}$.

On appellera enfin *galaxie de t* la classe d'équivalence $Gal(t) = t + Gal(0)$. On pourra se réferer à l'article en question pour voir démontré que l'espace $Gal(\mathbb{K})$ est un groupe additif ordonné.

Soit G un groupe additif ordonné. Posons $G_+ := \{x \in G; x \geq 0\}$. On munit G d'une distance généralisée, i.e. une distance pouvant prendre une valeur infinie. On peut alors définir des boules de centre $x \in G$ et de rayon $r \in \mathbb{R}_+ \cup \{+\infty\}$.

Chaque boule est elle-même un espace métrique et est appelée une composante métrique de G pour la distance d. On pose alors l'ensemble M_G des composantes métriques de G pour d, et on définit pour tout $x \in G$ la classe $C_G(x)$ des composantes de G telles que x soit un un élément de ces composantes.

On peut alors écrire :

$$\forall (x,y) \in G^2, d(x,y) = \begin{cases} d_i(x,y) \text{ si } y \in C_G(x) \\ +\infty \text{ sinon} \end{cases}$$

où $d_i(x,y)$ est la distance associée à $E_i = C_G(x)$.

En posant $G = Gal(\mathbb{K})$, on peut définir une distance galactique de la manière suivante :

$\Delta : \varepsilon \times \varepsilon \to Gal(\mathbb{K})$

$\Delta(x,x) = 0$

$\Delta(x,y) = \Delta(y,x)$

$\Delta(x,z) \leq \Delta(x,y) + \Delta(y,z)$

Il est maintenant possible de définir un espace galactique comme étant une structure (F, δ, Δ) où F est un ensemble, δ une distance généralisée sur F, et Δ une distance galactique sur l'ensemble M_F des composantes métriques de F pour δ.

L'article indique alors que la structure peut se comprendre comme un ensemble sur lequel on se donne deux niveaux de résolution : une résolution entre chaque point d'une même composante métrique - la distance généralisée - et une résolution portant sur la distance entre composantes métriques - la distance galactique proprement dite.

On appellera alors isométrie toute application qui préserve chacun des

niveaux de résolution. Ainsi, munis de deux espaces galactiques (F, δ, Δ) et (F', δ', Δ'), l'application ϕ sera une isométrie de F sur F' si et seulement si :

1. $\forall (x, y) \in F^2, \delta'(\phi(x), \phi(y)) = \delta(x, y)$
2. $\forall (G, H) \in M_F^2, \Delta'(\phi(G), \phi(H)) = \Delta(G, H)$

On appellera enfin contraction de rapport γ l'application f qui satisfait aux propriétés suivantes :

$$\delta(f(x_1), f(x_2)) = \begin{cases} pv(\gamma\delta(x_1, x_2)) & \text{si} \quad \delta(x_1, x_2) < \infty \\ pv(\gamma.\Delta(C_F(x_1), C_F(x_2))) & \text{si} \quad \delta(x_1, x_2) = \infty \end{cases}$$

$$\Delta(f[E_1, E_2]) = \gamma \cdot \Delta(E_1, E_2)$$

En posant pour une image I le rapport $I \simeq \frac{Gal(0)}{Hal(0)}$, on pourra distinguer dès lors entre de simples variations d'échelles sans boulersement topologiques, et des phénomènes plus brusques qui auront pour conséquence que :

$$\Delta(f[E_1, E_2]) \in Hal(0)$$

soit quelque-chose à comprendre comme un devenir-indiscernable des séparations, exigeant une redéfinition de la topologie de l'image, par modification de la base d'ouverts qui la définit.